

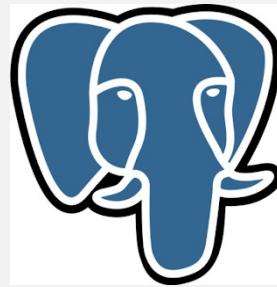
MICHELLE DIAZ

Data Analyst

Portfolio

San Francisco, CA

TECHNICAL SKILLS



SOFT SKILLS – CRITICAL THINKING, COMMUNICATION, PROBLEM SOLVING, GROWTH MINDSET, COLLABORATION

PROJECTS

-  **1. NWWs** : Slides 4-7 – Wastewater Surveillance for Sars-Cov-2
-  **2. Influenza** : Slides 8-11 – Allocate staffing resources in disproportional areas
-  **3. Game Co** : Slides 12-15 – Video game sales analysis
-  **4. Rock buster** : Slides 16-19 – Video rental performance analysis
-  **5. Instacart** - Slides 20-24 – Sales and marketing strategic analysis



WASTEWATER SURVEILLANCE – OBJECTIVE

With the rampant increase of COVID 19 antigen rapid tests, we are unable to track the virus within our community. By analyzing wastewater samples, we can identify how prevalent the virus is within the community and stop the propagation of the virus. More cases of COVID 19 in the community are associated with increased levels of SARS-CoV-2 in wastewater, meaning that data from wastewater analysis can be used as an indicator of the level of transmission of COVID 19 in the community.

- We can partner with Community Based Organizations to support the community. This data can provide an early warning of COVID-19's spread in communities

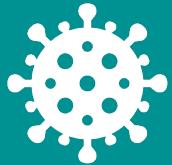
DELIVERABLES

[Tableau Dashboard](#)

[Data](#)

[Executive Summary](#)

[GitHub](#)



WASTEWATER SURVEILLANCE – OVERVIEW

DATA



- [National Wastewater Surveillance data](#)
- [Clean data set](#)

SKILLS

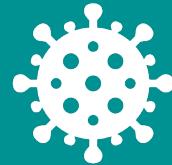


- Conceptualize time series data
- Created categorical plots
- Perform exploratory analysis
- Tested linear regression
- Dashboard creation in Tableau.

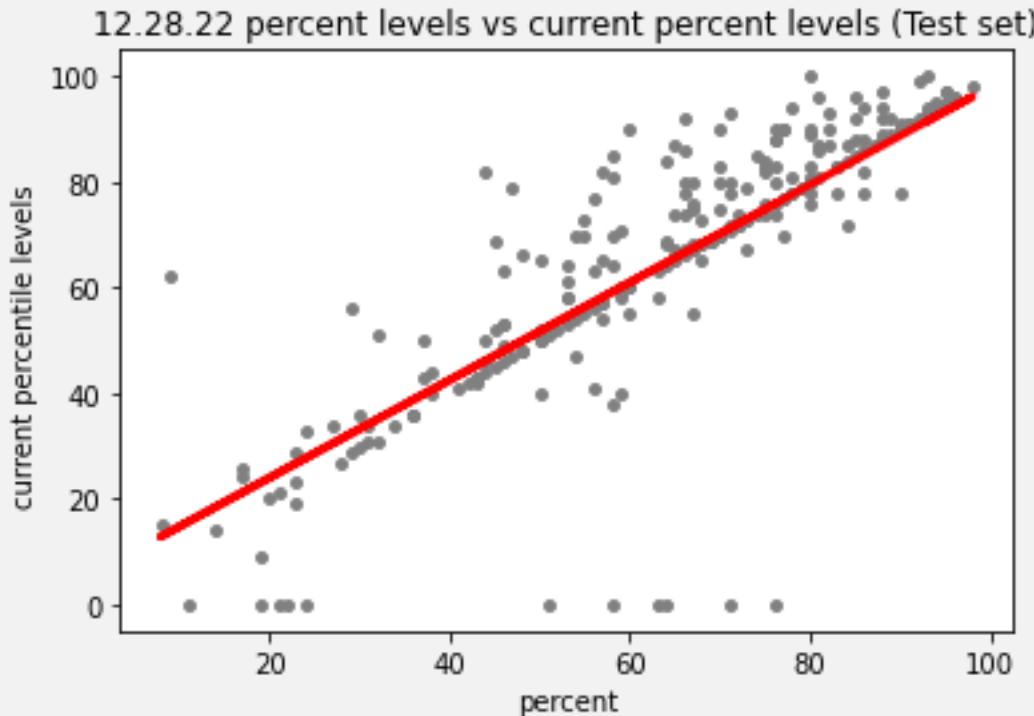
TOOLS



- Python
- Anaconda
- Jupyter Notebook
- Pandas
- NumPy
- Seaborn,
- Matplotlib
- SciPy
- Folium
- GitHub
- Tableau



WASTEWATER SURVEILLANCE – LINEAR REGRESSION AND CATEGORICAL PLOT

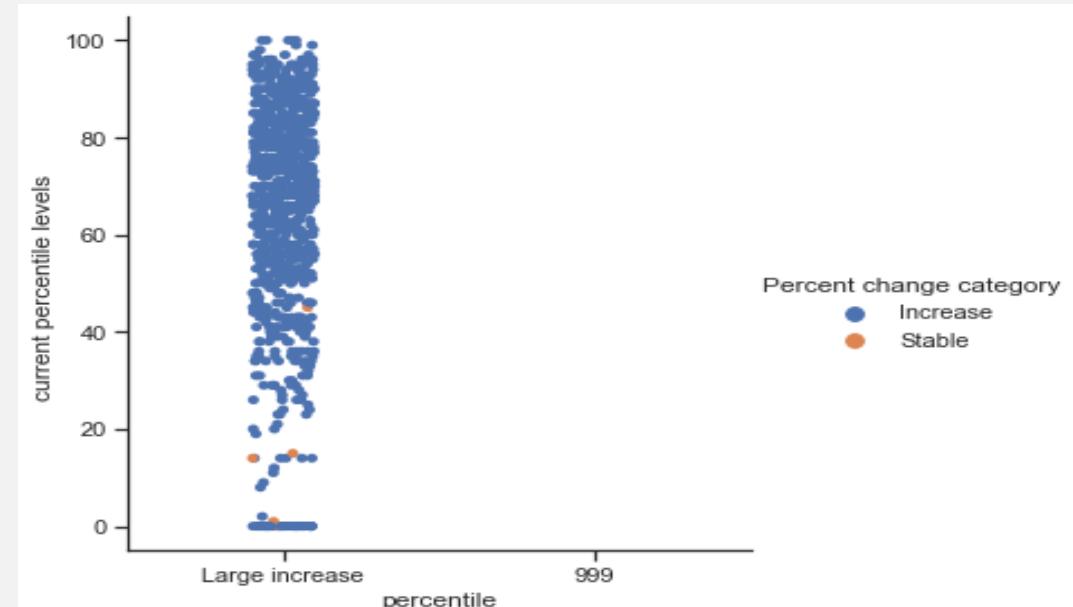


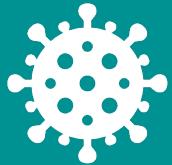
Linear Regression:

