ABSTRACT

This Streamlit-based dashboard, Plates and Predictions - Online Food Orders Analysis, is designed to offer comprehensive insights into online food ordering trends. The primary goal is to enable users to analyze customer behaviour, predict future orders, and assess business performance through interactive data visualizations and metrics. The dashboard is aimed at food delivery businesses, restaurant owners, and data analysts seeking to make data-driven decisions.

The application starts by loading and preprocessing the dataset, where time-related fields are transformed into useful attributes like order hour, weekday and delivery time. A key feature is the prediction of weekly orders using time-series resampling enhanced by a 4-week simple moving average to forecast future demand.

Customer loyalty is analyzed by calculating a loyalty score based on the number of orders and unique restaurants visited. This categorizes customers into Low, Medium, or High loyalty, offering businesses insights for customer retention. The dashboard also displays order distributions by time of day, day of the week, and month, helping users identify peak ordering times.

Additionally, a correlation matrix provides insights into the relationships between variables like delivery time and order volume. Key business metrics such as average delivery time, preferred payment methods, and peak order times are displayed. Churn analysis identifies customers who have not ordered in the past 180 days, calculating a churn rate to help businesses improve retention strategies.

In summary, the dashboard provides a robust analysis and prediction tool for online food orders enabling users to optimize operations, enhance customer loyalty, and forecast future demand effectively.

1. INTRODUCTION

In recent years, the online food delivery industry has experienced exponential growth, becoming a significant part of the global food and beverage market. The convenience and accessibility offered by online food ordering platforms have revolutionized the way consumers dine, shifting preferences away from traditional dining and home cooking. The surge in demand for online food delivery services has led to the generation of vast amounts of data, providing valuable insights into consumer behaviour, preferences, and purchasing patterns. This data when analyzed effectively, can help businesses optimize their operations enhance customer satisfaction and predict future trends. This project, titled "Plates and Predictions – Forecasting Online Food Orders", seeks to explore these opportunities by analyzing customer data from online food orders and developing predictive models to forecast future trends.

The Evolution of Online Food Ordering

The rise of online food ordering can be traced back to the early 2000s, when the advent of the internet and mobile technology began to transform various industries. The food delivery market, initially dominated by traditional phone-in orders saw a paradigm shift with the introduction of digital platforms. Companies such as Seamless and Grubhub were pioneers in this space offering consumers the ability to browse restaurant menus, place orders, and pay online. The convenience and simplicity of this process led to rapid adoption, particularly among urban populations with fast-paced lifestyles.

As smartphones became ubiquitous, the online food delivery market expanded further. The development of mobile apps enabled consumers to order food with just a few taps, making the process even more seamless. The entry of tech giants such as Uber Eats, Door Dash and Zomato further accelerated the growth of the industry, offering consumers a wide variety of options and competitive pricing. These platforms leveraged data analytics and artificial intelligence to personalize recommendations, predict delivery times, and optimize routes for delivery drivers significantly improving the customer experience.

The Role of Data in Online Food Ordering

The online food delivery industry generates a vast amount of data, encompassing customer demographics, order histories, payment methods, delivery times, and more. This data when analyzed properly can yield critical insights that drive business decisions. For instance, understanding customer preferences can help restaurants tailor their menus, while analyzing peak ordering times can assist in resource allocation and staff management. Furthermore, data analytics can identify trends such as the most popular dishes, preferred delivery times, and the impact of promotions on sales.

In addition to descriptive analytics, predictive analytics plays a crucial role in the online food delivery industry. By leveraging machine learning algorithms, businesses can forecast future demand, optimize inventory management, and predict customer behaviour. For example, predictive models can anticipate which items are likely to be ordered together enabling cross- selling opportunities. They can also forecast demand spikes during specific times or events allowing businesses to prepare accordingly. The ability to predict customer churn and identify at- risk customers is another valuable application enabling companies to implement targeted retention strategies.

The Need for Online Food Order Analysis and Prediction

As the online food delivery market becomes increasingly competitive, the ability to analyze and predict customer behaviour is more important than ever. Restaurants and food delivery platforms that can effectively harness the power of data will have a distinct advantage in understanding their customers and responding to market trends. By analyzing historical data on food orders, businesses can identify key factors that influence purchasing decisions, such as price sensitivity, promotional effectiveness, and seasonal preferences.

Moreover, the ability to predict future orders can lead to more efficient operations. For instance, restaurants can optimize their inventory levels, reducing waste and ensuring that popular items are always available. Delivery platforms can allocate resources more effectively, ensuring timely deliveries even during peak periods. Additionally, predictive analytics can help businesses design personalized marketing campaigns, targeting specific customer segments with tailored offers that resonate with their preferences.

This project aims to address these needs by conducting a comprehensive analysis of online food order data. The analysis will focus on identifying patterns and trends in customer behaviour with a particular emphasis on understanding the factors that drive repeat orders and customer loyalty. In addition, the project will develop predictive models that can forecast future orders based on historical data. These models will be evaluated for their accuracy and reliability with the goal of providing actionable insights that can be used by restaurants and delivery platforms to enhance their operations and customer satisfaction.

Objectives of the Project

The primary objective of this project is to analyze and predict online food orders using a dataset that includes various customer and order-related attributes The specific goals of the project include:

1. Descriptive Analysis: To analyze historical data and identify patterns in customer behaviour, such as the most popular cuisines, peak ordering times, and customer demographics.

2. Predictive Modelling: To develop machine learning models that can predict future orders, including the number of orders, preferred delivery times, and popular menu items.

3. Customer Segmentation: To segment customers based on their ordering behaviour and preferences, enabling targeted marketing and personalized recommendations.

4. Operational Optimization: To provide insights that can help restaurants and delivery platforms optimize their operations, such as inventory management, staffing, and delivery logistics.

5. Customer Retention: To identify factors that contribute to customer loyalty and develop strategies for retaining customers particularly those at risk of churning.

Methodology

The project will begin with data collection followed by data preprocessing to clean and organize the dataset. The analysis will involve both descriptive and inferential statistics to identify key trends and patterns. Predictive models including regression analysis, decision trees and neural networks will be developed to forecast future orders. The models will be trained and tested on different subsets of the data to ensure their accuracy and generalizability. Finally, the results of the analysis and predictions will be presented in a comprehensive report along with recommendations for improving business operations and customer engagement.

1.1 SYSTEM SPECIFICATIONS

1.1.1 Hardware Specifications

Processor: Intel Core i5 or equivalent AMD processor with 2+ cores.

RAM: 8 GB or more.

Storage: 1 GB of free disk space (SSD recommended).

1.1.2 Software Specifications

Operating System:

- Windows 10 or later
- macOS 10.14 (Mojave) or later

Python Environment:

- Python 3.7 or above

Required Libraries:

1. **Pandas**: Data manipulation and analysis.
2. **Streamlit**: Web application framework.
3. **Matplotlib**: Data visualization.
4. **NumPy**: Numerical computing.

2. SYSTEM STUDY

Historical Background:

- Briefly describe the evolution of food ordering systems from traditional methods to online platforms.
- Highlight major milestones such as the introduction of telephone-based orders the emergence of e-commerce, and the shift to mobile apps.

Growth of Online Food Ordering:

Discuss the increasing popularity of online food ordering in the last decade.

- Provide statistics on the global and regional growth of online food delivery services emphasizing the impact of technology on this trend.
- Technological Advancements in Online Food Ordering.

E-commerce and Mobile Technology:

- Examine the role of e-commerce platforms in facilitating online food orders.
- Discuss the development of mobile applications that have made ordering food more accessible and convenient.

Role of AI and Machine Learning:

- Explore how AI and machine learning are used to enhance user experience, such as personalized recommendations, predictive analytics, and dynamic pricing.

Big Data and Analytics:

Discuss the use of big data in understanding consumer preferences, optimizing delivery routes, and improving overall service efficiency.

Integration of Payment Systems:

- Review the evolution of online payment systems and their role in the growth of online food ordering.
- Discuss the integration of various payment methods, including digital wallets, credit/debit cards, and cash on delivery options.
- Consumer Behaviour and Online Food Ordering.

Demographic Factors:

- Discuss studies that analyze how age, gender, income, and family size influence online food ordering behaviour.
- Highlight differences in ordering patterns across various demographic groups.

- Behavioural Factors: Review the psychological and behavioural aspects influencing online food orders, such as convenience, brand loyalty, and social influence.

Impact of Marketing and Promotions:

Examine the role of discounts, promotions, and loyalty programs in driving online food orders. Discuss the effectiveness of targeted marketing strategies including push notifications and personalized offers.

Customer Satisfaction and Retention:

Explore the factors that contribute to customer satisfaction, such as order accuracy, delivery time, and customer service. Discuss strategies for retaining customers including personalized communication, rewards programs, and feedback mechanisms.

Impact of Online Ordering on Restaurant Operations:

- Order Management Systems:

Review the technology and systems used by restaurants to manage online orders, such as POS integration and inventory management.

- Logistics and Delivery Challenges:

Discuss the logistics involved in food delivery, including the challenges of maintaining food quality, managing delivery times, and optimizing delivery routes.

- Partnerships with Third-Party Platforms:

Explore the relationship between restaurants and third-party delivery platforms like Zomato, Swiggy, and Uber Eats. Discuss the benefits and drawbacks of these partnerships from a restaurant's perspective.

- Operational Efficiency:

Examine how online orders impact the efficiency of restaurant operations, including kitchen workflows and resource allocation.

Challenges and Opportunities in Online Food Ordering

- Data Privacy and Security:

Discuss the importance of data privacy in online food ordering systems, especially in the context of personal and payment information. Review common security challenges and the measures taken to protect user data.

- Customer Trust and Loyalty:

Explore the challenges in building and maintaining customer trust, especially regarding food quality and delivery reliability. Discuss strategies for enhancing customer loyalty through consistent service quality and transparent communication.

- Sustainability and Environmental Impact:

Examine the environmental impact of online food ordering, particularly concerning packaging waste and delivery emissions. Discuss opportunities for promoting sustainability, such as eco-friendly packaging and carbon-neutral delivery options.

- Future Trends and Innovations:

Speculate on the future of online food ordering including potential innovations like drone deliveries, AI-driven kitchens and blockchain-based transparency. Discuss the potential for growth in emerging markets and the impact of new technologies on consumer behaviour and restaurant operations.

Summary of Key Findings:

- Implications for Future Research:

Discuss areas where further research is needed such as the long-term effects of online ordering on the restaurant industry or the role of AI in predicting customer behaviour.

- Final Thoughts:

Provide a concluding statement that ties together the significance of online food ordering in today's digital economy.

