



# Puneet Badwal

DATA ANALYST PORTFOLIO




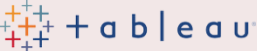







# About Me

- ▶ Education 
  - ▶ Bachelors in Marketing Management at Virginia Tech
  - ▶ CareerFoundry Data Analytics Program
  - ▶ CareerFoundry Data Specialization – Machine Learning
- ▶ Professional Experience 
  - ▶ Cisco Systems Inc. Intern (2017)
  - ▶ 5+ years as Marketing Specialist and Marketing Manager at Cisco Systems Inc. (2018-2023)



# Projects & Tools

<u>Project</u>	<u>Summary</u>	<u>Tools</u>
 GameCo Market Analysis	Analyzed global video game sales trends and presented insights and recommendations	 
 Influenza Staffing Forecast	Forecasting flu season to allocate medical staff and improve resource planning	 
 Rockbuster Stealth Analysis	Analyzed video rental service company to develop new strategies to get the edge on a changing market	  
 Instacart Grocery Analysis	Explored customer base and purchasing behavior for online grocery retailer	  
 Chocolate Bar Analysis	Analyzed chocolate ratings to determine the best factors for best chocolate consumption and production	  

# Game Co

## Market Analysis

Game Co wants to use data to inform the development of new games, by leveraging historical sales data

### Data & Tools:

- VGCharts Dataset (via Kaggle)
- Excel
- PowerPoint



### Objectives:

- ▶ Analyze game sales data from 1980-2016
- ▶ Identify the most popular genres, publishers and platforms
- ▶ Compare sales across regions to create geographic marketing strategy

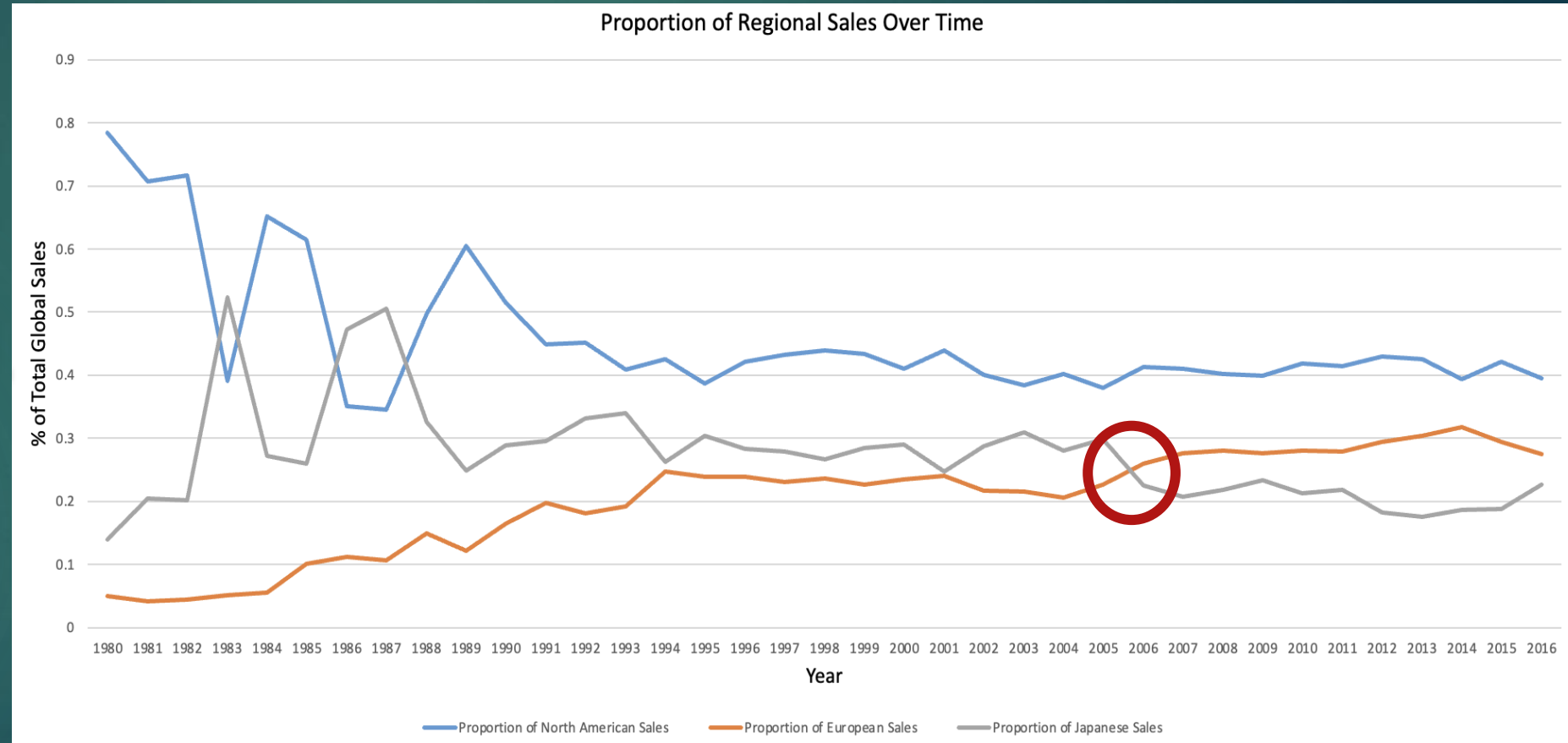
### Analysis Methods:

- ▶ Organized, cleaned, and prepared data in Excel
- ▶ Ran descriptive analysis to find trends, including pivot tables
- ▶ Visualized charts and graphs to display insights & findings

# Regional Market Share Proportions



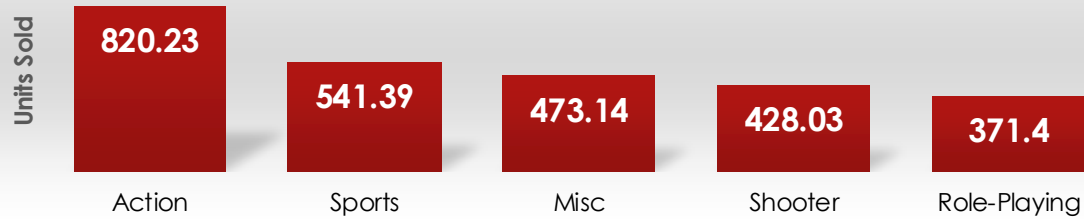
- North America has held a large proportion of total global sales, consistently hovering around 40% share
- The European market has recently surpassed the Japanese market (highlighted 2005), and has since remained in 2<sup>nd</sup> place
- Although the Japanese market was competing with NA in the early days of gaming (1981-1987), it has drastically decreased in recent years
- Based on the current climate, we will be looking into data from 2005-2016 in all our findings for relevant game sales data



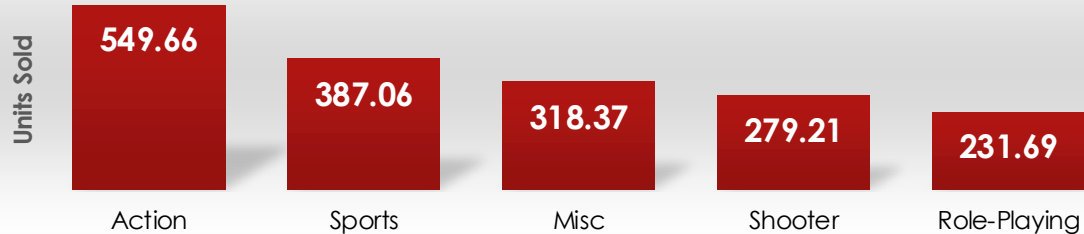
# Top Selling Genres by Region



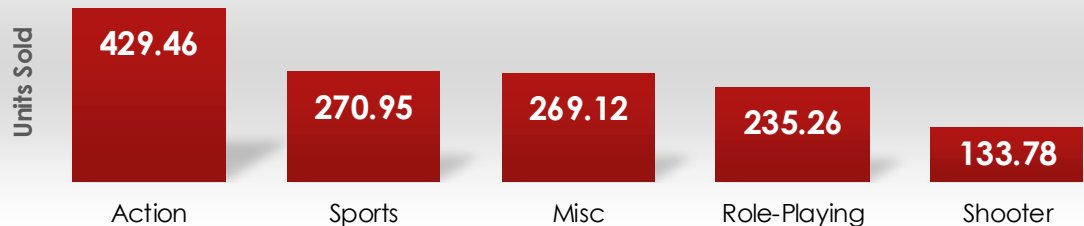
Top 5 Best Selling Genres in North America