- Tested the hypothesis that post holiday's would see an increased in virus levels and determined that there are many points that fall beyond the regression line and there is a high density of data points on a higher percentile

Categorical plot:

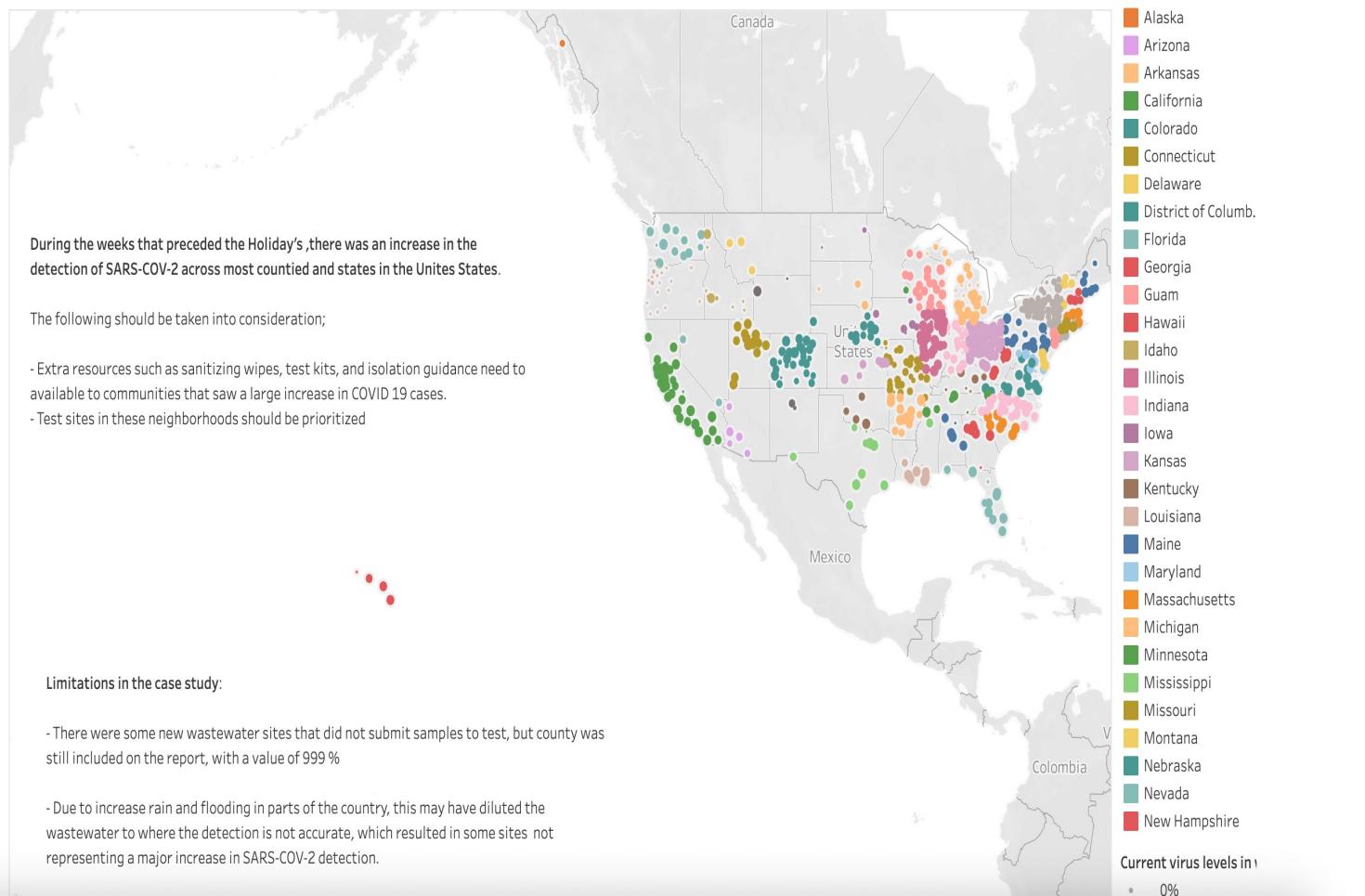
A **categorical plot** splits the data for the Percent variable. We can then compare the groups to uncover new patterns and noticed increase virus levels across all counties was prevalent, compared to the sites with stable virus levels.





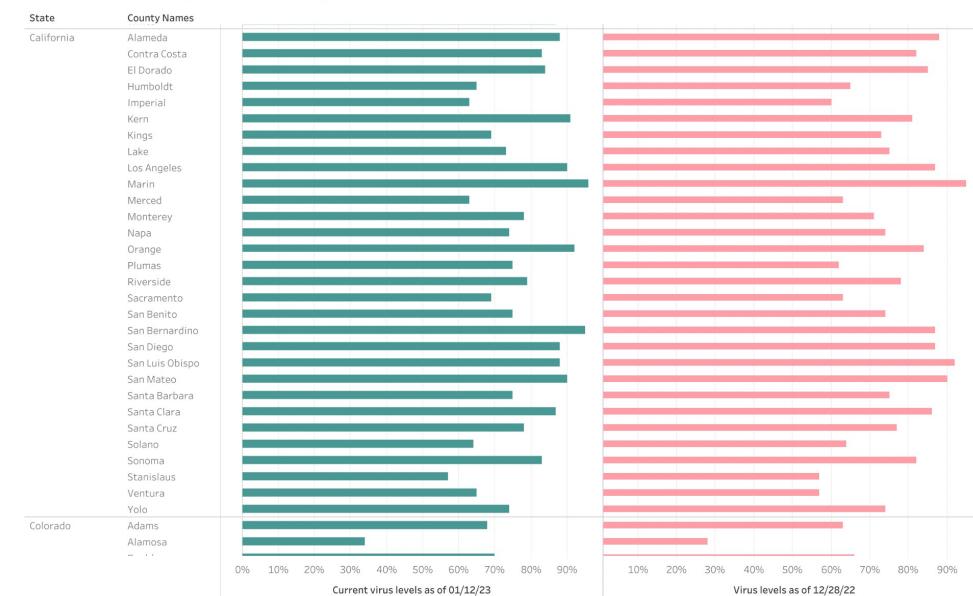
WASTEWATER SURVEILLANCE – ANALYSIS

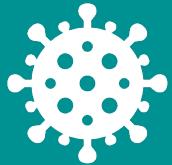
National Waste Water Surveillance



When comparing the virus levels by counties, there was an increase from 12/28/22 and 01/12/23.

Difference in levels by State and County





WASTEWATER SURVEILLANCE – KEY INSIGHTS AND RECOMMENDATIONS

KEY INSIGHTS

- States and counties that participated in the Wastewater Surveillance program saw an increase in virus levels of SARS-CoV-2 after the holidays (12/28/22).
- There were less wastewater surveillance sites in rural areas of the U.S.
- For sites with that were categorized having stable virus levels, meant that there were no fluctuations in the sample.
- Due to inclement weather patterns, such as storms and flooding, this diluted wastewater samples and therefore an accurate read of virus levels will need to be captured.

