Identifying Optimal Locations for New Home Construction

Matt Cannon, David Kaplan, Dan Quinn, and Roberto Reis

Project 01
July, 22 - 2024

GitHub Repository: https://github.com/dfquinn23/Project_1

Outline

- Goals
- Data Sources
  - Census Data, BEA, Zillow, FEMA
- Data Integration
- Main Outcomes
- Business Application
- Lessons Learned and Improvements

Project Overview

Home builders face the critical decision of selecting the best locations for new construction projects to maximize profitability and meet market demand.

This project aims to identify and evaluate the most suitable areas for building new homes by analyzing various factors such as market affordability and related fundamentals, demographic trends, and environmental risks.

This analysis will provide actionable insights to guide home builders in making informed decisions about where to build.

Objectives

1. Determine the top locations for new home construction based on comprehensive data analysis.
2. Analyze the impact of key factors such as land cost, market demand, infrastructure, demographics, and environmental risks on the suitability of a location.
3. Provide a comparative analysis of different regions to highlight their strengths and weaknesses for new home construction.
4. Offer actionable recommendations to home builders on the most promising areas for development.

Data Sources

01

United States® Census Bureau

American Community Survey (ACS) is the annual survey the Census Bureau conducts in between the decennial Census.

It is a survey that provides information across more than 20 datasets on an ongoing basis.

For our final analysis, we employed county-level data to provide granular data to assess promising areas for development.

In our initial research, we also reviewed metro, state, and national-level data.

ACS datasets we employed in our initial screens:

- Total population growth
- Housing Affordability
- Poverty Rates

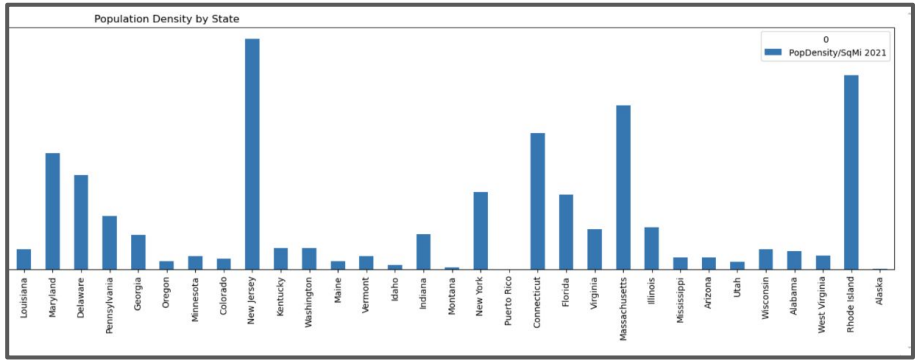
There were dozens of other datasets that we could have used to further refine our analysis, but these three provided our team with a strong foundation from which to proceed.

Had to ensure data frame column headings and data types matched across the various tables.

United States® Census Bureau

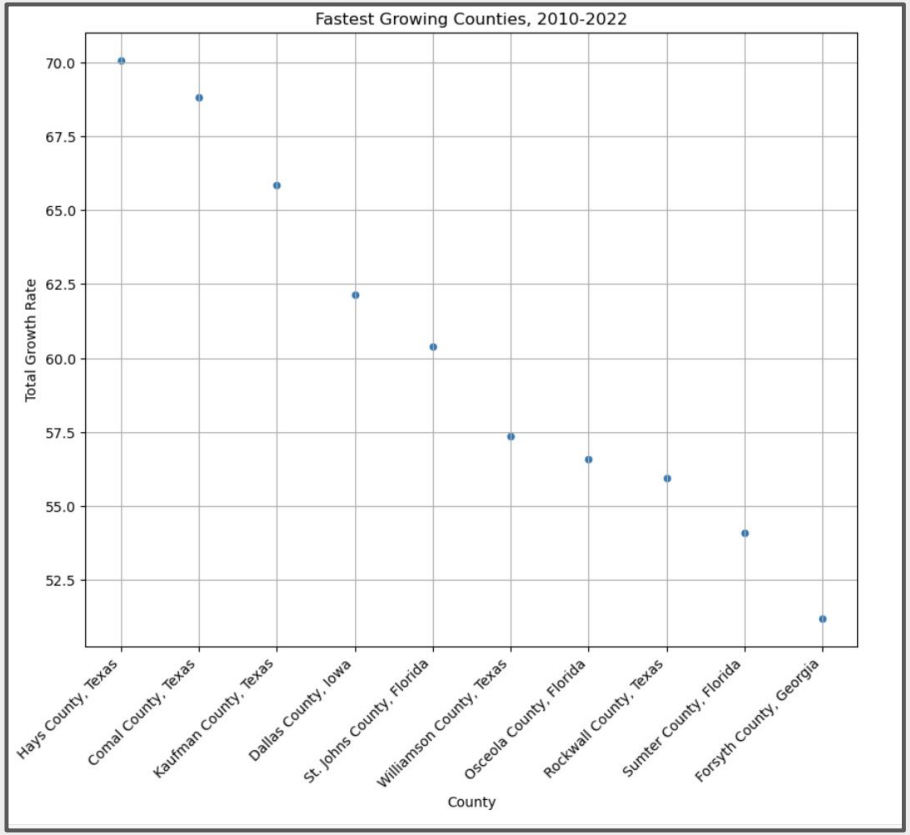
For example, one of our initial screens related to state population density in 2021.

As you can see in the cropped image below, we were able to easily ascertain which states had the highest population density per square mile:

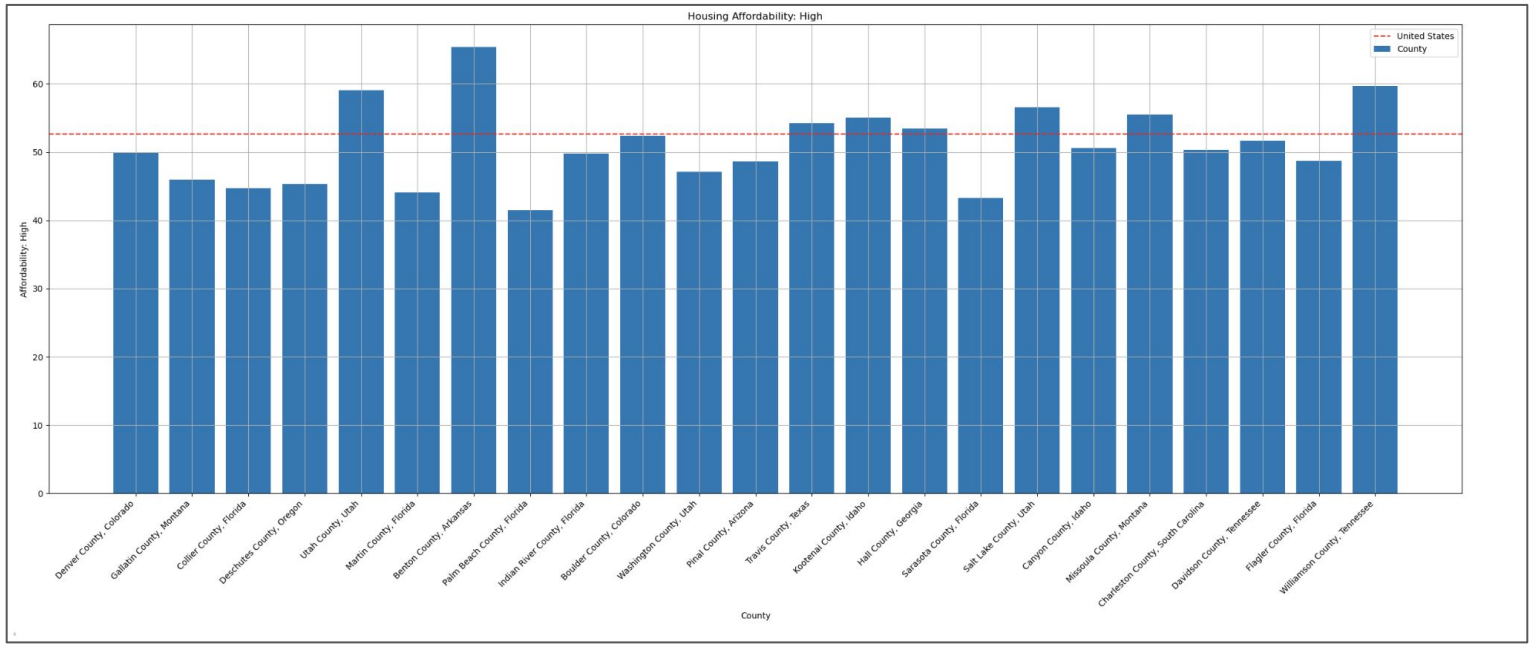


https://github.com/dfquinn23/Project_1.git

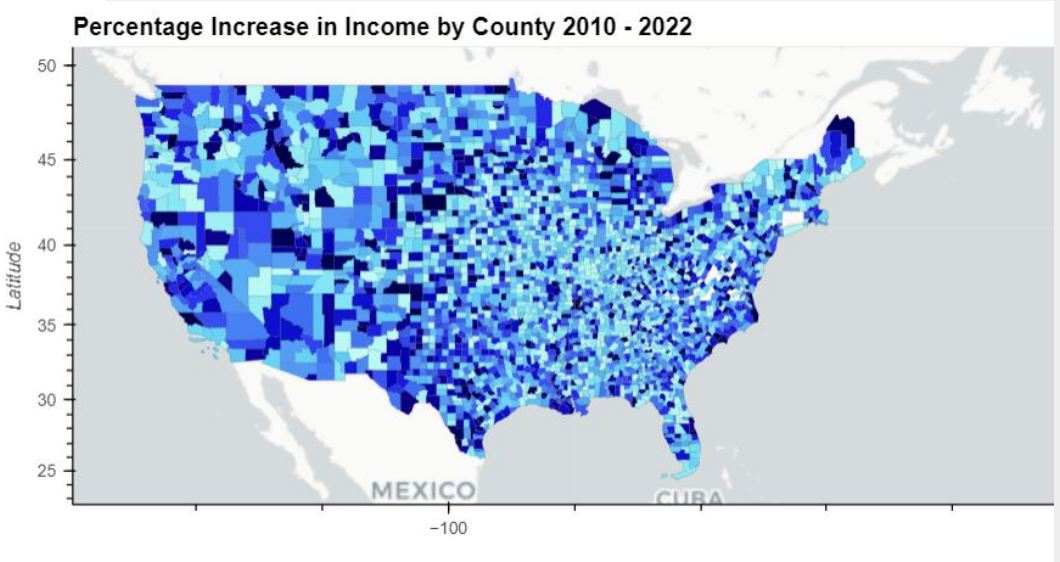
In another, we identified the fastest growing counties in the US, from 2010-2022:



Finally, these are a bit more interesting - housing affordability graphs (*High Affordability shown here - also have Mid- and Low*). These were based off of the percentage of household income spent on owner-occupied housing in our selected counties (2010-2022):

