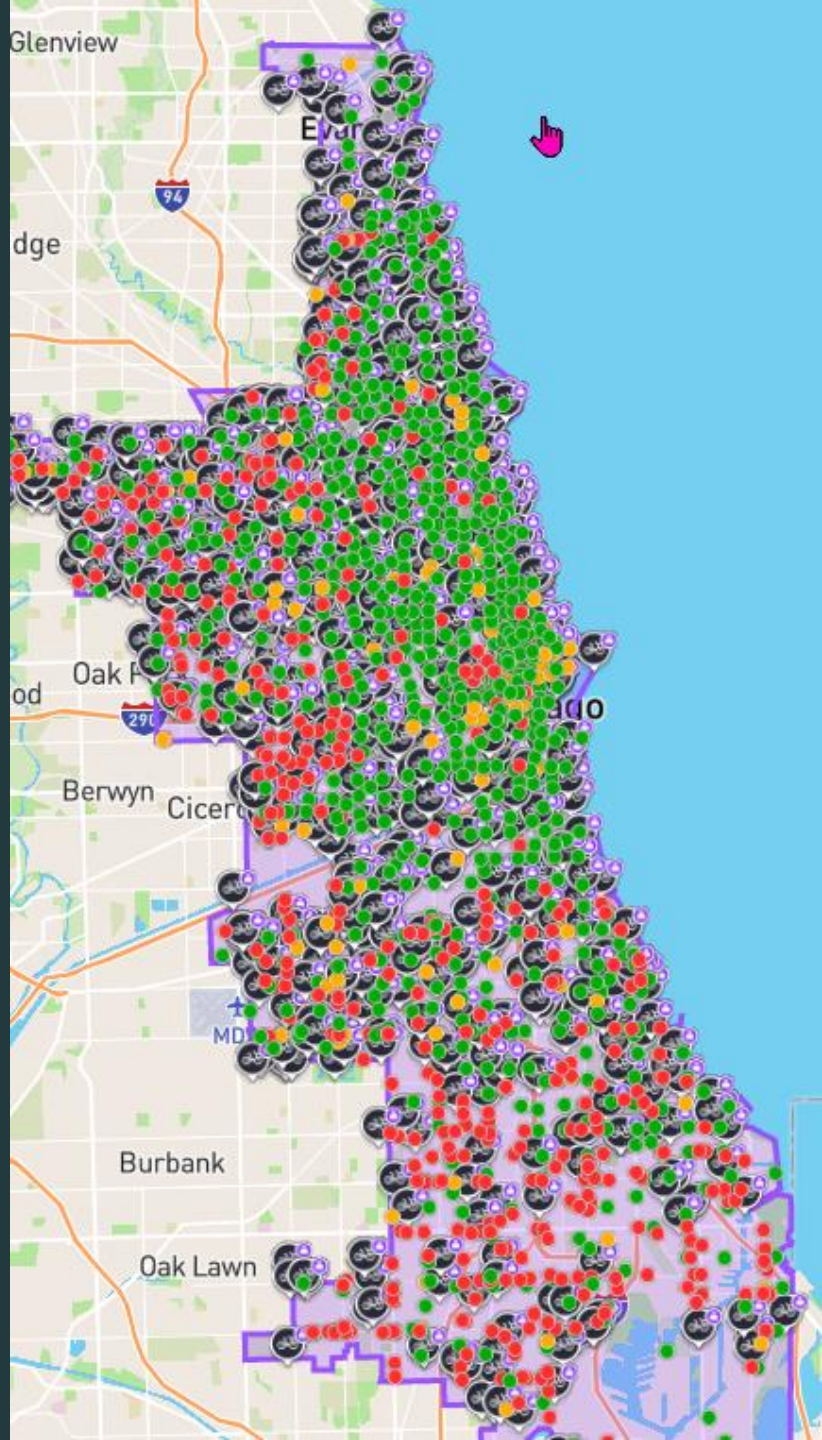
BIKE SHARING GEOSPATIAL DATA ANALYSIS DURING COVID-19

By Yarong Chen

INTRODUCTION

Divvy is the bicycle sharing system in the Chicago metropolitan area, which currently is serving the cities of Chicago and Evanston. The system is owned by the Chicago Department of Transportation and currently is operated by Lyft since 2019. As of July 2019, Divvy operated 5,800 bicycles and 608 stations, covering almost majority of the city. I found its data has been shared on Chicago Data Portal site (https://data.cityofchicago.org/Transportation/Divvy-Trips/fg6s-gzvg) and thought it's a great geospatial data source for my geo-programming course project. I wonder how the bike sharing system has been changed since COVID-19 hit in early 2020.



GOAL

The overall project goal is to explore how the bike sharing pattern and volume have been changed since COVID-19.