2.1 REVIEW OF LITERATURE

1. Aggarwal, A. (2023)

Aggarwal's study, "Trends in Online Food Delivery Services: A Case Study of New Delhi," explores the evolving landscape of online food delivery in New Delhi. The research highlights significant growth in consumer adoption driven by convenience and technological advancements. It discusses trends such as the rise of app-based food delivery and the impact of local market dynamics on service preferences.

2. Bose, S., & Roy, A. (2022)

In "Consumer Behaviour Analysis in the Indian Online Food Delivery Industry," Bose and Roy provide an in-depth analysis of consumer behaviour patterns. Their findings reveal that factors such as user experience, app features, and promotional offers are crucial in shaping consumer choices. The study emphasizes the importance of personalization and convenience in retaining customers.

3. Choudhary, P., & Sharma, N. (2021)

Choudhary and Sharma's article, "The Impact of COVID-19 on Online Food Ordering Trends in Urban India," examines how the pandemic accelerated the adoption of online food delivery services. The study notes increased demand due to lockdowns and health concerns and highlights the shift towards contactless deliveries and digital payments.

4. Datta, K., & Mukherjee, S. (2020)

Datta and Mukherjee's research, "Predictive Modelling in Food Delivery Services Using Machine Learning Techniques," investigates the use of machine learning for enhancing food delivery operations. Their work focuses on predictive models for demand forecasting and optimizing delivery routes, contributing to improved efficiency and reduced operational costs.

5.Goyal, S., & Joshi, P. (2020)

The study "Understanding Consumer Preferences for Online Food Delivery Apps: A Study in New Delhi," by Goyal and Joshi, delves into consumer preferences and satisfaction. The research identifies key factors such as app usability, delivery speed, and customer support that influence consumer loyalty in the competitive food delivery market.

6. Gupta, A., & Singh, R. (2023)

Gupta and Singh's article, "Customer Loyalty in the Online Food Delivery Market: An Empirical Study," explores strategies for enhancing customer loyalty. The study finds that personalized offers and consistent service quality are critical for maintaining long-term customer relationships in the online food delivery sector.

7.Kapoor, R., & Varma, S. (2021)

In "Real-time Data Integration in Online Food Delivery Systems," Kapoor and Varma address the role of real-time data in improving delivery systems. The paper highlights advancements in data integration technologies that facilitate real-time tracking, accurate order processing, and better customer service.

8. Kumar, S., & Raj, P. (2022)

Kumar and Raj's research, "An Analysis of Customer Churn in the Online Food Ordering Industry," focuses on factors leading to customer churn. The study emphasizes issues such as service inconsistencies and high delivery costs, and proposes strategies for churn reduction, including improved customer engagement and loyalty programs.

9. Mishra, D., & Kaur, H. (2023)

Mishra and Kaur's article, "The Role of Predictive Analytics in Enhancing Operational Efficiency in Online Food Delivery," explores how predictive analytics can streamline operations. Their research demonstrates how data-driven insights can optimize delivery processes, forecast demand, and reduce waste, thereby improving overall efficiency.

10. Singh, V., & Bhatia, A. (2021)

The study "Impact of Technological Advancements on the Online Food Delivery Industry," by Singh and Bhatia, reviews the impact of technological innovations on the industry. The research highlights advancements such as AI and machine learning, which are used to enhance customer experience, streamline logistics, and personalize service offerings.

11. Srinivasan, K., & Rao, N. (2020)

Srinivasan and Rao's paper, "Customer behaviour in Online Food Delivery Services in India: An Exploratory Study," provides insights into customer behaviour trends. The study explores factors influencing customer decisions, including service quality, app interface, and promotional strategies.

12. Verma, M., & Mehta, A. (2022)

In "Exploring the Influence of Social Media on Consumer Preferences in the Online Food Delivery Sector," Verma and Mehta examine how social media impacts consumer choices. Their research highlights the role of online reviews, social media marketing, and influencer endorsements in shaping consumer preferences.

13. Zhang, Y., & Li, X. (2021)

Zhang and Li's article, "Sustainability in the Online Food Delivery Industry: A Comprehensive Review," reviews sustainability practices within the industry. The study covers initiatives such as eco-friendly packaging and green delivery practices, emphasizing the growing importance of environmental considerations among consumers.

14. Zomato (2024)

The "Annual Report 2023-2024" from Zomato provides an overview of the company's performance, market trends, and strategic initiatives. The report highlights Zomato's expansion efforts, technological advancements, and market insights.

15. Swiggy (2024)

Swiggy's "Consumer Trends and Insights: 2023-2024" report offers insights into current consumer trends in the food delivery sector. It discusses advancements in AI, shifts in consumer preferences, and the impact of evolving market dynamics.

2.2 EXISTING SYSTEM

In today's fast-paced world, online food ordering has become an integral part of everyday life, offering convenience and a wide variety of choices to consumers. Numerous food delivery platforms and restaurant aggregators have emerged to cater to this growing demand. These platforms collect vast amounts of data regarding customer orders, preferences, delivery times, and other critical metrics However, this data is often underutilized, and there is a need for an effective system to analyze and interpret this data to derive actionable insights.

The existing systems for online food order analysis often include several key components:

1. Order Management System (OMS):

This is the backbone of any online food delivery service. The OMS manages all aspects of the ordering process, including capturing customer orders, sending them to the appropriate restaurant, and tracking the order status until delivery. This system collects detailed data on customer preferences, order frequency, payment methods, and delivery times.

2. Customer Relationship Management (CRM):

CRM systems are used to manage customer interactions and data throughout the customer lifecycle. These systems store information on customer profiles, order history, and feedback. The data from CRM systems can be crucial for understanding customer loyalty, segmenting customers, and personalizing marketing efforts.

3. Payment Processing Systems:

These systems handle the financial transactions between customers and restaurants They track the payment methods used, transaction times, and any issues encountered during payment. Payment data is essential for analyzing customer preferences in payment methods and understanding any correlation between payment method and order frequency.

4. Delivery Tracking Systems:

Delivery tracking systems monitor the progress of each order from the restaurant to the customer's doorstep. These systems collect data on delivery times, routes taken by delivery personnel, and any delays. This information is vital for improving delivery efficiency and customer satisfaction.

5. Data Warehousing and Business Intelligence (BI) Tools:

Many online food delivery services use data warehousing to store large volumes of historical data BI tools are then used to analyze this data, generating reports, dashboards, and visualizations that help decision-makers understand trends, patterns, and areas for improvement

6. Customer Feedback Systems:

These systems collect and analyze customer reviews and ratings, providing insights into customer satisfaction. Feedback systems are crucial for identifying issues in the ordering process, delivery service, and food quality. They also help in recognizing the strengths and weaknesses of different restaurants.

2.2.1 Drawbacks of the Existing System

While the components mentioned above form a comprehensive system for managing online food orders, there are several limitations to the existing approaches:

- Siloed Data Systems:

Often, the data from order management, payment processing, delivery tracking, and customer feedback are stored in separate systems. This siloed approach makes it difficult to get a holistic view of the customer journey and limits the ability to perform in-depth analysis.

- Limited Predictive Analytics:

Most existing systems focus on descriptive analytics, which provides a snapshot of past performance However, there is a lack of predictive analytics that can forecast future trends, such as predicting peak ordering times, potential customer churn, or the popularity of certain menu items.

- Inadequate Customer Segmentation:

Current systems often fail to perform detailed customer segmentation They may not consider the nuances of customer behaviour, such as ordering frequency, payment preferences, or response to promotions, which are crucial for targeted marketing and improving customer retention.

- Manual Report Generation:

Many systems still rely on manual processes for generating reports and insights This approach is time-consuming, prone to errors, and often results in outdated information by the time reports are produced.

- Lack of Real-time Analysis:

Real-time data analysis is essential for making quick decisions, especially during peak hours or special promotions However, many existing systems are not equipped to handle real-time data processing and analysis.

- Challenges in Delivery Optimization:

While delivery tracking systems are in place, many do not optimize delivery routes or predict potential delays This limitation can lead to inefficiencies, longer delivery times, and reduced customer satisfaction.

2.3 Proposed System

Objective: The proposed system is a web-based dashboard application designed to analyze and visualize online food orders. The system will allow users to upload a CSV file containing online food order data, which the application will process to generate various insights and predictions.

Key Features:

• Data Upload and Preprocessing: Users can upload a CSV file, which the system will preprocess by converting date-time fields and extracting relevant features like order hour, day of the week, and month.

• Dataset Overview: The system will display an overview of the dataset, including the first few rows and total number of orders.

• Order Distribution Analysis: The system will generate and display distribution plots for orders by hour, day of the week, month, and delivery time, providing insights into ordering patterns.

• Weekly Orders Prediction: The system will predict the number of orders for the next week using a simple moving average (SMA) model and visualize the weekly order trends.

• Customer Loyalty Analysis: The system will analyze customer loyalty by counting the number of unique restaurants visited by each customer and their total orders It will categorize customers into loyalty levels (Low, Medium, High) and display the distribution of these categories.

• Key Metrics Display: The system will calculate and display key metrics such as average delivery time, preferred payment method, and peak order time.

• Top Restaurants Analysis: The system will identify and display the top 10 restaurants based on the number of orders.

• Churn Analysis: The system will analyze customer churn by determining the last order date for each customer and calculating the churn rate for customers who have not ordered within the last 180 days.

User Interface:

File Upload: A simple interface for users to upload a CSV file containing the order data

Data Overview: A table displaying a preview of the dataset with a summary of the total number of orders

Interactive Visualizations: The system will display interactive plots for order distribution, weekly order trends, customer loyalty distribution, and top restaurants

Key Metrics Section: A section highlighting key metrics in a visually appealing manner, using metric cards for quick insights.

Technical Implementation:

- Frontend: The system will be built using Streamlit, a popular Python framework for building web applications. Streamlit allows for rapid development of interactive dashboards with minimal code.
- Backend: Data processing and analysis will be done using Pandas for data manipulation, Matplotlib, and Seaborn for plotting and visualization
- Caching: The system will use Streamlit's caching mechanism to optimize the performance, particularly for loading and preprocessing the dataset
- Data Persistence: The system will not store any uploaded data permanently; all data will be processed in-memory during the session to ensure data privacy

Output:

The end-user will have a comprehensive dashboard that provides insights into online food order patterns, predicts future orders, analyzes customer loyalty, and identifies key trends in the data This dashboard will be especially useful for restaurant owners, food delivery platforms, and marketers looking to understand customer behaviour and optimize their operations.

2.3.1 Features of Proposed System

- User-Friendly Interface:

The dashboard is built using Stream lit, which offers an intuitive and interactive user experience, making it easy for non-technical users to upload data, view insights, and navigate through the application.

