Project Report: Effect of Traffic on Uber's Business in Mumbai (2015–2017)

---

1. Introduction

Urban transportation networks in densely populated cities like Mumbai are heavily affected by fluctuating traffic volumes. For a real-time mobility service like **Uber**, traffic conditions directly impact **trip duration**, **fare amount**, **customer satisfaction**, and **driver availability**. This study investigates the **effect of traffic congestion**—along with weather and public events—on Uber's business performance between 2015 and 2017.

---

2. Objective

To quantify and analyze the **impact of hourly traffic volume** and related external factors on:

- Uber **trip duration**
- Uber **fare amount** and **surge multiplier**
- Spatial-temporal **demand fluctuations**
- Business-critical periods like holidays or concerts

---

3. Datasets and Integration

3.1 Datasets Used

Dataset	Key Features
Uber Trips (Synthetic)	Pickup/drop-off times & locations, fare, surge
Mumbai Traffic Data	Hourly vehicle counts per junction
Weather Dataset (Hourly)	Temperature, humidity, rainfall, wind speed
IPL Matches	Event dates and locations
Concerts & Holidays	Major city events impacting mobility
Public Demonstrations	Protest data by date

3.2 Integration Steps

- **Datetime Unification**: All datasets were converted to a uniform hourly datetime format.
- **Geo-Mapping**: Uber pickups were matched with nearest traffic junctions.
- **Event Engineering**: Created binary flags for is_sports_event, is_concert, is_holiday, is_protest.
- **Final Merge**: Combined on datetime and event date using left join.

---

4. Data Cleaning and Normalization

Task	Approach
Missing Values	Imputed using mean (numeric) and mode (categorical)
Duplicate Records	Removed duplicate entries
Column Normalization	Applied MinMaxScaler on numerical fields
Categorical Encoding	One-hot encoding of junctions and other relevant features
Feature Renaming	Standardized column names for clarity

---

5. Key Variables for Analysis

Feature	Description
trip_duration	Time between pickup and dropoff (in minutes)
fare_amount	Total trip fare (₹)
traffic_volume	Number of vehicles at pickup junction
surge_multiplier	Surge pricing factor
temperature, rainfall, humidity, wind_speed	Weather indicators
Event Flags	Binary flags indicating special events

---

6. Exploratory Analysis

6.1 Hourly Traffic and Trip Duration

- **Strong correlation** between high traffic volumes and longer trip durations.
- On average, **trip duration increased by 25%** during peak congestion hours (8–10 AM, 6–9 PM).
- Rainy days showed even more pronounced delays.

6.2 Traffic and Fare Amount

- Higher congestion correlated with **higher fare amounts** due to both **longer duration** and **surge pricing**.
- Median fare on high traffic days: ₹220
  Median fare on low traffic days: ₹145

6.3 Impact of Events

- Surge pricing spikes were observed on:
  - **Concert Days** in Bandra, Andheri
  - **Match Days** (IPL) near Wankhede Stadium
  - **Holidays** like New Year, Diwali
- Uber faced increased demand and longer wait times during events.

---

7. Statistical Insights

Relationship	Correlation (r)
Traffic Volume vs Trip Duration	**+0.63**
Rainfall vs Trip Duration	**+0.41**
Traffic Volume vs Surge Multiplier	**+0.52**
Holiday/Event vs Surge Multiplier	**+0.56**

---

8. Predictive Modeling (Optional Next Phase)

While this project focused on analysis, the dataset is now ready for:

- **Trip Duration Prediction** using Random Forest or Gradient Boosting
- **Fare Estimation Model** considering weather and events
- **Traffic-Aware Driver Dispatch Optimization**

---

9. Business Implications

9.1 For Uber Operations

- **Traffic-Aware Pricing**: Dynamic surge pricing models should incorporate traffic forecasts and not just demand.
- **Driver Allocation**: High-demand zones (e.g., Dadar, Lower Parel, Andheri) during peak hours should be prioritized.
- **Customer Communication**: Preemptive delay notifications can improve trust during heavy traffic or rain.

9.2 Recommendations

- Incorporate **real-time traffic APIs** into fare prediction systems.
- Offer **traffic-based incentives** to drivers during congested hours.
- Run **proactive campaigns** during known citywide events to manage expectations.

---

10. Conclusion

The effect of traffic on Uber's business is **significant and multi-faceted**:

- It increases **trip duration** and **cost**,
- Triggers **surge pricing** more frequently,
- Impacts **supply-demand equilibrium**, especially during events.

A data-driven approach to forecasting, resource allocation, and pricing adjustment can **optimize operational efficiency** and **enhance user satisfaction** in high-traffic urban markets like Mumbai.

✅ **Deliverables Generated**

- final_integrated_dataset_hourly.csv – Merged hourly Uber-Traffic-Weather-Event dataset
- final_merged_dataset_cleaned.csv – Cleaned and normalized dataset ready for modeling