Price and Popularity Analysis of Airbnbs in NYC

CUSP-GX 6001 Applied Data Science

Student Names: Anbo Guo, Letao Hou, Catherine Liu, Yuetong Zhou

Due Date: May 4, 2021

Word Count: 2328

Table of Contents

Abstract

Some features of Airbnb data, including but not limited to its high dimensionality and updating frequency, have attracted many people to study them. While other studies analyzed many different features, we not only analyzed different features, but also built a model to help new hosts predict their housing popularity. Our main focus would be to help the hosts increase booking rates and profitability. We utilized NYC airbnb daily open data for time series analysis and NYC Airbnb listings data for booking rate analysis in our project. Our main hypothesis is that seasonal trends exist, people prefer private properties over shared ones and location matters. We also build several machine learning models for price and popularity prediction. As a result, hosts can use our analysis results to adjust their pricing strategies to increase booking rates and profitability.

Introduction

Nowadays, there's a rising trend of people paying more attention to the sharing economy. For instance, Airbnb, the famous lodging company which provides vacation rental housing, started trading on Nasdaq in December last year and showed monster gains on its first day of trading. Along with an increasing number of investors entering the market, analyzing bookings becomes essential. In this project, we will use historical data to analyze past bookings. We want to explore subtle modes of Airbnb pricing and key attributes that make some Airbnb properties more popular to build some prediction models, so the company and its hosts may take reference and adjust their operations strategies in time in order to improve their bookings and thus profitability. We expect some seasonal trends and notable impacts brought by covid-19 on pricing. Our hypothesis on key attributes is that guests prefer private properties over shared ones, while locations and prices are the two features that affect their final decision.

Literature Review

Airbnb listings in Barcelona are analyzed according to multiple variables. The hosts' data includes the top 100 super hosts, depicting greater representation, and the listed hosts' profile influences

the ratings. Profile picture, phone, and identity verification contribute to the host's ratings (Gunter, 31). The listings details used to classify the listed rooms are instantly bookable, neighborhood status, and relative density. These details help to characterize the booking demand, based on bed type, room type, and the cancellation policy. According to Szotek, the most popular room types are apartments, guest suites, house and loft rooms, whereby multiple room types may be combined in a booking. The inclusion of bed type increases the booking rates. Furthermore, the neighborhood groups that emerged top of the listing include Iutat Vella, Sants Marti, and Eixample (Szotek). The cancellation policies include flexible and 14-day strict cancellation policies (Yao et al., 4526). Thus, these parameters determine the booking's popularity in Barcelona, which reflects the global popularity trends of the Airbnb property listings.

Operational strategies required to improve bookings and profitability of Airbnb range from fees, amenities rating, and host information. Fees charged include the price per person, cleaning, security deposit, cleaning fees, extra people. Another criterion for assessing popularity is the review scores and the minimum nights (Yao et al., 4526). The review scores depend on the cleanliness, communication, location, total amenities. The lower pricing per person had a higher booking rate for the top 100 reviewed cases. The security deposit, fees, and extra people fees affected the booking rates, whereby the lower the fees, the higher the booking rates. A higher number of amenities, higher review scores, and high location review scores attracted higher booking rates. Therefore, the booking analysis shows that the lower pricing, higher review scores, amenities, and higher locations' reviews attract more bookings.

Baseline (Naïve Bayes) and decision tree modeling may be used to forecast the properties that will be booked in the future. The parameters used in the model include pricing, number of rooms, location, hosts' ratings, and the neighborhoods. Szotek noted that the decision tree model gives better results, whereby the hoisting and listing duration indicate that host experience yields higher booking rates. The top listings had a strict cancellation policy, at least one bathroom, lower price per person, lower cleaning fees, higher security in the neighborhood, more amenities, and lower extra people fees (Szotek).

Hence, the model may be used to predict future booking trends, based on the aforementioned parameters, whereby pricing and host rating evinces to be the main determinants.

Analysis of the popularity of listings at different times of the year helps to predict the demand in different regions and seasons. Holidays attracted higher booking rates and higher prices, whereby popular holidays, such as Christmas and New Year attracted the highest booking rate and prices (Chawla). Moreover, the summer seasons depicted a higher booking rate than winter in the Seattle region. Furthermore, Chawla) notes that the average prices change with the season, whereby the demand/availability ratio determines the pricing in different regions. Hence, the dataset analyzed depicts that the main variables that determine the popularity and pricing in a given region are the seasons, events, and holidays, whereby the highly rated hosts with multiple amenities attract the highest number of bookings. Hence, predicting future booking rates and profitability relies on the key parameters of a room, pricing, host's reviews, time of the year, and the location.

Data

For analysing prices, we used NYC airbnb daily open data from October 2017 to October 2018 obtained from Airbnb Open Data for price analysis. Another dataset we explored is the NYC Airbnb listings data obtained from Inside Airbnb, which contains detailed data on Airbnb property listings.

We gathered two Airbnb calendar price datasets of 2017-2018 & 2021-2022, hoping to have deeper insight of price changes with and without the impact of COVID-19.

The NYC Airbnb listings data was used to analyze key features that contribute to the popularity of properties and to build models for predicting future bookings of properties. To define popularity in our case, we took the following steps:

1. host_since*availibility_365 gives us total available days for booking of each property since host

2. total_reviews/total available days gives us the approximate booking rate of each property
3. We calculate the median of booking rates, properties with booking rate higher than the median are defined as “popular”, with label =1, and vice versa.

For data cleansing, we did:

- pd.to_datetime() on date variables
- fillna(0) on numeric variables
- dropna() on categorical variables
- Convert categorical variables to numeric by making them dummy variable via get_dummy
- Replace boolean variables with f = 0, t = 1
- Normalization

In the end, all data of listings.csv within the dataframe are normalized numeric data, with shape (10163, 408).

Methods and Results

With the cleaned listing dataset, we performed two groups of correlation analyses: features and booking rates; features and prices. We obtained the top 20 features that have the most positive and negative correlations with prices and booking rates. Since we get dummies for categorical variables in the preprocessing step, we manually aggregate those top and bottom features into the previous categorical features before doing dummy variables. For booking rates, special requirements like minimum and maximum nights, luxury residential neighborhoods and shared amenities are features that have negative impacts on bookings, while places near visitors-crowded areas, host is superhost or hotel-room-type rooms will attract more guests(Figure 1&2). This is in accordance with our original hypotheses that guests prefer private properties over shared ones, and property location is another important factor of them. According to correlation between features and prices, sharing of private rooms, beds or bathrooms have

the most negative effects on prices, followed by property locations in Bushwick, Astoria and Elmhurst (Figure 3). In contrast, more bedrooms, beds and baths will greatly increase the prices, and properties in good locations like Soho and the Upper west side are also more expensive than others(Figure 4). We also compared the same categorical variables if they appear in both of the correlation tables and get insights from the results. For instance, we found that the longer the host response time, the lower the booking rates of their properties. We also calculated the average price for each neighborhood as reference for hosts of Airbnb.

