REMOTE WORK LOCATION RECOMMENDER

Cynthia Rodriguez

AGENDA

01

DATA WRANGLING

Collecting, cleaning, and combining the data.

02

EXPLORATORY DATA ANALYSIS

Looking for notable features and similarities between datasets.

03

FUNCTION BUILDING

Converting user input to values that can be fed into a recommender system.

04

STREAMLIT APPLICATION

Creating a usable app with multiple functions implemented.

PROBLEM STATEMENT

I work at ABC Marketing, Inc. and we just signed Airbnb as our latest client.

They've asked us to build a web application that is easy to use and will drive repeat traffic to their website.

PROBLEM STATEMENT

I work at ABC Marketing, Inc. and we just signed Airbnb as our latest client.

They’ve asked us to build a web application that is easy to use and will drive repeat traffic to their website.

Given that Forbes has projected 25% of all professional jobs in North America will be remote by the end of 2022, our goal is to create a ready-for-launch remote work location recommendation system that drives users directly to Airbnb’s page.

01

DATA WRANGLING

COLLECTING THE DATA

Datasets used:

1. **Airbnb Listings**

Airbnb data for listings in 30 US locations.

2. **Restaurant Chains**

Mapped data of 705,621 independent and chain restaurants across the US.

3. **Cost of Living/Socioeconomic Stats**

Cost of living compared the US average (overall and specific spending categories), population, median age, political lean in 2020 election.

4. **Temperature Trends**

Average monthly temperature for each location.

5. **Rain Trends**

Average monthly rain for each location.

