

Margo Shogren

Data Analyst

About Me

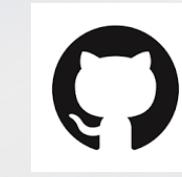
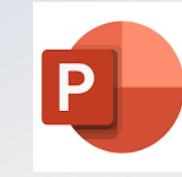
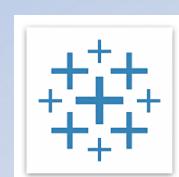
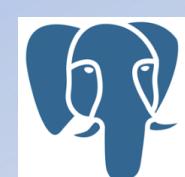
I am currently changing my career focus from accounting back to data analytics, which I have some background in conjunction with my accounting career. I have come to the realization that I was always the most passionate about projects in my career that involved problem solving with data. In my previous employment there was often a need for reporting which involved extracting and reporting on data from the accounting system and I quickly found I had a knack for this using Cognos, SQL, Excel and other tools.

I have just increased my skillset by completing the CareerFoundry online data analytics course with Python, Tableau, presentation skills and data analysis skills.

If you have data and you need answers, contact me!

Skills used

Excel, Postgres SQL, DB Visualizer, Tableau, Python, Jupyter Notebook, Github, Powerpoint



Project links & Analysis preparation

- 1) [**GameCo**](#) - Video game company
Preparation: Used Excel pivot tables for grouping, summarizing, descriptive analysis to understand data.
- 2) [**Preparing for the Influenza season**](#)- Medical staffing agency
Preparation: Excel Data cleaning, integration, hypothesis testing, forecasting the flu season. (Deliverable in Tableau)
- 3) [**Rockbuster Stealth, LLC**](#) - Video rental company
Preparation: SQL to join tables, filter data and clean data
- 4) [**Instacart**](#) -Online grocery delivery
Preparation: Used Python to wrangle, merge, group, and aggregate data
- 5) [**Superstore**](#) - Giant office supply retailer
Preparation: Python for regression analysis and Tableau for visualizations

GameCo

Video game descriptive analysis
to understand the current game market



GameCo

Goal

- Descriptive analysis of GameCo data to determine if sales have remained constant for each major geographic region and which games to promote.

Data

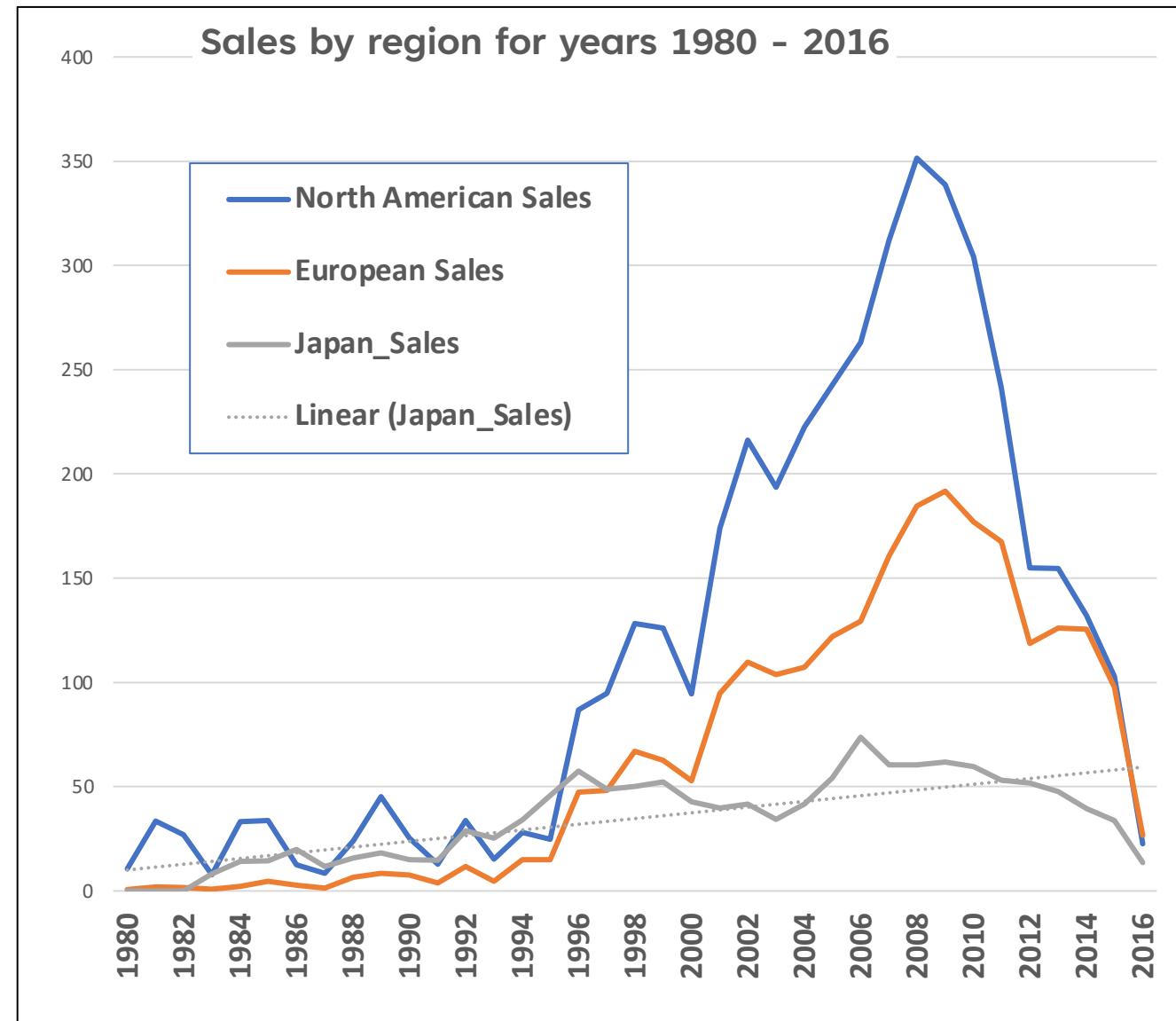
- Game Co. in house data of the historical sales of different genres, platforms, publishers across geographic regions.
- [Data Source: VGChartz](#)

Analysis

- What are the more popular games?
- What are the publishers with highest revenue?
- Have the games decreased or increased in popularity over time?
- How have their sales figures varied between geographic regions over time?

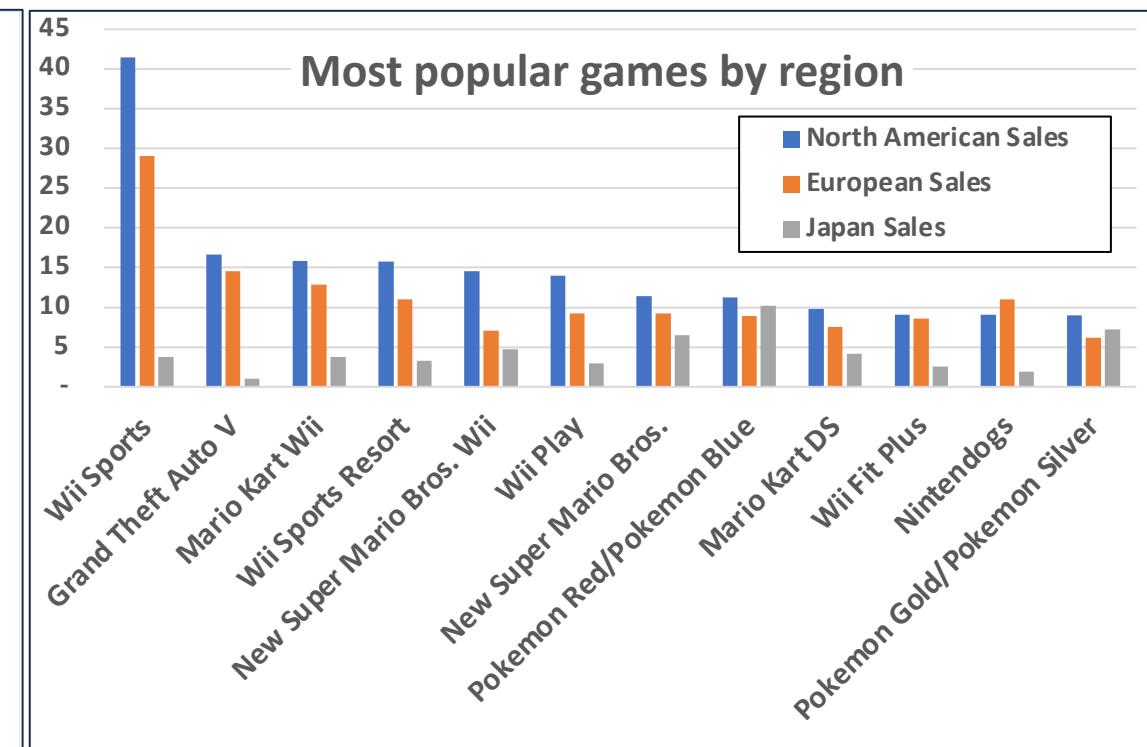
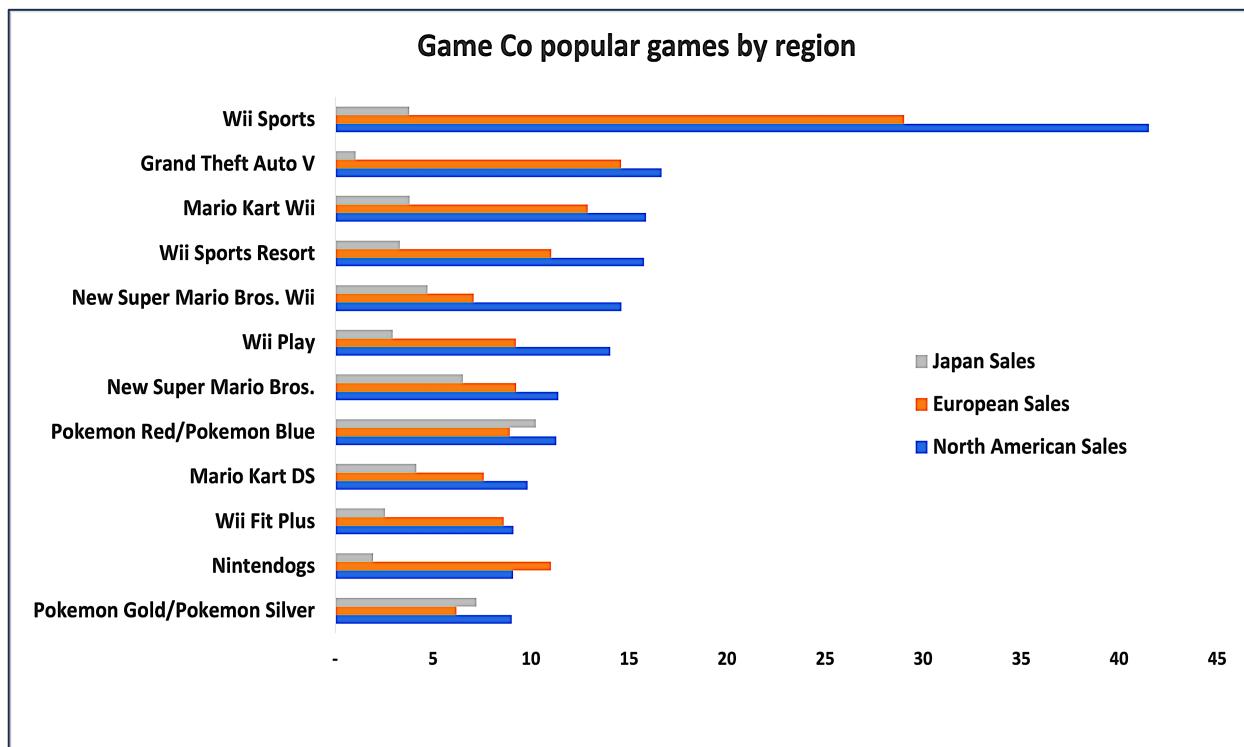
GameCo

GameCo video game sales for Europe follows a similar trend to North American through the years. Japan sales are much lower.