For further analysing prices, we performed time series analysis. After data cleaning, we have 19 million listings and a time series-based pivot table (Figure 5). The average price of Airbnb listings decreased sharply during the late 2017 and had a burst at the end of the year due to few listings made of luxury renting and holidays' impact. After that, the price increased step by step as more and more rentals were listed. The price after COVID-19 dropped as expected and gained back slowly as the pandemic got less severe. One thing worth paying attention is that all the prices fluctuate seasonally (Figure 6). After slicing the data into 2017-2018 & 2021-2022 two datasets, we did signal periodogram test and plot ACF test for datasets from 2017 to 2018 (Figure 7). The two tests suggested differently, which time series of price may have seasonality of three months and one week. After consideration of seasonal decomposition and comparing with the original time series, we determined one week as the frequency of seasonality (Figure 8). The result of simple regression indicates that neither of the two regressions is ideal since values of R-squared are both quite low and the regression with dummy variables only improves the result a bit (Figure 9). From the impact graph, Fridays and Saturdays have more influences on prices than other days of week, suggesting that people tend to book Airbnb on Friday and Saturday for weekend trips (Figure 10). After visualizing the time series of data with prediction, we found a similar trend between price of 2021-2022 and prediction. We plotted the graph again after abstracting the impact of COVID-19 (27.66), the trend simulation fit better than before with in-sample R-squared of 0.17 but out of sample R-squared of 0.5 (Figure 11). Next, we tried to fit the ARIMA model instead of simple OLS regression for the time series. After conducting ADF test and first order differencing, the time series reached

stationary. According to PACF and ACF plots, we fit an ARIMA (1,1,3) model and predicted the trend of price with 95% confidence interval. From the result, a few AR and MA terms are not significant according to the 95% confidence interval (Figure 12). To further improve our model, we fit an SARIMA (1,1,3) model with 7 days' seasonality and predicted the price. The prediction shows that the price will seasonally increase after 2021 June, reaching $155 on average in the end of the year, which comes back to the original price level as of the year 2018 before COVID-19 (Figure 13).

We performed PCA on the normalized listing dataset for further analysing booking rates and popularity. From the explained variance graph, we chose n_components = 290 as our leading principal components out of 406 features in total to obtain 80% explained variance (Figure 14). For the baseline model, we established logistic regression to predict popularity, with 0.759 accuracy. Then we added kernel PCA with logistic regression to improve the performance of our model, reaching 0.76 accuracy.In addition, considering tree models may give better results, we built a random forest model, with 0.743 accuracy. Then we implemented GridSearchCV in Sklearn to find the best parameters in random forest, finally achieving 0.775 in the out-of-sample test. We also plotted relative feature importance with Random Forest(Figure 15). It seems that review_score_rating and price are the top 2 factors which affect popularity of properties. Finally, we established the XGBoost model. Xgboost is a decision-tree-based ensemble machine learning algorithm that uses a gradient boosting framework. It has many advantages than ordinary tree models, such as parallelization, tree pruning, better handling missing values and so on. We first trained a XGBoost Classifier with 0.766 accuracy. In order to build a more robust one, We tuned hyperparameters and implemented cross validation and finally got 0.7794 accuracy, which was 2.68% higher than the baseline model. We also obtained feature importance(Figure 16).

Discussion

Due to the lack of pricing data from 2019 to 2020, the time series model we built is not that accurate since the trend study of price right before 2020 is missing, which causes the underestimation of the COVID-19 impacts. For the listings data, we omitted all columns with pure descriptive texts. Most of

those columns are not helpful for our analysis, while the column that describes amenities included is believed to be useful and has high correlation with booking rates. Besides, we have to get dummies for all categorical variables, and the number of columns increases from 20 to 408, which brings down the scale correlation coefficients of each feature, and hence causes an underestimation of correlations between each original feature before dummy. Besides, we should have used the borough columns in our raw listings dataset for visualizing booking rates and prices spatially, but we had to give up this step due to the large number of nan values in the columns. We used a smaller neighborhood column instead, getting limited characteristics compared to what we originally expected to get from boroughs.

Conclusion

From our model, we can find that some features are of vital importance to increase booking rate, such as responding time, location and min/max nights. Therefore, we suggest hosts respond in time, lower the price if location is not good, or loosen restrictions on min/max nights so as to enhance booking rate. We also developed some suggestions for the hosts to improve their profitability directly. Firstly, they should get to know the seasonal trend on pricing from the time series model, and rise up the prices accordingly. In addition, since location matters, properties in good locations with relatively high booking rates can raise prices as well. For new hosts, they can check out the average listing prices of other properties in the same neighborhood as a reference to set up their own prices.

Team Roles

Abstract: Letao Hou, Catherine Liu, Yuetong Zhou

Introduction: Letao Hou

Literature review: Catherine Liu

Data: Anbo Guo, Letao Hou, Catherine Liu, Yuetong Zhou

Methods and Results: Anbo Guo, Letao Hou, Yuetong Zhou

Conclusion: Catherine Liu, Yuetong Zhou

References

Chawla, Ravish. "Analyzing the Airbnb Dataset for Trends Using Data Visualizations And Modeling." *Medium*. N.p., 2019. Web. 30 Apr. 2021. Available: https://medium.com/ml2vec/data-analysis-on-the-airbnb-dataset-e0be9254eeb9

Gunter, Ulrich. "What makes an Airbnb host a superhost? Empirical Evidence from San Francisco and The Bay Area." *Tourism Management* Vol. 66 No. 1 (2018): 26-37. Available: https://www.modul.ac.at/article/view/what-makes-an-airbnb-host-a-superhost-empirical-evidence-from-san-francisco-and-the-bay-area/

Szotek, Maksym. "Understanding Data - Airbnb Listing Popularity Analysis Based On Barcelona Data." *Rstudio-pubs-static.s3.amazonaws.com*. N.p., 2018. Web. 30 Apr. 2021. Available: https://rstudio-pubs-static.s3.amazonaws.com/407929_afc5ef0f2ad648389447a6ca3f4a7cd4.html

Yao, Bin, et al. "Standing Out From the Crowd–An Exploration Of Signal Attributes Of Airbnb Listings." *International Journal of Contemporary Hospitality Management*. Vol. 1 No 1. (2019). 4520-4542 Available: https://www.emerald.com/insight/content/doi/10.1108/IJCHM-02-2019-0106/full/pdf

Tables and Figures

	features	bora_cor
14	calculated_host_listings_count	-0.037390
6	minimum_nights	-0.030905
288	host_neighbourhood_Upper East Side	-0.026678
406	host_response_time_within a few hours	-0.022905
7	maximum_nights	-0.017764
380	bathrooms_text_1 shared bath	-0.016634
381	bathrooms_text_1.5 baths	-0.016296
19	host_neighbourhood_Astoria	-0.012707
91	host_neighbourhood_East New York	-0.011998
374	room_type_Shared room	-0.011497
89	host_neighbourhood_East Flatbush	-0.010674
388	bathrooms_text_3 baths	-0.010673
167	host_neighbourhood_Long Island City	-0.010346
283	host_neighbourhood_Tribeca	-0.010084
404	host_response_time_a few days or more	-0.010074
47	host_neighbourhood_Cambridge	-0.009112
386	bathrooms_text_2.5 baths	-0.008568
98	host_neighbourhood_Elmhurst	-0.008274
377	bathrooms_text_0 shared baths	-0.008214
389	bathrooms_text_3 shared baths	-0.008181

