CLASS_PROJECT_GROUP_5

Introduction to the Analysis of Real Estate Price Trends

AI Bootcamp
Project 1

Project 1 – Class_Project_Group_5

Teacher: Tom Seeber

Asst Teacher: Enzo Zirotti

Team Members: Mathew Watkins, Michael Sheridan, Terry Hood

GitHub Repository:

- The team will work as a group to find and analyze a dataset of choice and work in the team repository created in GitHub https://github.com/MWatkins87/Class_Project_Group_5

- @MWatkins87's Group 5 Group Project has been created as a Kanban project in Gishu's Project section https://github.com/users/MWatkins87/projects/2/views/1

- Project 1 Overview and list of requirements can be found in the DU Bootcampspot.com website https://bootcampspot.instructure.com/courses/5432/pages/9-project-1-overview?module_item_id=1200592

Table of Contents

Study of Impacts to Housing Prices

Introduction:

- Provide an overview of the project and its objectives.

- Briefly describe the dataset and its context.

- Outline the structure of the report.

Project Overview

Introduction to the Analysis of Real Estate Price Trends

The recent surge in housing prices in the American real estate landscape presents a significant challenge, with conventional wisdom attributing it primarily to internal migration and a stagnant housing supply. However, contrary to expectations, regions experiencing population declines have not seen corresponding decreases in housing prices, suggesting additional influencing factors.

This study investigates alternative factors affecting housing price trends across ten selected cities, representing both booming and sluggish markets. These cities include Atlanta, Miami, Detroit, Tampa, and Las Vegas with strong market growth, and Baltimore, New Orleans, Washington D.C., San Francisco, and Portland with less robust markets.

Analysis spanning from 2012 to 2022 incorporates various variables such as population growth, homelessness, crime rates, college tuition costs, and mortgage rates to identify correlations explaining the divergent housing price trends observed.

Project Approach

Our project began with an intensive brainstorming session, wherein we delineated the project's objectives, delineated goals, and pinpointed key variables and social contexts likely to yield impactful datasets for comparison against our foundational dataset of median housing prices.

Slack communication Direct Message channel, Slack Huddle, Google Meets, Zoom, were all identified as team communication methodologies.

Task allocation was developed and populated in a Kanban style Project in GitHub Project, with team members assigned specific elements to navigate through the project's trajectory. Throughout the project lifecycle, we engaged in continuous comparison of dataset discoveries, collaborated to resolve coding intricacies, and conducted thorough reviews and evaluations of each other's work as part of a rigorous quality assurance process.

Collaboratively, we delved into the analysis of our findings, endeavoring to discern their implications and uncover correlations of significance.

Report Structure

This report will contain various sections of organization to inform the reader of information gathered and techniques utilized for data gathering and analysis for this study.

- Introduction
- Data Exploration
- Data Analysis
- Conclusions
- References
- Sources

Datasets Purpose

Multiple Datasets were utilized to pull variables into the study and determine the impacts they may have in Housing Price Growth.

1. Zillow
   - Foundation dataset that we used to compare the additional datasets against for correlation analysis.
2. College Tuition
   - This is the data collected for average college tuition costs. Does the cost of higher education impact home values in their respective cities? To try and see a correlation here I took the sum of all average tuition costs for the colleges in the states we are interested in and graphed based on the change in tuition cost from year to year. I figured that with the cost of tuition going that those graduating would be seeking out better jobs, having more money, and spending that money on real estate.
3. Scholastic Aptitude
   - This is the data collected to see if scholastic aptitude has an impact on Home Values in the 10 cities we are looking at. To get some measure of what "scholastic aptitude" can be graphed by, I took the average SAT scores for the colleges in the cities we were looking at and how much they changed over the same 10 years. My assumption is that if scholastic aptitude is going up, that those people would be more inclined to invest in things like real estate, and that could be a factor in the inflation of home values.
4. Printing of Money
   - This is the data collected to see if the printing of money is increasing real estate prices. My assumption is that if money is being printed at the same rate housing prices are going up, that there is correlation between the printing of money itself and the increase in home values. For this I gathered the annual production reports, totaled them, and graphed them over the same 10-year span.

5. Crime
   - This analysis examines the relationship between crime rates and real estate prices in ten major U.S. cities, using data sourced from the FBI's Uniform Crime Reporting (UCR) database. The dataset includes annual figures for both violent and property crimes, organized by year and state.
6. Population Change
   - This study examines the impact of population dynamics on real estate prices, utilizing a comprehensive dataset derived primarily from FBI crime statistics that include annual population data. This data encompasses decennial population counts from cities and towns, supplemented by annual estimates from the U.S. Census Bureau. In instances where local data was unavailable, Census Bureau estimates were used to ensure dataset completeness. Notably, the inclusion of the year 2021 marked the restoration of complete data coverage, enhancing the study's scope and reliability.
7. Homelessness
   - This study leverages data from the "HUD Point-in-Time (PIT) Counts by Continuum of Care (CoC)," a program designed to foster community-wide efforts to end homelessness. This initiative provides funding for efforts by nonprofit providers, states, and Indian tribes or tribally designated housing entities. CoCs— collaborations of governmental and non-governmental organizations—operate within designated metropolitan or rural areas and are required to perform biennial PIT counts of the homeless population. This data is systematically compiled into annual datasets available for analysis.

Data Exploration:

Datasets

(Describe the dataset, including its size, features, and any preprocessing steps.)

1. Zillow
   - Zillow Housing Data - [Housing Data - Zillow Research](#)
   - Licensing - [Public Records Data Terms | Zillow](#)
   - This input "'" delimited .csv file is 80,745k in size.
   - Git Large File Storge was required for handling this file.
     - https://git-lfs.com/
   - Additional Storage on GitHub was also required for use.
2. Average College Tuition Costs / Scholastic Aptitude
   - College Score Card - [Data Home | College Scorecard (ed.gov)](#)
   - Licensing - [Copyright Status Notice | U.S. Department of Education](#)
   - Inputs were from .csv files that were merged together during the data exploration process.

3. Engraving and Printing
   - Bep.gov - Annual Production Reports | Engraving & Printing (bep.gov)
   - Licensing - FOIA | Engraving & Printing (bep.gov)
   - URL = Annual Production Reports | Engraving & Printing (bep.gov)
     - introduced into dataset as html.
4. Crime / Population
   - Data.gov - Data.gov Home - Data.gov
   - Licensing - Data License for Access to Restricted Data | HUD USER
   - This input "'" delimited .csv file is 4k in size.
   - This data encompasses decennial population counts from cities and towns, supplemented by annual estimates from the U.S. Census Bureau.
   - The UCR data is provided voluntarily by local law enforcement agencies, which leads to some inconsistencies and gaps. Notably, in April 2015, Portland and 40 surrounding communities transitioned to a new $12.6 million reporting system, resulting in a lapse in data reporting for that year. To address this gap, we "healed" the dataset by averaging the crime figures from 2014 and 2016, thereby providing a continuous data set for Portland.
     - OregonLive report - Portland to be left out of FBI's 2015 crime report due to programming problems - oregonlive.com
   - Additionally, The Marshall Project highlights a broader issue: approximately 7,000 of the 18,000 law enforcement agencies, covering about 35% of the U.S. population, did not submit their crime data due to a transition to the FBI's new National Incident-Based Reporting System. Many municipalities and several states were unable to adapt to this change in time, leading to the exclusion of 2021 crime data from this study.
     - The Marshall Project report - The Problem With The FBI's Missing Crime Data | The Marshall Project
5. Homelessness
   - Hud.gov - Point-In-Time (PIT) Count | HUD.gov / U.S. Department of Housing and Urban Development (HUD)
   - Licensing - Data License for Access to Restricted Data | HUD USER
   - For this analysis, ten annual datasets spanning various years were consolidated into a single DataFrame. Specific data relevant to urban centers and homeless counts were then extracted into another DataFrame for detailed examination. Notably, data for the years 2018 and 2019 was unavailable, prompting their exclusion from the analysis.

Data Analysis:

(Present the main findings of your analysis.)

1. Zillow
   - Main Findings

- Below is the initial code used as a baseline for the overall question of the group: "What is driving home prices up?" We got the data from Zillow and used it as a base line for the comparisons we tried to make. I decided to look at education, scholastic aptitude, and the printing of money as factors I thought might be contributing to the rise in real estate prices.
- As you can see from the first 10 graphs, real estate prices have been going up in not only the top 5 real estate market cities, but the bottom 5 as well. From here I will be looking for other factors to show a steady increase as well as to show correlation.
- Variables – (Summarize the distribution of the variables (mean, median, mode, standard deviation, etc.)
  - Summarize the distribution of the variables:
    - The dataset includes housing price data from 2012 to 2022 for various regions, states, and metro areas.
    - Filtering was applied to extract relevant data for analysis, including specific states, cities, and metro areas of interest.
    - Missing values were filled with zeros and rounded to enhance readability.
    - The distribution of housing prices over time was examined to understand trends and variations.
  - Describe the methods and techniques used for analysis:
    - Data was imported from a Zillow dataset stored in a CSV file.
    - Initial filtering was performed to select columns necessary for analysis, focusing on region names, state names, metro areas, and monthly housing prices from 2012 to 2022.
    - Additional filtering was applied to extract data for specific states, cities, and metro areas of interest.
    - The dataset was cleaned by filling missing values with zeros and rounding numerical values.
    - Techniques such as transposing the dataset and converting it into a DataFrame were used to prepare for plotting.
    - Plots were generated for each city to visualize the median home prices over the 10-year period, providing insights into housing price trends and variations across different regions.

- Visualizations - (Visualize the data using histograms, box plots, scatter plots, etc.)
  - Sample plots for Zillow Baseline

