Time Series Analysis- Airbnb

This report provides a comprehensive analysis of time series data from the Airbnb dataset, spanning 2016-2017. The data consists of prices of various listings per day, reviews of listings and listing information. We will be taking a look into the foremost data listed above i.e. ‘calendar.csv’. We hope to uncover patterns and answer intriguing questions about the seemingly random nature of prices.

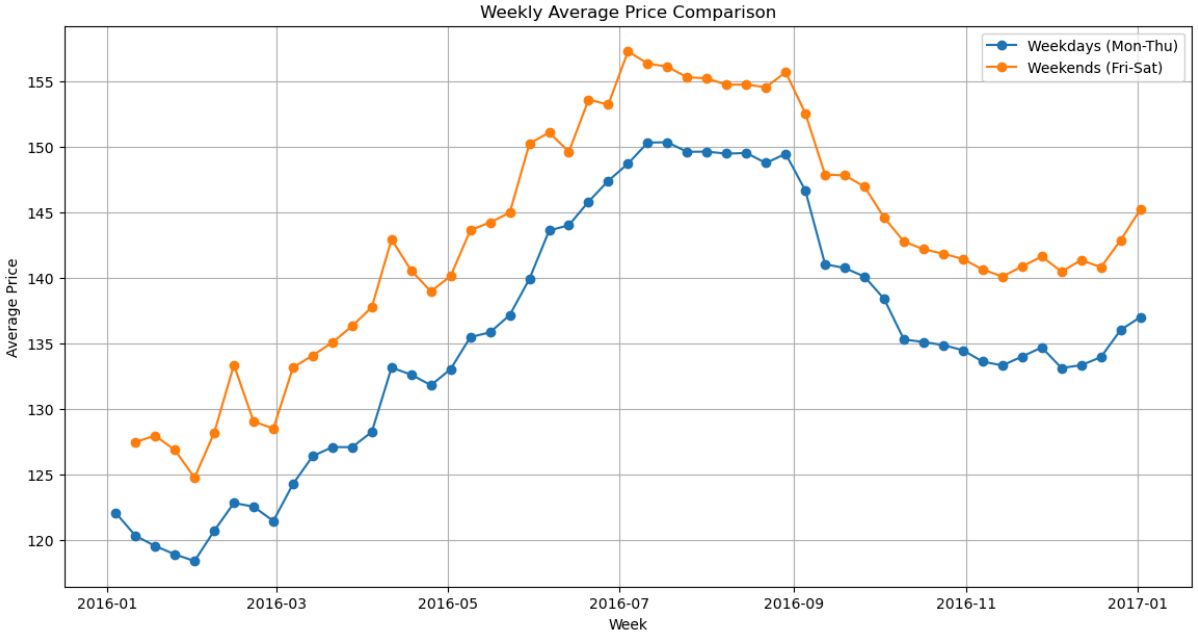
Preliminary Challenges

The data from “calendar.csv” is not in the desired form for analysis, we need to -

