Kasey Suszko

Business Understanding

The purpose of this study was to analyze the datasets 'Zip_Zhvi_2bedroom.csv', containing historical cost data for two-bedroom properties from Zillow, and 'listings.csv' from AirBnB, containing revenue data for properties listed on the site. A model was created to determine which zip codes in New York City would be best for a niche real estate company looking to facilitate short-term rentals to invest in.

Data Wrangling

Quality Check

- Some of the columns needed to be cleaned
    - Pricing columns in listings such as "price", had "$" sign in front of them and were strings
    - The city column in listings has different formats for "New York City"
    - The State column in listings has different formats for New York
- Meta data created and in table below

Data Munging

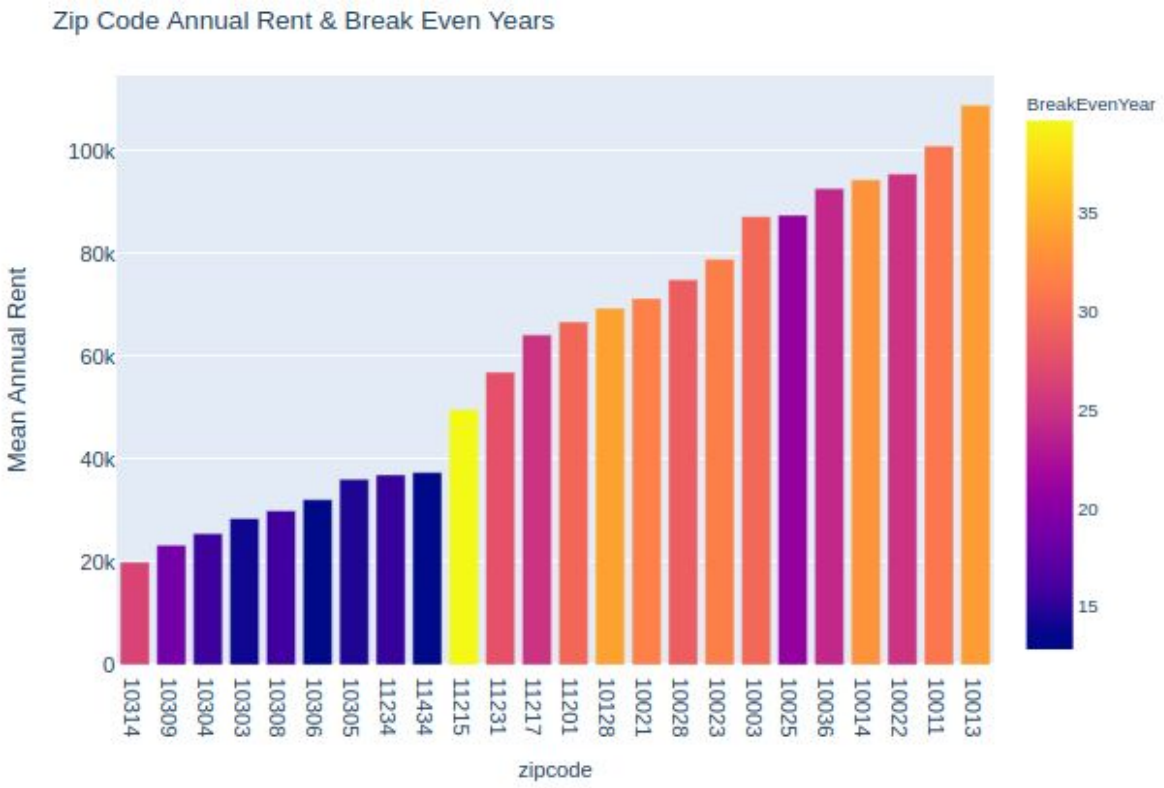
- In order to merge data, a common column had to be used to unite the data
    - I used "zipcode" for an inner join when merging
        - This allowed there to be no null values for zipcode, which was used for most data groupings
        - I do understand that I may have done this correctly. I sorted the data in Data_Processing_Zillow as the "city" being "New York", but at the end of my analysis I realized that I possibly should have included other cities being that not all zip codes in New York City may have been listed with " New York" as their cityu

Metadata		
Variable	**Formula**	**File**
AR_(Year)	ARIMA Forecast For Year	Data_Processing_Zillow
CAGR_3	CAGR For 3 years	Data_Processing_Zillow
Year_CAGR_3	Calculated Forecast With CAGR	Data_Processing_Zillow
price_equiv	Price value formatted	Data_Processing_Airbnb_Combination_ARIMA