GameCo

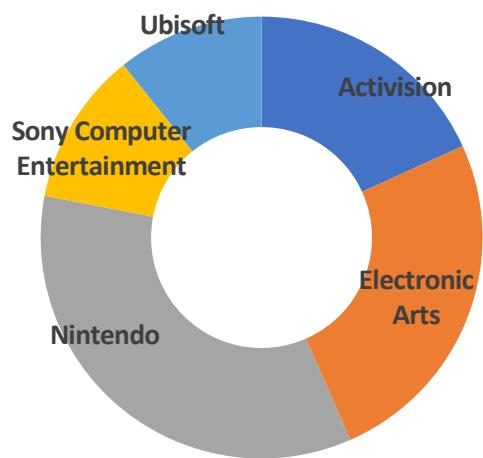
Popular video games by region shows that the most popular games are Wii Sports and Grand theft Auto V



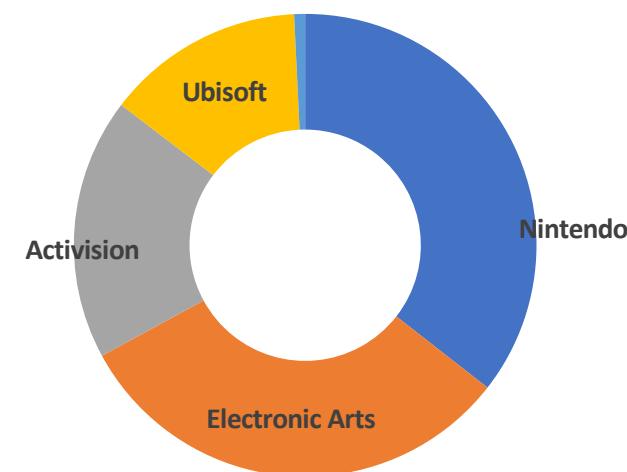
GameCo

These are the top 5 publishers for each of the 3 geographic regions: Nintendo has the most sales for all regions

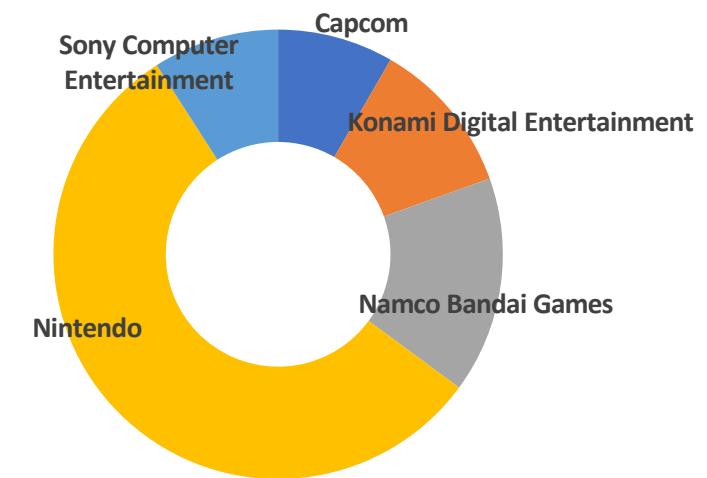
North America Top 5 publishers



European Top 5 Publishers



Japanese Sales Top 5 publishers



Summary and Recommendations

Summary

North American and Europe sales increased in 2008, then decreased. Japanese sales have been more constant, but lower. Since sales have been dropping overtime, we can expect that end to continue.

Recommendations

- GameCo should research the lower sales trend over time for North American sales
- GameCo could promote more games for younger audiences like the Wii games popular in 2008. European customers prefer the FIFA games.
- GameCo should invest in more European games.
- For Japanese customers, GameCo should promote different game types such as Pokemon.



Preparing for Influenza season

Analysis for a Temporary medical staffing agency to plan for flu season



Preparing for Influenza season

Goal

- Help a medical staffing agency proactively plan for the proper number of temporary workers to send to clinics & hospitals during highest demand of influenza.

Data

- [Data Source: Population Data](#)
Owned by US Census Bureau
This is census data of populations by geographic regions and age (years 2009-2017) from surveys.
- [Data Source: Influenza Deaths](#)
owned by CDC
Data is based on U.S. death certificates signifying underlying cause of death.

Analysis

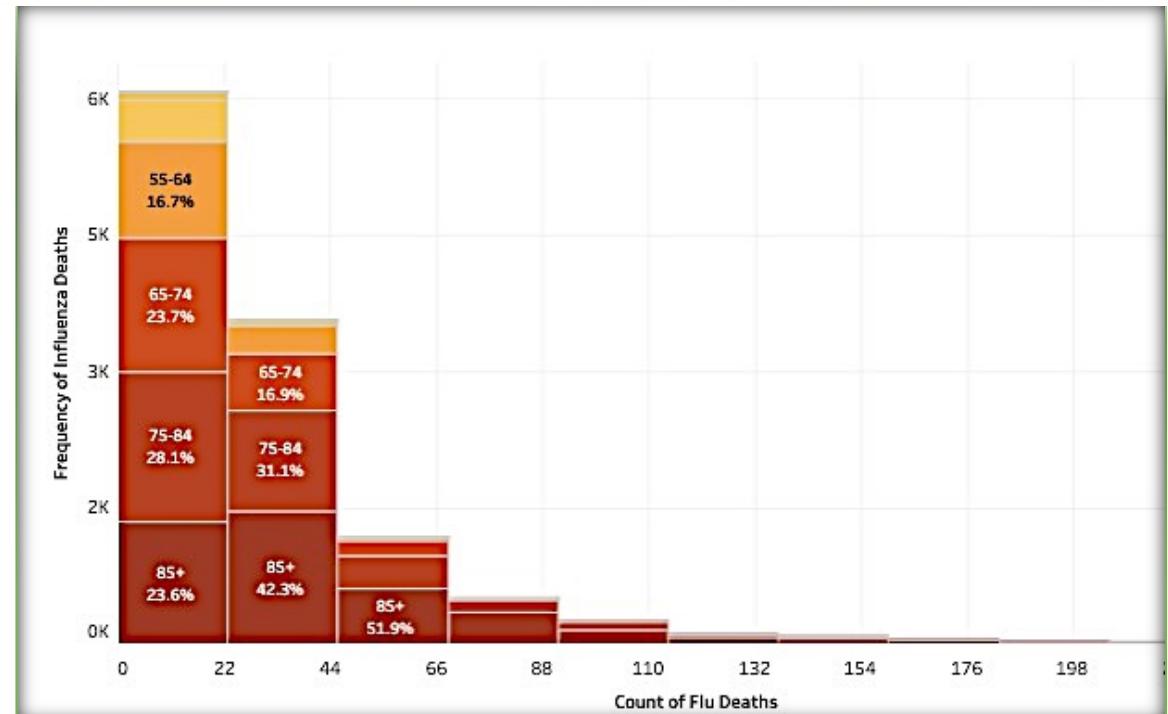
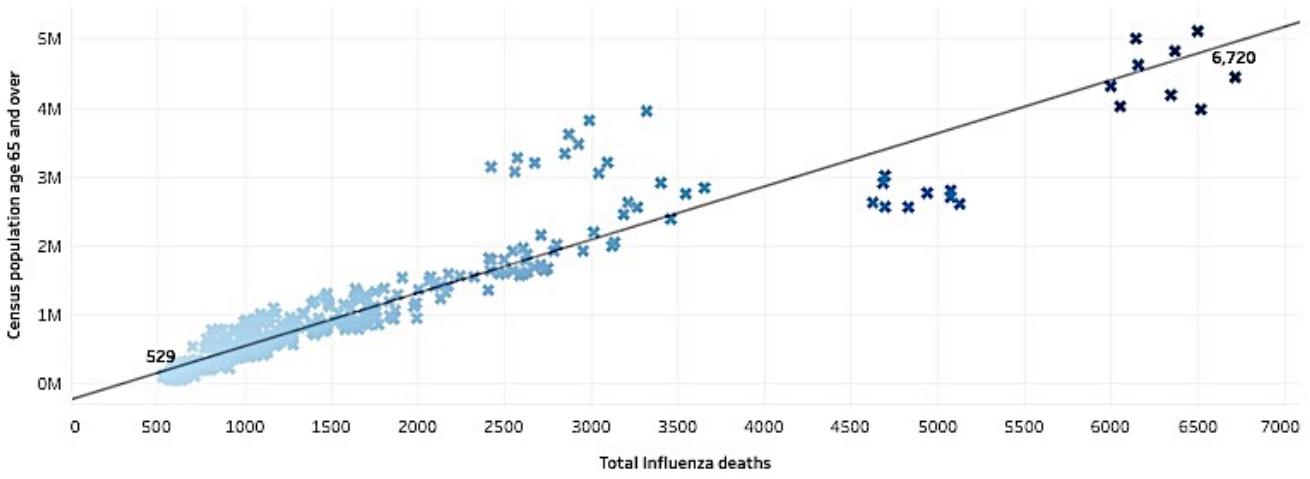
- Created a data profile for both data sources.
- Checked data integrity
- Examined the data quality
- Mapped and integrated the data with v-lookups and pivot tables
- Conducted statistical hypothesis testing

Preparing for Influenza Season Analysis

This scatterplot was used to test the hypothesis that the highest population for influenza deaths are for ages 65 and over.

