

# Zillow Analysis

## Project

---

Vishal Aiely, Sehee Hwang, Matt Mohandiss,  
Marc Muszik, Selena Xue





## TABLE OF CONTENTS

**01**

### **Project Overview**

Introduction &  
Objectives

**02**

### **TimeLine**

Timeline of the  
project

**03**

### **Process**

How we worked as a  
team

**04**

### **Project Approach**

How we were able to  
get the data

**05**

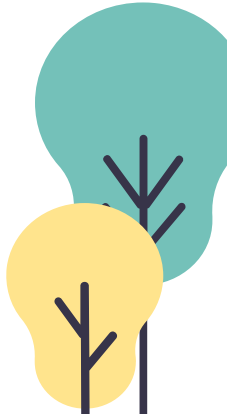
### **Data Analysis**

Models and results

**06**

### **Conclusion**

Overall thoughts,  
insights, challenges



# Introduction

Home ownership is a beacon of prosperity and a staple of the American Dream. Everyone takes part in the real estate market, whether buyer or seller, renter, or landlord. However, the real estate market can be difficult to follow as it is always in flux.

---

The objective for this project is to use digital archaeological techniques to retrieve data from Zillow and analyze trends in the housing market.



## Project Overview - Objectives

### 1 Zillow Scraper

Our goal was to develop a web scraping algorithm that will allow us to collect useful listing information from Zillow without getting blocked.

### 2 Data Collection

We collected any and all attributes that can affect a property's value, such as the location, number of bedrooms, bathrooms, square footage, lot size, and year remodeled, among others.

### 3 Data Analysis

We analyzed trends such as listing prices in cities, number of bedrooms in listings from various cities, median lot size by state, land prices in various states, trends in home type, median house prices, listings among seasons, and other regression analyses.



## TIMELINE

### Weeks 1-2

- Organize Team
- Successfully scrape the zillow web page



### Week 4

- Add feature to allow the scraper to accept zip codes
- Integrate Firestore and BigQuery for data storage



### Week 6

- Analyzing Data
- Visualizing Data



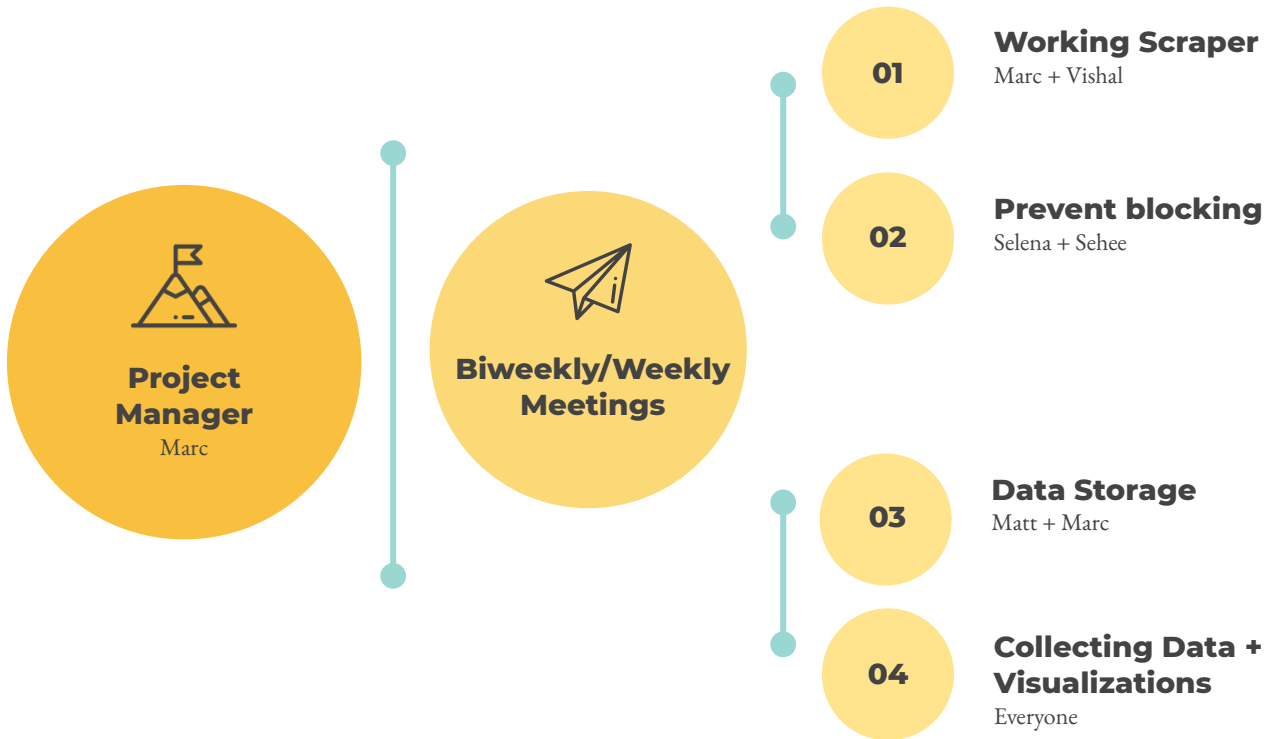
### Week 3

- Work on not getting blocked by Zillow
- Put the web scraper into a working function