Baltimore

Detroit

Tampa

Washington
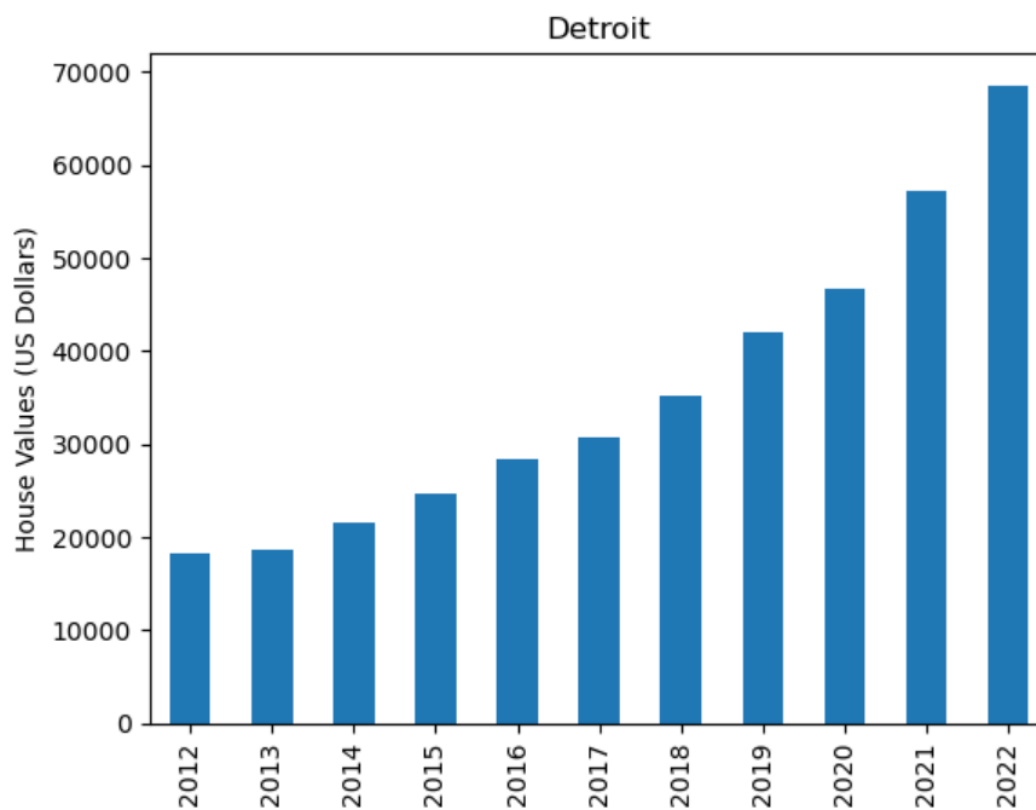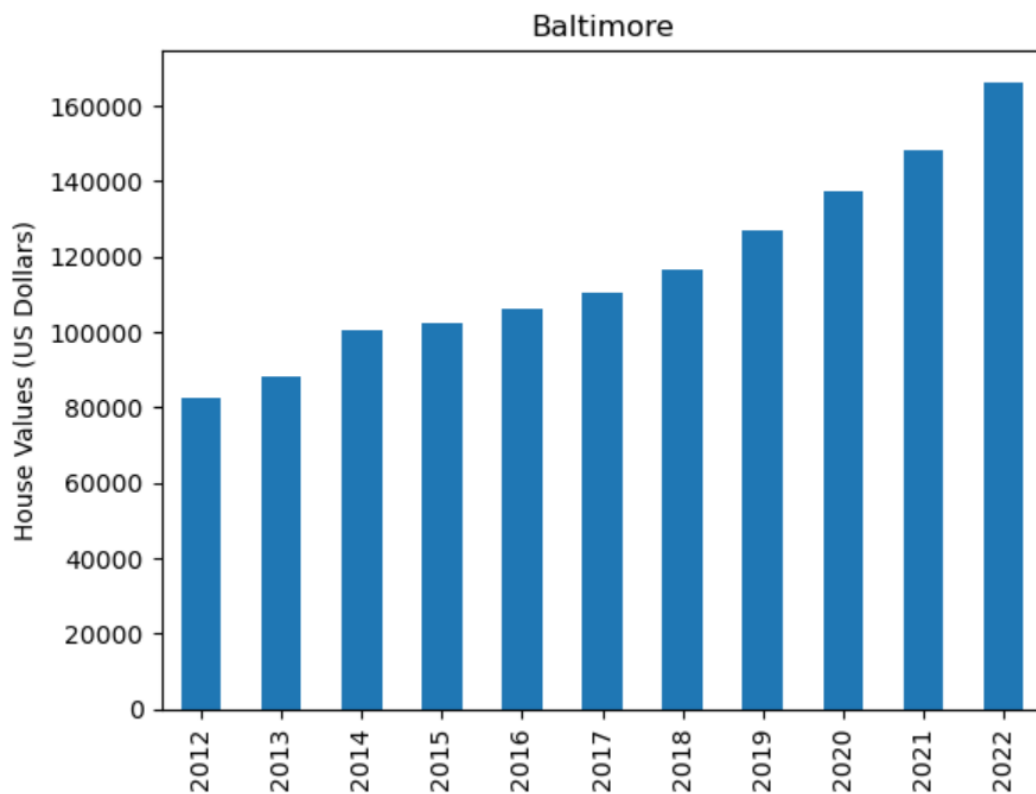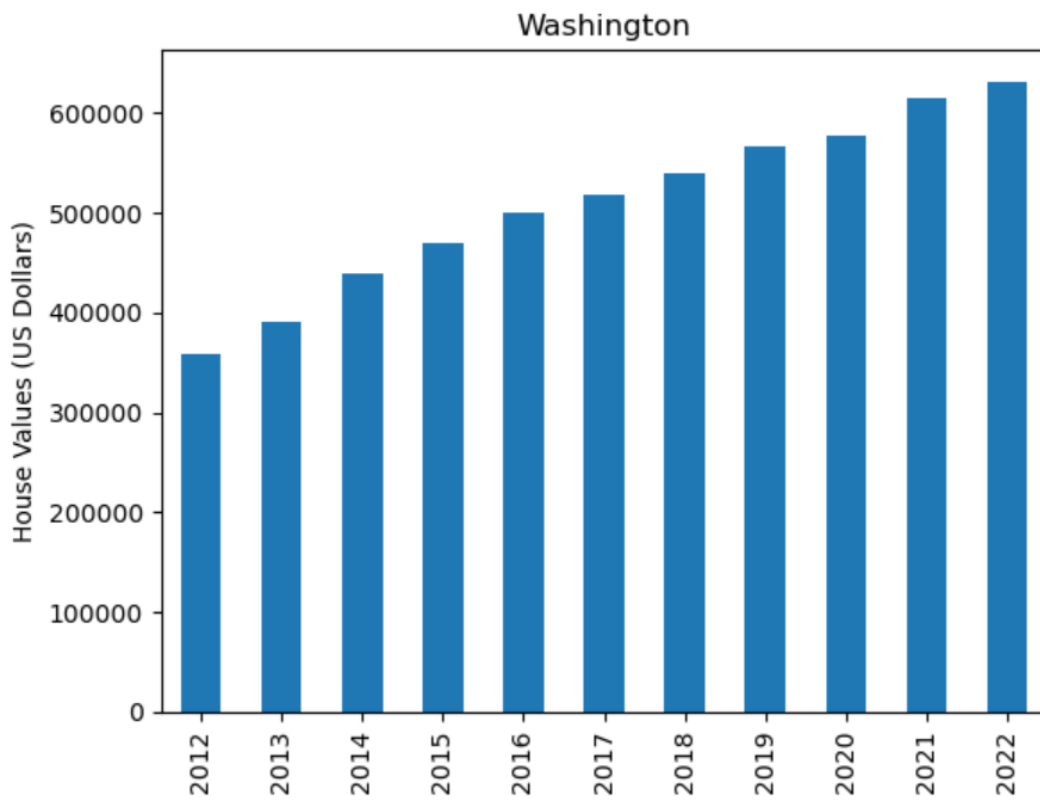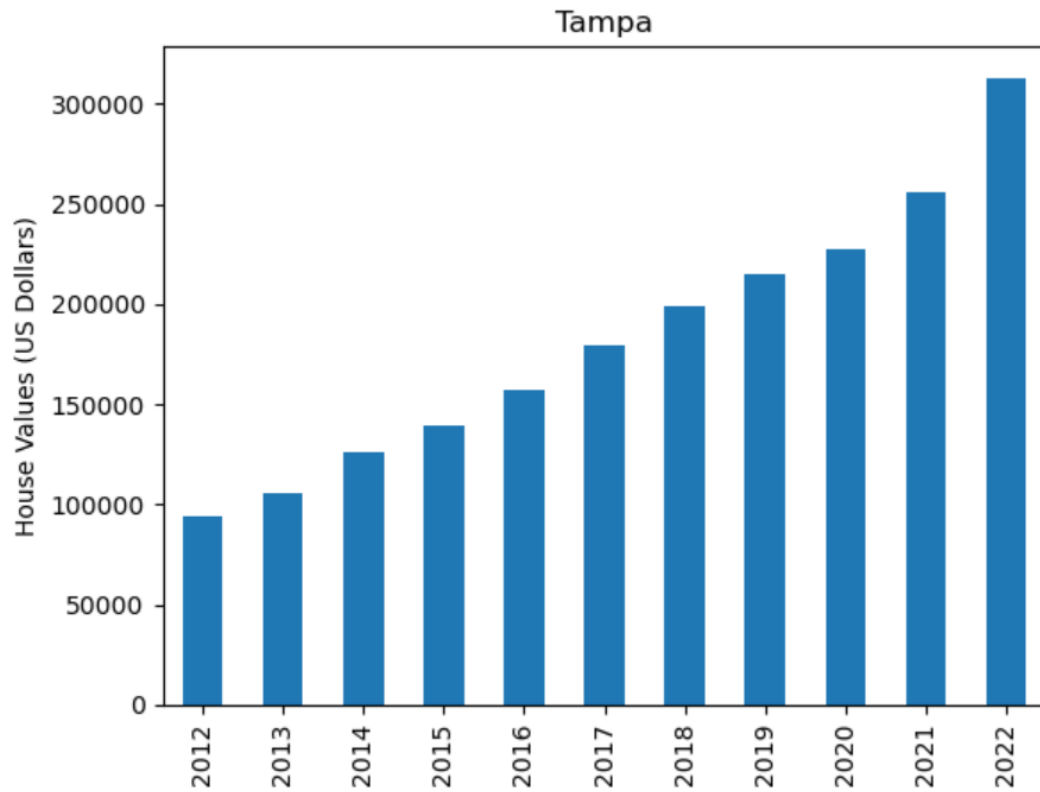
- Patterns-Trends-Outliers - (Identify any patterns, trends, or outliers in the data.)
  -

2. Average College Tuition Costs
   - Main Findings
     - This is the data collected for average college tuition costs. Does the cost of higher education impact home values in their respective cities? To try and see a correlation here I took the sum of all average tuition costs for the colleges in the states we are interested in and graphed based on the change in tuition cost from year to year. I figured that with the cost of tuition going that those graduating would be seeking out better jobs, having more money, and spending that money on real estate.
     - As you can see the average cost of College Tuition has gone up in each of the cities we investigated.
   - Variables – (Summarize the distribution of the variables (mean, median, mode, standard deviation, etc.)
     - Summarize the distribution of the variables:
       - The dataset includes average college tuition costs for various cities over a span of 10 years.
       - Initial data extraction involved importing data from the US Department of Education CSV file.
       - Filtering was performed to select relevant columns such as institution name, city, state, and tuition costs.
       - Data cleaning steps included removing unnecessary columns to focus on city-level tuition costs.
       - The mean tuition cost was calculated for each city to provide a representative measure of average tuition expenses.
       - Summary statistics such as mean, median, mode, and standard deviation were not explicitly provided but could be calculated from the filtered dataset to analyze the distribution of tuition costs over time.
     - Describe the methods and techniques used for analysis:
       - Data was imported from a CSV file using pandas.
       - Initial filtering was performed to extract relevant columns and create filtered DataFrames for states and cities.
       - Data manipulation techniques such as grouping by city and calculating the mean tuition cost were applied to analyze trends in college tuition expenses.
       - Data visualization techniques such as bar plots were used to visualize changes in average tuition costs for target cities over the 10-year period.
       - The concatenated dataset allowed for comparison of average tuition costs across multiple years for each city, providing insights into trends and variations in college expenses.
   - Visualizations - (Visualize the data using histograms, box plots, scatter plots, etc.)
     - Sample plots for Tuition Costs

Atlanta

Average Tuition Cost (US Dollars)

Las Vegas

Average Tuition Cost (US Dollars)

