Bike Sharing Geospatial Data Analysis During COVID-19

Course Project – Final Paper

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

1. What are major monthly and yearly differences on bike sharing volume between 2019 and 2022?
2. How the duration time for the overall bike sharing rides has been changed?
3. What is the daily bike sharing volume differences by ride types?
4. What has changed on daily time series data by membership type during COVID-19?
5. How does weekly bike-sharing changes during COVID-19?
6. Where is the hot spot for the ride start point? Where needs more bikes?
7. Where are the locations that bike-sharing customer tends to use bike longer?
8. Future demand time series forecasting? (This turns out to be not the core question in this analysis due to time and computation limitations, but I still included it here since I have done some experimental research) **

DATA COLLECTION AND DESCRIPTIVE STATISTICS

There are two types of data has been collected listed as below:

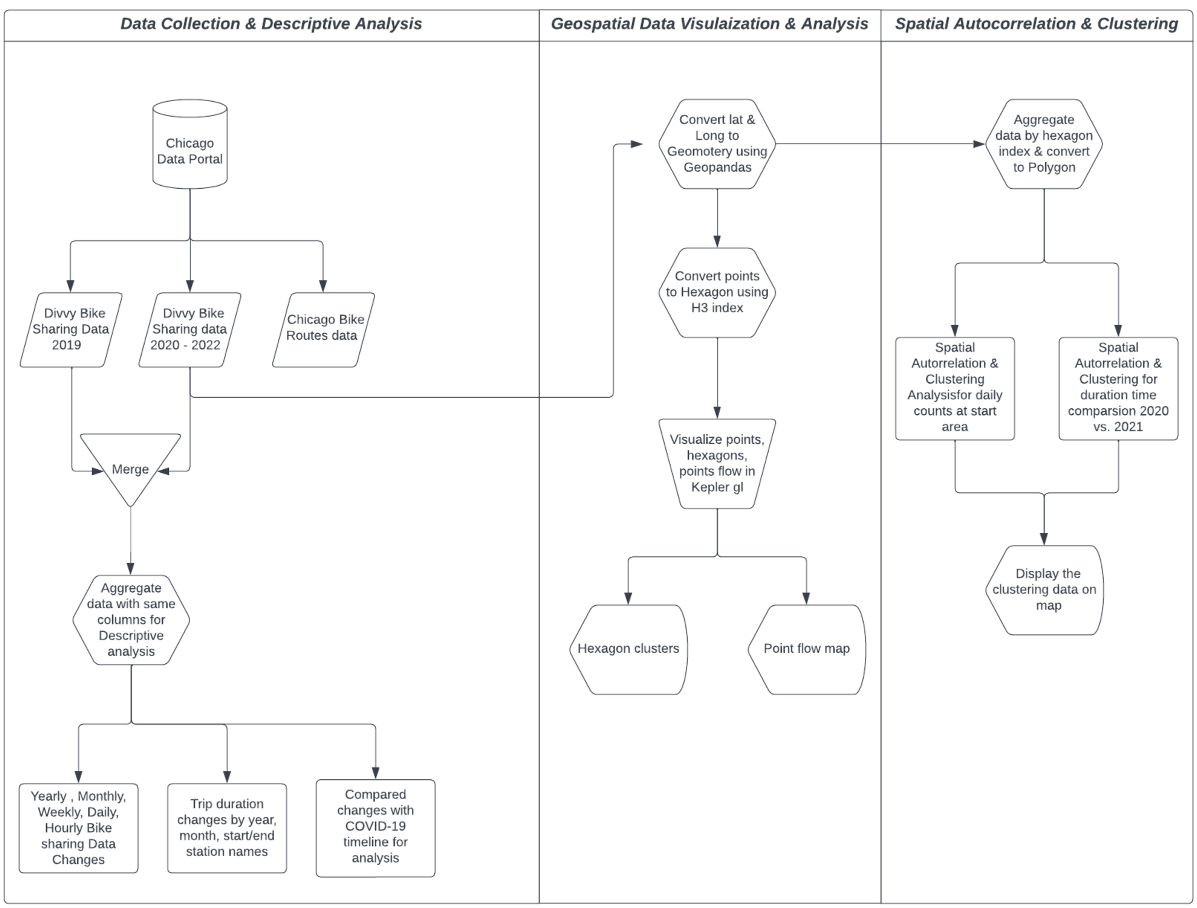
1. The bike sharing data is downloaded from Chicago Data Portal (https://data.cityofchicago.org/Transportation/Divvy-Trips/fg6s-gzvg).
    a. Bike sharing data of 2019. There are 3,818,795 rows and 16 columns. No latitude and longitude. No boundary type of data column included.
    b. Bike sharing data 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.
    c. **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
    d. 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.
    e. I was not able to upload all data to GitHub because the data size is close to 10GB. The below is the screenshot of the original datasets.

