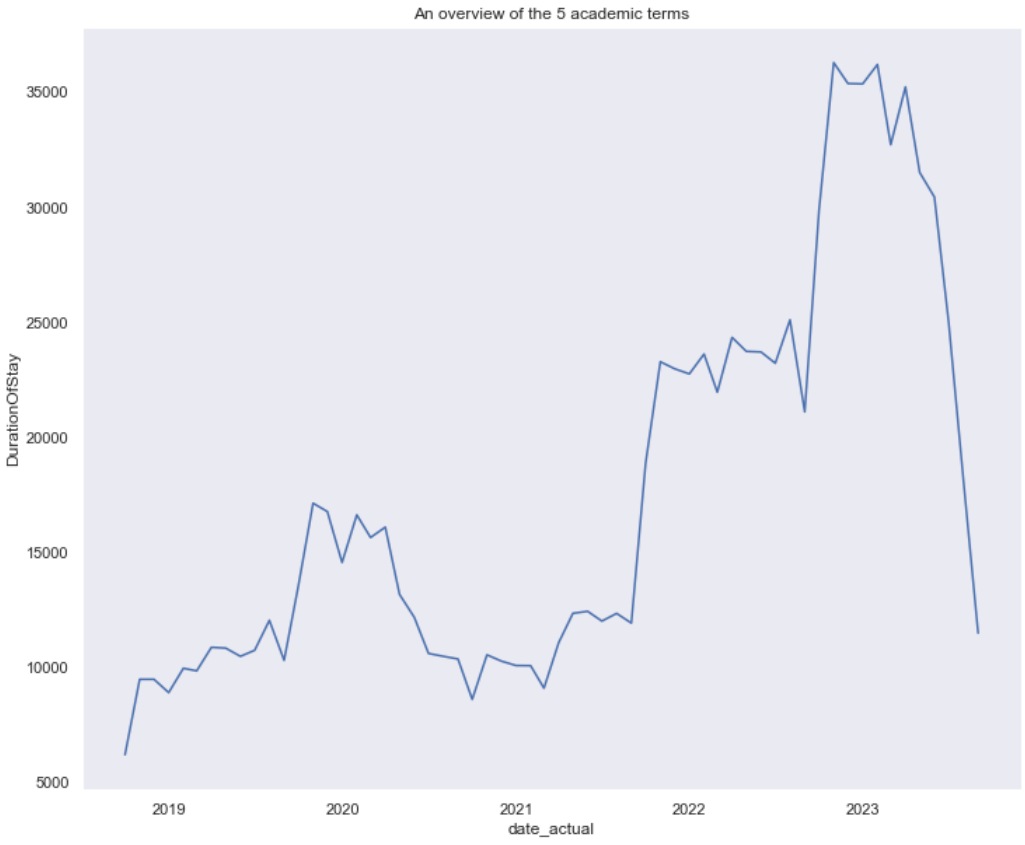
The aim of this study is to forecast occupancy levels for the coming weeks and months for Londonist. Via this we will be able to know the occupancy level for upcoming weeks and we will be able to use these forecasted occupancy levels in our machine learning dynamic pricing algorithm.

In these analyzes; from the 2018-2019 academic year to the 2022-2023 academic year, monthly total occupancy forecast calculations were made according to the data obtained from the Londonist residences/room types. In addition, the existence of the trend and the direction of the trend were determined by examining each academic years separately. In the continuation of the findings obtained from the analyzes, various forecasting were made for the monthly time series whose trend and seasonality were determined. These forecast models which we used are; Facebook Prophet model, Arima model and Sarima model.

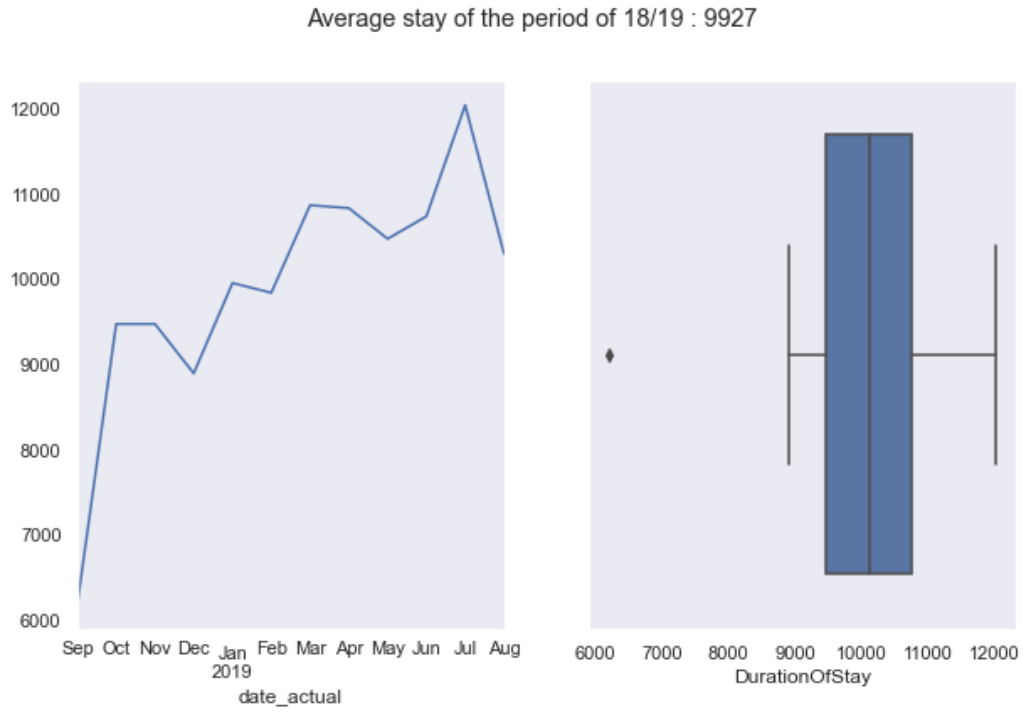
Findings are below;