San Francisco

Tampa
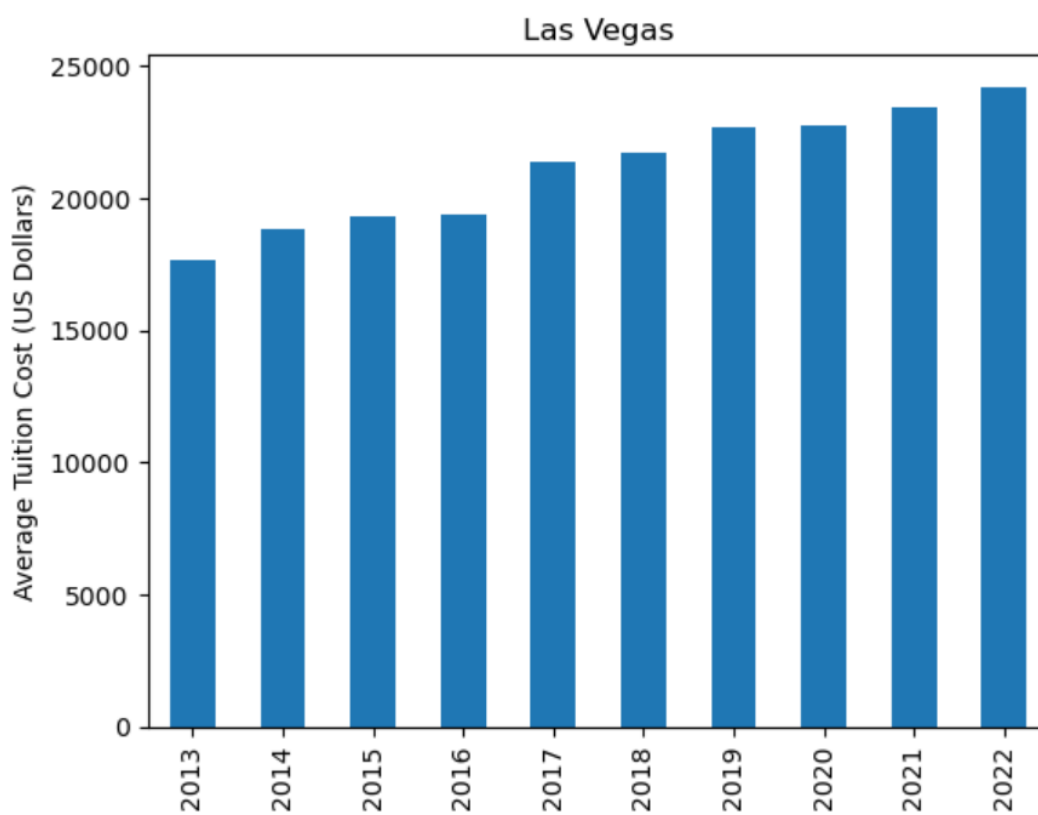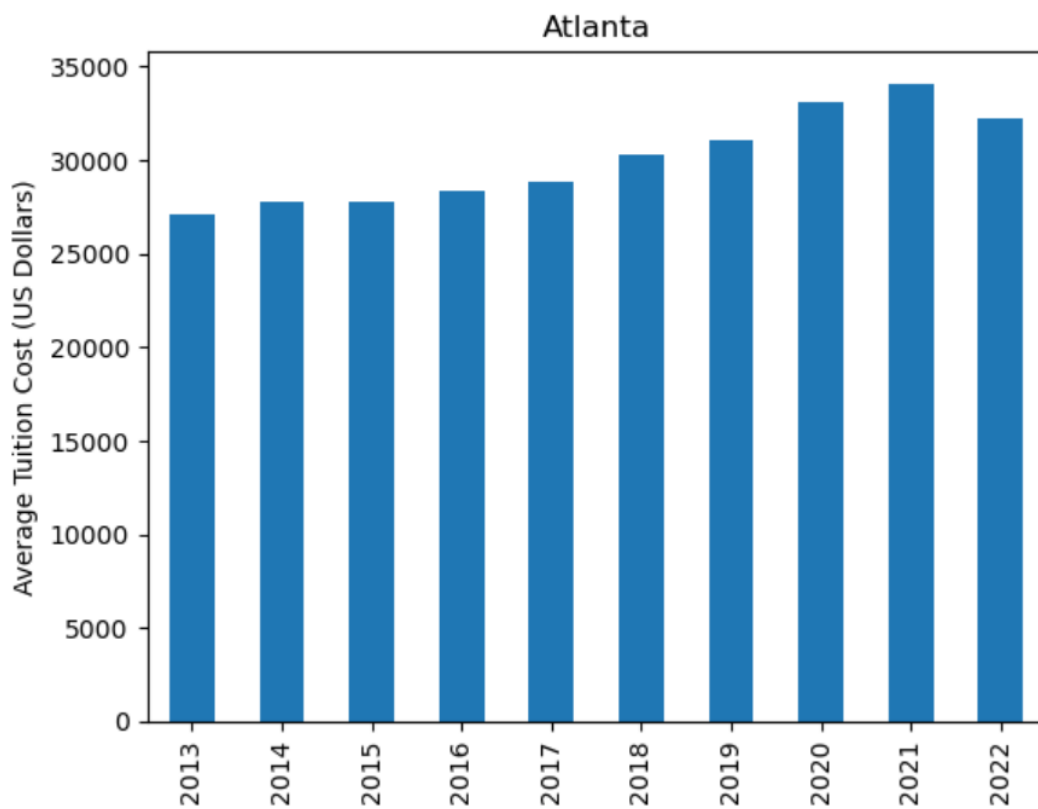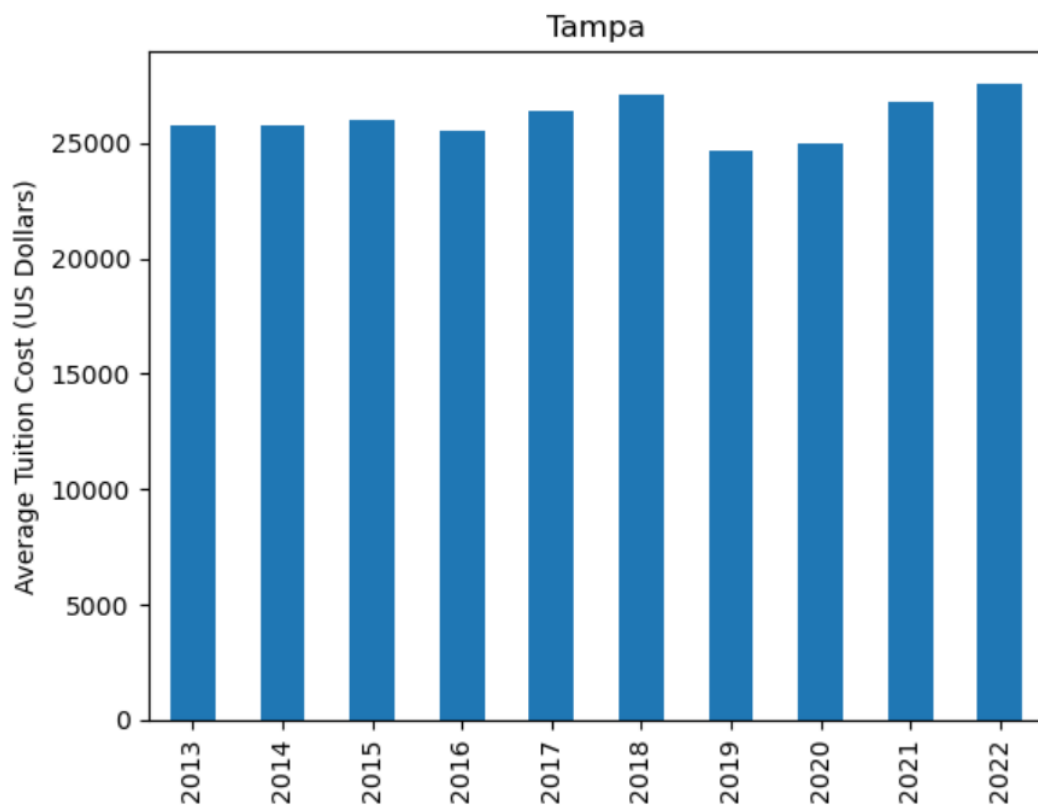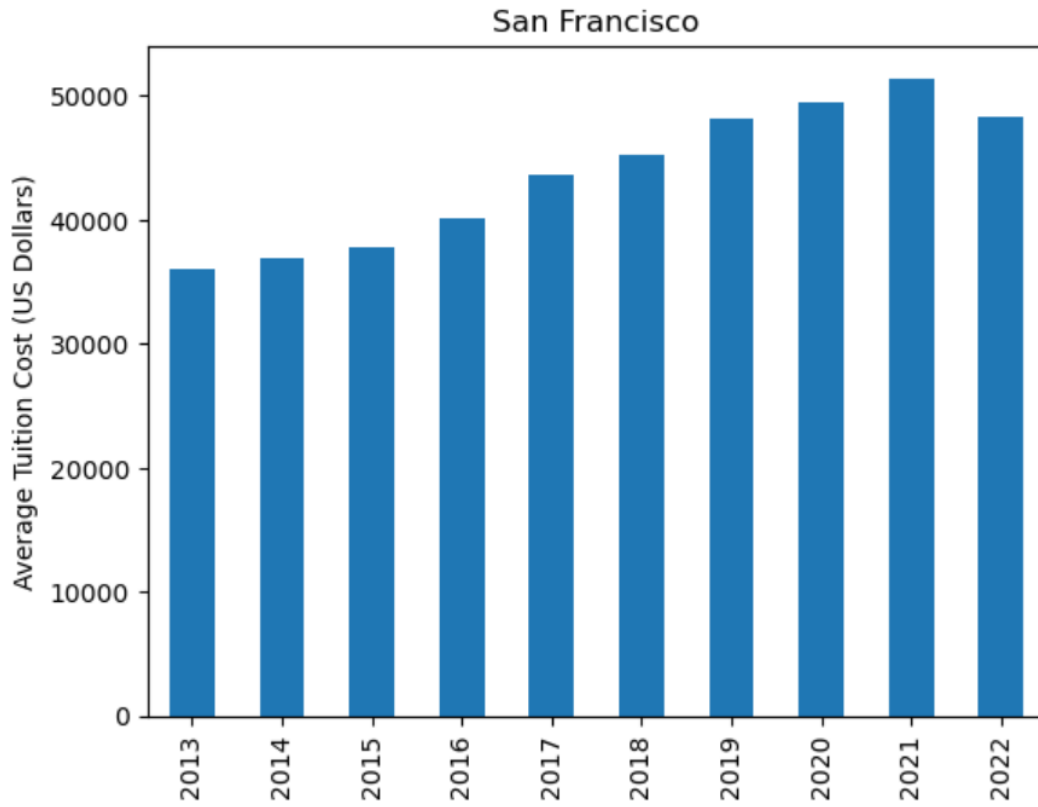
- Patterns-Trends-Outliers - (Identify any patterns, trends, or outliers in the data.)

- Though not as much of an increase as we saw in the real estate data set, which would have shown a strong correlation. My conclusion is that the cost of Higher education does seem to have an impact on the cost of homes in their respective cities, even if it's only a minor correlation.

3. Scholastic Aptitude
   - Main Findings
     - Now we wanted to see if scholastic aptitude has an impact on Home Values in the 10 cities we are looking at. To get some measure of what "scholastic aptitude" can be graphed by, I took the average SAT scores for the colleges in the cities we were looking at and how much they changed over the same 10 years. My assumption is that if scholastic aptitude is going up, that those people would be more inclined to invest in things like real estate, and that could be a factor in the inflation of home values.
     - What I found is that in every city beside Baltimore, there is a noticeable increase in the scholastic aptitude reflected in average SAT scores over the last 10 years. Higher performing home buyers might be a contributing factor in the increase in home values based on this analysis.

   - Variables – (Summarize the distribution of the variables (mean, median, mode, standard deviation, etc.)
     - Summarize the distribution of the variables:
       - The dataset includes average Scholastic Aptitude Test (SAT) scores for various cities over a span of 10 years.
       - Initial data extraction involved importing data from the US Department of Education CSV file.
       - Filtering was performed to select relevant columns such as institution name, city, state, and SAT scores.
       - Data cleaning steps included removing unnecessary columns and filtering out NaN values to focus on city-level SAT scores.
       - The mean SAT score was calculated for each city to provide a representative measure of average student performance.
       - Summary statistics such as mean, median, mode, and standard deviation were not explicitly provided but could be calculated from the filtered dataset to analyze the distribution of SAT scores over time.
     - Describe the methods and techniques used for analysis:
       - Data was imported from a CSV file using pandas.
       - Initial filtering was performed to extract relevant columns and create filtered DataFrames for states and cities.
       - Data manipulation techniques such as grouping by city and calculating the mean SAT score were applied to analyze trends in student performance.

- Data visualization techniques such as bar plots were used to visualize changes in average SAT scores for target cities over the 10-year period.
- The concatenated dataset allowed for comparison of average SAT scores across multiple years for each city, providing insights into trends and variations in student performance.

- Visualizations - (Visualize the data using histograms, box plots, scatter plots, etc.)

