NYC INVESTMENT ANALYSIS

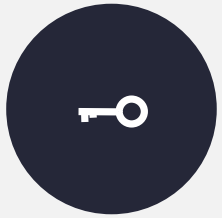
Xinran Tao

02/05/2019

Content

Business Context

Data Preparation & Quality Check



Model Methodology

Model Results & Insights

Further Analysis

Business Context

Business Context

Background

Our client is a real estate company, they have a niche in purchasing properties to rent out short-term as part of their business model specifically within New York City. The real estate company has already concluded that two-bedroom properties are the most profitable; however, they do not know which zip codes are the best to invest in.

The real estate company has engaged our firm to build out a data product and provide conclusions to help them understand which zip codes would generate the most profit on short term rentals within New York City.

Objective

Find out the most profitable zip code for short-term investment in NYC.

Data Preparation & Quality Check

Data Overview

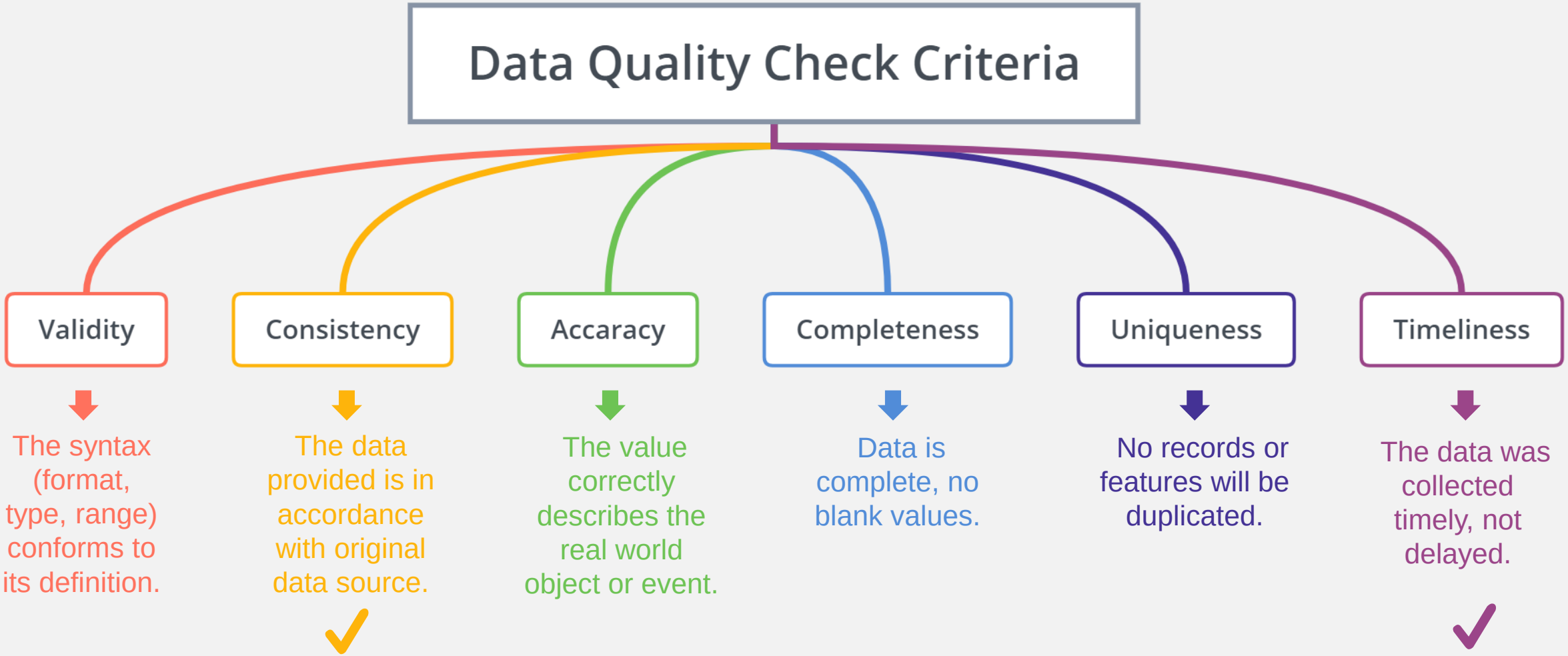
Airbnb Dataset — Revenue Data

AirBnB provides information on the listings with 97 features including **location** (City, State, Zipcode), **property features** (Bedrooms, Room Types) and the **rent price** (Daily Price, Weekly Price, Monthly Price) to lease out their property. The date on rent is in 2017-05.

Zillow Dataset — Cost Data

Zillow provides information of the **location** (Zipcode, City) and the historical median **property price** within that area in different years (Each month's price is a column). The time range is from 1996-04 ~ 2017-06.

Data Quality Check & Data Engineering



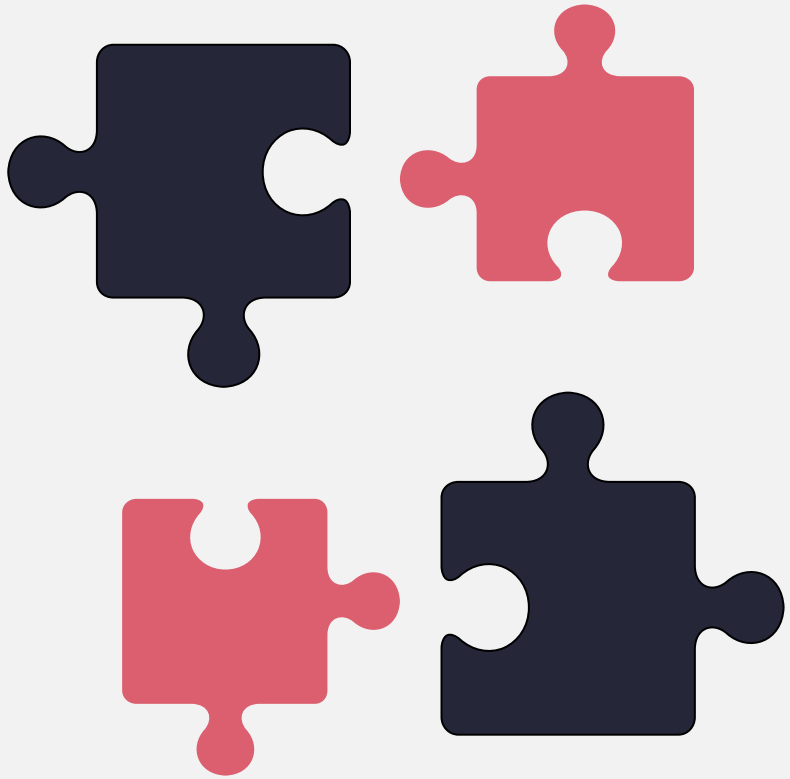
Data Quality Check & Data Engineering

Validity

- ✓ Data Types: Correctly reflect the definition (using .dtypes)
- ✓ Data Values: Only check relevant values for efficiency. The format and range are reasonable(check unique values, and outliers). The zipcode values are unique for each record.

Uniqueness

- ✓ Duplicates: No duplicated rows or columns(using .duplicated())



Accuracy

- ✓ Data Values: The values correctly describe the data (check unique values and descriptive statistics).

Completeness

- ✓ Missing Values: There are only missing values for the earlier years (1996-04 ~ 2007-05).
- ❖ How to fix: Delete the columns for the earlier years and keep the most recent 10 years data.

