

## Data Intake Report – EDA Project – Rodrigo Tumani

### Introduction

This report outlines the datasets used for the exploratory data analysis (EDA) of the cab service data.

### Datasets

#### 1. Cab\_Data.csv

- **Description:** Contains information related to cab transactions.
- **Columns:**
  - Transaction ID
  - Date of Travel
  - Company
  - City
  - KM Travelled
  - Price Charged
  - Cost of Trip

#### 2. Customer\_ID.csv

- **Description:** Provides customer details.
- **Columns:**
  - Customer ID
  - Gender
  - Age
  - Income (USD/Month)

#### 3. Transaction\_ID.csv

- **Description:** Includes transaction details.
- **Columns:**
  - Transaction ID
  - Customer ID

#### 4. City.csv

- **Description:** Contains information about different cities.
- **Columns:**
  - City

### Data Processing

1. **Duplicates Removal:** Duplicate records were removed from each dataset to ensure data accuracy.
2. **Merging:** Datasets were merged based on common columns (Transaction ID, Customer ID, City) to create a comprehensive dataset for analysis.

### Missing Values

- Checked for missing values in the merged dataset.
- No significant missing values found.

### Exploratory Data Analysis (EDA)

#### 1. Hypotheses Tested:

- Seasonality in the number of customers.
- Variations in profit margins over months.
- Differences in customer preferences across cities.

- Customer retention rates between Pink Cab and Yellow Cab.
- Impact of average trip distance and payment methods on profitability.

## **2. Visualizations:**

- Comparison of customer numbers between Pink Cab and Yellow Cab over months.
- Average profit margin trends for both cab companies.
- Customer preferences across different cities.
- Customer retention rates.
- Impact of trip distance and payment methods on profitability.

## **Conclusion**

The EDA provides valuable insights into customer behavior, profitability trends, and company performance. Further analysis or targeted strategies can be devised based on these findings.