CASE STUDY

AIRBNB

ANALYSIS OF RENTAL LISTINGS

TOOL: JUPYTER AND TABLEAU

LANGUAGE: PYTHON

Airbnb provides an online platform to rent lodging for tourism activities. With this analysis, I explored the listings located in Broward County, FL where big cities of Hollywood and Fort Lauderdale are in a close proximity to the beach, which makes this region highly attractive for travel.

In this project, an advanced exploratory analysis conducted in Python built an interactive dashboard in Tableau to showcase the results.

The open-source dataset comes from the www.insideairbnb.com, a project that uses web scraping methods to collect the data. For this analysis, I used the set with properties that were listed on September 27th, 2021. For the geographical visualization, I used the JSON file from the project's website.

The data set contains listing ids with localization along with their pricing, availability, and reviews.

When choosing this data, I mainly focused on the requirements specified in the project brief only to discover on the way to my results that a meaningful prediction cannot be performed on every set .

After sourcing and cleaning the data, a correlation diagram was created to interpret the relationships between variables.

In the set, none of the data showed direct dependency and at this point I recognized the wish to have more elements in the set that could show some influence on the price per night for the rented property.

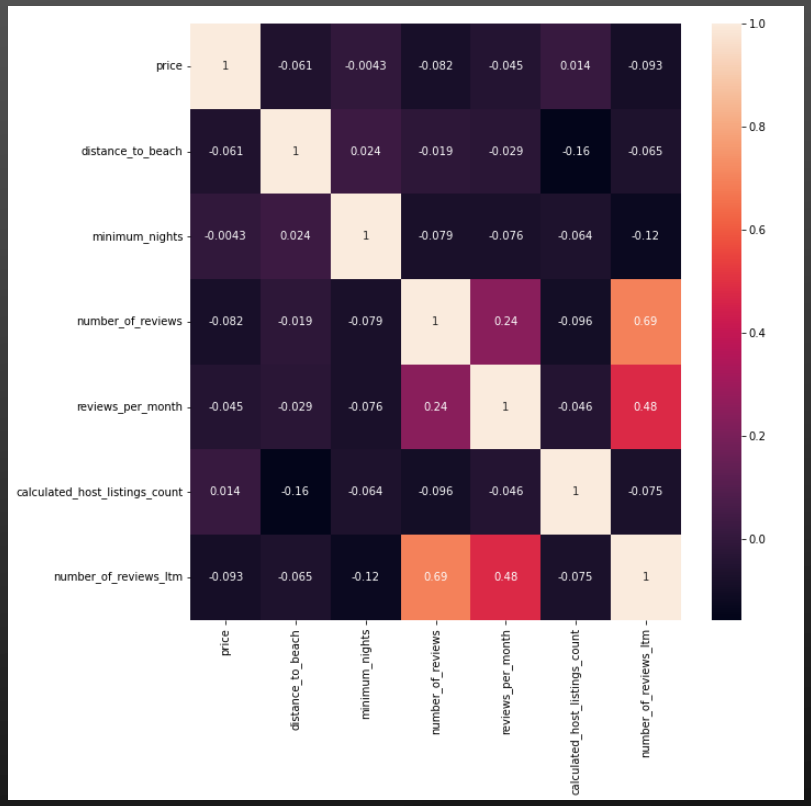

Fig. 1 Correlation diagram

Nevertheless, by creating a new data with a property's distance to the beach, I was able to explore this relationship by building a scatterplot and running a linear regression algorithm.

In the unsupervised machine learning algorithm, I used clusters to confirm that there is no visible linear connection between both variables.

