



LISA GREER
DATA ANALYST PORTFOLIO

Table of Contents

Projects:

#1 GameCo

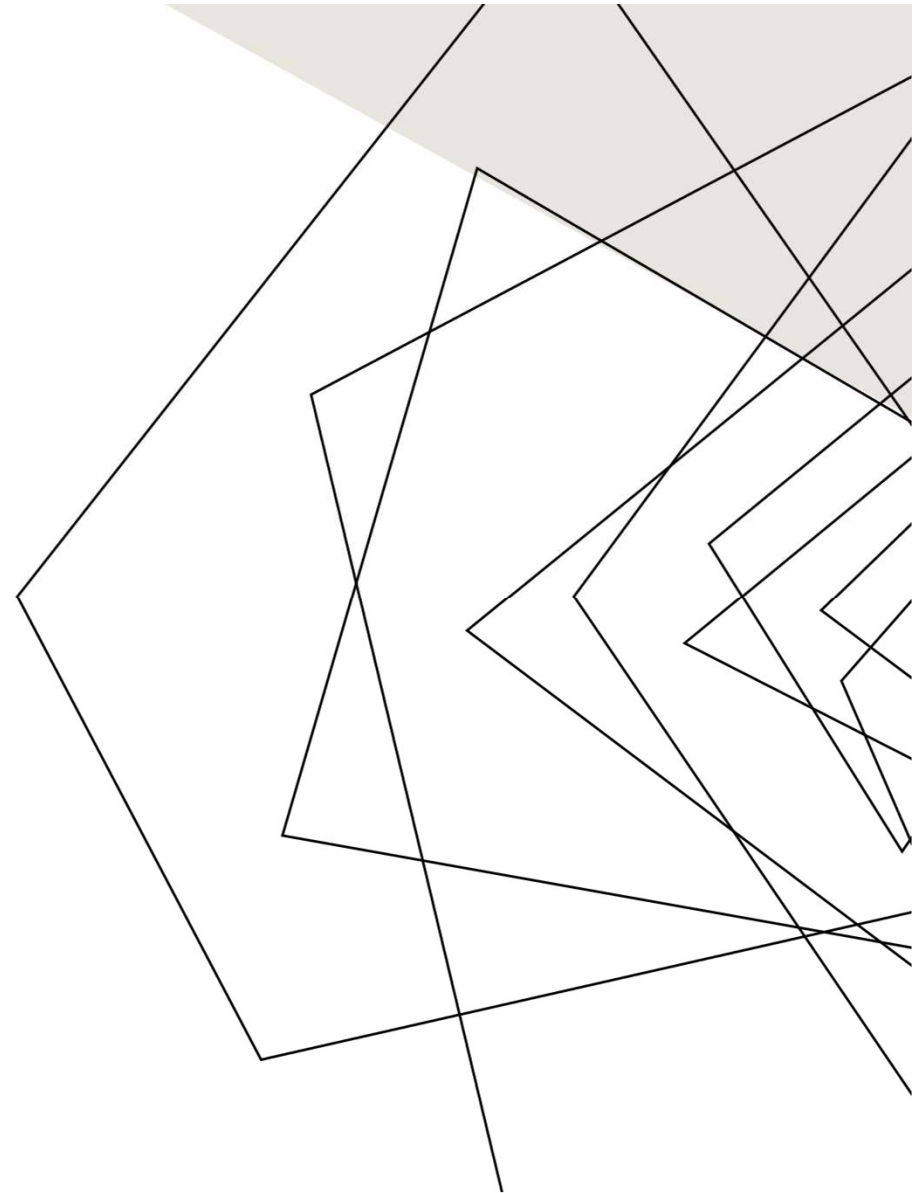
#2 Center for Disease Control

#3 Rockbuster Stealth LLC

#4 Instacart

#5 Pig E. Bank

#6 USA Real Estate Market



SKILLSETS



Project #1 – GameCo

Excel

Analyzing Global Video Game
Sales by Region



Business Goals and Data

Goals

- Assess regional differences in video game sales over time
- Assess trends in regional markets
- Provide insight into GameCo's next year's marketing budget based on region

Data

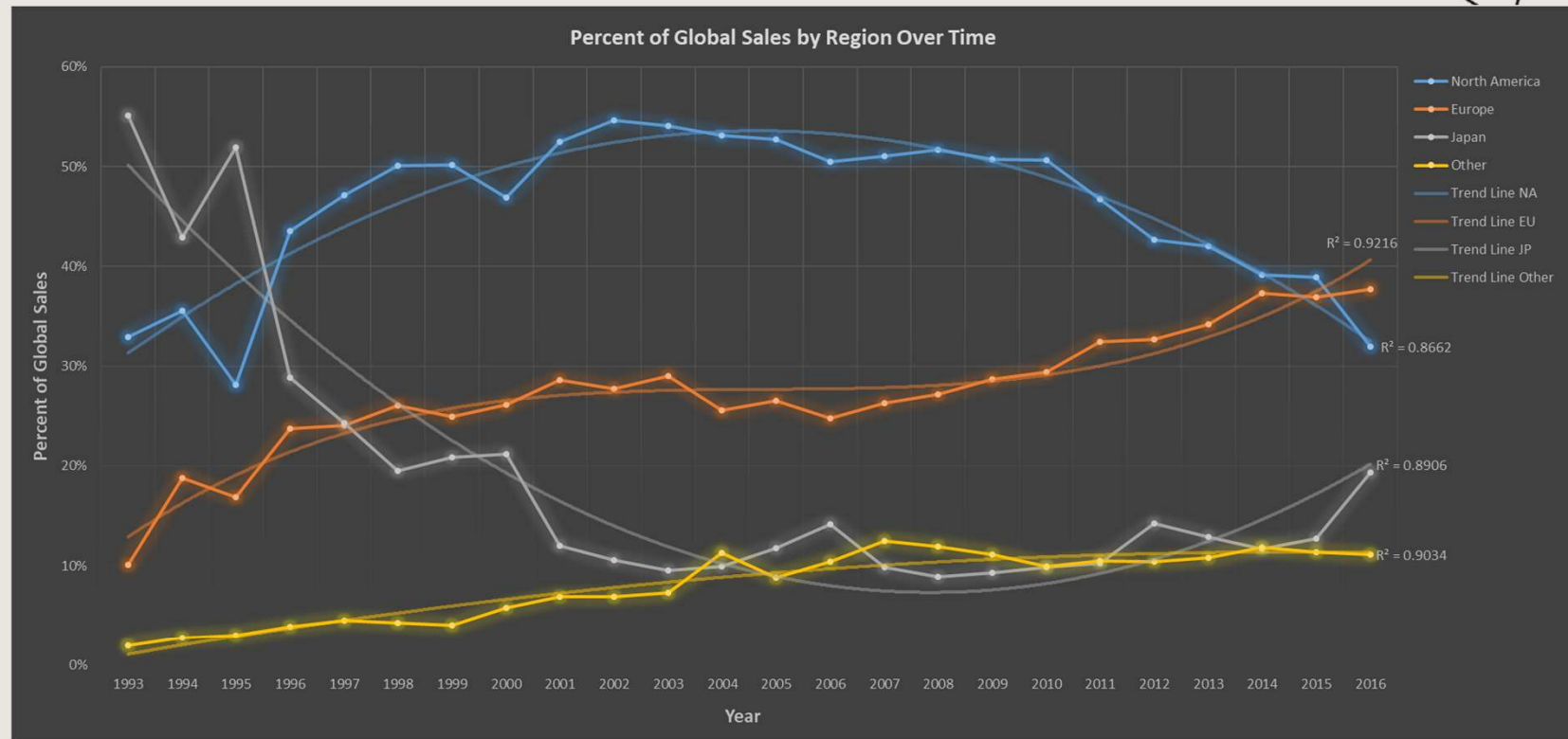
- ~16,000 game titles
- 1980 to 2016
- North America, European, Japanese, and Other regional sales categories measured in million units sold
- Data was cleaned, consolidated, and made uniform in Excel
- Pivot tables and graphs were used to derive key insights and conduct statistical analysis

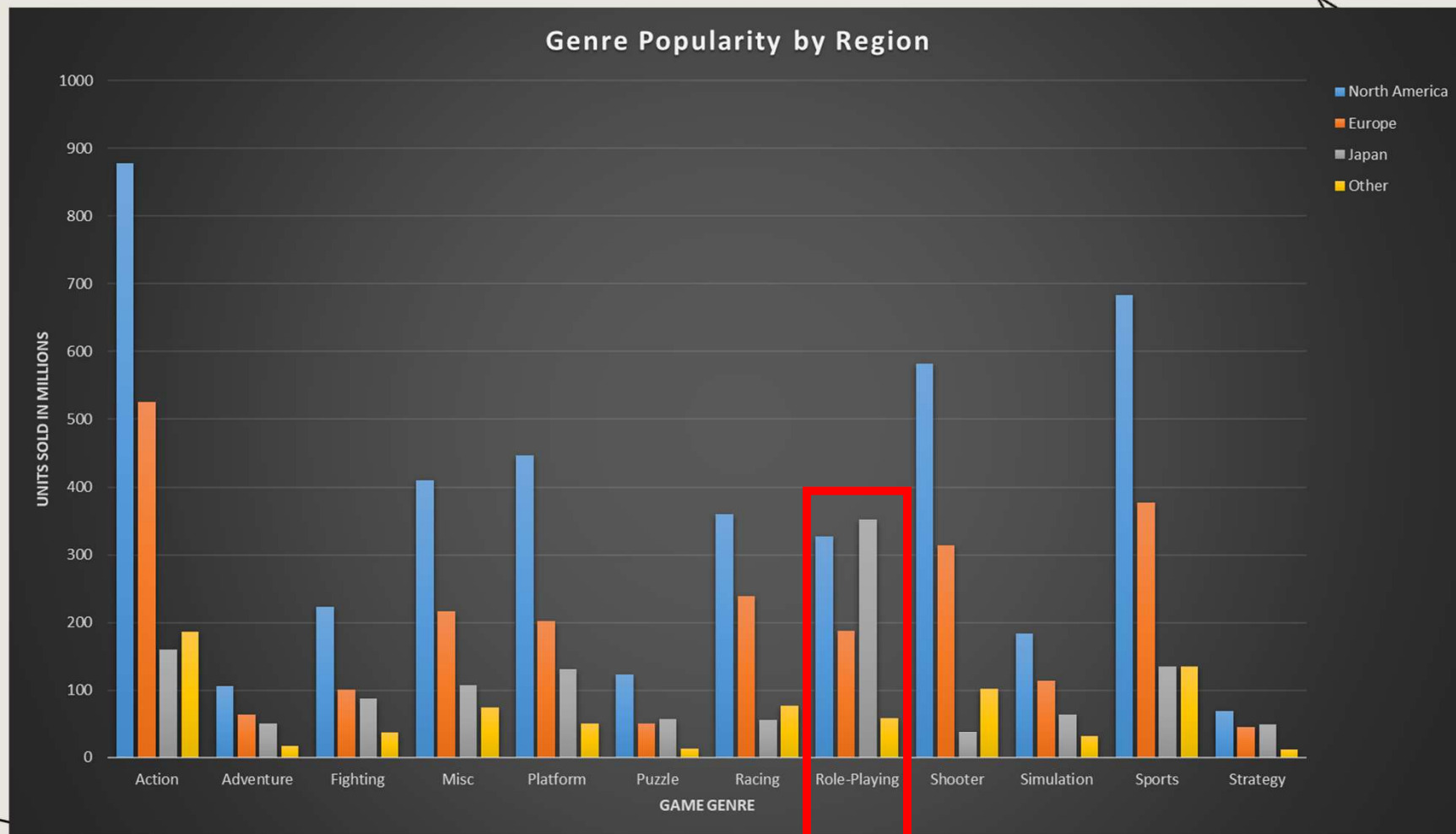
Key Insights:

- Japanese and European sales are rising
- North American sales are declining

Course of Action:

- Add resources to the North American budget depending on overall marketing strategy





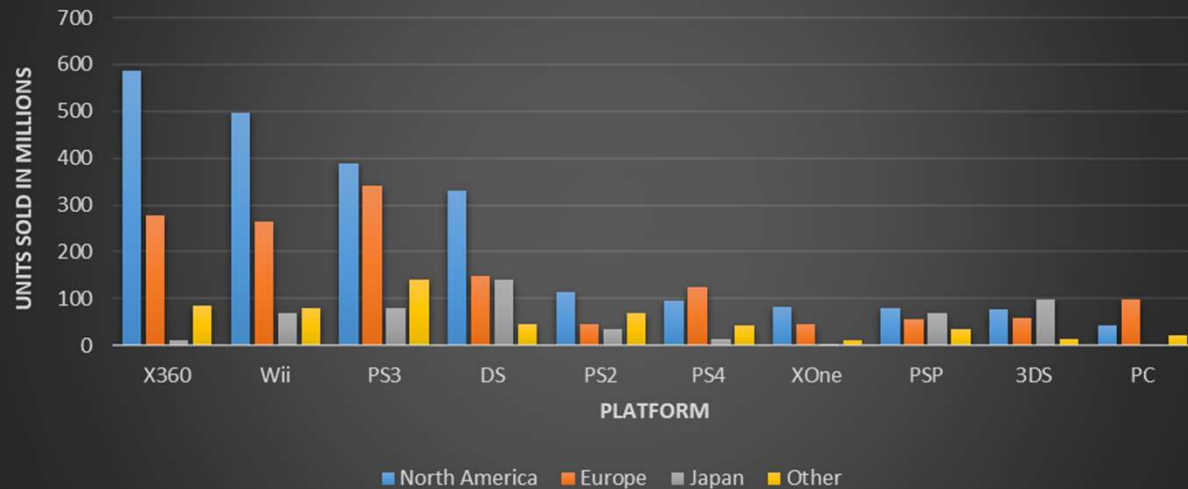
Key Insight:

- The outlying trend is the comparatively high ratio of role-playing games' popularity in Japan.

Course of Action:

- Target marketing role-playing games in Japan

Top 10 Most Popular Platforms from 2006-2016



Key Insights:

- X360, Wii, and PS3 are top selling platforms
- The DS sells particularly strong in North America

Course of Action:

- Focus marketing budget on top selling platforms

Platform	North America	Europe	Japan	Other
X360	587	278	12	84
Wii	497	264	68	79
PS3	389	340	79	141
DS	331	148	141	44
PS2	115	46	36	69
PS4	97	124	14	43
XOne	83	46	0	12
PSP	80	56	70	34
3DS	78	58	97	12
PC	43	98	0	21

Project #2 - Center for Disease Control (CDC)

Excel/Tableau

Finding Insights for a Medical Staffing Agency to Prepare for Influenza Season in the United States

[Link to Tableau Story](#)

[My Storytelling to the Problem](#)



Business Goals and Data

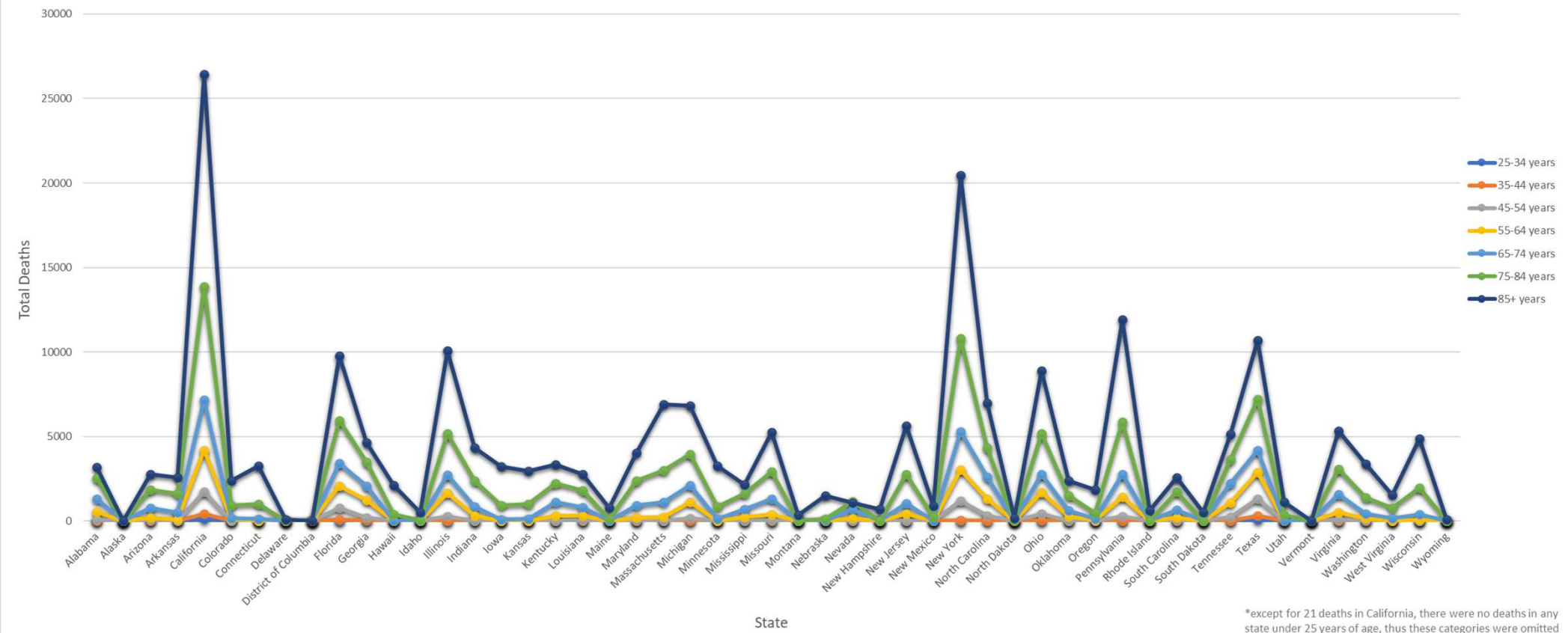
Goals

- Assess regional and seasonal differences in the flu's effects across the United States
- Assess trends in deaths from flu across time and region in the United States
- Provide insight into at risk populations and how they are affected
- Provide insight into where, when, and how much staff to send to each state

Data

- USA Census Data
 - Drawn from the US Government
 - 2009-2017 all U.S. county populations broken down by age group and year
- CDC Influenza Deaths Data
 - Drawn from the CDC
 - 2009-2017 all U.S. deaths by flu broken down by state, age group, month, and year
 - Death records with values between 0-9 were "suppressed" so death counts of this data were brought to a value of "0"
- Excel tools and Tableau, including pivot tables and statistical analysis and charts, were used to clean and extrapolate meaning from the data

Deaths per State by Age Group from 2009 - 2017



Key Insights:

- Regardless of state, 65+ years of age is at much higher risk of death from influenza

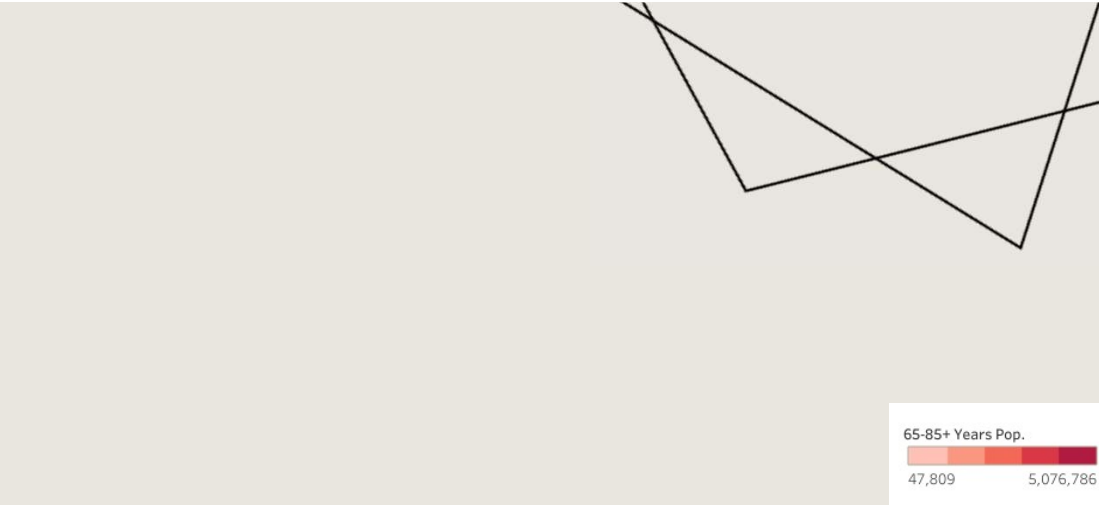
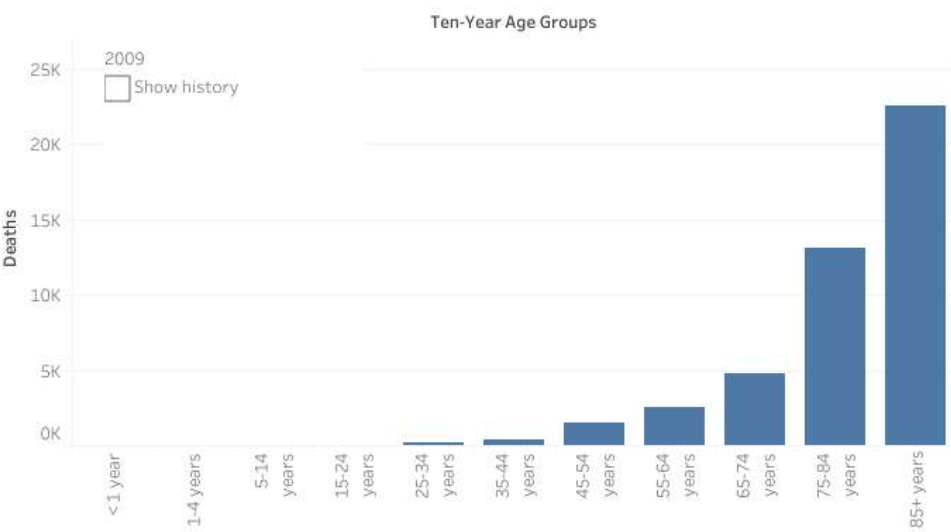
Course of Action:

