**STATISTICAL ANALYSIS   
AND NATURAL LANGUAGE PROCESSING FOR DELIVERY SERVICES IN INDIA  
USING PYTHON**

**1.INTRODUCTION**

Delivery services in India have seen significant growth in recent years, especially with the rise of e-commerce and on-demand delivery platforms. These services cater to a wide range of products, including groceries, essentials, electronics, food, and pharmaceuticals.

Natural Language Processing (NLP) plays a critical role in revolutionizing the delivery services landscape in India, particularly in the domains of e-commerce, logistics, and on-demand delivery platforms. With the rapid growth of online shopping and consumer demand for quick, efficient services, NLP technologies enable businesses to enhance customer experiences, streamline operations, and improve service delivery. This project explores the descriptive and explorative data analysis for the better understanding of the delivery services. Inferential analysis is done to understand the relationship of the attributes of the service providers. This project aims to analyse delivery service data in India using various NLP techniques. Specifically, we use text data from customer reviews to derive insights into service quality, sentiment, and other factors that affect delivery services. NLP allows us to process the text in customer feedback, extract useful patterns, and analyse service provider performance.

**2.DATA EXPLORATION**

**Data Source:** The dataset for this project is sourced from Kaggle, a popular platform for data science and machine learning competitions, which provides a variety of datasets. In this case, the dataset contains information related to delivery services in India, including details about different agents, customer feedback, delivery times, service ratings, and more. The data is used to analyze various performance metrics and predict future service outcomes.

**Database Atrributes Decription:** The dataset describes delivery performance data from four different delivery agents across ten different locations in India. This information helps in analyzing various Wperformance metrics and customer feedback based on delivery service providers and their geographical reach. Below is a detailed explanation of how the data is structured:

|  |  |  |  |
| --- | --- | --- | --- |
| **Attributes** | **Data Type** | **Description** | **Varibles as Examples** |
| Attribute 1 | Category | Agent Name | Delivery service provider (e.g., Zepto, JioMart, Blinkit) |
| Attribute 2 | Numerical | Rating | Customer rating of the service (1-5) |
| Attribute 3 | Text | Review Text | Customer feedback in text form |
| Attribute 4 | Numeric | Delivery Time (min) | Time taken to complete the delivery (in minutes) |
| Attribute 5 | Category | Location | Geographical location of delivery (e.g., Delhi, Lucknow) |
| Attribute 6 | Category | Order Type | Type of product (e.g., Grocery, Pharmacy, Food) |
| Attribute 7 | Category | Customer Feedback Type | Sentiment of the feedback (Positive, Neutral, Negative) |
| Attribute 8 | Category | Price Range | Price category of the order (Low, Medium, High) |
| Attribute 9 | Category | Discount Applied | Indicates if a discount was applied to the order (Yes/No) |
| Attribute 10 | Category | Product Availability | Availability status of the product (In Stock/Out of Stock) |
| Attribute 11 | Numerical | Customer Service Rating | Rating of the customer service interaction (1-5) |
| Attribute 12 | Category | Order Accuracy | Indicates whether the delivered order matched the original (Correct/Incorrect) |

**3.PROJECT OUTCOME**

* To understanding the Explorative data analysis of Delivery services in India.
* Assess the Relationship between Order Type and Product Availability.
* Evaluate the Impact of Order Type on Customer Service Ratings.
* Investigate the Association between Discount Application and Customer Feedback Type.
* Examine the Influence of Agent Name and Location on Customer Service Ratings.
* Analyse the Relationship between Price Range and Location.
* Explore the Association between Price Range and Order Type.
* Assess the Interaction between Location and Order Type on Delivery Time.
* Evaluate the Cosine Similarity between Sentiment for Review text and Customer Feedback Type.
* Analyse Sentiment Distribution Across Location and Order Type
* Perform Agent-wise Keyword Extraction using TfidfVectorizer and Generate Word Clouds

**4.METHODOLOGY**

**INFERENTIAL DATA ANALYSIS**

**1. CHI-SQUARE TEST FOR INDEPENDENCE TO TEST**

A chi-square (Χ2) test of independence is a type of Pearson’s chi-square test. Pearson’s chi-square tests are nonparametric tests for categorical variables. They’re used to determine whether your data are significantly different from what you expected. If two variables are related, the probability of one variable having a certain value is dependent on the value of the other variable.

**Contingency tables**

When you want to perform a chi-square test of independence, the best way to organize your data is a type of frequency distribution table called a contingency table. A contingency table, also known as a cross tabulation or crosstab, shows the number of observations in each combination of groups. It also usually includes row and column totals.

Pearson’s chi-square (Χ2) is the [test statistic](https://www.scribbr.com/statistics/test-statistic/) for the chi-square test of independence:

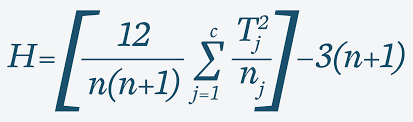
\begin{equation*}X^2 = \sum {\frac {(O-E)^2}{E}}$\end{equation*}

Where

* **Χ2** is the chi-square test statistic
* **Σ** is the summation operator (it means “take the sum of”)
* **O** is the observed frequency
* **E** is the expected frequency

**2. KRUSKAL-WALLIS H-TEST**

The Kruskal–Wallis test is a statistical test used to compare two or more groups for a continuous or discrete variable. The Kruskal-Wallis test is an extension of the Mann-Whitney *U* test. The test is the nonparametric analog of one-way analysis of variance and detects differences in distribution location. The test assumes that there is no *a priori* ordering of the *k* populations from which the samples are drawn

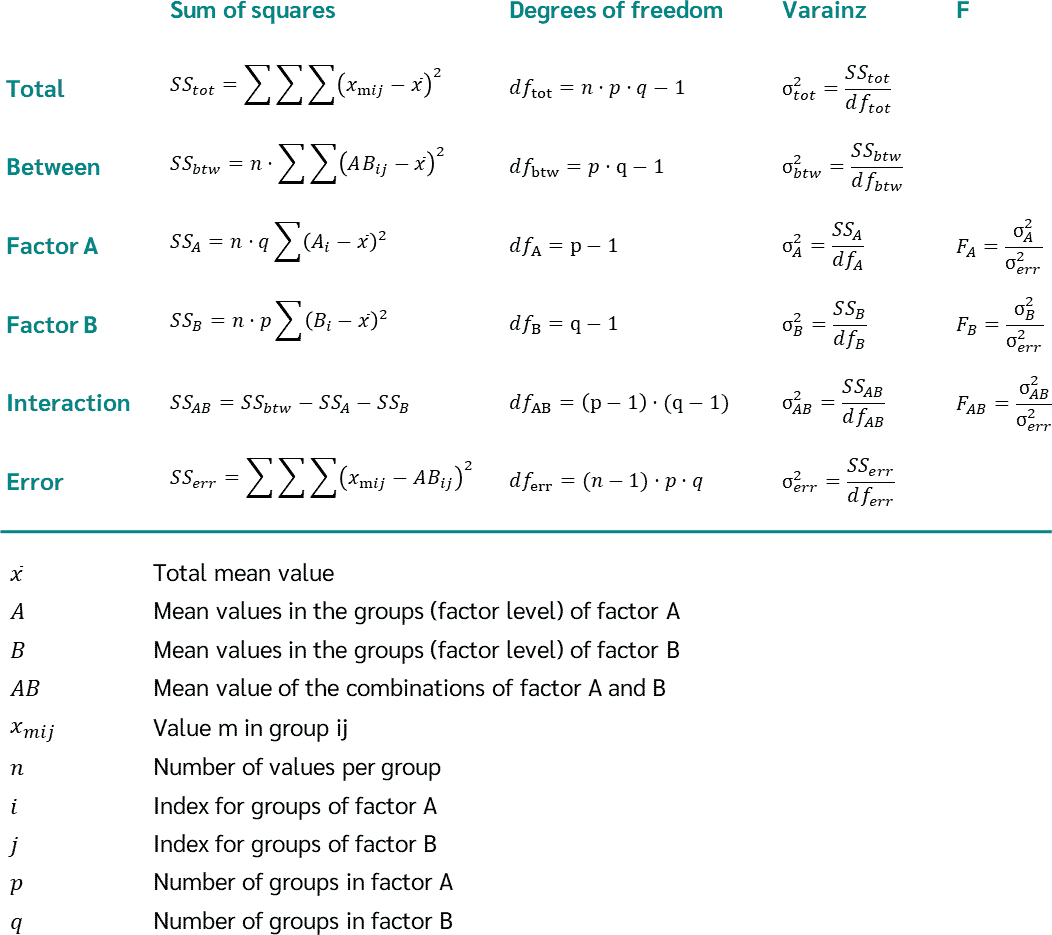
. 

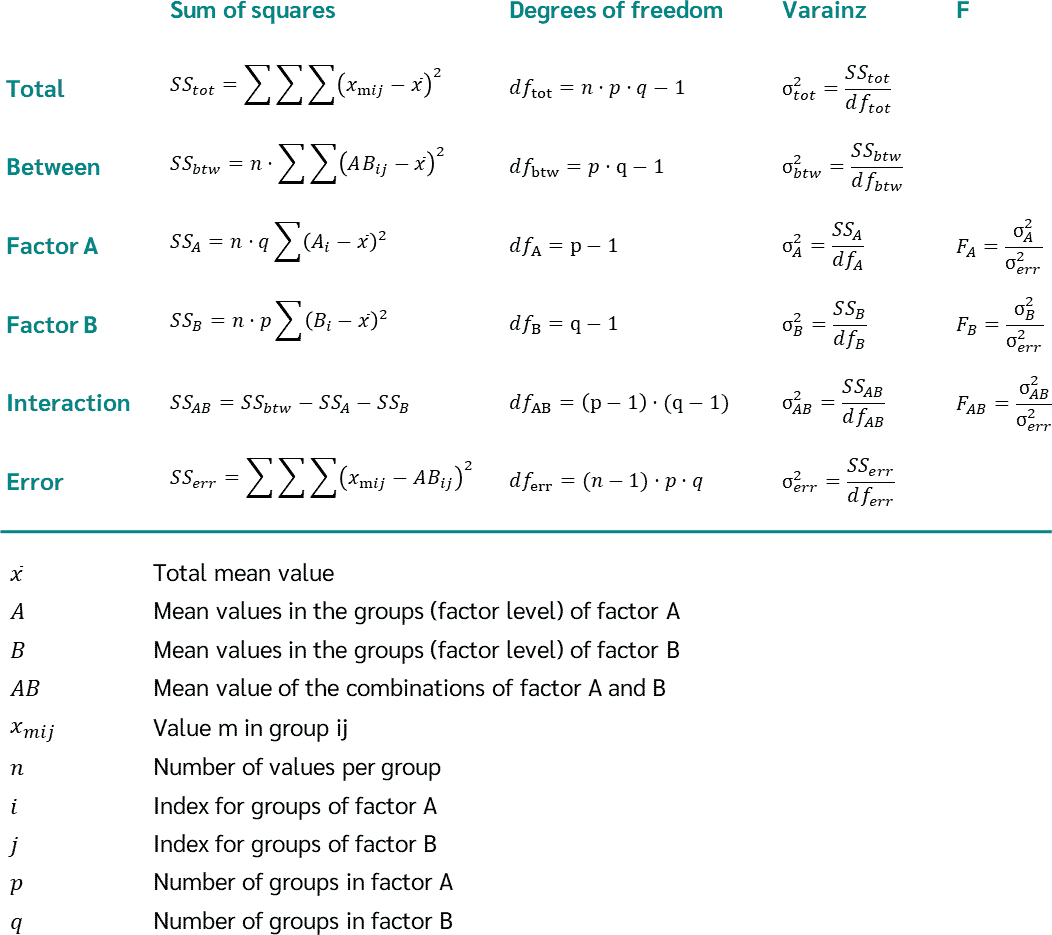
What the variables in the formula mean

* **N**: The total number of observations in the sample
* **k**: The number of groups being compared
* **Ri**: The sum of ranks for group i
* **ni**: The number of observations in group i