Top 5 Best Selling Genres in Europe



Top 5 Best Selling Genres in Japan



- Despite there being over a dozen genres, the top 5 genres from 2005 to 2016 were always Action, Sports, Misc, Shooter, and Role-Playing genres
- All 3 regions had the exact same top 3 best selling genres:
  - 1. Action
  - 2. Sports
  - 3. Misc
- North America and Europe have the exact same market for genres, while the only difference for Japan is switching Role-playing and shooter in the 4<sup>th</sup> and 5<sup>th</sup> spots

# Recommendations



This will be GameCo's largest market, and at least 40% of total global budget should be allocated based on current and historical market share



The top genres remain the same - Action, Sports, and Misc games should all be the primary focus for all new game development



For partnership recommendations, GameCo should think about producing games in alignment with Xbox/Microsoft related products and system releases, based on publisher sales

# Influenza Staffing Forecast



A medical staffing agency needs help planning resource distribution for the upcoming flu season

## Data & Tools:

- US Census Data (US County and State population, 2009-2017)
- CDC Influenza Deaths Data (Monthly US Influenza death counts, 2009-2017)
- Excel & PowerPoint
- Tableau



## Objectives:

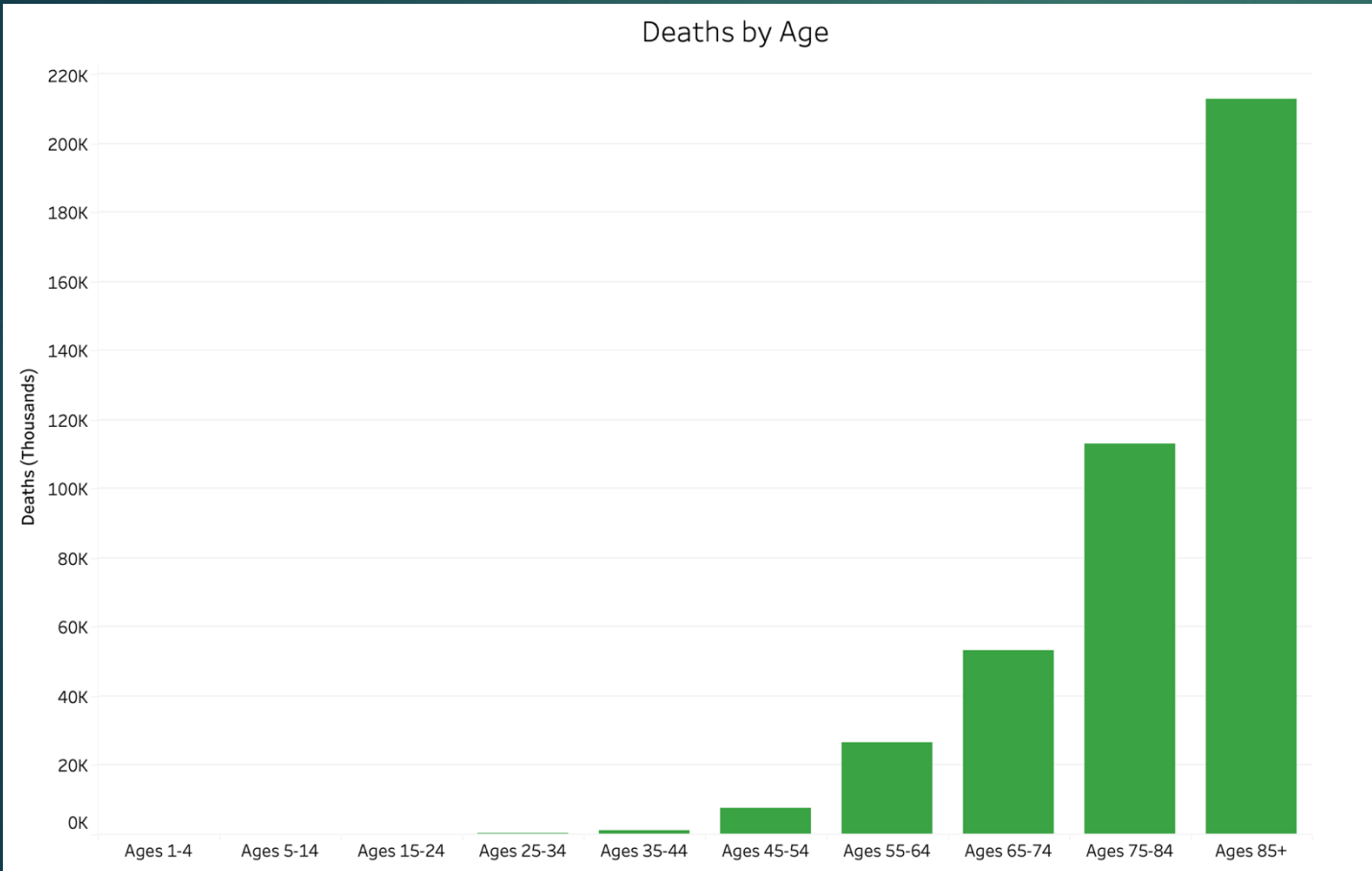
- ▶ Analyze and track Influenza trends based on CDC Flu deaths data (2009-2017)
- ▶ Identify the most vulnerable states based on death rate and population
- ▶ Forecast and optimize logistics for medical staffing agency

## Analysis Methods:

- ▶ Organized, cleaned, integrated and transformed from 2 different dataset (CDC & US Census data)
- ▶ Trending analysis to determine flu seasonality by state
- ▶ Statistical hypothesis testing
- ▶ Data storytelling with Tableau



# Influenza Deaths by Age

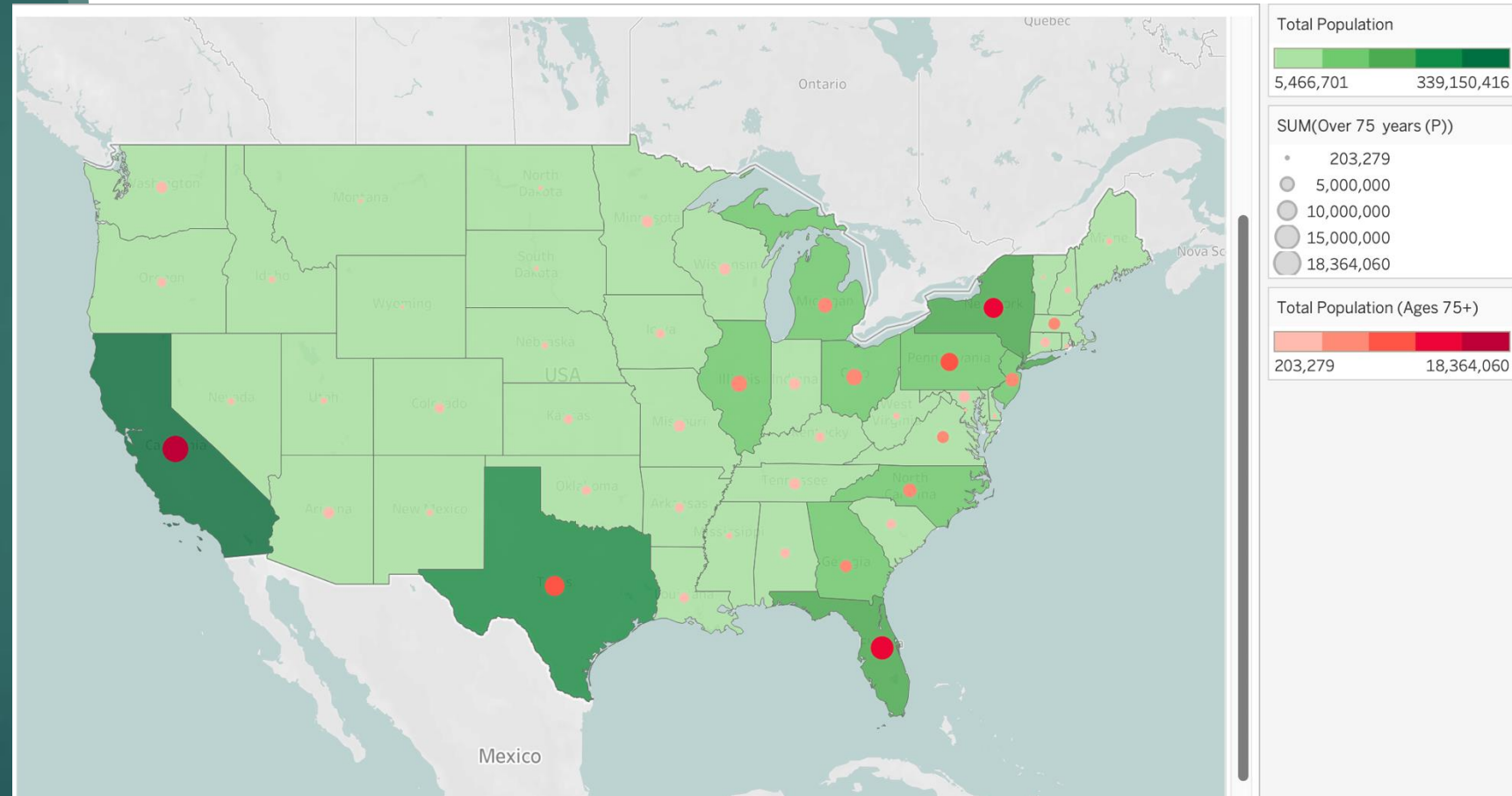


- ▶ We can see from this graphic that there is a drastic increase in Influenza related deaths after the age of 75
- ▶ From this we determine those states with older populations are considered *vulnerable populations*