Graphic: Analysis of 5 Academic Year



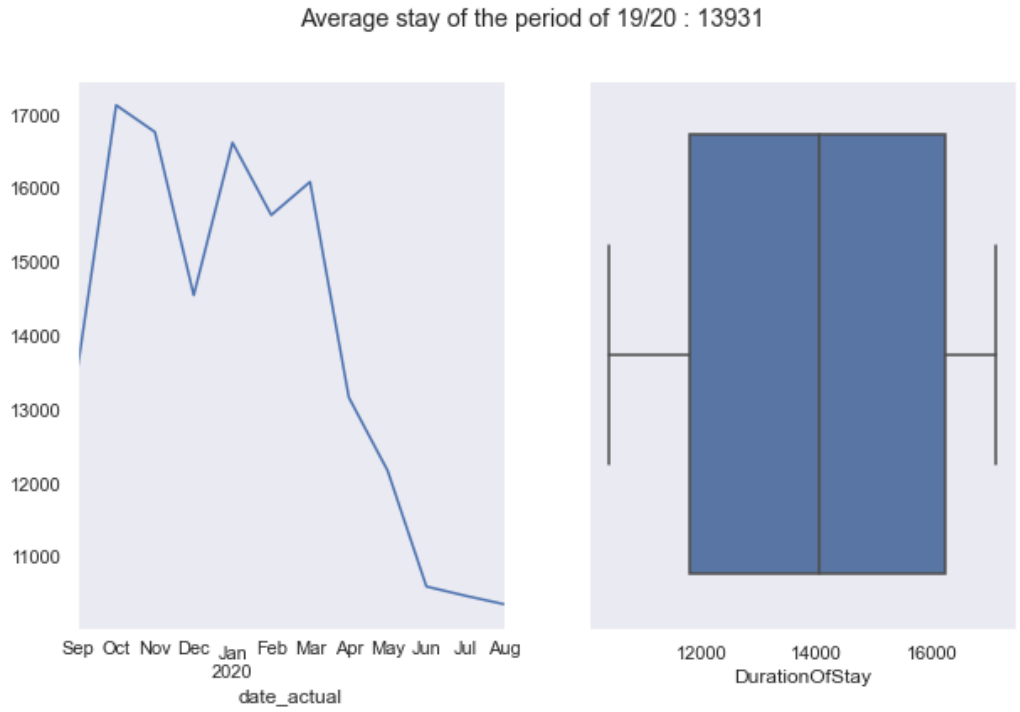
According to the line chart created with the occupancy rates from the 2018-2019 academic year to the 2022-2023 academic year; It can be said that the duration of stay outside the 2020-2021 academic year is also in an increasing trend. The reason for the decline in 2020-2021 is that this academic year coincides with the Covid-19 pandemic period. In addition, according to the graph, each new academic term has increased in terms of occupancy compared to the previous academic term. However, the reason for the decrease in the total occupancy rates in the last months of 2022-2023 is due to the fact that the stays in the 2022-2023 academic year have not yet been completed.

Graphic: Monthly sold number of rooms for the 2018-2019 Academic Year



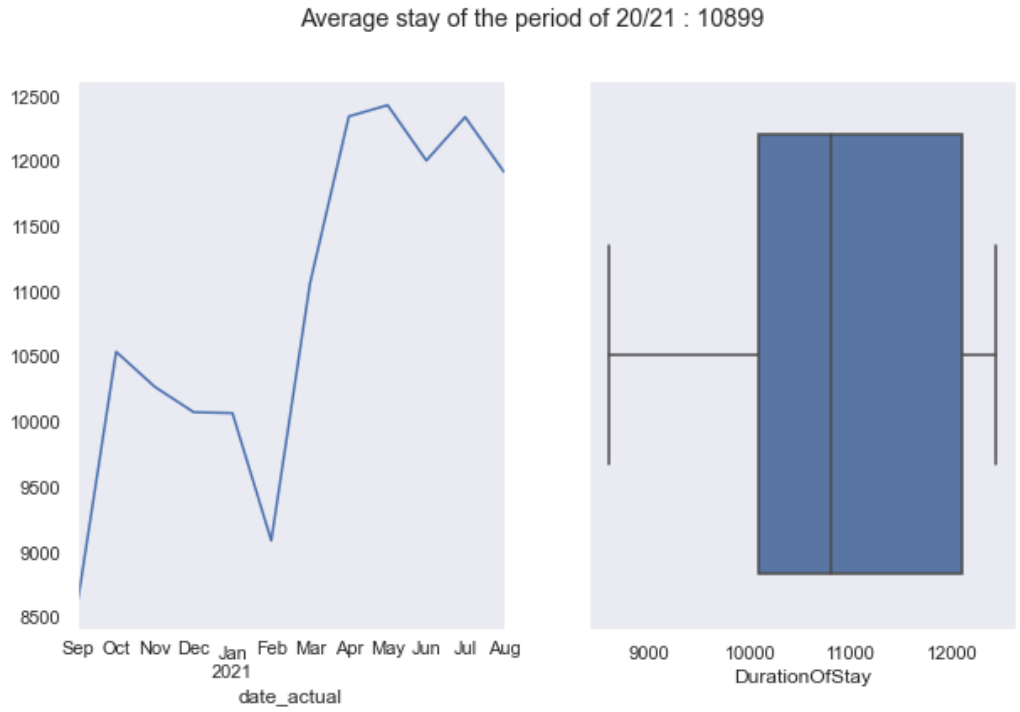
Examining the monthly occupancy rates in the 2018-2019 academic year; The total sold rooms started with 6000-7000 at the beginning of the period, reached 12000 at the end of the academic year. In addition, the average sold room number in the 2018-2019 academic year is 9927.