Data 7.06.20 PM	May 8, 2022 at 7:48 PM	--	Folder
2019	May 8, 2022 at 8:11 PM	--	Folder
Divvy_Trips_2019_Q4.csv	Jan 21, 2020 at 3:00 PM	97 MB	CSV Document
Divvy_Trips_2019_Q3.csv	Nov 4, 2019 at 3:51 PM	225.5 MB	CSV Document
Divvy_Trips_2019_Q2.csv	Jul 16, 2019 at 2:56 PM	152.6 MB	CSV Document
Divvy_Trips_2019_Q1.csv	Jul 16, 2019 at 2:35 PM	50.5 MB	CSV Document
2020_2021_2022	May 8, 2022 at 7:47 PM	--	Folder
202108-divvy-tripdata.csv	May 8, 2022 at 4:33 PM	151 MB	CSV Document
202107-divvy-tripdata.csv	May 8, 2022 at 4:32 PM	154 MB	CSV Document
202106-divvy-tripdata.csv	May 8, 2022 at 4:32 PM	136.4 MB	CSV Document
202105-divvy-tripdata.csv	May 8, 2022 at 4:32 PM	99.9 MB	CSV Document
202104-divvy-tripdata.csv	May 8, 2022 at 4:32 PM	64 MB	CSV Document
202102-divvy-tripdata.csv	May 8, 2022 at 4:32 PM	9.4 MB	CSV Document
202103-divvy-tripdata.csv	May 8, 2022 at 4:32 PM	43.6 MB	CSV Document
202012-divvy-tripdata.csv	May 8, 2022 at 4:32 PM	25 MB	CSV Document
202101-divvy-tripdata.csv	May 8, 2022 at 4:32 PM	18.4 MB	CSV Document
202011-divvy-tripdata.csv	May 8, 2022 at 4:32 PM	46.2 MB	CSV Document
202010-divvy-tripdata.csv	May 8, 2022 at 4:32 PM	69.3 MB	CSV Document
202204-divvy-tripdata.csv	May 3, 2022 at 8:28 AM	66.9 MB	CSV Document
202203-divvy-tripdata.csv	Apr 5, 2022 at 12:00 PM	51.7 MB	CSV Document
202202-divvy-tripdata.csv	Mar 2, 2022 at 9:46 AM	21.1 MB	CSV Document
202201-divvy-tripdata.csv	Feb 1, 2022 at 10:40 AM	19 MB	CSV Document
202112-divvy-tripdata.csv	Jan 5, 2022 at 12:22 PM	44.7 MB	CSV Document
202111-divvy-tripdata.csv	Dec 8, 2021 at 8:25 AM	65.4 MB	CSV Document
202110-divvy-tripdata.csv	Nov 4, 2021 at 1:20 PM	116 MB	CSV Document
202109-divvy-tripdata.csv	Oct 4, 2021 at 11:02 AM	141.1 MB	CSV Document
202009-divvy-tripdata.csv	Oct 13, 2020 at 2:04 PM	95.5 MB	CSV Document
202008-divvy-tripdata.csv	Sep 4, 2020 at 9:59 AM	111 MB	CSV Document
202007-divvy-tripdata.csv	Aug 11, 2020 at 8:36 PM	97.6 MB	CSV Document
202006-divvy-tripdata.csv	Jul 5, 2020 at 7:11 PM	60.8 MB	CSV Document
202005-divvy-tripdata.csv	Jun 1, 2020 at 9:44 AM	33.6 MB	CSV Document
202004-divvy-tripdata.csv	Jun 1, 2020 at 9:43 AM	14.3 MB	CSV Document
Divvy_Trips_2020_Q1.csv	May 26, 2020 at 7:13 PM	71.3 MB	CSV Document

2. Bike routes data is downloaded from Chicago Data Portal: https://data.cityofchicago.org/Transportation/Bike-Routes/3w5d-sru8