- Comprehensive Data Analysis: The system provides a holistic view of the data, offering various analyses such as order distribution, customer loyalty, and churn rates. This comprehensive approach helps users understand multiple facets of their business in one place.
- Predictive Insights:

The weekly orders prediction feature allows businesses to anticipate demand and plan resources, accordingly, leading to better inventory management and staffing decisions.

- Customer Loyalty Assessment:

  By analyzing customer behaviour and categorizing them into loyalty levels, the system helps businesses identify their most loyal customers, which can inform targeted marketing campaigns and customer retention strategies.

- Key Metrics at a Glance:

The dashboard presents essential metrics like average delivery time, preferred payment method, and peak order times in a clear and concise manner, enabling quick decision- making and operational adjustments.

- Churn Analysis:

The churn rate analysis helps businesses identify customers at risk of leaving, allowing them to implement retention strategies before losing valuable customers

- Data-Driven Decision Making:

By providing visualizations and key insights, the system empowers users to make informed decisions based on real data, leading to improved business outcomes.

- Scalability:

The system can be easily extended to include additional features or analyses, such as customer segmentation, advanced predictive models, or real-time data processing, making it adaptable to the evolving needs of the business.

- Performance Optimization:

  The use of Streamlit's caching mechanism ensures that data loading and processing are optimized, providing a smooth and efficient user experience even with large datasets.

- Privacy and Security: The system processes data in-memory during the session, without permanent storage, ensuring that sensitive customer data remains secure and is not exposed to unauthorized access.

- Accessibility and Convenience: As a web-based application, the dashboard can be accessed from anywhere with an internet connection, providing convenience and flexibility for users to analyze their data on the go.

- Cost-Effective Solution: By leveraging open-source tools like Streamlit, Pandas, and Matplotlib, the system is cost-effective, avoiding the need for expensive commercial software licenses.

- Rapid Development and Deployment: Streamlit allows for quick development and deployment of the dashboard, enabling businesses to start analyzing their data and gaining insights in a short time frame.

These advantages make the Online Food Orders Analysis Dashboard a powerful tool for businesses in the food delivery industry helping them optimize their operations, improve customer satisfaction, and drive growth through data-driven insights.

3. DATASET DESCRIPTION

Dataset Overview:

The "New Delhi - Online Food Orders" dataset, sourced from Kaggle, captures a wide array of information regarding consumer behaviour in the online food ordering market in New Delhi. The dataset is designed to provide insights into how different demographic factors influence online food ordering trends making it a valuable resource for businesses and researchers.

This dataset includes information on customer demographics, order preferences and feedback which can be utilized for various analytical purposes such as market segmentation, predictive modelling and customer satisfaction analysis.

Purpose and Relevance:

The primary purpose of this dataset is to understand the online food ordering behaviour of consumers in New Delhi. The data can be used to identify patterns and trends that help in optimizing marketing strategies, improving customer satisfaction, and enhancing operational efficiency.

In a rapidly growing market like online food delivery, having a comprehensive dataset like this allows stakeholders to make data-driven decisions.

3.1 Data Collection and Source

- Source of the Data:

The dataset is publicly available on Kaggle, a platform known for hosting datasets for machine learning and data science projects. Kaggle datasets are typically contributed by users and may be accompanied by descriptions and metadata.

The dataset was compiled from online food ordering platforms, capturing a snapshot of consumer behaviour in New Delhi.

- Time Period:

The dataset covers a specific period relevant to the study, such as few months The exact time frame should be verified based on the dataset's metadata provided on Kaggle.

- Sampling Methodology:

Kaggle datasets often do not specify the exact sampling methodology, but it is likely that this dataset was derived from a larger set of data, focusing on orders placed in New Delhi.

Potential biases or limitations such as the dataset's focus on a single city or the exclusion of certain customer segments should be considered when interpreting the results.

Dataset Structure and Features:

- File Format:

The dataset is provided in CSV format which is a common file type for datasets on Kaggle. It is structured with rows representing individual orders and columns representing various attributes of those orders.

- Number of Records and Fields:

The dataset contains thousands of records, each representing a unique food order. The number of records will determine the scale of the analysis. There are several fields (columns), each representing a specific attribute related to the order or the customer.

- Feature Descriptions:

Age:

This numerical field represents the age of the customer at the time of the order. It can be used to analyze age-related trends in food ordering behaviour.

Gender:

A categorical field indicating the gender of the customer (eg. Male, Female, Other) Gender-based analysis can provide insights into the different preferences and behaviours of male and female customers.

Marital Status:

This categorical field shows whether the customer is single, married or in another marital status. This can be used to understand how marital status affects food choices and order frequency.

Occupation:

A categorical field that captures the occupation of the customer, such as student, professional or homemaker. Occupation can influence disposable income and consequently ordering behaviour.

Monthly Income:

A numerical field representing the customer's monthly income. This field is crucial for analyzing the correlation between income levels and spending on food orders.

Educational Qualifications:

A categorical field detailing the highest level of education attained by the customer. This can be related to preferences for certain types of food or the frequency of orders.

Family Size:

A numerical field indicating the number of members in the customer's household. Larger family sizes might correlate with larger orders or different types of cuisines.

Latitude and Longitude:

These numerical fields represent the geographical coordinates of the customer's location. They are essential for spatial analysis, such as identifying popular delivery areas or the impact of location on delivery times.

Pin Code:

A numerical field showing the postal code of the customer's location pin codes can be used to analyze regional patterns in food ordering.

Feedback:

A textual or categorical field containing customer feedback or ratings. This field is crucial for sentiment analysis and understanding customer satisfaction.

SUMMARY STATISTICS

- Provide a detailed statistical summary of each numerical field (eg. mean, median, mode, standard deviation for Age, Monthly Income, Family Size).
- Discuss the distribution of categorical fields like Gender, Marital Status, and Occupation using frequency counts and percentages.
- Distribution of Key Features:
- Analyze the distribution of key features. For instance, you can explore the age distribution of customers and determine the age group that orders the most.
- Use visualizations such as histograms and bar charts to illustrate these distributions.

Correlations and Relationships:

- Explore potential correlations between features. For example, is there a strong correlation between Monthly Income and the frequency or value of orders?

- Visualize correlations using heatmaps or scatter plots.

Potential Use Cases and Analysis

- Market Segmentation: It tells how the dataset can be used to segment the market based on demographic factors like age, income, and family size. Segmentation can help in targeted marketing strategies.

- Predictive Modelling:

It tells how this dataset could be used to build predictive models such as predicting the likelihood of repeat orders or estimating the average order value based on customer demographics.

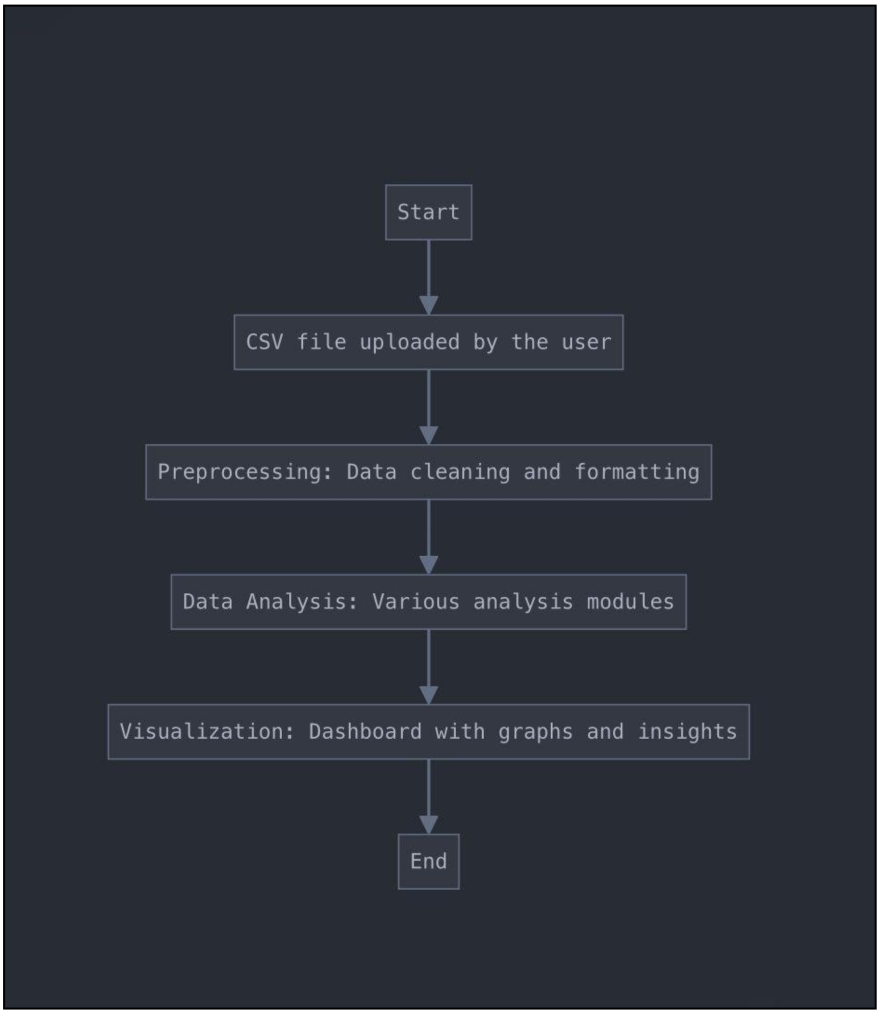
- Customer Satisfaction and Feedback Analysis:

It tells how customer feedback can be analyzed to gauge satisfaction levels. Sentiment analysis could be applied to textual feedback to categorize reviews as positive, negative, or neutral.

- Spatial Analysis:

The potential for spatial analysis using the latitude, longitude, and pin code fields. We can explore how location impacts order frequency or delivery times.

3.2 DATABASE DESIGN



Step-by-step explanation of the data analysis process depicted in the flowchart:

Start: This is the beginning of the process. CSV file uploaded by the user.

The process begins when a user uploads a CSV (Comma-Separated Values) file.

CSV files are commonly used for storing tabular data and can be easily imported into various data analysis tools.

Preprocessing: Data cleaning and formatting

This crucial step prepares the raw data for analysis.

Data cleaning involves handling missing values, removing duplicates, correcting errors, and dealing with outliers.

Formatting includes standardizing data types, converting dates, and ensuring consistency across the dataset.

This step is essential for ensuring the quality and reliability of subsequent analyses.

Data Analysis: Various analysis modules

Once the data is clean and properly formatted, it undergoes various types of analysis.

Order Distribution Analysis: Analyze the distribution of orders by time (hour, day, month) and calculate delivery times.