**3. TWO-WAY ANOVA**

ANOVA (Analysis of Variance) is a [statistical test](https://www.scribbr.com/statistics/statistical-tests/) used to analyze the difference between the means of more than two groups. A two-way ANOVA is used to estimate how the[mean](https://www.scribbr.com/statistics/mean/) of a [quantitative variable](https://www.scribbr.com/methodology/types-of-variables/) changes according to the levels of two categorical variables. Use a two-way ANOVA when you want to know how two independent variables, in combination, affect a dependent variable.





**NLP (NATURAL LANGUAGE PROCESSING)**

**1. SENTIMENT ANALYSIS**

Sentiment analysis is a process that involves analysing textual data such as social media posts, product reviews, customer feedback, news articles, or any other form of text to classify the sentiment expressed in the text. The sentiment can be classified into three categories: Positive Sentiment Expressions indicate a favourable opinion or satisfaction; Negative Sentiment Expressions indicate dissatisfaction, criticism, or negative views; and Neutral Sentiment Text expresses no particular sentiment or is unclear.

**2. COSINE SIMILARITY**

Cosine Similarity is a metric used to determine the cosine of the angle between two non-zero vectors in a multi-dimensional space. It is a measure of orientation and not magnitude, ranging from -1 to 1. In the context of text similarity, this metric provides a robust way to gauge the similarity between two sets of text data.

*Mathematical Definition:*Cosine Similarity is calculated as the dot product of two vectors divided by the product of their magnitudes.

**3. WORD CLOUD GENERATION**

Word Cloud is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance. Significant textual data points can be highlighted using a word cloud. Word clouds are widely used for analysing data from social network websites.

**4. TF-IDF (TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY)**

TF-IDF (Term Frequency-Inverse Document Frequency) is a way of measuring how relevant a word is to a document in a collection of documents. This is done by multiplying two metrics:

**Term Frequency (TF):** how many times a word appears in a document.

**Inverse Document Frequency (IDF):** the inverse document frequency of the word across a collection of documents. Rare words have high scores, common words have low scores. TF-IDF has many uses, such as in information retrieval, text analysis, keyword extraction, and as a way of obtaining numeric features from text for machine learning algorithms.

**5. ALGORITHM FRAMEWORK**

**5.1. Data Preprocessing:**

Data preprocessing ensures that the data is cleaned and properly formatted for analysis. Here's a summary of the preprocessing steps done in your code:

**1. Handling Column Names**:

* + The column names in the Data Frame are cleaned up by replacing spaces with underscores **(\_)** and replacing parentheses with underscores. This helps in making column names more readable and easier to work with in Python.

**2. Checking for Missing Values:**

* + The **isnull().sum()** method is used to check for missing (NaN) values in each column of the Data Frame. This helps in identifying any missing data that needs to be handled before performing analysis.

**3. Data Type Information**:

* + The **info()** method is called to provide detailed information about the Data Frame, such as column names, data types, and the number of non-null entries. This gives a sense of the data quality and what types of transformations may be needed.

**5.2. Text Preprocessing:**

1. **Lowercasing**:

* In the **preprocess\_text** function, the text is converted to lowercase using the **.lower()** method. This ensures that the text is uniform and that words like "Good" and "good" are treated identically.

2. **Removing Non-Alphabetical Characters**:

* A regular expression **(re.sub(r'[^a-z\s]', '', text))** is used to remove any characters that are not lowercase letters (a-z) or spaces. This eliminates numbers, punctuation, and other unwanted characters from the text, leaving only the words that are relevant for analysis.

3. **Removing Stop words**:

* The **preprocess\_text** function also removes common **"stopwords" (like "the", "and", etc.)** that don’t add much value to sentiment analysis or word cloud generation. This is done by using the **stopwords. words('english')** list from the **nltk.corpus** and filtering out these words.

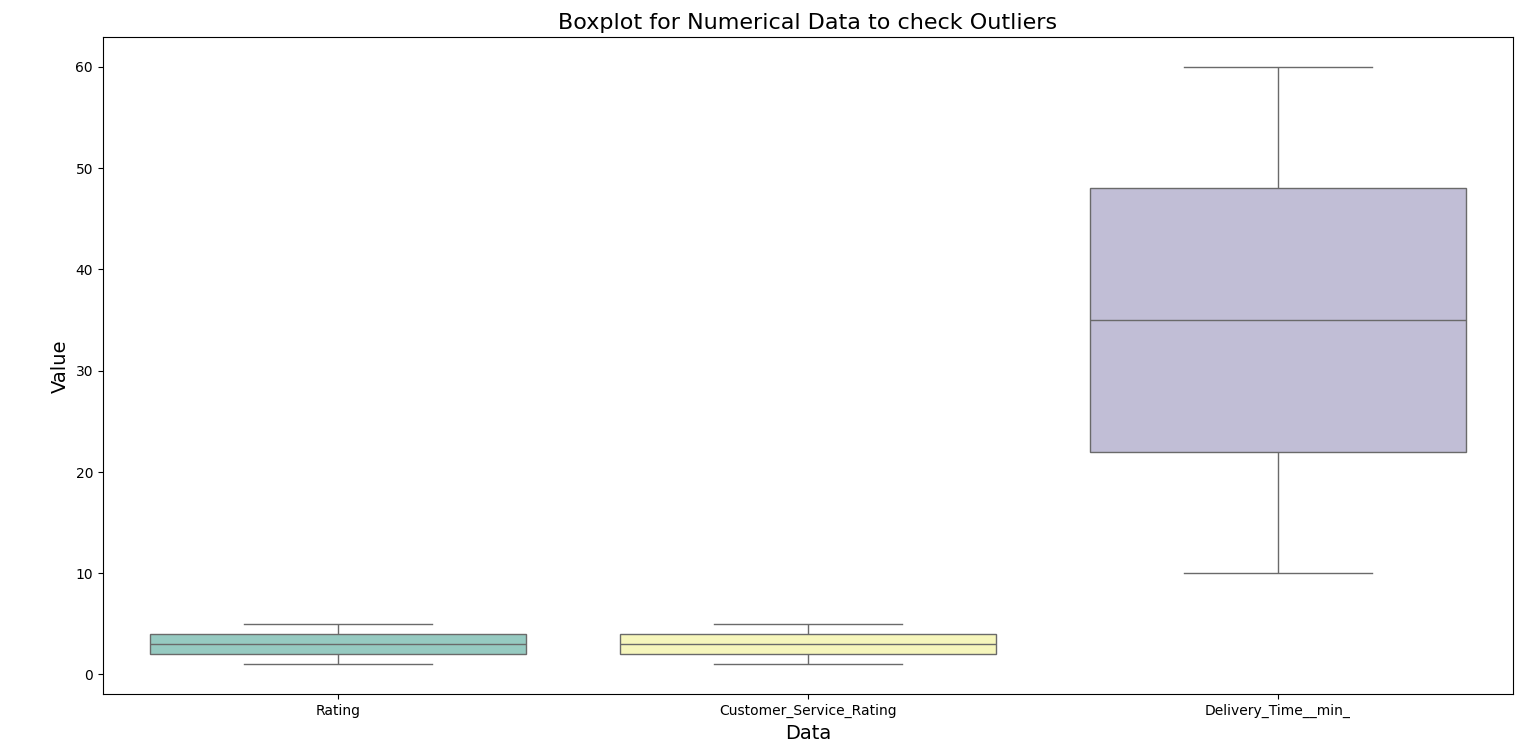
**5.2. Exploratory Data Analysis (EDA):**

Exploratory Data Analysis (EDA) is an analysis approach that identifies general patterns in the data. These patterns include outliers and features of the data that might be unexpected.

