Tableau Homework – Citi Bike Analysis

Terrence Cummings
August 3, 2020

This documents the process and observations taken in creating Tableau worksheets, dashboards, and story related to New York City Citibike data.

Data source:
https://www.citibikenyc.com/system-data

Data description:
- 12 Monthly CSV files from July 2019 through June 2020 representing 1 full year of data were downloaded.
    - Data was restricted to one calendar year do to slow calculation speed.
    - This encompasses about 15 million records, the limit of Tableau Public
    - Jersey City (JC) data was not used.

Tableau used:
- Tableau Public version 2020.2.2 64-bit for Mac OSx
- Link: https://public.tableau.com/profile/terrence.cummings#!/

Objective of Analysis:
This exercise was focused on three data points not contained in the standard dataset, and hence not normally considered in the analysis of Citibike rides in NYC. The three new fields are:
1. Distance between start and ending stations miles.
2. Speed of travel in mph.
3. A name for the route that reflects start and end stations.

These fields needed to be created by the formulas:
- Distance in miles = DISTANCE(MAKEPOINT(start_lat, start_long), MAKEPOINT(end_lat, end_long), "miles")
- Speed in mph = (DISTANCE/(trip_dur/60/60))
- Route_ID = start_ID+"_"+end_ID
- Route_Name = start_name+" to "+end_name

With these new fields added, the following areas were investigated:
- Is there a difference between the average speed and distance bike by males and females?
- Does the average speed and distance vary by age?
- How does speed and distance vary by month, day, and time of day?
    - Do people pedal faster and travel less distance when it is cold or hot?
    - Are there differences between weekday and weekend speed and distance due to different mix of casual versus commuting bikers?

- o   Are there observable differences during peak rush hour on a weekday?
- What are the longest routes biked?
- What routes have the fastest speed?

Additionally, the Tableau file contains worksheets and dashboards related to the most frequently used starting and ending stations including:
- Top 10 starting/ending stations by frequency during July 2019 to June 2020
- Dynamic map showing the monthly change in starting/ending station usage over 12 months to include all seasons.

The "story" in tableau is as follows:
1. Big picture – Which stations are the most popular to begin and end a trip?
    a. OBSERVATION: Generally, stations near the rivers appear to be most popular. This is likely due to commuters coming on the ferry, as well as being a popular point for tourists on weekends.
2. Moving from 'stations' to 'routes' – If we look at start and end stations as a specific route, which routes are longest? Which are taken at the highest speed?
    a. OBSERVATION: The longest routes are unsurprisingly from the far north part of Manhattan in Washington Heights to the southern tip near Battery Park. The speediest routes tend to be those that are a 'straight shot' down a major thoroughfare such as 5th Ave along Central Park.
3. Let's look at the 'when' dimension – Does month, day of week, and time of day impact speed and distance?
    a. OBSERVATIONS:
        i. People tend to bike faster and shorter distances in months where the weather is cold or hot.
        ii. People bike more leisurely on average during the weekend, at a slower pace and for longer distances.
        iii. There is an observable peak in distance and speed during morning rush hours. During evening rush hours the distance peaks again, but speed is more leisurely as people return home.
4. Let's look at the 'who' dimension – Does speed and distance vary by gender? By age?
    a. OBSERVATIONS:
        i. Women 25-30 tend to bike a little further than men.
        ii. Men tend to bike at a higher speed.
        iii. For men and women distance begins to decline with age at around age 35. Speed declines less so with age. Likely due to the nature of city bike not being focused on speed.

Please see the link to Tableau Public for worksheets, dashboards, and story.

--END--