RECOMMENDATIONS

- Continue to monitor the wastewater surveillance metrics for the period of 01/25/23.
- Follow-up on the New Sites and observe if there are any fluctuations in that region.
- Analyze the impact of rain in parts that experienced major flooding on the next collection date.
- Communities can track the presence of SARS-CoV-2, the virus that causes COVID-19, in wastewater samples. This data can provide an early warning of COVID-19's spread in communities.



INFLUENZA – OBJECTIVE

Assist staffing agencies allocate healthcare personnel based on location and mortality rate. Analyzing historical influenza illness mortality rate across the United States, merging datasets with population data by geography. Making predictions for staffing, based on relationships between variable

DELIVERABLES

[Management plan](#)

[Dashboard](#)

[Interim report](#)

[Presentation](#)



INFLUENZA – OVERVIEW

DATA



- [CDC](#) data set
- [Census](#) data set
- [Influenza Visits](#) data set
- [Lab Tests](#) Data Set

SKILLS



- Conceptualize business requirements & design a research project
- Data transformation & integration
- Statistical hypothesis testing & forecasting
- Create charts, graphs, statistical visualizations, spatial analysis in Tableau.

TOOLS



Excel

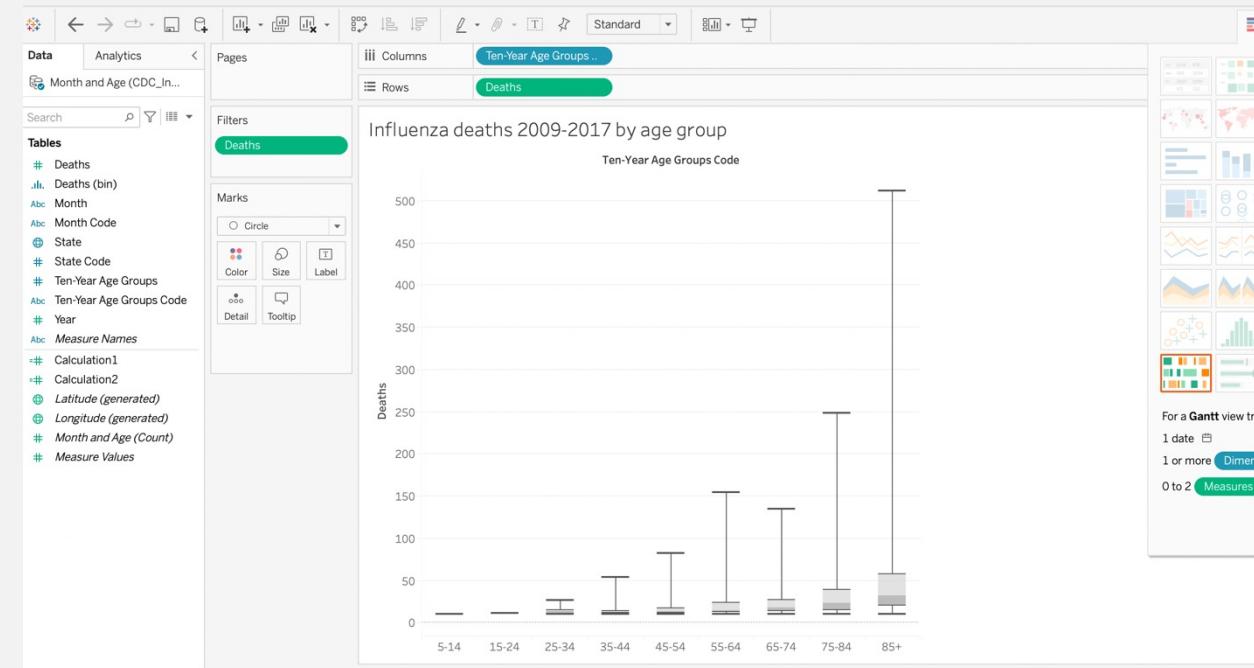
- Pivot Tables
- Formulas

Tableau

- Charts, graphs, tree-maps, bubble charts, scatterplots, spatial analysis



INFLUENZA – FORECASTING

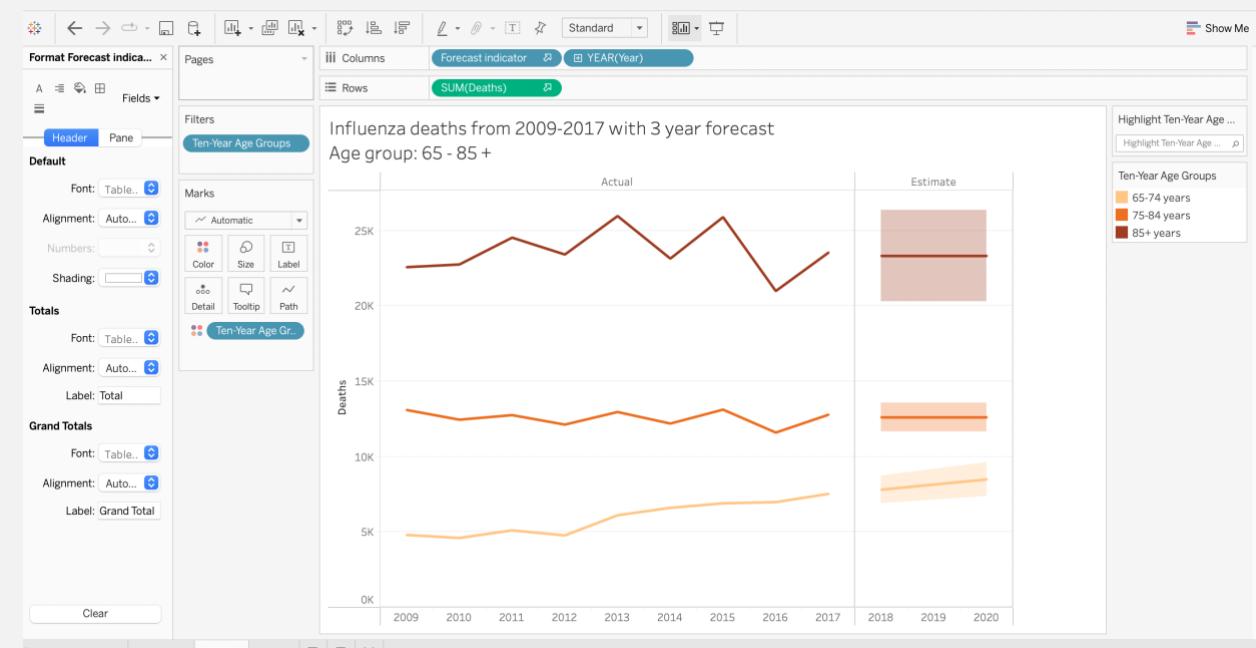


Statistical Visualization:

- Box and Whisker helped identify the influenza mortality rate by age.