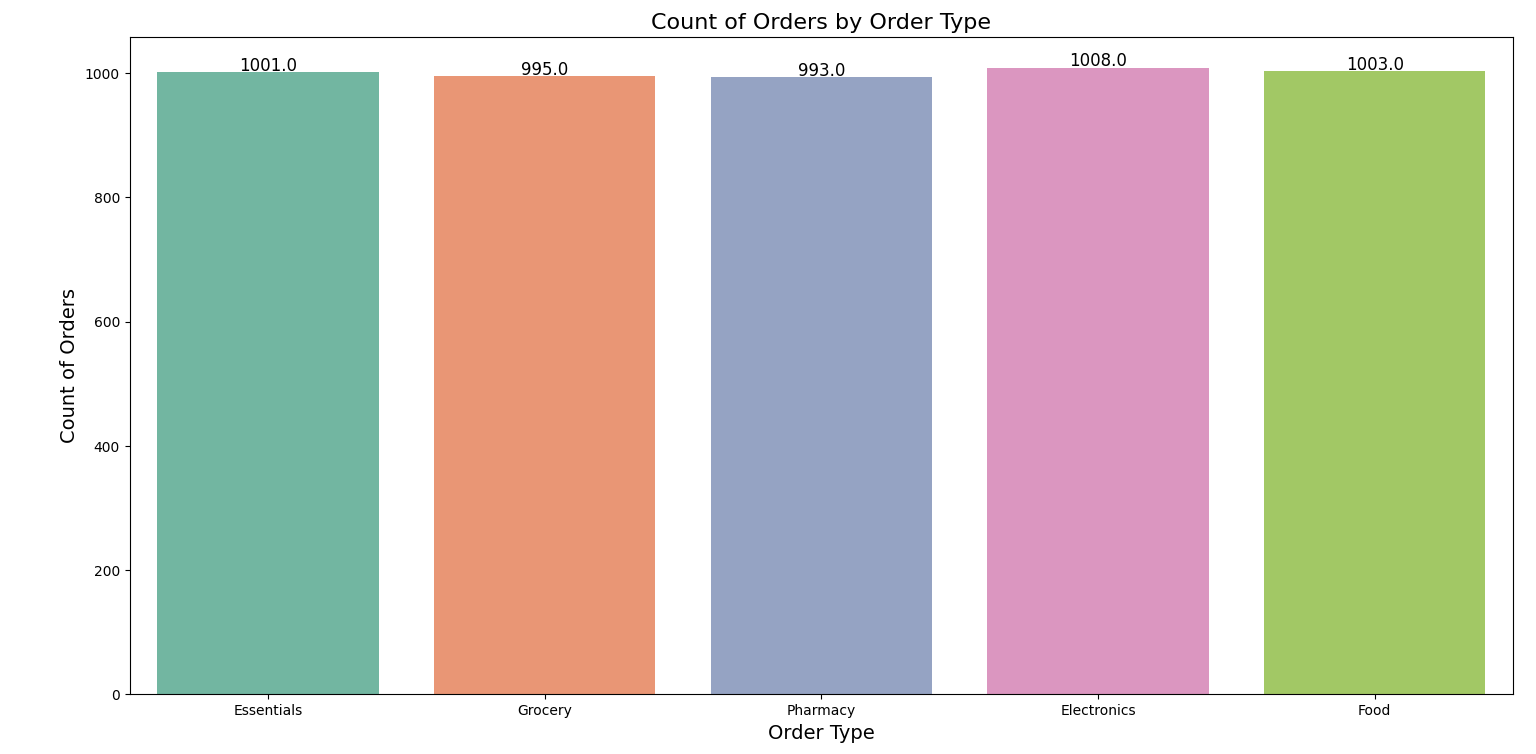
EDA is an important first step in any data analysis. Understanding where outliers occur and how variables are related can help one design statistical analyses that yield meaningful results. In biological monitoring data, sites are likely to be affected by multiple stressors.

**Data Visualization:**

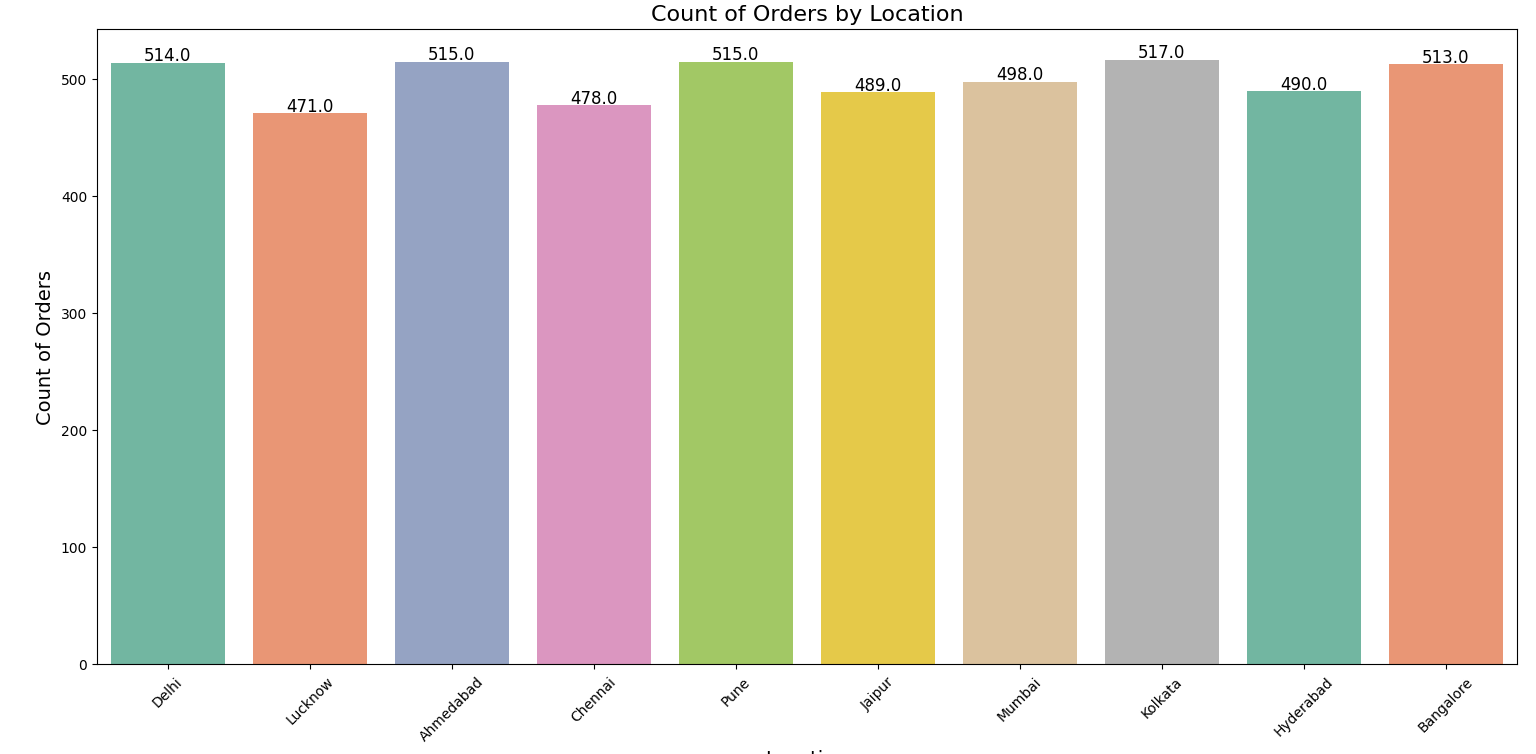
**1. Checking Outliers by boxplot (Numerical Data)**

****

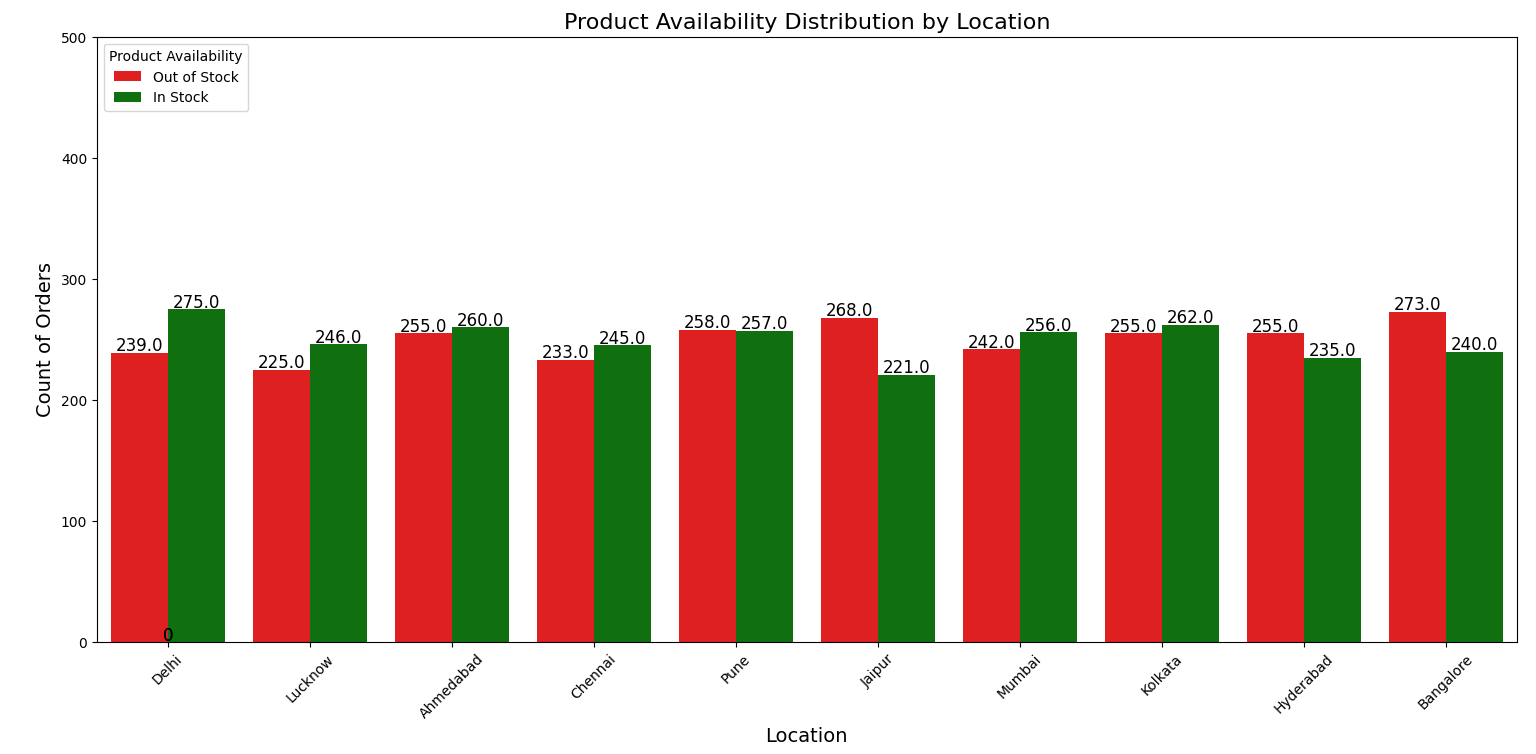
**2. Plotting Frequency Distribution for Order Type with count labels**

****

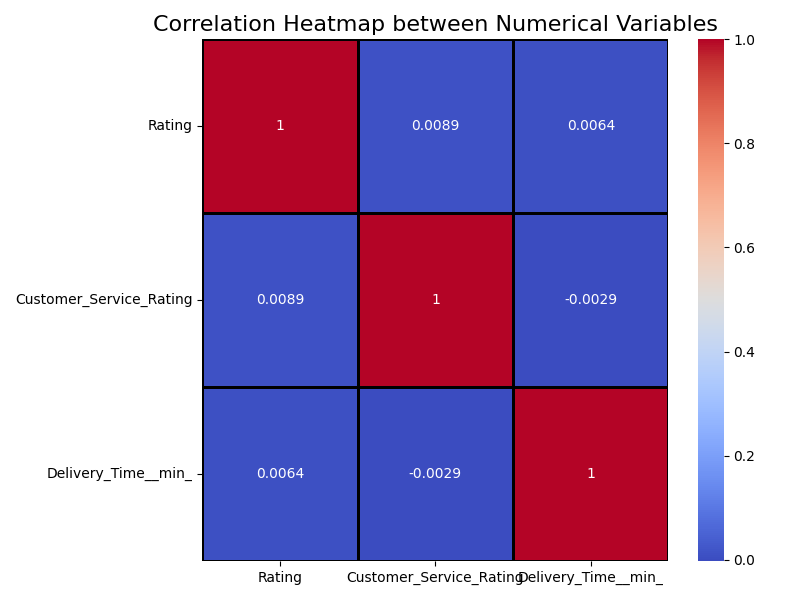
**3. Plotting Frequency Distribution for Location with count labels**