### Week 5

- Cleaning Data
- Scrapping housing data using the web scraper

## STRATEGY



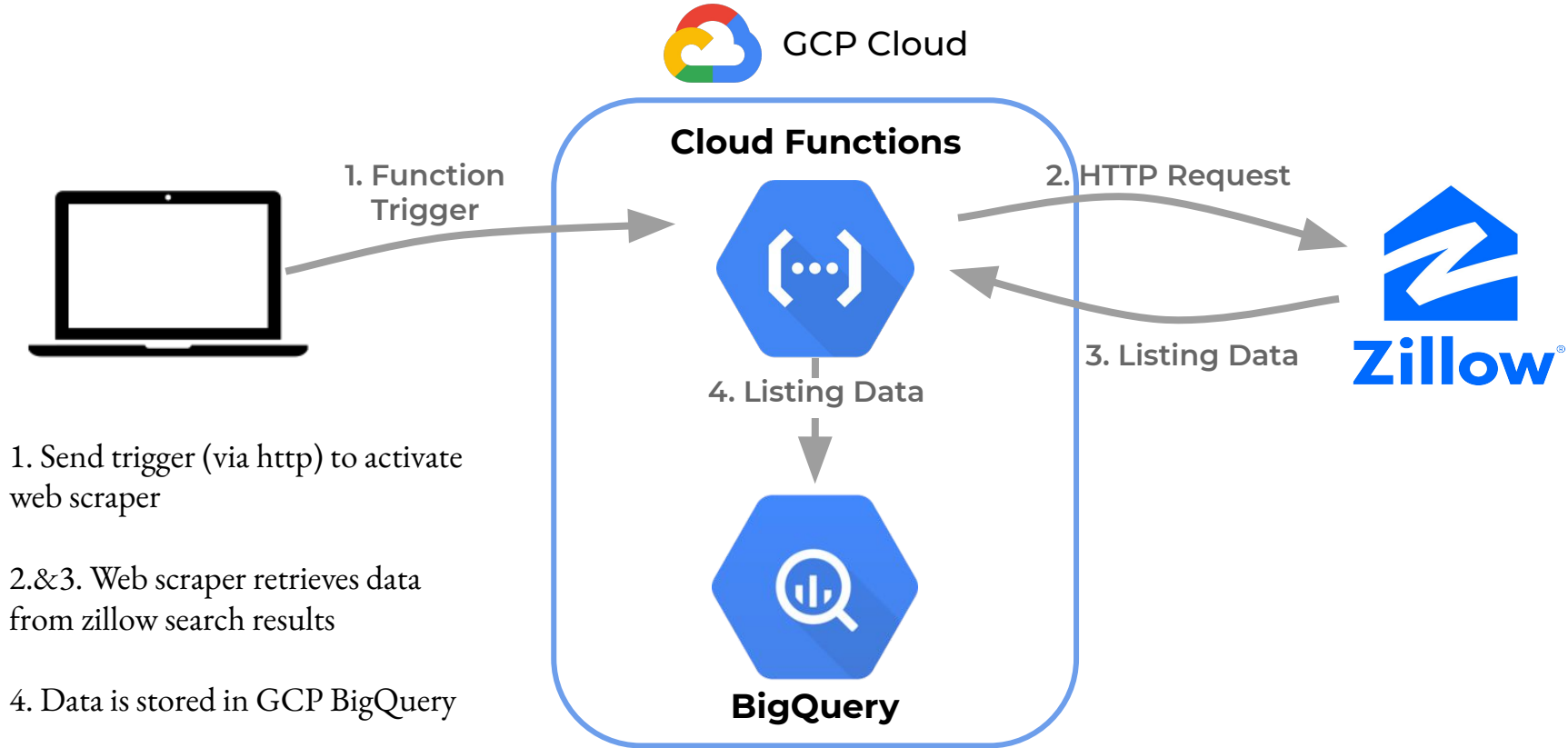


## Scraping and Parsing

```
vishalaiely at Vishals-MacBook-Pro in ~/D/c/f/C/ZillowAnalysis (main|...1)
» python3 parser.py
```



## Data Collection





## Data Summary



**174,657**

Listings



**18,009**

Zip codes

Team 1

60% completed

40% remaining

0

10000

\*as of 11/30/21 4:30pm

## Property Types



**80,259**

lots



**70,070**

single-family homes



**9,450**

condos



**5,841**

townhomes



**4,587**

multi-family homes



**4,168**

manufactured

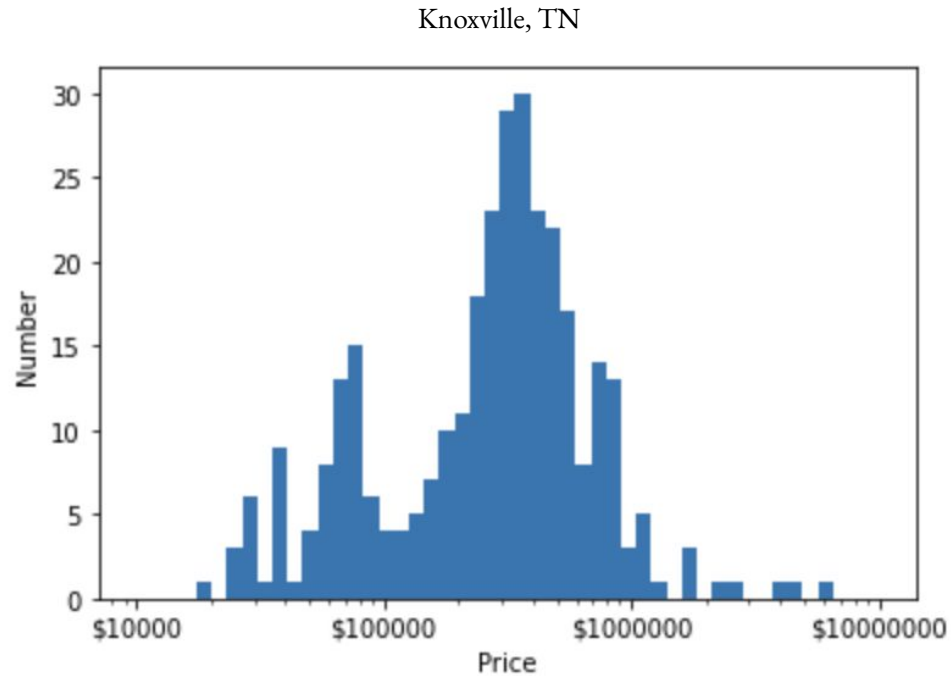


**282**

apartments

## Data Analysis - Knoxville Listing Prices

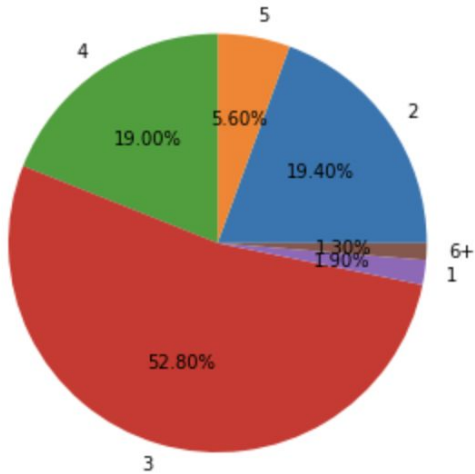
**What is the price range of listings in Knoxville, TN?**