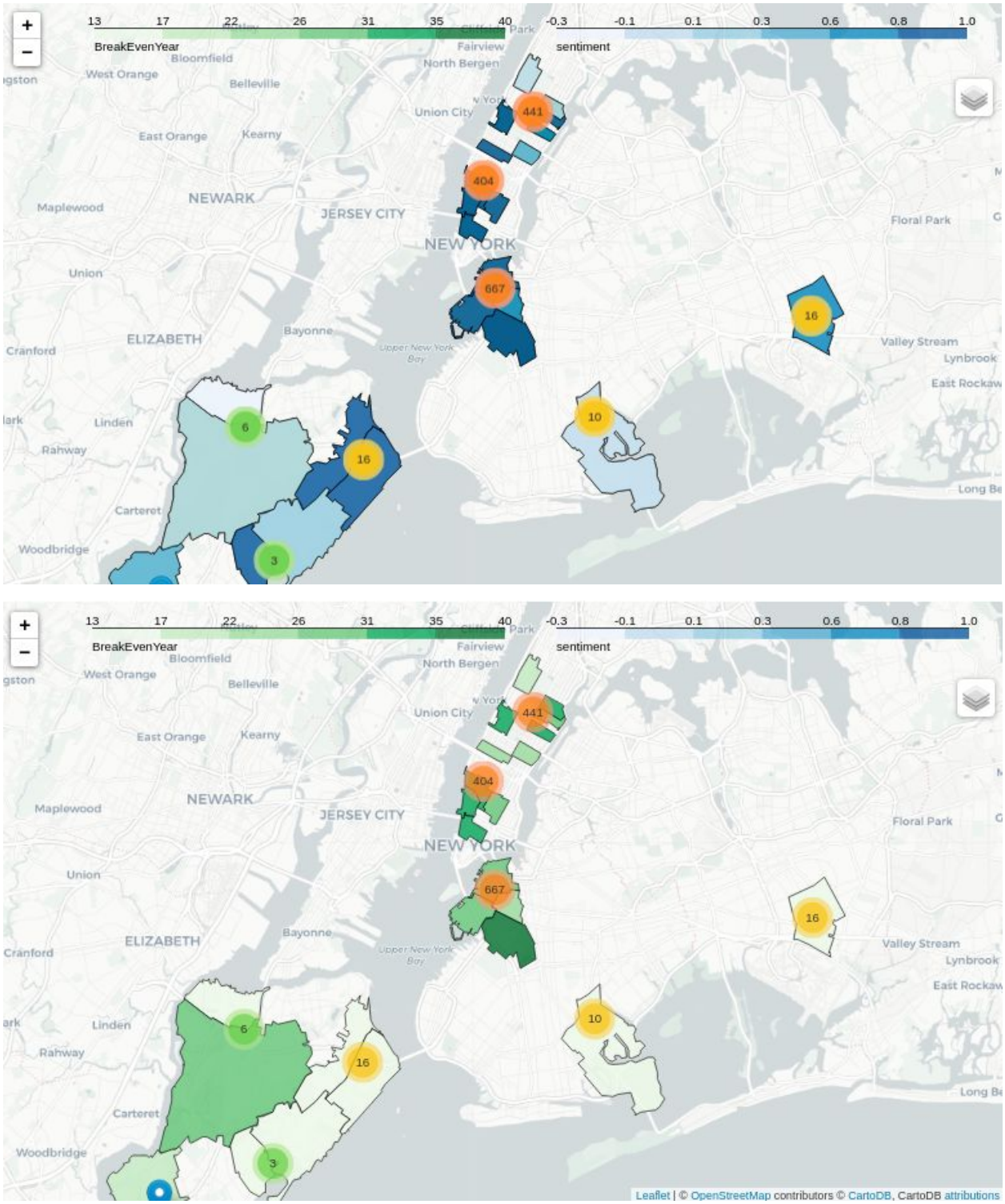
BreakEvenYear	BreakEvenYear= 2020 value / (0.75 * daily_price * 365)	Data_Processing_Airbnb_Combination_ARIMA
AnnualRent	(0.75 * daily_price * 365)	Data_Processing_Airbnb_Combination_ARIMA
sentiment	Vader Sentiment Score for summary of airbnb listing	Data_Processing_Airbnb_Combination_ARIMA
'Year'	Rental Profit - Property Cost at Purchase + Property Price if to sell that year	Data_Processing_Airbnb_Combination_ARIMA

Visualizations + Insights

Above, it can be seen that the 5 zip codes with the lowest breakeven years are 10036, 11434,10303,10305, and 11234. The average sentiment scores for airbnb rentals in the zip code is provided, giving light to which neighborhoods renters had the best experience in. Of the top 5 zip codes, 11434 and 10303 had the highest positive sentiment. Therefore, if the real estate company is aiming for a low break even year (short term investment) and security in rental prospects as indicated by renters' liking of the region, **I suggest investing in 11434 and 10303**.