****

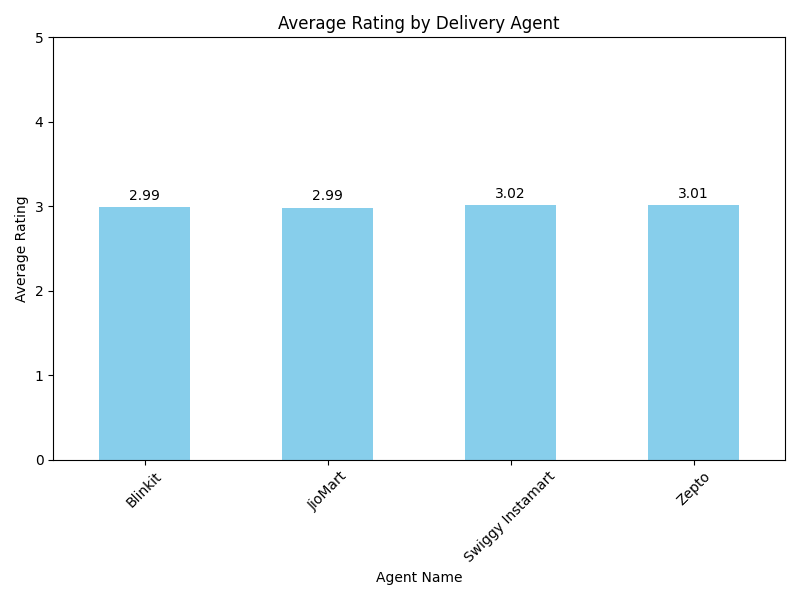
**4. Plotting Frequency Distribution for Product Availability by Location with count labels**

****

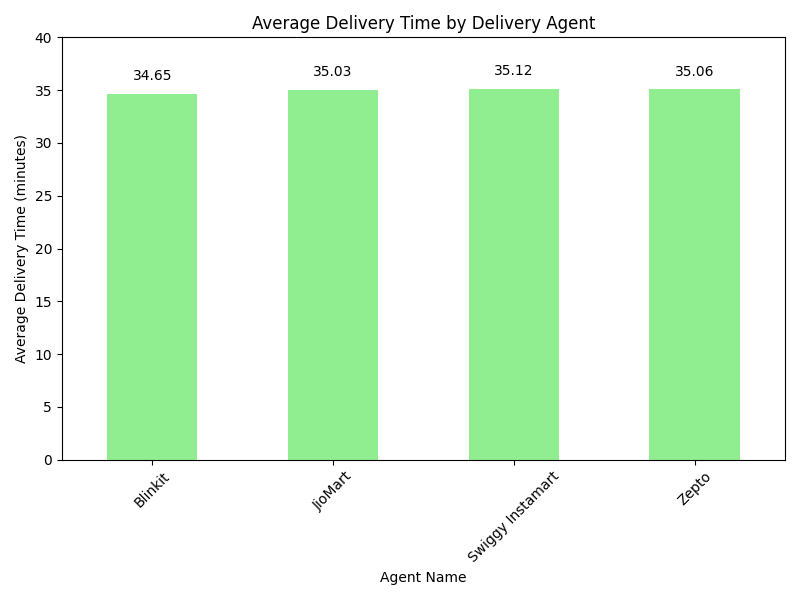
**5. Visualizing the correlation between numerical variables**

****

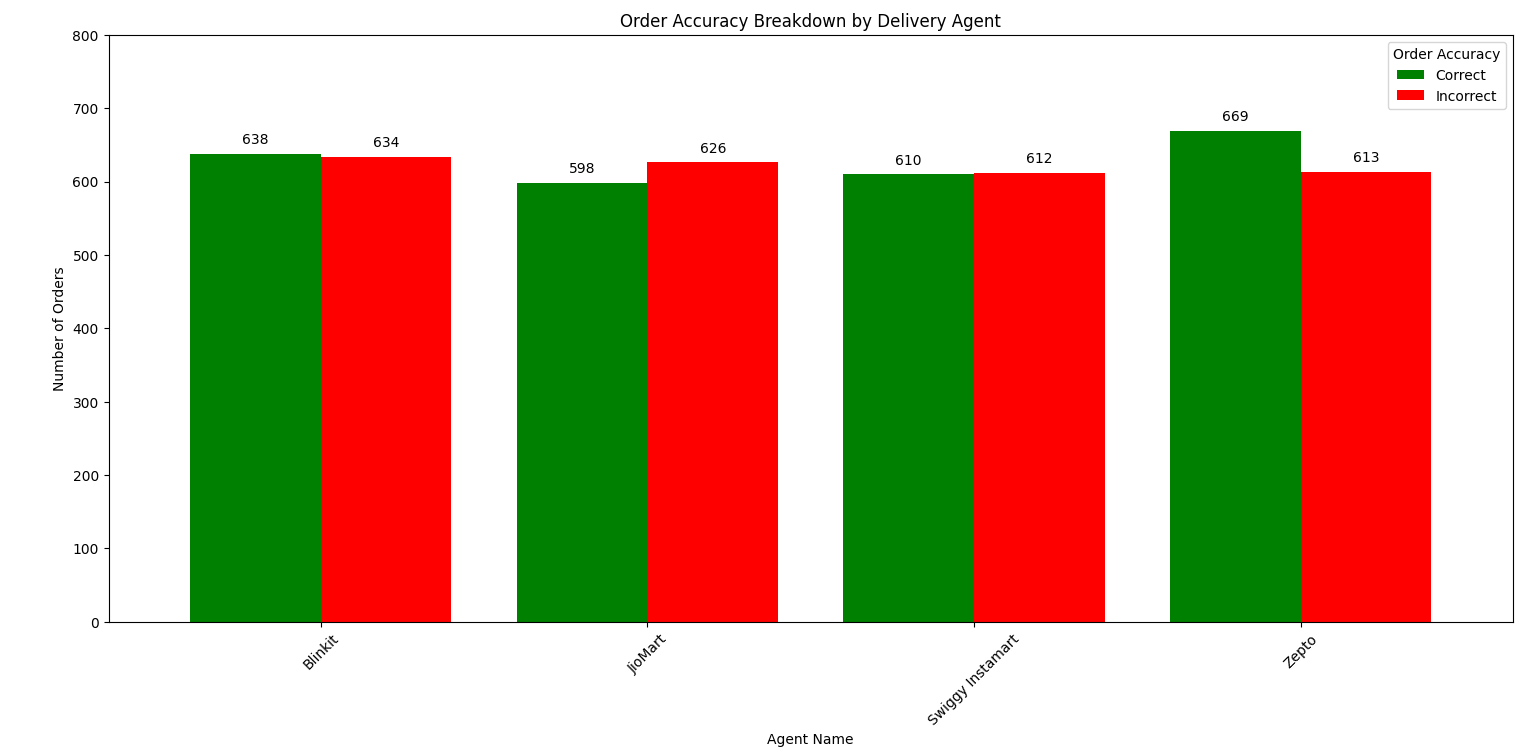
**6. Average Rating by Agent**

****

**7. Average Delivery Time by Agent**

****

**8. Order Accuracy Breakdown Agent wise**

****

**9. Group by Location and Agent Name to get the frequency**



**5.3. Model Building and Evaluation:**

**5.3.1. Inferential Data Analysis**

**1. Chi-Square Test for Independence to test whether there is an association between Order Type and Product Availability**

**Hypothesis Framing**

Null Hypothesis (H₀): There is no significant association between Order Type and Product Availability.

Alternative Hypothesis (H₁): There is a significant association between Order Type and Product Availability

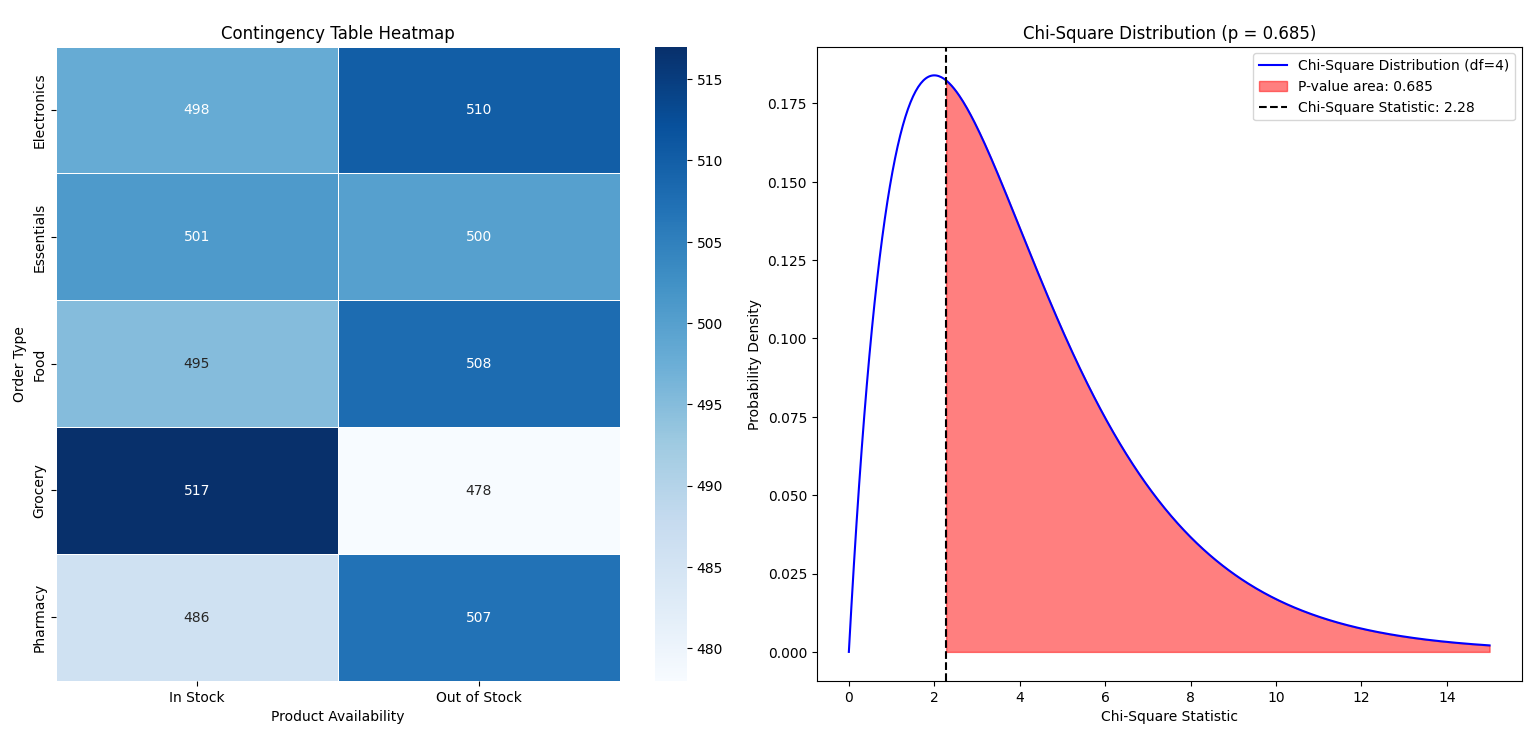
**Output**:

Chi-Square Statistic: 2.28

P-value: 0.685

**Interpretation of Results:**

Since your p-value (0.685) is much greater than the typical significance level of α = 0.05, we fail to reject the null hypothesis. This means that based on the data, there is no significant association between Order Type and Product Availability



**Explanation of the Plot**

* **Contingency Table Heatmap (Left Plot)** gives a **visual understanding** of how the two categorical variables (Order\_Type and Product\_Availability) are distributed in the data.
* **Chi-Square Distribution Plot (Right Plot)** provides the **statistical inference**: it shows the probability of obtaining the observed result by chance and helps you decide whether or not to reject the null hypothesis. If the p-value is smaller than your significance level (e.g., 0.05), it suggests that there is a **significant association** between Order\_Type and Product\_Availability.

