Assumptions & Conclusions on Aibnb\_zillow data challenge

# Conclusions

The conclusions that were drawn from Airbnb\_zillow challenge are as follows:

Top 3 zip codes that are profitable to invest in New York city are :

1. 11215 (Brooklyn)
2. 11231 (Brooklyn)
3. 10308 (Staten island)

These zip codes have relatively lower Break Even point while having relatively higher Revenue To Cost ratio. This implies that, if we invest in properties in these zip codes, it will take relatively lower time to break even and start making profits. These zip codes also have relatively lower prices as well.

If higher capital is available zip code 10028 (Manhattan) can also be a lucrative option. Though this zip code has high break-even point(40+) years, it also has highest revenue to cost ratio. If the properties here can be rented out at prices above mean, the break-even point can be lowered, thereby making this more lucrative option.

# Assumptions & Decisions for Analysis

**Assumption 1:** We chose mean value of price per night for all properties in zip code.

**Rationale**: This assumption was made to be conservative in calculations. Since, amenities provided, transport available, connectivity and other factors play significant role in renting price as well as cost of property, using mean value would result in conservative analysis thereby allowing scope of upward adjustments.

**Assumption 2:** We chose ARIMA model for predicting current prices(for Jan, 2020) of property in zip code

**Rationale:** Zillow data contains monthly prices of properties, which is limited up to June 2017. The data shows an upward trend implying that prices are rising over the years. Since the data isn’t stationary, ARIMA model would be a better fit to predict the future prices. Though property prices are seasonal in some aspect and other aspects like connectivity, proximity to market and other factors play a role in deciding property prices, for ease of calculations, those factors were not considered. Since, ARIMA model doesn’t consider seasonal or other aspects of data, it applies better on this property prices.

A jupyter code file testing the ARIMA model on one zip code data is attached for reference.

# Scope of Improvement

The above analysis has some scope of improvements. The break even point and revenue calculations can be refined based on analysis of:

1. If the square-feet data is imputed using classification and regression, it can refine the revenue calculations.
2. Reviewing amenities data provided for each property in zip code using word cloud, we can identify which amenities are more important for guests and thus factor it in revenue calculations.
3. We can improve current price estimates for zip-codes by using other prediction models like LSTM

# References:

1. <https://pandas.pydata.org/pandas-docs/stable/user_guide/merging.html>
2. <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.groupby.html>
3. <https://towardsdatascience.com/time-series-forecasting-arima-models-7f221e9eee06>
4. <https://uszipcode.readthedocs.io/01-Tutorial/>
5. <https://www.analyticsvidhya.com/blog/2017/12/fundamentals-of-deep-learning-introduction-to-lstm/>
6. Stack Overflow for some plots and pandas data-frame operations