# INSIGHTS INTO THE EFFECTS OF COVID-19 ON TAXI USE NEAR M2 METRO LINE IN ISTANBUL

by

Ece Özcan

B.S., Civil Engineering, Boğaziçi University, 2019

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Civil Engineering
Boğaziçi University

2020

# ACKNOWLEDGEMENTS

First, my profound thanks and respect to my supervisor, Assoc. Prof. Ilgın Gökaşar, who has helped me throughout not only my thesis but also my master's degree with her valuable ideas and guidance. Her energy and persistent effort to change everything for the better kept me on the rails.

I would also like to express my sincere appreciation to my thesis committee Assist. Prof. Gürkan Günay and Assist. Prof. Mehtap Işık for their comments as a part of my defense jury.

I would like to thank the Boğaziçi University Intelligent Transportation Laboratory team, without whom I would not have been able to complete this thesis, and without whom I would not have made it through my master's degree. Especially, to Alperen Timuroğulları, for his professional and personal support and encouragements; and Ali Atilla Arısoy, for his positive and constructive attitude.

The Bizimdurak application provides the data used in this thesis. I am most grateful for their support and efforts that made this thesis possible.

I would like to express my sincere gratitude and love to my parents for their endless support that allows me to step forward, even when I felt powerless.

Finally, I would like to express my special thanks to Niyazi Doğan Öncü, who supported me in every stage of my life since the day we met.

# ABSTRACT

## INSIGHTS INTO THE EFFECTS OF COVID-19 ON TAXI USE NEAR M2 METRO LINE IN ISTANBUL

COVID-19 has become one of the most significant events in this century and the effects are felt in daily life all around the world economically and socially. Therefore, investigating the effects of COVID-19 must be prioritized to minimize future damages. The goal of this thesis is to reveal different travel reactions to COVID-19 based on spatial and socioeconomic characteristics such as public transport connectivity, education level, female percentage, etc. Towards this goal, taxi GPS data is used and the M2 subway line in Istanbul is selected as the case study area since the line covers many residential and commercial centers. The prepared COVID-19 timeline is divided into five phases based on critical events such as the announcement of governments or unexpected peaks in daily case numbers. The analyses are conducted for the average trip counts in four time periods of a day, namely total, off-peak, morning, and evening. K-means clustering is used to observe the relationship between stations and the data is analyzed by ordinary least squares (OLS), spatial auto regression (SAR), and geographically weighted regression (GWR) models based on daily average trip counts and characteristics of stations. The best results are obtained by the GWR model. According to the results, the population size is one of the most significant parameters, that explains the change in trip counts. The morning peak shows a unique characteristic that can be explained by the socioeconomic index, which is a weighted average of many parameters including education and income level. In general, the decrease in taxi trips is higher for the areas with a higher socioeconomic index. Other significant variables are the number of shopping malls, the existence of another public transportation option, and the population density.

# ÖZET

## İSTANBUL M2 METRO HATTI YAKININDAKİ TAKSİ KULLANIMINA COVID-19 ETKİLERİ

COVID-19 içinde bulunduğumuz yüzyılın en önemli olaylarında biri haline gelmiş ve bütün dünyada günlük hayatı ekonomik ve sosyal olarak etkilemiştir. Bundan dolayı, gelecekteki zararları en aza indirmek için COVID-19'un etkilerinin araştırılması öncelik haline getirilmelidir. Bu tezin amacı, toplu taşıma bağlantısı, eğitim seviyesi, kadın oranı gibi mekansal ve sosyoekonomik özelliklere dayanarak COVID-19'a bağlı gelişen farklı seyahat davranışlarını ortaya çıkarmaktır. Bu amaçla, taksi GPS verileri kullanılmış ve pek çok yerleşim yeri ile ticaret merkezini kapsayan M2 metro hattı çalışma için seçilmiştir. Hazırlanan COVID-19 zaman çizelgesi, hükümetin yaptığı duyurular ve günlük vaka sayısındaki beklenmedik artışlar göz önünde bulundurularak beş aşamaya bölünmüştür. Analizler, ortalama seyahat sayıları için yapılmış ve gün, toplam, yoğun olmayan saatler, sabah ve akşam zirve saatleri olarak dörde bölünmüştür. İstasyonlar arasındaki ilişkiyi gözlemlemek amacıyla K-means kümeleme yöntemi kullanılmıştır ve data en küçük kareler (OLS), mekansal otoregresyon (SAR) ve coğrafi ağırlıklı regresyon (GWR) modelleri ile analiz edilmiştir. Sonuçlara göre, nüfus büyüklüğü, yolculuk sayılarındaki değişimi açıklayan en önemli parametrelerden bir tanesidir. Analizde, sabah zirve saatleri diğer zamanlardan farklılık göstermektedir ve eğitim ve gelir düzeyi dahil birçok parametrenin ağırlıklı ortalaması olan sosyoekonomik endeks ile açıklanabilmektedir. Genel olarak taksi seferlerinin azalması sosyoekonomik endeksi yüksek olan bölgelerde daha yüksektir. Analizde, Taksi seyahatlerindeki değişimi açıklayan diğer önemli parametreler ise alışveriş merkezi sayısı, başka bir toplu taşıma seçeneğinin varlığı ve nüfus yoğunluğudur.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF SYMBOLS

| | |
|---|---|
| $C_{earth}$ | Circumference of the Earth |
| $d_ik$ | The Distance Term |
| $h$ | The Bandwidth |
| $I$ | Moran's I value |
| $k$ | The Number of Regression Parameters |
| $n$ | Number of Equations |
| $r$ | Radius of Primary Circles |
| $R$ | Radius of Earth |
| $R^2$ | Coefficient of Determination |
| $r_e$ | Equatorial Radius of Earth |
| $X$ | Independent Variable |
| $Y$ | Dependent Variable |
| $W$ | Spatial Lag Term |
| $W_1$ | Diagonal Weight Matrix |
| | |
| $\alpha_ik$ | The Weight Term in Kernel Function |
| $\beta$ | Regression Coefficient Terms |
| $\varepsilon$ | Error Term |
| $\rho$ | Rho, Coefficient of Spatial Lag Term |

# LIST OF ACRONYMS/ABBREVIATIONS

| | |
|---|---|
| AIC | Akaike Information Criterion |
| APTA | American Public Transportation Association |
| BRT | Bus Rapid Transit |
| COVID-19 | Coronavirus disease 2019 |
| GPS | Global Positioning System |
| GWR | Geographically Weighted Regression |
| ID | Identification Number |
| LAT | Latitude |
| LON | Longitude |
| MER | Mean Error Sum of Squares |
| MERS | Middle East Respiratory Syndrome |
| MSR | Mean Regression Sum of Squares |
| OLS | Ordinary Least Square |
| PCA | Primary Catchment Area |
| RSS | Residual Sum of Square |
| SAR | Spatial Auto Regression |
| SARS | Severe Acute Respiratory Syndrome |
| SEGE | Socio-economic Index |
| SES | Socio-economic Status |
| SSE | Sum of Squared Error |
| SSR | Regression Sum of Squares |
| TURKSTAT | Turkish Statistical Institute |
| WHO | World Health Organization |

# 1. INTRODUCTION AND MOTIVATION

Throughout history, infectious diseases affected humankind. Among those diseases, a pandemic is a worst-case scenario that spreads beyond national boundaries. The plague in the 14th century, known as the black death, is one of the most known examples that killed one-third of the world population and spread from Asia to Europe via trade routes. Another extensive pandemic took place in 1918, namely Spanish flu. It resulted in 50 million total deaths in Europe, the US, Asia, and America. Even though the infectiousness is more controllable, AIDS is also one pandemic that killed 35 million people, and the cure is yet to be found [1]. SARS (Severe Acute Respiratory Syndrome) is a milestone happened in 2003 that waken the professionals to improve outbreak measures [2] which are also utilized to control later diseases like H1N1 (2009) [3], MERS (2012) [4] and Ebola (2014-16) [5]. Even though the death rate of infectious diseases has decreased over the years, the number of new ones is increasing enormously compared to previous centuries [6].

Each of those pandemics changes the economic situation and demographics of the society unpredictably and suddenly. The increasing internationalization of many sectors, including tourism, trade, production, etc. made the boarders of countries more and more invisible. It eased the spread of diseases to all continents in various ways. Now, history repeats itself with COVID-19, which will be perhaps the most remarkable disease, if not the deadliest. The first cases of the virus appeared in December 2019 in Wuhan, China, and spread 114 countries in three months [7] following the announcement of WHO (World Health Organization) on March 11, 2020, that the COVID-19 is a pandemic [8]. As an extensive crisis, the COVID-19 is neither the first pandemic nor will be the last. Therefore, investigating the effects of COVID-19 on today's cities and lifestyles must be prioritized to decrease the damages created by pandemics. In this thesis, taxi GPS data is investigated to reveal the relationship between taxi usage behavior based on various socioeconomic factors during the COVID-19. The data is spatially restricted with the M2 subway line that carries most of the passengers in the subway network in Istanbul.

## 1.1. Problem Statement

Pandemics have the potential to change human travel behavior and mode choice that might disturb the balance of a city's mobility. The studies about transportation during the COVID-19 outbreak showed that mobility habits might change permanently, especially for urban areas. The increase in on-demand services and individual mobility can cause extra traffic load on the network, but also can create alternative possibilities for practitioners. Besides, the reaction of society to extreme events like pandemics changes significantly because of socioeconomic differences among territories. Therefore, it is essential to understand the condition and intervene appropriately to decrease possible negative effects.

## 1.2. Goals and Objectives

The goal of this thesis is to reveal the change in the travel behavior of taxi users based on the spatial characteristics and to have an insight about different reactions to COVID-19 depending on the socioeconomic factors. The following objectives are completed to achieve this goal:

(i) To define COVID-19 phases based on specific events,
(ii) To cluster the areas according to their reactions,
(iii) To test the change in taxi trip numbers based on socio-economic

## 1.3. Organization of the Thesis

In Chapter 2, a literature review is provided related to COVID-19 and on-demand services by including taxi studies available. Following that, the theoretical background used in the thesis is explained in Chapter 3. Afterward, in Chapter 4, the case study area and data collection process are explained. In Chapter 5, the methodology of the thesis and analysis results are presented. Finally, in the 6th Chapter, the conclusion, comments, and recommendations are given.

# 2. LITERATURE REVIEW

As from the previous outbreaks, social isolation and quarantines are used as effective measures to slow down the spread of diseases. However, they also slow down economies by damaging many sectors. Here, the COVID-19 is the most extensive infectious disease in the last century that affected the world economy radically [9]. According to UNCTAD, the cost of the COVID-19 outbreak exceeds $1 trillion, and the disease still cannot be stopped. Some economies came to the edge of the fall [10]. In the time of COVID-19, many countries applied a nation-wide isolation procedure that includes shutting down all crowded places where transmission is possible. Most of the countries declared lockdown in crowded cities. The lockdowns target especially the mobility since it is one of the main contamination areas used by almost every citizen on a daily-basis [11]. Mostly, public transportation is strictly restricted in many countries [12], and the allowable number of passengers are reduced by less than half of the capacity of vehicles [13].

A study related to the 2009 H1N1 pandemic shows that the travel restrictions can only create a two-week delay in the spread even in the best-case scenario with the early intervention [14]. Therefore, to provide traveling conditions with sufficient social distance, there are suggestions such as staggered working hours to reduce the crowds during peak hours [15]. Still, there are ongoing discussions about the subways as the reason for most of the COVID-19 cases in New York City [16, 17]. A study criticizes this claim and shows that the infection rates increase with automobile mode share statistically, which can be explained as people preferred automobiles more after the outbreak [18].

The impacts and losses of such diseases can be even permanent for public transport [19] since it is a place where people are most likely to get infected [20]. According to a study, related to the SARS outbreak in 2002, the infectious diseases reduces the travel demand even though there isn't a lockdown going on [19]. Because of the high contamination risk and reduced travel options, the travel behavior and mode selection

of people change. For example, bicycle usage increased substantially in many cities as a travel mode during COVID-19 [21, 22]. As is known, walking and bicycling are the safest and most affordable travel modes [23], yet not entirely possible in many cities. All of which requires additional measures and policies for authorities to sustain equal and accessible mobility.

Travel restrictions are one of the first measures to slow down the spreading of infectious diseases. Yet, most of the time, efficiency cannot be achieved without implementing proper epidemic models [24]. Therefore, many spatial spread models for different scenarios that investigate the transmission of the diseases via travel behavior from a medical perspective are investigated in the literature [25, 26, 27, 28]. Mostly, the mobility pattern of the population is considered creating a framework for the disease spread [29], and human mobility patterns and spatial characteristics are taken into account to predict the spread of diseases. Most of the travel behavior study focuses on tourism-related travel activities because of economic concerns [30, 31]. One such study conducted in Bulgaria with 974 respondents found that most of the respondents are ready to travel within two months after travel restriction is over [32]. Especially, air travel-related researches are popular since they are necessary for today's economies, yet they spread diseases across continents quickly [33, 34, 35, 36]. Despite the considerable number of studies that cover mobility as a tool for disease spreading, the number of studies that investigates the effects of disease on mobility and travel behavior within the cities is quite limited. However, transportation must be carefully considered since it affects many areas, including the economy, tourism, and daily life of city-dwellers.

Using mobile phones to track population movements via GPS (Global Positioning System) is an effective way to decrease the infection rate and control the diseases. It is used in 2010 during the Haiti earthquake and cholera outbreak and proved to be effective [37]. It is also used effectively during 2005 cholera in Senegal [38]. A study used mobile phone calls to detect the changes in travel during the Ebola lockdown in Sierra Leone. The data revealed a dramatic decrease in human mobility for three days. There were 31%, 46%, and 76% decrease in individuals' relocations for distances below 15 km, 15-30 km, and above 30 km, respectively. So, for long-distance trips, there was

a significant decrease in travel. The results were space-sensitive, meaning that spaces where case numbers are higher, the impact was also high. Yet, the travel behavior turned to the previous state right after the restrictions were lifted [39].

A study conducted in France during COVID-19 with mobile phone data examines the travel behavior based on trip distance [40]. The residents are distinguished by age, daily, and weekly travel time. It is found that overall trips, both short and long duration, dropped 65%. The results of the study were like the Sierra Leone Ebola epidemic research considering the reductions in mobility. Another study conducted in Switzerland during COVID-19 uses GPS tracking via a mobile application to define the changes in travel behavior [41]. It is found that the number of trips during workdays decreased to 60% of the previous period. Besides, the kilometers traveled also decreased significantly. Even though trip durations decreased at first, then it almost reached "the old normal" again. Besides, the share of individual motorized modes (car, motorbike, taxi, Uber) increased compared to the reference period [41]. An interesting finding is that the decrease in the change in average kilometers traveled increases proportionally with income level. The most distinctive difference occurred for bicycle usage. It increased substantially, particularly for leisure activities at the weekends.

The effects of COVID-19 on travel behavior are investigated in New York by comparing two reopening scenarios (with or without transit capacity restrictions) [42]. Both of the scenarios suggest a lower transit ridership (64% and 73%, respectively) and an increase in car trips (142-143% of pre-pandemic levels). Besides, there was also an increase in taxi usage and bike and FHV (For-Hire-Vehicles). A white paper is published related to the study results that investigates the impact of COVID-19 on New York City (NYC) transportation system and potential policies for the system by using an agent-based simulation model built in the MATSim environment [43]. A pre-COVID model for standard travel behavior and a COVID model for travel behavior during the first month of the pandemic are developed as baseline models. In the calibrated pre-COVID period, the mode share of the taxi is 3.2%, and the transit mode share is 35%. In the first scenario, there is no transit capacity restriction, and in the second scenario, there is a 50% restriction in all transit capacity. Based on the four-phase reopening,

which is done under the consideration of the regional guidelines, in the fourth phase, the taxi mode share becomes 2.0% when there is no transit capacity restriction and 2.9% when there is 50% transit capacity reduction. When transit mode share is considered, the model estimates the transit mode share to be 26% for no transit capacity restriction scenario and 23% for 50% transit capacity restriction scenario.

In a recent study conducted in 2018, the travel mode choice behavior between taxi and subway is investigated in the city of Beijing [44]. The priority is given to the influence of traveling convenience. The origin-destination (OD) points of taxi trips in Beijing are examined, and the locations of these points have been compared with the respective nearest subway station. Based on the OD couples, 24.89% of all considered trips have no convenient subway connections, and thus taxi becomes the only option. When, again, all trips are considered, 60.26% of the trips the access distance from either the origin or destination point to the closest subway station is over 500 meters. This finding shows that walking distance has a significant role in travel mode choice. According to a survey conducted by the Institute for Transportation and Development Policy in March 2020, in China, 40% of public transport users shifted to individual modes, namely private cars, taxis, ride-hailing, walking, and biking. The survey found that the bus system had recovered only 50% - 60% of its previous ridership, and only 34% of previous metro and bus commuters continue using public transport. Besides, the usage of bike-sharing increased by 150% after reopening [45]. The same ridership statistics also exist in major cities in the U.S. and Europe that show a 50-90% reduction from pre-crisis levels [46]. Considering the studies, by considering the effect of COVID-19, even recent researches can undergo severe changes and lose their effectiveness radically.

Many European countries are analyzing to see the changes in the mobility of their residents. In Belgium, an average citizen reduced mobility by over 50%, and 80% of the population stays in the border of their neighborhood [47]. The reduction in mobility is considerably higher for cities compared to rural areas in Germany [48]. On average, there is a 38% decrease in mobility, and this reaches up to 54% for cities in March. Interestingly, the long-distance travel between districts reduced up to 70% below the

previous situation. Yet, the following report also shows that mobility turned back its previous state by increasing slowly. In April, the reduction in mobility compared to the baseline is detected as 27% lower [49]. Besides, there is a recovery in weekend trips meaning that people are turning back to normal for optional trips (for leisure, travel, etc.). Similar results are obtained for France [50]. A study used Facebook mobile users' data to observe the changes in mobility patterns [51]. The data represents 28 million users that are approximately 3% of the total age population of India. The results show that individuals' mobility decreased by almost 90% in Delhi and 40% in other states. Besides, some unusual long-distance relocations from cities to remote villages are observed in the research.

As a result, COVID-19 affects urban areas more, as suggested in different studies [48] [51]. Besides, the spatial differentiation of the effects, previous researches also show that the mode choice characteristics depend on factors like education level, income level, or travel period. A study analyzes the choice behavior for different passengers based on their travel distances, in the city of Nanjing, the capital of Jiangsu Province in China [52]. Three categories are defined: short distance (less than 3 km), medium distance (3 to 9 km) and long-distance (over 9 km) passengers. Binary logit models are developed for each category, and the factors affecting the passengers' choice are investigated. The models are constructed with only statistically significant factors affecting the decision of short-distance passengers which are education level (high school and university), travel purpose (visiting friends or family), preference of online car hailing service versus traditional taxi service , comfort level, travel period, and whether the travel time is ample. For medium distance passengers, age (18-25 and 26-35 years old passenger groups), preference, comfort level, and availability of companions of the passengers, are the factors significant in the decision making. For the long-distance passengers, monthly income (7000-9000 yuan), preference, safety level, service attitude, waiting time, travel period and availability of companions of the passengers, are the significant factors that influence the decision making the process of these passengers. Long-distance passengers, with a monthly income of 7000 to 9000 yuan, are more likely to choose online car-hauling services (OCS). With the increase of age, the number of companions, comfort level, and other significant factors of the medium distance

passenger group, the passengers become more likely to choose OCS. For short distance passengers, they become less likely to use OCS as their education level increases and more likely when the travel time is ample.

Another study also investigates the factors affecting the decision of travelers to use on-demand ride services such as Uber and Lyft, in California [53]. The millennials and Generation $X$ are the passengers that are considered in this study. The results show that young passengers, passengers with high levels of education, and non-Hispanic origin passengers are more likely to adopt on-demand services. Individuals that make long-distance business trips by plane are more likely to adopt on-demand ride services. Individuals that are using their private vehicles less than they used to and those who are planning to replace or part with their vehicles are also more likely to adopt on-demand ride services.

In a report that investigates California ridership trends in the 2010s, there is an eightfold increase in ride-hail (Lift, Uber, etc.) and taxi use [54]. It is found that ride-hail use is the highest in the evening and weekends, not in peak times. The current studies also showed that there is and will be an increase in individual transport modes because of COVID-19. Despite the indisputable superiority of private vehicles during COVID-19, the worldwide trends in the auto industry show that there is a non-negligible decrease in international car sales for the economic recession [55]. Therefore, it wouldn't be wrong to assume that people might prefer to use on-demand services such as Uber, Lift, and taxi.

By considering the lack of ride-hailing services in Istanbul, it would be reasonable to assume that taxis will gain popularity. Taxi drivers are at a high risk of infection after health care workers [56, 57, 58, 59], which created hesitation for the public during the first stages of the outbreak. However, after the mandatory measures stated by the government [60], the situation might change in favor of on-demand services. All in all, COVID-19, as one of the most extensive pandemics in the 21st century, can change the rules radically. Therefore, investigating its effects is essential not only to take lessons for tomorrow but also to manage the present.

# 3. THEORY

In this chapter of the thesis, the theoretical background for the used machine learning methods and regression models are presented. K-means clustering algorithm is preferred as a machine learning method because of its effectiveness. To analyze the data and predict the estimates of the models, different models are used and compared, namely ordinary least square (OLS), spatial auto regression (SAR), and geographically weighted regression (GWR).

## 3.1. K-means Clustering Algorithm

Machine learning methods are split into three categories, namely unsupervised, supervised, and reinforcement learning. Unsupervised learning categorizes data without a training data set and with minimum intervention. So, the only categorization relies on the data similarities without a predefined input or output. Clustering gives an intuition about the structure of the data as one of the simplest and widely used unsupervised machine learning algorithms. It groups data points according to their similarity by creating subgroups from the data and tries to homogenize subgroups such that they become as similar as possible. Clustering makes inferences from datasets by utilizing input vectors.

In this thesis, K-means clustering is used to create meaningful clusters and to observe the similarities between different locations. K-means clustering algorithm tries to partition the dataset into k non-overlapping subgroups iteratively. All data point belongs to one subgroup (cluster), meaning that subgroups are discrete. The primary aim is to create clusters that are as far as possible in terms of Euclidean distance. Euclidean distance is the distance between two points in Euclidean space that is usually two or three dimensions yet, can be generalized to higher dimensions. It is the absolute

value of the difference between coordinates and shown as;

$$d = |x - y| = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2} \tag{3.1}$$

where $x$ and $y$ are two points, and $n$ is the number of dimensions.

The number of K refers to the number of centroids, or the number of clusters, and K-means clustering creates K number of centroids and allocates the data points to the nearest cluster centroid. The centroid of a cluster is found by calculating the arithmetic mean of all data points in that cluster. The K-means clustering works in the following order [61]:

(i) A value for K is chosen as the total number of clusters.
(ii) K points as cluster centroids are randomly chosen.
(iii) The data points are assigned to the closest centroid.
(iv) A new cluster center is calculated for each subgroup.
(v) The iteration continues until no further improvement to minimize the sum of the squared distance between the data points, and the cluster's centroid is possible.

The most preferred measure to test K-means clustering is the Sum of Squared Error (SSE) that is the sum of the distance to the nearest centroid for all points and formulated as:

$$SSE = \sum_{i=1}^{K} \sum_{x \in C_i}^{d} ist^2 (m_i, x) \tag{3.2}$$

where $x$ is a data point in cluster $C_i$, and mi is the representative point for cluster $C_i$.

### 3.2. Ordinary Least Squares (OLS) Model

Regression analysis is used in this thesis to discover the relationships between the taxi trip change throughout the COVID-19 period and socioeconomic characteristics.

Regression analysis is used to represent the best relationship between a dependent variable and independent variables. As the simplest version, simple linear regression comprises a single independent variable and given with the following formula:

$$Y = \beta_0 + \beta_1 \tag{3.3}$$

where $\beta_0$ is the constant of the intercept and shows the point where the regression line crossed the y-axis, $\beta_1$ is the slope of the line that explains the relationship between $x$ and $Y$ and called the coefficient.

Yet, the error term is a natural result of reality, since any data cannot be correctly explained without a margin to deviate. So, for every individual observation i, the equation becomes:

$$y_i = \beta_0 + \beta_\mathrm{i} x_i + \varepsilon_\mathrm{i} \tag{3.4}$$

where $y_i$ is value of the dependent variable I, and $x_i$ is value of the independent variable $i$.

The equation for the best fit line or the predicted regression line is:

$$\hat{\mathrm{y}} = b_0 + b_i x \tag{3.5}$$

where $\hat{\mathrm{y}}$ is the predicted value of the model. Therefore, $b_0$ and $b_1$ are used as estimators. As a result, error terms ($\varepsilon_i$) are the difference between the observation and estimation, and formulated as below:

$$\varepsilon_\mathrm{i} = \mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i} \tag{3.6}$$

However, most of the regression models have more than one independent variable. So, OLS, which is known as the general linear model, is used to represent a dependent

variable as a function of independent variables, and the aim is to find a generalized coefficient matrix that can explain most of the data points. The general equation is as follows [62]:

$$y_i = \beta_0 + \beta_1 \mathrm{x}_{1\mathrm{i}} + \ldots + \beta_{\mathrm{k}} \mathrm{x}_{\mathrm{k\cdot i}} + \varepsilon_{\mathrm{i}} \tag{3.7}$$

where i (1,2,...n) is the number of variables, $\beta_i$'s are coefficients of independent variables, $\beta_0$ is the intercept, $y_i$ is the $i^{th}$ value of the dependent variable, $x_i$'s are the $i^{th}$ value of the $k$ independent variables and $\varepsilon_i$'s are the individual error terms.

Here, the equation can be generalized as:

$$\mathrm{y} = \mathrm{x}\beta + \varepsilon \tag{3.8}$$

where

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} x = \begin{bmatrix} 1\chi_{21} \cdots \chi_{k1} \\ 1\chi_{22} \cdots \chi_{k2} \\ \vdots \ \vdots \ \ \ldots \ \ \vdots \\ 1\chi_{2n} \cdots \chi_{kn} \end{bmatrix} \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{3.9}$$

To find if there is a statistically significant relationship between the dependent variable ($y_i$) and independent variables ($x_i$), the condition where $\beta_i = 0$ must be tested. In other words, if all coefficients are zero, then there is no relationship, and the following hypothesis is used:

$$H_0 = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.10}$$

For that purpose, F-test is used with the following formula:

$$F = \frac{SSR/k}{SSE/\left(n-k-1\right)} = \frac{MSR}{MER} = \frac{R^2/\left(k\right)}{\left(1-R^2\right)/\left(n-k-1\right)} \tag{3.11}$$

where

- SSR (Regression sum of squares) $= \sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}\right)^2$
- SSE (Error sum of squares) $= \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2$
- MSR (Mean regression sum of squares) $= SSR/k$
- MER (Mean error sum of squares) $= SSE/\left(n-k-1\right)$
- n= number of equations
- k= number of independent variables
- $R^2$ = coefficient of determination.

The value obtained from the formula is compared with the table value depending on the chosen confidence level. If the obtained value is higher than the table value, it can be concluded that there is a relationship between $y$ and $x$, meaning the rejection of the hypothesis. Here, the coefficient of determination ($R^2$) is an important parameter because it explains the adequacy of the model and is valued between 0 as no fit and 1 as the perfect fit.

The F-test reveals the existence of a relationship, yet it doesn't explain which one of the independent variables ($x_i$) affects the dependent variable ($y_i$). The t-statistics are used to compare model parameters bi and $\beta_i$ with the following formula:

$$t_{b_i} = \frac{\beta_i - b_i}{\sqrt{\varepsilon'\varepsilon/\left(n-k-1\right)c_{ii}}} \tag{3.12}$$

where $c_{ii}$ is the value from the $i^{th}$ row and $i^{th}$ column of the matrix $(x'x)^{-1}$

In just the same way as F-test, if the calculated t-value is higher than the table value, the hypothesis is rejected, and the coefficient isn't equal to zero. The procedure

is repeated for each coefficient separately.

There are four assumptions of the OLS model that are given below [63]:

(i) E($\varepsilon$)=0 ;
(ii) E($\varepsilon\varepsilon'$)=$\sigma^2$ $I_n$ where $I_n$ is an identity matrix;
(iii) $x$ is a set of fixed numbers;
(iv) $x$ has rank k < n.

The first assumption shows that negative and positive errors are equally likely. The second assumption, also called homoscedasticity, means that the error terms are uncorrelated with independent variables. The final assumption ensures that the number of variables is less than the number of observations. OLS is preferred because of its ease of use, yet it is not very accurate in all cases.

### 3.3. Spatial Auto Regressive (SAR) Model

If a model doesn't have a spatial lag, meaning that it is not localized, the OLS method is enough to explain the relationship. However, the spatial dependence of the model requires a spatial leg term. In this regards, Tobler's law of geography says that [64]:

"Everything is related to everything, but near things are more related than distant things".

This concept refers that nearby variables are more related than the ones far away. In SAR models, the outcomes of linear regression (explained in 3.2) in an area can be affected by:

(i) Outcomes in nearby areas
(ii) Covariates of nearby areas
(iii) Errors of nearby areas

The general equation for Spatial Auto Regression (SAR) model is presented as [65]:

$$y_i = \rho \mathrm{W} y_j + \mathrm{X}\beta + \varepsilon \tag{3.13}$$

where $y_i$ is the dependent variable, $\rho$ is the SAR coefficient, W is spatial weight matrix, $y_j$ is the neighbor point, $X$ is the independent variable, $\beta$ is the coefficient and $\varepsilon$ is the error term.

The main difference of SAR is that it is a more generalized regression that includes spatial effects by using $\rho$Wy as the spatial autocorrelation term and W as the spatial weight matrix. The model includes the effects of neighbor points, and the researcher selects the number of points that will be added to the model. Therefore, it is not systematic in that way, and outputs are not general.

The most popular measure for spatial autocorrelation is the Moran's I value, which shows the similarity between observations, and given as [66]:

$$\mathrm{I} = \frac{\sum_{\mathrm{i}}^{\mathrm{n}} \sum_{\mathrm{j}}^{\mathrm{n}} W_{\mathrm{ij}} \left(\mathrm{x_i} - \bar{\mathrm{x}}\right) \left(\mathrm{x_j} - \bar{\mathrm{x}}\right)}{\left(\mathrm{S}^2 \sum_{\mathrm{i}}^{\mathrm{n}} \sum_{\mathrm{j}}^{\mathrm{n}} \mathrm{W_{ij}}\right)} \tag{3.14}$$

where

- $\mathrm{S}^2 = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}}^{\mathrm{n}} \left(\mathrm{x_i} - \bar{\mathrm{x}}\right)^2$
- $x_{ij}$= Observed value at loction $i$,
- $x$ = Average of the $x_i$ over the n location
- $W_{ij}$= Spatial weight matrix.

If there is a similarity between the observed values within a certain distance, the Moran's I value is positive and vice versa if negative. If the Moran's I value is approximately zero, then the values are independent over space.

### 3.4. Geographically Weighted Regression (GWR) Model

The analysis of spatial dependencies requires spatial terms, as mentioned in the SAR model. Both of the techniques improve the general regression model to measure location-based effects. GWR is a local regression model to reveal spatially varying relationships. It is widely used to analyze local relationships within the climate, demographic, or environmental characteristics. The technique is useful to explore that allows spatial autocorrelation of variables. GWR creates separate OLS equations for each location in the dataset that includes dependent and explanatory variables. GWR is highly used in infectious disease researches, yet it might be a slow process because of the high number of regression models.