**2. Non-Parametric Kruskal-Wallis H-test to check significant difference between the Customer Service Ratings across Order Types.**

**Hypothesis Framing**

Null Hypothesis (H₀): There is no significant difference in Customer Service Ratings across the different Order Types.

Alternative Hypothesis (H₁): There is a significant difference in Customer Service Ratings across the different Order Types.

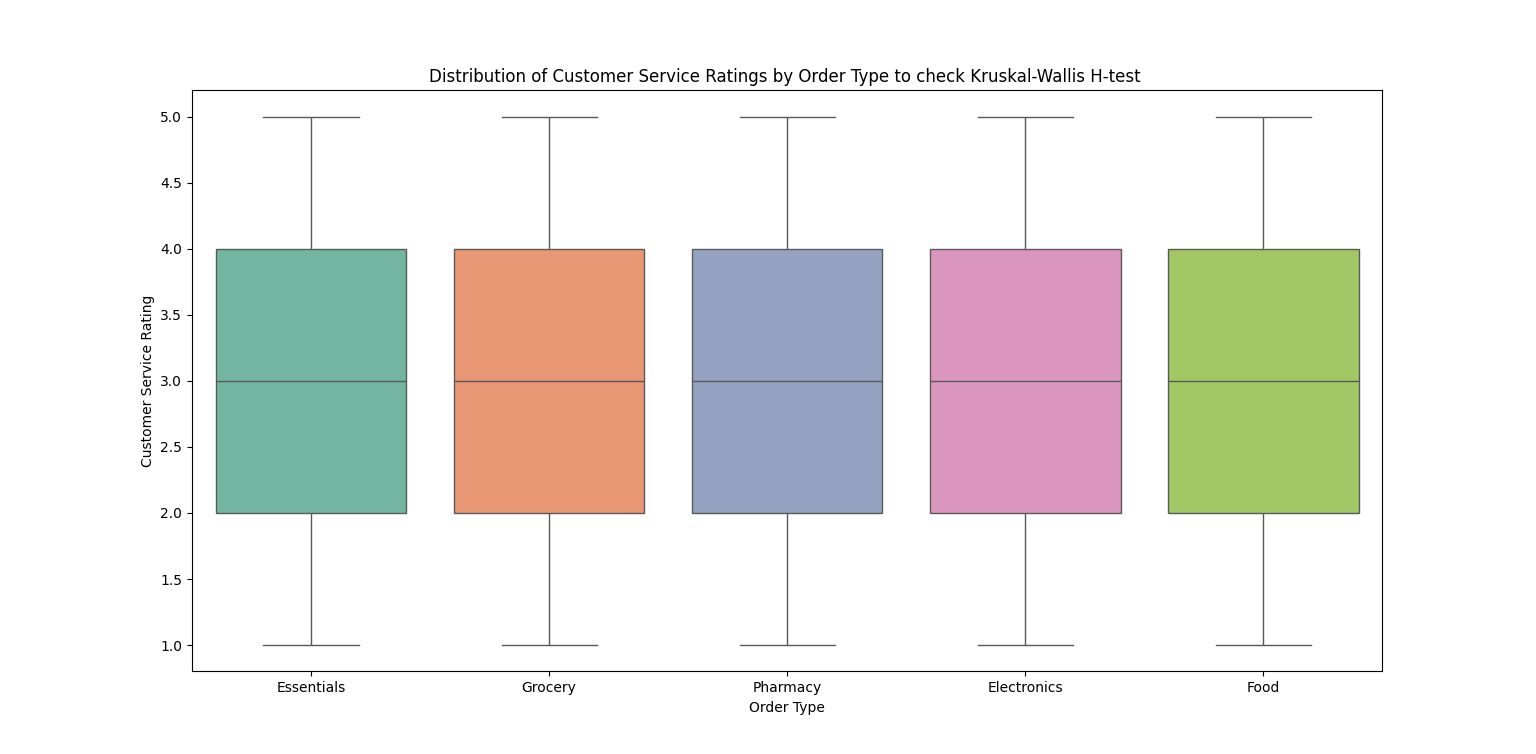
**Output**:

**Kruskal-Wallis H-statistic**: 1.17

**P-value**: 0.884

**Interpretation of Results:**

Since the **p-value** (0.8837) is **greater than 0.05**, we **fail to reject the null hypothesis**. This means that there is **no significant difference** in the **Customer Service Ratings** across the different **Order Types**.



**Explanation of the Plot**

* If the **medians** and **box sizes (IQRs)** are **similar across all order types**, and the **whiskers** and **outliers** are **comparable**, it suggests that there is **no significant difference** in **Customer Service Ratings** between the different **Order Types**.
* This aligns with the result from the **Kruskal-Wallis H-test**, where you found a **high p-value (0.8837)**, meaning the differences in **Customer Service Ratings** between the **Order Types** are **not statistically significant**.

**3. Chi-Square Test for Independence to test whether there is an association between Discount Applied and Customer Feedback Type.**

**Hypothesis Framing**

Null Hypothesis (H₀): There is no significant association between Discount Applied and Customer Feedback Type.

Alternative Hypothesis (H₁): There is a significant association between Discount Applied and Customer Feedback Type.

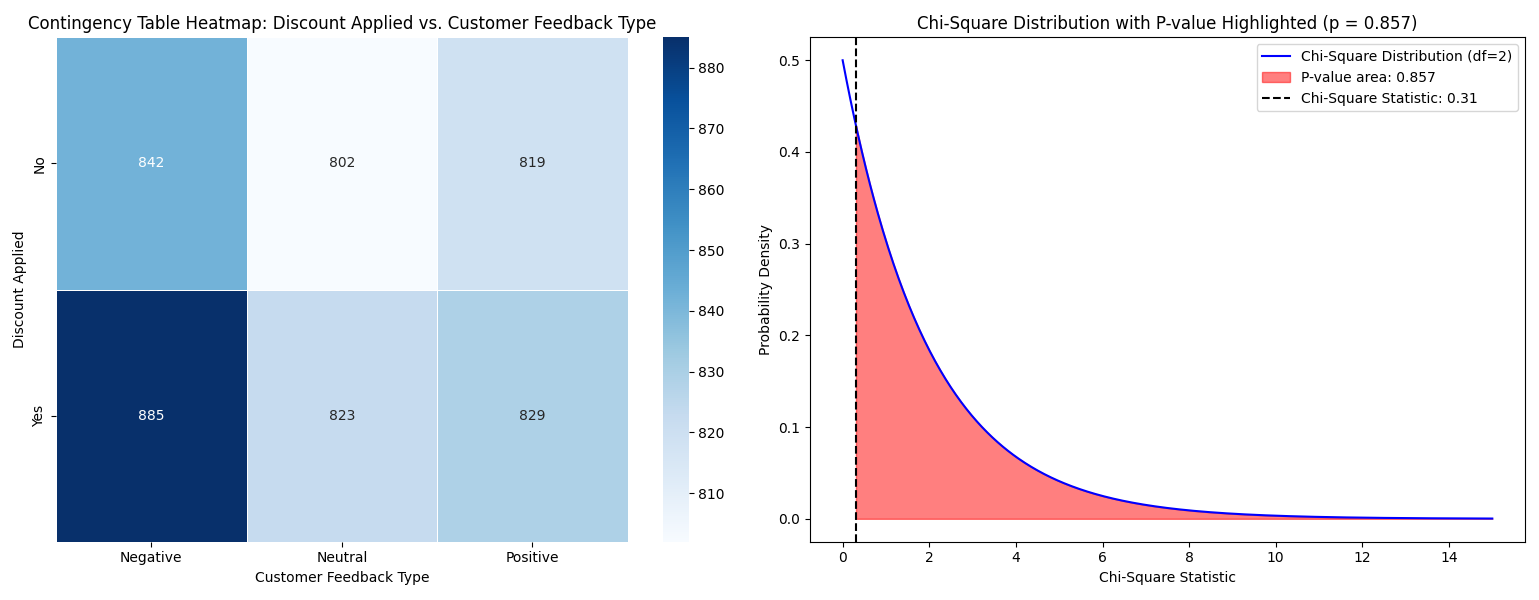
**Output**:

Chi-Square Statistic: 0.30757433117971655

P-value: 0.8574544974143203

**Interpretation of Results:**

* **Fail to reject the null hypothesis**: The p-value of 0.857 is much larger than the standard threshold of 0.05, meaning there is insufficient evidence to suggest that there is a significant association between Discount Applied and Customer Feedback Type.This indicates that Discount Applied does not appear to have a strong or significant impact on Customer Feedback Type in your dataset.

****

**Explanation of the Plot**

* **Contingency Table Heatmap**: It shows the observed frequencies for combinations of **Discount Applied** and **Customer Feedback Type**. The heatmap allows you to quickly visualize the distribution of feedback across different discount levels, and in this case, there is no clear pattern of association.
* **Chi-Square Distribution Plot**: This plot shows the distribution of Chi-Square statistics, with the observed statistic marked. Since the p-value is large, the statistic does not fall in the extreme tail of the distribution, reinforcing the conclusion that there is no significant association between **Discount Applied** and **Customer Feedback Type**.