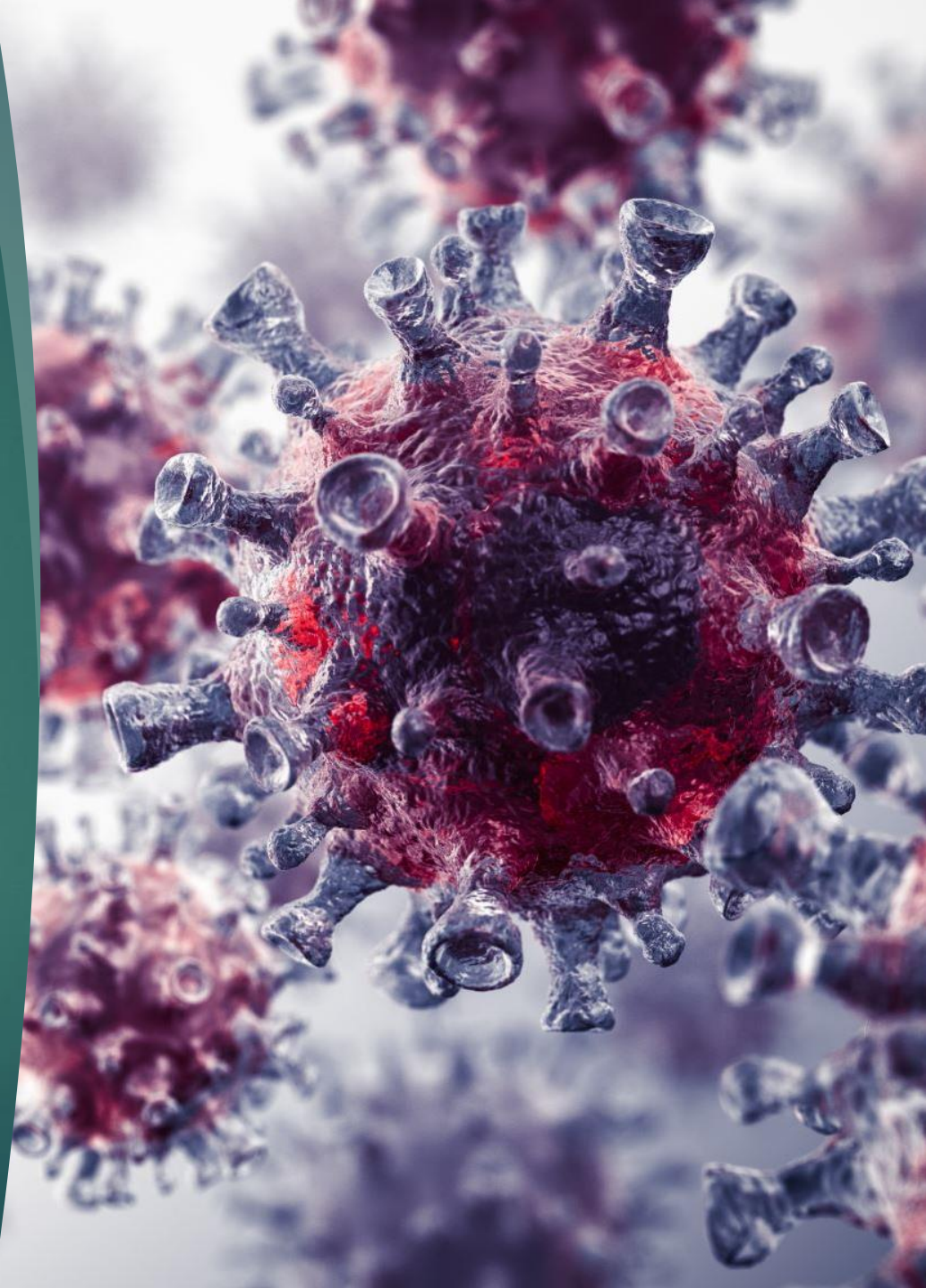
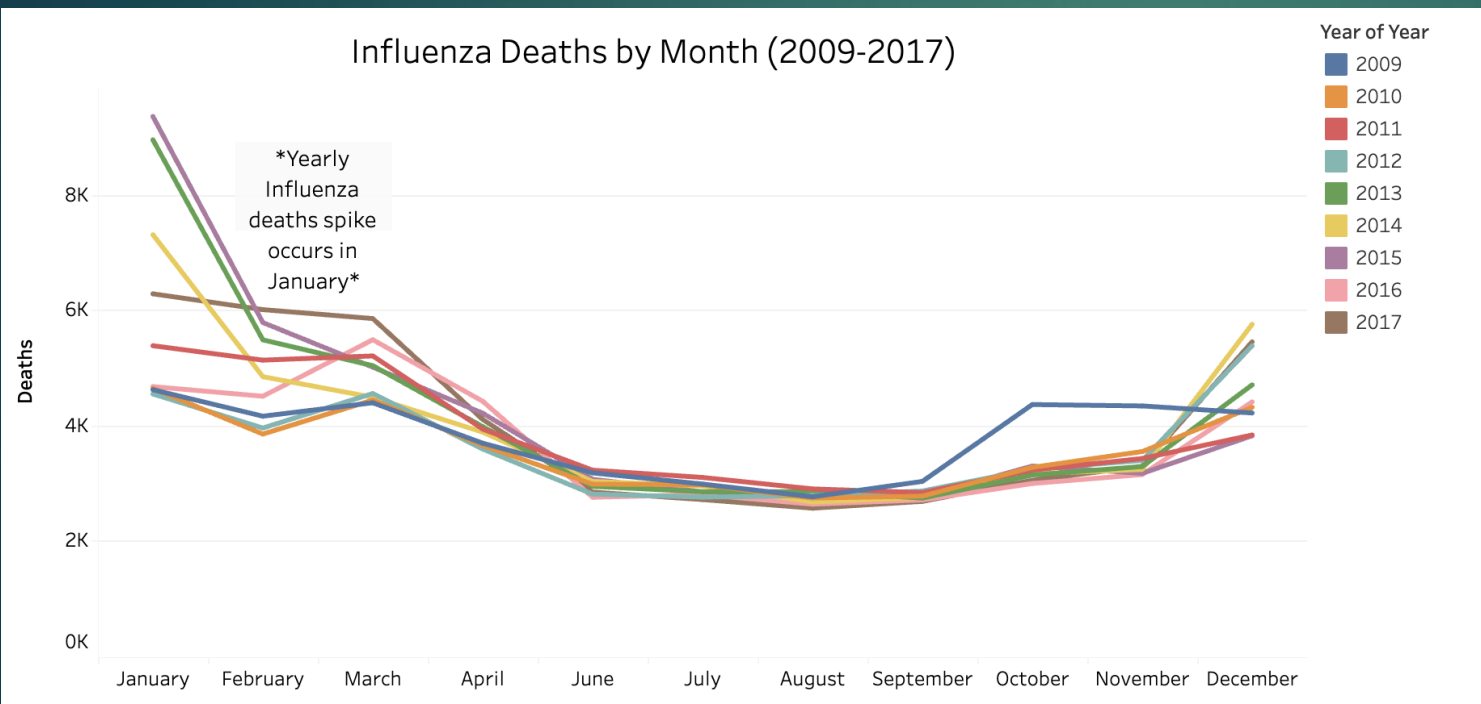
# Choropleth Population Map

- ▶ We can see the states with the highest overall populations share also contain the most number of us citizens over the age of 75
- ▶ These states include:
  - ▶ California
  - ▶ New York
  - ▶ Texas
  - ▶ Florida
  - ▶ Pennsylvania



# When is Flu Season?

Based on the cases of Influenza deaths by month, year over year, we see that Flu Season occurs annually from **November-January**