- Send more resources to states with higher populations

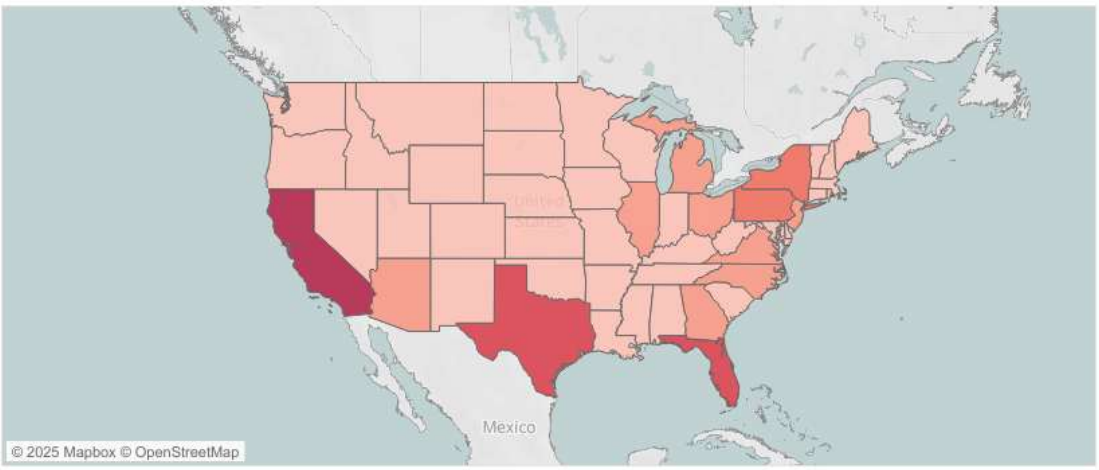
Preparing for Influenza Season

During influenza season there is high demand for additional medical staff to respond to patient needs across the United States. This analysis provides insight into when and where to send additional support to protect vulnerable populations (those 65+ years old) and minimize influenza's impact.

Deaths by Age Group 2009-2017



Vulnerable Population Demographics 2009-2017



The chart indicates the vulnerable population of 65+ years (as defined by the CDC) should be the priority for the upcoming flu season. Ages <64 are negligibly affected comparatively. The map indicates where this vulnerable population is predominantly located.

The Plan:

Based on 2017 populations, states can be prioritized based on vulnerable population size.

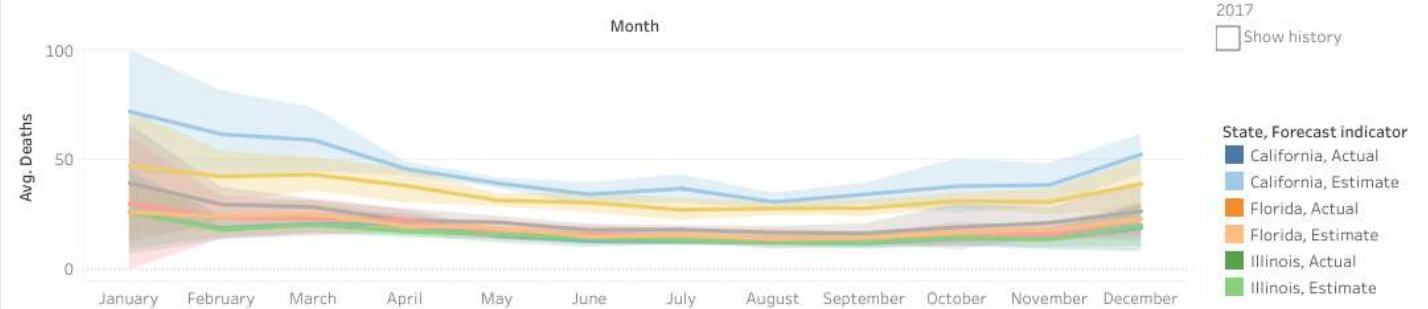
High Priority = California and Florida, **Medium Priority** = New York, Texas, Pennsylvania, Ohio, Illinois, **Low Priority** = All Others

Priority months for all states are January, February, March, November, and December.

Limitations:

Due to data limitations, the exact amount of staff to send per patient is prevented at this time.

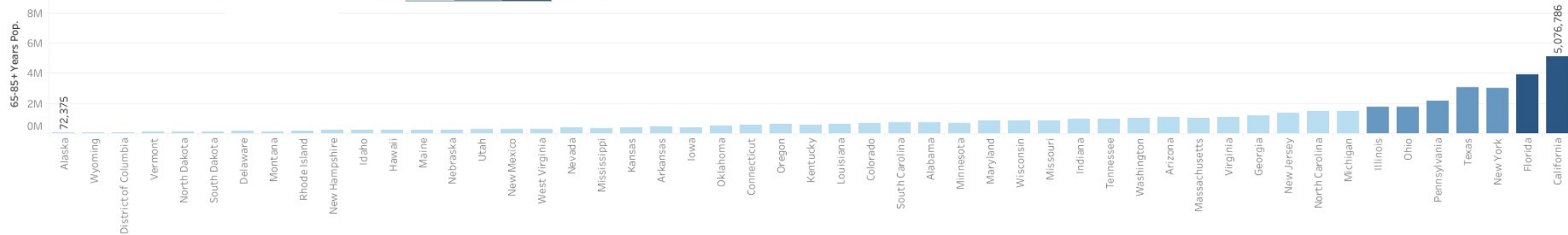
Priority States by Month - 2017



Priority States - 2017

2017
☐ Show history

65-85+ Years Pop.
47,809 5,076,786



Project #3 - Rockbuster Stealth LLC

SQL/ Tableau/Data Warehousing

Finding Data-Driven Answers to Business Questions for an Online Video Rental Company

[Link to GitHub Repository](#)



Business Goals and Data

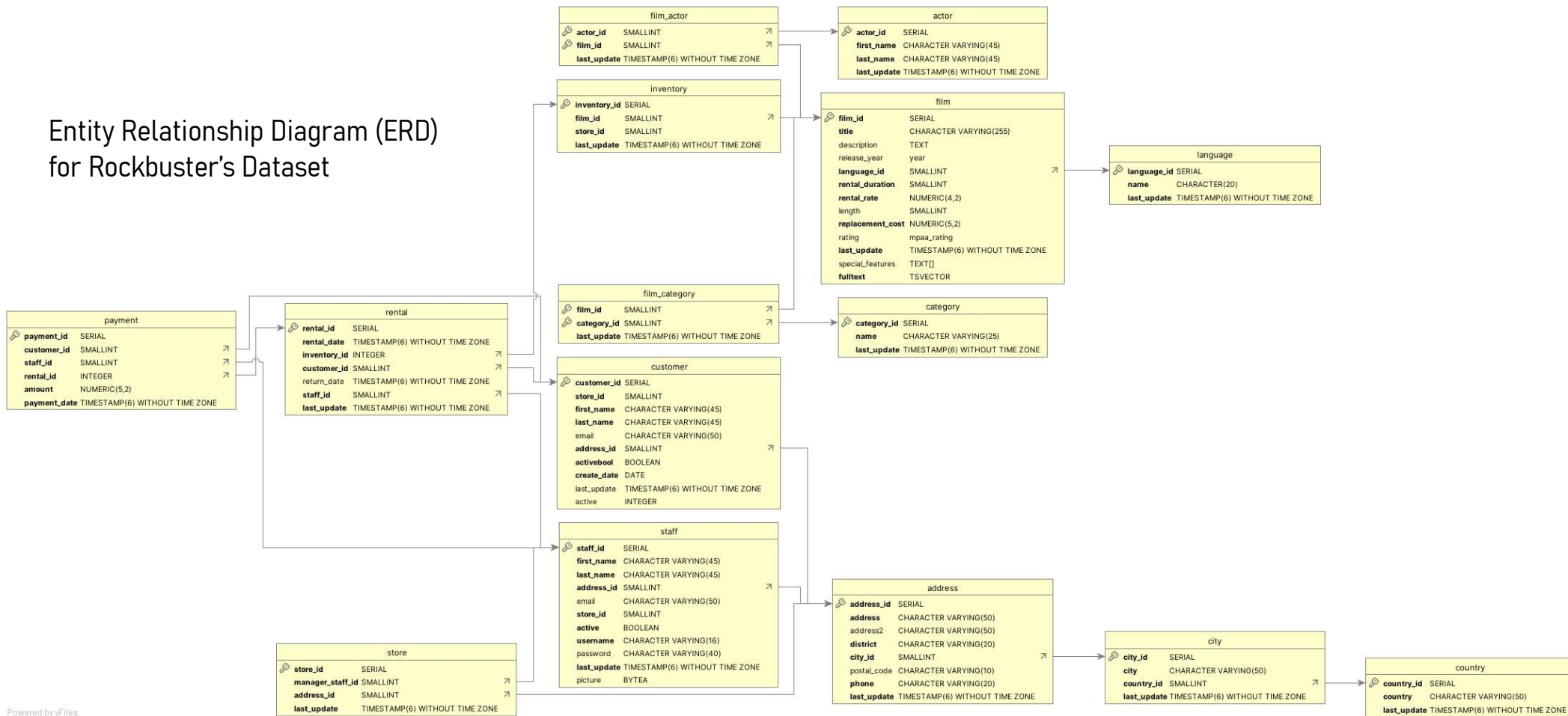
Goals

- Assess regional differences in sales and customers
- Assess most and least profitable movies
- Provide insight into customer rental trends

