CHANGE LOG

PROJECT NAME: GOOGLE DATA ANALYTICS CAPSTONE PROJECT (CYCLISTIC)

FEB 2023

This change log documents the process that data went through from extraction, cleaning, analysis and visualization throughout the life of the project. Most of the information will come directly from the project team that includes the Director of Marketing i.e Lily Moreno, the marketing analytics team, myself included. I recently joined the marketing analyst team as a junior data analyst.

Date	***Detail***	***Status***
Jan 23, 2023	Downloaded 12 months of data, i.e January to December 2022, from the link provided Index of bucket "divvy-tripdata"	
Jan 23, 2023	• Used Excel (Power Query) to get a preview of the data • Used SQL for further analysis and cleaning • Used Tableau and Excel for analysis and visualization • Used R for data cleaning, processing, analysis and visualization	
Jan 23, 2023	Lily Moreno believes the annual members are more profitable than casual riders and seeks to understand how the 2 client categories behave differently, in order to start a targeted marketing campaign aimed at converting casual riders to annual members since they have already chosen Cyclistic for their mobility needs.	
Jan 24, 2023	I raised the following questions concerns about the data with Lily Moreno: • Some 97 observations had details in started_at and ended_at columns reversed. • Some rides had impractical long hours but no break-times were provided.	

	• Out of the total number of observations i.e 5,667,717, there were 833,064 & 892,742 nulls in start_station_name & end_station_name columns. • 5,858 missing details in end_lng & end_lat columns. • Some station id's had numbers only while others had numbers and alphabets combined. • 37569 observations had a ride_length of 0 minutes.	
Jan 26, 2023	Lily Moreno's responses to the above concerns were: • Exclude the 97 obs from analysis since the data is collected and managed externally. The data has been made available for Cyclistic's use by Motivate International Inc. under the Data License Agreement | Divvy Bikes • She will suggest to data collection and tracking team whether breaks will be included in the data going forward. • For the observations with missing details in start_station_name, end_station_name and id columns, Cyclistic made an acquisition from Planet9 Bikes, another bike_share company. • Regarding the missing end_lng & end_lat details, I could delete the column if it doesn't have an impact in my analysis. • She advised it's safe to assume those bikes with 0 ride_length were not used.	
Jan 30, 2023	Based on the above, the following changes were made to the dataset in R: i. After importing the 12 files to RStudio, they were merged to one dataframe. ii. Data type conversion of started_at and ended_at from string character to Date type. iii. Deleted all start_lng, start_lat, end_lng & end_lat columns iv. Replaced all empty cells in start_station & end_station columns with Planet9 & their start_station_id and end_station_id's with NA. v. Extracted dates, month, day_of_month, day_of_week, hours from started_at column. vi. Calculated ride_length in both hours and mins vii. Filtered rows with negative ride_length, they are the 97 where started_at & ended_at were interchanged. viii. Filtered rows with 0 ride_length, it's assumed they did not move.	

	ix. Proceeded to analyze the 'good' data i.e total less obs with ride_length <=0. They were 5,630,051 observations. x. Saved the 5,630,051 observations in a new data frame. xi. Conducted descriptive analysis xii. Visualized the different subsets of data obtained from analysis.	
Feb 6, 2023	Based on the above analysis and visualization, the following recommendations were made: i. Marketing the electric bikes more among the casual riders especially towards/during the weekends and towards/during summer months starting from April. ii. Introduce a weekly subscription and make it easy to upgrade from being daily casual riders. After some time e.g quarterly performance can be reviewed before introducing monthly subscription in the 3rd quarter of the year. iii. Make the prices more flexible and introduce discounts depending on ride_length especially during weekends. iv. Partner with tourist attraction sites and leisure centers to get more casual riders to subscribe for annual membership at a discounted price.	