DISSERTATION ON

**A Comprehensive Study of Socio-Economic Factors, Migration Patterns And Business Sustainability of Street Vendors of Mumbai city.**

SUBMITTED

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTERS OF SCIENCE (STATISTICS)

UNIVERSITY OF MUMBAI DEPARTMENT OF STATISTICS SUBMITTED BY

Sana Kazi Rutika Khaire

Mrunali patil Ekta Shirsulla

During the Academic Year 2024 – 2025 Under the Guidance of

DR. SANTOSH P. GITE

MISS. PURVAJA SAWANT



**University of Mumbai**

### DEPARTMENT OF STATISTICS

#### Vidyanagari, Mumbai-400 098



**University of Mumbai**

## DEPARTMENT OF STATISTICS

#### Vidyanagari, Mumbai-400 098

**CERTIFICATION**

This is to certify that the following group of students of MSc Part-II has successfully completed the project “A Comprehensive Study of Socio-Economic Factors, Migration Patterns, and Business Sustainability of Street Vendors of Mumbai city.” during the academic year 2023-2024.

## Team members:

Sana Kazi

Rutika Khaire

Mrunali Patil

Ekta Shirsulla

### PROJECT MENTOR HEAD OF THE DEPARTMENT

Dr. Santosh P. Gite Dr. Santosh P. Gite

Miss. Purvaja Sawant

## Acknowledgement

We would like to extend our sincere thanks to our mentor **Dr. Santosh P. Gite sir** and **Miss. Purvaja Sawant** mam for their valuable guidance, support, and valuable insights throughout the duration of this project. Additionally, we acknowledge the contributions of Department of Statistics, University of Mumbai for their assistance, resources and support. We would also like to thank our families and friends for their unwavering support, encouragement and understanding. Without the collective efforts and support of all these individuals, this project would not have been possible.

Lastly, we would like to extend our sincere appreciation and gratitude to our fellow group members who dedicated their time, efforts, and expertise to the successful completion of this project.

## INDEX

|  |  |  |
| --- | --- | --- |
| Serial No. | Topics | Pg. No. |
| 2. | Introduction | 1-2 |
| 3. | Literature Review | 3-8 |
| 4. | Problem Statement | 9 |
| 5. | Objectives | 10 |
| 6. | Methodology | 11-14 |
| 7. | Statistical Techniques | 15 |
| 8. | Statistical Analysis | 16 |
| Exploratory Data Analysis (EDA) | 17-21 |
| Multiple Linear Regression | 22-33 |
| Cluster Analysis | 34-38 |
| Poisson Regression | 39-45 |
| 9. | Overall Conclusion | 46 |
| 10. | Suggestion | 47 |
| 11. | Project Codes | 48-54 |
| 12. | Questionnaire | 55-57 |
| 13. | References | 58-59 |

**Introduction**

The concept of the informal sector as we understand it today was largely popularized by Keith Hart, a British anthropologist, in the early 1970s while conducting a study in Accra, Ghana. His research focused on the labor market in developing countries, particularly in urban areas, where many workers were engaged in small-scale, unregistered, or self-employed activities outside of the formal economy. The terminology of ‘informal sector’ gained widespread acceptance after the International Labour Organization (ILO) used it to analyse economic activities in Kenya for an ILO Employment mission in 1972 conducted by a number of specialists. The ILO team noted that the workers in informal sector were unrecognizedunrecorded, unprotected and unregulated.

The **NCEUS (National Commission for Enterprises in the Unorganised Sector)** report (2007) defines the unorganized sector in India as a segment of the economy that includes small, unregistered, privately owned businesses (often run by individuals or families) with fewer than ten workers where most workers are casual laborers, earning low wages, and working in irregular jobs with little investment. According to **National Policy for Urban Street Vendors, 2009** A Street Vendor is defined as 'a person who offers goods or services for sale to the public in a street without having a permanent built-up structure’. They may operate fromstationary positions on pavements or other public/private spaces, they may be mobile, carrying their merchandise on push carts or in baskets.

In other words, street vendors are those who are unable to get a job and possess low educational qualification and skills. They try to live life by selling fresh vegetables, prepared food items, clothes, crafts, mobile accessories and many more. The poorer sections too are able to procure their basic necessities mainly through street vendors,

as the goods sold are cheap. Lower income groups spend a higher proportion of their income in making purchases from street vendors mainly because their goods are cheap and thus affordable. In addition, they help many small-scale industries to flourish by marketing the products that they manufacture. Thus, they help to sustain the urban economy to a great extent in terms of generation of employment and income, and provision of services to others.

Street vendors usually operates in high traffic areas where they can attract customers some common locations where street vendors setup their stalls includes Bazaars, Railway Stations, Bus stops, Pavements, Traffic signals, Roadside Junctions, Public Parks, Tourists spots, near school and college campus, office areas, religious places, slums and residential areas, fairs. These locations are chosen based on customer demand and accessibility.It has been estimated that around 30 per cent of the Mumbai workforce buys at least one meal a day from vendors (Bhowmik, 2001). Therefore, street vendors can be said to signify a viable solution to some of the problems of the poverty-stricken urban dwellers.

Transport hubs in Mumbai such as bus stations, train stations, or metro stops are prime locations for vendors who want to capture a broad range of customers. Commuters are often in a hurry and looking for quick, affordable items like snacks, beverages, or daily essentials. These transport hubs attract a variety of people, including office workers, students, tourists, and daily wage earners. They are not only pivotal in connecting the city with the rest of the country but also serve as centers of commerce and daily activity. These hubs act as gathering points for millions of people – office workers, daily commuters, tourists, and business travelers – all of whom rely on convenient access to food, snacks, beverages, and various goods while on the move.Many people migrate from villages and small towns to cities in search of work due to poverty and lack of jobs. However, they often have low skills and struggle to find well-paying jobs in the formal sector, where opportunities are also shrinking. As a result, many turn to the informal sector for survival.

Street vending becomes a common livelihood choice for the urban poor because it requires little investment and less skills. Many vendors are migrants who could not find other jobs. Though earnings are low, it provides a way to sustain themselves.

**Literature Review**

**1. Mika Febriani and Eni Setyowati (13)**

This study explores key factors influencing the income of street vendors in the informal sector around the Karangpandan Bus Terminal. Using data from 30 respondents and applying Ordinary Least Square (OLS) regression, the research identifies capital, working hours, and length of business as significant factors positively impacting income. Conversely, age and education level show no significant effect. The findings highlight that adequate capital enables better stock and operational management, while extended working hours and business longevity enhance customer reach and market understanding. These results underline the importance of supporting street vendors with financial and operational resources to maximize their income potential..

**2. Dr. A. Priya (6)**

According to this study, Street vending is an important part of the informal economy, helping many poor uncertainty and marginalized people earn a living. However, street vendors face many problems, such as legal, financial struggles, and harassment from authorities. Studies have shown that unclear laws and weak policies make vendors vulnerable to eviction and exploitation (Bhowmik, 2005). The study on street vendors in Coimbatore confirms these challenges, highlighting issues like bribery, unstable income, and difficulty in getting vending spaces. Urban planning often ignores street vendors, forcing them to operate in unsafe or overcrowded areas (Anjaria, 2006; Schindler, 2014). Health risks, long working hours, and lack of hygiene facilities also add to their struggles (Roever & Skinner, 2016). Although laws like the Street Vendors Act (2014) exist to protect them, poor implementation keeps vendors at risk (Bhowmik & Saha, 2012). This study shows that better policies and stronger government support are needed to improve the lives of street vendors.

**3. Tikam Singh (18)** :

This paper defines street vendor as individuals from the most marginalized, poor and vulnerable sectors of the urban informal sector. This paper focuses on the fact that providing vending zone in the city is the only solution to create facility and safety to street vendors. The research was based on a qualitative approach. The main aim of this research is to provide guidance and awareness to street vendors about their rights. This study has found that there are 2 main reasons for growth in street vending. Firstly, the push factor which is migration from rural areas in search of better employment opportunities and secondly the pull factors which has forced workers to join the informal sector. These are workers who were employed in the formal sector.

**4. Asema Siddiqui and shri Saratchandra Patra, (2)**

It examines the factors influencing the daily income of street vendors in Mumbai. The study identifies age, hours of work, and years since migration as variables affecting income, using a descriptive research design and data collected through questionnaires from 30 street vendors. Statistical analysis reveals that while age does not significantly impact income, longer working hours and years since migration positively influence earnings. The findings highlight the importance of the informal sector for migrants in urban areas, emphasizing the need for policies that recognize the contributions of street vendors and address their challenges, including lack of legal recognition and urban congestion.

**5. Dr.Sumanta Bhattacharya,Debashis sen and Bhacneet kaur Sachdev (8)**

According to this paper, Street vendors play a crucial role in urban economies by providing affordable goods and services. However, they face legal and economic challenges, including lack of recognition and exploitation. The COVID-19 pandemic worsened their situation, leading to financial distress. Street vending is a major source of income for many, yet policies often fail to support them. The Street Vendors Act, 2014, was introduced to protect their rights, but implementation remains weak. Government schemes like PM SVANidhi provided financial aid, but many vendors struggled to access benefits.