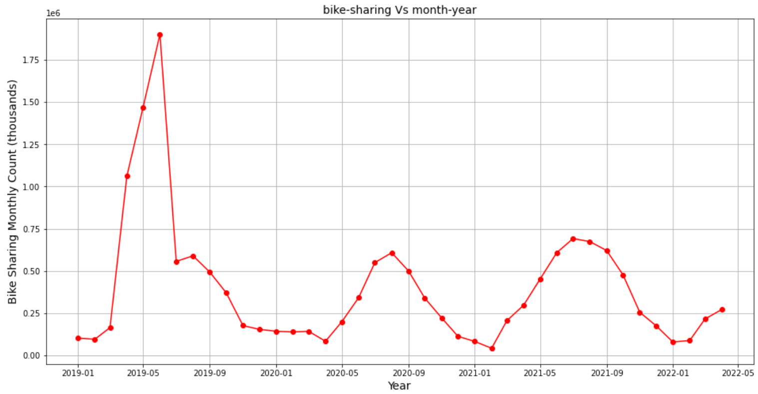
MODIFIED WORKFLOW MODEL

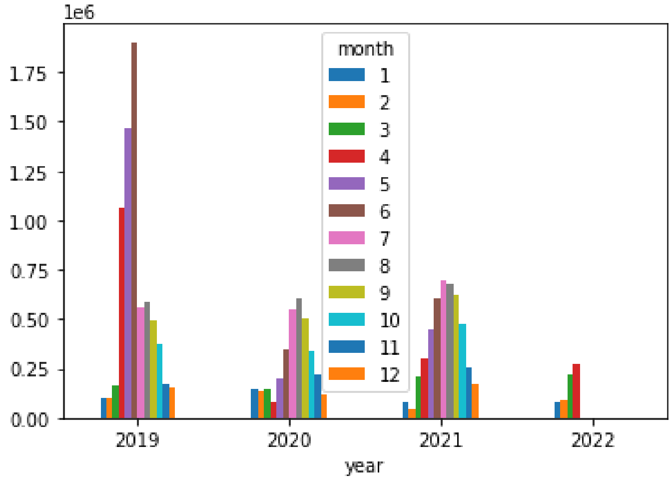
1. I converted start points (latitude & longitude) to hexagon using Uber H3 package in order to reduce the need of computation power.
2. The hexagon index and further converted to polygon in order to do spatial autocorrelation & clustering analysis.
3. Time series forecasting analysis is formally removed from original project proposal due to the time and computation limitations. But I still provide some results with attached notebooks in this project just for your reference.
4. A few new tools and packages used besides arcpy:
    - Uber H3 package & Kepler gl viz
    - Pysal package

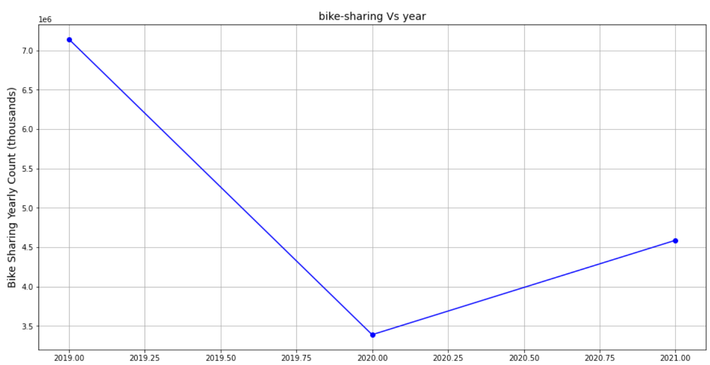


FINDINGS

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



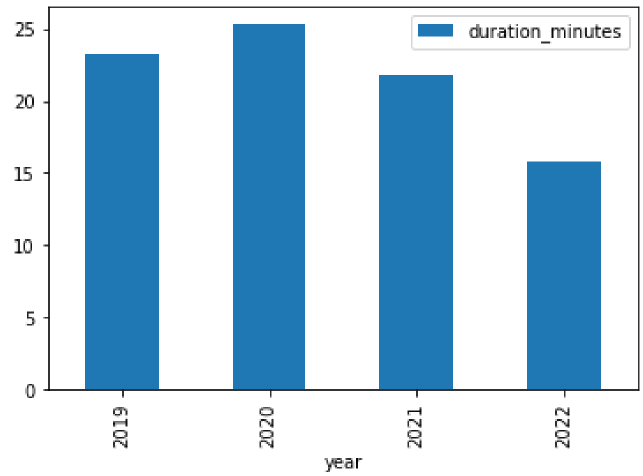


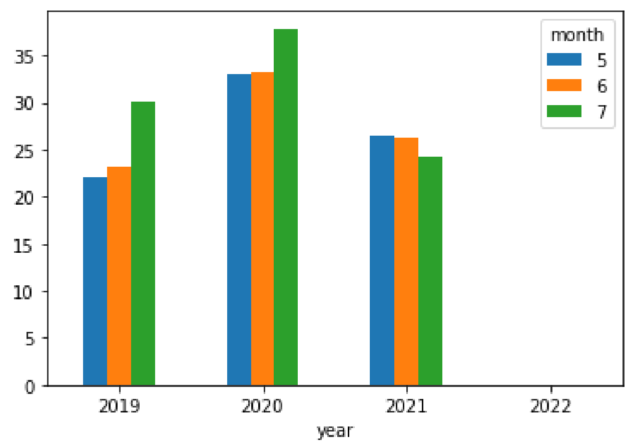


- Overall bike sharing volume is much higher before COVID-19 hit in 2019 compared with 2020 and 2021. The bike sharing in April and March of this year is higher than the same month in 2020, 2021, which showing some recovery signals.
- The bike sharing volume shows seasonality. There are not many bikes sharing demand in winter but more demand in summertime.
- Before covid hits, in 2019, the bike sharing volume is significantly higher than those in covid years.

- The bike sharing significantly decreased in April of 2020 right after the Covid began, whereas it usually increases in April in other years.
- The bike sharing volume is higher in 2021 compared with 2020.
- The bike sharing in April and March of this year is higher than the same month in 2020, 2021, which showing some recovery signals.