Graphic: Monthly sold number of rooms for the 2019-2020 Academic Year



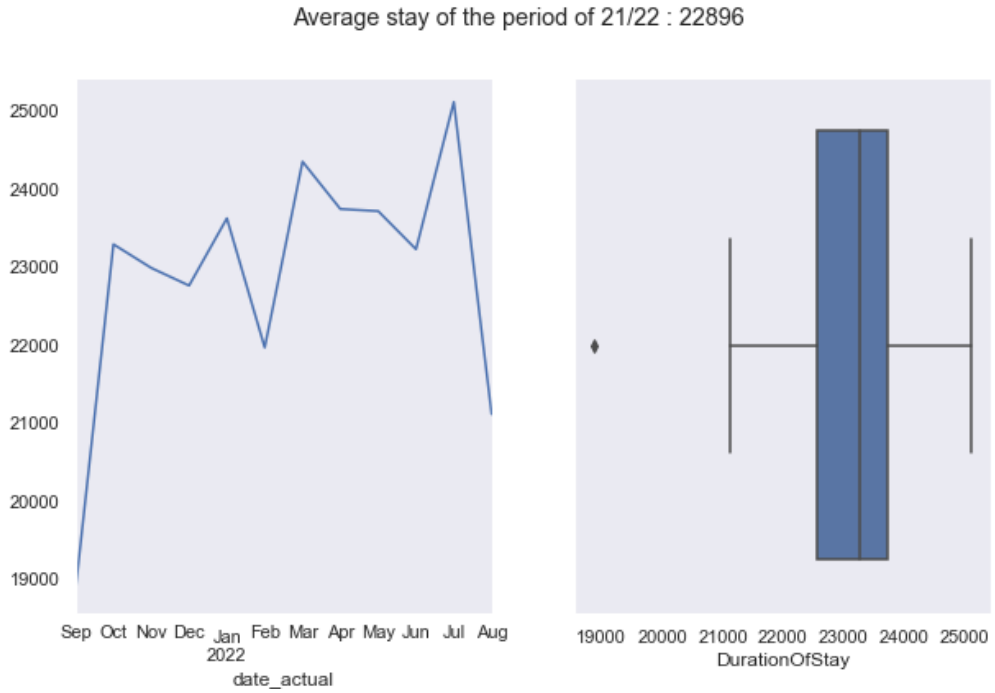
When the monthly number of sold rooms are examined in the 2019-2020 academic year; The monthly total sold numbre of rooms started with 14.000 at the beginning of the academic year, increased to 17.000 during the period, but decreased below 11.000 at the end of the academic year. In addition, the average number of sold rooms in the 2019-2020 academic year is 13,931. As seen in the graph, the total number of sold rooms started to decrease in March 2020 and the monthly total occupancy rate fell below the current occupancy rate at the beginning of the period because the Covid-19 pandemic period started in March 2020.

Graphic: Monthly sold number of rooms for the 2020-2021 Academic Year



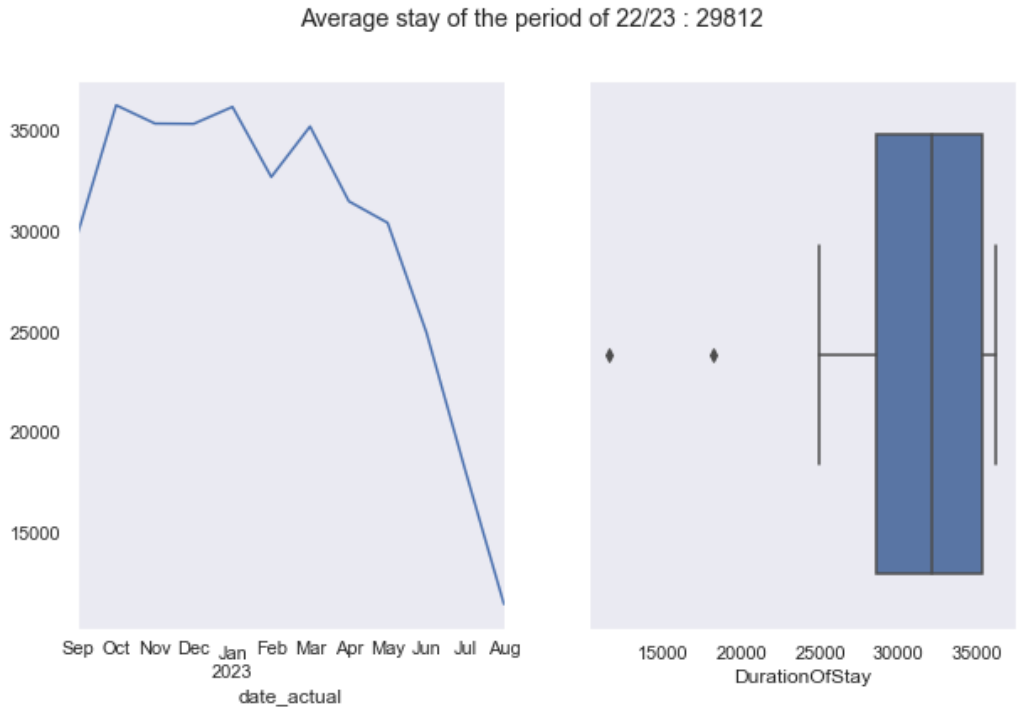
When the monthly number of sold rooms are examined in the 2020-2021 academic year; The monthly total number of rooms started with 8500-9000 at the beginning of the academic year, reached 12000 at the end of the academic yar. In addition, the average numbre of sold in the 2020-2021 academic year is 10,899.

