Data Analytics Portfolio

An Analysis of Data by:

Swati Gaikwad

PROJECTS

01. GameCo

Analyzing global video game sales.

02. Medical Staffing Agency

Analysis to determine allocation of medical staff for upcoming Influenza Season in US.

03. Rockbuster Stealth LLC

Answering business questions for an online video rental company.

GitHub Link

04. Instacart Basket

Marketing Strategy for an online grocery store.

GitHub Link

05. Employee Attrition Analysis

Perform Advanced Analytical techniques on Employee termination.

GitHub Link

01. GameCo

Overview

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.

Link To : Project Brief

Skills used: Cleaning , Grouping and Summarizing Data , Descriptive Analysis, Presentation of findings, Visualizations.

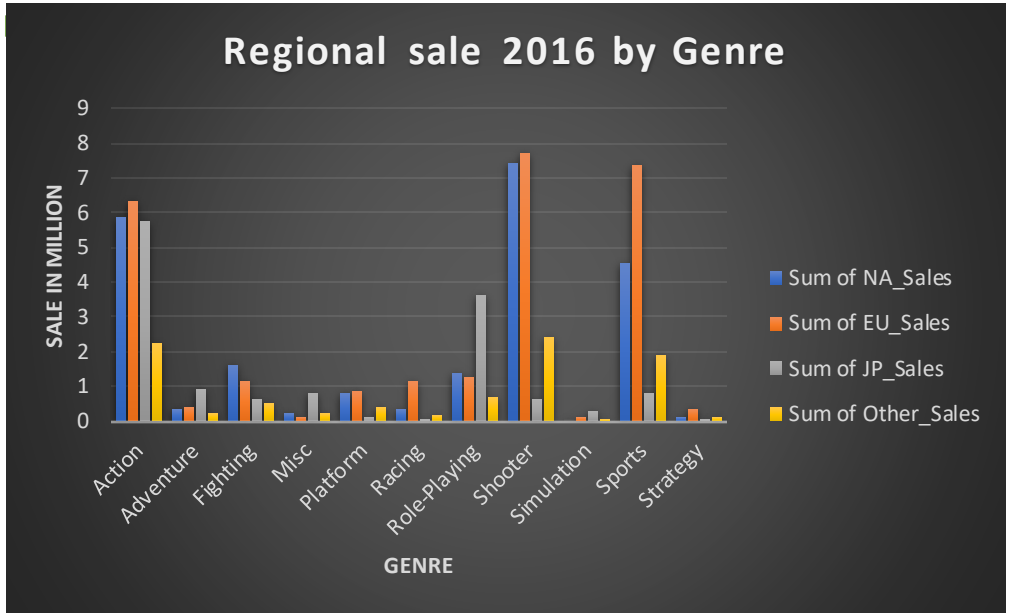
Data set: Link

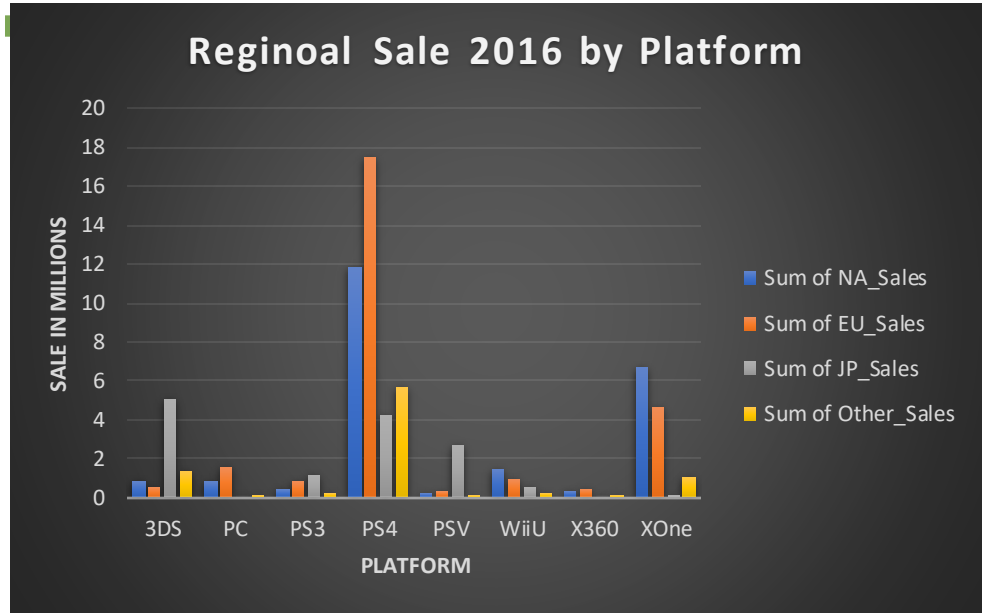
Tools Used:

PowerPoint

Excel

Analysis 2006-2016

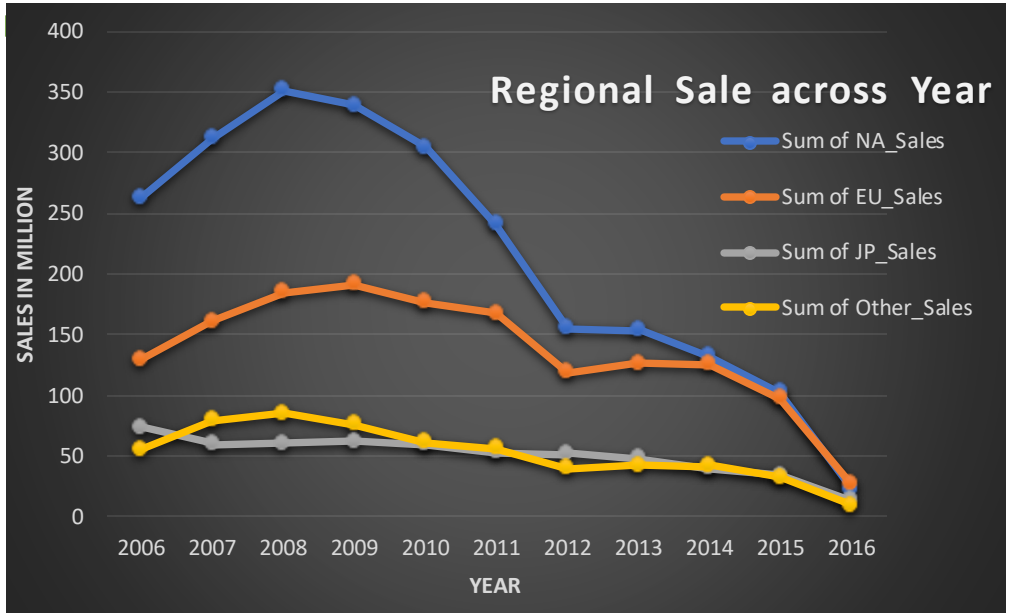




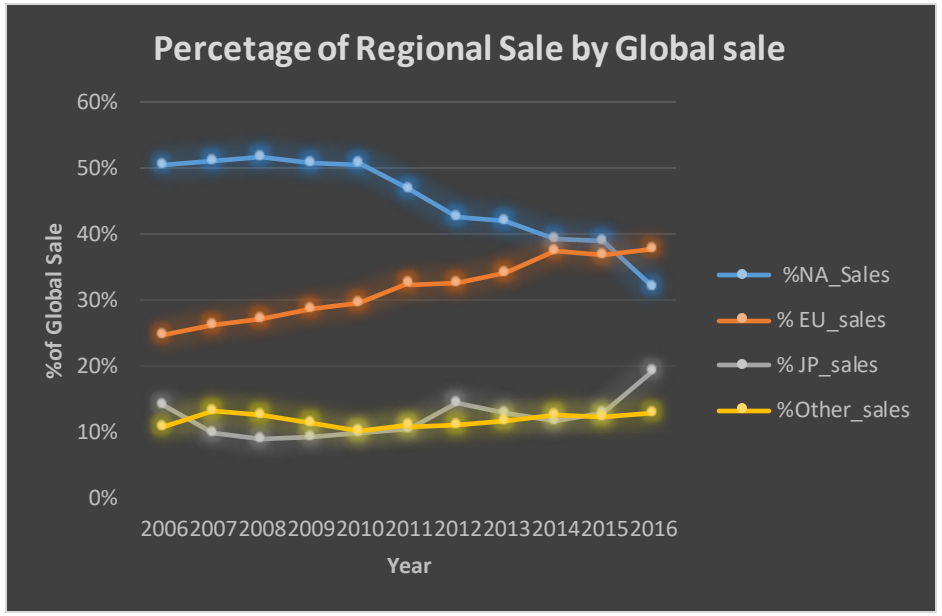
- North America is the most upcoming market followed by Europe where again chances of Action , Shooter ,and Sports games will be in focus.