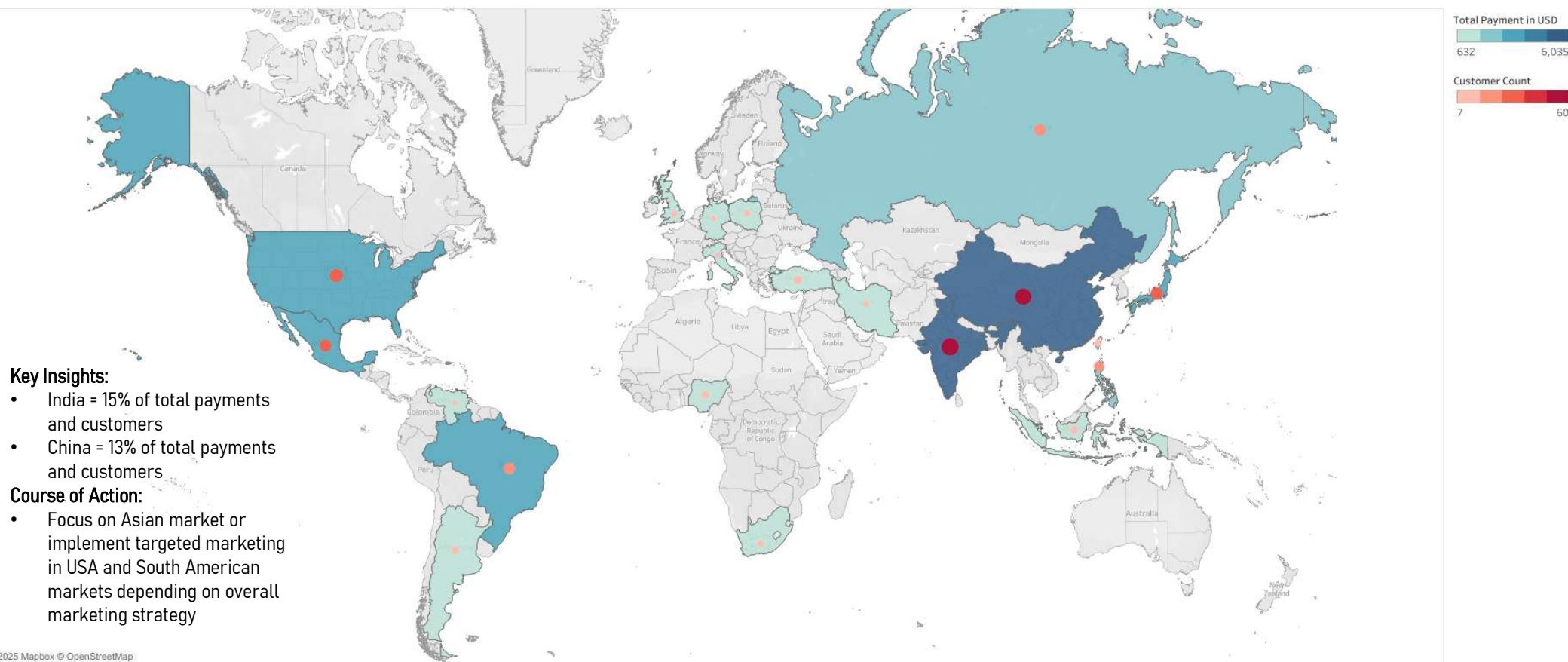
Data

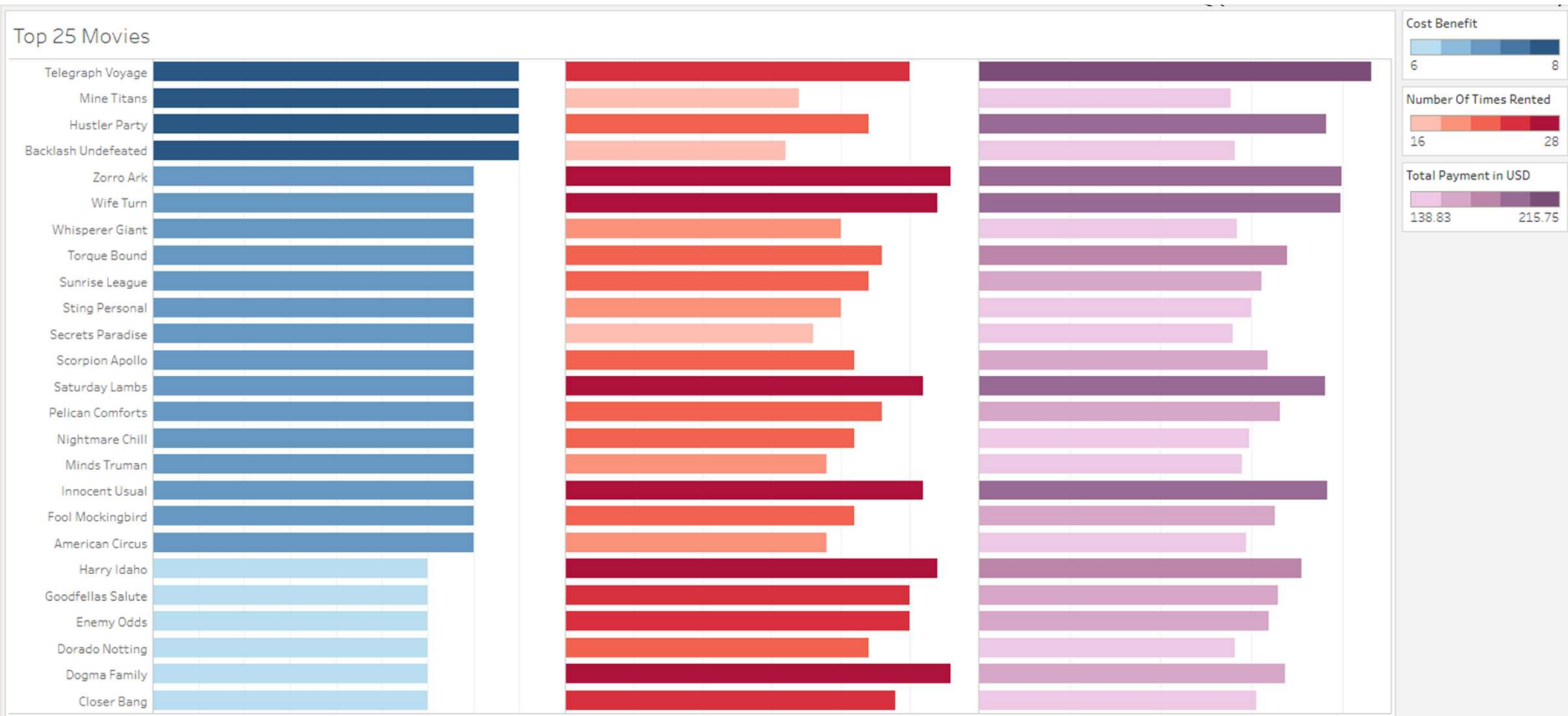
- Sales and customer data from Rockbuster
- Data was cleaned, formatted, and joined for statistical analysis and pattern identification
- Excel, Tableau, and SQL queries were used to clean, analyze, and provide insights into the data through visualizations

Entity Relationship Diagram (ERD) for Rockbuster's Dataset



Top 20 Countries by Customer Spending and Customer Count





Key Insights:

- Top 25 were calculated by a ratio of total payment/number of rentals to give a cost benefit

Course of Action:

- Focus on keeping stock of top 25 movies and minimize stock of others to lower overhead costs

Project #4 – Instacart

Python/Anaconda/Jupyter

Finding Insights for a Data-Driven Marketing Strategy for an Online Grocery Store

[Link to GitHub Repository](#)



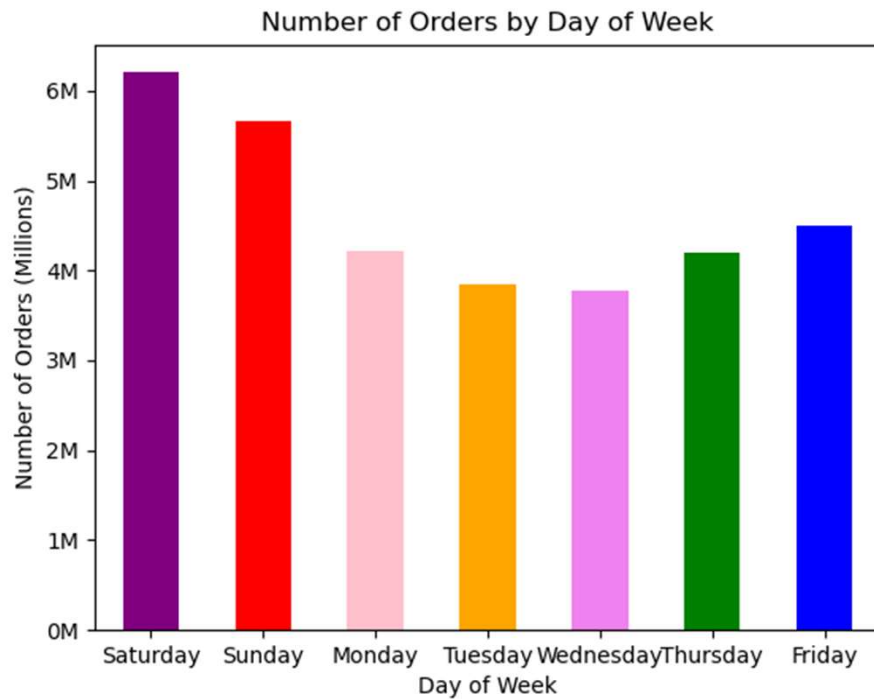
Business Goals and Data

Goals

- Assess busy times of days and weeks based on traffic and money spent
- Assess effective price range groupings
- Provide insights into customer demographics

Data

- Customer data set containing identifiable personal consumer information
- Order, products, and sales data from Instacart's 2017 sales year
- Excel and Anaconda/Jupyter were used to clean, merge, analyze, and provide insights into the data through visualizations
 - Python libraries including pandas, NumPy, os, matplotlib, scipy, and seaborn used to visualize and conduct statistical analysis on the data

