**Google Data Analytics Professional Certificate – Capstone Project**

**Case Study: Cyclistic Bike-Share Analysis**

**PHASE 2: PREPARE – Collect and Understand the Data**

**Data Source**

The dataset used for this analysis is publicly available from the Divvy bike-share program in Chicago:

<https://divvy-tripdata.s3.amazonaws.com/index.html>

It contains detailed trip-level information for each ride, including:

- Ride ID

- Rideable type (bike type)

- Start and end timestamps

- Start and end station names

- User type (member or casual)

This data is published by Motivate International Inc. under a public data license. The Cyclistic brand is fictional, but the dataset is real and credible.

**Data Organization**

I downloaded the most recent 12 months of trip data in `.csv` format and stored them in the `data\_raw` folder of the project directory.

A separate `data\_cleaned` folder will be used to store processed datasets after cleaning.

**Data Credibility and Privacy**

-Credibility: The data is provided by an official bike-share program operator.

- Privacy: No personally identifiable information (PII) is included.

- Accuracy: The dataset structure and columns are consistent across months.