Eric McClelland

Data Analytics Portfolio

Projects

GameCo

Analyze global video game sales to recommend market strategy.

Medical Staffing Company

Analyze influenza trends to recommend staffing assignments.

Rockbuster Stealth, LLC

Analyze movie rentals to recommend market strategy.

Instacart

Analyze grocery orders and customer profiles to recommend sales and promotions.

Texas Rain

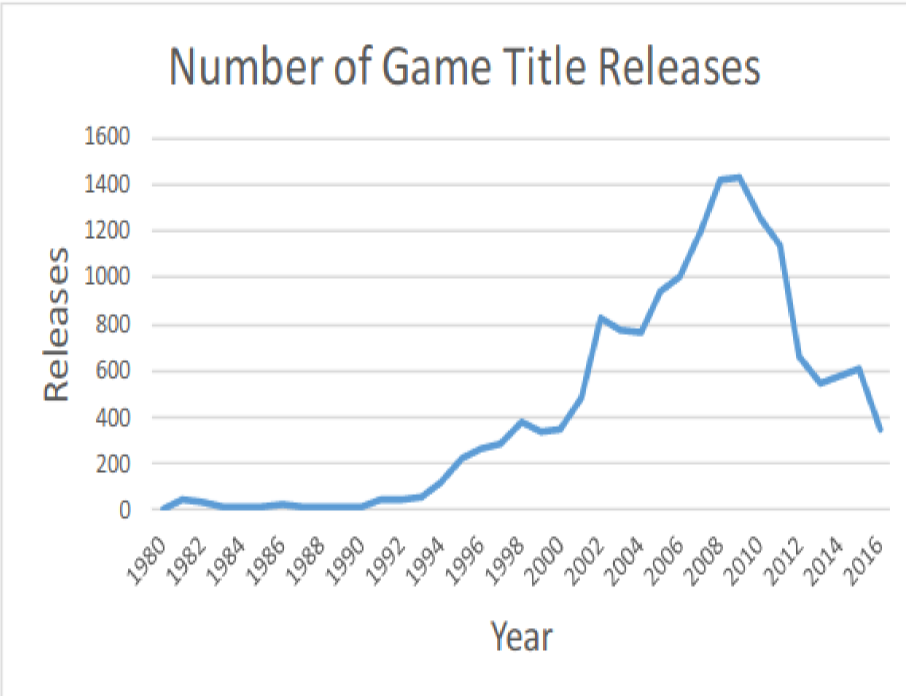
Analyze dataset of weather events to identify relationships.

GameCo

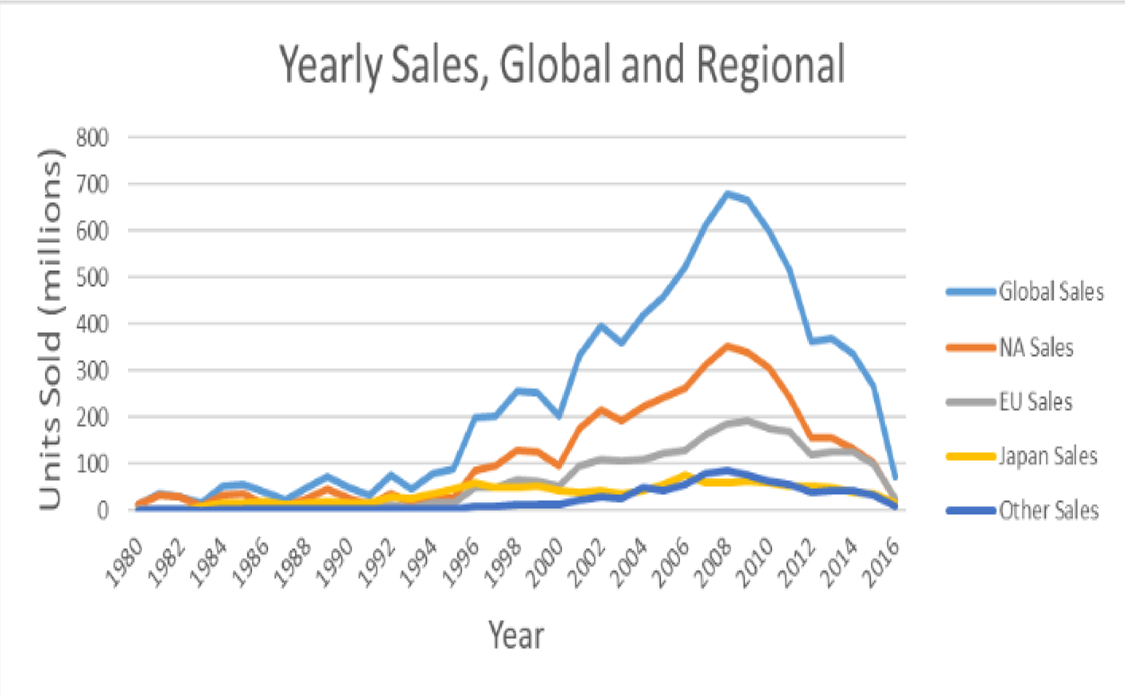
- **Objective**: Perform a descriptive analysis of a video game data set to foster a better understanding of how GameCo's new games might fare in the market.
- Project Brief
- **Skills Used**: Data Integrity checks and Cleaning; Grouping, Summarizing, and Visualizing data; Descriptive Analysis; Presentation of findings.
- **Dataset**: Video Game Sales from VGChartz.
- **Tools Used**: PowerPoint Excel

Analysis for 1980 - 2016: Plummeting Physical Market

Both the number of new game title releases - and their sales - have plummeted since their peak in 2009, at least for physical retail. Our data do not cover online purchases or digital delivery channels.



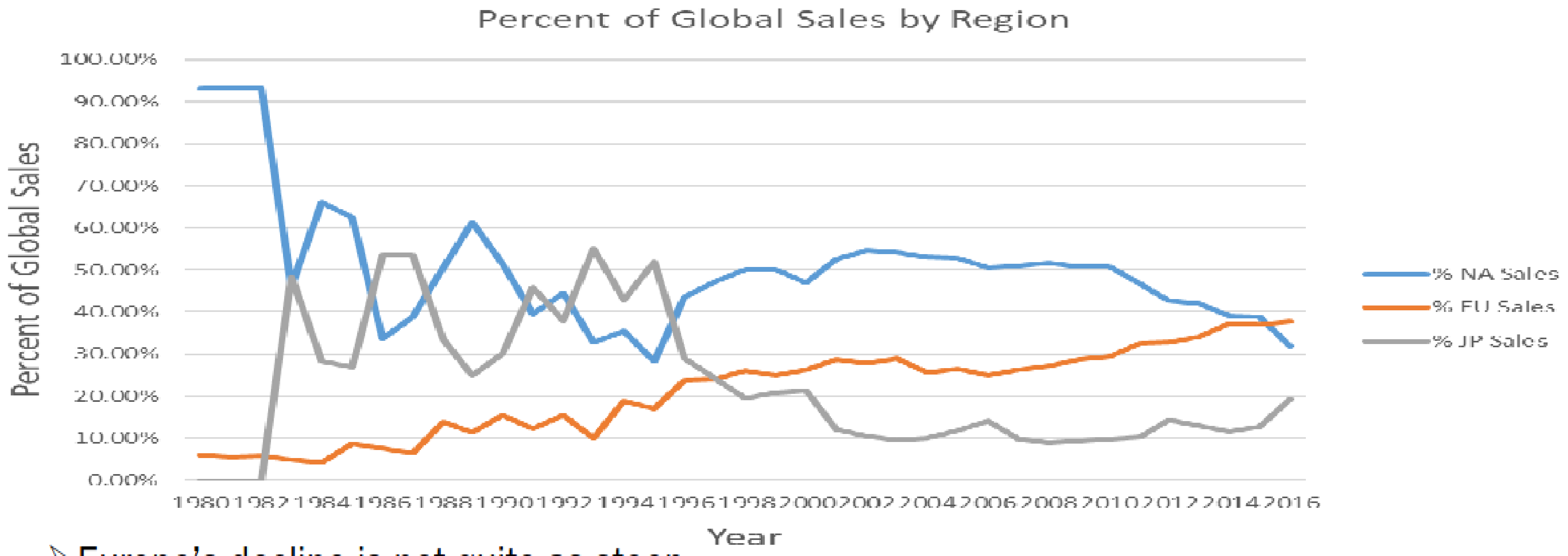
Number of Game Title Releases



Yearly Sales, Global and Regional

Analysis for 1980 – 2016: Regional Dominance Shifts Over Time

North America and Japan alternately dominated until the mid-1990s in terms of units sold, after which Japan fell to third place. However, among the worldwide declines since 2009, Europe’s has been the most gradual, such that Europe now appears to have the largest slice of this dwindling pie.

