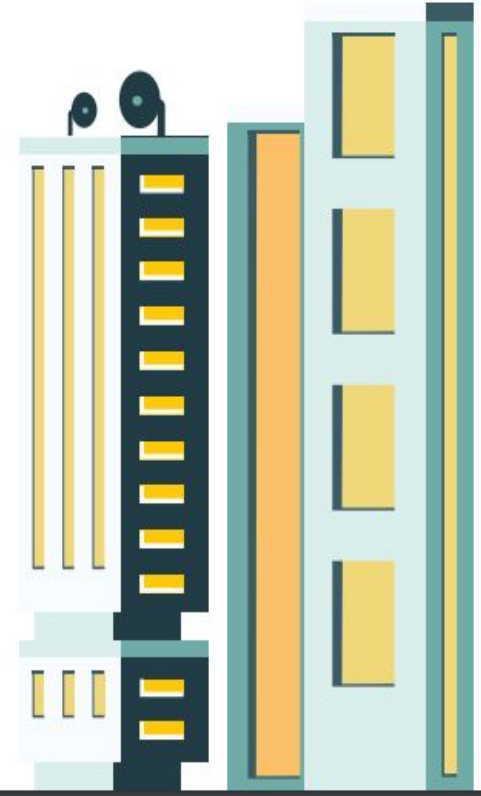


Real Estate Price Analysis

By : Devanshi Shah



Motivation

- Real Estate Price Analysis helps in -
 - Understanding the current trends of the housing market.
 - Track the consumer's buy/sell behavior.
 - Estimating property value based on location and condition.
 - Helping a potential customer decide where to invest/build a property.
 - Predict the most popular areas for buyers and sellers.



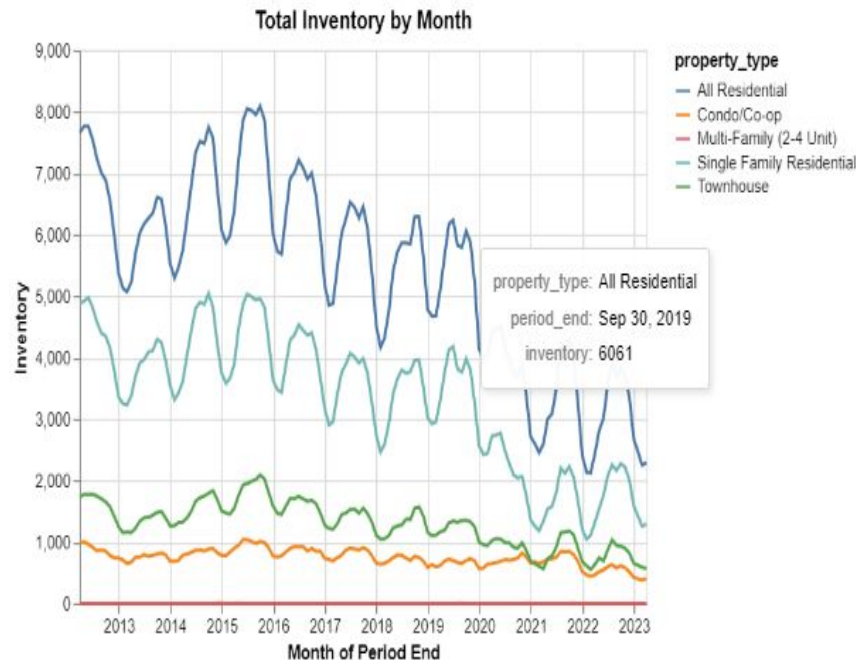
The Dataset

- Used Redfin Zip code dataset for analysis which is dynamic in nature.
- Dataset contains around 6.5 million rows.