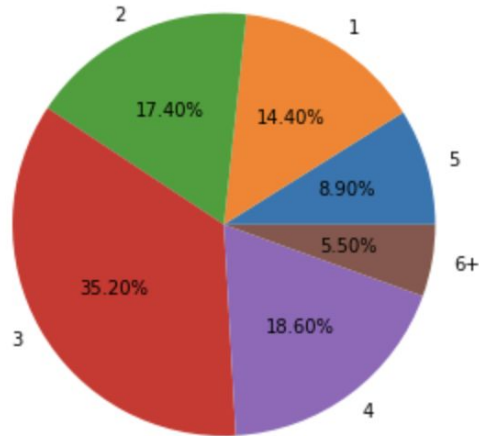
## Data Analysis - Bedrooms

How many bedrooms do the listings in the major cities of Tennessee have?

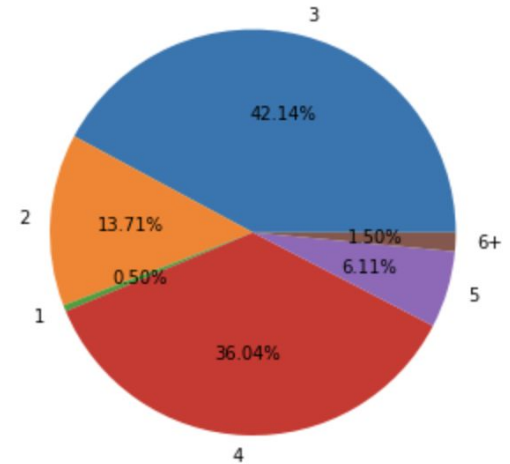
Memphis, TN Bedrooms



Nashville, TN Bedrooms



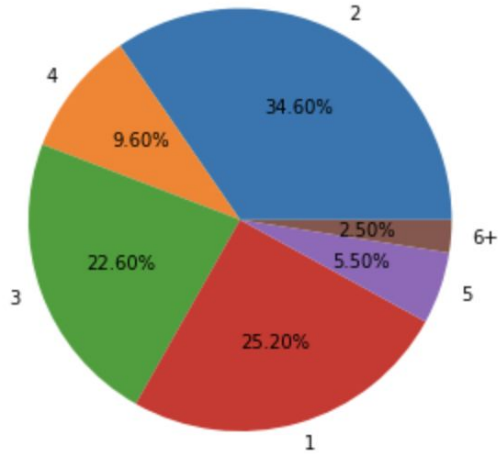
Knoxville, TN Bedrooms



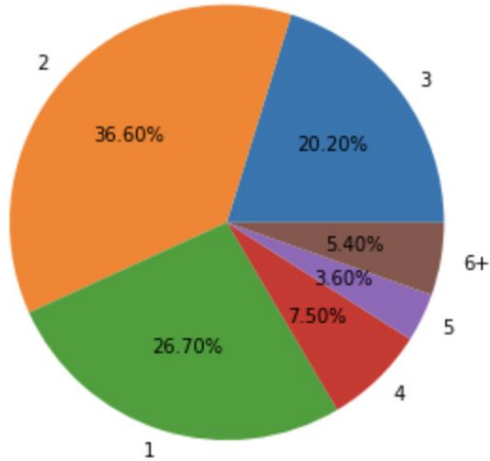
## Data Analysis - Bedrooms Part 2

How many bedrooms do the listings in the other major cities have?