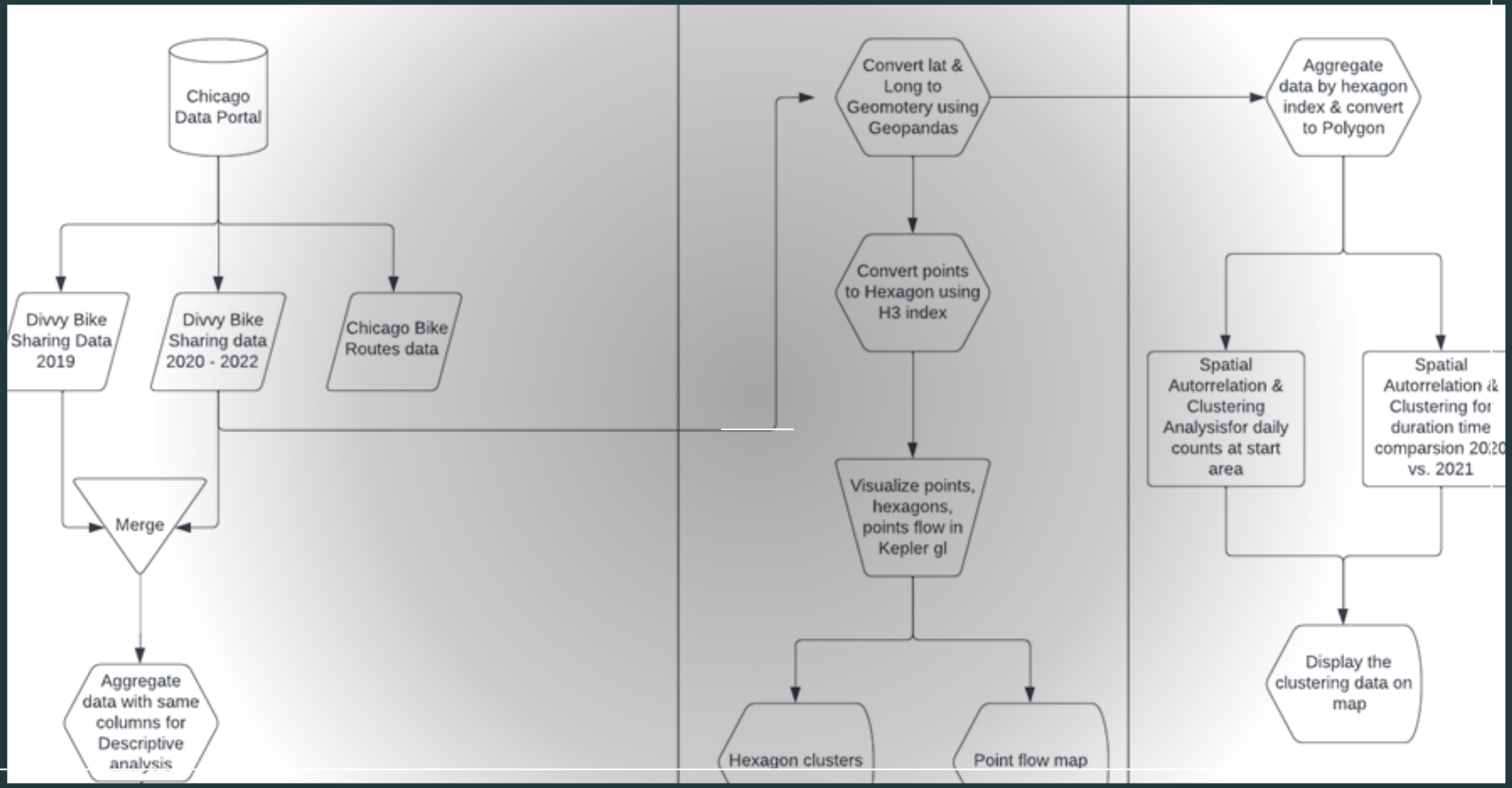
SPECIFIC OBJECTIVES AND QUESTIONS

- What are major monthly and yearly differences on bike sharing volume between 2019 and 2022?
- How the duration time for the overall bike sharing rides has been changed?
- What is the daily bike sharing volume differences by ride types?
- What has changed on daily time series data by membership type during COVID-19?
- How does weekly bike-sharing changes during COVID-19?
- Where is the hot spot for the ride start point? Where needs more bikes?
- Where are the locations that bike-sharing customer tends to use bike longer?

DATA COLLECTION AND DESCRIPTIVE STATISTICS

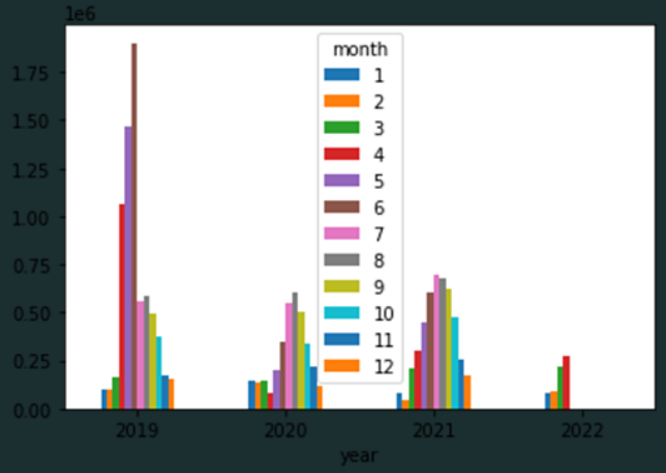
- The main data sets are described as below:
- The bike sharing data is downloaded from Chicago Data Portal (https://data.cityofchicago.org/Transportation/Divvy-Trips/fg6s-gzvg).
  - •Bike sharing data of 2019. There are 3,818,795 rows and 16 columns. No latitude and longitude. No boundary type of data column included.
  - •Bike sharing data between 2020 until the April of 2022. There are 27,282,580 rows and 22 columns. It contains start point and end point latitude and longitude. No boundary type of data column included.
  - •The Data for each trip is anonymized and includes:
    - Trip start day and time
    - Trip end day and time
    - Trip start station
    - Trip end station
    - Rider type (Member, Single Ride, and Day Pass)
    - Latitude
    - Longitude
  - •The combined dataset is huge! There are 37,253,960 rows for the data from 2019 and 27,282,580 from 2020 to 2022. It is not an idea way to analyze spatial data using individual ride geospatial points. My solutions to reduce computation power is aggregating data points to hexagon and the convert to polygon.

