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

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

**PHASE 3: PROCESS – Clean and Prepare the Data**

**Tools Used**

Python (pandas, numpy)

**Actions Taken**

1. Merged 12 months of raw trip data into one dataset.

2. Added new columns:

- `ride\_length` (trip duration in minutes)

- `day\_of\_week` (day name from trip start date)

3. Removed invalid data:

- Negative or zero ride lengths

- Missing start or end station names

- Test rides

4. Saved cleaned dataset to `data\_cleaned/cyclistic\_cleaned.csv`.

**Output**

- `cyclistic\_cleaned.csv` (ready for Phase 4 analysis)