- Top 2 Platform –PS4, Xone

Analysis 2006-2016



- Graph shows ,NA sales which had highest sale in past, decreasing continuously after 2008.
- It also shows that Europe has Top market position in this year 2016.



- We can see here the negative correction of NA sales across EU and JP sales, but other region sale is quite stable.

Recommendations

- I would recommend to focus on North America which will give highest sale followed by Europe in Action, Shooter and Sports games.
- I would like to recommend maximize the marketing opportunities in Europe ,as market in this region increases in it's global share of sales as we can see in the Percentage of Global Sales by Region chart.
- GameCo should invest more in Role playing games in Japan.
- I would also recommend to focus on marketing of Genres & Platforms aa per demand in specific region.

02. Medical Staffing Agency

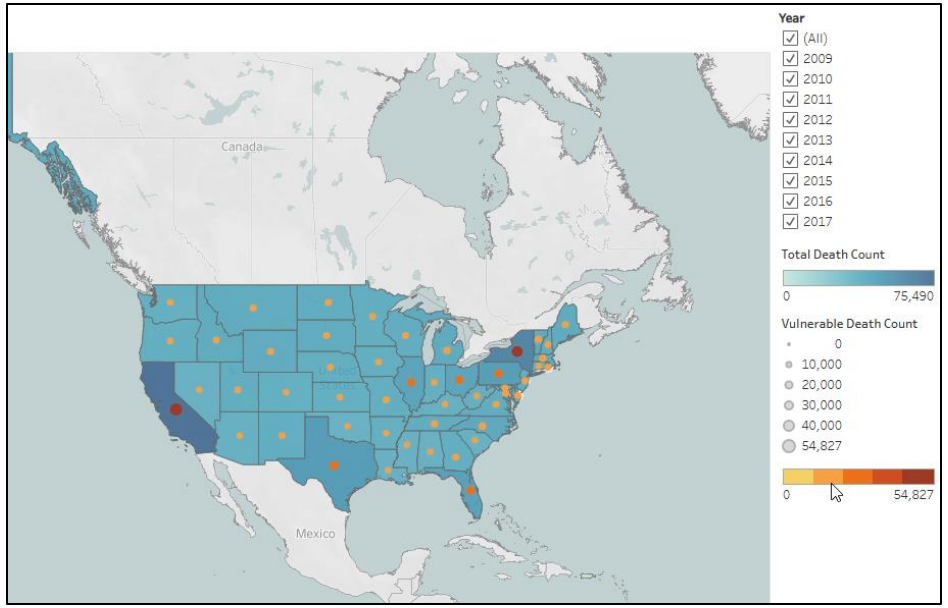
Overview

- **Objective:** Perform an analysis to determine when and how much additional medical staff need to send by a medical staffing agency in each state of US.
- **Link To :** Project Brief
- **Skills used:** Cleaning , Integrating and transforming data, Hypothesis Testing , Predictive analysis, Visualization , Storytelling using Tableau and presenting findings .
- **Link To:** CDC Dataset
- **Tools Used:**