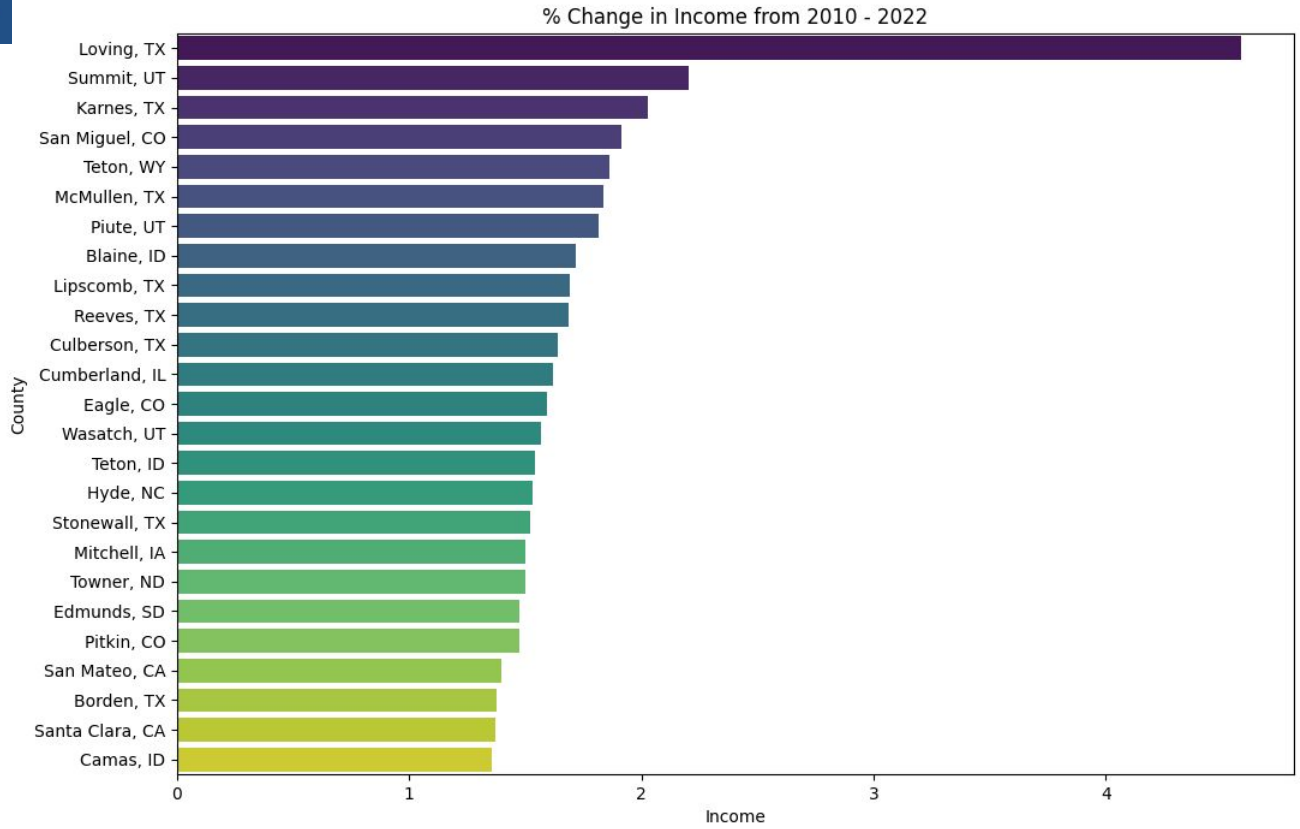


Data Sources
02

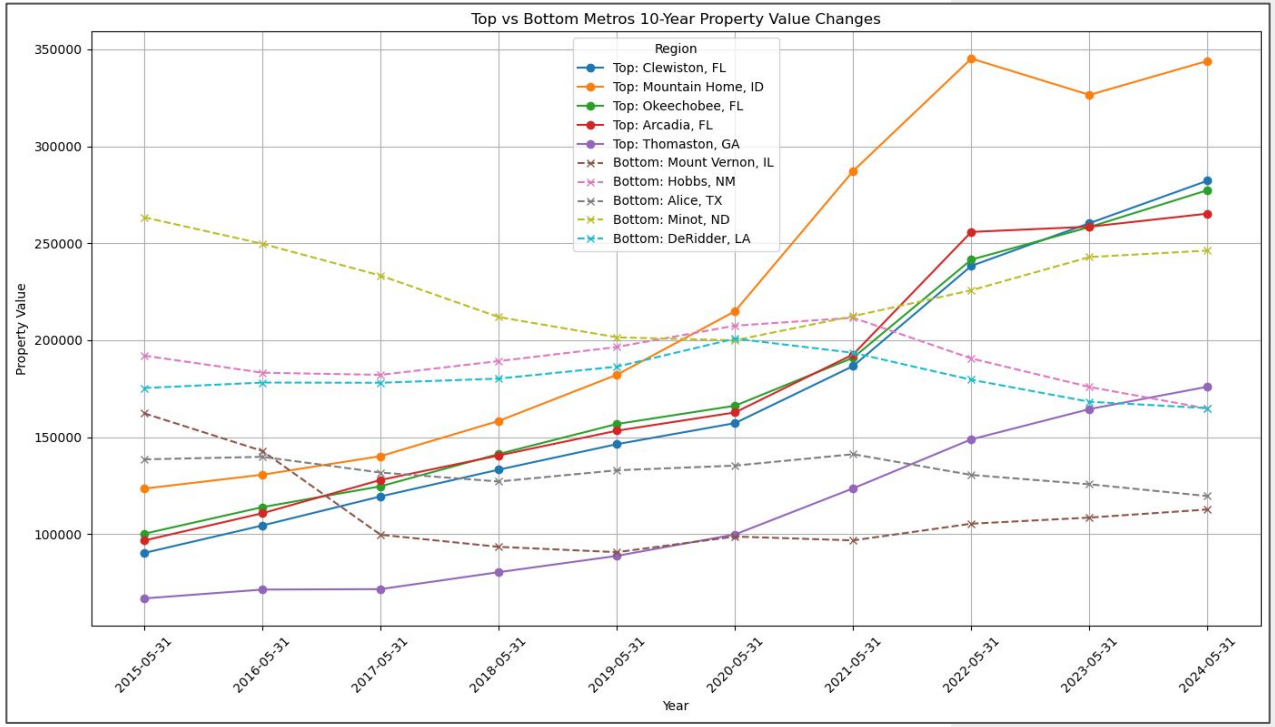


Bureau of Economic Analysis (BEA)

- **assess** the nations income growth by county
- **identify** counties with positive income growth
- **help** showcase counties with higher income growth


% Change in Income from 2010 - 2022
Loving, TX
Summit, UT
Karnes, TX
San Miguel, CO
Teton, WY
McMullen, TX
Piute, UT
Blaine, ID
Lipscomb, TX
Reeves, TX
Culberson, TX
Cumberland, IL
Eagle, CO
Wasatch, UT
Teton, ID
Hyde, NC
Stonewall, TX
Mitchell, IA
Towner, ND
Edmunds, SD
Pitkin, CO
San Mateo, CA
Borden, TX
Santa Clara, CA
Camas, ID
County
0
1
2
3
4
Income