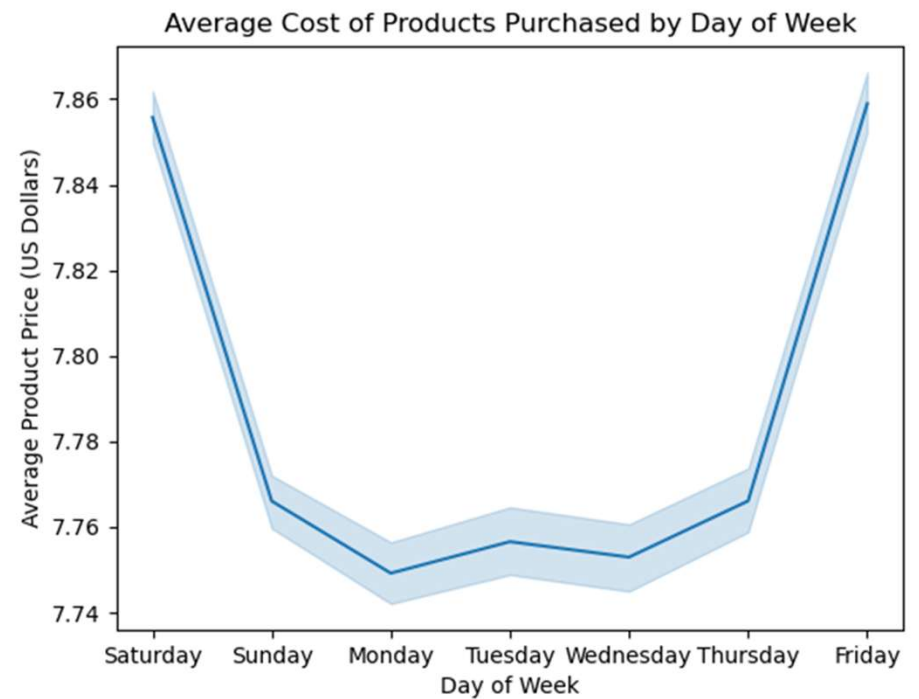


Key Insight:

- The highest volume of sales occurs on weekends.

Course of Action:

- Initiate sales on weekdays to drive traffic



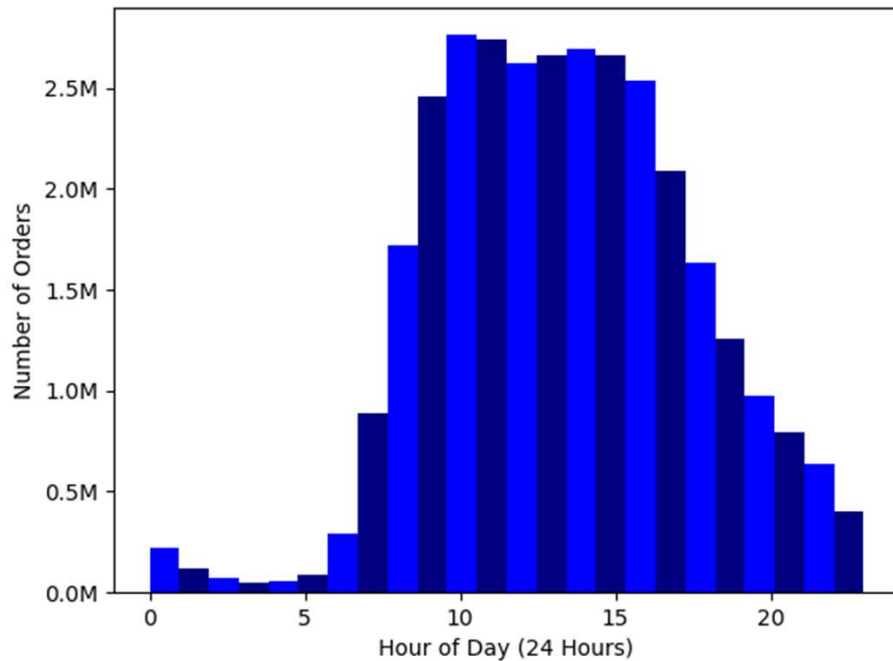
Key Insight:

- The ~12 cent difference from the low to the high end of the chart would indicate a noticeable difference in cost by day of the week, but may not be significant depending on financial goals.

Course of Action:

- May indicate people are more willing to pay for convenience on weekends, retain full prices on Fridays and Saturdays

Popularity of Orders by Time of Day



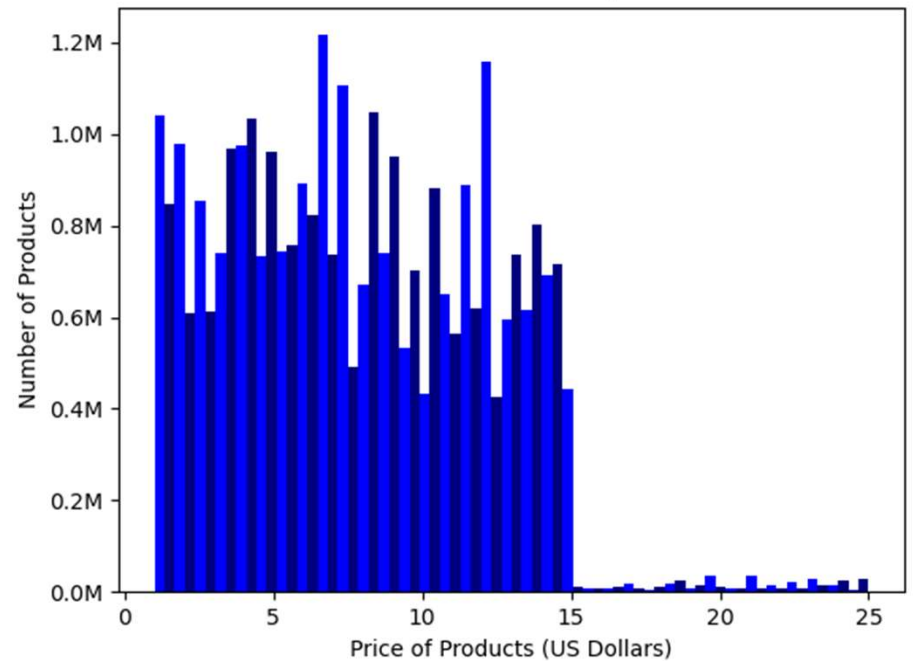
Key Insight:

- 9am to 5pm are when most orders are placed on Instacart.

Course of Action:

- Consider staffing more during busy times
- Push advertising on the shoulders of busy times to increase impulse buys

Number of Products by Price

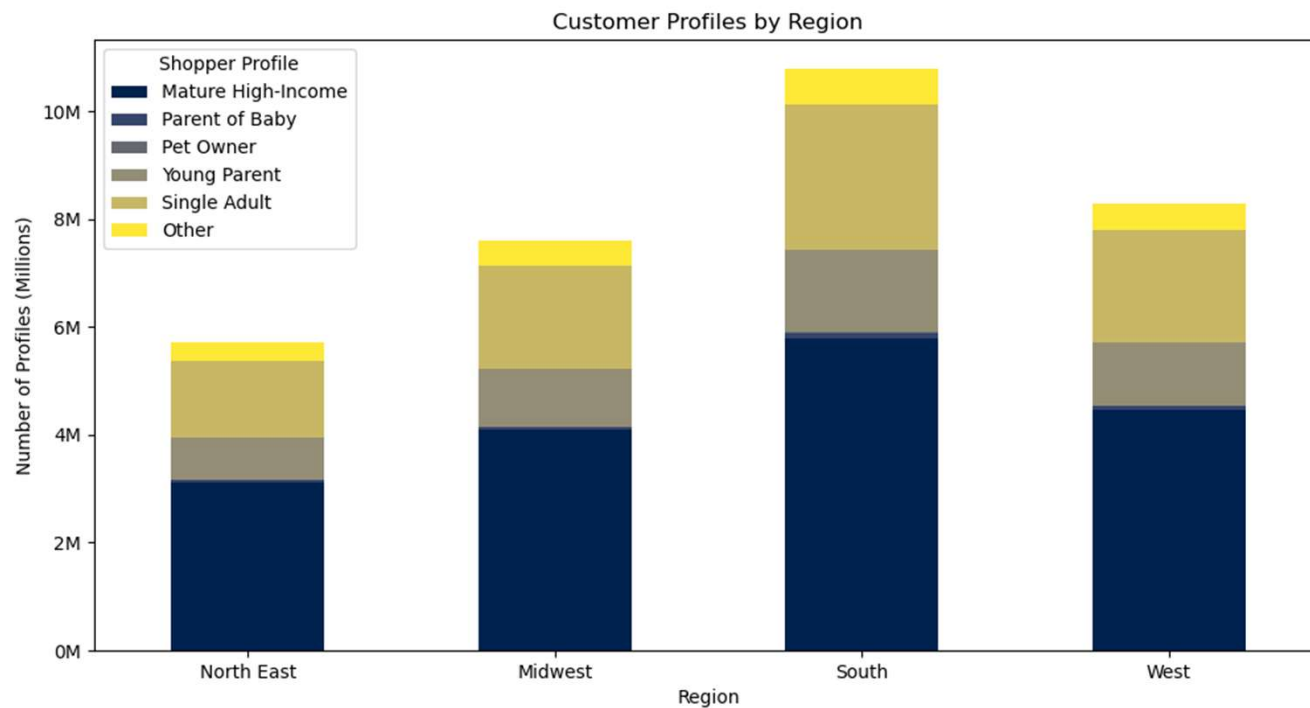


Key Insight:

- Products priced lower than \$15 sell in much higher volumes than those priced greater than or equal to \$15.

Course of Action:

- Products on cusp of \$15 dollars can be lowered below the threshold



North East:	Midwest:	South:	West:
Maine	Wisconsin	Delaware	Idaho
New Hampshire	Michigan	Maryland	Montana
Vermont	Illinois	District of Columbia	Wyoming
Massachusetts	Indiana	Virginia	Nevada
Rhode Island	Ohio	West Virginia	Utah
Connecticut	North Dakota	North Carolina	Colorado
New York	South Dakota	South Carolina	Arizona
Pennsylvania	Nebraska	Georgia	New Mexico
New Jersey	Kansas	Florida	Alaska
	Minnesota	Kentucky	Washington
	Iowa	Tennessee	Oregon
	Missouri	Mississippi	California
		Alabama	Hawaii
		Oklahoma	
		Texas	
		Arkansas	
		Louisiana	

Key Insights:

- Chart indicates shopper profiles by region.
- Shopper profile concentrations are consistent across regions.