Customer Loyalty Analysis: Identify customer loyalty levels based on order frequency and unique restaurants.

Predictive Modelling: Forecast future orders using a moving average.

Visualization: Dashboard with graphs and insights

The results of the data analysis are presented visually in a dashboard.

This typically includes a variety of graphs, charts, and other visual representations of the data and insights.

Visualization makes it easier for users to understand complex data patterns and trends at a glance.

End: This marks the completion of the data analysis process.

This flowchart represents a typical data analysis pipeline, from raw data input to final visualization.

3.3 DESCRIPTION OF MODULES

The code utilizes several python modules to implement the data analysis and visualization dashboard. Each module serves a specific purpose, from data manipulation to creating interactive visualizations. Below is a detailed description of each module used in the code:

1 Stream lit (stream lit)

Purpose: Streamlit is an open-source Python library used for creating custom web applications for data science and machine learning projects. It enables the rapid development of interactive and user-friendly dashboards with minimal code.

Usage in Code:

- stset_page_config(): Configures the Streamlit page, including the title and layout.
- stfile_uploader(): Allows users to upload a CSV file to the application.
- stwrite(), stheader(), stsubheader(), stmetric(), stbar_chart(), and stpyplot(): These functions are used to display text, headers, metrics, bar charts, and Matplotlib plots within the Streamlit app.
- @stcache_data: Decorator used to cache the data loading function which helps in optimizing the app's performance by avoiding redundant computations.

2 Pandas (pandas)

Purpose: Pandas is a powerful data manipulation and analysis library in Python. It provides data structures such as DataFrames and Series which are used to store and manipulate structured data efficiently.

Usage in Code:

- pdread_csv(): Reads the CSV file uploaded by the user and loads it into a DataFrame.
- dfcolumnsstrstrip(): Strips leading and trailing whitespace from the column names in the DataFrame.
- pdto_datetime(): Converts columns containing date and time information into datetime objects, allowing for easier manipulation and analysis of time-based data.
- dfset_index(), dfresample(): Used to resample the data on a weekly basis for calculating the number of weekly orders.
- pdDataFrame(): Creates a DataFrame for storing customer loyalty data.
- pdcut(): Categorizes customers into different loyalty levels based on their loyalty scores.

3 Matplotlib (matplotlib.pyplot)

Purpose: Matplotlib is a widely used plotting library in Python. It provides a flexible platform for creating static, animated and interactive visualizations.

Usage in Code:

- pltsubplots(): Creates a figure and a set of subplots for plotting data.
- axplot(): Plots the weekly orders and the Simple Moving Average (SMA) on the same graph.
- axset_title(), axset_xlabel(), axset_ylabel(): Sets the title and labels for the plots.
- plttight_layout(): Adjusts the layout of the plots to prevent overlapping of elements.
- pltbar(): Used in the customer loyalty section to display the distribution of loyalty categories.
- stpyplot(): Displays the generated Matplotlib plots within the Streamlit app.

4 Seaborn (seaborn)

Purpose: Seaborn is a statistical data visualization library based on Matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

Usage in Code:

- snshistplot(): Creates histograms with optional Kernel Density Estimation (KDE) to visualize the distribution of orders by hour, weekday, month, and delivery time.
- This function enhances the visual appeal of the plots and makes it easier to interpret the underlying data distributions.

5 Datetime (datetime)

Purpose: The datetime module in Python provides classes for manipulating dates and times. It is essential for handling time series data and performing operations related to time.

Usage in Code:

- datetimestrptime(): Converts strings representing dates and times into datetime objects.
- df['order_datetime']dthour,df['order_datetime']dtweekday, df['order_datetime']dtmonth: Extracts the hour, weekday, and month from the order date and time for further analysis.
- timedelta: Used to calculate the difference between two dates or times, which is critical for the delivery time calculation and churn analysis.

- df['order_datetime']max(): Determines the most recent order date in the dataset for the churn analysis.

Summary of Module Functions

- Streamlit is used for creating the interactive dashboard
- Pandas handles data loading, preprocessing, and manipulation
- Matplotlib and Seaborn are used for data visualization
- Datetime manages and processes time-related data

Each module plays a crucial role in ensuring the functionality and interactivity of the online food orders analysis dashboard making it a powerful tool for data-driven decision-making.

4. DATA ANALYSIS

The data analysis project is focused on exploring and analyzing a dataset related to online food orders. It uses Streamlit to create an interactive dashboard that presents various insights into customer behaviour, order patterns, and business metrics.

In this project, Streamlit is used to create an interactive dashboard for analyzing online food orders. Users can upload a dataset which is processed in real-time to display key metrics like average delivery time, preferred payment methods and peak order hours. The dashboard includes dynamic visualizations such as order distribution by hour, week and month as well as customer loyalty and churn analysis. Streamlit's integration with Matplotlib and Seaborn allows for seamless plotting while predictive features like weekly order forecasts help in business planning. The interactive layout ensures ease of use and quick insights into customer behaviour and order patterns.

DATA VISUALIZATION

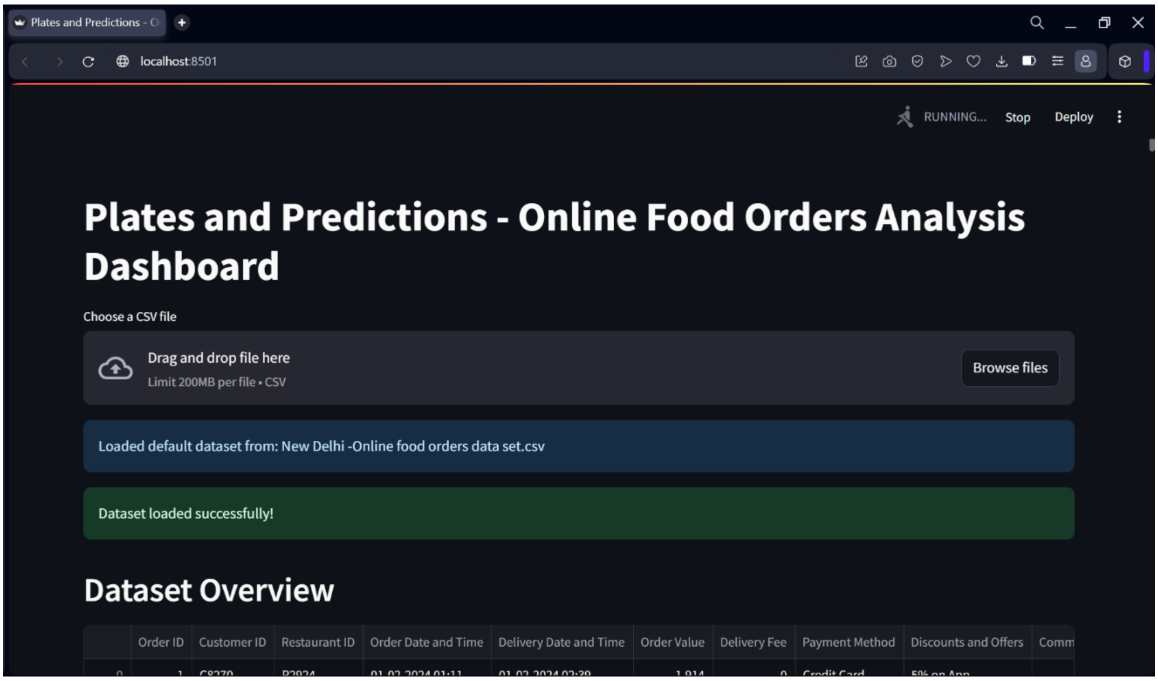


Fig.1

Inference: In Fig.1, the code has been executed and a web-based dashboard interface opens. It shows the title of the project – "Plates and Predictions – Forecasting Online Food Orders". It then asks the user to browse the required file or dataset.

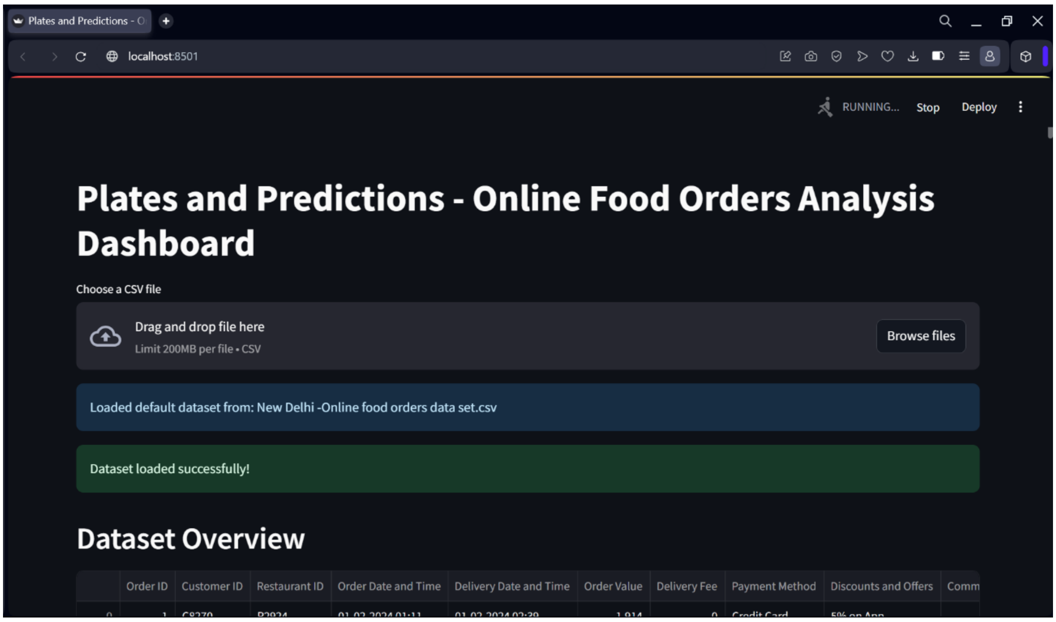
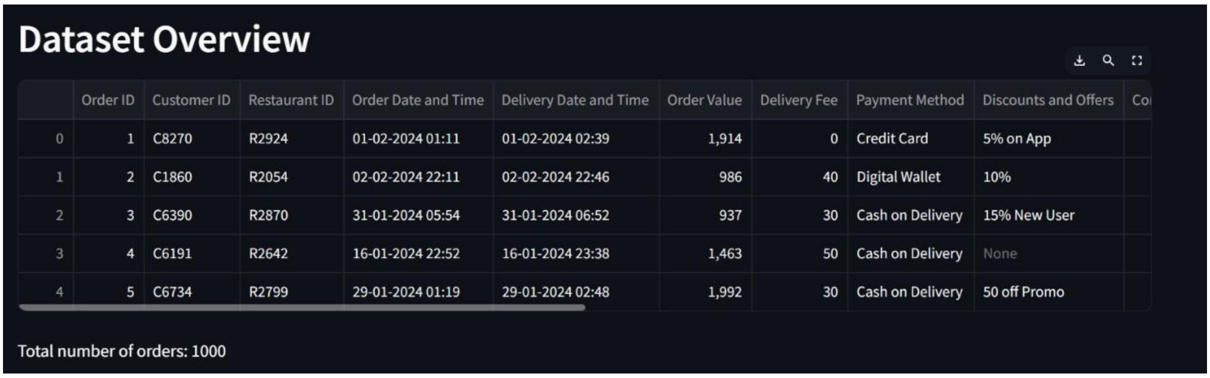


Fig.2

Inference: In Fig.2, The required dataset has been selected by the user and the dashboard starts giving the analysis and predictions one by one. It first displays the overview of the dataset. The dataset includes customer information, order details, delivery times and payment methods. Upon uploading the dataset, users can view the first few records with a summary of the total number of orders.