  - Sample plots for Scholastic Aptitude

Atlanta

Average SAT Scores

Baltimore

Average SAT Scores

Detroit

Average SAT Scores

**San Francisco**

Average SAT Scores
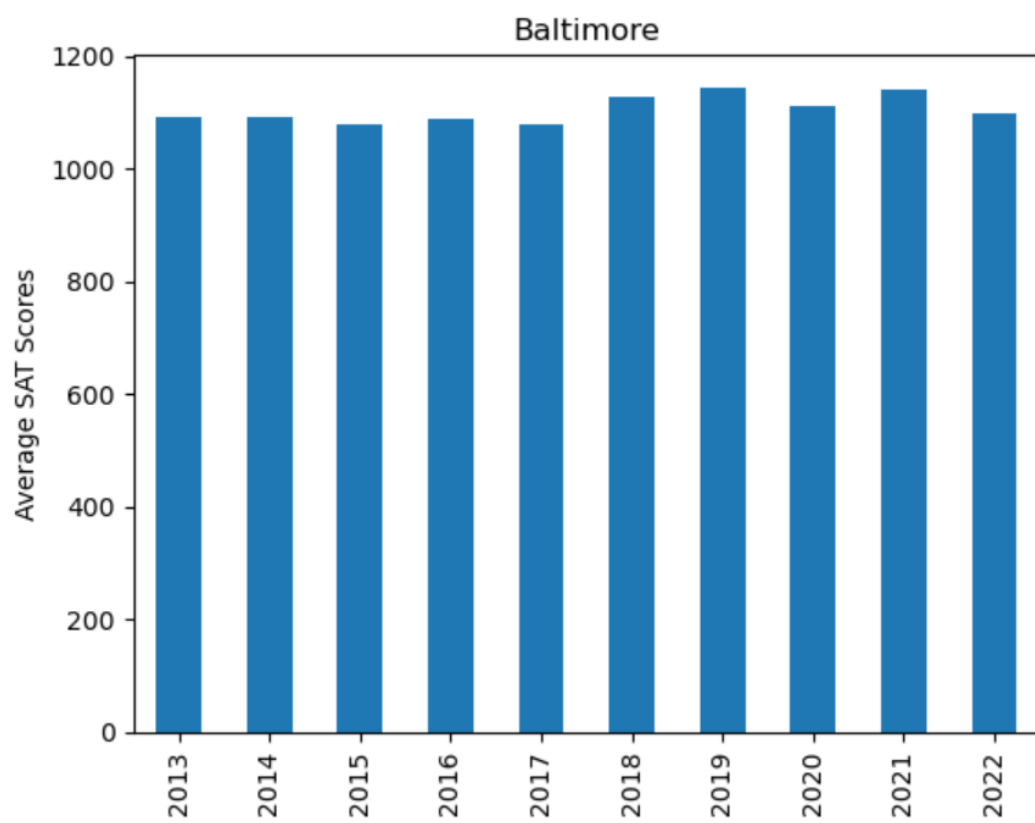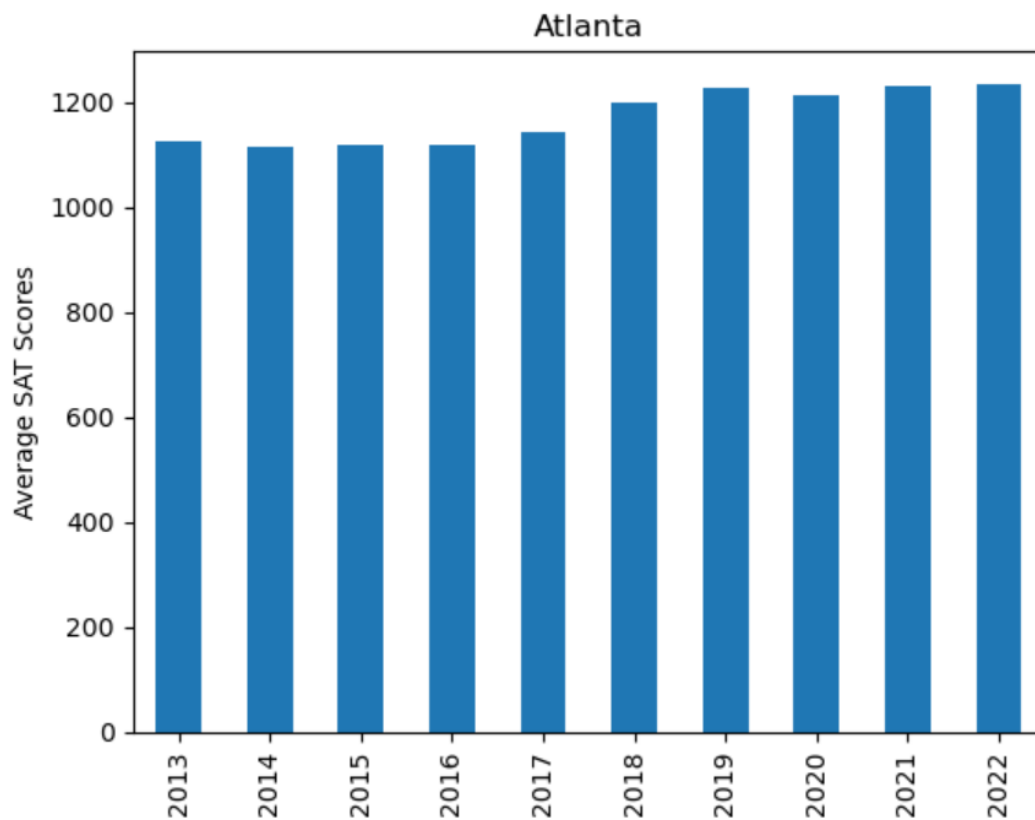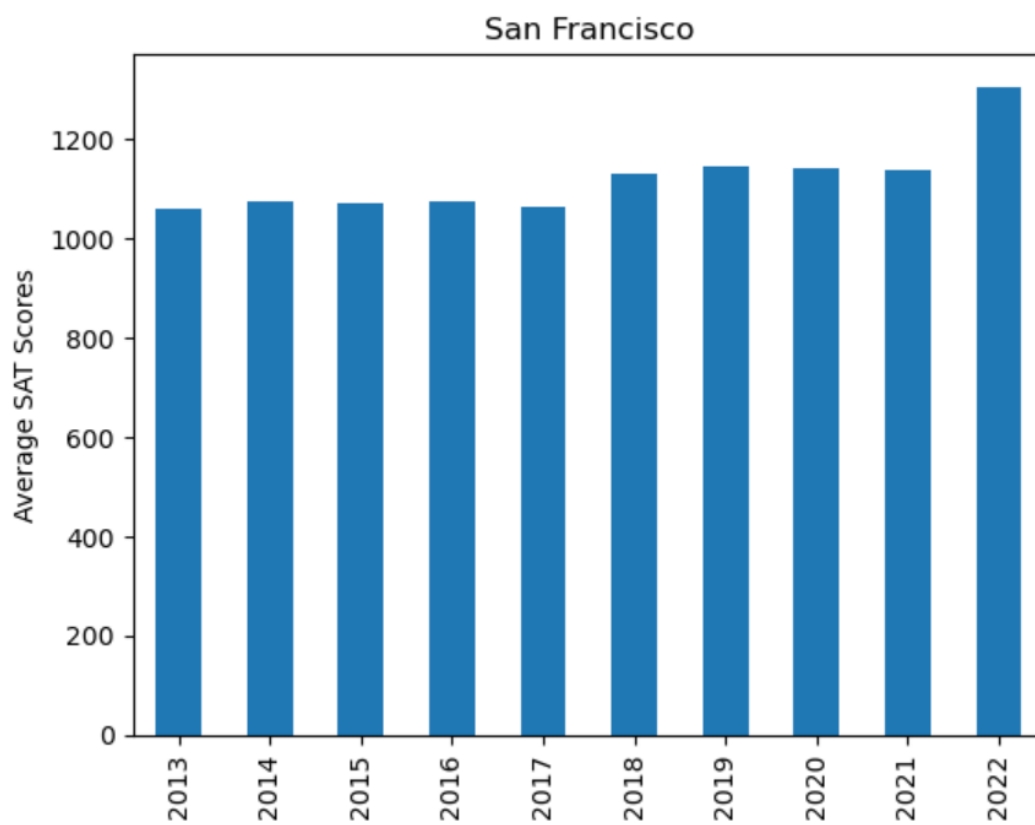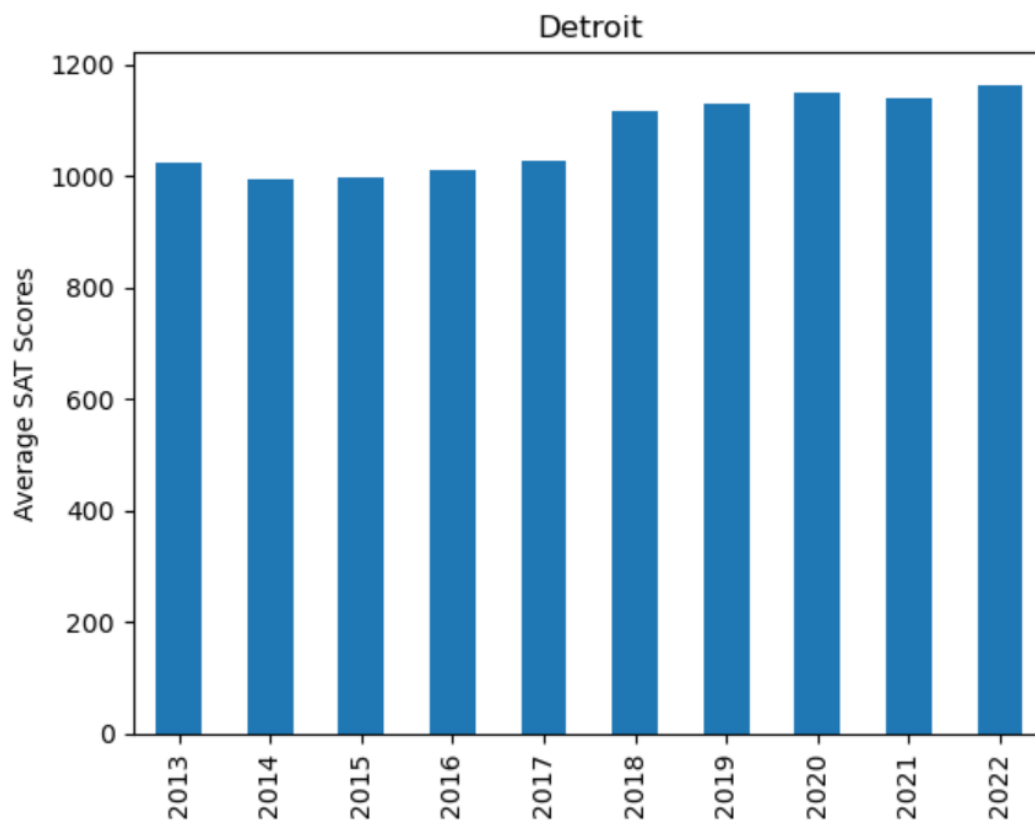
- Patterns-Trends-Outliers - (Identify any patterns, trends, or outliers in the data.)

4. Engraving and Printing
   - Main Findings
      - Now I wanted to ask the question: Is the printing of money increasing real estate prices? My assumption is that if money is being printed at the same rate housing prices are going up, that there is correlation between the printing of money itself and the increase in home values. For this I gathered the annual production reports, totaled them, and graphed them over the same 10-year span.
      - The graph shows that despite the popular opinion that printing money is the major cause of inflation, we have been printing less per year as a general trend. Not to believe it was just one ten years span I took the data from the previous 10 years as well.
   - Variables – (Summarize the distribution of the variables (mean, median, mode, standard deviation, etc.)
      - Summarize the distribution of the variables:
         - The dataset provides information on the annual production of printed money by the Bureau of Engraving and Printing.
         - Data includes production figures for different denominations of currency from fiscal year 2012 to fiscal year 2022.
         - Initial data extraction involved scraping data directly from the Bureau of Engraving and Printing website.
         - Data cleaning steps included merging multiple DataFrames, renaming columns, and selecting relevant columns for analysis.
         - The dataset consists of production figures for each fiscal year, allowing for an analysis of the distribution of printed money over time.
         - Summary statistics such as mean, median, mode, and standard deviation were not explicitly provided but could be calculated from the dataset to analyze the distribution of printed money production across fiscal years.
      - Describe the methods and techniques used for analysis:
         - Data was extracted directly from the Bureau of Engraving and Printing website using pandas' read_html function.
         - Multiple DataFrames were merged to combine production figures for different denominations of currency.
         - Filtering and renaming columns were performed to prepare the dataset for analysis and visualization.
         - The sum of printed money for each fiscal year was calculated to understand the total production figures.