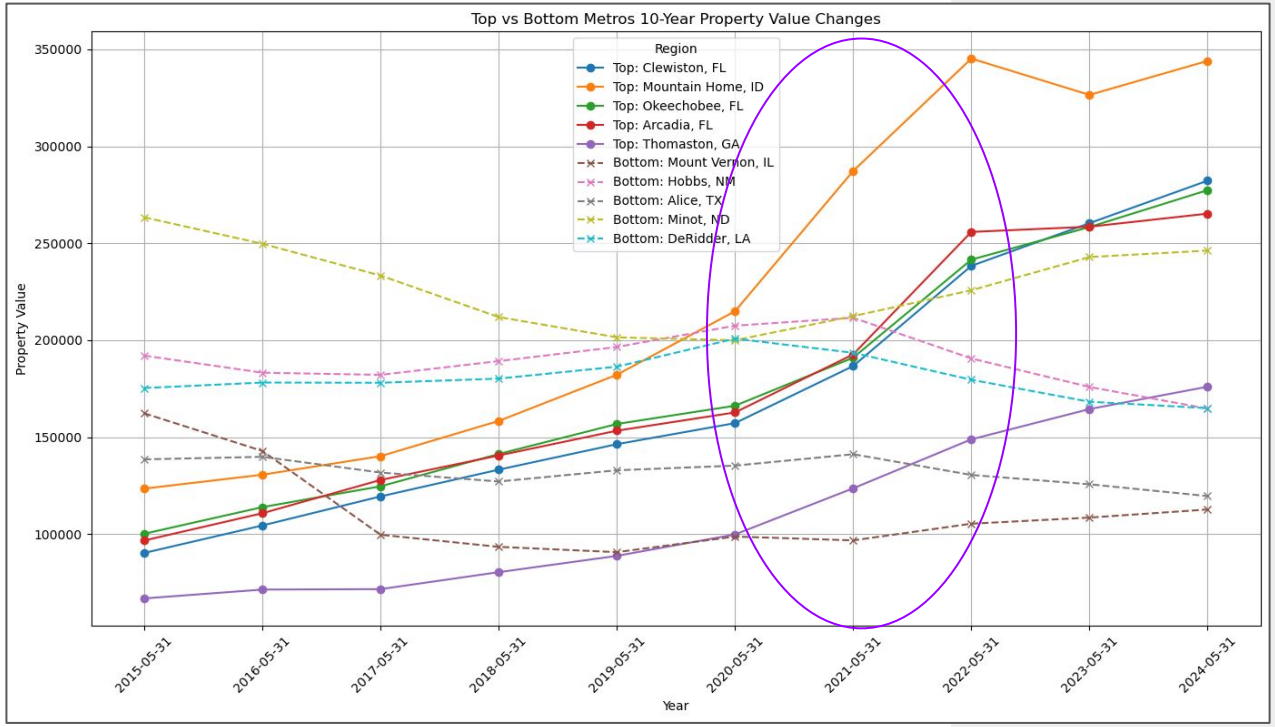
Zillow Home Values



Zillow Home Value Index (ZHVI):

- A measure of typical home value and market changes for a region
- Reflects the typical value for homes in the 35th to 65th percentile range
- The data is made available as a smoothed, seasonally adjusted measure

Zillow Home Values



The Covid Bump

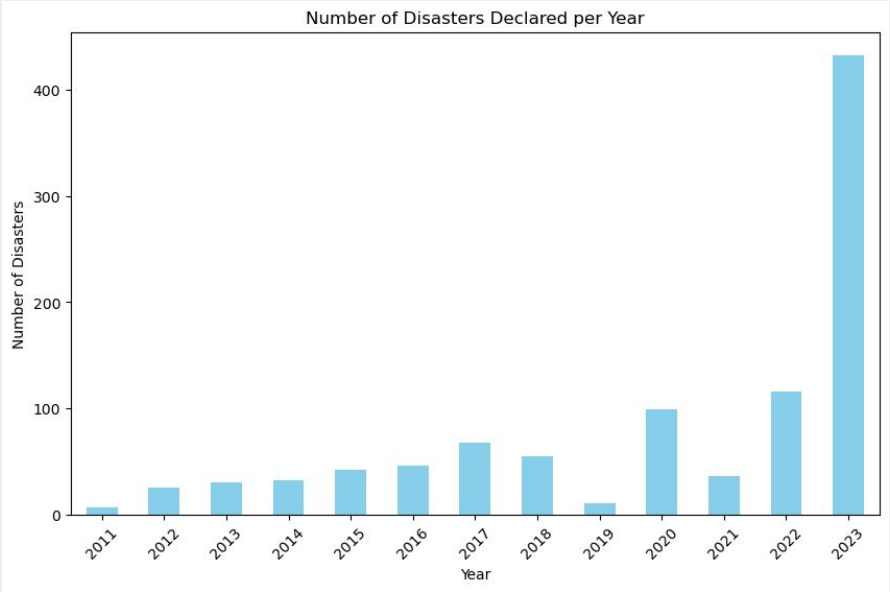
03/2020 - 06/2021

Data Sources

04

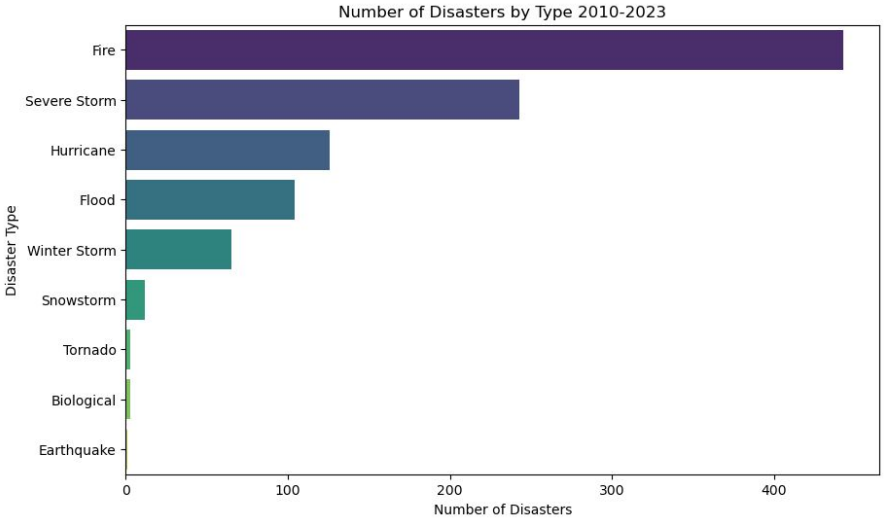
Federal Emergency Management Agency (FEMA)

