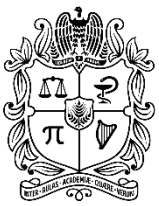


UNIVERSIDAD
NACIONAL
DE COLOMBIA

Food Services Freight (Trip) Attraction in Urban Areas of Emerging Economies

Carlos Andrés GranadaMuñoz

Universidad Nacional de Colombia
Facultad de Minas, Departamento de Ingeniería Civil
Medellín, Colombia
2023

Food Services Freight (Trip) Attraction in Urban Areas of Emerging Economies

Carlos Andrés Granada Muñoz

Tesis presentada como requisito parcial para optar al título de:
Maestría en Ingeniería - Infraestructura y Sistemas de Transporte

Director:
Carlos Alberto Gonzalez-Calderon, PhD.
Codirector:
John Jairo Posada Henao, PhD.

Línea de Investigación:
Transporte de Carga
Grupo de Investigación:
Grupo de investigación en Vías y Transporte (VITRA)
Semillero de Transporte de carga y logística

Universidad Nacional de Colombia
Facultad de Minas, Departamento de Ingeniería Civil
Medellín, Colombia
2023

Dedicatoria

A mis padres, que por su cuidado, cariño y dedicación soy y seré.
A mis profesores, por inculcar en mi la curiosidad, la lectura y el análisis.
A mis amigos, quienes han hecho esta jornada inolvidable durante todo el proceso.
A mi amor, por darme la motivación necesaria para culminar y continuar este camino.
A Dios por todas las bendiciones recibidas.

A aquel niño motivado que todo lo preguntaba y lo sigue haciendo para buscar soluciones, pero, sobre todo, le recuerdo que siempre busque la felicidad.

“Las cuatro condiciones para la felicidad según Edgar Allan Poe:
la vida al aire libre,
el amor de un ser,
la ausencia de toda ambición,
la creación.”

Albert Camus.

Declaración de obra original

Yo declaro lo siguiente:

He leído el Acuerdo 035 de 2003 del Consejo Académico de la Universidad Nacional. «Reglamento sobre propiedad intelectual» y la Normatividad Nacional relacionada al respeto de los derechos de autor. Esta disertación representa mi trabajo original, excepto donde he reconocido las ideas, las palabras, o materiales de otros autores.

Cuando se han presentado ideas o palabras de otros autores en esta disertación, he realizado su respectivo reconocimiento aplicando correctamente los esquemas de citas y referencias bibliográficas en el estilo requerido.

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto).

Por último, he sometido esta disertación a la herramienta de integridad académica, definida por la universidad.

Carlos Andrés Granada Muñoz

Nombre

Fecha 09/10/2023

Fecha

Agradecimientos

Quiero expresar mis agradecimientos a mis profesores Carlos A. Gonzalez-Calderon y John Jairo Posada Henao quienes me abrieron las puertas a la investigación y me han guiado, corregido y enseñado en este proceso de formación.

Al profesor Iván Sánchez-Díaz de la Universidad Tecnológica de Chalmers en Suecia, quien me abrió las puertas de su campus y que por su gestión pude finalizar este trabajo investigativo en las aulas de esa universidad y con su guía y conocimiento ayudaron a que siguiera creciendo personal, profesional y académicamente.

A la Universidad Nacional de Colombia que ha sido mi hogar en la última década y sus aulas me han formado como un ingeniero integral. Pero más importante, me han hecho una persona empática y crítica. Agradezco profundamente todo el apoyo que he recibido de la Universidad y lo transformadora que ha sido para mí.

De allí también han salido mis más grandes amigos y que gracias a ellos puedo sacar muy buenos recuerdo y aprendizajes profundos de la vida.

Al entrenador del equipo de balonmano de la Universidad, Juan David Cano Pozo, quién me enseñó tanto sobre el verdadero significado del deporte y de cómo construir un equipo.

Al ingeniero Rafael Agudelo y a la empresa Movintegral Consultores S.A.S. quien me acogió profesionalmente durante esta etapa y me permitió crecer como ingeniero y como tesista al mismo tiempo.

Resumen

Atracción de (viajes) carga en los servicios de alimentación en zonas urbanas de economías emergentes

Los sistemas eficientes de transporte de mercancías facilitan el intercambio de bienes y servicios, generando un crecimiento del mercado en áreas urbanas. El estudio tiene como objetivo caracterizar la atracción de viajes de carga de los servicios de alimentación en una zona urbana del Área Metropolitana del Valle de Aburrá (AMVA). Se utilizó una base de datos del Estudio de Transporte de Carga del AMVA. Esta clasificación contiene datos de 266 establecimientos. Para estimar la atracción de viajes de carga, el estudio desarrolló modelos de regresión lineal y no lineal con el método de regresión de mínimos cuadrados ordinarios. Se obtuvieron tres ecuaciones finales para modelar la relación entre los viajes de carga semanales y variables como las horas de operación, que es una nueva variable que contribuye a la literatura académica, y el área total, se encontraron que estas variables independientes que describen mejor la atracción de viajes de carga. Esto significa que el rango de los viajes de carga semanales obtenidos se encuentra entre 1 y 15 viajes de carga. Las horas de operación varían de 4 a 24 horas, y el área total varía de 4 a 670 metros cuadrados. Estos resultados ofrecen información útil para los tomadores de decisiones y propietarios de empresas de logística o establecimientos de servicios de alimentación para sugerir políticas públicas precisas relacionadas con usos de la tierra y flujos de carga, y adaptar estrategias comerciales para una mejor gestión de recursos en áreas urbanas.

Palabras clave: Atracción de viajes de carga, logística urbana, servicios de alimentación, economías en desarrollo

Abstract

Food services freight (trip) attraction in urban areas of emerging economies

Efficient freight transportation systems facilitate the exchange of goods and services and generate market growth in urban areas. The study aims to characterize the freight trip attraction (FTA) of Food Services (FS) in an urban area of the Medellin Metropolitan Area (MMA). The FS are defined by ISCI as the preparation and sale of food and drinks, mainly complete meals, for immediate consumption on the premises. It may or may not offer home delivery, alcoholic beverages, or some type of entertainment. A database from the Freight Transportation Study conducted in the case study city was used. This classification contains data for 266 establishments. To estimate FTA, the study will develop linear and non-linear regression models with the ordinary least squares (OLS) regression method. Three final equations were obtained to model the relationship between weekly freight trips and variables such as hours opened, which is a new variable to contribute to academic literature, and total area. These independent variables were found that better described the FTA. They mean the range of the weekly freight trips obtained is between 1 and 15 freight trips. The hours opened vary from 4 to 24 hours, and the total area ranges from 4 to 670 square meters. These results offer useful information to decision-makers and business owners of logistics companies or food service establishments to suggest accurate policies regarding land uses and freight flows and adapt business strategies for better resource management in urban areas.

Keywords: Freight trip attraction, urban logistics, food services, developing economies.

Contents

Pág.

Resumen ..... **IX**

List of figures ..... **XIV**

Lista of tables **1**

1. Introduction **3**
1.1 Objectives 6
1.1.1 General Objective 6
1.1.2 Specific Objectives 6

2. Background **7**

3. Methodology **11**
3.1 Modelling FTA 12
3.1.1 Sample estimation 15
3.1.2 Identifying Outliers 16
3.1.3 Handling outliers 18
3.1.4 Correlation analysis 18
3.1.5 Linear Regression Model and Ordinary Least Squares (OLS) 19
3.1.6 Non-linear regression models. 20
3.1.7 Considerations for Model Selection 22

4. Freight Trip Attraction of Food Services in Emerging Economies **25**
4.1 Case study 25
4.2 Summary statistics 27
4.2.1 Numeric variables 31
- **Outliers** 33
- **New dataset** 35

4.2.2 Correlation analysis 38
4.3 Modelling output 40
4.3.1 Models selected. 43

5. Analysis and practical implications **45**

6. Conclusions and recommendations **47**
6.1 Conclusions 47
6.2 Recommendations 48

References **49**

List of figures

Pág.

Figure 4-1: Location of MMA. 26
Figure 4-2: Histogram of weekly freight trips to FS establishments at MMA. 32
Figure 4-3: Box plot with outliers by IQR method of weekly freight trips..... 34
Figure 4-4: Box plot with outliers by IQR method of the independent's variables. 34
Figure 4-5: Histograms of dependent variable Windsorized..... 36
Figure 4-6: Histograms of independent variables Windsorized. 37
Figure 4-7: Correlation Heat Map of corrected numeric variables. 39
Figure 4-8. FTA calculated with models..... 44

List of tables

Pág.

Table 1-1. Definition of FS according to ISIC..... 4
Table 3-1: Growth of Freight transport in Colombia..... 12
Table 4-1: Trips and freight by ISIC type..... 28
Table 4-2: Table of variables description..... 30
Table 4-3: Type of variable in the database. 30
Table 4-4: First descriptive analysis of FS..... 32
Table 4-5: Number of outliers with each method. 33
Table 4-6: Final descriptive analysis of FS..... 35
Table 4-7: Models summary..... 42
Table 4-8: Models selected..... 44

1. Introduction

The rate of urban population growth can explain a significant portion of the number of trips issued for freight transportation in urban areas. This growth rate created a need to provide additional and better food, shelter, and clothing (Weeks, 2010). It is expected that by 2050, the urban population in Latin American cities will rise to 86% (Wilmsmeier et al., 2015). This growth rate and the implications of freight flows made this situation chaotic because freight trips grow at a faster rate than the population, complicating his study and enhancing the externalities produced.

Throughout history, the movement of goods has been a crucial pillar of welfare in urban centers. In recent years, transportation has directly influenced the economies of nations. Through efficient transportation systems for both freight and passengers, transportation has facilitated the exchange of goods and services and generated market growth in urban areas. For example, in Colombia, freight/logistics accounts for between 5 and 10% of GDP; cargo/logistics accounts for 10-20% of total employment; and logistics chains account for nearly 50% of employment (Area Metropolitana del Valle de Aburrá, 2018). That is, the distribution of goods is based on industries and economic activities (e.g., restaurants, offices, and construction) that rely on both producers and consumers.

The requirement to provide food to the populace is one of the oldest logistical sectors. This thesis work seeks to provide insight into the freight flows regarding the supply operations of the Food Services (FS) sector, typically known as restaurant sector in the urban area of the Medellin Metropolitan Area (MMA), Colombia.
What are the most important variables affecting FS trip attractions at their establishments? How can I foresee the freight trips in this commercial sector in the city's newest development areas?

Following the definition provided by the Chamber of Commerce of Medellin. This sector falls under the Macrogroup I - Accommodation and Food Services (AFS), but a subsection of it due the Internacional Standard Industrial Classification (ISIC) specifically categorized the FS "Restaurant, Cafeteria, and Mobile Food Service Activities" (ISIC code 561) and "Sale of alcoholic beverages for consumption within the establishment" (ISIC code 563). The specific definitions for these categories are outlined in Table 1-1.

Table 1-1. Definition of FS according to ISIC.

Description		ISIC
Restaurant, cafeteria, and mobile food service activities		561
Dispense to the table of prepared meals	This class includes: the preparation and sale of food and drinks, mainly complete meals, for immediate consumption on the premises. It may or may not offer home delivery, alcoholic beverages, or some type of entertainment. Examples include restaurants, cafes, and mobile food services.	5611
Self-service of prepared foods	This category involves the preparation and sale of food for immediate consumption, along with the sale of drinks that accompany the food, mainly under the self-service modality. These establishments may or may not offer home delivery and generally have a highly standardized decoration.	5612
Sale of prepared foods in cafeterias	This category includes the preparation and sale of light meals for immediate consumption, often not prepared on-site. These establishments may or may not serve alcoholic and non-alcoholic beverages and generally do not have a standardized decoration.	5613
Sale of alcoholic beverages for consumption within the establishment		563
Sale of alcoholic beverages for consumption within the establishment	This class includes: the sale of alcoholic beverages, mainly for consumption within the establishment with or without table service, offering the possibility of presenting some type of show. Additionally, they can provide restaurant service, and some tend to adapt a space for dancing. Bar service on board ships, when provided by independent units.	563

Source: (Cámara de Comercio de Medellín, 2023).

In MMA most of the cargo is transported in small trucks, autos, and motorcycles (79%). These vehicles produce 90% of the daily freight trips, producing negative externalities such as congestion, emissions, and noise (Gonzalez-Calderon et al., 2018). In addition to this, in the city of Medellin there are some 160,000 commercial establishments, and Accommodation and Food Services (AFS) represents approximately 9%–10% of the total.

But there is a lack of knowledge of how the cargo is moving only to the FS and how its supply chain works. Moreover, the study of consumption patterns of foods purchased at FS establishments is relatively new in developing countries, many of which are witnessing

explosive increases in consumption (Haven, 2015). Therefore, in Latin American cities, the modeling of this behavior has not been studied deeply yet, and this type of economy has its own particularities.