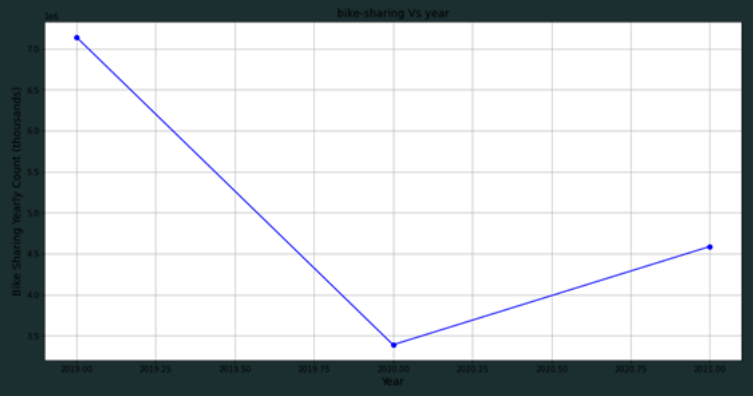
MODIFIED WORKFLOW MODEL

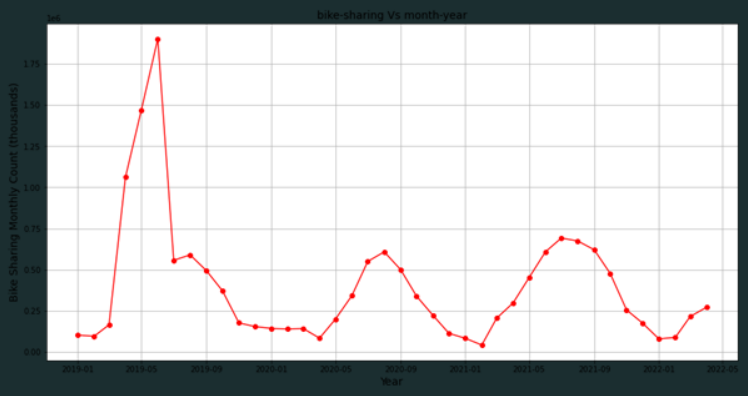


FINDINGS

- What are major monthly and yearly differences on bike sharing volume between 2019 and 2022?