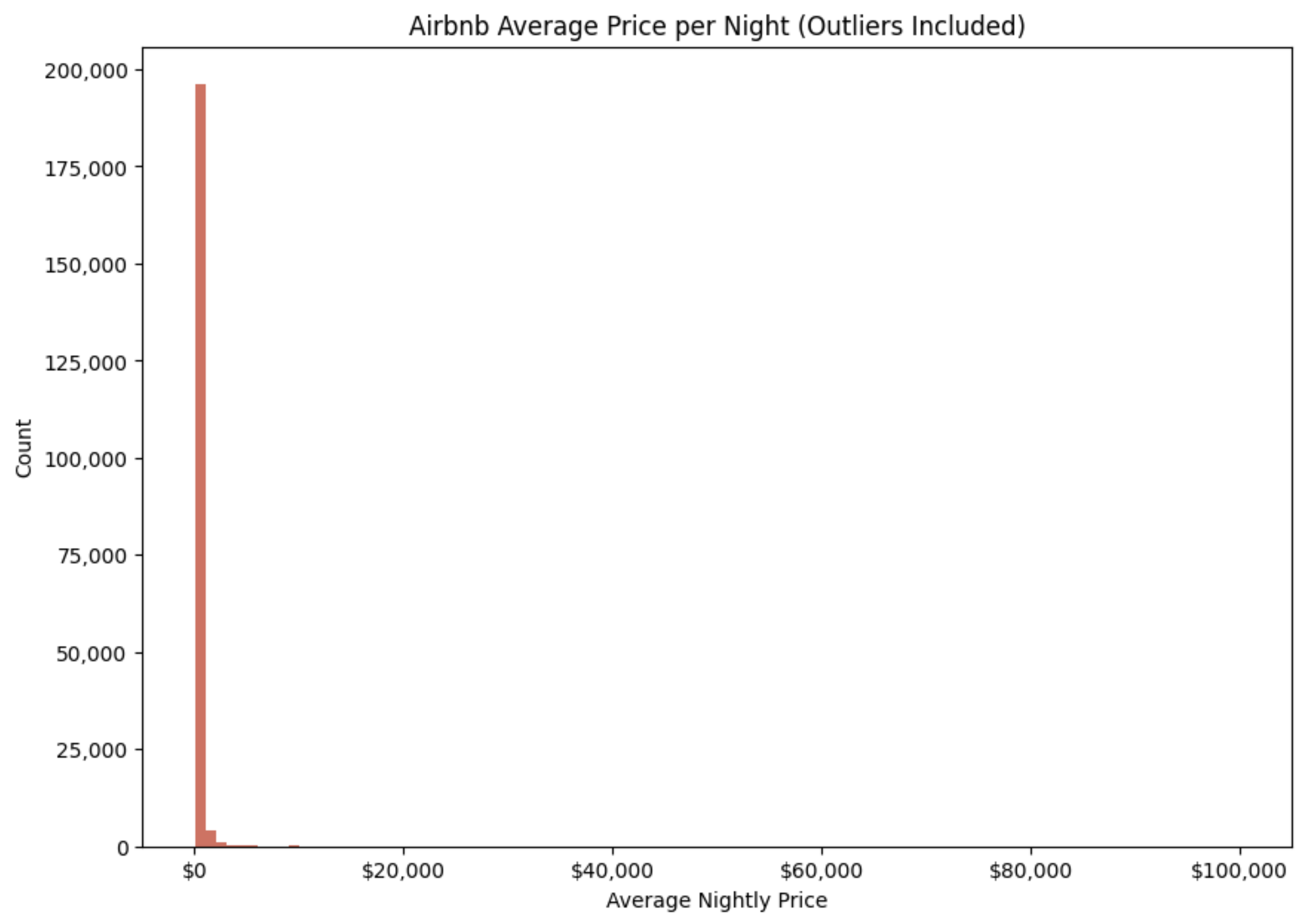
6. **Walkability**

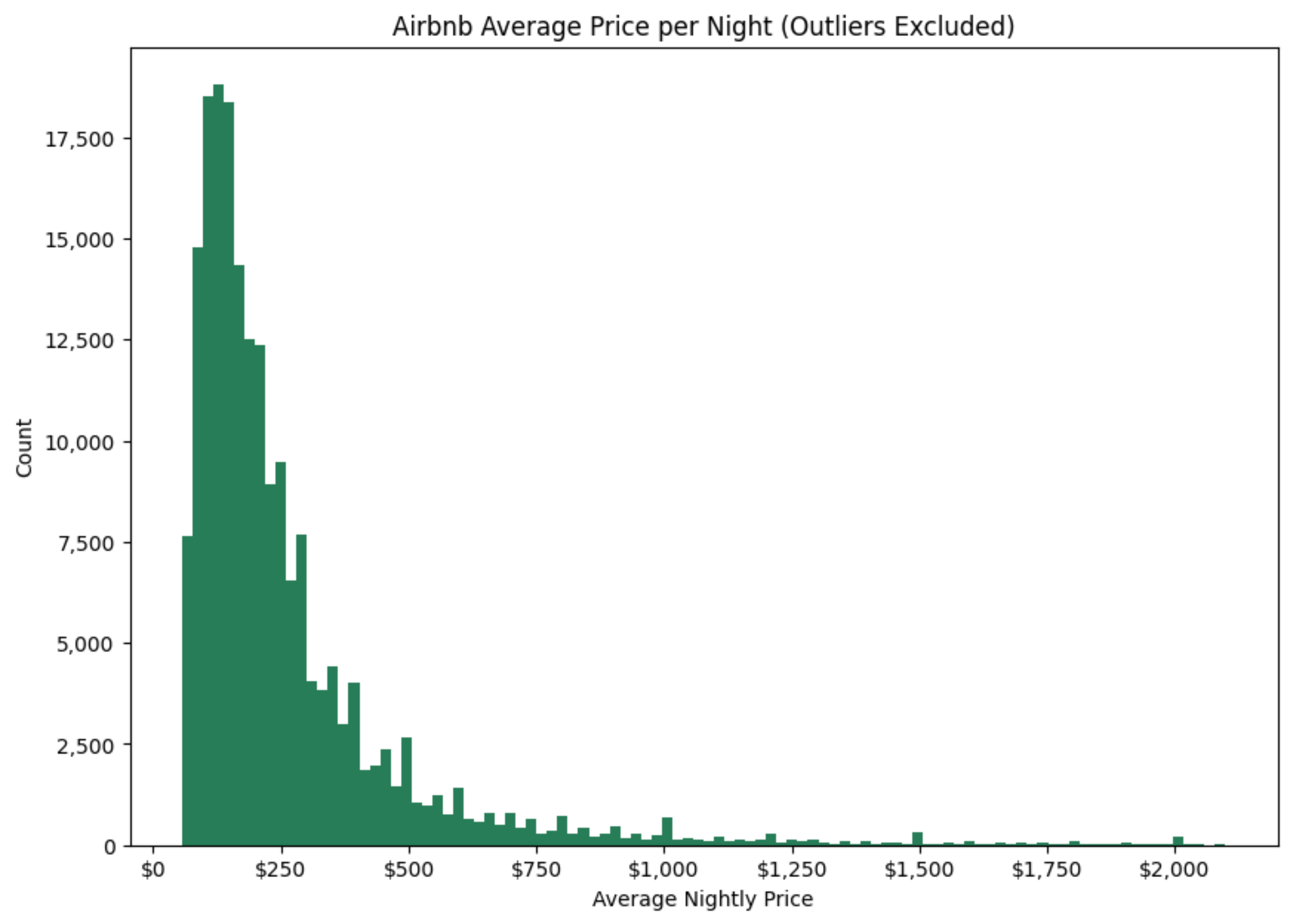
Walkability index, auto and transit access indices, and percent of people who do not own a vehicle.

CLEANING THE DATA

Airbnb Price Outliers

- Removed listings with prices greater than 99th percentile or less than 1st percentile


Airbnb Average Price per Night (Outliers Included)
Count
200,000
175,000
150,000
125,000
100,000
75,000
50,000
25,000
0
$0
$20,000
$40,000
$60,000
$80,000
$100,000
Average Nightly Price


Airbnb Average Price per Night (Outliers Excluded)
Count
0
2,500
5,000
7,500
10,000
12,500
15,000
17,500
$0
$250
$500
$750
$1,000
$1,250
$1,500
$1,750
$2,000
Average Nightly Price

CLEANING THE DATA

Airbnb price outliers

- Removed listings with prices greater than 99th percentile or less than 1st percentile

Airbnb listing types

- Working under the assumption that users looking for long-term WFH stays would be looking at ‘entire home/apartment’ listings only

CLEANING THE DATA

Airbnb price outliers

- Removed listings with prices greater than 99th percentile or less than 1st percentile

Airbnb listing types

- Working under the assumption that users looking for long-term WFH stays would be looking at ‘entire home/apartment’ listings only

Converting datasets to same scale

- The Airbnb, walkability, and chain datasets were neighborhood-specific, while all others were city-specific; some imputing required

CLEANING THE DATA

Airbnb price outliers

- Removed listings with prices greater than 99^{th} percentile or less than 1^{st} percentile

Airbnb listing types

- Working under the assumption that users looking for long-term WFH stays would be looking at ‘entire home/apartment’ listings only

Converting datasets to same scale

- The Airbnb, walkability, and chain datasets were neighborhood-specific, while all others were city-specific; some imputing required