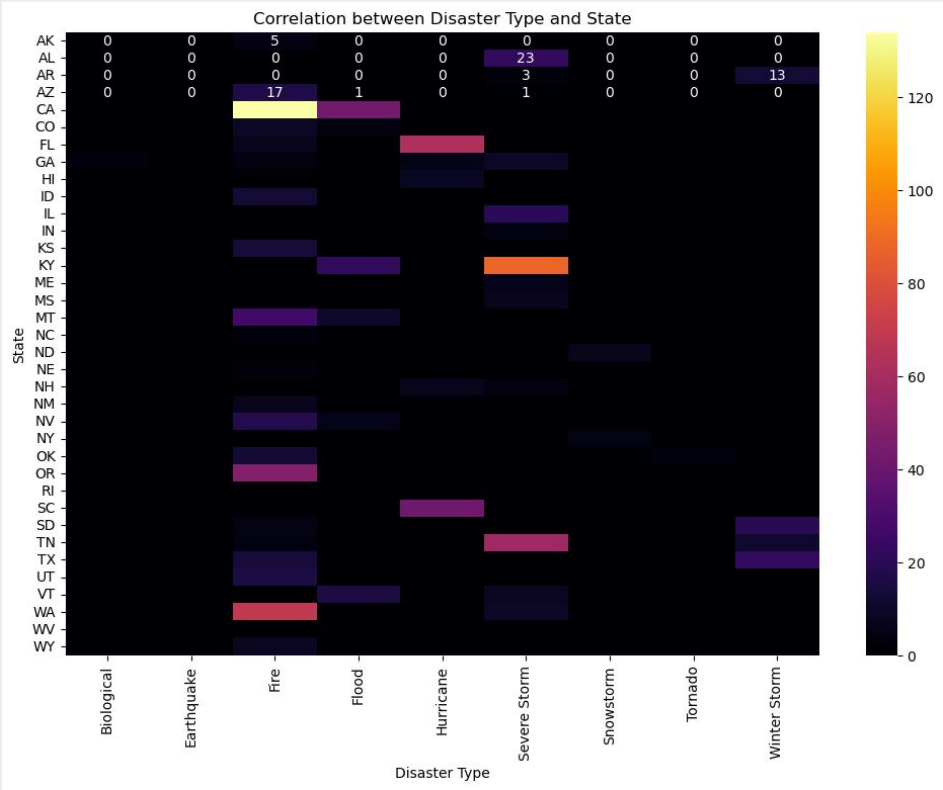
FEMA



- To **assess** an area's risk profile
- To **identify** locations where lower disaster risk correlates with stable or increasing property values
- To **help** balance the trade-off between safety and profitability

FEMA


Number of Disasters by Type 2010-2023
Fire
Severe Storm
Hurricane
Flood
Winter Storm
Snowstorm
Tornado
Biological
Earthquake
Disaster Type
0
100
200
300
400
Number of Disasters


Correlation between Disaster Type and State
AK 0 0 5 0 0 0 0 0 0
AL 0 0 0 0 0 23 0 0 0
AR 0 0 0 0 0 3 0 0 13
AZ 0 0 17 1 0 1 0 0 0
CA
CO
FL
GA
HI
ID
IL
IN
KS
KY
ME
MS
MT
NC
ND
NE
NH
NM
NV
NY
OK
OR
RI
SC
SD
TN
TX
UT
VT
WA
WV
WY
State
Biological
Earthquake
Fire
Flood
Hurricane
Severe Storm
Snowstorm
Tornado
Winter Storm
Disaster Type
120
100
80
60
40
20
0

Data Integration

Top 23 counties

- ➔ Combined BEA, Census, and Zillow data
- ➔ Sorted data sets to get top counties
- ➔ sliced data sets for 10 percent of counties
- ➔ Merged data sets to find county overlap

Data Cleaning

BEA

	GeoFips	GeoName	Income Change
0	48301	Loving, TX	4.581468
1	49043	Summit, UT	2.204708