Figure 1. Top 20 features with negative correlation to booking rates

	features	bora_cor
0	booking_rate	1.000000
1	popularity	0.095921
15	host_neighbourhood_Allerton	0.068734
308	host_neighbourhood_Windsor Terrace	0.039493
106	host_neighbourhood_Flatiron District	0.038395
9	host_is_superhost	0.038181
372	room_type_Hotel room	0.034426
46	host_neighbourhood_Bushwick	0.028523
407	host_response_time_within an hour	0.027780
28	host_neighbourhood_Bedford-Stuyvesant	0.024469
253	host_neighbourhood_South Beach	0.023754
186	host_neighbourhood_Midtown East	0.021085
162	host_neighbourhood_Lefferts Garden	0.020827
353	property_type_Private room in villa	0.019319
127	host_neighbourhood_Greenwich Village	0.018718
13	instant_bookable	0.018376
378	bathrooms_text_1 bath	0.017814
358	property_type_Room in serviced apartment	0.016830
385	bathrooms_text_2 shared baths	0.016102
355	property_type_Room in boutique hotel	0.016041

Figure 2. Top 20 features with positive correlation to booking rates

	features	pri_cor
373	room_type_Private room	-0.277343
380	bathrooms_text_1 shared bath	-0.232763
337	property_type_Private room in apartment	-0.203694
346	property_type_Private room in house	-0.120518
352	property_type_Private room in townhouse	-0.079409
385	bathrooms_text_2 shared baths	-0.078217
382	bathrooms_text_1.5 shared baths	-0.065834
374	room_type_Shared room	-0.056850
14	calculated_host_listings_count	-0.056320
407	host_response_time_within an hour	-0.051955
13	instant_bookable	-0.051073
359	property_type_Shared room in apartment	-0.050788
46	host_neighbourhood_Bushwick	-0.048563
1	popularity	-0.045752
19	host_neighbourhood_Astoria	-0.041697
98	host_neighbourhood_Elmhurst	-0.039346
89	host_neighbourhood_East Flatbush	-0.036654
108	host_neighbourhood_Flushing	-0.035766
91	host_neighbourhood_East New York	-0.033321
28	host_neighbourhood_Bedford-Stuyvesant	-0.031916

Figure 3. Top 20 features with negative correlation to price

	features	pri_cor
5	price	1.000000
2	accommodates	0.486090
3	bedrooms	0.421059
4	beds	0.357533
371	room_type_Entire home/apt	0.285180
384	bathrooms_text_2 baths	0.248369
395	bathrooms_text_5 baths	0.239291
390	bathrooms_text_3.5 baths	0.214501
386	bathrooms_text_2.5 baths	0.208358
332	property_type_Entire townhouse	0.189382
394	bathrooms_text_4.5 baths	0.155531
323	property_type_Entire condominium	0.139420
388	bathrooms_text_3 baths	0.139367
328	property_type_Entire house	0.130608
329	property_type_Entire loft	0.128376
251	host_neighbourhood_Soho	0.102206
348	property_type_Private room in resort	0.094399
289	host_neighbourhood_Upper West Side	0.093181
397	bathrooms_text_6 baths	0.090711
383	bathrooms_text_15.5 baths	0.085427

Figure 4. Top 20 features with positive correlation to price

	listing_id	date	available	price
0	2515	2018-10-01	t	$99.00
1	2515	2018-09-30	t	$89.00
2	2515	2018-09-29	t	$99.00
3	2515	2018-09-28	t	$99.00
4	2515	2018-09-27	t	$99.00
...	...	...	...	...
13464016	47939451	2022-01-31	f	$61.00
13464017	47939451	2022-02-01	f	$61.00
13464018	47939451	2022-02-02	f	$61.00
13464019	47939451	2022-02-03	f	$61.00
13464020	47939451	2022-02-04	f	$62.00

19044658 rows × 4 columns

date	2017-10-02	2017-10-03	2017-10-04	2017-10-05	2017-10-06	2017-10-07	2017-10-08	2017-10-09	2017-10-10	2017-10-11	...
listing_id											
2515	NaN	NaN	NaN	NaN	NaN	NaN	NaN	99.0	99.0	NaN	...
2539	150.0	150.0	150.0	150.0	99.0	99.0	150.0	150.0	150.0	150.0	...
2595	198.0	198.0	198.0	198.0	198.0	NaN	NaN	198.0	198.0	198.0	...
3330	NaN	NaN	NaN	NaN	NaN	NaN	70.0	70.0	70.0	70.0	...
3647	150.0	150.0	150.0	150.0	150.0	150.0	150.0	150.0	150.0	150.0	...
...	...	...	...	...	...	...	...	...	...	...	...
48033101	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
48033611	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
48038944	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
48039640	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...
48039776	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...

Figure 5. Daily data of Airbnb listings

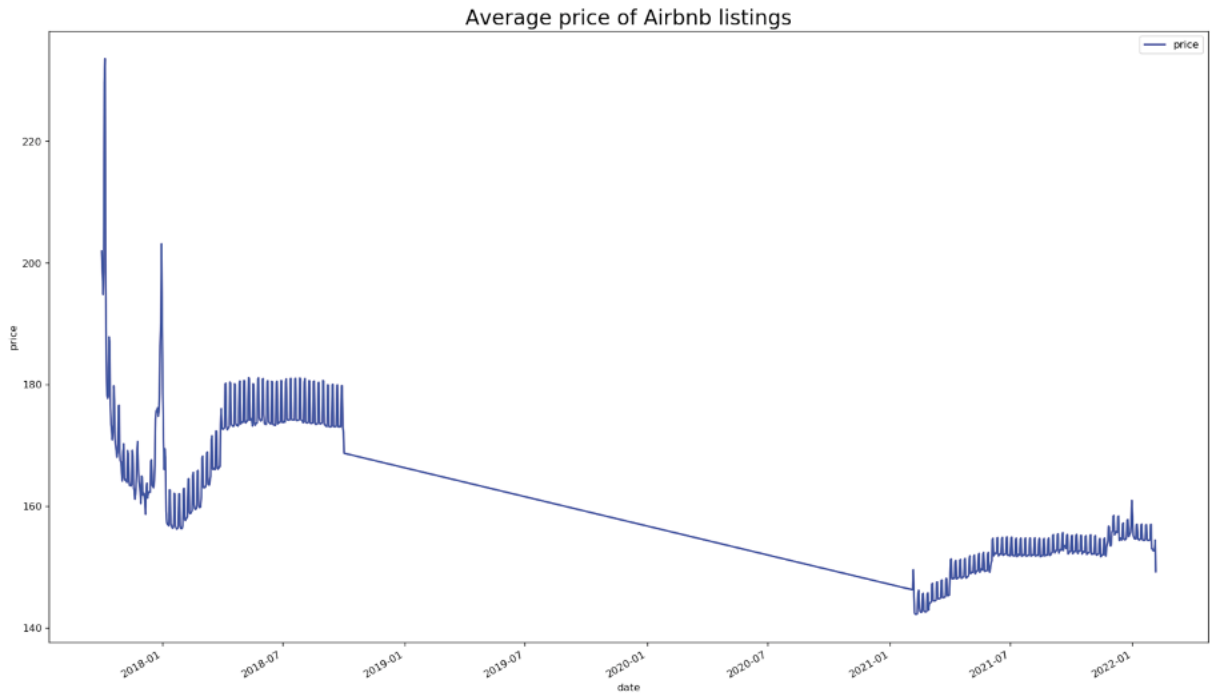
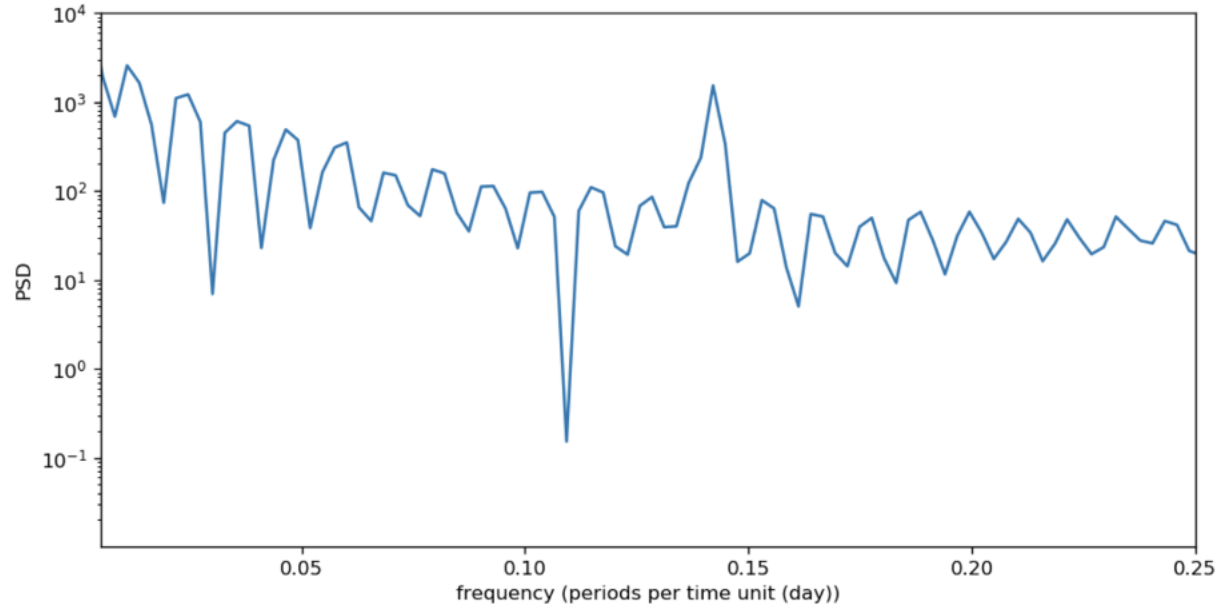