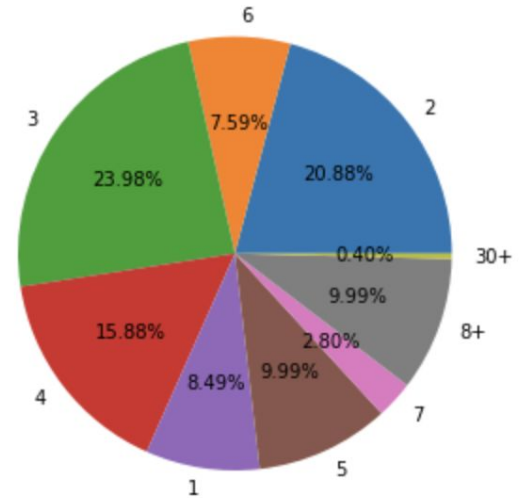
Seattle, WA Bedrooms



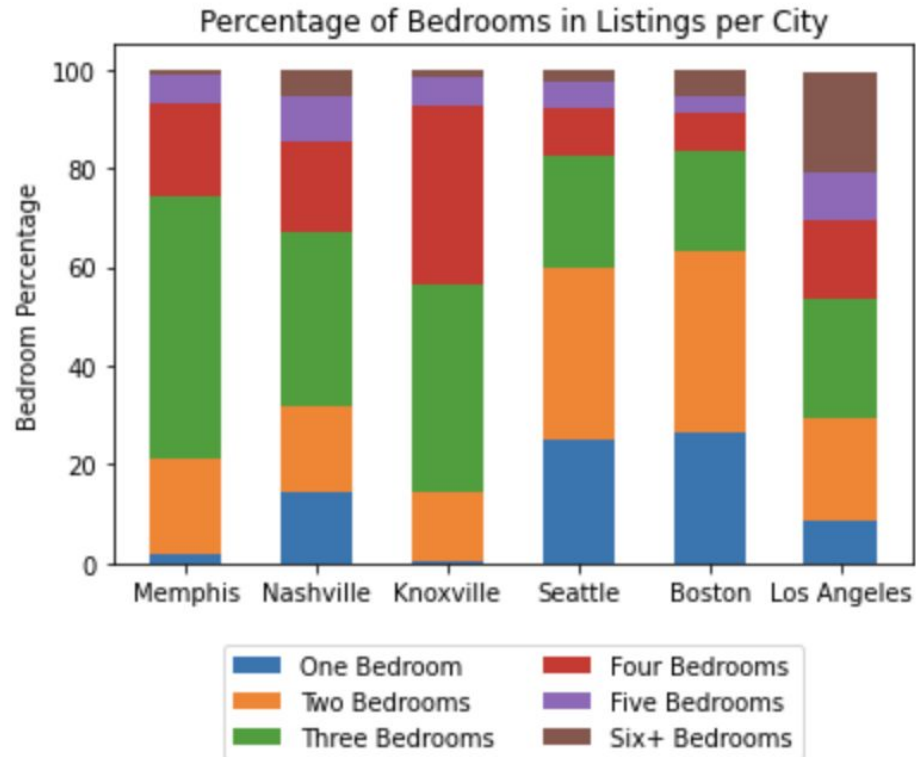
Boston, MA Bedrooms



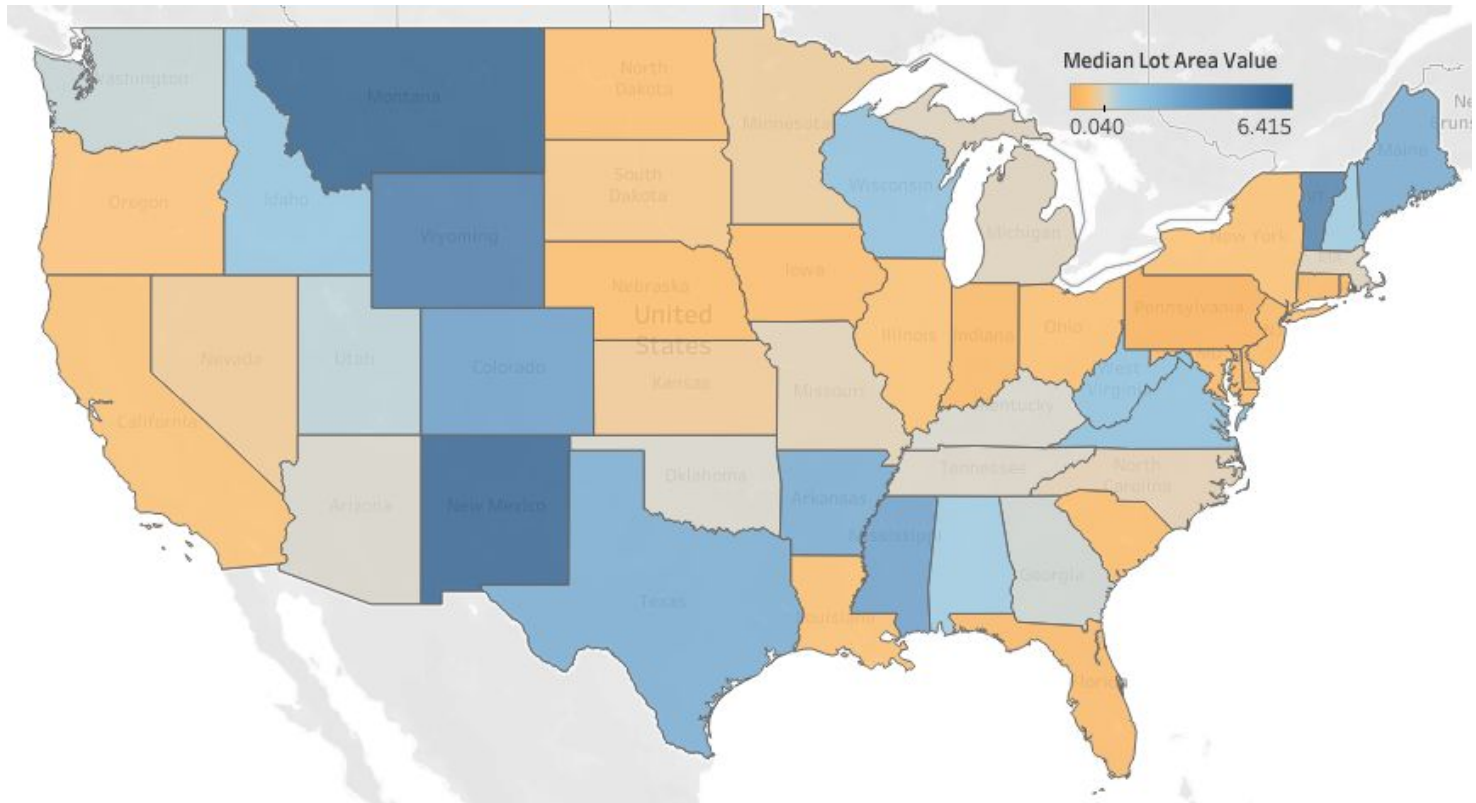
Los Angeles, CA Bedrooms



## Data Analysis - Bedrooms Part 3



## Median Lot Size by State



## Land Costs

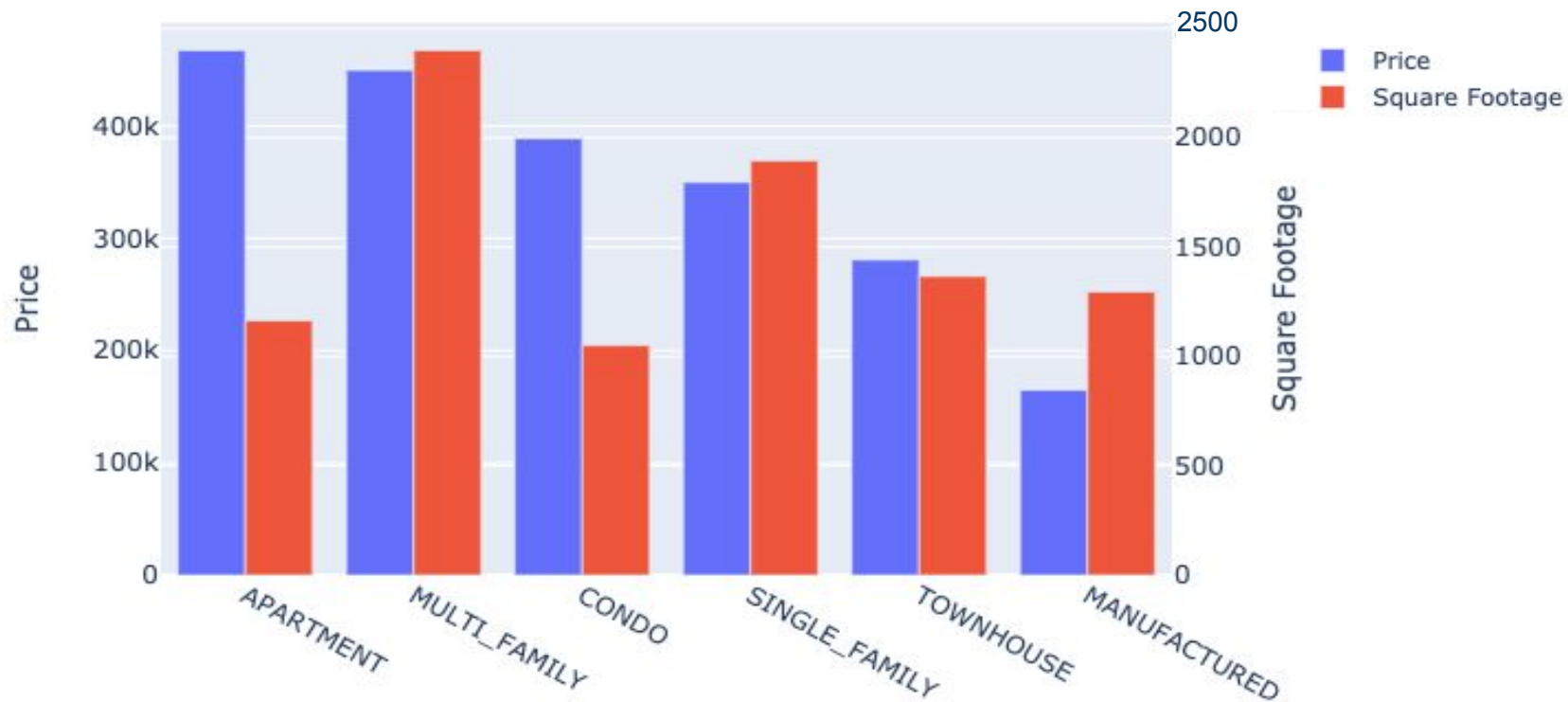
