**Project Title: Swiggy Instamart Data Analysis Report**

Project Report  
Prepared by: Ritika Roy  
Date: 23/05/2025

# 1. Introduction

This project aims to perform advanced analytics on Swiggy Instamart's customer and order data to uncover patterns, forecast future trends, segment customers, and provide strategic recommendations for operational and marketing optimization. The analysis includes exploratory data analysis (EDA), visualizations, forecasting models, and machine learning techniques.

To understand customer behavior, K-Means clustering is employed to segment users based on their total revenue and number of orders. This algorithm is effective for identifying distinct groups in the customer base, allowing businesses to create tailored engagement strategies for high-value, frequent buyers as well as one-time or low-engagement users.

For forecasting, the Seasonal Autoregressive Integrated Moving Average (SARIMA) model is used to predict monthly order volumes. SARIMA extends traditional ARIMA by incorporating seasonality, which is crucial for modeling fluctuations that follow consistent periodic patterns. By training the SARIMA model on monthly aggregated order data, the analysis forecasts future demand over a 12-month horizon and includes confidence intervals to reflect variability. These insights help in proactive inventory planning, workforce allocation, and budget forecasting.

# 2. Dataset(s) Used

**📂**

1. **CustomersWS.csv**
   * Contains customer details such as CustomerID, Name, Phone, Address, and City.
2. **OrdersWS.csv**
   * Holds order-level data including OrderID, CustomerID, OrderDateTime, DeliveryDateTime, TotalAmount, and DeliveryStatus.
3. **OrderDetailsWS.csv**
   * Describes individual product items per order with OrderDetailsID, OrderID, ProductName, Quantity, and PricePerUnit.

# 3. Project Objectives

**✅ What?**

* Cleaning and exploring customer, order, and product data
* Analyzing delivery performance and city-wise revenue
* Performing time-series forecasting for monthly order growth
* Clustering customers by purchase behavior
* Predicting delivery time using machine learning
* Visualizing findings with well-structured charts

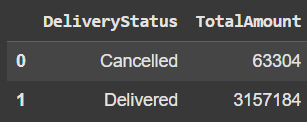
**✅ Why?**

* To identify high-value customer segments
* To optimize inventory and delivery operations
* To guide targeted marketing and retention campaigns
* To prepare for future demand through forecasting
* To reduce late deliveries with predictive insights

# 4. Visual Insights

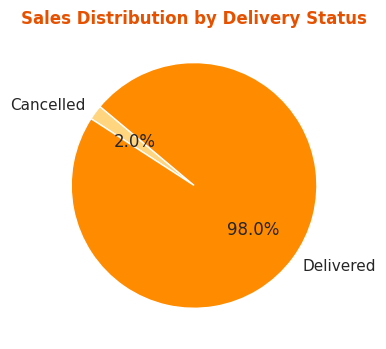
**a. Revenue by Delivery Status**

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



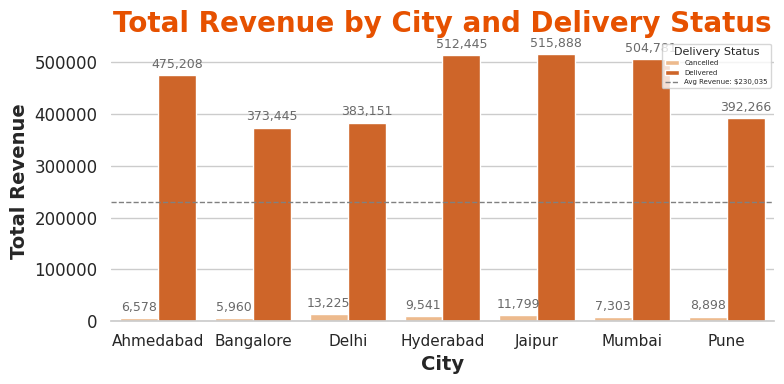
**b. Sales Distribution (Pie Chart)**

* Delivered: 98%
* Cancelled: 2%



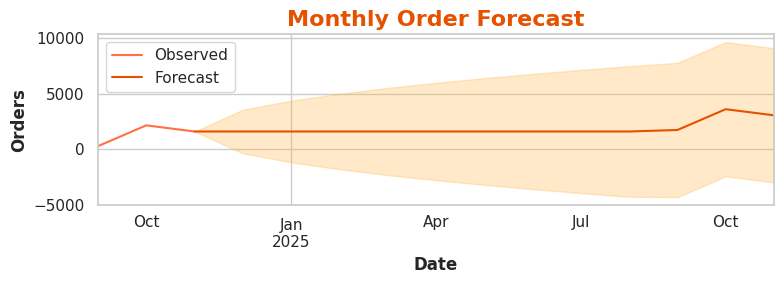
**c. Revenue by City and Delivery Status**

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



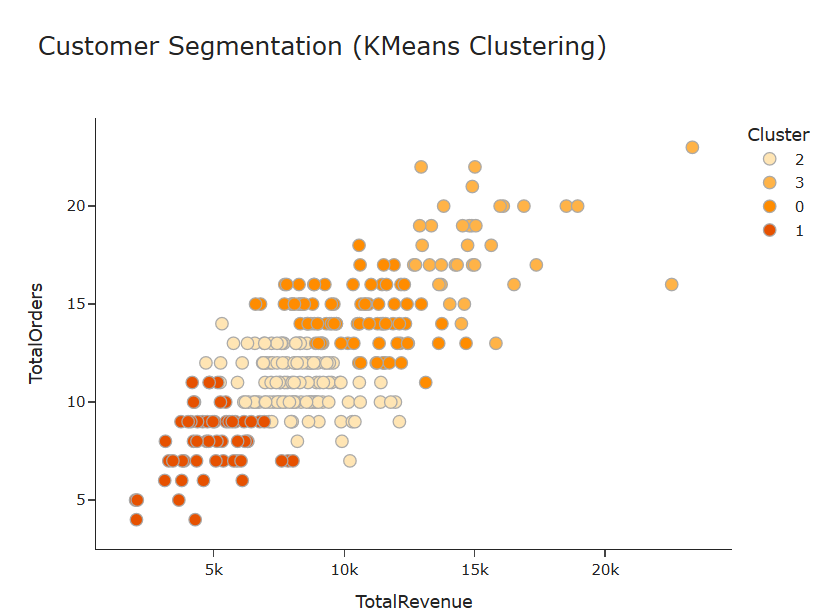
**d. Monthly Order Forecast (SARIMA)**

* 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



# 5. 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 |

# 6. Recommendations

* Focus promotional campaigns on Jaipur and Hyderabad
* Introduce a tiered loyalty program for repeat buyers
* Use SARIMA results for demand forecasting and resource planning
* Re-engineer delivery routes for high-delay cities
* Refine ML models with geographic and temporal features

# 7. Conclusion

This end-to-end analysis of Swiggy Instamart's data delivers both tactical and strategic insights. By combining statistical analysis, visualization, and machine learning, we identify actionable areas for improvement in customer retention, delivery optimization, and operational scalability.