Student name: Nargiz Guliyeva
Student number: ▮▮▮▮▮▮▮▮
Group 47
Instructor: Dr. Maher Elshakankiri
Course code: INF1340H
Course name: Programming for Data Science
Program: MI
Faculty of Information
University of Toronto
Final Project
Output file

Intro
Miami housing market historically has been dynamic and affected by the factors such as population growth, job opportunities, international investments and others. The analysis below examines trends in the Miami housing market between 2017 and 2023.

Descriptive

annual_sales.svg
  The line graph illustrates the fluctuation in the volume of homes sold in Miami, delineating changes in different categories: total homes sold, total homes sold with a price drop, new listings, and delisted homes, i.e., unsold home listings that were removed from the market during a given period.
  The graph shows that the number of homes with a price drop and delisted has been historically low, i.e., not exceeding 20,000 homes, and has been declining since 2017, reaching its lowest point in 2023. In contrast, total homes sold and new listing figures have been dynamically changing throughout the years.
  First, the notable disparity in the number of listed homes and total homes sold can be highlighted in 2017. This discrepancy suggests an influx of new listings, illustrated as a green line, outpacing the number of homes sold, illustrated as a blue line. This phenomenon might be attributed to many factors, including increased property development/construction and other factors. However, the demand and supply were getting narrower, peaking in 2021 and significantly dropping in 2023.
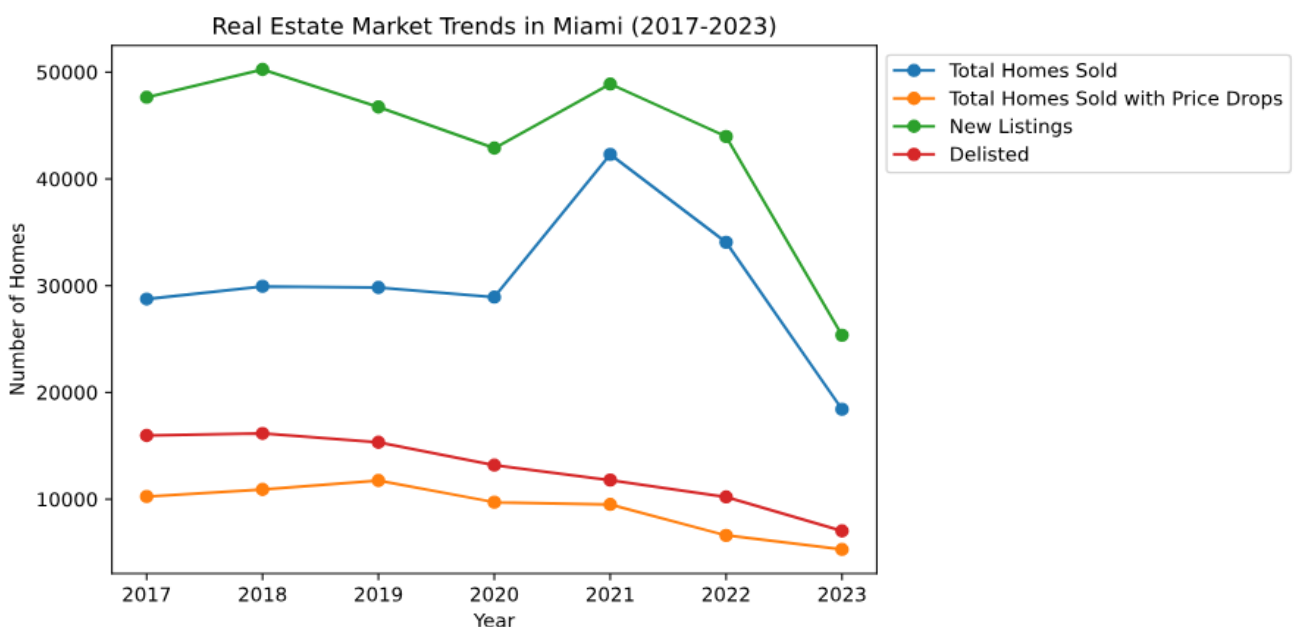
Monthly_means.csv

The table describes the monthly average value of 18 primary variables. When comparing the percent change between January 2021 and August 2023, significant changes can be observed in the averages of all reported variables. Notably, the median sales price has increased by almost 100%, the total homes sold increased by 20%, and the median new listing price has also increased by 65%. Downward trends can be observed in inventory dropped by about 40%, median days on the market decreased by about 30%, and the month of supply dropped by about 50%.

normality.csv

The normality test has been applied to assess whether the key variables follow a normal distribution. In this regard, the table computes the statistic (the degree to which the sample deviates from a normal distribution), p-value (probability of obtaining the observed results if the data were sampled from a truly normal distribution), skewness (measures the asymmetry of the distribution), kurtosis (tail of the distribution), and result. Based on the table, only 'total_homes_sold_with_price_drops' is normally distributed.