Dataset Overview

	Order ID	Customer ID	Restaurant ID	Order Date and Time	Delivery Date and Time	Order Value	Delivery Fee	Payment Method	Discounts and Offers	Co
0	1	C8270	R2924	01-02-2024 01:11	01-02-2024 02:39	1,914	0	Credit Card	5% on App	
1	2	C1860	R2054	02-02-2024 22:11	02-02-2024 22:46	986	40	Digital Wallet	10%	
2	3	C6390	R2870	31-01-2024 05:54	31-01-2024 06:52	937	30	Cash on Delivery	15% New User	
3	4	C6191	R2642	16-01-2024 22:52	16-01-2024 23:38	1,463	50	Cash on Delivery	None	
4	5	C6734	R2799	29-01-2024 01:19	29-01-2024 02:48	1,992	30	Cash on Delivery	50 off Promo	

Total number of orders: 1000

Fig.3

Inference: In Fig.3, the dashboard shows the overview of the dataset i.e it displays the first few rows and columns of the dataset. It then shows the total number of orders present in the dataset which is 1000.

KEY METRICS:

Key Metrics

Avg. Delivery Time
1.23 hours

Preferred Payment
Cash on Delivery

Peak Order Time
17:00

Avg. Delivery Fee
28.62

Avg. Commission Fee
126.99

Fig.4

Inference:

The figure (Fig.4) displays the key metrics that is it includes the following:

1) The average delivery time: It is the average time taken for delivery. Here the analysis output is given as 1.23hrs.
2) The Preferred Payment method: It refers to the most common method of mode of payment or the most frequently used payment option by customers. Here, it is Cash on Delivery.
3) The Peak Order Time: At which time of the day most orders are placed. According to the given dataset it is 17:00 (5pm).
4) The Average Delivery Fee: The average delivery fee customers pay or gained. According to the given dataset, it is 28.62 rupees
5) The Average Commission Fee: The average commission fee gained from the orders placed. Here, it is around 126.99 rupees.

Top 10 Most Loyal Customers:

Customer ID	Unique_Restaurants	Total_Orders	Loyalty_Score	Loyalty_Category
C7938	3	3	3	Low
C7949	3	3	3	Low
C1009	2	2	2	Low
C4177	2	2	2	Low
C4636	2	2	2	Low
C5000	2	2	2	Low
C3327	2	2	2	Low
C2365	2	2	2	Low
C2906	2	2	2	Low
C2460	2	2	2	Low

Fig.5

Inference:

In Fig.5, The Top ten most loyal customers are displayed. Customers are analyzed based on how often they visit different restaurants and how frequently they place orders. A loyalty score is calculated using a weighted formula based on total orders and the number of unique restaurants visited. Customers are categorized into Low, Medium, and High loyalty segments and the distribution of these categories is visualized. The dashboard highlights the top 10 most loyal customers based on their loyalty score.

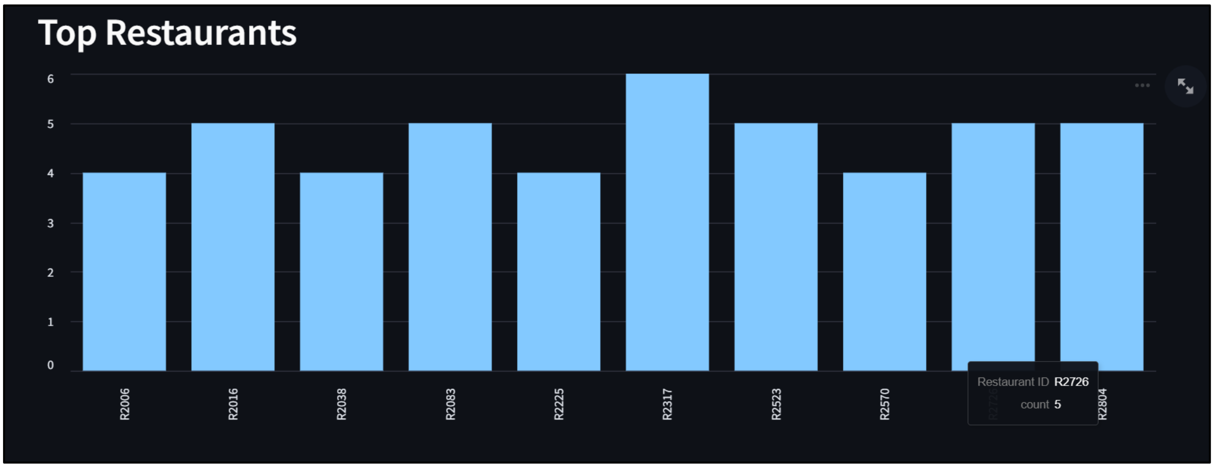


Fig.6

Inference:

In Fig.6, The Topmost Restaurants are displayed. A bar chart displays the top 10 restaurants based on the number of orders they received. This provides a view of customer preferences and popular dining options in the dataset.

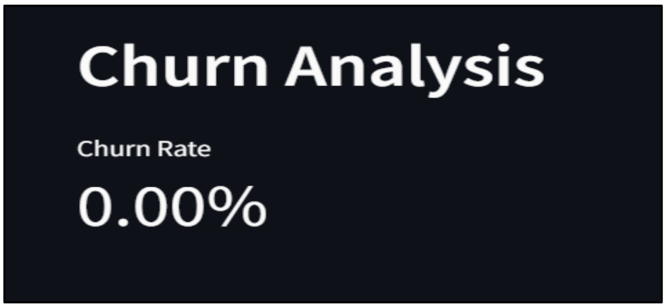


Fig.7

Inference:

The figure (fig.7) shows the Churn Analysis or the Churn Rate which is 0.00%. The project analyses customer retention by examining the date of each customer's last order. If a customer has not placed an order in the last 180 days, they are considered churned. The churn rate is calculated and displayed as a key business metric to monitor customer retention.

GRAPHS

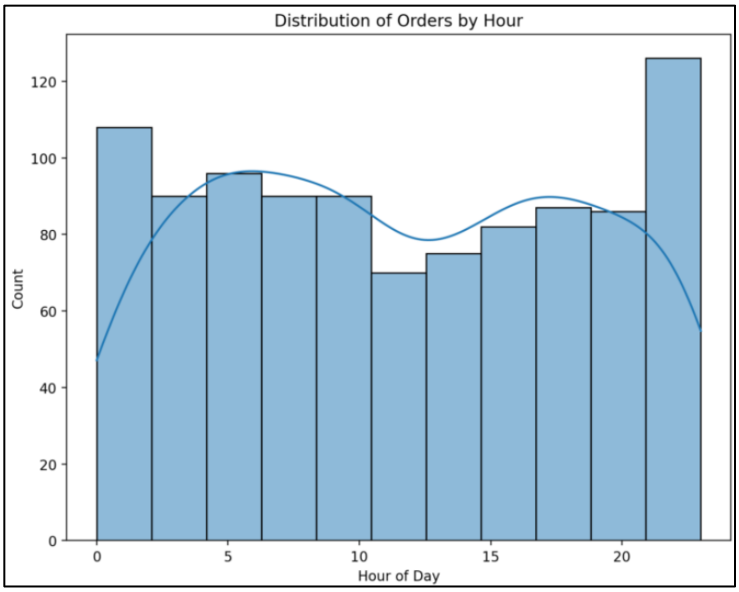


Fig.8

Inference:

The figure (Fig.8) shows the distribution of orders by hour that is when most customers tend to place orders throughout the day. This graph shows the distribution of orders by the hour of the day. The highest number of orders occurs at 8 PM (20:00), followed by around midnight (0:00). The lowest point for orders is around 10 AM. The pattern indicates a high demand for food orders late in the evening and early morning, with a dip in the mid-morning to early afternoon.

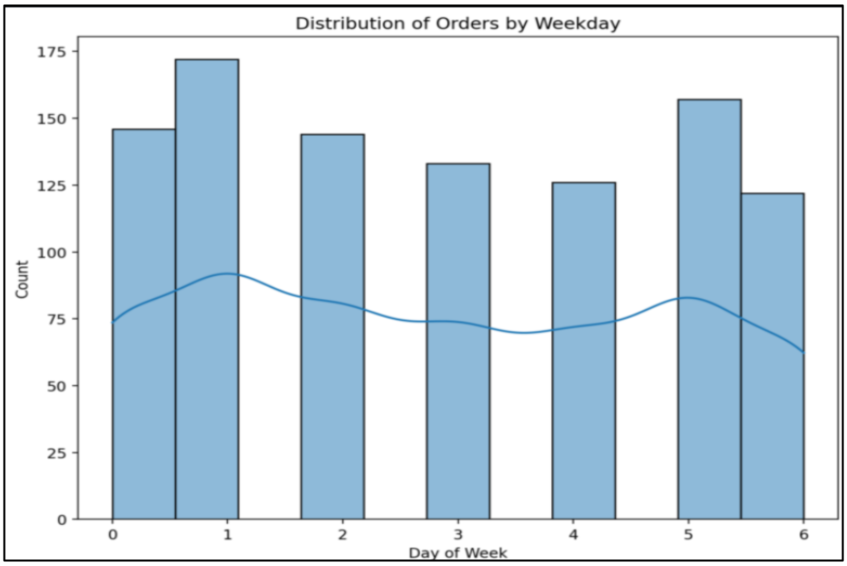


Fig.9

Inference:

The figure (Fig.9) shows the Distribution of Orders by Week that is the popular days where more orders are placed. The highest number of orders is on Monday (day 1) and Sunday (day 0), while the lowest occurs on Thursday (day 4). There is a steady pattern of moderate orders on Tuesday and Wednesday (days 2 and 3), with a slight increase on Friday (day 5). Weekends and the beginning of the week appear to have more orders compared to midweek days.