**4. Two-way ANOVA to examine how both the Agent Name and Location affect Customer Service Rating, including the interaction effect.**

**Effect of Agent Name (C(Agent\_Name)):**

* Null Hypothesis (H₀): There is no significant effect of Agent Name on Customer Service Rating.
* Alternative Hypothesis (H₁): There is a significant effect of Agent Name on Customer Service Rating.

**Effect of Location (C(Location)):**

* Null Hypothesis (H₀): There is no significant effect of Location on Customer Service Rating.
* Alternative Hypothesis (H₁): There is a significant effect of Location on Customer Service Rating.

**Interaction Between Agent Name and Location:**

* Null Hypothesis (H₀): There is no significant interaction effect between Agent Name and Location on Customer Service Rating.
* Alternative Hypothesis (H₁): There is a significant interaction effect between Agent Name and Location on Customer Service Rating.

**Output**:

**Effect of Agent Name (C(Agent\_Name)):**

* **F-statistic**: 0.656
* **p-value**: 0.579
* **Interpretation**: Since the p-value for Agent Name is **greater than 0.05**, we **fail to reject the null hypothesis**. This indicates that **Agent Name does not have a significant effect on Customer Service Rating**. In other words, the ratings do not significantly differ based on which agent was involved.

**Effect of Location (C(Location)):**

* **F-statistic**: 2.316
* **p-value**: 0.0135
* **Interpretation**: Since the p-value for Location is **less than 0.05**, we **reject the null hypothesis**. This indicates that **Location has a significant effect on Customer Service Rating**. Different locations have different customer service ratings, and the location of the service seems to play a role in how customers rate the service.

**Interaction Effect Between Agent Name and Location (C(Agent\_Name):C(Location)):**

* **F-statistic**: 0.974
* **p-value**: 0.503
* **Interpretation**: Since the p-value for the interaction effect is **greater than 0.05**, we **fail to reject the null hypothesis**. This means that there is **no significant interaction** between **Agent Name** and **Location**. The effect of the agent on customer service ratings does not depend on the location, nor does the effect of the location depend on the agent.

**Interpretation of Results:**

* Agent Name: No significant effect on Customer Service Rating.
* Location: Significant effect on Customer Service Rating.
* Interaction between Agent Name and Location: No significant interaction.



**Explanation of the Plot**

* The Location factor has an observable effect on Customer Service Ratings (as reflected by the varying average rating levels across locations).
* The Agent Name does not have a clear, distinct effect on the ratings (as reflected by similar lines for different agents).

**5. Chi-Square Test of Independence for Price Range vs Location**

**Hypothesis Framing**

**Null Hypothesis (H₀):** There is no significant relationship between Price Range and Location

**Alternative Hypothesis (H₁):** There is a **significant relationship** between **Price Range** and **Location**

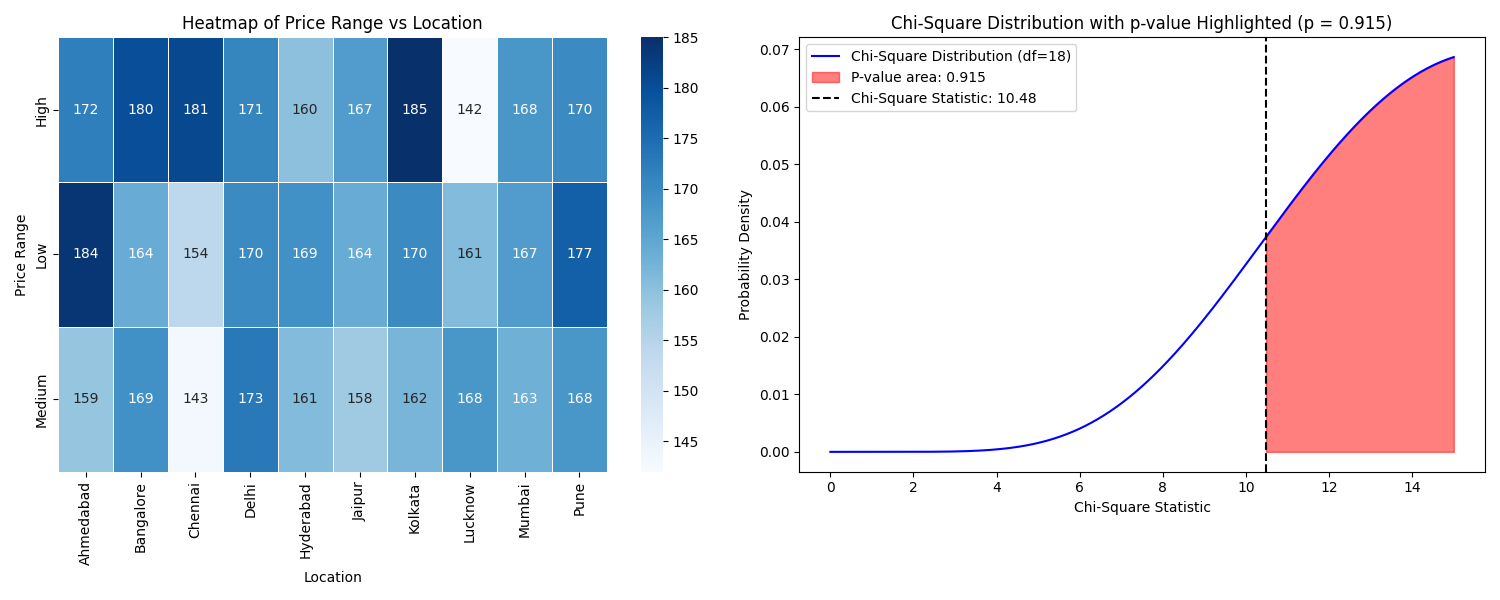
**Output**:

**Chi-Square Statistic**: 10.48

**p-value**: 0.9151

**Interpretation of Results:**

Since the **p-value (0.9151)** is much **greater than** the significance level (α = 0.05), we **fail to reject** the **null hypothesis (H₀)**.There is **no significant relationship** between **Price Range** and **Location** based on the data. In other words, the distribution of price ranges across different locations does not appear to be significantly different.



**Explanation of the Plot**

**Contingency Table Heatmap**: The heatmap visually shows that the observed frequencies are close to the expected frequencies, indicating no significant deviations.

**Chi-Square Distribution Plot**: The observed statistic (10.48) falls within the area where we would expect to see values under the null hypothesis, which further confirms that the null hypothesis should **not** be rejected (p-value = 0.9151).

**6. Chi-Square Test of Independence for Price Range vs Order Type**

**Hypothesis Framing**

**Null Hypothesis (H₀):** There is no significant relationship between Price Range and Order Type.

**Alternative Hypothesis (H₁):** There is a significant relationship between Price Range and Order Type.

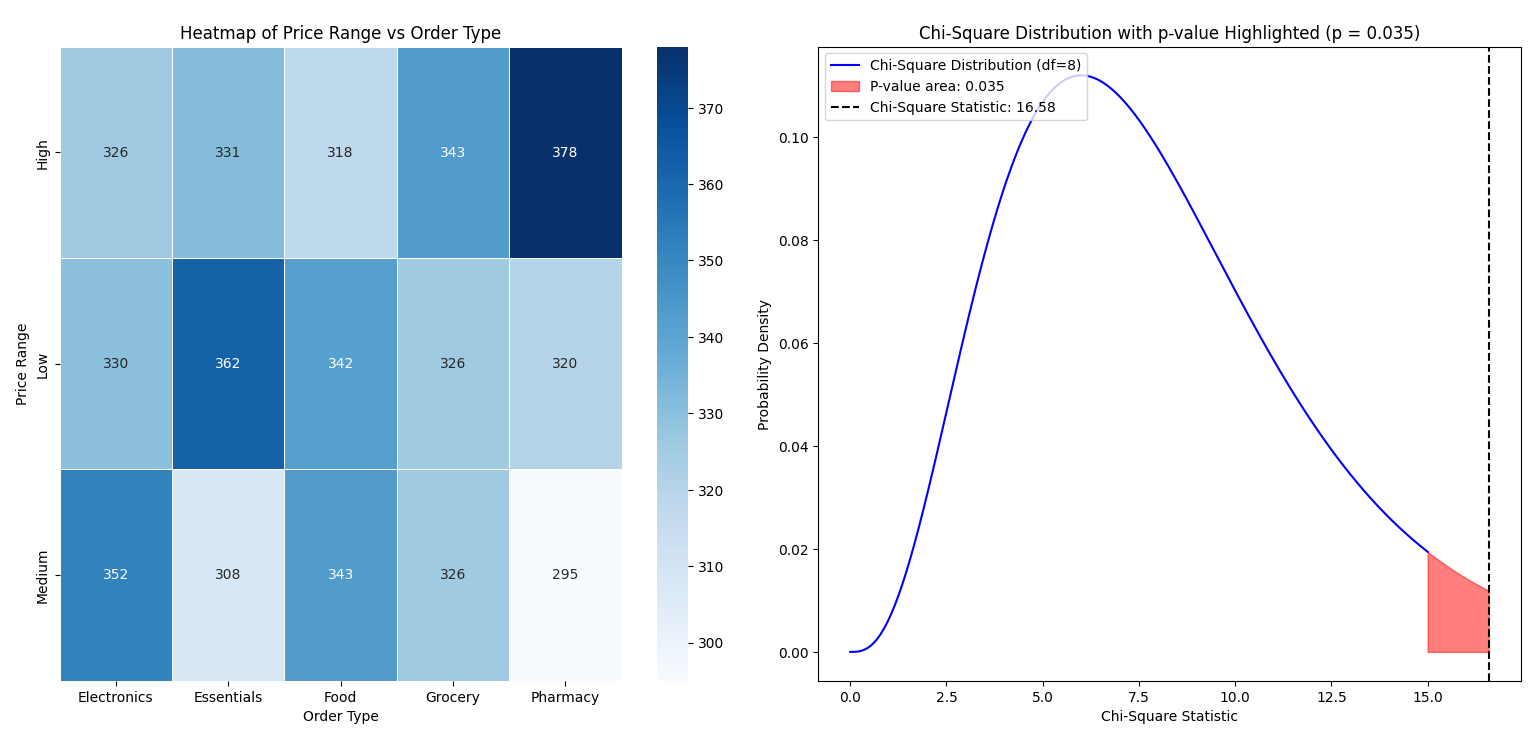
**Output:**

**Chi-Square Statistic:** 16.58

**p-value:** 0.0347

**Interpretation of Results:**

The results suggest that there is a **significant relationship** between **Price Range** and **Order Type**.This implies that the **distribution of prices** varies across different order types. For example, certain types of orders may be associated with higher or lower price ranges, which shows that **order type** might influence or be influenced by **price range**.