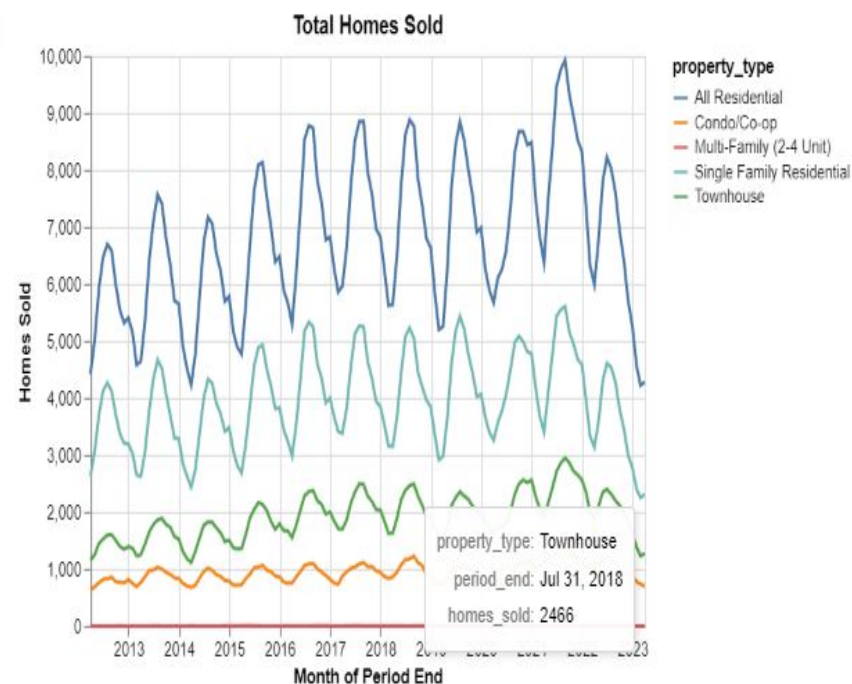
```
url = 'https://redfin-public-data.s3.us-west-2.amazonaws.com/redfin_market_tracker/zip_code_market_tracker.tsv000.gz'
```

	period_begin	period_end	period_duration	region_type	region_type_id	table_id	is_seasonally_adjusted	region	city	state	state_code	property_type	property_ty
0	2022-06-01	2022-08-31	90	zip code	2	30725	f	Zip Code: 71923	NaN	Arkansas	AR	All Residential	
1	2019-12-01	2020-02-29	90	zip code	2	21126	f	Zip Code: 49265	NaN	Michigan	MI	All Residential	
2	2012-07-01	2012-09-30	90	zip code	2	34119	f	Zip Code: 78727	NaN	Texas	TX	Townhouse	
3	2012-11-01	2013-01-31	90	zip code	2	13119	f	Zip Code:	NaN	Georgia	GA	Single Family Residential	