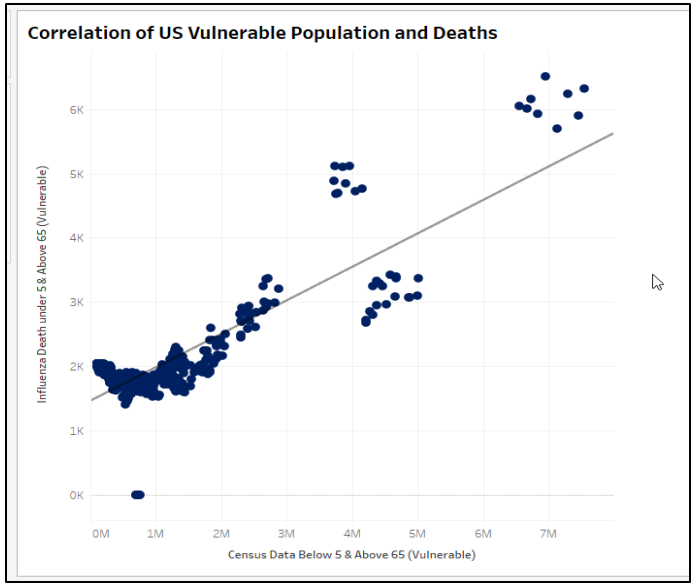

Excel


Tableau

Analysis

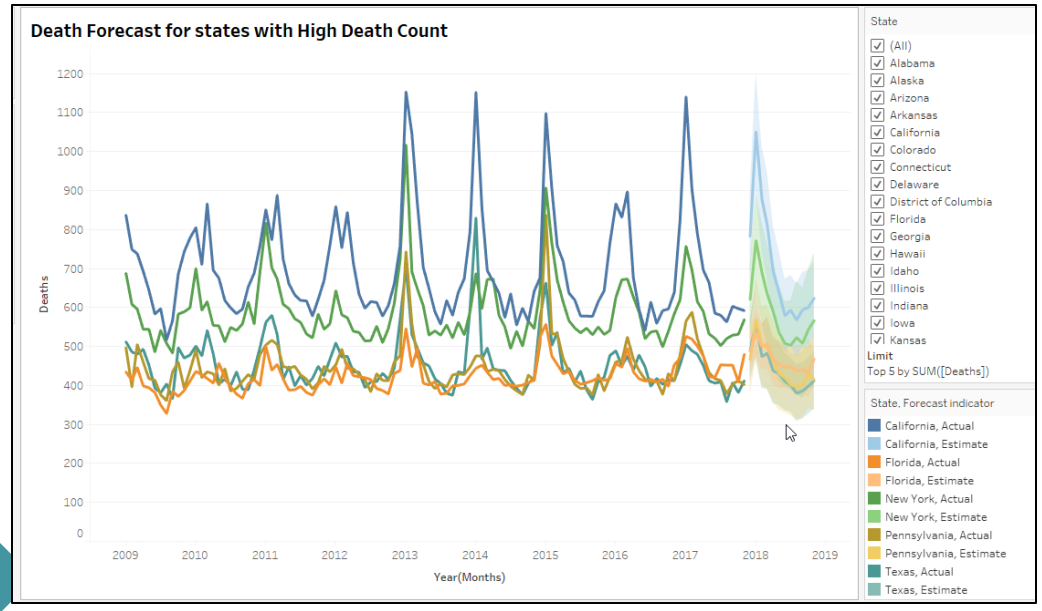


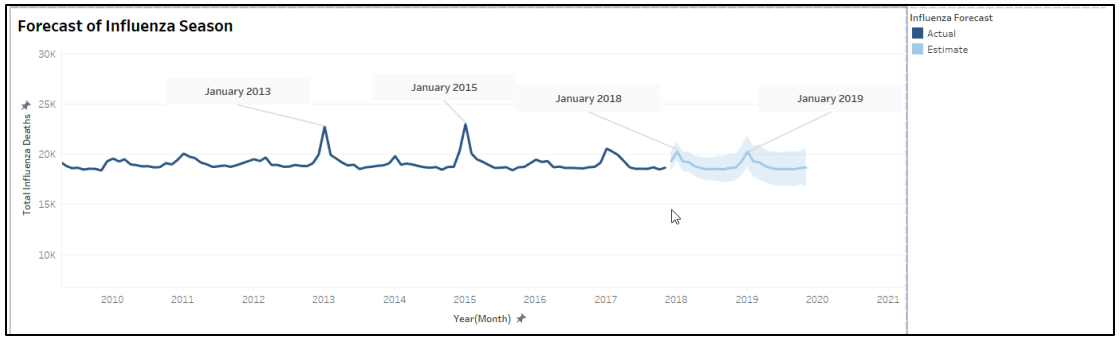
- California and New York are the states where number of vulnerable deaths are higher.