Graphic: Monthly sold number of rooms for the 2021-2022 Academic Year



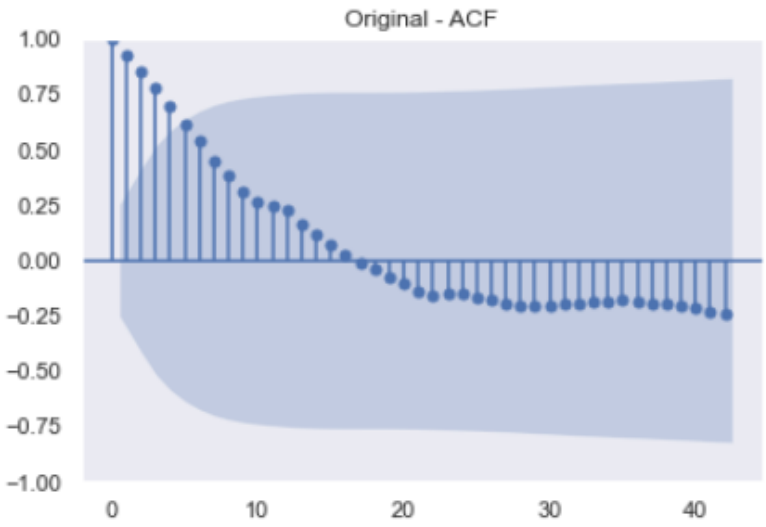
When the monthly number of sold rooms are examined in the 2021-2022 academic year; The monthly total number of sold rooms started between 19.000-20.000 at the beginning of the academic year, reached 25.000 during the academic year and ended with 21.000 at the end of the academic year. In addition, the average sold nımber of rooms in the 2021-2022 academic year is 22,896.

Graphic: Monthly sold number of rooms for the 2022-2023 Academic Year



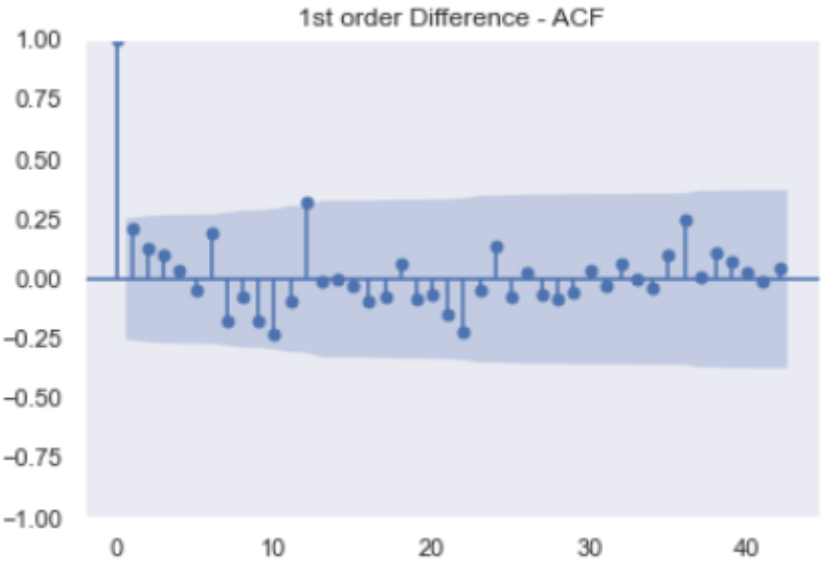
When the monthly number of sold rooms are examined in the 2022-2023 academic year; The monthly total number of sold rooms started in the range of 30,000-35,000 at the beginning of the academic yar, followed this range during the period. However, the decrease in total number of sold rooms in the second half of the period is due to the fact that the sales have not been completed yet. In addition, the average number of sold rooms in the 2021-2022 academic year is 29,812 instantly.

Graphic: Autocorrelation Analysis



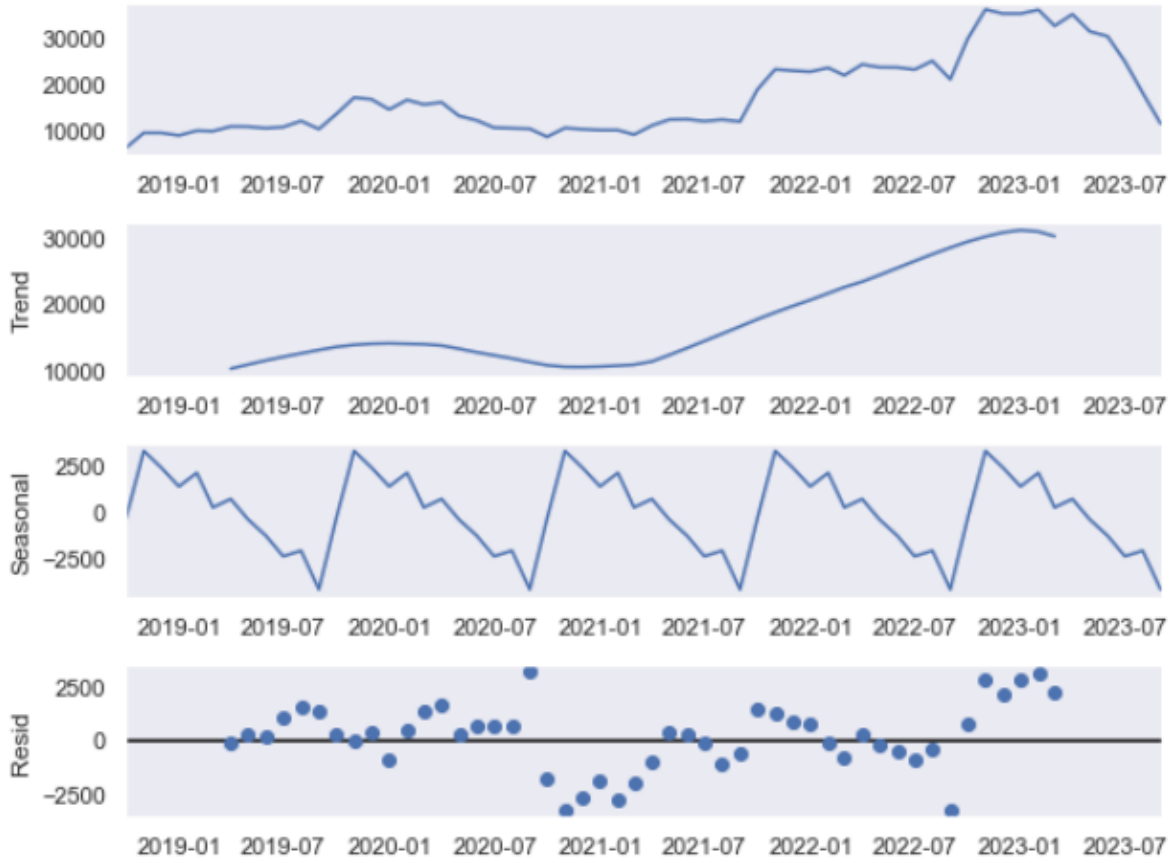
When the ACF graph created with the monthly occupancy rates of the above 5 academic periods is examined, it can be said that there is a clear trend in the monthly time series.

