Data Visualization of Uber Rides with Power Bi

INTRODUCTION

This project is carried out on a private dataset by Zeeshan-Ul-Hassan on Kaggle. We are going to analyze this dataset in Power BI without writing a single line of code.

This Visualization creates a view of uber rides in different cities from January to December 2016. It aims to provide consumers insights on how the number of uber rides changed within different day periods on different days in a month, in different cities.

The dataset includes start and end date, start and end location, purpose of trip, category and miles driven. Category includes business and personal. Purpose of trip includes charity, customer visits, errand/supplies, meeting, airport /travel,

moving, meal/entertainment and commute. Exploring this dataset, I decided to focus on the following questions:

1. What month had the highest number of trips?

2. What is the total number of rides?

3. What was the highest duration spent on a trip?

4. What is the most popular category and purpose of trip?

5. Where did most of the drop-offs occur?

6. Which weekday had the highest number of trips?

7. What was the average duration and time spent on rides?

DATA CLEANING AND PROCESSING

I began my data cleaning process by using power query. Next I looked at the dataset information. From the cell below, 'END DATE', CATEGORY, START, STOP and PURPOSE columns all had missing values. The PURPOSE column had the highest number of missing values.

The 'END DATE*', 'START DATE*', 'PURPOSE*', 'MILES*' AND 'CATEGORY*' columns were renamed for readability and to remove the asterisk (*) attached to each column names.

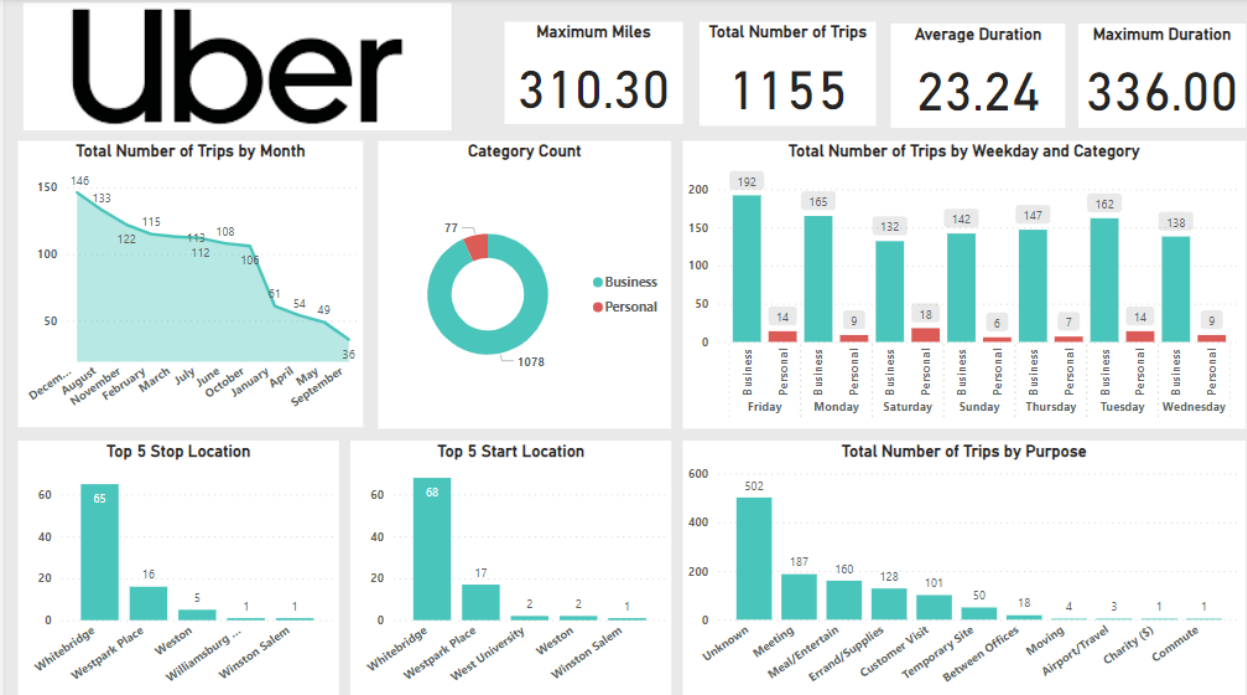
The missing values in the purpose column was replaced with ‘Unknown’ because dropping all the missing rows in the purpose column would mean working with just half dataset. Other missing rows in ‘END DATE’, ‘CATEGORY’, ‘START LOCATION’ and ‘STOP LOCATION’ were dropped respectively. The dataset was then inspected once again to ensure the missing values have all been treated.

Hour, Month, Day of the week and weekday columns were created from the Start date column. The hour attribute was applied on the Start date column to get the hour the ride was taken. The same was done to get the Month and Day of the week using the datetime Month and Day of the week attribute.

A new column ‘Duration’ was created as well. This is the resulting difference between ‘END DATE’ and ‘START DATE’. It gives us the amount of time spent travelling from start location to stop location for each trip.

VISUALIZATIONS

I calculated the total number of rides which is 1155. I checked for maximum and minimum miles travelled as well as maximum and minimum duration. The maximum miles travelled was 310.3 miles and the minimum was 0.5 miles. Maximum duration was 336 minutes from an unknown location to another unknown location for business purpose. Minimum duration was 0s, this might be because the trip was cancelled or the values were not entered accurately since each corresponding entry had a calculated mileage. The trip with the maximum miles was from Latta to Jacksonville, however it was not the trip with the maximum duration.



Conclusion

After my analysis, I can conclude the following:

I. Other than rides for unknown purposes, the top 5 purposes for rides were meetings- 187, meals and entertainment- 160, errands/supplies- 128, customer visits- 101, and temporary site-50.

II. Majority of the trips were for business purposes.

III. December was the busiest month with 146 rides and September was the least busy month with 36 rides.

IV. Fridays were the busiest days of the week for the rider.

Thank you for taking your time to read my article. Any comments, questions and suggestions would be greatly appreciated.