Which cities have the most expensive land?



Which cities have the cheapest land?



## Costs and Sizes by Home Type

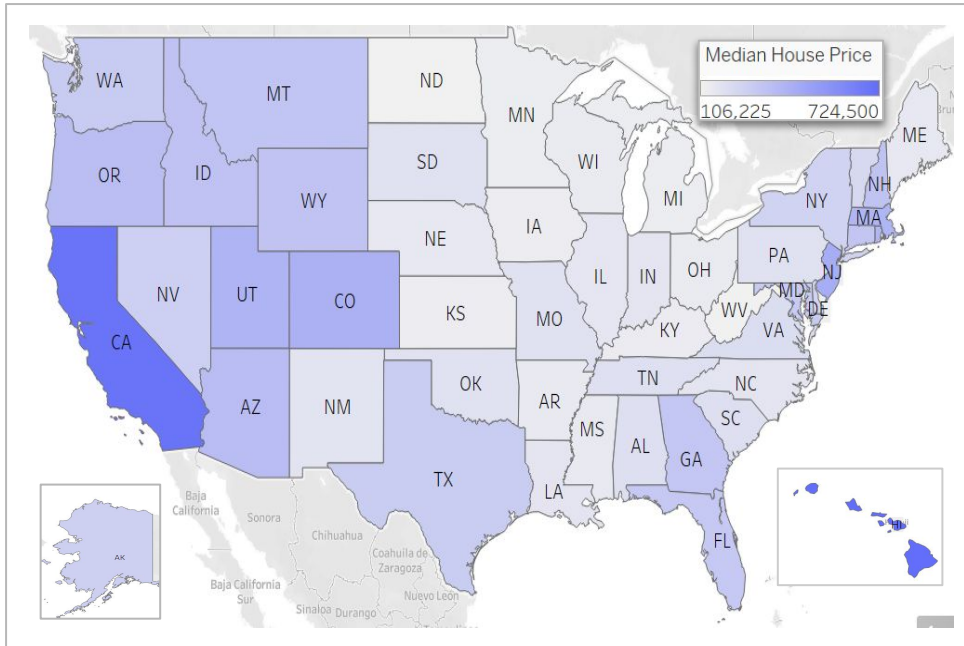




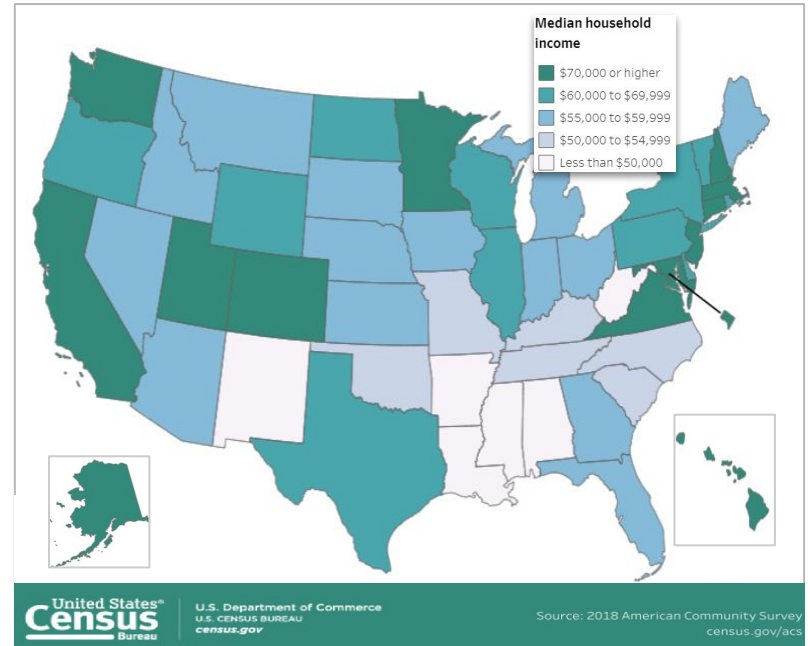
## Data Analysis - Median House Prices

➡ How do house prices vary across the US?