- Majority of Data points are very close to the trend line, what does it indicate a strong correlation between the Population above 65 years and Influenza Deaths .

Influenza Season :Forecast





- Flu tends to be more prevalent from the moths during December, January, February and March. We can see a pattern of repeating seasonality.

- According to the Forecast Chart the 2018 Influenza season should follow the patterns from the previous years.

Recommendations

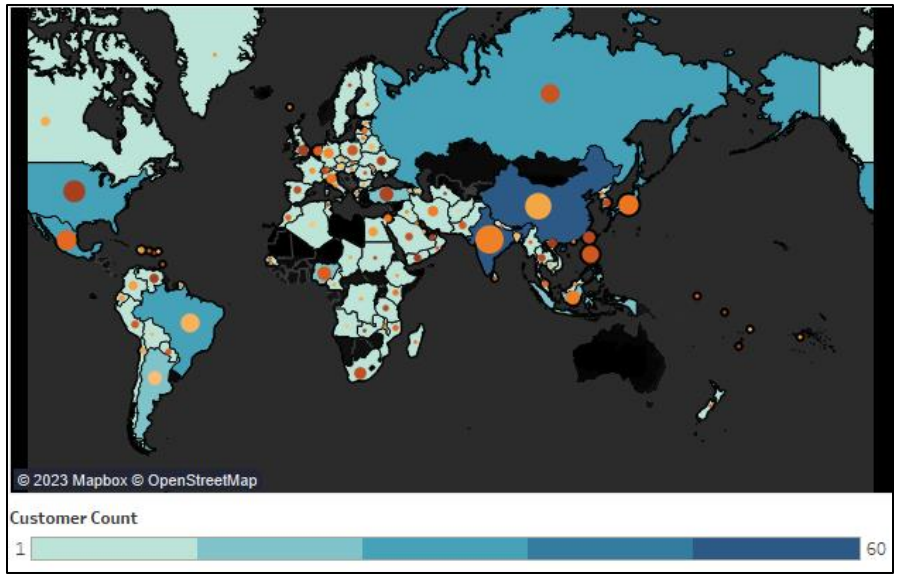
- Additional resources should send by November as Flu season occurs during December, January, February and December.
- California , New York, Texas, Pennsylvania, and Florida these are the states where number of deaths are higher and should receive the largest number of additional resources.
- Vulnerable populations below 5 and above 65 years need more additional resources as compared to other age group.
- Need further analysis in order to find staff-to-patient ratio in each state to send additional resources.
- **For Complete insights please visit :** Tableau Storyboard & Recording

03. Rockbuster Stealth LLC

Overview

- **Objective:** Rockbuster Stealth LLC is a movie rental company is planning to use its existing movie licenses to launch an online video rental service in order to stay competitive.
- As a data analyst for Rockbuster' s business intelligence department, I was asked with assisting various departments by giving data driven answers that they can use for their 2020 company strategy.
- **Link To:** Project Brief
- **Skills used :** Relational Databases, Cleaning ,Filtering and summarizing data, Joining tables, Database Querying ,Subqueries , Common Table Expressions.
- **Link To :** Dataset
- **Tools Used:**