The first expansion model is presented as a localized regression [67]:

$$y_i = \Sigma_{\mathrm{j}} \mathrm{x}_{\mathrm{ij}} \beta_{\mathrm{j}} \left(\mathrm{p}_{\mathrm{i}}\right) + \varepsilon_{\mathrm{i}} \tag{3.15}$$

where $y_i$ is the dependent variable, $\beta_j$ is the coefficient of the $j^{th}$ independent variable, $\varepsilon_i$ is the error term and $p_i$ is the geographical location of the ith case.

The main problem of Casetti's model is to provide the boarders of neighbor data that influence the regression. Therefore, a circle with radius r is drawn, yet the value of r remains questionable. Besides, the effect of weighting suddenly falls to zero outside of the circle. The first GWR model is introduced, and it uses Casetti's expansion model [67]. Kernel-Weighted regression is addressed to solve the continuity problem in Casetti's model, and the weights are given as [68]:

$$\alpha_{ik} = \begin{cases} 1, \ d_{ik} < r \\ 0, \ otherwise \end{cases} \tag{3.16}$$

where $\alpha_{ik}$ is the weight and $d_{ik}$ is the distance from observation point $i$ to an observation point $k$.

Equation (3.13) is transferred into a continues form to allow a better application of the Kernel-Weighted regression as:

$$\alpha_{ik} = \begin{cases} \left\{1 - \left(-\frac{d_{ik}^2}{h^2}\right)\right\}^2, \ d_{ik} < r \\ 0, \ otherwise \end{cases} \tag{3.17}$$

where $h$ equals the control range for the influence zone or bandwidth.

The continues form presented in Equation (3.15) is called Kernel function and denoted by $K(d_{ik})$, where $\alpha_{ik} = K(d_{ik})$. So, Kernel functions provide a gradual transition between zones meaning that the weight decreases with the distance. The likelihood estimate of $\beta$ Casetti's equation gives:

$$\beta = \left(\mathrm{X^T X}\right)^{-1} \mathrm{X^T} y \tag{3.18}$$

The calibrated version with the selected weighting function, the $\beta$ becomes:

$$\beta_{\mathrm{i}} = \left(\mathrm{X^T W_i X}\right)^{-1} \mathrm{X^T W_i} y \tag{3.19}$$

where

$$W_i = \begin{bmatrix} a_{i1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & a_{iN} \end{bmatrix} \tag{3.20}$$

The diagonal components of $W_i$ are the weights of the regression around $p_i$ and $N$ represents the total number of observations.

The bandwidth is the single parameter that is used to include the geographic effects, and the bandwidth of each location can be chosen manually or determined by the software and chosen as 'default' or 'adaptive'. GWR model assumes that all the

independent variables have a constant scale, meaning that all have the same bandwidth. Yet, this assumption creates a limitation by not allowing different and more optimal scales for different variables. The optimal bandwidth can be found by minimizing the Akaike information criterion (AIC) that accounts for the trade-off between complexity and accuracy of the model [69]. AIC is correlated with the model Residual Sum of Squares (RSS) as expressed in the following equation [70]:

$$\mathrm{AIC} = 2k + \mathrm{n}\log\left(\mathrm{RSS}/\mathrm{n}\right) \tag{3.21}$$

where $k$ is the number of regression parameters.

# 4. CASE STUDY: TAXI TRIPS NEAR M2 SUBWAY LINE IN ISTANBUL

Istanbul has the 9th, most congested traffic in the world [71]. Therefore, it is essential to develop alternative transportation options that disturb the traffic to a minimum, such as metro, tram, or BRT systems. The evaluation of such systems must be carried carefully, and the potential effects of significant externalities must be detected to prevent the increase in traffic congestion.

The city has seven subway (M1a, M1b, M2, M3, M4, M5, and M6), four trams (T1, T2, T3, and T4), two cable car (TF1 and TF2) lines and a BRT line that carries most of the passengers throughout the city (Figure 4.1) [72]. The subway lines, except M3, are in operation seven days and 24 hours, which makes them an important public transportation option that can compete with congestion-creating transportation means [73]. According to Istanbul Metropolitan Municipality (IBB), most of the subway passengers are carried with the M2 line according to the previous year's statistics since the line covers many residential and commercial centers throughout the city shown with the green line in Figure 4.1 [74].

To limit the study area, the M2 subway line is selected for the evaluation of the outbreak on taxi trip numbers based on the socioeconomic characteristics due to its significant coverage area and the high number of passengers. M2 line includes 15 stations namely Yenikapı, Vezneciler-İstanbul Ü., Haliç, Şişhane, Taksim, Osmanbey, Şişli-Mecidiyeköy, Gayrettepe, Levent, 4.Levent, Sanayi Mahallesi, İTÜ-Ayazağa, Atatürk Oto Sanayi, Darüşşafaka and Hacıosman. Those stations cover 67 neighborhoods. In this section, the taxi GPS data, stations' characteristic data, and COVID-19 data are presented.

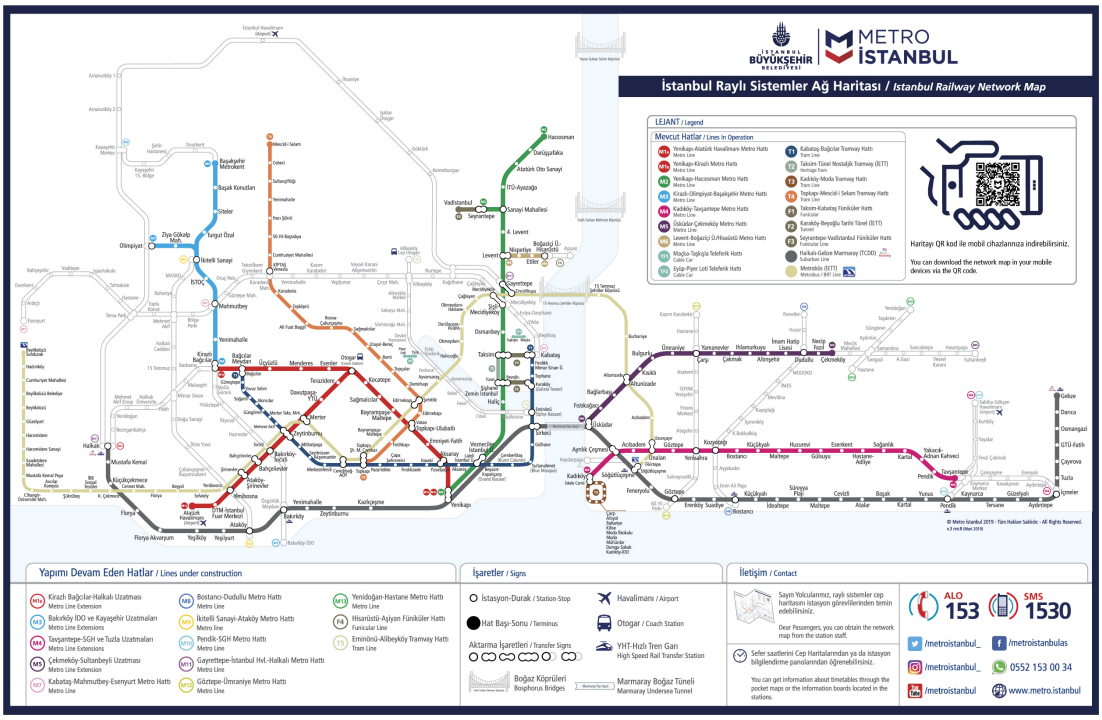


Figure 4.1. Istanbul railway network map.

## 4.1. Taxi GPS Data

Many mobile phone applications enable people to find and call a taxi online via smartphones. Some of the examples that are actively operated in Istanbul are BiTaksi, Careem, iTaksi, and so on. Apart from their advantageous features for users, they can provide significant information about the traffic network of the city via GPS tracking systems. The data in this study is taken from a mobile application, namely Bizimdurak [75], and it covers only some parts of Istanbul consisting of 499 taxis. The routes that are covered by the taxi data are shown in Figure 4.2 for the 15th of January, 2020. The GPS data starts on 31st December 2019 and continues until the 12th of June, 2020, except the 28th - 29th February and 1st - 2nd March. The data for those four days are missing, so 161 days are covered in total. For each day, there are more than 700000 data points.

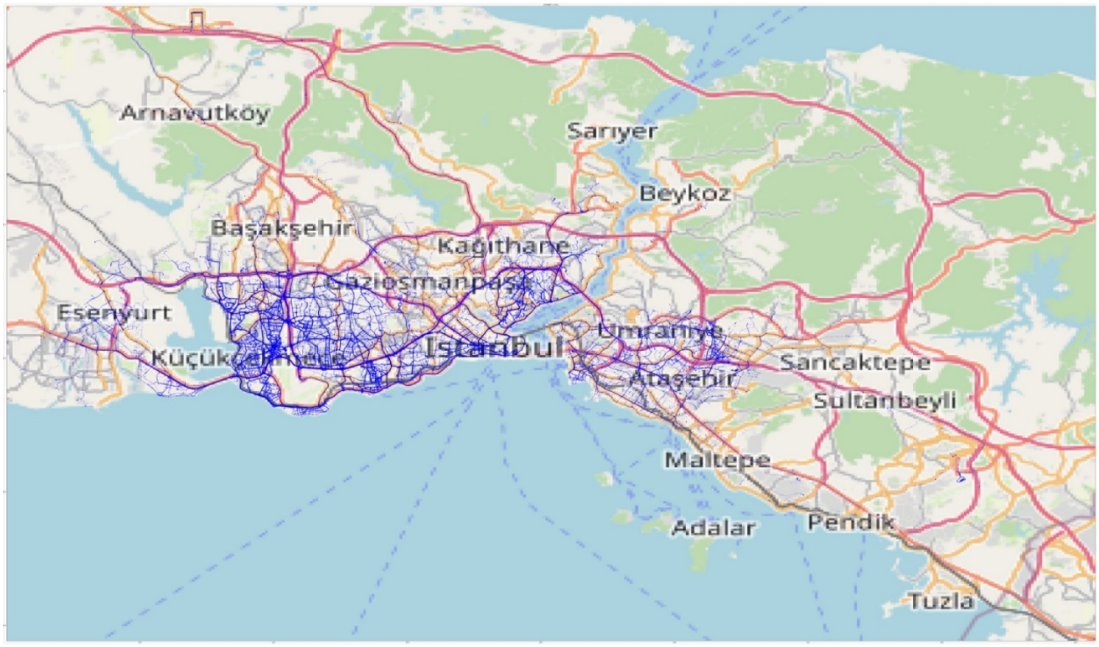


Figure 4.2. The GPS network for all taxi data on 15th January 2020 [76].

The raw data consists of a unique taxi ID, latitude, and longitude of the taxi, date, and time of the day (Table 4.1).

Table 4.1. The raw data sample

| **ID** | **Latitude** | **Longitude** | **Date** | **Time** |
|---|---|---|---|---|
| fb0de54b1719b85d85c5e78f176213db | 40.98549 | 28.87813 | 1/15/2020 | 0:00:00 |
| c32011b147f1d97574b60be02153ea3b | 41.00536 | 28.78546 | 1/15/2020 | 0:00:00 |
| d4b45c1eb6aa95b2f9e54b0d27965e32 | 41.01966 | 28.76796 | 1/15/2020 | 0:00:00 |
| a2f5bd1cc9f618196707ab2754dc4491 | 41.03421 | 28.77922 | 1/15/2020 | 0:00:00 |
| c51ca30d2d54ce3dbf8f62ea99e0a6ca | 41.01951 | 28.76802 | 1/15/2020 | 0:00:00 |

## 4.2. Characteristics Data of Each Station

The data for each station is prepared based on several categories. First, the latitude and longitude of each station are marked and transferred to Excel (Table 4.2). By doing so, a compatible relationship with the taxi data is obtained. So that circles with 800 meters radius can be drawn around each station as their coverage area.

Table 4.2. Latitudes and longitudes of each station.

| **Station Name** | **Lat** | **Lon** |
|---|---|---|
| **Hacıosman** | 41.139954 | 29.030778 |
| **Darüşşafaka** | 41.129659 | 29.025022 |
| **A. O. Sanayi** | 41.117980 | 29.024193 |
| **İTÜ** | 41.108113 | 29.020802 |
| **Sanayi** | 41.094241 | 29.004899 |
| **4.Levent** | 41.085796 | 29.006781 |
| **Levent** | 41.076496 | 29.014251 |
| **Gayrettepe** | 41.069092 | 29.010835 |
| **Şişli-Mecidiyeköy** | 41.064912 | 28.993250 |
| **Osmanbey** | 41.053559 | 28.987280 |
| **Taksim** | 41.038174 | 28.985638 |
| **Şişhane** | 41.028287 | 28.972828 |
| **Haliç** | 41.020495 | 28.964073 |
| **Vezneciler** | 41.012270 | 28.959842 |
| **Yenikapı** | 41.005593 | 28.951410 |

The demographic and socioeconomic data are mainly obtained from the "Mahallem Istanbul" project that is developed with the support of the Istanbul Development Agency [77]. The entire neighborhood that is covered by the project is 959, and the information is taken from various sources such as TURKSTAT, local public institutions, field research, and acquisitions. The project aims to reveal the socio-economic development level of each neighborhood in Istanbul and to create a socio-economic development index (SEGE) that covers indicators related to demographics, health, economic capacity, and transportation. The research uses the database categories, as shown in Table 4.3.

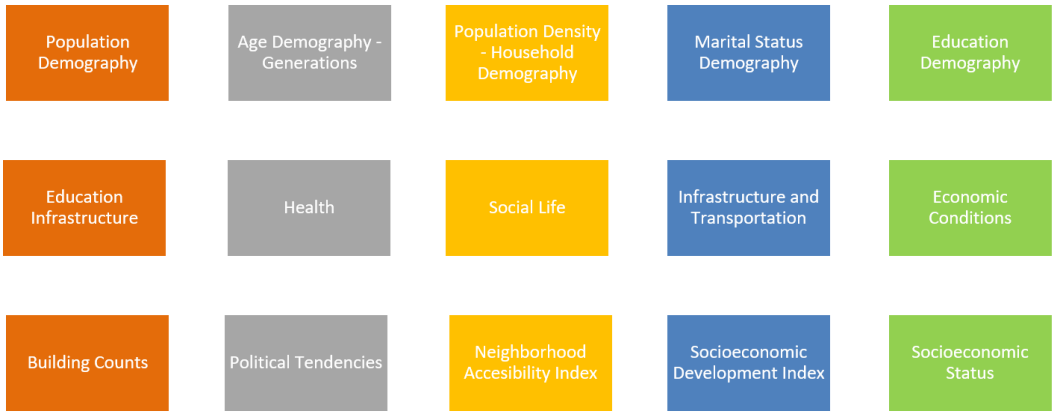


Figure 4.3. The database categories of the Mahallem Istanbul Project.

According to the study, there are two main indicators to understand the economic and social life of people that are education level and profession. Therefore, firstly a socioeconomic status (SES) research is performed in Istanbul to calculate SES Indexes, which is found with weighted averages. The calculated SES Indexes are compared with official SES Indexes and found highly similar. Then, the SEGE index is obtained from the collected data.

The population, population density, average age, average education years, university graduation ratio, household size, and SEGE (socioeconomic index) in the "Mahallem Istanbul" project are transferred to Excel manually. Hence, a data set is created, including the neighborhoods covered by each station (Table 4.3). However, the female percentage data is obtained from Endeksa [78] that is an online platform to create real estate information by using machine learning algorithms. Apart from socioeconomic indicators, two more parameters are also added to the analysis. One of them is the existence of another rail transport option. The value "1" is given if there is another possibility and "0" for the lack of possibility. As shown in the rail systems map in Table 4.1, Levent and Sanayi stations have a connection with another metro line.

In contrast, the connection in Şişli-Mecidiyeköy and Gayrettepe stations is with Metrobus that is a BRT line that carries a high number of passengers every day. Taksim station is one of the main hubs that carry not only city residents but also tourists as well. It has a connection with a funicular system. Şişhane station has

similar characteristics with Taksim and is connected with a nostalgic tram and tunnel. Vezneciler station locates in the old town of Istanbul, therefore a tourist attraction area connected with a tram. Lastly, Yenikapı is the most significant hub point among all stations that connects a suburban line, namely Marmaray and M1 metro line, that serves as a medium for two other lines.

The other parameter added to the analysis is the number of shopping malls around the station. By using the online map of Istanbul, the shopping malls are detected near the study area, and the numbers are recorded. Since, shopping malls are accepted as one of the main gathering areas of city dwellers in Istanbul, the closure of them is assumed to affect the travel counts of taxi substantially. Therefore, the addition of such information to the analysis might explain the relationship better.

The location of each station and covered neighborhoods for each primary catchment area are specified in the dataset along with total population, population density (population per 1000 square meter), average age, average education years, university graduation percentage, household size and calculated socio-economic development index (SEGE) as in Table 4.3. The data is averaged to obtain the characteristics of each station and is prepared for the regression analysis (Table 4.4).

Table 4.3. The independent variables for Hacıosman station.

| **Station Name** | **Lat** | **Lon** | **Neighborhood** | **Population** | **Pop. Density (per 1000m2)** | **Avg. Age** | **Avg. Education (in years)** | **University Grad. (%)** | **Household Size** | **SEGE** | **Female Percentage** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Hacıosman | 41.14 | 29.03 | Cumhuriyet | 11281 | 9.91 | 34.08 | 8.23 | 14.86 | 3.46 | 25.00 | 49.54 |
| | | | Çamlıtepe | 6186 | 4.24 | 33.83 | 8.02 | 12.76 | 3.58 | 31.70 | 50.61 |
| | | | PTT Evleri | 4956 | 1.12 | 33.64 | 8.29 | 14.49 | 1.67 | 24.20 | 45.63 |
| | | | Tarabya | 17945 | 7.05 | 37.53 | 10.01 | 26.00 | 3.04 | 64.40 | 51.89 |
| | | | **Average** | **10092** | **5.58** | **34.77** | **8.64** | **17.03** | **2.94** | **36.33** | **49.4175** |

Table 4.4. The final data for regression.

| Station Name | Lat | Lon | Population | Pop. Density (per 1000m2) | Avg. Age | Avg. Education (in years) | University Grad. (%) | Household Size | SEGE | Female Percentage | Public Transport Connection | Shopping Mall |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Hacıosman** | 41.14 | 29.03 | 10092 | 5.58 | 34.77 | 8.64 | 17.03 | 2.94 | 36.33 | 49.42 | 0 | 0 |
| **Darüşşafaka** | 41.13 | 29.03 | 7169.33 | 3.94 | 34.41 | 10.52 | 30.07 | 3.01 | 64.5 | 50.61 | 0 | 0 |
| **A. O. Sanayi** | 41.12 | 29.02 | 9021.33 | 7.68 | 33.99 | 10.47 | 29.87 | 3.01 | 70.37 | 50.44 | 0 | 0 |
| **İTÜ** | 41.11 | 29.02 | 10852.25 | 6.49 | 32.81 | 9.98 | 25.25 | 3.41 | 69.85 | 49.57 | 0 | 1 |
| **Sanayi** | 41.09 | 29.00 | 16055.00 | 26.80 | 33.73 | 9.59 | 25.27 | 3.30 | 59.32 | 50.10 | 1 | 1 |
| **4.Levent** | 41.09 | 29.01 | 19806.50 | 37.68 | 35.69 | 9.78 | 25.61 | 3.17 | 61.78 | 50.72 | 0 | 2 |
| **Levent** | 41.08 | 29.01 | 11485.00 | 20.36 | 39.72 | 10.93 | 32.50 | 2.77 | 82.15 | 53.13 | 1 | 4 |
| **Gayrettepe** | 41.07 | 29.01 | 12458.80 | 20.85 | 38.89 | 11.29 | 35.37 | 2.68 | 82.08 | 53.12 | 1 | 4 |
| **Şişli-Mecidiyeköy** | 41.06 | 28.99 | 14564.00 | 36.07 | 35.51 | 9.48 | 24.71 | 2.94 | 60.12 | 50.56 | 1 | 2 |
| **Osmanbey** | 41.05 | 28.99 | 6195.00 | 30.38 | 40.96 | 10.70 | 29.85 | 2.47 | 69.93 | 53.61 | 0 | 2 |
| **Taksim** | 41.04 | 28.99 | 2505.78 | 16.76 | 38.10 | 9.84 | 24.7 | 2.45 | 58.11 | 45.19 | 1 | 1 |
| **Şişhane** | 41.03 | 28.97 | 1575.90 | 15.91 | 37.80 | 9.14 | 22.44 | 2.64 | 43.99 | 41.55 | 1 | 2 |
| **Haliç** | 41.02 | 28.96 | 3088.00 | 12.26 | 35.43 | 6.97 | 8.44 | 3.07 | 24.55 | 37.09 | 0 | 1 |
| **Vezneciler** | 41.01 | 28.96 | 865.60 | 7.02 | 29.47 | 9.47 | 19.99 | 5.67 | 33.86 | 36.76 | 1 | 1 |
| **Yenikapı** | 41.01 | 28.95 | 3941.8 | 11.38 | 33.76 | 8.41 | 17.92 | 3.03 | 29.38 | 44.16 | 1 | 2 |

### 4.3. COVID-19 Data

Any data cannot be interpreted independently of its timeline. Therefore, the COVID-19 timeline is prepared based on the critical events that cover taxi-related events as well (Figure 4.6) [79]. The first case was seen on the 1st of December 2019. The effects of COVID-19 is felt around February 2020 throughout the world. However, the timeline and reactions have unique characteristics for different countries.

In Turkey, thermal cameras are installed at the airports on January 24th to determine infected people since fewer is one of the first symptoms of COVID-19. Then, on February 3rd, all flights from China are stopped following that other flights from high-risk countries such as Italy and South Korea are also suspended on the 29th of February. In the meantime, the disinfection of public places like buses, airports, and schools are started. Despite the taken measures, the first case in Turkey is seen on May 11, and the daily number of cases increased substantially in a short time. Until the first case, the conditions and the pace of life were normal. Therefore, this period is taken as Phase 1 and called "Pre-COVID".

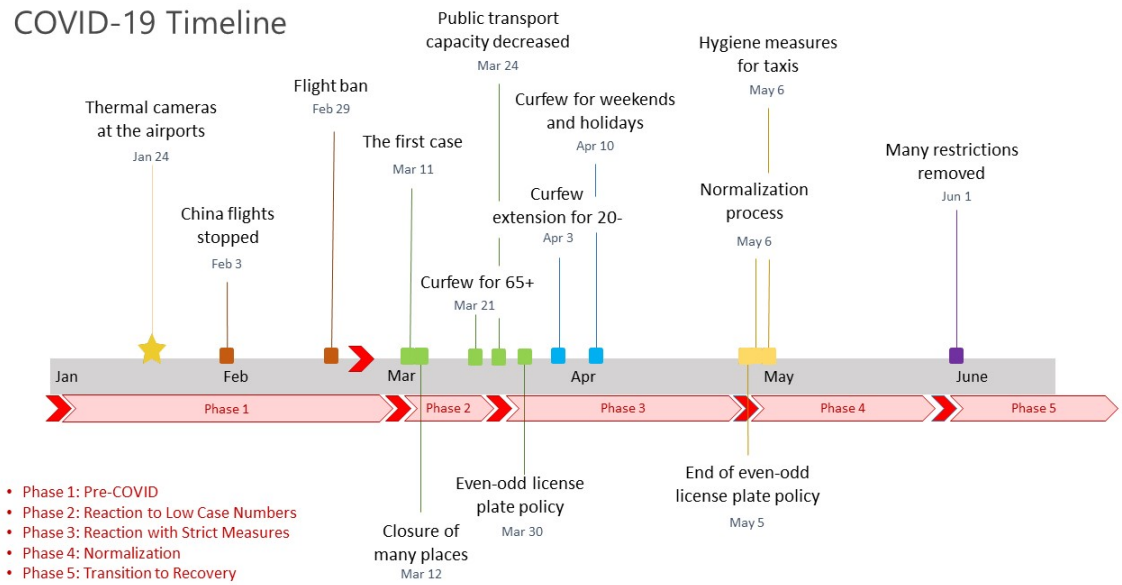


Figure 4.4. COVID-19 Timeline.

All in all, the timeline and phases are prepared by considering milestone events and case numbers, and each important date is inserted into the taxi data to observe the differences in taxi trip numbers. The date interval for each phase explained above is given in Table 4.5.

Table 4.5. COVID-19 phases.

| **Phase Number** | **Phase Name** | **Beginning** | **End** |
|---|---|---|---|
| 1 | Pre-COVID | 01.01.20 | 11.03.20 |
| 2 | Reaction to Low Case Numbers | 12.03.20 | 24.03.20 |
| 3 | Reaction with Strict Measures | 25.03.20 | 05.05.20 |
| 4 | Normalization | 06.05.20 | 31.05.20 |
| 5 | Transition to Recovery | 01.06.20 | 12.06.20 |

After Phase 1, there is a short period where the initial reactions to the pandemic are observed. Starting from the 12th of March, the government announced that the education would be suspended for one week in schools and three weeks in universities beginning on the 16th of March, and the education would be given online afterward. In the same period, all night clubs, pubs, and bars are closed, and crowded events, including public prayers, cinemas, gyms, concerts, weddings are canceled. Despite the closures of many events and public activities, the severity of the situation cannot be accepted by the public immediately, and the number of cases continued to increase. The first peak in the case numbers occurred on the 20th of March. Most were in Istanbul. A curfew for people older than the age of 65 and with chronic diseases is started. The reaction of the government to the increased number of cases came four days later, with a 50% decrease in public transportation capacity. Yet, starting from the 25th of March, the course of disease became more serious (Figure 4.4 and Figure 4.5) [80]. Therefore, Phase 2 covers the dates between 12th to 25th of March, and it is called “Reactions to Low Case Numbers”.

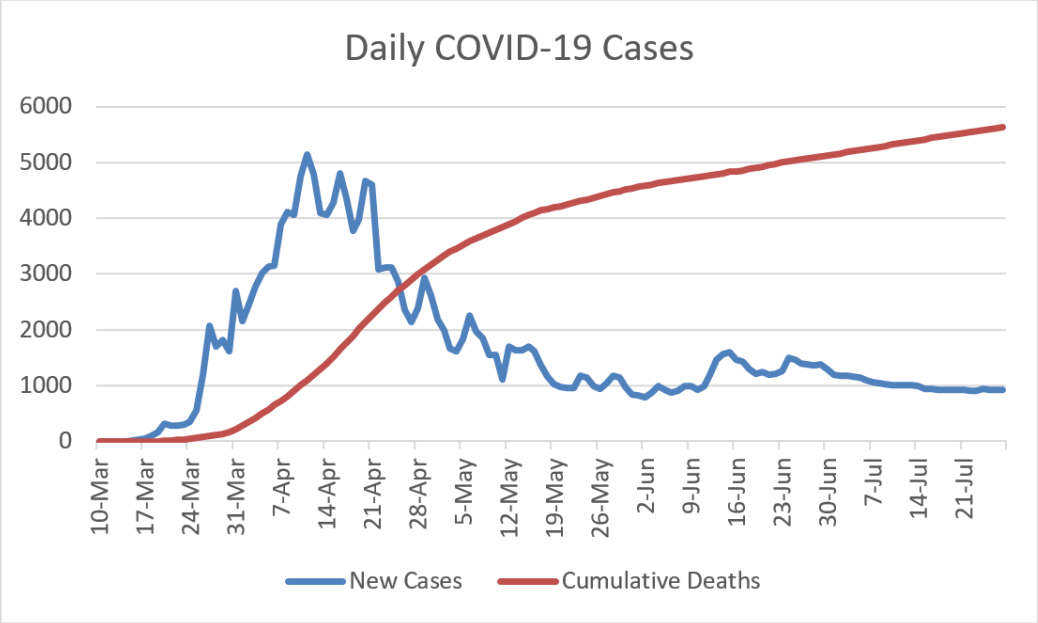


Figure 4.5. COVID-19 daily cases and deaths.

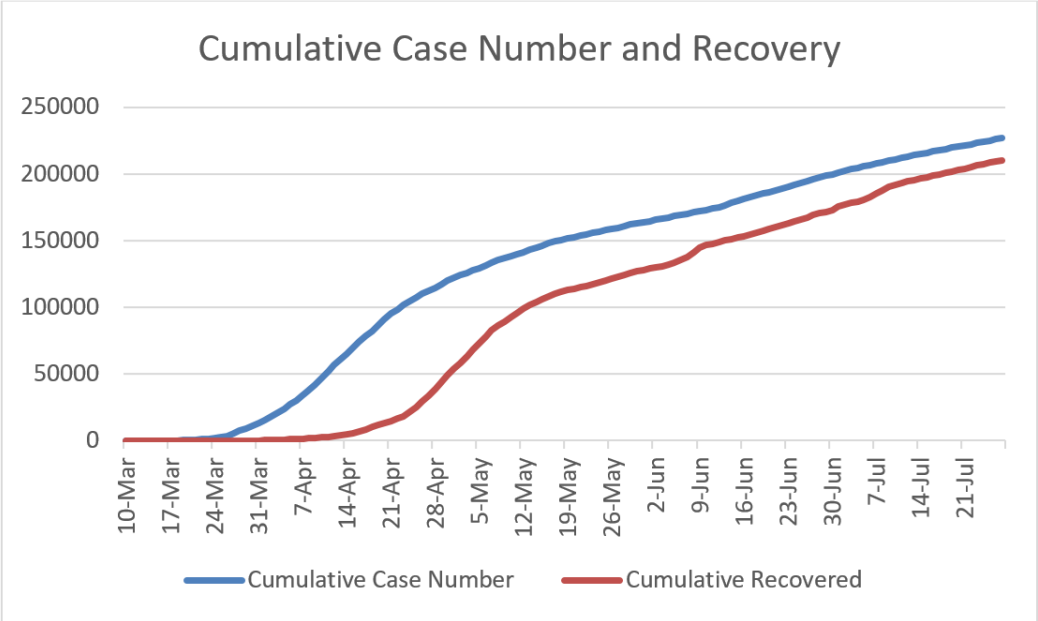


Figure 4.6. Cumulative COVID-19 cases and recoveries.

After Phase 2, a period with more strict measures started, an even-odd license plate policy that restricts taxis according to their plate number applied on the 30th of March, the curfew is extended for people below the age of 20. Soon afterward, a curfew for weekends and holidays enforced.

The period between the dates 25th of March and 6th of May is Phase 3 and called "Reaction with Strict Measures". The 3rd period was the most challenging time since the case numbers were increasing swiftly. The case numbers were almost fixed. The collective consciousness was formed with the intervention of media by the end of Phase 3. The normalization period is announced on the 4th of May. The taxi restriction with odd and even plate numbers has ended a day after. The government officially started the normalization period on the 6th of May with the announcement of the minister of health and defined it as the "new normal". With the "new normal", a circular letter is published by the ministry of interior for hygiene measures in taxis [60]. Therefore, the "Normalization" period, namely Phase 4, started on the 6th of May until the 1st of June. During the process, curfew periods are loosened, and specific times for people above 65 and below 20 years old are defined to go out. Hairdressers and shopping malls are reopened.