Graphic: 1st Differentiated ACf Graph of the Series



With the disappearance of the trend in the series, it can be said that seasonality is relatively less pronounced than the trend.

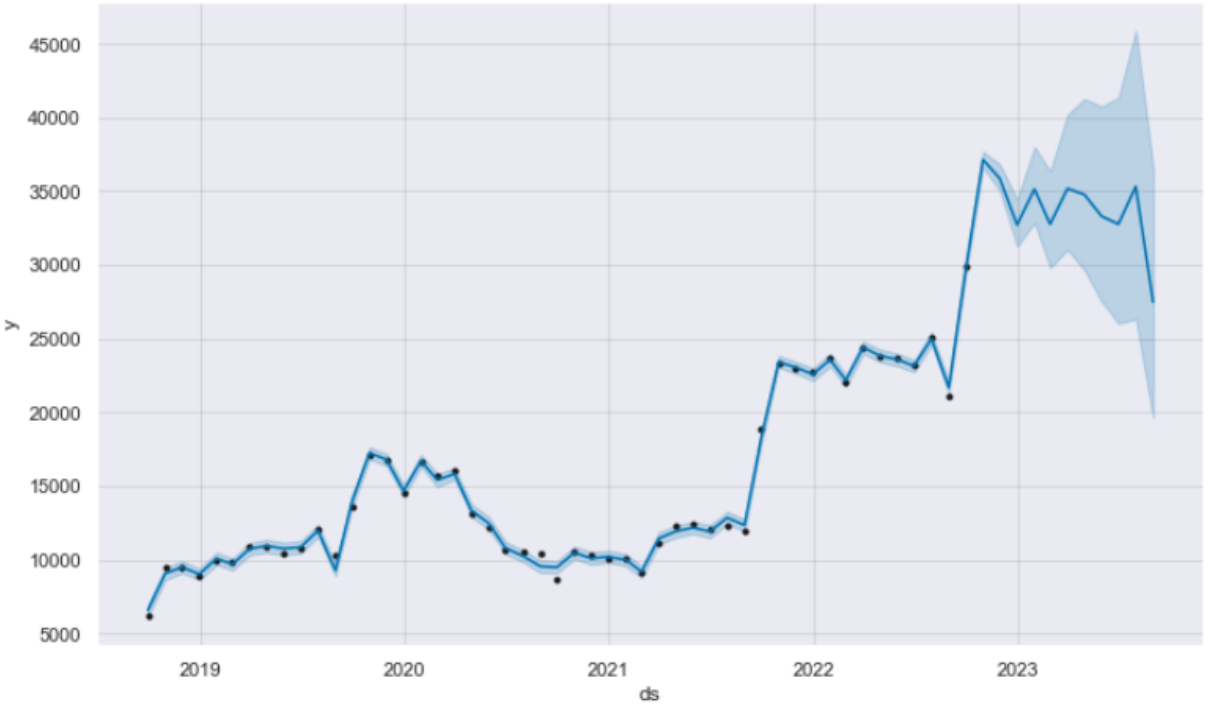
Graphic: Time Series Components



When the components of the monthly time series belonging to the academic years are examined; an increasing trend is mentioned. Also, the trend starts in January 2021. In other words, the trend continues to increase from January of the 2020-2021 academic term to the 2022-2023 academic year.

When the trend and noise are subtracted from the monthly time series created for the academic years, the seasonality in the series is clearly seen. As seen in the graph, the series repeats every 365 days (monthly, every 12 months). These dates coincide with the start and end times of the academic years.

