**Project on**

**Swiggy Instamart Data Analysis Report**

**Prepared by: Gourav Mohanty  
Date: 02/06/2025**

# Introduction

This project focuses on analyzing customer and order data from Swiggy Instamart to find useful patterns, predict future trends, group customers into segments, and offer recommendations to improve business operations and marketing. The analysis involves techniques like exploratory data analysis (EDA), data visualizations, forecasting models, and machine learning.

To better understand customer behavior, the K-Means clustering algorithm is used. It groups users based on how much revenue they generate and how often they order. This helps identify different types of customers — such as frequent, high-spending users or those who only order once — so Swiggy can tailor marketing and engagement strategies for each group.

For forecasting, the project uses the SARIMA (Seasonal Autoregressive Integrated Moving Average) model to predict monthly order volumes. This model is ideal because it accounts for repeating seasonal patterns in the data. By training it on monthly order summaries, it can forecast future demand for the next 12 months, including confidence ranges to show potential variation. These insights are useful for better planning inventory, assigning staff, and preparing budgets.

# Datasets Used

1. **CustomersWS.csv**
   * Contains customer details such as Customer ID, Name, Phone, Address, and City.
2. **OrdersWS.csv**
   * Holds order-level data including Order ID, Customer ID, Order Date Time, Delivery Date Time, Total Amount, and Delivery Status.
3. **OrderDetailsWS.csv**
   * Describes individual product items per order with Order Details ID, Order ID, Product Name, Quantity, and Price Per Unit.

# Project Objectives

**✅ What are the Objectives:**

* Clean and explore customer, order, and product datasets to ensure data quality and readiness for analysis.
* Analyze delivery efficiency and revenue distribution across various cities to uncover geographic performance trends.
* Apply time-series forecasting techniques to estimate monthly order growth and prepare for future demand.
* Use clustering algorithms to segment customers based on their purchase behavior for better personalization.
* Build machine learning models to predict delivery times and identify potential delays.
* Present all insights using intuitive and structured visualizations for better understanding and communication.

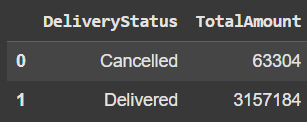
**✅ Why are these Objectives important?**

* To discover and understand high-value customer segments for focused marketing and engagement strategies.
* To enhance inventory and delivery logistics by identifying inefficiencies and making data-driven decisions.
* To create effective marketing and retention strategies by understanding user behavior and preferences.
* To anticipate future customer demand through forecasting and ensure readiness in operations and budgeting.
* To improve on-time delivery rates using predictive analytics, leading to better customer satisfaction.

# Visual Insights

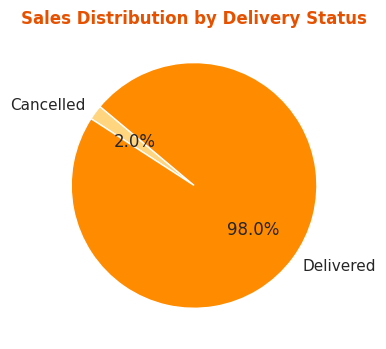
**a. Revenue by Delivery Status**

* Delivered orders account for over 98% of revenue
* Minimal revenue loss from cancellations