- Data visualization techniques such as bar plots were used to visualize the distribution of printed money production over the fiscal years, providing insights into trends and variations in currency production.

- Visualizations - (Visualize the data using histograms, box plots, scatter plots, etc.)
  - Sample plot representing Printing of Money over time.


Money Printed Each Year in Billions
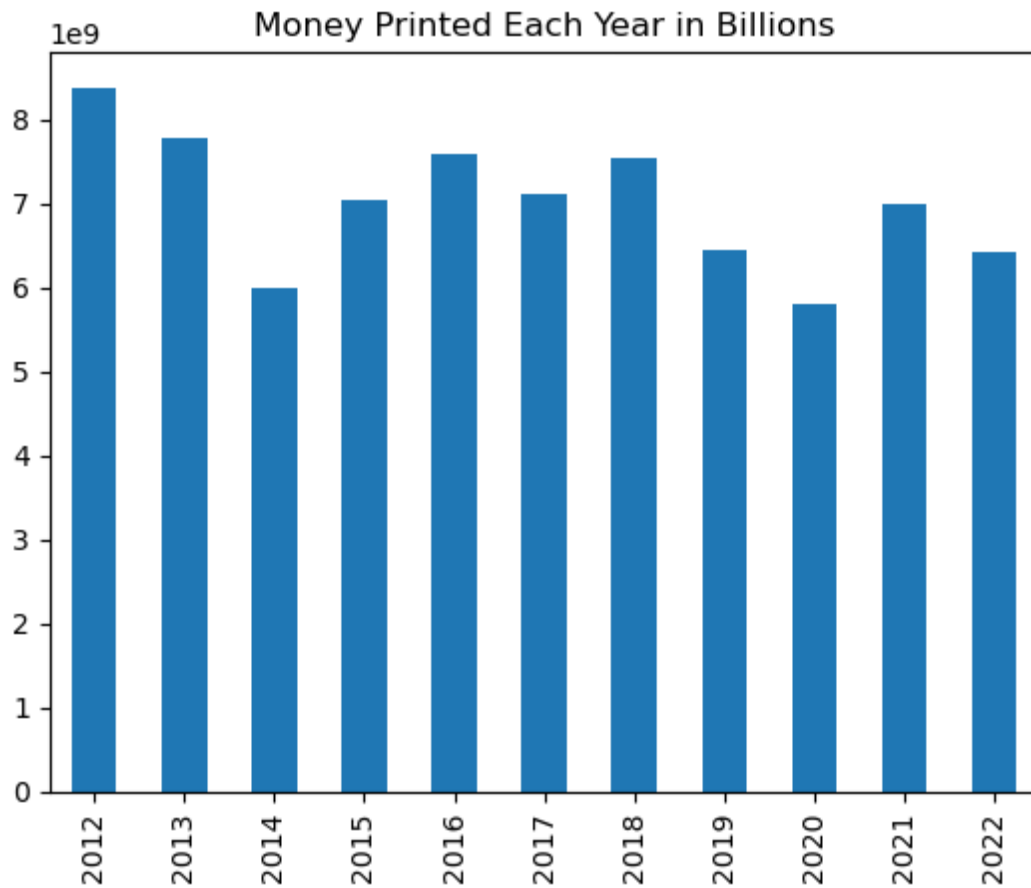
- Patterns-Trends-Outliers - (Identify any patterns, trends, or outliers in the data.)
  - Sample combination plots of converted graphs to z-score.

Atlanta

Baltimore

Detroit

New Orleans
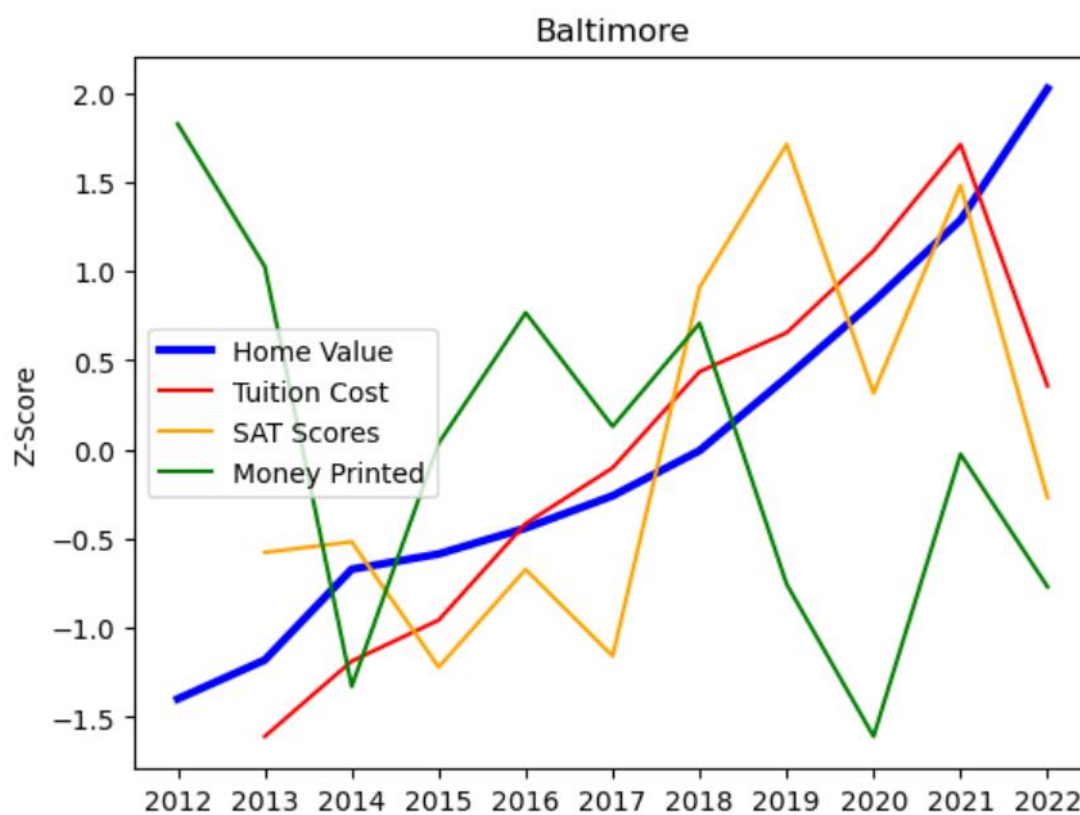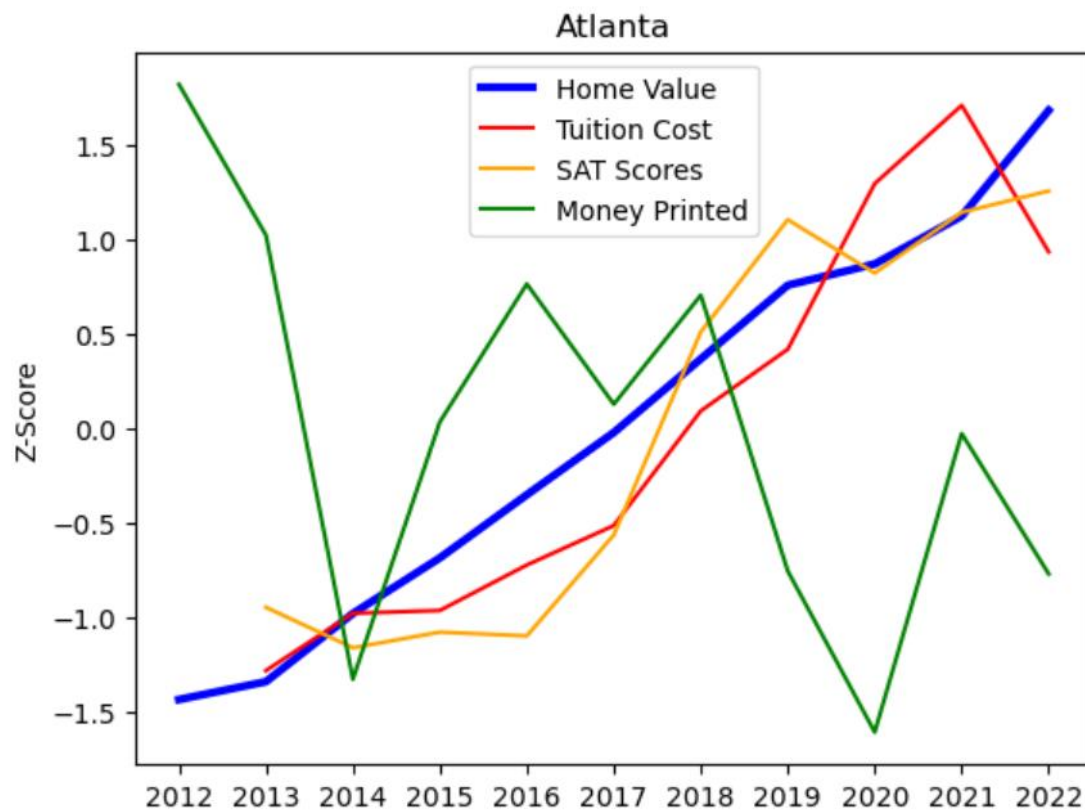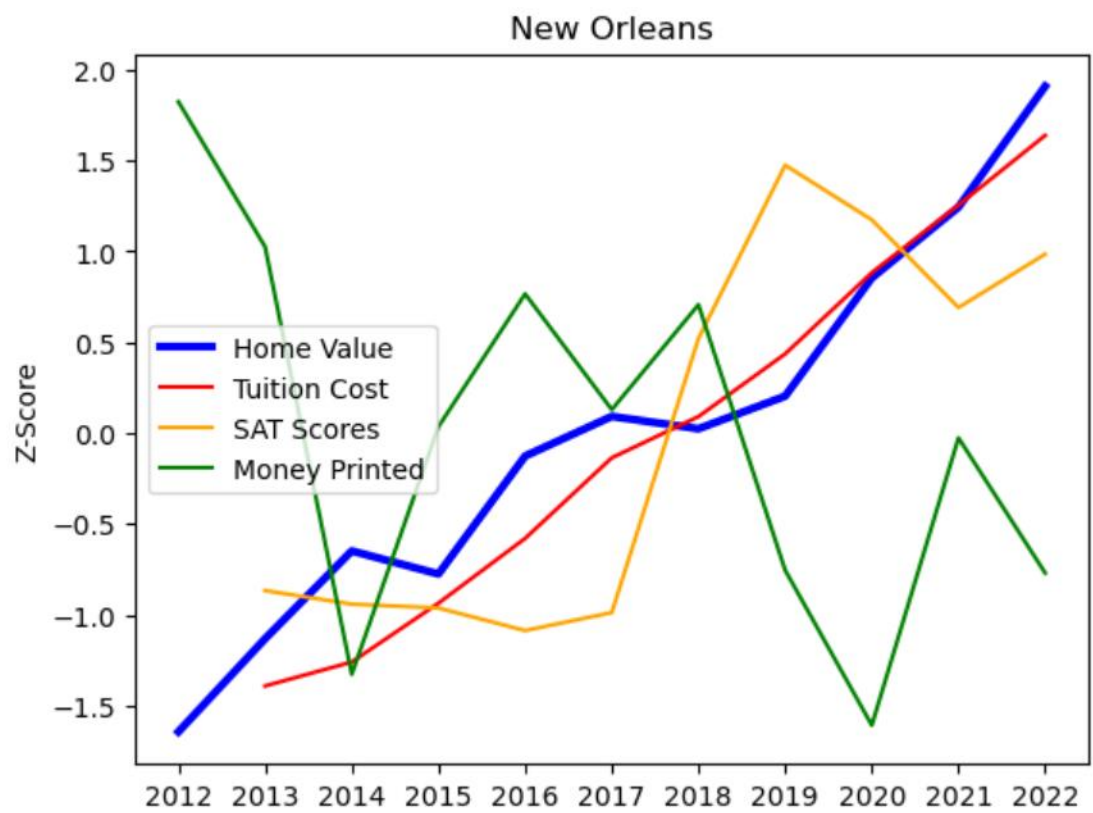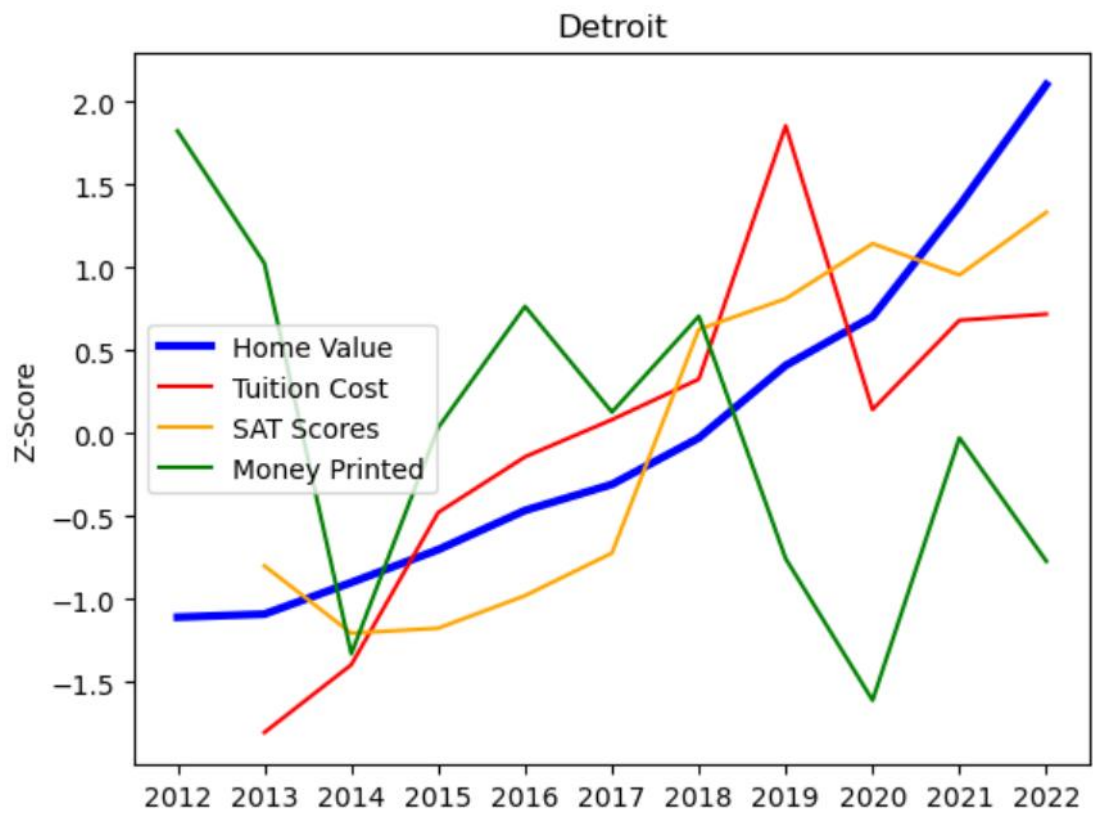
5. Crime
   - Main Findings
     - This analysis examines the relationship between crime rates and real estate prices in ten major U.S. cities, using data sourced from the FBI's Uniform Crime Reporting (UCR) database. The dataset includes annual figures for both violent and property crimes, organized by year and state.
     - To facilitate a comparative analysis across diverse urban populations, crime figures were normalized by the respective city populations provided in the FBI data. This adjustment allowed us to express the crime rate as a percentage of the population.
     - The guiding hypothesis for this study was that an increase in crime rates would correlate with a decrease in real estate prices in the examined cities. The subsequent sections detail the methods used for analyzing the data, the statistical tests employed, and the findings of this investigation.
     - Conclusion - The analysis suggests that while there are varied strengths and directions of correlation between crime rates and real estate prices, the universal increase in property values across these cities indicates that crime statistics alone are unlikely to be a significant determinant of real estate prices.
   - Variables – (Summarize the distribution of the variables (mean, median, mode, standard deviation, etc.)
     - Summarize the distribution of the variables.
       - Mean, median, mode, standard deviation, etc., for both real estate prices and crime rates are not explicitly calculated in the provided script. These statistics can provide insights into the central tendency, variability, and distribution of the data, which are essential for understanding the characteristics of each variable.
     - Describe the methods and techniques used for analysis.
       - The primary technique used for analysis seems to be correlation analysis, specifically Pearson correlation coefficients. This technique helps in understanding the linear relationship between real estate prices and crime rates.
       - Regression analysis could be another useful technique to explore the relationship between these variables, especially if we want to predict real estate prices based on crime rates or vice versa.
       - Classification or clustering techniques are not explicitly mentioned in the provided script, but they could be explored to identify patterns or groups within the data if relevant.