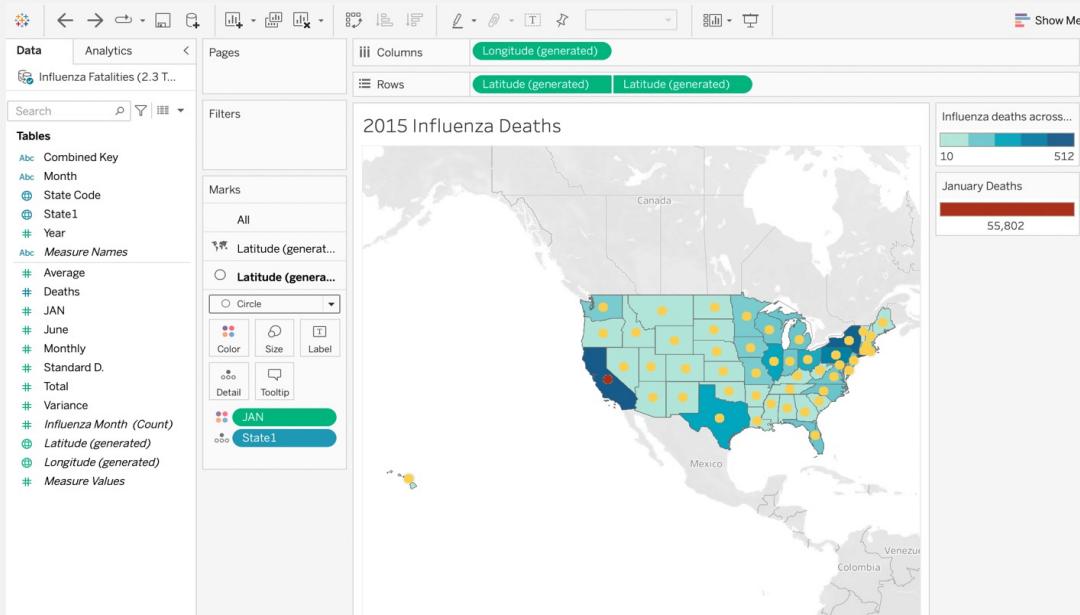
Temporal visualization and forecasting:

- 3 year forecast of influenza mortality rate for those 65 and older.





INFLUENZA – ANALYSIS

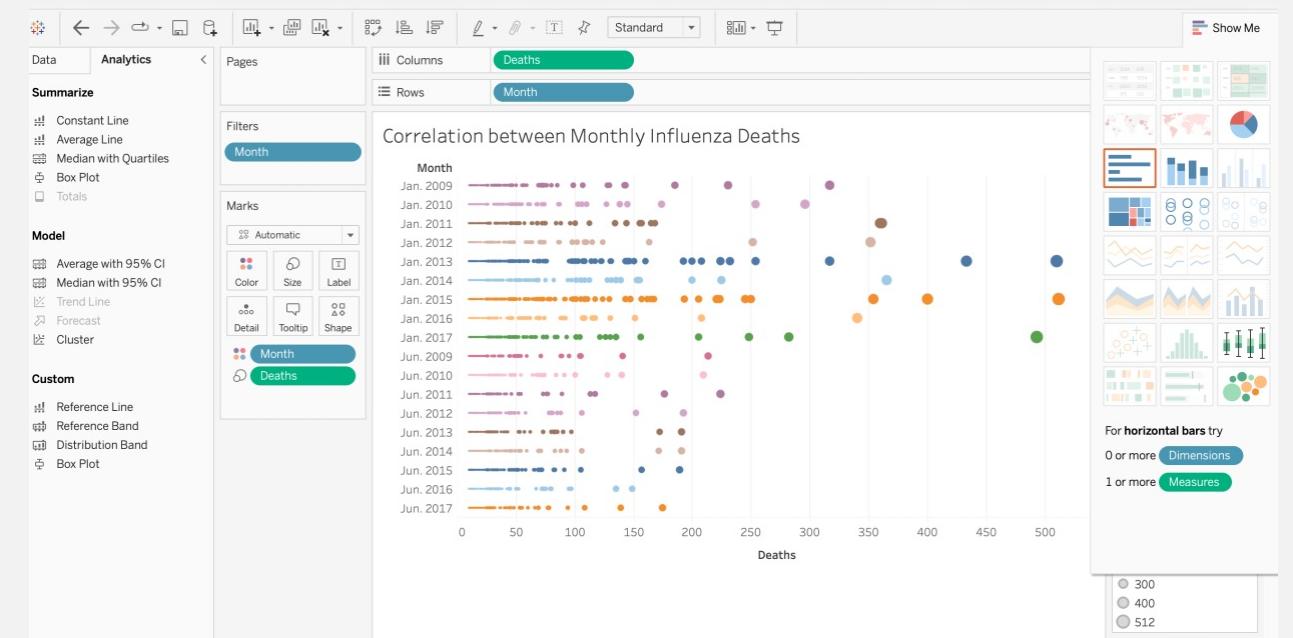


Spatial trends:

- California has the highest deaths rates at 512 and Wyoming has the lowest with only 10.
- A tree map identified data for all 50- states that have the highest mortality.

Bubble Chart and Scatterplot

- Correlation coefficient was 1.when comparing Influenza mortality rates in January to June rates.





INFLUENZA – KEY INSIGHTS AND RECOMMENDATIONS

KEY INSIGHTS

The most affected population by influenza deaths:

- 85+ year-olds

The top 5 states with the highest influenza deaths:

- California, New York, Texas, Pennsylvania, and Florida

Months with highest mortality rate:

- December – March

RECOMMENDATIONS

- Allocate extra staff during the winter surge for the states that have the highest mortality rate.
- Healthcare facilities should prepare for the winter surge: December – March
- Partner with Local Health Jurisdictions to determine what cities need health care workers during Flu season.
- Continue to monitor the staff capacity during the winter surge to ensure that hospitals have adequate staff support



GAME CO. - OBJECTIVE

Game Co. is planning a marketing budget for 2017 and is assuming all sales have remained constant from 2016.

The goal is to analyze the sales trend to see if they have remained consistent. This will determine whether the marketing budget will need to be distributed and allocated across all regions in efforts to maximize the return on investments.

DELIVERABLES

- [VG Sales](#)

[Clean VG Sales](#)

[Presentation](#)



GAME CO. – OVERVIEW

DATA



- [VG Sales](#) data set
- [Clean VG Sales](#) data set
- Sales history
- Market share since 1980
- Regional Analysis
- Target: Top genres and platforms by region
- 16.5k titles

SKILLS



- Data integrity, quality, and accuracy
- Cleaning, grouping, and summarizing data
- Addressing missing values
- Conduct statistical analysis
- Visualize and deliver insights

TOOLS



Excel

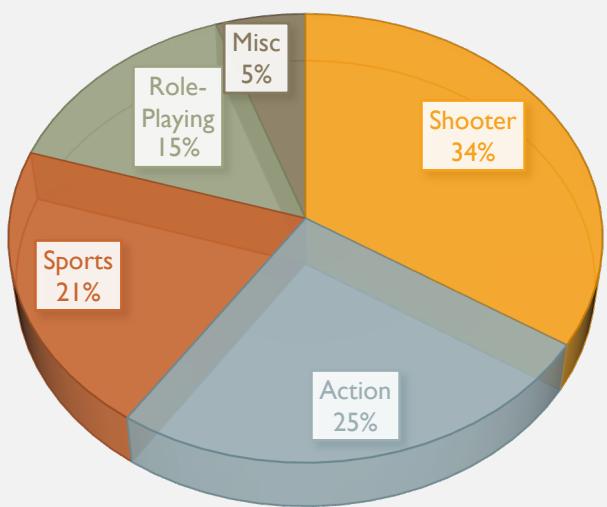
- Pivot Tables
- Formulas
- Charts
- Graphs



