IBM DATA SCIENCE CAPSTONE PROJECT

Yaxuan XUE

Introduction

I have been always longing to visit Orlando, Florida.

In this project, I will use **Foursquare** to choose our ideal hotels that are close to our destination—Universal Studios Florida, and at the same time has a relatively good rating.

Then, I will make a list of nearby restaurants that are located walking-distance from our selected hotel.

Finally,I will build an **ARIMA** model using the active cases data of United States to predict when this COVID-19 can be under control.

Data preparation

1) address of the Universal Studios Florida:

   6000 Universal Blvd, Orlando, FL

2) active COVID-19 cases in the United States

   source: https://www.coronanet-project.org/download

   Total Active Cases = Total Positive Cases - Total Deaths - Total Recovered

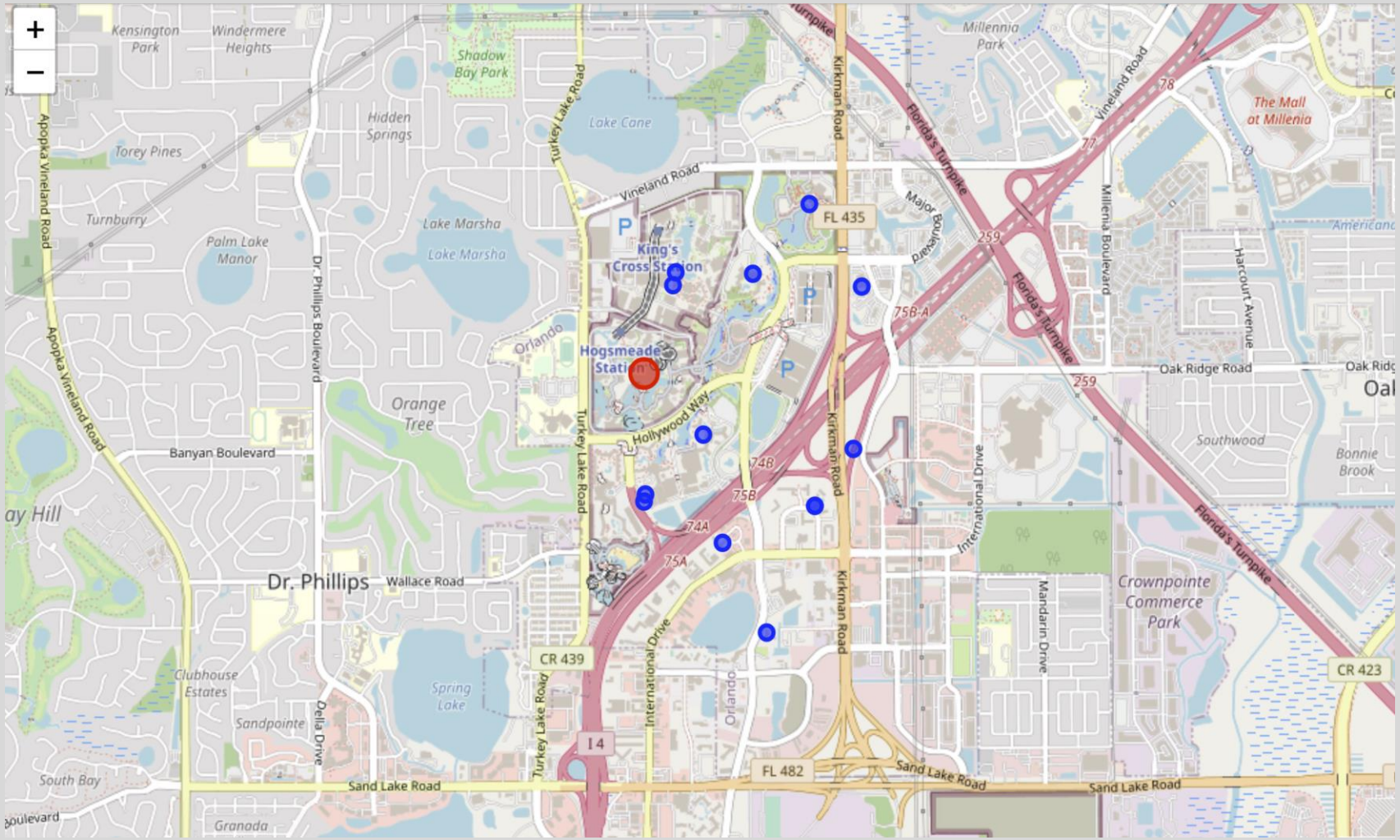
Accommodation

- **Getting latitude and longitude coordinates**

```
address = '6000 Universal Blvd, Orlando, FL'
geolocator = Nominatim(user_agent="foursquare_agent")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print(latitude, longitude)
```