Data Quality Check & Data Engineering

Validity

- ✓ Data Types: ‘zipcode’ and ‘price’ has wrong data types because of the special characters like – , $.
- ❖ How to fix: Using user-defined functions to delete special characters or extract strings.
- ✓ Data Values: Same process as Zillow. Format and range are reasonable. Detected Outliers(<1%.)
- ❖ How to fix: Delete outliers.

Uniqueness

- ✓ Duplicates: No duplicated rows or columns(using .duplicated())

Accuracy

- ✓ Data Values: Messy and wrong values in ‘State’ and ‘City’ columns. There are several formats to represent the same state or city.
- ❖ How to fix: Not use these column to select zipcode. Using USPS provided zipcode list.

Completeness

- ✓ Missing Values: All the columns with missing values have missing ratio of over 78% (including weekly price and monthly price).
- ❖ How to fix: Delete the columns. Use daily price as revenue metric.

Data Manipulation

Airbnb Dataset

- Select zip code in New York City using USPS NYC zipcode list rather than using City or State to choose the zipcode to make sure our selection is accurate.
- Select *Bedrooms* = 2 & *RoomType* = Entire Home to make sure the property type meet the client's investment requirement.
- Delete irrelevant columns and only keep the columns of zipcode and daily price.
- Aggregate by zipcode by using the mean value of daily price.

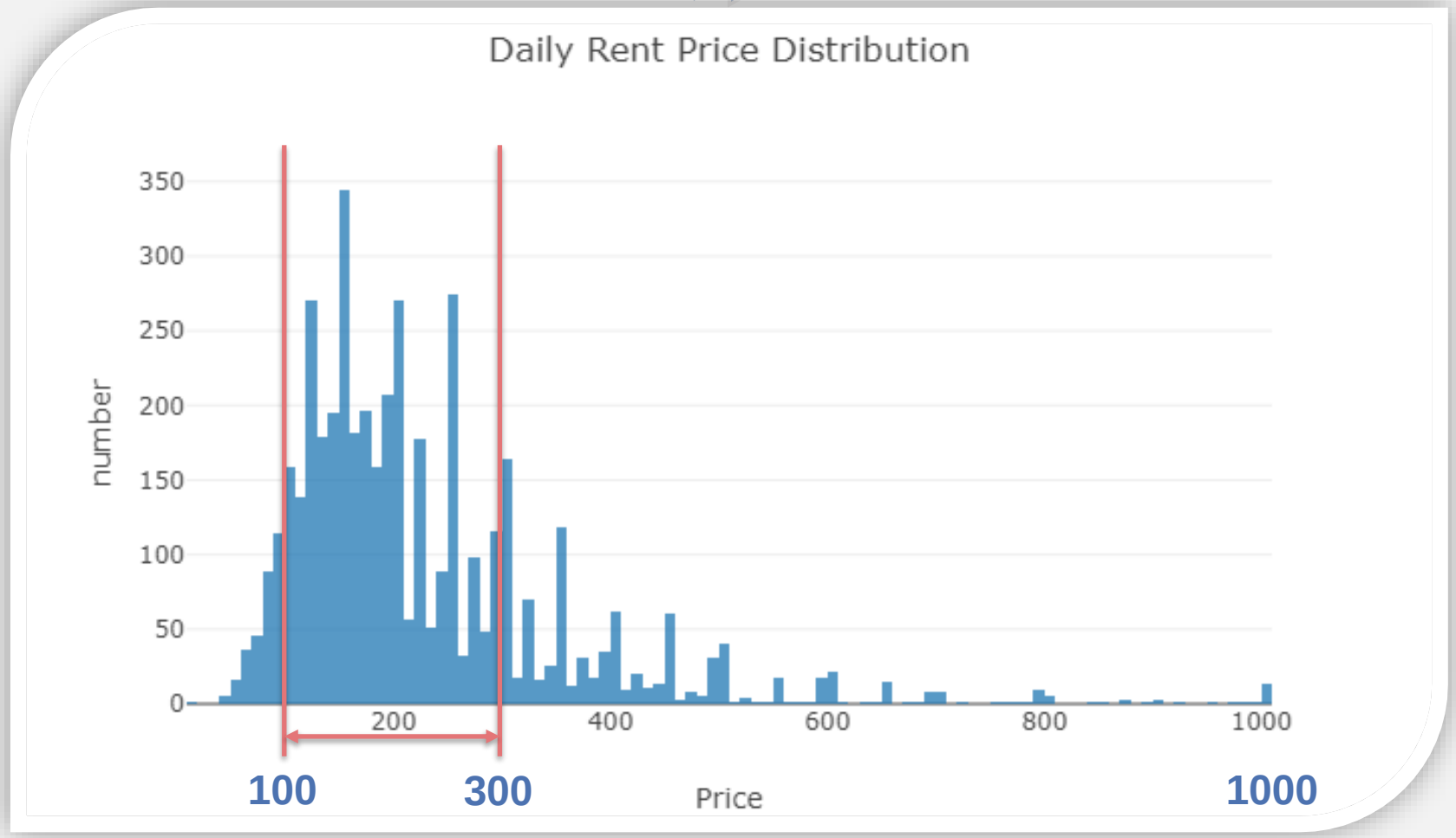
Zillow Dataset

- Select zip code in New York City using USPS NYC zipcode list rather than using City to choose the zipcode to make sure our selection is accurate.
- Delete irrelevant columns and only keep the columns of zipcode and property price column of 2007-07 ~ 2017-06.
- No aggregation needed for the zipcode because every record has a unique zipcode.

Merging Dataset

- Inner join the two datasets on zipcode and finally got 22 zipcodes.