Lastly, the 5th Phase began after the announcement of the opening of all restaurants, cafes, gyms, and parks on June 1st and it is called the "Transition to Recovery". Right after the beginning of the phase, the curfews are removed. During the process, some restrictions related to social distance and hygiene are forced, and the authorities control the inspections for the rules. At the beginning of the phase, there is a transition period from the "old normal" to the "new normal. Here, "old normal" defines the life before COVID-19, whereas "new normal" defines a more cautious life with masks and social distance. This period is followed by a recovery period where people started to return to their daily life.

# 5. METHODOLOGY

The studies related to pandemics are quite limited in the literature. Especially COVID-19, as a hot topic that affected many areas all around the world, offers limitless options to investigate. The case study area explained in section 4 will be the main focus of this thesis. Based on the literature review and the data examination, the road map and flowchart are drawn and presented in Figure 5.1. The data collection of the study is explained in section 4 in detail. In this section of the thesis, first, the concept of the primary catchment area (PCA) and its application are introduced to restrict the study area within one of the M2 subway stations. Afterward, the preprocessing of the data is explained to give insights about the data. The preprocessing mainly consists of three parts that include the data clearance based on the PCA, data separation based on daily periods, and calculation of daily average trip count changes. The correlation matrix is utilized to reduce the number of variables as a part of the preliminary analysis. Besides, the data plots of the average daily trip count, events, and daily COVID-19 cases are created to have an insight. Then, the obtained information from the data is introduced to present a broader perspective, and k-means clustering is used to observe possible relations between different zones. Finally, the data is analyzed by using various techniques, namely ordinary least squares (OLS), spatial auto regression (SAR), and geographically weighted regression (GWR). The result of the analysis is divided into two sections as clustering and regression. The findings are presented and compared based on statistical performances.

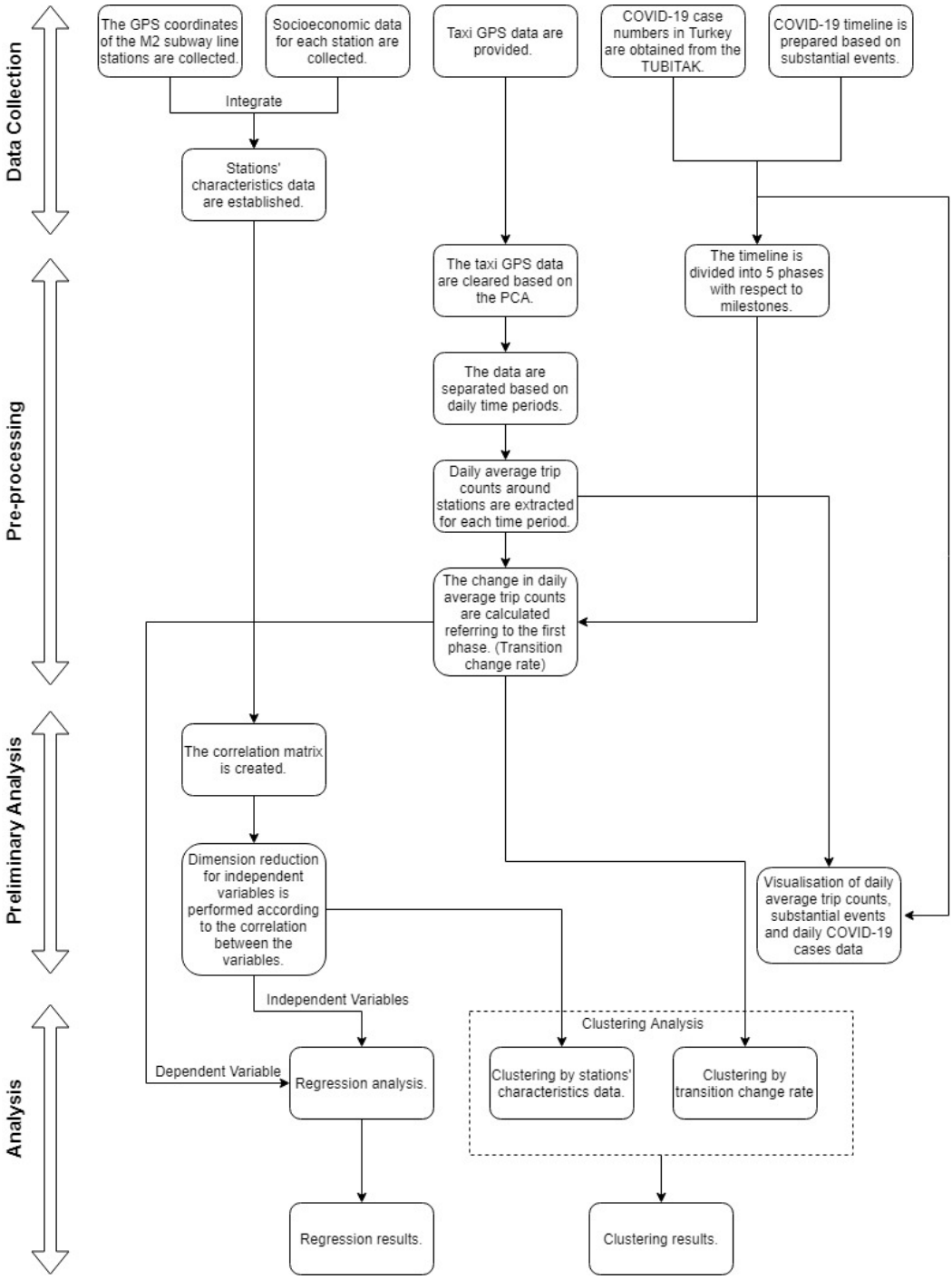


Figure 5.1. The flowchart of the methodology.

## 5.1. Primary Catchment Area

In the literature, different distances are considered as the reachable distance to a point. In public transportation, the reachable distance concept is used to define

the demand or number of users who can reach by walking to a station and named as the catchment area of the station. The difference between accepted catchment areas is affected by land-use characteristics of the region, such as stairways or slopes [81]. Besides, population, employment, population density, and land use also plays a role in the catchment distance [82].

According to the recommended practice by American Public Transportation Association (APTA), spatial areas are investigated under three categories, namely core station area, primary catchment area (PCA), and secondary catchment area, as in Figure 5.2 [83]. As mentioned in the standard, the secondary catchment area requires a means of transit to reach. In contrast, PCA is considered as the main area of influence where pedestrian access is possible. The PCA focuses on land use, pedestrian, and mobility access. Therefore, it is considered as the reference distance in this study.

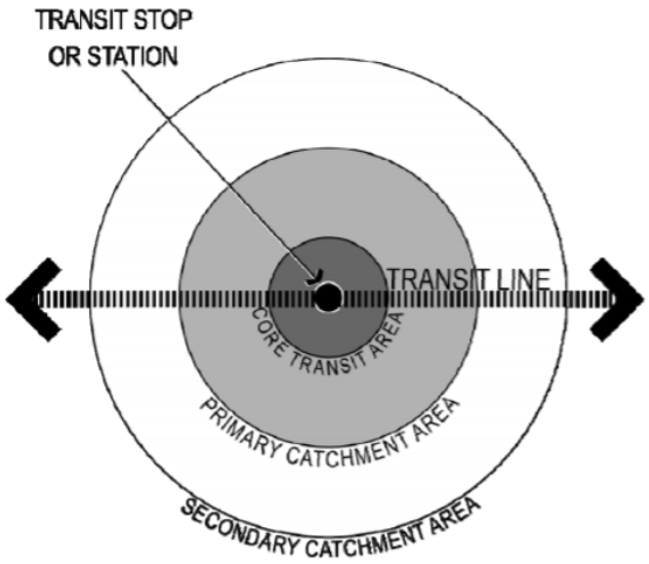


Figure 5.2. The concept of catchment area.

The standard assumes that the topography is relatively flat, and street networks have grid-like connections while deciding on the PCA radius. The distances are decided because a typical adult should reach within ten minutes' walk. According to the assumptions and considerations, the PCA should be 1/2 mile for semi-rapid and regional transit and 2/3 mile for rapid transit. By considering the hilly land conditions

and irregular urbanization of Istanbul, the assumptions are taken into consideration for the case study, and the PCA radius is taken as 800 meters (1/2 mile).

In this study, the primary catchment areas for each station on the M2 subway line are circled to reveal the neighborhoods covered by the subway. The radius for primary catchment areas is taken as 800 meters, and each circle is drawn in Google Earth Professional, as shown in Figure 5.3. The neighborhoods that are covered by the stations are determined separately from the Istanbul Metropolitan Municipality online city map, which consists of each neighborhood in the city [84]. The socio-economic data for each neighborhood data are averaged to obtain a value for each station.

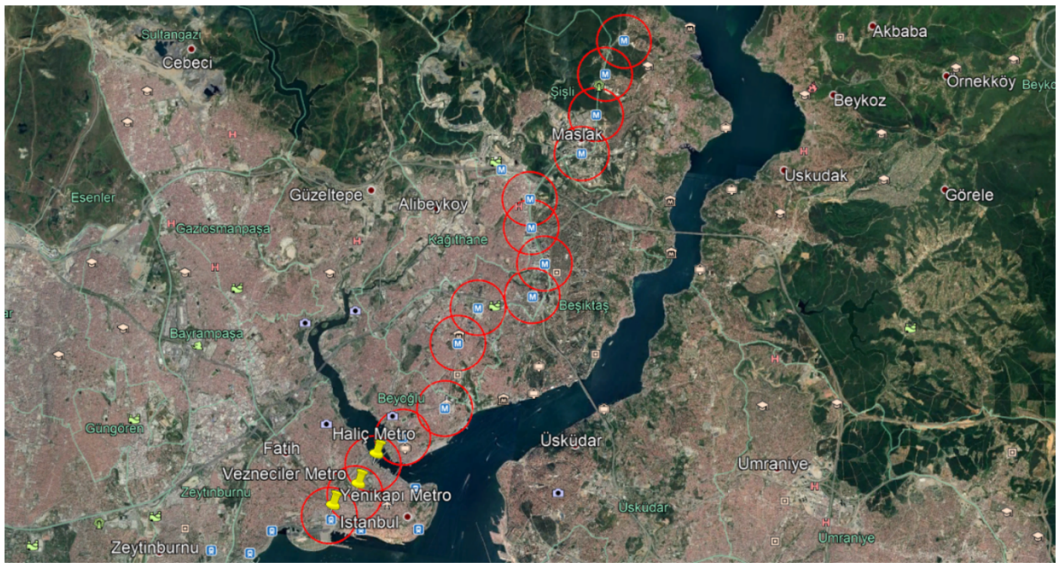


Figure 5.3. The catchment areas for each station [76].

## 5.2. Preprocessing of Taxi Data

Understanding the data is an important step to find out the proper methodology. Therefore, some assumptions are made throughout the process to reveal the pattern of the data. First, the primary catchment areas are used to limit the taxi data for each station. Afterwards, the number of taxi trips is calculated within the limited area in four different segments that are the total trip counts, trip counts at the morning peak hour, trip counts at the evening peak hour, and off-peak trip counts. Finally, the K-means clustering is used to observe similarities between stations, phases, and

socio-economic characteristics.

### 5.2.1. Primary Catchment Area Calculation

First, to be able to limit the taxi data with the metro stations, a circle with an 800 meters radius is drawn for each station. Here, the latitude and longitude for each station are taken as the center of the circle. Then, an approximation assuming a circle has 36 edges is used. In other words, each circle is assumed as a polygon with 36 angles, and each side has an angle of 10°. The assumption is accurate for distances below 100 km. The latitudes and longitudes are found by using the equatorial radius of the earth that is 6,378,137 m for GPS coordinates [85]. The equations to find each corresponding points in terms of latitude and longitude are given below [86]:

$$\theta = \frac{2\pi \mathrm{k}}{\mathrm{N}} \tag{5.1}$$

$$\mathrm{dx} = \mathrm{r}\cos(\theta) \tag{5.2}$$

$$\mathrm{dy} = \mathrm{r}\sin(\theta) \tag{5.3}$$

$$\mathrm{a} = \mathrm{cLat} + \left(\frac{180}{\pi}\right)\left(\frac{\mathrm{dy}}{\mathrm{r_e}}\right) \tag{5.4}$$

$$\mathrm{b} = \mathrm{cLon} + \frac{\left(\frac{180}{\pi}\right)\left(\frac{\mathrm{dy}}{\mathrm{r_e}}\right)}{\cos\left(\mathrm{cLat}\frac{\pi}{180}\right)} \tag{5.5}$$

where $N$ (number of edges) equals 36, $r$ (radius) equals 800 meters, $cLat$ is the latitude of the center, $cLon$ is the longitude of the center, $r_e$ (equatorial radius of earth) equals 6,378,137 meters, $a$ and $b$ are the latitudes and longitudes for an edge that corresponds to the respective center.

The above formulas are applied to each station iteratively, and 36 edges are defined separately for 15 stations. Then, the process is applied for each month to include the taxi data that remains inside the circles. The data is restricted to the stations.

### 5.2.2. Trip Count Calculations

After defining the area and the taxi data inside the area, the trip numbers are identified for each station. In the taxi data, the information related to the duration of a trip is not specified. Therefore, an assumption is made by considering the speeds of taxis and the primary catchment area. The great circle distance between two points on earth are found by using Haversine Formula [86]:

$$\mathrm{t} = \mathrm{t}_2 - \mathrm{t}_1 \tag{5.6}$$

$$d_{lat} = \mathrm{lat}_2 - \mathrm{lat}_1 \tag{5.7}$$

$$d_{lon} = \mathrm{lon}_2 - \mathrm{lon}_1 \tag{5.8}$$

$$\mathrm{a} = \sin\left(\frac{\mathrm{dlat}}{2}\right)^2 + \cos\left(\mathrm{lat}_1\right)\cos\left(\mathrm{lat}_2\right)\sin\left(\frac{\mathrm{dLon}}{2}\right)^2 \tag{5.9}$$

$$\mathrm{c} = 2\arctan\left(\frac{\sqrt{(1-\mathrm{a})}}{\sqrt{\mathrm{a}}}\right) \tag{5.10}$$

$$\mathrm{d} = \frac{\mathrm{Rc}}{10} \tag{5.11}$$

where, $t$ equals time difference for the same taxi ID, $d_{lat}$ equals latitude difference for the same taxi ID in $t_2$ and $t_1$, $d_{lon}$ equals longitude difference for the same taxi ID in $t_2$ and $t_1$, $R$ (Radius of the earth in km) equals 6373.

After the distances are calculated, the speeds are found by using the following formula:

$$v = \frac{d}{t} 3600 \tag{5.12}$$

where, $v$ equals speed of the taxi between $t_2$ and $t_1$

As a result, the average speed for all of the taxi data is calculated as 30 km/h. Theoretically, the maximum distance that can be covered inside the primary catchment area is the perimeter of the circle that is found 5 km with the following formula:

$$P = 2\pi r \tag{5.13}$$

where, $P$ equals perimeter of the circle, $r$ (radius of the circle) equals 800 m.

The distance of 5 km can be covered in 10 minutes with 30 km/h speed. Besides, an average trip in the same circle is expected to take shorter than 10 minutes. Therefore, each trip around a station is taken as 10 minutes, and taxi data are divided trips that continue for 10 minutes. Then trip counts are averaged for each day, and each station separately (Table 5.1).

Table 5.1. Taxi trip counts around each station on 1st January 2020.

| | **Station** | **Month** | **Day** | **Taxi Count** |
|---|---|---|---|---|
| **0** | Hacıosman | 01 | 01 | 2 |
| **1** | Darüşşafaka | 01 | 01 | 7 |
| **2** | Atatürk Oto Sanayi | 01 | 01 | 10 |
| **3** | İTÜ | 01 | 01 | 37 |
| **4** | Sanayi | 01 | 01 | 72 |
| **5** | 4.Levent | 01 | 01 | 75 |
| **6** | Levent | 01 | 01 | 427 |
| **7** | Gayrettepe | 01 | 01 | 1365 |
| **8** | Şişli-Mecidiyeköy | 01 | 01 | 1949 |
| **9** | Osmanbey | 01 | 01 | 3193 |
| **10** | Taksim | 01 | 01 | 7316 |
| **11** | Şişhane | 01 | 01 | 3977 |
| **12** | Haliç | 01 | 01 | 1732 |
| **13** | Vezneciler | 01 | 01 | 4761 |
| **14** | Yenikapı | 01 | 01 | 6257 |

### 5.2.3. K-means Clustering

In clustering, the main focus is to investigate the different reactions in stations to COVID-19 phases. The phases are explained before in detail in section 4.3 with Table 4.5. To be able to observe the decrease due to the pandemics in each station, the taxi trip counts are compared. The Pre-COVID phase is taken as the base for all transitions because the main aim is to compare the difference in changes to detect the effects of different factors on those changes. Therefore, four different comparisons are created with the following formula:

$$T_1 = \frac{P_2 - P_1}{P_1} 100 \tag{5.14}$$

$$\mathrm{T}_2 = \frac{\mathrm{P}_2 - \mathrm{P}_1}{\mathrm{P}_1} 100 \tag{5.15}$$

$$\mathrm{T}_3 = \frac{\mathrm{P}_2 - \mathrm{P}_1}{\mathrm{P}_1} 100 \tag{5.16}$$

$$\mathrm{T}_4 = \frac{\mathrm{P}_2 - \mathrm{P}_1}{\mathrm{P}_1} 100 \tag{5.17}$$

where $T_1$ is the trip count decrease between Phase 1 and 2, $T_2$ is the trip count decrease between Phase 1 and 3, $T_3$ is the trip count decrease between Phase 1 and 4, $T_4$ is the trip count decrease between Phase 1 and 5, $P_1$ is Pre-COVID (Phase 1), $P_2$ is Reaction to Low Case Numbers (Phase 2), $P_3$ is Reaction with Strict Measures (Phase 3), $P_4$ is Normalization (Phase 4) and $P_5$ is Transition to Recovery (Phase 5).

Observing latitudes and longitudes on the $x$- and $y$-axis would result in a scattered plot without an accurate representation of the distances between stations. To be able to represent the distances between stations properly in a scaled plot, the fact that the distance between two consecutive latitudes always equals 111.32 km is used. From that, the distances for longitudes are calculated as:

$$\mathrm{Lon} = \frac{\cos\left(\mathrm{Lat}\right)}{\mathrm{nLat}} \mathrm{C}_{earth} \tag{5.18}$$

where $nLat$ (total number of latitudes) equals 360, $C_{earth}$ (circumference of the earth) equals 40075 km and $dLat$ (distance between two consecutive latitudes) equals $\frac{\mathrm{C}_{earth}}{\mathrm{nLat}}$ = 111.32 km.

So, by using the above formula, the distances between stations are represented in exact order.

## 5.3. Analysis of the Data

After the preprocessing of data is completed, the data is divided into different segments to reveal the patterns of taxi trip usage. Segmentation is the process of dividing the data into smaller and more meaningful subgroups to define similar characteristics of the groups rather than accepting all the data as one. Here, two primary separations are considered clarifying the differences between the change in travel behavior and socio-economic characteristics. One of the essential variables in traffic is the peak hours that define the busiest hours of the traffic. During peak hours, the traffic demand is at its top, and many characteristics of traffic tend to change than its standard. Therefore, peak hours are investigated separately in terms of taxi trip counts. There are four different analyses in this chapter, namely morning peak hour, evening peak hour, off-peak hour, and total taxi trip counts.

First, the total trip counts are calculated, and then they are divided into two durations for peak and off-peak periods. The peak hours are taken as the morning peak between 6 to 9 am and evening peak between 5-8 pm. The times except peak hours are taken as the off-peak trip counts. The differentiation between durations can show different characteristics in terms of socio-economic parameters and clustering because they serve different purposes. For example, the morning peak mainly covers work-related trips, whereas evening trips also include recreational activities. In Istanbul, the morning peak covers the hours between 6 to 9 am, and the evening peak is between 5 to 8 pm. Mostly, the evening peak is more congested than the morning peak in Istanbul. Two of the peak hours are investigated separately in the analysis. The Python language is utilized for the analysis of the data in this thesis.

### 5.3.1. Total Trip Counts

In this section, first, the plots for each month are investigated separately by considering the timeline of COVID-19 to reveal the pattern of total trip counts. Then, clustering for each transition period is presented according to the station trip count changes. Finally, the regression results are presented for OLS, SAR, and GWR.

5.3.1.1. Trip Count Plots. The counted taxi trip data is combined in the plots for each month and station. Here, the primary purpose is to observe the general trends in the data. So that insight for the public reaction can be obtained during the COVID-19 period.

Below in Figure 5.4, the trip numbers in every station are given for January. The dashed line represents thermal camera installation in airports on January 24. There is a slight decrease in most of the stations after the thermal cameras, yet the reason for this decrease might also be the precipitation levels that are higher in January [87]. Previous studies have shown that people use taxis more when it is raining [88]. Therefore, the possible effects of rainfall must also be considered for the January period. The red-colored parts on the lines represent weekends, namely Saturdays and Sundays. A striking trend is that at weekends the trip counts decrease in every station. On Sundays, it reaches the lowest taxi trip of the week. Meaning that the taxi usage for recreational activities is lower around those stations, and the preferences are business focused. The trend for weekends can be observed for all the months.

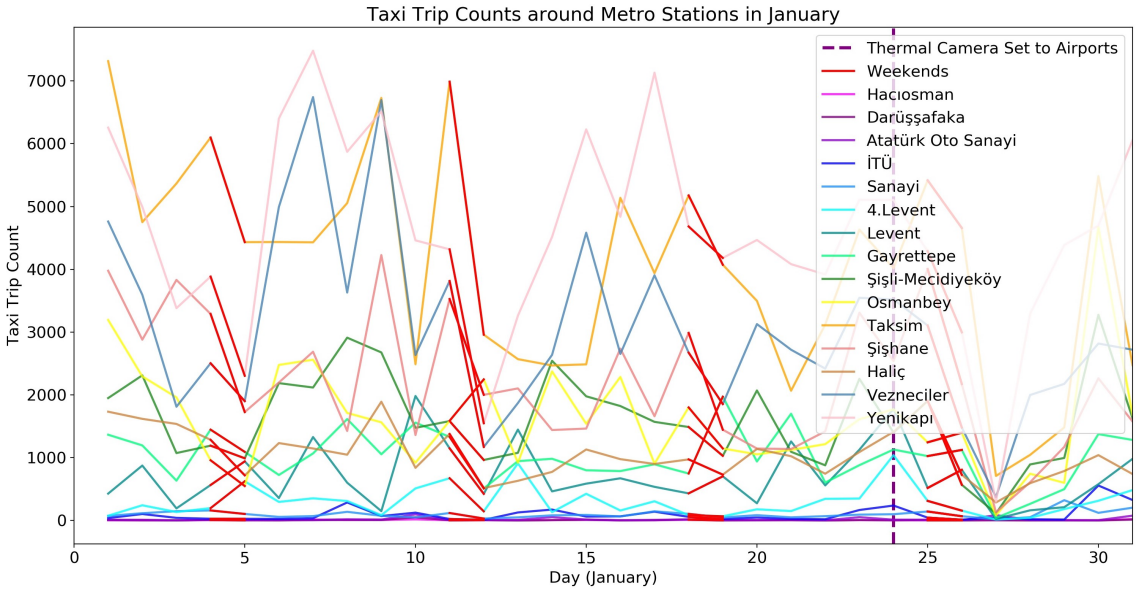


Figure 5.4. Total January counts.

In February, there are two main events namely the ban of China flights on the 3rd and the ban of flights for high-risk countries on the 29th of the month. Yet, both don't

significantly affect the taxi trip counts as in Figure 5.5. Besides, the latter cannot be observed, since the data between 27th February and 3rd March is missing.

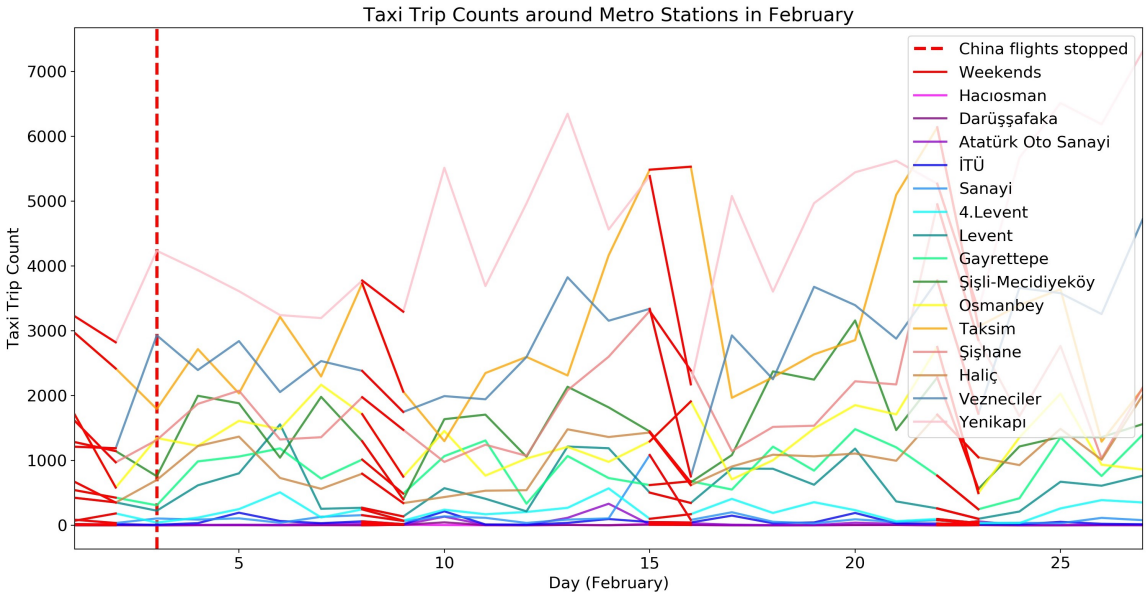


Figure 5.5. Total February counts.

The pre-COVID period ends on March 11 with the first COVID case in Turkey. Right after the first case, many places are closed, and public activities are stopped, as explained in Section 4.3. After the closures, a substantial amount of decrease in taxi trips is observed for every station. On the 25th March, there is a peak in the number of cases. Therefore, the Phase 2 ends, and a new phase with strict measures starts. The effects of the critical increase in case numbers can be observed in the taxi trip count as well (Figure 5.6). At the end of the month, the application for even and odd license for taxis is performed. Yet, the trip numbers were already relatively low by then.

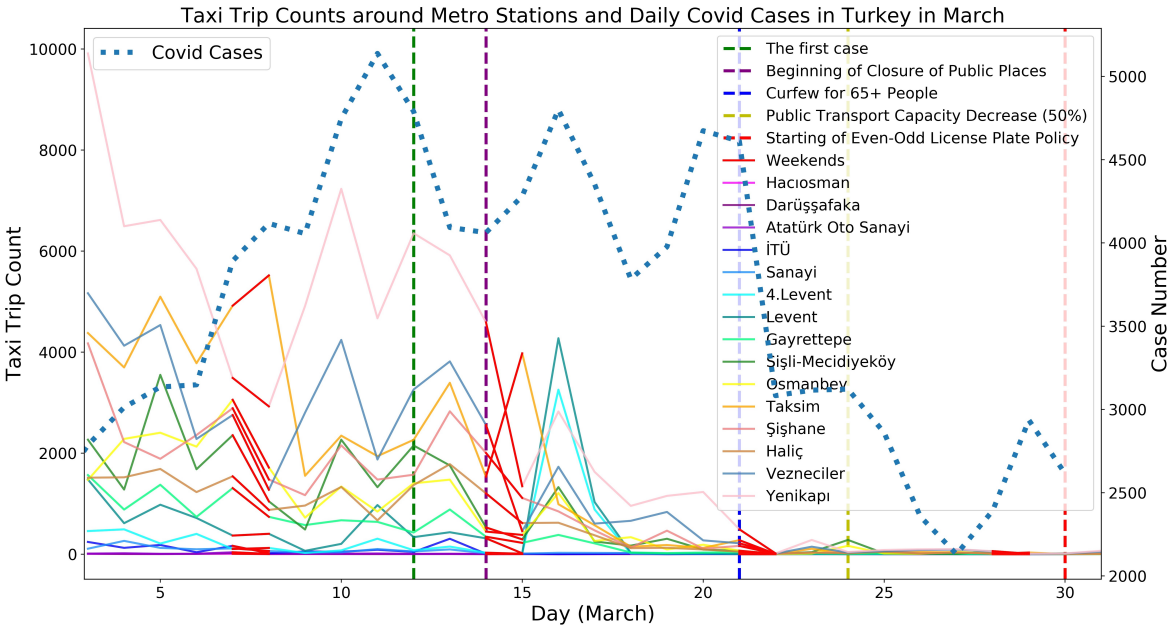


Figure 5.6. Total March counts.

The effects of the curfew for holidays and weekends on April 10 can be observed in taxi trip counts (Figure 5.7). The daily COVID-19 cases decrease in this month substantially as a result of the taken measures.

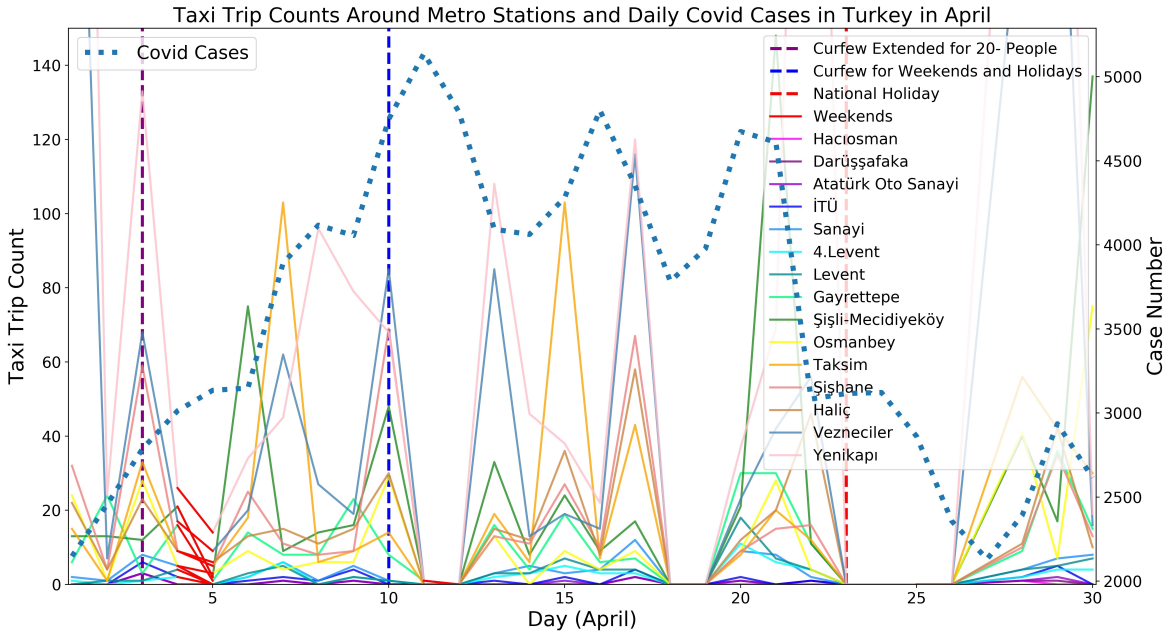


Figure 5.7. Total April counts.