outliers.svg

The boxplots display the outliers before and after treatment, done with the cap method. This is an essential pre-step for building and running a predictive model.
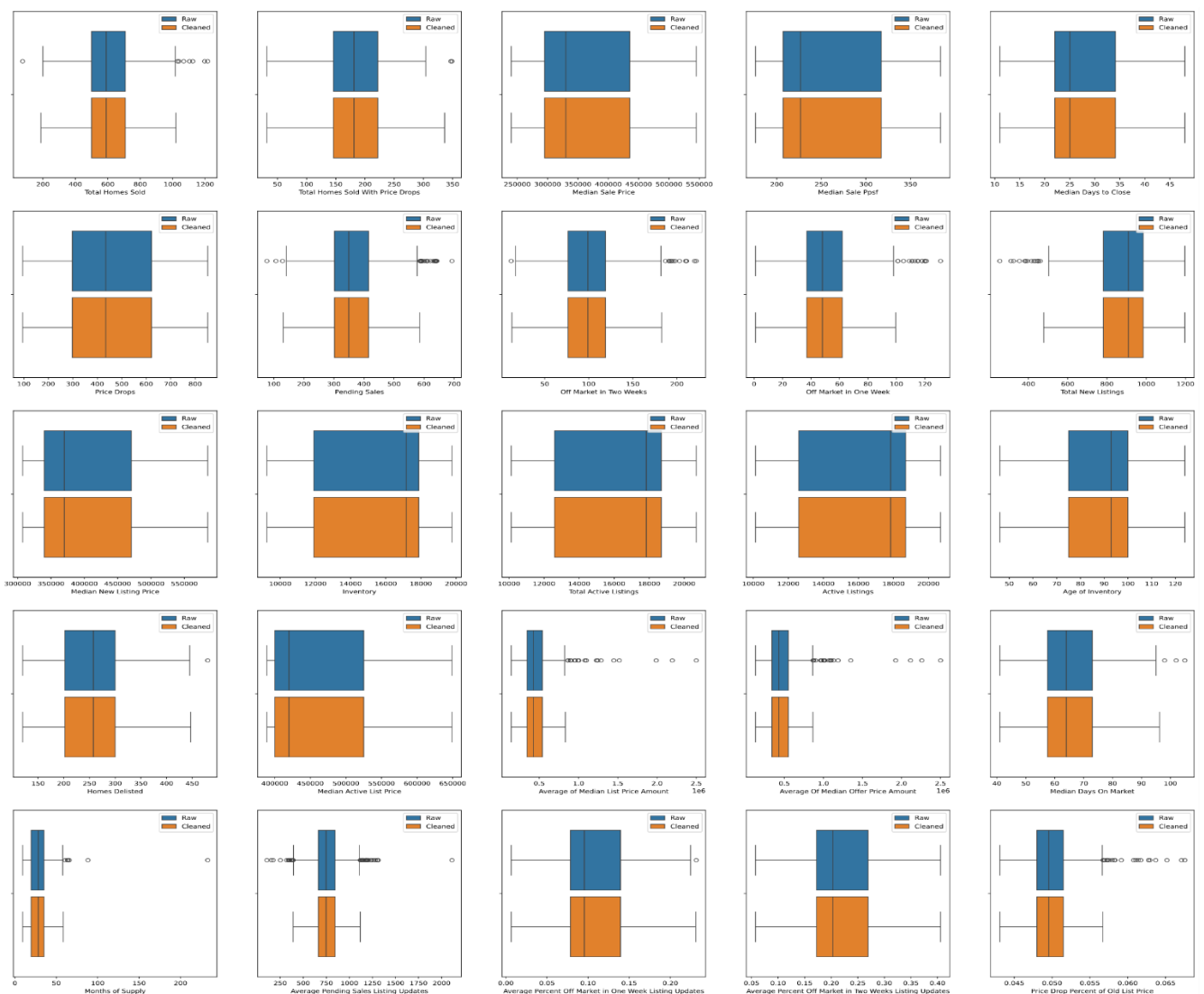
summary.csv

The table displays measurements such as mean, median, max, range, standard deviation, and standard error of the mean for the 19 key variables. The highest standard deviation is for 'median sale price' (305,000) and 'median new listing price' (277,505).

variance.csv

The table displays measurements (statistic, p-value, result) on the variance for 19 variables. According to the table, 'age_of_inventory' is the only variable with an equal variance.