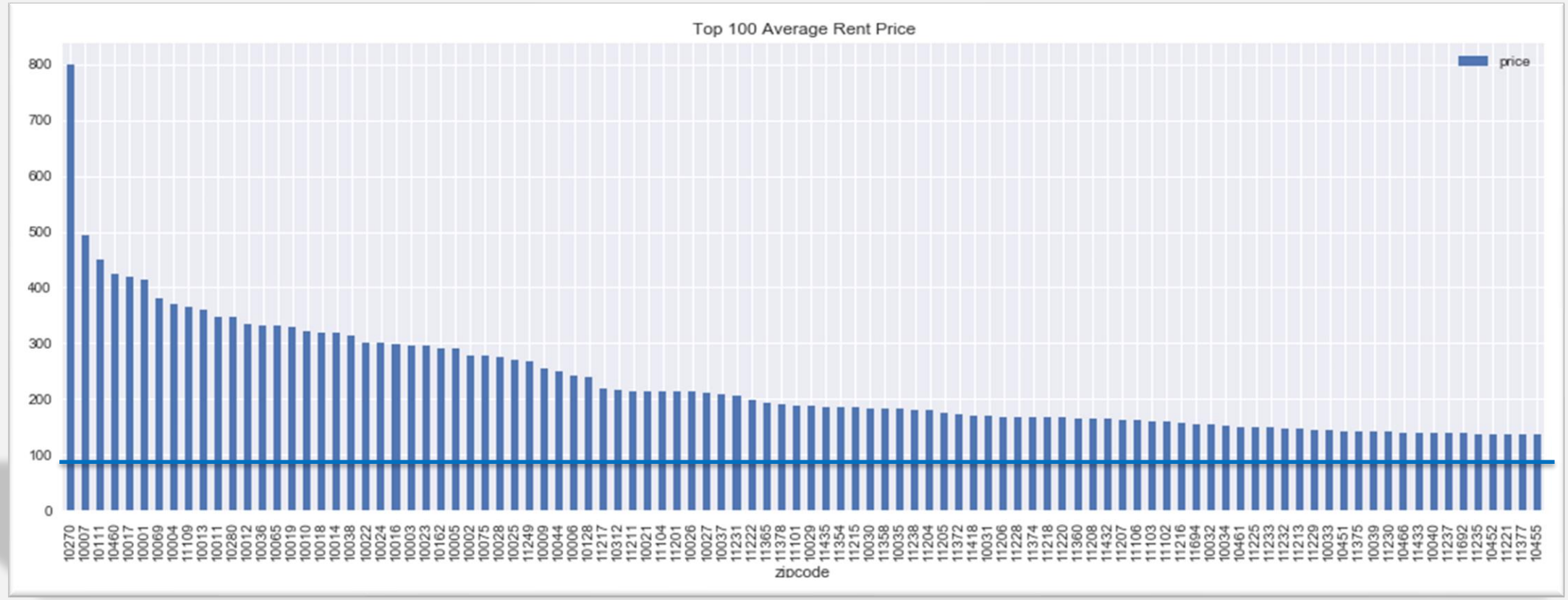
Dataset Visualization



Most: $100 ~ $300
Highest: $1000

Figure 1 NYC Daily Rent Price distribution

Dataset Visualization



$800

$100

Figure 2 NYC Top 100 Zipcodes with Highest Average Daily Rent Price

1st Highest:10270 Wall Street **2st Highest: 10007 Tribeca** **3st Highest: 10111 Rockefeller Center**

Dataset Visualization

- **Generally increasing**
- **Higher price increase faster**
- **Decrease in 2010 because the influence of financial crisis**

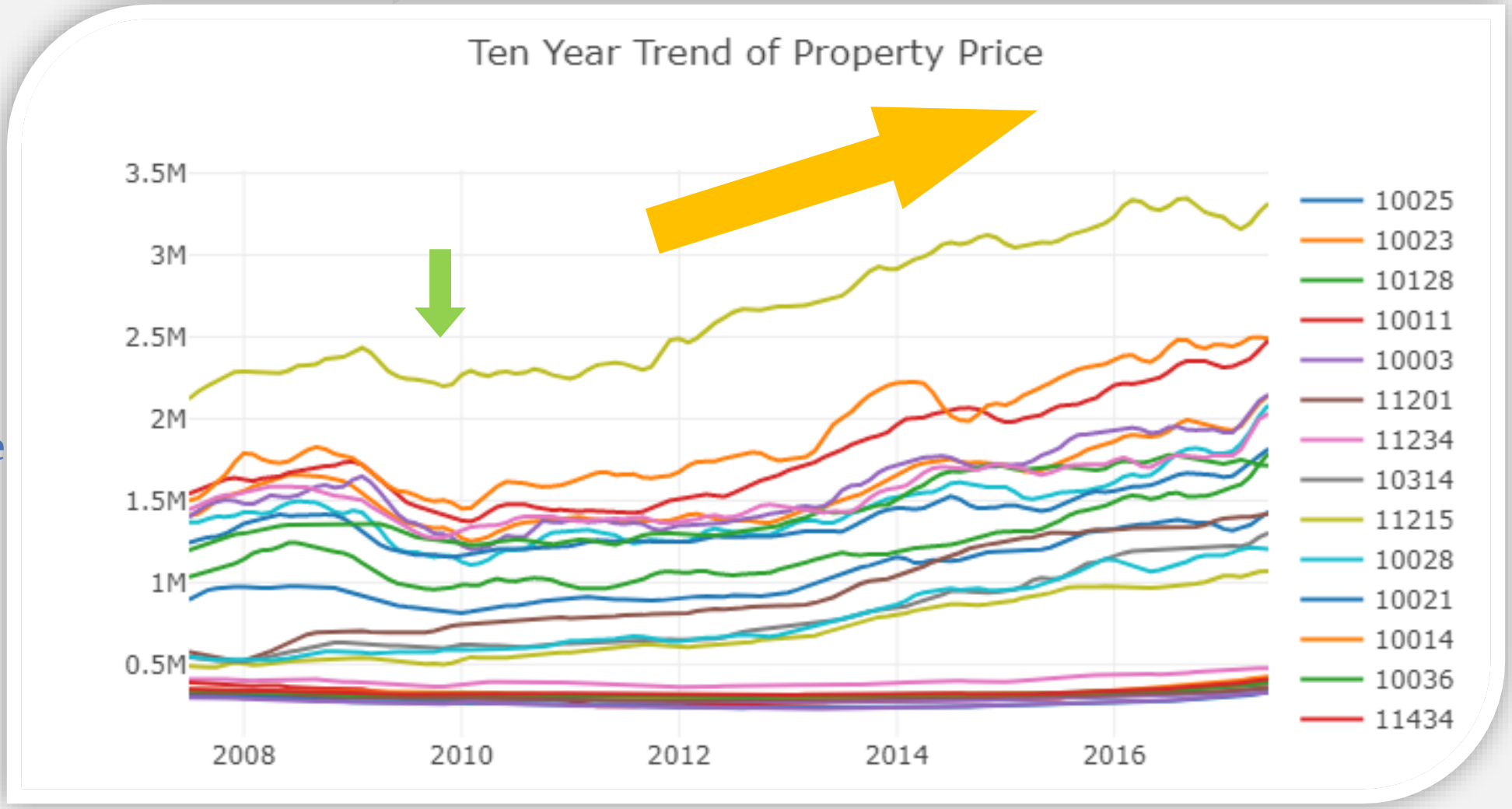
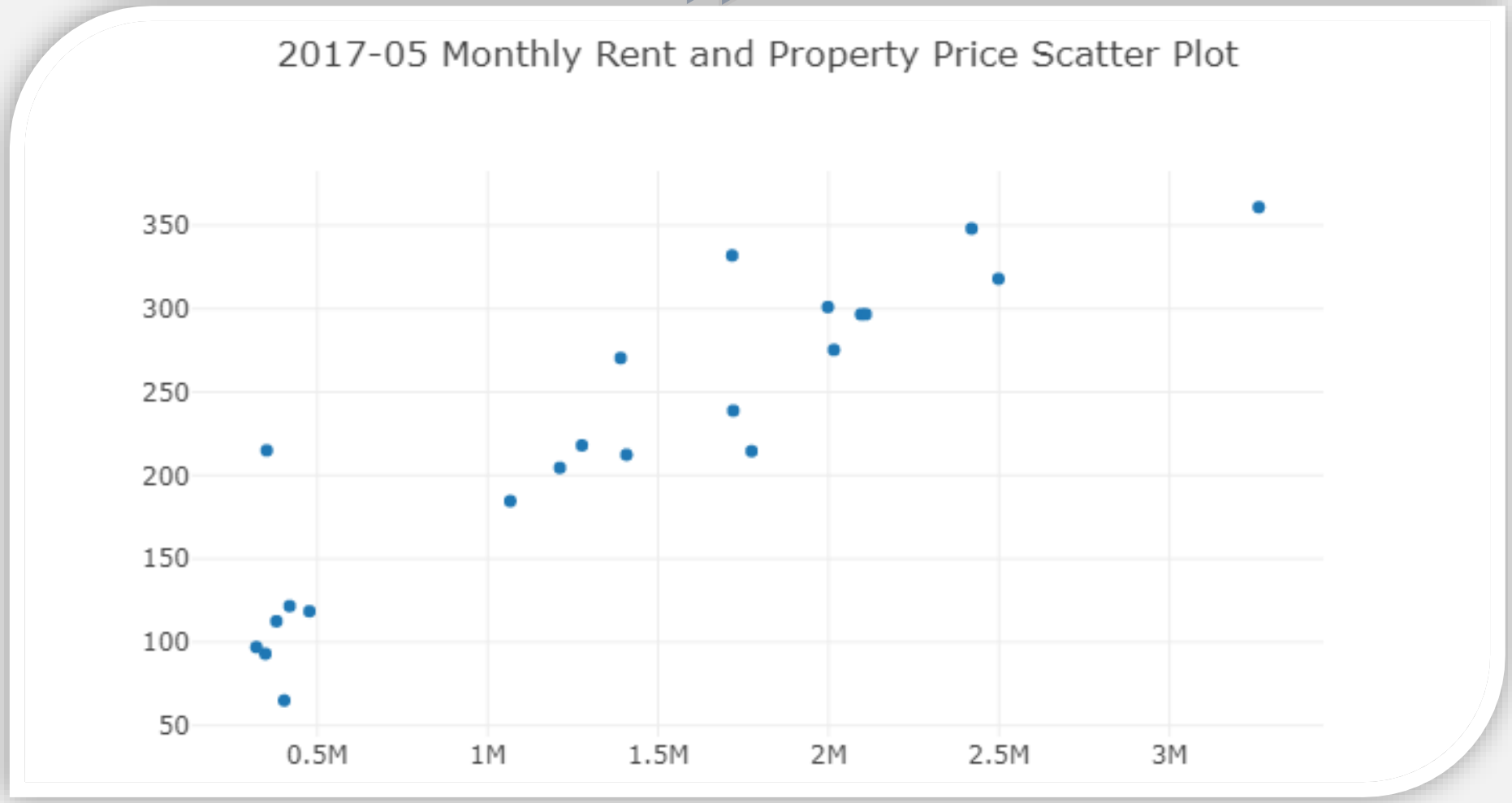