# Recommendations



It was determined that the “Vulnerable” populations consisted of those over the **age of 75** in the US. Medical staff should focus on patients over this age threshold when prioritizing patients in hospitals to be proactive and help prevent the spread and death of the virus



Based on the US Census population data, the largest populated states should be prioritized, and a larger percentage of medical personnel should be allocated to **California, Texas, Florida, New York, and Pennsylvania**



The Influenza “season” was determined to be from **November-January**. Medical staff should be sent to their respective states prior to the start of this period to alleviate some of the extra traffic coming within the following months. This can cut off the majority of the cases before they start

[\\*Tableau – Storyboard\\*](#)

[\\*Vimeo – Final Presentation\\*](#)



# Rockbuster Stealth Analysis

Rockbuster Stealth LLC, once a global leader in movie rentals is leveraging its existing movie license to launch a new online video rental service

## Data & Tools:

- Rockbuster Stealth dataset which includes: Film Inventory, Customer details, Payment transactions, rental history, and more
- PostgreSQL (pgAdmin4)
- dbVisualizer
- Tableau



## Objectives:

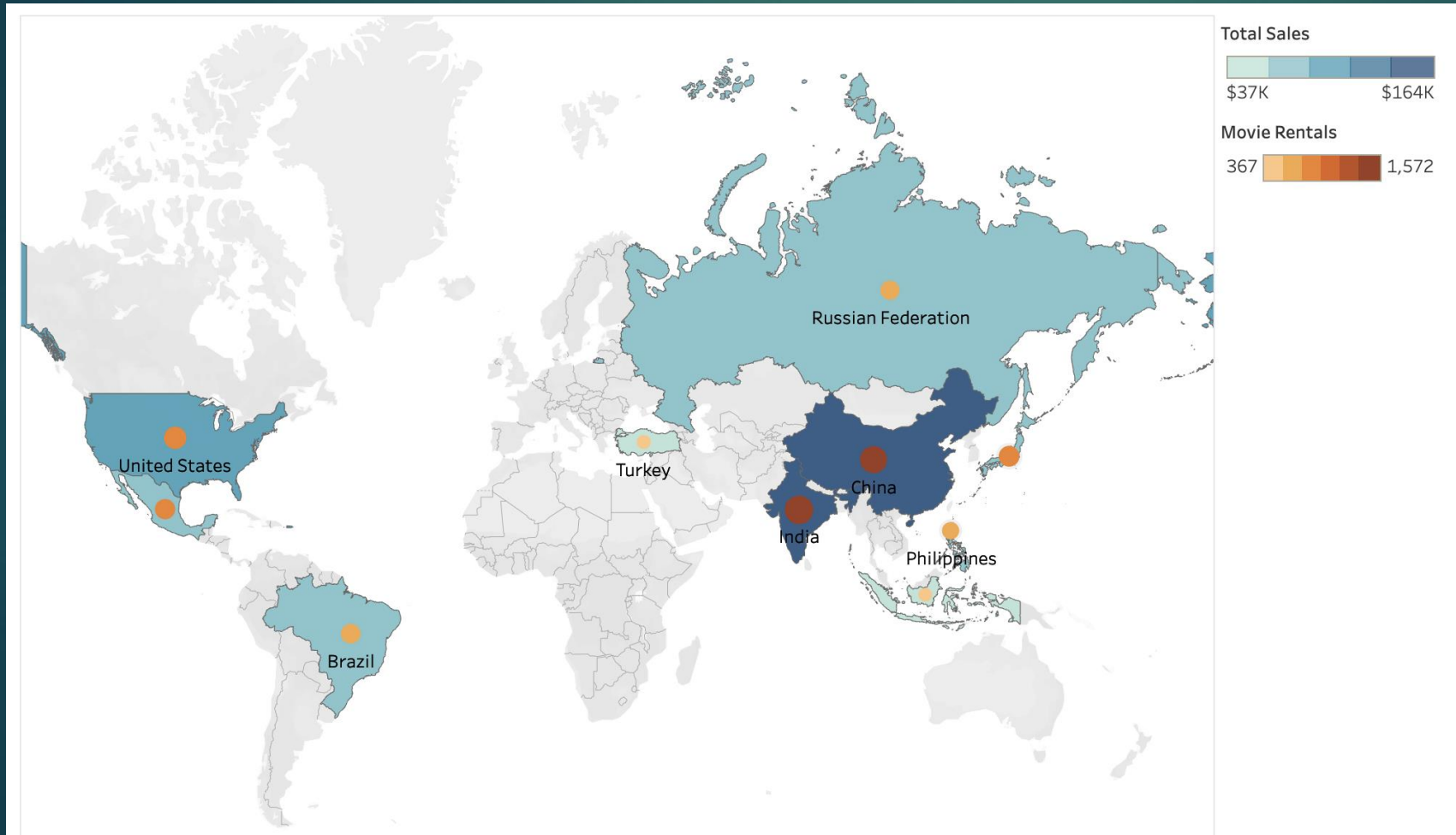
- ▶ Identify most profitable countries and regions, and top spending clients
- ▶ Determine which genres, titles, and ratings generated revenue
- ▶ Recommend strategy for growth based on analysis of market position

## Analysis Methods:

- ▶ SQL & dbVisualizer for data management and querying
- ▶ Perform **Table Joins, CTEs, Subqueries, Aggregations, and filtering** with SQL to extract CSV files
- ▶ Use SQL query outputs and transform into Tableau visualizations for insights



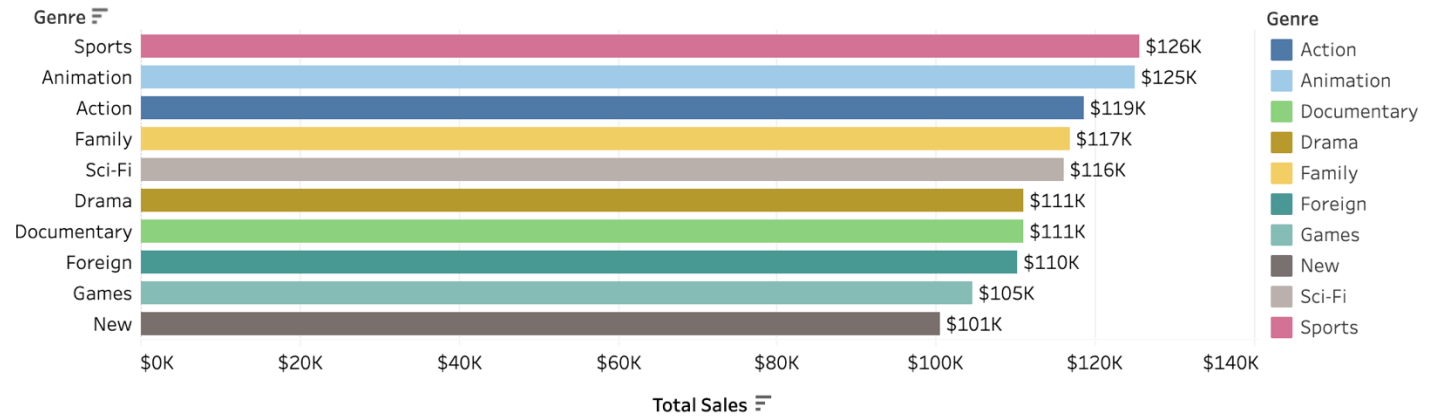
# Rockbuster Stealth's Sales by Country



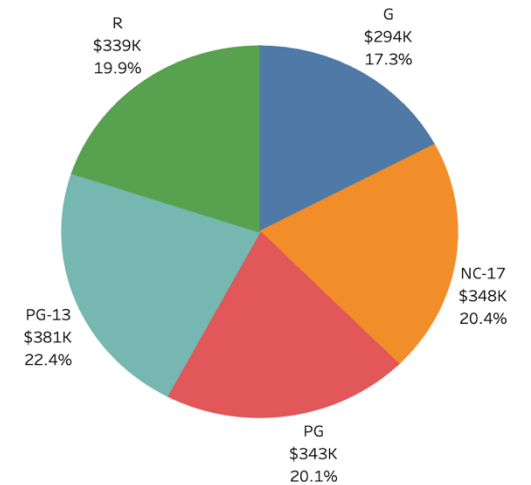
- ▶ The top 3 countries with the most rentals and profits are India, China, and the US, all with over \$100K in sales and over 950+ total rentals.
- ▶ These same 10 top profitable countries, are the same 10 countries with the most customers