weekly_sales.svg

The chart compares the annual difference between homes sold in Miami within the 1st or 2nd weeks of listing. Based on the chart, historically, homes are primarily sold within the 2nd week of listing, and sales within the 1st week are much lower. Overall, sales within both the 1st and 2nd weeks peaked in 2021, which is also validated by other findings in this analysis.
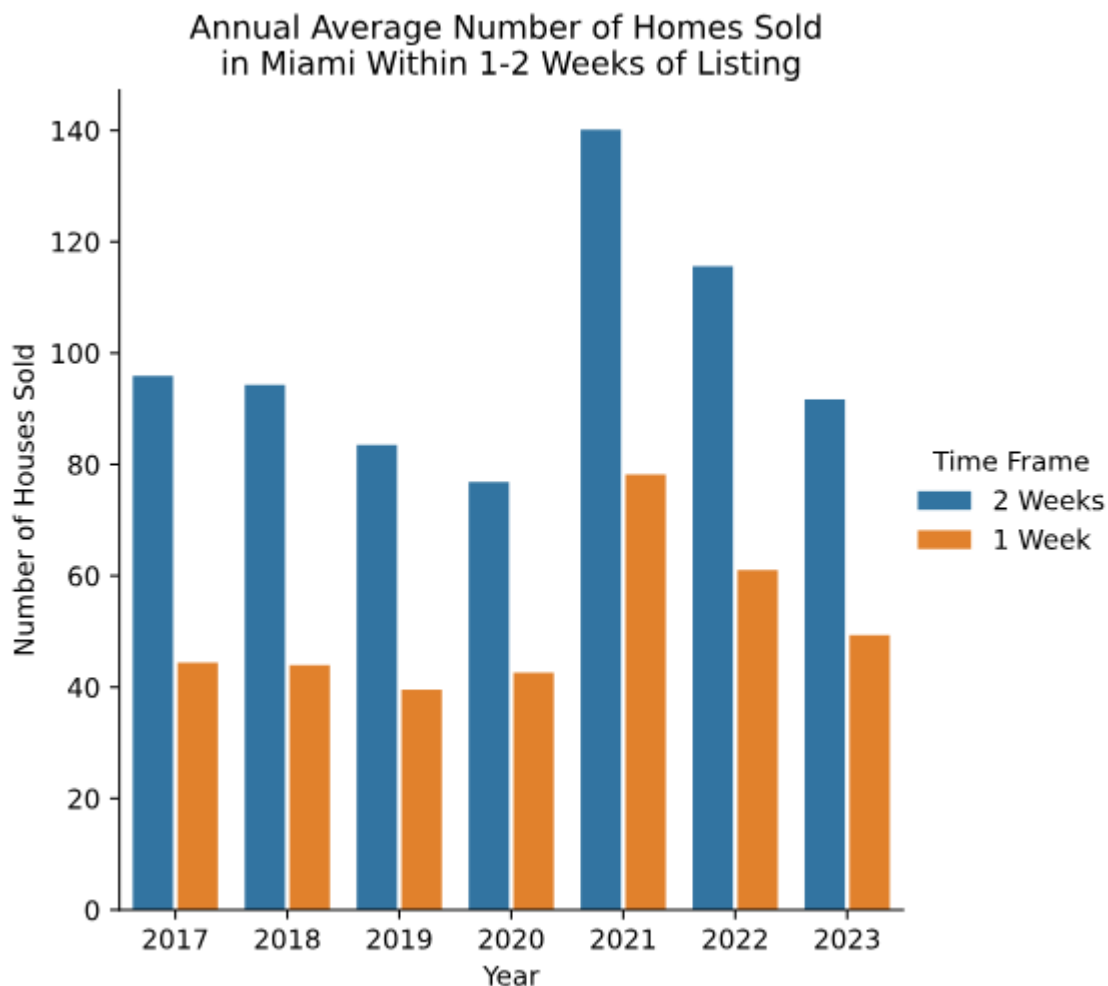
Diagnostic
# correlations_strong.svg
The correlation heatmap explores and illustrates the strong correlation (0.6 and above) between key variables in New York, Seattle, and Miami. The highest negative correlation for the dataframe on Miami has been observed between variables such as 'total homes sold,' 'total homes sold with a price drop,' 'age of inventory,' and 'monthly supply'; 'new listing prices,' 'median sale price,' and 'inventory.' The positive strong correlation is observed between 'total homes sold' and 'off-market in two weeks,' 'off-market in one week.'