U.S. Median House Price (Zillow)



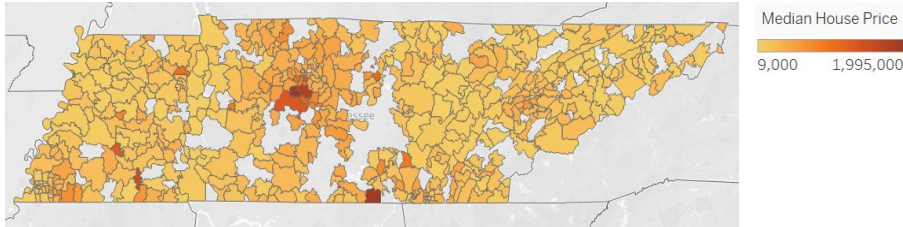
2018 U.S. Median Household Income (Census)



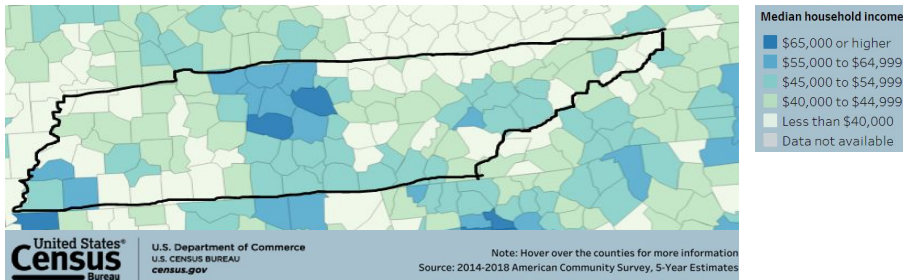
## Data Analysis - Median House Prices

➡ How do house prices vary across Tennessee?

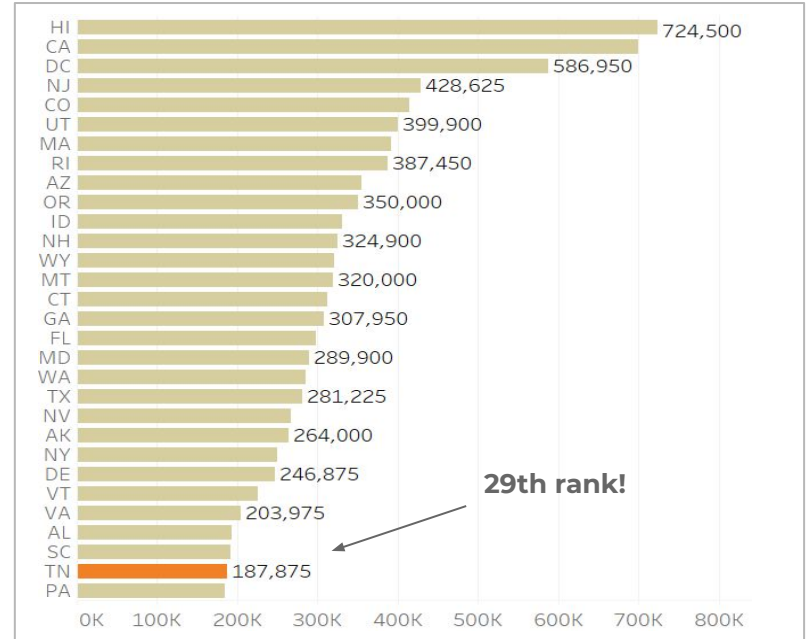
TN Median House Price (Zillow)



TN Median Household Income (Census)



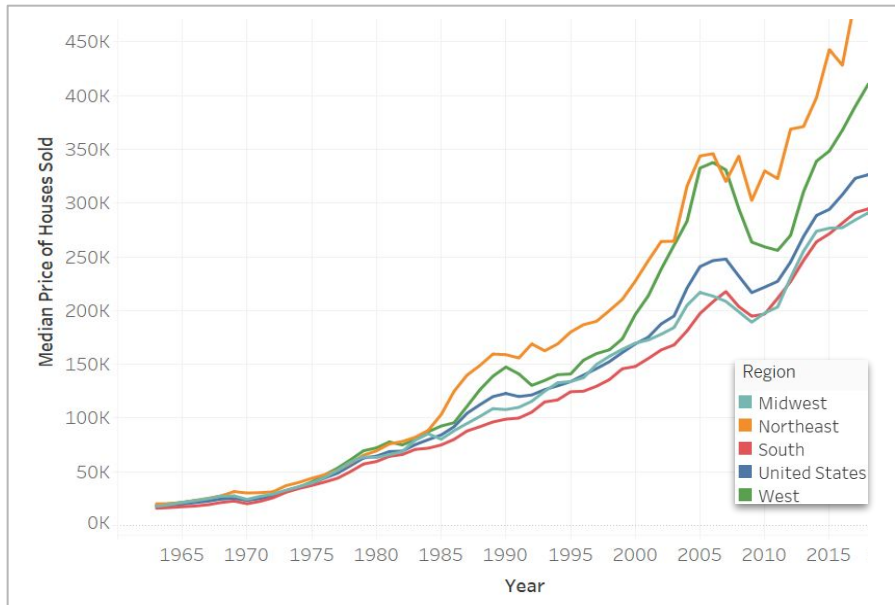
Top 30 States - Median House Price (Zillow)