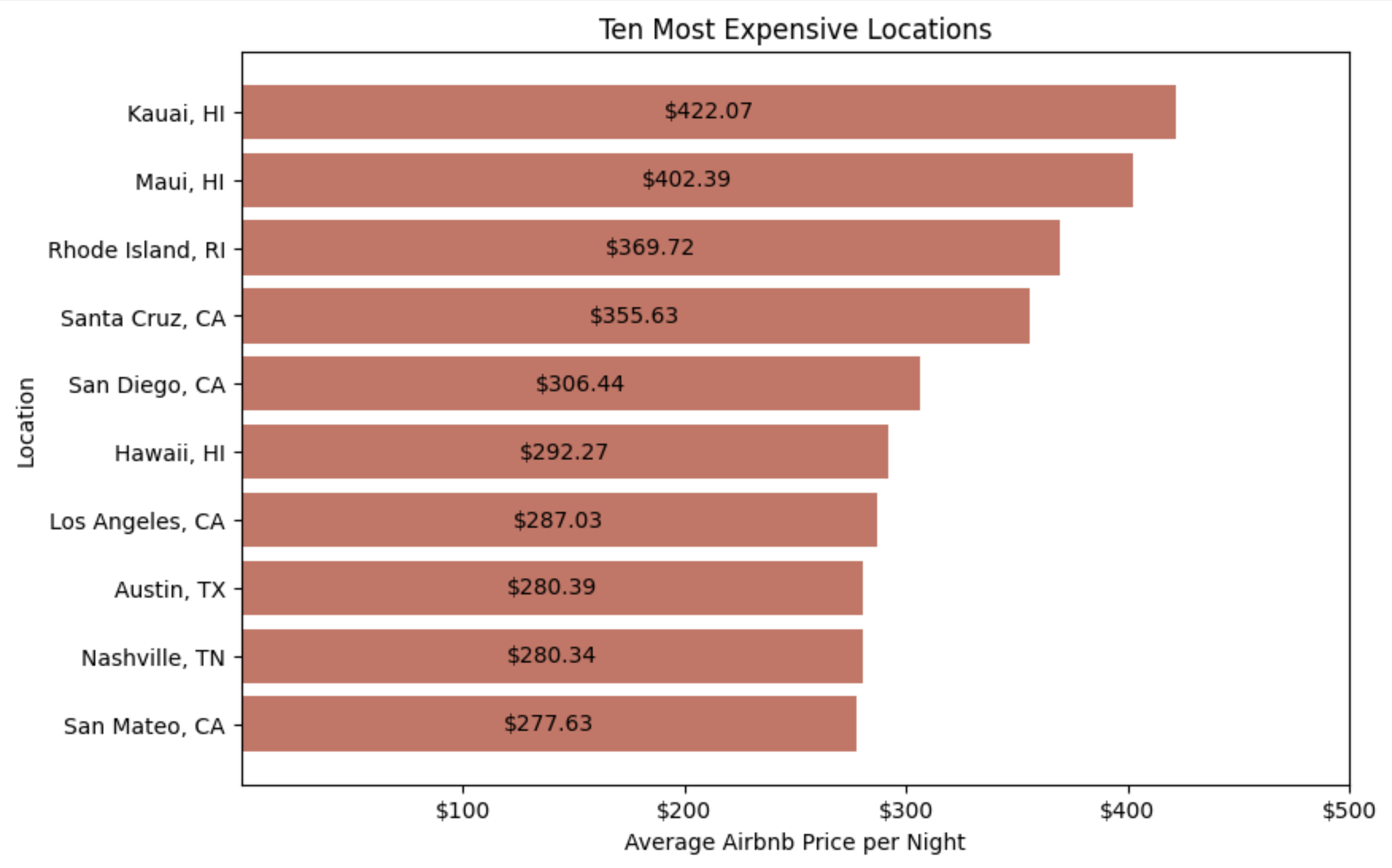
Removed pre-filled nulls

- Some nulls had been imputed by the data providers; dropped these observations or replaced the values