Figure 3 Ten Year Trend of Property Price for all Zipcode

Dataset Visualization



Price

Rent

Figure 4 2017-05 Monthly Rent and Property Price

Model Methodology

Problem Break-down

1. Measure the profitability and define a metric. According to the metric we can get the top 5 zipcodes to bid for our client's investment.
2. Analyze the property price of the target zipcodes in the long-term range to see the potential of the property appreciation within that zipcode area so that we will know the trend of the investment cost.
3. Analyze the rent price of the target zipcodes and combine with other demographic factors to see the potential of future rent.
4. Select the zip code with the highest investment worth based on the analysis above.

Assumptions for Analysis

The first four assumptions come from the client:

- The investor will pay for the property in cash (i.e. no mortgage/interest rate will need to be accounted for).
- The time value of money discount rate is 0% (i.e. $1 today is worth the same 100 years from now).
- All properties and all square feet within each locale can be assumed to be homogeneous.
- The renting occupancy rate in NYC is 75%.

The other assumptions for the analysis are as follow:

- The renting price and property price are in accordance with the date it shows with no delay or mistake.
- The rent price in AirBnb is representative in all the renting market and the property price in Zillow is representative in all the house purchasing market in NYC.
- The rent price for each property remains same within the month.
- The total revenue for the investment mainly comes from the rent, other earnings from the rent process is not counted in this case.
- The total cost of the investment mainly comes from the purchasing of the property, other costs and losses will not count in this case.
- The days within each month is simplified to 30 days.

Define Metric

Rent-to-Price ratio = Total revenue within the month/Total cost

It means the revenue they earned from monthly rent for each dollar they invested in the property. So the higher the ratio, the higher the profitability.

The total revenue within the month is the monthly rent of 2017-05, which is calculated as:

*Monthly Rent = Daily Rent*Days within a Month*Occupancy Rate*

*= Daily Rent*30*0.75*

The total cost is the property price of 2017-05.

Therefore, the optimal formula is:

*Rent-to-Price ratio = Daily Rent*30*0.75/Property Price*

Model Results & Insights

Insight of Rent-to-Price Rate

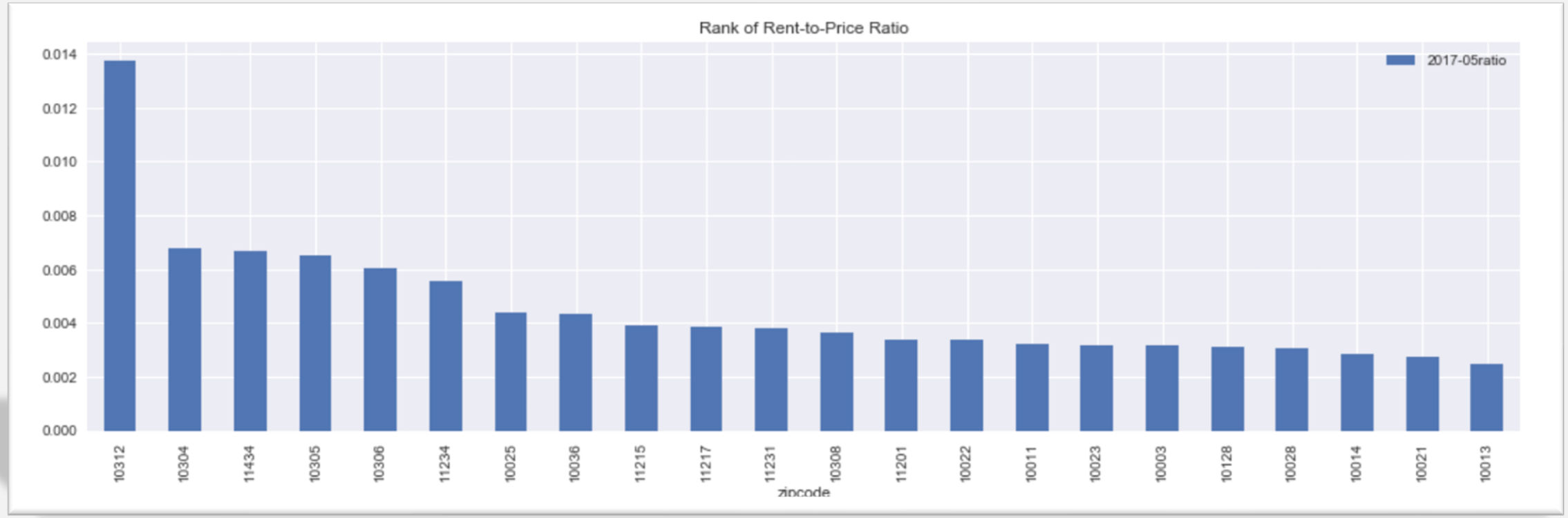


Figure 5 Rank of Rent-to-Price Ratio

Top 5: 10312, 10304, 11434, 10305, 10306

Insight of Rent-to-Price Rate

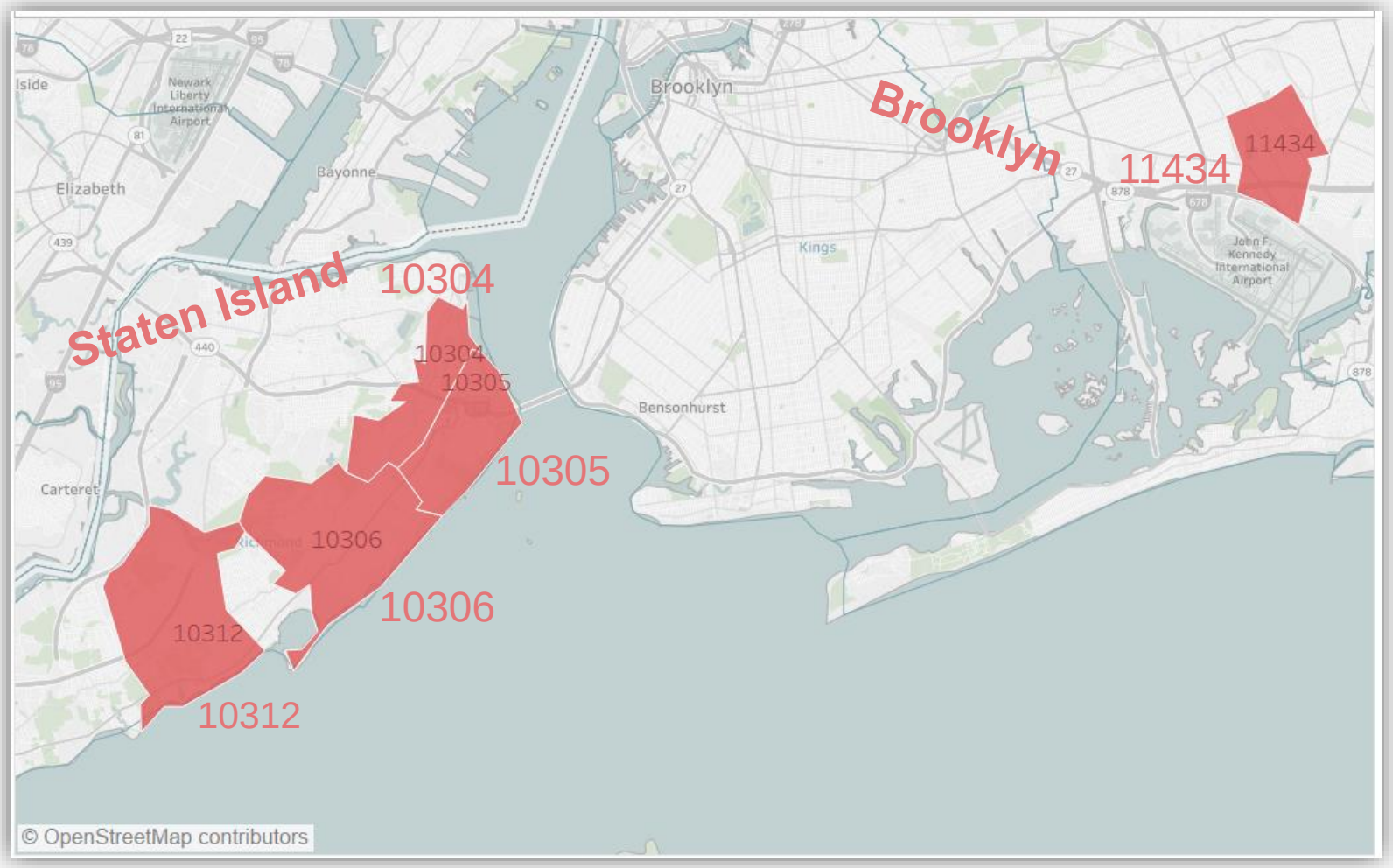


Figure 6 Geographic Distribution of the Five Zipcodes