On May 5, the even and odd plate policy ends, and the day after, the normalization process began officially. The effects of decreasing case numbers and the normalization process can be observed in taxi trip counts. The trips are increasing except weekends and holidays, because the period between 16-19 May and 23-26 May are national and religious holidays, respectively. During those times, there are still curfews. Therefore, no taxi trip exists (Figure 5.8).

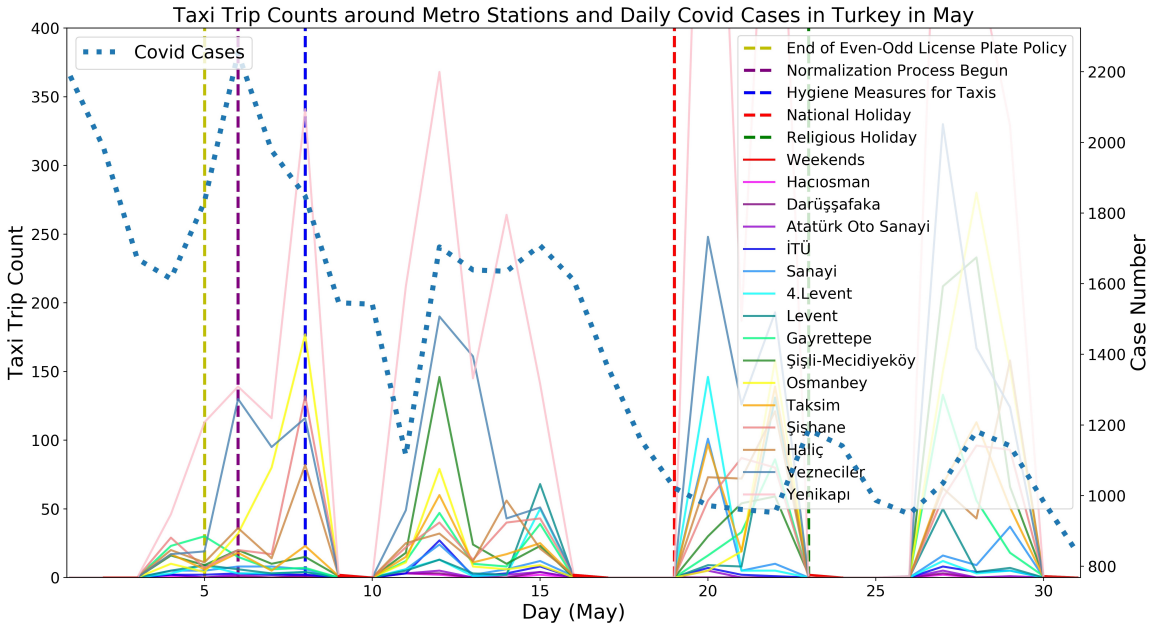


Figure 5.8. Total May counts.

At the beginning of June, many restrictions are removed, and public places are reopened. The phase after June 3 is accepted as the "Transition to Recovery". In this phase, the recovery cannot be achieved yet, because the case numbers are again started to increase. Yet, with the removal of restrictions, taxi trip counts increased, as in Figure 5.9.

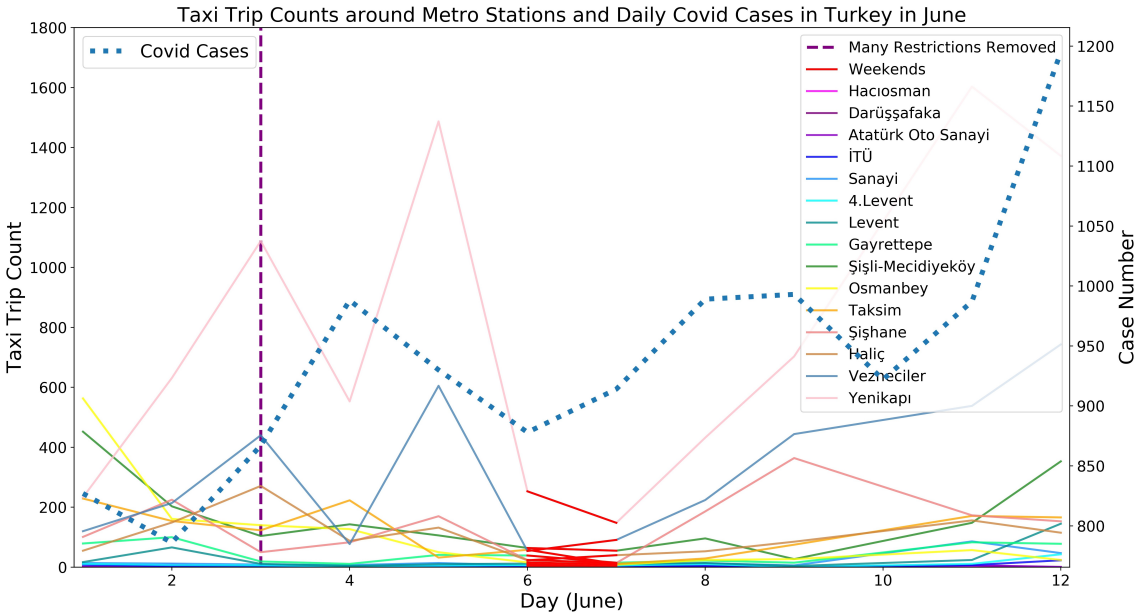


Figure 5.9. Total June counts.

5.3.1.2. Clustering. The results for each station and transition period are given in Table 5.2. The most remarkable decrease can be observed in Transition 2. The reaction to the strict measures has most of the effect on taxi trip counts, whereas the least effect is observed in Transition 1.

Table 5.2. The decrease in transitions for total trip counts.

| **Station Name** | **Pre-COVID** | **Reaction to Low Case Numbers** | **Reaction with Strict Measures** | **Normalization** | **Transition to Recovery** | **Transition1** | **Transition2** | **Transition3** | **Transition4** |
|---|---|---|---|---|---|---|---|---|---|
| Hacıosman | 3.97 | 0.67 | 0.30 | 0.40 | 1.70 | -83.21 | **-92.37** | -89.92 | **-57.18** |
| Darüşşafaka | 5.73 | 1.00 | 0.44 | 0.56 | 1.40 | -82.54 | **-92.33** | -90.22 | -75.56 |
| Atatürk Oto Sanayi | 24.61 | 2.08 | 0.71 | 1.16 | 2.00 | -91.53 | -97.11 | -95.29 | -91.87 |
| İTÜ | 81.82 | 29.25 | 1.79 | 2.92 | 5.50 | -64.25 | -97.81 | -96.43 | -93.28 |
| Sanayi | 108.79 | 19.83 | 5.82 | 9.72 | 20.50 | -81.77 | **-94.65** | -91.07 | -81.16 |
| 4.Levent | 259.91 | 364.00 | 5.17 | 10.24 | 9.90 | **40.05** | -98.01 | -96.06 | -96.19 |
| Levent | 653.18 | 511.17 | 6.41 | 12.28 | 30.10 | **-21.74** | -99.02 | -98.12 | -95.39 |
| Gayrettepe | 931.48 | 183.50 | 13.26 | 18.88 | 42.30 | -80.30 | -98.58 | -97.97 | -95.46 |
| Şişli-Mecidiyeköy | 1594.56 | 426.50 | 26.35 | 36.16 | 129.90 | -73.25 | -98.35 | -97.73 | -91.85 |
| Osmanbey | 1505.80 | 386.50 | 23.74 | 45.40 | 63.40 | -74.33 | -98.42 | -96.98 | -95.79 |
| Taksim | 3528.12 | 943.33 | 20.88 | 25.80 | 104.20 | -73.26 | -99.41 | -99.27 | -97.05 |
| Şişhane | 2059.35 | 683.00 | 20.80 | 31.24 | 144.20 | -66.83 | -98.99 | -98.48 | -93.00 |
| Haliç | 1039.85 | 419.92 | 20.33 | 31.08 | 111.20 | -59.62 | -98.04 | -97.01 | -89.31 |
| Vezneciler | 2900.86 | 944.92 | 56.94 | 75.88 | 343.00 | -67.43 | -98.04 | -97.38 | -88.18 |
| Yenikapı | 4690.06 | 1706.08 | 116.77 | 185.8 | 827.00 | -63.62 | -97.51 | -96.04 | -82.37 |

To detect the optimal $k$ value, the elbow method is used, and the number of clusters is determined according to the sum of squared distances (SSD). For example, in Figure 5.10 sum of squared distances for different cluster numbers is given. After 4 clusters, the change in SSD becomes insignificant. Therefore, $k$ is taken as 4. The same method is used for the selection of cluster number in every transition period in this thesis.

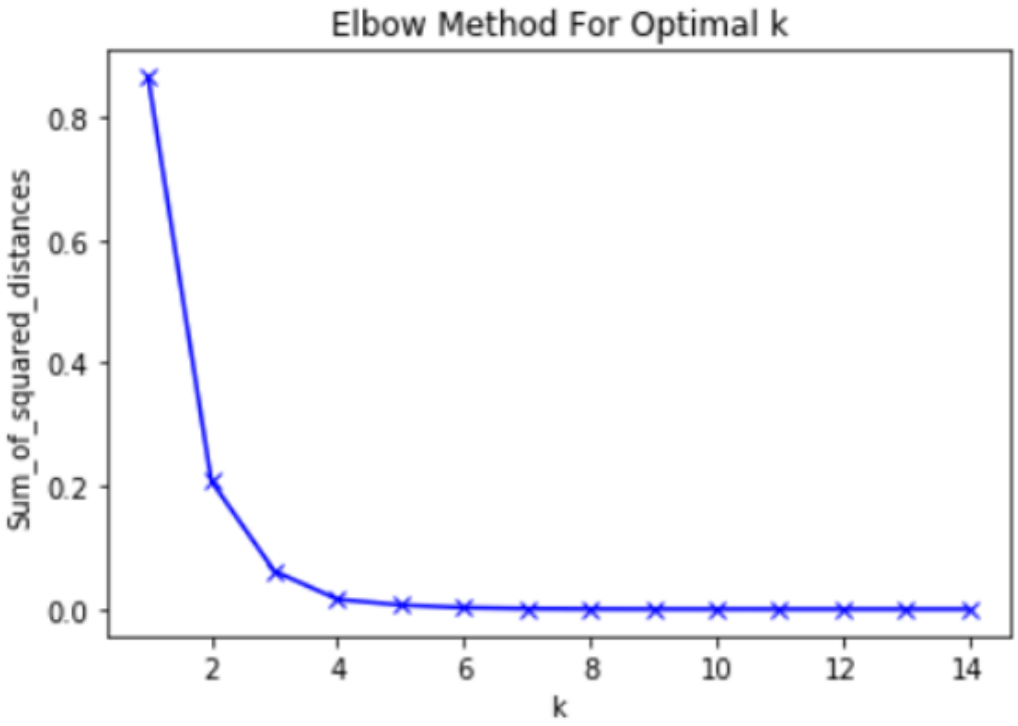


Figure 5.10. The elbow method for Transition 1.

As a result of the clustering, four different subgroups are obtained, and two of the subgroups only contain a single station. As observed below in Figure 5.11, Levent and 4.Levent stations show unique characteristics. In Table 5.2, 4.Levent is the only station where taxi trip counts increase rather than decrease in Transition 1 with 40%, and Levent has the lowest percentage of decrease compared to other stations with 21%. Therefore, they are creating their clusters by separating from the rest of the stations. The stations with the most reduction in taxi trip numbers are Hacıosman, Darüşşafaka, Atatürk Oto Sanayi, Sanayi, and Gayrettepe.

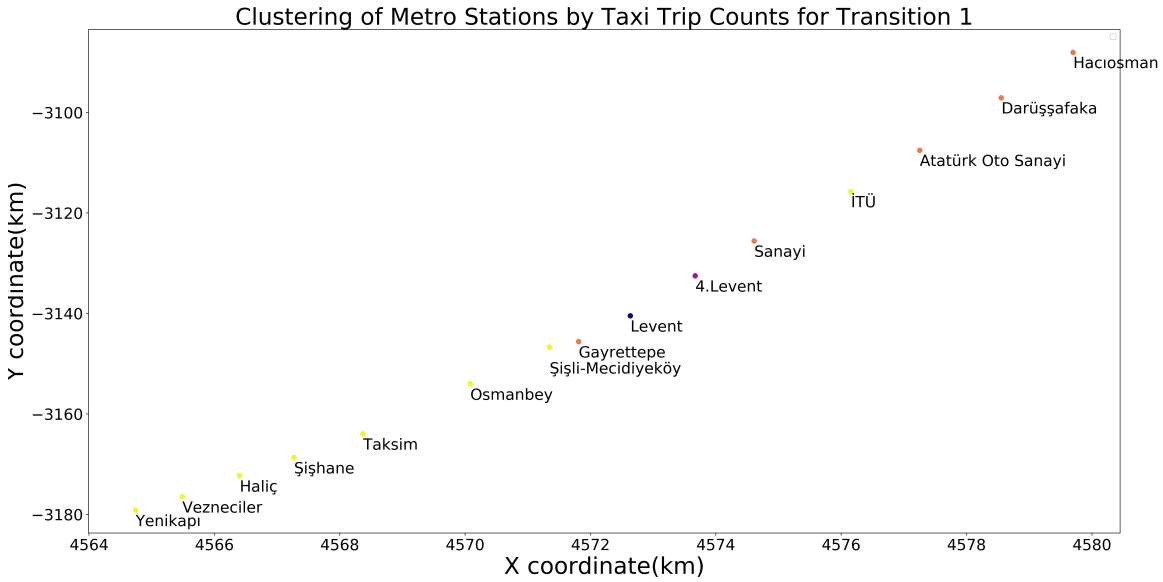


Figure 5.11. Clustering of stations for Transition 1 according to total trip counts.

As a result of the elbow method for Transition 2, the required number of clusters is detected as 4. The most significant drop characteristic in the taxi trip number is observed in Sanayi station (Figure 5.12). Therefore, it created a separate cluster by differentiating from the rest of the stations. Hacıosman and Darüşşafaka stations created a separate cluster that has the least reduction in trip counts. Yet, as shown in Table 5.2, the decrease in trip numbers is around 98% for most of the stations. Besides, the range in maximum and minimum values is less than 8%, meaning that the difference among clusters is not as significant as Transition 1.

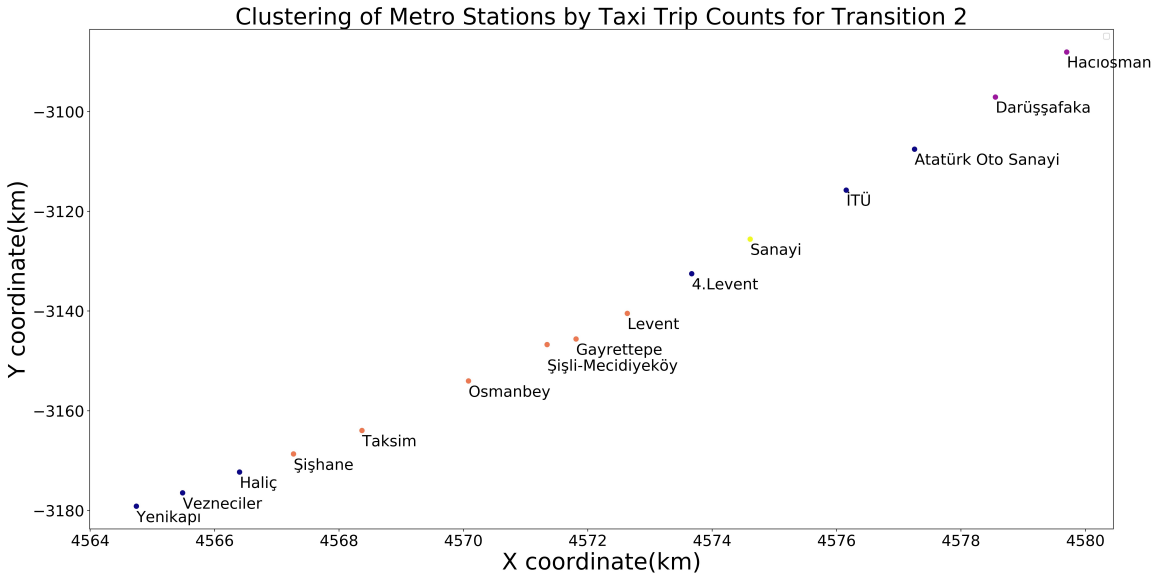


Figure 5.12. Clustering of stations for Transition 2 according to total trip counts.

The elbow method indicates 3 clusters for Transition 3. Hacıosman, Darüşşafaka, and Sanayi stations create a cluster with the lowest decrease in taxi trip counts. Then, the rest of the stations are divided into two subgroups (Figure 5.13).

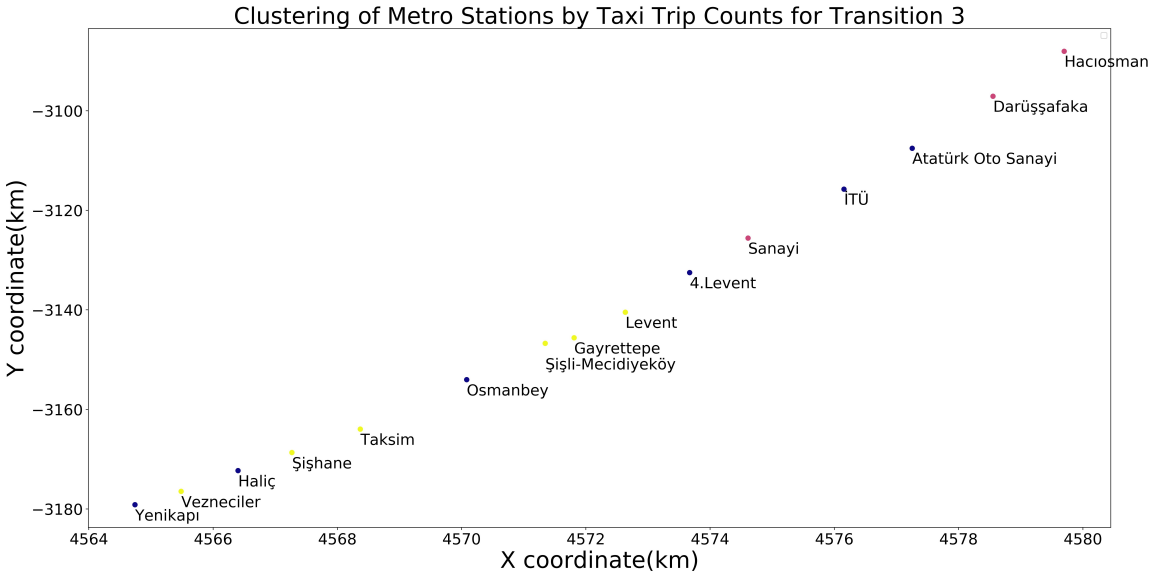


Figure 5.13. Clustering of stations for Transition 3 according to total trip counts.

In Transition 4, there are 4 clusters according to the elbow method. The most significant cluster is created by Hacıosman station that has the lowest decrease in trip

counts or the highest recovery compared to other stations. Then, the other cluster consists of Darüşşafaka, Sanayi, and Yenikapı stations with a high recovery percentage as in Figure 5.14.

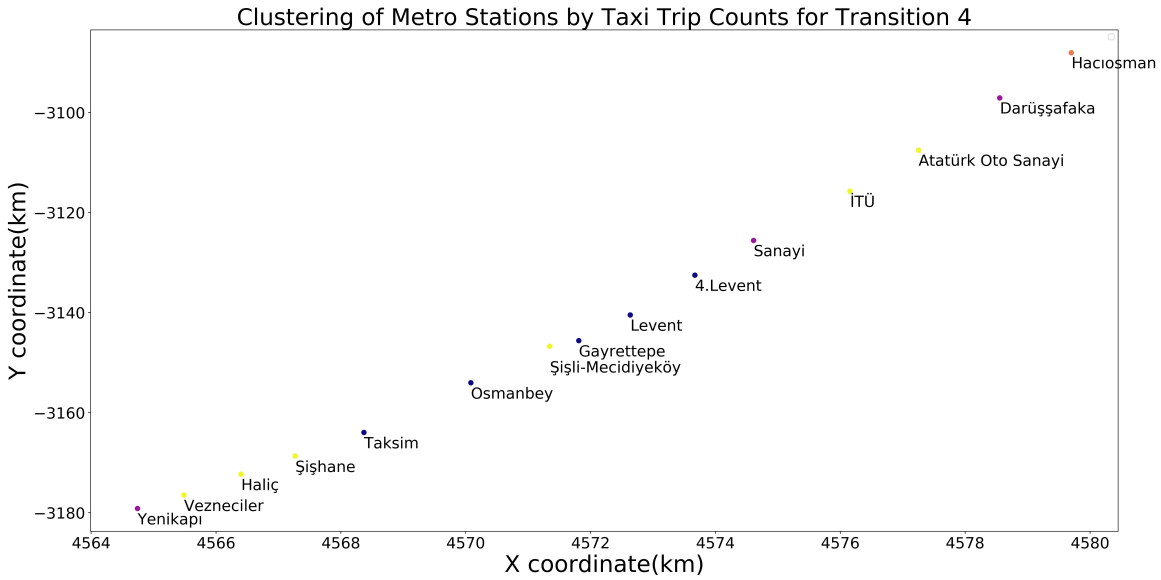


Figure 5.14. Clustering of stations for Transition 4 according to total trip counts.

A summary of the clustering results is given in Table 5.3. The stations that have different characteristics in taxi trip count decrease are indicated in red. There are 4, 4, 3, and 4 clusters for each transition period, respectively, as explained.

Table 5.3. The clusters of each station in every transition period for total trip counts.

| | **K-means for transition periods** | | | |
|---|---|---|---|---|
| **Station Name** | **Transition 1** | **Transition 2** | **Transition 3** | **Transition 4** |
| Hacıosman | 3 | 3 | 1 | **2** |
| Darüşşafaka | 3 | 3 | 1 | 1 |
| Atatürk Oto Sanayi | 3 | 1 | 2 | 3 |
| İTÜ | 1 | 1 | 2 | 3 |
| Sanayi | 3 | **0** | 1 | 1 |
| 4.Levent | **0** | 1 | 2 | 0 |
| Levent | **2** | 2 | 0 | 0 |
| Gayrettepe | 3 | 2 | 0 | 0 |
| Şişli-Mecidiyeköy | 1 | 2 | 0 | 3 |
| Osmanbey | 1 | 2 | 2 | 0 |
| Taksim | 1 | 2 | 0 | 0 |
| Şişhane | 1 | 2 | 0 | 3 |
| Haliç | 1 | 1 | 2 | 3 |
| Vezneciler | 1 | 1 | 0 | 3 |
| Yenikapı | 1 | 1 | 2 | 1 |

5.3.1.3. Regressions. First, the variables with similar patterns must be detected and eliminated so that this relationship wouldn't misdirect the results of the regression analysis. The correlation matrix is utilized for that purpose. Correlation analysis measures the linear association between variables [89] to prevent the possibility of multicollinearity, which might cause the overfitting of the model. Therefore, one variable that have higher Pearson correlation coefficients with each other must be eliminated from the regression to obtain more accurate models.

As a rule of thumb, the correlation between 0.9 to 1 is accepted as very high. The values between 0.7 and 0.9 indicate a high correlation between variables [90]. Therefore, in this thesis, any correlated variable above 0.7 is eliminated from the regression. The correlation matrix in Table 5.4 shows that there is a very high positive correlation (above 0.9) between SEGE, average education, and university graduation rate. Since the SEGE results from many elements in a neighborhood and reflects the socioeconomic

level by considering various factors, including education, the very high correlation between those variables is reasonable. The SEGE value is used for regression due to its multifaceted structure that covers many aspects, whereas average education and university graduation are kept out of the analysis. Yet, it must be considered that they are highly correlated with SEGE. Therefore, a significance that is applicable for SEGE will also be valid for those two variables. Another variable with a high correlation (above 0.7) with SEGE is the female percentage also removed from the analysis. It indicates that women prefer to live in neighborhoods with high socioeconomic and education levels. Besides, the female percentage is also found highly correlated with population and university graduation rates. The last significant correlation is between university graduation rate and average age. The neighborhoods with high university graduation rates have a lower average age, so they are negatively correlated.

Table 5.4. The correlation matrix for stations' characteristics.

| | **Population** | **Pop. Density (per 1000m2)** | **Avg. Age** | **Avg. Education (in years)** | **University Grad. (%)** | **Household Size** | **SEGE** | **Female Percentage** | **Public Transport Connection** | **Shopping Mall** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Population** | 1.000 | 0.588 | 0.075 | 0.343 | 0.410 | -0.187 | 0.536 | **0.710** | -0.139 | 0.226 |
| **Pop. Density (per 1000m2)** | 0.588 | 1.000 | 0.468 | 0.199 | 0.271 | -0.292 | 0.335 | 0.396 | 0.210 | 0.512 |
| **Avg. Age** | 0.075 | 0.468 | 1.000 | 0.372 | 0.432 | **-0.793** | 0.485 | 0.470 | 0.075 | 0.596 |
| **Avg. Education (in years)** | 0.343 | 0.199 | 0.372 | 1.000 | **0.982** | -0.155 | **0.914** | 0.736 | 0.088 | 0.367 |
| **University Grad. (%)** | 0.410 | 0.271 | 0.432 | 0.982 | 1.000 | -0.271 | **0.937** | **0.783** | 0.123 | 0.422 |
| **Household Size** | -0.187 | -0.292 | -0.793 | -0.155 | -0.271 | 1.000 | -0.367 | -0.523 | 0.118 | -0.241 |
| **SEGE** | 0.536 | 0.335 | 0.485 | 0.914 | 0.937 | -0.367 | 1.000 | **0.834** | -0.018 | 0.419 |
| **Female Percentage** | 0.710 | 0.396 | 0.470 | 0.736 | 0.783 | -0.523 | 0.834 | 1.000 | -0.183 | 0.280 |
| **Public Transport Connection** | -0.139 | 0.210 | 0.075 | 0.088 | 0.123 | 0.118 | -0.018 | -0.183 | 1.000 | 0.525 |
| **Shopping Mall** | 0.226 | 0.512 | 0.596 | 0.367 | 0.422 | -0.241 | 0.419 | 0.280 | 0.525 | 1.000 |

After the variables with high correlation are eliminated, the remaining variables for the analysis are population, population density, average age, SEGE, public transport connection, and the number of shopping malls in the area. According to the selected parameters, the following model is obtained:

$$\begin{aligned} y = {} & \beta_0 + \beta_1\,(\text{population}) + \beta_2\,(\text{pop.density}) \\ & + \beta_3\,(\text{avg.age}) + \beta_4\,(\text{SEGE}) + \beta_5\,(\text{public transport connection}) \\ & + \beta_6\,(\text{shopping malls}) \end{aligned} \tag{5.19}$$

First, to obtain a first insight into the data, the clustering is made. However, because of the elbow method, an apparent breakdown cannot be detected (Figure 5.15). Therefore, cluster number is taken as four by accepting an SSD around 25%.

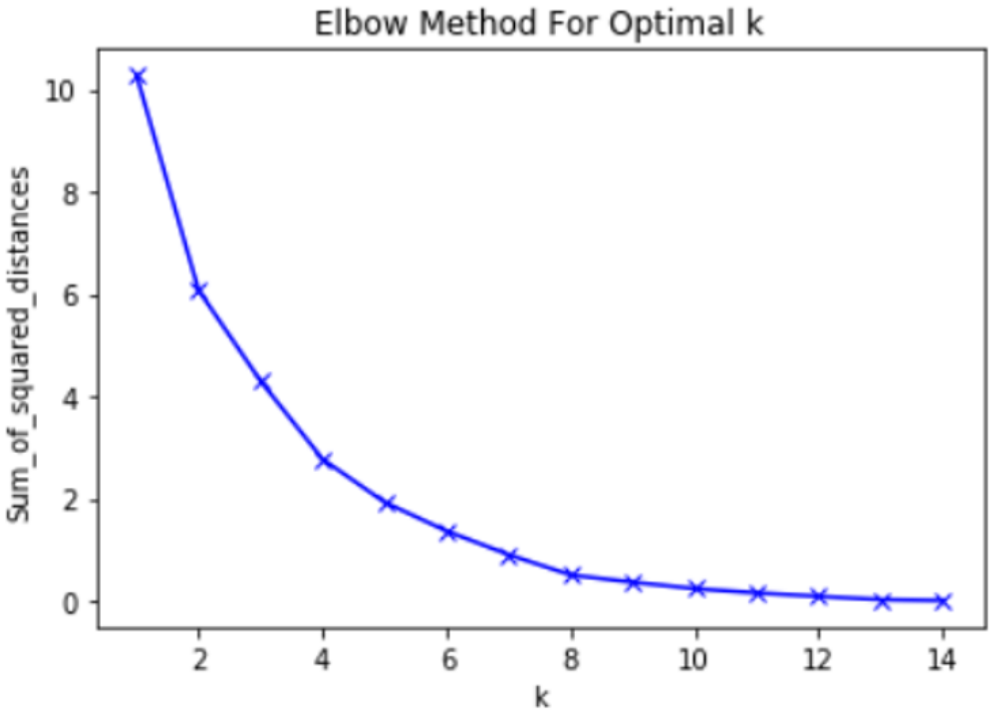


Figure 5.15. The elbow method for stations' characteristics.

As a result of the clustering, clusters in Figure 5.16 is obtained. The most remarkable characteristic of the clustering is that the clusters are separated based on their location. The ones that are closer to each other shows similar characteristics such as

Yenikapı, Vezneciler, Haliç, Şişhane, and Taksim cluster. Therefore, using spatial autocorrelation as a measure would be wise for the analysis like SAR. Yet, some stations deviate from this inference like Osmanbey. So, considering spatial non-stationarity would be reasonable to explain the relationships which cannot be explained only by a global linear regression that can be done by using the GWR.

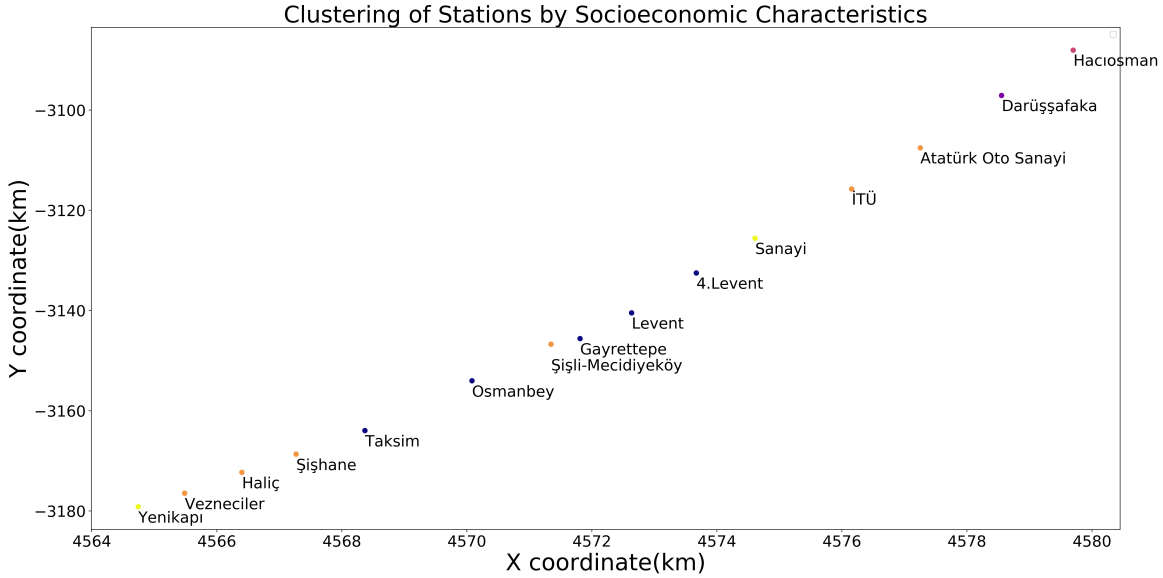


Figure 5.16. Clustering of stations based on socioeconomic characteristics.