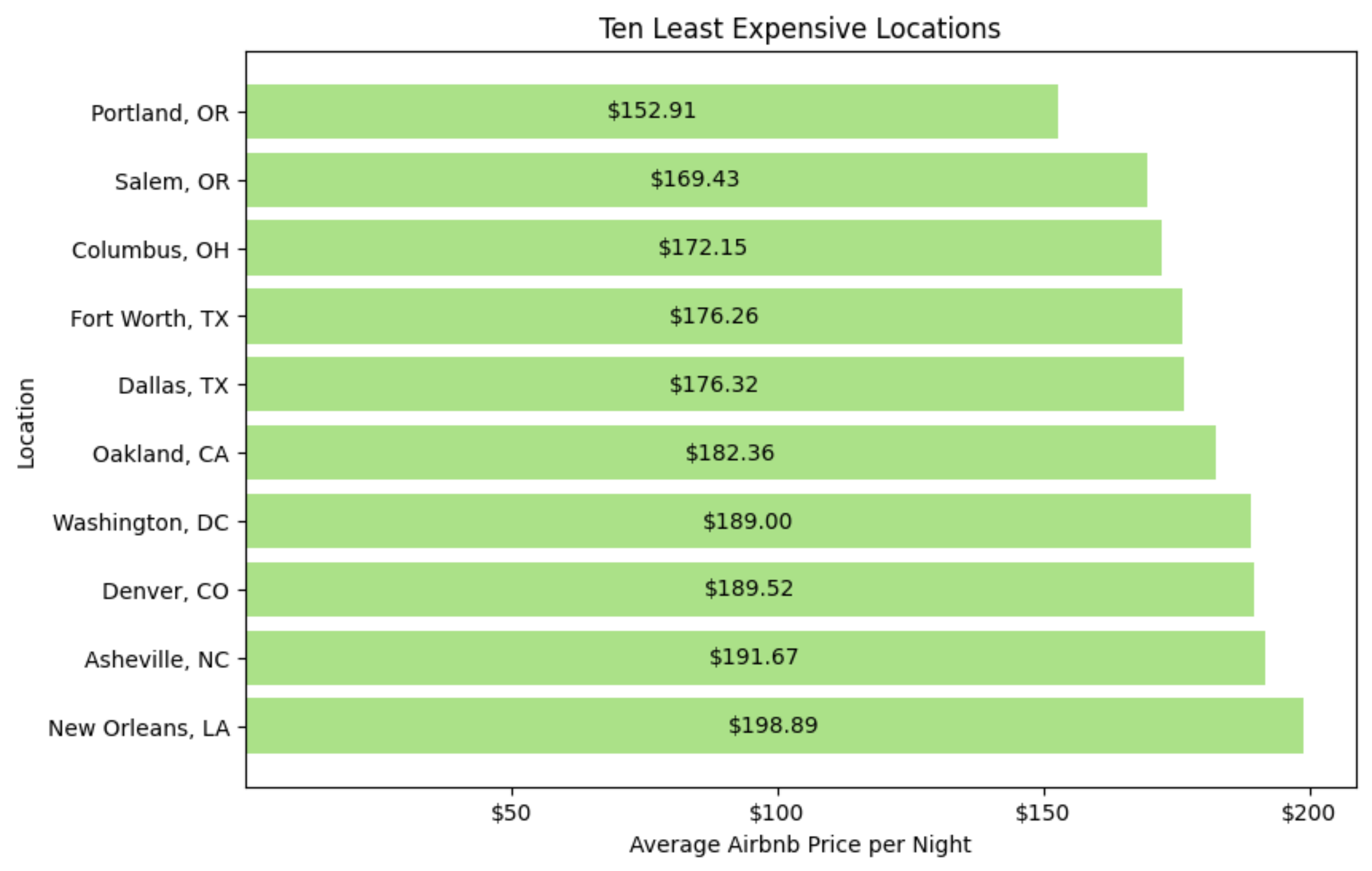
02

EXPLORATORY DATA ANALYSIS

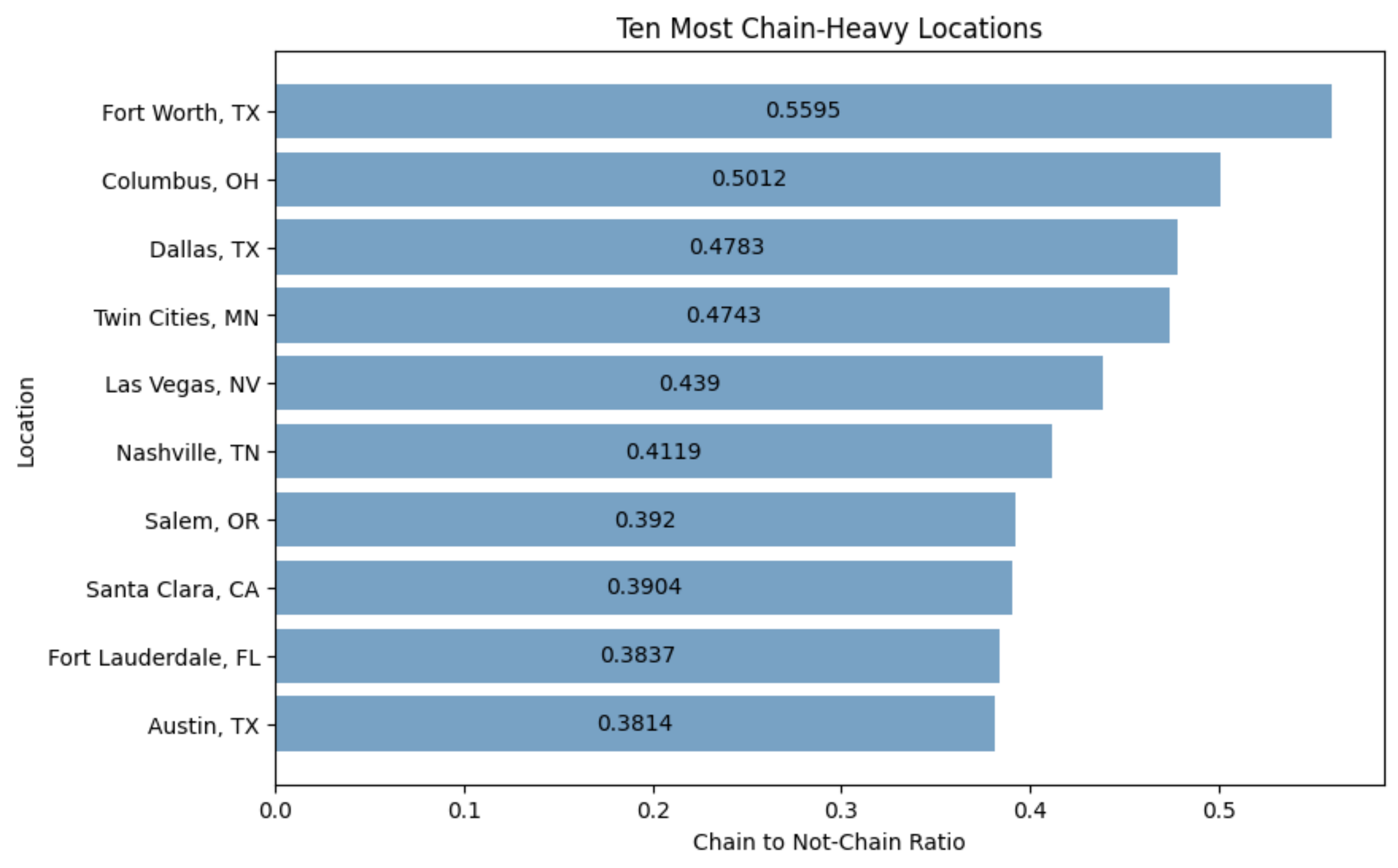
WHICH LOCATIONS HAVE THE MOST EXPENSIVE AIRBNB LISTINGS?



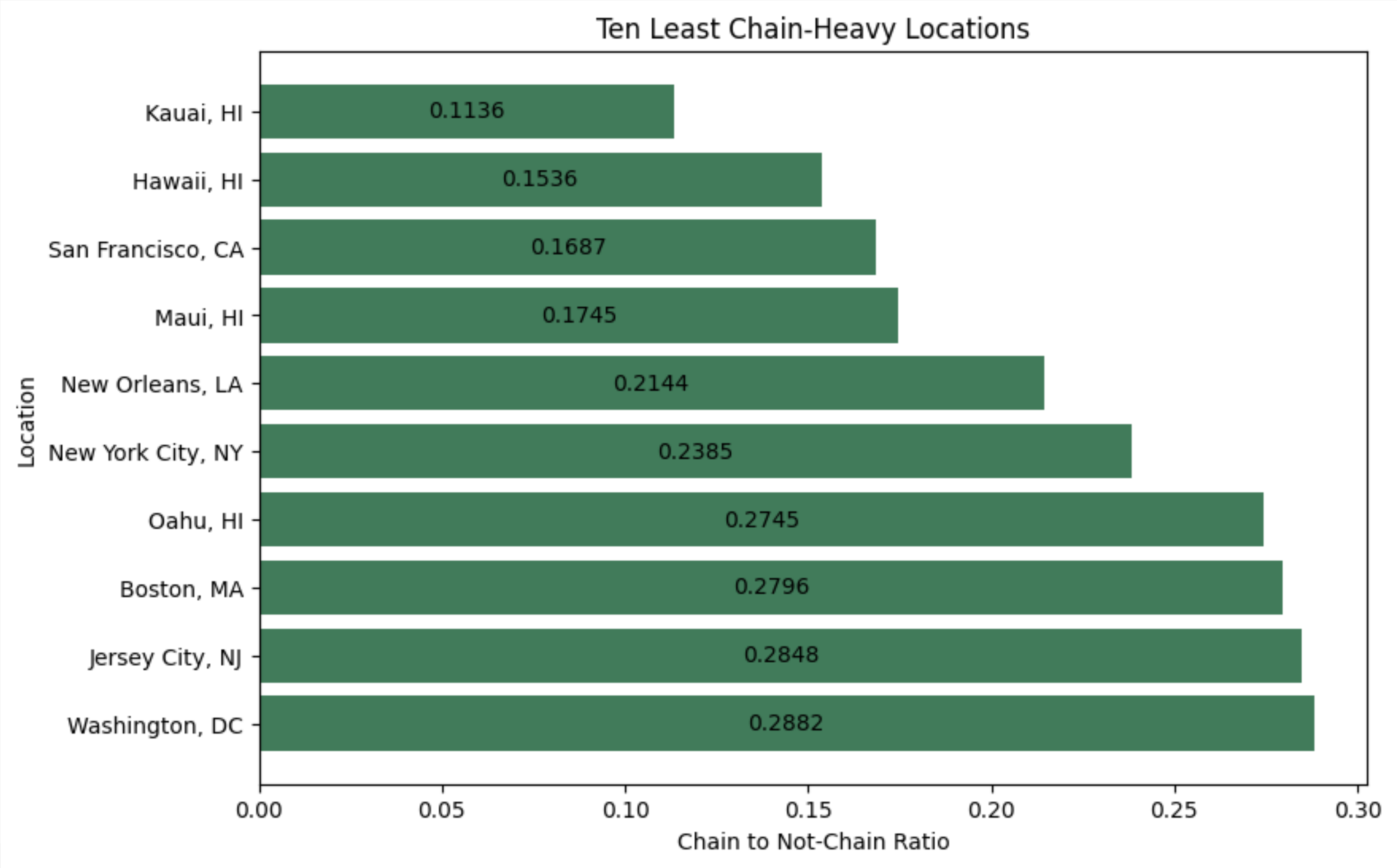
WHICH LOCATIONS HAVE THE LEAST EXPENSIVE AIRBNB LISTINGS?