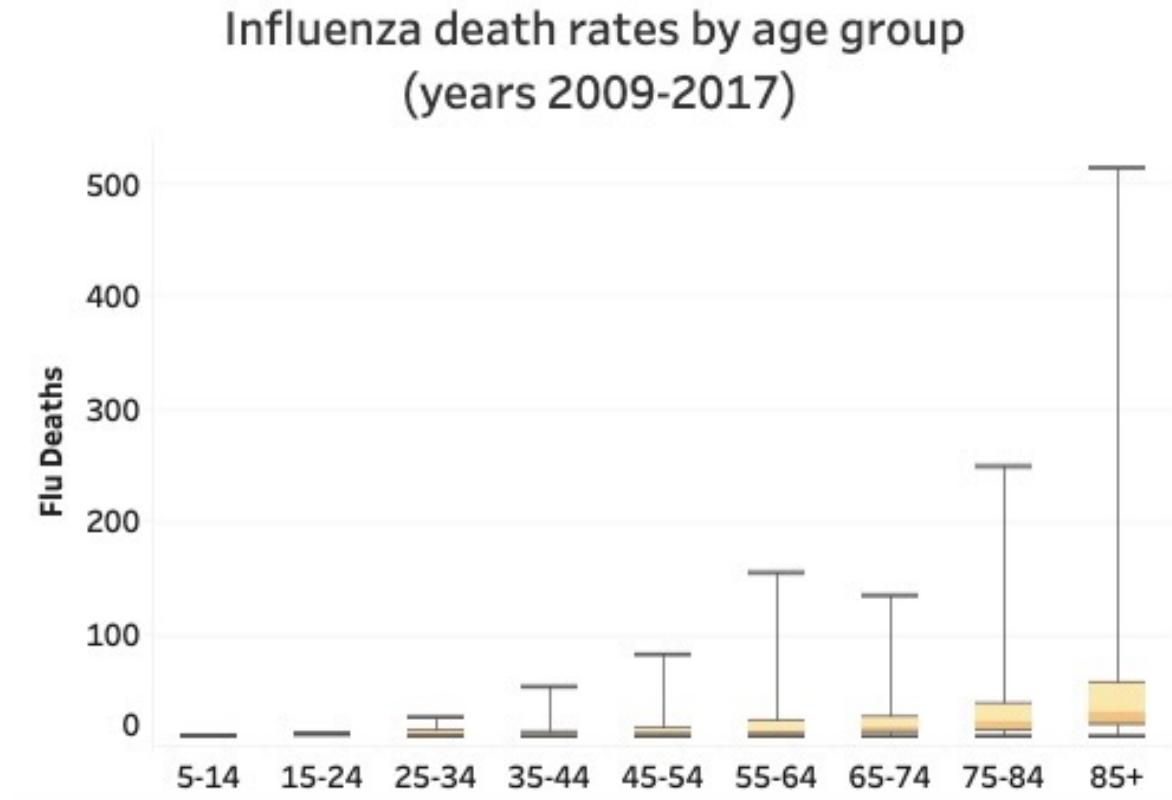
The correlation coefficient is .94355 showing a very strong relationship.

Influenza deaths frequency for different age groups



Preparing for Influenza Season

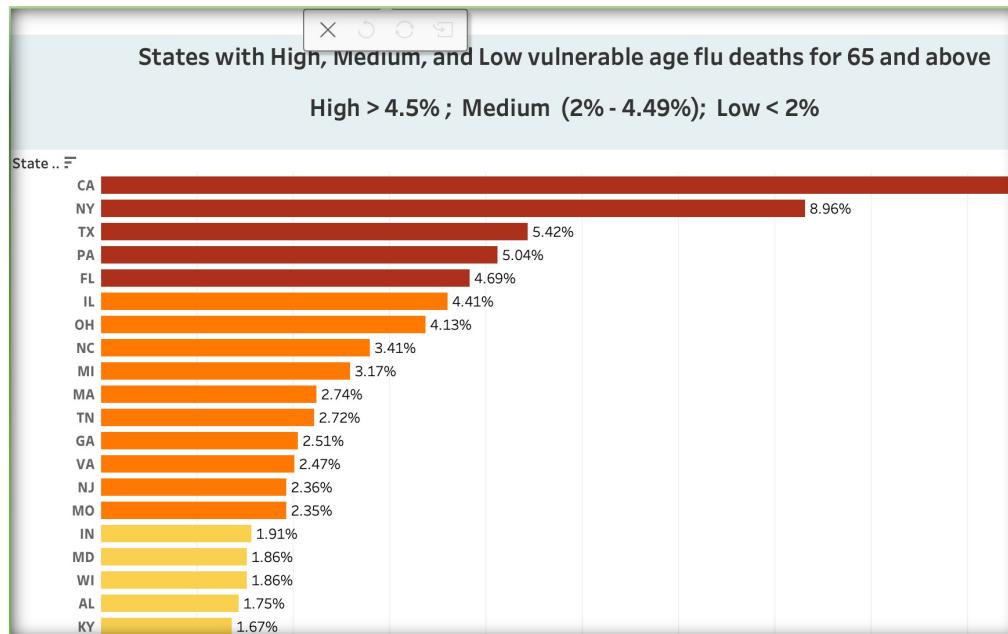
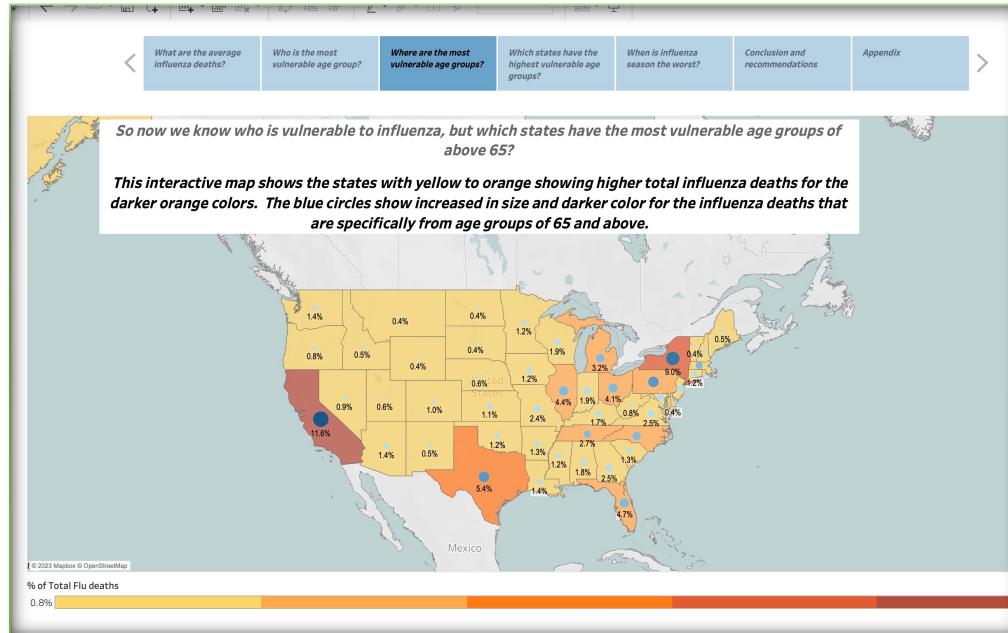
- This box plot was part of a statistical analysis to analyze the possibility of other ages having a high risk of influenza deaths and supports the 65 and over age group as the most at risk.*



Preparing for Influenza Season

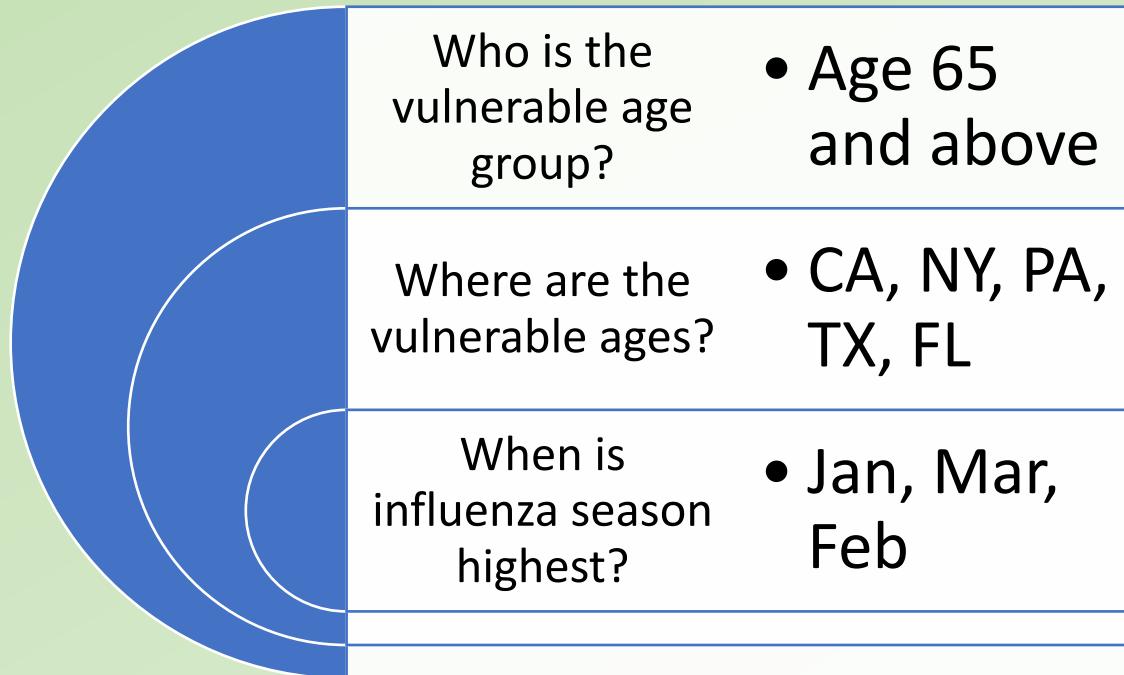
An overview of influenza deaths for the vulnerable age groups

- *Influenza deaths analysis drilled down by state*



Recommendation

Prioritize the medical staff to these states during these months



[Link to Tableau Report](#)