The current post-pandemic stage, where businesses are now focusing on online and digital functions to carry on their business, is another significant factor to be mentioned (Tasnim, 2020). Given that more people are getting access to the Internet and can use various platforms and channels of commerce to buy goods, mobile e-commerce may prove to be a successful strategy for increasing sales at FS establishments (Haven, 2015).

According to Brettmo & Sanchez-Diaz (2021) there is a common understanding among practitioners, supported by the literature, that the negative impacts of urban freight can be alleviated with better knowledge about the system and by identifying and engaging influential stakeholders. In this sense, the use of transport modeling is necessary to develop a good city plan. Also, due to the changes in the FS inside residential areas and the amount and frequency of cargo that this specific sector demands, it becomes a necessity for the city to analyze the patterns of Freight Trip Generation (FTG) by the FS sector in urban areas.

More importantly, in developing economies, this problem becomes more critical where cities have experienced the negative consequences of unplanned urban sprawl as well as the lack of adequate transportation infrastructure, making the problem worse since volatile emerging economies and unstable political situations are common in Latin American cities (de Oliveira et al., 2017; Puente-Mejia et al., 2020). Thus, as in the case of MMA, it must introduce additional limitations for strong logistics solutions. However, it is important that these outputs of the study of freight traffic flows in the FS sector be accompanied by statistics to catch the attention of decision makers and transportation managers and facilitate the discussion on transportation policies to improve the overall behavior of the cities.

1.1 Objectives

1.1.1 General Objective

To estimate statistical models to forecast the Freight Trip Attraction (FTA) of the Food Services (FS) sector in urban areas of emerging economies.

1.1.2 Specific Objectives

- To identify the most influential variables for forecasting the Freight Trip Attraction (FTA) of the Food Services (FS) sector in urban areas of emerging economies.
- To estimate the FTA of the Food Services sectors in a case study using the developed models.

2. Background

It has been found in the literature that there are different approaches by which these freight generation models can be analyzed. It was found that Freight Generation (FG) and Freight Trip Generation (FTG) are different concepts. According to Holguín-Veras, Jaller, et al. (2014) while FG is the amount of cargo generated by an establishment, FTG is the number of freight vehicle trips required to transport the cargo. The difference between FTG and general traffic counting is that FTG is more accurate since in establishments deliveries and pick-ups can be done at the same time and one vehicle can deliver to several establishments at the same stop (Sánchez-Díaz, 2018).

Therefore, the FTG will be treated in this work, which is defined below according to Holguín-Veras et al (2017) and Sánchez-Díaz (2018):

- Freight Trip Attraction (FTA) is defined as the number of trips attracted by the establishment during a certain time, for example, the deliveries of goods.
- Freight Trip Production (FTP) is the number of trips produced by the establishment, for example, the sending of goods to another establishment or returns.
- Freight Trip Generations (FTG) is the sum of FTA, FTP, and trips that involve both goods deliveries and pick-ups; this parameter can be used for the estimation of the traffic congestion generated by a certain establishment.

The relationships between FTA, FTP, FTG, business size, and some sub-sectors could be captured using regression analyses. The random sampling implemented enabled the use of regression models for inferential statistical analyses.

There are some studies that were focused on AFS establishments. (Sánchez-Díaz, 2018) estimated FTG for this sector, but jointly with the accommodation sector. He found that accommodation establishments generate on average more than twice as many trips as FS

(18.4 vs. 8.5 trips/week), and the main finding is that belonging to a chain has an incremental effect on FTG (1.6 to 1.9 times larger). Another interesting finding is that the effect of business size on FTA is larger for accommodation establishments than for food service establishments. But research on the effects of FS models is very limited, and the impact of these services on FTG is practically nonexistent.

The most related study found in the literature regarding outputs for the FS sector to the FTG model was made in Belo Horizonte, Brazil. They found that goods are delivered between five and six times a week in 66% of the FS establishments, except for meat, fish, and fresh food that have daily deliveries, and the most important variables taking into account in the modeling process are area of establishments and number of employees (de Oliveira et al., 2017). Some authors have found FTG model outputs for the AFS industry regarding FTG models. For example, the whole city of Stockholm shows that the AFS sector generates 24.648 freight trips every week, or about 15% of the city's FTG (Sánchez-Díaz, 2017). While in other European cities, this type of establishment accounts for 31% of freight moved in urban areas. This is related to the results found in some USA metropolitan areas, which account for 14%-17% of total freight traffic in the main metropolitan areas of some counties in the United States (Sánchez-Díaz, 2018). Also, some results show that retailers of perishable goods have the highest FTG per-establishment, followed by public sector offices and education establishments, retailers of non-perishable goods, and restaurants (Holguín-Veras et al., 2018)

According to (Patil et al., 2021) the interaction variable formed by employment and vehicle ownership is used in the FTA model. But this is just one example of the many combinations of variables that can be obtained from the models, which in turn have a direct relationship with the quality of the data. Current FTG modeling frameworks need data files of establishments that are standard and easy to access in developed countries such as the United States and many European countries but are incomplete or even nonexistent in Latin American countries (Puente-Mejia et al., 2020).

In MMA a total of 193,702 daily freight trips are made to all the establishments, and on average 350 kg of cargo is transported on each trip (Area Metropolitana del Valle de Aburrá, 2018), which presents a complex operating model for the logistic chain that can be

inefficient and expensive for freight transportation because of the low capacity of the vehicles (Kin et al., 2018).

In emerging economies such as Colombia, there is a significant lack of investigation related to FTG and mostly to FTA in the FS sector. By conducting studies in this area, FTA modelling can gain a deeper understanding of the factors influencing freight movements, promote economic development, optimize resource allocation, reduce environmental impact, enhance trade relations, and inform evidence-based policy and planning.

The supplier operations of the FS establishments not only suffer the impacts of traffic, but they are also a major contributor to it (Sánchez-Díaz, 2018). The FS sector is interested in efforts that aid in better understanding the behavior of its freight cargo attractiveness since it is searching for answers to its challenges, which is why it is crucial to comprehend how the sector behaves. Some initiatives enable more reliable delivery times, facilitate access for freight vehicles, promote clean vehicles, and advance freight delivery partnerships (Area Metropolitana del Valle de Aburrá, 2018). Demand for fewer freight trips means reduced emissions, together with a smaller environmental footprint for the building and better air quality in the general areas (Brettmo & Sanchez-Diaz, 2021).

The interest of this thesis is to analyze FTA applied to the movement of goods in the FS sector, which reveals fundamental information that could be used by planners and decision-makers to propose effective solutions in urban areas. According to city planning, it is critical for the city administration to spearhead these changes in specific city sectors. Thus, looking to establish a course of action that is expected to contribute to achieving a predefined set of objectives (by the city administration), planning is necessary for good urban growth and reducing negative externalities (Holguín-Veras, Wang, et al., 2014).

3. Methodology

In the first instance, a literature review about the topic was conducted. This is done with the aim of increasing the understanding of the problem and getting insights to accomplish the research and find gaps in the literature.

To estimate statistical models to forecast FTA patterns in the FS sector in urban areas of developing economies, the first step is to divide the objective into different approaches to discovering the FS sector's behavior.

This study first identified relevant establishments in order to analyze Freight Trip Attraction (FTA). Data for this analysis was sourced from the Freight Transportation Study conducted in the Medellín Metropolitan Area (MMA) in 2018. Initially, a freight survey in 2012 collected data from 2,947 commercial establishments to study cargo movements in the area. In 2018, a follow-up survey was conducted, combining data from both 2012 and 2018, resulting in a more extensive database comprising 4,382 commercial establishments within MMA. The dataset underwent rigorous cleaning and validation processes to ensure establishments were correctly categorized as Food Services (FS) as described in the Introduction. Any necessary corrections were made by reviewing the original surveys and cross-checking for errors. Subsequently, the FTA of FS establishments was estimated using this refined database derived from the freight survey.

It is important to note that the data was obtained in 2018. Thus, to have more accurate values, a forecast of the variables' growth is required. This forecast will be made with the base data of the growth of freight transport in Colombia since 2018 and apply that growth to the data. These data can be observed in **Table 3-1**. The growth percentage of freight transport in Colombia between 2018 and 2022 in terms of tons is approximately 21% and in terms of trips, it is 24%. These factors will be used to adapt the database to the current year of the investigation.

Table 3-1: Growth of Freight transport in Colombia.

Year	Tons	% Growth	Trips	% Growth
2018	112,39		8303729	
2019	118,24	5,21%	8918744	6,90%
2020	108,66	-8,10%	8206252	-8,68%
2021	123,63	13,78%	9282054	11,59%
2022	135,76	9,81%	10284945	9,75%

Source: Data accessed 05/06/2023.
(Ministerio de Transporte de Colombia, 2023).

3.1 Modelling FTA

Academic authorities have recently used FTG models to better understand the behavior of freight transportation in urban areas. This can be done for measurable variables that can be found in the sector of interest to be analyzed with the intention of producing practical results with which decision makers can generate public policies and carry out better planning of the needs of urban centers in terms of cargo movements that are required and not oversize or undersupply the solution proposals. The implication is that improving the efficiency of supply chains would directly benefit half the economy and indirectly benefit the other half (Holguín-Veras et al., 2018).

The present work will treat the FTA of the FS establishments. The logistics systems of this sector can be structured according to three elements. First, standard elements include logistics concepts such as warehousing locations, transport capacity, and intermodal setups, as well as many other transportation and distribution concepts. Second, specific elements characterize food distribution systems; it could be long-haul, short-haul, or shorter distances but in a regional context, such as those pertaining to cold storage and freshness requirements for food, determining the transport timeline, and defining equipment characteristics, which usually increase the operational cost levels in food distribution. Third, the last mile and urban delivery from the supplier store or warehouse to the FS establishment (Klumpp & Ruiner, 2018).

The focus of this thesis work will be on the second and third elements. Regarding those trips, they are the attraction trips for FS establishments needed for their operations. After determining FS establishments, an analysis of the relevant variables will be performed.

The following tools from literature could be used to accomplish this. Models are essential for forecasting freight movements and planning sustainable freight transportation. Most existing models use ordinary least squares (OLS) regression. FTA models have been used to assess the links between FTG rates and accessibility, and establishment type and size have been found to affect FTA mostly.

When the dependent variable and the independent variables have a linear relationship, OLS regression is used. In this method, the goal is to find the best-fit line that minimizes the sum of the squared differences between the predicted values and the actual values of the dependent variable. This is done by calculating the regression coefficients for each independent variable, which represent the change in the dependent variable for each unit change in the independent variable, holding all other independent variables (if there are any) constant. (Holguín-Veras, Jaller, et al., 2014; Ortúzar & Willumsen, 2011; Washington et al., 2020).

Non-linear regression is used when the relationship between the dependent variable and the independent variables is not linear. In this method, the goal is to find the best-fit curve that minimizes the sum of the squared differences between the predicted values and the actual values of the dependent variable. This is done by using non-linear functions (such as exponential, logarithmic, or polynomial functions) to model the relationship between the dependent variable and the independent variables (Holguín-Veras, Jaller, et al., 2014; Ortúzar & Willumsen, 2011; Washington et al., 2020).

Both linear and non-linear regression can be used to analyze relationships between variables that affect freight movement in the context of freight transportation. Overall, the regression method should be chosen based on the nature of the relationship between the variables of interest. Whether the relationship is linear or non-linear, regression can be a valuable tool for understanding the factors that impact freight transportation and making informed decisions related to logistics and FTA for FS.

FTA models using OLS in linear regressions and non-linear regression models will be developed in this thesis using statistical techniques to model the relationship between a dependent variable and one or more independent variables.

The analysis and prediction of transportation demand associated with the movement of goods, specifically within the FS industry, is involved in modeling data for FTA in restaurants. The factors that influence freight trips in the context of restaurant operations are identified using statistical and mathematical techniques in this process. The size of the restaurant, the number of employees, the hours open, and the size of his warehouse, among others, are all important factors to consider.

The following steps can be taken to model data for FTA using the OLS method explained in the next chapters:

- Data preparation, which includes tasks like handling missing values and outlier detection, is crucial to assuring data quality and compatibility.
- Key variables that are likely to impact FTA will be identified; a set of independent variables should be chosen for the model and an algorithm was developed to generate combinations of the variables needed for modeling.
- The model specification step comprises developing a linear and non-linear equation that expresses mathematically the link between the independent variables chosen, the dependent variable (weekly freight trips), and the OLS model.
- The model estimation process involves utilizing the OLS method to estimate the coefficients of the independent variables in the model equation.
- The model's accuracy is then assessed by looking at a variety of statistical indicators, including the significance of the coefficients, the R-squared value, the F-Fisher, and the t-test.
- Diagnostic tests are also performed to confirm that the assumptions of the OLS method are met, including the absence of multicollinearity, and homoscedasticity. Changing the variables' selection or altering the variables, such as by adopting logarithmic forms to account for non-linear correlations, can help the model if it needs to be improved.
- Model validation compares forecasts to actual observed values to assess the correctness of a model. This guarantees the model's performance is trustworthy and generalizable.