## Data Analysis - Median House Prices

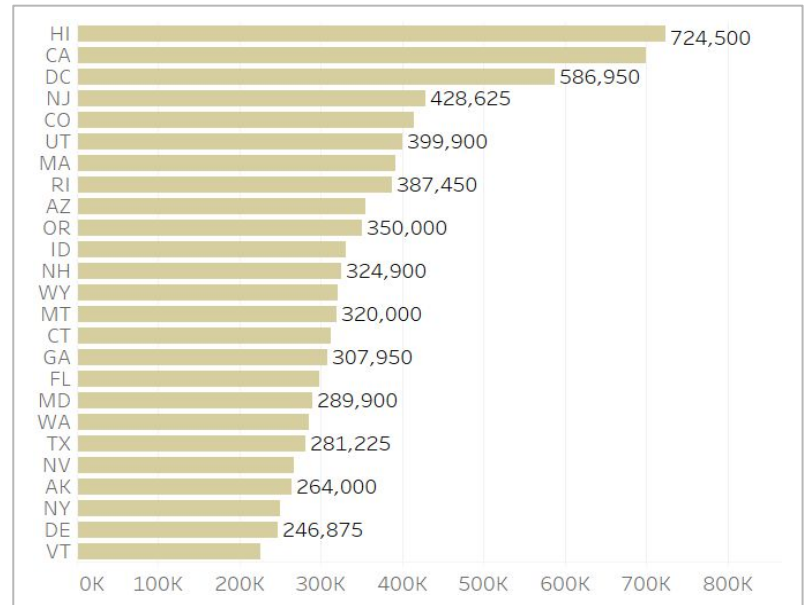
➡ Does our data follow historical trends?

Median Sales Price of Houses Sold by Region (Census)



[https://www.census.gov/construction/nrs/historical\\_data/index.html](https://www.census.gov/construction/nrs/historical_data/index.html)

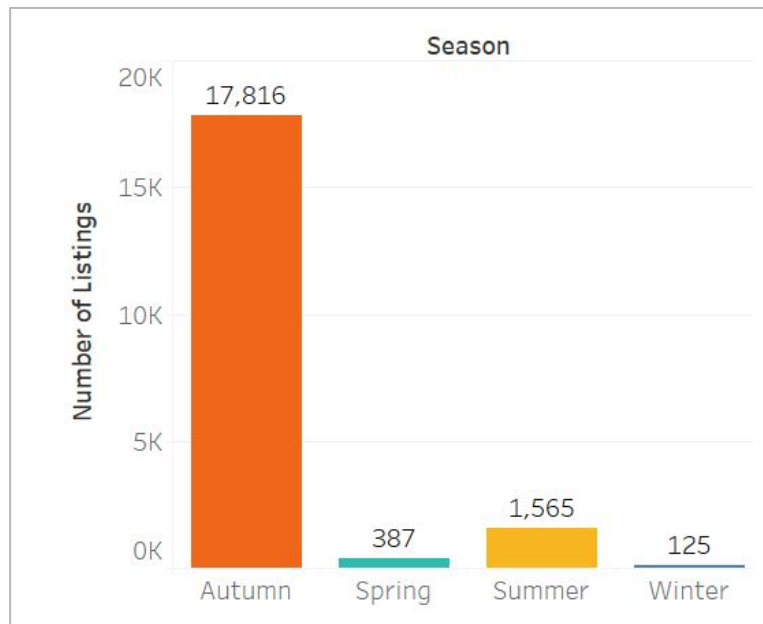
Top 25 States - Median House Price (Zillow)



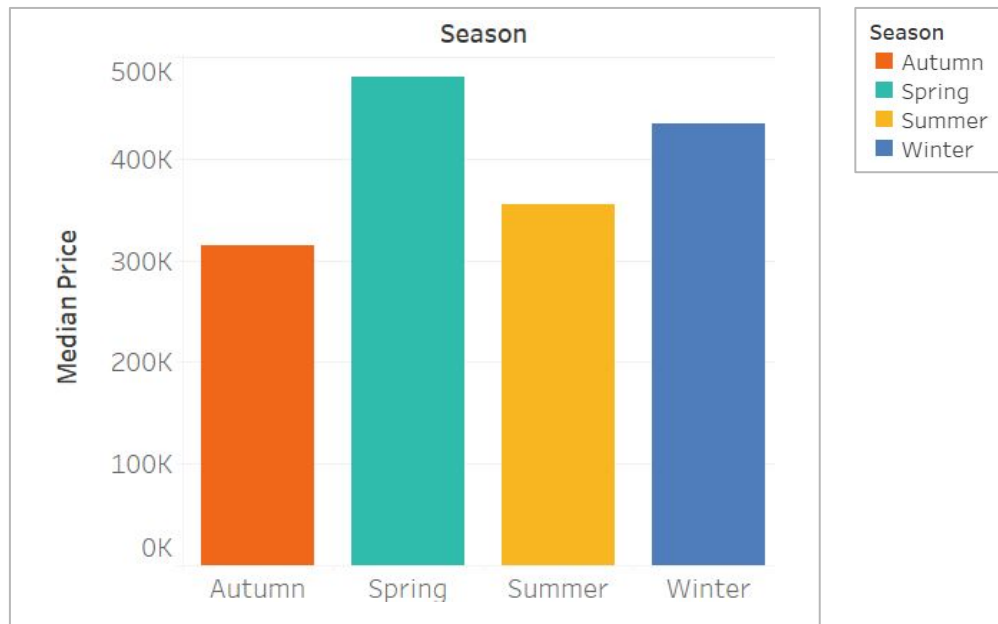
## Data Analysis - Season

➡ How do listing counts and house prices vary among seasons?