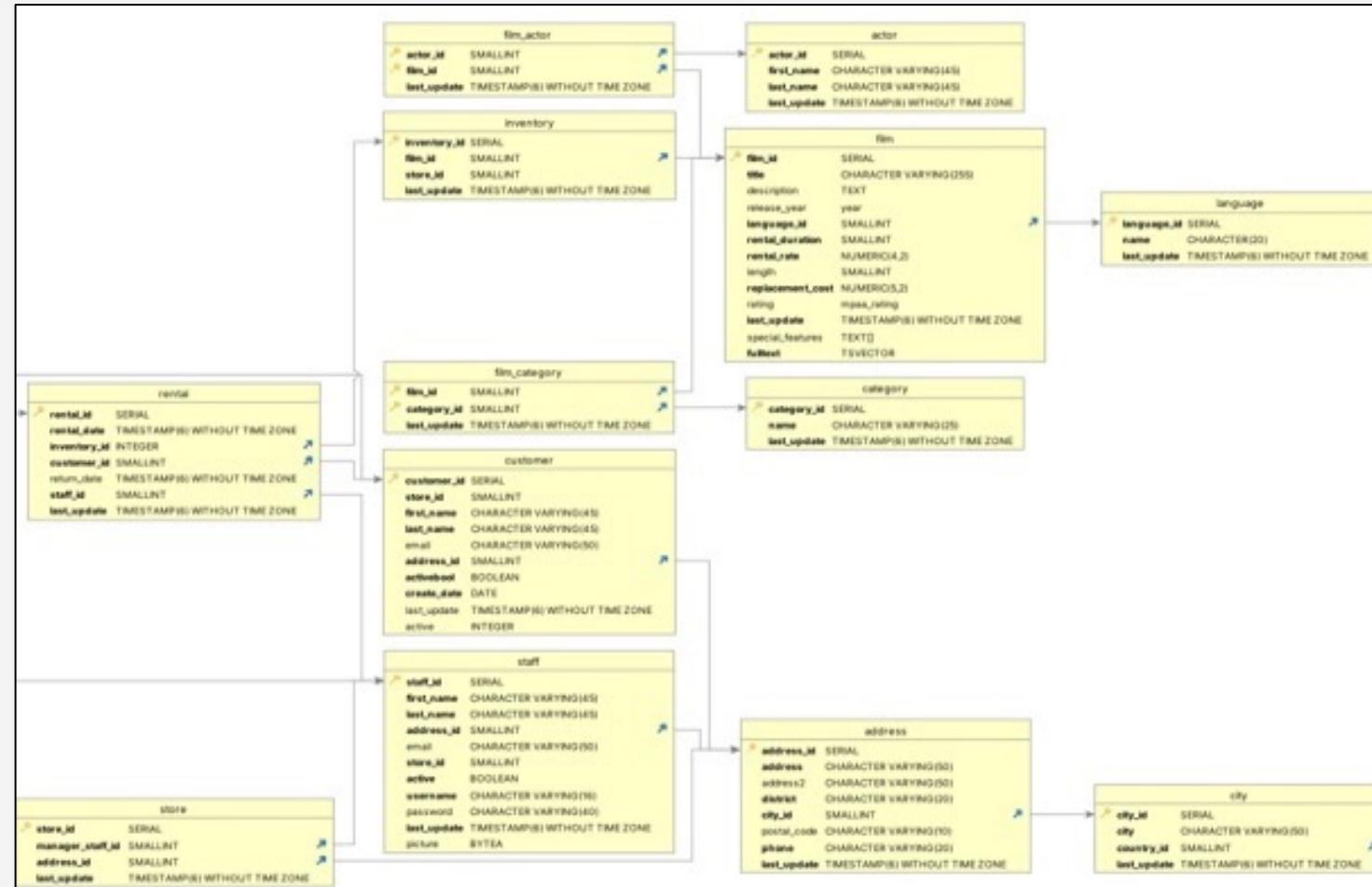
Rockbuster Stealth, LLC

A movie video rental company

Data Analysis to access the RDBMS and provide insight for launching a streaming service.



Rockbuster ERD



Rockbuster Stealth, Inc

Goal

Rockbuster, a video rental company needed to understand their customer database to stay competitive with online streaming options

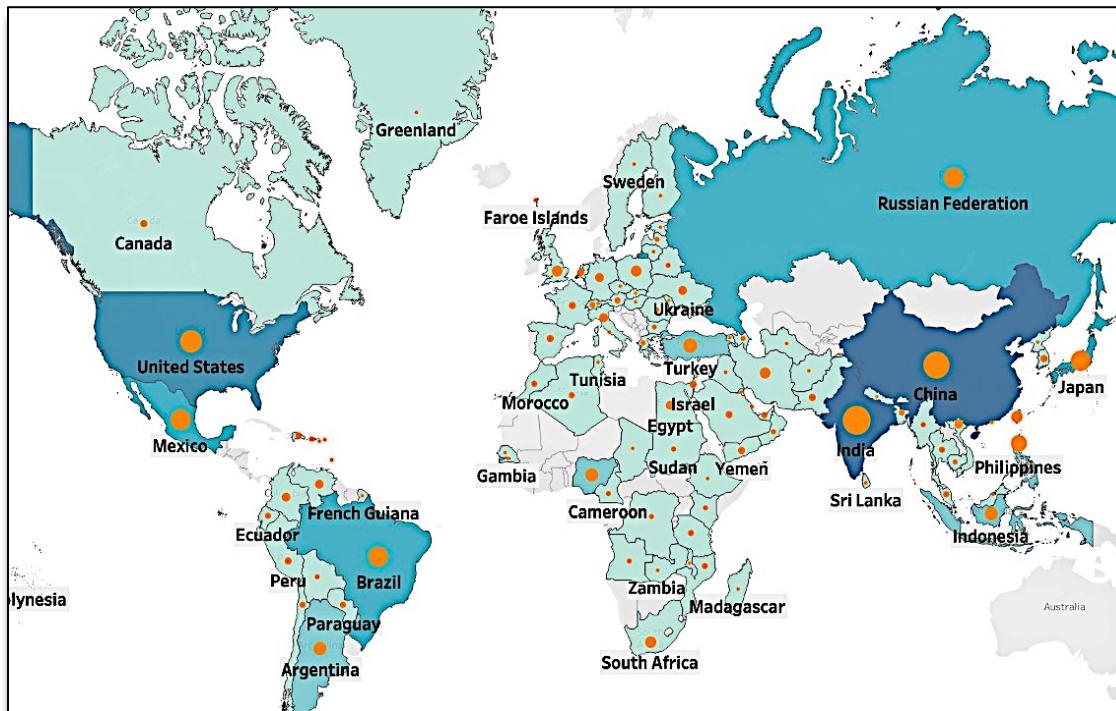
Data

- The Rockbuster database consists of sales, products, customers, pricing tables.
- [Data Source: Rockbuster data set](#)

Analysis

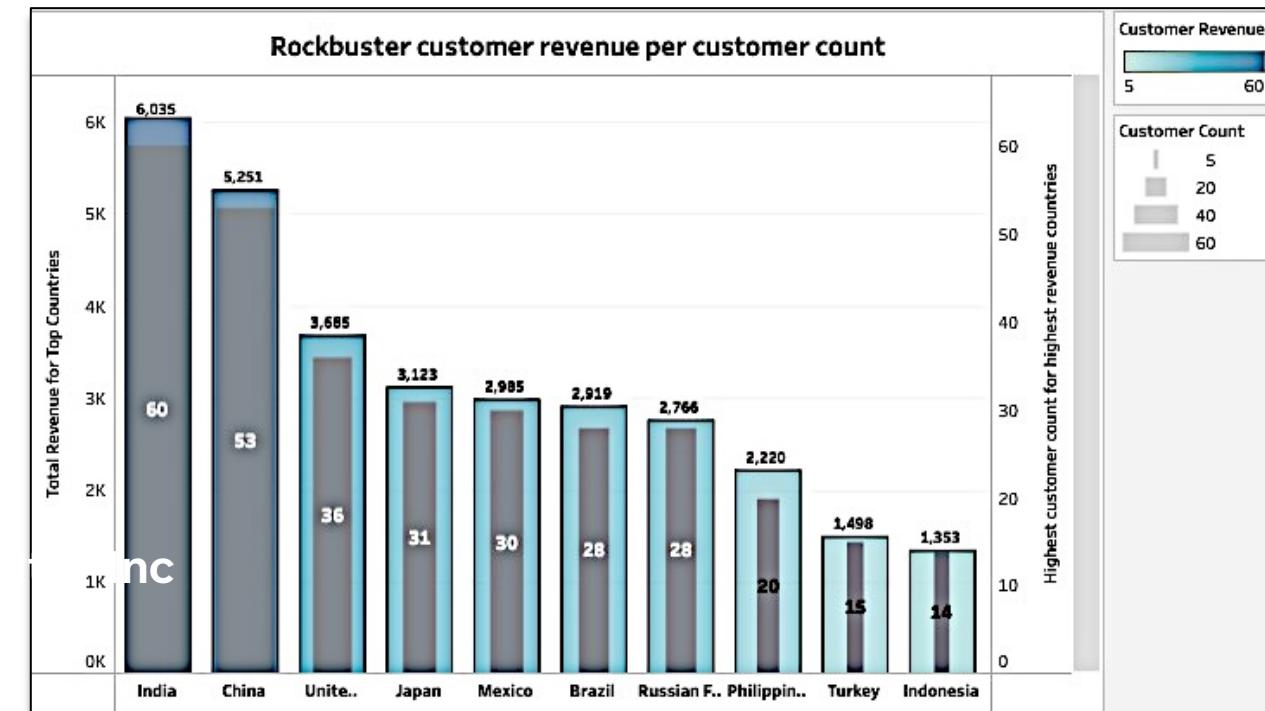
- Where are Rockbuster customers?
- What movie titles are Rockbuster customers renting?
- How does Rockbuster rental pricing affect the revenue?

Rockbuster Stealth, LLC



These are the countries where Rockbuster customers provide the highest revenue.

*Tools used:
Postgres SQL queries, Tableau*



India and China had the customers with highest revenue AND highest customer count.