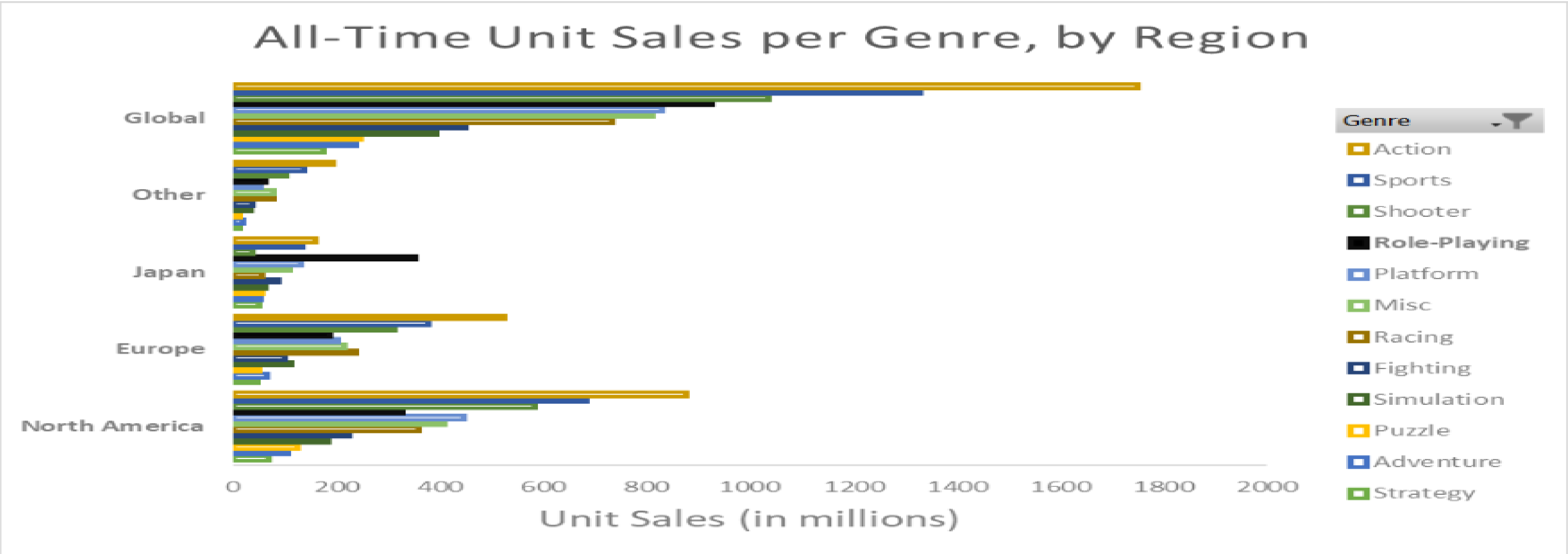


- Europe’s decline is not quite as steep.

Analysis for 1980 – 2016: Genre Dominance

Action, Sports, and Shooter genres are the top three globally, and everywhere except Japan.

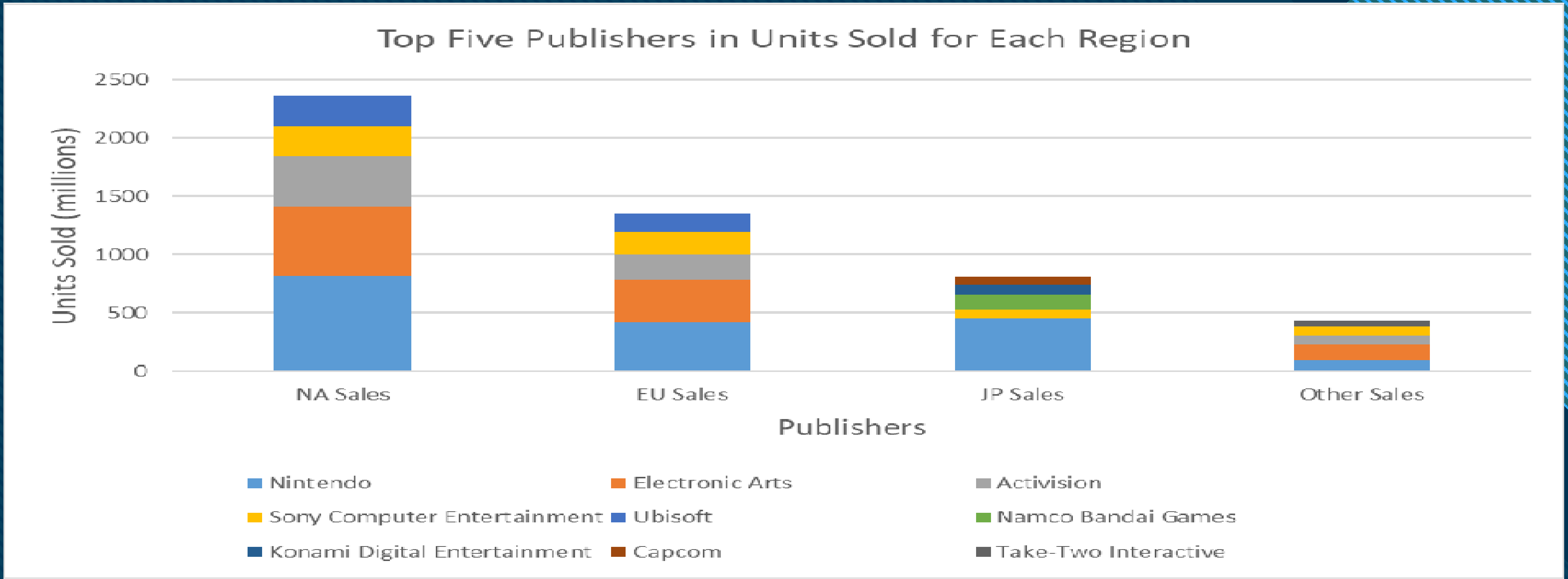
In Japan, Role-Playing games (highlighted in black) \dominate.



Analysis for 1980 - 2016:
Regional Competition: Top Five Publishers

Nintendo has dominated all three main regions.

Electronic Arts a strong second in North America and Europe, but not top five in Japan.



Recommendations

The physical retail market for games appears to be in serious decline!

- Since its peak in 2008 and 2009, physical retail game sales have plummeted, and the likely cause is that purchasing - and to some extent delivery - has shifted online.
- Urgently investigate online markets.
- We lack data about online markets, so we should acquire and analyze such data immediately.

To the extent that we remain invested in the physical retail markets:

- Europe appears to be the best region for investment right now, with North America a falling second and Japan a rising third.
- Europe is declining alongside the rest of the world, but its decline is not quite as steep.
- Action, Sports, and Shooter genres are generally most popular, except for Japan, where Role-Playing dominates.
- Nintendo is the publisher to beat.

Medical Staffing Agency

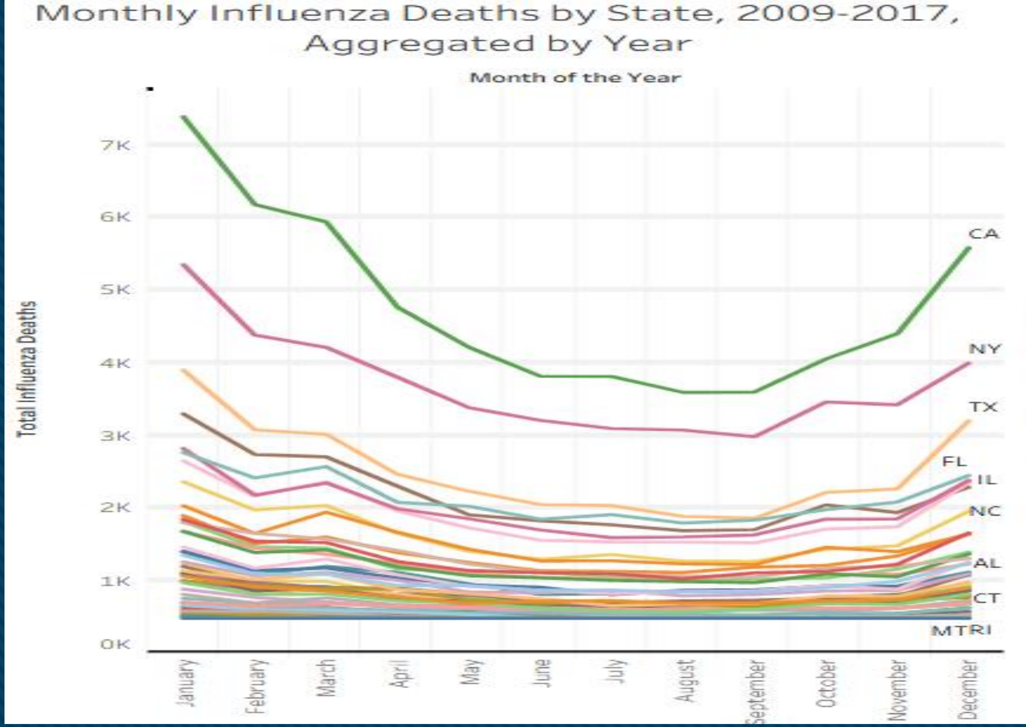
- **Objective**: Determine when to send staff, and how many, to each state, in preparation for the upcoming influenza season.
- Project Brief
- **Skills Used**: Data Integrity checks and Cleaning; Data Integration; Data Transformation; Statistical Hypothesis Testing; Visualization; Forecasting; Storytelling.
- **Dataset**: Influenza Death Statistics from the CDC.
- **Dataset**: Population Statistics from the U.S. Census Bureau.
- **Tools Used**:

Excel Tableau

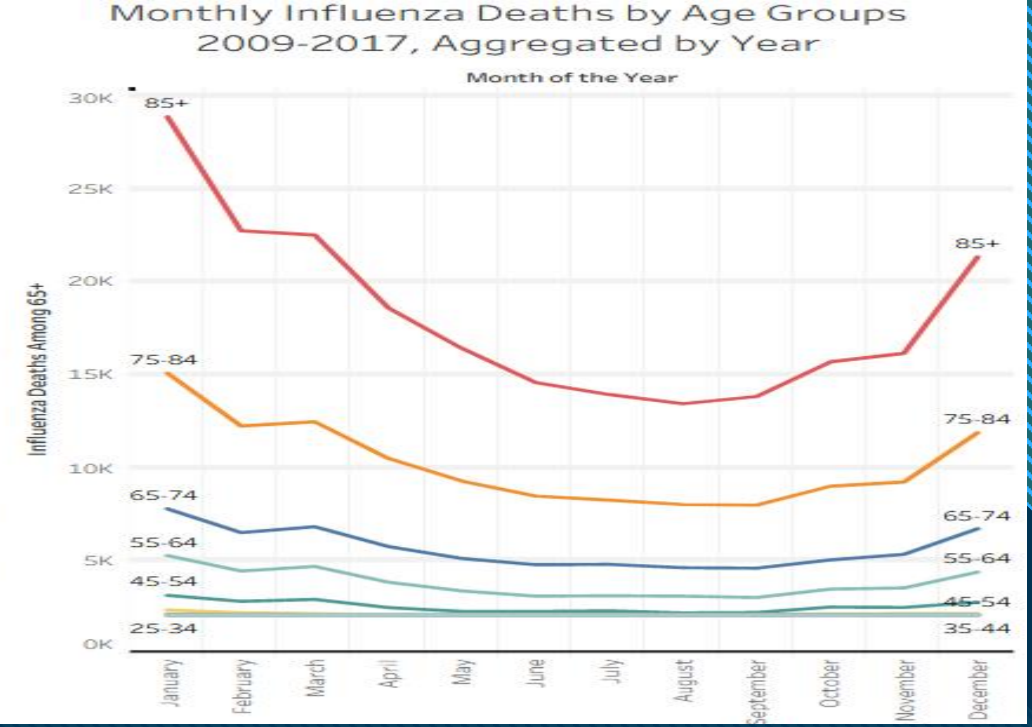
Analysis for 2009 – 2017: Seasonality, Ages of Influenza Deaths

Flu deaths generally rise across all states in October, peaking in January and tapering off in May.

Adults aged 65 years and over make up the majority of fatalities.



Monthly Influenza Deaths by State, 2009-2017, Aggregated by Year

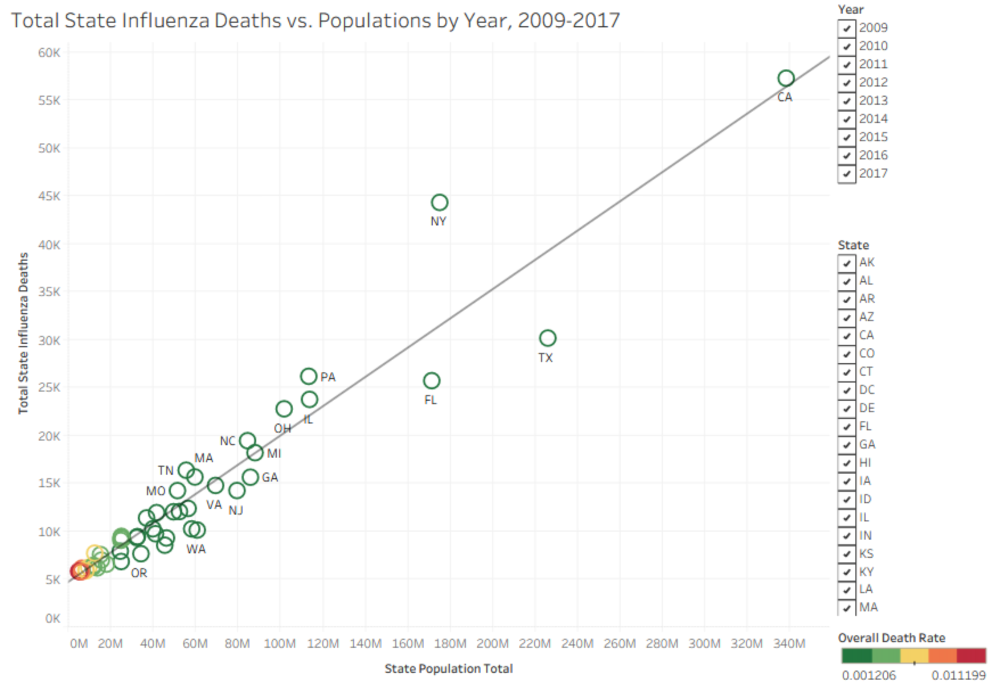


Monthly Influenza Deaths by Age Groups 2009-2017, Aggregated by Year

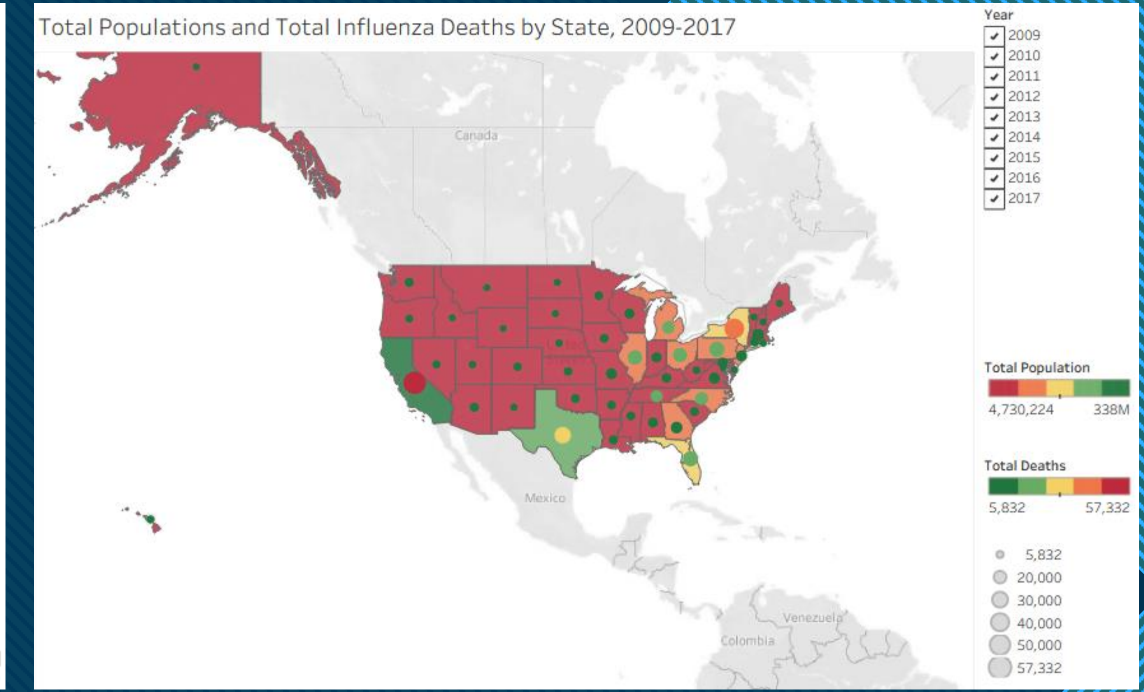
Analysis for 2009 - 2017: Geographic Distribution of Flu Deaths

Absolute counts of influenza deaths are highest in states with largest populations.