Correlation Heatmap for Miami
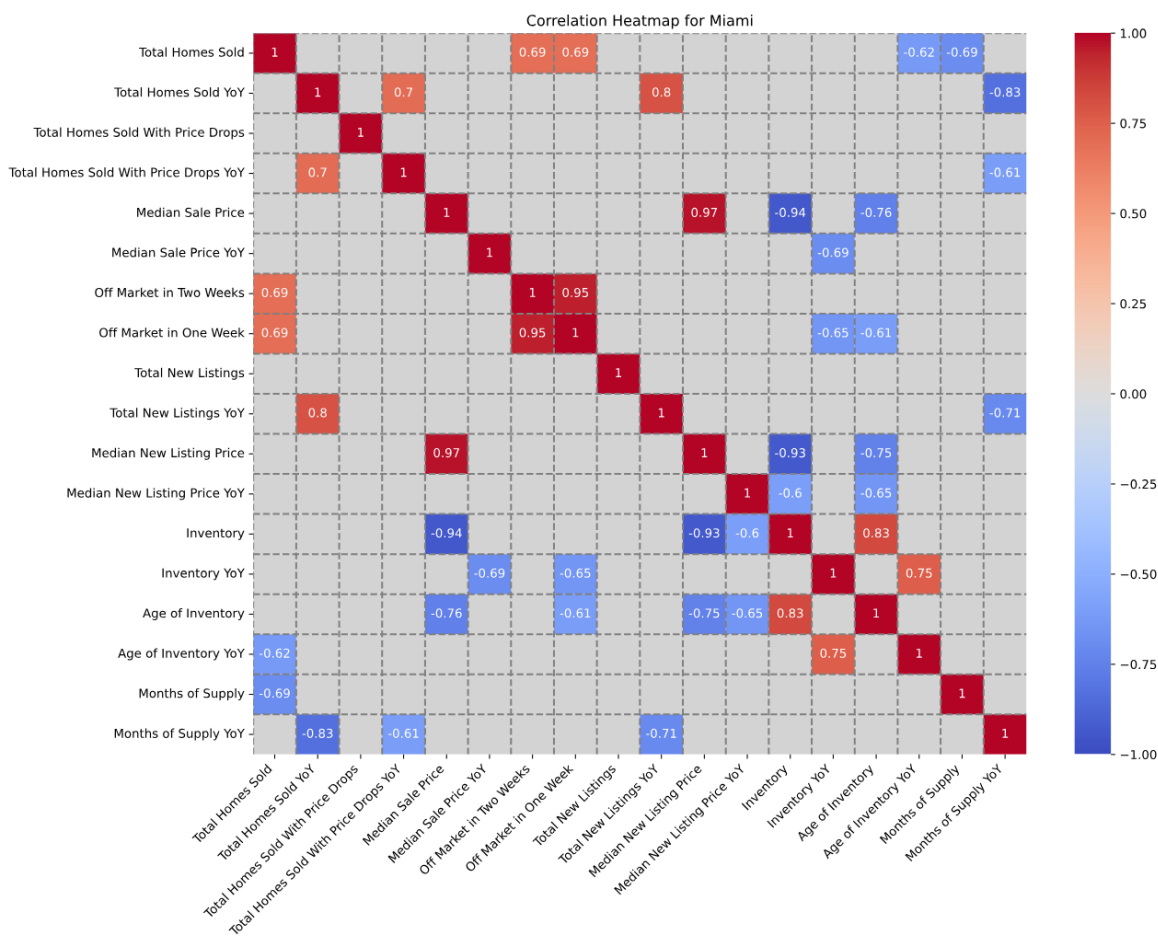
#distributions.svg
The histogram illustrates median days to close in Miami in compiled (marked as pink) and then in annual (marked as blue) formats. The compiled chart for the median days to close in Miami has a Bernoulli distribution shape with peaks at around 20 days and about 35 days. Further examination of the data, as evidenced by the annual chart, reveals that the median days to close the transaction between 2017 and 2019 was about 35 days. It dropped to 23 days in 2020 and has continued decreasing in recent years.

Median Days to Close in Miami (2017-2023) | Median Days to Close in 2017 | Median Days to Close in 2018 | Median Days to Close in 2019
Median Days to Close in 2020 | Median Days to Close in 2021 | Median Days to Close in 2022 | Median Days to Close in 2023
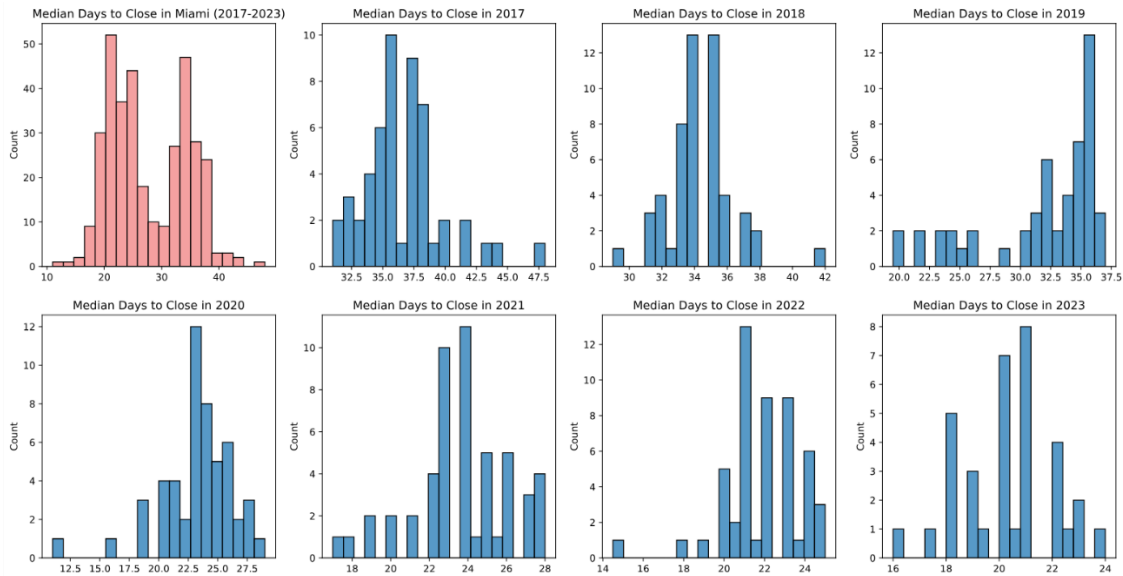
homes_sold_months_supply.svg
The scatterplot aims to compare the relationship between 'months of supply' and 'total homes sold' in three selected cities. Based on the plot, Seattle has the lowest value for the months of supply and total homes sold (this might be explained by the population size and overall demand); New York has more homes sold with the overall months having a higher minimum of 18 and a maximum of 60 months. Miami has a middle value with the lowest homes sold ranging from 15 up to 61 months of supply.

