# Data Analysis Project

## Data Analyst: Pedro P Paz

## Client/Sponsor: Uber Operations & Strategy Team

## Purpose:

*This project aims to analyze Uber's operational data to drive actionable improvements in vehicle performance, cancellation reduction, revenue recovery, and premium service quality. Goals include:*

1. *Optimize resource allocation by identifying high-performing vehicle types.*
2. *Reduce driver cancellations during peak hours.*
3. *Quantify and mitigate revenue leakage in high-traffic routes.*
4. *Enhance customer experience for premium services.*

## Scope / Major Project Activities:

| Activity | Description |
| --- | --- |
| **Data Extraction & Cleaning** | Collect and validate data from Uber databases (booking status, vehicle type, cancellation reasons, geolocation, ride distance, ratings). Handle missing values and outliers. |
| **Vehicle Performance Analysis** | Calculate ride completion rates and average booking value per vehicle type (Auto, Go Sedan, Bike) for Q2 2024. |
| **Cancellation Root-Cause Analysis** | Identify top 3 reasons for driver-initiated cancellations in "Go Mini" and "Auto" rides during peak hours (6 PM–10 PM) in 2024. |
| **Revenue Loss Assessment** | Quantify potential revenue lost from incomplete rides (status ≠ "Completed") on Delhi-Gurgaon routes (e.g., Cyber Hub, IGI Airport) in 2024. |
| **Customer Experience Correlation** | Compute Pearson correlation between ride distance and driver ratings for completed Premier Sedan rides in 2024. |
| **Insight Synthesis & Reporting** | Compile findings into actionable recommendations for operational efficiency, cancellation reduction, and service quality. |

## This project does not include:

* *Real-time data processing or dashboard development.*
* *Analysis of data outside 2024 (unless baseline comparison is critical).*
* *Driver/rider interviews or external market research.*

## Deliverables:

| Deliverable | Description/ Details |
| --- | --- |
| **Analytical Report** | PowerPoint/PDF detailing methodology, results, and strategic recommendations for all 4 SMART questions. |
| **Data Validation Summary** | Documentation of data quality checks, assumptions, and cleaning steps. |
| **Revenue Recovery Plan** | Prioritized list of high-leverage routes/vehicle types for reducing revenue loss. |
| **Cancellation Mitigation Playbook** | Actionable steps to address top cancellation reasons (e.g., driver training, incentives). |

## Schedule Overview / Major Milestones:

| Milestone | Expected Completion Date | Description/Details |
| --- | --- | --- |
| ***Data Validation Complete*** | *[Start Date + 1 Week]* | *Cleaned dataset ready for analysis.* |
| ***Interim Analysis Review*** | *[Start Date + 3 Weeks]* | *Preliminary results for Q1 (Vehicle Performance) and Q2 (Cancellations).* |
| ***Full Analysis Draft*** | *[Start Date + 5 Weeks]* | *Complete findings for all 4 questions.* |
| ***Final Deliverables Submitted*** | *[Start Date + 6 Weeks]* | *Client-ready report and supplemental materials.* |

## \*Estimated date for completion:

***[Start Date + 6 Weeks]***