Figure 6. The average price of Airbnb listings




```
Strongest period length = 91.5
```

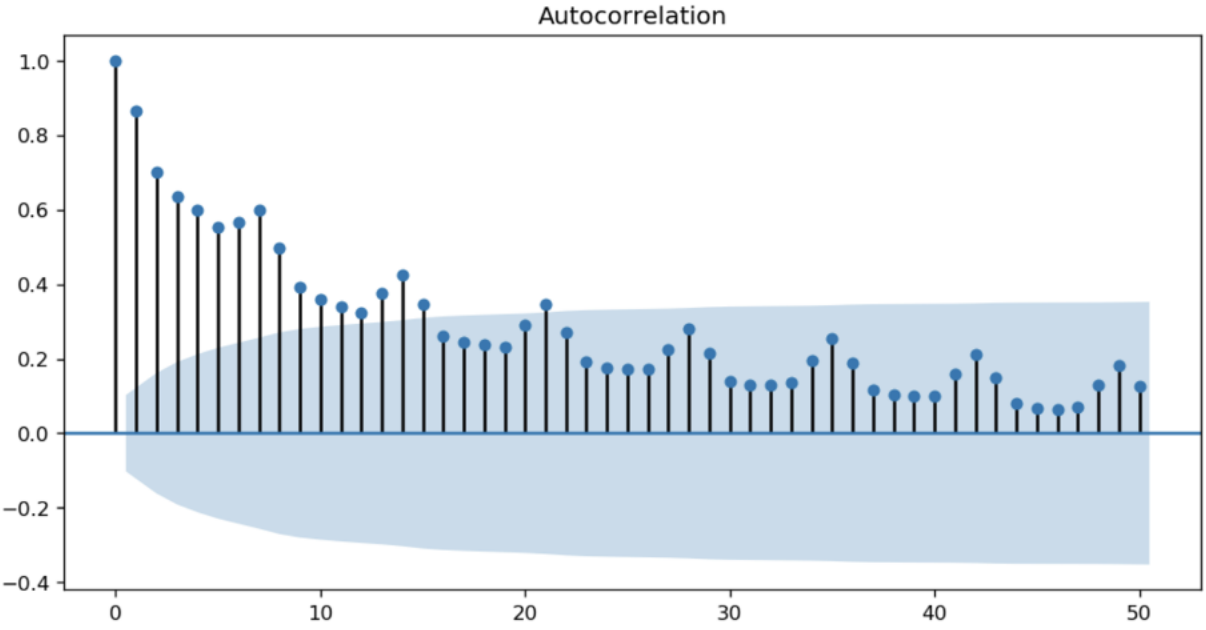


Figure 7. Signal periodogram and ACF graph

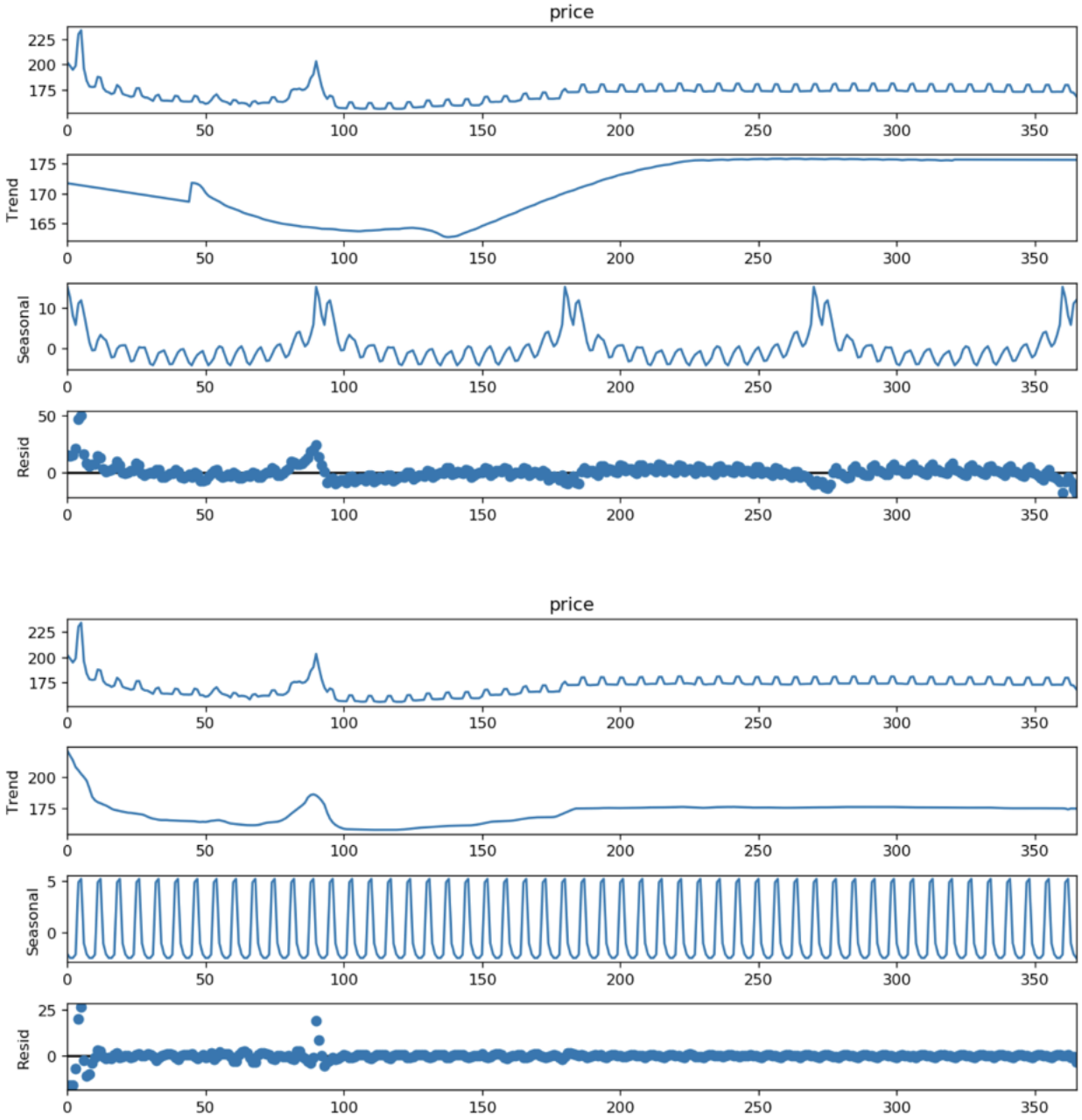


Figure 8. Seasonal decomposition of three months and one week

OLS Regression Results

Dep. Variable:	price	R-squared:	0.058
Model:	OLS	Adj. R-squared:	0.056
Method:	Least Squares	F-statistic:	22.60
Date:	Fri, 30 Apr 2021	Prob (F-statistic):	2.88e-06
Time:	20:02:00	Log-Likelihood:	-1320.8
No. Observations:	366	AIC:	2646.
Df Residuals:	364	BIC:	2653.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	167.7024	0.934	179.467	0.000	165.865	169.540
x1	0.0211	0.004	4.754	0.000	0.012	0.030

Omnibus:	244.838	Durbin-Watson:	0.254
Prob(Omnibus):	0.000	Jarque-Bera (JB):	3667.435
Skew:	2.579	Prob(JB):	0.00
Kurtosis:	17.625	Cond. No.	421.

OLS Regression Results

Dep. Variable:	price	R-squared:	0.168