Exploratory Data Analysis



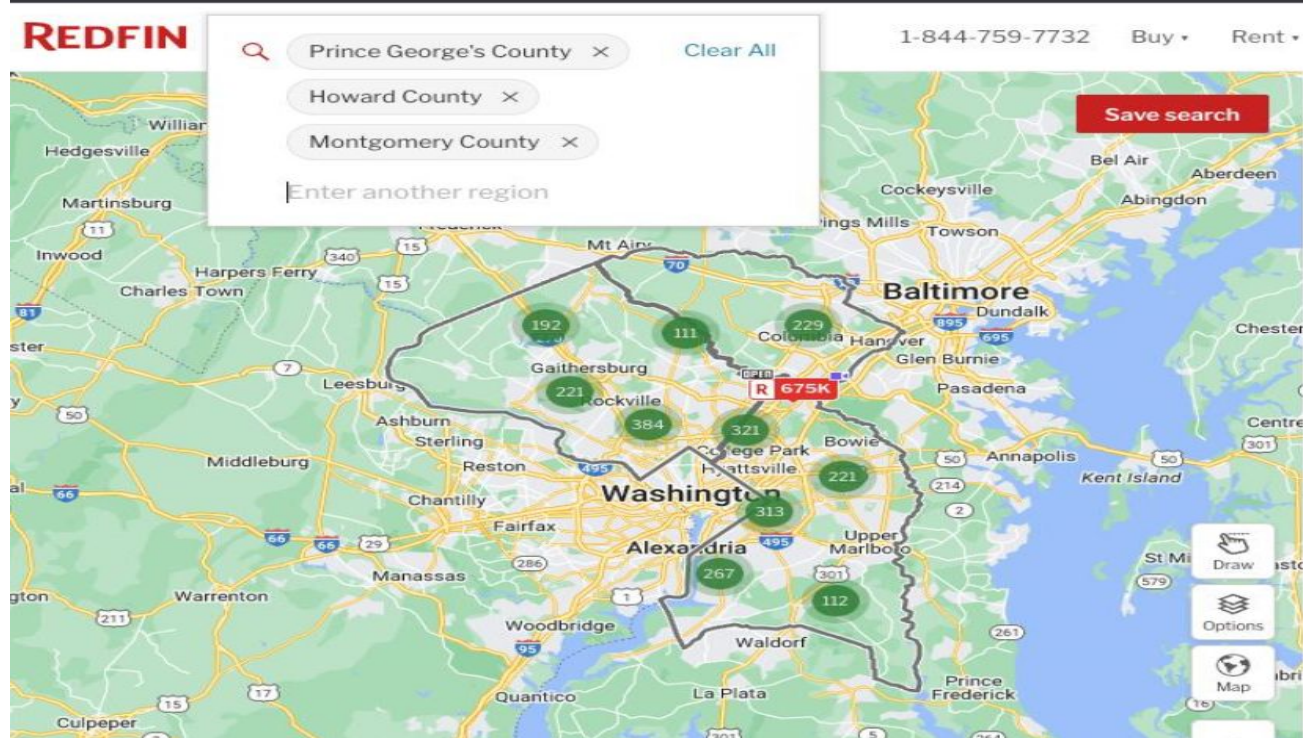
Total Inventory By Month



Total Homes Sold by Month

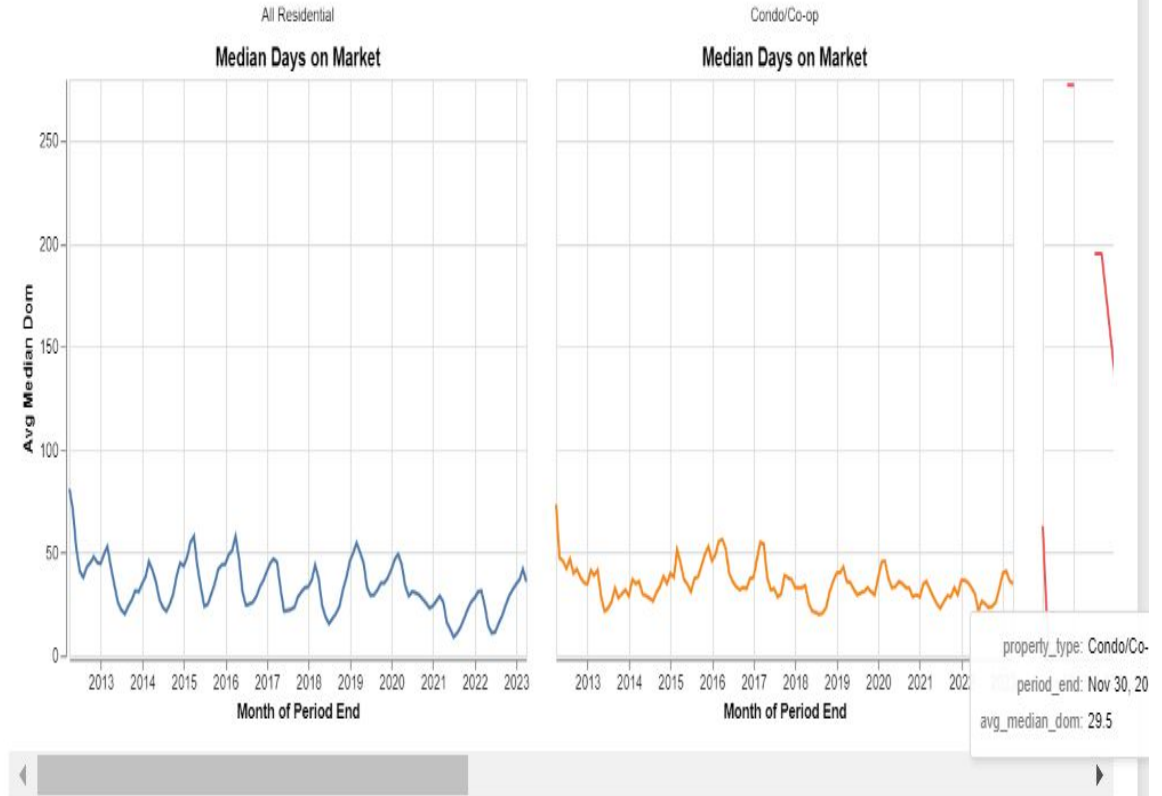
The Map

- Focused on Maryland(PG, Montgomery and Howard County) for analysis.



