**Table of Contents**

**Executive Summary**

**Introductory Section**

“The housing market suck”, “Now it is so difficult to buy a home”, and “It used to be so easy to buy a house”, are all things people keep saying, but how true are these claims? With data and statistics, I aim to either reject or support these claims by answering “how the affordability of a home has changed over the years”.

My objectives are to create a statistic that describes the affordability of a home, gather related data based on year and state so that the question is answered at a per-state level, and finally analyze the data over the years.

**Methodology**

Objectives:

* Collect Data
* Process and normalize data
* Perform analysis of data
* Create visuals to aid in communicating analysis

Steps Taken:

**Script: load\_data.ipynd**

1. Initialize an SQLite database and tables to understand what data will be needed and used fully
2. Gather related found on the web

* Data Sources:
  + Federal Reserve Bank of St. Louis.: Info on median salary and mortgage rates
  + Zillow Research.: Info on typical house sale

1. Use numpy and pandas to normalize data and load it into the database

* Dropped data on Zillow typical house sale on all dates before 2009 because some states did not have data until then
* Nulls were spotted for two entries; to minimize skewing the data, they were replaced with the midpoint of a given state's next and previous month.
* Zillow gave data per month instead of per year, so data was aggregated by average to get the desired format

**Script: analysis.ipynd**

1. Create a statistic to determine what an “affordable house” is

* p = m / s
  + p: Percentage of mortgage will consume
  + m: median monthly mortgage rate
  + s: median monthly salary
* This statistic will automatically normalize the effect of price level on the view of affordability

1. Gather statistics that reflect the change in house affordability

* Change in median salary
* Change in median house cost
* Change in average mortgage interest rate

**Script: get\_visualization.ipynb**

1. Collect a GeoJson file on the US state's shape

* Data source: Kaggle (Author: Kate Gallo)

1. Using the shapes and plotly create a heat map of the state's affordability of houses for 2009 and 2023
2. Using plotly create a bar graph of the top 10 most expensive states to buy a house in for 2009 and 2023
3. Using plotly create a line plot of the change in affordability of houses in the US from 2009 - 2023

**Results**

**Discussion**

**Conclusion**

**References**

Federal Reserve Bank of St. Louis.

(n.d.). Economic indicators: Table for populating "salary", “us\_salary",

"us\_house\_sale", and us\_mortgage\_rate tables in housing.db.

Retrieved December 27, 2024, from

<https://fred.stlouisfed.org/release/tables?eid=259515&rid=24>,

<https://fred.stlouisfed.org/series/MSPUS>,

<https://fred.stlouisfed.org/series/MORTGAGE30US>

files: All csv files in Median Income and "United States.csv"

Zillow Research.

"Zillow Home Value Index (ZHVI) and Other Housing Market Data.":

Table for populating "house\_sale" table in housing.db

Accessed December 27, 2024.

https://www.zillow.com/research/data/  
 files: "MedianStates.csv"

Kaggle (Author: Kate Gallo)

USA states GeoJson: Used in creating visualizations of the US

Accessed December 31, 2024.

<https://www.kaggle.com/datasets/pompelmo/usa-states-geojson>

files: us-states.json

**Acknowledgment**