# Top titles, genres, and ratings

- The Sports, Animation, and Action genres produced the most sales from the Rockbuster library
- There is a relatively even distribution between the 5 MPAA ratings, with a slight edge going to PG-13 movies making the most sales
- While only 1 movie (Rugrats Shakespeare) falls under one of the top 3 genres (Action), 50% of the top 10 movies are rated **PG-13**



Movie Title	Rating	Genre	Total Sales	Times Rented
Ridgemont Submarine	PG-13	New	\$3,543	843
Apache Divine	NC-17	Family	\$3,507	827
Forward Temple	NC-17	Games	\$3,443	817
Bucket Brotherhood	PG	Travel	\$3,435	808
Pulp Beverly	G	Horror	\$3,420	800
Harry Idaho	PG-13	Drama	\$3,411	803
Rugrats Shakespeare	PG-13	Action	\$3,398	783
Massacre Usual	R	Games	\$3,390	757
Network Peak	PG-13	Family	\$3,388	795
Rocketeer Mother	PG-13	Foreign	\$3,380	801



# Recommendations



India, China, and the US lead the way in terms of strongest lifetime customer base. These 3 countries should be the primary focus for all future campaigns and planning for new releases and promotions



More marketing efforts and resources should be tailored around Asian countries as well, because 7 of the top 10 most profitable countries are in Asia



Rockbuster might consider acquiring more titles in the Sports, Animation, and Action genres, and those that are rated PG-13 as well. These are the most rented and popular types of titles from a global perspective.

[\\*Tableau – Rockbuster Storyboard\\*](#)

[\\*GitHub – Rockbuster Project Files\\*](#)



# Instacart

## Basket Analysis

Instacart would like to uncover more information and insights about their sales patterns and customer base demographics

### Data & Tools:

- Instacart Online shopping data from 2017, via CareerFoundry
- Python (Matplotlib, Seaborn, Pandas, NumPy)
- Excel
- Jupyter Notebook



### Objectives:

- ▶ Identify peak shopping times and spending patterns
- ▶ Understanding customer segments and their buying behaviors
- ▶ Examine customer loyalty behavior and refine targeted marketing strategies by region

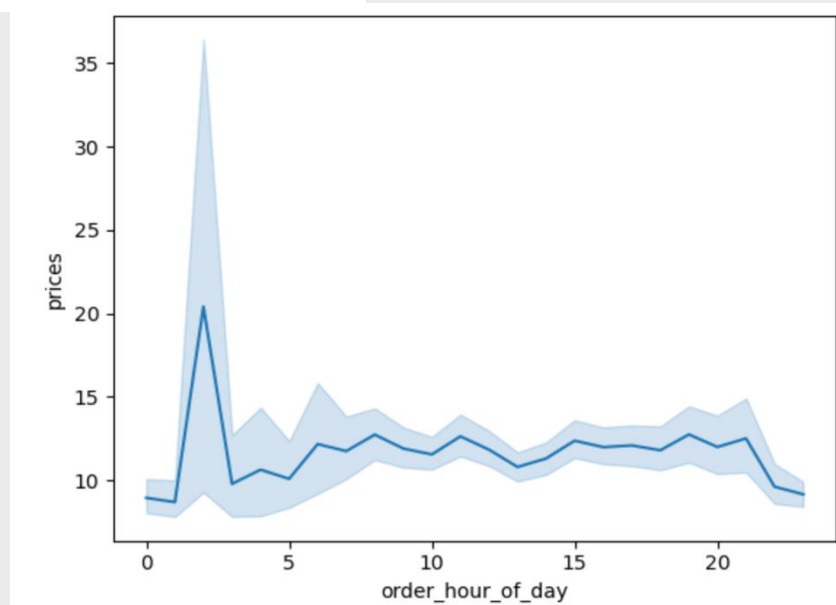
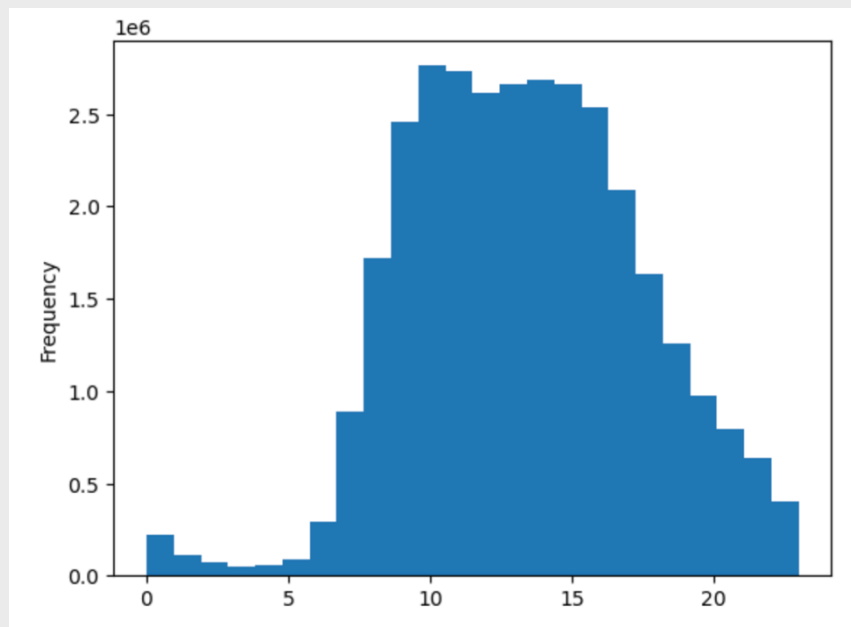
### Analysis Methods:

- ▶ **Python** (pandas, NumPy, matplotlib, seaborn) for analysis and data visualization
- ▶ Data cleaning, merging, wrangling, subsetting, grouping and aggregation in Python
- ▶ Exploratory data analysis and organizing population flow of multiple data files in Python



# Busiest Hours of the Day

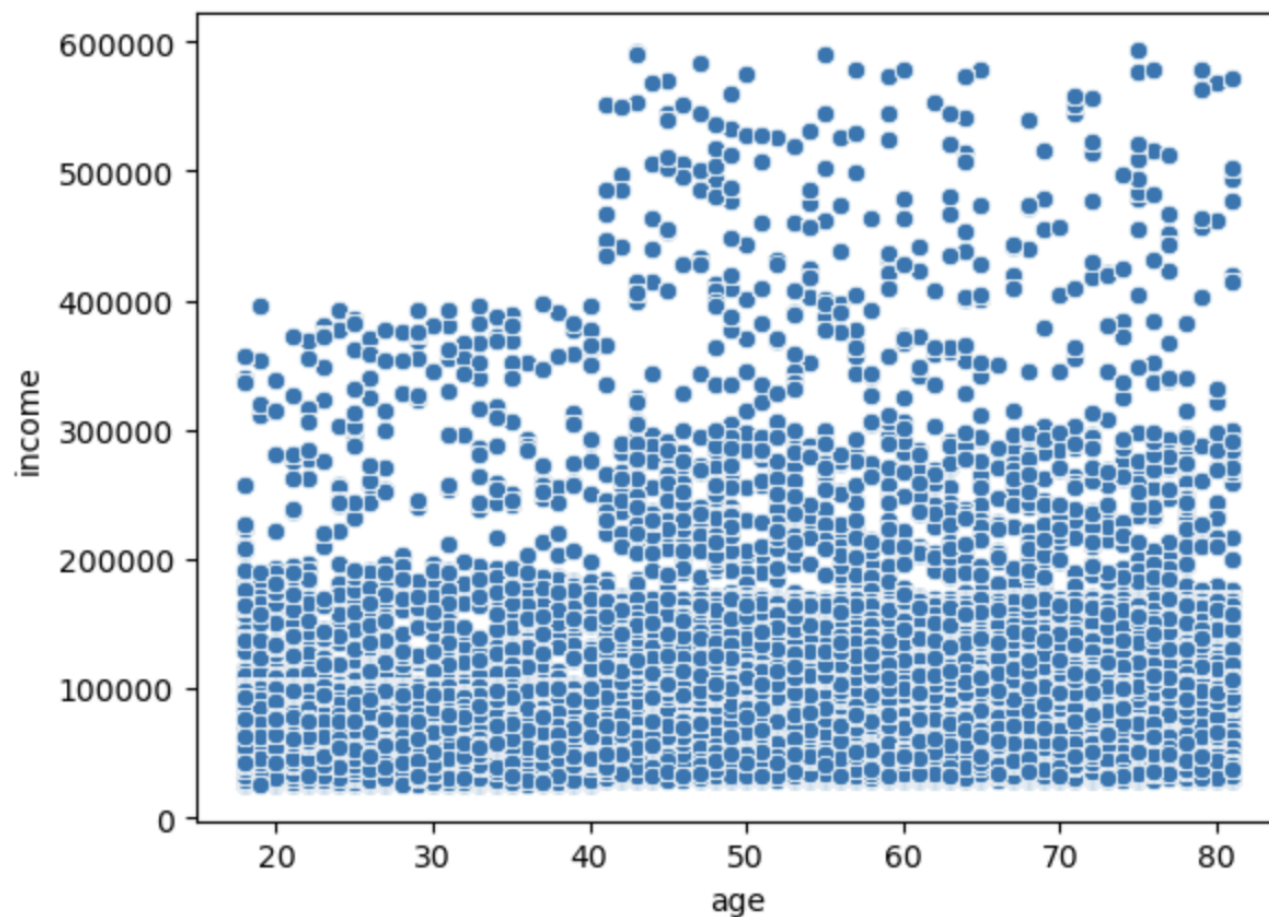
- ▶ The most shopping on Instacart gets done between the hours of 8AM and 4PM, with peak purchasing coming at Noon
- ▶ Interestingly, we see the most money spent happens late at night between 1:30-3:00 AM