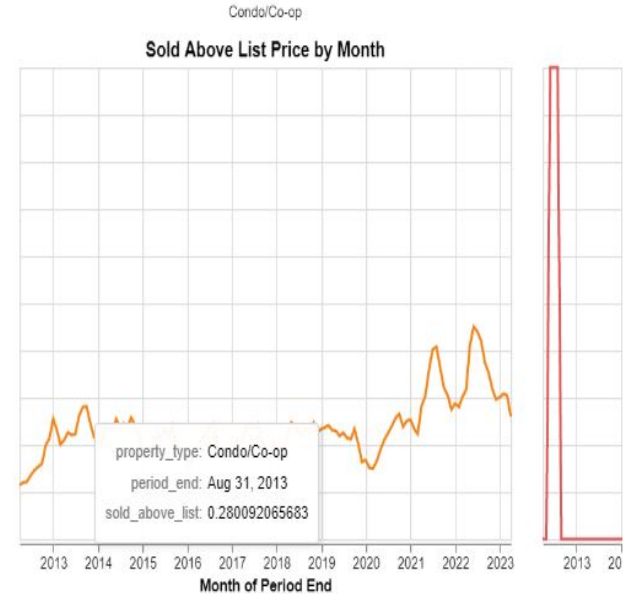
Median Days on Market By Month

- Properties stay for around 60 days in market after they are listed for sale.
- All property types stay in the market for around 40 days.
- Multi-family properties are the least popular property type for customers, staying in the market for around 80 days.
- The trend of days on market is declining.



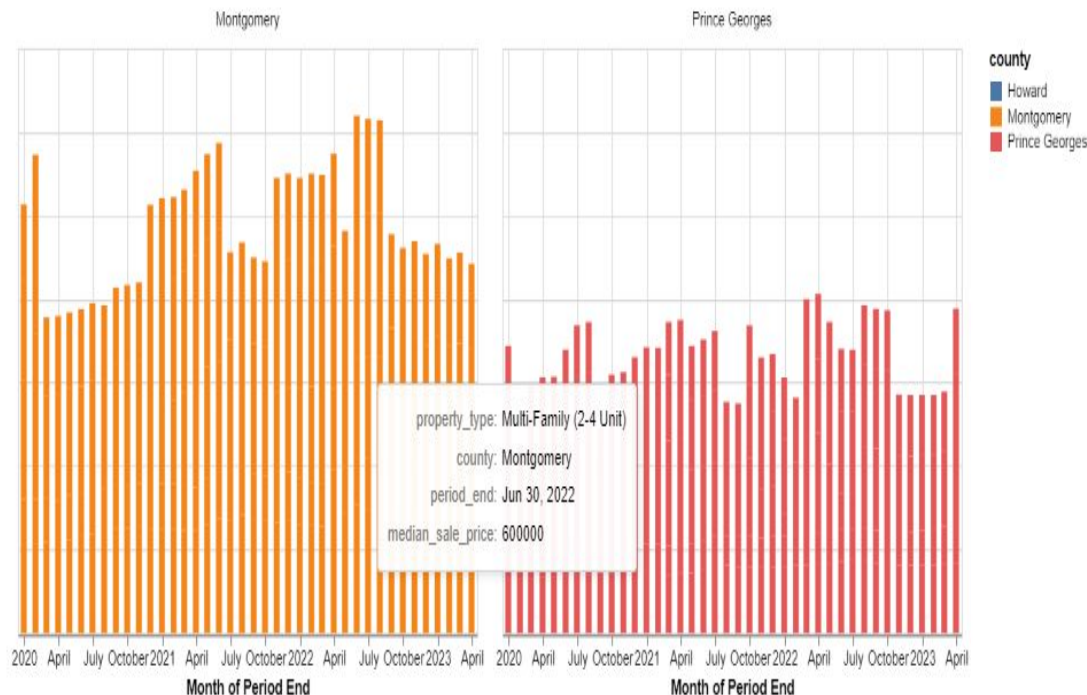
Sold Above List Price By Month

- Initially around 2013, there was a modest 24% increase in property prices.
- However, there has been a significant increase in real estate prices in recent years (2022-2023)
- 60% of properties were sold above the listed price during 2022-23