Total Homes Sold vs. Months of Supply

New York City
Seattle
Miami

Months of Supply
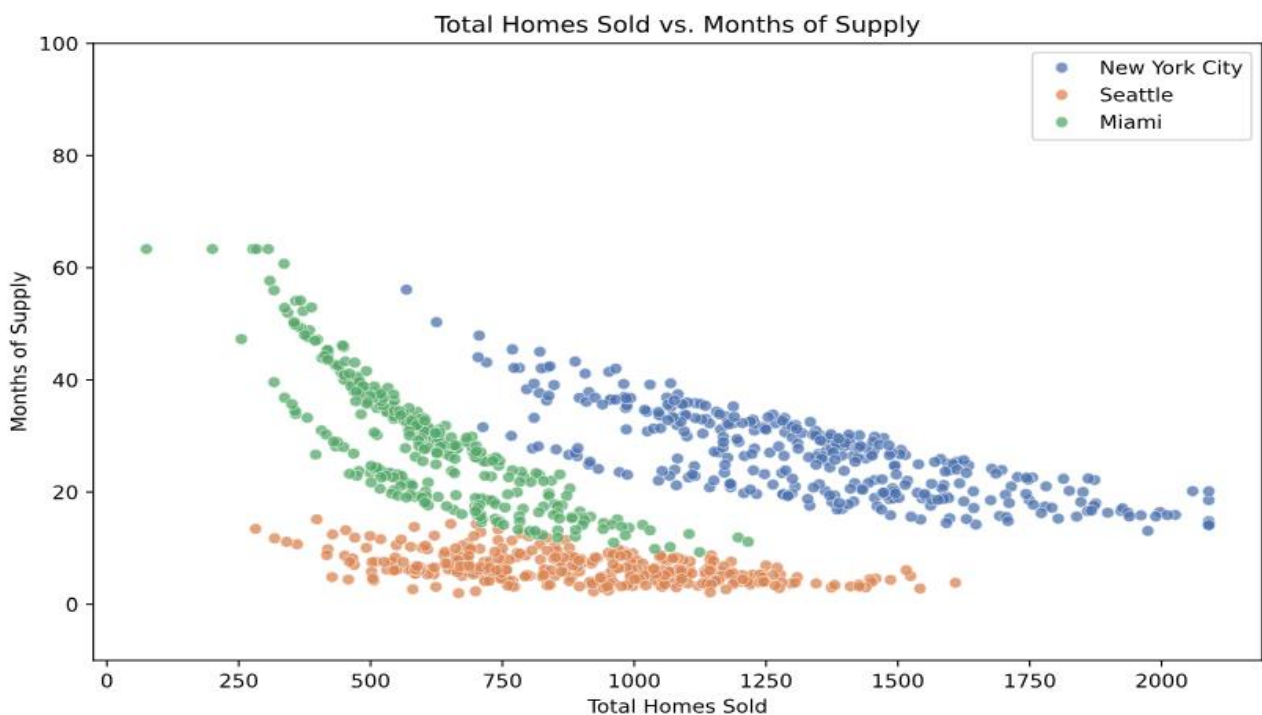
Total Homes Sold

price_trends.svg
The line graphs and bar chart illustrate the relationship between 'median sales price,' 'median sales price of new listings,' and 'total homes sold in Miami.' The chart highlights several important trends in