Insights of Property Price

Within Low Property Price Area

Relatively Low Investment Cost

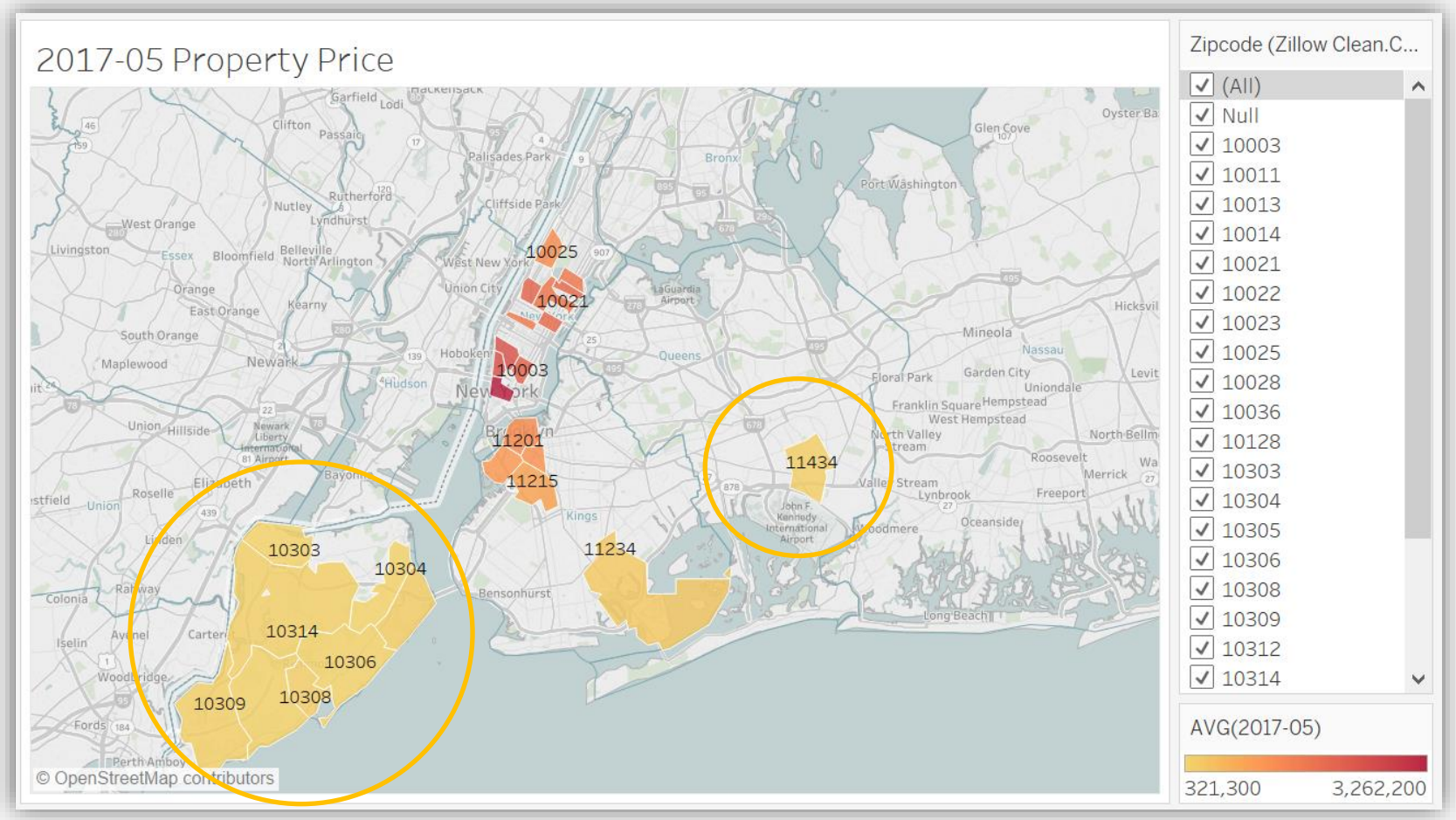


Figure 6 Property Price Level of the Five Zipcodes

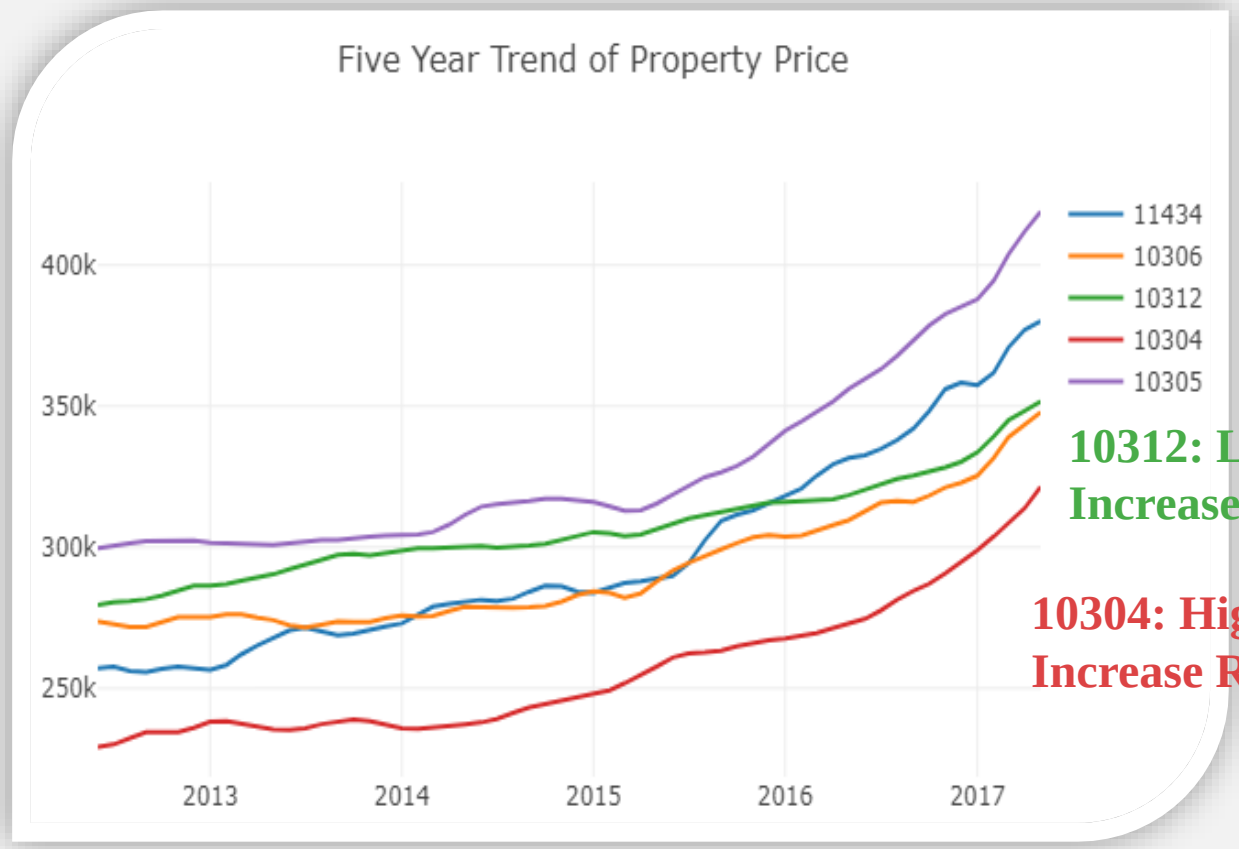
Insights of Property Price

10312: Median

10304: Lowest

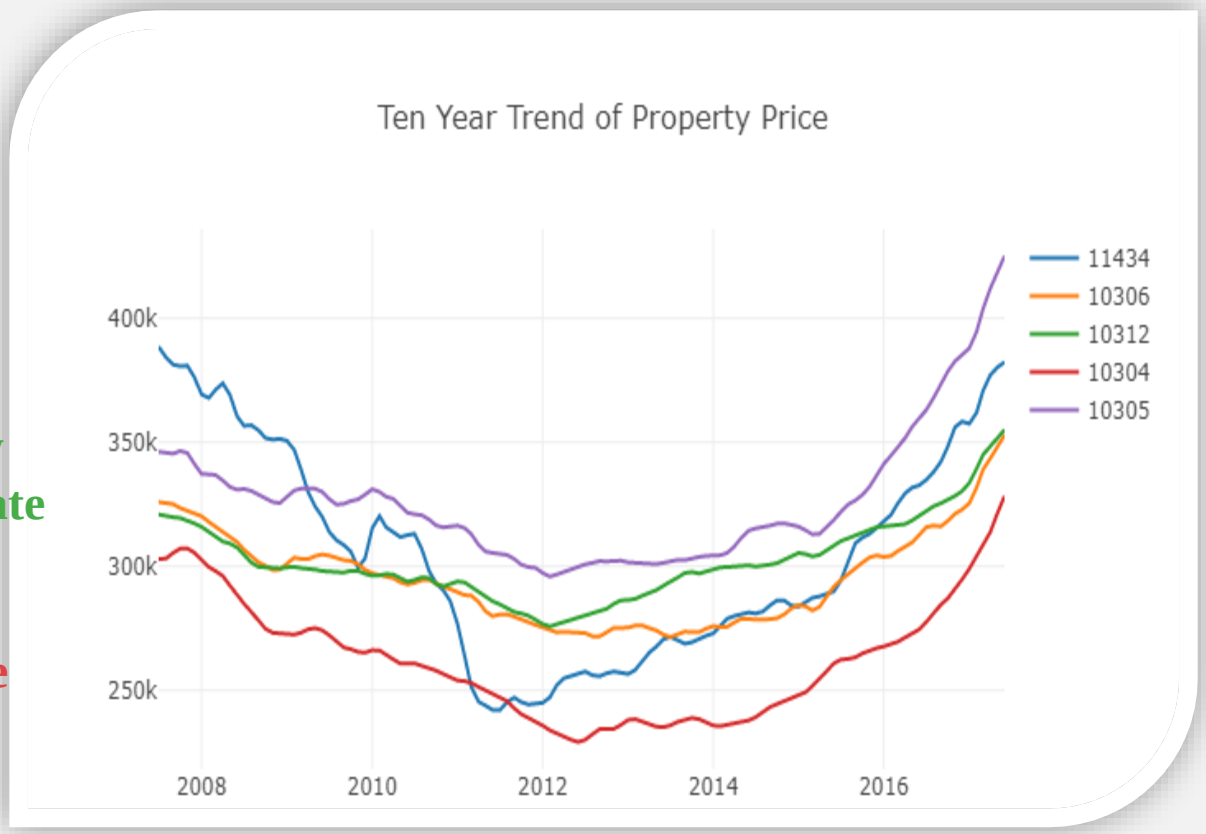
Figure 6 One Year Trend of Property Price