However, the death *rates* are higher among less-populated states.



Total State Influenza Deaths vs. Populations by Year, 2009-2017



Total Populations and Total Influenza Deaths by State, 2009-2017

Recommendations

Times & Locations for Staff Placement

- More staff should be sent to states with higher mortality counts, particularly California, New York, Texas, Pennsylvania, and Florida. Urgently investigate online markets.
- Staff should be in place by October.
- Link to Tableau Presentation: Here

Obtain More Data:

- Missing Data:
  - Agency staff numbers.
  - Required staff-to-patient ratios.
  - Past years’ staffing performance (over- vs. under-staffing).
  - Past years’ hospitalization rates.
- Research possible causes for inverse relationship between Total Population and Overall Death Rate.
- Conduct surveys, both during and after this season, to evaluate patient outcomes, as well as staff and patient satisfaction.

Rockbuster Stealth, LLC

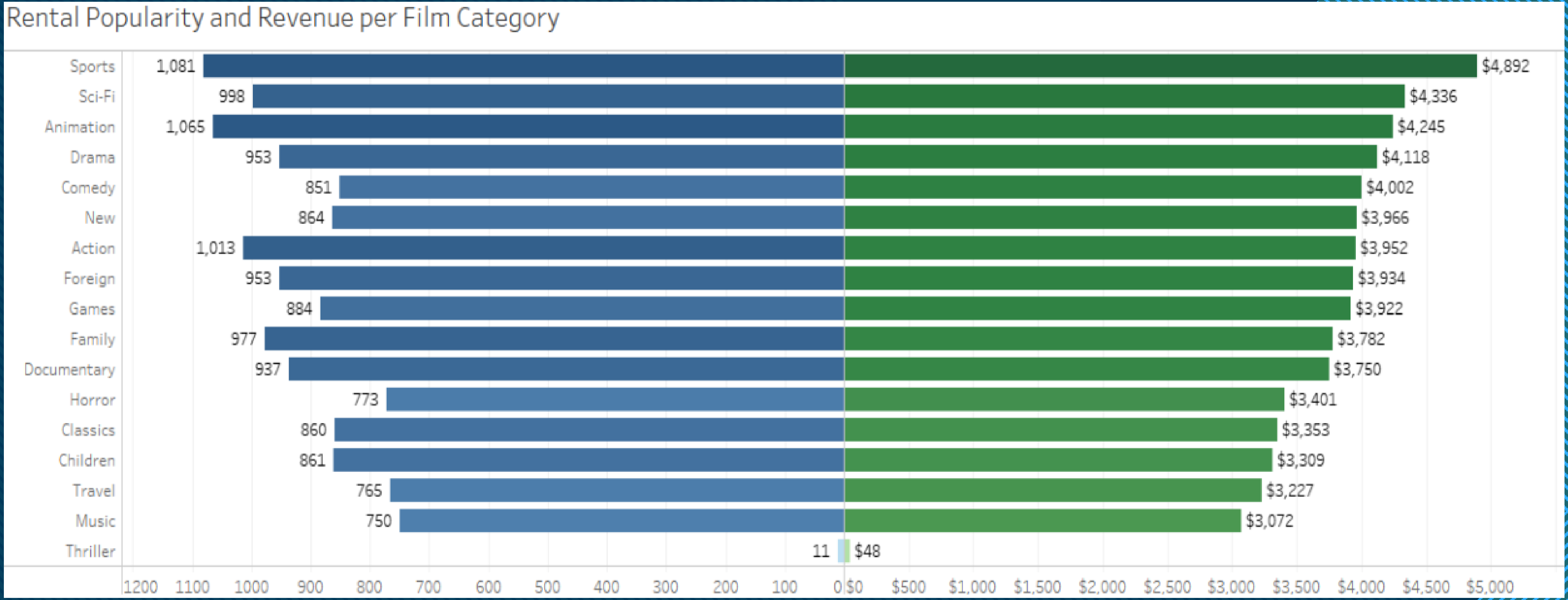
- **Objective**: Provide data-driven answers to a series of questions from executives; the answers will be used for their 2020 company strategy.
- Project Brief
- **Skills Used**: SQL Database Usage with PostgreSQL: Data Integrity checks and Cleaning; Data Filtering and Summarization; Joining Tables; Subqueries; Common Table Expressions; Views
- **Dataset**: Rockbuster dataset
- **Tools Used**:

PostgreSQL Excel Tableau PowerPoint

Analysis

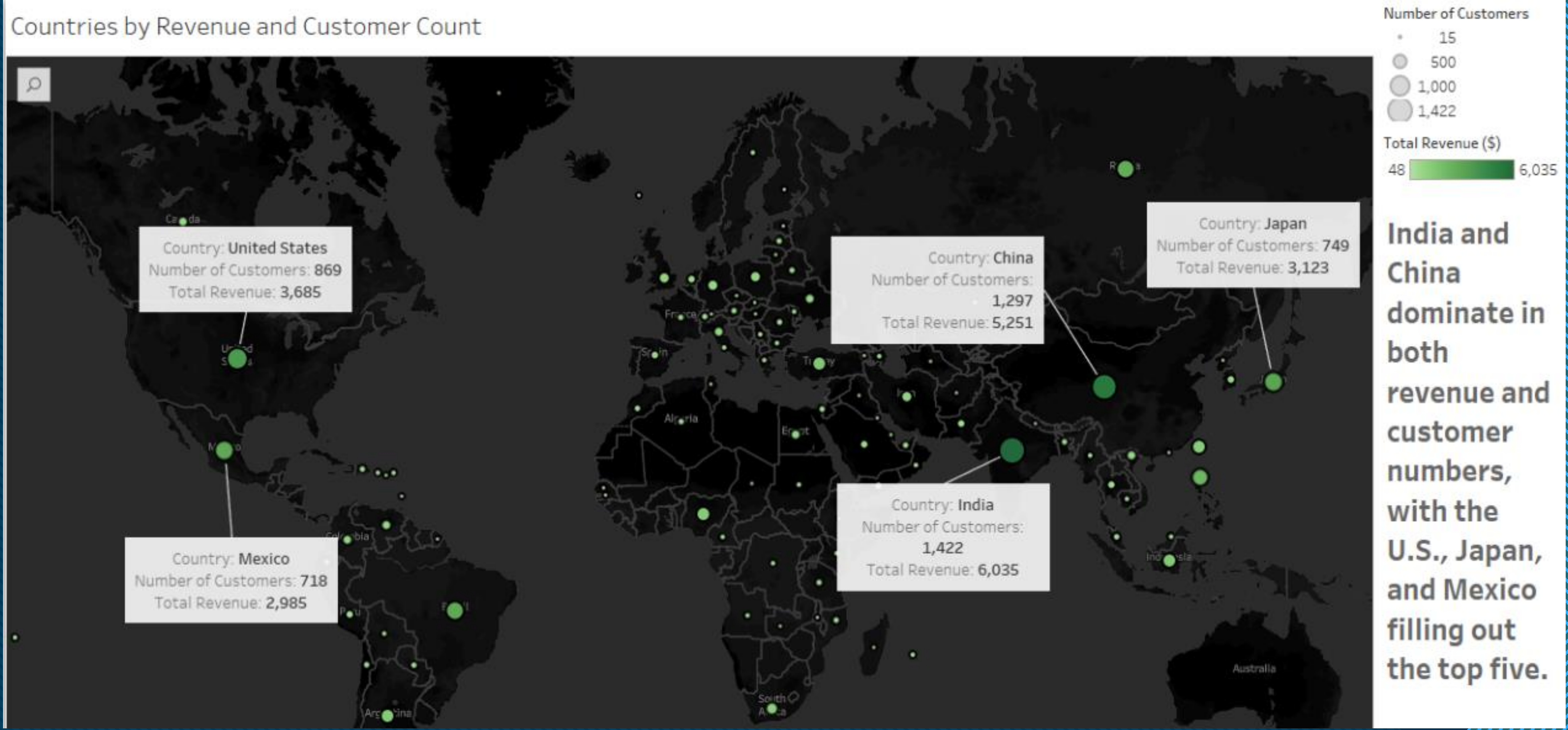
Sports, SciFi, and Animation are the Top Three Categories for both revenue and rental popularity.