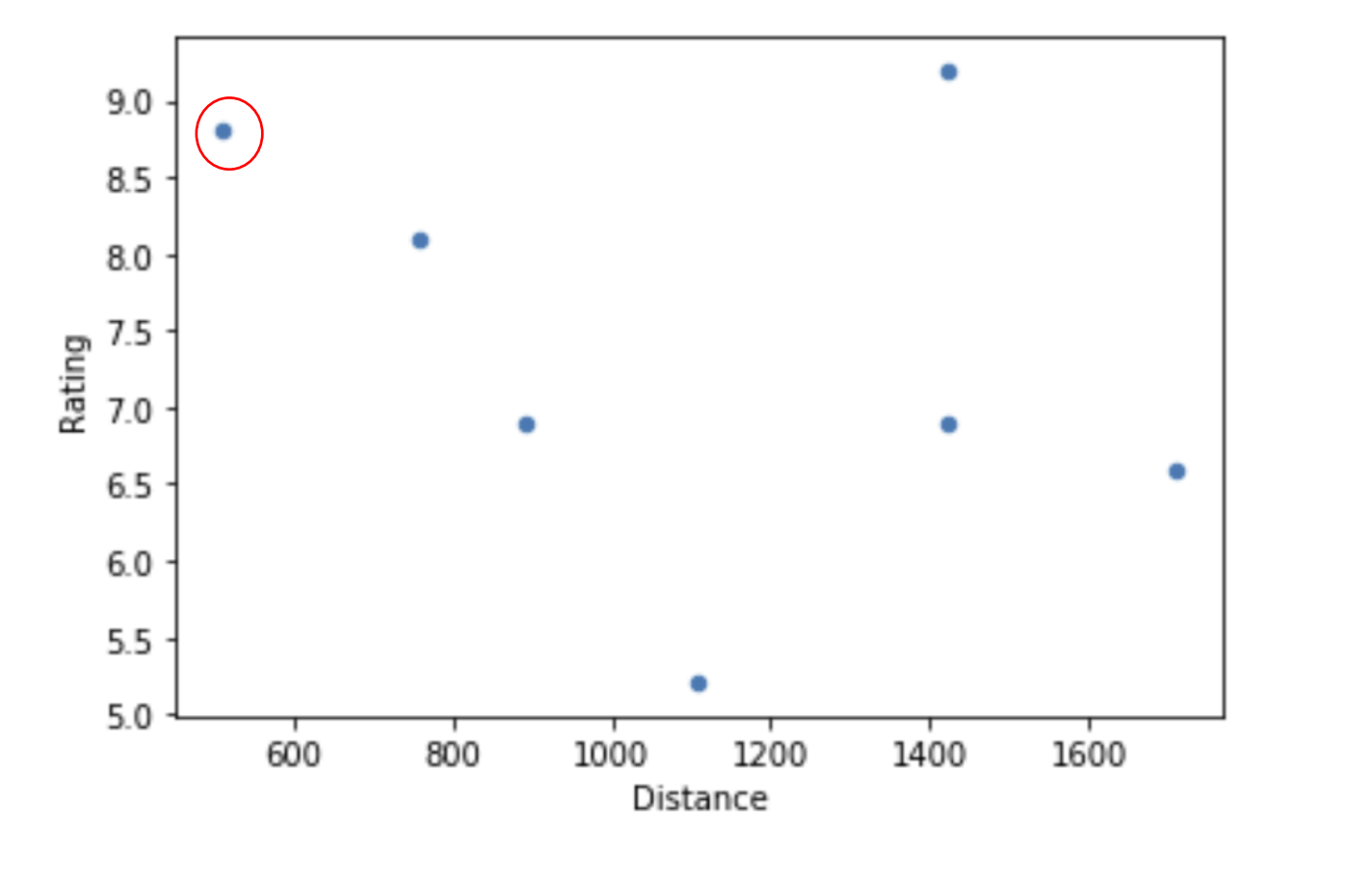
```
28.471878949999997 -81.47121064527349
```