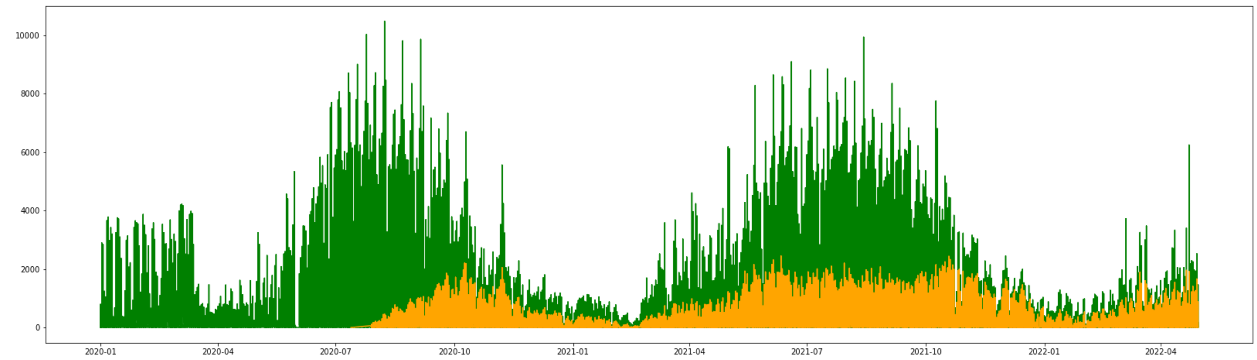
2. How the duration time for the overall bike sharing rides has been changed?





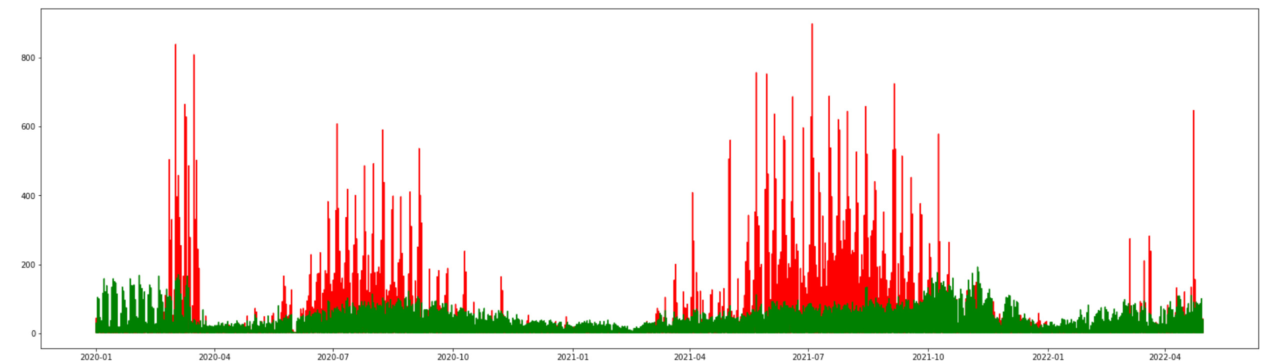
- Average bike sharing duration in minutes dropped since COVID-19 began in 2020.
- Average bike sharing show overall decrease in 2021 compare with 2020.

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



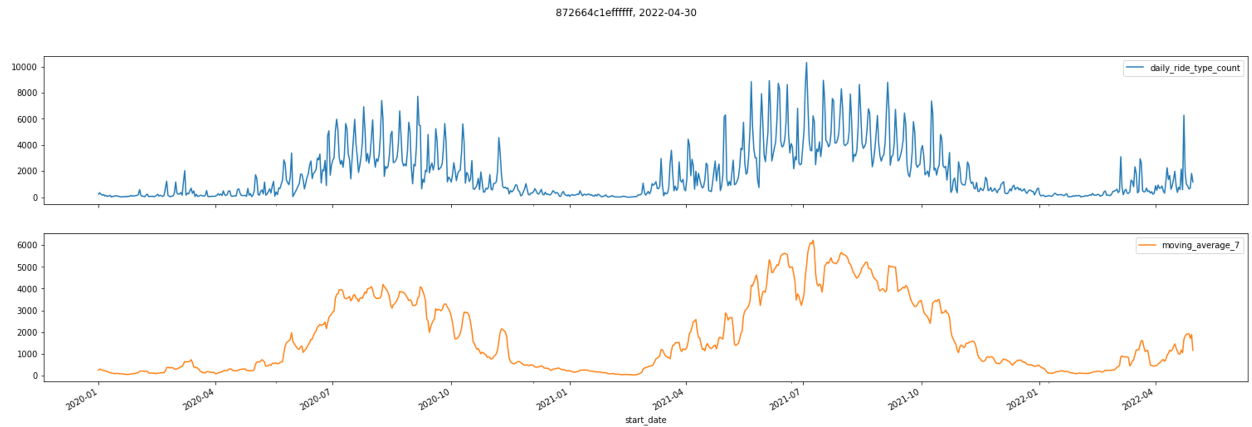
- It shows strong seasonality: more in summer, less in winter
- Electric bike is the new riding type since July of 2020. It shows more flat seasonality compared with traditional bike. (Green below is traditional bike, orange is electric bike)
- Bike sharing in winter is much higher before COVID hits, which remains low in the past two years.