*Tools used:
Postgres SQL queries, Tableau*

Rockbuster Stealth, LLC

Most and least rented movies

tool used:

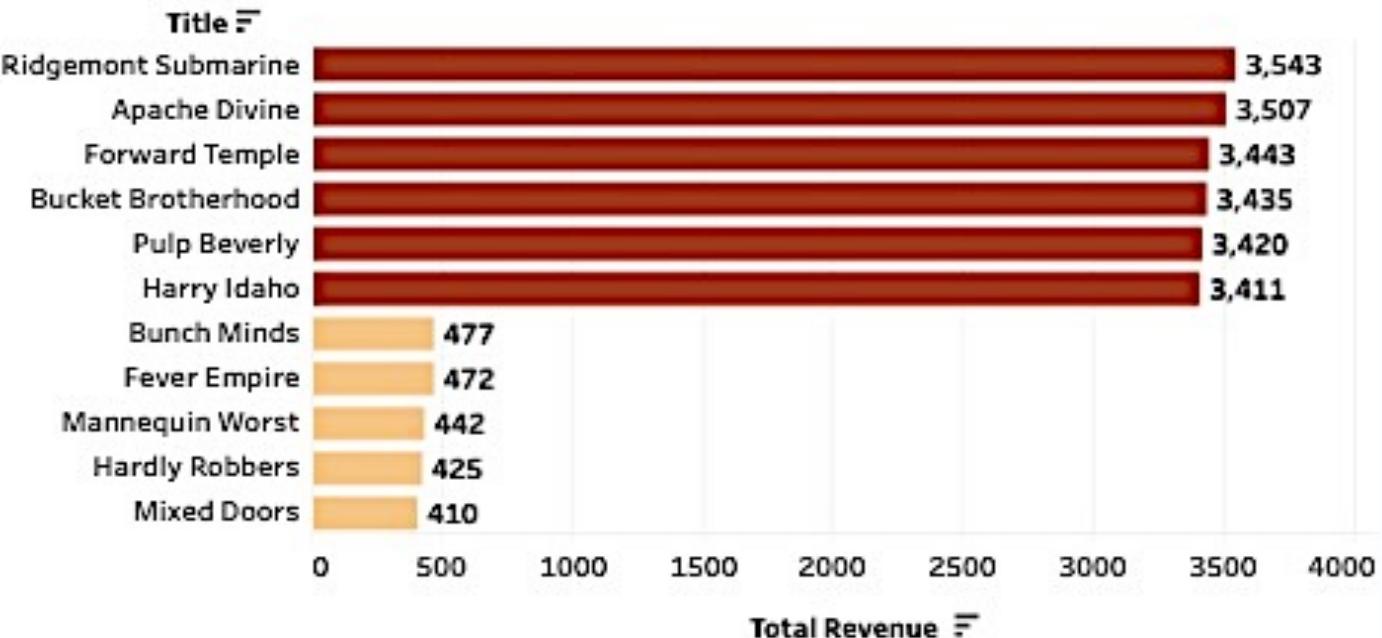
Postgres SQL and Tableau

*Most frequently rented videos
with the highest revenue per
film rating.*

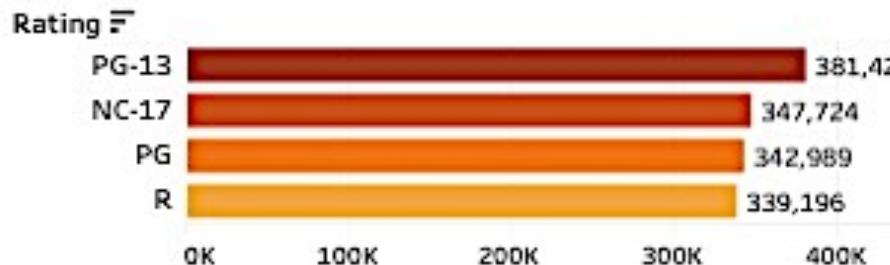
tool used:

Postgres SQL and Tableau

Most rented and least rented films with Rockbuster

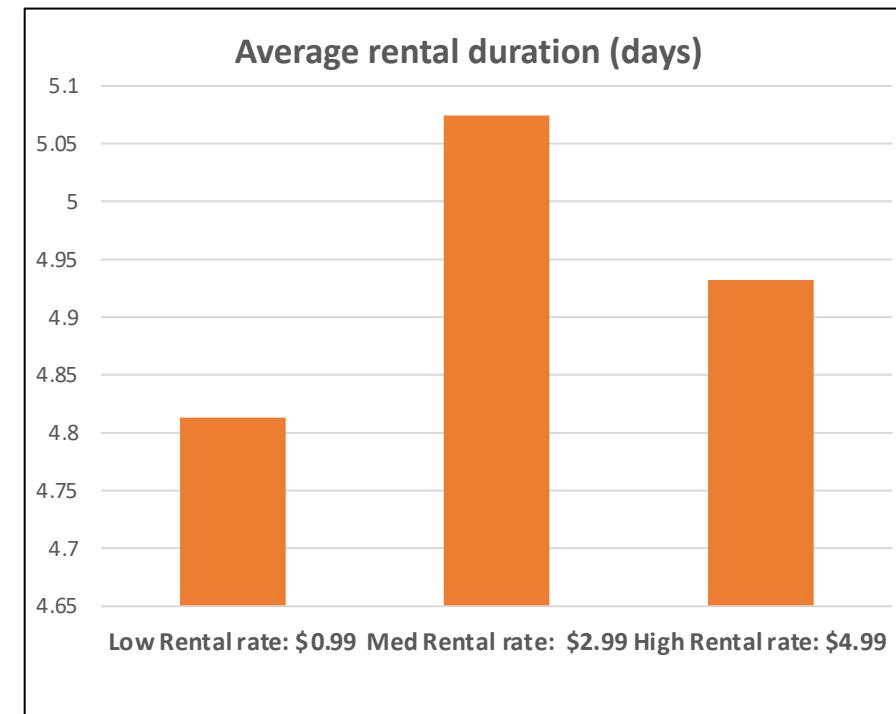
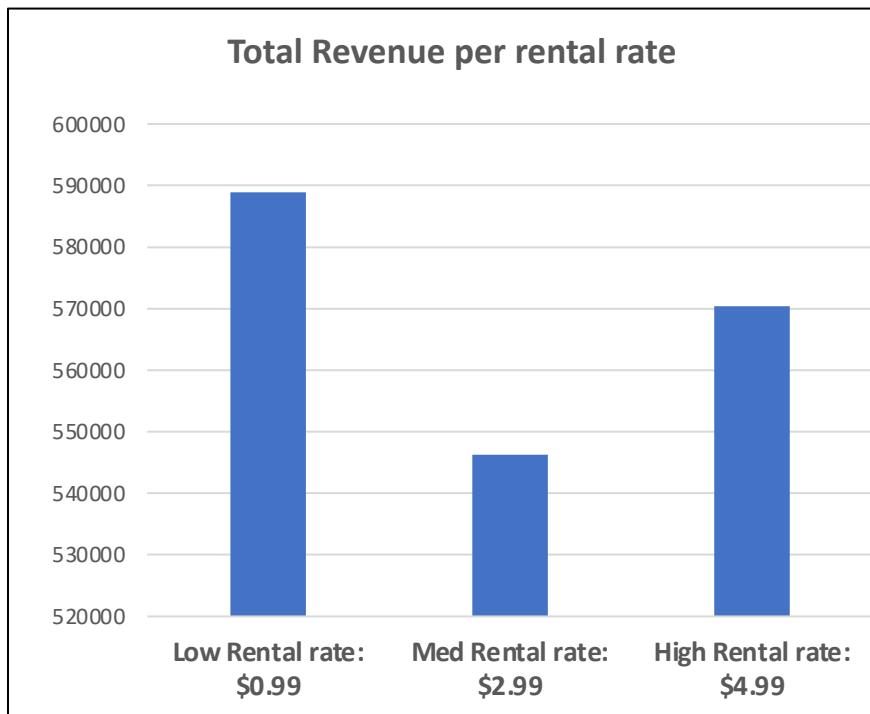


Revenue by film ratings



Rockbuster Stealth, LLC

Lowest rental prices generate the highest revenue and mid-priced rentals are rented the longest duration.



Rockbuster Stealth, LLC

Recommendations

- Locations

Rockbuster currently receives the highest revenue from the countries of India and China. Rockbuster should continue to concentrate marketing in those countries.

Rockbuster can also target market resources in remote areas where customers have fewer streaming opportunities.

- Products

Rockbuster should offer movie titles that streaming services do not offer. Review the popular titles that are rented and offer a wide variety of categories and languages.

- Pricing

Rockbuster could experiment with changing pricing options or extra day rental charges and assess the impact on revenue.



Sales Pattern Analysis

Instacart wants to understand who the current customers are in their database and what they purchase to make marketing decisions and target new customers.



Sales Pattern Analysis

Goal

Understand trends and patterns of the customer data set and products ordered using Python.

Data

[Data Source: Customer data, Orders and products](#)

“The Instacart Online Grocery Shopping dataset 2017”

[Data Source: Data Dictionary](#)

Analysis

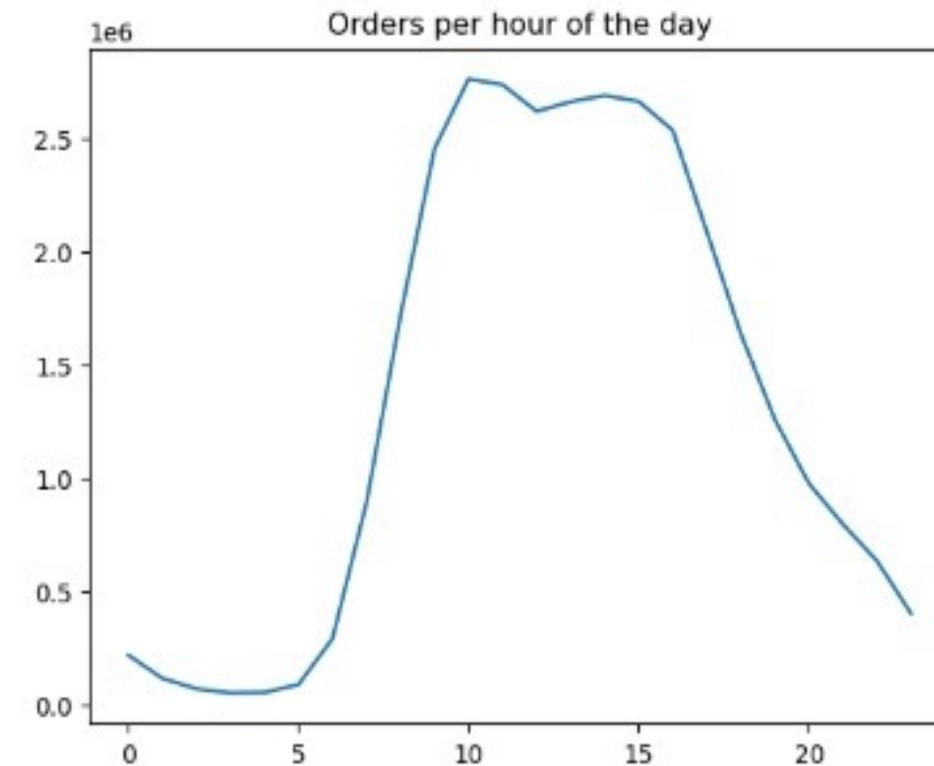
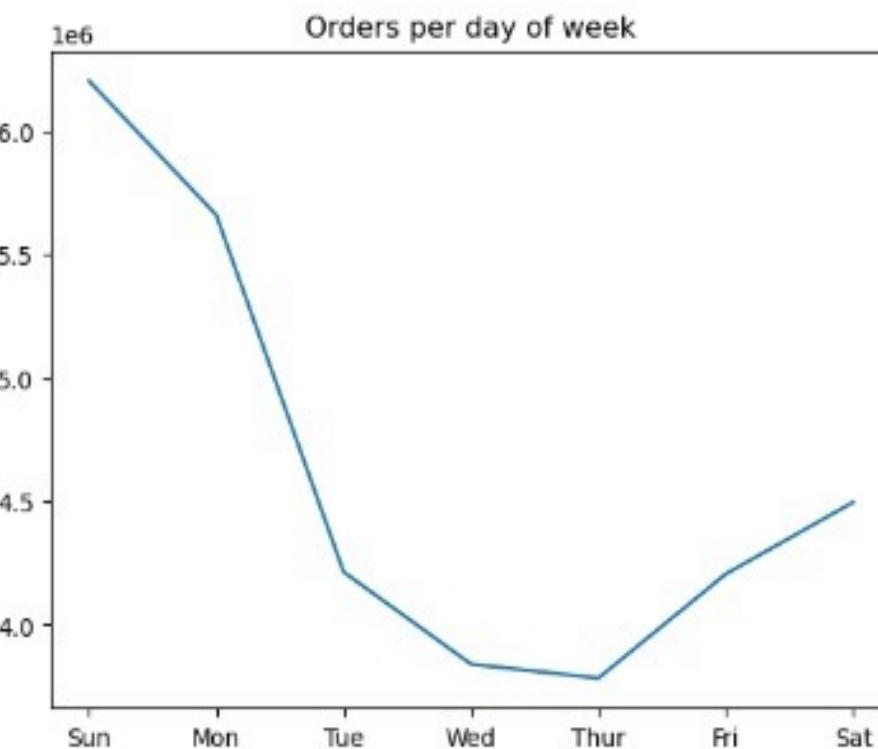
- Answered stakeholders' key questions using Python (Jupyter notebook): pandas, numpy, os, matplotlib, scipy, seaborn. Merged, grouped and aggregating data frames.

