AirBnB & Zillow Data Challenge

This project aims at finding the most profitable zipcode location in NYC area. Recommendations will be given regarding which zipcodes are most profitable for investing for short-term rental. Additionally, suggestions for further research will be provided at the end. Data is provided by AirBnB (revenue data) and Zillow (cost data).

Authors

Xiaojun Yao

email: xiaojunyao1199@outlook.com

1.0 Getting Started

Within the XYAO_work.zip file, should contains the following:

- Folders
  - Data

    ```
    Zip_Zhvi_2bedroom.csv
    XYAO_nyc_Zipcode_Latitude_and_Longitude.csv
    ```

  - Output (data and graphs generated from code)

    ```
    XYAO_profit.csv
    XYAO_revenue.csv
    XYAO_cost.csv

    XYAO_nyc_Zipcode_Latitude_and_Longitude.csv

    Boxplot of Three measures.png
    Top 10 Zipcode Rating by Net Operating Income.png
    Top 10 Zipcode Rating by price.png
    Top 10 Zipcode Rating by Return On Investment.png
    ```

- Code and documentation

  ```
  README.md
  XYAO_code.ipynb
  XYAO_functions.py
  XYAO_Metadata.pdf
  README_backup.pdf
  ```

- ReadMe Backup
    - an additional pdf format of README just in case README.md failed to open

      `README_backup.pdf`

- Additional data (in the `Data` folder)
    - an additional dataset generated from Airbnb dataset which contains Latitude and Longitude information for each zipcode

      `XYAO_nyc_Zipcode_Latitude_and_Longitude.csv`

- File should be opened in the following order:

```
1. README.md (Source documentation)
2. XYAO_code.ipynb (Data Analysis & Reporting)
3. XYAO_functions.py (Class & Functions)
4. XYAO_Metadata.pdf (Metadata for output table `full_table.csv`)
```

1.1 Programming Tools

- Anaconda Python Distribution: Jupyter Notebook
- Python 3.6.5

1.2 Installing and Importing Libraries

- Some packages might need to be installed, for example:

```
!pip install folium
!pip install ipywidgets
!pip install gmplot
```

- Most importanly, import XYAO_functions.py:

```
from XYAO_functions import *
```

2.0 Assumptions

The following assumptions are made:

2.1 Airbnb Dataset

- properties `price` has considered operating cost: meaning `price` per night of stay is calculated by its revenue/night minus its operating expenses/night.
- `neighbourhood_group_cleansed` and `neighbourhood_cleansed` are most correct location detail for each zipcode.
- all listed properties in this table are properties in New York City (recognized based on `city` and `state`)
- properties with missing `zipcode` do not have major effect on final result
- when table is grouped by `zipcode` , taking *median is an appropriate representative for the group

2.2 Zillow Dataset

- cost of properties in `2017-06` (June 2017) is the most up-to-date cost value and still valid for today
- cost of properties before `2017-06` represent the past values and should not be considered.

3.0 Measurement Assumption and Calculation

3.1 Revenue (Net Operating Income)

- 75% Occupancy rate: meaning that a property is occupied about 273 days every year (365 days/year)
- Calculate Annual Net Operating Income (NOI)
    i. it is assumed that price/night of stay is calculated by its revenue/night minus its operating
    expenses/night.
    ii. Daily price is a small and narrative representation of revenue of short-term rental.
    iii. it is also hard for us to calculate profit if we use price/night: Thus we use annual NOI
    iv. NOI = ['price'] x 365 days x 75%

3.2 Cost

- June 2017 costs are the most up-to-date records
- June 2017 costs have the same validity as the Airbnb data (2019-07-08).
- past cost values are not considered in this case.

3.3 Profit (Return On Investment)

- Since these properties are for short-term rental, it is hard to measure profit in an easy way
- In this case, we consider to calcualte `Capitalization Rate` (which is the same as `Return on Investment`)
- For an investment property to remain profitable as time goes by,its net operating income must increase either at the same rate as its market value, or at a greater rate. The capitalization rate is a strong measure of whether a property is becoming more or less profitable.
- For example, if ROI=15.5%, a property is rented a year (75% occupancy rate), the company would stand to earn 12.5% of the property's value as profit each year, assuming that NOI and market value remain constant.
- ROI = NOI / investment cost

4.0 Insights and Markdown

- Each step of data analysis is ended with an `Insights` part.
- `Insights` include:
  - key findings from data or graphs
  - why and how are assumptions maded
  - explanation of calculations
  - detailed layout of data analysis steps

5.0 Recommendations and Future Research

Recommendations and Future Research are included at the end of `XYAO_code.ipynb`.