****

**Explanation of the Plot**

* Contingency Table Heatmap: The heatmap will visually show the frequencies for each Price Range x Order Type combination. Darker shades would indicate higher frequencies, while lighter shades represent lower frequencies. This helps to spot the most common combinations of Price Range and Order Type.
* Chi-Square Distribution Plot: The plot of the Chi-Square distribution will highlight the area representing the p-value. This helps to visually confirm that the Chi-Square statistic (16.58) lies in the tail of the distribution, supporting the rejection of the null hypothesis.

**7. Two-Way ANOVA to Compare Delivery Time differs by Location and Order Type**

**Hypothesis Framing**

**Effect of Location(C(Location)):**

* Null Hypothesis (H₀) : There is no significant effect of Location on Delivery Time.
* Alternative Hypothesis (H₁) : There is a significant effect of Location on Delivery Time.

**Effect of Order Type(C(Order Type)):**

* Null Hypothesis (H₀) : There is no significant effect of Order Type on Delivery Time.
* Alternative Hypothesis (H₁): There is a significant effect of Order Type on Delivery Time.

**Interaction between Location** and **Order Type:**

* Null Hypothesis (H₀): There is no significant interaction effect between Location and Order Type on Delivery Time.
* Alternative Hypothesis (H₁): There is a significant interaction effect between Location and Order Type on Delivery Time.

**Output**:

**Effect of Location (C(Location))**:

* **F-Statistic**: 0.683430
* **p-value**: 0.724676
* **Interpretation**: Since the p-value (0.724676) is greater than the significance level (α = 0.05), we fail to reject the null hypothesis. This means that there is no significant effect of Location on Delivery Time.

**Effect of Order Type (C(Order\_Type))**:

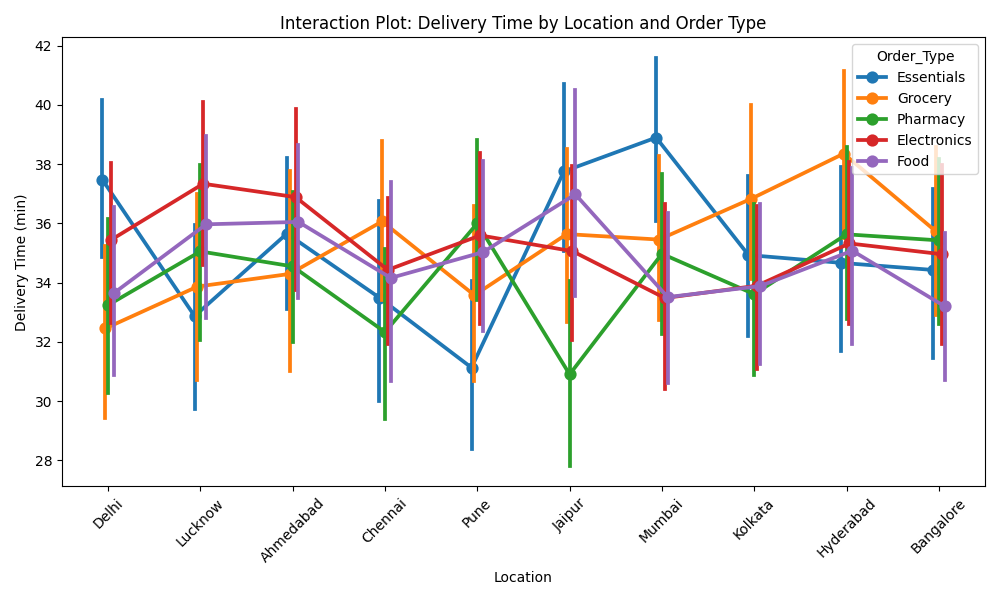
* **F-Statistic**: 0.984051
* **p-value**: 0.414814
* **Interpretation**: Since the p-value (0.414814) is greater than the significance level (α = 0.05), we fail to reject the null hypothesis. This means that there is no significant effect of Order Type on Delivery Time.

**Interaction Effect (C(Location):C(Order\_Type))**:

* **F-Statistic**: 1.456133
* **p-value**: 0.038539
* **Interpretation**: Since the p-value (0.038539) is less than the significance level (α = 0.05), we reject the null hypothesis. This means that there is a significant interaction effect between Location and Order Type on Delivery Time.

**Interpretation of Results:**

* For Location: No significant effect on Delivery Time (p = 0.7247 > 0.05).
* For Order Type: No significant effect on Delivery Time (p = 0.4148 > 0.05).
* For Interaction: Significant interaction between Location and Order Type on Delivery Time (p = 0.0385 < 0.05).

****

**Explanation of the Plot**

* If the lines are non-parallel, this supports the rejection of the null hypothesis for the interaction effect, meaning Location and Order Type have a significant combined effect on Delivery Time.
* If the lines are parallel or there’s little variation in Delivery Time for different Order Types at each Location, it supports failing to reject the null hypothesis for the Location and Order Type effects individually.

**5.3.2. NLP (Natural Language Processing) Analysis**

**1. Sentiment analysis on the Review Text and then determine if there is association between Sentiment and Customer Feedback Type.**

Null Hypothesis (H₀): There is no significant association between Sentiment analysis and Customer Feedback Type.

Alternative Hypothesis (H₁): There is a significant association between Sentiment analysis and Customer Feedback Type.

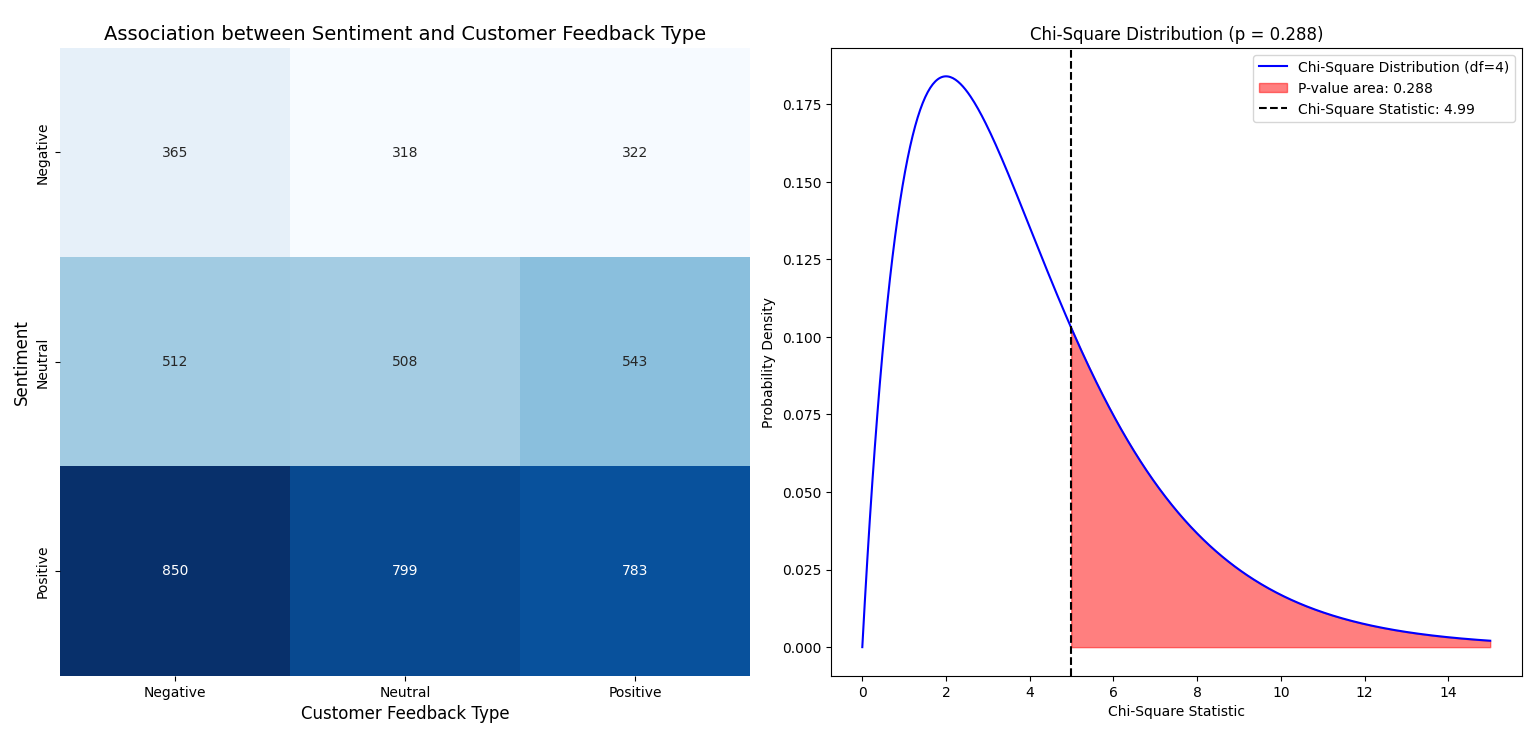
**Output:**

**Chi-Square Statistic:** 4.99

**p-value:** 0.288

**Interpretation of Results:**

Since the p-value is greater than 0.05, we fail to reject the null hypothesis. This means that there is no significant association between Sentiment analysis and Customer Feedback Type based on the data. The relationship observed in the data is not strong enough to suggest a statistically significant connection.

****

**Explanation of the Plot**

**Heatmap:** The heatmap visualized the frequency distribution of different combinations of Sentiment and Customer Feedback Type, with darker colors indicating higher occurrences and lighter colors indicating lower occurrences. This gave an intuitive view of how the two variables are related in the dataset.

**Chi-Square Distribution Plot:** The distribution plot visually confirmed the result. The Chi-Square statistic was not in the extreme tail of the distribution, and the shaded area (representing the p-value) further indicated a lack of statistical significance.