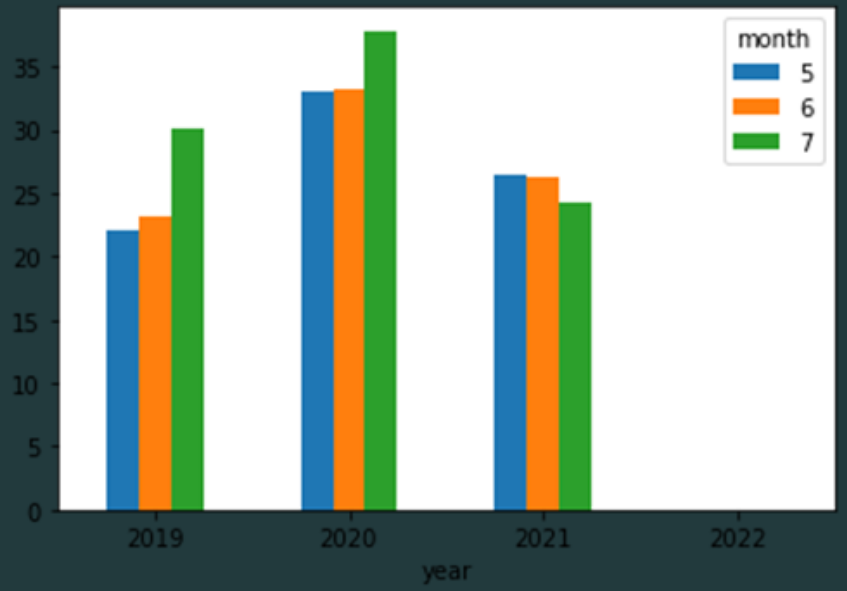


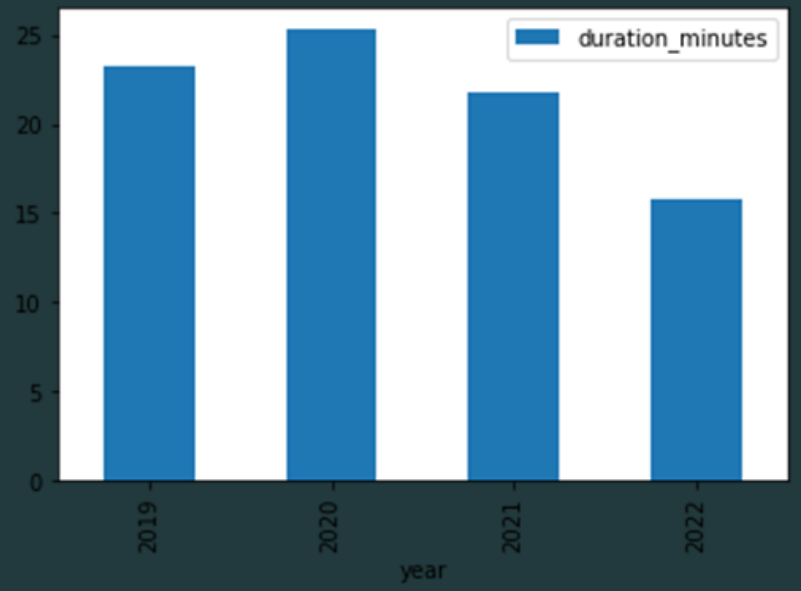




FINDINGS

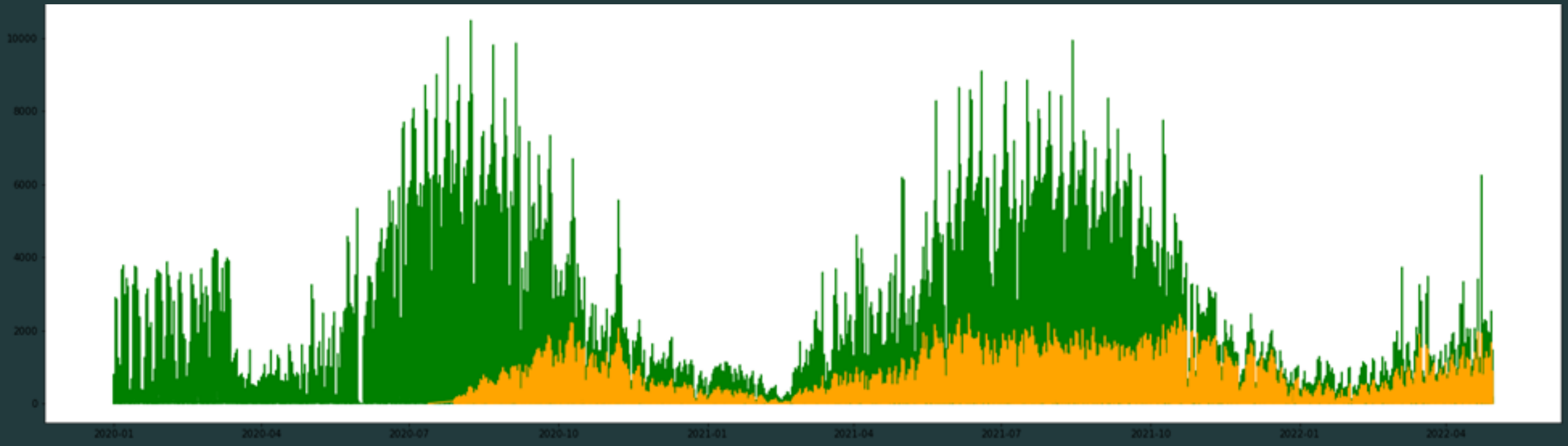
- How the duration time for the overall bike sharing rides has been changed?





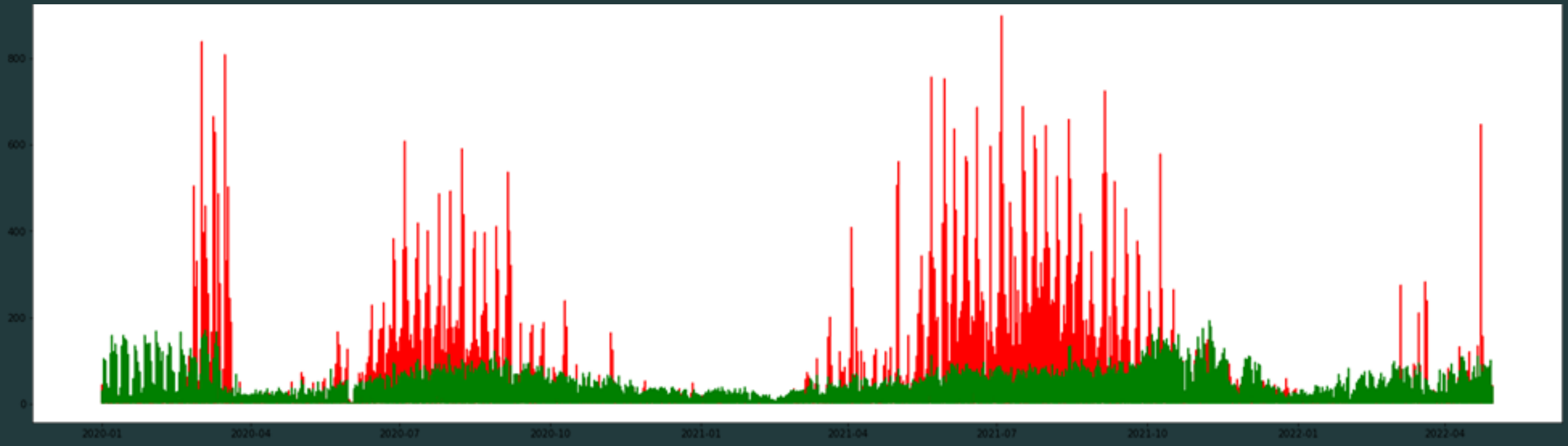
FINDINGS

- What is the daily bike sharing volume differences by ride types?