the housing market in Miami: (1) the new listing price is above the median price year after year; (2) housing prices grow steadily in the last years, with a sharp increase during the last 3 years; (3) the number of homes sold is negatively correlated with the price in 2023: with the highest price, the number of homes sold is the lowest. This can be explained that the data for 2023 is only until August.
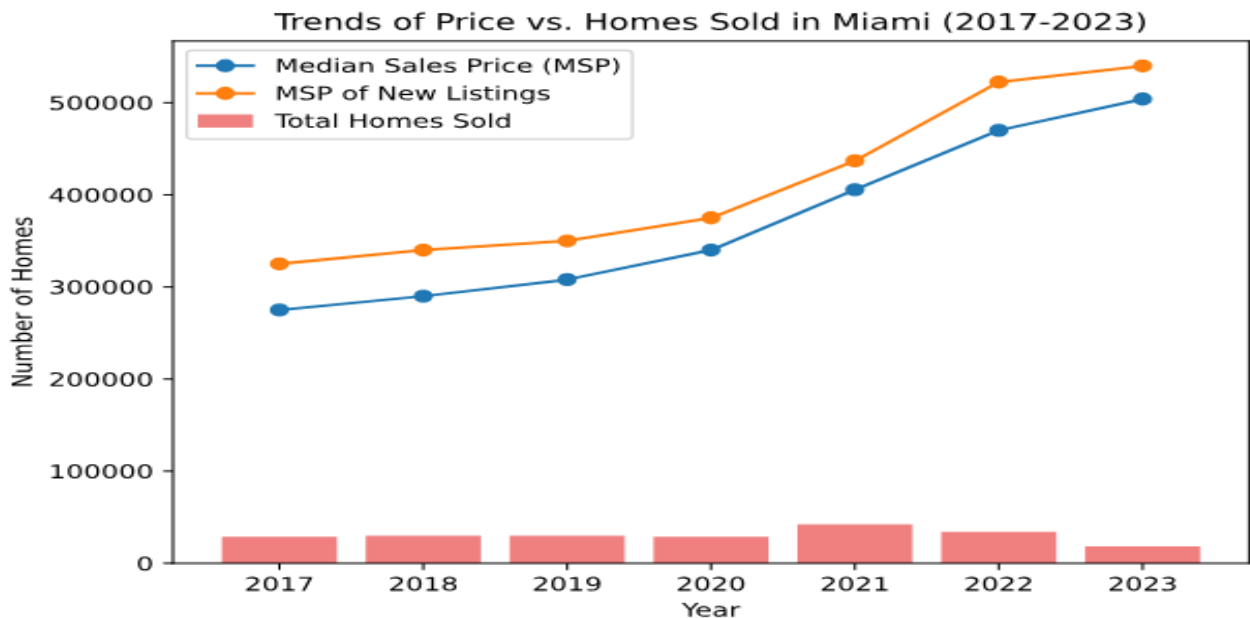
sale_price_days_to_close.svg
The scatter plot examines the median sales price and days to close the transaction in Miami in detail. The plot shows that the median price was fluctuating between USD 250,000 and USD 300,000, and the median days to close the transaction ranged between 30 and 45 days, with a drop in 2019. The number of days for the close transaction remained comparatively stable between 2019 and 2023, with the median prices gradually growing up to USD 550,000.
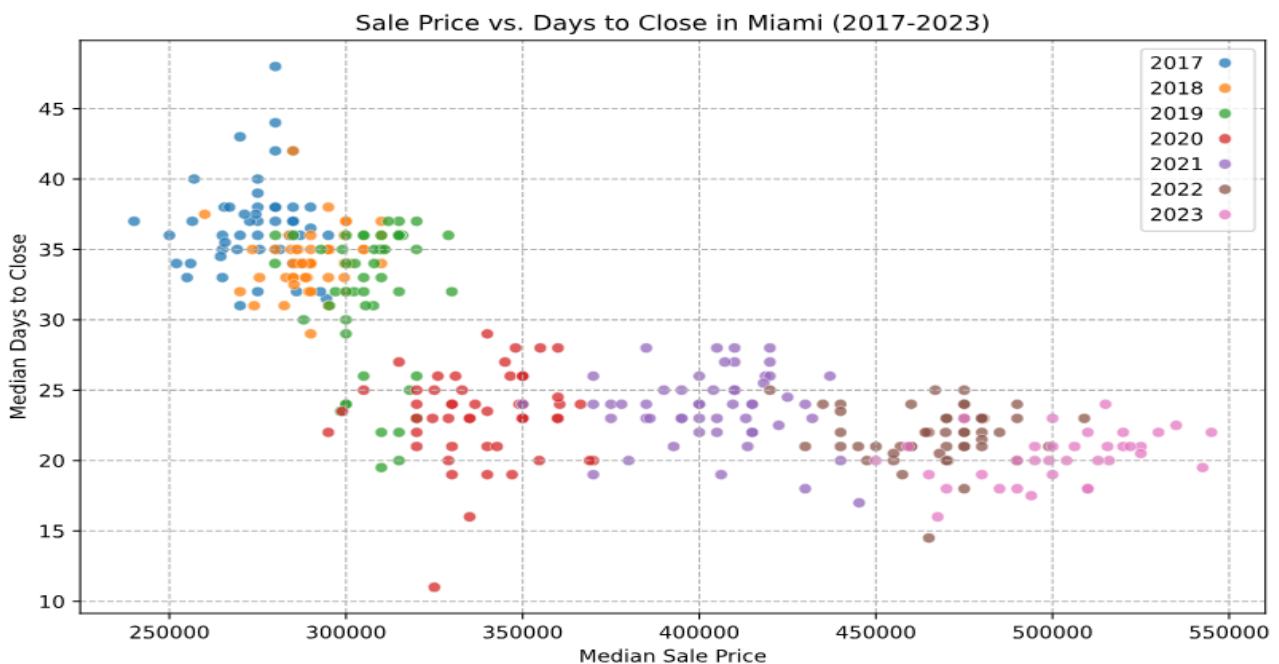
sale_price_median_days_to_close.svg