   - Visualizations - (Visualize the data using histograms, box plots, scatter plots, etc.)
     - Sample Plots for Crime data analysis.

Pearson Correaltion Real Estate and Crime 2012-2022

Atlanta Real Estate and Crime
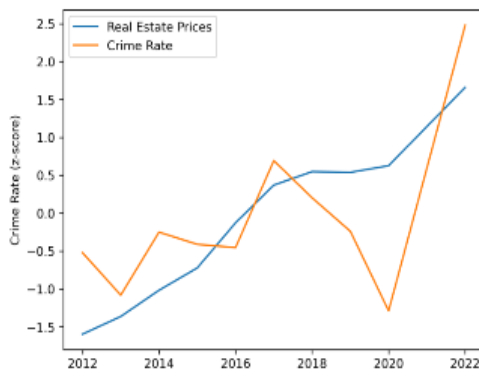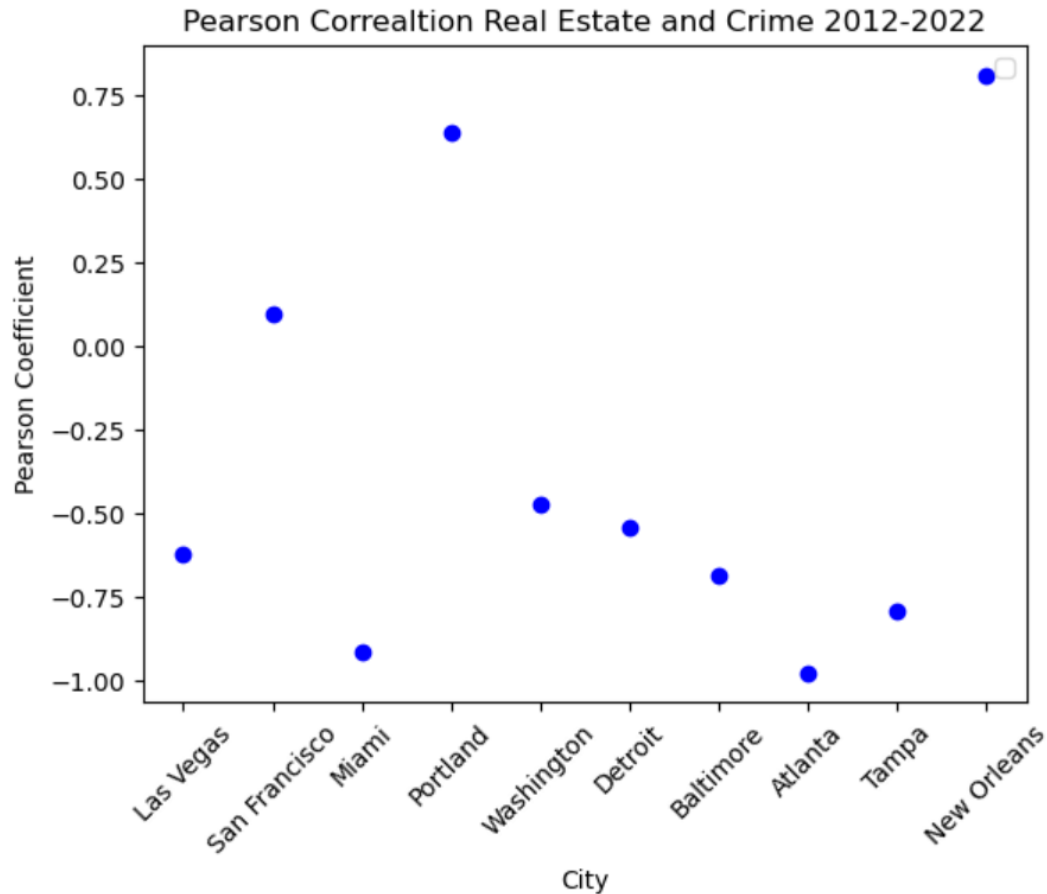
- Patterns-Trends-Outliers - (Identify any patterns, trends, or outliers in the data.)
  - Contextual Overview: - While Pearson coefficients can indicate the nature of the relationship between two variables, the consistent increase in real estate prices across all cities suggests that factors other than crime may be influencing real estate values. The coefficients range from -1 to 1, where values close to 1 or -1 indicate strong positive or negative correlations, respectively, and values near 0 suggest a lack of linear correlation.
  - City-Specific Correlation Insights:
    - Las Vegas (-0.621): Despite a moderate negative correlation, real estate prices in Las Vegas have risen.

- San Francisco (0.094): The very weak positive correlation in San Francisco is nearly negligible. Regardless of the increase in crime, it seems that crime rates have little to no linear impact on the increasing real estate prices.
- Miami (-0.913): Although there is a very strong negative correlation, the continuous increase in real estate prices indicates that crime rates might not be a significant deterrent to property value appreciation in Miami.
- Portland (0.639): The positive correlation here is more aligned with rising real estate prices. However, despite the massive spike in recent crime, real estate prices continue to climb.
- Washington (-0.473) and Detroit (-0.541): Both cities show moderate negative correlations.
- Baltimore (-0.687): A moderate negative correlation is observed.
- Atlanta (-0.977): This extremely strong negative correlation might imply a significant relationship, but the overall increase in property values across the board indicates that other variables are more pivotal in influencing the real estate market.
- Tampa (-0.791): Again, Despite the strong negative correlation, the increase in real estate prices points to other factors playing a more crucial role than crime rates in driving market dynamics.
- New Orleans (0.811): A strong positive correlation suggests that increases in crime could coincide with higher property values.

6. Population
   - Main Findings
      - The initial hypothesis posited that increases in population would correspond with rises in housing prices. This assumption is based on the economic principle that limited housing supply coupled with increased demand—driven by population growth—would naturally lead to higher real estate prices. The subsequent sections of this report detail the methods used to analyze the data and the results of our statistical tests.
      - Summary and Implications
         - Strong Positive Correlations (Las Vegas, Miami, Portland, Washington, Atlanta, Tampa): In these cities, the variable studied alongside housing prices seems to either contribute to or increase in conjunction with housing prices, suggesting a potential area for further investigation into causative factors or reinforcing trends.
         - Strong Negative Correlations (Detroit, Baltimore): These indicate that the rising housing prices are occurring despite decreases in the other variable, which could highlight resilience in the housing market or effectiveness of counteracting policy measures.
         - Negligible Correlations (San Francisco, New Orleans): Here, housing price increases appear to be decoupled from the changes in the

studied variable, suggesting that other factors (economic, demographic, or policy-driven) are influencing the housing markets.

- Variables – (Summarize the distribution of the variables (mean, median, mode, standard deviation, etc.)
    - Summarize the distribution of the variables.
        - Mean, median, mode, standard deviation, etc., for both population changes and real estate prices are not explicitly calculated in the provided script. These statistics would offer insights into the central tendency, variability, and distribution of the data.
    - Describe the methods and techniques used for analysis.
        - The primary technique used here is correlation analysis, specifically Pearson correlation coefficients. This helps in understanding the linear relationship between population changes and real estate prices.
    - Data Scaling and Preparation:
        - Population data is extracted and scaled using z-score normalization to make it comparable with real estate prices.
        - The scaled population data is then combined with the real estate dataset into a single DataFrame.
    - Correlation Analysis:
        - Pearson correlation coefficients are calculated between real estate prices and population changes for each city.
        - The correlation coefficients provide insights into the linear relationship between these two variables.
    - Visualization:
        - Scatter plot: Displays Pearson coefficients for each city, representing the correlation between real estate prices and population changes.
        - Line graphs: Show the trends of real estate prices and population changes for each city over the years.