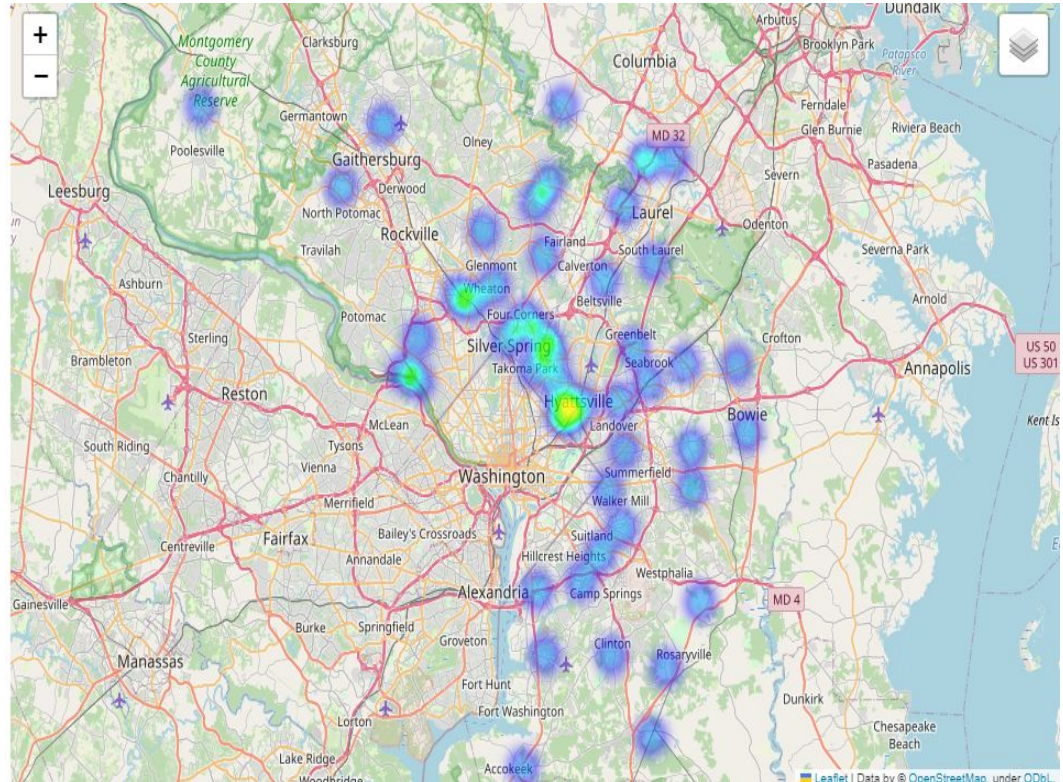
Median Sale Price by Month for different counties after Dec 2019

- Property prices vary across counties
- Montgomery and Howard have higher prices than PG county
- Better schools, hospitals, and lower crime rates could be factors
- Additional data and analysis needed to confirm these factors.



Sale to List Price Ratio Heatmap

- Sale-to-list price ratio varies across different counties.
- Prince George's county has the lowest ratio.
- Montgomery county has the highest ratio
- Darker shade indicates a higher ratio, while the lighter shade indicates a lower ratio.