Public perception of street vendors is mixed, with concerns about congestion and hygiene. However, they contribute to food security and local economies. A balanced approach is needed, including designated vending zones and better infrastructure. To support street vendors, policies should focus on legal recognition, improved sanitation, and financial inclusion. Digital platforms and cooperative models can enhance their resilience. Strengthening protections and creating inclusive policies will ensure their long-term sustainability.

**6. Ms. Thimmaiah Jyothsna (14)**

According to this study Street vending is a vital source of income for the urban poor in India, but vendors face legal insecurity, financial instability, and harassment from authorities. Many operate without licenses, making them vulnerable toeviction and extortion**.** They also struggle withlow earnings, tough competition fromretail stores, and unsafe working conditions**.** Though the Street Vendors Act, 2014, was introduced to protect their rights, poor implementation leaves many unaware of their legal protections. Strengthening enforcement, providing financial support, and creating designated vending zones can help vendors work safely and sustain their livelihoods**.**

**7. Hilmi Martin (10)**

This paper defines Street food vendors play a vital role in the informal economy, especially in low-income areas, by providing affordable food and jobs. They rely on local networks, customertrust, and market knowledge instead of formal marketing. Key traits include low costs, risk-taking,and strong relationships with buyers**.** Despite challenges like legal issues and unstable earnings**,** their ability to adapt, create value, and use word-of-mouth marketing helps them survive. Understanding these strategies can help improve food businesses and support vendors in growing their livelihoods.

**8. Saihjpal Meenu, Saihjpal Ashish, Kapoor Shashi,(17)**

This paper defines, Street vendors are an important part of cities, providing affordable goods and jobs. Many migrate from villages for better opportunities but struggle due to lack of skills and legal protection. They face harassment, eviction, and financial instability, often paying bribes to keep their selling spots. Despite this, they support the economy by serving low-income customers and small businesses. The *Street Vendors Act, 2014* was created to protect them, but its implementation is weak. Better policies, proper vending zones, and financial support can help improve their lives and ensure their contributions are recognized.

**9. Manucha Taneyaa, Singh Kuljit, (12)**

This paper defines, The Street Vendors (Protection of Livelihood and Regulation of Street Vending) Act, 2014was introduced to protect street vendors from eviction and harassment while regulating their trade. It provides for vending certificates, Town Vending Committees (TVCs), and grievance redressal mechanisms**.** However, poor implementation, corruption, lack of clear property rights, and exclusion of railway vendors weaken its effectiveness. Many vendors still face police harassment, forced evictions, and lack of social security**.** To make the Act truly effective,strong enforcement, fair zoning, digital registration, and anti-corruption measuresare needed.

**10. P.B.Narendra Kiran and Dr.G.N.P.V.Babu (15)**

According to this study, Street vendors play an important role in cities by providing affordable goods and services. However, they work in an informal sector with little legal protection. Many face problems like financial struggles, lack of proper space, and frequent harassment from authorities. Even though the Indian government introduced the *Street Vendors Act, 2014* to protect them, its implementation remains weak.The informal sector has grown due to urban migration and a lack of formal jobs. Street vendors often suffer from uncertain workspaces, limited access to credit, and harsh working conditions. They work long hours in difficult environments and face constant threats of eviction or bribes from officials.

Despite these challenges, street vendors contribute significantly to local economies. They create jobs and serve as a distribution channel for small-scale businesses. However, government policies often ignore their importance, treating them as a problem rather than a valuable part of the city.To improve their situation, studies suggest providing vendors with proper licenses, designated vending zones, financial support, and better working conditions. Recognizing their role and integrating them into urban planning can create a more balanced and inclusive economy. More research and policy efforts are needed to protect and empower street vendors.

**11. Dr. D.Indira and Dr. Y. Ramakrishna prasad, (7)**

This paper defines Street vending is an important part of city life, providing jobs and affordable goods, especially in developing countries. However, vendors face many problems, such as legal restrictions, financial struggles, and constant harassment from authorities. street vendors contribute to the economy but often work under difficult conditions. In India, about 2% of the urban population is engaged in street vending, yet they lack proper support and recognition. Vendors use smart business strategies to survive, such as buying materials daily, using display boards for marketing, and adjusting prices based on customer bargaining. Most rely on daily earnings and informal credit to run their businesses. Despite the challenges, street vendors play a vital role in urban economies, and governments should create fair policies to support them rather than displace them.

**12. Parikshit Chakraborty and Samarpita Koley,(16)**

This paper defines Street vendors are an important part of India's urban economy, providing essential goods and self-employment. However, they face many challenges, including legal issues, financial instability, and harassment from authorities. Most vendors work long hours in poor conditions and lack access to credit and social security. Despite their contributions, they are often ignored in urban planning. The *Street Vendors Act, 2014* aims to protect them, but enforcement is weak. To improve their situation, the study suggests creating designated vending zones, offering financial support, and raising awareness about their rights. Better policies can help integrate them into the formal economy while ensuring fair opportunities.

**13. Erni Febrina Harahap, (9)**

This studyfocuses on the informal sector, specifically street vendors in Padang City, Indonesia. It examines the factors influencing the income of street vendors and how income impacts their welfare. Key variables such as capital, working hours, turnover, and operating costs are analyzed using descriptive and regression methods. The study finds that higher turnover, working capital, and costs positively impact income, while longer time in business does not necessarily increase earnings. It also highlights the role of income in improving welfare, as reflected in education, health, and consumption standards among street vendors' families. The findings emphasize the importance of supporting the informal sector through better policies, access to training, and resource allocation to enhance the economic and social well-being of this group.

**14. Bhowmik Sharit K. (4)**

This study is the extension of earlier study since it includes more cities. The cities are: Bhubaneswar, Bengaluru, Delhi, Hyderabad, Imphal, Indore, Jaipur, Lucknow, Mumbai and Patna. Some of the cities like Bhubaneswar, Bengaluru, Imphal, Mumbai and Patna were covered in the earlier study. The study was able to compare the state of street vendors then(2000) and now(2009). The author has taken 2 years to write the report due to personal reasons. This paper gives the reason of why street vendors are identified as unsuccessful or are known as a person who are unable to get regular jobs. This research shows that low-income people spend a major part of their income on buying goods from street vendors because the prices of their goods are affordable. If there are no street vendors in the cities, the plight of the urban poor will get worse. In this way a section of the urban poor who is working as street vendor indirectly helps the other poor section in their livelihood. The study throws light on the condition of the work culture of the street vendors in Mumbai and highlights the role of member-based organization or unions.

**15. Debdulal S (5)**

According to this study, Street vending has been one of the easiest ways to survive for working poor and is wide spread in the urban informal sector. This study has considered “decent work” indicators such as employment opportunities, social security and social dialogue. This study attempts to explore the conditions work within the new frame of ‘decent work’ by taking into consideration all the mentioned indicators. The study has used the 3 stage (cluster) sampling unit and it includes areas as Dharavi, Jogeshwari, Kandivali, Powai, Vile parle. The data was collected by using questionnaire and both personal and group interview methods were followed and sample size was 200. The study reveals that street vendors need social protection and are having poor working condition in terms of excessive working hours. All the policies are just present In paperwork but are not executed or implemented by few organizations. This paper concludes that there is decent work deficit in working life of street vendors in Mumbai.

**16. Bhowmik Sharit K. (3)**

This paper tries to understand the problems of street vendors and tries to find possible solutions to overcome these problems. The conclusion of the meeting which was held in Mumbai in 1998, which comprised of some of activists of NASVI was that there is a need of long term perspective on hawkers at national level. The cities selected were Patna, Calcutta, Bangalore, Mumbai, Ahmedabad, Imphal, Bhubaneswar and Indore. But the study could not proceed beyond initial stage in Indore and it was dropped. The study covers the 5 aspects such as social composition, Income, working conditions and employment and Harassment and Bribes. The conclusion of the research was that street vendors face many difficulties which makes their livelihood more harder. One solution given by the author to this is that while formulating urban plans it is necessary to take into account the right of hawkers to public space. In other words, all urban plans demarcate public spaces for specific purposes such as parks and gardens, educational institutions, hospitals etc. Hawking also needs to be included in this. These are usually the most convenient spots for consumers for example railway station.

**Problem Statement**

The studies conducted on street vendors of Mumbai region majorly focused on their working life, socioeconomic characteristics and the problems faced by them. These studies were mainly qualitative. Also did not include how income is influenced by socioeconomic factors, customer flow, their business structure and sustainability. This project aims to investigate dynamics of street vending across Mumbai’s transportation hubs examining how socio-economic factors influence income, the causes behind the migration of street vendors to Mumbai, and the patterns of customer flow. In our study as population is unknown, we will divide Mumbai region based on Mumbai railway network. By focusing on this, the study will employ stratified sampling to gain insights into the types of vendors operating in specified area of Mumbai region.

**Objectives**

1. To study the influence of socio-economic factors on average monthly income of street vendors.
2. To study the migration patterns among street vendors based on socio-economic factors and employment characteristics.

3. To model daily customer count based on influential vendor characteristics.

**Methodology**

To ensure a comprehensive and representative dataset, this study employed a **stratified sampling** survey method that systematically captures the diversity of street vendors across Mumbai's urban geography and business categories.

**1. Sampling Design**

**Population Considered**: The study focuses on the street vendor population within the **Mumbai Metropolitan Region**, particularly along the **Mumbai suburban railway network**, which serves as a proxy for key urban and commercial zones.