$R^2$ indicates the goodness of fit; in other words, how many percent of the dependent variable can be explained by the model. So, the higher the R2 is, the better the model explains the relationship. As the best-performed model, GWR explains 76.8% of the data (Table 5.6). Yet, just looking at the R2 is not enough to select the best model, because an increase in the number of independent variables would cause higher R2 values that don't reflect the real improvement. Therefore, AIC, which is commonly used to select the best model with the chosen variables and log-likelihood values, must also be considered. As a general rule, a lower AIC and higher log-likelihood values also indicate a better-performed model. But if the improvement obtained with a more complicated model is not highly significant, choosing the simple model is always better. Therefore, preferring the OLS model could also be satisfying to obtain the necessary results.

Typically, 95% confidence interval is accepted for a significant relationship in the OLS model, however, if there are many independent variables compared to the data, then taking the confidence interval slightly lower might be accepted. Yet, the acceptable confidence interval in this thesis is taken as 95%. The value of Moran's I in the SAR model indicates the spatial distribution of the variables, as indicated in Table 5.5. The rho value is the spatial autoregressive coefficient of the model and expresses the influence of neighboring observations with a significance of p-value. In contrast to the OLS and SAR models, GWR doesn't only consider a single set of constant values; it estimates statistics at each point and varies over space. In other words, the SAR model reveals the relationship between a station and the stations around it without considering other stations, whereas, GWR considers all of the stations and adapts the model for each station to obtain a better representation.

Table 5.5. Moran's I index

| **Moran's I** | **Spatial Distribution** |
|---|---|
| Score $> 0$ | Clustered |
| Score $= 0$ | Random |
| Score $< 0$ | Dispersed |

The analysis results for OLS, SAR, and GWR methods for the Transition 1 period are presented in Table 5.6. The SAR model performs better than the OLS model. For Transition 1 of total trip counts, shopping mall number and population are the two significant variables with a positive correlation for the SAR model. The GWR model resulted slightly better than OLS and SAR models, and the coefficients of GWR for the population (0.721) and shopping mall (0.499) indicate the same relationship with SAR. In the GWR, SEGE and public transport connection also plays an essential role in the model.

Table 5.6. The analysis results of the stations' characteristics for Transition 1 in total trip counts.

| Primary Catchment Area (0-800m) of the subway for Transition 1 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | OLS | | SAR | | GWR | | |
| Variables | Coefficient | P>t value | Coefficient | P>t value | Min | Mean | Max |
| Constant | 0.000 | 0.000 | 0.315 | 0.013 | -0.332 | 0.018 | 0.317 |
| Population | 0.125 | 0.782 | 0.579 | **0.013** | 0.415 | **0.721** | 2.314 |
| Pop. Density (per 1000m2) | 0.387 | 0.355 | 0.162 | 0.436 | -1.262 | -0.065 | 0.347 |
| Avg. Age | -0.337 | 0.430 | -0.325 | 0.112 | -0.800 | 0.187 | 1.882 |
| SEGE | -0.262 | 0.483 | -0.037 | 0.833 | -2.135 | -0.527 | -0.176 |
| Public Transport Connection | -0.542 | 0.120 | -0.323 | 0.059 | -0.854 | -0.580 | -0.291 |
| Shopping Mall | 0.749 | 0.074 | 0.668 | **0.001** | 0.148 | **0.499** | 0.786 |
| $R^2$ | 0.514 | | 0.688 | | 0.768 | | |
| AIC | 45.756 | | 43.114 | | 42.523 | | |
| Log-likelihood | -15.878 | | -12.557 | | -10.316 | | |
| | | | Rho: -1.826 | p-val: 0.010 | Bandwidth = 14.000 | | |
| | | | Morans' I: -0.073 | | Adj. Alpha (95%) = 0.035 | | |

The analysis results for OLS, SAR, and GWR methods for the Transition 2 period are presented in Table 5.7. The OLS model performs better than the SAR model. Here, the population has a highly significant positive correlation with the taxi trip change meaning that the decrease in taxi trip counts is smaller for more populated areas. The population density and shopping mall number are also significant for Transition 2 with a negative correlation. So, the decrease in taxi trip counts is more for stations with more shopping malls and population density. It can be concluded that people were using taxis to arrive at a shopping mall, but with the closure of shops, the taxi usage decreased significantly. The GWR model gives better results, as with Transition 1. The R2 and log-likelihood values are higher. Even though AIC is not significantly better than the OLS, the main reason is the high number of variables. Therefore, the model can be considered better. The coefficients of GWR for the population (0.981), population density (-0.607), and shopping mall (-0.765) indicates the same relationship with OLS. Besides, the average age has a robust positive coefficient in the GWR model.

Table 5.7. The analysis results of the stations' characteristics for Transition 2 in total trip counts.

| Primary Catchment Area (0-800m) of the subway for Transition 2 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | OLS | | SAR | | GWR | | |
| Variables | Coefficient | P>t value | Coefficient | P>t value | Min | Mean | Max |
| Constant | 0.000 | 0.000 | -0.067 | 0.656 | -0.371 | -0.269 | -0.198 |
| Population | 0.974 | **0.007** | 1.118 | 0.000 | 0.380 | **0.981** | 1.530 |
| Pop. Density (per 1000m2) | -0.756 | 0.024 | -0.961 | 0.001 | -1.002 | **-0.607** | -0.157 |
| Avg. Age | 0.489 | 0.152 | 0.527 | 0.031 | -0.057 | **0.690** | 1.341 |
| SEGE | -0.366 | 0.22 | -0.298 | 0.175 | -0.408 | -0.245 | -0.094 |
| Public Transport Connection | 0.217 | 0.436 | 0.189 | 0.339 | -0.015 | 0.229 | 0.465 |
| Shopping Mall | -0.701 | **0.036** | -0.768 | 0.003 | -1.335 | **-0.765** | -0.107 |
| $R^2$ | 0.690 | | 0.701 | | 0.824 | | |
| AIC | 38.988 | | 42.459 | | 38.386 | | |
| Log-likelihood | -12.494 | | -12.229 | | -8.247 | | |
| | | | Rho: -0.416 | p-val: 0.467 | Bandwidth = 14.000 | | |
| | | | Morans' I: 0.294 | | Adj. Alpha (95%) = 0.035 | | |

The analysis results for OLS, SAR, and GWR methods for the Transition 3 period are presented in Table 5.8. The OLS model performs better than the SAR model. The population, population density, and shopping mall numbers follow the same pattern as in Transition 2, and the GWR model is superior to others. The coefficients of GWR for the population (1.176), population density (-0.629), and the number of shopping mall (-0.890) indicates the same relationship with OLS. Besides, the average age has a robust positive coefficient in the GWR model.

Table 5.8. The analysis results of the stations' characteristics for Transition 3 in total trip counts.

| Primary Catchment Area (0-800m) of the subway for Transition 3 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | OLS | | SAR | | GWR | | |
| Variables | Coefficient | P>t value | Coefficient | P>t value | Min | Mean | Max |
| Constant | 0.000 | 0.000 | -0.048 | 0.742 | -0.382 | -0.297 | -0.180 |
| Population | 1.017 | **0.004** | 1.192 | 0 | 0.503 | 1.176 | **1.765** |
| Pop. Density (per 1000m2) | -0.663 | **0.044** | -0.893 | 0.004 | -0.978 | **-0.629** | -0.194 |
| Avg. Age | 0.439 | 0.192 | 0.481 | 0.048 | -0.028 | **0.785** | 1.439 |
| SEGE | -0.361 | 0.219 | -0.292 | 0.186 | -0.417 | -0.218 | -0.120 |
| Public Transport Connection | 0.186 | 0.498 | 0.165 | 0.398 | -0.067 | 0.222 | 0.504 |
| Shopping Mall | -0.693 | **0.036** | -0.756 | 0.003 | -1.481 | **-0.890** | -0.123 |
| $R^2$ | 0.699 | | 0.709 | | 0.835 | | |
| AIC | 38.559 | | 42.063 | | 37.457 | | |
| Log-likelihood | -12.280 | | -12.031 | | -7.783 | | |
| | | | Rho: -0.398 | p-val: 0.481 | Bandwidth = 14.000 | | |
| | | | Morans' I: 0.326 | | Adj. Alpha (95%) = 0.035 | | |

The analysis results for OLS, SAR, and GWR methods for the Transition 4 period are presented in Table 5.9. The OLS model performs better than the SAR model. On top of population and population density, the average age and SEGE are also significant in Transition 4. The decrease in taxi trip counts is smaller for the stations with a higher average age. The probable reason for that could be the end of curfew for 65+. The analysis also shows that the recovery in taxi usage is less in places with a higher socioeconomic level (SEGE). Here, the GWR model also resulted better than the others, and it explains 92.6% of the data. The coefficients of GWR for the population (0.825), population density (-0.615), average age (0.633), and SEGE (-0.627) indicate the same relationship with OLS.

Table 5.9. The analysis results of the stations' characteristics for Transition 4 in total trip counts.

| Primary Catchment Area (0-800m) of the subway for Transition 3 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | OLS | | SAR | | GWR | | |
| Variables | Coefficient | P>t value | Coefficient | P>t value | Min | Mean | Max |
| Constant | 0.000 | 0.000 | -0.110 | 0.457 | -0.400 | -0.225 | 0.011 |
| Population | 1.091 | **0.001** | 1.118 | 0.000 | 0.249 | **0.825** | 1.452 |
| Pop. Density (per 1000m2) | -0.936 | **0.002** | -1.073 | 0 | -1.172 | **-0.615** | -0.108 |
| Avg. Age | 0.657 | **0.035** | 0.667 | 0.002 | -0.059 | **0.633** | 1.363 |
| SEGE | -0.802 | **0.003** | -0.734 | 0.000 | -1.055 | **-0.627** | -0.174 |
| Public Transport Connection | 0.328 | 0.195 | 0.285 | 0.110 | 0.103 | 0.277 | 0.448 |
| Shopping Mall | -0.541 | 0.076 | -0.595 | 0.007 | -0.899 | -0.449 | 0.093 |
| R2 | 0.743 | | 0.758 | | 0.926 | | |
| AIC | 36.169 | | 39.300 | | 25.388 | | |
| Log-likelihood | -11.085 | | -10.650 | | -1.748 | | |
| | | | Rho: -0.495 | p-val: 0.351 | Bandwidth = 14.000 | | |
| | | | Morans' I: 0.108 | | Adj. Alpha (95%) = 0.035 | | |

### 5.3.2. Morning Peak Trip Counts

The morning peak covers the hours between 6 to 9 am, and it is investigated separately in this section of the analysis.

5.3.2.1. Trip Count Plots. Below in Figure 5.17, the morning peak trip numbers in every station are given for January and February. After the thermal camera installation in airports on January 24, there is a decrease in all of the stations. There is a substantial increase on Saturdays on taxi trips, followed by a slight decrease. The main reason could be the decrease in public transportation frequency on Saturdays. There is a substantial increase in morning peak taxi trips in 20th February for Vezneciler, Yenikapı and Şişli-Mecidiyeköy stations all have another connection with public transport.

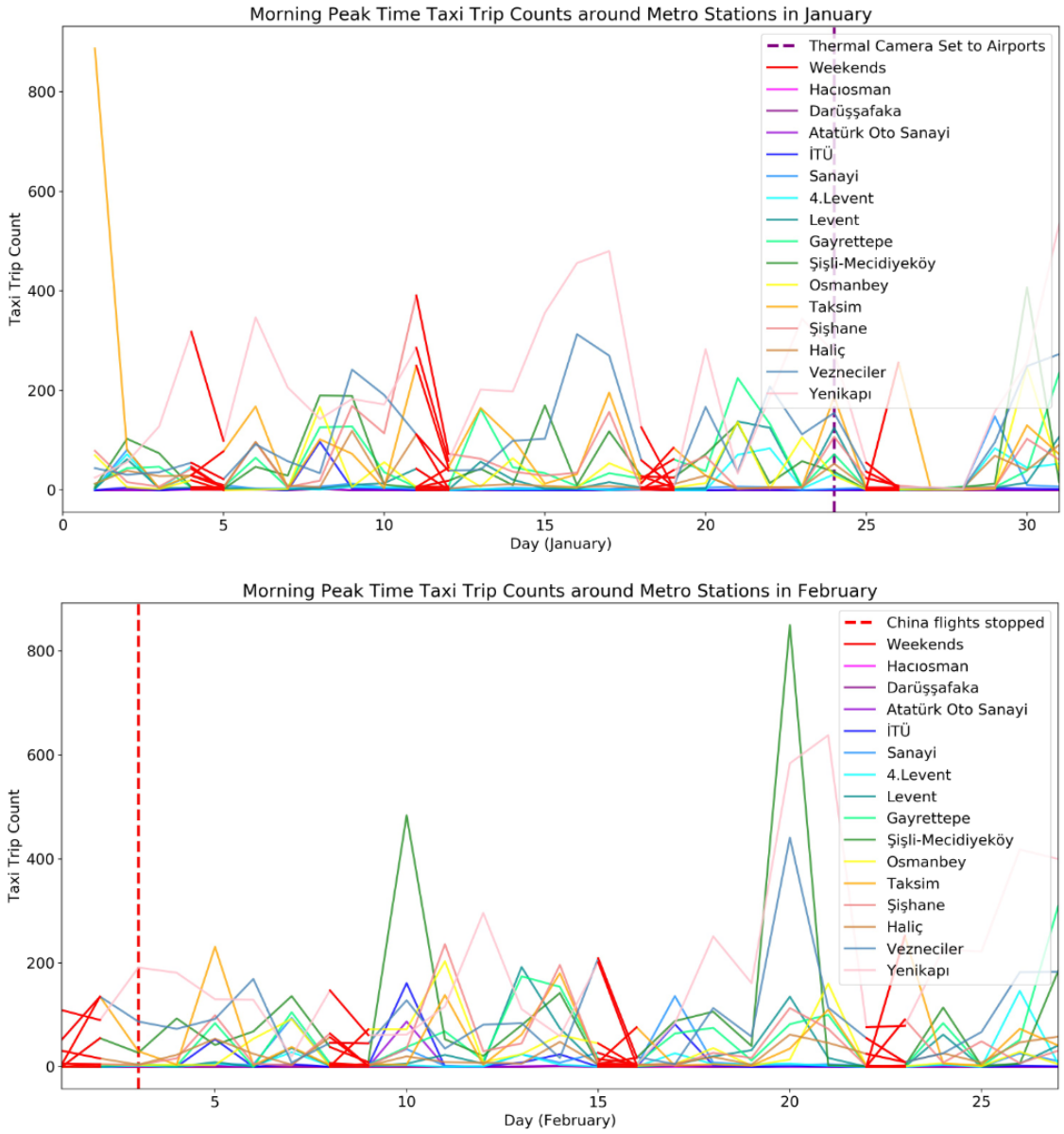


Figure 5.17. Morning peak of January and February counts.

Below in Figure 5.18, the morning peak trip numbers in every station are given for March and April. After the 19th of March, the trip numbers decrease below 30 trips per day, and the effects of COVID-19 can be observed clearly, and the low trip numbers continue in April as well.

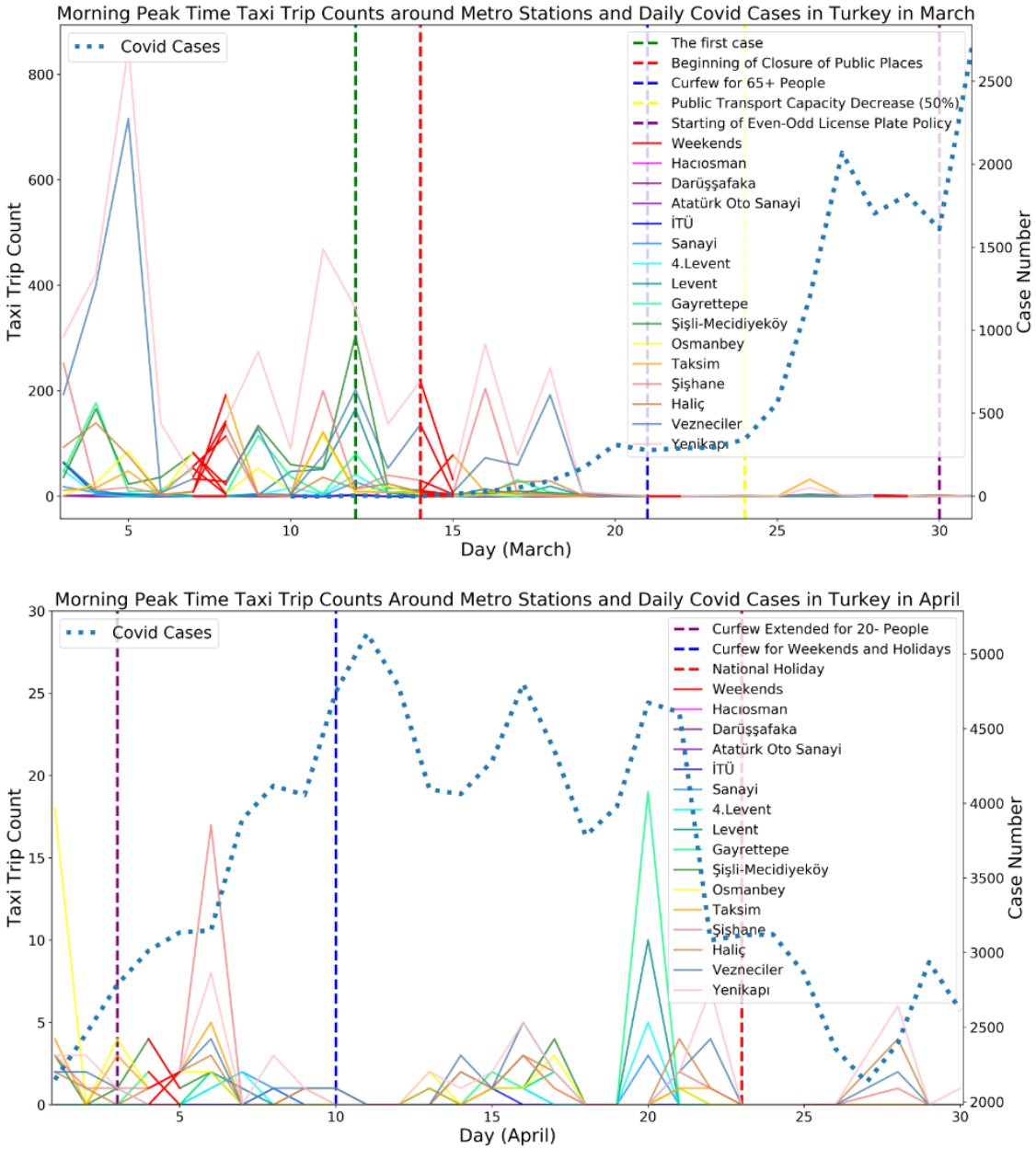


Figure 5.18. Morning peak of March and April counts.

Below in Figure 5.19, the morning peak trip numbers in every station are given for May and June. Most probably, many working places continued to work from home even after normalization is announced officially on May 6. Then, the numbers increase slightly for some stations in the second half of May.

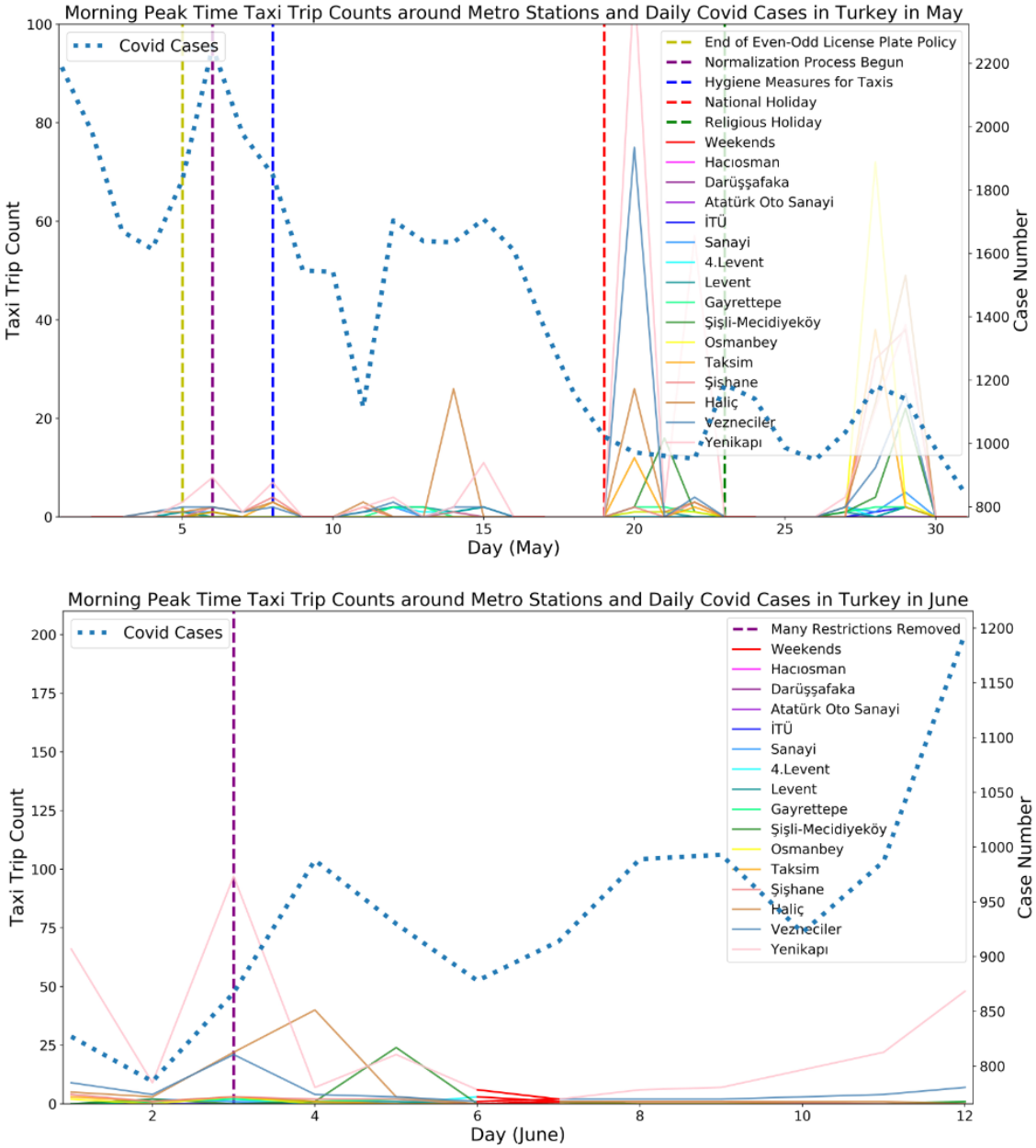


Figure 5.19. Morning peak of May and June counts.

5.3.2.2. Clustering. The results for each station and transition period for morning peaks are given in Table 5.10. The decrease in Transition 1 is lower for the stations with higher initial trip counts, and there is a significant difference between stations in trip changes for Transition 1.

Table 5.10. The decrease in transitions for morning peak trip counts.

| **Station Name** | **Pre-COVID** | **Reaction to Low Case Numbers** | **Reaction with Strict Measures** | **Normalization** | **Transition to Recovery** | **Transition1** | **Transition2** | **Transition3** | **Transition4** |
|---|---|---|---|---|---|---|---|---|---|
| Hacıosman | 0.12 | 0.00 | 0.00 | 0.00 | 0.00 | -100.00 | -100.00 | -100.00 | -100.00 |
| Darüşşafaka | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | -100.00 | -100.00 | -100.00 | -100.00 |
| Atatürk Oto Sanayi | 1.85 | 0.08 | 0.00 | 0.00 | 0.00 | -95.49 | -100.00 | -100.00 | -100.00 |
| İTÜ | 8.55 | 0.17 | 0.11 | 0.16 | 0.10 | -98.05 | -98.76 | -98.13 | -98.83 |
| Sanayi | 12.55 | 1.42 | 0.47 | 0.56 | 0.40 | -88.71 | -96.26 | -95.54 | -96.81 |
| 4.Levent | 13.42 | 1.50 | 0.39 | 0.44 | 0.70 | -88.83 | -97.07 | -96.72 | -94.79 |
| Levent | 24.98 | 3.75 | 0.55 | 0.52 | 0.70 | -84.99 | -97.82 | -97.92 | -97.20 |
| Gayrettepe | 53.47 | 5.33 | 0.70 | 0.56 | 0.80 | -90.03 | -98.7 | -98.95 | -98.50 |
| Şişli-Mecidiyeköy | 77.55 | 4.50 | 1.05 | 1.84 | 2.90 | -94.20 | -98.65 | -97.63 | -96.26 |
| Osmanbey | 35.74 | 2.92 | 1.79 | 3.12 | 0.30 | -91.84 | -95.00 | -91.27 | -99.16 |
| Taksim | 60.76 | 9.08 | 1.79 | 2.36 | 0.50 | -85.05 | -97.06 | -96.12 | -99.18 |
| Şişhane | 53.55 | 24.33 | 1.98 | 3.32 | 1.10 | -54.56 | -96.29 | -93.8 | -97.95 |
| Haliç | 29.67 | 8.50 | 2.68 | 5.48 | 7.20 | **-71.35** | **-90.96** | **-81.53** | **-75.73** |
| Vezneciler | 110.15 | 44.33 | 2.59 | 5.08 | 5.00 | -59.75 | -97.65 | -95.39 | -95.46 |
| Yenikapı | 208.06 | 84.42 | 5.15 | 10.44 | 22.50 | -59.43 | -97.52 | -94.98 | **-89.19** |

The clustering of each station is given in Table 5.11. The stations mostly show similar characteristics location-wise. Most of the clusters consist of the stations that are close to each other. Yet, Haliç station creates a different cluster by itself in every transition period in terms of the decrease of taxi trips (Table 5.10). The station has 71.35%, 90.96%, 81.53%, and 75.73% decrease in taxi trips for each transition period, respectively. Besides, Yenikapı station is also creating a different cluster in Transition 4 that indicates a better recovery with an 89.19% decrease in trip numbers for the morning peak.

Table 5.11. The clusters of each station in every transition period for morning peak.

| | **K-means for transition periods** | | | |
|---|---|---|---|---|
| **Station Name** | **Transition 1** | **Transition 2** | **Transition 3** | **Transition 4** |
| Hacıosman | 1 | 1 | 0 | 0 |
| Darüşşafaka | 1 | 1 | 0 | 0 |
| Atatürk Oto Sanayi | 1 | 1 | 0 | 0 |
| İTÜ | 1 | 1 | 0 | 0 |
| Sanayi | 2 | 0 | 2 | 1 |
| 4.Levent | 2 | 0 | 2 | 1 |
| Levent | 2 | 0 | 0 | 1 |
| Gayrettepe | 2 | 1 | 0 | 0 |
| Şişli-Mecidiyeköy | 1 | 1 | 0 | 1 |
| Osmanbey | 2 | 0 | 2 | 0 |
| Taksim | 2 | 0 | 2 | 0 |
| Şişhane | 0 | 0 | 2 | 0 |
| Haliç | **3** | **2** | **1** | **2** |
| Vezneciler | 0 | 0 | 2 | 1 |
| Yenikapı | 0 | 0 | 2 | **3** |

5.3.2.3. Regressions. The analysis results for OLS, SAR, and GWR methods for Transition 1 in the morning peak are presented in Table 5.12. The OLS model performs better than the SAR model. The population, SEGE, and shopping mall are significant variables for the OLS model. The population and SEGE have a negative correlation, whereas the shopping mall is positively correlated with the decrease in taxi trip counts. So, around stations with higher socioeconomic level and more population, the decrease in taxi trips are more, whereas around stations with more shopping mall the decrease in trips is smaller. In Transition 1, all three of the models perform successfully by explaining more than 85% of the data. The R2 and log-likelihood values are higher for

the GWR, and the AIC value is lower. Therefore, The GWR model results better than other models. The coefficients of GWR for the population (-0.258), SEGE (-0.559), and shopping mall (-.401) indicates the same relationship with OLS.

Table 5.12. The analysis results of the stations' characteristics for Transition 1 in morning peak trip counts.

| Primary Catchment Area (0-800m) of the subway for Transition 1 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | OLS | | SAR | | GWR | | |
| Variables | Coefficient | P>t value | Coefficient | P>t value | Min | Mean | Max |
| Constant | 0.000 | 0.000 | 0.008 | 0.937 | -0.375 | -0.108 | 0.179 |
| Population | -0.632 | **0.010** | -0.529 | 0.055 | -0.501 | **-0.258** | 0.582 |
| Pop. Density (per 1000m2) | 0.254 | 0.267 | 0.195 | 0.364 | -0.682 | 0.025 | 0.240 |
| Avg. Age | -0.364 | 0.120 | -0.342 | 0.049 | -0.194 | -0.035 | 0.795 |
| SEGE | -0.434 | **0.034** | -0.394 | 0.017 | -1.396 | **-0.559** | -0.070 |
| Public Transport Connection | 0.087 | 0.647 | 0.091 | 0.510 | 0.063 | 0.120 | 0.195 |
| Shopping Mall | 0.570 | **0.013** | 0.538 | 0.003 | 0.186 | **0.401** | 0.600 |
| $R^2$ | 0.854 | | 0.856 | | 0.935 | | |
| AIC | 27.678 | | 31.510 | | 23.532 | | |
| Log-likelihood | -6.839 | | -6.755 | | -0.820 | | |
| | | | Rho: 0.177 | p-val: 0.681 | Bandwidth = 14.000 | | |
| | | | Morans' I: 0.524 | | Adj. Alpha (95%) = 0.035 | | |

The analysis results for OLS, SAR, and GWR methods for Transition 2 in the morning peak are presented in Table 5.13. The OLS model performs better than the SAR model. In Transition 2, there isn't any variable that improves the model significantly for OLS. Here, the GWR performs significantly better than the other models with an explanation of 90.4%. In the GWR model, the coefficient of SEGE (-0.505) indicates a negative correlation with the decrease in trip numbers. So, for stations with a higher socioeconomic level, the decrease in taxi trips is more.