- Visualizations - (Visualize the data using histograms, box plots, scatter plots, etc.)
    - Sample plots of Population correlations.

Pearson Correaltion Real Estate and Population Change 2012-2022
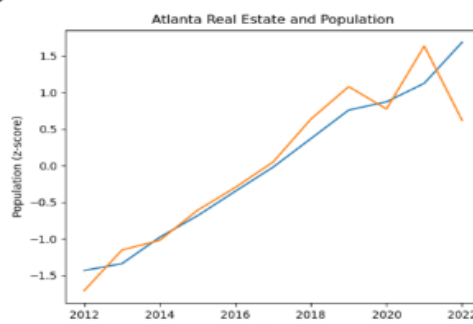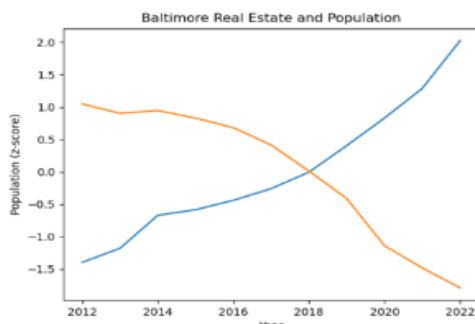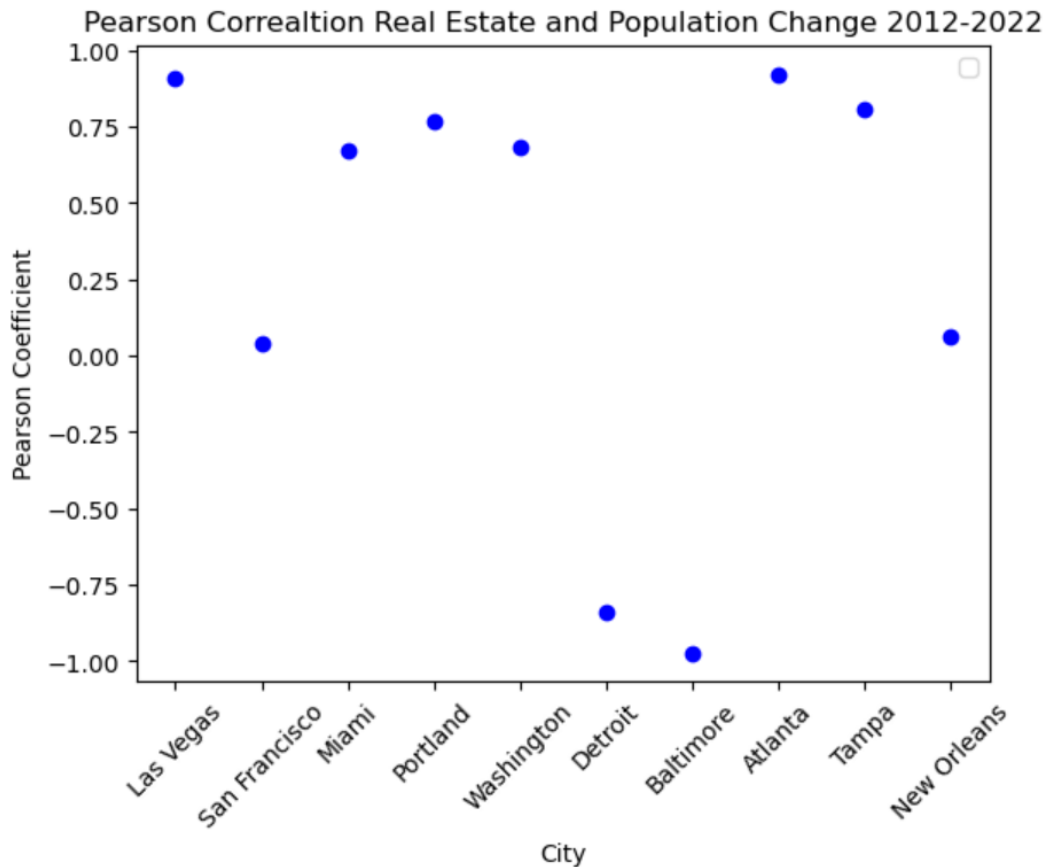
- Patterns-Trends-Outliers - (Identify any patterns, trends, or outliers in the data.)
  - Refined Analysis of Correlation Coefficients
    - Las Vegas (0.909): The very strong positive correlation in Las Vegas indicates that the increase in housing prices is closely aligned with the increase in the second variable. This suggests that both variables may be driven up by similar factors or that they positively reinforce each other.
    - San Francisco (0.043): Despite the general rise in housing prices, the negligible correlation in San Francisco suggests that the other

variable does not significantly impact or is not impacted by housing prices. This implies independent factors influencing each.

- Miami (0.674): A moderate to strong positive correlation in Miami indicates that as housing prices increased, the second variable also generally increased, suggesting a possible synergistic or parallel influence between the two.
- Portland (0.770): With a strong positive correlation, Portland's data implies that rises in the second variable might have complemented the housing price increases, suggesting a possible contributory factor to the housing market's dynamics.
- Washington (0.685): This strong positive correlation suggests that the factors driving up housing prices in Washington are potentially linked to increases in the second variable, indicating a related or mutually reinforcing relationship.
- Detroit (-0.837) and Baltimore (-0.972): These cities show strong negative correlations despite rising housing prices, indicating that the other variable moved inversely relative to housing prices. This could imply that rising housing prices might be happening despite negative trends in the other variable, possibly due to external economic or policy factors.
- Atlanta (0.923): Almost a perfect positive correlation, suggesting that in Atlanta, the variable closely tracked with the rise in housing prices, potentially indicating a very strong complementary relationship.
- Tampa (0.810): Another example of a strong positive correlation, where the increase in the second variable might be seen as a contributor to or moving in tandem with the increase in housing prices.
- New Orleans (0.065): Similar to San Francisco, the very weak correlation here suggests that the rise in housing prices is largely independent of the other studied variable, indicating that other factors are at play.

- Summary and Implications
    - Strong Positive Correlations (Las Vegas, Miami, Portland, Washington, Atlanta, Tampa): In these cities, the variable studied alongside housing prices seems to either contribute to or increase in conjunction with housing prices, suggesting a potential area for further investigation into causative factors or reinforcing trends.
    - Strong Negative Correlations (Detroit, Baltimore): These indicate that the rising housing prices are occurring despite decreases in the other variable, which could highlight resilience in the housing market or effectiveness of counteracting policy measures.

- Negligible Correlations (San Francisco, New Orleans): Here, housing price increases appear to be decoupled from the changes in the studied variable, suggesting that other factors (economic, demographic, or policy-driven) are influencing the housing markets.

7. Homelessness
   - Main Findings
     - For this analysis, ten annual datasets spanning various years were consolidated into a single DataFrame. Specific data relevant to urban centers and homeless counts were then extracted into another DataFrame for detailed examination. Notably, data for the years 2018 and 2019 was unavailable, prompting their exclusion from the analysis.
     - To ensure accurate comparative analysis across different scales, the data was normalized using z-scores. These standardized values were then analyzed using a Pearson Correlation Matrix to identify potential correlations between homelessness rates and real estate values in each city. The resulting correlation coefficients were visually represented in a scatter plot to provide an initial overview of the relationships.
     - Conclusion - The analysis highlights a clearer and more consistent pattern of negative correlations between homelessness and housing prices outside of pandemic years. This finding suggests that typical market dynamics, free from emergency interventions and temporary housing measures, may more reliably predict housing market responses to changes in homelessness. Policymakers and stakeholders should consider these underlying trends when designing long-term strategies to address housing and homelessness issues.

   - Variables – (Summarize the distribution of the variables (mean, median, mode, standard deviation, etc.)
     - Summarize the distribution of the variables.
       - The script does not explicitly compute mean, median, mode, or standard deviation for the variables. These statistics would provide insights into the central tendency and variability of both homelessness rates and real estate prices.
     - Describe the methods and techniques used for analysis.
       - The primary technique used here is correlation analysis, specifically Pearson correlation coefficients, to understand the relationship between homelessness rates and real estate prices.
     - Data Preparation:
       - The script fetches homelessness data for specific years and cities, then filters and processes it to create a DataFrame.
       - Homelessness numbers are adjusted to percentages of the population for normalization.

- Data Scaling and Integration:
  - Homelessness data is scaled using z-score normalization to make it comparable with real estate prices.
  - The scaled homelessness data is then combined with the real estate dataset into a single DataFrame.
- Correlation Analysis:
  - Pearson correlation coefficients are calculated between real estate prices and homelessness rates for each city.
  - These coefficients provide insights into the linear relationship between real estate prices and homelessness rates.
- Visualization:
  - Scatter plot: Displays Pearson coefficients for each city, representing the correlation between real estate prices and homelessness rates.
  - Line graphs: Show the trends of real estate prices and homelessness rates for each city over the years.

- Visualizations - (Visualize the data using histograms, box plots, scatter plots, etc.)
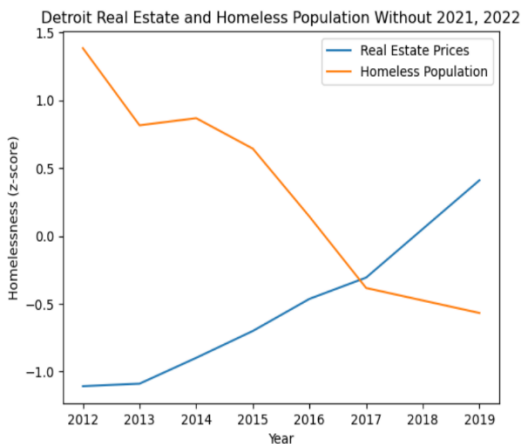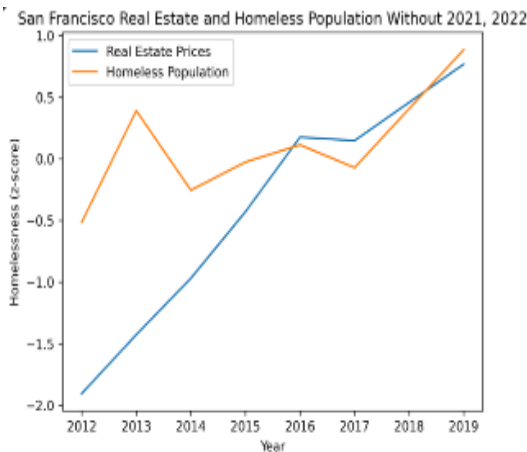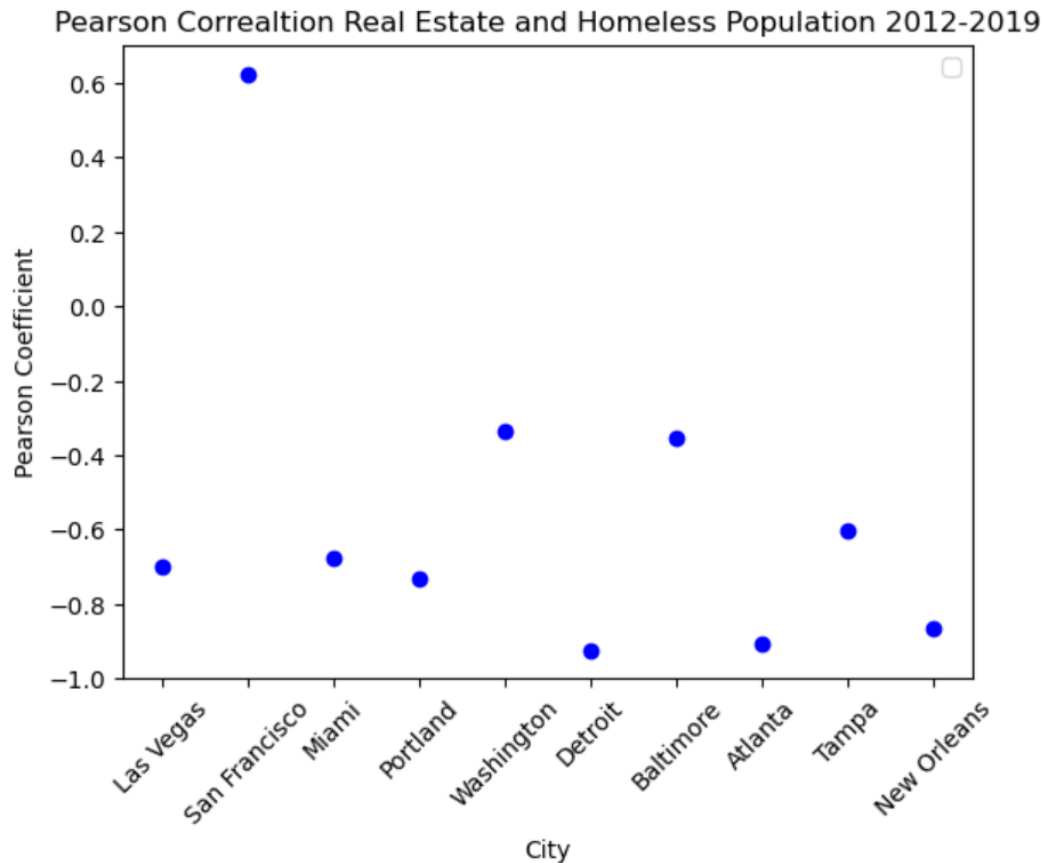 - Sample plots showing homelessness correlation with Post Pandemic.