Listing Counts Per Season

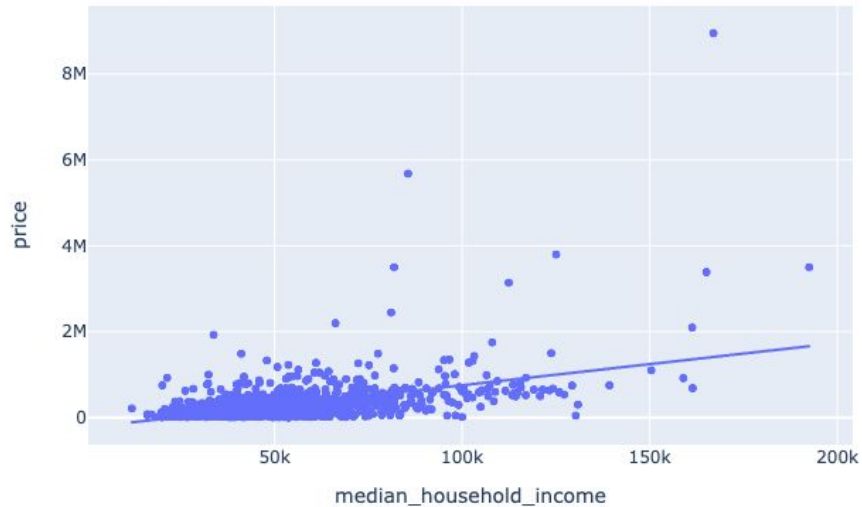


Median House Price Per Season



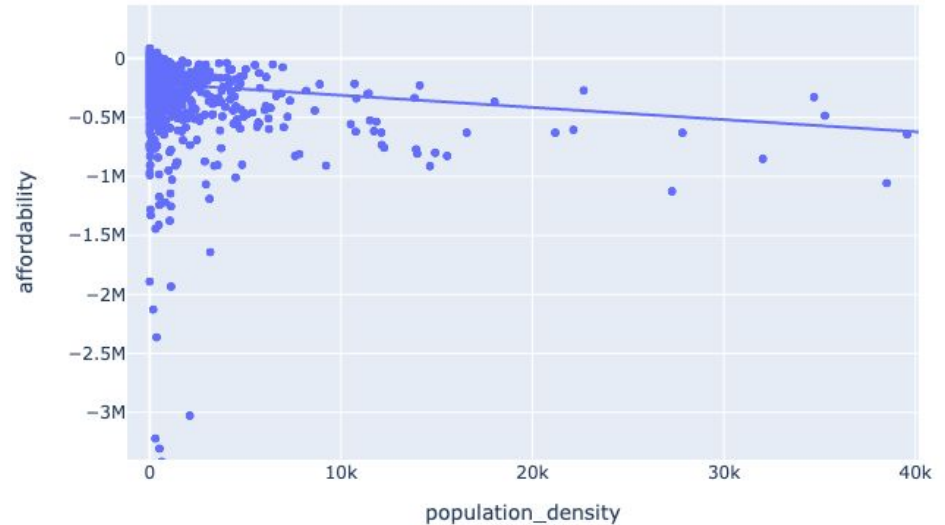
## Regression Analyses

Is the list price related to median household income?



p-value	0.000
R^2	0.445

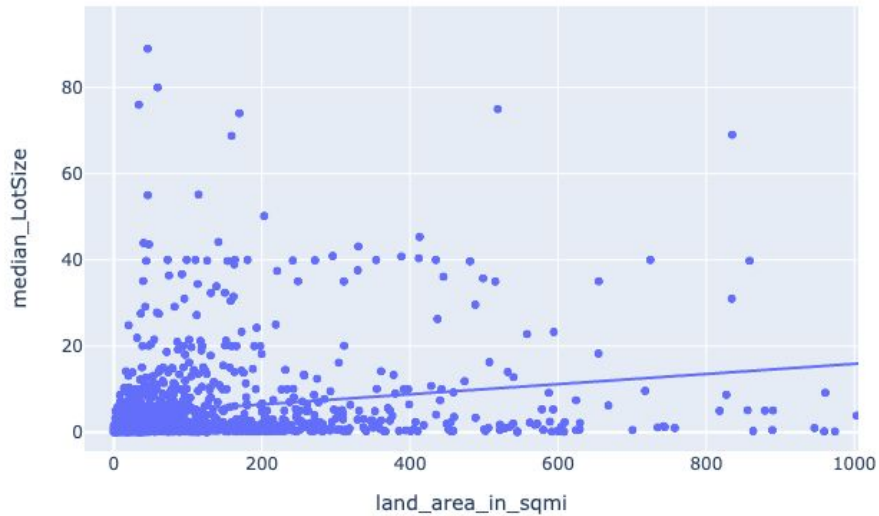
Is property more unaffordable in cities with high population densities?



p-value	0.000
R^2	0.055

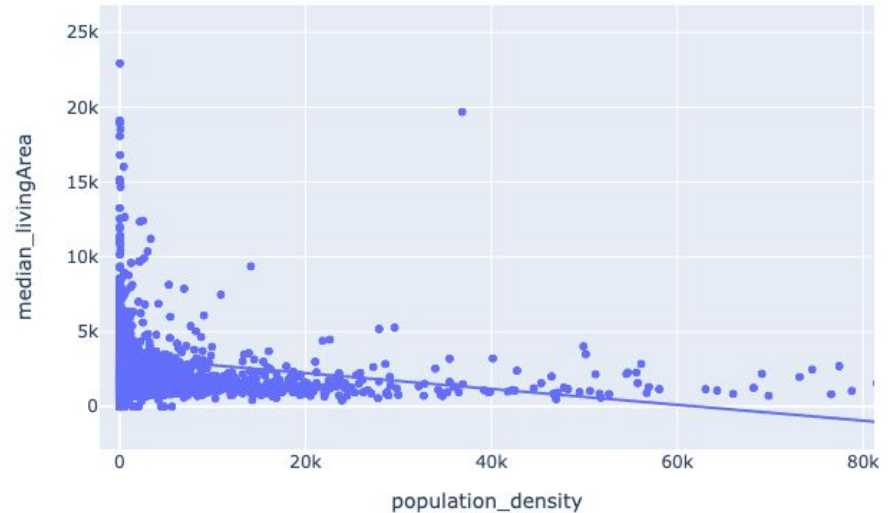
## Regression Analyses

Is a property's lot area related to a city's land area?



p-value	0.000
R^2	0.003

Do square footage decrease as population density increases?



p-value	0.579
R^2	0.000

# Conclusions

- Knoxville housing prices are mostly around concentrated around \$200k-\$300k
- Suburban areas that are less populated have listings with more bedrooms.
  - Outlier: One of the most populous cities (LA) has listings with a lot of bedrooms.
- House prices are higher in the Northeast and Western regions of the U.S.
  - House prices and household income are related



# THANKS

Does anyone have any questions?



CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

**Please keep this slide for attribution.**