Graphic: Monthly Number of Sold Rooms Forecast with Prophet Model

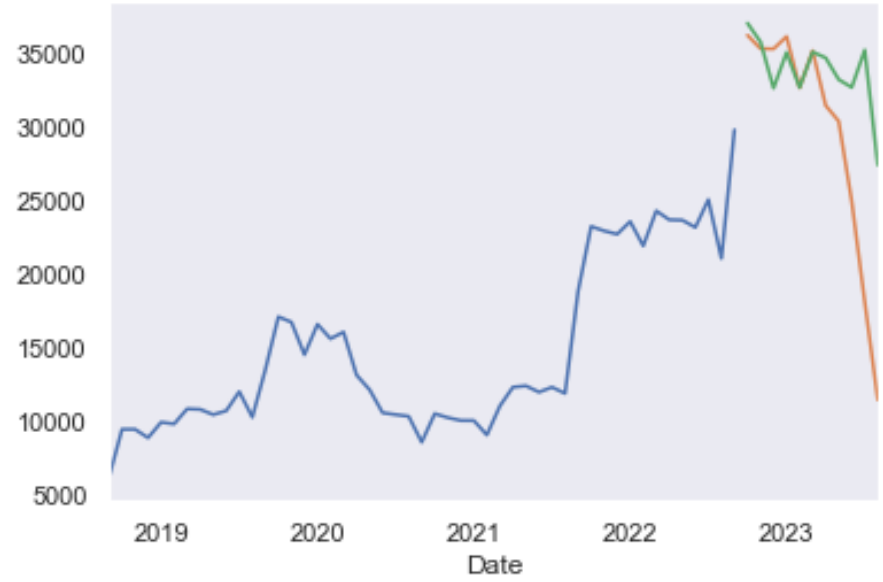


In the Prophet Model developed by Facebook, decomposable time series models are established with 3 main components. These components are in the series; trends, seasonality and holidays.

$Y(t) = g(t) + s(t) + h(t)$

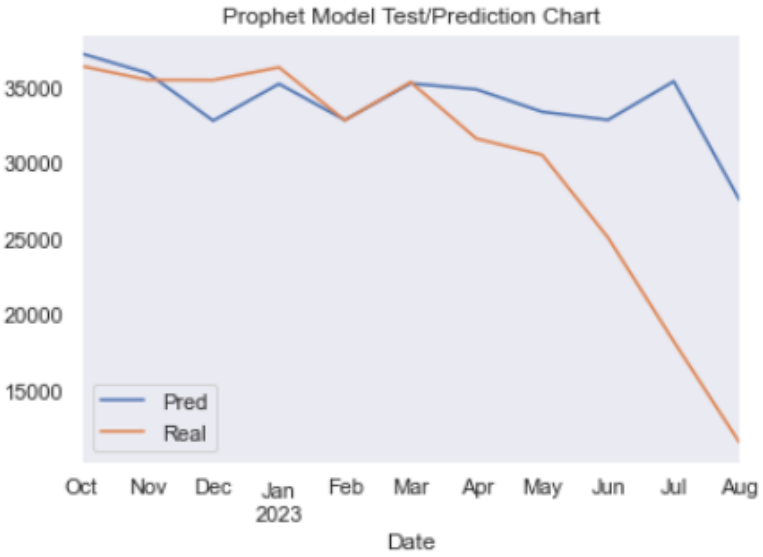
In the monthly total occupancy model prepared for Londonist; Monthly data from 2018-2019 academic year to September of 2022-2023 academic year were accepted as train data. With the data in this range, the trend, seasonality and holiday components of the model were learned and monthly occupancy estimates were made from October 2022-2023 until the end of the period until August.

Graphic: Prophet Model Train Data, Test Data and Forecasting Data



The graph created with 11-month forecasts and actual values for the 2022-2023 academic term is above. According to the graph, while the forecasts were close to the real values in some months, the difference was widened in some months.

Graphic: Prophet Model Test / Forecast



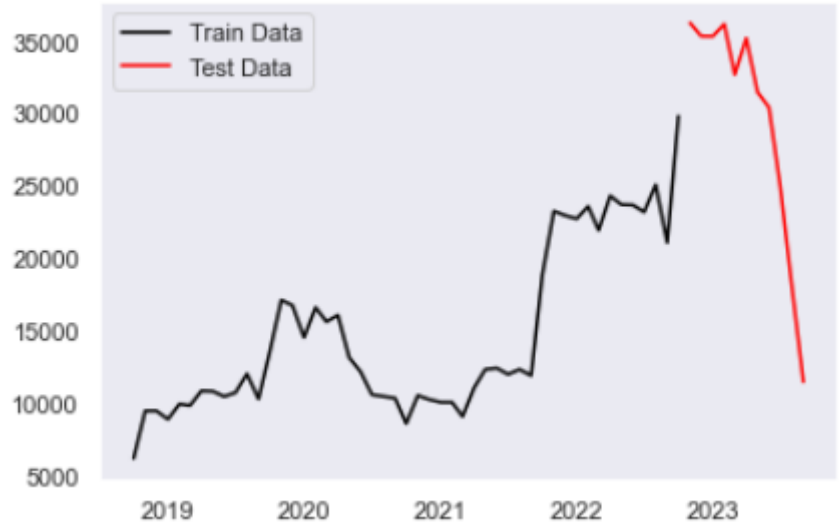
When we focus on closely at the forecasted months and the actual values in these months, although the forecast and actual values are quite close, especially until March, there is a decrease in the actual values after March. As can be seen from the forecast curve, although there was no sharp decrease in the forecast values, an increase was experienced in June and the total occupancy values decreased as of the end of July.

Table: Prediction of Occupancy Rates Based on Actual and Forecasting Values

	Date	Pred	Real	Capacity	Pred Occupancy	Real Occupancy
0	2022-10-31	37126.771600	36303	37200	0.998031	0.975887
1	2022-11-30	35849.923563	35391	36000	0.995831	0.983083
2	2022-12-31	32703.353331	35376	37200	0.879122	0.950968
3	2023-01-31	35132.492174	36219	36990	0.949784	0.979157
4	2023-02-28	32748.488701	32734	33404	0.980376	0.979943
5	2023-03-31	35163.050321	35277	36983	0.950790	0.953871
6	2023-04-30	34765.639088	31554	35790	0.971379	0.881643
7	2023-05-31	33288.141248	30484	36983	0.900093	0.824271
8	2023-06-30	32754.116541	25031	35790	0.915175	0.699385
9	2023-07-31	35297.175958	18188	36983	0.954416	0.491794
10	2023-08-31	27465.204814	11513	32211	0.852665	0.357424

The table includes the calculation of the last 11 months of forecast data for the 2022-2023 academic year, and the actual values, monthly capacities and occupancy rates. When the real values looked at instantly, as can be seen in the graphics, there is a decrease in the occupancy rates as of May. When the predicted values are examined, this decrease did not decrease as sharply as the real values, on the contrary, it continued to increase with the end of June. According to the occupancy rate prediction created with the forecasting values and capacity data; The 2022-2023 academic term will end in August with an occupancy rate of 85.26%.