- Sample plots showing homelessness correlation with Post Pandemic.

Pearson Correlation Real Estate and Homeless Population 2012-2019
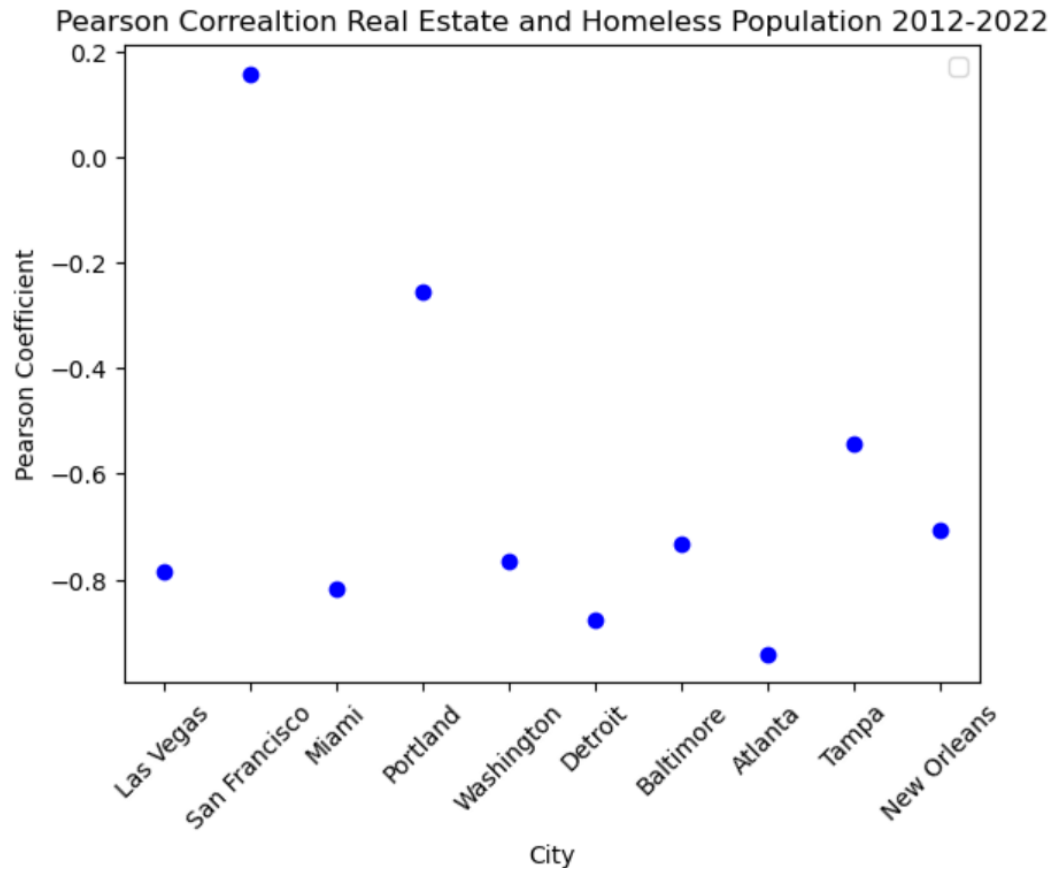
San Francisco Real Estate and Homeless Population Without 2021, 2022

Detroit Real Estate and Homeless Population Without 2021, 2022
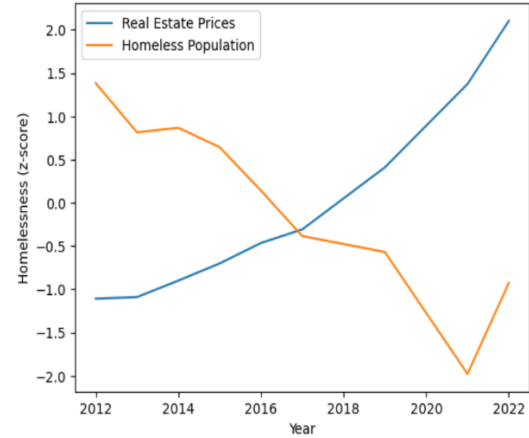
- Patterns-Trends-Outliers - (Identify any patterns, trends, or outliers in the data.)
  - Current Study Results (Excluding Pandemic Years)
    - Las Vegas (-0.698): Shows a strong negative correlation, indicating that as homelessness decreases, housing prices likely increase. This correlation is somewhat less negative compared to the pandemic-influenced data, suggesting less volatility in non-pandemic conditions.

- San Francisco (0.625): Displays a strong positive correlation, a significant change from the previous negligible correlation, indicating that both homelessness and housing prices may rise in tandem under normal circumstances.
- Miami (-0.677): Exhibits a strong negative correlation, consistent with previous findings, suggesting a robust inverse relationship between homelessness and housing prices.
- Portland (-0.734): A strong negative correlation, more pronounced than during the pandemic, indicating a clearer inverse relationship between homelessness and housing prices when pandemic distortions are removed.
- Washington (-0.337): Shows a moderate negative correlation, much weaker than in other cities, suggesting that other factors may be influencing housing prices beyond homelessness.
- Detroit (-0.923): This extremely strong negative correlation, stronger than in the pandemic data, underscores a significant inverse relationship, likely reflecting more stable market dynamics outside pandemic conditions.
- Baltimore (-0.354): A weak negative correlation, indicating that the impact of homelessness on housing prices is less pronounced.
- Atlanta (-0.907): Very strong negative correlation, consistent with the earlier findings, suggesting a stable pattern where increases in homelessness are associated with decreases in housing prices.
- Tampa (-0.603): Shows a strong negative correlation, indicating a consistent inverse relationship between homelessness and housing prices.
- New Orleans (-0.864): A very strong negative correlation, indicating a robust inverse relationship, more pronounced than during the pandemic years.

- Comparative Insights
  - The removal of pandemic years generally reveals more pronounced negative correlations between homelessness and housing prices, except for San Francisco which turned strongly positive. This suggests that:
  - Under normal conditions, the relationship between homelessness and housing prices is typically more stable and consistently negative, indicating that increases in homelessness tend to correspond with decreases in housing prices.
  - Pandemic-related interventions might have temporarily altered or obscured these trends, especially in cities like San Francisco where the correlation reversed.
- Conclusion
  - The analysis highlights a clearer and more consistent pattern of negative correlations between homelessness and housing prices

outside of pandemic years. This finding suggests that typical market dynamics, free from emergency interventions and temporary housing measures, may more reliably predict housing market responses to changes in homelessness. Policymakers and stakeholders should consider these underlying trends when designing long-term strategies to address housing and homelessness issues.

Conclusion:

Key Findings

Conclusions:

Real Estate Market Dynamics:

Across all ten markets analyzed, real estate prices consistently increased each year. However, this growth was not uniform; the five underperforming markets experienced modest increases compared to the explosive growth observed in the top-performing markets. The lack of a control group makes it challenging to establish robust correlations or causations between observed trends and specific variables.

Data Reliability Concerns:

The study faced significant challenges with data reliability:

FBI Crime Data: This data is reported voluntarily by municipalities, which frequently fail to submit annual updates, compromising data completeness.

Homelessness Counts: Accurately collecting homelessness data proved difficult as it largely depends on NGOs to organize and report counts, which can vary widely in methodology and completeness.

Public Perception vs. Reality:

Our findings challenge several prevailing public perceptions regarding crime, homelessness, monetary expansion, and tuition increases. Contrary to popular narratives, these variables either declined or remained stable, suggesting that common assumptions about their trends may be unfounded.

Influence of Inflation:

The study period was marked by rapid inflation, complicating the analysis as the general increase in prices across the board made it difficult to isolate specific causes for changes in real estate prices. This inflationary context adds ambiguity to the results, as it is challenging to distinguish the effects of individual variables.

Correlation Analysis:

The study did not uncover any definitive correlations that could comprehensively explain the housing cost crisis. However, a significant finding was the strong positive correlation between housing prices and population growth, observed in six of the ten markets with Pearson Coefficients of 0.6 or greater. This suggests that housing price increases in these areas could largely be driven by supply and demand dynamics.

In the analysis, a conspicuous pattern emerges: a robust negative correlation between homelessness and housing prices. Among the ten markets scrutinized, three exhibited a modest negative correlation, while seven displayed a pronounced negative correlation. It's plausible that this correlation doesn't imply causation; rather, housing prices might be escalating due to distinct factors, while homelessness experiences a decline owing to unrelated causes.

Summary of Results:

Tuition Prices: Exhibited a weak positive correlation with housing prices.

SAT Scores: Also showed a weak positive correlation.

Monetary Expansion: No significant correlation was found.

Crime Rates: No significant correlation was observed.

Population Increase: Demonstrated a strong positive correlation, supporting a supply-demand explanation for housing price increases.

Homelessness: Strong negative correlation

Overall Insights:

The analysis underscores the complexity of the real estate market and points to population growth as a significant factor influencing housing prices. The absence of correlations with other variables like crime and homelessness highlights the need for a nuanced understanding of what truly drives housing markets. This study suggests that economic fundamentals such as population dynamics may have a more direct impact on housing prices than previously thought.

Limitations and Future Study

Our study was constrained by several limitations stemming from the scope and timeframe of the project. Firstly, the number of markets analyzed was limited due to logistical constraints and the complexities of capturing diverse market influences within the project's deadline.

Future endeavors could expand the scope by encompassing a comprehensive analysis of all markets across the United States, offering a more extensive perspective on potential correlations.

Moreover, our examination was restricted to only six potential influences on the market, overlooking crucial factors such as mortgage rates, short-term rentals, housing construction starts, and properties owned by investors. The absence of accessible data, particularly at the city level and within the timeframe of our study, hindered the exploration of these additional variables.

Additionally, recent developments in the market, such as the Federal Reserve's actions to temper the housing market by increasing interest rates, have introduced significant changes. It's essential to recognize that adjustments in market variables often exhibit a lag in influence. Thus, it remains pertinent to monitor the impact of recent occurrences, such as population declines in many urban centers, on the real estate market.

Findings in Relation to Project Objectives

The analysis underscores the complexity of the real estate market and points to population growth as a significant factor influencing housing prices. The absence of correlations with other variables like crime and homelessness highlights the need for a nuanced understanding of what truly drives housing markets. This study suggests that economic fundamentals such as population dynamics may have a more direct impact on housing prices than previously thought.

References:

- Refer to Dataset section for links to study data.
- Refer to inline links to study data formatted in text.

Sources:

- Python for Data Analysis - O'Reilly Books
- Class Lessons
- Instructor
- Teacher Assistant
- Classmates
- Xpert Learning Assistant
- GitHub Copilot
- Tabnine
- (python)
- (pandas)
- (ChatGPT 3.5)
- (COPILOT)