Course of Action:

- Regionally based marketing may not be necessary based on customer profile

Project #5 - Pig E. Bank

Excel

Assisting with Customer Projects
at a Global Bank



Business Goals and Data

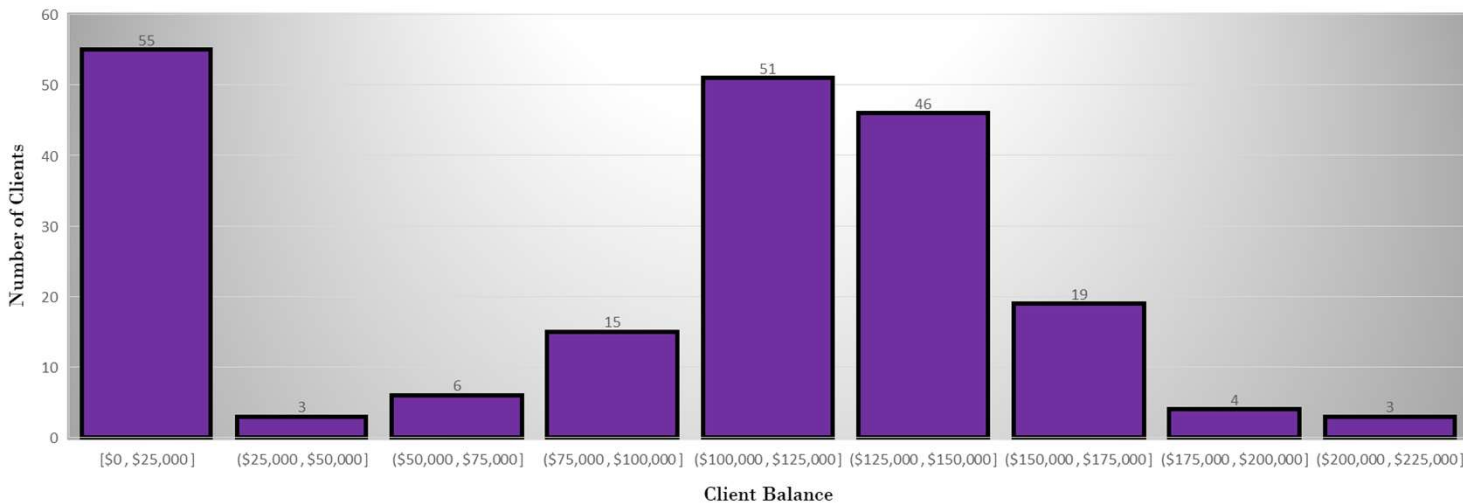
Goals

- Identifying reasons why clients might leave or remain with the bank

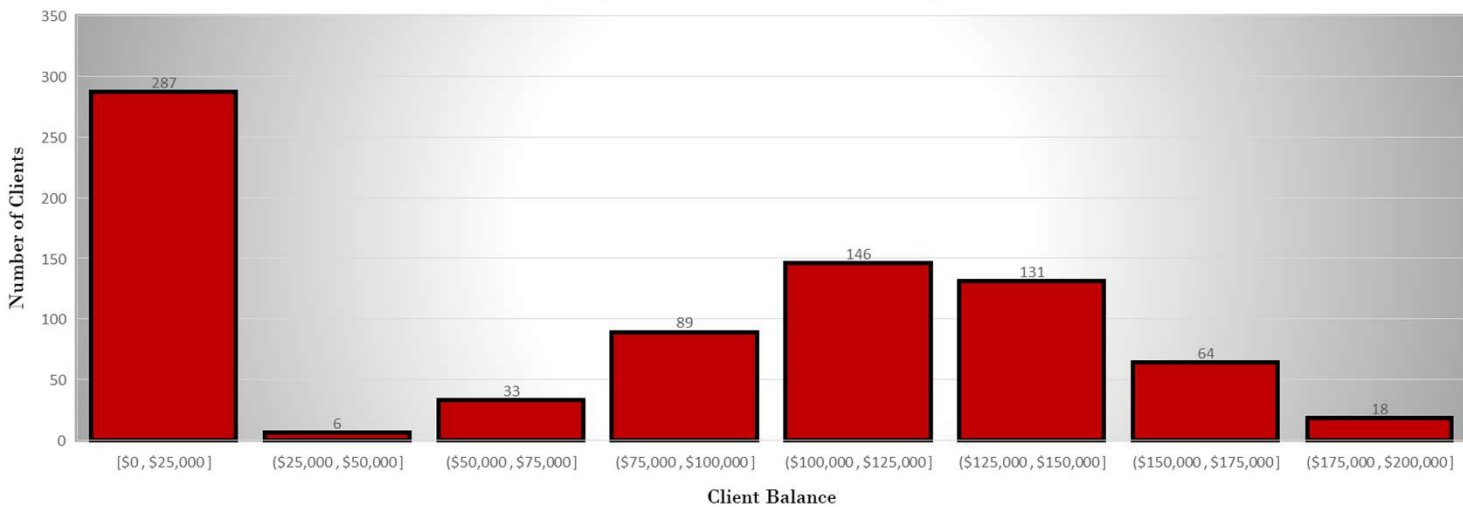
Data

- Data set provided by Pig E. Bank including client information of credit scores, country, gender, age, tenure, salary, and bank usage
- Excel was used to clean, anonymize, analyze, and provide insights into the data through visualizations, pivot tables, and statistical analysis

Frequency of Balances of Clients Who Left



Frequency of Balances of Clients Who Stayed

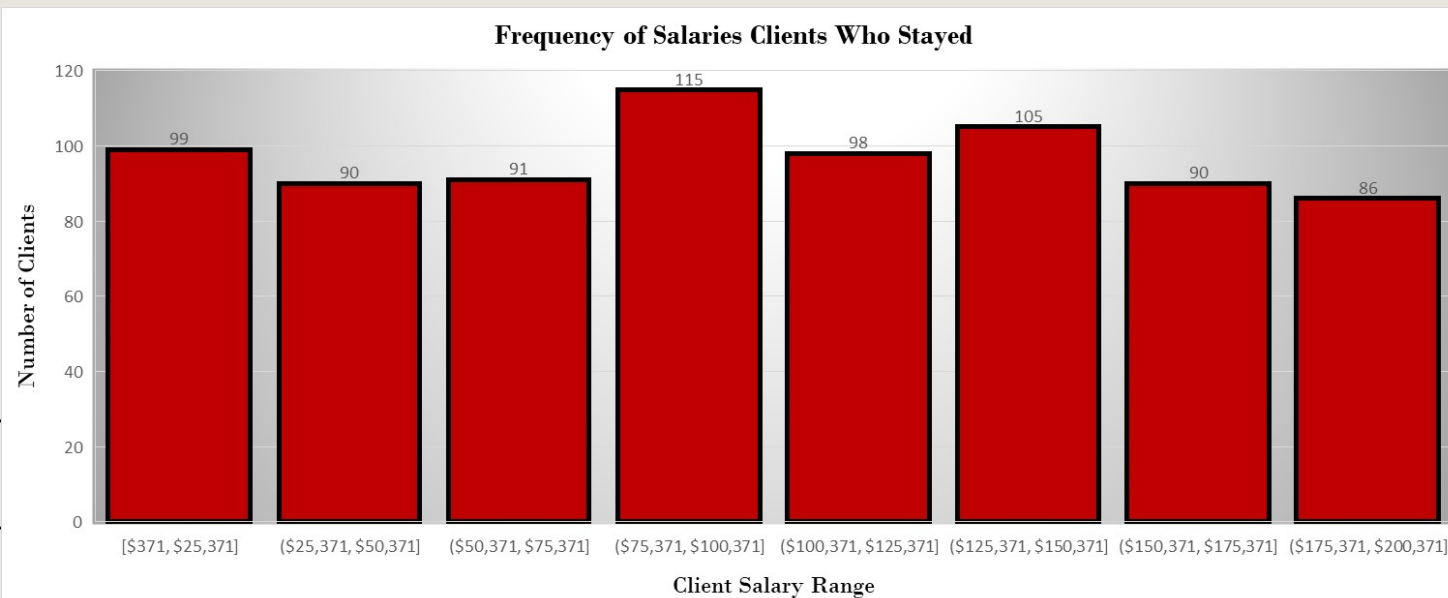
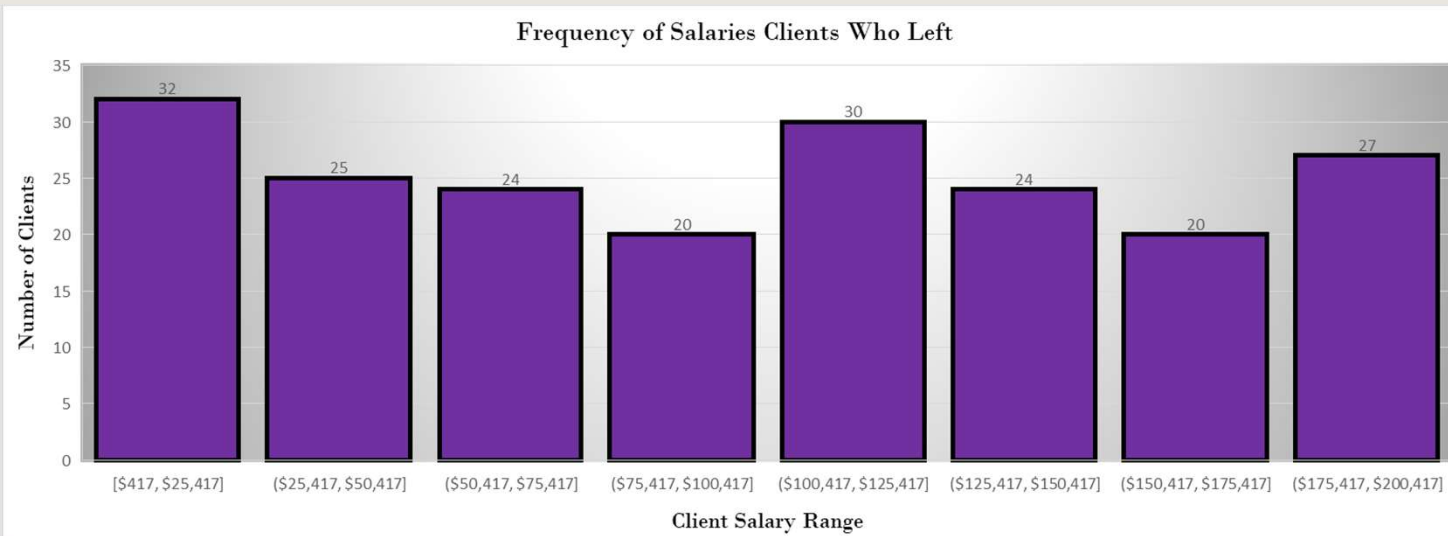


Key Insights:

- These charts identify if client balances were an indication of staying with or leaving the bank.
- The similar shape of the data indicates it was not a factor.