Model:	OLS	Adj. R-squared:	0.152
Method:	Least Squares	F-statistic:	10.32
Date:	Fri, 30 Apr 2021	Prob (F-statistic):	8.30e-12
Time:	20:01:57	Log-Likelihood:	-1298.1
No. Observations:	366	AIC:	2612.
Df Residuals:	358	BIC:	2643.
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
date	0.0210	0.004	4.988	0.000	0.013	0.029
0	166.1943	1.394	119.182	0.000	163.452	168.937
1	165.6097	1.397	118.567	0.000	162.863	168.357
2	165.3892	1.400	118.103	0.000	162.635	168.143
3	165.5282	1.403	118.010	0.000	162.770	168.287
4	172.4140	1.405	122.719	0.000	169.651	175.177
5	172.6456	1.407	122.684	0.000	169.878	175.413
6	166.3480	1.410	118.015	0.000	163.576	169.120

Omnibus:	253.031	Durbin-Watson:	0.111
Prob(Omnibus):	0.000	Jarque-Bera (JB):	3888.169
Skew:	2.694	Prob(JB):	0.00
Kurtosis:	18.031	Cond. No.	1.11e+03

Figure 9. OLS Regression results

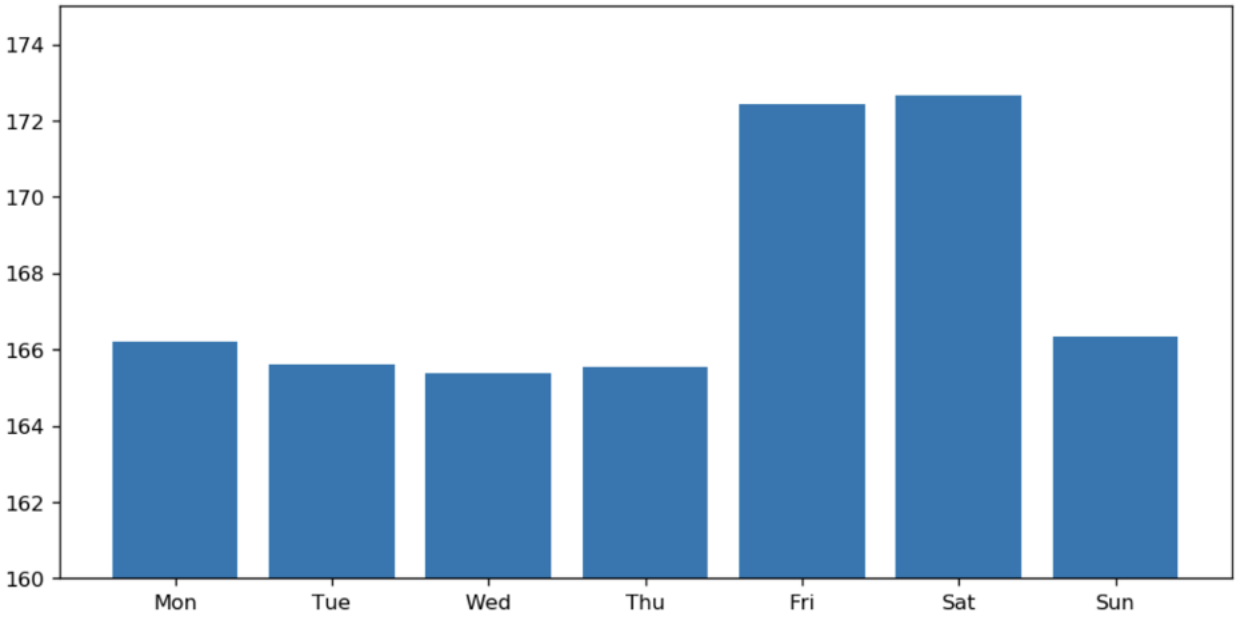
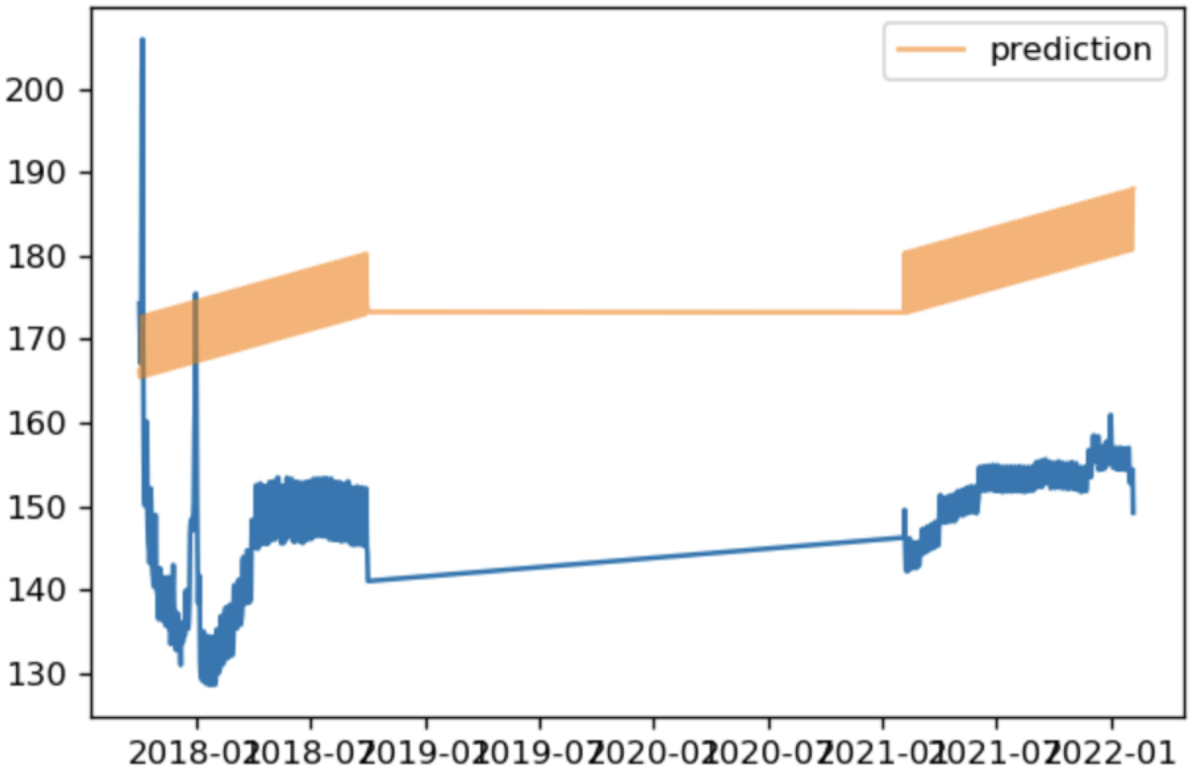


