Cyclistic bike-share services

01.10.2025, Cracow Poland

**Data analyst**: Piotr Paczek

**Report recipient**: executive team at Cyclist company

**Stakeholder’s problem**:

Cyclistic’s finance analysts have concluded that annual members are much more profitable than casual riders. In order to create a strategy that would incentivize casual riders to become membership drivers, we need to know what differs casual user from a membership user.

**Business task**:

**How do annual members and casual riders use Cyclistic bikes differently?**

**Data sources**:  
<https://divvy-tripdata.s3.amazonaws.com/index.html>  
2024\*\*-Divvy-tripdata.zip   
where \*\* = {01,02,03,04,05,06,07,08,9,10,11,12}  
 (on license of Google Data Analytics Professional Certificate programme,  
 The data has been made available by Motivate International Inc.)

**Data manipulation:**

Data analysis was conducted on a cleaned and standardized version of the original trip dataset. The following steps were performed to prepare the data for analysis:

* Removed duplicate records by keeping only the most recent entry per ride\_id (based on trip end time).
* Trimmed unnecessary characters such as quotation marks and whitespace from text fields.
* Converted all date and time columns from text to proper datetime formats (with seconds precision).
* Converted geographical coordinates (latitude and longitude) from text to decimal(9,6) format, ensuring consistent numeric precision.
* Standardized categorical fields (rideable\_type, member\_casual) to consistent lowercase values.
* Filtered out invalid trips, e.g. where trip start time was later than end time, or where trip duration was less than 60 seconds or greater than 24 hours.
* Restricted dataset to trips within 2024 and to coordinates falling inside the Chicago city boundary.
* Standardized station IDs by mapping each station name to its canonical (most frequent) station ID, correcting inconsistencies across records.
* Enriched the dataset with additional derived fields such as trip duration in seconds/minutes, ride date, weekday name/number, season, and roundtrip indicator.

**Data analysis:**