- **Searching nearby hotels**

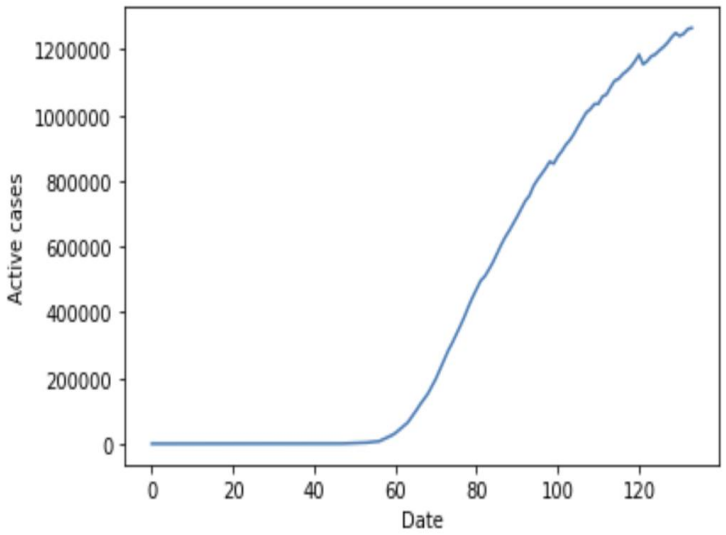


- **Comparing the ratings**

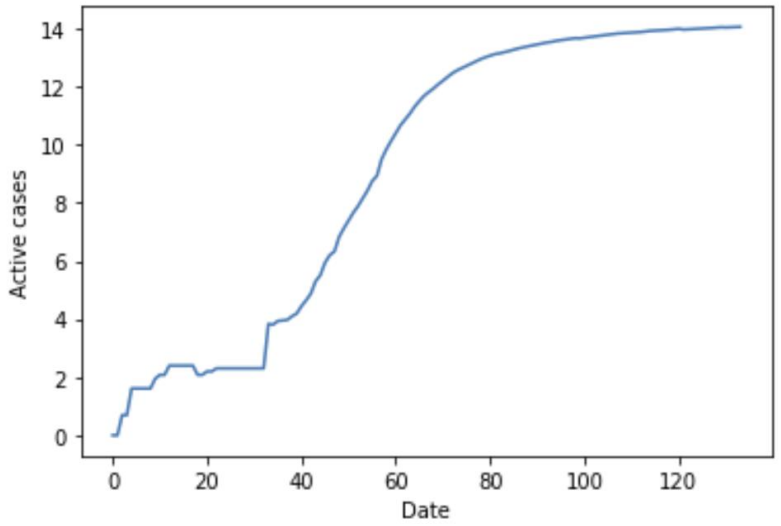
	Name	Rating	Distance
0	Loews Portofino Bay Hotel at Universal Orlando	9.2	1423
1	Loews Royal Pacific Resort	8.8	511
2	Universal's Aventura Hotel	8.1	757
3	Club 7 at the Hard Rock Hotel	6.9	891
4	Holiday Inn & Suites Across From Universal Orl...	6.9	1421
5	Best Western Orlando Gateway Hotel	6.6	1709
6	Orlando Continental Plaza Hotel	5.2	1107



- From the plot, the point in red circle which represents Loews Royal Pacific Resort seems to be the most ideal hotel that meets our requirements.



Active cases



log(active cases)

```
ADF Statistic: -1.645296370982438
p-value: 0.4594946462270914
Critical Values:
        1%: -3.4816817173418295
        5%: -2.8840418343195267
        10%: -2.578770059171598
```

COVID-19 Active Cases Study

1. **ADF test**
- **Active cases**
- **Log(active cases)**

```
ADF Statistic: -4.192485632973042
p-value: 0.0006784665366878 41
Critical Values:
        1%: -3.4816817173418295
        5%: -2.8840418343195267
        10%: -2.578770059171598
```

The first difference of log(active cases) is stationary. We can then use this series to run ARMA regression.