The scatterplot examines the relations between median sales price and median days to close, comparing three selected cities. Based on the plot, Miami has prices scattered between USD 250,000 and USD 550,000 and days to close transactions between 25 to 45 days, as also evidenced from the previous scatterplot. Notably, the housing price historically is lower in Miami than in Seattle and New York. Specifically, the feature of the housing market in Seattle is growing constantly and reaching the highest points in comparison with the other cities, remaining stable in terms of the median close transaction between 30 to 40 days. The housing market in New York City is highly skewed in comparison with Miami and Seattle: the price has increased from USD 550,000 to USD 750,000 with the highest median days to close the transaction above 60 days.
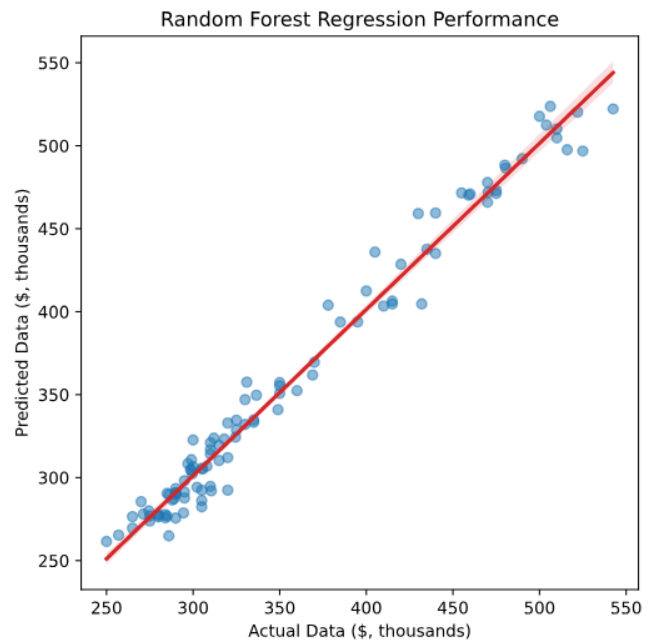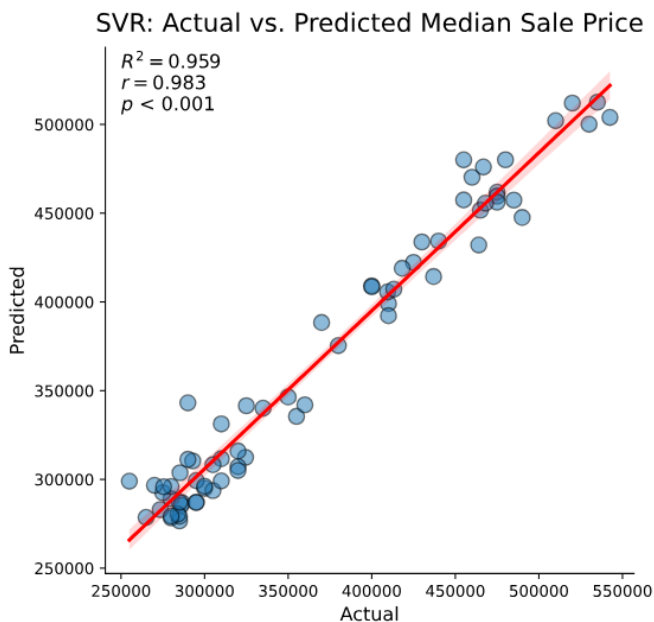
Predictive and Prescriptive Analytics

For the purposes of predictive and prescriptive analytics, Support Vector Regression(SVR) and Random Forest (RF) models were applied. The limitation of the dataframe and potential challenges in the models have been described in the README file.

In the SVR model, R-squared value of 0.959 indicates that approximately 95.9% of the variability in the median sales price can be explained by the model. The correlation coefficient (r) of 0.983 displays strong positive linear relationship between the predicted and actual median sales prices. The associated p-value, being less than 0.001, highlights the statistical significance of this relationship. These findings suggest that the SVR model demonstrates a high level of accuracy in predicting median sales prices in the Miami housing market.