The Thriller genre is strikingly unpopular.



Rental Popularity and Revenue per Film Category

Analysis



Countries by Revenue and Customer Count

Country: United States
Number of Customers: 869
Total Revenue: 3,685

Country: Mexico
Number of Customers: 718
Total Revenue: 2,985

Country: China
Number of Customers: 1,297
Total Revenue: 5,251

Country: Japan
Number of Customers: 749
Total Revenue: 3,123

Country: India
Number of Customers: 1,422
Total Revenue: 6,035

India and China dominate in both revenue and customer numbers, with the U.S., Japan, and Mexico filling out the top five.

Recommendations

- Invest in more titles in popular categories, particularly Sports, Sci-Fi, and Animation.
- Focus on countries with highest revenues and greatest numbers of customers, particularly India and China, but also the U.S., Mexico, and Japan.
- A customer refer-a-friend program might boost membership.
- Bulk discounts, e.g. rent one, rent a second at half price, might boost revenues among existing customers.
- Survey customers about why they choose Rockbuster.
  - Gather detailed demographic information, e.g. age.
  - Ask whether they would use referral programs and bulk rental discounts.

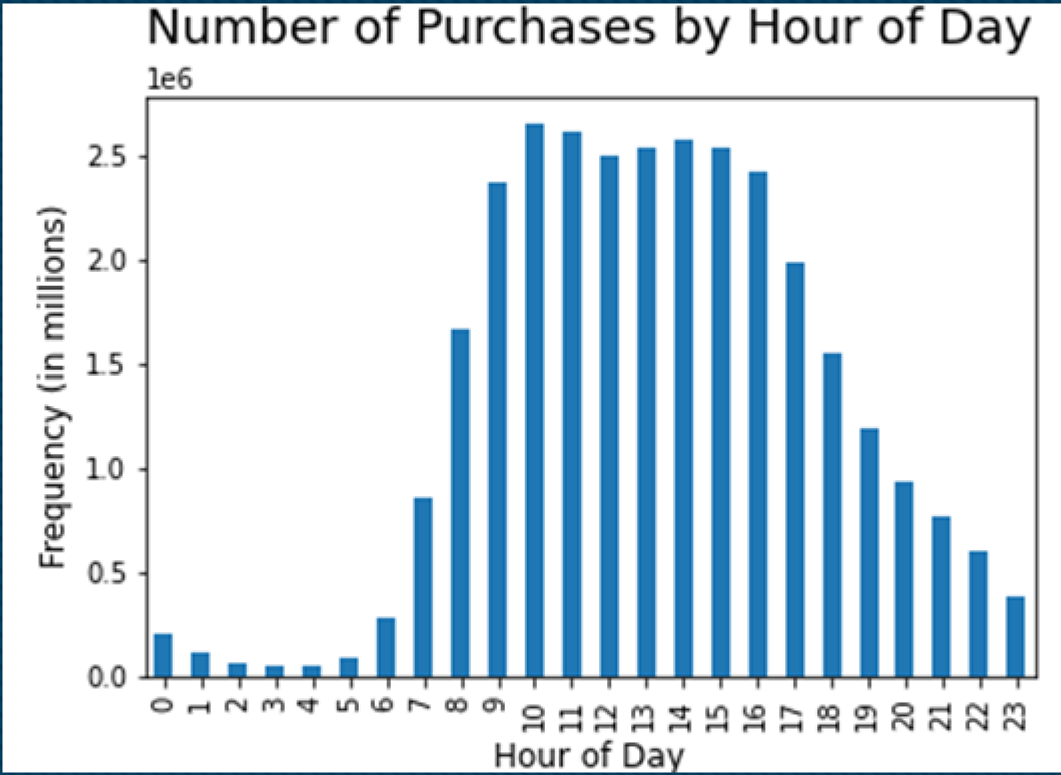
- Link to GitHub Repository: Here

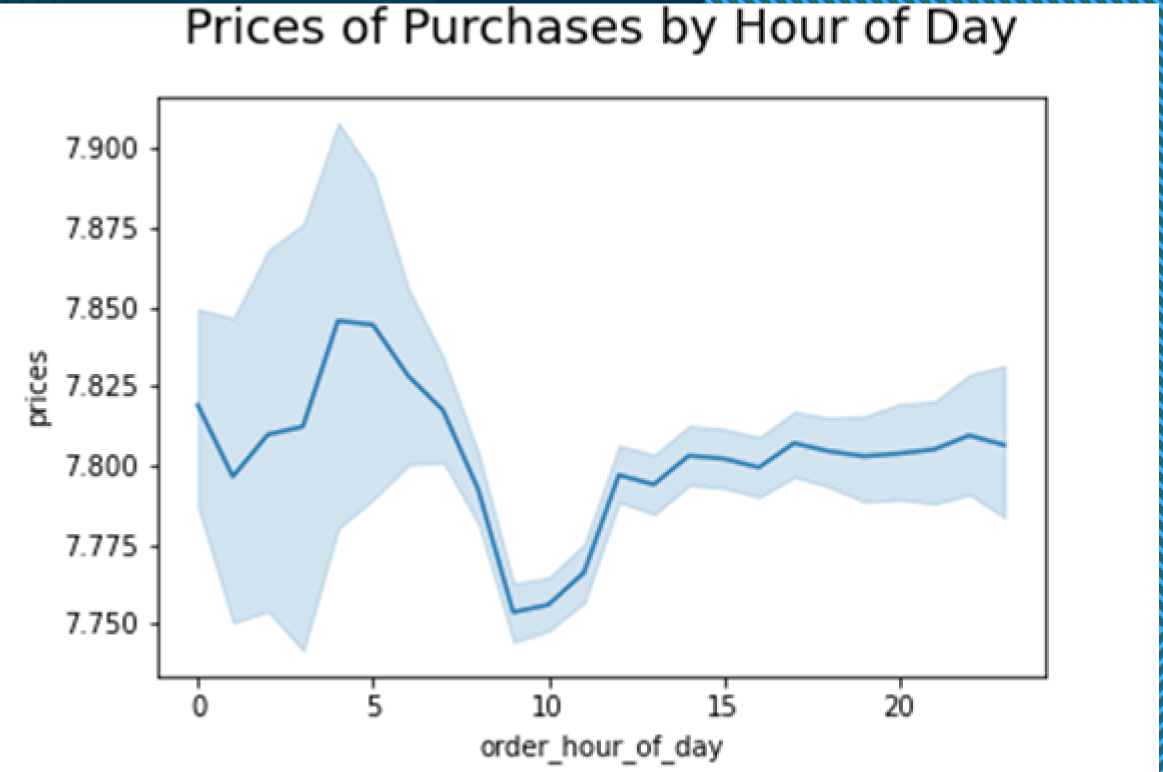
Instacart

- **Objective**: Perform an initial exploratory analysis of some of Instacart’s data, in order to derive insights and suggest strategies for better segmentation based on the provided criteria.
- Project Brief
- **Skills Used**: Data Analysis with Python (particularly pandas, numpy, matplotlib, seaborn, and scipy) and Jupyter: Data Integrity checks and Cleaning; Data Merging and Wrangling; Deriving variables; Grouping and Aggregating data; Visualizations; FOR loops; Iteration with itertools.
- **Dataset**: Customer Dataset
- **Dataset**: Data Dictionary
- **Tools Used**: Python Excel

Analysis

The busiest hours are generally 9am to 4pm. However, purchases made between 4am and 6am have significantly higher prices, albeit with a much broader price range.