Figure 10. Impact graph of day of week



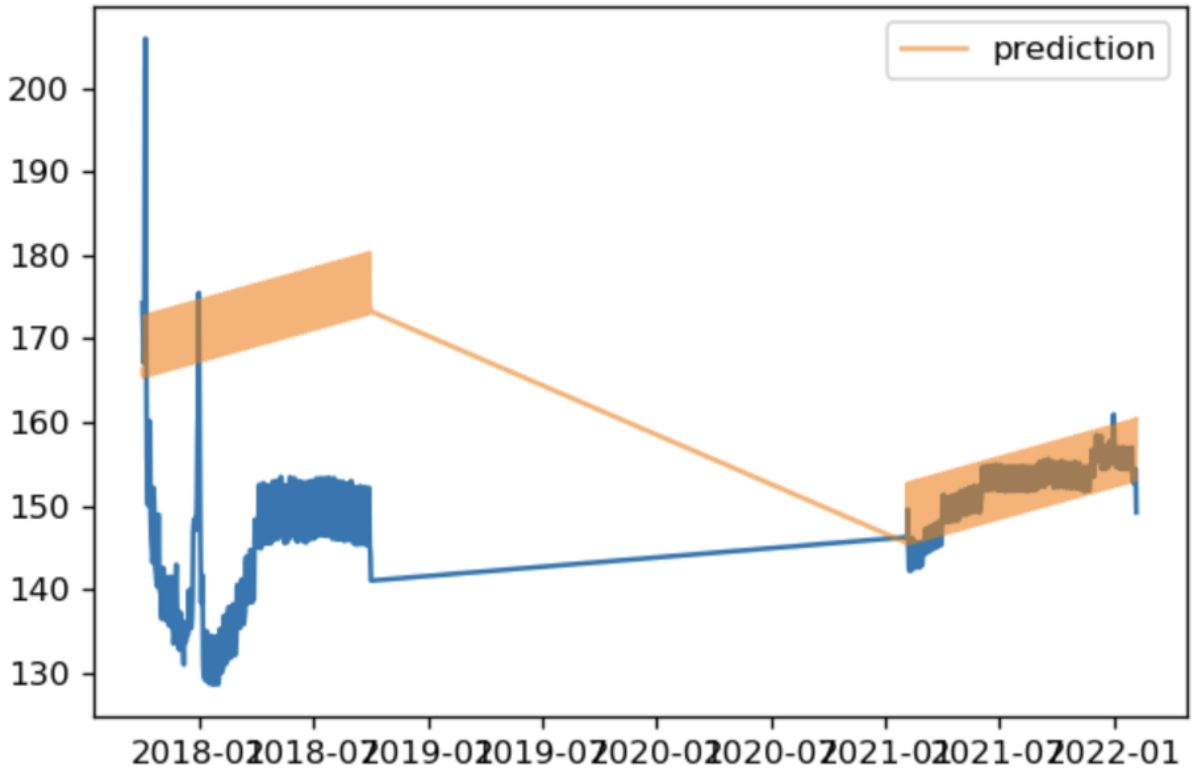
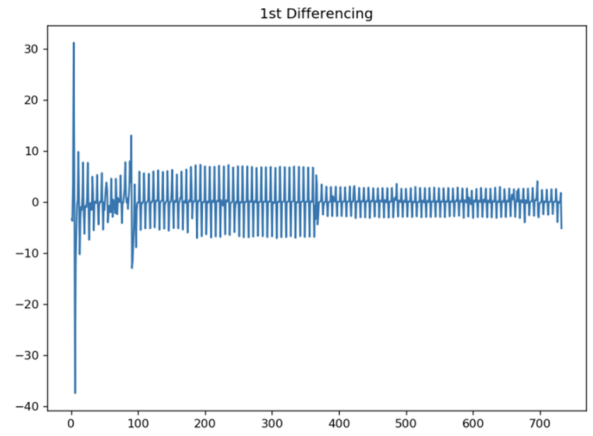
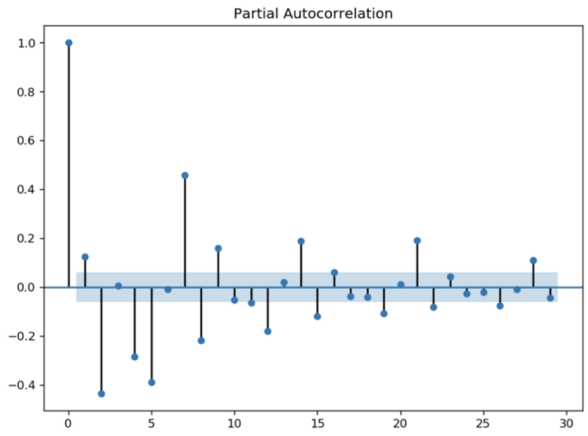


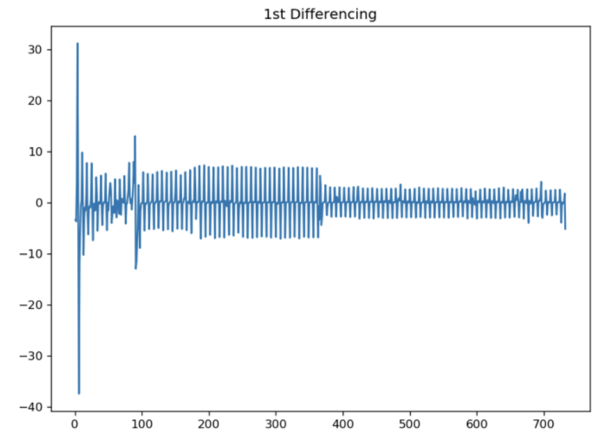
Figure 11. Prediction with and without of impact of COVID-19