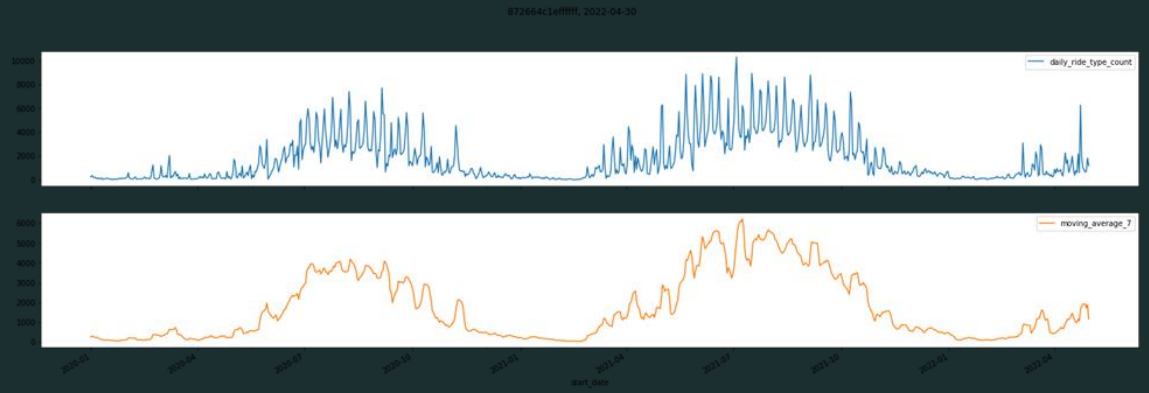
FINDINGS

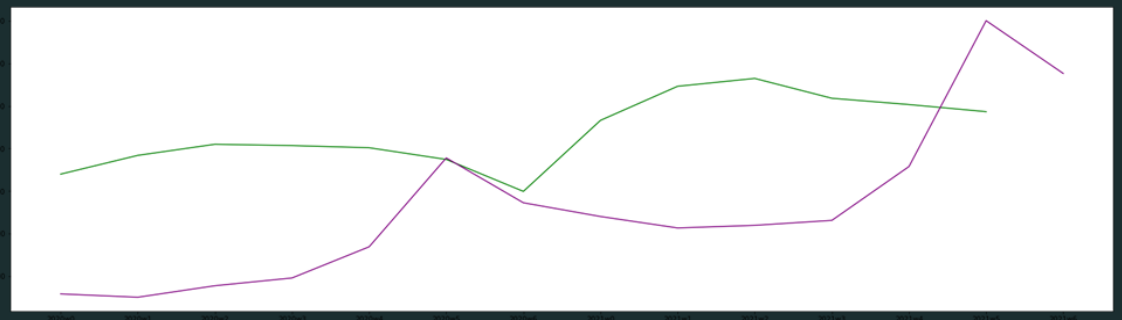
- What has changed on daily time series data by membership type during COVID-19?



FINDINGS

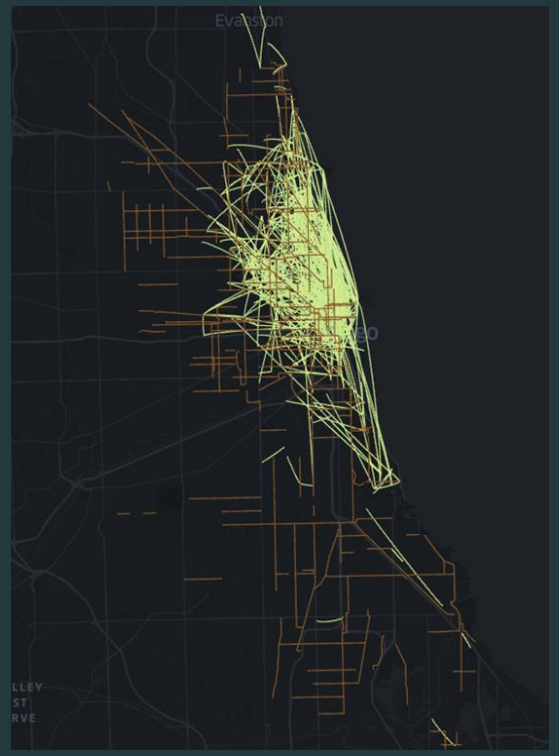
- How does weekly bike-sharing change during COVID-19?

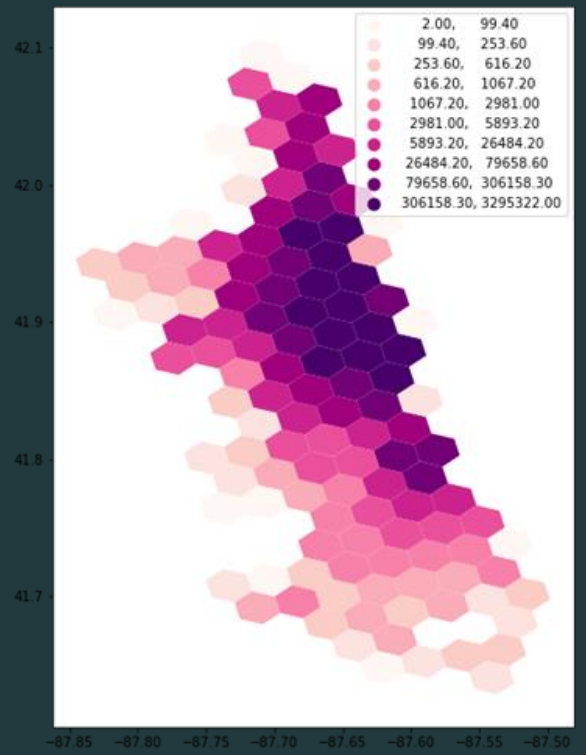


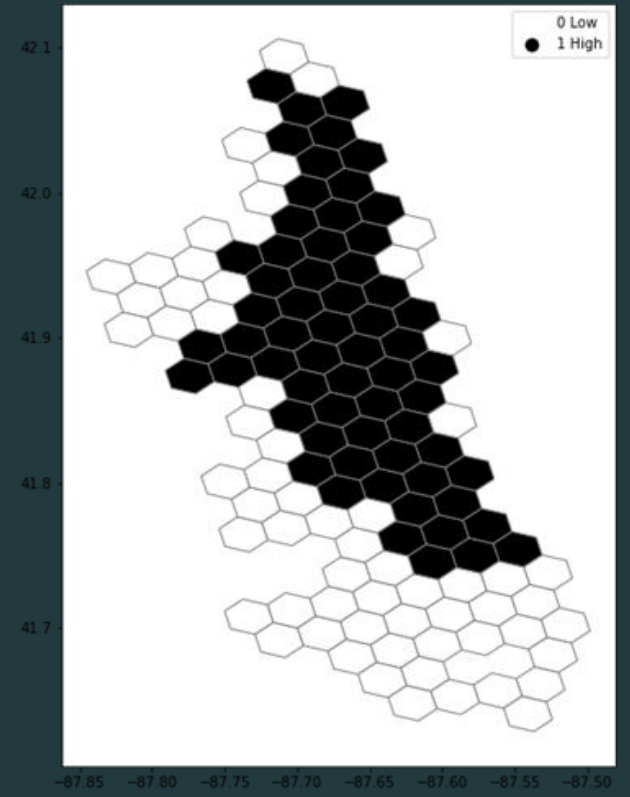


FINDINGS

- Spatial Autocorrelation & Hot/Cold Spot clustering analysis for average daily counts by individual hexagon. (Where is the hot spot for the ride start point? Where needs more bikes?)