Graphic: Training and Test Set



The training and test datasets created for the Arıma and Sarıma models are located in the fields shown in the chart above.

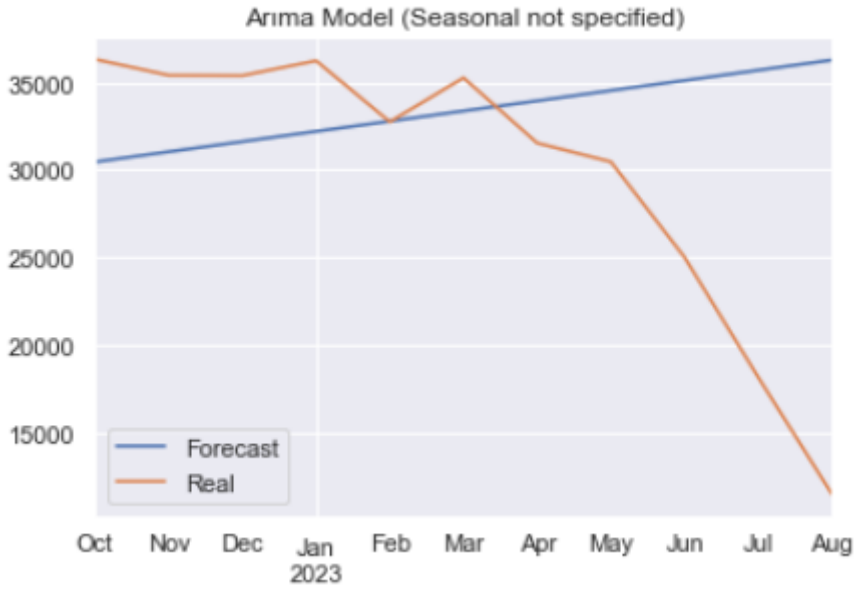
The Arima model is a statistical autoregressive model that forecast the future trend based on past values. Seasonality is ignored in the Arima model, as seasonality is a parameter that reduces the forecasting efficiency. In addition, the Sarima model, which is a model that does not ignore seasonality when there is a seasonal component in our monthly time series, is also used.

Graphic: Arima Model Train Data, Test Data and Forecast Data



The chart with 11-month forecasts and actual values created with the Arima model for the 2022-2023 academic year is above.

Graphic: Arima Model Test / Forecasting Graph



When we look closely at the forecasting months, it is seen that the forecast and actual values are close to each other only in certain months. An effective forecasting performance was not observed in this model, in which seasonal effects were not included.

Table: 11-month forecast values and actual values with the winding model

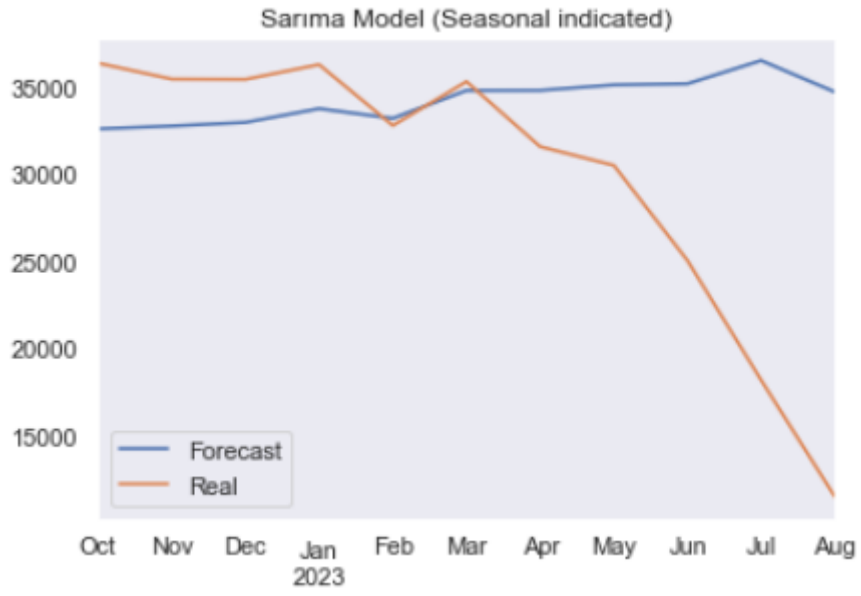
	Forecast	Real
2022-10-31	30443.618717	36303
2022-11-30	31025.237435	35391
2022-12-31	31606.856152	35376
2023-01-31	32188.474869	36219
2023-02-28	32770.093587	32734
2023-03-31	33351.712304	35277
2023-04-30	33933.331022	31554
2023-05-31	34514.949739	30484
2023-06-30	35096.568456	25031
2023-07-31	35678.187174	18188
2023-08-31	36259.805891	11513

Graphic: Sarima Model Train Data, Test Data and Forecasting Data



Above is the graph created with the Sarıma model of the 2022-2023 academic period, with 11-month forecasts and actual values.