3.1.1 Sample estimation

The sample size can be determined using the statement formulated by (López-Roldán & Fachelli, 2017), which is related to other studies where surveys have been identified. Starting with a formulation for a known population, the sample is defined as Equation 3-1:

$$n = \frac{Z^2 * p * q * N}{(N-1) * \epsilon^2 + Z^2 * p * q}$$ Equation 3-1

N= Total population.
Z = Confidence level of probability that the true value of what is being studied in the population is found in the calculated sample.
p = the proportion (or percentage) of individuals that have a characteristic.
q = the proportion (or percentage) of individuals that do not have the characteristic.
ϵ= Maximum estimation error

Set the value of the deviation unit z by convention at levels of 95%, that is, with z = 1.96. The total population of restaurants in Medellín is 9144 (N), according to Secretaría de Salud Medellin, (2023), data accessed 12/05/2023.

How the total population is known will be fixed by fixing the N and seeing the associated error, or, conversely, by fixing the error and determining the associated size.

It is always true for any pair P and Q that the multiplication of these values is always less than or equal to a maximum of 0.25, which occurs when P = 0.5 and Q = 0.5; in this case, all other values are lower; that is, the variance and error that it entails are always lower. When considering this extreme situation, it is said to be in a state of maximum uncertainty or maximum indeterminacy. This is the case that will be considered when the data is processed.

With these values established, the associated error for this sample is calculated using the tool of the goal seeker to find the error term regarding the n value.

3.1.2 Identifying Outliers

Outliers are data points that exhibit unusual behavior compared to the majority of the data, potentially skewing statistical analyses. Detecting and handling outliers is crucial. Common methods for identifying outliers include:

- **Z-Scores:**

Z-scores (Equation 3-2) measure deviations from the mean (Equation 3-3) in terms of standard deviations (Equation 3-4). High absolute z-scores indicate significant deviations. Defined the z-score threshold |Z| =3 (Washington et al., 2020)

$$Z = \frac{X - \mu}{\sigma}$$ Equation 3-2

Z= Data point being analyzed.
μ = meaning of the dataset.
σ = standard deviation of the dataset.

$$\mu = \frac{1}{n} \sum_{i=1}^{n} X_i$$ Equation 3-3

μ = mean of the dataset.
n=number of data points.
X_i= individual data points.

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2{}_i}$$ Equation 3-4

σ=standard deviation of the dataset.
n=number of data points.
X_i= individual data points.

- **IQR (Interquartile Range) Method**

In statistical analysis, quartiles are pivotal measures that divide a dataset into four equal parts, each encompassing 25% of the data. These quartiles, namely Q1, Q2 (the median), and Q3, serve as essential tools for assessing data distribution and variability. (Kwak & Kim, 2017)

Lower Quartile (Q1):

The lower quartile, Q1, represents the value below which 25% of the data points fall. To ascertain Q1, one must first arrange the data in ascending order and then locate the value at the 25th percentile. Essentially, Q1 demarcates the threshold beneath which the lowest 25% of the data is situated.

Upper Quartile (Q3):

Conversely, the upper quartile, Q3, delineates the value below which 75% of the data points lie. Similar to Q1, determining Q3 involves sorting the data in ascending order and identifying the value at the 75th percentile. In essence, Q3 segregates the upper 25% of the data from the rest.

Interquartile Range (IQR):

The interquartile range signifies the dispersion within the middle 50% of the data and is calculated as the difference between Q3 and Q1, expressed as Equation 3-5. It provides insights into the variability confined to the central portion of the dataset, excluding any extreme values.

$$IQR = Q_3 - Q_1 \qquad \text{Equation 3-5}$$

IQR= Interquartile Range
Q_3= Upper quartile
Q_1= Lower quartile

Quartiles offer a visual representation of data distribution and play a pivotal role in outlier detection techniques, such as the IQR method. By assessing values in the Outlier below limit (Equation 3-6) and in the Outlier above limit (Equation 3-7) as potential outliers, quartiles aid in pinpointing extreme observations that deviate from the central data distribution. Consequently, quartiles and the IQR method serve as invaluable tools for data exploration and outlier identification in statistical analyses.

$$\text{Outlier below limit} = Q_1 - 1{,}5 * IQR \qquad \text{Equation 3-6}$$

$$\text{Outlier above limit} = Q_3 + 1{,}5 * IQR \qquad \text{Equation 3-7}$$

IQR= Interquartile Range
Q_3= Upper quartile
Q_1= Lower quartile

3.1.3 Handling outliers

Alternatively, more robust statistical measures, such as the median and quartile range, can provide valuable insights because they are less affected by extreme values. Furthermore, the use of box plots provides a visual tool for identifying outliers. Outliers are data points that lie beyond the upper or lower fence lines in this box plot representation.

Data trimming is one potential approach for dealing with outliers, in which the analysis is performed on a dataset that excludes the outliers. While trimmed estimators such as the mean help to reduce data variance, they can introduce bias by underestimating or overestimating values. It should be noted that excluding outliers from the analysis may be deemed insufficient, as these outliers are still valid observations. (Kwak & Kim, 2017).

For that reason, it is recommended to use Winsorization is the process of adjusting the weights assigned to outliers or replacing the values under scrutiny for outliers with expected values. The weight adjustment technique allows for weight changes while preserving outlier values, limiting the impact of outliers. In contrast, the value modification technique allows outlier values to be replaced with either the highest or second-smallest value from observations that do not include outliers.

3.1.4 Correlation analysis

A correlation analysis is carried out to verify how dependent the variables are on each other, so that the model carried out is adequate for the identification of the most optimal variables. The following expression gives it:

$$r_{xy} = \frac{\frac{\sum XY}{N} - \bar{X}\bar{Y}}{S_x\, S_y}$$ Equation 3-8

XY = the variables under consideration
$\bar{X}\,\bar{Y}$ = Their mean
$S_x\, S_y$ =standard deviation.

The result of this equation oscillates, in absolute terms, between 0 and 1.
Correlations were generated between the data of the different categories, allowing initially to discard highly correlated variables, that is, with an r greater than |0,5|. Carrying out these analyses turns out to be valuable to be able to carry out the realization of the models, since by this method it is really known without subjectivity which variables can enter a model, in addition to allowing a sweep of the variables that would definitely not yield data.

3.1.5 Linear Regression Model and Ordinary Least Squares (OLS)

Linear regression is a statistical technique used to model the relationship between a dependent variable Y and one or more independent variables X. The relationship is assumed to be linear, meaning that the change in Y is proportional to the change in X. The goal of linear regression is to find the line that best fits the data, which is called the regression line. The regression line is characterized by its slope and intercept, which are estimated from the data using a method called ordinary least squares (OLS). The slope represents the change in Y per unit change in X, while the intercept represents the value of Y when X is zero. It is an important tool for both descriptive and predictive modeling.

As mentioned, OLS is a method for estimating the parameters of a linear regression model, and according to Ortúzar & Willumsen, (2011), the advantages of OLS models are that they're able to identify relations pertaining to demand generation and can be used not only to forecast future demand but also to establish the linkage between variables The least squares method, which is a popular line-fitting technique, minimizes the value of $\sqrt{\varepsilon_i}$. By introducing the variable $x_i = X_i - \tilde{X}$, where $\tilde{X}$, is the mean of X. A way to explain the OLS is the following:

In the first step, we must collect a sample of data consisting of pairs of values for variables X and Y.

Must make a model specification specifying the functional form of the linear regression model, which can be written as can be seen in Equation 3-9. Here β_0 and β_1 are the intercept and slope of the regression line, respectively, and $\in$ is the random error term.

$$\boldsymbol{Y} = \beta_0 + \beta_1 \boldsymbol{X} + \in \qquad \text{Equation 3-9}$$

β_0=Constant term (if is required)
β_1 =Calculated parameter of the variable
X= Independent variable of the model
$\in$= Random error term

Make certain assumptions about the errors, such as that they are normally distributed and have constant variance. These assumptions will be necessary later to conduct formal tests for the model. The error term includes both measurement and specification errors, as is typical in statistical models.

Use the OLS method to estimate the parameters by minimizing the sum of squared errors between the predicted values of Y and the actual values of Y in the sample.
The errors e_i (residuals) are the difference between the observed and estimated values of Y and the coefficients β_0 and β_1 are known as least-squares estimators.

3.1.6 Non-linear regression models.

The process of transforming linear regressions into non-linear regressions based on logarithmic transformation is highly demanded by academics, following the process described by Newman, (1993).
Firstly, it's important to understand that linear regression models the relationship between a dependent variable and one or more independent variables using a linear equation. However, in many cases, the relationship between the dependent variable and the independent variable(s) is not linear.
One approach to addressing this issue is to transform the data using a logarithmic transformation, which can help to linearize the relationship between the variables. To transform a linear regression into a non-linear regression using a logarithmic transformation, you can take the following steps:

- Take the logarithm of the dependent variable (Y) and/or the independent variable(s) (X) as appropriate, using either the natural logarithm (ln) or the base-10 logarithm (log). The choice of logarithm will depend on the specific problem and data being analyzed.

The transformation can have the form presented in Equation 3-10 for the Log-Log form (assuming natural logarithm):

$$\ln Y = \beta_0 + \beta_1 \ln X + \in \qquad \text{Equation 3-10}$$

And for the Log-Lin (assuming Log base-10) the form can be seen in Equation 3-11:

$$\log Y = \beta_0 + \beta_1 X + \in \qquad \text{Equation 3-11}$$

- Fit a linear regression model to the transformed data, as explained in the last section. This involves estimating the coefficients of the linear equation that relates the transformed dependent variable to the transformed independent variable(s).
- Interpret the results of the linear regression in terms of the original variables. This requires exponentiating the coefficients and/or predicted values as appropriate to

obtain estimates of the relationship between the original dependent variable and independent variable(s).

Make the transformation to his original expression using the logarithm and potential properties obtain the transformed equations for Log-Log form in the Equation 3-12 and for Log-Lin form in the Equation 3-13.

$$Y = e^{(\beta_0)} * X^{(\beta_1)} * e^{(\in)} \quad \text{Equation 3-12}$$

$$Y = 10^{(\beta_0)} * 10^{(\beta_1 X)} * 10^{(\in)} \quad \text{Equation 3-13}$$

However, it's important to note that when transforming the data using a logarithmic transformation, the resulting estimates may be biased. This is because the logarithmic transformation compresses the range of the data, which can lead to an overestimation of small values and an underestimation of large values. (Newman, 1993)

To correct for this bias, we can use a bias correction factor, which involves adding a constant to the predicted values to adjust for the compression caused by the error in the residuals in the model. The specific form of the bias correction factor it's denoted with $\in$**.**

Residuals serve as indicators of the model's errors or discrepancies. The observed (actual) values are subtracted from the model's predicted values to obtain residuals. Subtracting the observed (actual) values from the model's predicted values yields residuals. This subtraction process aids in identifying any underlying systematic patterns or biases present in the model's performance. Once the analysis of residuals is complete and patterns or biases are identified, the next step is to formulate a correction method based on this analysis. This correction technique aims to correct the model's forecasts to lessen the detected systematic errors. This correction method aims to adjust the model's predictions to mitigate the observed systematic errors. Two distinct strategies can be applied, depending on the nature of the residuals. (Newman, 1993)

If the residuals have a normal distribution, bias can be corrected using an estimation based on this distribution.

$$\in = \frac{MSE}{2} \quad \text{Equation 3-14}$$

MSE=Mean Square error

$$MSE = \frac{\sum_{i=1}^{N} e_i^2}{N-2}$$ Equation 3-15

e = Regression residual i.
N= Total number of pairs.

In cases where the residuals do not conform to a normal distribution, it is advisable to employ the "smearing estimate of bias." When the residuals display non-normal characteristics, this method is especially useful for identifying the bias in predictions.

$$\in = \frac{\sum_{i=1}^{N} e_i}{N}$$ Equation 3-16

e = Regression residual i.
N= Total number of pairs.

The Bias correction is calculated regarding the residuals of the models calculated Where MSE are the Mean Square Error.

3.1.7 Considerations for Model Selection

Selecting the appropriate model for modeling FA in FS using the calculation models is critical for accurate and reliable predictions. The student's t-test and the Fisher test are two commonly used tests in this context. The student's t-test aids in determining the significance of individual coefficients in the model, indicating whether a specific independent variable has a significant impact on FA. The Fisher test, on the other hand, evaluates the joint significance of multiple coefficients to determine the overall significance of the model. Both tests are critical in determining the model's adequacy and provide valuable insights into the factors influencing freight trips in the FS sector. Selecting the most appropriate model for accurately predicting and analyzing freight trip generation in restaurants by carefully considering the results of these tests.