GAME CO – REGIONAL ANALYSIS

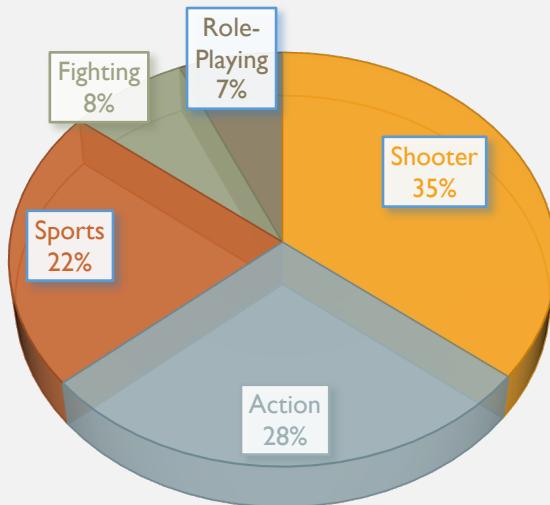
1. Shooter
2. Action
3. Sports
4. Role-Playing
5. Fighting

2015



North American Sales by genres

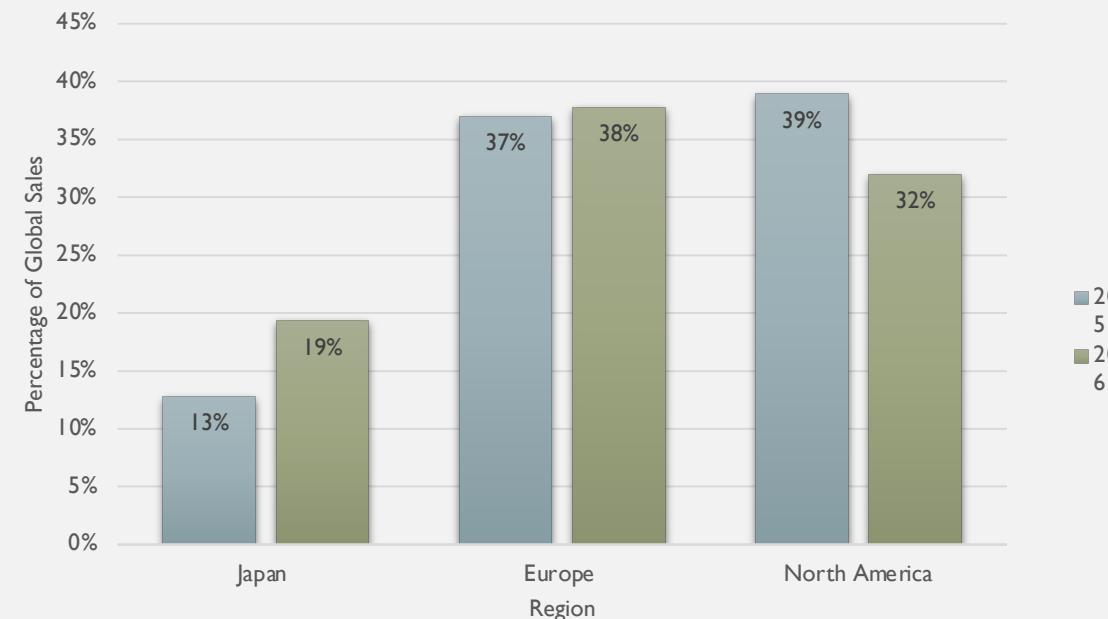
2016



Global Sales:

We observed North America having the highest sales we can see that Europe's sales are surpassing North America's sales in the past year (at 38% of sales Globally).