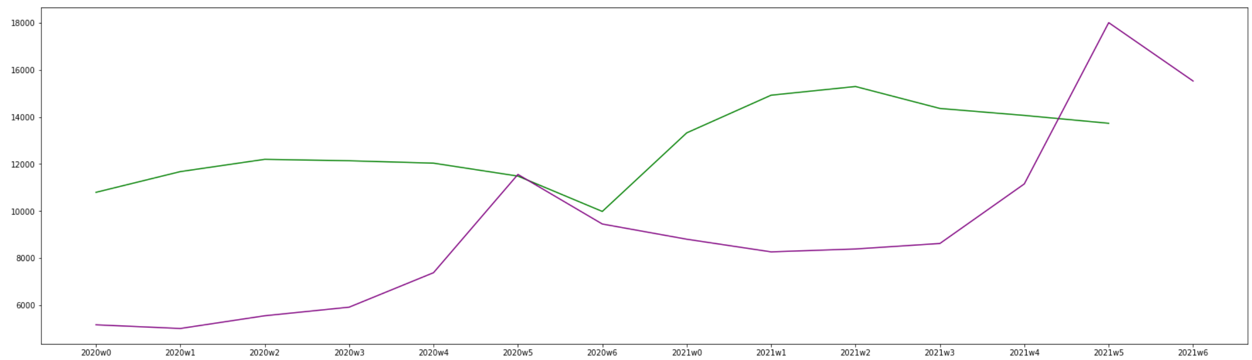
4. What has changed on daily time series data by membership type during COVID-19?



- Green above is with memberships, red is casual. Bike sharing with membership with less variability.
- Bike sharing with casual visit showing more demand than member demand during the summertime.
- It's interesting to see the peek for red, is the bottom for green, the same is true for green. Which makes sense that more causal visit during the weekend but more demand for weekdays for memberships.
- Bike sharing demand for winter is much lower for member after COVID hits in green.
- Casual demand in read is much lower in 2020 than 2021 in the summer.

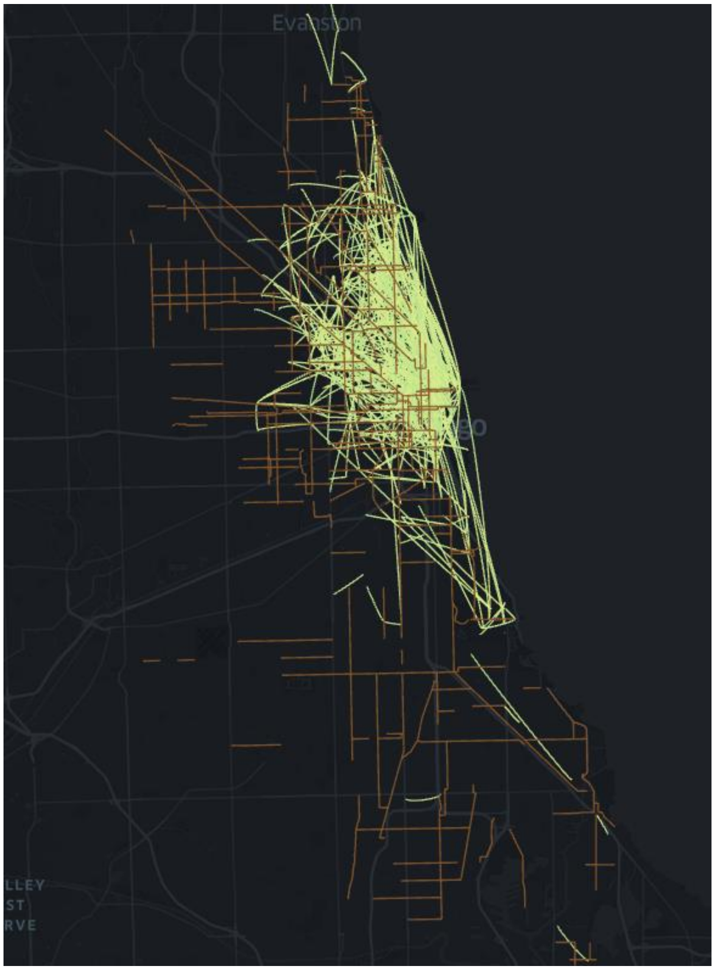
5. How does weekly bike-sharing change during COVID-19?