Course of Action:

- Other factors must be evaluated



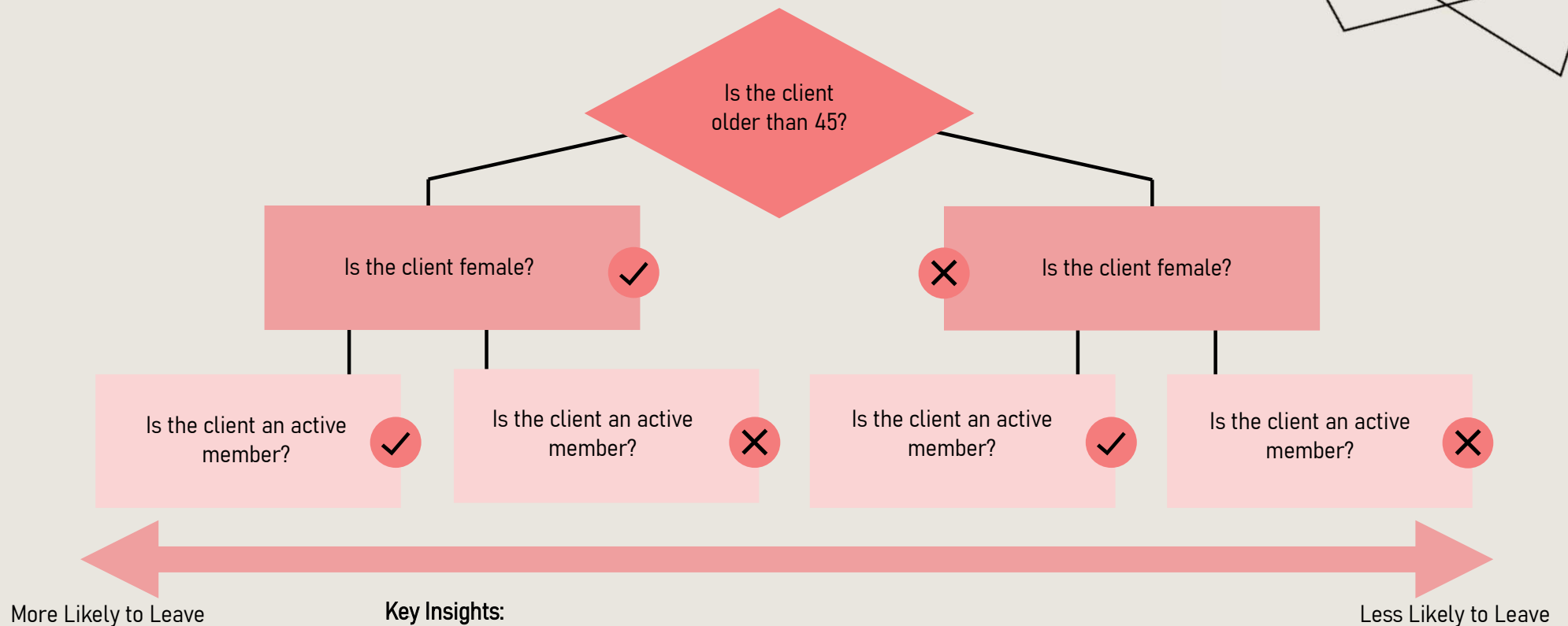
Key Insights:

- These charts identify if client salaries were an indication of staying with or leaving the bank.
- The proportions are different enough to warrant need for further exploration.

Course of Action:

- Widen the search to include more clients and their salaries, if possible, to evaluate if this is a significant factor

Will a client leave Pig E. Bank?



Key Insights:

- Statistical significance of variables was derived using a chart and calculating the P values to figure out what actually correlated to differences between clients who stayed and clients who left the bank.
- Age = statistically significant, P value = 0.0001
- Gender = statistically significant, P value = 0.041
- Active User = statistically significant, P value = 0.048

Project #6 – USA Real Estate Market

Python and Excel

Assessing Factors that Affect
USA Housing Prices

[Link to GitHub Repository](#)

[Link to Tableau Story](#)



Business Goals and Data

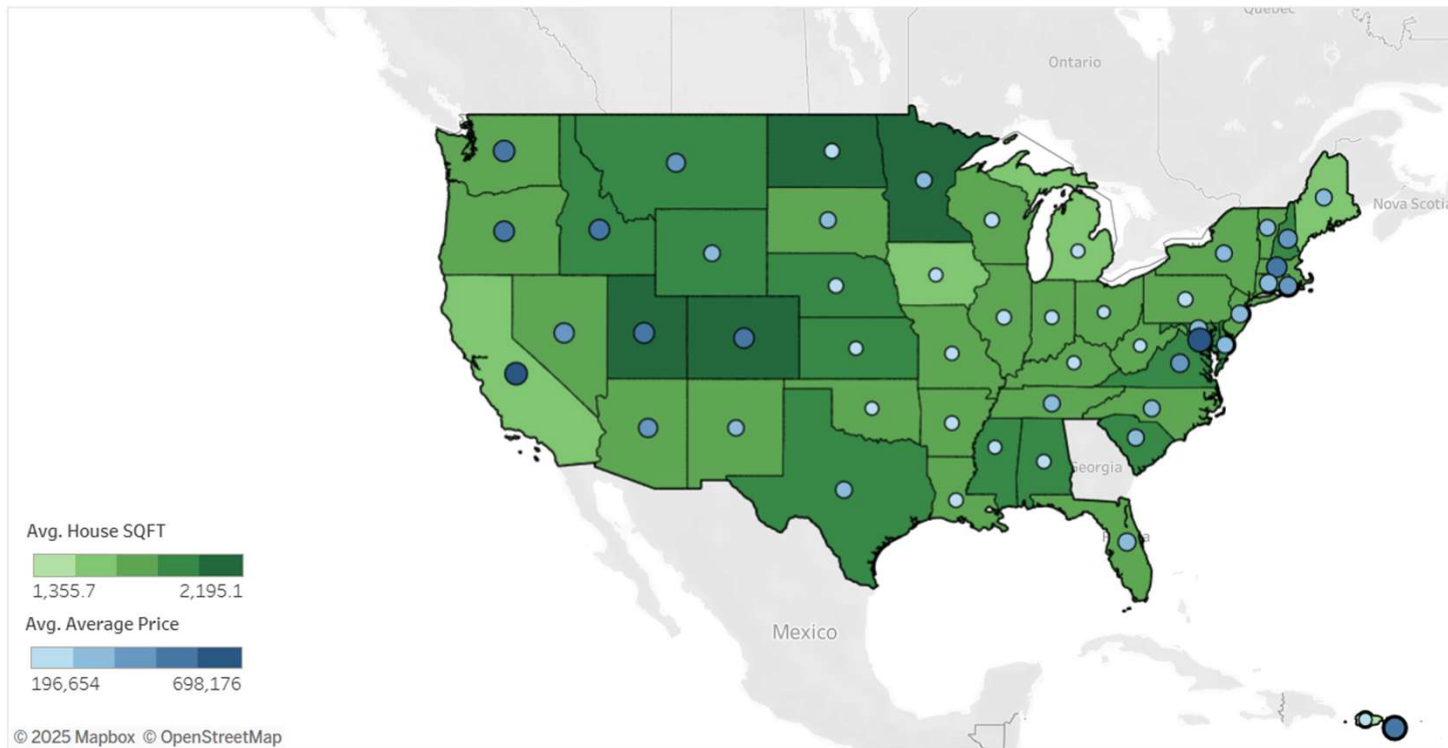
Goals

- Identifying reasons why certain houses are more expensive than others
- Provide insights for sellers and buyers into the factors affecting their housing costs

Data

- 2,226,382 data entries from across the USA and its territories including house price, # of beds, # of baths, acreage, square footage, city, state, and sell date
- Python was used to conduct cleaning, wrangling, exploratory analysis through visualizations, geospatial analysis, regression analysis, cluster analysis, time-series analysis, and analysis narrative and final results presented in a Tableau dashboard

Map of House Square Footage and Average Price



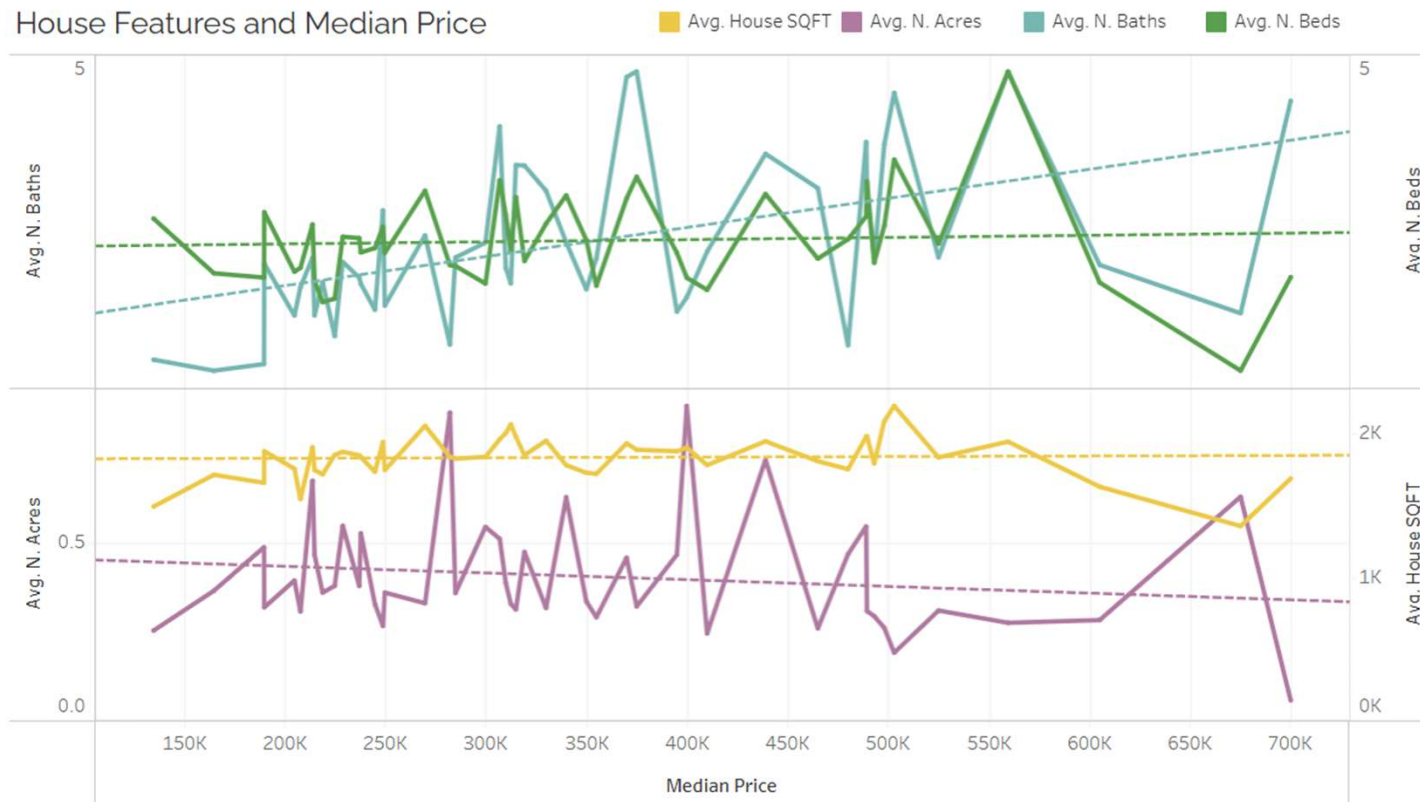
Location, location, location. We all know the age-old adage that location is the biggest factor driving real estate prices.