- Created Excel report deliverable with the population flow, consistency, wrangling, derivation steps and python visualizations and recommendations



Sales Pattern Analysis

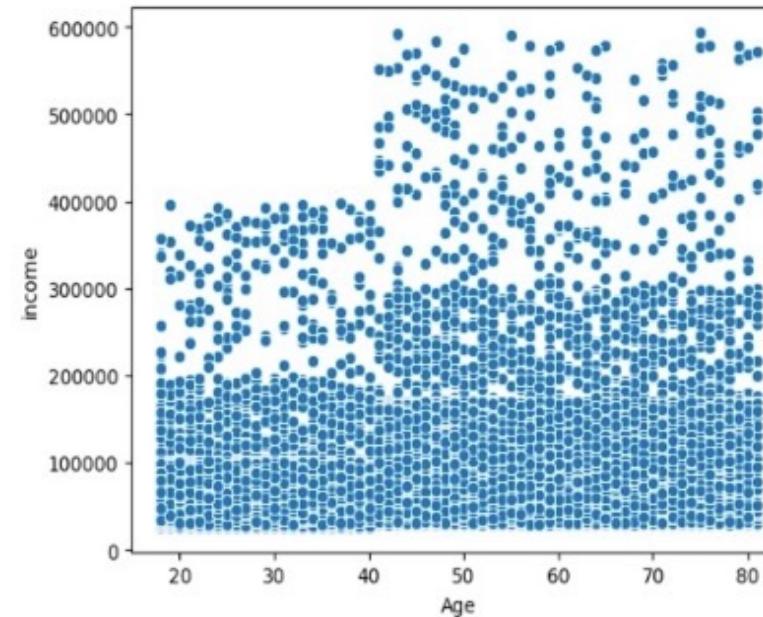
These charts created in Python find the busiest day of week for customer orders is Sunday and the busiest hour is 10am. The least busy day is Thursday and least busy hour is 4am to 5am.



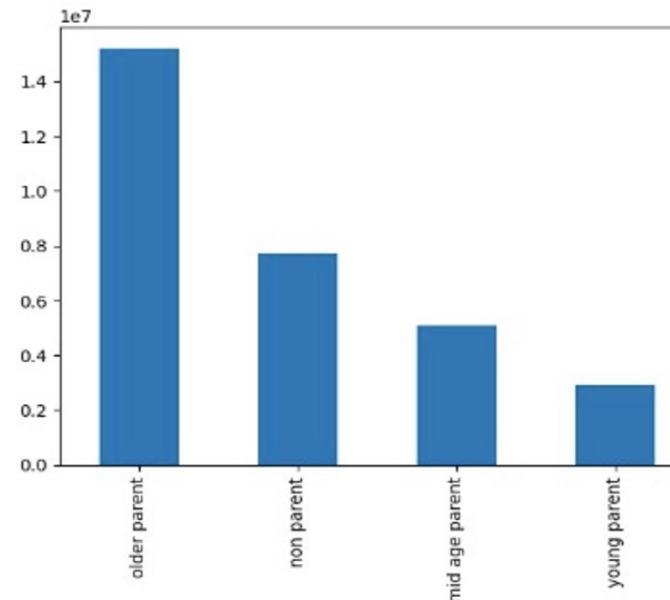
Instacart Sales Pattern Visuals created in Python

The scatterplot shows most of the customers are age over 40 and income under \$200k.

Scatterplot to explore whether there's a connection between age and spending power (income).

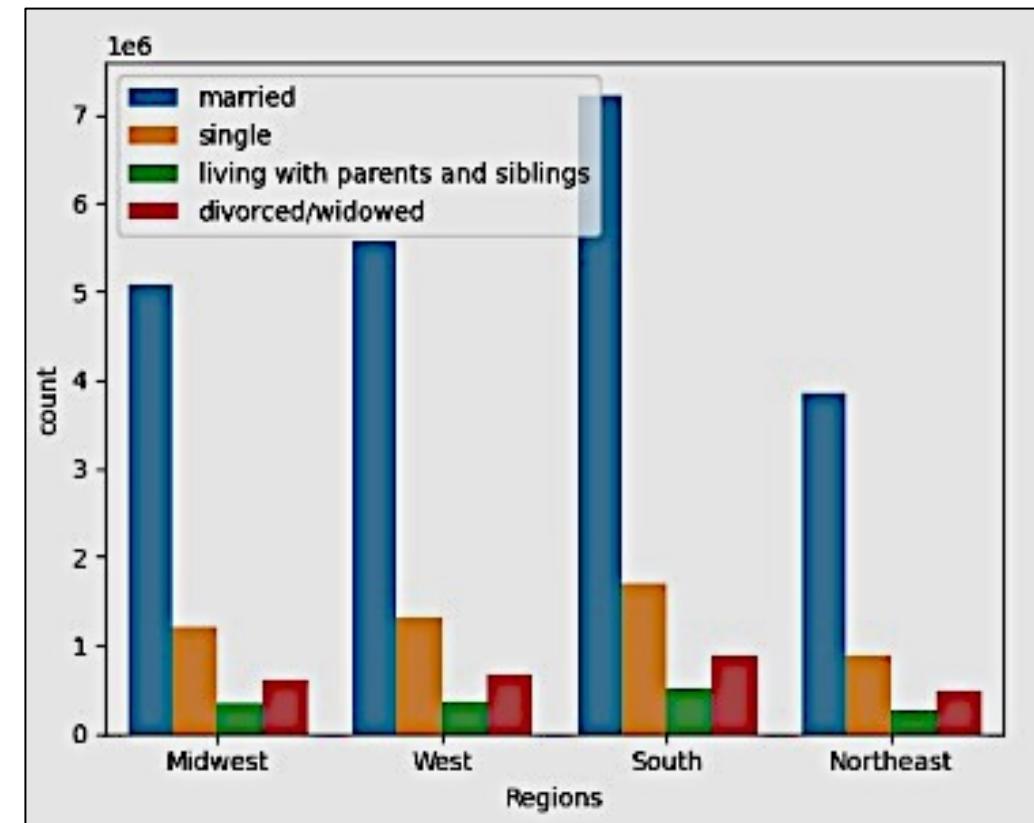
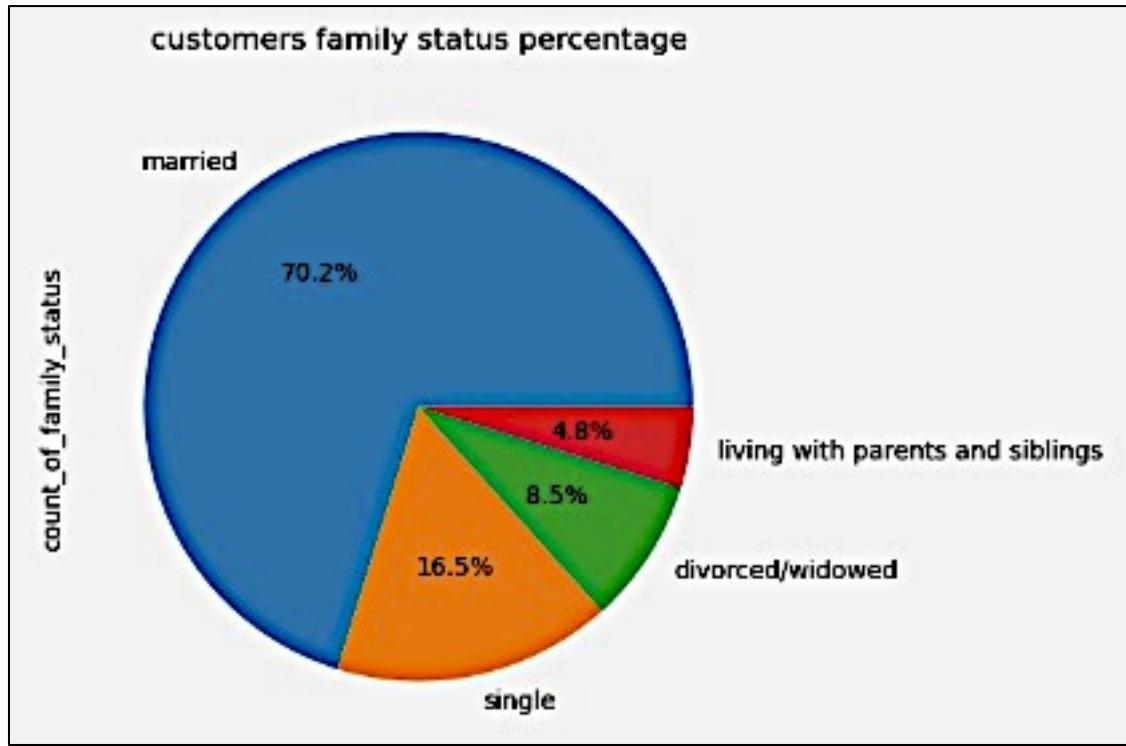


The sales analysis includes several categorical visuals of customer profiles such as family status and age.