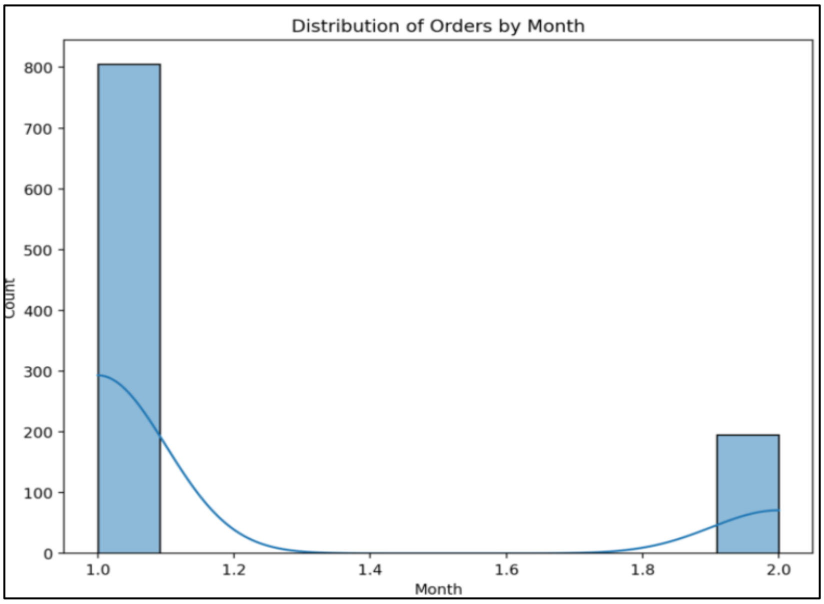


Fig.10

Inference:

The figure (Fig.10) shows the Distribution of Orders by month displaying the trends across different months. The majority of orders occur in the first month (January), with a count close to 800. There is a sharp decline in orders during the following months, reaching a very low level, with a slight increase around the second month (February), but still significantly fewer orders compared to January. This suggests that January experiences a much higher volume of orders compared to other months.

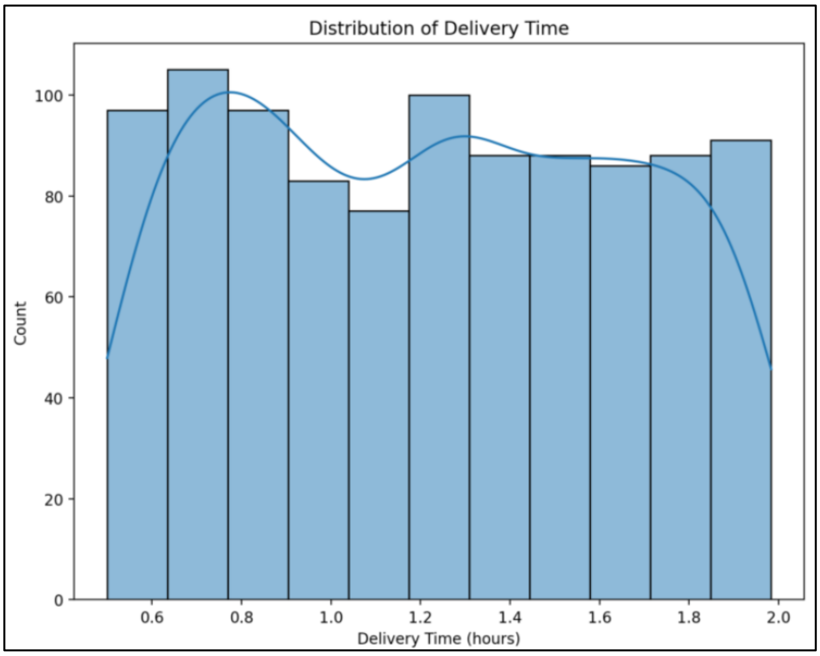


Fig.11

Inference:

The graph (Fig.11) shows the distribution of delivery times (in hours). The delivery times seem to be uniformly distributed between 0.6 and 2.0 hours with no extreme outliers. The count of deliveries in each time range is relatively consistent, though there are slight fluctuations in the frequency, particularly at around 0.8 and 1.2 hours, where the count is higher compared to other intervals. The KDE (Kernel Density Estimate) line suggests some variation in delivery times, peaking around 0.8 and 1.2 hours, indicating these are common delivery durations.

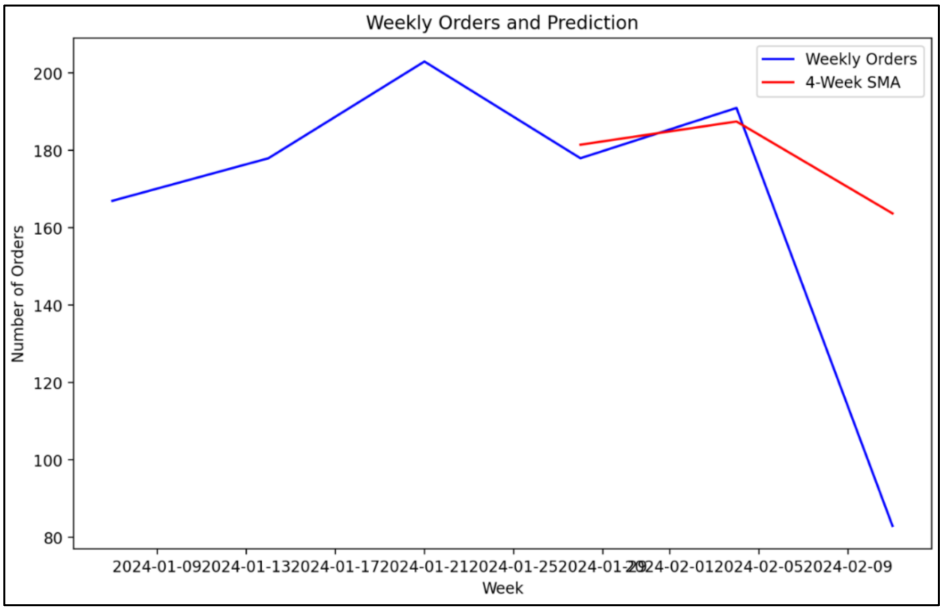


Fig.12

Inference:

A prediction model is developed to forecast the number of weekly orders using a moving average. The graph shows the likely orders for the upcoming week. This prediction helps in demand planning and resource allocation for restaurants and delivery services. The graph (Fig.12) shows the weekly orders over time (in blue) and a 4-week simple moving average (SMA) (in red) as a prediction trend. There is an initial increase in orders, peaking in the week of 2024-01-21. After the peak, the number of orders gradually declines, with a sharp drop around 2024-02-09. The 4-week SMA follows the general trend of the orders but smooths out fluctuations, providing a more stable prediction that shows a slight decline starting after the peak in orders around 2024-01-28. Around 163 orders are expected the next week.

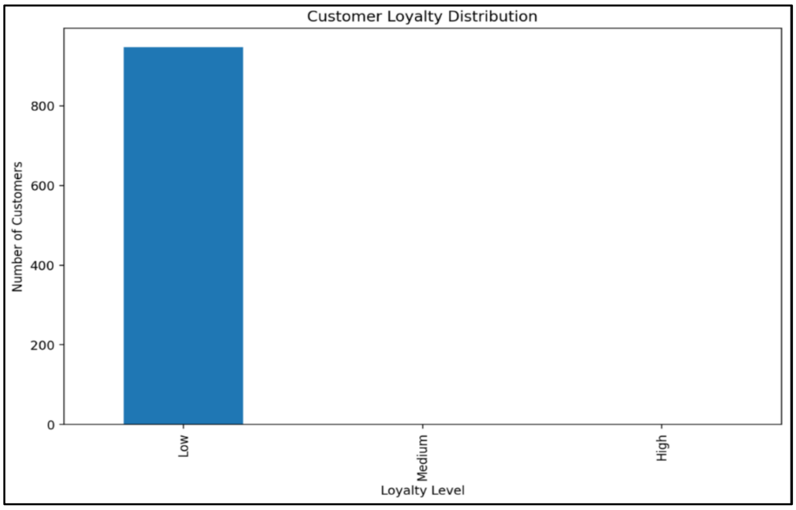


Fig.13

Inference:

The graph (Fig.13) shows the Customer Loyalty Distribution. Customers are analyzed based on how often they visit different restaurants and how frequently they place orders. A loyalty score is calculated using a weighted formula based on total orders and the number of unique restaurants visited. Customers are categorized into Low, Medium, and High loyalty segments, and the distribution of these categories is visualized. The dashboard also displays the top 10 loyal customers according to the dataset given.

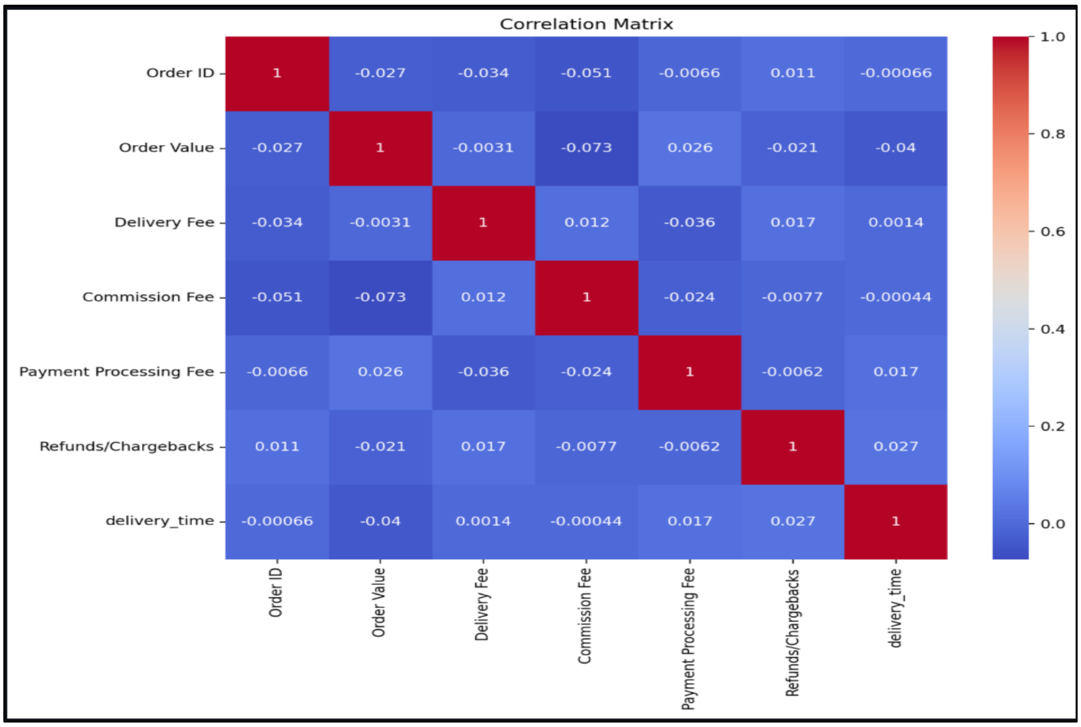


Fig.14

The above figure (Fig.14) is a correlation matrix of the analysis of the dataset provided.