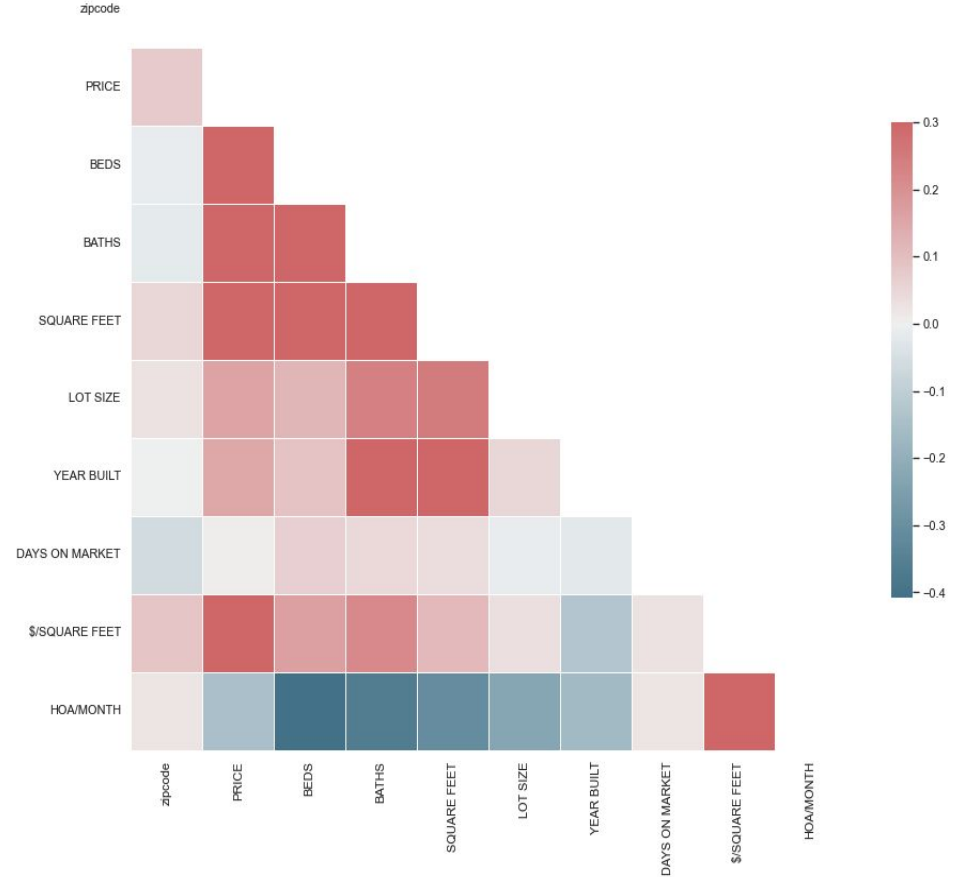


Prediction Modelling

- We were able to gather two different types of data.
- The historical real estate data was used primarily used for Data Analysis.
- Leverage active listing information to predict the list price of a desired property using linear regression.
- Developed two Linear Regression models

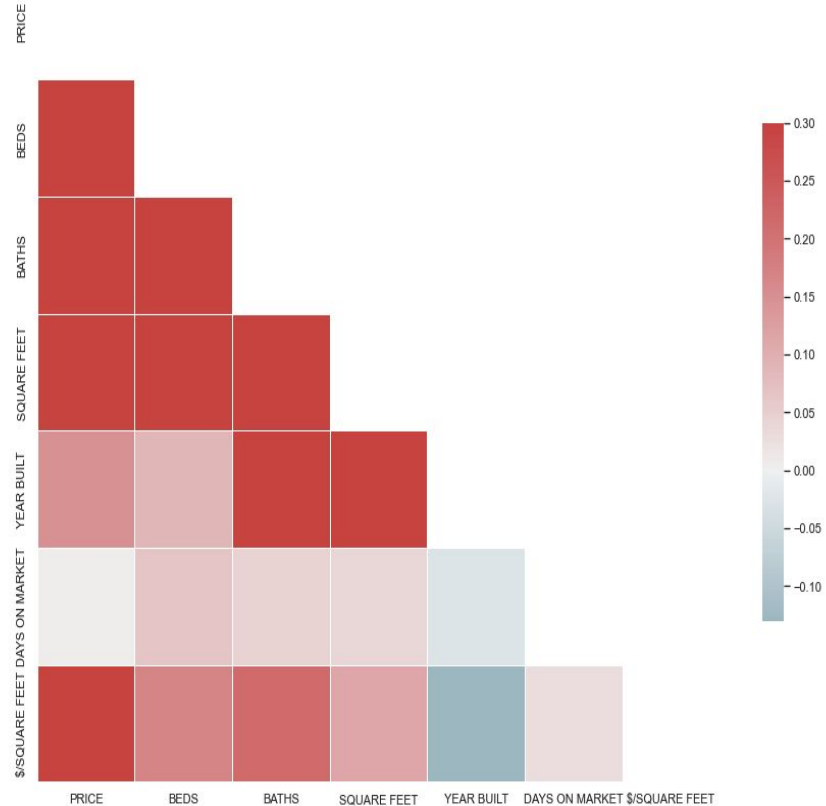
Linear Regression - Naive

- Considered all available features for listings
- RMSE: 1240002.6662823753
- R^2 score:
-6.742976906309445



Linear Regression - Improved

- Considered only numerical features for listings in addition to property type.
- RMSE: 348075.6980953351
- R^2 score:
0.7709235249116182



Challenges

- Difficult to scrape data from the websites like Zillow and Realtor.com as these websites use anti-scraping techniques like captchas, IP blocking to prevent its data from scraping.
 - We used data available from MLS listings from Redfin to solve this issue.
- Limited availability of memory
- Federal Information Processing System (FIPS) Codes for States and Counties FIPS codes are numbers which uniquely identify geographic areas. It was difficult to extract fips code through the Maryland County GeoJson file and add them to the main dataframe.
 - Difficult to create heatmaps with county boundaries.
- Combining data available in different sources.



Future Scope

- Perform time series analysis to determine seasonality accurately.
- Implement forecasting techniques to predict future price trends.
- Add sql querying capability by implementing in-process database like (Duck Db) to access data.
- Develop a web application to replace the python notebook.

THANK YOU !!