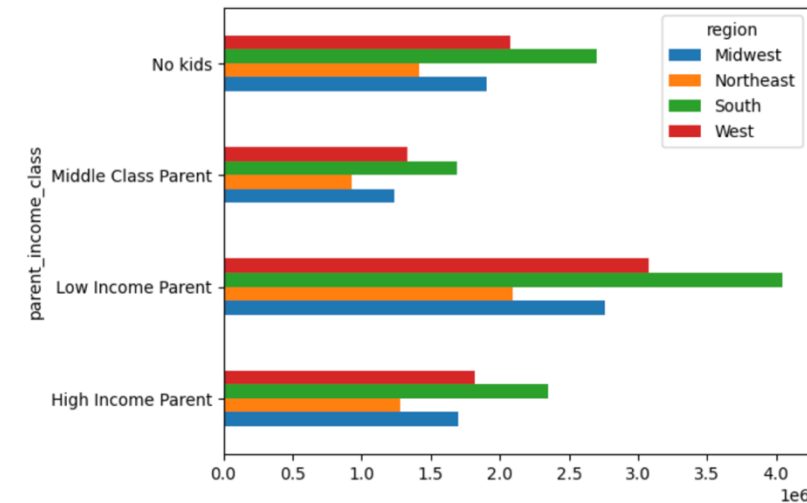
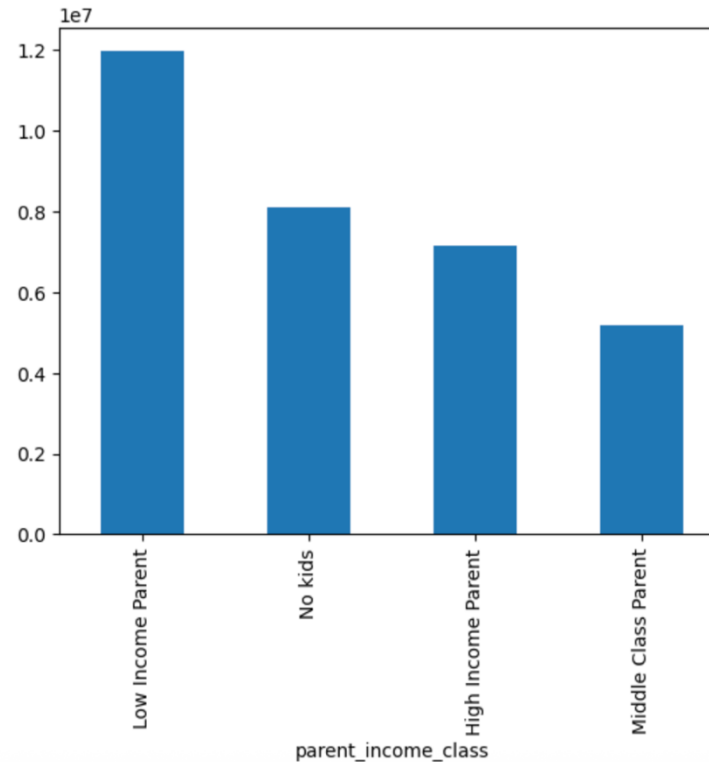
# Age and Income Correlation

- ▶ We can see an obvious pattern for those above the age of 40. They have significantly more wealth and purchasing power
- ▶ There isn't a single customer under the age of 40 that is in the high income bracket (making over \$400K)



# Parent and Income Class Segment

- ▶ We can see the majority of Instacart's customer base are parents, especially those in lower income (under \$96K income) classes
- ▶ The Southern region of the US produces the most sales, no matter the status of parent or income class



# Recommendations



While the most shopping is done during midday, Instacart should promote more ads/sales around Noon on their website to sell more. Instacart can also discount bulk order shopping done at night for large orders placed after midnight and before 4AM



While not the largest customer base, the most wallet share opportunity is for Instacart's clients over the age of 40. They have the most income available, and Instacart can promote products that are more adult focused for their older audience



The majority of Instacart's customers are parents. While the primary regions of focus should be the south, Instacart can also promote their "low-range" product prices, for those under \$10 for lower income families.

[\\*GitHub – Instacart File Repository\\*](#)

# Chocolate Bar Analysis



Exploring patterns in chocolate bar ratings using the 'Flavors of Cacao' dataset. A deep data analysis dive to determine the most influential factors on the best tasting chocolate

## Data & Tools:

- "Flavors of Cacao" dataset (2006-2017), via Kaggle
- Python (Matplotlib, Seaborn, Pandas, NumPy, etc.)
- Tableau
- Jupyter Notebook

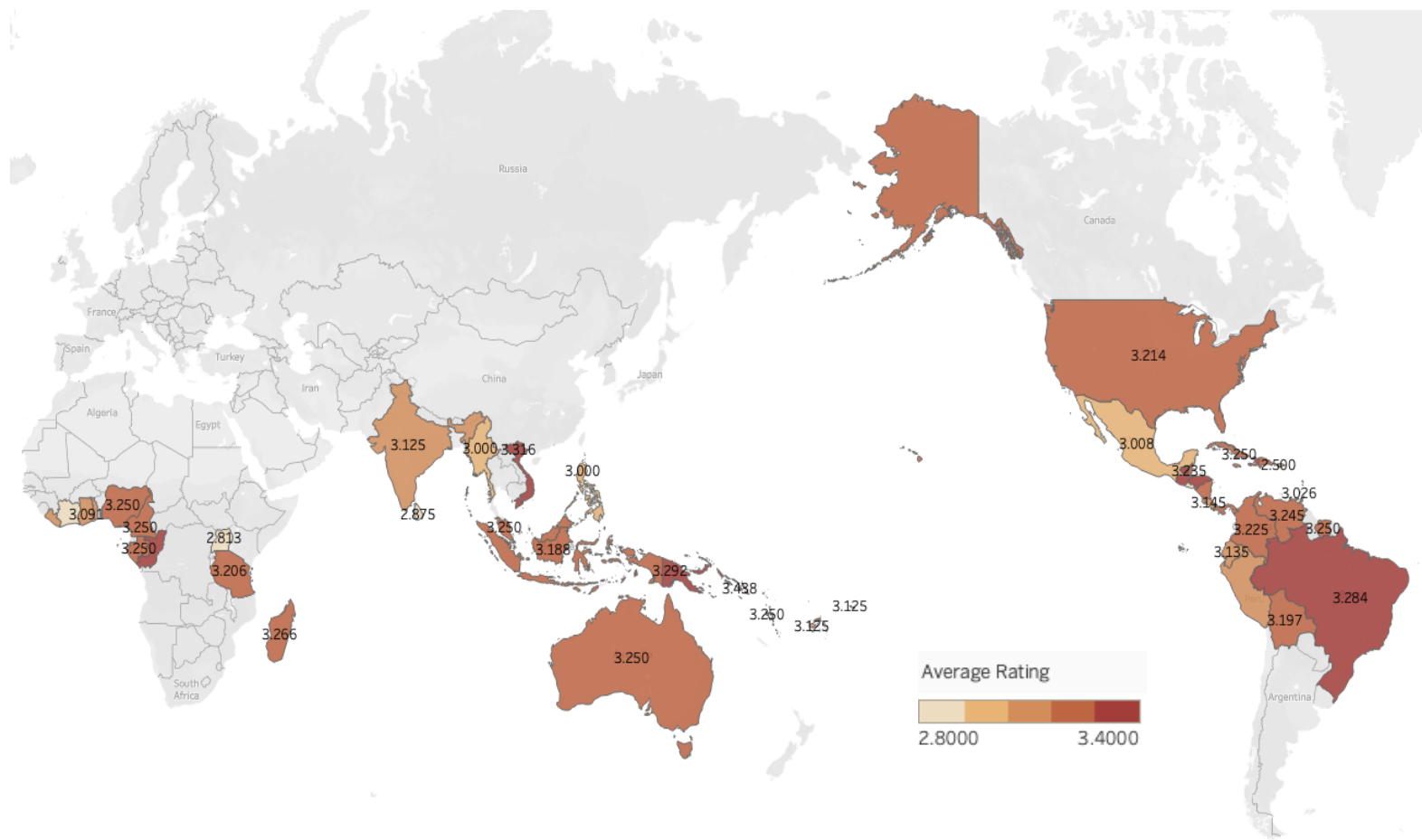


Objectives:

- ▶ Identify influencing categorical factors for best tasting chocolate (bean type, company location, etc.)
- ▶ Use linear regression and clustering to determine correlation/relationships between numerical factors (cocoa % vs. ratings)
- ▶ Examine regional trends to pinpoint best producing chocolate regions of the world

Analysis Methods:

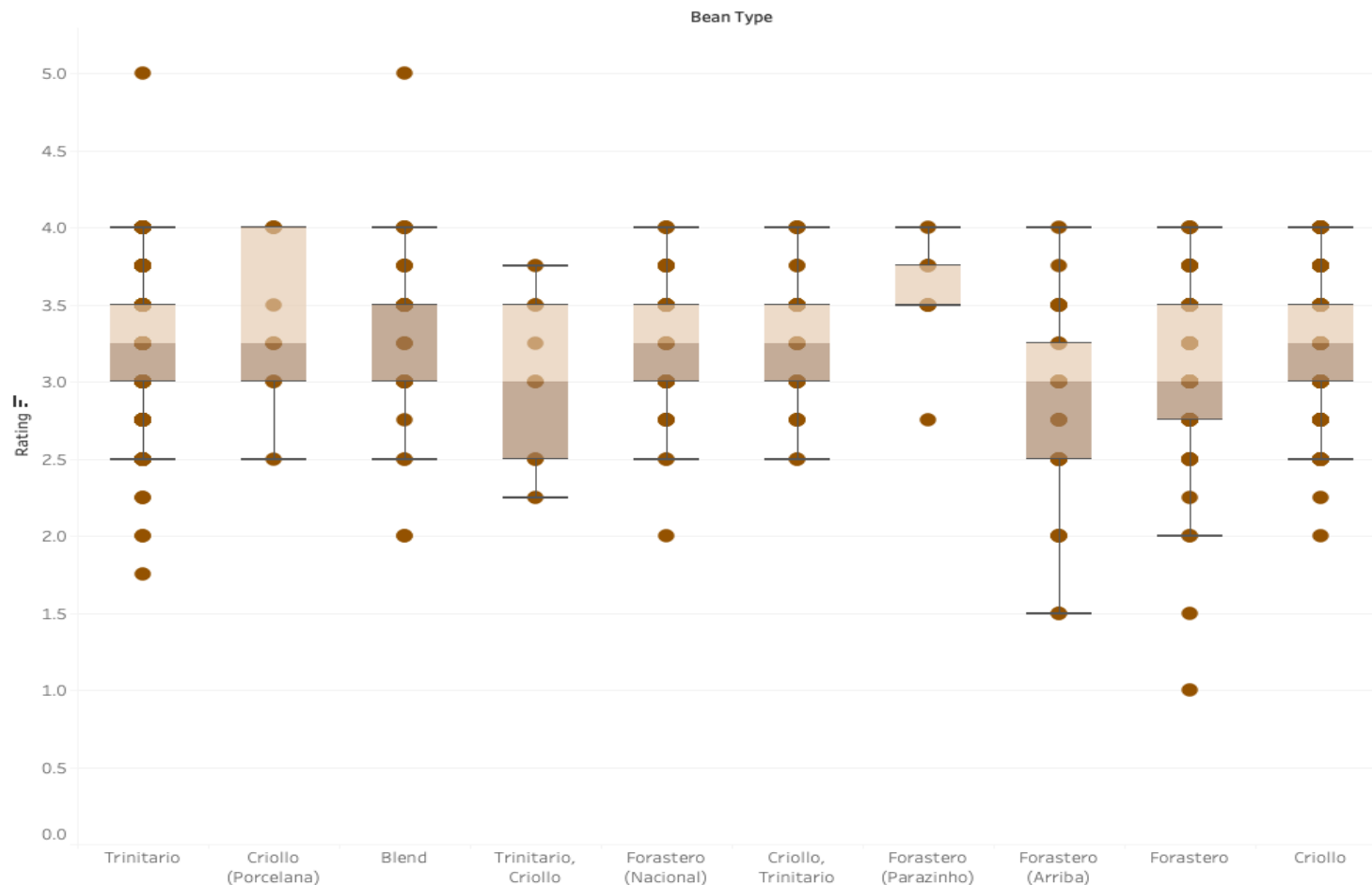
- ▶ **Python and Tableau** (pandas, NumPy, matplotlib, seaborn) for analysis and data visualization
- ▶ Supervised Machine Learning (Regression), Unsupervised ML (Clustering), geographical visualizations, time series analysis in Python
- ▶ Exploratory data analysis and cleaning and consistency checks (data wrangling and transforming) in Python



# Geospatial Analysis

- The top bean producing countries by average chocolate rating are **Solomon Islands, Haiti, Honduras, and Guatemala**, producing better than a 3.3 average chocolate rating
- The South and Central American regions produce some of the best bean quality, alongside some smaller countries in Southeast Asia
- The worst bean quality comes from **Puerto Rico, El Salvador, Ivory Coast, and Uganda**, producing lower than an average 2.8 chocolate rating

# Bean Types



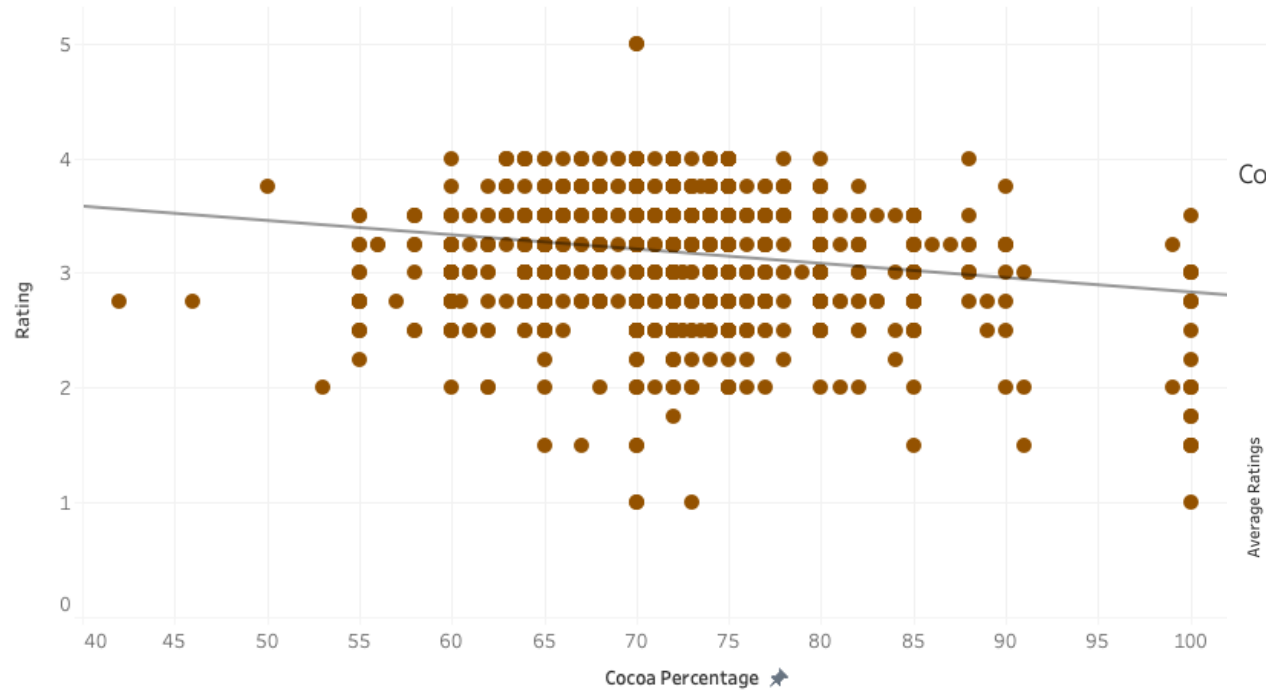
- The top 3 bean types that produce the highest average ratings are **Forastero (Parazinho)**, **Criollo (Porcelana)**, and **Blend**