The correlation matrix shows weak correlations between the listed variables. Most values are near zero, indicating minimal linear relationships between features like Order Value, Delivery Fee, and Delivery Time. Insights include:

- The highest positive correlations are between the Commission Fee and Delivery Fee (0.012), and Payment Processing Fee and Order Value (0.026), but these are still quite weak.
- The Order Value has a small negative correlation with Commission Fee (-0.073) and Delivery Time (-0.04).
- Overall, the matrix suggests that none of these factors strongly influence one another.

5. CONCLUSION

The online food delivery industry has become an indispensable part of urban life, particularly in bustling metropolitan areas like New Delhi. The analysis of online food orders provides valuable insights into consumer behaviour, market dynamics, and operational efficiencies. Through the development and use of data analytics dashboards, such as the one created using Streamlit for this project, stakeholders can gain a deeper understanding of key metrics and trends that drive the industry.

Insights Gained

This project has successfully demonstrated the power of data-driven decision-making in the online food delivery sector. By analyzing various dimensions of the dataset, including order distribution, customer loyalty, and predictive analytics, several critical insights have been uncovered:

- Order Patterns: Analysis of order distribution by time of day, day of the week, and month reveals peak periods, which are essential for resource allocation and marketing strategies. Understanding when customers are most likely to place orders enables platforms to optimize operations, ensuring timely deliveries and maximizing revenue.

- Customer Loyalty: The customer loyalty analysis has provided a clear view of how often customers return to the same restaurants and how many restaurants they typically order from. This information is crucial for developing targeted loyalty programs and retention strategies that can enhance customer satisfaction and reduce churn.

- Predictive Analytics: The prediction of weekly orders using moving averages has shown the potential of predictive analytics in anticipating future demand. Such forecasting models can be further refined and integrated into real-time decision-making processes, helping food delivery platforms to better manage their inventory, staffing, and marketing efforts.

Challenges Addressed

Several challenges were addressed during this project, including the handling of missing data, the preprocessing of time-related fields and the integration of various analytical functions within a user-friendly dashboard. By overcoming these challenges, the project not only enhanced the quality of insights derived from the dataset but also demonstrated the flexibility and power of Streamlit as a tool for building interactive data analysis applications.

5.1 Future Enhancements

While this project has provided a comprehensive analysis of online food orders, there are several areas where further enhancements can be made. These enhancements would improve the accuracy of predictions, enrich the insights generated, and expand the applicability of the dashboard to other aspects of the online food delivery ecosystem.

Advanced Predictive Modelling

The current approach to predicting weekly orders uses a simple moving average model. While effective, more sophisticated predictive models, such as machine learning algorithms (e.g., ARIMA, LSTM, or random forests) could be implemented to improve the accuracy of predictions. These models can capture more complex patterns in the data such as seasonality and trends and can be trained to account for external factors like holidays, weather conditions and special events that impact order volumes.

Real-time Data Integration

Incorporating real-time data feeds into the dashboard would significantly enhance its utility for operational decision-making. Real-time data integration would allow the platform to provide up-to-the-minute insights such as current order volumes, delivery times, and customer feedback. This capability would be particularly valuable during peak periods when quick decision-making is critical to maintaining service levels and customer satisfaction.

Enhanced User Interface and User Experience

The user interface of the dashboard could be enhanced to improve user experience and accessibility. Features such as customizable dashboards, interactive visualizations and mobile responsiveness would make the tool more user-friendly for a broader range of users from data analysts to marketing managers. Additionally, incorporating natural language processing (NLP) capabilities could allow users to interact with the dashboard using plain language queries further simplifying the data analysis process.

6. BIBLIOGRAPHY

1) Aggarwal, A. (2023). "Trends in Online Food Delivery Services: A Case Study of New Delhi." International Journal of Commerce and Management, 29(2), pp. 112-130.

2) Bose, S., & Roy, A. (2022). "Consumer Behavior Analysis in the Indian Online Food Delivery Industry." Journal of Consumer Research, 45(7), pp. 1458-1475.

3) Choudhary, P., & Sharma, N. (2021). "The Impact of COVID-19 on Online Food Ordering Trends in Urban India." Asian Journal of Economics and Business Studies, 14(1), pp. 54-70.

4) Datta, K., & Mukherjee, S. (2020). "Predictive Modeling in Food Delivery Services Using Machine Learning Techniques." Journal of Applied Data Science and Machine Learning, 7(3), pp. 233-249.

5) Goyal, S., & Joshi, P. (2020). "Understanding Consumer Preferences for Online Food Delivery Apps: A Study in New Delhi." International Journal of Marketing and Business Research, 9(4), pp. 233-245.

6) Gupta, A., & Singh, R. (2023). "Customer Loyalty in the Online Food Delivery Market: An Empirical Study." Journal of Retailing and Consumer Services, 64, 102745.

7) Kapoor, R., & Varma, S. (2021). "Real-time Data Integration in Online Food Delivery Systems." Journal of Information Systems and Technology Management, 18(2), pp. 112-126.

8) Kumar, S., & Raj, P. (2022). "An Analysis of Customer Churn in the Online Food Ordering Industry." Indian Journal of Marketing, 51(6), pp. 26-34.

9) Mishra, D., & Kaur, H. (2023). "The Role of Predictive Analytics in Enhancing Operational Efficiency in Online Food Delivery." Journal of Operations and Strategic Planning, 10(1), pp. 89- 105.

10) Singh, V., & Bhatia, A. (2021). "Impact of Technological Advancements on the Online Food Delivery Industry." Journal of Digital Innovation in Food and Beverage, 6(2), pp. 112-124.

11) Srinivasan, K., & Rao, N. (2020). "Customer Behavior in Online Food Delivery Services in India: An Exploratory Study." Journal of Business Research and Development, 8(3), pp. 78-95.

12) Verma, M., & Mehta, A. (2022). "Exploring the Influence of Social Media on Consumer Preferences in the Online Food Delivery Sector." Journal of Marketing and Digital Transformation, 5(4), pp. 98-110.

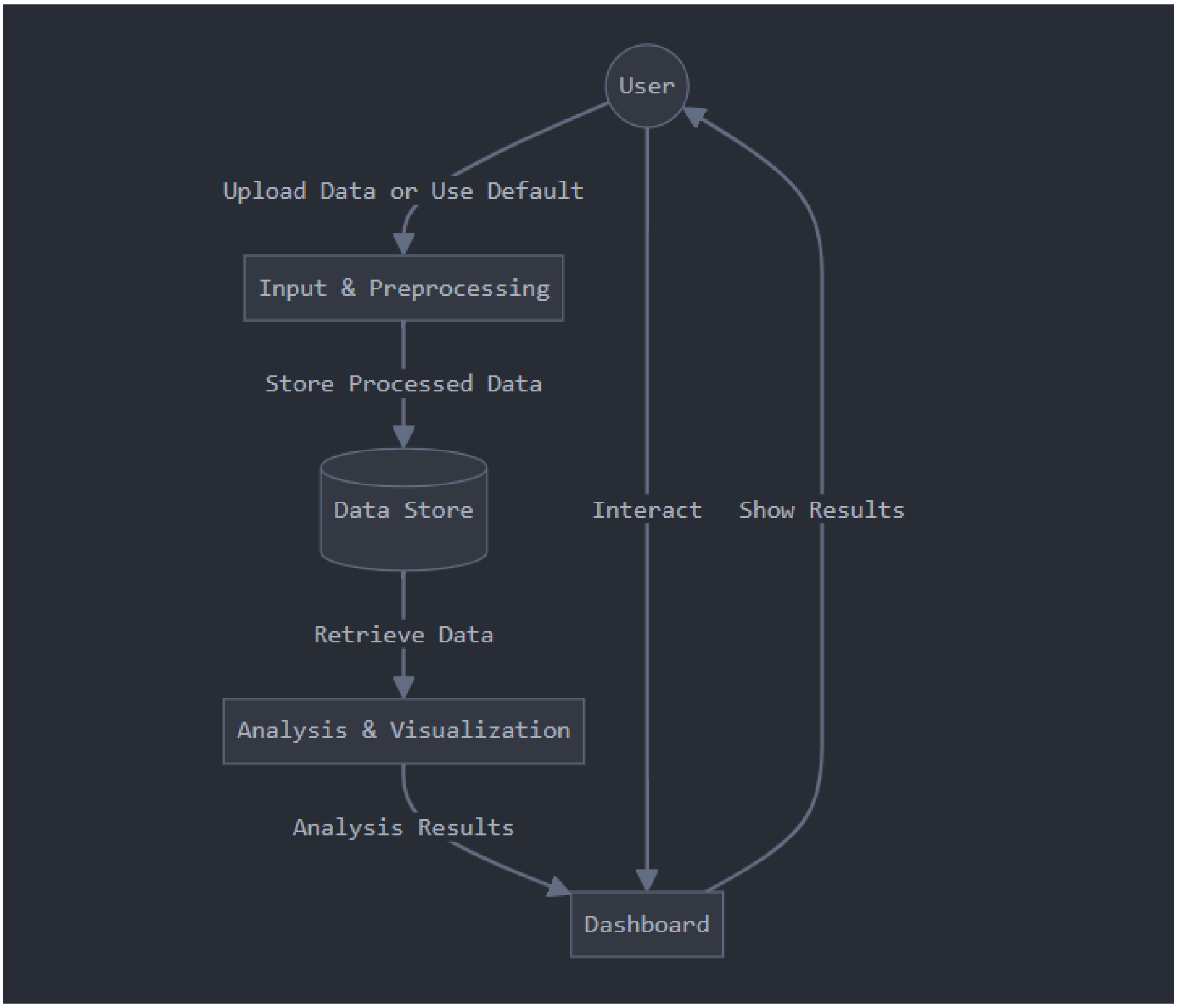
13) Zhang, Y., & Li, X. (2021). "Sustainability in the Online Food Delivery Industry: A Comprehensive Review." Journal of Environmental Sustainability and Business Ethics, 16(3), pp. 173-191.