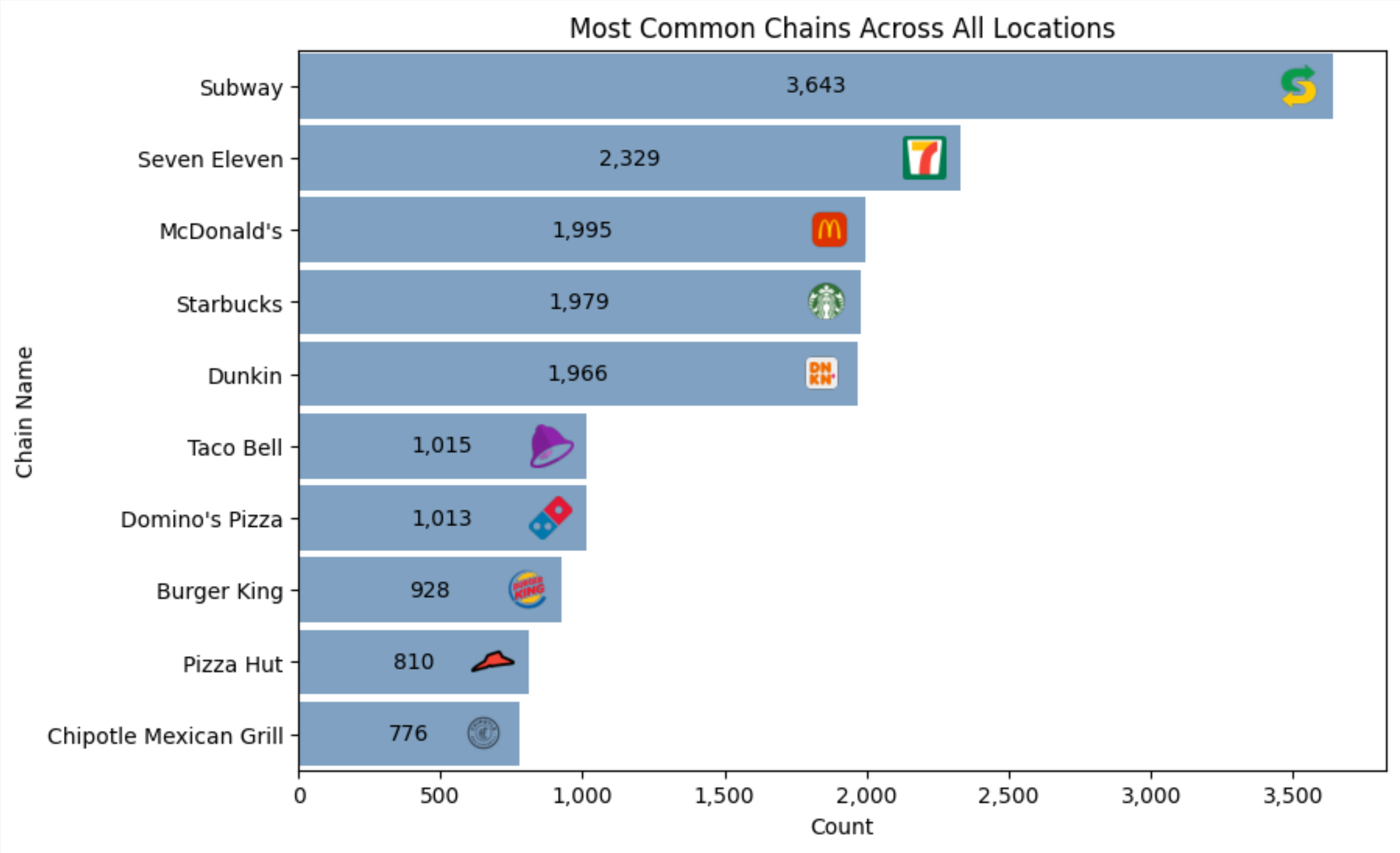
WHICH LOCATIONS HAVE THE HIGHEST RATIOS OF CHAIN RESTAURANTS?