**Sampling Method**: **Stratified Sampling:** Stratified sampling, a probability-based technique, was used to divide the population into meaningful subgroups (strata) based on shared characteristics. This method enhances representativeness and reduces sampling bias by ensuring balanced data collection across different spatial, socio-economic, and occupational categories.

**2. Sampling Technique**

**The stratification process was carried out in four hierarchical stages:**

**First Stratum – Railway Zones:**

* **Western Line**
* **Central Line**
* **Harbour Line**
* **Trans Harbour Line**

**Second Stratum – Selected Stations**:

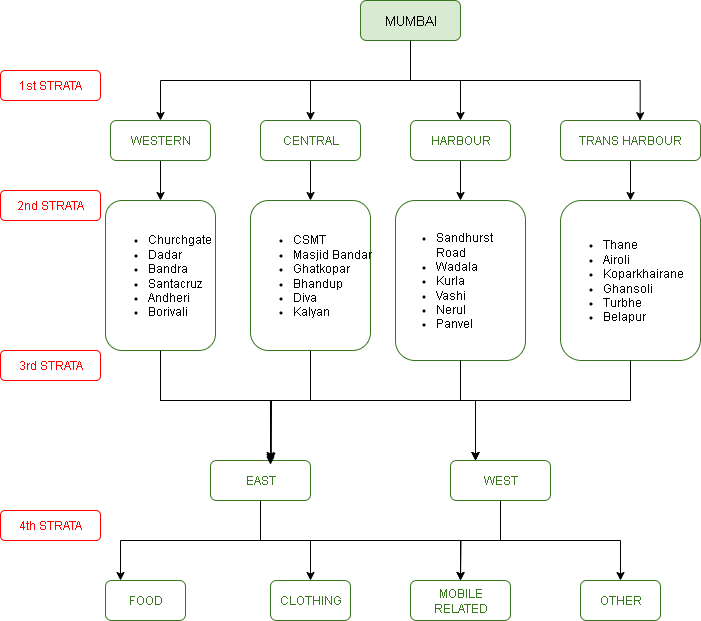
* **Western Line**: Churchgate, Santacruz, Dadar, Bandra, Andheri, Borivali
* **Central Line**: CST, Masjid Bandar, Ghatkopar, Bhandup, Diva, Kalyan
* **Harbour Line**: Sandhurst Road, Wadala, Kurla, Vashi, Nerul, Panvel
* **Trans Harbour Line**: Thane, Airoli, Koparkhairane, Ghansoli, Turbhe, Belapur

The 3rd stratum divides each station of the 2nd strata into East and West regions.

The 4th stratum categorizes street vendors based on their type of business:

* **Food Vendors**
* **Clothing Vendors**
* **Mobile Accessories Vendors**
* **Other Vendors**: Sellers like mehendi, perfumes, books, newspapers,etc.

This **multi-level stratified sampling design** enables the study to capture the heterogeneity of street vendors across geographical locations, operational zones, and business types within Mumbai.



**Sample size:**

Since the total population of street vendors was unavailable.The population of street vendors in the Mumbai region, according to the 2011 census, is approximately 250,000. However, no subsequent research has been conducted to provide valid evidence regarding the growth rate of street vendors. Therefore, to calculate the sample size, we cannot rely on estimated sampling methods. Instead, we will use the sample size formula for case where the population is unknown. The formula used is:

**n₀ =**

This formula is derived from studies *“****Importance of the Size of Sample and its Determination in the Context of Data Related to the Schools of Guwahati,”******published in the Bulletin of the Gauhati University Mathematics Association, Vol. 12, 2012, and “An Investigation on the Effect of Bias on Determination of Sample Size Based on Data Related to the Students of Schools of Guwahati,” published in the International Journal of Applied Mathematics and Statistical Sciences, Vol. 2, Issue 1, 2013****.*

Willaim Gemmell Cochran (1977) developed this formula to calculate a representative sample size for proportions:

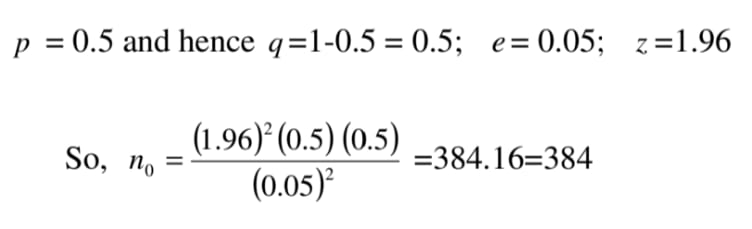
**n₀ =**

Where:

* **n₀** is the initial sample size for large populations (before any adjustments for finite populations).
* **Z** is the selected critical value for the desired confidence level (e.g., 1.96 for 95% confidence and 2.58 for 99% confidence).
* **p** is the estimated proportion of the attribute present in the population (use 0.5 if unknown, representing maximum variability).
* **q** is defined as **1-p**.
* **e** is the desired level of precision or margin of error. ( 5% - 20%)

To calculate the sample size for a large population with unknown variability, assuming maximum variability (**p = 0.5**), a 95% confidence level (**Z = 1.96**), and ±5% precision

(**e = 0.05**), the required sample size can be determined using the formula.



So, the sample size under consideration will be 384.

Therefore, no=384

Since the sample size for the project is 400, the distribution across the strata will be as follows:

* For the **1st stratum**, we will allocate the sample across the four lines of the **Mumbai suburban railway network**:  
  **384 / 4 = 96**
* For the **2nd stratum**, the sample for each railway line will be distributed across the six specified stations/regions:  
  **96 / 6 = 16**
* For the **3rd stratum**, the sample for each station/region will be divided between the east and west sections:  
  **16 / 2 = 8**
* This stratified distribution ensures balanced and representative sampling across all levels of the Mumbai District.

**STATISTICAL TECHNIQUES**

**1. EXPLORATORY DATA ANALYSIS**

We will the explore the data by using descriptive statistics, understanding nature of the variables, encoding the categorical variables, transforming the data, visualize the data. We will identify and correct errors, inconsistencies, and missing values.

**2. RERESSION ANALYSIS**

We will use this technique in the 1st objective where The dependent variable is Average monthly income, while the independent variables include socio-economic factors. Firstly we will check the assumption of regression techniques. We will use R software. Further we will use Stepwise regression to help identify key determinants affecting street vendors income.

**3. CLUSTER ANALYSIS**

We will use this technique to group the street vendors based on migration location, migration status, employment characteristics and the socio- economic factors. This will help us to identify the pattern of migration amongst street vendors .

**4. POISSON REGRESSION**

To model the number of customers served per day we will use Poisson Regression as the dependent variable under study is count and independent variables includes

**Location**: East and West

**Product Type**: Food, Clothing, Mobile accessories and Others.

**Software:**

R, Excel, Python

**Statistical Analysis**

**EXPLORATORY DATA ANALYSIS**

The dataset is based on primary data collected through a structured survey questionnaire targeting street vendors.

It includes a total of **20 variables**, capturing responses from individuals with varied demographic, economic, and occupational backgrounds.

The data contains a mix of:

* **Categorical variables**(e.g.,Gender, Location of Stall, Type of Product Sold, Migrated or Not, Education Level)
* **Numerical variables**(e.g*.,* Monthly Average Income, Working Hours, Average count of customers)

Each row in the dataset represents a **single street vendor**, and each column corresponds to a specific variable collected through the survey.

The graph shows that most street vendors in Mumbai are between 25 and 44 years old, with 153 vendors aged 25–34 and 158 vendors aged 35–44. Fewer vendors are younger than 24 or older than 55.

This suggests that street vending is mainly a source of livelihood for people in their working years.

The chart shows that street vending in Mumbai is mostly done by men, who make up 86% of all vendors. Women account for only 14%. This large difference suggests that men are much more involved in street vending activities than women

The most common types of street stalls in Mumbai are "Ready to eat" (24%), "Mobile accessories" (20%), and "Clothes" (19%). These three categories make up the majority of stalls.Other stalls include items like shoes, flowers, and juices (each about 6%), while stalls selling electronics (1%) and food items like fruits (4%) and vegetables (5%) are fewer. "Other" types of stalls account for 7%.

This shows that street vending in Mumbai mainly focuses on food and daily-use items.

Most street vendors in Mumbai have primary education (186 people) or secondary education (123 people). Fewer vendors have no education (40), higher education like college or university (39), or are still studying (15).

The chart shows that 58% of street vendors in Mumbai sell non-food items, while 42% sell food items.

This means more vendors are involved in selling products other than food.

Most street vendors in Mumbai are married. Fewer vendors are single, divorced, or widowed.

This shows that street vending is mainly done by people who have families.

The chart shows the total number of street vendors in different regions of Mumbai. The Western region has the highest number of vendors, with 105 individuals engaged in street vending. This is closely followed by the Trans Harbour region, which has 103 vendors. The Central region has slightly fewer vendors, totaling 100. The Harbour region has the lowest number, with 95 vendors.

**Objective 1**

To study the influence of socio-economic factors on the income of street vendors.

(Regression Analysis)

**REGRESSION ANALYSIS**

Regression analysis is a statistical method used to examine the relationship between a dependent variable and one or more independent variables. It helps in understanding how changes in independent variables affect the dependent variable and in making predictions. It is given by,

y= β₀ + β₁x₁ + β₂x₂ + ... + βkxk + ε

Where,

* y is the dependent variable.
* x₁, x₂, ... xₙ are independent variables.
* β₀ is the y-intercept.
* β₁, β₂ ..., βk are the coefficients for each independent variable, ε is error

**Assumptions:**

1) Linearity: A linear relationship must exist between the independent variables and the dependent variable.