Student's t-test: It is for variables, and it is calculated using Equation 3-17. It is a test statistic to approve or reject a hypothesis. The most commonly used criterion is the value of |1.96|, which indicates that the null hypothesis is accepted with a significance of 95%. The significance test of the regression coefficients (T-test) is a tool for testing statistical

hypotheses, and it consists of a procedure that uses sample results to verify the truth of a null hypothesis. In the language of significance tests, it is said that a test statistic is statistically significant if the value of the test statistic falls in the critical region; otherwise, the null hypothesis is not rejected.

$$t = \frac{\bar{X} - \mu}{\frac{S_x}{\sqrt{n}}}$$ Equation 3-17

$\bar{X} = \text{Population mean}$
$\mu = \text{Sample mean}$
$S_x = \text{Standard deviation}$
$\sqrt{n} = \text{Number of data}$

After meeting the requirement of being greater than |1.96|, the "t" indicates how representative a variable is or is not within a model. According to the 5% significance level, when this value is greater than the other variables, it indicates that said variable is representative, which is important when selecting a model that includes it.

Fisher test: It is for models. It is used to compare models that have been fitted to a dataset and identify the best one. The Fisher's F-test consists of a quantitative procedure to measure the relationship between the variances of the different variables in the given model and the variance of the data within each variable. The F-test explains the ability of the variables in a model to explain a phenomenon. See Equation 3-18.

$$F = \frac{V_E}{V_D}$$ Equation 3-18

$V_E =$ Variance between variables
$V_D =$ Variance within variable

So, if there is little similarity between the variances of the variables, this indicates that there is more dispersion between the variables than within them. This is exactly what a transportation model requires: for the variance between the various variables incorporated to be greater to estimate trips, as the greatest number of trips can be considered this way.

When the variance between variables is greater than the variance within variables, it indicates that the variables are quite different from one another. This is what a model requires because higher variability between variables results in a model that may cover more characteristics than a model with low variance variables. To select a better-quality

model that can represent the greatest number of trips given in the study area, it is necessary to associate an "F" for Fisher with a high value in comparison to the other models.

As a summary, for the models, it is important to identify if they are conceptually valid, that is, if all the coefficients are positive values, to avoid obtaining negative model results and misinterpreting inconsistent trips since there are no negative trips in any situation since it would not be possible. In addition, they must be statistically significant, which is determined by obtaining, as previously stated, that the "t" is greater than |1.96| to obtain 95% confidence in the model, and the other parameter is F, which should be remembered: the higher this value is, the theoretically determines between one model and another, which would be better to use.

The coefficient of determination (R^2) is used as a measure of goodness of fit, and it measures the percentage of the total proportion of Y explained by a regression model (See Equation 3-19). It is commonly used in linear models when there is only one variable, and it determines how far the data is from a proposed regression and how well it models a set of data. However, for transportation modeling, when using, in many cases, more than one linear variable, this R^2 value does not determine whether a model is good or not (Ortúzar & Willumsen, 2011).

$$R^2 = \frac{\sum_{t=1}^{N}(\hat{Y}_t - \bar{Y})^2}{\sum_{t=1}^{N}(Y_t - \bar{Y})^2}$$ Equation 3-19

N = Total number of pairs
$\hat{Y}_t$ = Value calculated with the model.
$\bar{Y}$ = Mean value of the data
Y_t = Value in the position t

4. Freight Trip Attraction of Food Services in Emerging Economies

This chapter provides an in-depth explanation of FS FTA, about the movement of food-related goods required for serving urban restaurants. It seeks to pinpoint the variables that influence this freight's attraction. Through careful examination, it seeks to provide insights into the challenges and potential of managing FS freight.

In an effort to provide light on the transit of freight for FS, this chapter respectfully contributes to the discussion of urban logistics in developing economies. In the first section, a description of the case study will be given. Following that, a description of the data processing and modeling in relation to the Methodology section will be noted. Finally, the models developed will be displayed.

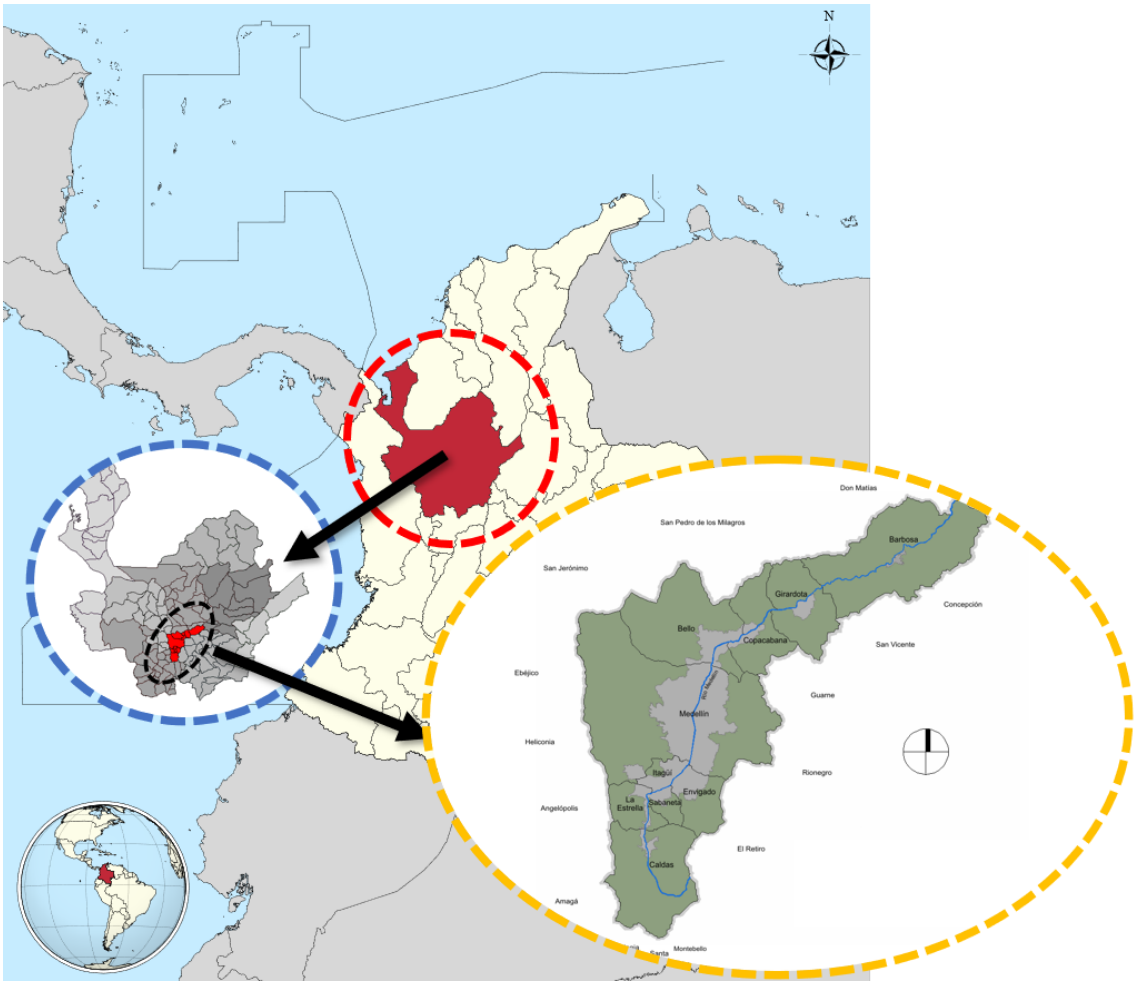
4.1 Case study

The characteristics of the urban region where the methodology is used, as well as the procedure for gathering, revising, and classifying data, are presented in this part. The study focuses on Colombia's emerging economy, particularly that of the Medellin Metropolitan Area.

The 10 cities that make up the Medellin Metropolitan Area (MMA), which has a population of around 4.2 million, are home to about 160,000 registered commercial establishments, according to data from the city's chambers of commerce (Area Metropolitana del Valle de Aburrá, 2018). Notably, Medellin stands out among these 10 cities because it is Colombia's second-largest city after the capital, Bogota. According to Area Metropolitana del Valle de Aburrá (2018), the MMA contributes significantly to the nation's GDP, accounting for almost

11% of it. Additionally, a significant international airport that is necessary for both import and export of commodities is located near the MMA, according to Gonzalez-Calderon et al., (2018). Figure 4 1 depicts the location of the case study region.

Figure 4-1: Location of MMA.



Source: (Área Metropolitana del Valle de Aburrá, 2010)

4% of the MMA's establishments are located in the Northern part of the zone, despite having 15% of the population. Gonzalez-Calderon et al. (2018). Because of the lower income levels in this region. In contrast, despite having the same proportion of population, 10% of enterprises are concentrated in the Southern part. 85% of the businesses and 70% of the population of Medellin are, however, concentrated in the city's central area. Even along the main road corridor that runs through the city, the city's central business district is greatly influenced by freight traffic.

To analyze the FTA, the establishments must be identified. The general data to use in this thesis comes from the Freight Transportation Study conducted in the Medellín Metropolitan Area in 2018 (Area Metropolitana del Valle de Aburrá, 2018). In 2012, a freight survey was conducted in MMA on a sample of 2,947 commercial establishments to analyze the cargo movements and patterns in the area. In 2018, another freight survey was conducted, which included data from 2018 as well as the 2012 data, resulting in a new database. This new study used a larger sample of 4,382 commercial establishments to characterize freight movement and estimate FG models, specifically for establishments in MMA.

The database was modified to produce the optimal dataset for FS modeling. This included a cleaning process that checked each institution's FS classification to ensure it was accurate.

4.2 Summary statistics

ISIC (International Standard Industrial Classification) codes are used to illustrate the movement of freight in the Medellin Metropolitan Area (MMA) in Table 4-1. The codes are arranged hierarchically, with the "Section," a broad category marked by a letter, followed by the "Division," which is denoted by two numeric numbers, and then the "Group," which is denoted by three numeric digits.

A noteworthy 62.6% of freight transportation is accounted for by Group G, which covers "Wholesale and retail trade; repair of motor vehicles and motorcycles," making it the dominant industry. Group F, which stands for "Construction," is in third place, followed by Group C, which stands for "Manufacturing," which contributed 13.4%. Additionally, the table provides data on the Freight Trip Generation (FTG) across several industrial sectors. Closely trailing Group G, which generates the most trips with 44.9% of the total, the Manufacturing sector accounts for 15.8% of all journeys.

Group I "Accommodation and Food Services" (AFS) contributes by 12,1% of the trips in the MMA, which is less than the values observed in the literature when compared to cities in the USA and Europe as seen in the Background of this document.

Table 4-1: Trips and freight by ISIC type.

ISIC Type	Trips/Day	Freight/Day (ton)	ton/Trip	Trips/Day (%)	Freight/Day (%)
A Agriculture, forestry, and fishing	1365,1	1420,3	1	0,7	2,1
B Mining and quarrying	375,2	71,8	0,2	0,2	0,1
C Manufacturing	30671,8	9119,2	0,3	15,8	13,4
D Electricity, gas, steam, and air conditioning supply	4,7	0,1	0	0	0
E Water supply; sewerage, waste management and remediation	946,1	759	0,8	0,5	1,1
F Construction	10605,9	7460,8	0,7	5,5	11
G Wholesale and retail trade; repair of motor vehicles and motorcycles	86931,6	42522,7	0,5	44,9	62,6
H Transportation and storage	5614	951,7	0,2	2,9	1,4
I Accommodation and food service activities	**23388,7**	**1987,9**	**0,1**	**12,1**	**2,9**
J Information and communication	3040,6	102,6	0	1,6	0,2
K Financial and insurance	2534,6	95,1	0	1,3	0,1
L Real estate activities	3448,7	224,8	0,1	1,8	0,3
M Professional, scientific, and technical activities	10857,8	1273,7	0,1	5,6	1,9
N Administrative and support service activities	3716,3	1198,7	0,3	1,9	1,8
O Public administration and defense	29	5,9	0,2	0	0
P Education	607,9	42,7	0,1	0,3	0,1
Q Human health and social work activities	1525,8	119,9	0,1	0,8	0,2
R Arts, entertainment, and recreation	1663,4	152,6	0,1	0,9	0,2
S Other service activities	6374,5	377	0,1	3,3	0,6
T Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use	0,4	0,1	0,2	0	0
Total	193702,1	67886,6	0,4	100	100

Source: (Area Metropolitana del Valle de Aburrá, 2018)

The FS sector associated with restaurants will be examined in this work. The Chamber of Commerce of Medellin, where the case study will be conducted, will utilize its description of this sort of establishment as its guide. The International Standard Industrial Classification (ISIC) divides the different types of commercial businesses into several groups, and the Accommodation and Food Services (AFS) group has a division devoted solely to the Food Services (FS).