Insights of Property Price



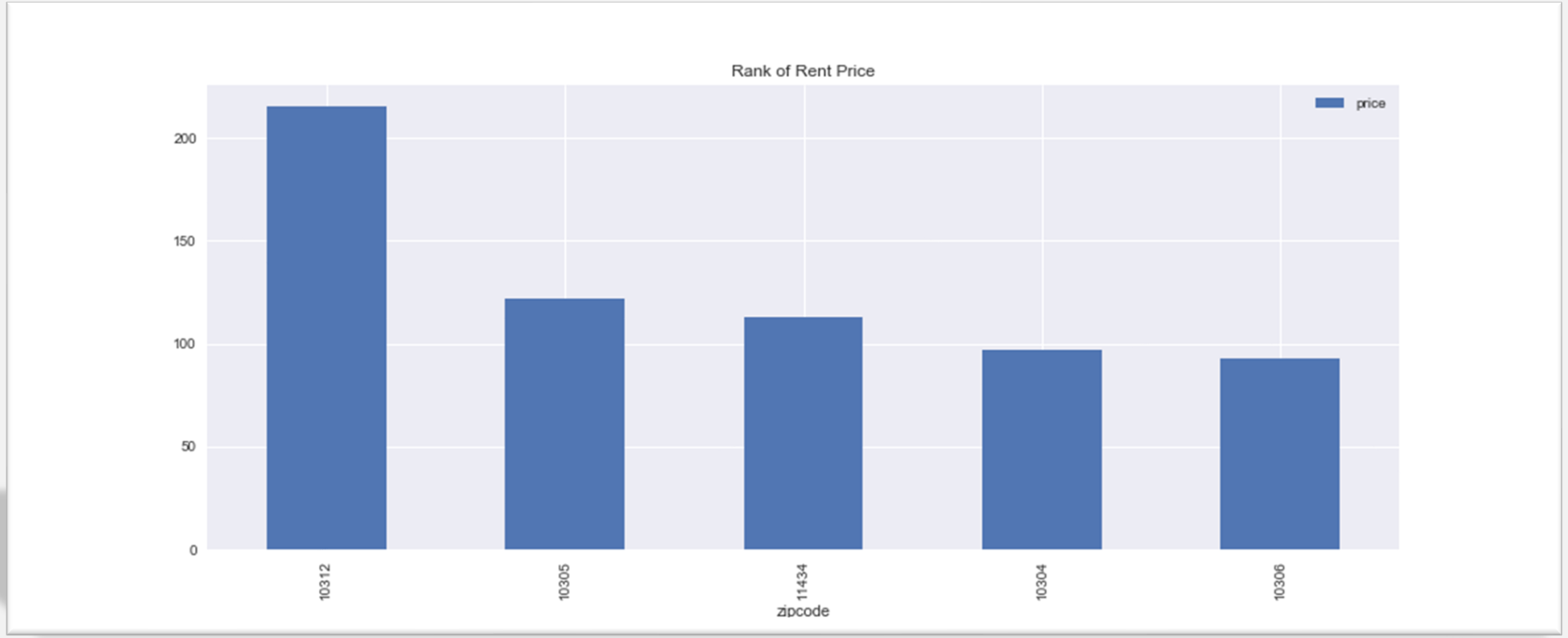
10312: Low Increase Rate

10304: High Increase Rate



10312: Low Investment Cost & Low Increase Rate. Good for investment!

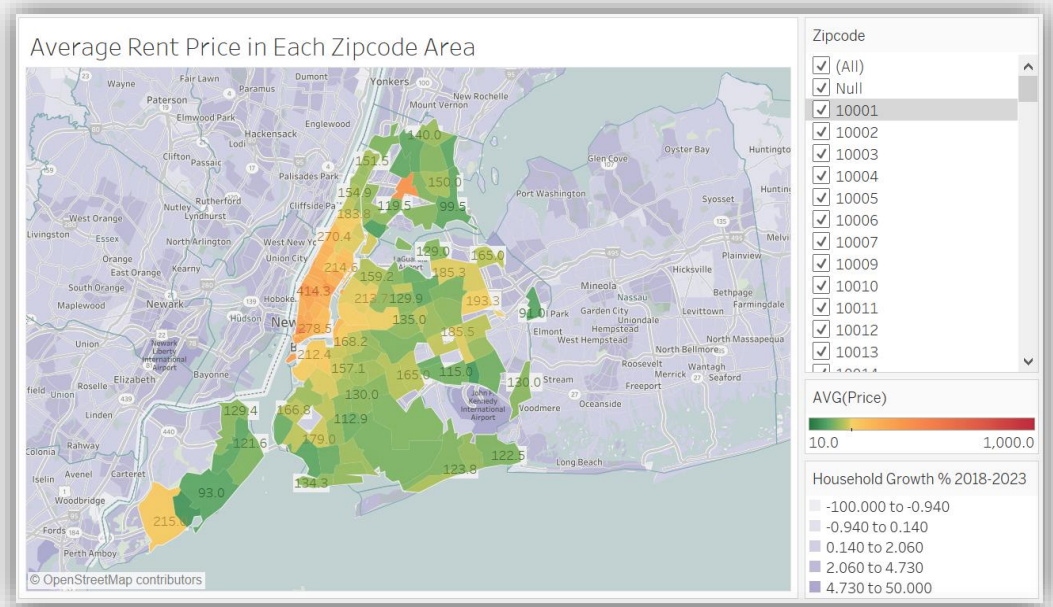
Insights of Monthly Rent

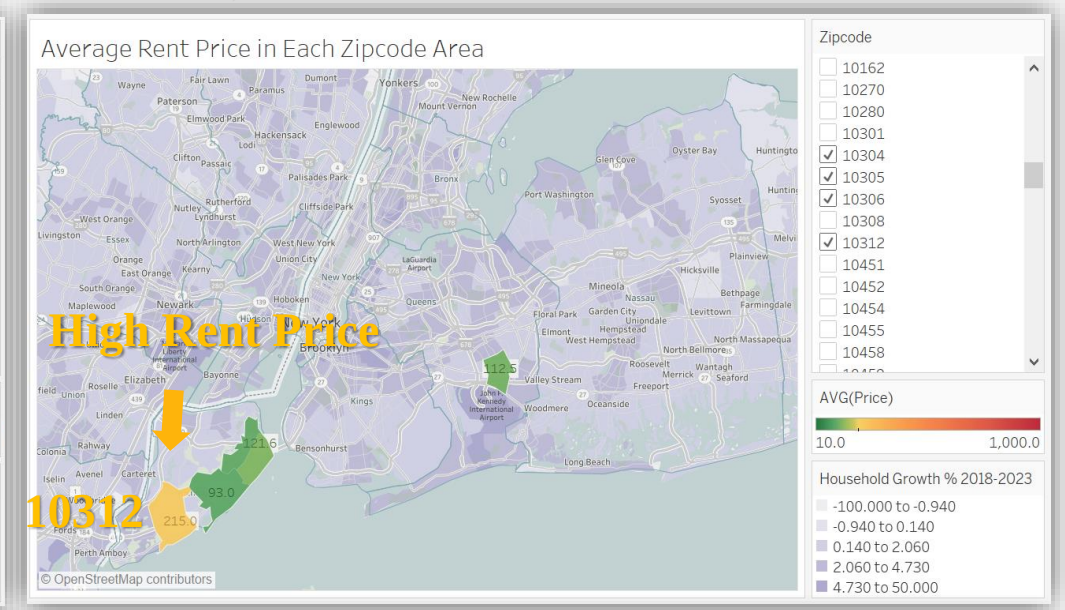


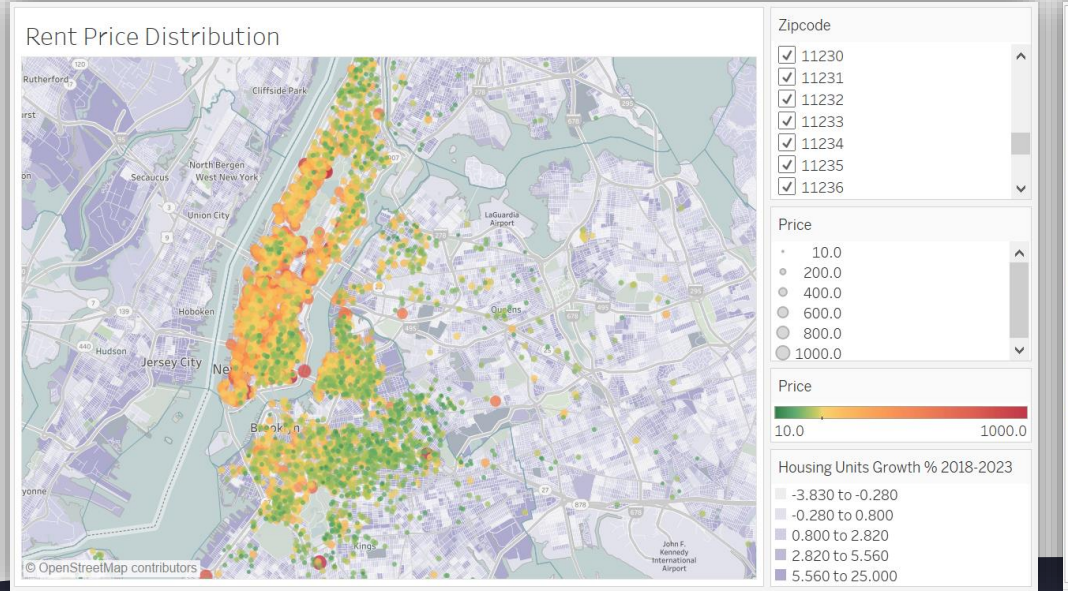
Highest: 10312. Also the only one of the top five in the "NYC top 100 zipcode with highest rent price"