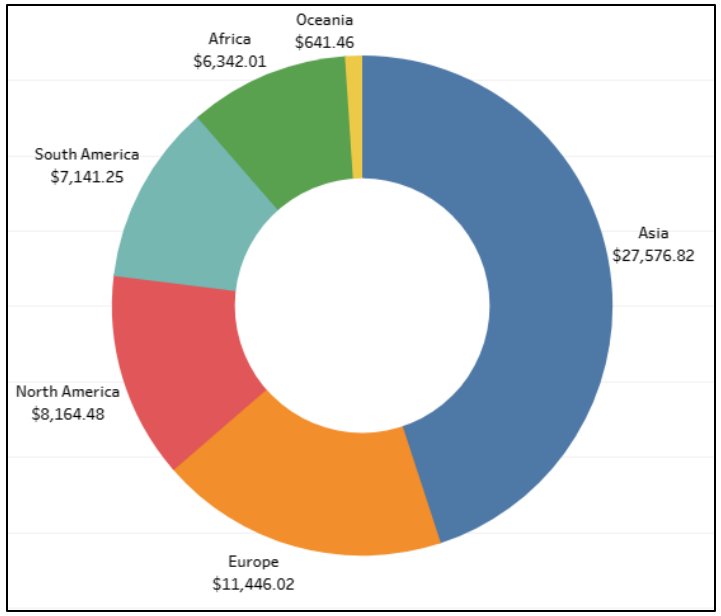
Rockbuster Stealth : Analysis



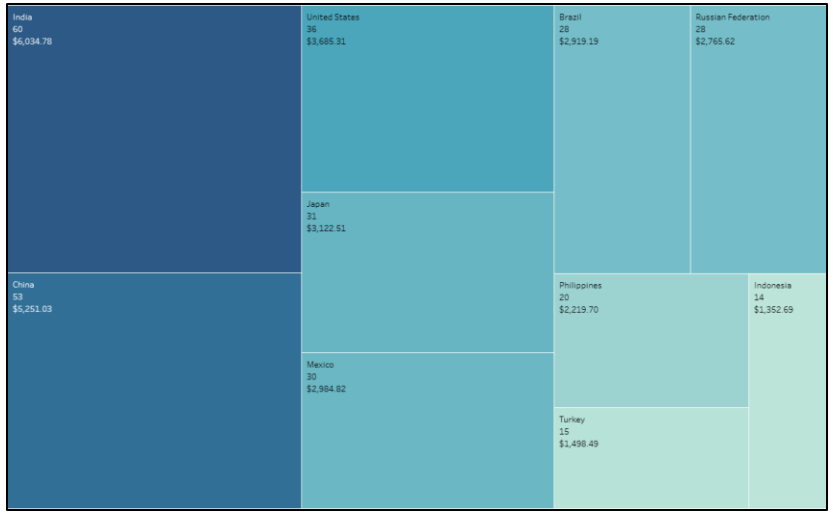
- As shown in Map , Rockbuster total 584 customers are based in 108 countries.
- **Top 10 Countries:**

country	customer_count
India	60
China	53
United States	36
Japan	31
Mexico	30
Brazil	28
Russian Federation	28
Philippines	20
Turkey	15
Indonesia	14

Rockbuster Stealth : Analysis



- Total Revenue $61,312
- Chart shows the Regional revenue contribution



- Most Profitable Regions: Asia, North America, South America
- Countries with high Customers Count: India and China

Statistical Analysis

RENTAL RATE

- Minimum: $0.99
- Maximum: $4.99
- Average: $2.98

RENTAL DURATION

- Minimum: 3 days
- Maximum: 7 Days
- Average: 5 Days

MOVIE LENGTH

- Minimum: 46 Mins.
- Maximum: 185 Mins.
- Average: 115 Mins.

REPLACEMENT COST

- Minimum: $9.99
- Maximum: $29.99
- Average: $19.84

Recommendations

- Rockbuster should continue to promote online services in Asia regions specifically India and China.
- Rockbuster can provide subscription offers or discount on rental to our High value customers.
- It is recommended to increase promotional services in region like Africa.
- Inventory related to the Movies with highest revenue should increase.
- We can recommend subscription offers or discount on rental to our High value customers.