- The **Trinitario** and **Blend** bean types are the only 2 bean types to produce a rating of 5, which is considered "Elite" tasting. The only two instances out of 1785 ratings

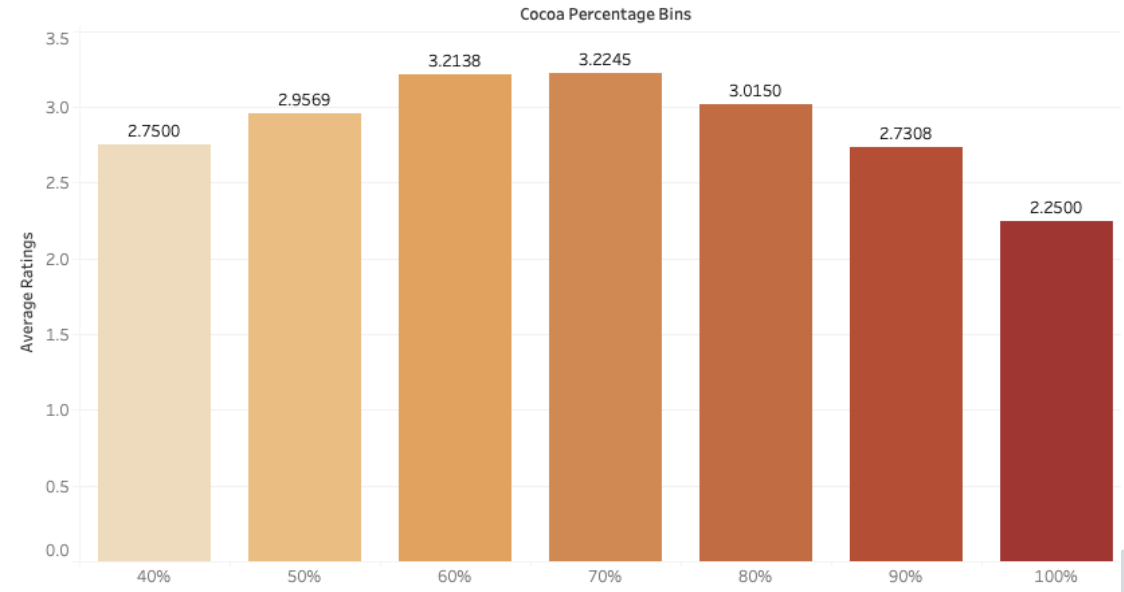
- **Forastero (Arriba)** shows the widest spread of scores ranging from 1.5-4.0. **Forastero** bean type is the only bean to produce a rating of 1.0, the lowest possible rating, considered to be unpalatable and unpleasant.



Cocoa Percentage vs. Chocolate Rating



Cocoa Percentage Bins Average Ratings



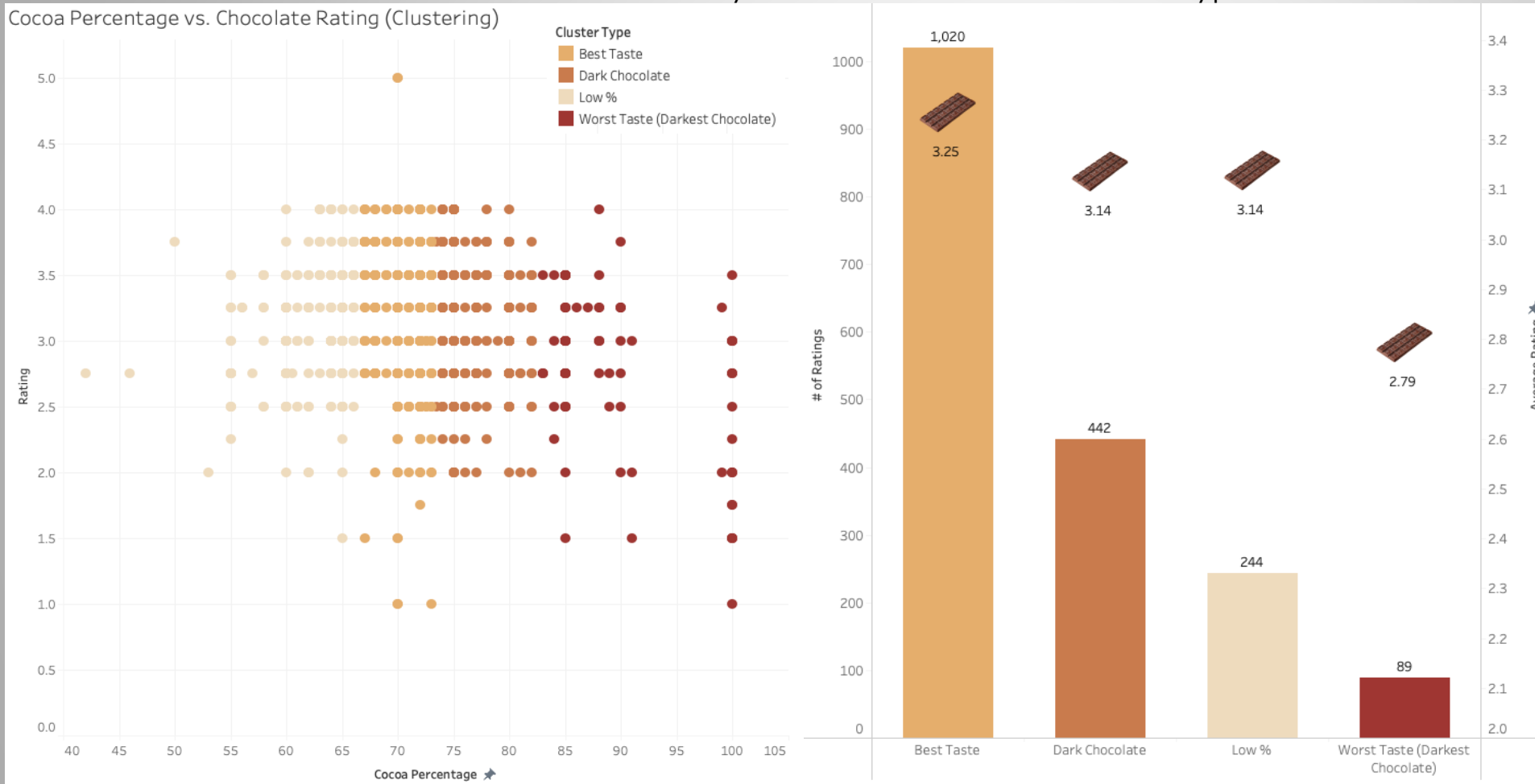
# Linear Regression

- The bar chart displays average ratings per cocoa percentage bin, while the scatterplot with the linear trend line displays every rating per cocoa percentage
- We see a downward slope suggesting an inverse relationship hypothesis: As cocoa percentages increase, chocolate ratings decrease
- The results of a low r-squared value, indicating that cocoa percentage is a weak predictor of bar rating. Further cluster analysis is needed for more insight....

# Cluster Analysis



\*The cluster analysis reveals 4 different cluster types\*



## Best Taste Cluster:

This cluster is the "best in class" tasting chocolate with a mean rating of 3.25. It is defined as chocolate with cocoa % between 65-75%

## Worst Taste Cluster:

This cluster is the darkest chocolate with a cocoa % over 85%. Due to the bitter taste and niche flavor, it receives the lowest mean rating of 2.79

**Dark Chocolate & Low% Clusters:** Both these clusters are interchangeable as the second-best tasting cluster with a mean rating of 3.14. They are chocolate bars with cocoa percentages of between 75-85% & below 65%, respectively.

# Recommendations



Chocolate bar manufacturers should stay close to the Americas (primarily central and South), and Southeast Asian regions to produce the best tasting chocolates using superior bean quality from those regions



The best bean types to use are Forastero (Parazinho), Criollo (Porcelana), and Blend. These beans types have the highest average ratings meaning superior taste and flavor profiles.



The “sweet” spot for cocoa percentages are between 65-75%, for the best tasting and highest rated chocolate. The most sales can be produced with the best tasting cocoa percentage. Producing chocolate bars with over 85% cacao should be avoided due to extreme bitterness and lower taste ratings.

[\\*GitHub – CBA Repository\\*](#)

[\\*Tableau – CBA Storyboard\\*](#)

# END



- ▶ Contact Information

- ▶ Email: [badwalpuneet@gmail.com](mailto:badwalpuneet@gmail.com)
- ▶ [LinkedIn](#)
- ▶ [GitHub](#)
- ▶ [Tableau](#)