Number of Purchases by Hour of Day

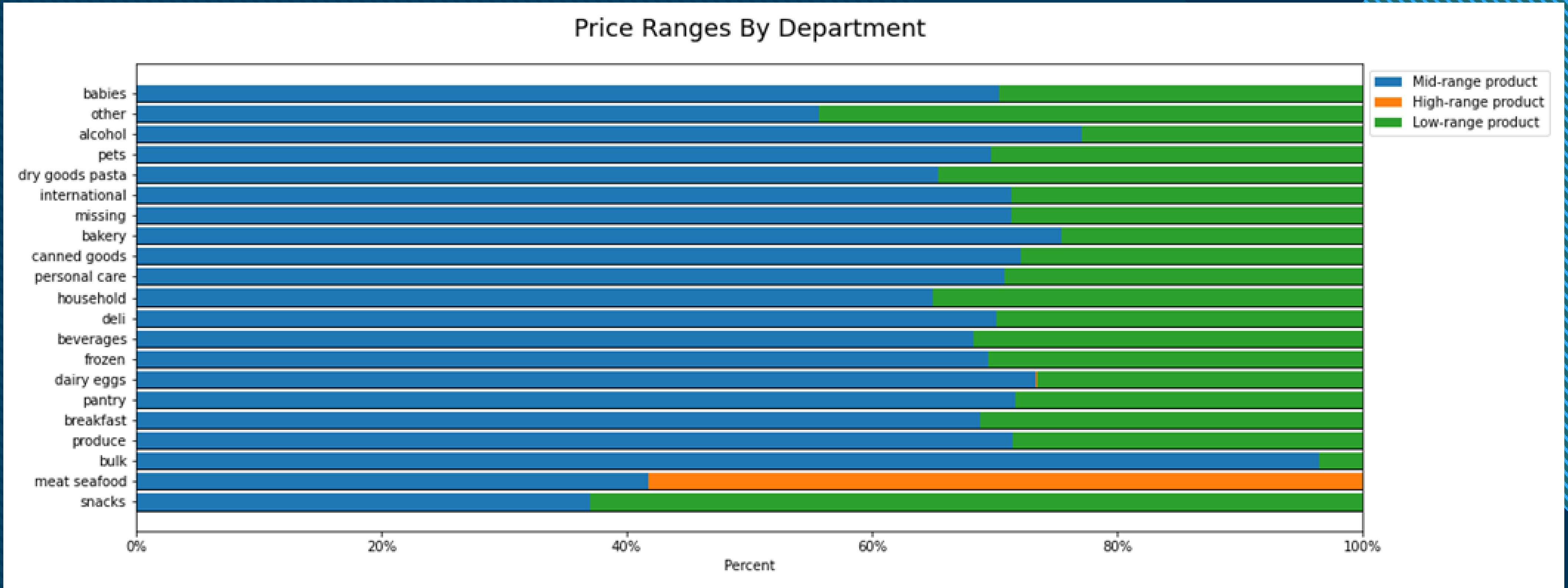

Prices of Purchases by Hour of Day

Analysis

Prices have been divided into three ranges:

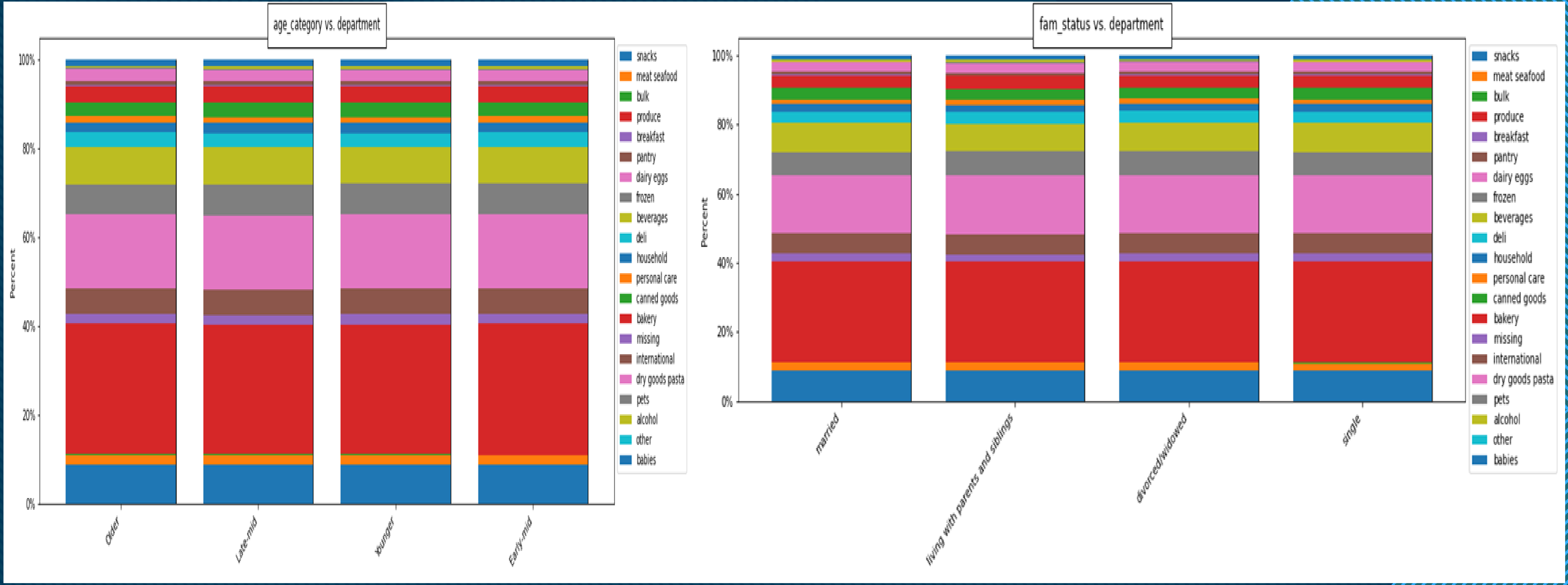
Low-range products: <=$ 5 Mid-range products: > $5 & <= $15

High-range products: > $15



Analysis

There is essentially no group classification of appreciable size suggested by purchasing behavior; comparing the different options for a given demographic against each other shows nearly identical patterns to those seen in the previous two questions.



Recommendations

Pricing & Promotions

- The expensive early-morning purchases may be driven by urgency or impulse. People may be more willing to pay more for products at those hours out of desperation, e.g. if other stores are not open.
- Scheduling ads on Tuesday (Day 3) or Wednesday (Day 4) would target the least busy days, and scheduling ads between 2am and 4am would target the least busy hours. However, it may be better to target Thursdays and Fridays, and 6am to 7am hours, to catch people just before the busier days and times, when they are most likely to be on the verge of making purchases already: they may be more receptive to the ads.

Obtain Real Data:

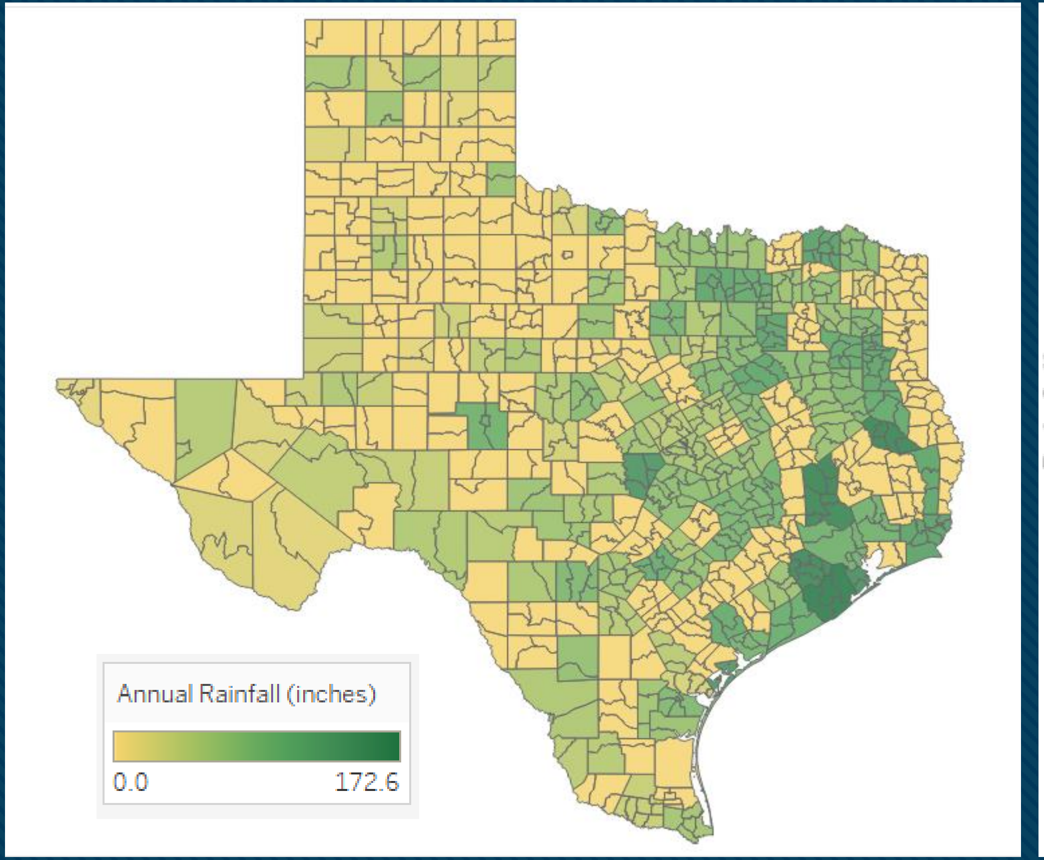
- The data is nearly uniform across all demographics: its artificial nature makes true insights nearly impossible.
- Data should be gathered on cost (and therefore profit margins) of products.
  - Gain better context.
  - Understand whether less-popular categories may be worth popularizing due to higher margins.
- Link to GitHub Repository: Here
- Link to Client Report: Here