In the Random Forest model, the R-squared score of 0.997 further underscores the model's exceptional accuracy, explaining approximately 99.7% of the variance in predicting median sales prices in the Miami housing market between 2017 and 2023. Additionally, Root Mean Squared Error (RMSE) of 4746.55 is also comparatively low. These suggest suggest that the Random Forest model is also highly effective in capturing and modeling the complex patterns in the housing market in Miami.

Insights on the Prescriptive Analytics

Overall, the housing market in Miami can be described as unaffordable. The analysis and predictive model presented above offer a glimpse into the dynamics of the market between January 2017 and August 2023. Market prices soared in 2019 and have continued to grow significantly since then. Furthermore, the initial high difference between supply (new listings) and demand (number of homes sold) has narrowed in recent years. These insights propel further investigation into the housing market in Miami to better understand the case in a larger context and provide recommendations on literature on the subject.

Wijburg (2021) compares the governance of affordable housing in Amsterdam and Miami, arguing that Miami relies heavily on market forces and private contributions[1]. However, turning points can be made with strong public commitments through an Affordable Housing Trust Fund. In a recent study, Weiss (2023) highlights that Miami is experiencing a housing affordability crisis, becoming increasingly unaffordable in the USA due to rising rates, record prices, insurance premiums, and unprecedented migration to Southeast Florida[2]. In this regard, the policy recommendations on affordable housing in Seattle provided as part of this assignment are relevant for Miami as well.

Apart from affordability, one aspect highlighted in the research is the high vulnerability of Miami to climate change and its potential effects on the housing market [Rodziewicz et. al (2022)[3], Gosain et. al (2022)[4], Keys & Mulder (2020)[5]], which developers, policymakers, and others must consider when acting upon the housing market in Miami.

Forecasting the housing market has been widely explored both in academic literature and corporate non-academic sources. Some models presented below contrast SVM and RF models above. Møller et al. (2023) developed the Housing Search Index (HIS) with strong predictive power over subsequent changes using Miami as a case.[6] Treyz (2023) presents REMI, a comprehensive economic/demographic forecasting model, for 20 U.S metropolitan areas, including Miami, containing assumptions on housing price elasticities[7]. Hsieh & Lin (2021) present a Generative Adversarial Network (GAN), an unsupervised machine learning algorithm, and Long Short-Term Memory (LSTM), a neural network with long-term memory, for forecasting house prices. [8] In comparison with academic literature, some online commercial publications provide generalized insights into the Miami housing market, claiming that rising construction costs will significantly impact home pricing, low unemployment in Miami will affect the demand and purchasing power of buyers, and other predictions.

Given all of the above, additional recommendations include:

- Enhancing public-private dialogue in Miami to strengthen human-centered housing resilience towards disaster risk reduction. This will involve dialogue supported by policymakers to address potential housing crises induced by climate change.

---

[1] Retrieved from https://www.sciencedirect.com/science/article/pii/S0016718520302979
[2] Retrieved from : https://dspace.mit.edu/handle/1721.1/150284
[3] Retrieved from: https://link.springer.com/article/10.1007/s10669-022-09842-6
[4] https://www.sciencedirect.com/science/article/pii/S2212420922002801?casa_token=VMCkH2NzvCoAAAAA:-Dnsjwm8gY75wCijKK6wsHnltc-grb1q0Xh76J_DOF7jyXvfDw2FqxoeFo3oKe5SXmbXF5G7Wg
[5] Retrieved from: https://www.nber.org/papers/w27930
[6] Retrieved from: https://pubsonline.informs.org/doi/full/10.1287/mnsc.2023.4672?casa_token=i-wIAkPwWjsAAAAA%3AORiTxXd6iRc0zP2FFdg-7KeML2KOXyfHKhb-OfjNIaNqjKvQ6PbpFeTdEXW7QsMGCrmbB_Q_fA
[7] Retrieved from https://journals.sagepub.com/doi/full/10.1177/08912424221145186?casa_token=VTm2H-UGXuAAAAAA%3AeP-u62ZZ-qvphvPHBywNXIsr27GVjIrSVgKVVI2nS9Rv3lHHxSos2DbYGX-DXPxdDv_5GZ-Y_PWK
[8] Retrieved from: https://ieeexplore.ieee.org/abstract/document/9778012?casa_token=vRgq-YlDXkAAAAAA:yh3cYoyYz15LMPUlgfEwTuLJnJaxZ3gM-qKhozLESVUDiJKEOYmKhu6-6zDgpBJ1sWcXQQD2

- Strengthen public commitment towards affordable housing through existing structures such as the Affordable Housing Trust Fund, and encouragement of development to increase the margin for affordable housing.

Future research on the housing market in Miami can significantly benefit from merging Redfin data with dataframes that describe demographic context and applying new models for forecasting housing prices.