2) Independence: The residuals (the differences between the observed and predicted values) should be independent of each other.

3) Homoscedasticity: The variance of the residuals should be constant across all levels of the independent variables.

4) Normality: The residuals should be approximately normally distributed.

5) No Multicollinearity: Independent variables should not be highly correlated with each other.

Dependent Variable: Average Income

Independent Variables:

• Age • Region • Location • Gender • Marital Status • Product type

• Number of dependents • Level of Education • Working Hours

• Migration Status

**Tool used**: RStudio

**Full Model:** A full model refers to a model that includes all the variables or the parameters of interest. It provides a comprehensive representation. The coefficients of full model are:

**Output:** Regression Coefficients Table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 58148.1 | 11888.36 | 4.891 | 1.49e-06 |
| RegionHarbour | 2062.14 | 3347.51 | 0.616 | 0.538250 |
| RegionTrans Harbour | 2896.93 | 3443.09 | 0.841 | 0.400669 |
| RegionWestern | 11360.45 | 3290.01 | 3.453 | 0.000617 |
| LocationWest | -2264.46 | 2281.59 | -0.992 | 0.321593 |
| Age | -73.73 | 163.9 | -0.45 | 0.653070 |
| Gender Male | 2943.11 | 3232.34 | 0.911 | 0.363128 |
| Marital Status Married | 10114.68 | 7178.94 | 1.409 | 0.159675 |
| Marital Status Single | 5203.89 | 8002.05 | 0.65 | 0.515881 |
| Marital Status Widowed | 12103.6 | 17107.08 | 0.708 | 0.479679 |
| Number of dependents1 | -10463.84 | 6462.92 | -1.619 | 0.106269 |
| Number of dependents2 | -11697.5 | 5769.13 | -2.028 | 0.043302 |
| Number of dependents3 | -10277.51 | 5664.82 | -1.814 | 0.070429 |
| Number of dependents4 | -8947.92 | 5842.74 | -1.531 | 0.126492 |
| Number of dependents5 | -5408.87 | 7095.88 | -0.762 | 0.446383 |
| Number of dependents6 | -11308.9 | 7709.25 | -1.467 | 0.143228 |
| Number of dependents7 | -16800.48 | 7089.52 | -2.37 | 0.018301 |
| Level of Education: No education | -3792.98 | 5145.37 | -0.737 | 0.461480 |
| Level of Education Primary education | 778.88 | 4028.45 | 0.193 | 0.846793 |
| Level of Education Secondary education | -812.01 | 4018.55 | -0.202 | 0.839974 |
| Level Of Education Studying | -5067.45 | 7412.93 | -0.684 | 0.494649 |
| Product Type Non-Food | -37033.89 | 2275.24 | -16.277 | < 2e-16 |
| Working Hours Half day | 3067.8 | 2555.38 | 1.201 | 0.230688 |
| Working Hours Quarter day | 4131.55 | 3319.18 | 1.245 | 0.213995 |
| Migration Status Non-migrants | -844.28 | 2346.62 | -0.36 | 0.719208 |

In equation form,

Average Income =  58148.10 +11360.45(RegionWestern) − 2264.46(LocationWest) − 73.73(Age) + 2943.11(GenderMale) + 10114.68(Marital.StatusMarried) + 5203.89(Marital.StatusSingle) + 12103.60(Marital.StatusWidowed) − 10463.84(Dependents1) − 11697.50(Dependents2) − 10277.51(Dependents3) − 8947.92(Dependents4) − 5408.87(Dependents5) − 11308.90(Dependents6) − 16800.48(Dependents7) − 3792.98(NoEducation) + 778.88(PrimaryEducation) − 812.01(SecondaryEducation) − 5067.45(Studying) − 37033.89(ProductTypeNonFood) + 3067.80(HalfDay) + 4131.55(QuarterDay) − 844.28(NonMigrants)

On the basis of P-Value, we can say that Harbour region, Western region, Number of dependents 2, Number of dependents 7 and product type Non-Food are significant predictors.

**Performance Metrics**:

Residual standard error: 21460 on 378 degrees of freedom

Multiple R-squared: 0.4631

Adjusted R-squared: 0.4291

F-statistic: 13.59 on 24 and 378 df,

p-value: < 2.2e-16

**Interpretation**:

42.91% of the variation in the dependent variable is explained by independent variables. P-Value < 0.05, the model is statistically significant.

**Checking Multicollinearity**:

Output:

|  |  |  |  |
| --- | --- | --- | --- |
|  | GVIF | DF | GVIF^(1/(2\*Df)) |
| Region | 1.778725 | 3 | 1.100740 |
| Location | 1.135008 | 1 | 1.065367 |
| Age | 1.691415 | 1 | 1.300544 |
| Gender | 1.126346 | 1 | 1.061294 |
| Marital Status | 2.438924 | 3 | 1.160201 |
| Number of dependents | 2.775729 | 7 | 1.075647 |
| Level of Education | 2.091462 | 4 | 1.096620 |
| Product Type | 1.102931 | 1 | 1.050205 |
| Working Hours | 1.438325 | 2 | 1.095126 |
| Migration Status | 1.175119 | 1 | 1.084029 |

Interpretation:

Since all VIF values are < 5, there is no multicollinearity present.

We want to identify the predictors influencing Average Income of street vendors. So we use stepwise regression.

**Stepwise Regression**

Stepwise regression is a technique used to select the most relevant predictors (independent variables) for a linear regression model by automatically adding or removing variables based on statistical criteria usually AIC, BIC, or p-values to find the optimal combination. The best model is the model with minimum AIC value

**Output and Interpretation:**

Stepwise regression results

Model 1

Start: AIC=8063.23

Average.Income ~ Region + Location + Age + Gender + Marital.Status + Number.of.dependents + Level.Of.Education + Product.Type + Working.Hours + Migration.Status

Model 2

Step: AIC=8057.27

Average.Income ~ Region + Location + Age + Gender + Marital.Status +

Number.of.dependents + Product.Type + Working.Hours + Migration.Status

Model 3

Step: AIC=8051.87

Average.Income ~ Region + Location + Age + Gender + Marital.Status +

Product.Type + Working.Hours + Migration.Status

Model 4

Step: AIC=8048.72

Average.Income ~ Region + Location + Age + Gender + Product.Type +

Working.Hours + Migration.Status

Model 5

Step: AIC=8046.75

Average.Income ~ Region + Location + Age + Gender + Product.Type +

Working.Hours

Model 6

Step: AIC=8044.82

Average.Income ~ Region + Location + Gender + Product.Type +

Working.Hours

Model 7

Step: AIC=8042.94

Average.Income ~ Region + Location + Gender + Product.Type

Model 8

Step: AIC=8041.95

Average.Income ~ Region + Location + Product.Type

Model 9

Step: AIC=8041.44

Average.Income ~ Region + Product.Type

Best model is model 9 which is

**Average.Income ~ Region + Product.Type**

Stepwise AIC Table for Model Selection

Model: Average.Income ~ Region + Product.Type

AIC = 8041.44

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Df | Sum of Sq | RSS | AIC |
| <none> |  |  | 1.8215e+11 | 8041.4 |
| + Location | 1 | 6.7465e+08 | 1.8147e+11 | 8041.9 |
| + Gender | 1 | 3.1344e+08 | 1.8183e+11 | 8042.7 |
| + Working.Hours | 2 | 1.0683e+09 | 1.8108e+11 | 8043.1 |
| + Migration.Status | 1 | 6.6486e+07 | 1.8208e+11 | 8043.3 |
| + Age | 1 | 1.9927e+07 | 1.8213e+11 | 8043.4 |
| + Marital.Status | 3 | 1.2267e+09 | 1.8092e+11 | 8044.7 |
| - Region | 3 | 5.0280e+09 | 1.8717e+11 | 8046.4 |
| + Level.Of.Education | 4 | 1.1200e+09 | 1.8103e+11 | 8047.0 |
| + Number.of.dependents | 7 | 3.1010e+09 | 1.7904e+11 | 8048.5 |
| - Product.Type | 1 | 1.3701e+11 | 3.1996e+11 | 8266.5 |

We can conclude that only Region and Product type are significant predictors influencing Average income of street vendors.

**Regression Coefficients Table**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 58560 | 2470 | 23.7 | <2e-16 \*\*\* |
| RegionHarbour | 1093 | 3067 | 0.36 | 0.7218 |
| RegionTrans Harbour | 2161 | 3006 | 0.72 | 0.4726 |
| RegionWestern | 8955 | 2990 | 3.0 | 0.0029 \*\* |
| Product.TypeNon-Food | -37612 | 2168 | -17.35 | <2e-16 \*\*\* |

Therefore, the best fitted model is given by

Average Income= 58559.752+1092.674\* RegionHarbour + 2161.316\* RegionTrans.Harbour + 8954.974\* RegionWestern

37611.846\* Product.TypeNon-Food

**Assumption checking**:

1) Normality of residuals

A graph with a red line

AI-generated content may be incorrect.

Majority of observations fall on straight line indicating residuals are normally distributed.

