Project Report:

Ford Bike

By: Gurjot Dhillon

INSTRUCTIONS

The report presents the findings of Data analysis performed on bike ride data in San Francisco Bay area. The project was undertaken as part of Udacity's Data Analysis Nanodegree.

First Observation

The data included 16 columns. ['duration_sec', 'start_time', 'end_time', 'start_station_id', 'start_station_name', 'start_station_latitude', 'start_station_longitude', 'end_station_id', 'end_station_name', 'end_station_latitude', 'end_station_longitude', 'bike_id', 'user_type', 'member_birth_year', 'member_gender', 'bike_share_for_all_trip'].

Objectives:

Some of the questions that I wanted to answer are:

- 1. What is the average age of a rider?
- 2. What Stations witness the most traffic?
- 3. Gender based analysis.
- 4. Hourly and Daily Traffic.
- 5. Which category of customers has more riders?

Data Cleaning

Upon inspection, I found the following issues with data: incorrect data types, missing data etc.

- I decided to change the format of start_time and end_time to datetime,
- I created new columns for days, hours, and age.
- I dropped columns for age > 80.

Exploratory Analysis

- I analyzed the data using python language.
- I employed various plots such as hist, counts, bar plot to study age, gender, days, and hours.

Observations

Age:

The average age of a rider is around 34 years.

Gender:

Of the total riders, 130345 are males, 40771are females and 3644 are others.

Time: Hours

The plot was bimodal for weekdays, with most rides starting at 8 am and around 5pm. The traffic peaked in the morning and the evening, but it differed for the weekends when hourly traffic was peaking in the afternoon.

Time: Days

Most rides began on Thursday and Tuesday while Saturday had the least number of rides.

Categories:

Subscribers form the biggest customer base with 158.2K rides while non-subscribers booked a ride 16.5K times.

Correlation between Age and time of travel:

Theres a strong negative correlation between the age and time at which the rides start. Young riders are likely to start early in the day while old riders prefer to travel later in the day.

Stations:

Most rides start at the station Market St at 10th St while most rides end at the station San Francisco Caltrain Station 2

FUTURE SCOPE:

- We can try to explore the duration of trips to understand how it is distributed for genders, age group, categories of customers etc.
- With given latitudes and longitudes, we can create a map and try to explore the geography of stations.
- Using the above map and duration, we can measure the average speed of a ride as well.

Summary

Most riders are on average 34 years old, male, subscribers and are likely to travel in the morning and evening, mostly starting on Thursday.