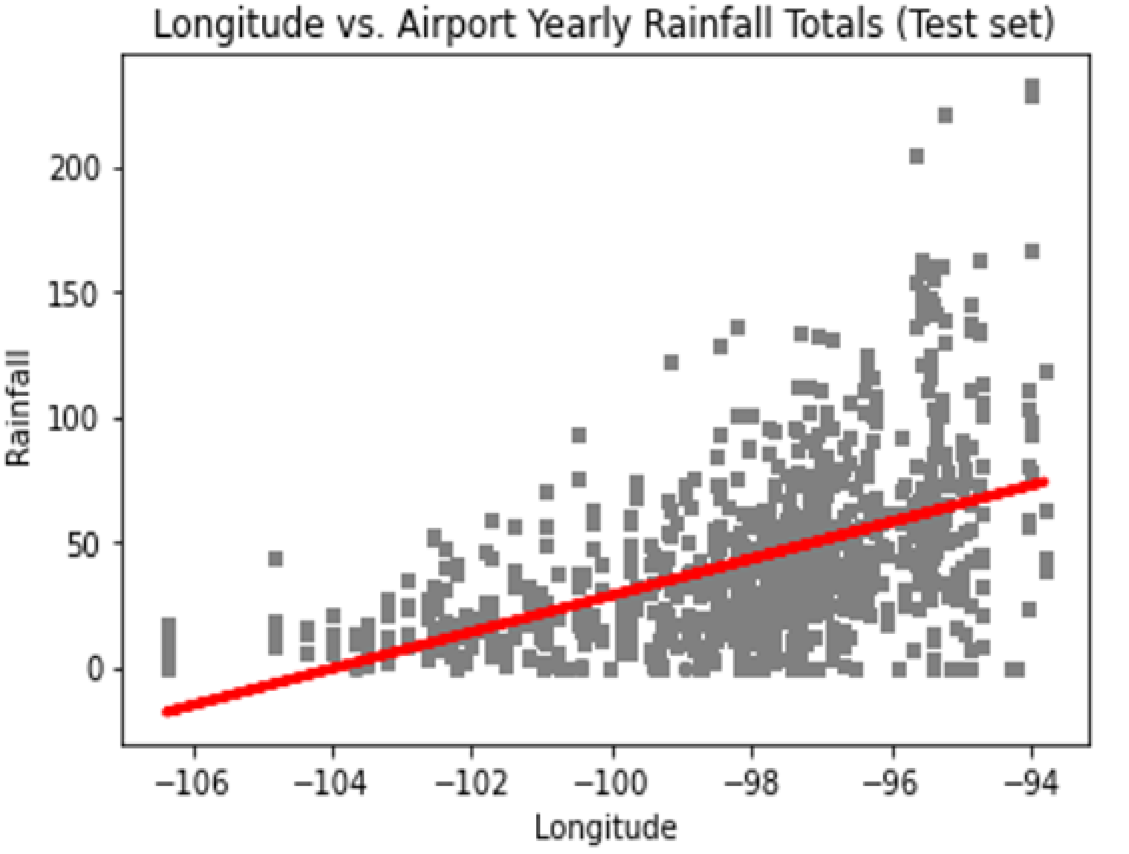
Rain in Texas

- **Objective**: Perform an initial exploratory analysis of chosen dataset from Kaggle, in order to derive insights and on trends and relationships in weather in the U.S. State of Texas.
- Project Brief
- **Skills Used**: Data Analysis with Python (particularly pandas, numpy, matplotlib, seaborn, scipy, scikit-learn, quandl, and folium) and Jupyter: Data Integrity checks and Cleaning; Data Merging and Wrangling; Deriving variables; Grouping and Aggregating data; Visualizations; Geospatial Analysis; Linear Regression; k-means Clustering; Time Series Analysis.
- **Dataset**: Primary Weather Dataset from Moosavi, et al. (2019)
- **Dataset**: Average Temperature in Texas by Year
- **Tools Used**: Python Tableau

Analysis

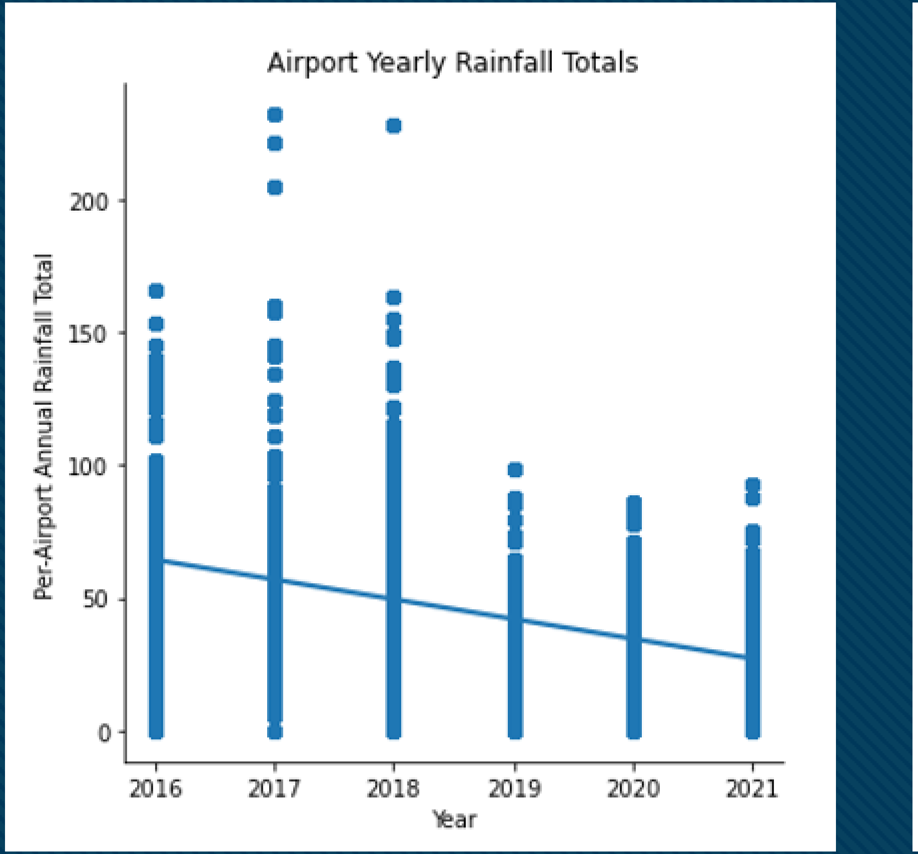
Total rainfall amounts generally increase towards the eastern part of the state and the Gulf of Mexico, and away from the Chihuahuan Desert. However, the relationship is not linear.