Table 5.13. The analysis results of the stations' characteristics for Transition 2 in morning peak trip counts.

| Primary Catchment Area (0-800m) of the subway for Transition 2 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | OLS | | SAR | | GWR | | |
| Variables | Coefficient | P>t value | Coefficient | P>t value | Min | Mean | Max |
| Constant | 0.000 | 0.000 | 0.047 | 0.771 | -0.156 | 0.209 | 0.655 |
| Population | -0.540 | 0.194 | -0.740 | 0.041 | -0.475 | 0.037 | 0.748 |
| Pop. Density (per 1000m2) | 0.639 | 0.098 | 0.727 | 0.012 | -0.527 | 0.048 | 0.663 |
| Avg. Age | 0.043 | 0.914 | -0.025 | 0.928 | -0.240 | 0.384 | 1.271 |
| SEGE | -0.496 | 0.149 | -0.552 | 0.024 | -1.157 | **-0.505** | -0.068 |
| Public Transport Connection | -0.417 | 0.194 | -0.351 | 0.128 | -0.637 | -0.250 | -0.016 |
| Shopping Mall | 0.332 | 0.389 | 0.418 | 0.134 | -0.414 | -0.105 | 0.174 |
| $R^2$ | 0.587 | | 0.605 | | 0.904 | | |
| AIC | 43.286 | | 46.624 | | 29.329 | | |
| Log-likelihood | -14.643 | | -14.312 | | -3.719 | | |
| | | | Rho: -0.484 | p-val: 0.416 | Bandwidth = 14.000 | | |
| | | | Morans' I: 0.166 | | Adj. Alpha (95%) = 0.035 | | |

The analysis results for OLS, SAR, and GWR methods for Transition 3 in the morning peak are presented in Table 5.14. The OLS model performs better than the SAR model. As in Transition 2, there isn't any significant variable in Transition 3 as well. Yet, SEGE and public transport connections have a significance with a 90% confidence level with a negative correlation. As mentioned before, 90% of confidence can also be accepted due to the high number of variables. The results of the GWR are significantly better than other models with higher log-likelihood, lower AIC, and 92.6% explanation. In the GWR model, the coefficient of SEGE (-0.525) indicates a negative correlation with the decrease in trip numbers as in Transition 2.

Table 5.14. The analysis results of the stations' characteristics for Transition 3 in morning peak trip counts.

| Primary Catchment Area (0-800m) of the subway for Transition 3 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | OLS | | SAR | | GWR | | |
| Variables | Coefficient | P>t value | Coefficient | P>t value | Min | Mean | Max |
| Constant | 0.000 | 0.000 | 0.016 | 0.912 | -0.083 | 0.188 | 0.550 |
| Population | -0.618 | 0.103 | -0.845 | 0.014 | -0.677 | -0.214 | 0.197 |
| Pop. Density (per 1000m2) | 0.578 | 0.100 | 0.666 | 0.011 | -0.313 | 0.145 | 0.632 |
| Avg. Age | -0.030 | 0.933 | -0.088 | 0.729 | -0.359 | 0.162 | 0.651 |
| SEGE | -0.541 | 0.084 | -0.613 | 0.007 | -0.998 | **-0.525** | -0.127 |
| Public Transport Connection | -0.516 | 0.078 | -0.437 | 0.040 | -0.719 | -0.347 | -0.114 |
| Shopping Mall | 0.383 | 0.276 | 0.450 | 0.072 | -0.133 | 0.052 | 0.247 |
| $R^2$ | 0.657 | | 0.675 | | 0.926 | | |
| AIC | 40.502 | | 43.695 | | 25.324 | | |
| Log-likelihood | -13.251 | | -12.848 | | -1.716 | | |
| | | | Rho: -0.493 | p-val: 0.369 | Bandwidth = 14.000 | | |
| | | | Morans' I: 0.220 | | Adj. Alpha (95%) = 0.035 | | |

The analysis results for OLS, SAR, and GWR methods for Transition 4 in the morning peak are presented in Table 5.15. The OLS model performs better than the SAR model, and the only significant variable in the OLS model is the socioeconomic index (SEGE). It indicates a negative correlation with the decrease in taxi trips, similar to Transition 1. Here, GWR gives much better results than other models with a 93.4% explanation. The coefficient of GWR for SEGE (-0.805) indicates the same relationship with OLS.

Table 5.15. The analysis results of the stations' characteristics for Transition 3 in morning peak trip counts.

| Primary Catchment Area (0-800m) of the subway for Transition 3 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | OLS | | SAR | | GWR | | |
| Variables | Coefficient | P>t value | Coefficient | P>t value | Min | Mean | Max |
| Constant | 0.000 | 0.000 | -113.280 | 0.014 | -0.253 | 0.120 | 0.575 |
| Population | -0.098 | 0.811 | 0.000 | 0.471 | -0.156 | 0.396 | 1.017 |
| Pop. Density (per 1000m2) | 0.156 | 0.681 | 0.107 | 0.476 | -0.892 | -0.273 | 0.256 |
| Avg. Age | -0.169 | 0.662 | -0.428 | 0.454 | -0.364 | 0.105 | 0.555 |
| SEGE | -0.819 | **0.015** | -0.285 | 0.000 | -1.443 | **-0.805** | -0.132 |
| Public Transport Connection | -0.460 | 0.145 | -4.74 | 0.087 | -0.530 | -0.240 | -0.067 |
| Shopping Mall | 0.656 | 0.084 | 3.421 | 0.011 | 0.065 | 0.253 | 0.395 |
| $R^2$ | 0.602 | | 0.619 | | 0.934 | | |
| AIC | 42.752 | | 46.108 | | 23.739 | | |
| Log-likelihood | -14.376 | | -14.054 | | -0.923 | | |
| | | | Rho: -0.481 | p-val: 0.422 | Bandwidth = 14.000 | | |
| | | | Morans' I: 0.067 | | Adj. Alpha (95%) = 0.035 | | |

### 5.3.3. Evening Peak Trip Counts

The evening peak covers the hours between 5 to 8 pm. Mostly, it is more congested than the morning peak. Therefore, it has the potential to have a different pattern. The evening peak is investigated in this section.

5.3.3.1. Trip Count Plots. Below in Figure 5.20, the evening peak trip numbers in every station are given for January and February. After the thermal camera installation in airports on January 24, there is the same decrease in taxi trips that is observed in total counts and morning peaks. There is a substantial increase on Saturdays on taxi trips in February, followed by a slight decrease.

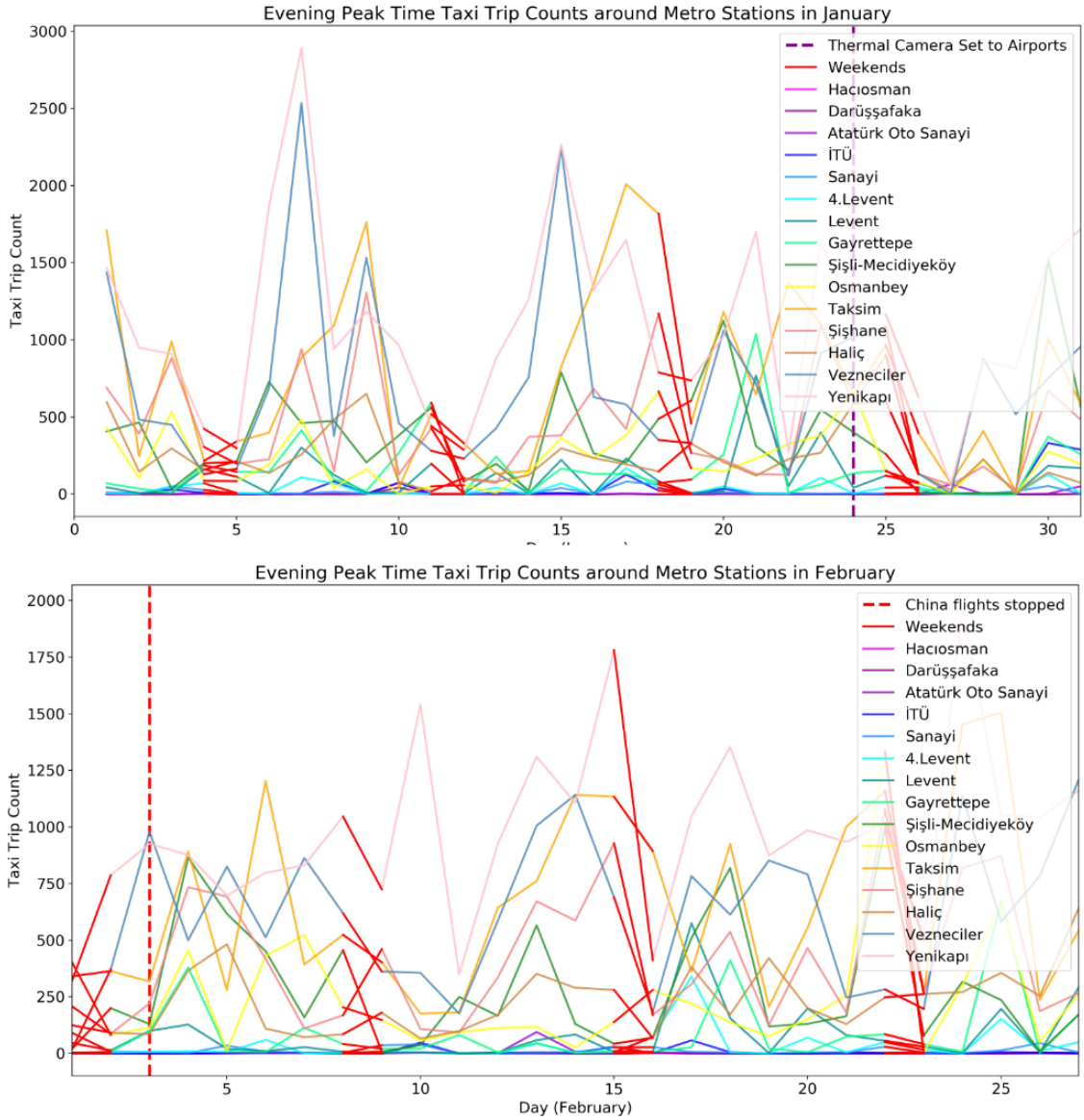


Figure 5.20. Evening peak January and February counts.

Below in Figure 5.21, the evening peak trip numbers in every station are given for March and April. The pattern for March and April counts is similar to total, and evening counts accept the peaks for Saturdays in March is sharper for evening peak counts. Besides, the trip counts are relatively low in April during weekdays.

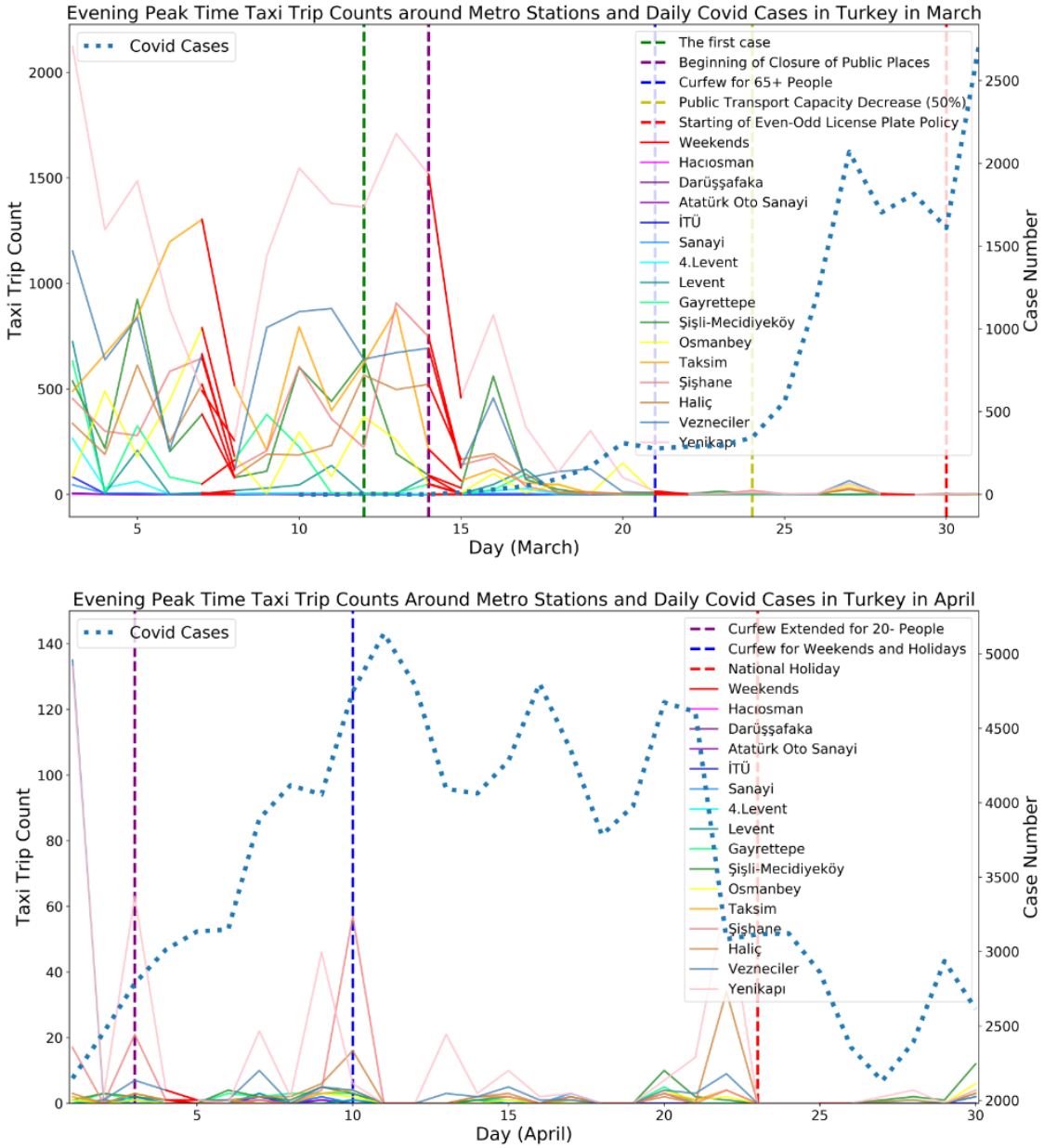


Figure 5.21. Evening peak March and April counts.

Below in Figure 5.22, the evening peak trip numbers in every station are given for May and June. There is a substantial recovery in weekday evening trips in the second half of May. Besides, there is an increase in weekend trips from Saturday to Sunday that is usually the other way around for total and morning peak counts.

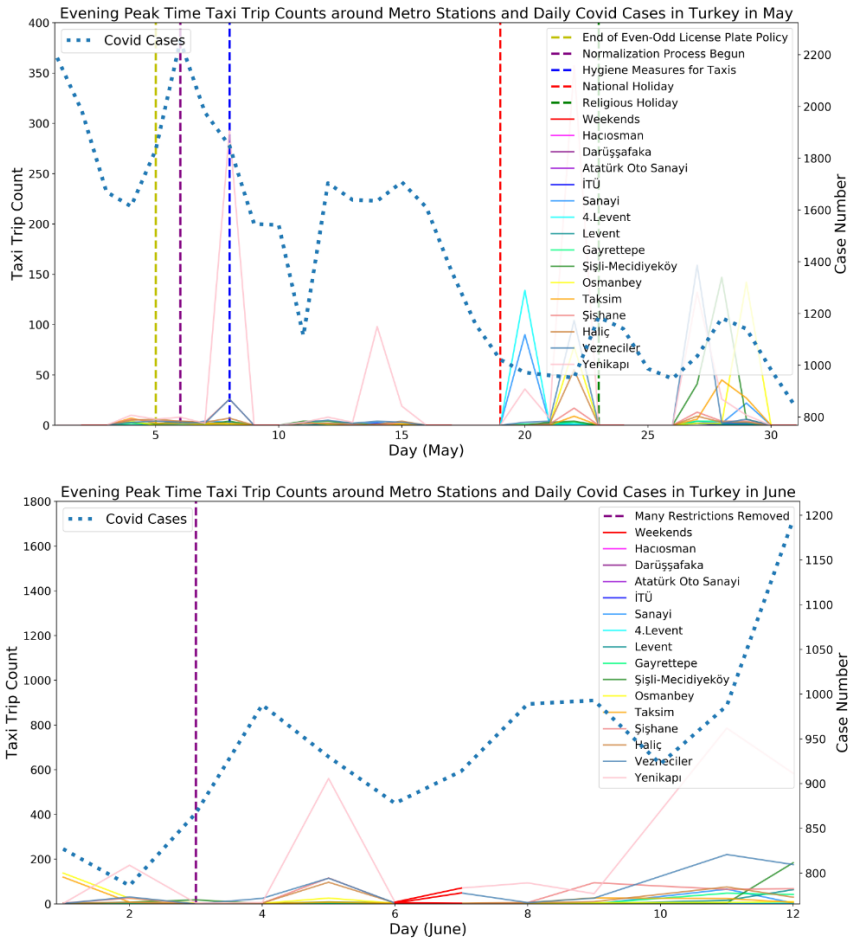


Figure 5.22. Evening peak May and June counts.

5.3.3.2. Clustering. The results for each station and transition period for evening peak trip counts are given in Table 5.16. In Transition 1, there isn't any significant cluster, and clusters are aligned based on their location (Table 5.17). Yet, cluster 2, which consists of Şişhane, Haliç and Yenikapı stations, has a lower decrease (below 60%) in taxi trip numbers. The reaction in these stations came later than the others. Sanayi station creates a cluster by itself in Transition 2, 3, and 4. The decreases in taxi trips around Sanayi station are 81.84%, 58.91%, and 37.74%, respectively, that are always lower than other stations. Hacıosman and 4.Levent also show different characteristics in Transition 2 and 3. The decrease in taxi trip counts is slightly lower than other stations in this cluster.

Table 5.16. The decrease in transitions for evening peak trip counts.

| **Station Name** | **Pre-COVID** | **Reaction to Low Case Numbers** | **Reaction with Strict Measures** | **Normalization** | **Transition to Recovery** | **Transition1** | **Transition2** | **Transition3** | **Transition4** |
|---|---|---|---|---|---|---|---|---|---|
| Hacıosman | 0.76 | 0.00 | 0.08 | 0.08 | 0.20 | -100.00 | -90.00 | -89.44 | -73.60 |
| Darüşşafaka | 1.94 | 0.00 | 0.08 | 0.08 | 0.30 | -100.00 | -96.09 | -95.88 | -84.53 |
| Atatürk Oto Sanayi | 6.21 | 0.00 | 0.14 | 0.24 | 0.40 | -100.00 | -97.80 | -96.14 | -93.56 |
| İTÜ | 20.98 | 0.17 | 0.30 | 0.48 | 0.50 | -99.21 | -98.56 | -97.71 | -97.62 |
| Sanayi | 12.85 | 1.83 | 2.33 | 5.28 | 8.00 | -85.73 | **-81.84** | **-58.91** | **-37.74** |
| 4.Levent | 35.91 | 4.50 | 2.39 | 5.76 | 1.30 | -87.47 | -93.33 | -83.96 | -96.38 |
| Levent | 102.53 | 22.08 | 0.44 | 0.60 | 9.00 | -78.46 | -99.57 | -99.41 | -91.22 |
| Gayrettepe | 142.92 | 16.25 | 1.08 | 1.08 | 10.70 | -88.63 | -99.25 | -99.24 | -92.51 |
| Şişli-Mecidiyeköy | 341.03 | 83.00 | 4.36 | 8.64 | 23.60 | -75.66 | -98.72 | -97.47 | -93.08 |
| Osmanbey | 237.41 | 52.50 | 4.62 | 9.28 | 8.70 | -77.89 | -98.05 | -96.09 | -96.34 |
| Taksim | 700.77 | 116.83 | 2.61 | 3.84 | 8.90 | -83.33 | -99.63 | -99.45 | -98.73 |
| Şişhane | 419.32 | 171.58 | 3.62 | 3.04 | 38.40 | -59.08 | -99.14 | -99.28 | -90.84 |
| Haliç | 233.74 | 126.67 | 3.29 | 3.60 | 23.70 | -45.81 | -98.59 | -98.46 | -89.86 |
| Vezneciler | 661.00 | 190.33 | 9.18 | 12.92 | 65.30 | -71.21 | -98.61 | -98.05 | -90.12 |
| Yenikapı | 1053.79 | 448.08 | 22.64 | 39.76 | 233.30 | -57.48 | -97.85 | -96.23 | -77.86 |

Table 5.17. The clusters of each station in every transition period for evening peak.

| | **K-means for transition periods** | | | |
|---|---|---|---|---|
| **Station Name** | **Transition 1** | **Transition 2** | **Transition 3** | **Transition 4** |
| Hacıosman | 0 | 1 | 0 | 0 |
| Darüşşafaka | 0 | 2 | 1 | 0 |
| Atatürk Oto Sanayi | 0 | 2 | 1 | 1 |
| İTÜ | 0 | 2 | 1 | 1 |
| Sanayi | 1 | **0** | **2** | **2** |
| 4.Levent | 1 | 1 | 0 | 1 |
| Levent | 1 | 2 | 1 | 1 |
| Gayrettepe | 1 | 2 | 1 | 1 |
| Şişli-Mecidiyeköy | 1 | 2 | 1 | 1 |
| Osmanbey | 1 | 2 | 1 | 1 |
| Taksim | 1 | 2 | 1 | 1 |
| Şişhane | 2 | 2 | 1 | 1 |
| Haliç | 2 | 2 | 1 | 1 |
| Vezneciler | 1 | 2 | 1 | 1 |
| Yenikapı | 2 | 2 | 1 | 0 |

5.3.3.3. Regressions. The analysis results for OLS, SAR, and GWR methods for Transition 1 in the evening peak are presented in Table 5.18. The OLS model performs better than the SAR model. The population and SEGE are negatively correlated with a confidence level above 95% for the OLS model, whereas population density and shopping malls are positively correlated. So, the decrease in taxi trips is more for stations with a higher population and socioeconomic level, and it is less for denser areas with more shopping malls. Here, the OLS model explains 86.5% of the data, but GWR performs better than other models in terms of R2, AIC, and log-likelihood. The coefficients of GWR for SEGE (-0.704) and shopping mall (0.442) indicate the same relationship with OLS. Even though the correlation for population and population density show parallelism with OLS, the power of coefficients is lower.

Table 5.18. The analysis results of the stations' characteristics for Transition 1 in evening peak trip counts.

| Primary Catchment Area (0-800m) of the subway for Transition 1 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | OLS | | SAR | | GWR | | |
| Variables | Coefficient | P>t value | Coefficient | P>t value | Min | Mean | Max |
| Constant | 0.000 | 0.000 | -0.004 | 0.963 | -0.120 | 0.020 | 0.119 |
| Population | -0.560 | **0.019** | -0.751 | 0.013 | -0.689 | **-0.362** | -0.027 |
| Pop. Density (per 1000m2) | 0.458 | **0.038** | 0.576 | 0.010 | 0.059 | **0.313** | 0.564 |
| Avg. Age | -0.103 | 0.648 | -0.154 | 0.367 | -0.352 | -0.017 | 0.322 |
| SEGE | -0.655 | **0.001** | -0.699 | 0.000 | -1.135 | **-0.704** | -0.218 |
| Public Transport Connection | -0.105 | 0.568 | -0.098 | 0.456 | -0.089 | 0.019 | 0.110 |
| Shopping Mall | 0.578 | **0.009** | 0.650 | 0.000 | 0.367 | 0.442 | 0.503 |
| $R^2$ | 0.865 | | 0.869 | | 0.974 | | |
| AIC | 26.564 | | 30.067 | | 9.545 | | |
| Log-likelihood | -6.282 | | -6.034 | | 6.173 | | |
| | | | Rho: -0.284 | p-val: 0.481 | Bandwidth = 14.000 | | |
| | | | Morans' I: 0.541 | | Adj. Alpha (95%) = 0.035 | | |

The analysis results for OLS, SAR, and GWR methods for Transition 2 in the evening peak are presented in Table 5.19. The OLS model performs better than the SAR model. As in Transition 1, population and shopping mall numbers have a significant level in the OLS model. The decrease in taxi trips is lower for high populated stations, and it is higher for stations with more shopping malls. The results of the GWR model is better than other models. The coefficients of GWR for the population (1.575) and shopping mall (-1.186) indicate the same relationship with OLS.

Table 5.19. The analysis results of the stations' characteristics for Transition 2 in evening peak trip counts.

| Primary Catchment Area (0-800m) of the subway for Transition 2 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | OLS | | SAR | | GWR | | |
| Variables | Coefficient | P>t value | Coefficient | P>t value | Min | Mean | Max |
| Constant | 0.000 | 0.000 | 0.037 | 0.786 | -0.447 | -0.154 | 0.170 |
| Population | 1.091 | **0.004** | 1.448 | 0.000 | 0.803 | **1.575** | 2.907 |
| Pop. Density (per 1000m2) | -0.224 | 0.524 | -0.678 | 0.020 | -1.436 | -0.572 | -0.298 |
| Avg. Age | 0.405 | 0.259 | 0.475 | 0.047 | 0.149 | 0.964 | 2.505 |
| SEGE | -0.472 | 0.133 | -0.313 | 0.163 | -1.121 | -0.333 | -0.127 |
| Public Transport Connection | 0.477 | 0.103 | 0.408 | 0.036 | 0.140 | 0.518 | 0.839 |
| Shopping Mall | -0.759 | **0.031** | -0.761 | 0.001 | -1.924 | **-1.186** | -0.579 |
| $R^2$ | 0.657 | | 0.698 | | 0.823 | | |
| AIC | 40.508 | | 42.633 | | 38.513 | | |
| Log-likelihood | -13.254 | | -6.034 | | -8.310 | | |
| | | | Rho: -0.837 | p-val: 0.171 | Bandwidth = 14.000 | | |
| | | | Morans' I: 0.042 | | Adj. Alpha (95%) = 0.035 | | |

The analysis results for OLS, SAR, and GWR methods for Transition 3 in the evening peak are presented in Table 5.20. The OLS model performs better than the SAR model, and the only significant variable in the OLS models is the population with a positive correlation. Besides, the results obtained with the GWR are better than other models. The coefficient of GWR for the population (1.516) indicates the same relationship with OLS. Besides, shopping mall (-1.211) also has a powerful negative correlation in GWR.

Table 5.20. The analysis results of the stations' characteristics for Transition 3 in evening peak trip counts.

| Primary Catchment Area (0-800m) of the subway for Transition 3 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | OLS | | SAR | | GWR | | |
| Variables | Coefficient | P>t value | Coefficient | P>t value | Min | Mean | Max |
| Constant | 0.000 | 0.000 | 0.112 | 0.445 | -0.522 | -0.131 | 0.283 |
| Population | 0.850 | **0.038** | 1.302 | 0.000 | 0.856 | **1.516** | 3.078 |
| Pop. Density (per 1000m2) | 0.080 | 0.832 | -0.448 | 0.135 | -1.467 | -0.419 | -0.151 |
| Avg. Age | 0.169 | 0.663 | 0.259 | 0.311 | 0.155 | 0.846 | 2.646 |
| SEGE | -0.278 | 0.412 | -0.125 | 0.591 | -1.185 | -0.171 | 0.116 |
| Public Transport Connection | 0.454 | 0.151 | 0.382 | 0.062 | 0.159 | 0.536 | 0.856 |
| Shopping Mall | -0.663 | 0.081 | -0.615 | 0.011 | -2.022 | **-1.211** | -0.636 |
| $R^2$ | 0.601 | | 0.652 | | 0.804 | | |
| AIC | 42.794 | | 44.716 | | 40.022 | | |
| Log-likelihood | -14.397 | | -13.358 | | -9.065 | | |
| | | | Rho: -0.980 | p-val: 0.149 | Bandwidth = 14.000 | | |
| | | | Morans' I: -0.011 | | Adj. Alpha (95%) = 0.035 | | |

The analysis results for OLS, SAR, and GWR methods for Transition 4 in the evening peak are presented in Table 5.21. The OLS model performs better than the SAR model. Here, on top of the population, public transportation connection also plays a significant role in the OLS model. Both of the variables have a positive correlation meaning that the decrease in taxi trips is lower for the stations with a higher population and with another public transportation connection option. The GWR model results better than other models. The coefficients of GWR for the population (1.475) and public transport connection (0.856) indicate the same relationship with OLS. Yet, the public transport connection is not as powerful as the OLS compared to the GWR. Instead, the shopping mall coefficient (-1.126) indicates a powerful negative correlation in the GWR model.

Table 5.21. The analysis results of the stations' characteristics for Transition 4 in evening peak trip counts.

| Primary Catchment Area (0-800m) of the subway for Transition 4 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | OLS | | SAR | | GWR | | |
| Variables | Coefficient | P>t value | Coefficient | P>t value | Min | Mean | Max |
| Constant | 0.000 | 0.000 | -0.021 | 0.897 | -0.514 | -0.223 | 0.118 |
| Population | 0.997 | **0.024** | 1.177 | 0.000 | 0.871 | **1.475** | 2.510 |
| Pop. Density (per 1000m2) | -0.433 | 0.292 | -0.740 | 0.008 | -1.380 | -0.726 | -0.507 |
| Avg. Age | 0.414 | 0.324 | 0.415 | 0.131 | 0.266 | 0.975 | 2.205 |
| SEGE | -0.524 | 0.153 | -0.414 | 0.096 | -0.763 | -0.322 | -0.051 |
| Public Transport Connection | 0.735 | **0.032** | 0.581 | 0.013 | 0.357 | **0.856** | 1.298 |
| Shopping Mall | -0.637 | 0.122 | -0.539 | 0.045 | -1.844 | **-1.126** | -0.316 |
| $R^2$ | 0.531 | | 0.583 | | 0.768 | | |
| AIC | 45.219 | | 47.447 | | 42.569 | | |
| Log-likelihood | -15.610 | | -14.724 | | -10.339 | | |
| | | | Rho: -0.889 | p-val: 0.183 | Bandwidth = 14.000 | | |
| | | | Morans' I: -0.101 | | Adj. Alpha (95%) = 0.035 | | |

### 5.3.4. Off-peak Hour Trip Counts

5.3.4.1. Trip Count Plots. Below in Figure 5.23, the off-peak trip numbers in every station are given for January and February. After the thermal camera installation in airports on January 24, there is a sharp decrease in all of the stations. The pattern of decreasing trips on Sundays is also happening in off-peak counts.