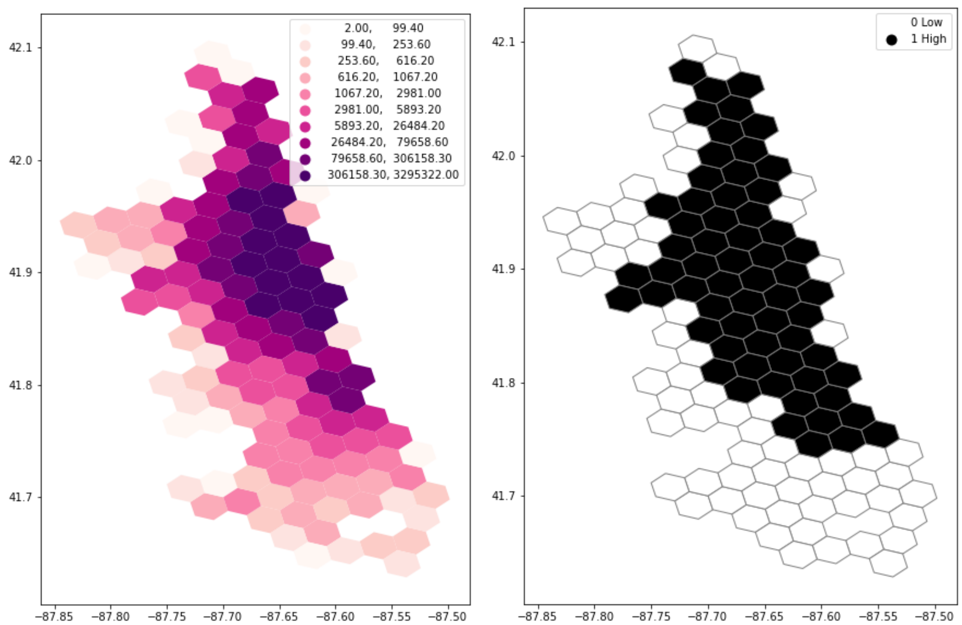


- The individual hexagon follows the overall time series pattern with trend and seasonality (both weekly and monthly).
- It shows more stable demand for memberships (in green) than casual type (in purple), especially in winter to sprint time.
- Overall, it has higher demand during weekdays, especially on Tuesdays, Wednesdays, and Fridays.
- Weekly demand is much higher in 2021 than 2020.
- the member and casual bike sharing demand in a week is trending in an opposite way. For member bike sharing, the demand is mostly during the weekdays, the demand for casual bike sharing is mostly during the weekend and the peek is Fridays.

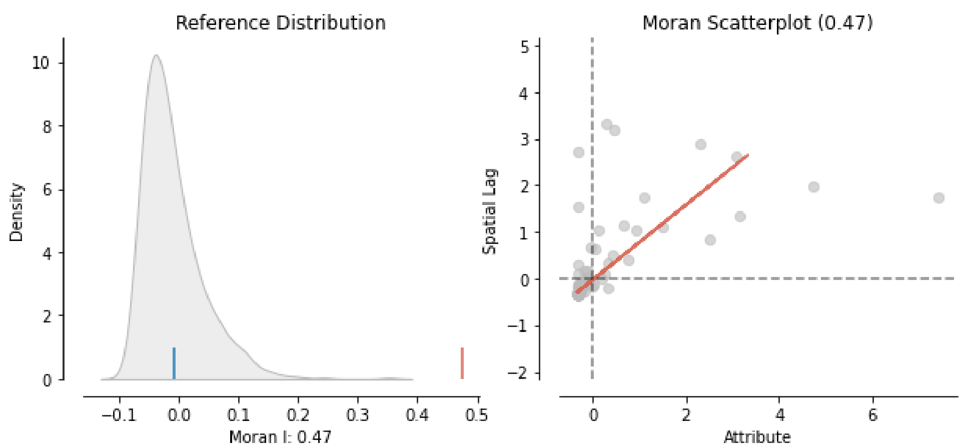
6. Spatial Autocorrelation & Hot/Cold Spot clustering analysis for average daily counts by individual hexagon. (Where is the hot spot for the ride start point? Where needs more bikes?)



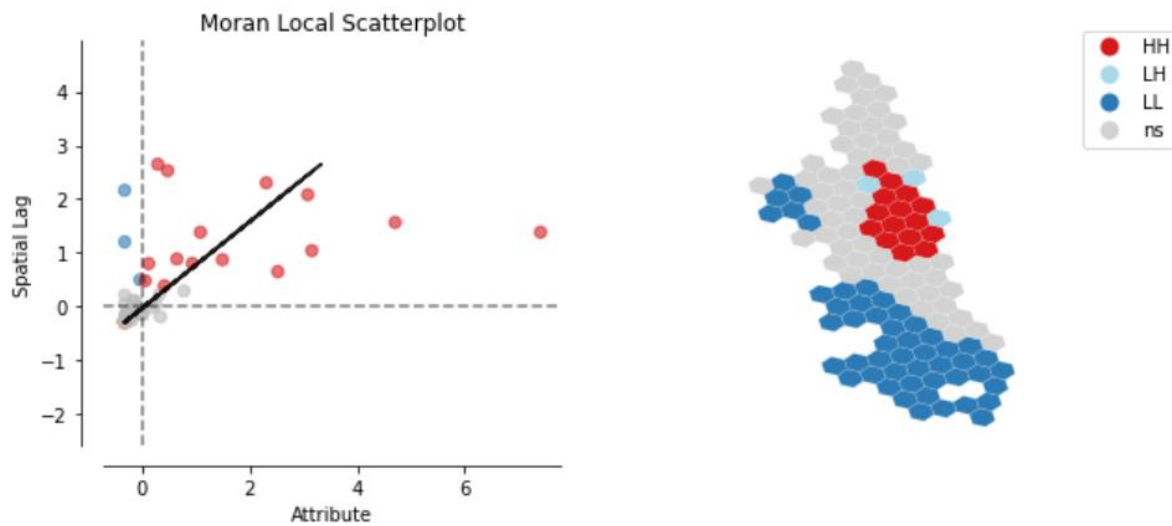
- The above sample (only contains 1,000 rides) start-point to end-point visualization is generated using kepler gl. It shows in some areas are busier that others. But it needs to be further analyzed.