1. **ADF test**
- **First difference of log(active cases)**

ARMA Model Results

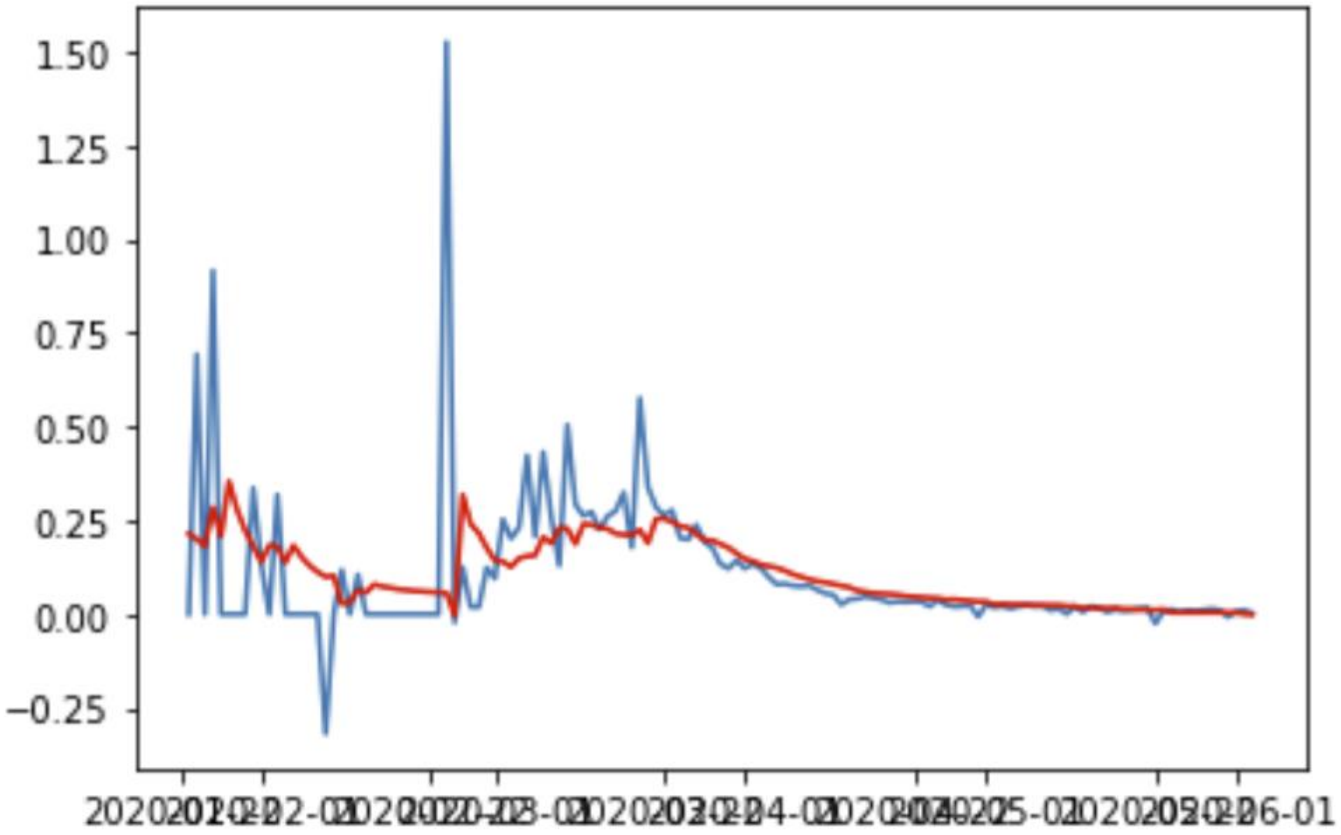
Dep. Variable:	log_active	No. Observations:	133
Model:	ARMA(2, 1)	Log Likelihood	41.984
Method:	css-mle	S.D. of innovations	0.176
Date:	Wed, 01 Jul 2020	AIC	-71.968
Time:	00:37:58	BIC	-54.626
Sample:	01-23-2020 - 06-03-2020	HQIC	-64.921