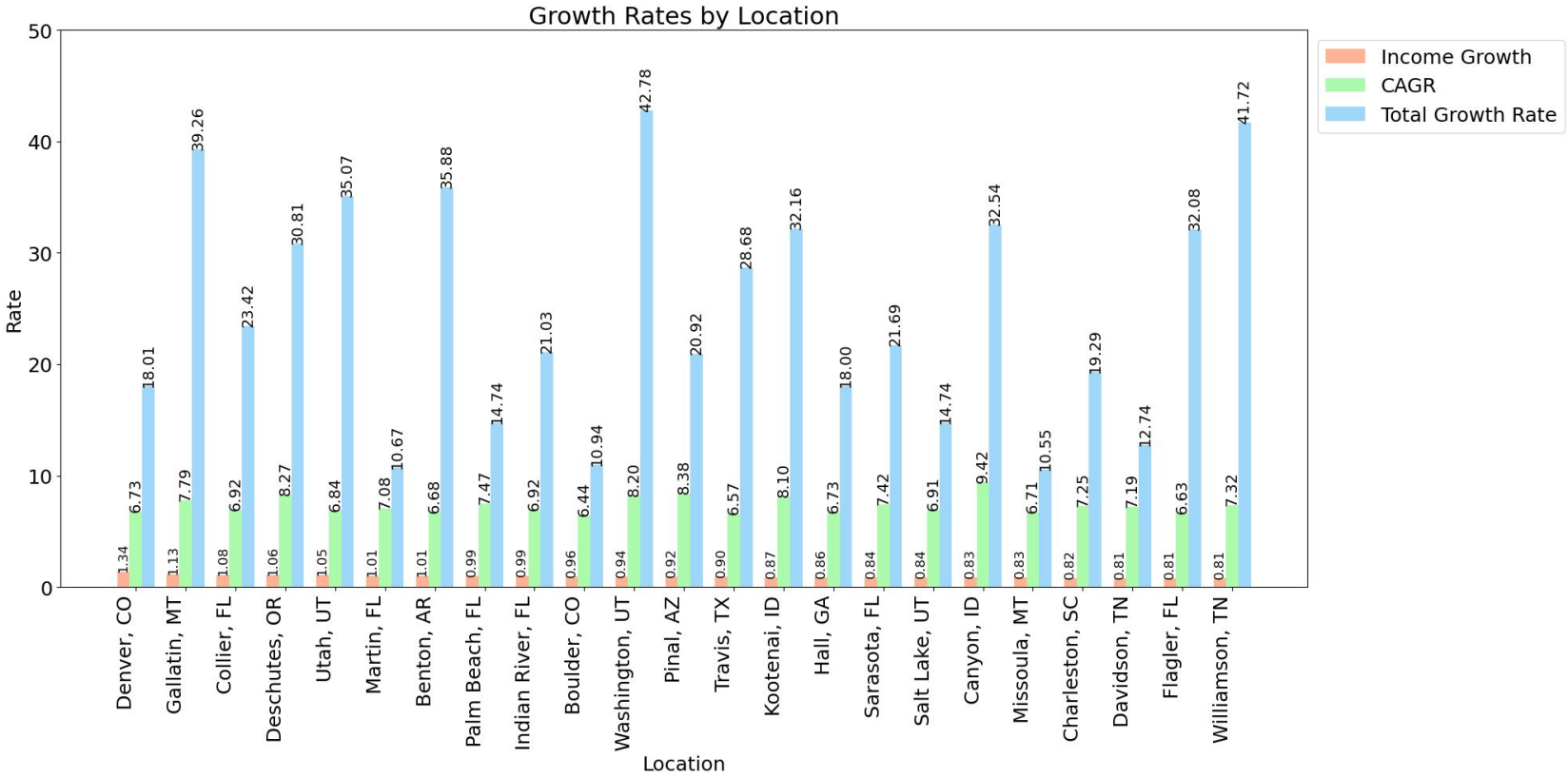
Zillow

	County_State	CAGR
0	Petersburg City, VA	11.300335
1	Clayton County, GA	10.520823

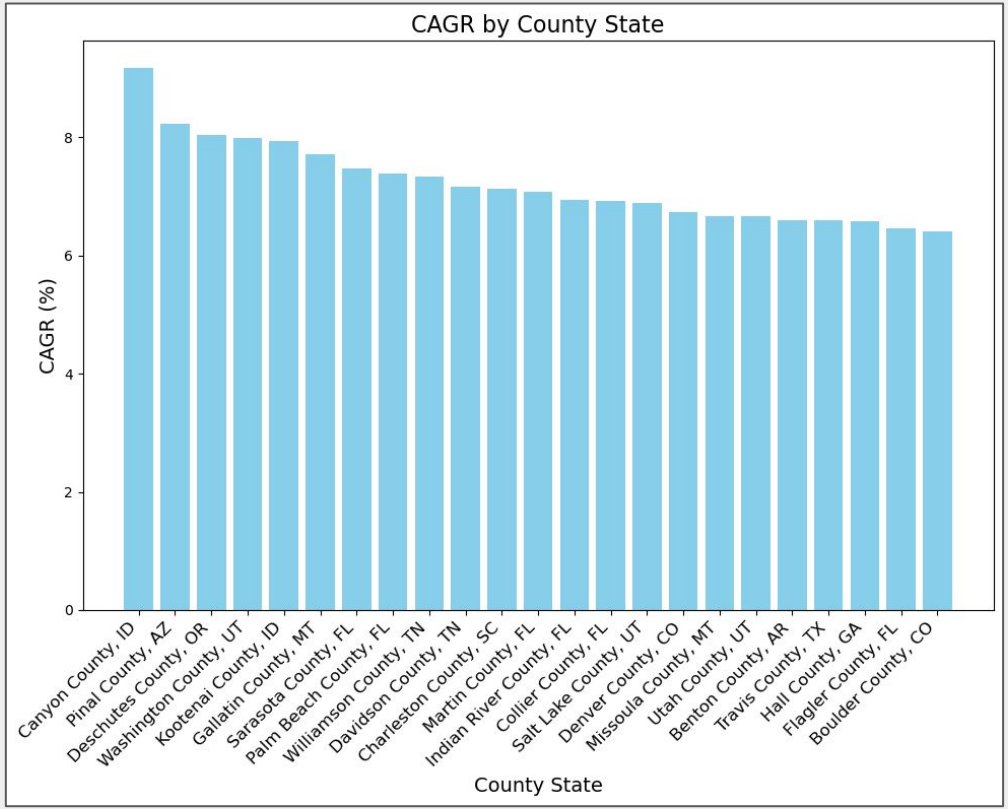
Census

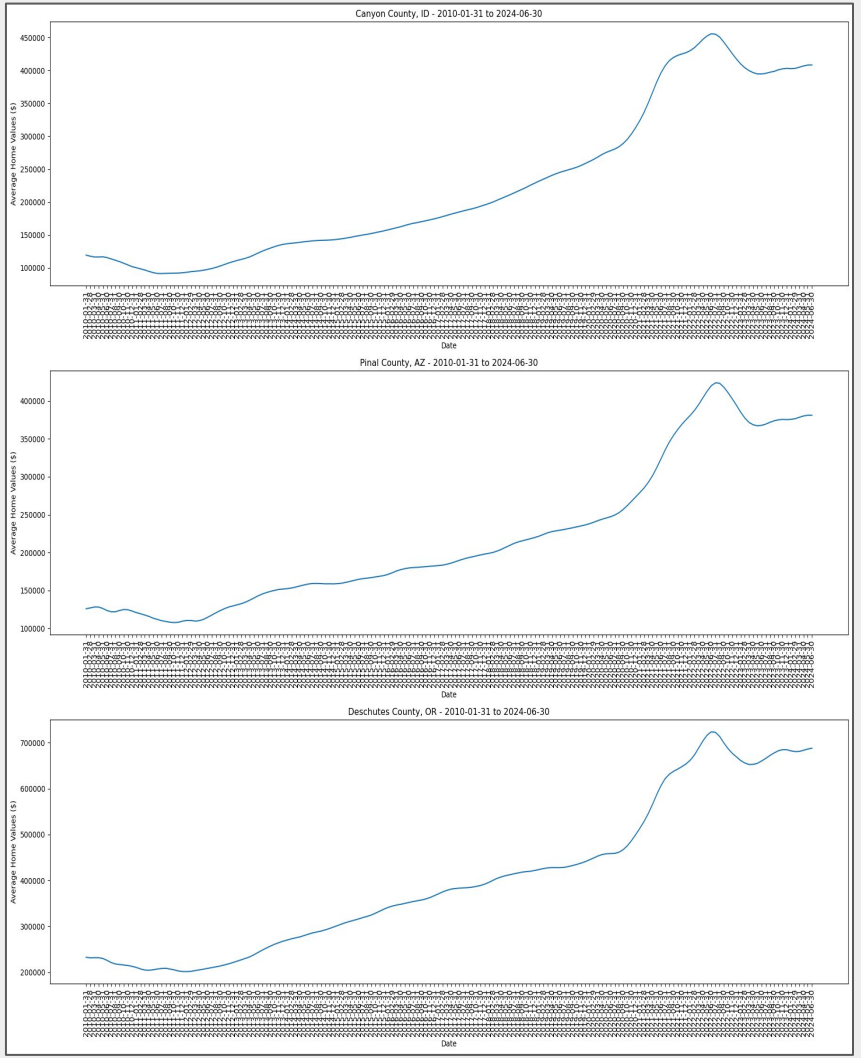
	County	State Code	County Code	Total Growth Rate
0	Tuscaloosa County, Alabama	1	125	0.214032
1	St. Clair County, Alabama	1	115	0.121429

