Dhaka Ride-Hailing Analytics: Strategic Optimization through Data-Driven Insights (2021–2025)

1. Executive Summary

This project presents a comprehensive analysis of a synthetic yet hyper-realistic ride-hailing dataset for Dhaka city, spanning five years (2021–2025). By utilizing Python for complex data modeling and SQL for deep-dive analytics, we have uncovered critical drivers of profitability and operational efficiency.

Key Highlights:

- **Surge Dependency:** Surge pricing is the dominant revenue driver, contributing **83.67%** of total platform revenue, a figure significantly higher than global averages due to Dhaka's gridlock.
- **Moto Dominance:** Motorbikes ("Moto") act as the primary mode of transport, holding a consistent **~48% market share**, proving that speed outweighs comfort in Dhaka.
- **Seasonal Volatility:** The Monsoon season (June–Sept) triggers a **35–40% drop in average speeds**, yet remains highly profitable per trip due to elevated surge multipliers.

2. Problem Statement & Objectives

Dhaka is characterized by extreme traffic congestion and unpredictable weather patterns. For a ride-hailing platform to succeed, it must balance high demand with severe supply constraints.

Project Objectives:

1. **Replicate Reality:** Generate a synthetic dataset that mathematically models Dhaka's inflation, traffic decay, and seasonal waterlogging effects.
2. **Profitability Analysis:** Identify which vehicles, routes, and time slots generate the highest margins.
3. **Operational Efficiency:** Analyze the impact of "Peak Hours" and "Monsoon Season" on service reliability and speed.

3. Data Pipeline & Methodology

A. Programming & Simulation:

- **Python:** Core language used for the synthetic data generation engine.
- **Pandas & NumPy:** Used for high-performance data manipulation and vectorized calculations.
- **Jupyter Notebook:** Interactive environment for iterative coding and feature engineering.

Database & Querying:

- **SQL:** Used for exploratory data analysis (EDA), aggregations, and deriving key business metrics (KPIs).
- **Window Functions:** Applied for ranking top drivers and calculating year-over-year growth.

Business Intelligence (BI) & Visualization:

- **Microsoft Power BI:** Used to build the interactive Executive Dashboard.
- **DAX (Data Analysis Expressions):** Used to create calculated measures like *Surge Revenue %* and *Cancellation Rate*.
- **Data Modeling:** Star schema implementation to link Rides and Cancellations tables.

B. End-to-End Data Generation (Python)

Using `pandas` and `numpy`, we engineered a dataset of **50,000 trips** and **15,000 cancellations**. The generation logic incorporated specific "Dhaka Contexts":

- **Inflation Factor:** Fares were logically scaled from **0.70x (2021)** to **1.00x (2025)** to simulate fuel price hikes and economic inflation.
- **Traffic Decay:** Average traffic speeds were programmed to decline year-over-year, simulating worsening city congestion.
- **Vehicle Physics:** "Moto" and "CNG" were programmed with higher speeds during `Gridlock` conditions to simulate lane-filtering capabilities.

C. Feature Engineering

To enable granular analysis, the following features were engineered:

- **`is_weekend`:** Custom logic applied to mark **Friday & Saturday** as the weekend, adhering to the Bangladesh workweek structure.
- **`time_of_day`:** Trips categorized into "Morning Rush," "Office Hours," "Evening Rush," and "Night".
- **`Start_of_Month`:** A DAX-calculated column used in Power BI to correct visual trends and avoid data noise.

4. Strategic Business Insights (SQL Analysis)

The following insights were derived from 13 targeted SQL queries:

I. Financial Performance

- **Surge is King:** Surge pricing accounts for **83.67%** of total revenue. This confirms that the base fare alone is insufficient for profitability in Dhaka; the business model relies heavily on "High Traffic" and "Gridlock" multipliers.
- **Revenue Growth:** The platform saw explosive early adoption, with monthly completed trips growing **15x** within the first 10 months of 2021.
- **Fare Inflation:** The operational cost increase is reflected in fare pricing. For example, **UberXL** per-km rates grew by **63%** from 2021 to 2025, while Moto rates remained the most affordable with a 44% increase.

