➢ Introduction

understand the factors affecting delivery performance, customer experience, and operational efficiency.
The dataset includes delivery time, distance, weather, traffic levels, courier experience, vehicle type, and time of day.

The goal is to

- *Clean and prepare the dataset*
- *Perform exploration data analysis (EDA)*
- *Discover patterns using association rules*
- *Identify clusters of deliveries*
- *Extract insights and provide operational recommendations*

➢ Dataset Overview

The dataset contains variables

- *Distance_km → Distance traveled by courier*
- *Weather → Weather condition during delivery*
- *Traffic_Level → Low / Medium / High*
- *Time_of_Day → Morning / Afternoon / Evening / Night*
- *Vehicle_Type → Bike / Scooter / Car*
- *Preparation_Time_min → Restaurant food preparation time*
- *Courier_Experience_yrs → Years of courier experience*
- *Delivery_Time_min → Final time from order to delivery*

➢ Data Cleaning Process

A structured cleaning workflow was applied

Handling Missing Values

- *Empty strings ("") were converted to NA*
- *Complete-case approach: rows with missing critical fields were removed*

Data Type Conversion

- *Weather, Traffic_Level, Time_of_Day, and Vehicle_Type → converted to categorical factors*
- *Distance_km, Delivery_Time_min, Rating → converted to numeric*

Removing Impossible / Infinite Values

- *Filtered rows with non-finite values*
- *Ensured distance > 0 and time > 0*

Outlier Removal

- *Using boxplot statistics:*
    - *Extreme outlier delivery times removed*
    - *Extreme distances removed*

Feature Engineering

- *Three new features were created*

➢ *✓Speed_kmph*

Speed = Distance / (Time/60)

Shows courier performance.

➢ *✓Late_Delivery*

- *Late if delivery time > 30 minutes*
- *OnTime otherwise*

➢ *✓Distance Bins*

➢ *Created using quantiles:*

- *Short*
- *Medium*
- *Long*

These help in modeling and segmentation.

➢ Exploratory Data Analysis (EDA)

Distribution of Delivery Time

- Delivery time ranges from 8 minutes to 140 minutes
- Most deliveries are between 25–45 minutes

Delivery Time by Vehicle Type

- Scooters deliver the fastest
- Cars tend to be slower due to traffic impact
- Bikes show high variance

Correlation Analysis

- Positive correlation between Distance and Delivery Time
- Negative correlation between Rating and Delivery Time
- Traffic_Level strongly impacts delivery time

Weather & Traffic Impact

- Rainy/Foggy weather increases delivery time
- High traffic doubles the delay in long-distance deliveries
- Evening orders show highest delays

Association Rule Mining

Using Apriori algorithm with minimum support = 0.01 and confidence = 0.4

Top Rules

Common pattern discovered

- {High Traffic} → {Late Delivery}
- {Long Distance, Evening} → {Late Delivery}
- {Rainy Weather} → {Low Rating}

Interpretation

Rules confirm that weather and traffic significantly influence customer satisfaction and punctuality.

Clustering Analysis

K-means clustering was applied using

- Distance
- Delivery Time
- Rating
- Preparation Time

Identified Clusters

Cluster 1 — Short & Fast Deliveries

- Low distance
- Low delivery time
- High ratings

Cluster 2 — Medium Distance / Moderate Time

- Balanced characteristics
- Medium traffic impact

Cluster 3 — Long & Late Deliveries

- Long distances
- High delivery time
- Majority occur during traffic peaks

This segmentation helps the company optimize courier allocation and scheduling.

Key Insightsy

- ✓ Insight 1: Traffic has the strongest impact on delivery time
  High traffic increases average delivery time by **40–60%**
  ✓ Insight 2: Weather conditions influence both time and customer rating
  Rainy/Foggy → lower ratings + slower deliveries.
  ✓ Insight 3: Long-distance orders should not use slow vehicles
  Scooters outperform bikes and cars under pressure.
  ✓ Insight 4: Distance_bins help predict delay risk
  Long-distance + high traffic → very high probability of late delivery.

Recommendations

Based on the analysis:

Operational Improvements

- Assign fast vehicles (scooters) to long-distance or peak-hour orders
- Improve route planning during high traffic
- Pre-warn customers during bad weather

Customer Experience

- Reduce prep times for orders far from the restaurant
- Offer discounts for long-distance deliveries during traffic peak

Business Strategy

- Employ more couriers during evening peak
- Focus expansion in low-traffic zones

Conclusion

This project gave us a clear understanding of how different factors affect food delivery performance. After cleaning the dataset and exploring the patterns, we found that the most important elements influencing delivery time are **distance, traffic level, weather conditions, and the type of vehicle used**.These findings help explain why some orders arrive quickly while others are delayed.
By looking at clusters, rules, and visual patterns, we were able to identify the common situations that lead to longer delivery times.

Overall, the analysis provides useful insights that can help improve delivery planning, reduce delays, and offer customers a more reliable experience.