- It shows above that if the dark hexagon areas only analyzed in global level, then the dark points represent hot spots in the relative in the middle areas.

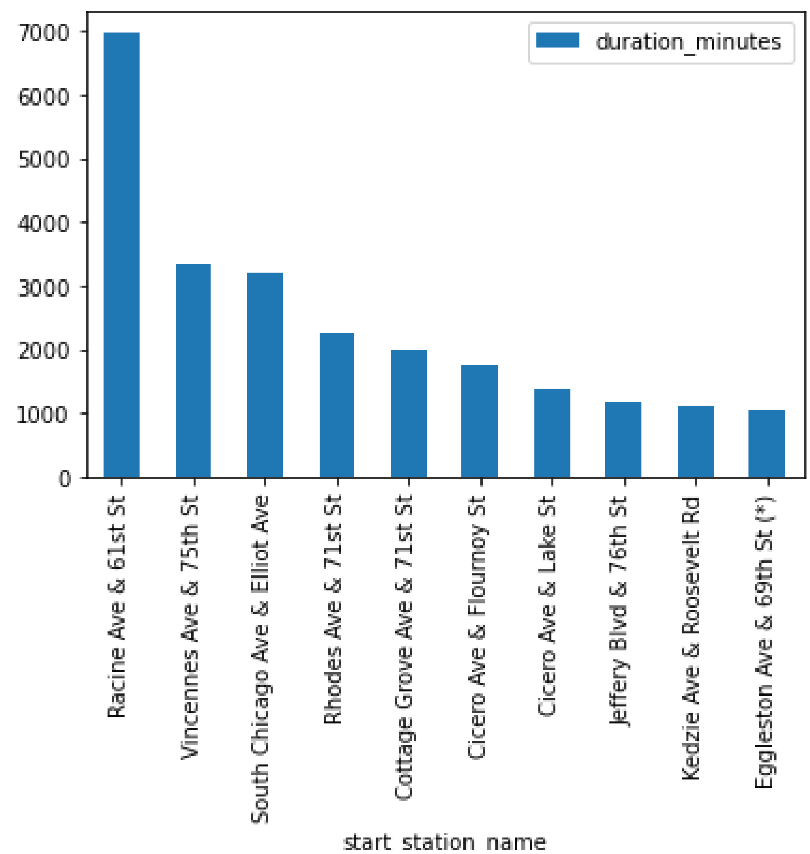


- The Moran's I analysis shows significant autocorrelation with correlation value of 0.47 and p-value=0.001



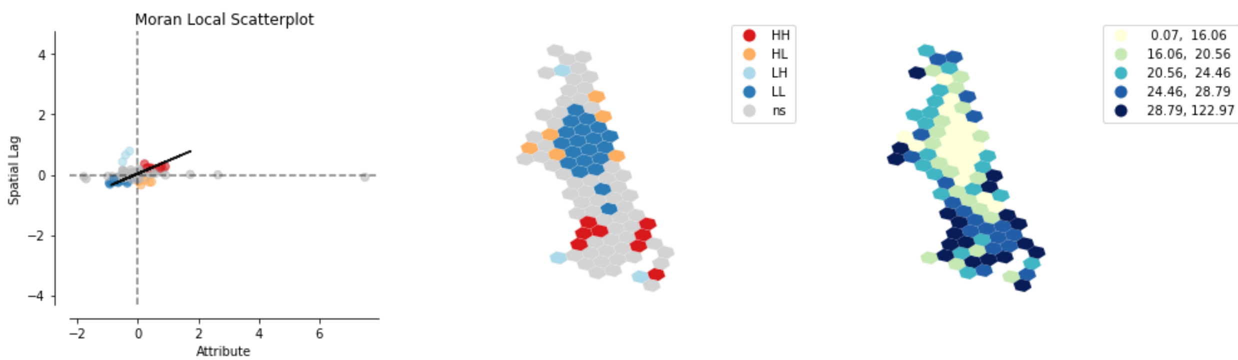
- The red shows hot pot and blue shows cold spot. The hot spot areas on the real map shows a lot of attraction spots such as navel pier, Willis tower, field museum, which makes hot spot meaningful and making sense.

7. 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?)