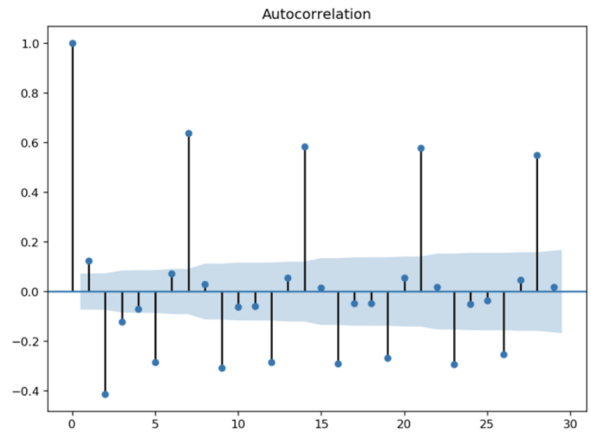
```
ADF Statistic: -2.704897
p-value: 0.073195
Critical Values:
        1%: -3.440
        5%: -2.866
        10%: -2.569
```

```
1st order differencing
ADF Statistic: -6.551318
p-value: 0.000000
2nd order differencing
ADF Statistic: -11.994548
p-value: 0.000000
```









ARIMA Model Results

Dep. Variable:	D.price	No. Observations:	513
Model:	ARIMA(1, 1, 3)	Log Likelihood	-1348.560
Method:	css-mle	S.D. of innovations	3.350
Date:	Fri, 30 Apr 2021	AIC	2709.120
Time:	20:03:45	BIC	2734.561
Sample:	1	HQIC	2719.092

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.0413	0.071	-0.584	0.559	-0.180	0.097
ar.L1.D.price	0.1722	0.223	0.772	0.440	-0.265	0.609
ma.L1.D.price	-0.0127	0.221	-0.058	0.954	-0.445	0.420
ma.L2.D.price	-0.5996	0.048	-12.597	0.000	-0.693	-0.506
ma.L3.D.price	0.0053	0.136	0.039	0.969	-0.261	0.272

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	5.8078	+0.0000j	5.8078	0.0000
MA.1	-1.2945	+0.0000j	1.2945	0.5000
MA.2	1.2881	+0.0000j	1.2881	0.0000
MA.3	112.4518	+0.0000j	112.4518	0.0000

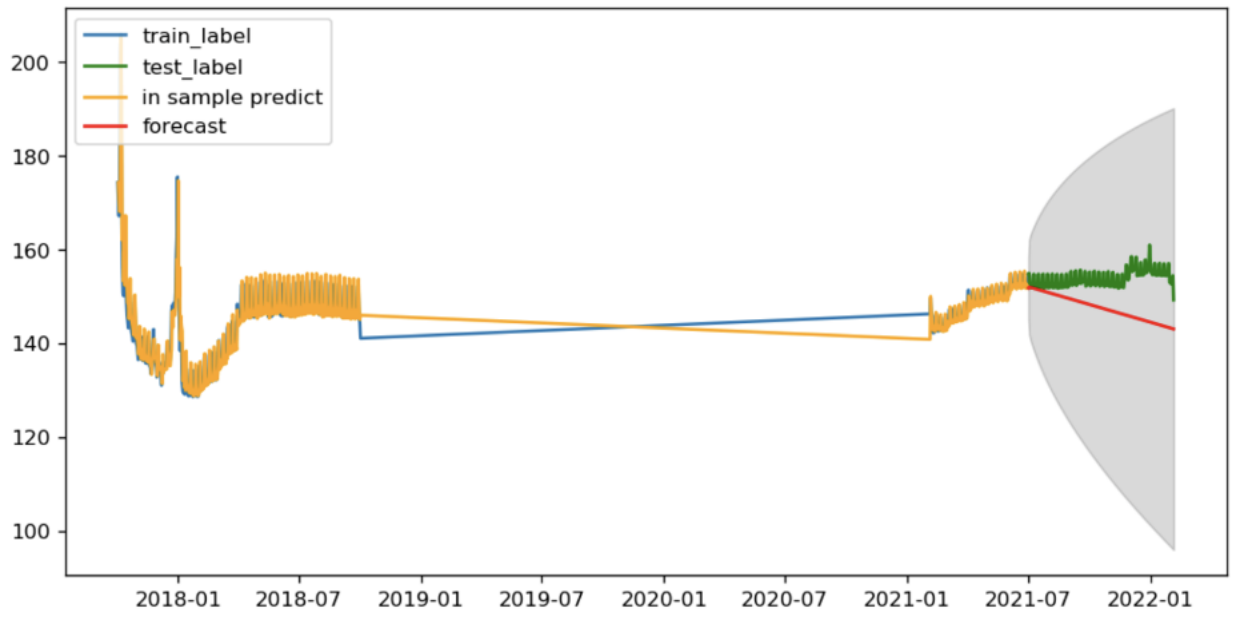


Figure 12. Result of ARIMA (1,1,3) model

SARIMAX Results

Dep. Variable:	price	No. Observations:	514
Model:	SARIMAX(1, 1, 3)x(1, 1, [1], 7)	Log Likelihood	-1132.905
Date:	Fri, 30 Apr 2021	AIC	2279.809
Time:	20:22:34	BIC	2309.395
Sample:	0	HQIC	2291.413
	- 514		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.2279	0.120	1.904	0.057	-0.007	0.463
ma.L1	0.0386	0.115	0.337	0.736	-0.186	0.263
ma.L2	-0.3330	0.036	-9.316	0.000	-0.403	-0.263
ma.L3	-0.1041	0.040	-2.627	0.009	-0.182	-0.026
ar.S.L7	0.0725	0.038	1.907	0.056	-0.002	0.147
ma.S.L7	-0.6241	0.037	-17.021	0.000	-0.696	-0.552
sigma2	5.1225	0.099	51.652	0.000	4.928	5.317

Ljung-Box (Q):	75.77	Jarque-Bera (JB):	20172.63
Prob(Q):	0.00	Prob(JB):	0.00
Heteroskedasticity (H):	0.12	Skew:	-0.38
Prob(H) (two-sided):	0.00	Kurtosis:	33.92

