Ford GoBike - Data Visualization Project

Program: Udacity Data Analyst, Nondegree (DAND)
Student: Laman Mammadli
Project: Data Visualization

Data set source

As part of the Udacity Data Visualization project, I chose Ford GoBike project. I downloaded the data from the Udacity Project Details page.

Summary of Findings on Exploratory Analysis

In the exploratory part, I conducted univariate, bivariate and multivariate analysis. Before starting to the analysis, I made an assessment and found several quality issues to deal with. After solving the quality issues, I created several variables from the existing variables to make data interesting for analysis. For instance:

- Created months, years, hours, week_days, period_of_day variables from the start_time variable;
- Created distance_km variable by means of 'haversine' formula (took function from stackoverflow) on latitude and longitude variables;
- Created duration_min variable by dividing duration_sec variable to 60.

Univariate Data Analysis

By conducting univariate analysis, I found that:

- Male users are more than female users
- Subscriber user type is more than customer
- Rides are more on Thursday than other days of the week
- Rides are less on weekend (on Saturday and Sunday)
- In my data set, rides are on only February 2019
- Rides are more during the morning than afternoon and night, rides are lower than other times during the night
- The average age is 36
- The average duration of trips is 11 minutes
- The average distance is 1.47 km

Bivariate Data Analysis

By conducting bivariate analysis, I found that:

- Duration of rides are more in younger generation
- Distance is more in younger generation
- Male subscribers are more than female ones
- Rides by males are more than females during the days of the week, including weekends
- Rides by males are more than female during the day (morning, afternoon, night)
- Find 5 top start and end station names which take most time
- Average age of subscriber and customer is almost the same
- Average age of male and female is almost the same

Multivariate Data Analysis

By conducting multivariate analysis, I found that:

- Rides are less on weekend (on Saturday and Sunday)
- During the weekend, at night, the riders are the youngest.
- During the morning and night, on Tuesday and Wednesday the duration is minimum, however, on Saturday and Sunday mornings the duration is maximum.
- During the weekends, afternoon duration average is the highest

Summary of findings on Explanatory Analysis

In the explanatory analysis, I want to focus on these issues:

- Rides are more during the morning than afternoon and night, rides are lower than other times during the night
- Younger people rides longer distance than older ones
- Males are more prone to be subscriber than females
- Males ride more than females
- During the weekends, duration of the rides are the highest on afternoons, however, in general, rides during the weekends are the lowest.