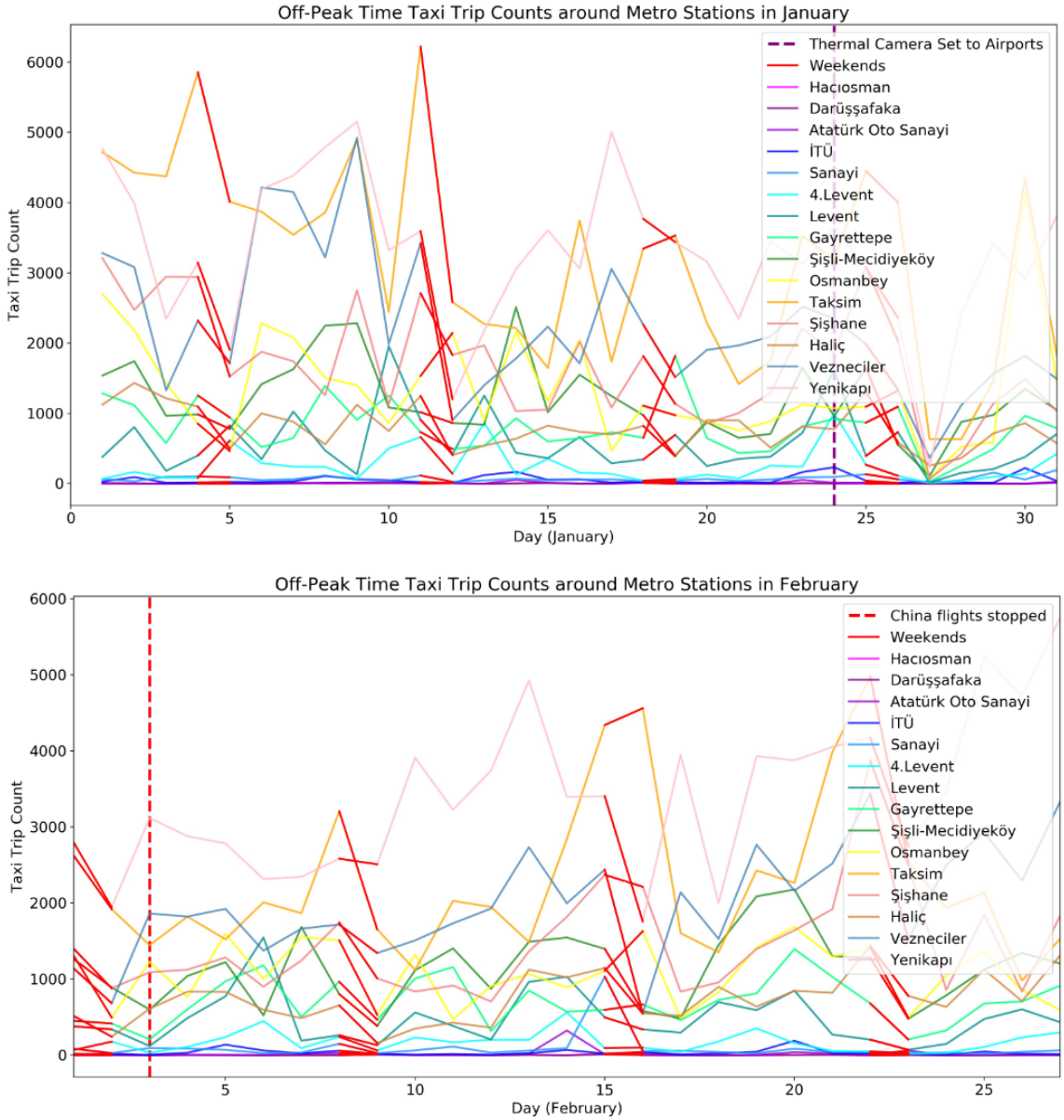


Figure 5.23. Off-peak January and February counts.

Below in Figure 5.24, the off-peak trip numbers in every station are given for March and April. The pattern that exists in total counts can also be applied for off-peak counts.

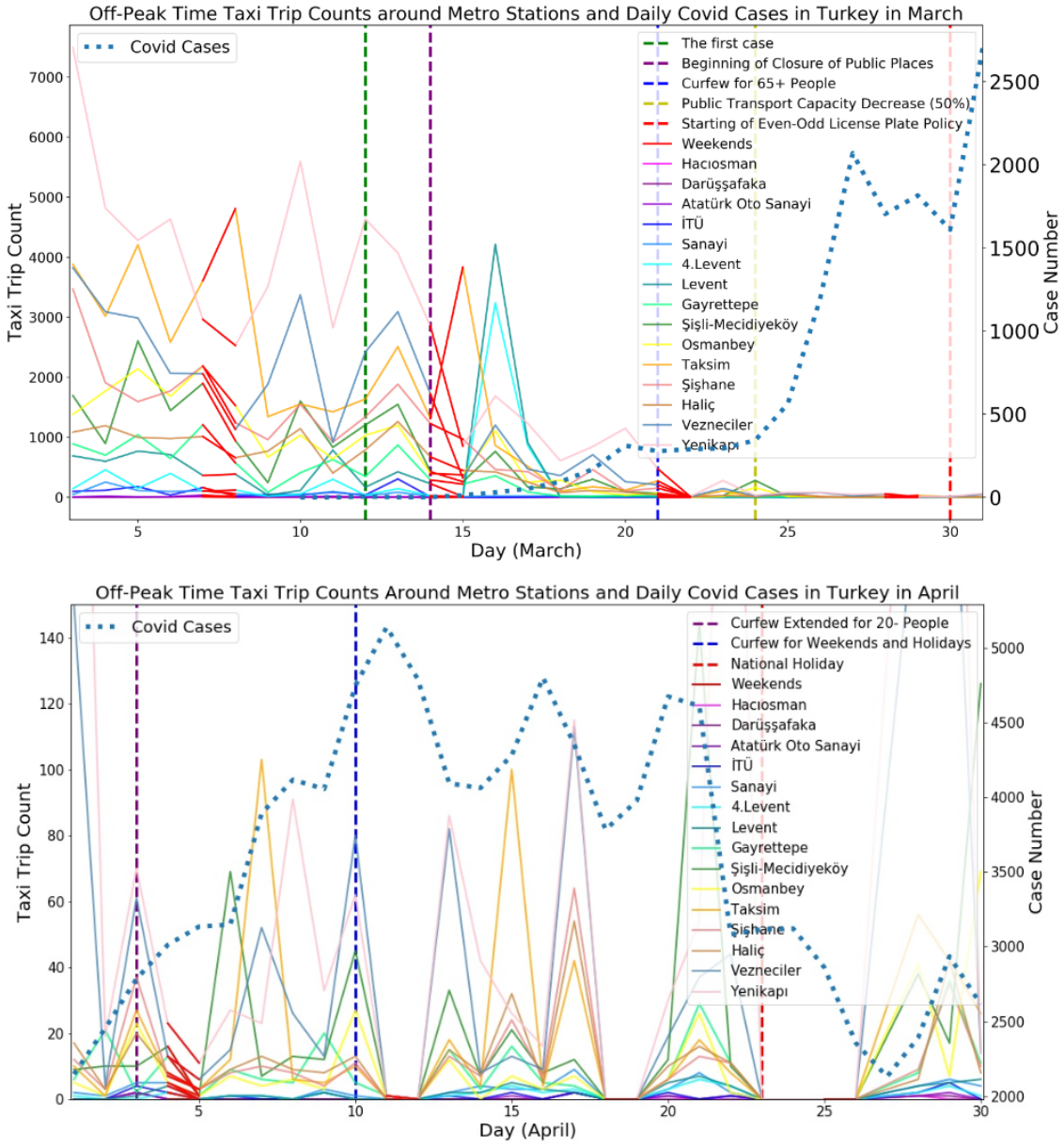


Figure 5.24. Off-peak March and April counts.

Below in Figure 5.25, the off-peak trip numbers in every station are given for May and June. Even though the trip numbers recovery came in the second half of May for peak periods, the situation is different for off-peak counts. The increase in trip counts can be observed from the beginning of May and continues in June except for weekends and holidays.

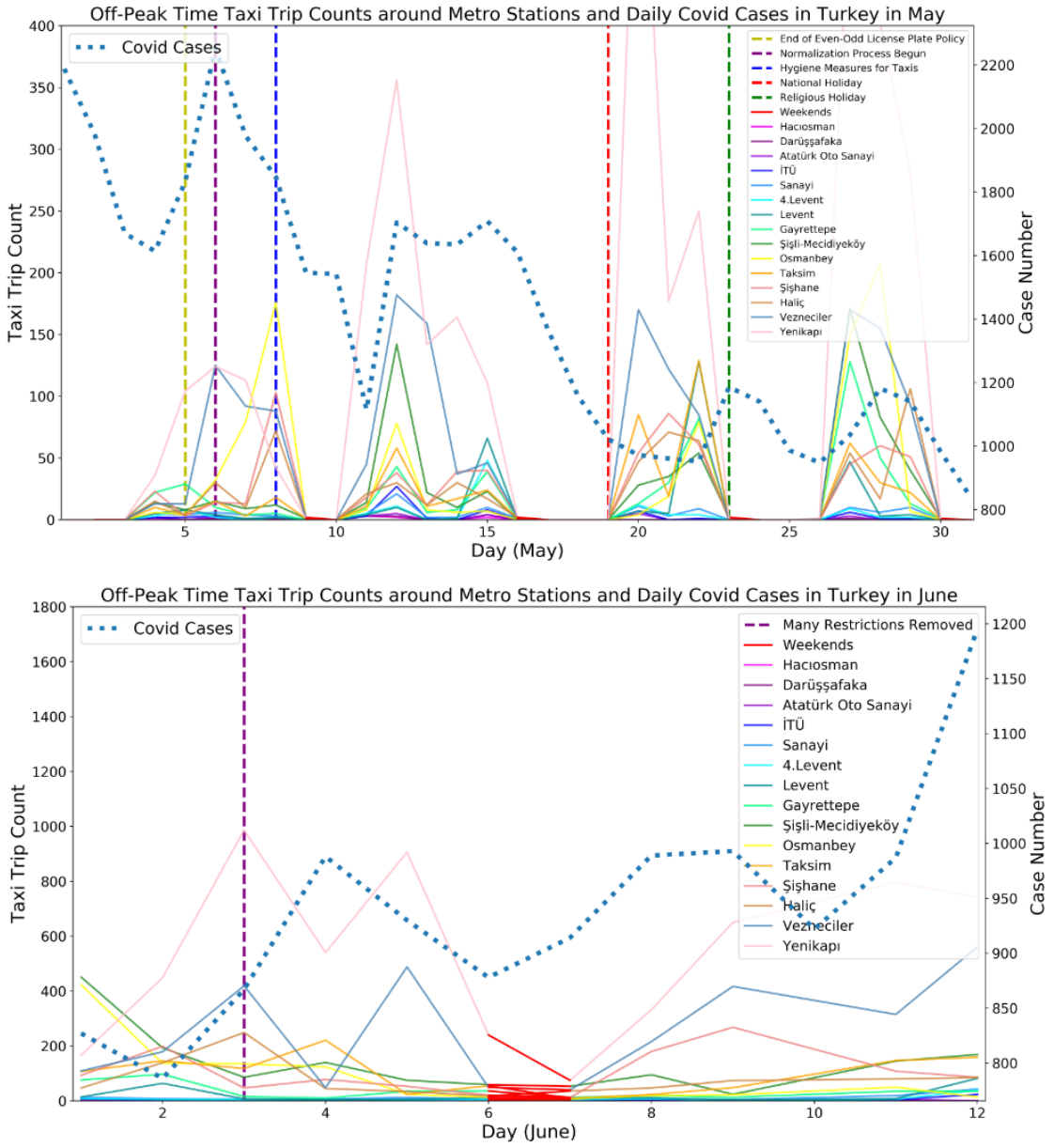


Figure 5.25. Off-peak May and June counts.

5.3.4.2. Clustering. Off-peak hours are the hours outside of the morning and evening peaks, meaning that the hours except 6-9 am and 5-8 pm. The trips in off-peak hours are obtained for each period. The change in transitions is presented in Table 5.22.

Table 5.22. The decrease in transitions for off-peak trip counts.

| **Station Name** | **Pre-COVID** | **Reaction to Low Case Numbers** | **Reaction with Strict Measures** | **Normalization** | **Transition to Recovery** | **Transition1** | **Transition2** | **Transition3** | **Transition4** |
|---|---|---|---|---|---|---|---|---|---|
| Hacıosman | 3.09 | 0.67 | 0.23 | 0.32 | 1.50 | -78.43 | **-92.65** | **-89.65** | **-51.47** |
| Darüşşafaka | 3.62 | 1.00 | 0.36 | 0.48 | 1.10 | -72.38 | **-89.96** | **-86.74** | **-69.62** |
| Atatürk Oto Sanayi | 16.64 | 2.00 | 0.58 | 0.92 | 1.60 | -87.98 | -96.54 | -94.47 | -90.38 |
| İTÜ | 52.47 | 28.92 | 1.39 | 2.28 | 4.90 | **-44.89** | -97.34 | -95.65 | -90.66 |
| Sanayi | 83.79 | 16.58 | 2.97 | 3.88 | 11.90 | -80.21 | -96.46 | -95.37 | -85.80 |
| 4.Levent | 210.95 | 358.17 | 2.38 | 4.04 | 7.90 | **69.78** | -98.87 | -98.08 | -96.26 |
| Levent | 526.15 | 485.42 | 5.42 | 11.12 | 20.60 | **-7.74** | -98.97 | -97.89 | -96.08 |
| Gayrettepe | 735.85 | 162.33 | 11.47 | 17.24 | 30.90 | -77.94 | -98.44 | -97.66 | -95.80 |
| Şişli-Mecidiyeköy | 1176.61 | 339.33 | 21.12 | 25.88 | 103.90 | -71.16 | -98.20 | -97.80 | -91.17 |
| Osmanbey | 1233.45 | 331.75 | 17.44 | 33.12 | 54.60 | -73.1 | -98.59 | -97.31 | -95.57 |
| Taksim | 2767.45 | 817.83 | 16.56 | 19.64 | 95.00 | -70.45 | -99.40 | -99.29 | -96.57 |
| Şişhane | 1587.38 | 487.50 | 15.23 | 24.96 | 105.20 | -69.29 | -99.04 | -98.43 | -93.37 |
| Haliç | 777.18 | 285.08 | 14.48 | 22.16 | 80.70 | -63.32 | -98.14 | -97.15 | -89.62 |
| Vezneciler | 2130.65 | 710.58 | 45.24 | 57.92 | 273.20 | -66.65 | -97.88 | -97.28 | -87.18 |
| Yenikapı | 3429.32 | 1174.25 | 89.23 | 135.76 | 571.70 | -65.76 | -97.4 | -96.04 | -83.33 |

As a result of the clustering given in Table 5.23, 3 different subgroups are obtained for Transition 1. 4.Levent station creates a cluster on its own, and it is the single station that has an increase with 69% in taxi trip counts for Transition 1. Then, Levent and İTÜ stations are creating a cluster, because the decrease in trip counts in those stations are smaller than the others, with 7.7% and 44.8%, respectively. The rest of the stations have similar decrease rates around 70%.

For Transition 2, the decrease in trip counts is 97% on average, and the range between maximum and minimum values is relatively narrow. Therefore, only two clusters are detected, and the difference between clusters are not as significant as the Transition 1. One of the clusters only consists of Hacıosman and Darüşşafaka stations and includes a smaller decrease than the other cluster.

In Transition 3, there are 3 clusters. Hacıosman and Darüşşafaka stations are creating one of those clusters similar to Transition 1. The decrease in those is smaller than other clusters and below 90%. Then, the 2nd cluster consists of Atatürk Oto Sanayi, İTÜ, Sanayi, and Yenikapı stations, and the decrease is around 95%, whereas, the final cluster has a decrease of around 98%.

The pattern in Transition 4 is similar to Transition 2 and 3. There are 4 clusters. Two of them consist of a single station. Hacıosman and Darüşşafaka are creating separate clusters by their own with 51.5% and 69.6% decreases in trip counts. So, the recovery of trips in those areas is more than other stations.

Table 5.23. The clusters of each station in every transition period for off-peak.

| | **K-means for transition periods** | | | |
|---|---|---|---|---|
| **Station Name** | **Transition 1** | **Transition 2** | **Transition 3** | **Transition 4** |
| Hacıosman | 0 | **1** | **1** | **1** |
| Darüşşafaka | 0 | **1** | **1** | **3** |
| Atatürk Oto Sanayi | 0 | 0 | 0 | 2 |
| İTÜ | **2** | 0 | 0 | 2 |
| Sanayi | 0 | 0 | 0 | 2 |
| 4.Levent | **1** | 0 | 2 | 0 |
| Levent | **2** | 0 | 2 | 0 |
| Gayrettepe | 0 | 0 | 2 | 0 |
| Şişli-Mecidiyeköy | 0 | 0 | 2 | 2 |
| Osmanbey | 0 | 0 | 2 | 0 |
| Taksim | 0 | 0 | 2 | 0 |
| Şişhane | 0 | 0 | 2 | 0 |
| Haliç | 0 | 0 | 2 | 2 |
| Vezneciler | 0 | 0 | 2 | 2 |
| Yenikapı | 0 | 0 | 0 | 2 |

5.3.4.3. Regressions. The analysis results for OLS, SAR, and GWR methods for Transition 1 in the off-peak period are presented in Table 5.24. The SAR model performs better than the OLS model. In the SAR model, the population and shopping mall numbers have high significant values. They have a positive correlation with the decrease in trip counts. So, the decrease in taxi trip numbers is smaller for stations with a high population and more shopping malls. The GWR model results slightly better than the other two models. The coefficients of GWR for the population (0.699) and shopping mall (0.511) indicate the same relationship with the SAR. Yet, public transport connection has a more powerful negative correlation (-0.621) than the shopping mall, in the GWR model.

Table 5.24. The analysis results of the stations' characteristics for Transition 1 in evening peak trip counts.

| Primary Catchment Area (0-800m) of the subway for Transition 1 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | OLS | | SAR | | GWR | | |
| Variables | Coefficient | P>t value | Coefficient | P>t value | Min | Mean | Max |
| Constant | 0.000 | 0.000 | 0.306 | 0.023 | -0.230 | 0.054 | 0.345 |
| Population | 0.178 | 0.697 | 0.801 | **0.003** | 0.438 | **0.699** | 2.227 |
| Pop. Density (per 1000m2) | 0.323 | 0.446 | -0.011 | 0.961 | -1.237 | -0.089 | 0.319 |
| Avg. Age | -0.367 | 0.398 | -0.297 | 0.174 | -0.245 | 0.080 | 1.703 |
| SEGE | -0.146 | 0.699 | 0.098 | 0.602 | -1.979 | -0.427 | -0.051 |
| Public Transport Connection | -0.563 | 0.111 | -0.354 | 0.052 | -0.931 | **-0.621** | -0.290 |
| Shopping Mall | 0.701 | 0.099 | 0.549 | **0.009** | 0.185 | **0.511** | 0.853 |
| $R^2$ | 0.501 | | 0.667 | | 0.762 | | |
| AIC | 46.151 | | 44.077 | | 42.917 | | |
| Log-likelihood | -16.075 | | -13.038 | | -10.513 | | |
| | | | Rho: -1.735 | p-val: 0.014 | Bandwidth = 14.000 | | |
| | | | Morans' I: -0.032 | | Adj. Alpha (95%) = 0.035 | | |

The analysis results for OLS, SAR, and GWR methods for Transition 2 in the off-peak period are presented in Table 5.25. The OLS model performs better than the SAR model, and the only significant variable is the population density around the station in the OLS model. The stations with a denser population have a higher decrease in taxi trip counts. Even though the AIC value isn't higher in the GWR, it explains the data more than other models. Besides, the reason for low AIC value is the high number of variables. The coefficient of GWR for the population (0.547) indicates the same relationship with the OLS. Besides, population and shopping also have significant coefficients in the GWR model.

Table 5.25. The analysis results of the stations' characteristics for Transition 2 in evening peak trip counts.

| Primary Catchment Area (0-800m) of the subway for Transition 2 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | OLS | | SAR | | GWR | | |
| Variables | Coefficient | P>t value | Coefficient | P>t value | Min | Mean | Max |
| Constant | 0.000 | 0.000 | -0.108 | 0.516 | -0.391 | -0.265 | -0.106 |
| Population | 0.679 | 0.085 | 0.837 | 0.007 | 0.093 | **0.547** | 0.995 |
| Pop. Density (per 1000m2) | -0.770 | **0.035** | -0.992 | 0.001 | -0.922 | **-0.486** | -0.070 |
| Avg. Age | 0.357 | 0.339 | 0.406 | 0.123 | -0.162 | 0.412 | 1.055 |
| SEGE | -0.129 | 0.691 | -0.068 | 0.767 | -0.193 | -0.066 | 0.158 |
| Public Transport Connection | 0.089 | 0.769 | 0.058 | 0.784 | -0.033 | 0.103 | 0.258 |
| Shopping Mall | -0.576 | 0.115 | -0.680 | 0.014 | -1.108 | **-0.510** | 0.045 |
| $R^2$ | 0.63 | | 0.648 | | 0.768 | | |
| AIC | 41.670 | | 44.917 | | 42.521 | | |
| Log-likelihood | -13.835 | | -13.458 | | -10.315 | | |
| | | | Rho: -0.543 | p-val: 0.385 | Bandwidth = 14.000 | | |
| | | | Morans' I: 0.290 | | Adj. Alpha (95%) = 0.035 | | |

The analysis results for OLS, SAR, and GWR methods for Transition 3 in the off-peak period are presented in Table 5.26. The OLS model performs better than the SAR model, and the only significant variable is the population density around the station in the OLS model. The stations with a denser population have a higher decrease in taxi trip counts as in Transition 2. The GWR explains the data more than other models, and the AIC value shows that the performance of the GWR is slightly better than the OLS. The coefficient of GWR for population density (-0.529) indicates the same relationship with the OLS. On top of population density, the population also has a high coefficient (0.554) in the GWR model.

Table 5.26. The analysis results of the stations' characteristics for Transition 3 in evening peak trip counts.

| Primary Catchment Area (0-800m) of the subway for Transition 3 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | OLS | | SAR | | GWR | | |
| Variables | Coefficient | P>t value | Coefficient | P>t value | Min | Mean | Max |
| Constant | 0.000 | 0.000 | -0.070 | 0.661 | -0.419 | -0.281 | -0.125 |
| Population | 0.689 | 0.063 | 0.801 | 0.008 | 0.098 | **0.554** | 1.009 |
| Pop. Density (per 1000m2) | -0.821 | **0.017** | -0.969 | 0.001 | -0.971 | **-0.529** | -0.113 |
| Avg. Age | 0.373 | 0.288 | 0.404 | 0.111 | -0.135 | 0.436 | 1.082 |
| SEGE | -0.129 | 0.673 | -0.088 | 0.692 | -0.160 | -0.052 | 0.188 |
| Public Transport Connection | 0.006 | 0.983 | -0.008 | 0.968 | -0.117 | 0.017 | 0.172 |
| Shopping Mall | -0.512 | 0.137 | -0.588 | 0.031 | -1.061 | -0.455 | 0.137 |
| $R^2$ | 0.673 | | 0.680 | | 0.814 | | |
| AIC | 39.814 | | 43.471 | | 39.243 | | |
| Log-likelihood | -12.907 | | -12.735 | | -8.676 | | |
| | | | Rho: -0.353 | p-val: 0.558 | Bandwidth = 14.000 | | |
| | | | Morans' I: 0.335 | | Adj. Alpha (95%) = 0.035 | | |

The analysis results for OLS, SAR, and GWR methods for Transition 4 in the off-peak period are presented in Table 5.27. The OLS model performs better than the SAR model. Population, population density, and SEGE have high significant values in the OLS model. The population has a positive correlation with the decrease in trip counts, whereas the correlation is negative for population density and SEGE. The GWR model results much better than the other two models. The coefficient of GWR for the population (0.586), population density (-0.526), and SEGE (-0.549) indicates the same relationship with the OLS.

Table 5.27. The analysis results of the stations' characteristics for Transition 4 in evening peak trip counts.

| Primary Catchment Area (0-800m) of the subway for Transition 4 | | | | | | | |
|---|---|---|---|---|---|---|---|
| | OLS | | SAR | | GWR | | |
| Variables | Coefficient | P>t value | Coefficient | P>t value | Min | Mean | Max |
| Constant | 0.000 | 0.000 | -9.16E-02 | 5.56E-01 | -0.410 | -0.191 | 0.132 |
| Population | 0.981 | **0.004** | 1.01E+00 | 4.35E-05 | -0.005 | **0.586** | 1.198 |
| Pop. Density (per 1000m2) | -0.939 | **0.003** | -1.04E+00 | 3.42E-05 | -1.135 | **-0.526** | 0.013 |
| Avg. Age | 0.611 | 0.059 | 6.25E-01 | 0.006422 | -0.205 | 0.458 | 1.175 |
| SEGE | -0.686 | **0.015** | -6.37E-01 | 0.00191 | -1.066 | **-0.549** | 0.001 |
| Public Transport Connection | 0.215 | 0.415 | 1.90E-01 | 0.311775 | 0.038 | 0.134 | 0.231 |
| Shopping Mall | -0.514 | 0.105 | -0.5738 | 0.015256 | -0.678 | -0.302 | 0.113 |
| $R^2$ | 0.723 | | 0.732 | | 0.928 | | |
| AIC | 37.336 | | 40.823 | | 24.988 | | |
| Log-likelihood | -11.668 | | -11.412 | | -1.548 | | |
| | | | Rho: -0.390 | p-val: 0.474 | Bandwidth = 14.000 | | |
| | | | Morans' I: 0.158 | | Adj. Alpha (95%) = 0.035 | | |

## 5.4. Results

In this section, the results of clustering and regression analysis are presented by comparing each period to obtain a bigger picture of the findings.

### 5.4.1. Clustering Results

As a result of the clustering, 4.Levent is the only station where trip counts increase in Transition 1 for off-peak periods and consecutively for total counts since most of the total counts consist of off-peak counts. The increase is exceptionally robust for off-peak hours, with 69.78%. When the station characteristics of 4.Levent are examined, it is the station with the highest population density (Table 5.28). The regression results indicate a significant negative correlation with the decrease in taxi trip counts for population density.

İTÜ and Levent stations also create different clusters for the off-peak period in Transition 1 with a lower decrease in trip counts. The decrease for İTÜ and Levent stations is 44.89% and 7.74%, respectively. Both of the 4.Levent and Levent stations are located in one of the most crowded business centers of the city, So, by considering that the home office period hasn't started yet and there are only soft measures during the period, the low decrease is logical. The same applies to İTÜ station, which locates near one of the most crowded universities in the city and also a business center. An interesting observation is that for İTÜ, Levent, and 4.Levent stations, there is a peak in taxi trip counts right after the first case is seen as indicated with a red arrow in Figure 5.26. The dates correspond to the 2nd Phase in the COVID-19 timeline.

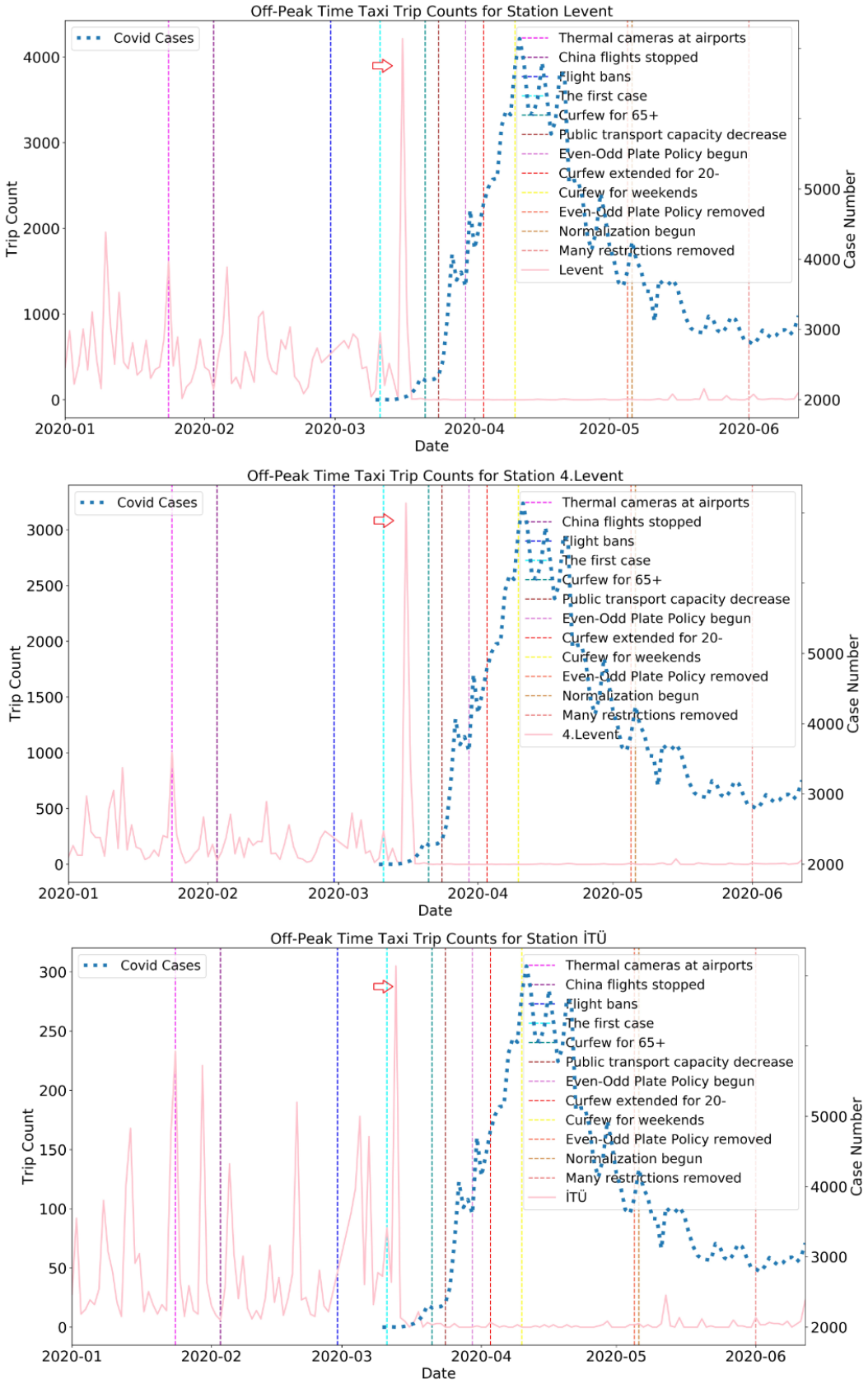


Figure 5.26. Off-peak taxi trip counts of İTÜ, 4.Levent and Levent stations during the process.

Hacıosman and Darüşşafaka are two stations that have different characteristics in Transition 2, 3, and 4 for off-peak counts, and Transition 2 and 4 for total counts. Both of the stations create a different cluster with a smaller decrease in trip counts compared to other clusters. However, the reason for the different patterns of those stations is probably the low trip numbers, because both of the stations have the lowest trip counts in every period.

The results of the clustering for total trip counts indicate a relationship between Hacıosman, Darüşşafaka, and Sanayi stations. In all of the transition periods except Transition 1, the decrease in trip counts for those stations is lower than other stations, and Taksim station has the highest trip count decrease. So, the clustering indicates a relationship between Transition 2, 3, and 4 for total trip counts, whereas, Transition 1 shows different characteristics than others for total trip counts.

Morning and evening peaks show different patterns than off-peak and each other. Haliç is the only station that creates a different cluster in every transition period during morning peak hours. It has the lowest decrease in Transition 2, 3, and 4, with 90.96%, 81.53%, and 75.73%. The socioeconomic data shows that Haliç has the lowest SEGE, university graduation ratio, average education, and the second-lowest female percentage (Table 5.28). In the correlation matrix, all those variables are found correlated with socioeconomic index. So, the low socioeconomic level could result in less caution for COVID-19. Yenikapı also creates a cluster by itself in Transition 4 during morning peak hours with 89.19% that is the second-lowest decrease after Haliç. Besides, the station has the second-lowest average education and SEGE after Haliç station.