The AFS are represented by 396 businesses, or 9% of the overall sample. It consists of the ISIC codes 551, 561, 562, and 563, as defined by the Medellin Chamber of Commerce. For the establishments of FS and because many restaurants offer alcoholic beverages on their menus, the categories 561 ("Restaurant, Cafeteria, and Mobile Food Service Activities") and 563 ("Sale of alcoholic beverages for consumption within the establishment") are taken into account to create the new dataset (the precise definitions are in Table 1-1**)**. In this classification, the sample was converted into a dataset of 266 establishments. The remaining codes in Category I of AFS are 551 "short-stay accommodation activities" and 562 "catering activities for events and other food services."

Due to the procedure of the sample estimation, the author knows that there is a fixed population of restaurants in MMA, so the sample of 266 observations constitutes an error of 5.2% with the process described in the section Sample estimation.

In this database, there are a variety of characteristics for internal classification and to describe how these businesses behave when it comes to moving freight. The database has a total of 19 columns, but only 10 remain after determining the most useful variables for modeling.

The variables were categorized based on their nature, including continuous, categorical, and binary attributes, as well as by their characteristics, such as freight quantity, establishment attributes, and delivery specifics. And this database derived from the freight survey is used to calculate the FTA of FS establishments and could be observed in the **Table 4-2** and **Table 4-3**.

Table 4-2: Table of variables description.

Variable	Description
Year	Year of the survey
List Number	Identification in the survey
MMA Zone	The zone where the establishment is located is the establishment inside the MMA
Mu	Code of the municipality of the survey
ISIC Description	Description of the ISIC
Division Group	The group division inside the Chamber of commerce of Medellin
Macrogroup	The letter assigned to a type of establishment
MG Description	The macrogroup description
Establishment Name	The name of the establishment
Hours Open (h)	The hours of operation of the establishment
Total Employees	The number of total employees of the establishment
Establishment Size	Categorical variable of the size of establishment
Total Area (m^2)	The total area of the establishment
Has Warehouse	a dummy variable that takes the value of 1 if the establishment has a warehouse and 0 if not
Warehouse Area (m^2)	The total area of the warehouse
Has Parking	a dummy variable that takes the value of 1 if the establishment has parking and 0 if not
Parking Minutes (min)	The time of the freight trip while downloading
Weekly Trips (trips/week)	The numbers of freight weekly trips received by the establishment (FTA)
Weekly Kilograms (kg/week)	The freight amount of cargo received by the establishment in a week (FG)

Table 4-3: Type of variable in the database.

Variable	Type
Weekly Trips	Continuous variables
Hours Open	
Total Employees	
Total Area	
Warehouse Area	
Parking Minutes	
Weekly Kilograms	
Establishment size	Categorical
Has Warehouse?	Binary
Has Parking?	

4.2.1 Numeric variables

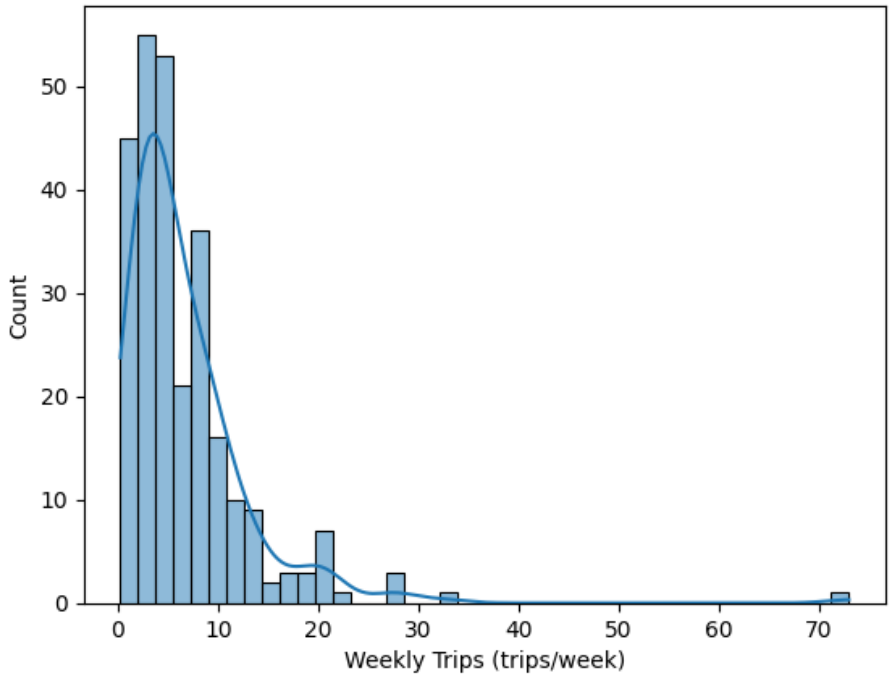
In terms of quantitative variables, continuous variables are those that have an infinite number of possible values within a given range. In simpler terms, a continuous variable can assume any numeric value within a defined interval. These variables are distinguished by having a well-defined order, meaningful intervals between values, and a measurement method other than counting. (Washington et al., 2020)

To forecast freight attraction to the FS establishments, the analysis's selection of "weekly trips" as the dependent variable is crucial. "Weekly Trips" quantifies the number of trips made by customers or deliveries to establishments over a week. The dependent variable, which is the main subject of the study, is affected by changes in other variables, or independent variables, and variations in this variable can be predicted or explained. In this instance, the goal is to comprehend the aspects or characteristics of the FS sector that affect how many weekly trips they produce. The "Hours Open," "Total Employees," "Total Area," "Warehouse Area," "Parking Minutes," "Weekly Kilograms," variables may be among these influencing factors are the independent variables. And a descriptive analysis of this variables is shown in the **Table 4-4**.

The factors influencing customer or delivery activity within the sector can be better understood by analyzing the independent variables and the dependent variable ("Weekly Trips"). This first analysis could be observed in the Figure 4-2 and Figure 4-3 show below.

One can see that the number of trips made by the FS sector each week varies between less than one (once per month) and more than 70, but the majority of trips made for freight transportation are between 0 and 10 per week. The data are distributed according to a gamma distribution.

Figure 4-2: Histogram of weekly freight trips to FS establishments at MMA.



A thorough analysis of continuous variables is conducted, with an emphasis on comprehending the dynamics and traits of the establishments within the sector under study. While "Hours Open" reflects the accessibility of the businesses. “Total Area” and “Total Employees” provide information on the size of the workforce and the amount of space occupied, respectively. "Warehouse Area" identifies the precise area set aside for warehousing. “Weekly Kilograms” measures material handling, while "Parking Minutes" examines how freight trips behave in parking lots to supply the establishment.

Table 4-4: First descriptive analysis of FS.

	Weekly Trips (trips/week)	Hours Open (h)	Total Employees	Total Area (m^2)	Warehouse Area (m^2)	Parking Minutes (min)	Weekly Kilograms (kg/week)
count	266.00	266.00	266.00	266.00	266.00	266.00	266.00
mean	6.69	12.39	3.65	79.28	34.54	11.51	644.78
std	6.74	4.33	4.15	132.95	73.09	6.54	1640.97
min	0.25	4.00	1.00	4.00	0.00	1.00	1.10
25%	3.00	10.00	2.00	25.00	0.00	7.50	117.75
50%	5.00	12.00	3.00	45.00	16.00	10.00	255.90
75%	8.75	14.00	4.00	80.00	48.00	15.00	532.69
max	73.00	24.00	42.00	1450.00	1000.00	36.67	17610.00

According to the analysis, these FS establishments make an average of 6.69 trips per week; some businesses report as few as 0.25 trips; this value could be explained as in a month only on freight trips, while others report as many as 73 trips, work a total of 12.39 hours per day, and employ roughly 3.65 people. The total area of the establishment accounts for a mean of 80 square meters and has a high variation with one of the highest standard deviations of 132.95 compared to the other variables. While the warehouse space has an average of 34.54 units, the total area varies greatly. Businesses also typically allot 11.51 minutes per trip for parking, with a weekly transportation volume of 644.78 kilograms on average.

- **Outliers**

Obtaining this relevant information, it important analyze through the methods described in the chapter 3.1.2 the outliers presented in each numeric variable of the sample. These are the observations of outliers consolidated in **Table 4-5** and graphically of the outliers with the IQR method (**Figure 4-4**).

Table 4-5: Number of outliers with each method.

Variable	IQR	%	Z Score	%
Hours Open	19	7,145	0	0
Total Employees	20	7,52	3	1,13
Total Area	23	8,65	6	2,26
Warehouse Area	11	4,14	2	0,75
Parking Minutes	13	4,89	3	1,13
Weekly Trips	16	6,02	5	1,88
Weekly Kilograms	24	9,02	5	1,88

The IQR method, which has the highest percentage of outliers, will be used to determine the ranges, and the values that fall outside of those ranges will be replaced by the Windsorized process explained in the 3.1.3 section.

Figure 4-3: Box plot with outliers by IQR method of weekly freight trips.

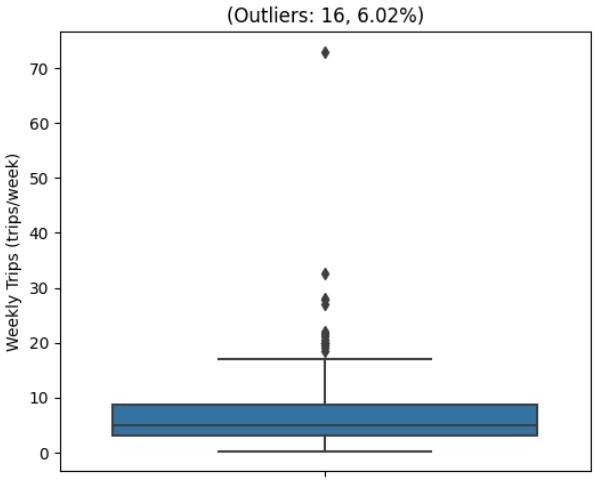
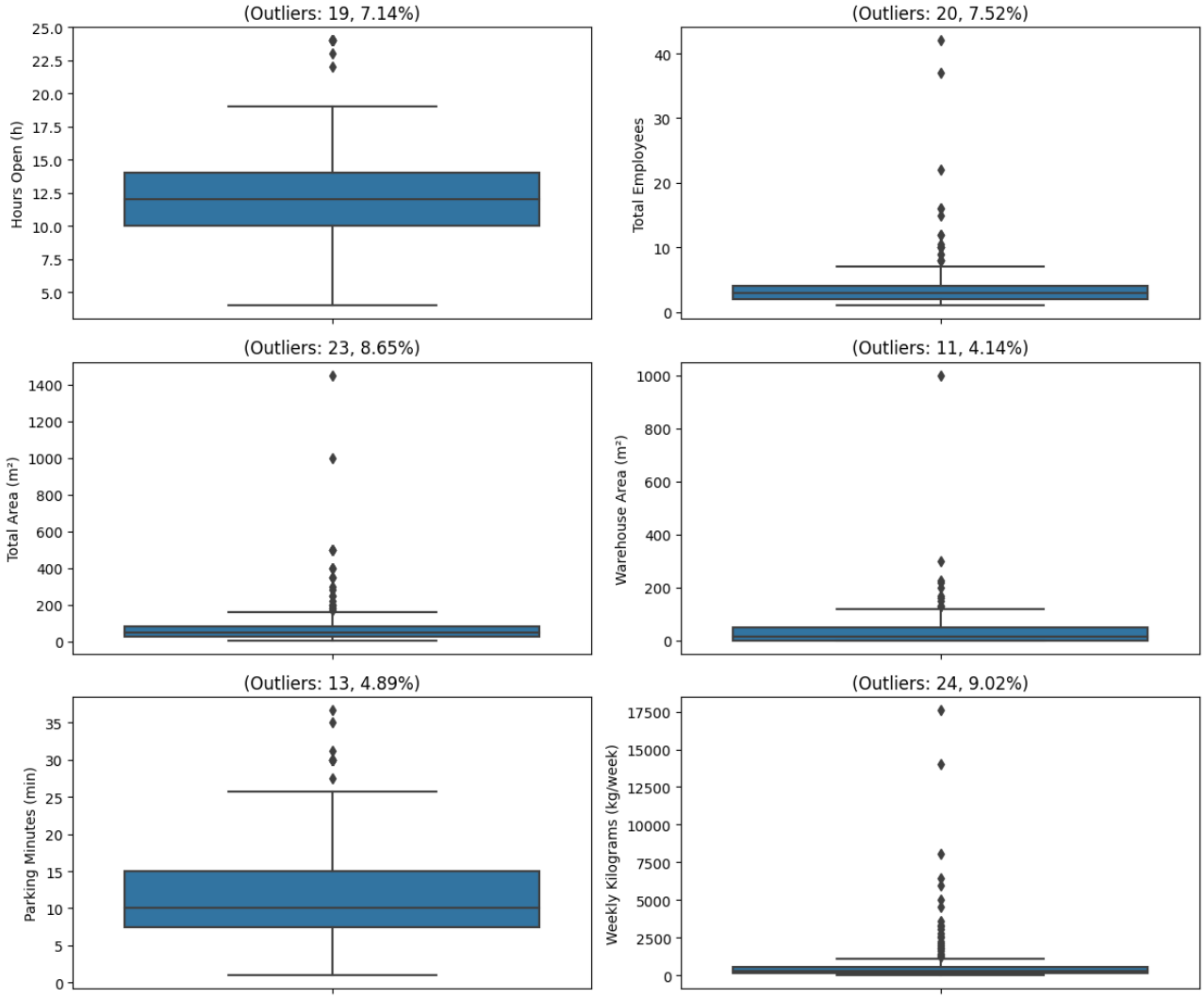


Figure 4-4: Box plot with outliers by IQR method of the independent's variables.