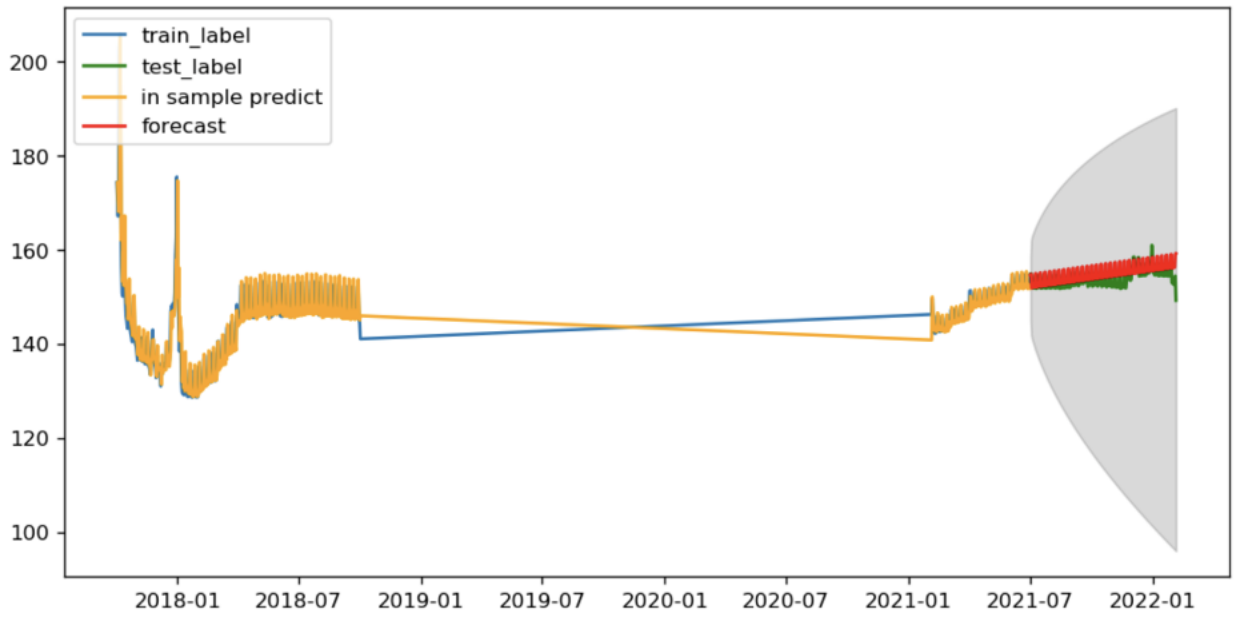


Figure 13. Result of SARIMA (1,1,3,7) model

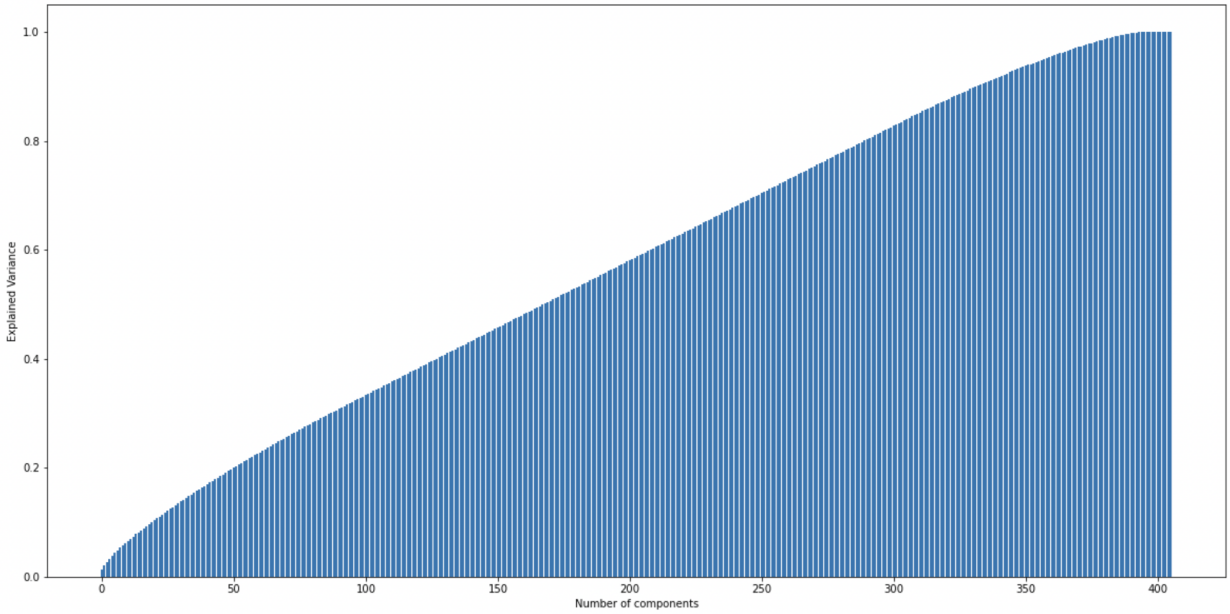


Figure 14. Explained Variance vs. # of leading components

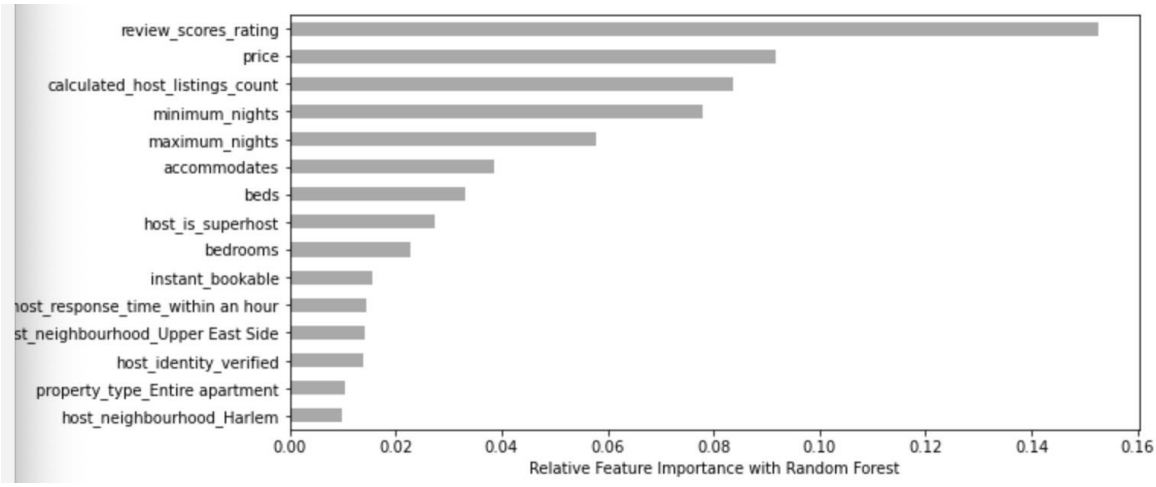


Figure 15. feature importance of Random Forest

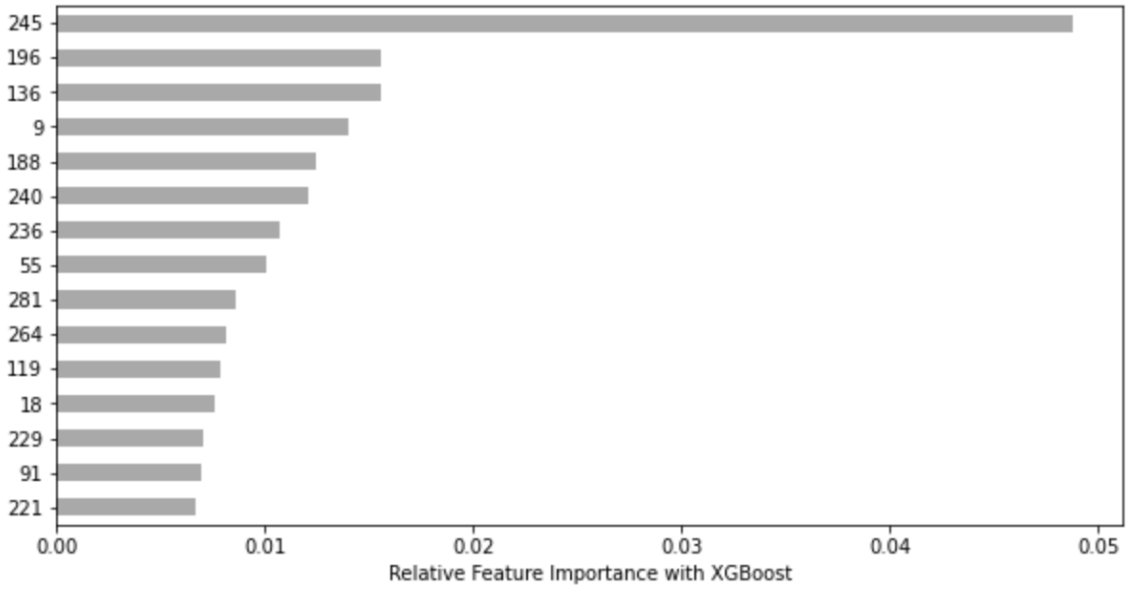


Figure 16. feature importance of XGBoost