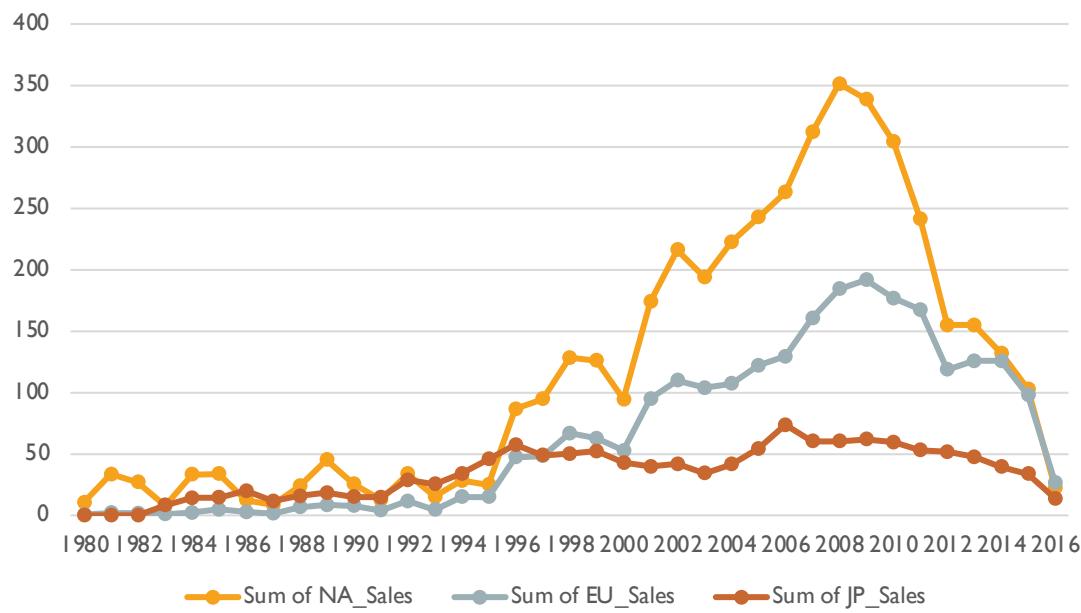
Sales By Region For 2015 vs. 2016





GAME CO. – ANALYSIS

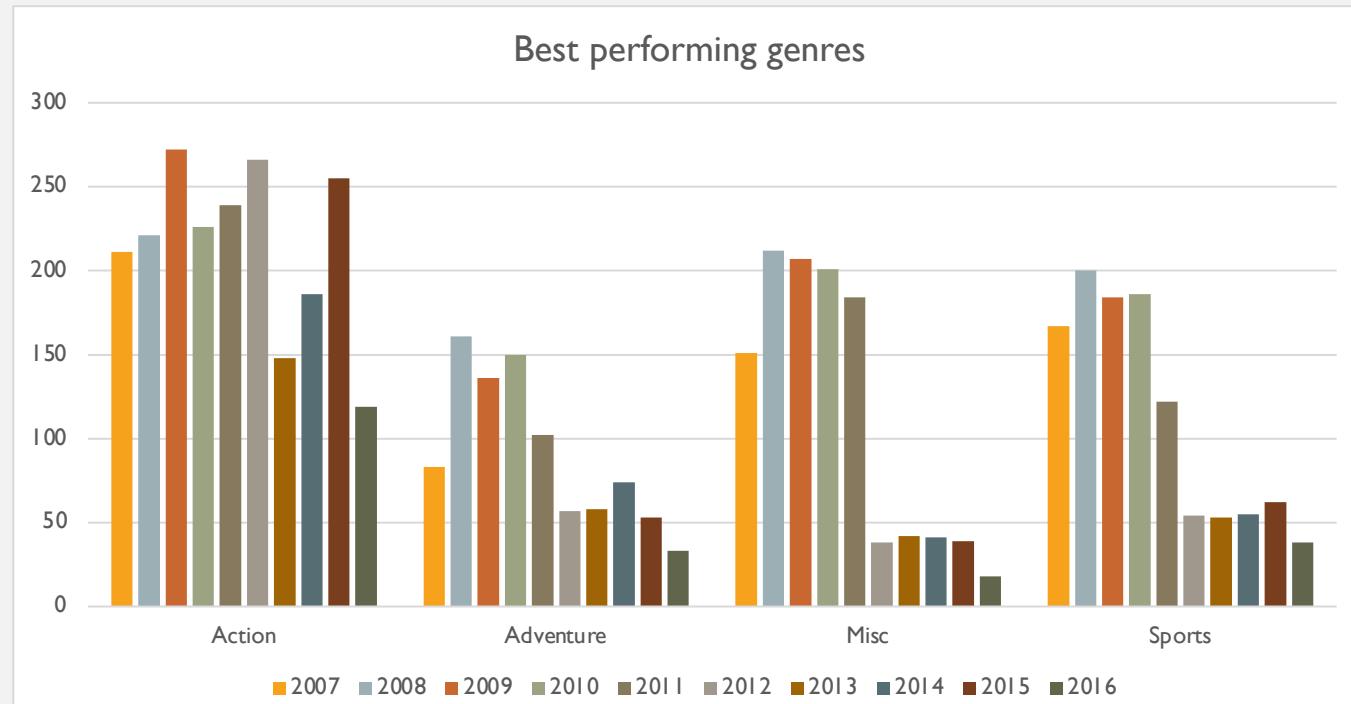
1980-2006 Sales by unit



Time Series Analysis:

2016 had an unforeseeable downturn in sales. As we can see from the 2015 comparisons, there was a substantial decline in sales in 2016. This data suggest that sales have not remained the same over time.

Genres: Action, Racing, Shooter, Sports and Simulation performed poorly. Perhaps allocating some resources to help market this category will help boost sales units for the genres listed.





GAME CO. – KEY INSIGHTS AND RECOMMENDATIONS

KEY INSIGHTS

Most popular genres & platforms in the 3 biggest markets – North America, Europe, and Japan

- In 2015, European sales ranked North American sales for the first time since 1980 and was trending upwards
- Japan sales has an increase since 2015
- The trends for years 2014, 2015 and 2016 in genres and platforms do not alter our understanding of their respective popularity

RECOMMENDATIONS

- Allocate the marketing budget for 2017 based on the most recent successful genre for each region
- Allocate the budget accordingly: EU: 38% ; NA 32%, and 19% to Japan
- Invest in further research to understand WHY are global sales decreasing rapidly and what the next trend might be



ROCK BUSTER STEALTH OBJECTIVE

Rockbuster is a movie rental service. The management team is seeking insights in efforts to make effective business decisions according to the following variables; titles, geographies, popular, genres etc. Their database was analyzed using SQL, performing complex queries such as CTEs, subqueries, and inner joins. Data Dictionary and Entity relationship diagram were created during the process.

DELIVERABLES

[Data dictionary and queries](#)

[Presentation](#)

[Spatial Analysis](#)



ROCK BUSTER STEALTH

DATA



- [Rockbuster](#) dvd rental data set
- [Final Analysis](#)

SKILLS



- Understand relational databases, storage and structures
- Postgres
- DB Visualizer ERD
- Conduct SQL commands
- CTE – common table expressions
- Subqueries, Join, Filter, Order, and summarize
- Visualizations in Tableau

TOOLS



- Excel
- SQL
- DB Visualizer
- Tableau



ROCK BUSTER STEALTH

TOP 10 Countries

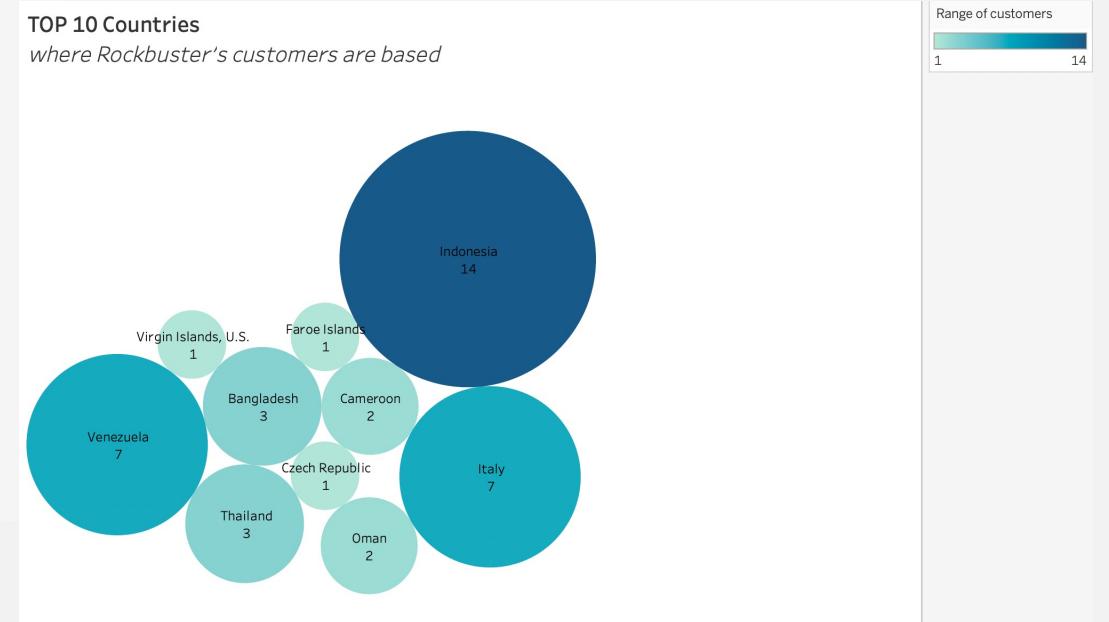
where Rockbuster's customers are based



Country
Bangladesh
Cameroon
Czech Republic
Faroe Islands
Indonesia
Italy
Oman
Thailand
Venezuela
Virgin Islands, U.S.

