

# Data Intake Report

Name: Anar Mammadli

Date: October 13<sup>th</sup>, 2024

Project Title: Cab Industry Analysis

## Intoduction:

In this report, I will provide an overview of the datasets used for the analysis of the cab industry, especially focusing on two competing cab companies, Yellow Cab and Pink Cab. By comprehending these datasets is crucial for deriving useful and important information to provide the decisions on investment for the XYZ firm.

## Overview of Datasets:

1. Cab\_Data.csv

Description: This dataset is about the transaction details of the users for Pink and Yellow cab companies, which includes the price charges and cost of trip.

The features:

- Transaction\_ID: The unique identifier for each transaction
- Date of Travel: Shows us the year and date of the ride for either pink or yellow cab by a user.
- Company: The names of the two companies, Pink and Yellow cab.
- City: This shows in which cities the two cabs being used for a particular number of trips.
- KM Travelled: It specifies how many kilometers or distances were travelled by the cabs.

- **Price Charged:** The final amount that the customer pays for the ride
- **Cost of Trip:** The price of the ride

## 2. City.csv

Description: This dataset describes about the number of users that used the particular cabs across different cities of the United States.

Features:

- **City:** Names of cities where cab services are used.
- **Population:** The total population of the city.
- **Users:** The number of users who utilized the cab services in that city.

## 3. Customer\_ID.csv

Description: This dataset provides the unique identifiers of the customers with particular genders and age, in which it specifies each of their income that can

provide information about which cab service they would prefer.

Features:

- **Customer ID:** The unique identifier of the customers.
- **Gender:** The type of sex a customer is, either male or female
- **Age:** It specifies how old you are
- **Income (USD/Month):** This refers to the monthly income of the customers.

#### 4. Transaction ID.csv

Description: This dataset contains unique identifiers for each transaction, represented by the **Transaction ID**. Each transaction is linked to a specific customer through the **Customer ID**. Additionally, the **Payment\_Mode** column

indicates the method of payment used for each transaction, which can be either **Card** or **Cash**. This information helps in analyzing payment preferences and transaction patterns among customers of the cab service.

#### Features:

1. Transaction ID: The unique identifier for each transaction
2. Customer ID: The unique identifier of the customer
3. Payment\_Mode: This provides the information about the payment methods used by customers for their cab rides either with cash or card.