Title: Research Report: Effect of Traffic on Uber Fare Prices and Data Integration Methodology

1. Introduction

Ride-sharing platforms like Uber use **dynamic pricing models** where fares are influenced by demand, supply, and road conditions. Among these, **traffic congestion** plays a critical role in determining pricing through surge multipliers, increased travel time, and longer distances due to rerouting.

The purpose of this report is two-fold:

1. To analyze how traffic impacts Uber's fare pricing.
2. To present a methodology for **data collection, integration, and feature engineering** to build a dataset capable of predicting traffic influence on ride fares.

---

2. Effect of Traffic on Uber's Fare Pricing

2.1 Direct Impacts

- **Increased Travel Time:** More time in traffic means higher **time-based charges**.
- **Surge Pricing:** Traffic congestion leads to a demand-supply imbalance, activating surge pricing.
- **Route Deviations:** Congestion often requires alternate routes, increasing trip distance.

2.2 Impact on Stakeholders

- **Passengers:** Face higher fares and longer waiting times.
- **Drivers:** Benefit from higher earnings due to surge pricing, but also face fatigue from traffic delays.

2.3 External Conditions Worsening Traffic

- **Weather:**
  - Rain → slippery roads, reduced visibility, longer trip duration.
  - Fog (from high humidity) → low visibility, slower speeds.
  - Temperature extremes → may affect demand and vehicle performance.
  - Wind → extreme cases cause delays, diversions.
- **Special Events:**

- Sports matches, concerts, and festivals increase localized traffic demand.
- Public holidays and demonstrations lead to irregular traffic flow and road closures.

---

3. Data Collection & Integration

3.1 Sources

- **Traffic Data:** Provided hourly vehicle count per junction.
- **Weather Data:** Temperature, precipitation, humidity, wind speed.
- **Event Data:** Concerts, sports events, public holidays, demonstrations.
- **Weekend Indicator:** Captures weekly travel behavior differences.

3.2 Cleaning Process

- **Handled Missing Values:** Forward fill used for time-series gaps.
- **Removed Duplicates:** Ensured data integrity.
- **Datetime Conversion:** Unified format for merging datasets.

3.3 Integration

- All datasets merged on **timestamps** to ensure alignment.
- Created a **single processed dataset** (Trafficproject_processed.csv) that is synchronized and ready for modeling.

---

4. Data Preprocessing and Feature Engineering

- **Normalization:** Applied StandardScaler to bring numerical features to a common scale.
- **Feature Engineering:**
  - Extracted **Hour, DayOfWeek, Month** from datetime.
  - Created **Lag Features** (1h, 3h, 6h) to capture traffic patterns over time.
- **Final Dataset Columns:**
  - Traffic: Vehicles
  - Weather: Temperature_C, Rain
  - Events: IsEvent, IsWeekend

- Time Features: Hour, DayOfWeek, Month
- Lag Features: Lag_1h, Lag_3h, Lag_6h

---

5. Results

5.1 Correlation Analysis

- Strong correlation observed between **Vehicles** and **Lag features (Lag_1h, Lag_3h, Lag_6h)**.
- Hourly patterns also show moderate correlation, reflecting rush-hour traffic.
- Weak correlation with weather and events → meaning they influence but not as strongly as lagged traffic.