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.2158	0.072	3.003	0.003	0.075	0.357
time	-0.0016	0.001	-1.772	0.076	-0.003	0.000
ar.L1.log_active	0.7034	0.127	5.532	0.000	0.454	0.953
ar.L2.log_active	0.2011	0.095	2.115	0.034	0.015	0.387
ma.L1.log_active	-0.7422	0.103	-7.239	0.000	-0.943	-0.541

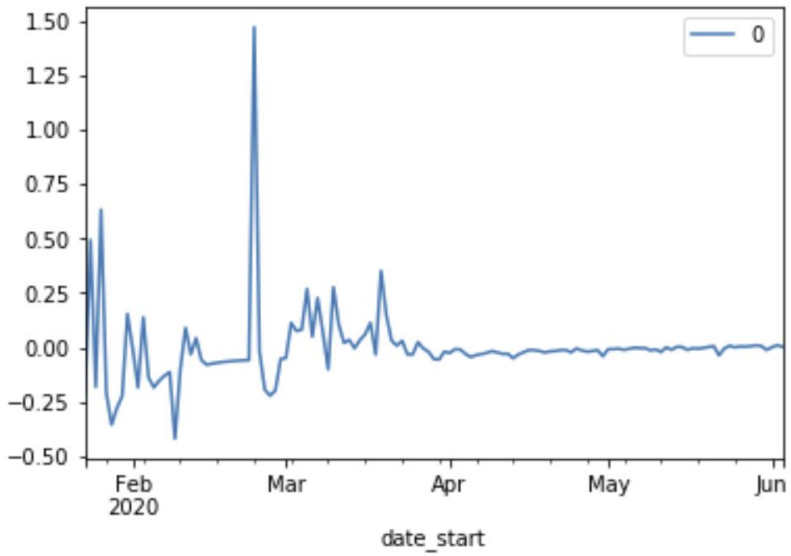
Roots

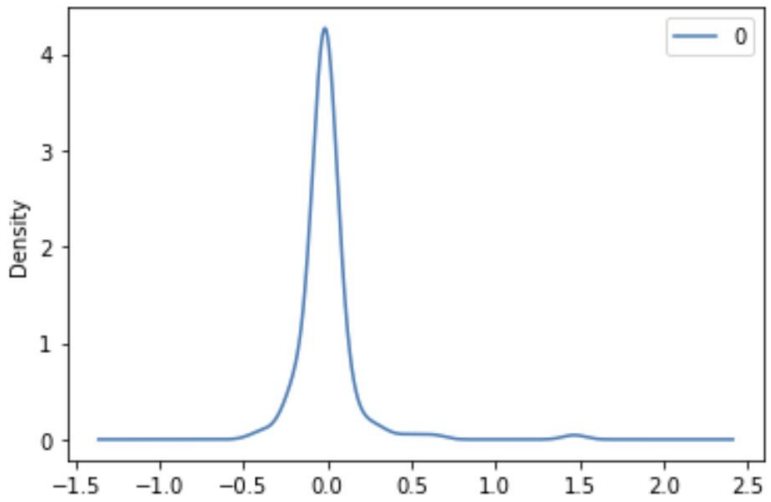
	Real	Imaginary	Modulus	Frequency
AR.1	1.0851	+0.0000j	1.0851	0.0000
AR.2	-4.5828	+0.0000j	4.5828	0.5000
MA.1	1.3474	+0.0000j	1.3474	0.0000