- Data cleaned, using CRUD method
- Used SQL commands to filter, group and summarized to determine key insights

TOP 10 Countries
where Rockbuster's customers are based



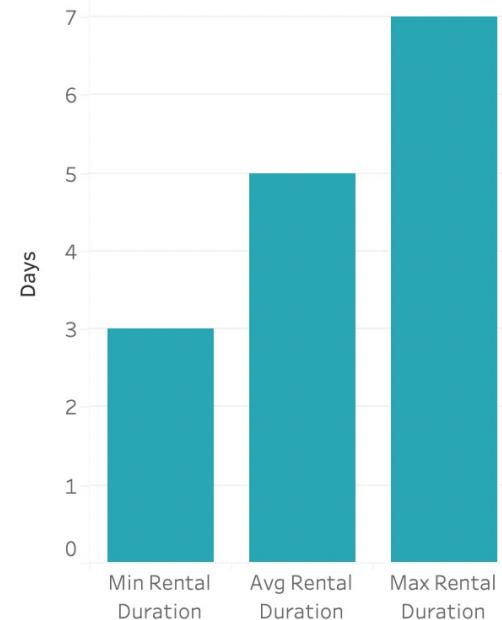


ROCK BUSTER STEALTH

Rental Rate



Rental Duration



Location of Top 5 customers

*Personal Identifiers of these 5 individuals was not disclosed. If this PII is necessary, feel free to send me a private message *



Data was concluded by conducting the following;

- CTE – common table expressions
- Subqueries
- Joining



ROCK BUSTER STEALTH

KEY INSIGHTS

- India and China are the biggest markets, therefore making sure we provide efficient service is of great importance.
- Despite identifying the most popular genres and ratings, the strength of Rockbuster movie database is having titles that can satisfy different tastes. • Focus on having a broad movie selection, rather than focusing on certain genres.
- We suggest modifying offerings based on the popularity of genres in each country. This can be leveraged during seasonally.

RECOMMENDATIONS

- Launch a marketing campaign in the Top 10 cities with highest revenues
- Initiate an Ambassador program and recruit the Top 5 customers
- Monitor the data to see if there is any growth after 6 months of the Marketing campaign launch
- Have the acquisitions team send a mass survey to all customers in non- performing regions



INSTACART - OBJECTIVE

Instacart is an open-source data set, that contains information about customers and their purchase habits. The goal is to analyze purchasing habits of different socio-economic groups of customers and advise marketing strategy

DELIVERABLES

[Project brief](#)

[Report](#)

[GitHub](#)



INSTACART - OVERVIEW

DATA



- [Project Brief](#)
- [Scripts](#)
- [Final Analysis](#)

SKILLS



- Python, Pandas, NumPy, Matplotlib, SciPy, Seaborn
- Jupyter Notebooks
- Data wrangling & subsetting
- Data consistency checks & deriving new variables
- Combining, grouping & aggregating data
- Data visualizations using matplotlib

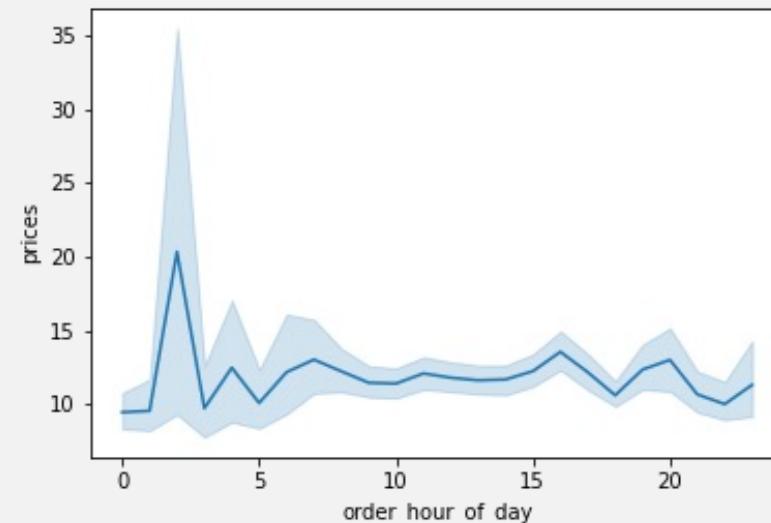
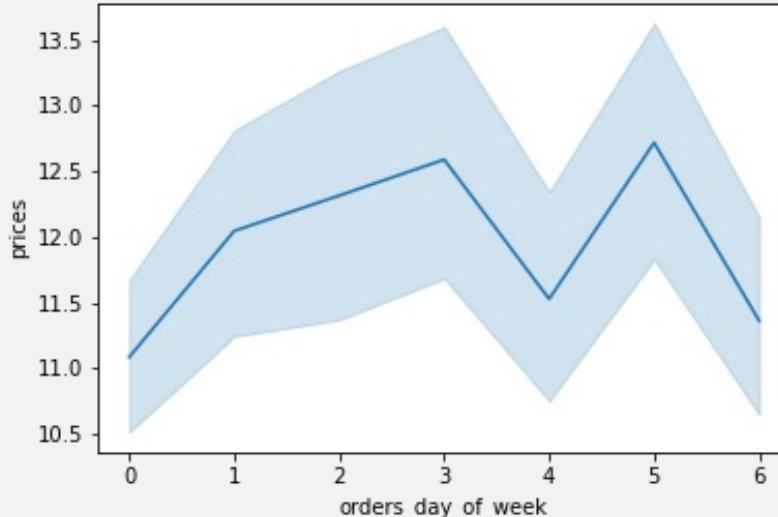
TOOLS



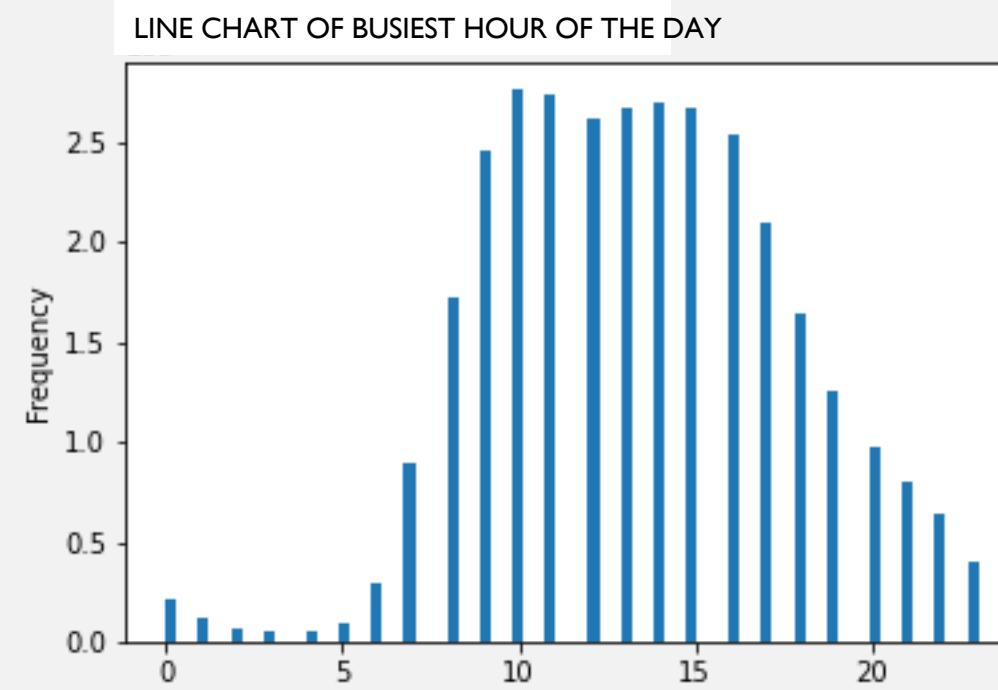
- Python
- Anaconda
- Jupyter Notebook
- Pandas
- NumPy
- Seaborn,
- Matplotlib
- SciPy
- GitHub