What if there is more to it?

This analysis attempts to find out if house features can be an equally important factor in predicting housing prices. Or is there another factor entirely worth looking into?

Using data from realtor.com (a real estate listing website that is the second most visited of its kind in the United States as of 2024, with over 100 million monthly active users) this analysis will dive into factors that could potentially drive housing prices.

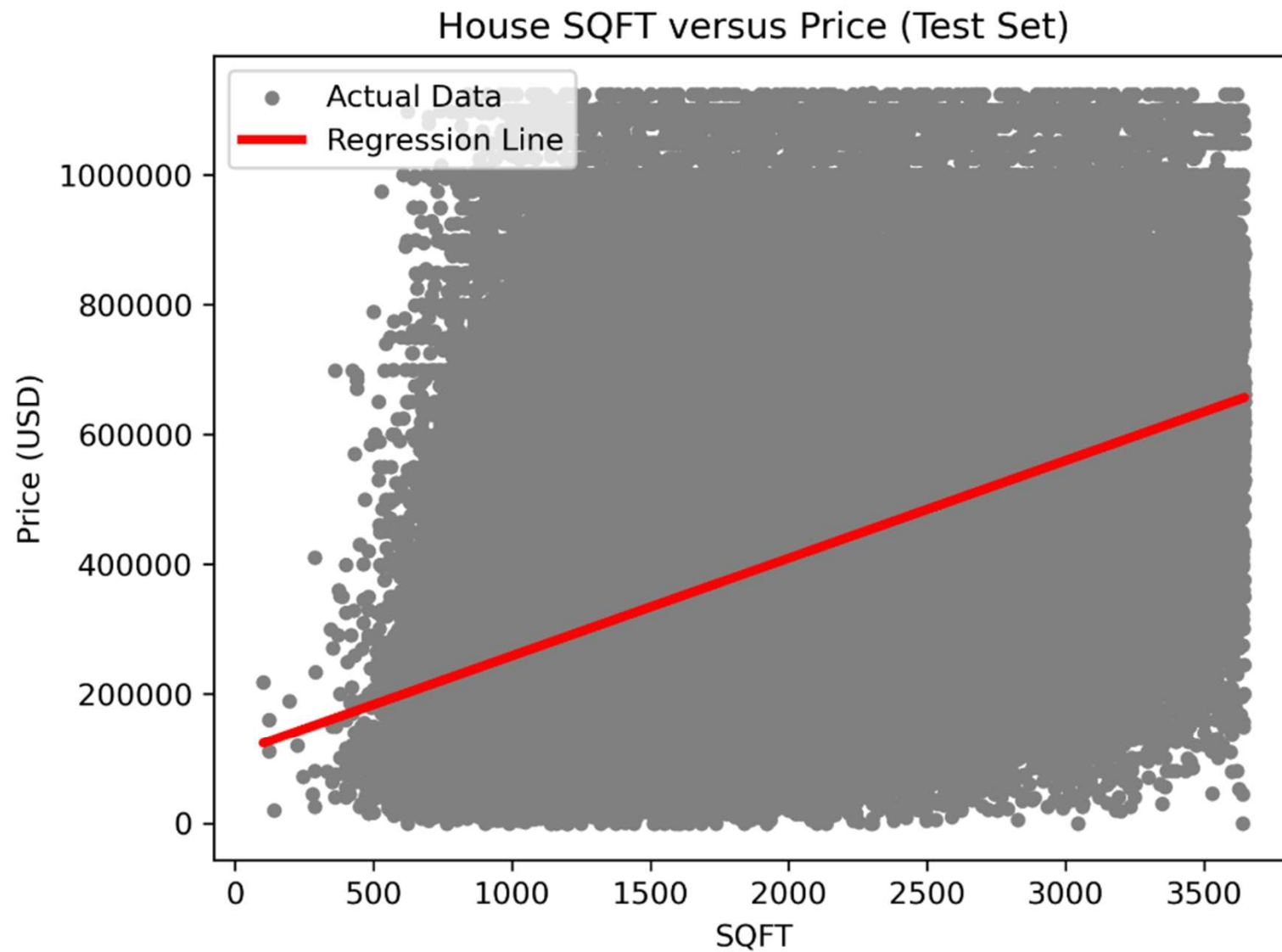
House Features and Median Price



To start the exploration, we started by looking for any **linear relationship** between variables.

As this chart shows **there is not any linear dependence** between number of baths or bedrooms or size of the house or land. You can see that the median cost of the house is not significantly affected by any of these factors.

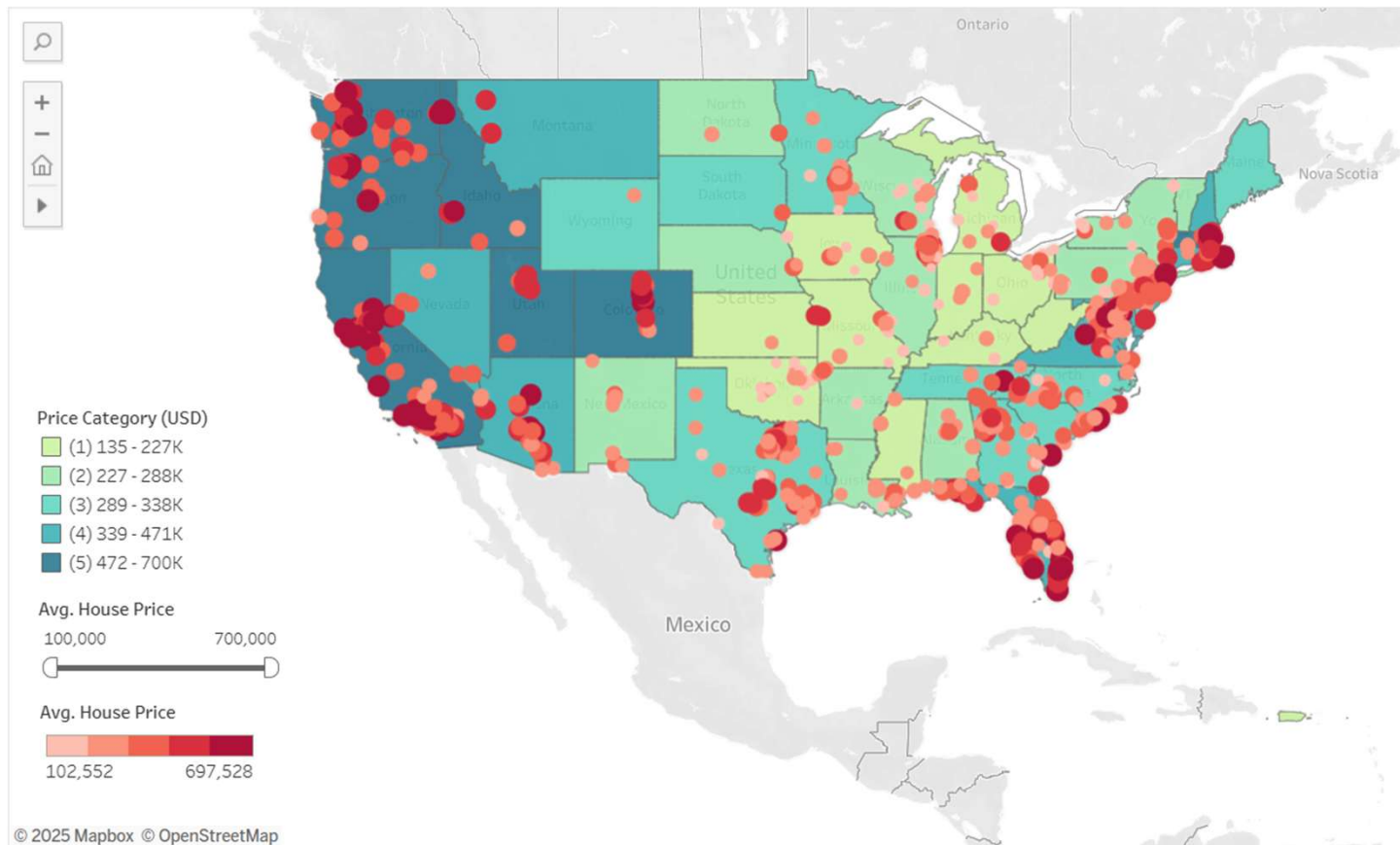
This led us to conclude there must be another factor at work here.



To confirm the hypothesis that house features (house square footage being the example here) did not significantly impact house price a linear regression was conducted.

The results indicated that only **19%** of house cost could be attributed to square footage. Therefore a linear regression was not enough to explain house costs.

State and City Average House Prices



The age-old adage appears to still be accurate. The data reveals that **location is in fact the largest factor** when navigating house prices.

As you can see by the map; large, dense cities and popular places to live (coasts, islands, and metropolitan areas) are still the most expensive parts of the USA real estate market.

Contact Information

- Phone: +1 (719)-351-6330
- E-Mail: lisa.greer@comcast.net
- Linked In: TBD
- GitHub: <https://github.com/lgreer929>
- Website: TBD