14) Zomato. (2024). "Annual Report 2023-2024." Retrieved from https://www.zomato.com/annual-report.

15) Swiggy. (2024). "Consumer Trends and Insights: 2023-2024." Retrieved from https://www.swiggy.com/trends-report

7. APPENDICES

A) Data Flow Diagram(DFD)



B) TABLE STRUCTURE

Order ID	Customer ID	Restaurant ID	Order Date and Time	Delivery Date and Time	Order Value	Delivery Fee	Payment Method	Discounts and Offers	Commission Fee	Payment Processing Fee	Refunds/Charg
1	C8270	R2924	01-02-2024 01:11	01-02-2024 02:39	1914	0	Credit Card	5% on App	150	47	
2	C1860	R2054	02-02-2024 22:11	02-02-2024 22:46	986	40	Digital Wallet	10%	198	23	
3	C6390	R2870	31-01-2024 05:54	31-01-2024 06:52	937	30	Cash on Delivery	15% New User	195	45	
4	C6191	R2642	16-01-2024 22:52	16-01-2024 23:38	1463	50	Cash on Delivery	None	146	27	
5	C6734	R2799	29-01-2024 01:19	29-01-2024 02:48	1992	30	Cash on Delivery	50 off Promo	130	50	
6	C7265	R2777	25-01-2024 04:36	25-01-2024 05:27	439	20	Cash on Delivery	10%	92	27	
7	C1466	R2457	12-01-2024 23:55	13-01-2024 00:48	303	30	Digital Wallet	5% on App	144	12	
8	C5426	R2978	26-01-2024 22:46	27-01-2024 00:36	260	0	Credit Card	10%	55	19	
9	C6578	R2877	02-01-2024 18:29	02-01-2024 20:23	1663	40	Cash on Delivery	5% on App	116	48	
10	C9322	R2161	05-01-2024 00:50	05-01-2024 02:10	491	40	Digital Wallet	10%	189	10	
11	C2685	R2379	01-01-2024 17:02	01-01-2024 18:21	868	0	Cash on Delivery	5% on App	149	36	
12	C1769	R2992	09-01-2024 17:15	09-01-2024 17:55	1800	20	Credit Card	15% New User	61	36	
13	C7949	R2086	05-01-2024 16:43	05-01-2024 18:08	674	0	Cash on Delivery	None	115	20	
14	C3433	R2054	04-02-2024 18:48	04-02-2024 20:28	633	30	Digital Wallet	15% New User	51	14	
15	C6311	R2475	03-02-2024 00:49	03-02-2024 02:00	1193	0	Credit Card	None	192	16	
16	C6051	R2799	15-01-2024 20:46	15-01-2024 21:54	992	0	Digital Wallet	15% New User	200	28	
17	C7420	R2177	04-02-2024 17:48	04-02-2024 19:26	504	20	Credit Card	15% New User	130	38	
18	C2184	R2390	13-01-2024 15:41	13-01-2024 16:58	707	30	Credit Card	10%	184	24	
19	C5555	R2348	31-01-2024 17:27	31-01-2024 18:53	1798	50	Credit Card	50 off Promo	191	44	
20	C4385	R2007	13-01-2024 01:32	13-01-2024 02:22	1714	50	Cash on Delivery	5% on App	119	46	

C) SAMPLE CODE

```
import streamlit as st

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from datetime import datetime, timedelta


# Set page configuration (This must be placed at the top of the script)

st.set_page_config(page_title="Plates and Predictions - Online Food Orders Analysis",
layout="wide")

# Function to load and preprocess the dataset

@st.cache_data

def load_dataset(file):

  df = pd.read_csv(file)

  df.columns = df.columns.str.strip()


  if 'Order Date and Time' in df.columns and 'Delivery Date and Time' in df.columns:

    df['order_datetime'] = pd.to_datetime(df['Order Date and Time'], format='%d-%m-
%Y %H:%M', dayfirst=True)

    df['delivery_datetime'] = pd.to_datetime(df['Delivery Date and Time'], format='%d-
%m-%Y %H:%M', dayfirst=True)


    df['order_hour'] = df['order_datetime'].dt.hour

    df['order_weekday'] = df['order_datetime'].dt.weekday

    df['order_month'] = df['order_datetime'].dt.month

    df['delivery_time'] = (df['delivery_datetime'] -
df['order_datetime']).dt.total_seconds() / 3600
```

```
    return df

# Function to predict weekly orders
def predict_weekly_orders(df):
    weekly_orders = df.set_index('order_datetime').resample('W').size()
    sma_window = 4
    weekly_orders_sma = weekly_orders.rolling(window=sma_window).mean()
    last_sma = weekly_orders_sma.iloc[-1]

    fig, ax = plt.subplots(figsize=(10, 6))
    ax.plot(weekly_orders, label='Weekly Orders', color='blue')
    ax.plot(weekly_orders_sma, label=f'{sma_window}-Week SMA', color='red')
    ax.set_title('Weekly Orders and Prediction')
    ax.set_xlabel('Week')
    ax.set_ylabel('Number of Orders')
    ax.legend()

    return fig, int(last_sma)

# New function for customer loyalty analysis
def analyze_customer_loyalty(df):
    customer_restaurant_count = df.groupby('Customer ID')['Restaurant ID'].nunique()
    customer_order_count = df['Customer ID'].value_counts()

    loyalty_data = pd.DataFrame({
        'Unique_Restaurants': customer_restaurant_count,
        'Total_Orders': customer_order_count
```

```
    }).fillna(0)

    loyalty_data['Loyalty_Score'] = (loyalty_data['Total_Orders'] * 0.7) + (loyalty_data['Unique_Restaurants'] * 0.3)
    loyalty_data['Loyalty_Category'] = pd.cut(loyalty_data['Loyalty_Score'], bins=[0, 5, 10, float('inf')], labels=['Low', 'Medium', 'High'])

    return loyalty_data

# Function to create distribution plots
def create_distribution_plots(df):
    fig, axs = plt.subplots(2, 2, figsize=(15, 12))

    sns.histplot(df['order_hour'], kde=True, ax=axs[0, 0])
    axs[0, 0].set_title('Distribution of Orders by Day')
    axs[0, 0].set_xlabel('Hour of Day')

    sns.histplot(df['order_weekday'], kde=True, ax=axs[0, 1])
    axs[0, 1].set_title('Distribution of Orders by Week')
    axs[0, 1].set_xlabel('Day of Week')

    sns.histplot(df['order_month'], kde=True, ax=axs[1, 0])
    axs[1, 0].set_title('Distribution of Orders by Month')
    axs[1, 0].set_xlabel('Month')

    sns.histplot(df['delivery_time'], kde=True, ax=axs[1, 1])
    axs[1, 1].set_title('Distribution of Delivery Time')
    axs[1, 1].set_xlabel('Delivery Time (hours)')
```

```python
    plt.tight_layout()
    return fig


# New function for correlation analysis
def correlation_analysis(df):
    # Select numerical columns only
    numerical_cols = df.select_dtypes(include=['float64', 'int64']).columns
    corr_matrix = df[numerical_cols].corr()

    fig, ax = plt.subplots(figsize=(10, 8))
    sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', ax=ax)
    ax.set_title('Correlation Matrix')
    return fig

# Main Streamlit app
def main():
    st.title("Plates and Predictions - Online Food Orders Analysis Dashboard")

    # File uploader or default dataset path
    uploaded_file = st.file_uploader("Choose a CSV file", type="csv")

    # Default dataset path
    dataset_path = "New Delhi -Online food orders data set.csv"

    if uploaded_file is not None:
```

```
    df = load_dataset(uploaded_file)
    st.success("Dataset loaded successfully!")
else:
    df = load_dataset(dataset_path)
    st.info(f"Loaded default dataset from: {dataset_path}")
    st.success("Dataset loaded successfully!")

st.header("Dataset Overview")
st.write(df.head())
st.write(f"Total number of orders: {len(df)}")

st.header("Order Distribution Analysis")
dist_fig = create_distribution_plots(df)
st.pyplot(dist_fig)

st.header("Weekly Orders Prediction")
weekly_fig, predicted_orders = predict_weekly_orders(df)
st.pyplot(weekly_fig)
st.write(f"Predicted Orders for Next Week: {predicted_orders}")

st.header("Customer Loyalty Analysis")
loyalty_data = analyze_customer_loyalty(df)
loyalty_summary = loyalty_data['Loyalty_Category'].value_counts().sort_index()

fig, ax = plt.subplots(figsize=(10, 6))
loyalty_summary.plot(kind='bar', ax=ax)
ax.set_title("Customer Loyalty Distribution")
```

```
ax.set_xlabel("Loyalty Level")
ax.set_ylabel("Number of Customers")
st.pyplot(fig)

st.write("Top 10 Most Loyal Customers:")
st.write(loyalty_data.sort_values('Loyalty_Score', ascending=False).head(10))

st.header("Correlation Analysis")
corr_fig = correlation_analysis(df)
st.pyplot(corr_fig)
st.header("Key Metrics")
col1, col2, col3 = st.columns(3)
with col1:
    st.metric("Avg. Delivery Time", f"{df['delivery_time'].mean():.2f} hours")
with col2:
    preferred_payment = df['Payment Method'].value_counts().idxmax()
    st.metric("Preferred Payment", preferred_payment)
with col3:
    most_ordered_time = df['order_hour'].mode()[0]
    st.metric("Peak Order Time", f"{most_ordered_time}:00")
# Calculate and display average delivery fee
if 'Delivery Fee' in df.columns:
    avg_delivery_fee = df['Delivery Fee'].mean()
else:
    avg_delivery_fee = "N/A"
        st.metric("Avg. Delivery Fee", f"{avg_delivery_fee:.2f}" if isinstance(avg_delivery_fee, (int, float)) else avg_delivery_fee)
```

```python
    # Calculate and display average commission fee
    if 'Commission Fee' in df.columns:
        avg_commission_fee = df['Commission Fee'].mean()
    else:
        avg_commission_fee = "N/A"
    st.metric("Avg. Commission Fee", f"{avg_commission_fee:.2f}" if isinstance(avg_commission_fee, (int, float)) else avg_commission_fee)

    st.header("Top Restaurants")
    top_restaurants = df['Restaurant ID'].value_counts().head(10)
    st.bar_chart(top_restaurants)

    st.header("Churn Analysis")
    last_order_date = df.groupby('Customer ID')['order_datetime'].max()
    churn_threshold = df['order_datetime'].max() - timedelta(days=180)
    churned_customers = last_order_date < churn_threshold
    churn_rate = churned_customers.mean() * 100
    st.metric("Churn Rate", f"{churn_rate:.2f}%")

if __name__ == "__main__":
    main()
```