

Data Dictionary

This data dictionary covers terminology associated with the analysis of Cyclistic bike ride data.

Columns in Dataset

Column Name	Description
ride_id	Unique identifier for each bike ride
rideable_type	Type of rideable bike used for the ride (e.g., electric bike, docked bike, etc.).
day_of_week	Day of the week when the ride started
started_at	Date and time when the ride started.
ended_at	Date and time when the ride ended.
start_station_name	Name of the station where the ride started.
start_station_id	Unique identifier for the station where the ride started.
end_station_name	Name of the station where the ride ended.
end_station_id	Unique identifier for the station where the ride ended.
member_casual	Indicates whether the rider is a member or a casual user.
ride_duration	Duration of the ride in minutes, calculated as the difference between the start time and end time.
time_of_day	Time of day when the ride started.
season	Season in which the ride took place (winter, spring, summer, fall).
start_lat	Latitude coordinate of the starting station.
start_lng	Longitude coordinate of the starting station.
end_lat	Latitude coordinate of the ending station.
end_lng	Longitude coordinate of the ending station.

Geographic Coordinates

The geographic coordinates of the starting and ending stations are captured in the start_lat, start_lng, end_lat, and end_lng columns, allowing for spatial analysis and visualization.

KPIs

Key Performance Indicators (KPIs) are essential metrics used to measure the performance and success of a project. In the Google Capstone Project for Cyclistic, the following KPIs are identified to analyze the differences between annual members and casual riders and guide the marketing strategy:

Total number of rides

This KPI measures the overall usage of Cyclistic bikes by both annual members and casual riders. It provides insights into the popularity and demand for the service, allowing the marketing team to identify trends and target specific customer segments.

Percentage of rides

This KPI calculates the proportion of rides taken by annual members and casual riders. It helps in understanding the distribution of users and their contribution to the total rides, enabling the marketing team to focus on increasing annual memberships.

Average ride duration

This KPI determines the typical length of rides for both customer groups. By comparing the average ride durations, the marketing team can identify if there are differences in usage patterns. It can help tailor marketing strategies based on the specific needs and preferences of each group.

Average distance per ride

This KPI measures the average distance traveled during rides by annual members and casual riders. It provides insights into the usage patterns and preferences related to ride length. Understanding these differences can aid in designing targeted promotions or incentives to convert casual riders into annual members.

Average speed

This KPI calculates the average speed during rides for both customer segments. It helps evaluate the efficiency and effectiveness of the bike-sharing service. By analyzing the speed differences between annual members and casual riders, the marketing team can identify factors that may influence membership conversion.

These KPIs matter to the project as they provide quantitative measurements of user behavior and preferences. By analyzing these metrics, the marketing team can gain a deeper understanding of how annual members and casual riders utilize Cyclistic bikes differently. These insights will guide the development of a new marketing strategy aimed at converting casual riders into annual members, ultimately maximizing the number of annual memberships and ensuring the future success of Cyclistic.

R Packages

R packages used for the data analysis project.

tidyverse

Provides a collection of packages for data manipulation and visualization. It includes ggplot2, dplyr, tidyr, and others. This package is used to manipulate and visualize data in the project.

janitor

Provides functions for cleaning up messy data by performing operations like renaming, removing duplicated rows, and formatting variable names. This package is used to clean data in the project.

data.validator

Provides a set of functions to validate data and check if it meets certain criteria. This package is used to validate the data in the project.

lubridate

Provides functions for working with dates and times. This package is used to manipulate date and time columns in the data.

data.table

Provides a data manipulation framework for efficient data processing. This package is used to manipulate data in the project.

plyr

Provides tools for data manipulation, including aggregating and transforming data. This package is used to manipulate data in the project.

knitr

Provides tools for creating dynamic reports in R. This package is used to create reports in the project.

readr

Provides functions for reading in data files. This package is used to read in data files in the project.

summarytools

Provides functions to summarize and visualize data. This package is used to summarize data in the project.

DataExplorer

Provides functions for exploratory data analysis. This package is used to explore data in the project.

skimr

Provides a compact summary of data. This package is used to summarize data in the project.

ggplot2

Provides tools for creating graphics in R. This package is used to visualize data in the project.

cluster

Provides functions for clustering data. This package is used to cluster data in the project.

geosphere

Provides functions for measuring distances and areas on the Earth's surface. This package is used to calculate distances between coordinates in the data.

Heading

More Coming Soon

Heading

More Coming Soon