1. Convert prices into integers(remove dollar sign, remove commas)
2. Convert date to ‘datetime’ datatype

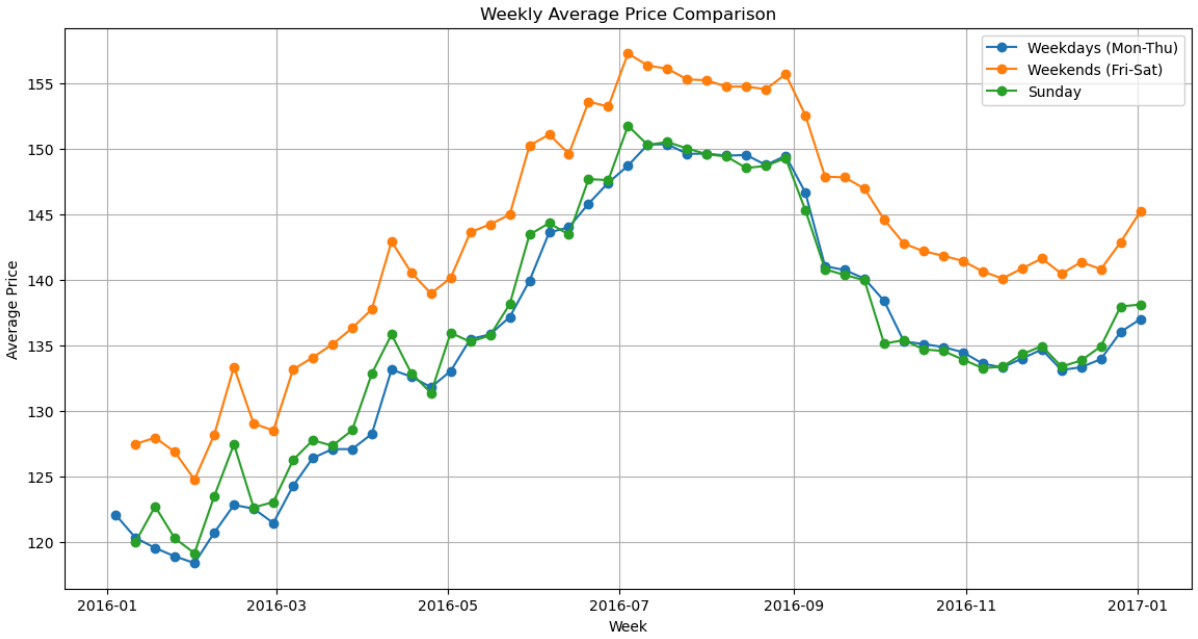
Weekly Trends of Data

Airbnb being a hospitality service provider would see costs go up when the demand for travelling is higher. Our hypothesis- the cost of Airbnb rentals would go up during the weekends as people have time off from work and would prefer to travel on the weekends more than weekdays. To test the hypothesis we plot a graph of weekly average prices between Monday to Thursday and between Friday to Saturday.



We see that throughout the year weekend prices are consistently above weekday prices hence confirming our hypothesis.

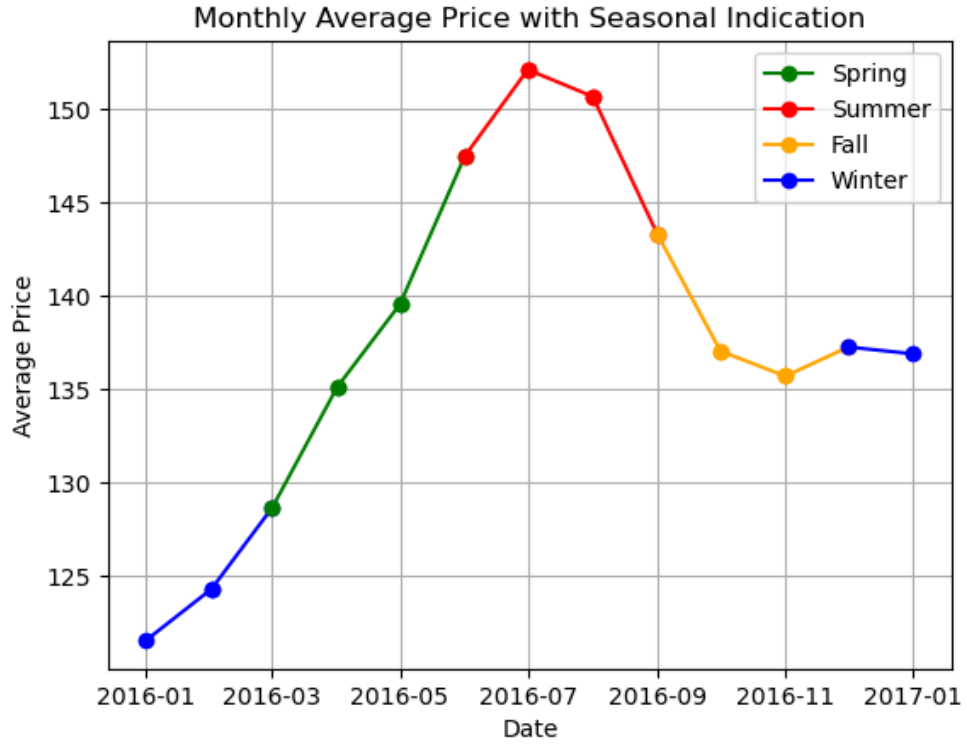
Note:- As the work week resumes on Monday, many travellers prefer to be at home Sunday night leading to lesser demand than on Friday and Saturday. This hypothesis is confirmed by the following graph