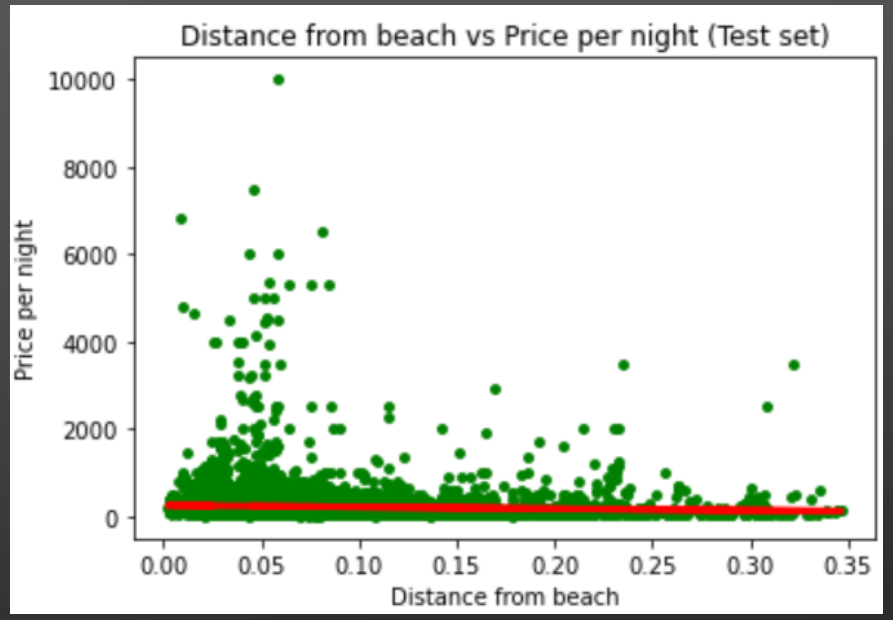

Fig. 2 Regression analysis

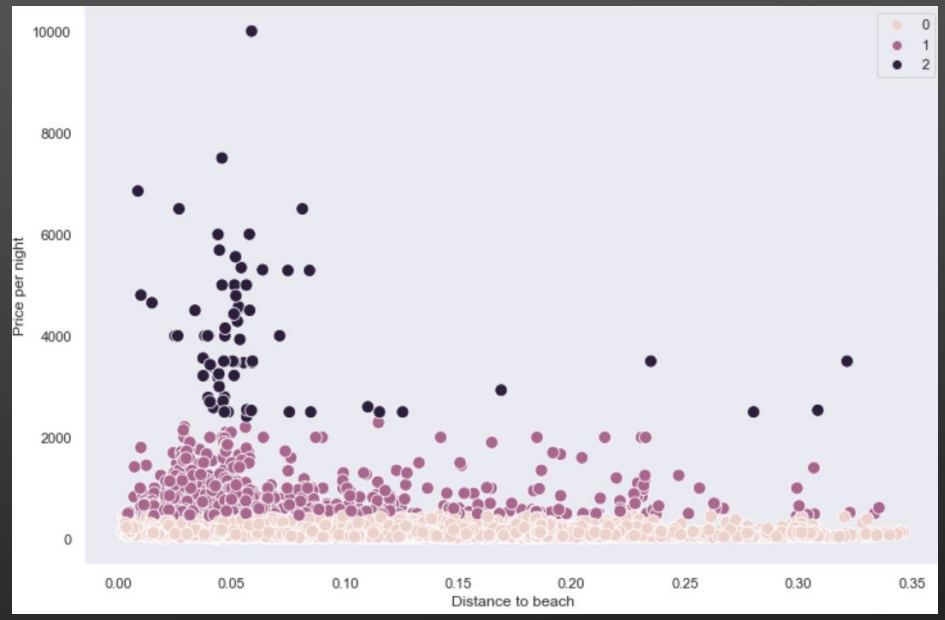

Fig. 3 Clusters

The most pricey rental properties are concentrated closer to the ocean, as is the majority of places with medium high prices. The rentals with the lowest prices per night are distributed across all the distances.