The only significant cluster during the evening peak consists of Sanayi station, and the decrease in trip counts is significantly lower than the average with 81.84%, 58.91%, and 37.74%, respectively, for Transition 2, 3, and 4.

Table 5.28. The characteristics of stations.

| **Station Name** | **Population** | **Pop. Density** | **Avg. Age** | **Avg. Education** | **University Grad.** | **Household Size** | **SEGE** | **Female Percentage** | **Public Transport Connection** | **Shopping Mall** |
|---|---|---|---|---|---|---|---|---|---|---|
| Hacıosman | 10092.00 | 5.58 | 34.77 | 8.64 | 17.03 | 2.94 | 36.33 | 49.42 | 0 | 0 |
| Darüşşafaka | 7169.33 | 3.94 | 34.41 | 10.52 | 30.07 | 3.01 | 64.5 | 50.61 | 0 | 0 |
| Atatürk Oto Sanayi | 9021.33 | 7.68 | 33.99 | 10.47 | 29.87 | 3.01 | 70.37 | 50.44 | 0 | 0 |
| İTÜ | 10852.25 | 6.49 | 32.81 | 9.98 | 25.25 | 3.41 | 69.85 | 49.57 | 0 | 1 |
| Sanayi | 16055 | 26.80 | 33.73 | 9.59 | 25.27 | 3.30 | 59.32 | 50.10 | 1 | 1 |
| 4.Levent | 19806.50 | 37.68 | 35.69 | 9.78 | 25.61 | 3.17 | 61.78 | 50.72 | 0 | 2 |
| Levent | 11485 | 20.36 | 39.72 | 10.93 | 32.50 | 2.77 | 82.15 | 53.13 | 1 | 4 |
| Gayrettepe | 12458.80 | 20.85 | 38.89 | 11.29 | 35.37 | 2.68 | 82.08 | 53.12 | 1 | 4 |
| Şişli-Mecidiyeköy | 14564 | 36.07 | 35.51 | 9.48 | 24.71 | 2.94 | 60.12 | 50.56 | 1 | 2 |
| Osmanbey | 6195 | 30.38 | 40.96 | 10.70 | 29.85 | 2.47 | 69.93 | 53.61 | 0 | 2 |
| Taksim | 2505.78 | 16.76 | 38.10 | 9.84 | 24.70 | 2.45 | 58.11 | 45.19 | 1 | 1 |
| Şişhane | 1575.9 | 15.91 | 37.80 | 9.14 | 22.44 | 2.64 | 43.99 | 41.55 | 1 | 2 |
| Haliç | 3088 | 12.26 | 35.43 | 6.97 | 8.44 | 3.07 | 24.55 | 37.09 | 0 | 1 |
| Vezneciler | 865.60 | 7.02 | 29.47 | 9.47 | 19.99 | 5.67 | 33.86 | 36.76 | 1 | 1 |
| Yenikapı | 3941.8 | 11.38 | 33.76 | 8.41 | 17.92 | 3.03 | 29.38 | 44.16 | 1 | 2 |

### 5.4.2. Regression Results

The GWR model always resulted in better for "reaction to low case numbers", "normalization", and "transition to recovery" periods (Transition 1, 3, and 4). The only analysis that the GWR doesn't give significantly better results are the total, and off-peak trip counts in the period of "reaction with strict measures" (Transition 2). In those analyses, the OLS models are slightly superior. Since the period with strict measures is a duration in which people are forced to stay at home without their preferences, it is expected not to have a location-wise relationship. Besides, it is the period with the highest decrease in taxi trip counts as a result of curfews and even-odd license policy.

The second model that is used to compare the significance of variables is the model with the highest performance after the GWR. In every period except "the reaction to low case numbers' (Transition 1), the OLS outperforms over the SAR. Yet, in the first transition, the SAR resulted better for total and off-peak trip counts. In most of the SAR results, Moran's I value is more than 0, meaning that the distribution of variables is clustered. So, there is a spatial correlation. Yet, this correlation isn't randomly distributed over space. Therefore, assuming spatial heterogeneity is reasonable, that is proved to be right by the GWR results. In other words, the stations have a location-wise correlation, yet this location variable is not stable everywhere.

In Table 5.29, the significant variables for each transition and trip count period are given based on the GWR and the OLS models. The population is the most dominant variable in the models except the morning peak. In general, the decrease in taxi trips is lower for more populated stations as expected. Interesting is that in Transition 1 during peak hours, there is a negative correlation for the population.

So, the decrease in trips is more in highly populated areas for peak hours. Population density is also an essential variable for off-peak and total trip counts. In Transition 2, 3, and 4, the taxi trips around the densely populated stations decrease more than other stations. The reason for negative correlations for higher populated and denser

areas is that people might be more hesitant to use taxis in crowded neighborhoods due to the higher contamination risk.

The number of shopping malls is also a critical variable because the decrease in trips is lower for stations with more shopping malls, in Transition 1 in which measures were softer and the shopping malls were still open. Afterward, the trip counts negatively affected by the existence of shopping malls. As a result, people prefer a taxi as a transport mode to go to shopping malls. Therefore, the closure of shopping malls affected taxi trips negatively. Besides, the station with the highest number of shopping malls is Levent and Gayrettepe, and they also have the highest socio-economic index (SEGE) among all stations, 82.15 and 82.08, respectively. In contrast, the average is 47.73 for all stations.

The SEGE (socio-economic index) is a variable which covers many aspects related to socio-economic characteristics of a neighborhood. As explained in Section 5.3.1.3, average education and university graduation rate are directly correlated with SEGE over 90%. The female percentage also has an 83.4% correlation with SEGE. According to the analysis results, SEGE is the only significant variable for morning peak trips in Transition 2, 3, and 4. The negative correlation shows that in stations with higher socio-economic level, the decrease in taxi trips is more. Considering that the morning peak mainly comprises people who are traveling to work, it can be concluded that people with higher socioeconomic levels either use their private car or they have the choice of not traveling to work. Besides, SEGE has a negative correlation for the "transition to recovery" period (Transition 4), so the private car option might also affect the higher decrease in taxi trips. Moreover, due to a higher education level, people might be more sensitive to COVID-19.

The existence of another public transport option near the area has a negative correlation in Transition 1 for off-peak and total counts. So, the decrease in taxi trips is more if there is another public transportation option for the period of "reaction to low case numbers". By considering that the period is a phase in which people are just started to take the situation seriously, selecting public transport is more reasonable

when there are many options. But in the "transition to recovery" period (Transition 4), the situation is the opposite for evening counts. The existence of another public transport connection affected the taxi trip counts positively. Evening peaks mostly consist of people who are obliged to arrive at a point in that specific period. So, people either prefer to use taxis as a median to arrive public transportation or think that taxis have a lower risk than public transport in terms of infection as a result of applied hygiene measures.

The positive correlation of average age is significant in Transition 4 for total and evening peak counts. By considering that the curfew is removed for people +65 years old in Transition 4, it is expected to have a smaller decrease in taxi trip counts for that period. It indicates that taxis are found safer than public transport during the "transition to recovery" period in areas where the average age is higher.

Table 5.29. The significant variables for each transition and trip count.

| | **Transition 1** | **Transition 2** | **Transition 3** | **Transition 4** |
|---|---|---|---|---|
| **Total trip counts** | Population (+)<br>SEGE (-)<br>PT connection (-)<br>Shopping mall (+) | Population (+)<br>Population density (-)<br>Shopping mall (-) | Population (+)<br>Population density (-)<br>Average age (+)<br>Shopping mall (-) | Population (+)<br>Population density (-)<br>Average age (+)<br>SEGE (-) |
| **Off-peak trip counts** | Population (+)<br>PT connection (-)<br>Shopping mall (+) | Population (+)<br>Population density (-)<br>Shopping mall (-) | Population (+)<br>Population density (-) | Population (+)<br>Population density (-)<br>SEGE (-) |
| **Morning peak trip counts** | Population (-)<br>SEGE (-)<br>Shopping mall (+) | SEGE (-) | SEGE (-) | SEGE (-) |
| **Evening peak trip counts** | Population (-)<br>Population density (+)<br>SEGE (-)<br>Shopping mall (+) | Population (+)<br>Average age (+)<br>Shopping mall (-) | Population (+)<br>Shopping mall (-) | Population (+)<br>Average age (+)<br>PT connection (+)<br>Shopping mall (-) |

# 6. CONCLUSION

In this thesis, the effects of the COVID-19 around the stations on the M2 metro line is investigated using the GPS data of taxis. Initially, a comprehensive literature review is performed to understand the effects of pandemics and COVID-19, specifically around the world. Besides, the literature is used to determine the parameters that might affect taxi usage. Then, the concept of the primary catchment area restricted the study area in terms of socio-economic parameters and taxi data.

The M2 subway line is selected for this restriction since it carries most of the passengers in the network as a public transport mode that is not affected by the traffic congestion in the city. The neighborhoods that are covered by the primary catchment area of stations are detected and selected parameters are found for each neighborhood separately. Then, the parameters in neighborhoods are averaged to obtain a value for each station.

The trips are counted for each station's primary catchment area separately, and a timeline is prepared based on COVID-19 events in Turkey. The trip counts are categorized as total trip counts, off-peak trip counts, morning and evening trip counts into four segments. The plots for each segment are drawn by showing significant events during the COVID-19 process to understand the pattern of travel behavior. Afterwards, the timeline is divided into five phases according to the critical milestones and daily increase in case numbers.

The comparison for taxi trip counts is performed between the last four phases and the first phase to obtain the changes in trip counts and named as Transition 1, 2, 3, and 4. Then, k-means clustering is used to cluster the stations according to the decrease in taxi trip counts to determine the ones with different characteristics. Finally, the decreases in each Transition are analyzed using ordinary least squares (OLS), spatial auto regression (SAR), and geographically weighted regression (GWR) those which the latter resulted in better.

The results show that:

(i) Population size and population density are two dominant parameters that affect the changes in taxi trips. The reaction to COVID-19 arrived late in the areas with a higher population except for peak hours. Besides, the business centers are also felt by the effects later, and the initial reaction in those areas was an increase in taxi usage. The situation is the opposite of areas with higher population density, which resulted in a higher decrease in taxi trips after Transition 1.

(ii) The areas that already have low taxi trip counts are affected less from the COVID-19, and they recovered faster than other areas. Whereas, the areas that host tourists are possibly affected more from COVID-19, assuming that tourists dominate most of the long trips from the airport.

(iii) Morning peak hours have a significantly unique pattern than other durations. The variable that dominates the morning peak is the socio-economic index, namely the higher the SEGE is, the higher the decrease in the trip count is. It shows an obvious difference among class distinctions since the morning reflects the people who are traveling to work.

(iv) The areas with lower socio-economic levels also have a low female percentage, education level, and university graduation rate. In those areas, the decrease in trip numbers is lower, and the difference is especially significant for the "transition to recovery" period. So, people with higher socioeconomic levels are more sensitive to COVID-19, and the recovery comes slower for them.

(v) The existence of another public transport option increases the decrease in taxi trip counts in transition 1; however, the situation changes in Transition 4. So, the decrease in taxi trips gets lower that might be a result of the decrease in public transport capacity.

(vi) Location wise, the least significant period is Transition 2, where there are strict measures forced by the authorities. So, people don't prefer with their own will. Therefore, regression results are similar to each other and independent from spatial variables. Moreover, the characteristics of the stations become spatially insignificant.

(vii) In the areas with more shopping malls, the decrease in taxi trips is first lower

in Transition 1, and then higher. The results indicate that the purpose of some trips is arriving at the shopping mall and when they are closed, the trips are negatively affected. Besides, the areas with more shopping malls have a higher socioeconomic index that might also explain the high decrease of taxi trips in the area.

The contribution of the thesis is that it reveals some travel behavior characteristics with a pandemic which requires immediate action of authorities to prevent massive and permanent damages in the economy. The finding has the potential to enlighten the reaction of society to a rare yet severe event that captivates the entire world suddenly. For further studies, the methodology can apply to other transport modes and private vehicle usage to obtain a broader perspective about the travel behavior characteristics of any society.

The major limitation of this thesis is that the trips are approximated based on their durations. The information about the existence of a customer in a taxi was not available in the dataset. The data only covers a specific duration that limits the comparison of the trips with previous periods. As a natural outcome of social studies, the results cannot be generalized for other countries directly because of cultural differences.

# REFERENCES


1. History.com Editors, *Pandemics That Changed History*, 2019, https://www.history.com/topics/middle-ages/pandemics-timeline, accessed in September 2020.

2. Little B., *SARS Pandemic: How the Virus Spread Around the World in 2003*, 2020, https://www.history.com/news/sars-outbreak-china-lessons, accessed in September 2020.

3. Wikipedia, *Swine influenza*, 2020. https://en.wikipedia.org/wiki/Swine_influenza, accessed in May 2020.

4. WHO, *Middle East respiratory syndrome coronavirus (MERS-CoV)*, 2020 https://www.who.int/news room/ fact-sheets/ detail/middle-east-respiratory-syndrome-coronavirus-(mers-cov), accessed in September 2020.

5. WHO, *Ebola virus disease*, 2020, https://www.who.int/news-room/fact-sheets/detail/ebola-virus-disease, accessed in September 2020.

6. BBC, *Covid-19: The history of pandemics*, 2020, https://www.bbc.com/ future/article/20200325- covid-19-the-history-of-pandemics , accessed in September 2020.

7. Rodríguez-Morales J., K. Macgregor, S. Kanagarajah,, D. Patel and P. Schlagenhauf. "Going Global - Travel and the 2019 Novel Coronavirus", *Travel Medicine and Infectious Disease*, Vol. 33, pp. 101578, 2020.

8. WHO, *Timeline of WHO's response to COVID-19*, 2020, https://www.who.int/news-room/detail/29-06-2020-covidtimeline, accessed in September 2020.

9. Nwaeze G., "COVID-19 Outbreak and the Transportation Industry: Effects,

Prospects and Challenges", *SSRN Electronic Journal*, 2020.

10. UNCTAD, *Coronavirus will cost global tourism at least $1.2 trillion*, 2020 https://unctad.org/en/pages/newsdetails.aspx-OriginalVersionID=2416, accessed in September 2020.

11. Vinceti M., T. Filippini, K. J. Rothman, F. Ferrari, A. Goffi, G. Maffeis and N. Orsini, "Lockdown timing and efficacy in controlling COVID-19 using mobile phone tracking", *EClinicalMedicine*, Vol. 2, pp. 100457, 2020.

12. Wikipedia, *Impact of the COVID-19 pandemic on public transport*, 2020, https://en.wikipedia.org/wiki/Impact_of_the_COVID-19_pandemic_on_public_transport, accessed in September 2020.

13. Vally H. *As coronavirus restrictions ease, here's how you can navigate public transport as safely as possible*, 2020, https://theconversation.com/as-coronavirus-restrictions-ease-heres-how-you-can-navigate-public-transport-as-safely-as-possible-138845, accessed in September 2020.

14. Bajardi P., C. Poletto, J.J. Ramasco, M. Tizzoni, V. Colizza, and A.Vespignani, "Human Mobility Networks, Travel Restrictions, and the Global Spread of 2009 H1N1 Pandemic", *PLoS ONE*, Vol. 6, No. 1, 2011.

15. Ponkshe A., *Staggered timings, the new way ahead for socially distanced mobility*, 2020, https://www.orfonline.org/expert-speak/staggered-timings-the-new-way-ahead-for-socially-distanced-mobility-69485/, accessed in September 2020.

16. Harris J.E., "The Subways Seeded the Massive Coronavirus Epidemic in New York City", *SSRN Electronic Journal*, 2020.

17. Gordon A., *It's Easy, But Wrong, to Blame the Subway for the Coronavirus Pandemic*, 2020, https://www.vice.com/amp/en_us/article/qjdy33/its-easy-but-wrong-to-blame-the-subway-for-the-coronavirus-pandemic, accessed in September

2020.

18. Furth S., *Automobiles Seeded the Massive Coronavirus Epidemic in New York City, Market Urbanism*, 2020, https://marketurbanism.com/2020/04/19/automobiles-seeded-the-massive-coronavirus-epidemic-in-new-york-city/, accessed in September 2020.

19. Beutels P., N. Jia, Q.Y. Zhou, R. Smith, W.-C. Cao and S.J.D. Vlas, "The economic impact of SARS in Beijing, China", *Tropical Medicine; International Health*, Vol. 14, pp. 85-91, 2009.

20. Bell D.M., "Public Health Interventions and SARS Spread, 2003", *Emerging Infectious Diseases*, Vol. 10, pp.1900-1906, 2004.

21. Davies N., C. Cornes, G. Sherriff, *How major cities are trying to keep people walking and cycling, The Conversation*, 2020, https://theconversation.com/how-major-cities-are-trying-to-keep-people-walking-and-cycling-137909, accessed in September 2020.

22. Caballero S., *COVID-19 made cities more bike-friendly - how to keep them that way*, 2020, https://www.weforum.org/agenda/2020/06/covid-19-made-cities-more-bike-friendly-here-s-how-to-keep-them-that-way/, accessed in September 2020.

23. Litman T., "Pandemic-Resilient Community Planning: Practical Ways to Help Communities Prepare for, Respond to, and Recover from Pandemics and Other Economic, Social and Environmental Shocks", *Victoria Transport Policy Institute*, 2020.

24. Bajardi P., C. Poletto, J.J. Ramasco, M. Tizzoni, V. Colizza, A. Vespignani, "Human Mobility Networks, Travel Restrictions, and the Global Spread of 2009 H1N1 Pandemic", *PLoS ONE*, Vol. 6, 2011.

25. Apolloni, C. Poletto, J.J. Ramasco, P. Jensen, V. Colizza, "Metapopulation Epidemic Models with Heterogeneous Mixing and Travel Behaviour", *Theoretical Biology and Medical Modelling*, Vol. 11, No. 3, pp. 3-29, 2014.

26. Meloni S., N. Perra, A. Arenas, S. Gómez, Y. Moreno, A. Vespignani, "Modeling Human Mobility Responses to the Large-scale Spreading of Infectious Diseases", *Scientific Reports*, Vol. 1, pp. 62-70, 2011.

27. Belik V., T. Geisel, D. Brockmann, "Natural Human Mobility Patterns and Spatial Spread of Infectious Diseases", *Physical Review X*, Vol. 1, pp.5-10, 2011.

28. Rizzo, B. Pedalino, M. Porfiri, "A Network Model for Ebola Spreading", *Journal of Theoretical Biology*, Vol. 394, pp. 212-222, 2016.

29. Chirombo J.J., *Modelling Spatial Processes of Infectious Diseases*, Ph.D. Thesis, Lancaster University, 2018.

30. Wen J., M. Kozak, S. Yang, F. Liu, "COVID-19: Potential Effects on Chinese Citizens' lifestyle and travel", *Tourism Review*, Vol. 1, pp. 62-71, 2020.

31. Li J., T.H.H. Nguyen, J.A. Coca-Stefaniak, "Coronavirus Impacts On Post-Pandemic Planned Travel Behaviours", *Annals of Tourism Research*, Vol. 1, pp. 102964, 2020.

32. Ivanova M., I.K. Ivanov, S.H. Ivanov, "Travel Behaviour After the Pandemic: the Case of Bulgaria", *IDEAS*, Vol. 1, pp. 1-11, 2020.

33. Viboud C., M.A. Miller, B.T. Grenfell, O.N. Bjørnstad, L. Simonsen, "Air Travel and the Spread of Influenza: Important Caveats", *PLoS Medicine*, Vol. 3, pp. 11-17, 2006.

34. Hollingsworth T.D., N.M. Ferguson, R.M. Anderson, "Will travel restrictions control the international spread of pandemic influenza?", *Nature Medicine*, Vol. 12, pp. 497-499, 2006.

35. Epstein J.M., D.M. Goedecke, F. Yu, R.J. Morris, D.K. Wagener, G.V. Bobashev, "Controlling Pandemic Flu: The Value of International Air Travel Restrictions", *PLoS ONE*, Vol. 2, pp. 5-12, 2007.

36. Poletto C., M.F. Gomes, A.P.Y. Piontti, L. Rossi, L. Bioglio, D.L. Chao, *et al.*, "Assessing the Impact of Travel Restrictions on International Spread of the 2014 West African Ebola epidemic", *Eurosurveillance*, Vol. 19, pp.42-53, 2014.

37. Vinceti M., T. Filippini, K.J. Rothman, F. Ferrari, A. Goffi, G. Maffeis, et al., "Lockdown timing and efficacy in controlling COVID-19 using mobile phone tracking", *EClinicalMedicine*, Vol. 2, pp. 100457, 2020.

38. Finger F., T. Genolet, L. Mari, M. Constantin, Guillaume, A. Rinaldo, E. Bertuzzo, "Modeling the Spread of Cholera using Human Mobility Estimates Derived from Mobile Phone Records", *Epidemics V, Clearwater, FL*, 2015.

39. Peak C.M., A. Wesolowski, E.Z. Erbach-Schoenberg, A.J. Tatem, E. Wetter, X. Lu, *et al.*, "Population Mobility Reductions Associated with Travel Restrictions During the Ebola Epidemic in Sierra Leone: Use Of Mobile Phone Data", *International Journal of Epidemiology*, Vol. 47, pp. 1562-1570, 2018.

40. Pullano G., E. Valdano, N. Scarpa, S. Rubrichi, V. Colizza, "Population Mobility Reductions During COVID-19 epidemic in France under lockdown", *The Preprint for Healt Science*, Vol. 1, pp. 27-35, 2020.

41. Molloy J., C. Tchervenkov, B. Hintermann, K.W. Axhausen, *Tracing the Sars-CoV-2 Impact: The First Month in Switzerland*, Transport Findings, 2020.

42. Wang D., B. He, J. Gao, J. Chow, K. Ozbay, S. Iyer, "Impact of COVID-19 Behavioral Inertia on Reopening Strategies for New York City Transit", *General Economics*, Vol. 2, pp. 13368, 2020.

43. Wang D., F. Zuo, J. Gao, B. He, Z. Bian, S.B. Duran, C. Na, J. Wang, K. Ozbay,

J. Chow, S. Iyer, H. Nassif, X. Ban, "White Paper: Agent-based Simulation Model and Deep Learning Techniques to Evaluate and Predict Transportation Trends around COVID-19", *ResearchGate*, Vol. 5, 2020.

44. Li L., S. Wang, M. Li, J. Tan, "Comparison of Travel Mode Choice Between Taxi and Subway Regarding Traveling Convenience", *Tsinghua Science and Technology*, Vol. 23, pp. 135-144, 2018.

45. Xinhua News Agency, *The Impact of the New Crown Epidemic on Urban Transportation and Future Trends*, 2020, http://www.itdp-china.org/news/?newid=155&lang=0, accessed in September 2020.

46. Bloomberg, *How U.S. Public Transit Can Survive Coronavirus*, 2020, https://www.bloomberg.com/news/articles/2020-04-24/how-u-s-public-transit-can-survive-coronavirus, accessed in August 31, 2020.

47. Telenet, *COVID-19: Belgium Analyses Telecom Data to Measure the Impact of Confinement*, 2020, https://press. telenet. be/ covid- 19- belgium- analyses-telecom- data- to- measure- the- impact-of-confinement, accessed in August 2020.

48. COVID-19 Mobility Project, *First Report: Mobility in Germany and Social Distancing*, 2020, https://www.covid-19-mobility.org/reports/first-report-general-mobility/, accessed in September 2020.

49. COVID-19 Mobility Project, *Second Report: Mobility on the rise*, 2020, https://www.covid-19-mobility.org/reports/second-report/, accessed in September 2020.

50. Santepubliquefrance, *COVID-19 : Point épidémiologique du 16 avril*, 2020, https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies - et - infections - respiratoires/ infection - a - coronavirus/ documents/ bulletin- national/ covid- 19- point- epidemiologique- du- 16- avril- 202, accessed in September 2020.

51. Géographe E.D., O. Telle, S. Benkimoun, *Mapping the Lockdown Effects in India: How Geographers Can Contribute to Tackle Covid-19 Diffusion*, 2020, https://theconversation.com/mapping-the-lockdown-effects-in-india-how-geographers-can-contribute-to-tackle-covid-19-diffusion-136323, accessed in September 2020.

52. Wang T., Y. Zhang, X. Fu, X. Li, L. Liu, "Finding Taxi Service Management Opportunities Based on the Analysis of Choice Behavior for Passengers with Different Travel Distances", Research in Transportation Business and Management, Vol. 33, pp. 100457, 2019.

53. Alemi F., G. Circella, S. Handy, P. Mokhtarian, "What Influences Travelers to Use Uber? Exploring The Factors Affecting The Adoption of on-Demand Ride Services in California", *Travel Behaviour and Society*, Vol. 13, pp. 88-104, 2018.

54. Taylor B.D., E. Blumenberg, J. Wasserman, M. Garrett, A. Schouten, H. King, J. Paul, M. Ruvolo, *Transit Blues in the Golden State: Analyzing Recent California Ridership Trends*, UCLA Institute of Transportation Studies. June, 2020

55. Wagner, *Global car sales 1990-2020*, 2020, https://www.statista. com/ statistics/ 200002/ international- car- sales- since- 1990/, accessed in September 2020.

56. Koh D., "Occupational risks for COVID-19 infection", *Occupational Medicine*, Vol. 70, pp. 3-5, 2020.

57. Ooi P.L., S. Lim, S.K. Chew, "Use of Quarantine in the Control of SARS in Singapore", *American Journal of Infection Control*, Vol. 33, pp. 252-257, 2005.

58. Ponkshe A., *Public transit in post COVID19 India: where do we go from here*, 2020, https://www.orfonline.org/expert-speak/public-transit-in-post-covid19-india-where-do-we-go-from-here-64735/, accessed in September 2020.

59. Deurenberg-Yap M., L.L. Foo, Y.Y. Low, S.P. Chan, K. Vijaya, M. Lee, "The

Singaporean Response to the SARS Outbreak: Knowledge Sufficiency versus Public Trust", *Health Promotion International*, Vol. 20, pp. 320-326, 2005.

60. T.C. İçişleri Bakanlığı Bilgi İşlem Dairesi Başkanlığı, *81 İle Ticari Taksi Hijyen Tedbirleri Genelgesi*, 2020, https://www.icisleri.gov.tr/81-ile-ticari-taksilere-iliskin-genelgesi, accessed in September 2020.

61. Lloyd S., "Least Squares Quantization in PCM", *IEEE Transactions on Information Theory*, Vol. 28, No. 2, pp. 129-137, 1982.

62. Hutcheson G.D., L. Moutinho, "Ordinary Least-Squares Regression", *The SAGE Dictionary of Quantitative Management Research*, pp 224-228, 2011.

63. Johnston J., *Econometric Methods*, 2nd Edition, McGraw-Hill, New York, 1972.

64. Waters N., "Tobler's First Law of Geography", *International Encyclopedia of Geography: People, the Earth, Environment and Technology*, pp. 1-13, 2017.

65. Kazar B.M. and M. Celik, *Spatial Auto Regression (SAR) Model Parameter Estimation Techniques*, Springer, Boston, MA, USA, 2012.

66. Bao S., *Literature review of spatial statistics and models China Data Center*, 1998, http://www.umich.edu/ iinet/chinadata/docs/review.pdf, accessed in September 2020.

67. Casetti E., "Generating Models by The Expansion Method: Applications to Geographical Research", *Geographical Analysis*, Vol. 4, No. 1, pp. 81-91, 1972.

68. Fotheringham S., M.E. Charlton, C. Brunsdon, "Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis", *Environment and Planning A: Economy and Space*, Vol. 30, No.11, pp. 1905-1927, 1998.

69. Akaike H., "Information Theory and an Extension of the Maximum Likelihood

Principle", in B. N. Petrov, and F. Csaki (Eds.), *Proceedings of the 2nd International Symposium on Information Theory*, pp. 267-281, 1973.

70. Burnham K.P., D. R. Anderson, "Multimodel Inference: Understanding AIC and BIC in Model Selection", *Sociological Methods and Research*, Vol. 33, pp.261-304, 2004

71. TomTom Traffic Index, *Istanbul traffic report*, 2020, https://www. tomtom. com/ en_gb/ traffic- index/ istanbul- traffic/, accessed in September 2020.

72. İETT. *İstanbul'da Toplu Ulaşım*, 2020, https://www.iett.istanbul/tr/main/pages /istanbulda-toplu-ulasim/95, accessed in September 2020.

73. Metro Istanbul, *Metro Istanbul*, 2020, https://www.metro.istanbul/en/ Yolcu Hizmetleri /AgHaritalari, accessed in September 2020.

74. Metro Istanbul, *Yolcu İstatistikleri*, 2020, https://www.metro.istanbul/yolcu hizmetleri/ yolcuistatistikleri, accessed in September 2020.

75. Bizim Durak, *Taksi ve Taksi Durak Uygulaması ve BasKonuş Uygulaması*, 2019, https://www.bizimdurak.com/, accessed in September 2020.

76. Google Maps, *İstanbul*, 2020, https://www.google.com/maps, accessed in September 2020.

77. Mahallem Istanbul Project, *Istanbul Project*, 2020, http://www.mahallemistanbul .com/MahallemSEGE_/, accessed in September 2020.

78. Endeksa, *Endeksa*, 2020, https://www.endeksa.com/tr/analiz/istanbul/ fatih/ nisanca/demografi, accessed in September 2020.

79. Wikipedia, *Timeline of the COVID-19 pandemic in Turkey* 2020, https://en.wikipedia.org/wiki/Timeline_of_the_COVID-19_pandemic_in_Turkey, accessed in September 2020.

80. COVID-19 *Türkiye Web Portalı*, 2020, https://covid19.tubitak.gov.tr/turkiyede-durum, accessed in September 2020.

81. Andersen J.L.E., A. Landex, *Catchment Areas for Public Transport*, Urban Transport XIV, 2008.

82. Tu J., Z. Xia, "Examining Spatially Varying Relationships Between Land Use and Water Quality Using Geographically Weighted Regression I: Model Design and Evaluation", *Science of The Total Environment*, Vol. 407, pp. 358-378, 2008.

83. American Public Transportation Association (APTA), *Defining Transit Areas of Influence*, American Public Transportation Association, Washington, USA, 2009.

84. İBB Şehir Haritası, *İstanbul Şehir Haritası* 2020, https://sehirharitasi.ibb.gov.tr/, accessed in September 2020.

85. Wikipedia, *Earth Ellipsoid*, 2020, https://en.wikipedia.org/wiki/Earth_ellipsoid, accessed in September 2020.

86. Veness C.W., *Movable Type Scripts* 2020, https://www.movable-type.co.uk/scripts/latlong.html, accessed in September 2020.

87. T.C. Tarim Ve Orman Bakanliği Meteoroloji Genel Müdürlüğü, *2020 Yılı Ocak Ayı Sıcaklık ve Yağış Değerlendirmesi*, Ankara, February, 2020.

88. Chen D., Y. Zhang, L. Gao, N. Geng, X. Li, "The Impact of Rainfall on the Temporal and Spatial Distribution of Taxi Passengers", *Plos One*, vol. 12, pp. 9-17, 2017.

89. Rogerson P.A., *Statistical Methods for Geography*, 4th edition, Sage Publication, London, 2004.

90. Hinkle D.E., W. Wiersma, S.G. Jurs, *Applied Statistics for the Behavioral Sciences*, Wadsworth, Belmont, 2003.