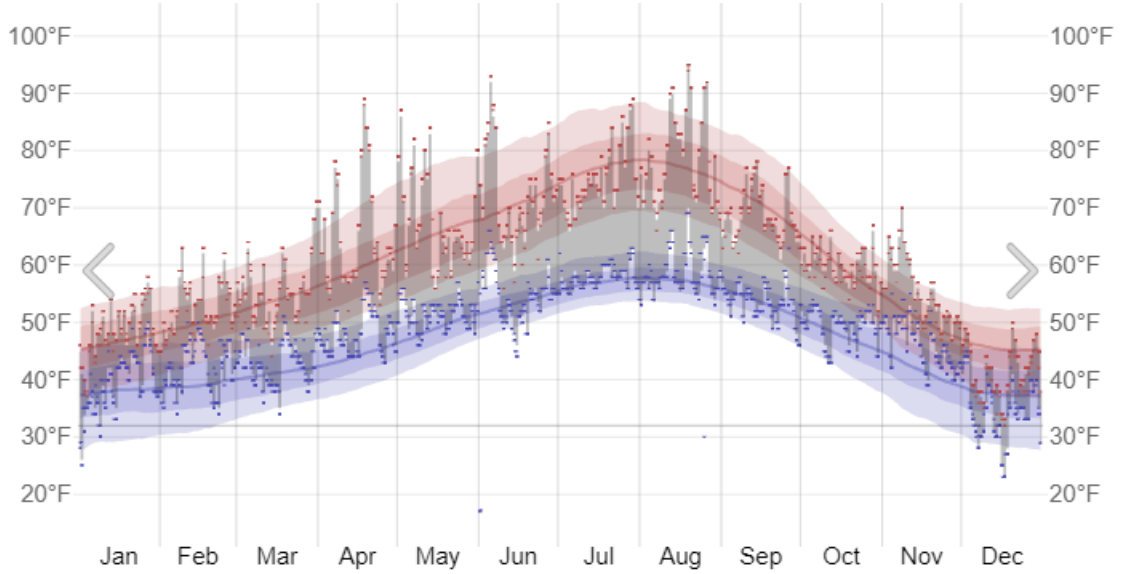
We can see that the average price on Sundays more closely matches weekday prices than weekends although there is some deviation.

Seasonal Trends Of Data

The data spanning a full calendar year gives the unique opportunity to explore trends in travelling with respect to the seasons. Our hypothesis- Prices of Airbnb rentals in Seattle will tend upward during the summer and spring months as winter months are too cold to attract tourists to the area. Another reason could be that children have their summer breaks thus families travel more during these months. To test our hypothesis we plot the monthly average of prices and see the differences across seasons.



We can see a clear peak during mid-summer and winter seems to be when prices are at their lowest. There is a significant difference between average prices in January 2016 versus January 2017, this can be explained by the overall rise in popularity of Airbnb in 2016.