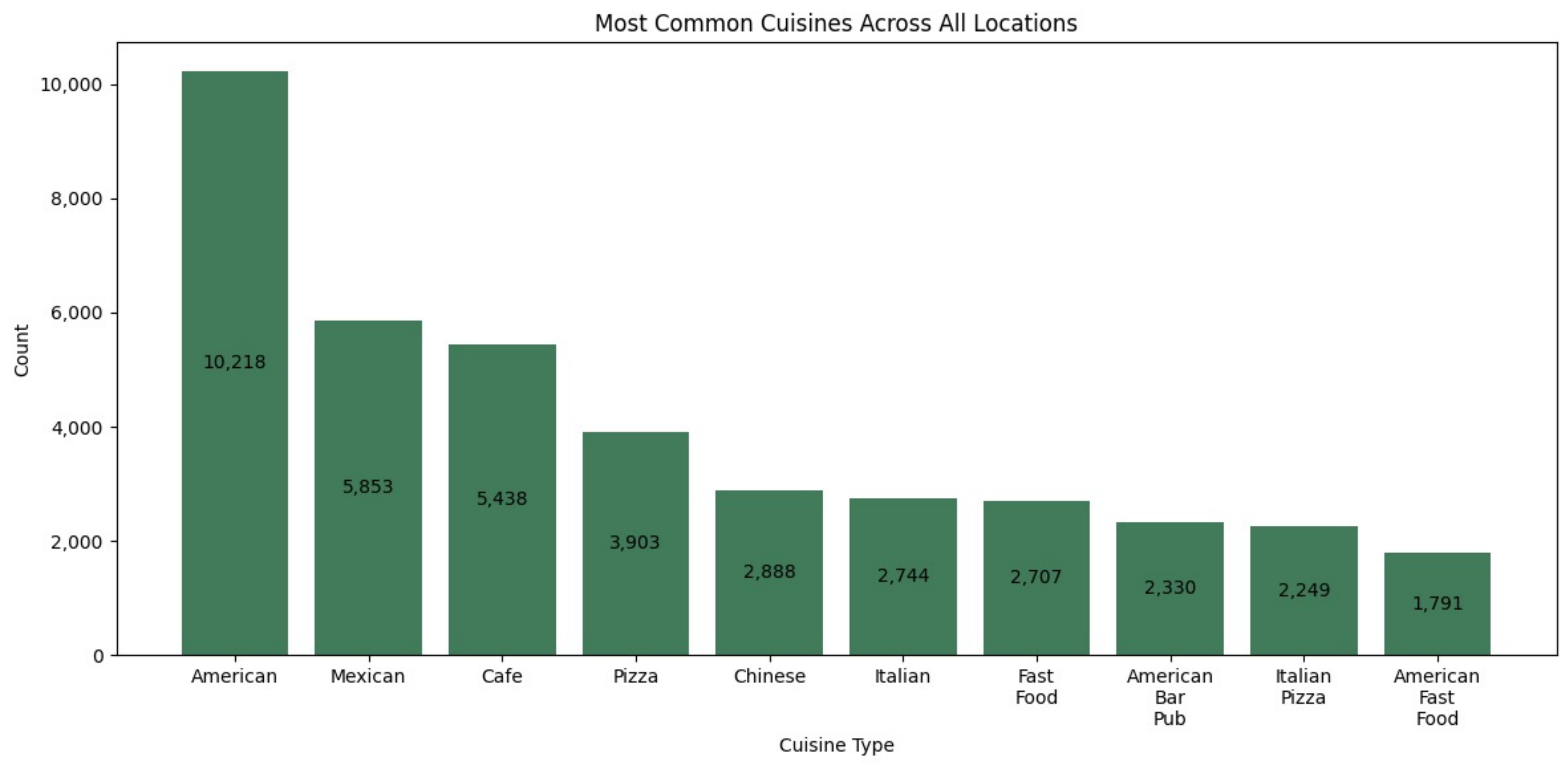
WHICH LOCATIONS HAVE THE LOWEST RATIOS OF CHAIN RESTAURANTS?



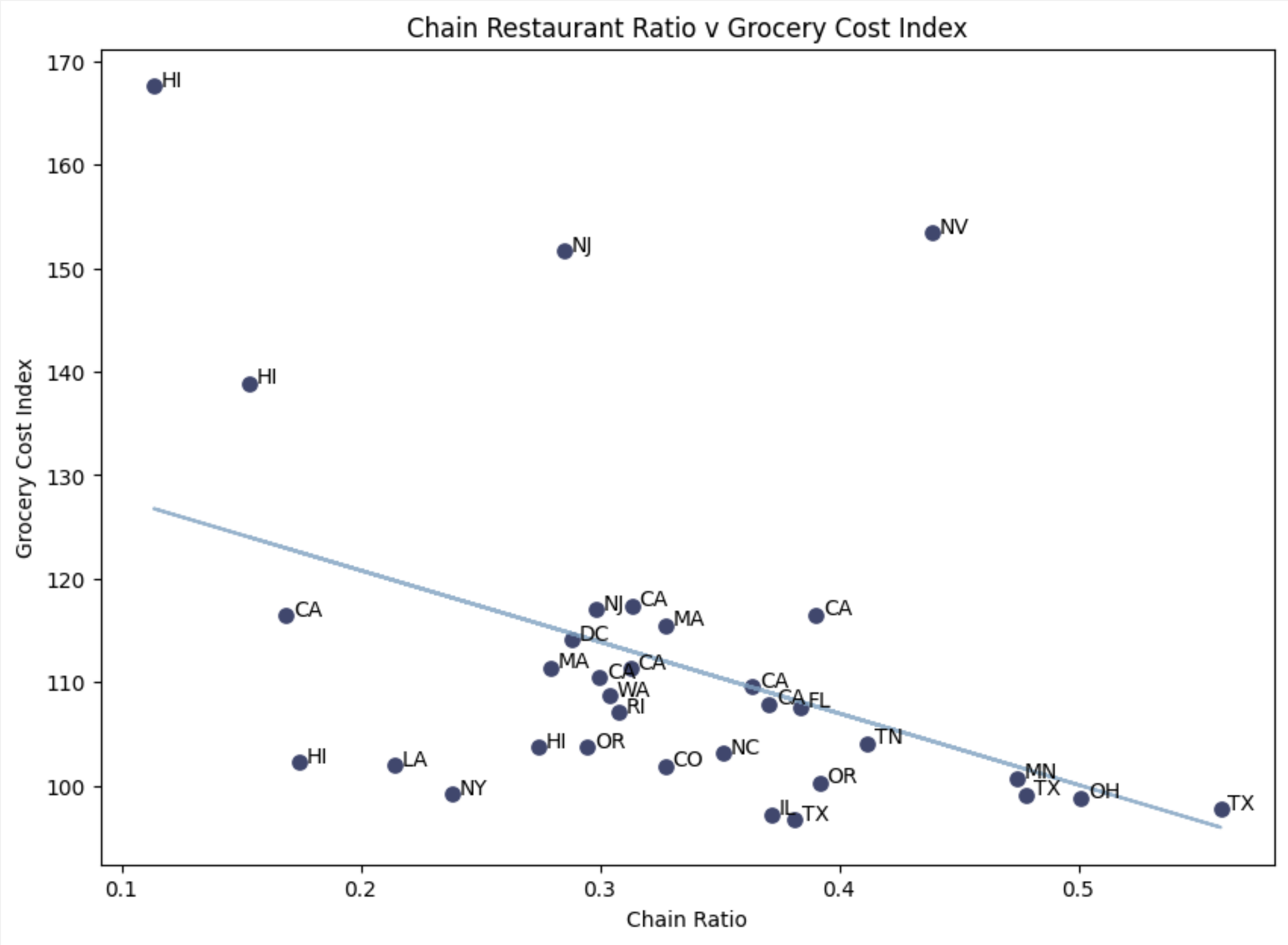
WHAT ARE THE MOST COMMON CHAINS ACROSS ALL LOCATIONS?



WHAT ARE THE MOST COMMON CUISINES ACROSS ALL LOCATIONS?