- **New dataset**

Using the method of Winsorization to replace the outliers explained in section 3.1.3, A new data set is obtained to make the statistical models and the descriptive statistics could be appreciate in the **Table 4-6**.

Table 4-6: Final descriptive analysis of FS.

	Weekly Trips (trips/week)	Hours Open (h)	Total Employees	Total Area (m^2)	Warehouse Area (m^2)	Parking Minutes (min)	Weekly Kilograms (kg/week)
count	266	266	266	266	266	266	266
mean	6,59	12,39	3,51	75,01	31,69	11,51	562,39
std	5,92	4,33	3,10	99,84	44,46	6,54	1016,71
min	0,25	4,00	1,00	4,00	0,	1	1,10
25%	3,00	10,00	2,00	25,00	0	7,5	117,75
50%	5,00	12,00	3,00	45,00	16	10	255,90
75%	8,75	14,00	4,00	80,00	48	15	532,69
max	46,94	24,00	21,00	666,25	271,88	36,67	6044,55

Significant changes in the descriptive statistics of several variables are observed after applying a Winsorization process to the dataset. First, the mean value of the variable 'Weekly Trips' falls slightly from 6.69 to 6.59, indicating a decrease in the average number of weekly trips. The standard deviation falls from 6.74 to 5.92, indicating less variability in the trip count. The data for this variable is distributed according to a Gamma distribution, with the majority of observations falling between zero and ten weekly freight trips. This distribution could be seeing in The 'Total Employees,' the mean decreases from 3.65 to 3.51, and the standard deviation decreases from 4.15 to 3.10, indicating a more concentrated distribution of employee counts.

The 'Total Area' and 'Warehouse Area' variables also show reductions in their means and standard deviations, reflecting more consistent values after Winsorization. 'Parking Minutes' remains relatively stable.

However, the most noticeable change occurs in the 'Weekly Kilograms' variable, where the mean decreases from 644.78 to 562.39, and the standard deviation reduces from 1640.97 to 1016.71. This suggests that Winsorization leads to a more tightly clustered distribution of weekly kilograms.

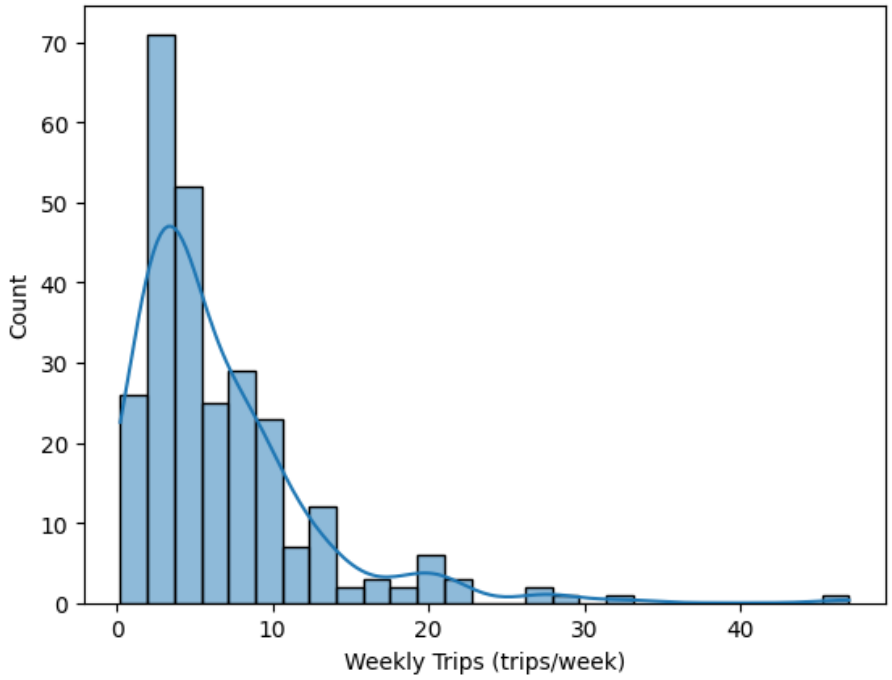
Figure 4-5.

The 'Total Employees,' the mean decreases from 3.65 to 3.51, and the standard deviation decreases from 4.15 to 3.10, indicating a more concentrated distribution of employee counts.

The 'Total Area' and 'Warehouse Area' variables also show reductions in their means and standard deviations, reflecting more consistent values after Winsorization. 'Parking Minutes' remains relatively stable.

However, the most noticeable change occurs in the 'Weekly Kilograms' variable, where the mean decreases from 644.78 to 562.39, and the standard deviation reduces from 1640.97 to 1016.71. This suggests that Winsorization leads to a more tightly clustered distribution of weekly kilograms.

Figure 4-5: Histograms of dependent variable Windsorized.

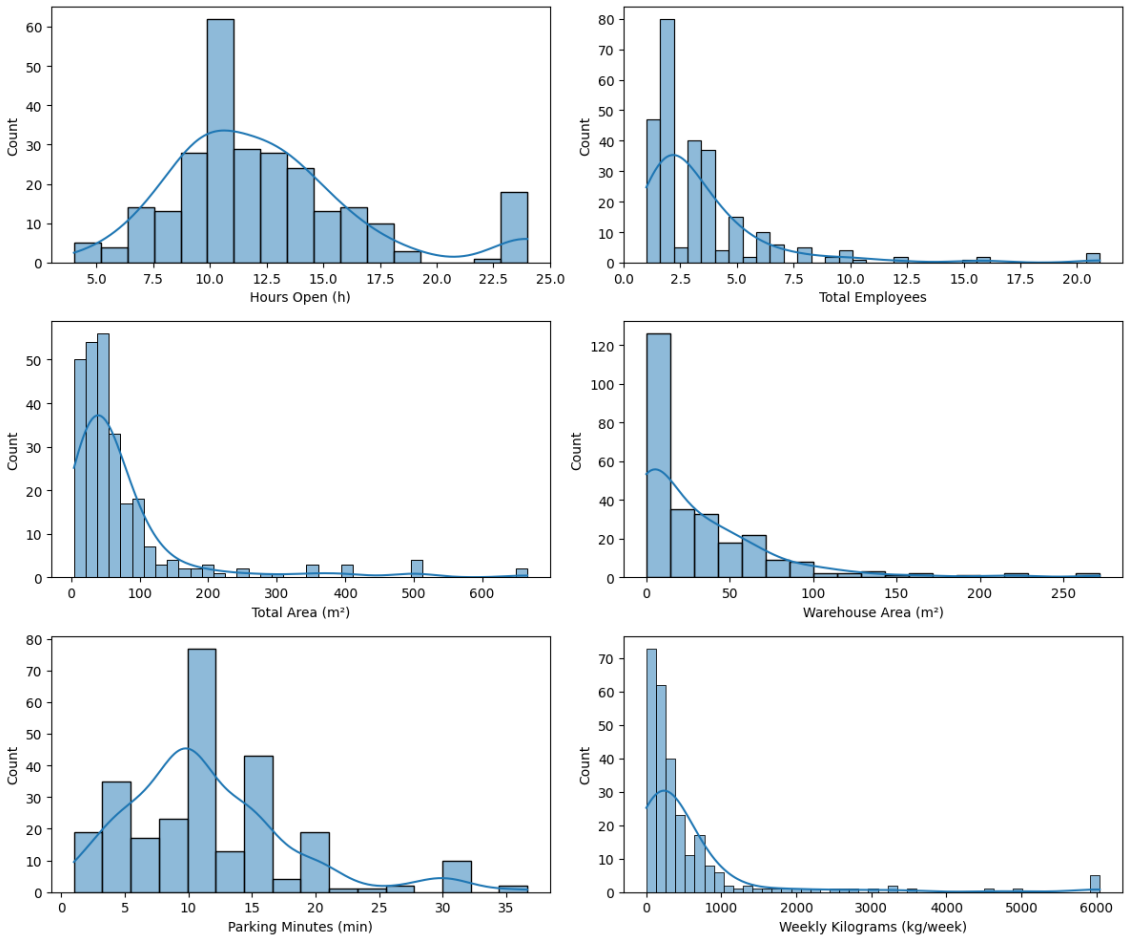


By minimizing the effect of outliers on the descriptive statistics, the dataset created by the Winsorization process has fewer extreme values and lower variability, which may make it more suitable for subsequent analyses.

The variables Total Employees, Total Area, Warehouse Area, and Weekly Kilograms all exhibit the Gamma distribution. The distribution of the data for the variables Parking time and hour opened, which have normal distributions, can be seen in **Figure 4-6**.

The dataset reveals valuable insights into several key metrics. Weekly trips exhibit a wide range, with an average of approximately 6.59 trips and a standard deviation of 5.92. Business hours average around 12.39 hours per day, with moderate variability and a range from 4 to 24 hours. Workforce sizes are diverse, with an average of 3.51 employees and a standard deviation of 3.10, ranging from 1 to 21 employees. Total and warehouse areas vary considerably, with typical values of 75.01 and 31.69 square meters, respectively, and substantial standard deviations of 44.46 and 99.84. Parking times fluctuate, with an average of 11.51 minutes and a standard deviation of 6.54 minutes, spanning from 1 to 36.67 minutes. Weekly kilograms reflect significant variability, averaging 562.39 kilograms with a substantial standard deviation of 1016.71 and a wide range from 1.10 to 6044.55 kilograms. These statistics provide a comprehensive overview of the dataset, capturing the diversity and dispersion of the included metrics.

Figure 4-6: Histograms of independent variables Windsorized.



4.2.2 Correlation analysis

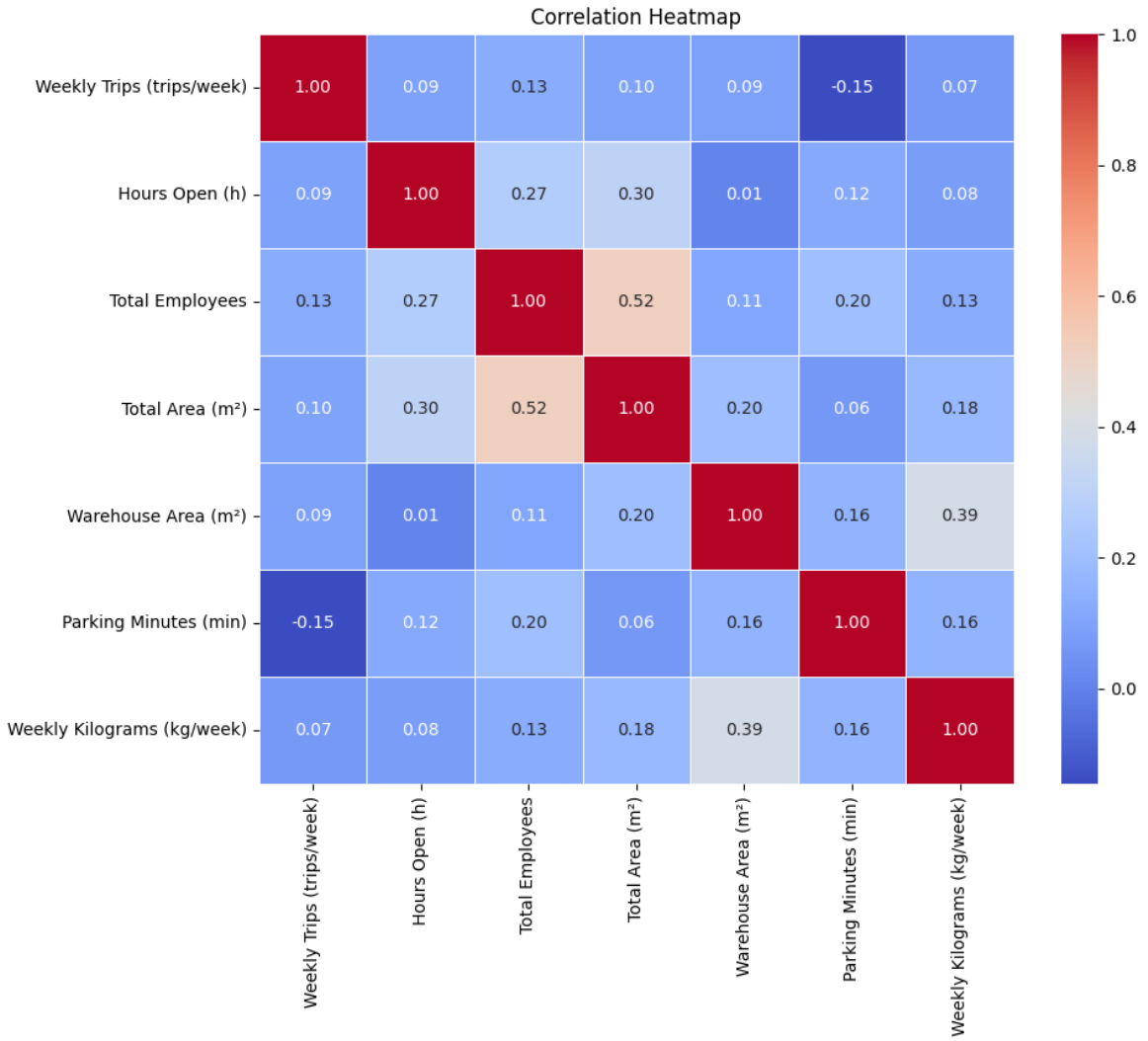
The correlation matrix was made following the process described in the 3.1.4 section and could be seen in the **Figure 4-7**. This provides insightful information about the interrelationships between the variables in the dataset. From weak to moderate, the correlations range in strength. One interesting finding is the correlation between the two variables, "Total Employees" and "Total Area (m2)," which is 0.52, indicating a moderately strong positive relationship. This suggests that bigger businesses with bigger spaces typically employ more people, which is the anticipated outcome and gives modeling insights so that these variables are not used in the same model.