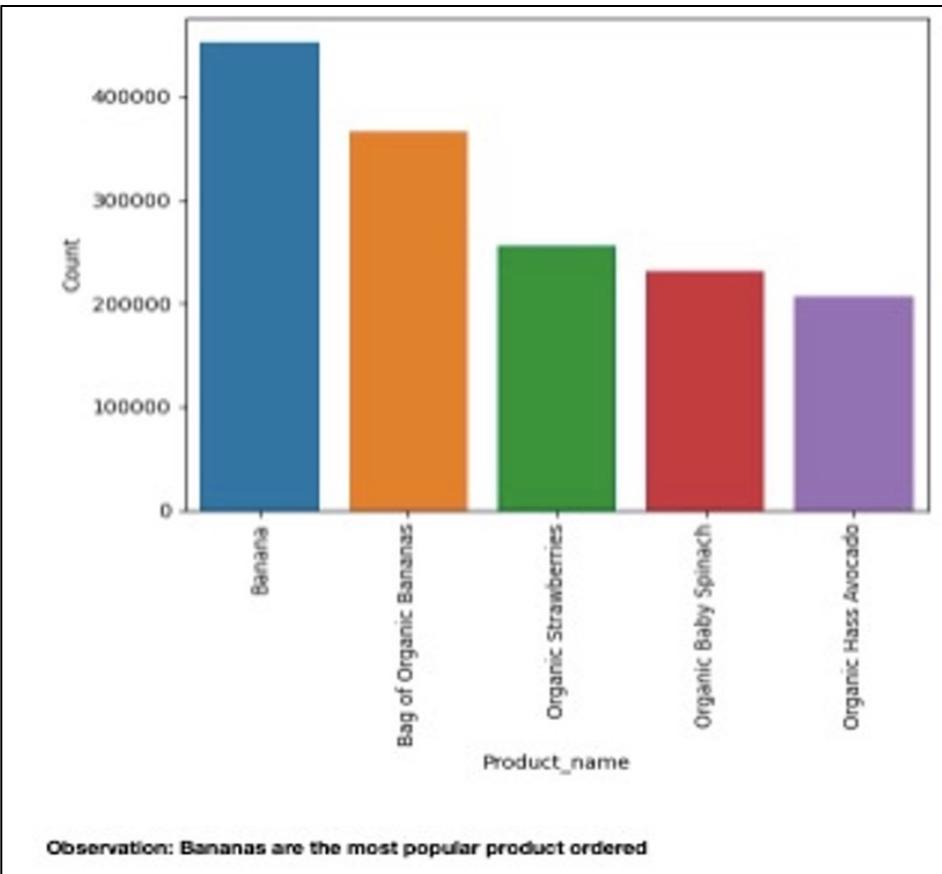


These charts created in Python show that most customers are married and most are in the South region

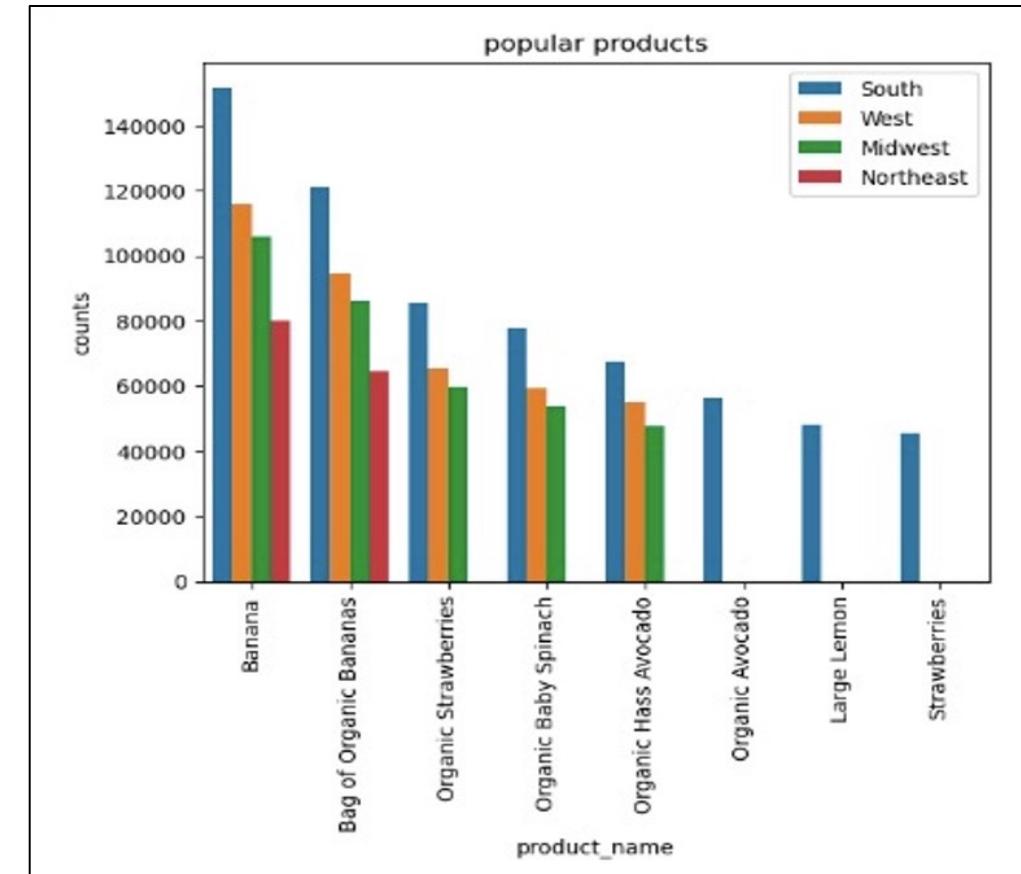




A graph of the 5 most popular products overall



By grouping on regions and products, we can see that bananas are popular with all regions, especially the South.





Insights and Recommendations

Instacart should target ads to customer with product purchases with the following profiles.

- The busiest day for customer purchases is Saturday, the busiest hour is 10am. Customers spend the most money between 4am and 5am.
- The price point for customers is on average under \$10
- The most popular products purchased are bananas
- The most frequent geographic region for customer purchases is the South
- 46% of the customers are considered ‘loyal’ and return more, however ‘loyal’ customers spend less than ‘regular’ customers.
- The average age of customers is 49, and income is \$50-150k





Superstore Sales Analysis

A Superstore Giant is seeking knowledge in understanding what works best for them. They would like to understand which products, regions, categories and customer segments they should target or avoid.

Exploratory Data Analysis reveals insights about the Superstore data.

The data also shows several items to explore such as:

- large deviation of the sales
- 3 sales segments, 3 product categories
- regions of sales
- negative profits
- correlation between sales and profit
- 4 shipping modes
- sales trends



Superstore Sales Analysis

Goal

Using Python and Tableau to do an EDA and find sales insights.

Data

Data Source: [Superstore data](#)

data with customer orders, products, sales, discount, profit, states for years 2014-2017.

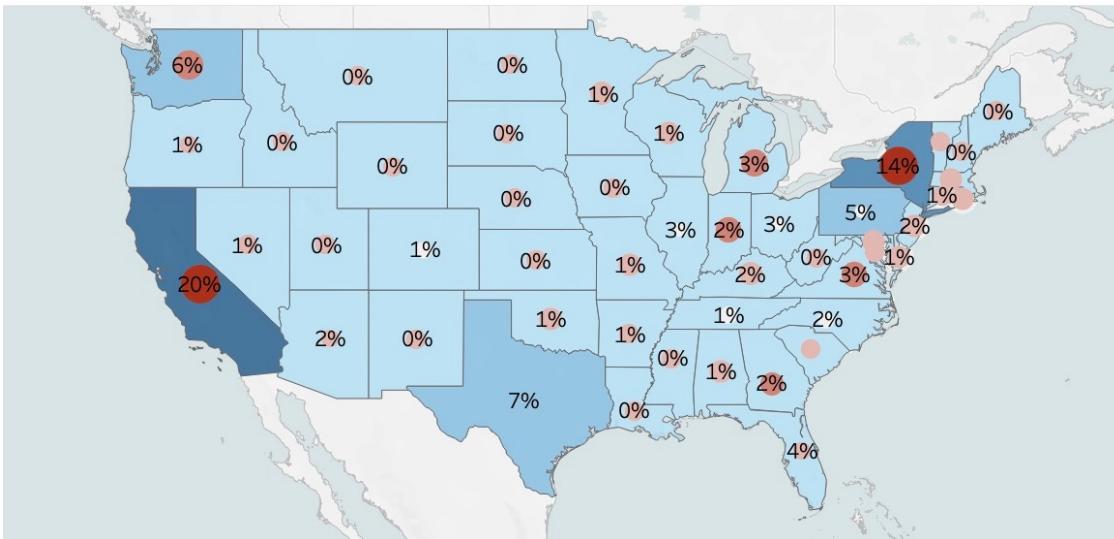
Analysis

- Cleaned data and performed Exploratory data analysis
- Regression analysis of sales and profit, cluster analysis of discounts
- Analysis of products, geographic profitability