Graphic: Sarima Model Test / Forecasting Graph



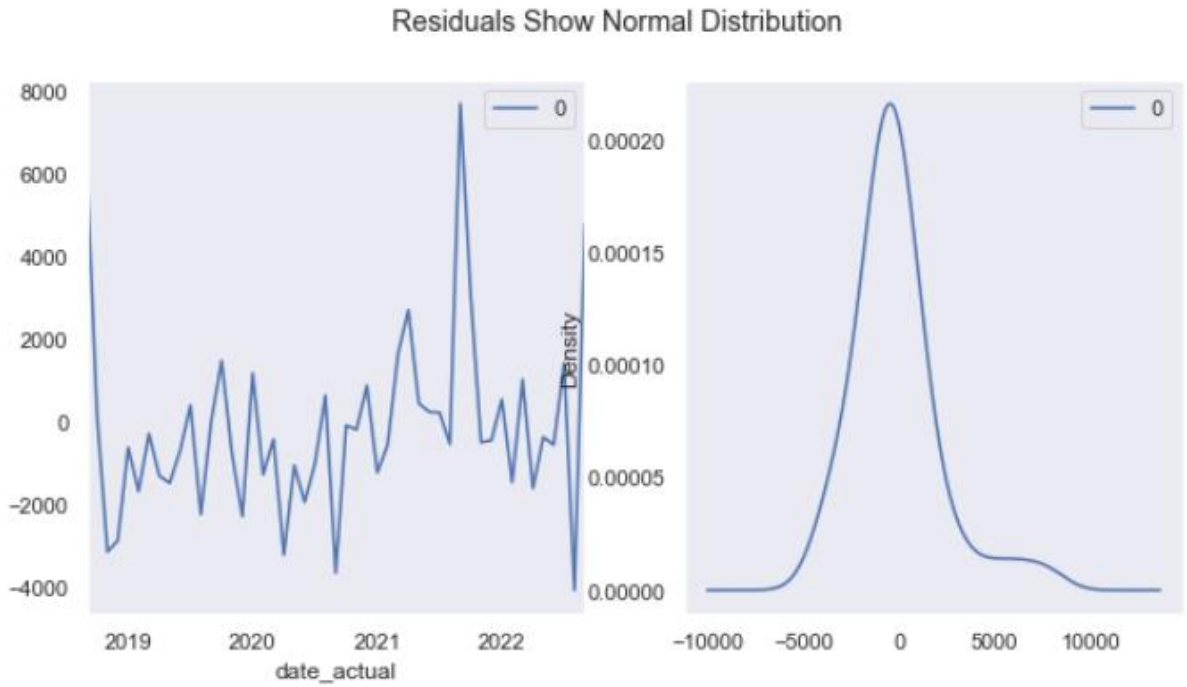
When we look closely at the estimated months, it is seen that the estimated and actual values in February and March are very close, while the difference is not much in other months. In

addition, as seen in the predictions made in the prophet model, there are no sharp decreases in the predictions made with the Sarıma model after May.

Table: 11-month forecast values and actual values with the winding model

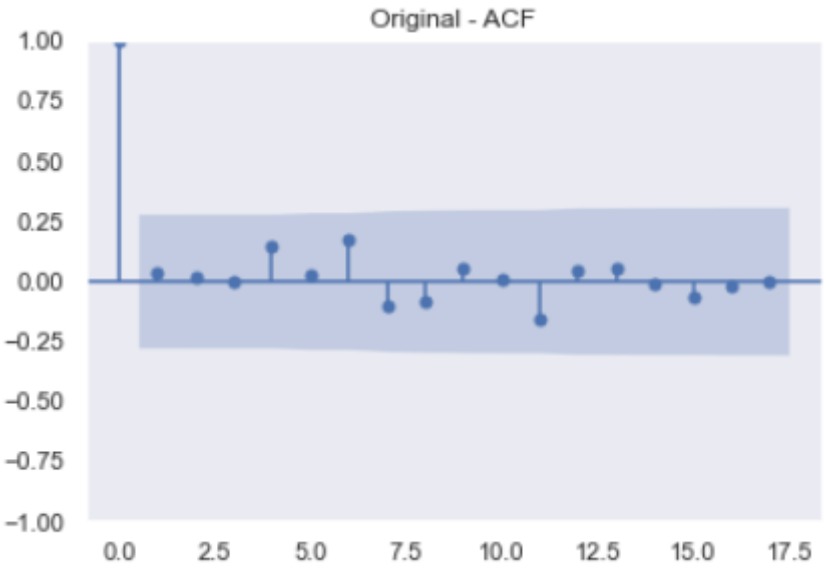
	Forecast	Real
2022-10-31	32538.871550	36303
2022-11-30	32704.997049	35391
2022-12-31	32912.073077	35376
2023-01-31	33697.774757	36219
2023-02-28	33144.340967	32734
2023-03-31	34738.416720	35277
2023-04-30	34742.867404	31554
2023-05-31	35055.776617	30484
2023-06-30	35122.982657	25031
2023-07-31	36453.805011	18188
2023-08-31	34653.241483	11513

Graphic: Distribution of Residuals



The common component of all established prediction models is the error term. It can be said that the residuals term mean of the Sarima model is close to zero and the error term distribution is a normal distribution.

Graphic: Autocorrelation Analysis Between Residuals



Looking at the ACF graph created to determine whether there is a relation between the residuals, it is mentioned that there is no relation between the residuals. In addition, it can be said that the models established here are statistically significant.

Summary

Forecasting of the upcoming weeks and months is very critical for machine learning because of the fact that we decide prices based on this information in our dynamic pricing model. Until today we were calculating occupancy rates of upcoming weeks and months manually (for example we knew the current occupancy rate of May 2023. However, we didnt know what it will be by the end of May. Now we have machine learning model which can forecast the occupancy levels of the future)

According to the analysis findings; The Prophet model deviation is only %1 (It means we can forecast the occupancy level for the future with %1 deviation), the Arima model deviation was %14 and finally the Sarıma model deviation was %8.

Based on these analysis, we decided to use The Prophet Model in our dynamic pricing algorithm.