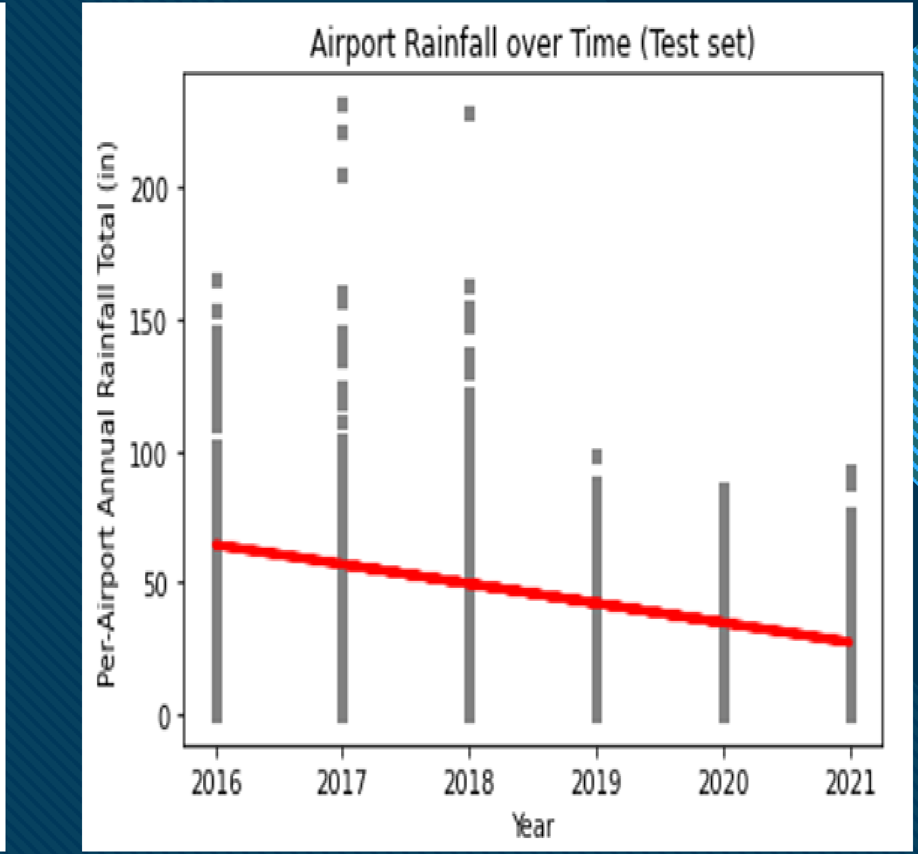




Analysis

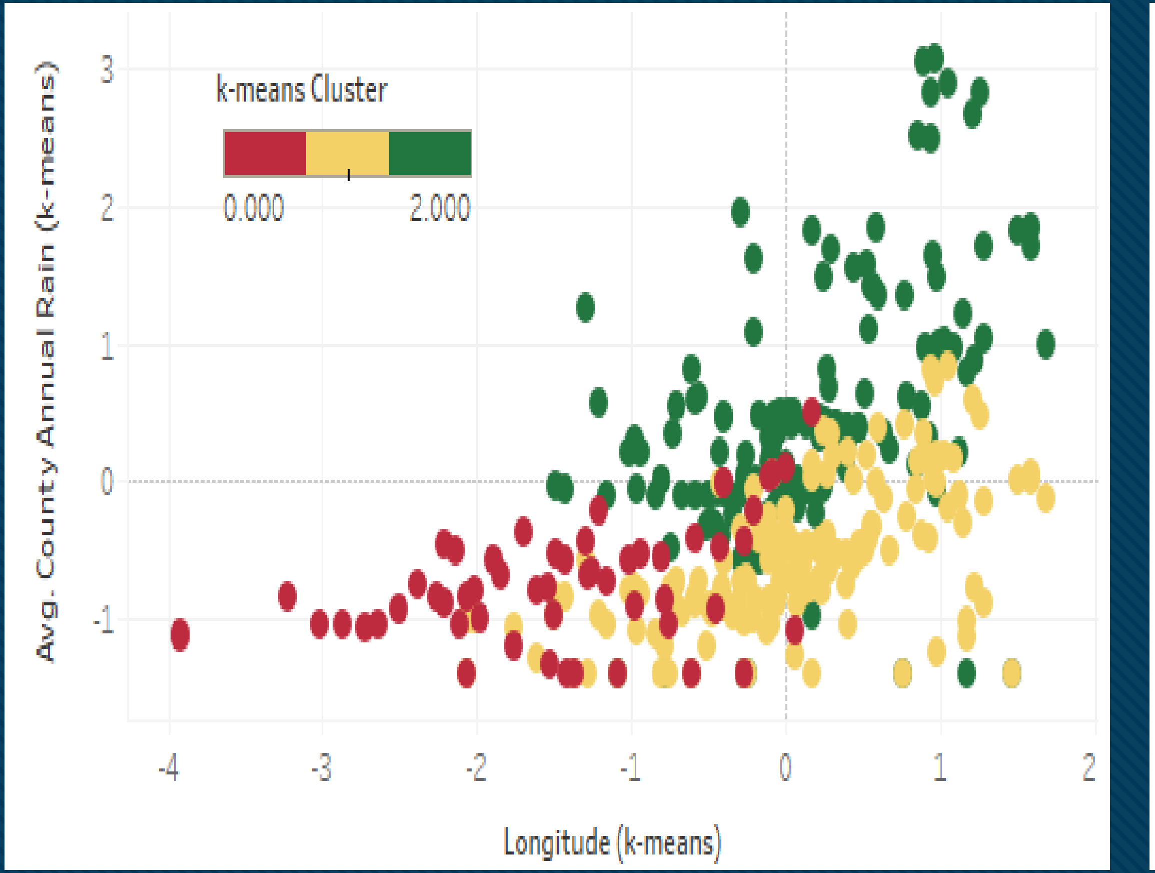
Rainfall totals reported at airports appear to be decreasing over the 2016-2021 time period. However, this relationship is not linear, either.

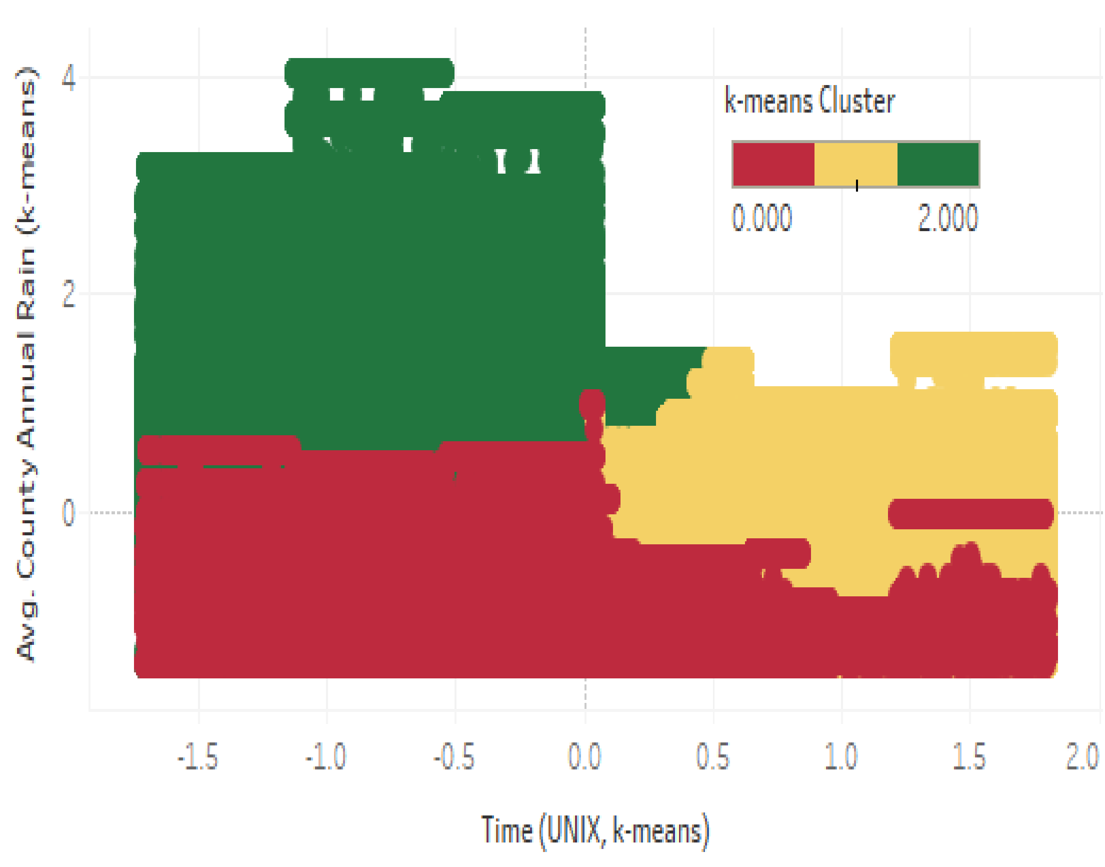




Analysis

We found three clusters of data points. The two larger clusters, colored Green and Yellow here, overlap greatly in longitude, but show near-total separation in the times data were collected, as well as some separation in precipitation amounts.





Findings & Next Steps

Findings & Discussion

- While relationships between longitude, rainfall amounts, and time appear to exist, they are non-linear.
- The primary dataset contains continuous measurements solely for precipitation; data for other climatic dimensions such as wind speed and temperature are categorical and extremely limited in scope.
- Unexpected chronological separation of clusters with strong longitudinal overlap raises questions of data integrity.

Next Steps

- Additional data sources should be analyzed with the following goals:
  - Gain better context.
  - Perform multivariate analysis to explain non-linear relationships.
  - Corroborate or discount the data from the original data source.

- Link to GitHub Repository: Here
- Link to Tableau Presentation: Here

Thank You