**Confirmatory test:**

Shapiro-Wilk test

H₀ : Data is normally distributed.

H₁ : Data is not normally distributed.

|  |  |
| --- | --- |
| Shapiro-Wilk test statistic | P-Value |
| 0.9626 | 0.00000001303 |

p-value < 0.05 , residuals are not normally distributed.

2) Homoscedasticity:

Brush Pagan Test

H0: Variance of the residuals is constant across all levels of the independent variables VS

H1: Variance of the residuals is constant across all levels of the independent variables

Output:

|  |  |  |
| --- | --- | --- |
| BP | DF | P-Value |
| 24.476 | 4 | 0.00006412 |

Interpretation:

P value < 0.05 , we reject H0. It implies that heteroscedasticity is present. Therefore, we use Box-Cox transformation and transform dependent variable.

**Box-Cox transformation**:

It is a statistical technique used to identify the most appropriate power transformation for the response variable in order to improve the validity of linear regression assumptions—particularly normality of residuals and homoscedasticity (constant variance of errors). This transformation yields lambda value as **0.3030303**. Based on the lambda value, the transformation used is

f (y) = ((y)^λ - 1)/λ, y > 0, λ ≠ 0

**Output:**

Regression Output Summary Table

Coefficients Table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 85.654 | 1.611 | 53.168 | < 2e-16 \*\*\* |
| RegionHarbour | 1.625 | 2.0 | 0.812 | 0.41703 |
| RegionTrans Harbour | 1.083 | 1.96 | 0.552 | 0.58107 |
| RegionWestern | 6.241 | 1.959 | 3.201 | 0.00148 \*\* |
| Product.TypeNon-Food | -24.596 | 1.413 | -17.401 | < 2e-16 \*\*\* |

**Model Statistics**

Residual Standard Error: 13.95 on 398 degrees of freedom

Multiple R-squared: 0.4418

Adjusted R-squared: 0.4362

F-statistic: 78.76 on 4 and 398 DF

Model p-value: < 2.2e-16

Interpretation:

The fitted model with transformed dependent variable is

Average Income= 85.653806+ 1.625014\* RegionHarbour + 1.082683 \* RegionTrans.Harbour + 6.240668 \* RegionWestern - 24.596214\* Product.TypeNon-Food

Assumptions checking of new fitted model:

1) Normality of residuals

A graph with a red line

AI-generated content may be incorrect.

All observations fall on straight line indicating residuals are normally distributed.

Histogram of residuals:

A graph of a number of blue bars

AI-generated content may be incorrect.

We can observe that there is bell shaped curve which indicates data is normally distributed

**Confirmatory test**:

Shapiro-Wilk test

H₀ : Data is normally distributed.

H₁ : Data is not normally distributed.

|  |  |
| --- | --- |
| Shapiro-Wilk test statistic | P-Value |
| 0.99316 | 0.06391 |

p-value > 0.05 , residuals are normally distributed.

2) Homoscedasticity:

NCV test (Non-Constant Variance test)

H₀ : The residuals have constant variance (homoscedasticity)

H₁ : The residuals have non-constant variance (heteroscedasticity)

|  |  |  |
| --- | --- | --- |
| Chisquare | Df | P-Value |
| 3.482434 | 1 | 0.062023 |

Since p value > 0.05 , the residuals have constant variance.

All assumptions of the transformed model are met now.

The best fitted model is given by,

Average Income= 85.653806+ 1.625014\* RegionHarbour + 1.082683 \* RegionTrans.Harbour + 6.240668 \* RegionWestern - 24.596214\* Product.TypeNon-Food

**Performance Metrics:**

Residual standard error: 13.95 on 398 degrees of freedom

Multiple R-squared: 0.4418

Adjusted R-squared: 0.4362

F-statistic: 78.76 on 398 DF

P-value: < 2.2e-16

Interpretation:

43.62% of the variation in the dependent variable is explained by independent variables. P-Value < 0.05, the model is statistically significant. The model is moderate fit .

**Discussion:**

Interpretation of Significant Variables:

1) Intercept:

The intercept is the baseline average income when all categorical variables are at their reference levels (e.g., Region = Central, Product Type = Food). It is highly significant (p < 0.001), indicating that the base income level is statistically reliable.

2) RegionWestern:

Street Vendors located in the Western region earn more than those in the reference region (Central). This coefficient is statistically significant at the 1% level (\*\*), suggesting the region has a meaningful effect on income. This may reflect higher economic opportunities, cost of living, or industry presence in the Western region.

3) Product.TypeNon-Food

Selling non-food products is associated with a significantly lower income compared to those selling food products (reference category). The strong negative effect and very high significance level (\*) indicate this is a crucial factor influencing income. This could be due to higher demand or better margins in food-related sales, or challenges in marketing or storage for non-food items.

**Objective 2**

To study the causes for migration among street vendors based on similar migration patterns, socio-economic factors and employment characteristics.

(Cluster Analysis)

**Cluster Analysis**

Cluster analysis, or clustering, is a data analysis technique aimed at partitioning a set of objects into groups such that objects within the same group (called a cluster) exhibit greater similarity to one another than to those in other groups (clusters). It is also called as an unsupervised machine learning algorithm where there is no labelled data. Clustering categorical data involves grouping similar data points based on categories rather than numerical values. Common approaches include using distance metrics like Hamming distance, or algorithms like k-modes, which are specifically designed for categorical data.

**K-modes clustering**: This algorithm is specifically designed for categorical data.  It uses the "mode" (most frequent value) as the cluster center instead of the mean. It replaces the Euclidean distance with a dissimilarity measure suitable for categorical values.

**Hamming distance:** It is used to measure the dissimilarity between data points, specifically for categorical data. It quantifies the number of mismatches (differences) between corresponding attributes (features) of two data points. Essentially, it counts how many times the values of the attributes differ between two data points.

**Objective** : To study the causes for migration among street vendors based on similar migration patterns, socio-economic factors and employment characteristics, we have used K-modes clustering.

**Tool used**: Python

Variables for K-modes clustering:

1) Age group

2) Education status

3) Migration location

4) Product type

5) Working hours

6) Income range

7) Number of dependents

**Steps involved:**

STEP 1: Converting the categories into numbers by encoding them,

STEP 2: Determining the optimal number of clusters by using Elbow Method.

STEP 3: Fitting the model with optimal number of clusters as obtained in step 2.

STEP 4: Predicting the cluster assignment for each row (observation) in the data.

STEP 5: Exploring the cluster profile to find out what makes each cluster unique.

STEP 6: Giving common names to each cluster based on its characteristics.

STEP 7: Building Dendrogram using Hamming distance to find out similarity between clusters.

**Elbow method:**

A graph with a line

AI-generated content may be incorrect.

Interpretation:

By elbow method, optimal number of clusters is 3

**Cluster profile:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Cluster 0 | Cluster 1 | Cluster 2 |
| Common Names | Economic Migrants | Stable Vendors | Opportunity  Seekers |
| Age group | Middle | Middle | Middle |
| Education Status | Educated | Educated | Educated |
| Migration Location | Uttar Pradesh | Mumbai | Mumbai |
| Migration Status | Migrant | Non migrant | Migrant |
| Product Type | Non- food | Non-food | Food |
| Working Hours | Full day | Full day | Full day |
| Income Range | Low | Low | Average |
| Number Dependents | 3 | 2 | 4 |

Interpretation of each cluster:

**Economic Migrants**: The vendors in this group are likely migrating due to economic necessity. Their low income and moderate family size indicate that they may be forced to move to different areas in search of higher demand for their goods. They sell Non-food products and work long hours to support their families.

**Stable vendors**: The vendors in this group are more stable in their businesses. They do not migrate often, likely because they have established themselves in high-demand urban areas. They focus on selling non-food products but earn low which is likely sufficient to sustain nuclear family and likely do not need to move to sustain their business.

**Opportunity seekers**: These vendors migrate for business opportunities rather than out of economic necessity. They are seeking better locations or markets for their goods. Their average income and moderately large family size suggest that they are relatively more financially stable but are still motivated by business growth.

**Hamming distance matrix**:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Economic Migrants | Stable Vendors | Opportunity Seekers |
| Economic Migrants | 0 | 0.62 | 0.38 |
| Stable Vendors | 0.62 | 0 | 0.62 |
| Opportunity Seekers | 0.38 | 0.62 | 0 |

Discussions:

Interpretation:

The Hamming distance between Economic Migrants and Opportunity Seekers is 0.38, showing they are quite similar in characteristics. In contrast, the Hamming distance between Stable Vendors and the other two groups is 0.62, indicating they are significantly different. This highlights two distinct clusters: one of mobile vendors (Migrants + Seekers) and one of settled vendors (Stable Vendors).

A graph with blue and orange lines

AI-generated content may be incorrect.

**Dendrogram**:

Interpretation:

The dendrogram reveals a clear two-level cluster hierarchy among the vendor profiles. Economic Migrants and Opportunity Seekers are relatively close, forming a sub-cluster merging at a low Hamming distance(0.38). Only 38% of features differ between them which means they share 62% similarity which Indicates similar socio-economic attributes, such as age group, education status, migration location. Stable Vendors are significantly different, merging only at a higher distance (0.62). This suggests stronger permanence, deeper local integration, and business stability.