The variable "Parking Minutes (min)" shows a negative correlation with the variable "Weekly Trips (trips/week)" of -0.15. Although it would seem logical to assume that more weekly trips would be associated with longer parking times, this negative correlation points in the opposite direction.

Furthermore, "Weekly Kilograms (kg/week)" exhibits a reasonably strong positive correlation of 0.18 with "Total Area (m2)," indicating that larger establishments with more total area tend to handle a greater weekly weight of goods, making them a good candidate for the modeling process. Additionally, there is a strong positive correlation between the variables "Weekly Kilograms (kg/week)" and "Warehouse Area (m2)" of 0.39, indicating that larger warehouse spaces are recommended for use in the models as well.

All of the variables, in general, have lower correlations and are suitable for modeling.

Figure 4-7: Correlation Heat Map of corrected numeric variables.



4.3 Modelling output

The MMA conducted two cargo surveys that provided the data for this study, which was explained in detail in the previous chapters. Finding information about the behavior traits of FS establishments is the goal of the models.
For modeling purposes, the models were made with the numeric and continuous variables present in the database. These variables are explained and analyzed in detail in Section 4.2.1.

Ordinary least squares regression was used to estimate linear and non-linear models, as explained in sections 3.1.5 and 3.1.6.. The validation of the models was made due to the information in Section 3.1.7. All possible combinations of models, like the Lin-Lin, Lin-Log, Log-Log, and Log-Lin models, were created for the purposes of this study.

Employer-based and total area-based modeling techniques were used to create models for the findings in the literature and the analysis's findings. It was found that for the type of establishment regarding FS, the total area and the number of employees have the biggest effects on weekly cargo trips.

The conceptual validation of the existence or absence of a constant in the model is only considering the variables 'Total Area (m²)', 'Total employees' and 'Hours Open (h)' If one of these values is zero, it couldn't be possible to attract cargo.

Additionally, the categorical and binary variables that represent a few operational facets of the FS establishment (like the accessibility of parking and storage spaces and the time needed to load and unload cargo) were put to the test. But in each case, it was found that they were not significant for the models.

From a list of six variables, including the variable weekly trips, an algorithm was created to produce combinations of the variables required for modeling for both approaches. Total area and total employees were not included at the same time in any of the combinatorics generated. These combinations come in a range of sizes, from combinations of two variables to combinations of all six variables.

- The distribution of these combinations can be seen by counting them:
- 15 different size 2 combinations are produced by choosing pairs of variables.
- Triplet combinations of variables are available in 20 different ways.
- For sets of four variables, there are 15 different combinations.
- There are six different combinations for sets of five different variables.
- When choosing all six variables at once, there is only one possible combination.

The algorithm generates a full set of combinations consisting of 57 unique combinations per approach in the modeling; a total of 456 were tested in the process, each of which represents a different arrangement of predictor variables for modeling purposes.

A process of modelling with the programming language Python using the library of stats models was taking place. This is a Python module that offers functions for estimating a variety of statistical models, running statistical tests, and exploring statistical data. For each estimator, a comprehensive list of result statistics is available.

The projected models for weekly cargo trip attractions for FS are presented in the Table 4-7 that follow. The models' various functional configurations underwent testing, as shown in the same table.

The main criteria considered for model selection were the students t of each parameter and, most importantly, the Fisher test mentioned in Section 3.1.7. In the table below, the colors in the column regarding F-Test have a color scale from green to red, where green represents the highest values of this test and degrades to less strong and reddish colors, which represent fewer F values.

Table 4-7: Models summary.

Model	Variable	Coefficient	t-value	R2	F-value	RMSE
Lin-Lin	**Hours open**	0,488	17,066	0,522	291,300	6,110
Lin-Lin	**Total Employees**	1,160	12,730	0,380	162,000	6,970
Lin-Lin	Hours open	0,444	13,081	0,530	150,800	6,040
	Warehouse area	0,019	2,343			
Lin-Lin	Total Employees	0,411	9,308	0,529	150,563	6,048
	Hours open	0,283	2,285			
Lin-Lin	Hours open	0,382	8,279	0,535	103,000	5,999
	Total Employees	0,248	1,998			
	Warehouse area	0,017	2,063			
Lin-Lin	**Total Area**	0,035	9,401	0,247	88,38	7,66
Lin-Log	lnHours Open (h)	2,671	18,430	0,560	339,700	5,858
Lin-Log	lnWeekly Kilograms (kg/week)	1,167	18,057	0,550	326,100	5,925
Lin-Log	**lnTotal Employees**	5,161	16,193	0,495	262,200	6,274
Lin-Log	lnHours Open (h)	1,914	7,198	0,577	182,100	5,736
	lnTotal Employees	1,840	3,370			
Lin-Log	**lnTotal Area (m²)**	1,681	18,471	0,561	341,200	5,856
Lin-Log	lnHours Open (h)	1,310	2,175	0,567	175,400	5,799
	lnTotal Area (m²)	0,881	2,326			
Log-Log	**lnHours Open (h)**	**0,624**	**29,477**	**0,765**	**868,882**	**0,856**
Log-Log	lnWeekly Kilograms (kg/week)	0,273	28,417	0,752	87,500	0,879
Log-Log	lnWarehouse Area (m²)	0,434	24,664	0,782	608,300	0,831
Log-Log	**lnTotal Employees**	1,168	21,549	0,635	464,340	1,067
Log-Log	lnHours Open (h)	0,501	12,974	0,776	462,900	0,834
	lnTotal Employees	0,298	3,753			
Log-Log	**lnTotal Area (m²)**	**0,390**	**28,855**	**0,758**	**832,600**	**0,870**
Log-Log	lnHours Open (h)	0,370	4,236	0,772	451,900	0,841
	lnTotal Area (m²)	0,164	2,987			
Log-Lin	Hours Open (h)	0,113	25,291	0,706	639,600	0,958
Log-Lin	Hours Open (h)	0,098	14,291	0,714	332,500	0,943
	Total Employees	0,055	2,853			
Log-Lin	**Total Employees**	0,266	16,148	0,494	260,753	1,256
Log-Lin	**Total Area (m²)**	0,008	11,132	0,316	123,900	1,461
Log-Lin	Total Area (m²)	0,006	7,297	0,420	97,300	1,343
	Warehouse area	0,012	6,962			

4.3.1 Models selected.

The statistical models developed to forecast the weekly cargo freight trip attraction behavior of the MMA FS industry are listed in the Table 4-7. Various combinations of variables are represented by the models that are given, along with the corresponding statistics. Among the crucial considerations are the establishment's total area, number of employees, hours of operation, warehouse space, parking duration and weekly transported weight. For both linear and non-linear models, the models are chosen based on conceptual validity and statistical significance. Non-linear models perform well across a range of evaluation metrics, especially when using log-log transformations with surface area as the explanatory variable.

The Log-Log Model (Hours Open), Log-Log Model (Weekly Kilograms) and Log-Log Model (Warehouse Area) were the models with the highest statistical significance. Because not all restaurants have this capability, the size of the warehouse was not considered at the final stage of model selection. And because the purpose of this thesis is FTA and not FG, the final awards did not consider the weekly load that was only obtained in that model.

Moreover, to continue with the conceptual validity was chosen the models with the variable of total area and a model with the combination of hours open and total area.

After adding the BIAS correction to the Log-Log models, in the **Table 4-8** are displayed the final form of the models in Equation 4-1, Equation 4-2 and Equation 4-3.

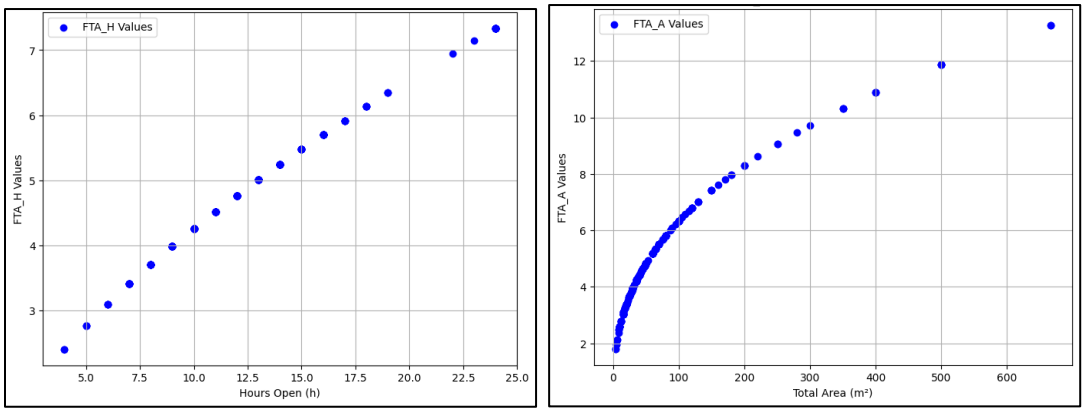
Table 4-8: Models selected.

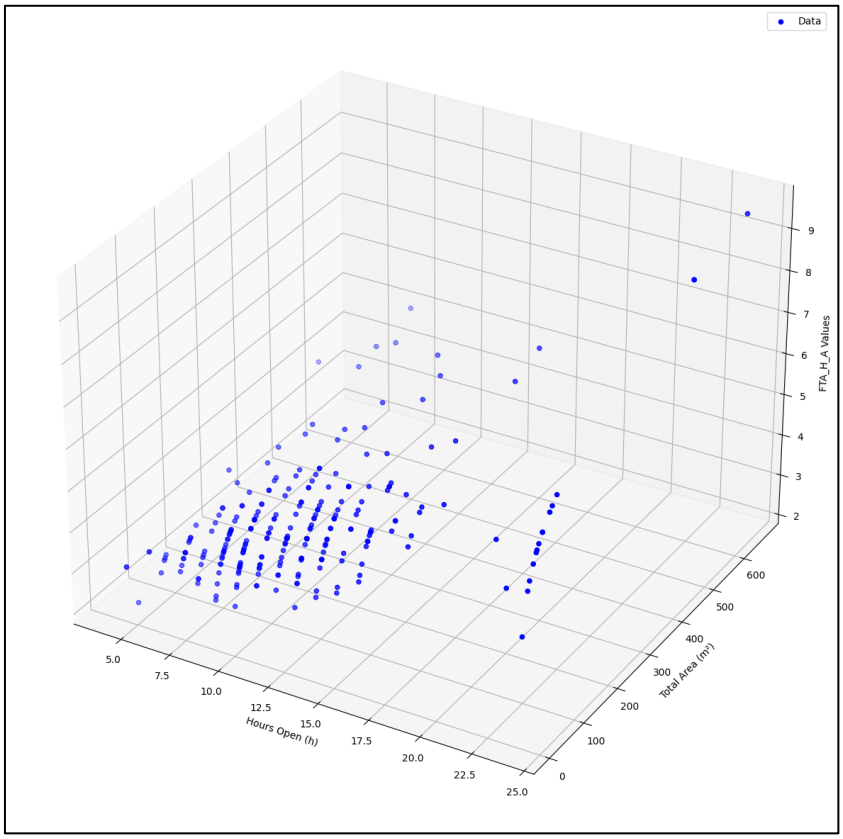
$$FTA_H(week) = Hours\ open^{(0,624)} * e^{(0.01)} \quad \text{Equation 4-1}$$

$$FTA_A(week) = Total\ area^{(0,390)} * e^{(0.05)} \quad \text{Equation 4-2}$$

$$FTA_H_A(week) = Hours\ open^{(0,370)} * Total\ area^{(0,164)} * e^{(0.007)} \quad \text{Equation 4-3}$$

Figure 4-8. FTA calculated with models.