2. Regression



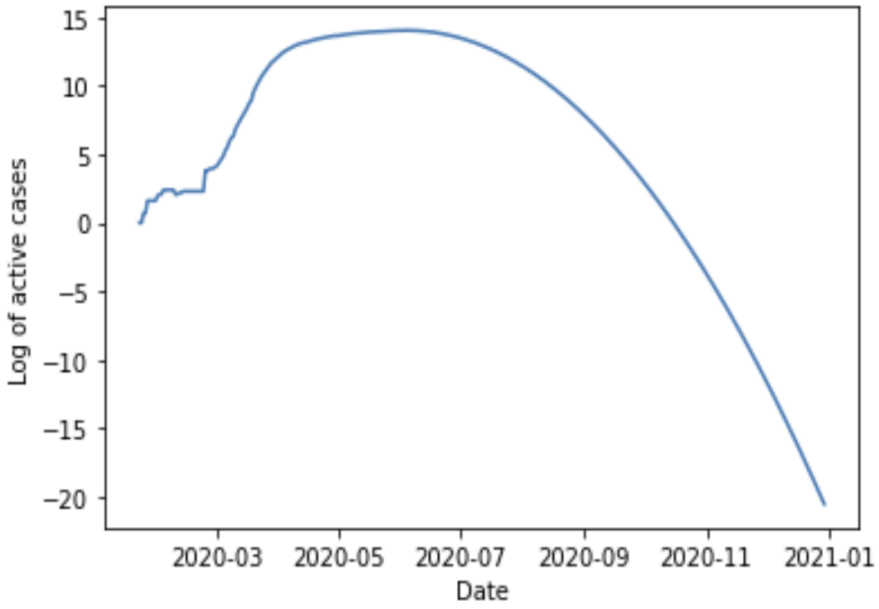
3. Fitting the model

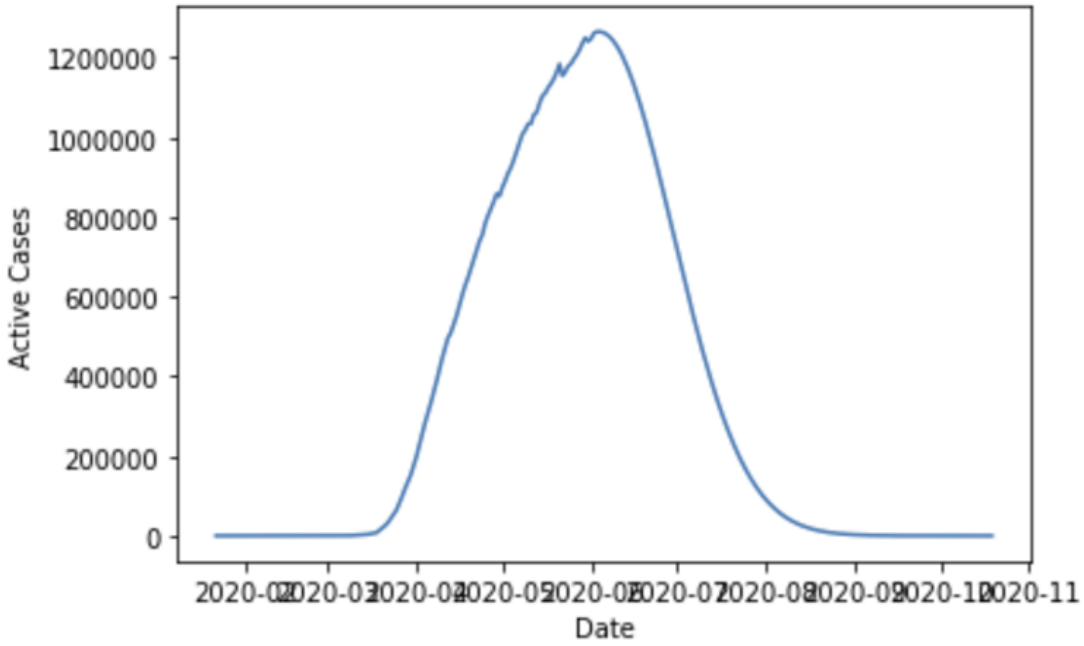




3. Fitting the model

Residuals—White noise





4. Forecasting

RESULTS

Thanks to Foursquare, I found our ideal hotel for the vacation. ***Loews Royal Pacific Resort*** seems to be the most suitable hotel for my friend and I to stay in. And around the hotel, there are 15 restaurants that serve cuisines from all over the world.

If nothing unexpected happens and everything goes smoothly as right now, from the ARIMA model I built, the active cases of the U.S. should come to zero in late October.

CONCLUSION

By using data collected in Foursquare, now I have some ideas about my future trip to Orlando, such as which hotel to stay in so that we can be extremely close to the Universal Studios and enjoy excellent services, and which restaurants to go to that has cuisines from all over the world. I believe this report can also be a good small guidance to people who have the same interests and plans as I do.

With the help of several Python libraries, I have built an ARIMA model to predict when the active cases in the United States can be zero and when it may be safe to travel around the country again. If everything goes as I expected, according to the model, the situation should be alleviated by the end of October.


THANK YOU!!!