2018 Fordgo bike system datasets Project

By Kadiri Aiyegbeni Abdulkadir

The data was obtained from one of the udacity dataset called Ford Gobike System data

The source : ttps://www.lyft.com/bikes/bay-wheels/system-data

The 2018 fordgobike dataset consist of 94802 entries with 16 columns,with 2 datatimes , 5 floats, 4 int64 and 7strings

The main features include:
duration_sec,start_timeend_time,start_station_id,start_station_name,start_station_latitude,start_station_longitude,end_station_id,end_station_name,end_station_latitude,end_staton_longitude,bike_id,user_type,member_birth_year,member_gender,bike_share_for_all_trip.

I am investigating the impact of gender in fordgobike system data of the first month of 2018, how gender availability in each trip?. The number of user that registered that are male or female for the trip ,and how long its was the trip based on the gender.

I did some data wrangling. I dropped some columns, change start time and end type datatype , added the day. I finally had the following features that include:
duration_sec,start_timeend_time,start_station_id,bike_id,user_type,member_birth_year,member_gender,bike_share_for_all_trip and day . then is started my exploration .

Summary of Findings

During my exploration on the datasets, there was a strong relationship between the member_gender,user_type,day and duration_min . the male counterpart spend more time on the trips ,are fully registered whether through subscribe or customer than the female. With steps to be display in my presentation.

Key Insights for Presentation

For the presentation, I focus on just the influence of member_gender variable on day ,user_type and duration_min variables. I start by introducing the member_gender variable,user_type,and day variable with countplot step by step while an histogram plot was used for duration_sec but a transformation was needed .which I did first by creating a new column called duration_sec through duration_sec. then I did my scale transform.

Afterwards I did plot a member_gender and duration_min,looking at which gender spend more time on the trip. Then plotted member_gender|day and member_gender|user_type in order to drawn the availability of gender and how many gender are register in term of subscriber and customer.

Looking more insights on the impact of gender,I did multivariate plot involving member_gender|day|duration_min and member_gender|user_type|duration_min using a stripplot.