INSTACART – EXPLORATORY ANALYSIS

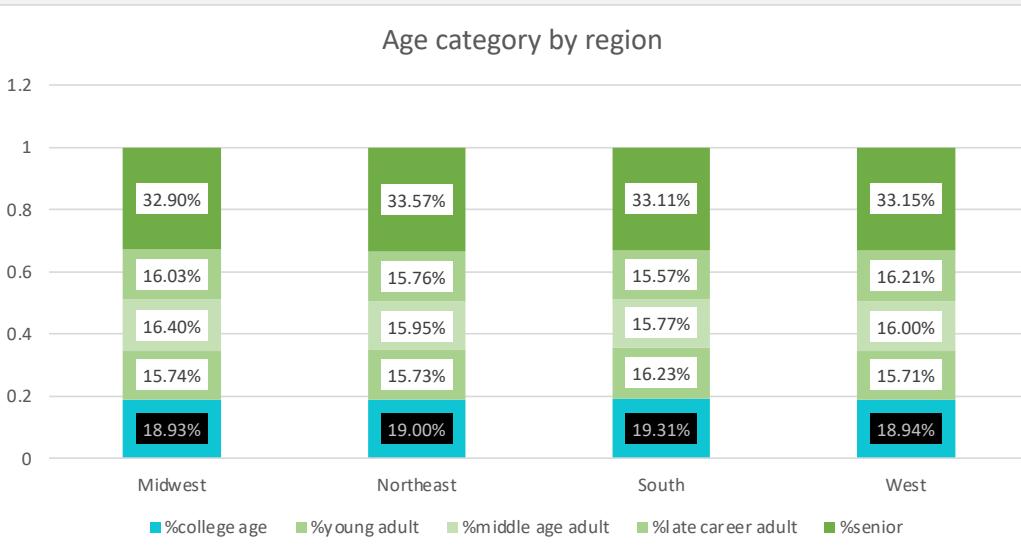


- Merged data sets
- Explored variables such as busiest days, days with most orders or busiest hours of the day
- Used Python commands to wrangle and clean data
- Performed commands to group, aggregate and summarize



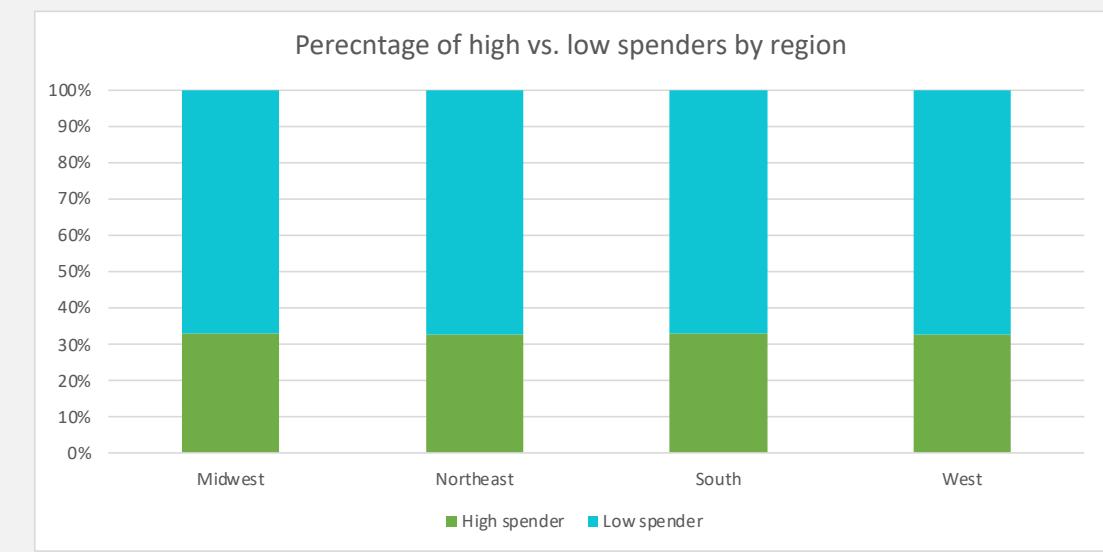
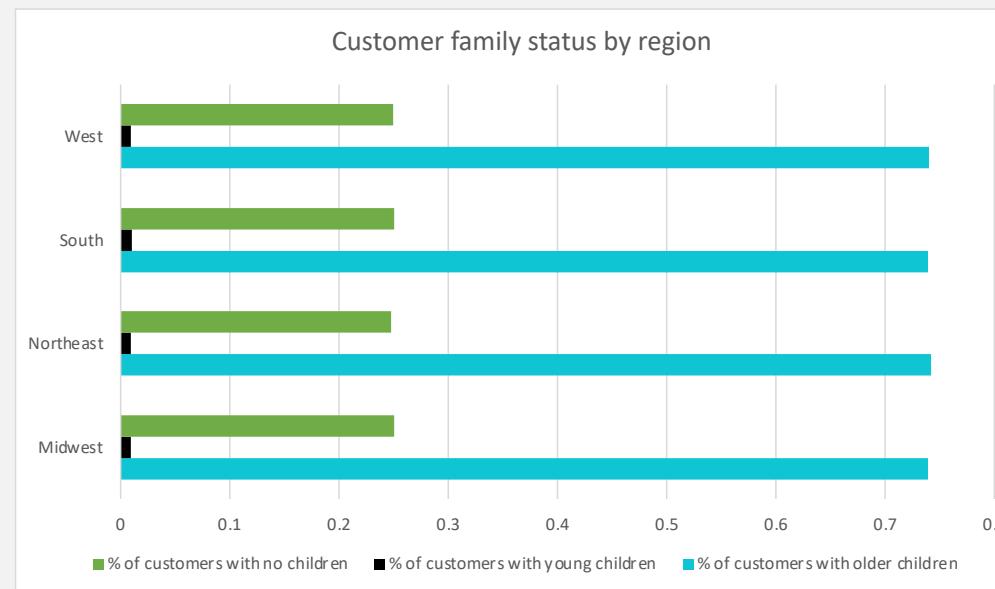


INSTACART – CUSTOMER PROFILE



Customer profiles were divided by age, marital status, income, and number of dependents

- They were further grouped by regions
- New variables were defined using; For cycles, Loc functions, If statements
- This provided insights into customer's spending habits, based on different socioeconomic backgrounds



Across all regions, they all have the same proportion of low spender vs high spender



INSTACART – CUSTOMER PROFILE

KEY INSIGHTS

- There is a correlation between first time customers first and buy high range product.
- Customer Loyalty program customers only purchase mid and low range products.
- Regions across all boards have mid range products, followed by low range products. There are very few high range products in each region.
- Every region has the same proportioned of high and low spenders.

RECOMMENDATIONS

- The operations team should inquire and target where the biggest margins are.
- Identify the high range products that sell the most and allocate marketing efforts accordingly.
- Marketing ads should be launched on the weekends and between 9am to 5pm during the week.