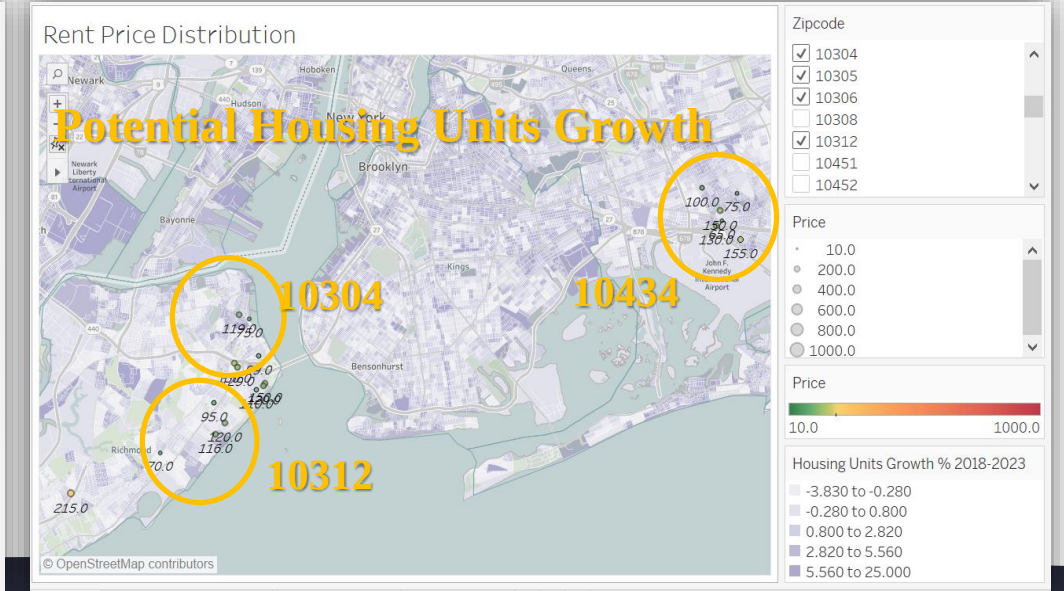
Good for Investment!

Insights of Monthly Rent









10312: High Revenue & Potential Market Growth. Good for investment!

Best Zipcode to Invest

10312

- *Low Investment Cost*
- *low Cost Increasing Rate*
- *High Revenue*
- *Potential Market Growth*

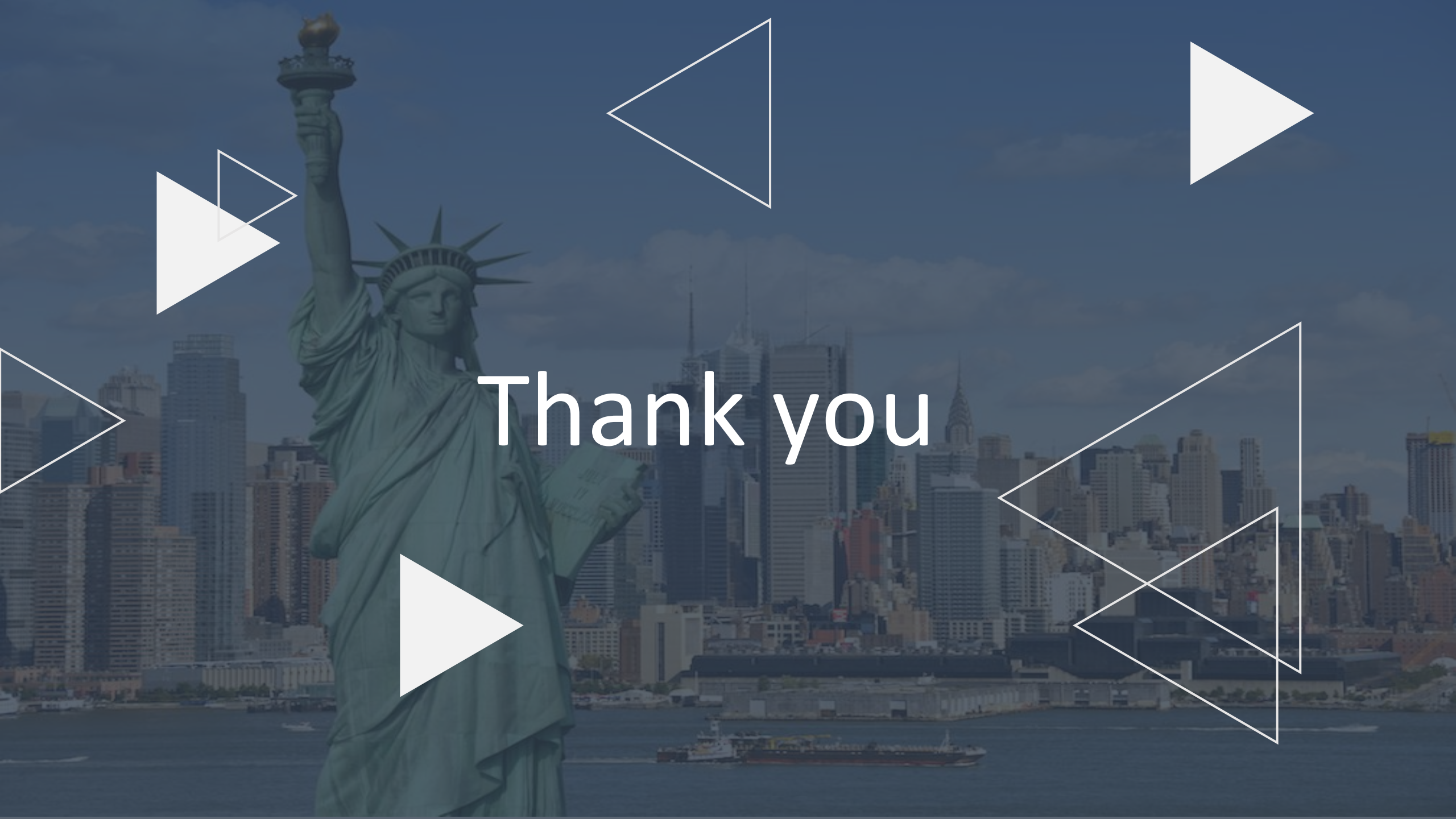
Further Analysis

Further Analysis

- Time Series Analysis for Price-to-Rent Ratio to evaluate an area's long-term investment potential if having the data of the historical rent price.
- Get external economy and social science data and combine with property feature's data to build machine learning models to investigate the important factors that will affect property price & rent price, so that we are able to better evaluate an area's overall investment potential.
- Competitive Analysis for the investment market.

Appendix

- Tableau Public https://public.tableau.com/profile/xinrantao#!/
- Main Code
- Web Scrapping code
- Analysis Report


Thank you