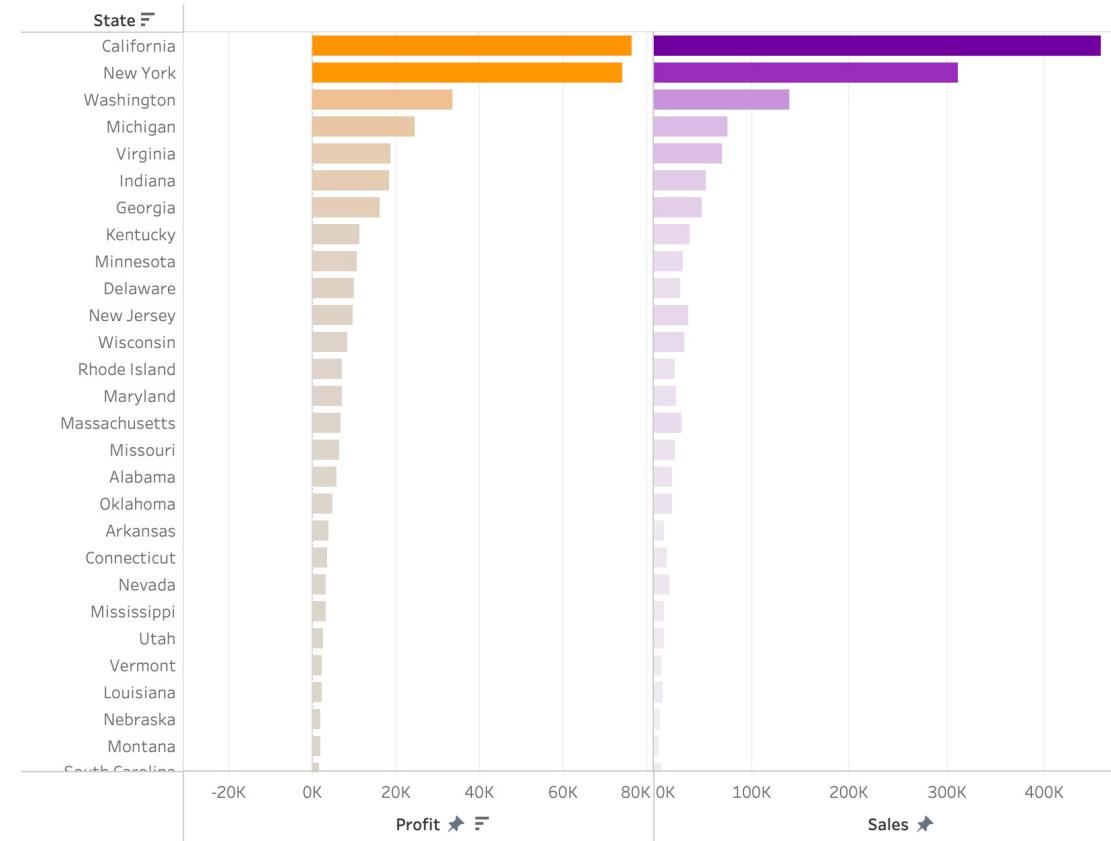


Superstore Sales Analysis

Superstore
Sales and Profit by State

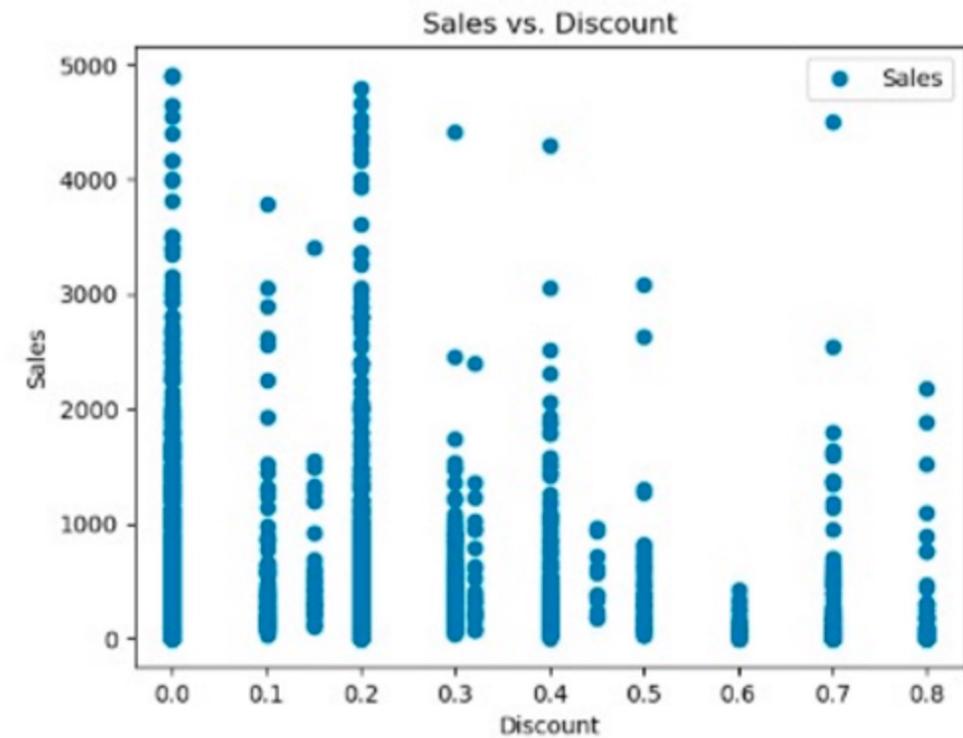
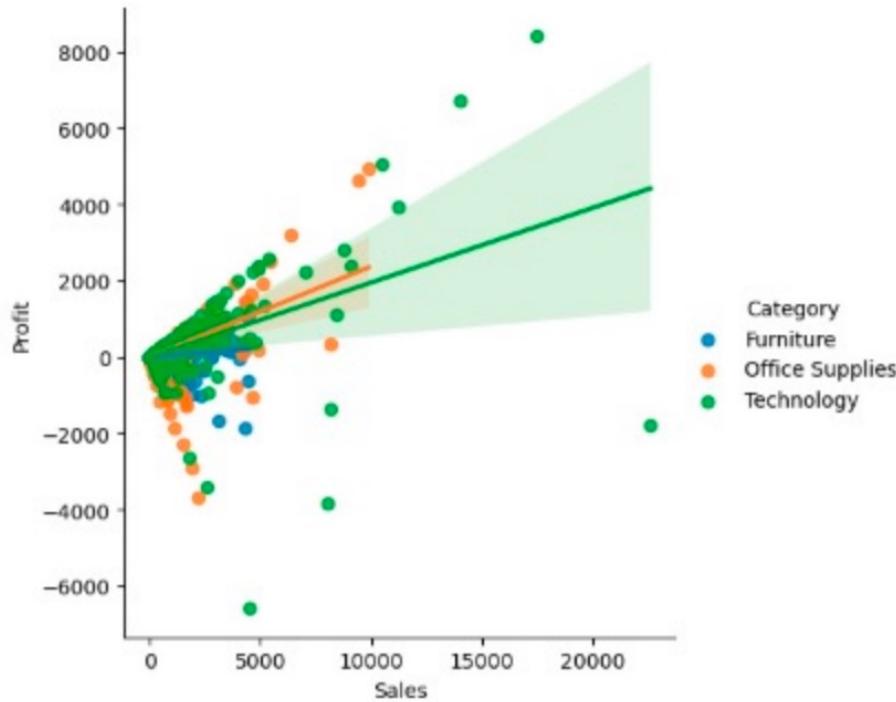


Profit and Sales per State





Superstore Sales Analysis

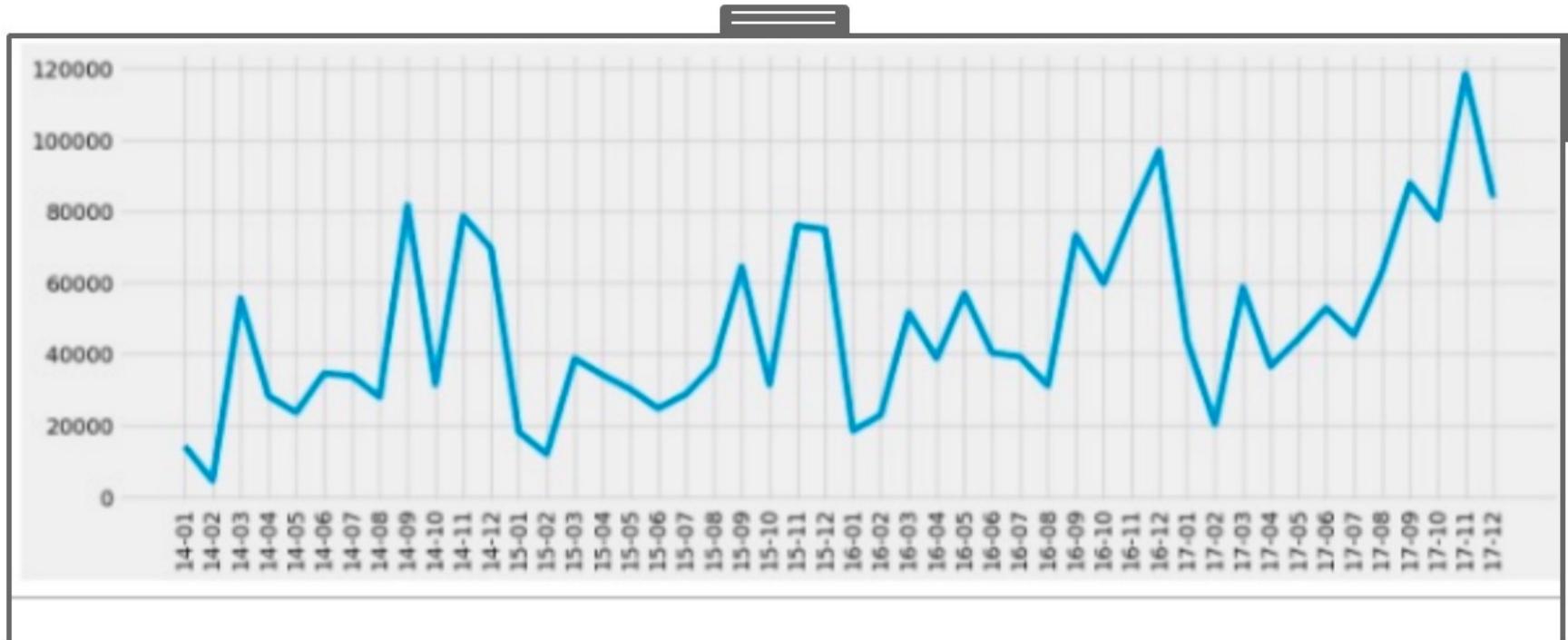


Using a cluster analysis, we can see that sales were only higher with the 20% discount.



Superstore Sales Analysis

Sales over time shows a slight upward trend from 2014 to 2017. There is a peak in the seasonality of the sales in months of September and November and a consistent downturn in February.





Superstore Sales Analysis

Results and Recommendations

The Superstore could look through the product profitability and research the reason some products are getting losses and either change them or try different pricing.

We see the top profit products to target are Staple envelopes and paper and staples. The products to avoid are the binding system, office refrigerator and surge suppressor.

The Superstore can do more research into applying discounts such as the 20% for products with low sales.

The most common shipping mode used was standard.

Research can also be done to understand why some states have losses such as Texas and Ohio versus state that have profits like California and New York.

The Superstore should understand the high and low month seasons and offer promotions with that timing considered.

**For more information and for your data analysis needs
please contact me:**



Margo.Shogren@gmail.com



<https://github.com/mshogren64>