5. Analysis and practical implications

To forecast the Freight Trip Attraction (FTA) in Food Services (FS) or restaurants to urban areas of developing economies, three final equations were obtained to model the relationship between Weekly Freight trips and variables such as hours opened and total area.. These equations, denoted as Equation 4-1, Equation 4-2 and Equation 4-3, represent different approaches to understanding the influence of specific variables on FTA.

According to the results of the statistical procedure, the best model was one where the number of hours open was the primary variable in the log-log form. The more hours they are open, the more sales they make and the more cargo they need. This result supports the hypothesis that longer business hours might encourage more trade and logistics activities and consequently, more freight trips.

The modeling process provides the best models when hour-opened variables are included, which is different from what is reported in the literature. Nevertheless, there are useful models that consider the variables of total area and numbers of employees. The employment-based models do not have the highest statistical significance. possibly because it is challenging to estimate the precise number of personnel in the FS industry. The employment base models behave poorly during the modeling process, maybe as a result of their frequent rotation and the likelihood that although the employment level may be the same, the hours worked at these businesses may differ greatly with the same number of employees in each journey.

There is importance to the establishment's total area, as related in the literature, because they may have more operational capacity, in this case, more tables to offer customers, and larger businesses may tend to draw more freight trips. The use of logarithmic functions suggests that scale economies may be at play.

The other variables evaluated were loading times, storage space, and parking spaces, but none of them achieved statistical significance. This demonstrates the significance of total area, hours open and total area in the cargo transportation industry, but more significantly, the variable hours open with regard to the model's operating hours take huge importance in the modeling process; it has conceptual significance as related to the FS sector.

The model that combines the total area and the operating hours shed light on the fact that extending business hours and expanding the size of a location both increase the number of freight trips to the establishment. Furthermore, the lower exponents in the error term compared to the equations from the other models imply that the combined effect of hours open and total area might be more gradual.

As the mean values range between 1 and 15 freight trips each week, the data obtained by the models are consistent with the original database, with the effect being bigger in the model based on total area. And the hours opened varies from 4 to 24 hours and the total area from 4 to 670 square meters. These equations therefore provide crucial insight into the dynamics of FTA in still-growing urban areas. The allure of freight trips can be significantly impacted by extended business hours, larger facility capacity, or a combination of the two.

The logistics companies may need to adopt operational guidelines that take these factors into account due to these variables of the organizations they work for. For better resource management and customer service, they might also need to alter their fleet or operating schedule.

For FS establishments, planning his freight trips based on these variables and calculating the budget and logistical standards within the organization may be helpful solutions. The same could be true for new companies looking to grow their current brands or launch new restaurants.

The FTA in urban areas of developing economies is a practical insight with broad implications for FS. These models could be used to estimate the number of freight trips in areas with a high density of food services, which would allow planners and freight policy controllers to implement the appropriate measures based on the specific circumstances of each zone. These insights guide policymakers and business executives in designing urban environments and transportation systems.

6. Conclusions and recommendations

6.1 Conclusions

- A novel finding of this study was the determination of the significance of the hours open in food service establishments, considering the behavior of this industry sector, where nearly all of the hours were spent performing the same task, which is to provide food to the customers, in contrast to other economic sectors. The conclusions in this regard have conceptual validity and may be used by policymakers, owners of restaurants, or those in the logistics industry who deal with this commercial category.

- The results for weekly trips are consistent with the literature found of similar studies in Latin America, but they also appear to be a little lower than economies like Brazil, the United States, and Europe. This is because developing economies behave differently when it comes to logistics operations. The development of this research is crucial in order to learn more about logistics operations in the current environment, which is one of its most significant contributions.

- The results offer useful information to decision-makers and business owners of logistics companies or food service establishments in developing economies. By regulating operating hours and facility sizes, businesses can boost management, enhance customer service, and forecast their weekly freight trips, all of which will contribute to the expansion of the economy. These findings also point to the potential for more precise urban planning, whether it be in terms of land use or the formulation of freight regulations or policies, in developing economies. This plan may lead to less congestion, fewer emissions, and increased economic competitiveness in residential areas, where food services are primarily located in urban areas.

- The study successfully identified key factors that predict Freight Trip Attraction (FTA) in the Food Services (FS) sector in urban areas of emerging economies. Three distinct FTA models were developed as a result of the research, each focusing on a different aspect to forecast weekly freight trips and highlighting the significance of facility size and extended operating hours. These models offer useful instruments for forecasting and improving FTA in comparable urban areas in emerging economies.

6.2 Recommendations

- When you use this model, consider that you cannot simply multiply this number by each establishment to determine the freight trip attraction. Due to the possibility that different trips may be made by the same companies, or the trips of various establishments could be made by the same truck, it must be careful not to overestimate the number of freight trips in a zone.

- Consider the variance in days and hours of the freight trips calculated in a week when planning the use of urban transportation. Examine the effects of weekday variations and seasonality on FTA. Understanding how FTA evolves can help forecasting and resource allocation become more accurate. Future research will need to look at these variables in relation to each context.

- Urban planners can make more environmentally friendly choices by weighing the social and environmental impacts of more freight travel. To gain a deeper understanding, future research might broaden the set of variables included in FTA modeling. by considering extra elements like population density, the standard of the road network, and land-use patterns or sustainability metrics.

References

Área Metropolitana del Valle de Aburrá. (2010). *Atlas Metropolitano*.

Area Metropolitana del Valle de Aburrá. (2018). *Estudio de transporte de carga en el Area Metropolitana del Valle de Aburra*. https://www.metropol.gov.co/movilidad/Documents/Estudio-de-transporte-de-carga-en-el-Area-Metropolitana-del-Valle-de-Aburra.pdf

Brettmo, A., & Sanchez-Diaz, I. (2021). Property owners as possible game changers for sustainable urban freight. *Research in Transportation Business and Management*. https://doi.org/10.1016/J.RTBM.2021.100745

Cámara de Comercio de Medellín. (2023). *Consulta de actividades económicas*. https://virtuales.camaramedellin.com.co/consultas/ciiu.html

de Oliveira, L. K., Nóbrega, R. A. de A., Ebias, D. G., & Corrêa, B. G. e. S. (2017). Analysis of freight trip generation model for food and beverage in Belo Horizonte (Brazil). *Region*, *4*(1), 17–30. https://doi.org/10.18335/REGION.V4I1.102

Gonzalez-Calderon, C. A., Sánchez-Díaz, I., Sarmiento-Ordosgoitia, I., & Holguín-Veras, J. (2018). Characterization and analysis of metropolitan freight patterns in Medellin, Colombia. *European Transport Research Review*, *10*(2). https://doi.org/10.1186/s12544-018-0290-z

Haven, A. (2015). Consumer Behavior and the Growth of the Fast Food Industry in a Small Emerging Country. *International Journal of Advances in Management and Economics Allahar Haven*, *4*(3), 180–189. www.managementjournal.Info

Holguín-Veras, J., Campbell, S., González-Calderón, C. A., Ramírez-Ríos, D., Kalahasthi, L., Aros-Vera, F., Browne, M., & Sanchez-Diaz, I. (2018). Importance and Potential Applications of Freight and Service Activity Models. In *City Logistics 1* (pp. 45–63). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119425519.CH3

Holguín-Veras, J., Jaller, M., Sánchez-Díaz, I., Campbell, S., & Lawson, C. T. (2014). Freight Generation and Freight Trip Generation Models. In *Modelling*

Freight Transport (pp. 43–63). Elsevier. https://doi.org/10.1016/B978-0-12-410400-6.00003-3

Holguín-Veras, J., Lawson, Catherine., Wang, C., Jaller, M., González-Calderón, C., Campbell, S., Kalahashti, L., Wojtowicz, J., Ramirez-Ríos, D., National Academies of Sciences, E., National Cooperative Freight Research Program, American Association of State Highway and Transportation Officials, & United States. Federal Highway Administration. (2017). *NCFRP Research Report 37: Using commodity flow survey microdata and other establishment data to estimate the generation of freight, freight trips, and service trips - Guidebook* (37th ed.). Transportation Research Board. https://trid.trb.org/view/1459616

Holguín-Veras, J., Wang, C., Browne, M., Hodge, S. D., & Wojtowicz, J. (2014). The New York City Off-hour Delivery Project: Lessons for City Logistics. *Procedia - Social and Behavioral Sciences*, *125*, 36–48. https://doi.org/10.1016/J.SBSPRO.2014.01.1454

Kin, B., Ambra, T., Verlinde, S., & Macharis, C. (2018). Tackling Fragmented Last Mile Deliveries to Nanostores by Utilizing Spare Transportation Capacity—A Simulation Study. *Sustainability*, *10*(3), 653. https://doi.org/10.3390/su10030653

Klumpp, M., & Ruiner, C. (2018). Digitalization and Work Organization in New Urban Food Delivery Systems. *International Journal on Food System Dynamics*, *9*(5), 399–408. https://doi.org/10.18461/IJFSD.V9I5.952

Kwak, S. K., & Kim, J. H. (2017). Statistical data preparation: Management of missing values and outliers. In *Korean Journal of Anesthesiology* (Vol. 70, Issue 4, pp. 407–411). Korean Society of Anesthesiologists. https://doi.org/10.4097/kjae.2017.70.4.407

López-Roldán, P., & Fachelli, S. (2017). El diseño de la muestra. *Metodología de La Investigación Social Cuantitativa*. https://ddd.uab.cat/record/185163

Ministerio de Transporte de Colombia. (2023, June 5). *Registro Nacional de Despacho de Carga - RNDC. Data accesed 05/06/2023*. https://rndc.mintransporte.gov.co/

Newman, M. C. (1993). REGRESSION ANALYSIS OF STATISTICAL BIAS AND ITS CORRECTION LOG-TRANSFORMED DATA. In *Environmental Toxicology and Chemistry* (Vol. 12).

Ortúzar, J. de D., & Willumsen, L. (2011). *Modelling transport*. https://books.google.es/books?hl=es&lr=&id=qWa5MyS4CiwC&oi=fnd&pg=PT7&dq=modelling+transport+ortuzar+willumsen&ots=tw_nZn7EYy&sig=pEfJx_zh7qUxazOxte_Z_6WZB80

Patil, G. R., Thadoju, S., Sahu, P. K., Mondal, A., & Bajpai, V. (2021). Data Collection and Modeling of Restaurants' Freight Trip Generation for Indian Cities. *Transportation in Developing Economies*, *7*(1). https://doi.org/10.1007/s40890-021-00114-7

Puente-Mejia, B., Palacios-Argüello, L., Suárez-Núñez, C., & Gonzalez-Feliu, J. (2020). Freight trip generation modeling and data collection processes in Latin American cities. Modeling framework for Quito and generalization issues. *Transportation Research Part A: Policy and Practice*, *132*, 226–241. https://doi.org/10.1016/J.TRA.2019.10.013

Sánchez-Díaz, I. (2017). Modeling urban freight generation: A study of commercial establishments' freight needs. *Transportation Research Part A: Policy and Practice*, *102*, 3–17. https://doi.org/10.1016/J.TRA.2016.06.035

Sánchez-Díaz, I. (2018). Potential of Implementing Urban Freight Strategies in the Accommodation and Food Services Sector. *Https://Doi.Org/10.1177/0361198118796926*, *2672*(9), 194–203. https://doi.org/10.1177/0361198118796926

Secretaría de Salud Medellin. (2023). *En Medellín ya no hay restaurantes a manteles. Data accessed 12/05/2023*. https://vivirenelpoblado.com/restaurantes-a-manteles-en-medellin/

Tasnim, Z. (2020). Disruption in Global Food Supply Chain (FSCs) Due to Covid-19 Pandemic and Impact of Digitalization Through Block Chain Technology in FSCs Management. *European Journal of Business and Management*. https://doi.org/10.7176/ejbm/12-17-08

Washington, S., Karlaftis, M., Mannering, F., & Anastasopoulos, P. (2020). *Statistical and Econometric Methods for Transportation Data Analysis*. https://www.crcpress.com/go/ids

Weeks, J. R. (2010). Defining urban areas. In *Remote Sensing and Digital Image Processing* (Vol. 10, pp. 33–45). Springer International Publishing. https://doi.org/10.1007/978-1-4020-4385-7_3

Wilmsmeier, G., Johansson, L., & Jallow, D. (2015). *The complex urban freight puzzle*. https://repositorio.cepal.org/handle/11362/38960