**Objective 3**

To Model Daily Customer Count Based on Influential Vendor Characteristics

(Poisson Regression)

**POISSON REGRESSION**

Poisson regression is a type of generalized linear model (GLM) used when the dependent variable represents **count data** (e.g,number of events, customers,etc)

* **Assumptions:**
* The response variable Yfollows a **Poisson distribution**.
* Mean and variance are equal: *E(Y)=μ, Var(Y)=μ*
* The log of the mean is modeled as a linear function of the predictors:  
   *log⁡(μ)=β0+β1X1+⋯+βkXk*

**Use Case:** Best for count data with no overdispersion (i.e., variance ≈ mean).

**Quasi-Poisson Regression** is an extension of Poisson regression that **adjusts for overdispersion** (when the variance is greater than the mean).

**Assumptions:** The mean structure is the same as Poisson: *log⁡(μ)=β0+β1X1+⋯+βkXk*

But the variance is: Var(Y)=ϕμ where *ϕ* is an overdispersion parameter ( ϕ >1)

**Use Case:** Used when data is overdispersed but a full likelihood model (like negative binomial) is not required.

**Negative Binomial Regression** is a GLM that models overdispersed count data using a Negative Binomial distribution, which explicitly introduces an extra parameter to model the overdispersion.

**Assumptions:**The variance is modeled as: Var(Y)=μ+ αμ^2 where *α\* is the dispersion parameter.

* The mean is still modeled using the log link: l*og⁡(μ)=β0+β1X1+⋯+βkXk\*

**Use Case:** Better than quasi-Poisson when overdispersion is strong and a proper likelihood-based approach is needed (e.g., for AIC, likelihood ratio tests).

**Objective**: To Model Daily Customer Count Based on Influential Vendor Characteristics

Dependent variable: Average Customers served per day

Independent variable: Location: (East and West)

Working Hours : (Full Day, Half Day, Quarter Day)

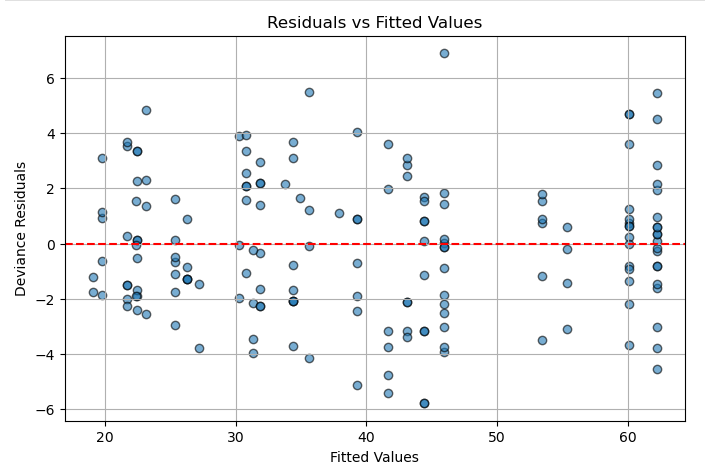
Description of stall: (Ready to east, Raw Food, Clothing, Mobile Accessories and Others,etc)

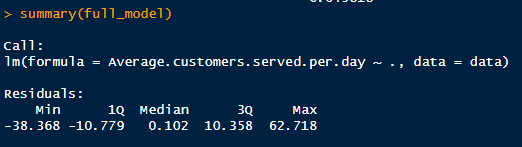
Peak time of Business: (Morning, Afternoon, Evening, Night, EN, MA, Varies throughout the day)

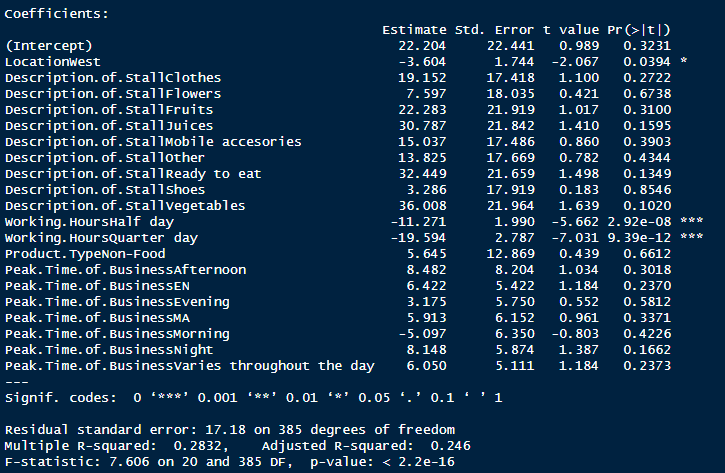
**Tools used** : Python and R software

### Key Assumptions of Poisson Regression:

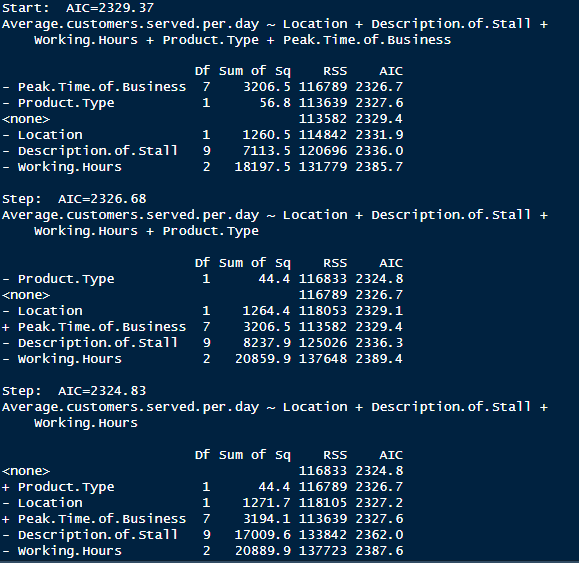
1. **Count data as response** (non-negative integers): Since the Average number of customers served per day is count data
2. **Mean ≈ Variance** (equidispersion) (Mean: 40.15, Variance: 463.90)
3. **No overdispersion**
4. **Independence of observations**







Through Stepwise Regression, we will identify the necessary significant predictors affecting the customer count.

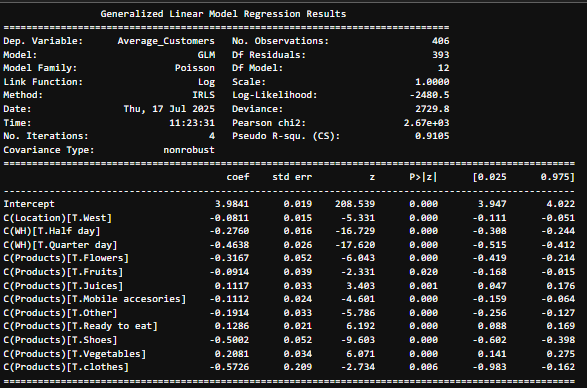


Best Fit Model:

Average.Customers.served.per.day ~ Location + Description.of.stall + Working.Hours

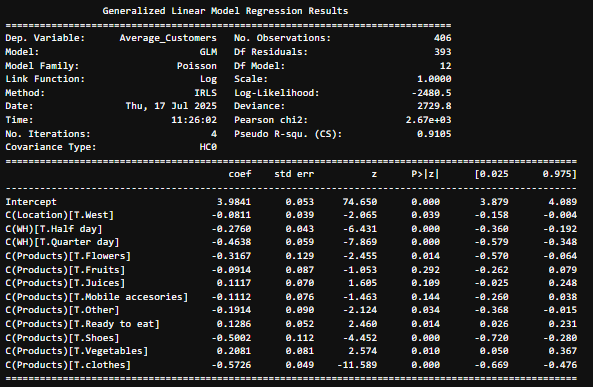
Proceeding for :

1. **Poisson Regression:**



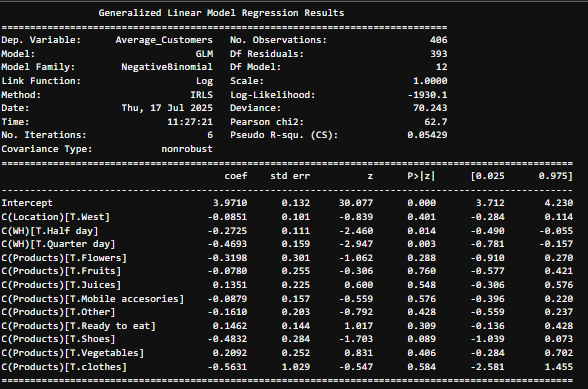


1. **Quasi-Poisson Regression:**





1. **Negative Binomial Regression:**



Dispersion parameter: 0.16 AIC: 3886.216006015199

The model equation is Average\_customers =exp(3.9710 β0 - 0.0851 β1 – 0.2725 β2 -0.4693 β3 – 0.3198 β4 – 0.0780 β5 + 0.1351 β6 – 0.0879 β7 – 0.1610 β8+ 0.1462 β9 – 0.4832 β10 + 0.2092 β11 – 0.5631 β12

Result:

Dependent Variable: Average\_Customers

Significant Predictors (p < 0.05):

C(WH)[T.Half day]: Coefficient = -0.2725, p = 0.014

➤ Working half-day reduces customer count significantly.

C(WH)[T.Quarter day]: Coefficient = -0.4261, p = 0.001

➤ Working quarter-day further reduces customer count significantly.