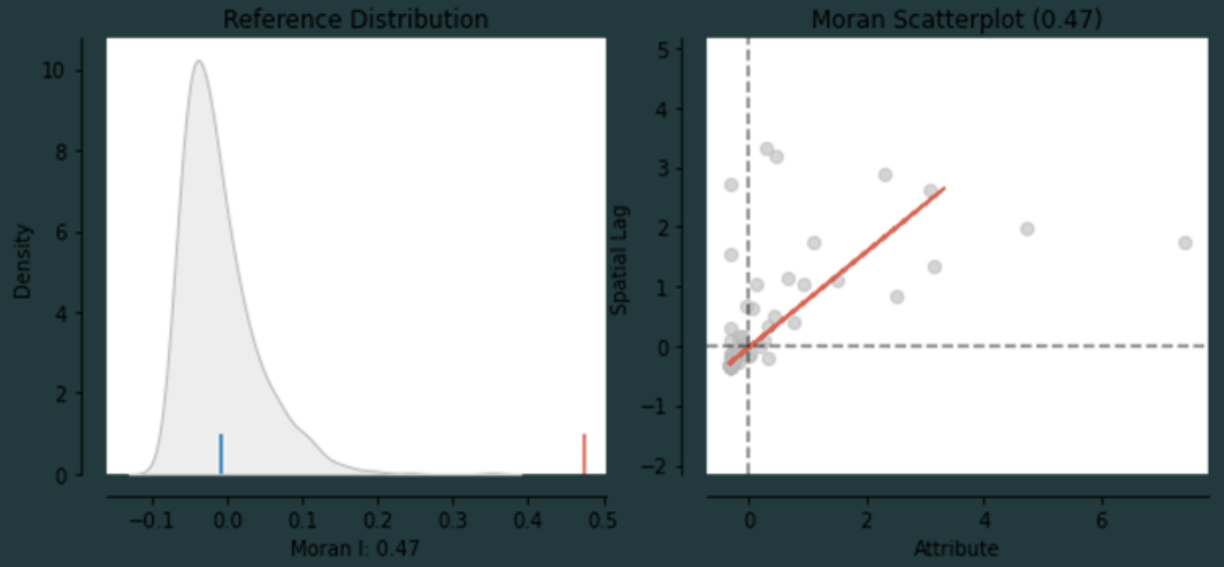


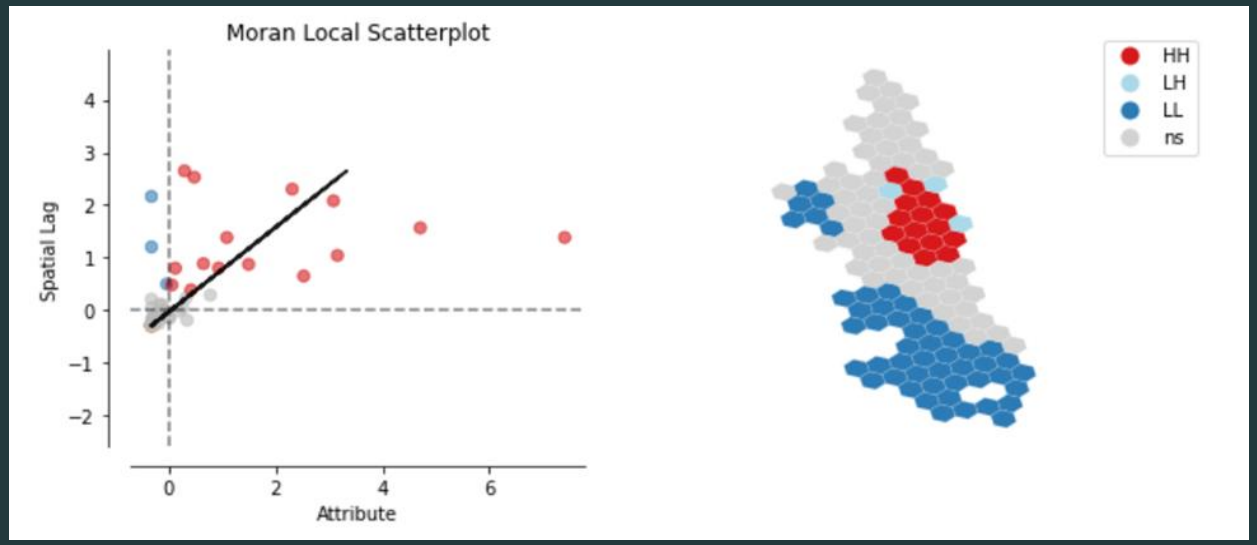




FINDINGS

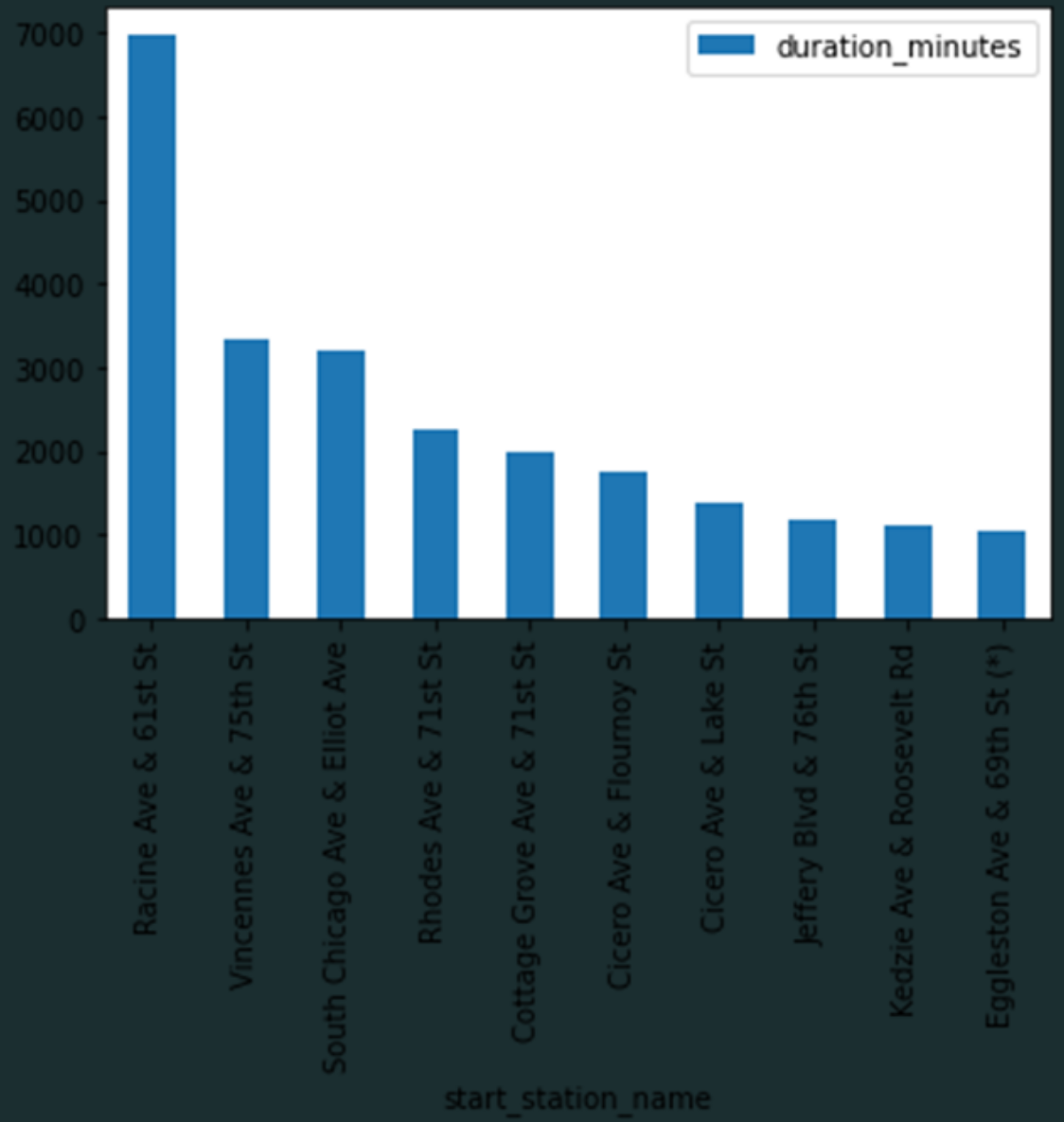
- Spatial Autocorrelation & Hot/Cold Spot clustering analysis for average daily counts by individual hexagon. (Where is the hot spot for the ride start point? Where needs more bikes?)





FINDINGS

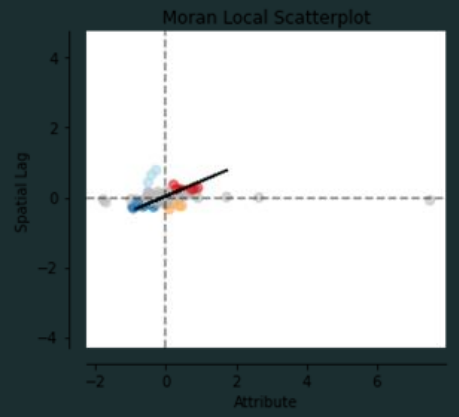