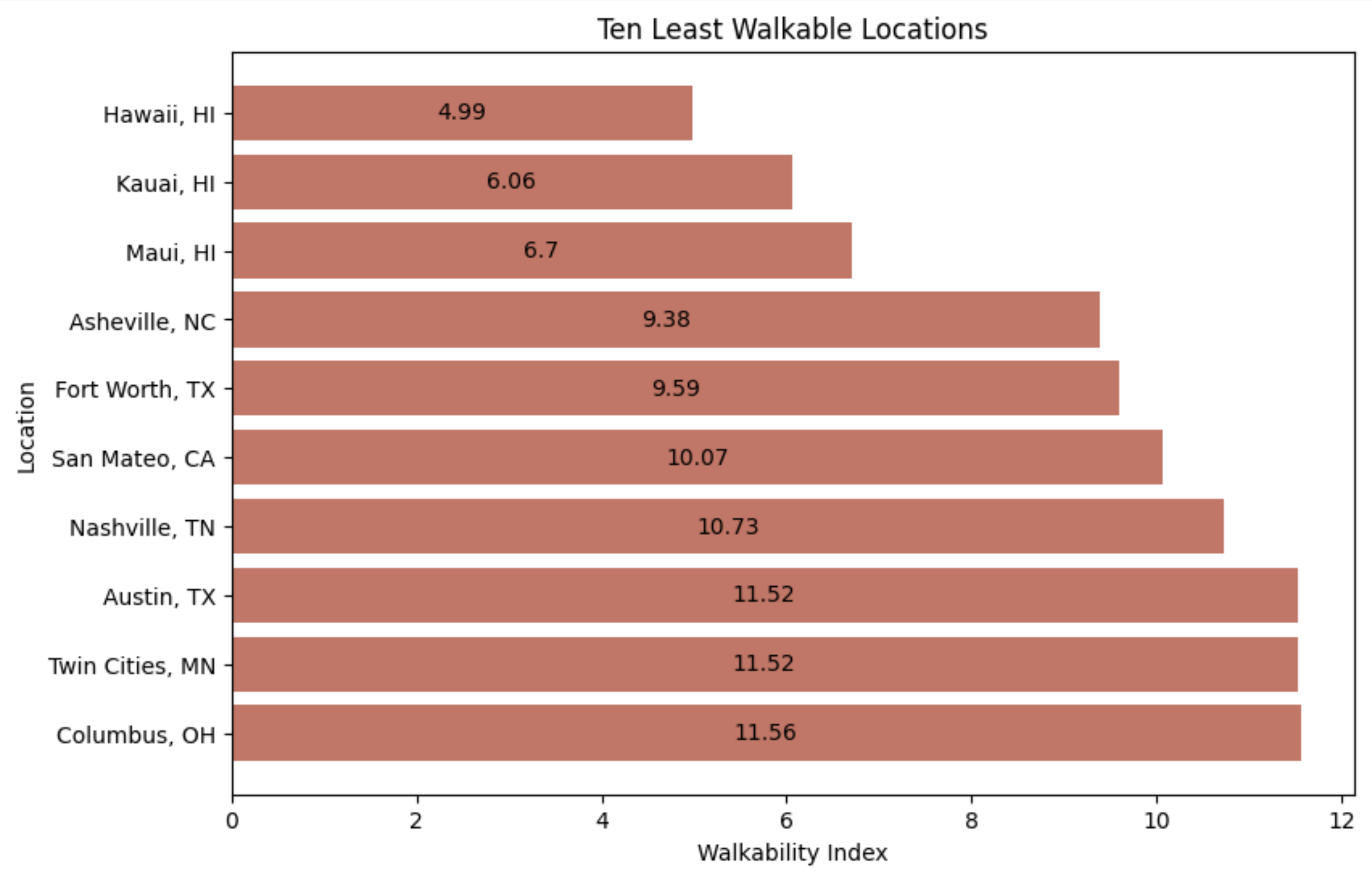


ARE THERE MORE CHAINS IN LOCATIONS WHERE THE GROCERY COST INDEX IS HIGHER?

