

# Housing Costs and Inflation



## Project Summary

### Project Overview

#### Motivation:

This project is to better understand the United States housing market by looking at the cost, size, and location of homes for sale across the country side by side with average income levels per area.

#### Objective:

Interested parties will be able to see the most cost-efficient areas to purchase a home to help influence their personal real-estate choices.

#### Scope:

This project must inform potential buyers looking at homes within the United States.

#### Stakeholders:

- Data Analyst
- Career Foundry Tudor
- Career Foundry Mentor
- Potential Real-Estate Customers
- Real-Estate Agents

#### Success Factors:

The project will be considered a success:

- If potential buyers can see which areas are likely to have homes within their budgets.
- If the project stays within the geographic region of the United States.

#### Assumptions:

- Larger homes will cost more.
- Areas with lower income levels will have lower housing costs.
- Older homes will cost less than newer ones.

### Constraints

- The project must adhere to the legal licensing rights of the data used.
- All tools and resources used must be either already owned by the analyst or free to use.

### Requirements:

- Identify which of the available variables affect the cost of a home.
- Create a Tableau storyboard to display the results of the analysis.
- Create a GitHub Repository for the project.

## Data Source

### Housing

- **Contents:** Current real estate listings in the United States by state and zip code, last updated in 2023.
- [Link to Data Source](#)
- **Usage rights:** Educational Purposes Only
- **Collection Method:** scraped from data on [www.realtor.com](http://www.realtor.com)
- **Reason For Inclusion:** This data set is both up to date and extensive. It shows relevant information for the needs of the analysis, and is already relatively clean.

### Income

- **Contents:** United States federal income tax data from the year 2020.
- [Link to Data Source](#)
- **Usage Rights:** Public Domain
- **Collection Method:** Government filing and taxes.
- **Reason for Inclusion:** This data set comes from a reliable source and is no more than 3 years old. It includes the data needed to compare listing price to income levels.

# Data Profile



## Project Preparation

### Data Cleaning

#### Housing

- Removed rows with blank values and status column from housing data set. (Kept blank values in “previously sold” column)
- Created two new dataframes: One with only homes that have no previous sale date, and one with only homes that do have previous sale dates.
- Removed duplicate rows from all 3 dataframes.

#### Income

- Removed irrelevant columns.
- Added state name to each row.
- Removed totals and blanks.
- Tidied formatting.

### Consistency Checks

#### Housing

- Checked for mixed data types. prev\_sold\_date is the only column with mixed data types.
- Checked for null values. prev\_sold\_date is the only column with null values, likely due to having no previous sale on those houses.
- Checked for and removed duplicate rows.

#### Income

- Merged sets into single dataframe.
- Checked for duplicate rows.
- Checked for null values
- Checked for mixed data types, removed single row of data to fix it.

# Descriptive Statistics

## Housing

	Unnamed: 0	bed	bath	acre_lot	zip_code	house_size	price
count	51960.00	51960.00	51960.00	51960.00	51960.00	51960.00	51960.00
mean	226637.87	3.71	2.69	13.49	8290.77	2376.16	778465.09
std	119142.66	1.82	1.83	890.45	5397.89	6927.25	1769405.04
min	0.00	1.00	1.00	0.00	601.00	122.00	500.00
25%	120325.75	3.00	2.00	0.10	4364.00	1309.00	274900.00
50%	284142.50	3.00	2.00	0.23	7436.00	1849.00	439000.00
75%	324824.75	4.00	3.00	0.70	10348.25	2700.00	749900.00
max	413082.00	99.00	198.00	100000.00	95652.00	1450112.00	169000000.00

## Income

	ZIP_Code	Num_Individuals	Total_Income
count	30947.00	30947.00	30947.00
mean	10855.76	1647.75	86702.80
std	6777.49	2723.72	288772.03
min	1001.00	0.00	0.00
25%	5660.00	210.00	8107.00
50%	11755.00	630.00	24695.00
75%	15344.00	1950.00	77809.50
max	99999.00	54570.00	12899619.00

## Data Limitations

### Housing

There are many missing cells throughout the data set. The listings are also only based on what is currently for sale, and thus any historical analysis would be limited to the few homes that have been sold previously. The previously sold listings only show last sale dates and do not include past sale amounts.

There are only listings available in 14 states, 4 of which are unable to be included in the analysis, either due to too few records, or due to having no income data to use with it. The analysis will have to be limited to the remaining 10 states.

### Income

Tax returns are filed manually on a large scale from multiple different data sources and are therefore subject to human error. Further information on the data can be found in Publication 55-B (Rev. 3-2023) ([irs.gov](https://www.irs.gov/publications/p55b))

## Ethical Considerations

### PII

It's possible that the income data provided may be traced back to individuals, and therefore it's important to ensure that any information used in the analysis takes privacy into account.

### Location

Some locations may be expected to be more costly than others, which could allow for bias in the analysis. Maintaining an open-minded perspective on the zip codes and state data will help prevent this.

# Questions to Explore



## Project Guide

- ? Is there a connection between income and the cost of housing?
- ? Out of the available data, what are the 10 most and 10 least expensive zip codes to buy a home?
- ? Out of the available data, what are the 3 most and least expensive states to buy a home?
- ? How much does square footage and lot size affect listing price?
- ? How do bedroom and bath count affect square footage?
- ? Do lot size or square footage have a greater affect on listing price than the other?
- ? How do older homes with previous sales vary in price from brand new ones?
- ? Does the duration of time between last sale of older homes affect current listing price?
- ? What states have the most previously sold homes?
- ? What states have the most brand-new homes?
- ? What percentage of homes in each zip code are within the average income level?
- ? What percentage of homes in each state are within the average income level?
- ? Is there a difference in house size over time for the homes with previous sales?