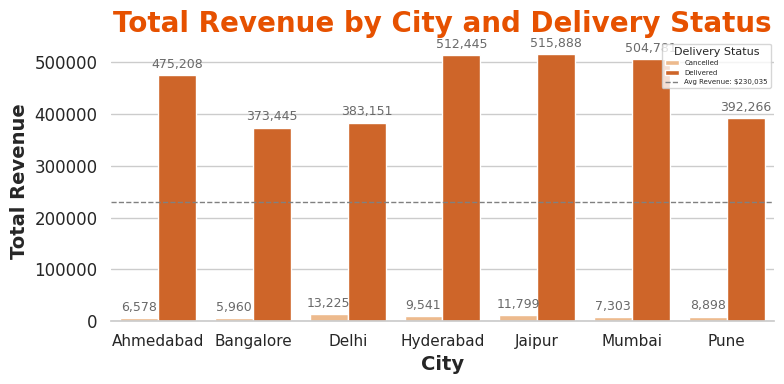
**b. Sales Distribution (Pie Chart)**

* Delivered: 98%
* Cancelled: 2%



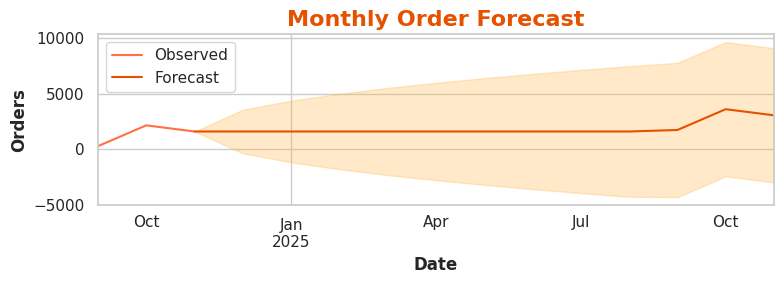
**c. Revenue by City and Delivery Status**

* Jaipur, Hyderabad, and Mumbai lead in revenue
* Pune and Delhi have higher average revenue per customer



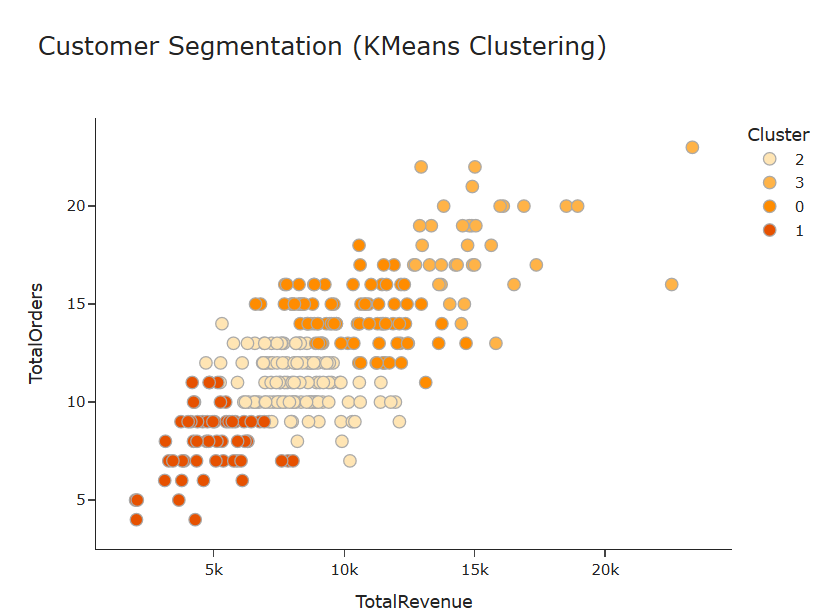
**d. Monthly Order Forecast (SARIMA)**

* Rising order volume trend
* Forecast shows continued growth with low variance.



**e. Customer Segmentation (K-Means)**

* 4 distinct clusters
* Top segment: high revenue + frequent orders



# Insights Summary

|  |  |
| --- | --- |
| **Area** | **Insight** |
| Top Cities | Jaipur, Hyderabad, and Mumbai have most customers and revenue |
| Delivery Status | ~98% of orders are successfully delivered |
| Forecasting | Monthly order volume is steadily increasing |
| Clustering | Segment identified for loyalty marketing |
| Delivery Time | Variation across cities suggests local logistics optimization needed |

# Recommendations

* Prioritize marketing and promotional efforts in high-performing cities like **Jaipur** and **Hyderabad** to maximize customer acquisition and engagement.
* Launch a **tiered loyalty program** to reward repeat customers, encouraging continued purchases and long-term retention.
* Leverage insights from the **SARIMA forecasting model** to anticipate demand trends and improve **inventory and workforce planning**.
* Optimize logistics by **redesigning delivery routes** in cities experiencing frequent delays, aiming to improve service speed and reliability.
* Enhance **machine learning models** by integrating **geographical and time-based features** for more accurate delivery time predictions.

# Conclusion

This comprehensive end-to-end analysis of Swiggy Instamart’s data provides a blend of both tactical and strategic insights to support data-driven decision-making. By integrating statistical methods, data visualizations, and machine learning techniques, the study uncovers key opportunities to enhance customer retention, streamline delivery operations, and scale business processes efficiently. These findings not only highlight immediate areas for improvement but also lay the foundation for long-term operational excellence and customer engagement strategies.