NY CitiBike

Rental Analytical Report

GitHub | Tableau Report

Goal

Analyzing bike's rental patterns based on trip duration by user type and gender.

Overview

Citibike which launched in May 2013, is the nation's largest bike share program operating out of New York City.

This bike share systems with 27,000 bikes and over 1,700 stations across Manhattan, Brooklyn, Queens, the Bronx, Jersey City and Hoboken, has become an essential part of transportation network for New Yorkers at any day and time of the year.

This analysis will address the following questions;

1. What timeframe and day has the most bike's usage?
2. Does gender have any impact on length of rentals or membership status?
3. Which user type group leads in total rentals?

Tools Used:

Jupyter Notebook | MS Excel | Python

Resources:

Dataset Source: Citibike Dataset

Overview: Citibike NYC

Timeframe: September 2013

The Process

NY Citibike

Dataset Preparation

Performed the following actions:

1. Sourced NY Citibike Dataset
2. Data Quality Checks
3. Wrangling, Subsetting and Deriving Variables.
4. Grouping and Aggregating Data
5. Formulating Hypothesis

The Analysis

Performed the following actions:

- **Ran Charts:** Scatterplots, Heatmaps.
- **Conducted** Exploratory Analysis, Regression Analysis, Cluster Analysis and Time Series Analysis.
- Conclusions

The Results

Shared final visual report via:

1. Electronic PDF
2. Tableau
3. GitHub.

Skills Learned:

- Data Cleaning and Quality Checks
- Wrangling and Subsetting
- Exploratory Analysis
- Linear Regression; Supervised Machine Learning
- K-Means Clustering; Unsupervised Machine Learning
- Time Series Analysis

Few challenges encountered;

1. Rental Pricing Column Missing. Harder to Measure Sales Generated.
2. Over 5000 Missing Values on "Birth_Year'
3. Represents only one-month observation.

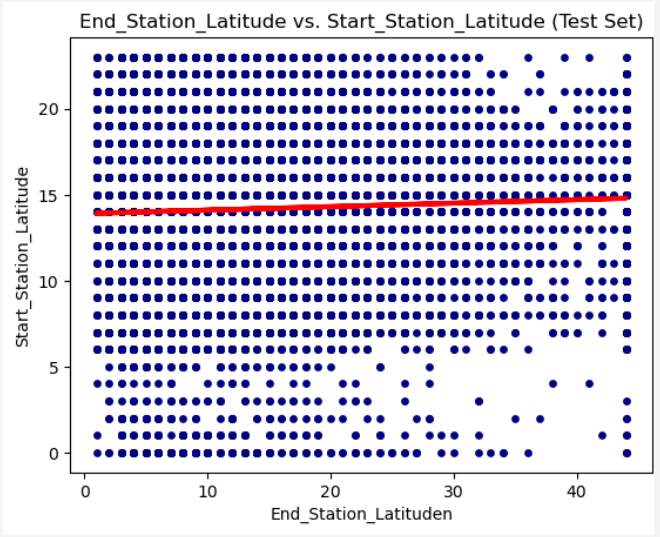
EXPLORATORY ANALYSIS

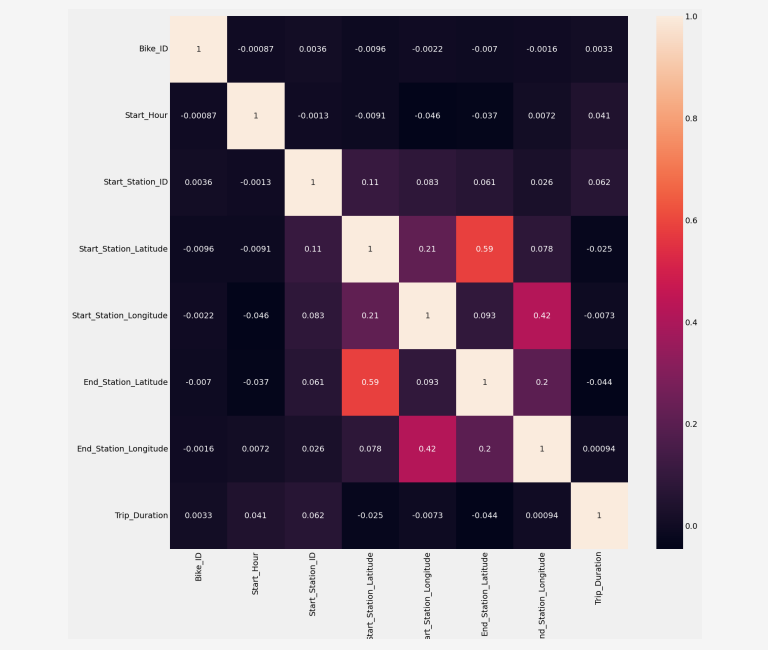
Explored relationships between variables using different methods such as Scatterplots, Heatmaps, and Linear Regression

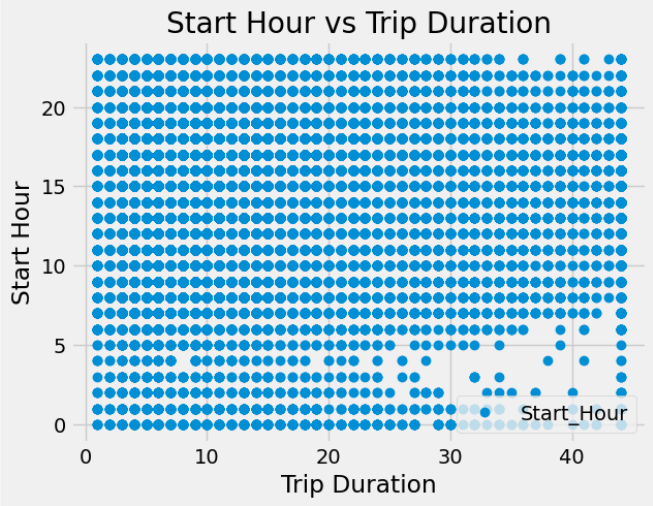
Testing the formulated Hypothesis: "**The length of trip turation is determined by the start hour**".