These are statistically significant, indicating shorter working hours negatively impact customer visits

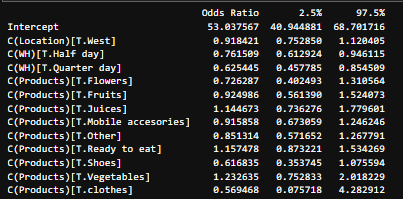
Non-significant Predictors (p > 0.05):

Location (West): No significant effect on customers (p = 0.401).

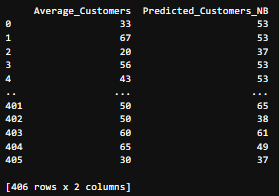
Product types (Flowers, Fruits, Juices, etc.):

All p-values > 0.05 → not statistically significant, though coefficients show weak trends.

Odds Ratio:



**Predicticted number of cutomers**



**Conclusions:**

**1.** The model identifies Western Region and Product Type Non-Food as significant predictors of income. The Western Region positively impacts income, while engaging in Non-Food product sales drastically reduces it.

**2.** The cluster analysis clearly demonstrates that migration among street vendors is differentiated and structured, rather than uniform. On one hand, Economic Migrants and Opportunity Seekers form a closely linked cluster reflecting a common migration pattern characterized by Socio-economic factors and temporary or permanent mobility. On the other hand, Stable Vendors, who are markedly different represent a segment of street vendors who are settled in Mumbai and have quite stable business

This distinction indicates that migration decisions among street vendors are shaped by diverse factors — such as economic necessity, migration location, access to opportunities, and family size.

3. For the Customer Count: Work duration is the strongest factor influencing average customer count. Time of Work (WH) has a clear and significant effect: Vendors working half-day or quarter-day experience significantly fewer customers. Product Type generally does not show significant effects, except weak trends: Selling "Vegetables" or "Juices" may attract slightly more customers. Selling "Shoes" or "Clothes" may lead to fewer customers . Location does not have a strong effect

**Suggestions**

**1.** Non-food vendors, who face significantly lower earnings compared to their food-selling counterparts, require support in skill development training and marketing assistance. This can help improve their product appeal, market access, and business sustainability. The difference in income between regions especially the higher income seen in the Western region shows that we need to make things fairer across all areas. To do this, the government should improve basic facilities, build better markets, and give financial support in lower-income regions. These findings are highly valuable for evidence-based urban planning, particularly for designing targeted interventions under the Smart Cities Mission and PM SVANidhi scheme. These insights can help urban local bodies allocate resources, define vending zones, and deliver benefits more efficiently to the most vulnerable vendor segments.

**2.** Recognizing these patterns allows for targeted policy interventions: while mobile vendors may need support with housing, loans, business expansion awareness and settled vendors may benefit more space rights and digital empowerment.

**3**. To maximize customer footfall and increase the number of customers served per day, vendors are encouraged to operate for full-day shifts rather than shorter durations. Extended working hours have shown a significant positive impact on customer volume.In terms of product strategy, vendors may benefit from focusing on items such as ready-to-eat foods, vegetables, or juices, as these categories tend to be associated with higher customer engagement. Conversely, caution should be exercised when selling products like shoes or clothes, which may attract fewer customers unless there is strong supporting demand or the location is particularly favorable. Regarding location, there is no strong evidence suggesting that operating in the West zone negatively impacts customer count. Therefore, vendors can prioritize improving their operational strategies over relocating.

**R Code**

**1. R code for Regression analysis:**

#Code1 for converting to factors

> data$Region=as.factor(data$Region)

> data$Location <- as.factor(data$Location)

> data$Gender <- as.factor(data$Gender)

> data$Marital.Status <- as.factor(data$Marital.Status)

> data$Level.Of.Education <- as.factor(data$Level.Of.Education)

> data$Product.Type <- as.factor(data$Product.Type)

> data$Working.Hours <- as.factor(data$Working.Hours)

> data$Migration.Status <- as.factor(data$Migration.Status)

> data$Number.of.dependents<- as.factor(data$Number.of.dependents)

> str(data)

#Fitting full regression model code

> full\_model <- lm(Average.Income ~., data = data)

> coef(full\_model)

Stepwise regression code

> step\_model <- step(full\_model, direction = "both",trace = 2)

Code for best fitted model

>model1=lm(Average.Income ~ Region + Product.Type,data=data)

> summary(model1)

>coef(model1)

#Code for assumptions:

>qqnorm(residuals(model1), main = "Q-Q Plot of Residuals")

> qqline(residuals(model1), col = "red", lwd = 2)

> shapiro.test(resid(model1))

>library(lmtest)

> bptest(model1)

#Box-cox transformation code

> library(MASS)

> boxcox\_model=boxcox(model1,lambda = seq(-2,2,0.1),plot=TRUE)

>best\_boxcox\_lambda=boxcox\_model$x[which.max(boxcox\_model$y)]

> print(best\_boxcox\_lambda)

> lambda=best\_boxcox\_lambda

> lambda

Code for transformed model

> model\_boxcox=lm(Average.Income ~ Region + Product.Type,data=data)

> summary(model\_boxcox)

**2. Cluster analysis Python code:**

pip install kmodes

import pandas as pd

Code for converting these columns to categorical codes

for col in cat\_cols:

df[col] = df[col].astype('category').cat.codes

data\_for\_clustering = df[cat\_cols]

Code to Initialize and fit k-modes clustering

km = KModes(n\_clusters=3, init='Huang', n\_init=5, verbose=1)

clusters = km.fit\_predict(data\_for\_clustering)

df['Cluster'] = clusters

costs = []

for k in range(2, 10):

km = KModes(n\_clusters=k, init="Huang", n\_init=5, verbose=0)

km.fit\_predict(df)

costs.append(km.cost\_)

import matplotlib.pyplot as plt

plt.plot(range(2, 10), costs, marker='o')

plt.xlabel('Number of clusters')

plt.ylabel('Cost')

plt.title('Elbow Method For Optimal K')

plt.show()

km = KModes(n\_clusters=3, init='Huang', n\_init=5, verbose=1,random\_state=42)

clusters = km.fit\_predict(data\_for\_clustering)

Codes for dendrogram:

from sklearn.preprocessing import OrdinalEncoder

from scipy.spatial.distance import pdist

from scipy.cluster.hierarchy import linkage, dendrogram

import matplotlib.pyplot as plt

encoder = OrdinalEncoder()

X = encoder.fit\_transform(data)

Code to compute Hamming distance

dist\_matrix = pdist(X, metric='hamming')

Z = linkage(dist\_matrix, method='average')

Code to Plot dendrogram

plt.figure(figsize=(8, 5))

dendrogram(Z, labels=data.index.tolist())

plt.title("Dendrogram of Cluster Profiles (Hamming + Average Linkage)")

plt.xlabel("Clusters")

plt.ylabel("Hamming Distance")

plt.grid(True, axis='y', linestyle='--', alpha=0.5)

plt.show()

**3. Poisson Regression R code:**

#converting to factors

data$Working.Hours=as.factor(data$Working.Hours)

data$Location <- as.factor(data$Location)

data$Description.of.Stall <- as.factor(data$Description.of.Stall)

data$Product.Type <- as.factor(data$Product.Type)

data$Peak.Time.of.Business <- as.factor(data$Peak.Time.of.Business)

str(data)

full\_model <- lm(Average.customers.served.per.day ~., data = data)

coef(full\_model)

summary(full\_model)

step\_model <- step(full\_model, direction = "both",trace = 2)

**Python Code:**

import pandas as pd

import statsmodels.api as sm

import statsmodels.formula.api as smf from sklearn.model\_selection

import train\_test\_split from sklearn.metrics

import mean\_absolute\_error, mean\_squared\_error

target = 'Average\_Customers'

features = F1.columns.drop(target)

X = F1[features]

y = F1[target]

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

train\_F1 = X\_train.copy() train\_F1[target] = y\_train test\_F1 = X\_test.copy() test\_F1[target] = y\_test

formula = 'Average\_Customers ~ C(Location) + C(WH) + C(Products)'

poisson\_modelF1 = smf.glm(formula=formula, data=F1, family=sm.families.Poisson()).fit() print(poisson\_modelF1.summary()) # POISSON REGRESSION

mean\_val = F1['Average\_Customers'].mean()

var\_val = F1['Average\_Customers'].var()

print(f"Mean: {mean\_val:.2f}, Variance: {var\_val:.2f}")

if var\_val > mean\_val:

print("Overdispersion detected → Poisson may not be appropriate")

else: print("No overdispersion → Poisson may be appropriate")

print(f"AIC: {poisson\_modelF1.aic}")

# Calculate dispersion (manually)

resid = poisson\_modelF1.resid\_deviance

dispersion = sum(resid\*\*2) / poisson\_modelF1.df\_resid

print(f"Dispersion: {dispersion:.2f} (if > 1, overdispersion exists)")

if dispersion > 1.5: print("Strong overdispersion detected — Consider Negative Binomial") elif dispersion > 1.0: print("Mild overdispersion detected — Consider Negative Binomial") else: print("Dispersion within Poisson assumptions")

import statsmodels.api as sm

import statsmodels.formula.api as smf

formula = 'Average\_Customers ~ C(Location) + C(WH) + C(Products)'

# Fit Poisson with robust (quasi-Poisson-style) standard errors

quasi\_poisson\_modelF1 = smf.glm( formula=formula, data=F1, family=sm.families.Poisson() ).fit(cov\_type='HC0') # This simulates quasi-Poisson