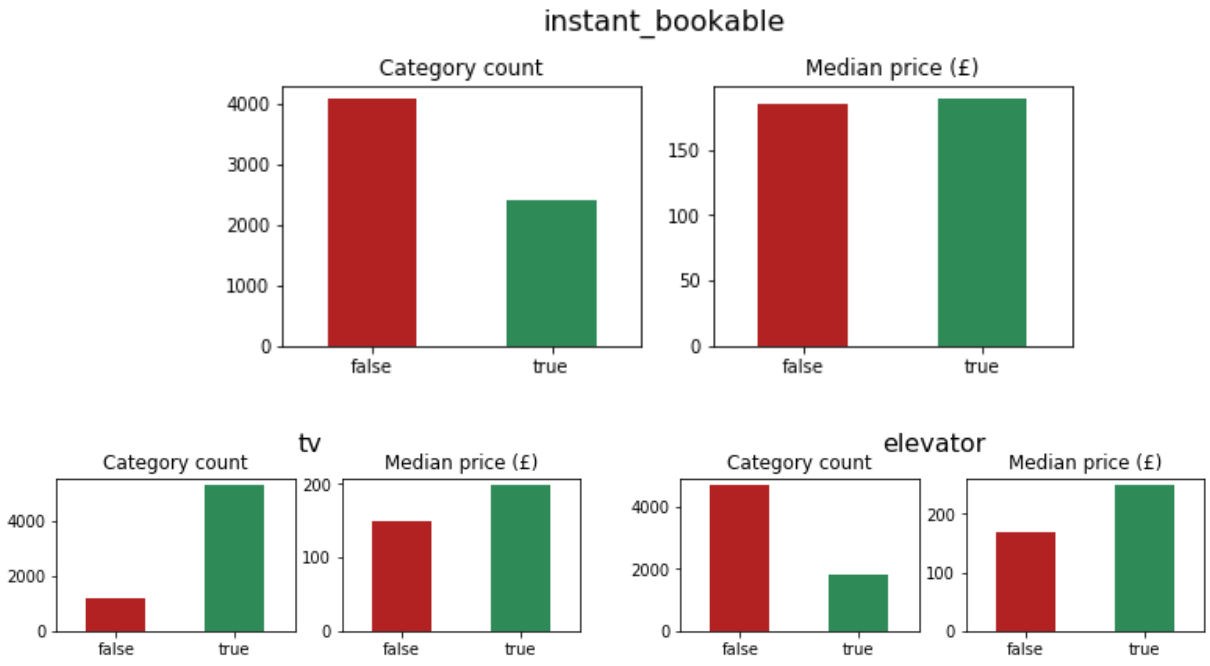
Above, it can be seen that the 5 zip codes with the highest mean annual rent are 10013, 10011,10022,10014,10036. The corresponding Break Even Years for these zip codes are lowest for 10022 and 10036. If the real estate company is aiming for a high rental profit and low Break Even Year (medium-term investment), **I suggest investing in zip codes 10022 and 10036**.If the real estate company is looking for their best long-term investment regardless of years to the Break Even Point, **I suggest investing in zip codes 10013 and 10011.**

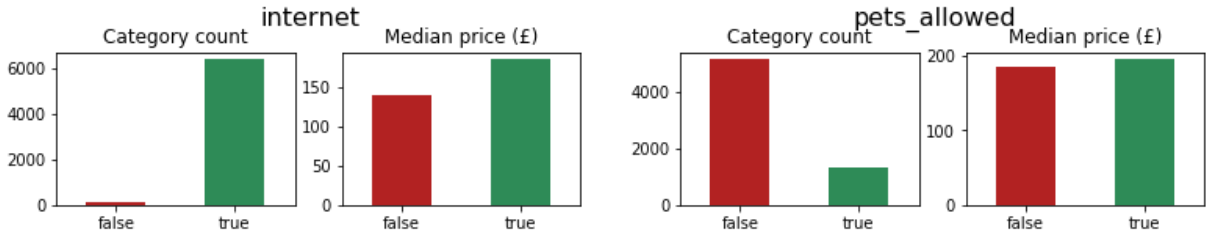


Above, interactive choropleth maps are shown for the zip codes. These maps are colored by Break Even Year and sentiment scores, but going to the individual airbnb listings shows information on that properties long term investment potential, with predicted overall profit for 2020, 2025, and 2030 if

the company were to invest in the property today. This interactive feature is useful for giving to the real estate company to select properties they would potentially be interested in.

Above, the change per year in the nightly price of Airbnb listings in New York City can be seen. They remain constant over time, which justifies assuming that the rent will be constant over time.





Above, different amenities and components of Airbnb listings along with their representative impact on price are plotted. It can be seen that the listing being instant_bookable, having TV, an elevator, internet, and pets_allowed have the highest positive impact on the price of the property. **I suggest that the real estate company make sure that any properties that they invest in have the amenities listed above in order to optimize the rental price.**

Insight Summary

For Short Term Investment: I suggest investing in 11434 and 10303
For Medium Term Investment: I suggest investing in zip codes 10022 and 10036
For Long Term Investment: I suggest investing in zip codes 10013 and 10011
Property and Listing Should Include: instant_bookable, tv, elevator, internet, pets_allowed

What's Next

The final data table that I used in my analysis in "Data_Processing_Airbnb_Combination_ARIMA" contains several variables which were not used. These variables were included because I had originally intended to perform the following analysis below. I suggest that this analysis be performed in the future for an improved product to give the client.

1. Create neural network to predict airbnb prices
2. Use neural network to predict given zillow property rental prices
3. Plot predicted zillow rental prices and property costs in D3.js interactive choropleth map