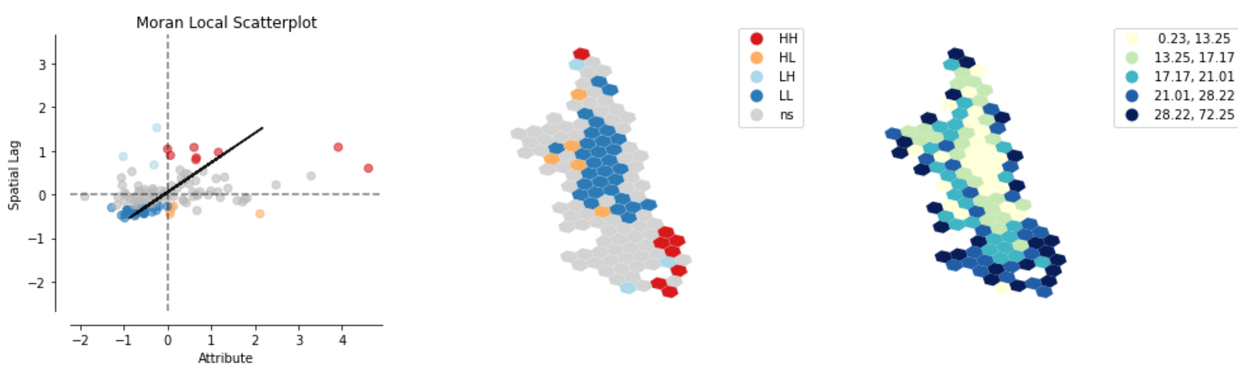


- The above graph shows the top 10 start point location for people do bike sharing.
- The blow two screens are showing hot/cold clustering analysis for 2020 and 2021 respectively.

2020 →

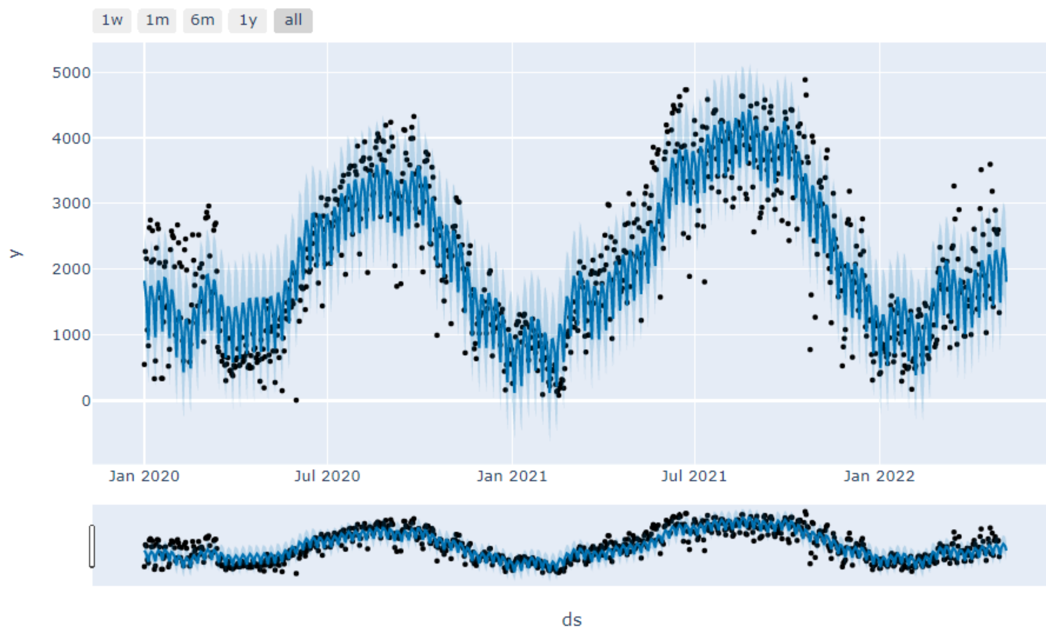


2021→



- Comparing the clustering differences, I see the hot spot locations have changed in 2021. One big hot spot on the left disappeared in 2021. The cold spots remain similar pattern in 2021.
- Comparing the quantile values between 2020 and 2021 for median time duration, we can see that overall, it's higher in 2020 than 2021. It could relates to the electric bike became available later 2020, which can largely reduce the bike-sharing duration time.

8. Bike-sharing demand time series forecasting at start area



The forecasted data looks like above. To zoom in, the forecasted demand for the next week is looking like below (the original only to April, 2022).

```
# Python
forecast = m.predict(future)
forecast[['ds', 'yhat', 'yhat_lower', 'yhat_upper']]
```

	ds	yhat	yhat_lower	yhat_upper
0	2020-01-01	1825.776218	1115.321970	2540.129565
1	2020-01-02	1696.748200	998.481449	2389.442515
2	2020-01-03	1551.541752	814.573848	2305.611534
3	2020-01-04	1194.529628	515.693868	1882.470530
4	2020-01-05	924.075465	273.642497	1653.698893
...	...	...	...	...
851	2022-05-03	2319.550915	1590.608311	2968.102888
852	2022-05-04	2330.531370	1609.556456	3014.652709
853	2022-05-05	2242.656170	1556.312749	2989.042843
854	2022-05-06	2133.725647	1452.854304	2882.704383
855	2022-05-07	1808.103996	1081.688674	2461.424601

856 rows × 4 columns

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

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