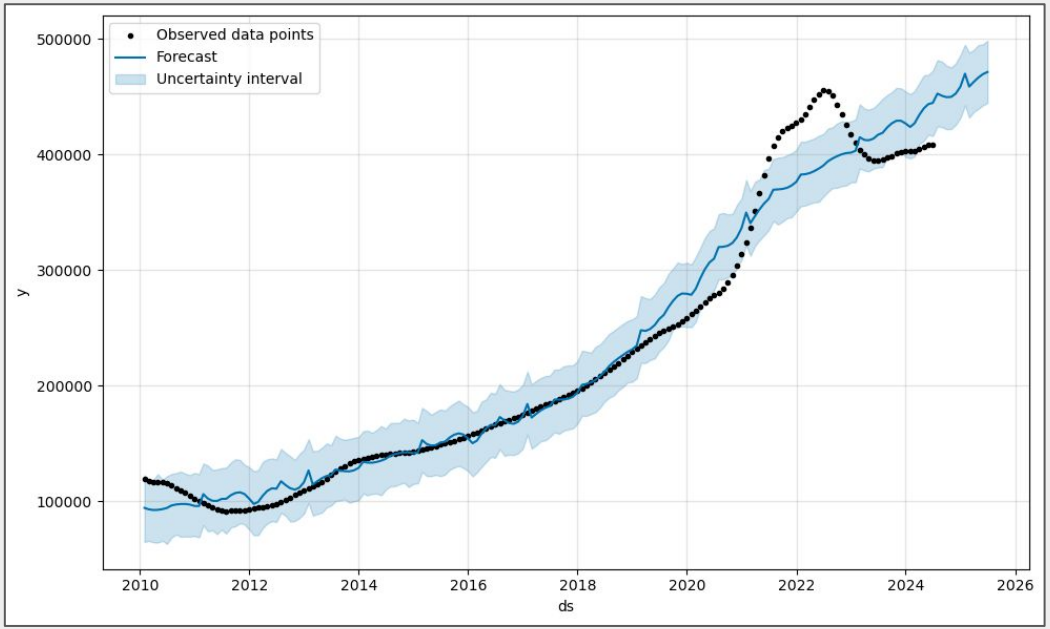
Output

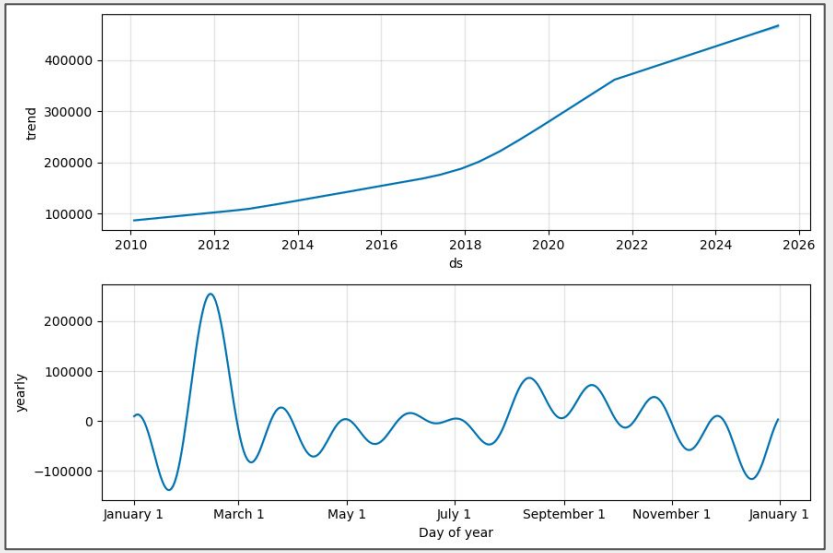

Growth Rates by Location
Income Growth
CAGR
Total Growth Rate
Rate
Location
0
10
20
30
40
50
Denver, CO 1.34 6.73 18.01
Gallatin, MT 1.13 7.79 39.26
Collier, FL 1.08 6.92 23.42
Deschutes, OR 1.06 8.27 30.81
Utah, UT 1.05 6.84 35.07
Martin, FL 1.01 7.08 10.67
Benton, AR 1.01 6.68 35.88
Palm Beach, FL 0.99 7.47 14.74
Indian River, FL 0.99 6.92 21.03
Boulder, CO 0.96 6.44 10.94
Washington, UT 0.94 8.20 42.78
Pinal, AZ 0.92 8.38 20.92
Travis, TX 0.90 6.57 28.68
Kootenai, ID 0.87 8.10 32.16
Hall, GA 0.86 6.73 18.00
Sarasota, FL 0.84 7.42 21.69
Salt Lake, UT 0.84 6.91 14.74
Canyon, ID 0.83 9.42 32.54
Missoula, MT 0.83 6.71 10.55
Charleston, SC 0.82 7.25 19.29
Davidson, TN 0.81 7.19 12.74
Flagler, FL 0.81 6.63 32.08
Williamson, TN 0.81 7.32 41.72

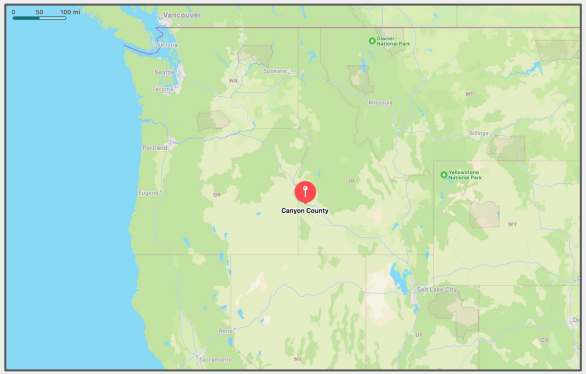
Zillow


CAGR by County State
CAGR (%)
0
2
4
6
8
Canyon County, ID
Pinal County, AZ
Deschutes County, OR
Washington County, UT
Kootenai County, ID
Gallatin County, MT
Sarasota County, FL
Palm Beach County, FL
Williamson County, TN
Davidson County, TN
Charleston County, SC
Martin County, FL
Indian River County, FL
Collier County, FL
Salt Lake County, UT
Denver County, CO
Missoula County, MT
Utah County, UT
Benton County, AR
Travis County, TX
Hall County, GA
Flagler County, FL
Boulder County, CO
County State