- Spatial Autocorrelation & Hot/Cold Spot clustering analysis for median duration in minute by individual hexagon. (Where is the hot spot for the ride start point? Where needs more bikes?)

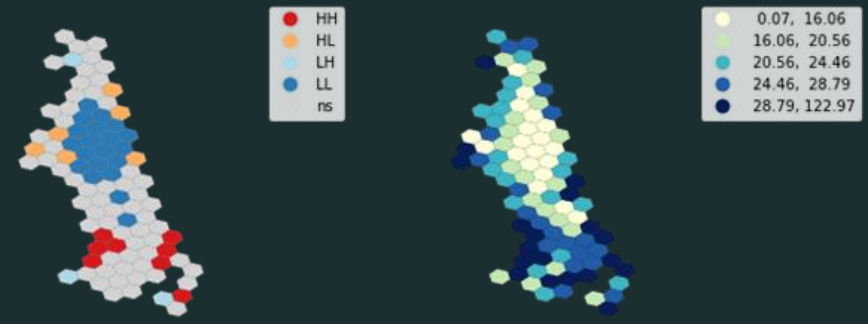


FINDINGS

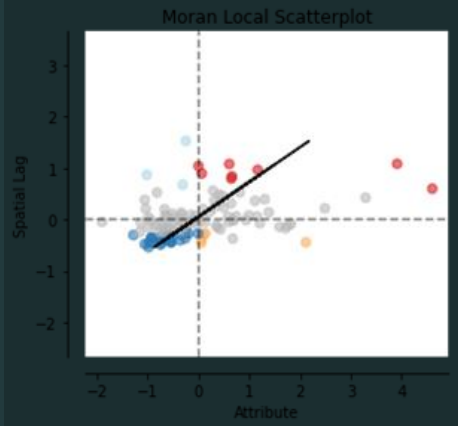
- Spatial Autocorrelation & Hot/Cold Spot clustering analysis for median duration in minute by individual hexagon. (Where is the hot spot for the ride start point? Where needs more bikes?)