# Show summary

print(quasi\_poisson\_modelF1.summary())

print(f"AIC: {quasi\_poisson\_modelF1.aic}")

pearson\_chi2 = quasi\_poisson\_modelF1.pearson\_chi2

df\_resid = quasi\_poisson\_modelF1.df\_resid

dispersion = pearson\_chi2 / df\_resid

print(f"Dispersion parameter: {dispersion:.2f}")

# Negative Binomial model

import statsmodels.api as sm

import statsmodels.formula.api as smf

formula = 'Average\_Customers ~ C(Location) + C(WH) + C(Products)'

nb\_modelF1 = smf.glm( formula=formula, data=F1, family=sm.families.NegativeBinomial(alpha=1) ).fit()

print(nb\_modelF1.summary())

# Calculate Dispersion

pearson\_chi2 = nb\_modelF1.pearson\_chi2

df\_resid = nb\_modelF1.df\_resid

dispersion = pearson\_chi2 / df\_resid

print(f"Dispersion parameter: {dispersion:.2f}")

import numpy as np

coefficients = nb\_modelF1.params

odds\_ratios = np.exp(coefficients)

conf\_int = np.exp(nb\_modelF1.conf\_int())

odds\_table = pd.DataFrame({ "Odds Ratio": odds\_ratios, "2.5%": conf\_int[0], "97.5%": conf\_int[1] })

print(odds\_table)

print(f"AIC: {nb\_modelF1.aic}")

F1['Predicted\_Customers\_NB'] = nb\_modelF1.predict(F1)

F1['Predicted\_Customers\_NB'] = F1['Predicted\_Customers\_NB'].round().astype(int)

# Display the first few predictions

print(F1[['Average\_Customers', 'Predicted\_Customers\_NB']]) pd.set\_option('display.max\_rows', 250)

**Questionnaire**

## Section A: General Information

1. Station Name:  
 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2. Region:  
 ☐ Central ☐ Western ☐ Harbour ☐ Trans Harbour

3. Location of Stall:  
 ☐ East ☐ West ☐ Only East ☐ Only West

4. Type of Stall (Select one or more):  
 ☐ Ready to Eat ☐ Fruits ☐ Vegetables ☐ Juices ☐ Flowers ☐ Clothes ☐ Mobile Accessories ☐ Shoes ☐ Electronics

☐ Other: \_\_\_\_\_\_\_

## Section B: Personal Details

5. Age Group:  
 ☐ Under 18 ☐ 18-24 ☐ 25-34 ☐ 35-44 ☐ 45-54 ☐ 55+

6. Gender:  
 ☐ Male ☐ Female

7. Marital Status:

☐ Single ☐ Married ☐ Divorced ☐ Widowed

8. Number of Dependents:  
 ☐ None ☐ 1–2 ☐ 3–4 ☐ 5+

9. Level of Education:  
☐ No Education ☐ Primary ☐ Secondary ☐ Higher (College/University) ☐ Currently Studying

## Section C: Business Information

10. Products Sold:  
 ☐ Food & Beverages ☐ Clothing & Accessories ☐ Mobile Accessories ☐ Other: \_\_\_\_\_\_\_

11. Years as a Street Vendor:  
 ☐ Less than 1 year ☐ 1–5 years ☐ 6–10 years ☐ More than 10 years

12. Average Working Hours per Day:  
 ☐ Less than 4 hours ☐ 4–6 hours ☐ 7–9 hours ☐ More than 9 hours

☐ Full day

## Section D: Migration

13. Have You Migrated?:  
 ☐ Yes ☐ No

14. If Yes, When Did You Migrate?:  
☐ Less than 1 year ago ☐ 1–3 years ago ☐ 4–6 years ago ☐ More than 6 years ago ☐ Not Applicable

15. From Where?:  
 ☐ Urban ☐ Rural ☐ Not Applicable

16. Type of Migration:  
 ☐ Within Country ☐ Within State ☐ Not Applicable

17. State Name (If migrated):  
 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

18. Reason for Migration:  
 ☐ No job in village ☐ Less income in village ☐ More opportunities in city ☐ More customers in city ☐ Social network ☐ For studies ☐ Due to children ☐ Not applicable

19. Income Change After Migration:  
 ☐ Increased ☐ Remained the Same ☐ Decreased ☐ Not Applicable

20. Did Migration Help You?:  
 ☐ Yes ☐ No ☐ Not Sure

21. Challenges Faced During Migration:  
 ☐ Lack of money ☐ Hard to find a place to sell ☐ Trouble getting a license

☐ Housing problems ☐ Not Applicable

## Section E: Business Operation

22. Busiest Time of Day (Customer Flow):  
 ☐ Early morning (6–9 AM) ☐ Late morning (9 AM – 12 PM) ☐ Afternoon (12 PM – 4 PM) ☐ Evening (4 PM – 7 PM) ☐ Night (7 PM – 10 PM) ☐ Varies throughout the day ☐ Full time

23. Customer Loyalty:  
 ☐ Very few (less than 10%) ☐ Some (10–30%) ☐ Many (30–50%)

☐ Almost all (more than 50%) ☐ Not regular

24. Average Customers Served Per Day (Food Vendors):  
 ☐ < 100 ☐ 100–200 ☐ 200–300 ☐ 300–400 ☐ 400–500 ☐ > 500

25. Average Customers Served Per Day (Non-Food Vendors):  
 ☐ < 50 ☐ 50–100 ☐ 100–150 ☐ 150–200 ☐ > 200

## Section F: Income & Sustainability

26. Monthly Income (Food Vendors):  
 ☐ ₹10,000–20,000 ☐ ₹20,000–30,000 ☐ ₹30,000–40,000 ☐ ₹40,000–50,000 ☐ Above ₹50,000 ☐ Other: \_\_\_\_\_\_\_

27. Monthly Income (Non-Food Vendors):  
 ☐ ₹5,000–10,000 ☐ ₹10,000–15,000 ☐ ₹15,000–20,000 ☐ ₹20,000–25,000 ☐ Above ₹25,000 ☐ Other: \_\_\_\_\_\_\_

**REFERENCES**

1. An Investigation on the Effect of Bias on Determination of Sample Size Based on Data Related to the Students of Schools of Guwahati published in the International Journal of Applied Mathematics and Statistical Sciences, Vol. 2, Issue 1, 2013.
2. Asema Siddiqui and shri Saratchandra Patra,2021, A study of the income of street vendors in ward g in Mumbai city,Vol-9,Issue-1
3. Bhowmik , Sharit K., (2000), Study of Street Vending in Seven Cities, National Association of Street Vendors of India, Delhi (Mimeograph)
4. Bhowmik, Sharit K. 2011, Street Vendors in India: A study in Ten Cities, National Association of Street Vendors of India, Delhi (mimeograph)
5. Debdulal S, (2011), “Working life of street vendors in Mumbai”. The Indian Journal of Labour Economics, VOL.54, No.2
6. Dr A.Priya, 2022,A study on problems faced by street vendors with reference to oimbatore districts,Vol-10,Issue-8 Aug .
7. Dr. D.Indira and Dr. Y. Ramakrishna prasad,2018, A Study on the Challenges faced by Street Vendors(Hyderabad),Vol-5,Issue-3 July
8. Dr.Sumanta Bhattacharya,Debashis sen and Bhacneet kaur Sachdev,2021, A study on the role Played by street vendors in our daily life and impact of the covid-19 on the street vendors,Vol-5, Issuse-6
9. Erni Febrina Harahap,2017,Analysis of determinants of income and efforts to improve the welfare of street vendors in Padang, vol-9,Issue-3.

10) Hilmi Martin ,2020,Literature Review on Street Food Vendors’ Marketing in Subsistence Marketplaces,Vol-9,Issue-2

11) Importance of the Size of Sample and its Determination in the Context of Data Related to the Schools of Guwahati published in the Bulletin of the Gauhati University Mathematics Association, Vol. 12, 2012

12) Manucha Taneyaa, Singh Kuljit,2019, A shop without name: a critical analsyis of the street vendors (protection of livelihood and regulation of street vending) act, 2014 ,Vol-6,Issue-3 Mar

13)Mika Febriani and Eni Setyowati,2023, Analysis of factors affecting the income of street sellers at Karangpandan,Vol-(ICOEBS-22-2),Issue-4.

14) Ms. Thimmaiah Jyothsna,2021,Literature Review on Street Vendors inBangalore , Vol-23,Issue-4 Ms. Thimmaiah Jyothsna,2021,Literature Review on Street Vendors inBangalore ,Vol-23,Issue-4

15)P.B.Narendra Kiran and Dr.G.N.P.V.Babu,2019, Problems and prospects of street vendors: a study with reference to visakhapatnam city ,Vol-9,Issue-6

16) Parikshit Chakraborty and Samarpita Koley,2018, Socio-Economic View on Street Vendors: A Study of a Daily Market at Jamshedpur,Vol-5,Issue-1

17)Saihjpal Meenu, Saihjpal Ashish, Kapoor Shashi,2019, Life and Livelihoods of Street Vendors,Vol-4,Issue-7 July

18)Singh, Tikam,(2022), “Importance of street vendors in India.” International Journal of Advance Research and Innovative Ideas in Education, Vol-8 Issue-5.