Findings:

- Heatmap shows the relationship between variables was very weak.
- Close strength existed on 'End_Station_Latitude' and 'Start_Station_Latitude'.





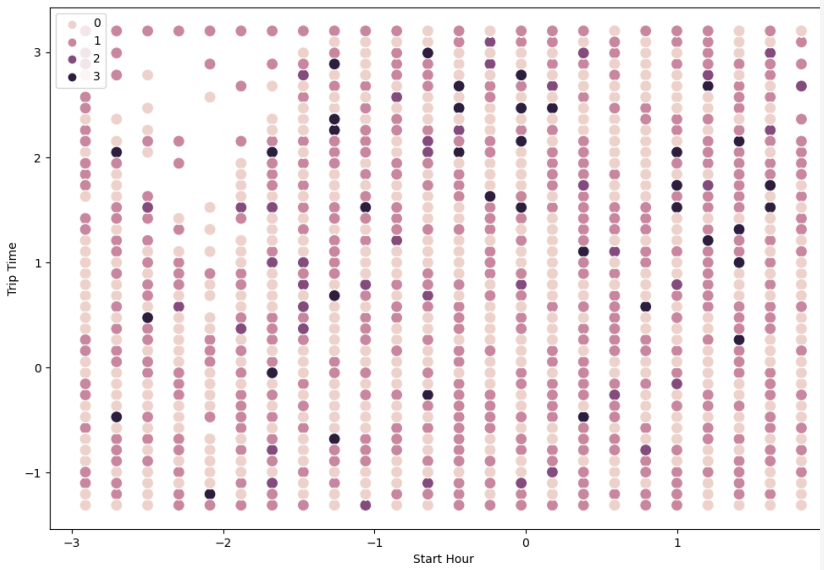


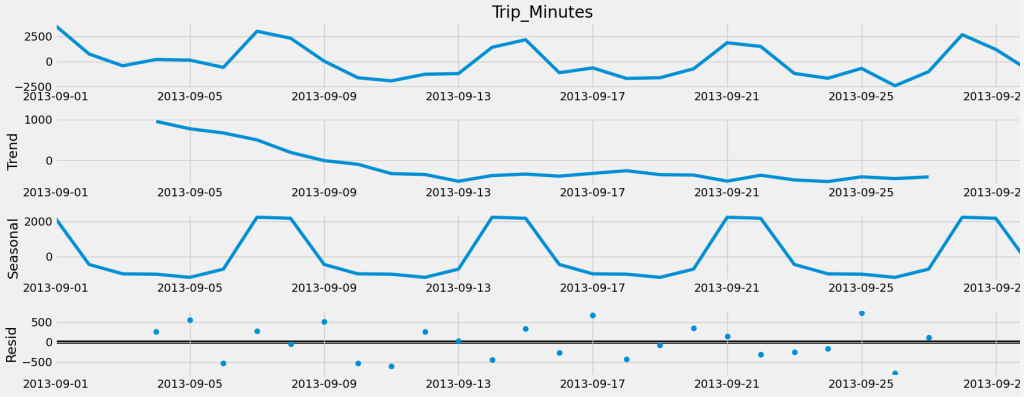
FURTHER ANALYSIS

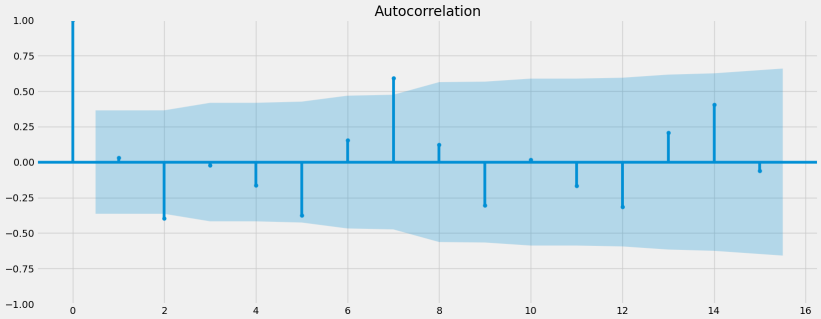
Further analysis was conducted to analyze relationship between the two variables and time trends.

Tri minutes seem to be longer in the first day of the month and probably the third day of the week.

Minutes tend to decline over time with minimal bumps in between the weeks







RESULTS

FINDINGS

> Start hour has no bearing on trip duration hence rider's trip length is not determined by the start hour.

> The overall trend line on trip minutes is downwards with September 21 being the lowest.

On the other note;

> Highest bike's rental happens between 17:00 – 18:00pm.

> Weekdays seems to experience the highest number of bike's rental than weekends with Monday leading in rental counts.

NEXT STEP

> Having a complete dataset with more than one month will help in analyzing the patterns for better predictions.

> The ages of the riders were missing especially on all the over 6700 casual riders limiting insights on age factor.

On the other note;

> Ensure more bikes are available for use during rush hours and busier days.

GitHub Repository Tableau Storyboard