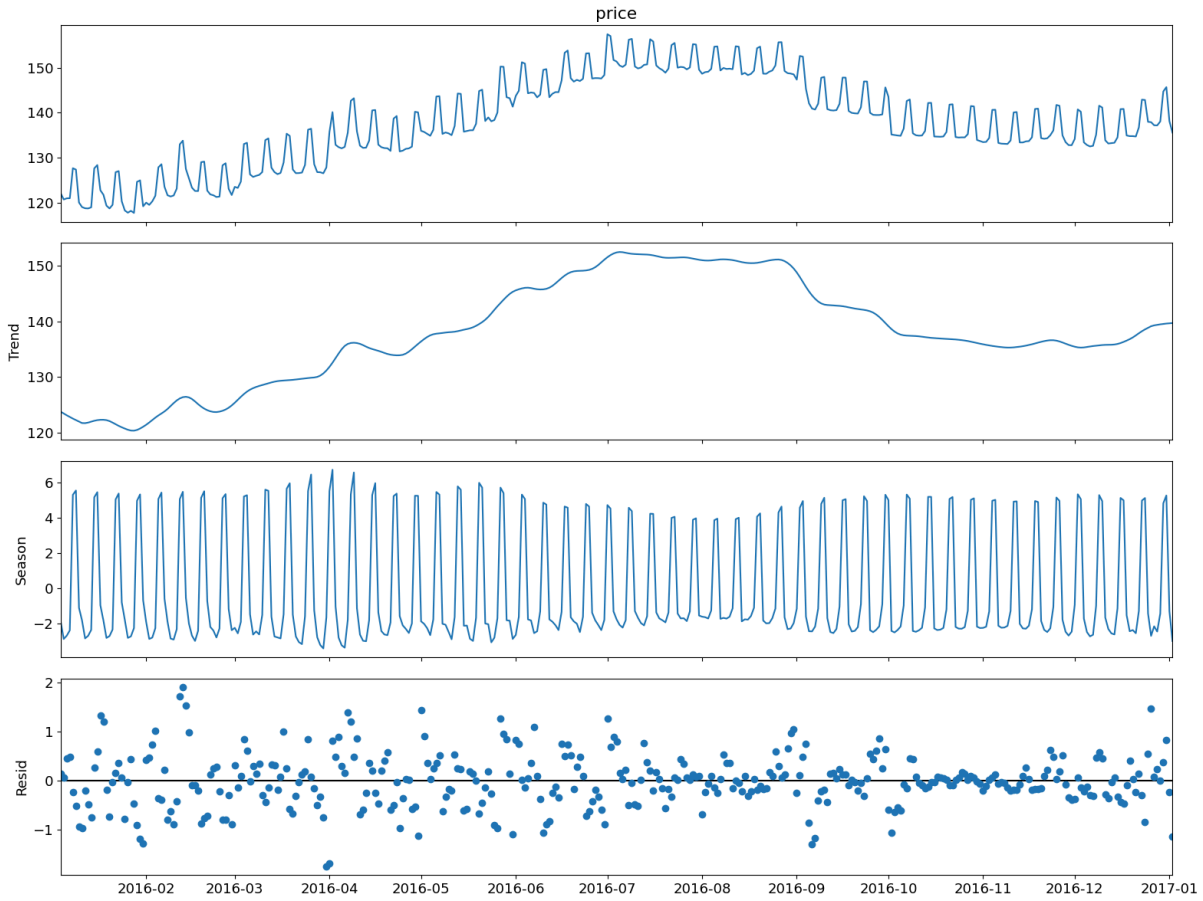
This graph shows the temperature of Seattle in 2016. As we can see in January and December the temperatures are below freezing and it corresponds to the lowest average months. July- August is when Seattle is warmest and it correlates with the peak of prices. The temperature is within 20-27 which is quite comfortable and invites tourists.

STL Analysis

Seasonal decomposition is a statistical technique for breaking down a time series into its essential components, which often include the trend, seasonal patterns, and residual (or error) components. The goal is to separate the different sources of variation within the data to understand better and analyze each component independently. The fundamental components are:

1. Trend: The underlying long-term progression or direction in the data.
2. Seasonal: The repeating patterns or cycles that occur at fixed intervals like daily, monthly or yearly.
3. Residual: The random fluctuations or noise in the data that cannot be attributed to the trend or seasonal patterns.

Seasonal-trend decomposition using Loess (STL) is a specific decomposition method that employs the Loess technique to separate a time series into its trend, seasonal, and residual components. Locally Weighted Scatterplot Smoothing or Loess is a non-parametric regression method used for smoothing data



Understandings from the above-

1. Trend- Shows the average prices tending upward from January to August and then going downwards after September
2. Seasonality- The period was set to 7 i.e. weekly. We see that the prices fluctuate substantially during the week, reaching a peak during the weekends and then taking a sharp turn downwards. We see this weekly fluctuation occurs every week and roughly to the same extent as the peaks reach a similar height.
3. Residual- Shows the data points that are irregular and cannot predicted by the above two

This report analyzed the Airbnb dataset to uncover key trends in weekly pricing, seasonal variations, and decomposed time series components using STL (Seasonal-Trend decomposition using Loess). The weekly analysis revealed that prices peak on Fridays and Saturdays due to increased demand for weekend stays, with a noticeable dip on Sundays. Seasonal trends highlighted distinct patterns, with higher prices during peak vacation months such as summer and lower prices during the winter months. The STL decomposition provided a clear separation of the trend, seasonal, and residual components, offering valuable insights into underlying patterns and irregularities. These findings can guide hosts in optimizing pricing strategies and improving occupancy rates throughout the year, ultimately enhancing the profitability and attractiveness of their listings.