Canyon County, ID - 2010-01-31 to 2024-06-30
Average Home Values ($)
Date
Pinal County, AZ - 2010-01-31 to 2024-06-30
Average Home Values ($)
Date
Deschutes County, OR - 2010-01-31 to 2024-06-30
Average Home Values ($)
Date

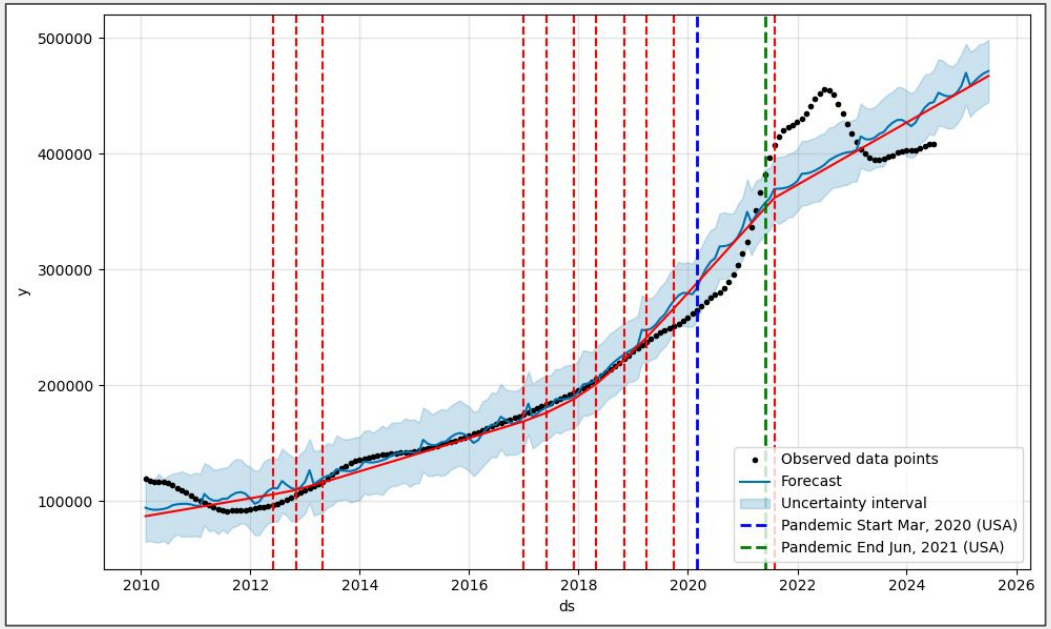


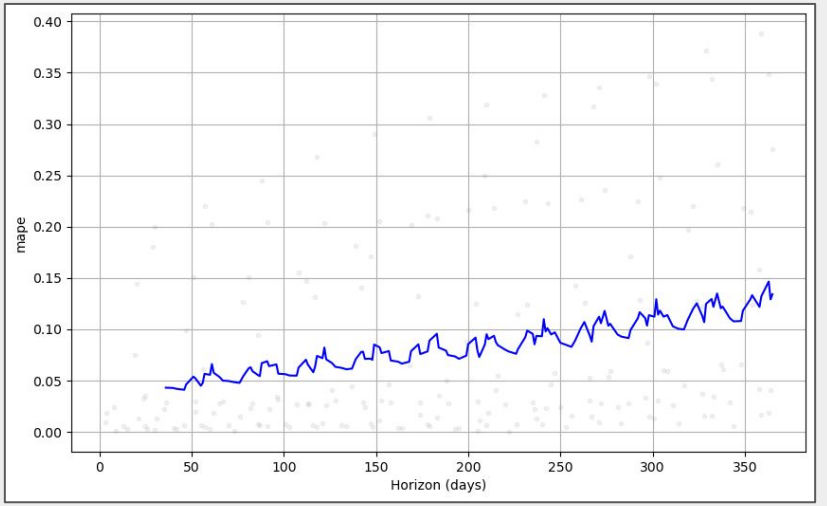




Characteristics for Canyon County, ID:

- Experienced a Covid bump
- Forecast is up and to the right
- Yearly market changes are more volatile in Q1, and then trend down from Aug/Sept through the end of the year





Trend Changepoints:

- Identify changes in the trend's trajectory
- Picks up the end of Covid
- Only inferred for first 80% of the time series

Prediction Accuracy

Mean Absolute Percentage Error (MAPE)

- The average deviation between the forecasted value and the actual value.
- This forecast typically errors around 5% for the first 100 days and then errors increase up to around 15% for predictions that are a year out.

Summary

- There are >3,200 counties in the US
- Based on the parameters used to merge our Census, Zillow, and BEA data, we were able to narrow a select client list of 23 counties
- Within the selected list, the environmental risk is negligible

Business Application

The project has to be saleable to a realistic client

Base assumption - 2-week project

Let’s assume our “company” does 20 2-week projects per year:

- [(Salaries ($150K x 5) / 20] + $1,000 ($20K in company overhead annually, divided by 20 – we run a tight ship.)
- Add in a profit margin of 25%
- We need to do a cost-benefit analysis for the client:
    - What problem are we solving?
    - How does our solution enable more precise decision-making for our client?
    - Can we justify our total fee against what the ultimate value is for the client?

Total Project Fee: $50,000

Job Market Analysis

linkedin.com
Glassdoor.com
ChatGPT.com

7,999 Data Scientist jobs in the US

Requirements:

- Educational Background: Typically requires a Bachelor's degree in data science, statistics, computer science.
- Technical Skills: Proficiency in programming languages such as Python or R. Familiarity with data manipulation tools and libraries like Pandas and NumPy, and experience with data visualization tools.
- Specialized Knowledge: Understanding of time series analysis, experience with machine learning models for forecasting.

Experience:

- Experience with large datasets and real-time data processing.
- Background in statistical modeling and regression analysis.
- Prior work with cross-validation and predictive modeling techniques.

Expectations:

- Ability to translate complex datasets into actionable insights.
- Strong analytical skills to monitor and forecast market trends.
- Capable of developing models that predict real estate values and market dynamics.

Salary:

According to Glassdoor, the average base pay for data scientists in the U.S. is around $113,000 to $160,000 a year. Senior data scientists might earn between $160,000 and $200,000 annually.

Job Outlook:

Employment of data scientists is expected to grow significantly due to the high demand for data-driven decision-making and the advancement in AI and machine learning technologies.

Lessons Learned and Improvements

Lessons Learned:

1. Getting sync'd up earlier on data integration
2. Picking the right data from a mountain of data

Improvements:

1. Fail faster and better expectations of data sifting
2. Clear integration plans for the data

Questions?