II. Operational Dynamics

- **The "Monsoon Effect":**
    - During the rainy season (June–Sept), average speed plummets to **12.8 km/h** (a 28% drop compared to dry months).
    - While trip volume drops by **47%** due to supply constraints, the average surge multiplier rises to **1.59x**, partially offsetting the revenue loss.
- **Peak Hour Concentration:** A concise **8-hour window** (7–10 AM & 5–10 PM) generates over **55%** of daily revenue. Optimizing supply during these windows is the highest-leverage operational lever.

- **Cancellation Trends:** The cancellation rate hovered around **6–7%** historically but spiked to **7.53% in 2025**, correlating with worsening traffic conditions and longer driver arrival times.

III. Market & Customer Behavior

- **Vehicle Preference:**
    - **Moto (48%)** and **CNG (25%)** dominate the market. Together, they account for nearly 73% of all trips.
    - Premium vehicles (Premier/UberXL) hold only ~9% share, dropping even lower during Peak Hours as users sacrifice comfort for speed.
- **Digital Payment Adoption:** There has been a massive shift in user behavior. In 2021, **Cash (35.7%)** was a major payment method. By 2025, **Mobile Money (42.9%)** became the leader, reflecting Bangladesh's digital transformation.
- **Top Routes:** The **Mirpur 10 → Mohakhali** corridor is the highest-grossing route. Cross-city routes connecting residential hubs (Mirpur/Uttara) to business districts (Gulshan/Mohakhali) dominate the Top 10 list.

5. Visual Analytics (Power BI Dashboard)

The Executive Dashboard visualizes the dataset with a high-contrast "Dark Theme" for clarity.

- **Monthly Revenue & Speed Trend:**
    - *Observation:* The Red Line (Speed) shows distinct "V-shaped" dips every mid-year. This visually confirms the **Monsoon Impact**, where speeds crash due to waterlogging while Revenue (Blue Bars) continues to trend upward due to surge pricing.
- **Vehicle Market Share:**
    - *Observation:* A clear visualization showing **Moto (47.69%)** occupying half the chart, validating the strategy to focus on two-wheelers.
- **Payment Method Evolution:**
    - *Observation:* The "Pink" area (Mobile Money) expands significantly from 2021 to 2025, visually representing the decline of Cash usage.
- **Route Profitability:**
    - *Observation:* The top routes are clearly identified, allowing the operations team to focus driver incentives on specific corridors like **Mirpur 10** and **Gulshan 2**.

6. Recommendations

Based on the SQL and Python analysis, the following strategic actions are recommended:

1. **"Monsoon Moto" Incentives:** Since speed drops drastically in rain, 4-wheelers become stuck. Launch aggressive incentives for **Moto drivers** during June–September to capture high-surge demand that cars cannot fulfill.
2. **Peak Hour Supply Dynamic:** Implement "Dynamic Vehicle Routing" during the 8-hour peak window. Prioritize dispatching Motos and CNGs in gridlocked zones (like Farmgate) to keep the completion rate high.
3. **Cashless Transition:** With Mobile Money overtaking Cash in 2025, the platform should introduce "Cashless-Only" queues at airports or premium zones to further reduce operational friction.
4. **Route-Specific Pricing:** The **Mirpur to Mohakhali** route is a goldmine. Introduce "Route Passes" or subscription models for daily commuters on this specific corridor to lock in loyalty.

7. Conclusion

This project successfully modeled the complex ride-hailing ecosystem of Dhaka. By combining Python's robust data generation with SQL's analytical power, we established that **Dhaka is a "Time-Sensitive" rather than "Comfort-Sensitive" market**. Success in this region depends entirely on mastering **Surge Pricing mechanics** and maintaining a robust **Moto fleet** to navigate the city's unique traffic challenges.