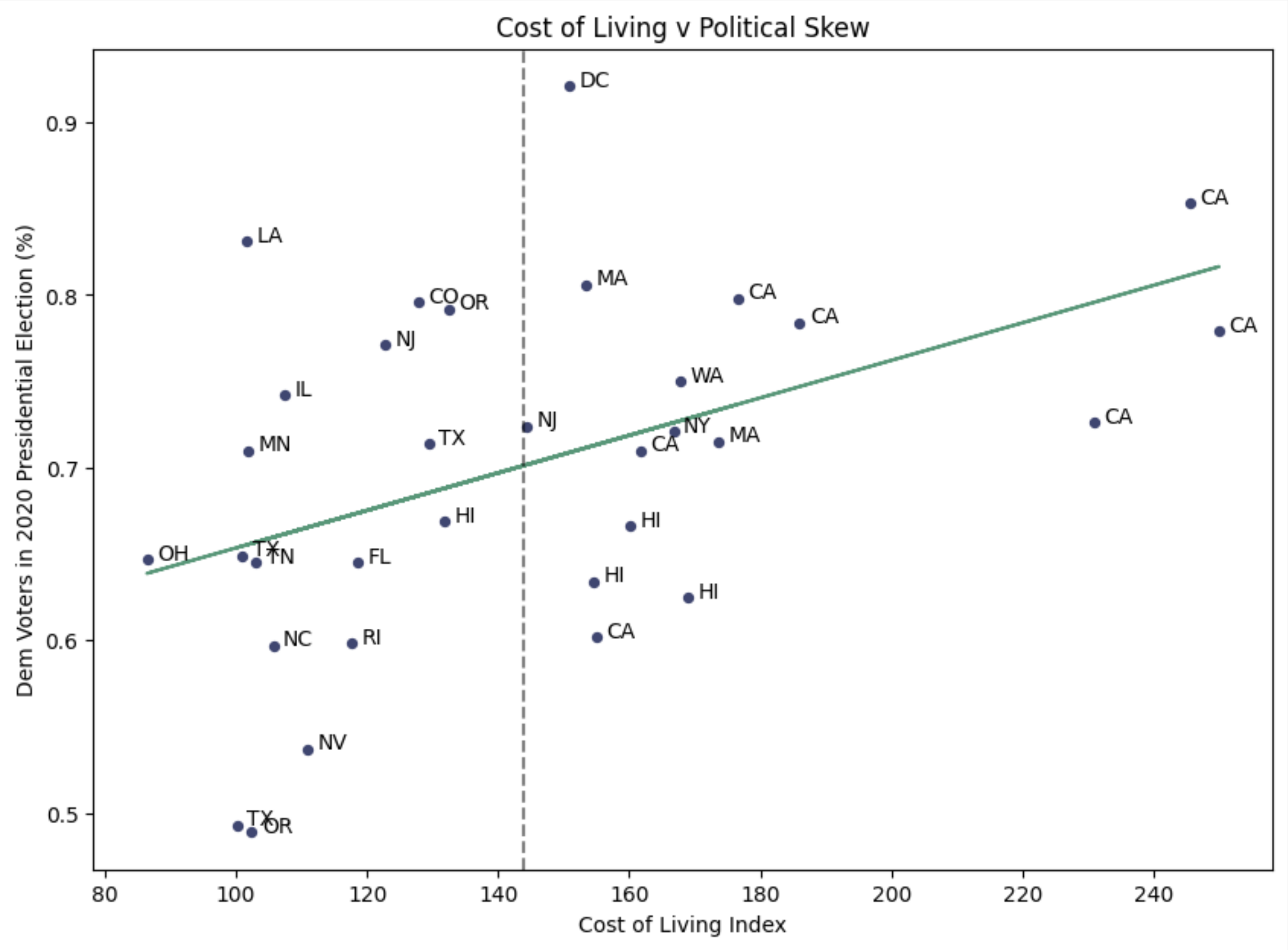


WHICH LOCATIONS HAVE THE HIGHEST WALKABILITY INDICES?

WHICH LOCATIONS HAVE THE LOWEST WALKABILITY INDICES?



DO WE SEE A RELATIONSHIP BETWEEN COST OF LIVING AND POLITICAL LEAN?



03

FUNCTION BUILDING

FUNCTION BUILDING

1. **User Input Conversion**

Meant to take in responses from a live user and convert into usable metrics.

FUNCTION BUILDING

1. **User Input Conversion**

Meant to take in responses from a live user and convert into usable metrics.

2. **Cosine Similarity for Single Recommendation**

Finds the single location with the nearest cosine similarity and returns as a recommendation.

FUNCTION BUILDING

1. **User Input Conversion**

Meant to take in responses from a live user and convert into usable metrics.

2. **Cosine Similarity for Single Recommendation**

Finds the single location with the nearest cosine similarity and returns as a recommendation.

3. **Cosine Similarity for Multiple Recommendations**

Finds the five locations with nearest cosine similarity and returns all five in recommendation.

FUNCTION BUILDING

1. **User Input Conversion**

Meant to take in responses from a live user and convert into usable metrics.

2. **Cosine Similarity for Single Recommendation**

Finds the single location with the nearest cosine similarity and returns as a recommendation.

3. **Cosine Similarity for Multiple Recommendations**

Finds the five locations with nearest cosine similarity and returns all five in recommendation.

4. **KMeans Cluster for Randomized Recommendation**

Uses KMeans to group the locations, then returns a random recommendation from the list of locations that were in the same cluster.

FUNCTION BUILDING

1. **User Input Conversion**

Meant to take in responses from a live user and convert into usable metrics.

2. **Cosine Similarity for Single Recommendation**

Finds the single location with the nearest cosine similarity and returns as a recommendation.

3. **Cosine Similarity for Multiple Recommendations**

Finds the five locations with nearest cosine similarity and returns all five in recommendation.

4. **KMeans Cluster for Randomized Recommendation**

Uses KMeans to group the locations, then returns a random recommendation from the list of locations that were in the same cluster.

04

STREAMLIT APPLICATION

STREAMLIT APP BUILDING

What was helpful?

STREAMLIT APP BUILDING

What was helpful?

- A lot of features are templatized

STREAMLIT APP BUILDING

What was helpful?

- A lot of features are templatized
- Easy importing of dataframes and functions

STREAMLIT APP BUILDING

What was helpful?

- A lot of features are templatized
- Easy importing of dataframes and functions
- Different ways of collecting user input

STREAMLIT APP BUILDING

What was helpful?

- A lot of features are templatized
- Easy importing of dataframes and functions
- Different ways of collecting user input
- Syntax is intuitive

STREAMLIT APP BUILDING

What was helpful?

- A lot of features are templatized
- Easy importing of dataframes and functions
- Different ways of collecting user input
- Syntax is intuitive

What were the obstacles?

STREAMLIT APP BUILDING

What was helpful?

- A lot of features are templatized
- Easy importing of dataframes and functions
- Different ways of collecting user input
- Syntax is intuitive

What were the obstacles?

- A lot of features are templatized

STREAMLIT APP BUILDING

What was helpful?

- A lot of features are templatized
- Easy importing of dataframes and functions
- Different ways of collecting user input
- Syntax is intuitive

What were the obstacles?

- A lot of features are templatized
- Page reloads after any value modification by user

STREAMLIT APP BUILDING

What was helpful?

- A lot of features are templatized
- Easy importing of dataframes and functions
- Different ways of collecting user input
- Syntax is intuitive

What were the obstacles?

- A lot of features are templatized
- Page reloads after any value modification by user
- Clearing page after submission

STREAMLIT APP BUILDING

What was helpful?

- A lot of features are templatized
- Easy importing of dataframes and functions
- Different ways of collecting user input
- Syntax is intuitive

What were the obstacles?

- A lot of features are templatized
- Page reloads after any value modification by user
- Clearing page after submission
- Adding personalization

05

CONCLUSIONS AND NEXT STEPS

CONCLUSIONS

Combining multiple datasets comes with challenges

- Some datasets were pulled at the neighborhood-specific level of each location, while others were pulled at the city level
- This does slightly diminish the interpretability of our findings

CONCLUSIONS

Combining multiple datasets comes with challenges

- Some datasets were pulled at the neighborhood-specific level of each location, while others were pulled at the city level
- This does slightly diminish the interpretability of our findings

There are likely patterns in the data that were overlooked

- Unable to fully explore each dataset due to timeline and scope of project
- Probable that trends exist within the data that could be useful in web app/functions

CONCLUSIONS

Combining multiple datasets comes with challenges

- Some datasets were pulled at the neighborhood-specific level of each location, while others were pulled at the city level
- This does slightly diminish the interpretability of our findings

There are likely patterns in the data that were overlooked

- Unable to fully explore each dataset due to timeline and scope of project
- Probable that trends exist within the data that could be useful in web app/functions

The app works, but could be improved

- While we were able to build an application that provides and recommendation and directs users toward the Airbnb website, there are a few necessary revisions before presenting to the client and going to market

NEXT STEPS

1. **Additional locations**

- Ideally, there would be various cities recommended from every state and eventually, international locations

2. **NLP on the Airbnb listing names and descriptions**

- Run the words used in each listing name and description through Count Vectorizer/Tfidf, and from there, generate a list of selected words the user can choose from that match their preferences

3. **More data**

- Local communities and activities (e.g. music scene if user wants to live somewhere that has multiple venues or more opportunities to see live music), nightlife, population diversity, etc.