**Business Requirement Document (BRD)**

## **Project Overview**

## This document outlines the business requirements for developing a **Data Engineering Project** and a **Data Analysis Project** using the provided rideshare dataset. The dataset includes tables for **Driver**, **Location**, **Rides**, **User**, and **Vehicle**, which will be used to derive insights and build a robust data pipeline.

## **Business Objectives:**

1. **Data Engineering**

* To design and implement an efficient data pipeline that ingests, transforms, and stores data from the rideshare system.
* To ensure data is clean, consistent, and available for downstream analytical purposes.
* To automate data workflows for scalability and reliability

**2. Data Analysis**

* To derive actionable insights from the rideshare data.
* To identify key performance indicators (KPIs) for business operations.
* To visualize data to aid decision-making and improve customer and driver satisfaction

## **Scope of Work**

### **Data Engineering**:

1. **Data Ingestion**:
   * Extract data from the five tables (Driver, Location, Rides, User, and Vehicle).
   * Load data into a cloud data warehouse (e.g., Azure ADLS, One lake).
   * Support multiple file formats (e.g., CSV, Excel, XML, and JSON) and ensure schema consistency.
2. **Data Transformation**:
   * Standardize and clean the data
   * Create derived columns for better analysis
   * Normalize and join the tables to establish relationships (e.g., User-Driver, Rides-Vehicle).
3. **Data Storage**:
   * Design a data model that supports both OLAP (Online Analytical Processing)
4. **Automation and Scheduling**:
   * Use a workflow orchestration tool (Databricks/ ADF) to schedule daily or real-time updates.
   * Monitor data pipeline performance and implement error handling.
5. **Documentation**:
   * Provide detailed documentation of the pipeline architecture and processes.

### **Data Analysis:**

1. **Data Exploration**:
   * Perform exploratory data analysis (EDA) to understand data distributions, trends, and anomalies.
   * Identify key metrics (e.g., average ride distance, revenue per trip).
2. **KPIs**:
   * Total rides completed.
   * Driver ratings and performance.
   * Revenue trends (daily, monthly, and yearly).
   * Customer retention rates.
   * Vehicle utilization rates.
3. **Visualization**:
   * Build dashboards using Power BI or Tableau to display:
     + Trip statistics (e.g., number of rides, distances, durations).
     + Driver leader board (ratings, trips completed).
     + Revenue by location and time.
   * Create interactive filters for better insights.
4. **Insights and Reporting**:
   * Provide actionable insights (e.g., peak hours for rides, underperforming drivers).
   * Develop reports for stakeholders to track business performance.

## **Stakeholders**

1. **Business Team:**
   * Define KPIs and validate insights.
2. **Data Engineering Team:**
   * Build and maintain the data pipeline.
3. **Data Analysts:**
   * Perform analysis and create dashboards.
4. **IT Team:**
   * Manage infrastructure and security.

## **Deliverables**

1. **Data Engineering Project:**
   * Fully operational data pipeline.
   * Cleaned and transformed data in the warehouse.
   * Detailed documentation and monitoring tools.
2. **Data Analysis Project:**
   * EDA findings and recommendations.
   * Interactive dashboards. (Power BI Reports)
   * Business performance reports.

## **Timeline**

1. **Week 1-2:** Data ingestion and cleaning.
2. **Week 3-4:** Pipeline design and transformation.
3. **Week 5:** Dashboard and reporting.
4. **Week 6:** Testing and deployment.

## **Approval**

**Prepared By:** [Your Name]  
**Date:** [Date]

**Reviewed and Approved By:**

* Business Lead: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
* Data Engineering Lead: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
* Data Analyst Lead: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_