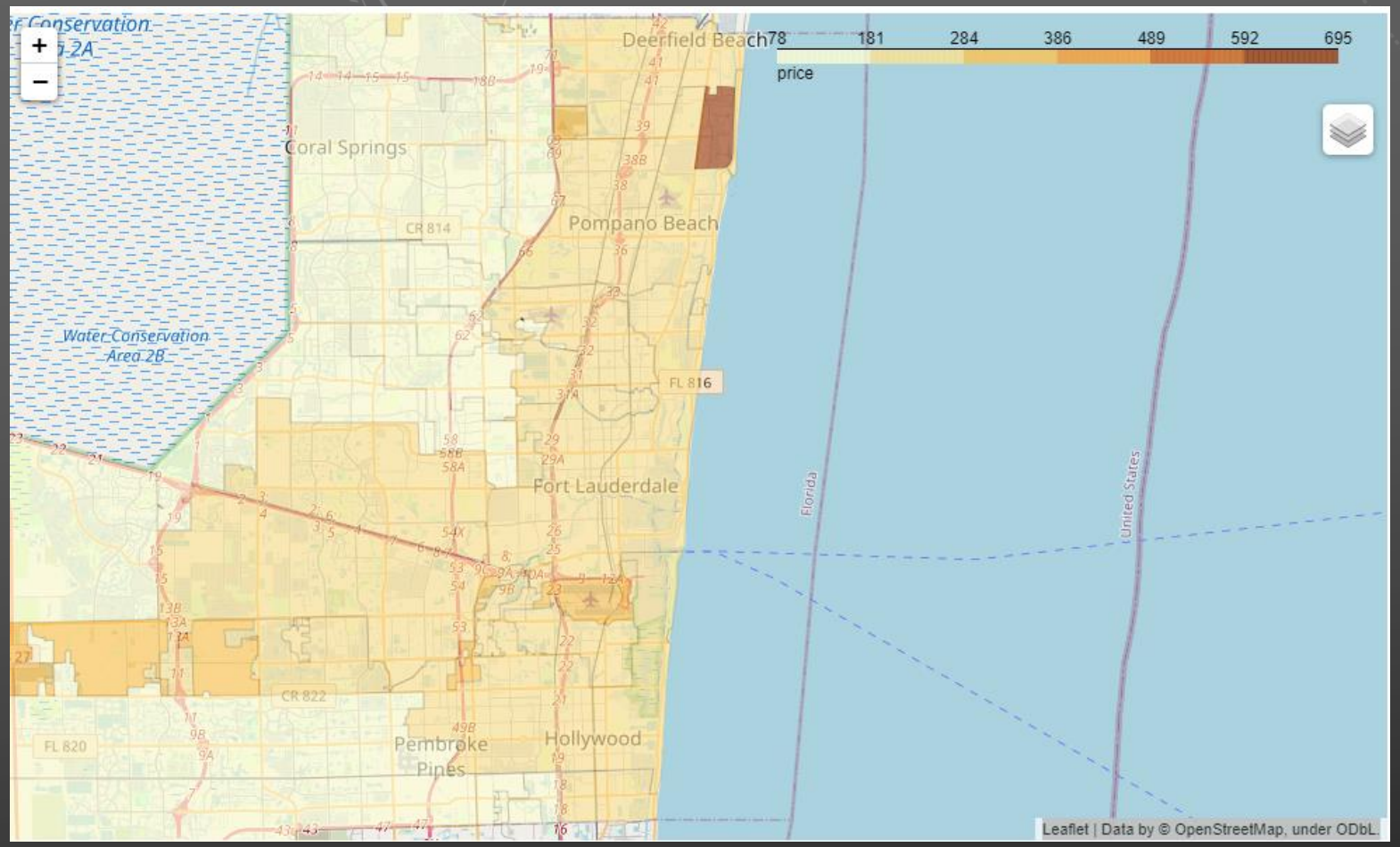

Fig. 4 Map of neighborhoods with average price per night

Using the geo JSON file, I created a map that represent the regions and shows the average price per night. It was interesting to uncover the most pricey neighborhood in Lighthouse Point (the darkest orange) or the more pricey airport area surrounded by average priced places. The map only confirmed the great variability of listings' prices.

To plot the time series analysis of the data, I grouped the numbers of reviews left in the last 12 months and build a data frame with data for each day of the last 3 months.

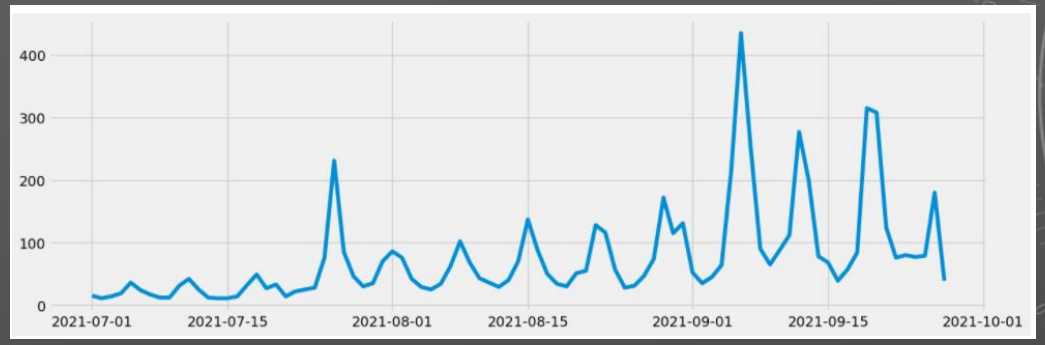

Fig. 5 Timeseries

The Dickey-Fuller test was performed to test the data's stationarity to check whether they are ready for predictions. A few differencing operation were performed to stationarize the data.

```
# The adfuller() function will import from the model from statsmodels for the test; however, running it will only return
# an array of numbers. This defines a function that prints the correct output from that array.

from statsmodels.tsa.stattools import adfuller # Import the adfuller() function

def dickey_fuller(timeseries): # Define the function
    # Perform the Dickey-Fuller test:
    print ('Dickey-Fuller Stationarity test:')
    test = adfuller(timeseries, autolag='AIC')
    result = pd.Series(test[0:4], index=['Test Statistic','p-value','Number of Lags Used','Number of Observations Used'])
    for key,value in test[4].items():
           result['Critical Value (%s)'%key] = value
    print (result)

# Apply the test using the function on the time series
dickey_fuller(df_2021_07_08_09['no-of_reviews'])
```

```
Dickey-Fuller Stationarity test:
Test Statistic                 -1.832094
p-value                         0.364648
Number of Lags Used             5.000000
Number of Observations Used    83.000000
Critical Value (1%)            -3.511712
Critical Value (5%)            -2.897048
Critical Value (10%)           -2.585713
dtype: float64
```

Fig. 6 The Dickey-Fuller test

When exploring the listings published on the Airbnb platform, it's easy to recognize that they are characterized by many more factors than only those given in the set.

Between those given in the set, no strong dependency could be confirmed. Within the measured correlations, they show either extremely weak or no relationship.

In fact, more pricey rentals can be found in the smaller proximity to the beach but so can the listings at a lower price. This made any trials for predictions of rental price difficult and more factors need to be considered.

The Python scripts are available here in my GitHub repository.
An interactive dashboard can be accessed in Tableau.