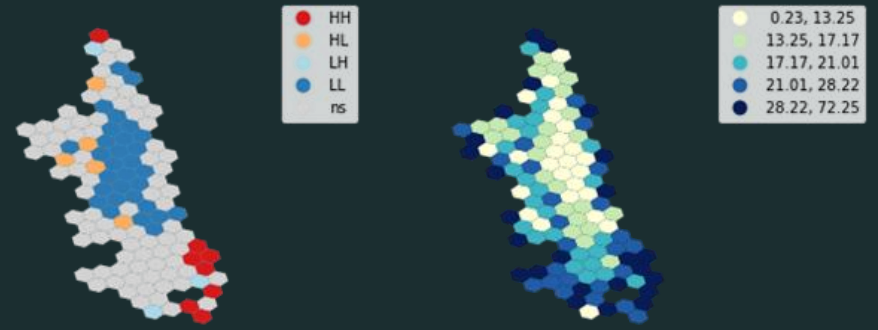
2020



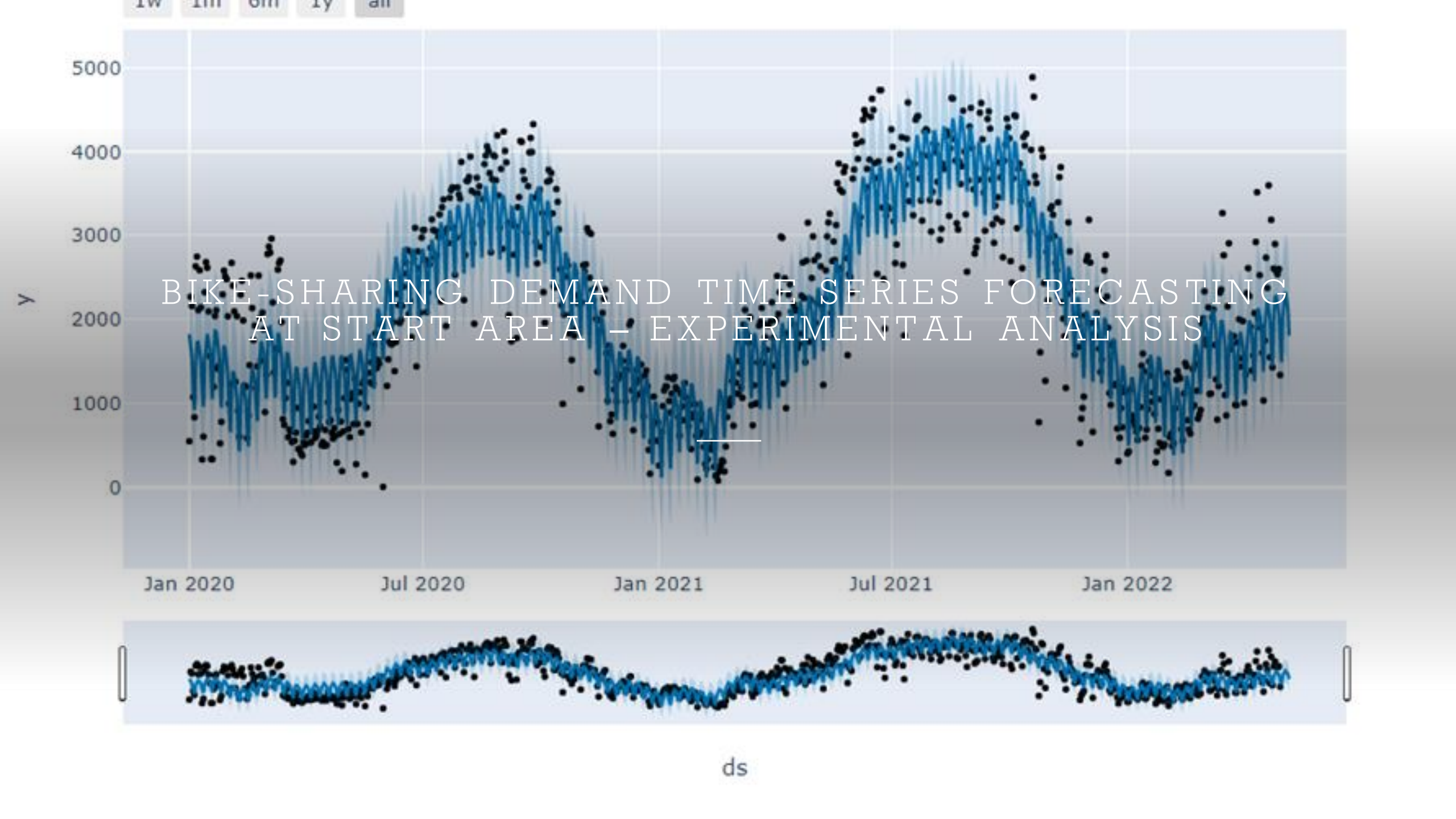


2021





BIKE-SHARING DEMAND TIME SERIES FORECASTING AT START AREA – EXPERIMENTAL ANALYSIS



CONCLUSION

- Since COVID-19 began, overall, yearly, monthly, weekly bike-sharing volume has decreased compared with 2019. 2021 shows certain level of recovery trend but still not as good as 2019.
- Overall ride duration time has reduced. My inference is that electric bike largely reduced the duration time but could also be due to other factors, such as lower demand for tourists. The median of bike duration time shows significant spatial autocorrelations. The clustering for hot/cold spots changed between 2020 and 2021.
- The mean of daily bike-sharing shows significant spatial autocorrelation and clustering trends, using this type of information can help the company who manages the bike-sharing to optimize the bikes into different locations.

LIMITATIONS

- The bike-sharing points data is aggregated on hexagon level. It may lose some good information in this process.
- The spatial autocorrelation and clustering analysis are also limited because many other factors can impact on the overall clustering patterns. It needs future analysis to have more holistic view of the changes.

REFERENCES

- Bike share responses to COVID-19: https://www.sciencedirect.com/science/article/pii/S2590198221000609
- H3: Uber's Hexagonal Hierarchical Spatial Index: https://eng.uber.com/h3/
- Pysal spatial autocorrelation: https://pysal.org/notebooks/viz/splot/esda_morans_viz.html