**2. Cosine Similarity between Sentiment analysis of review text and Customer Feedback Type**

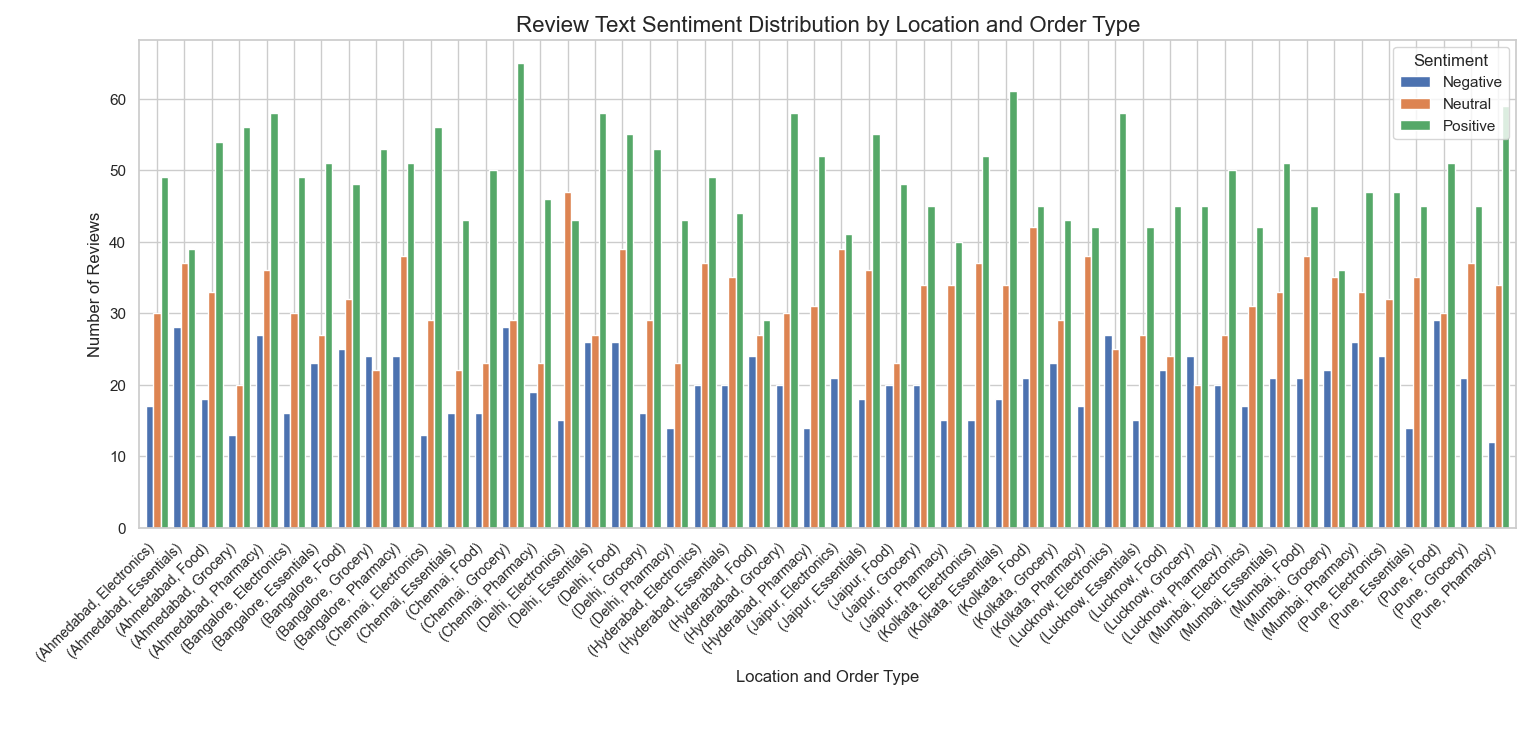
**Output:**

Mean Cosine Similarity: 0.33133784

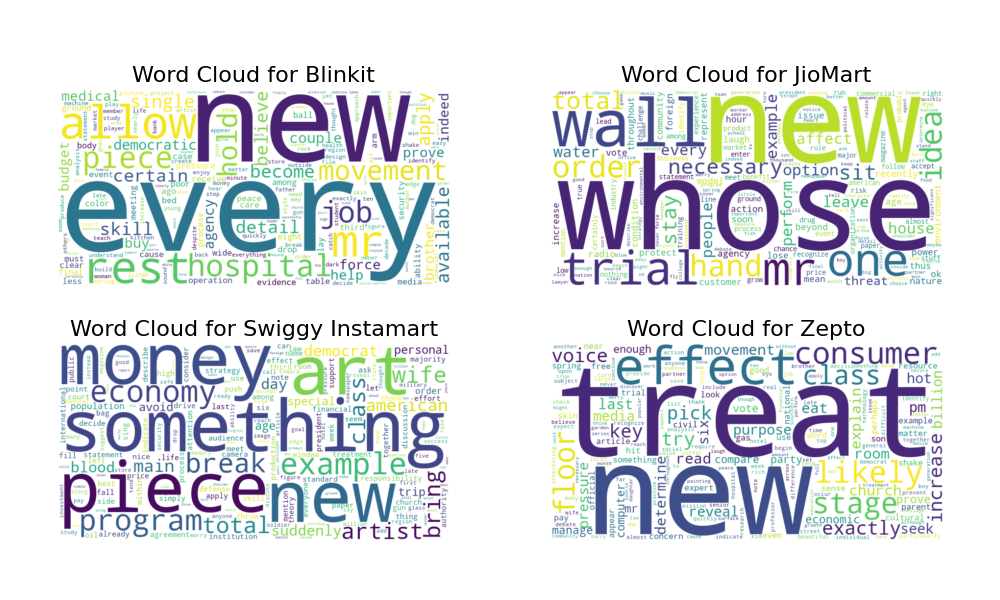
Low Similarity: A mean similarity of 0.33 suggests that there isn't a strong direct relationship between the sentiment expressed in the reviews and the type of customer feedback. In other words, the sentiment might not always align with the feedback type, and the two columns may be driven by different factors.

Cosine Matrix: The individual similarity values show varying degrees of similarity. Some might be close to 1, indicating that for some rows, sentiment and feedback are similar. However, the matrix shows a lot of disparity, suggesting that the two columns do not always align.

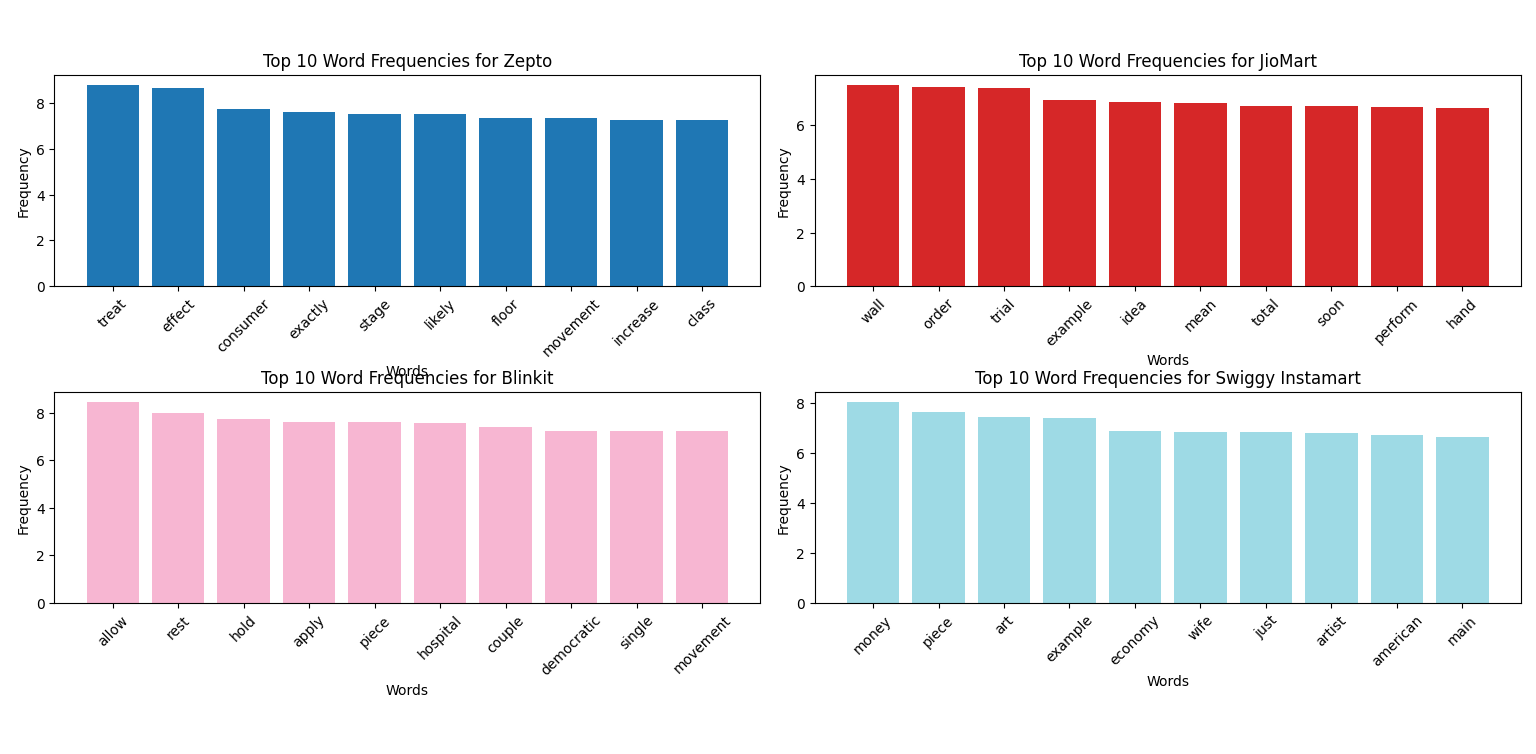
**3. To analyse Review text sentiment distribution by Location and Order Type**

****

**4. Creating Word cloud for Agent Names**



**5. Agent wise key word extraction by TfidfVectorizer**



**Results and conclusion**

The Chi-Square test for independence conducted between Order Type and Product Availability showed no significant association, meaning that the type of order does not impact product availability. Similarly, the Kruskal-Wallis H-test for Customer Service Ratings across different Order Types revealed no significant differences, indicating that the method of placing an order does not influence the way customers rate their service. A Chi-Square test was also applied to assess the relationship between Discount Applied and Customer Feedback Type, and the results showed no significant relationship, suggesting that the application of discounts does not affect the type of feedback customers provide. In a Two-Way ANOVA examining the effects of Agent Name and Location on Customer Service Ratings, it was found that location significantly affected customer ratings, while the agent did not. There was no significant interaction between the agent and location, suggesting that customer ratings are more influenced by location than the specific agent handling the service.

Another Chi-Square test was conducted for Price Range and Location, which showed no significant relationship, indicating that price distribution across locations does not vary significantly. However, a Chi-Square test for Price Range and Order Type revealed a significant relationship, suggesting that different order types are associated with different price ranges, possibly reflecting customer preferences or pricing strategies. A Two-Way ANOVA examining the effects of Location and Order Type on Delivery Time showed that while location and order type individually did not significantly affect delivery time, their interaction did.

This indicates that the effect of order type on delivery time can vary depending on the location.

In the realm of Natural Language Processing (NLP), sentiment analysis showed no significant association with Customer Feedback Type, implying that sentiment in reviews does not strongly influence the feedback customers provide. The Cosine Similarity between sentiment and feedback type was low, suggesting that these two variables are not closely related and may be driven by different factors. Sentiment also varied across Location and Order Type, revealing important insights into customer satisfaction across different regions and product categories.

Word Clouds for each agent highlighted the most common themes and keywords mentioned by customers, offering insights into agent-specific feedback. Additionally, the Agent-wise Keyword Extraction using TfidfVectorizer identified key terms related to each agent, helping understand the specific customer concerns associated with different agents.