- **For Complete insights please visit :** Tableau & Project Presentation

04. Instacart Basket

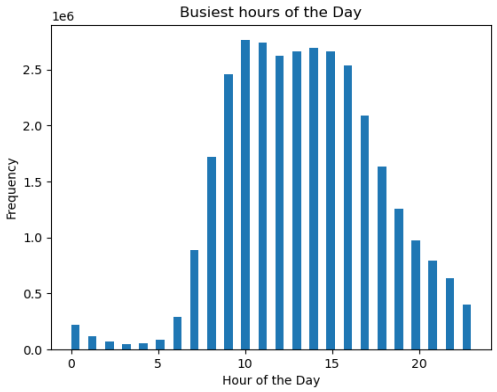
Overview

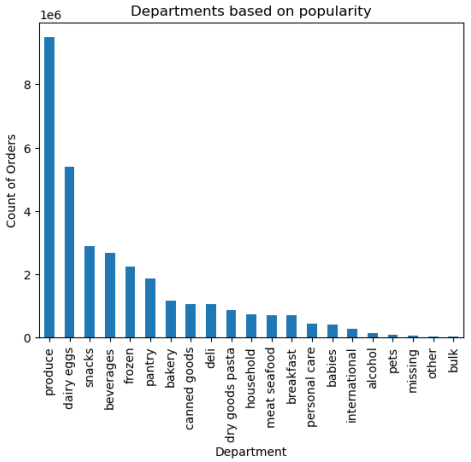
- **Objective:** Perform an initial data and exploratory analysis of some of Instacart data which is an online grocery store, in order to derive insights and suggest strategies for better segmentation based on the provided criteria.
- **Link To:** Project Brief
- **Skills used :** Data Consistency Check, Data merging & Wrangling, Deriving Variables, Grouping & Aggregating Data using Python libraries (pandas, numpy, matplotlib, scipy, seaborn), Creating visualization & Reporting in Excel
- **Link To :** Dataset
- **Tools Used:**

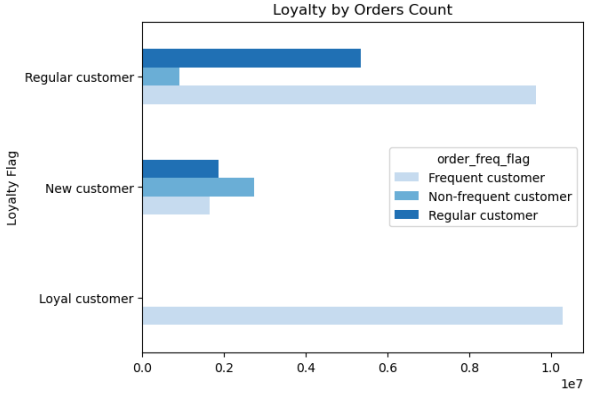

Python

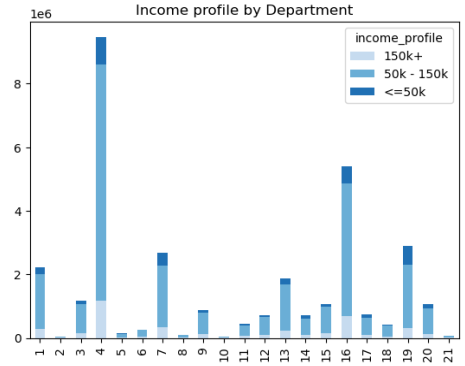

Excel

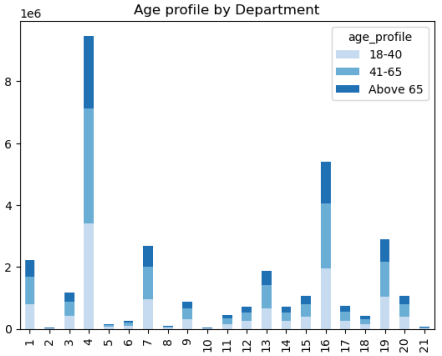
Analysis

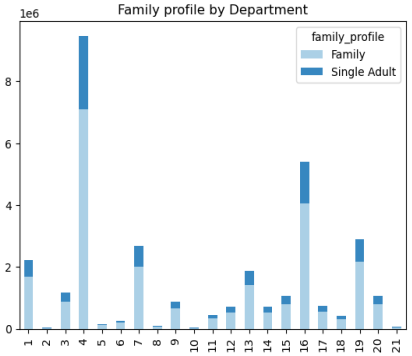






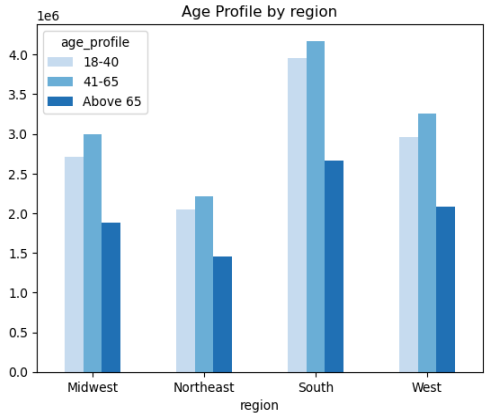


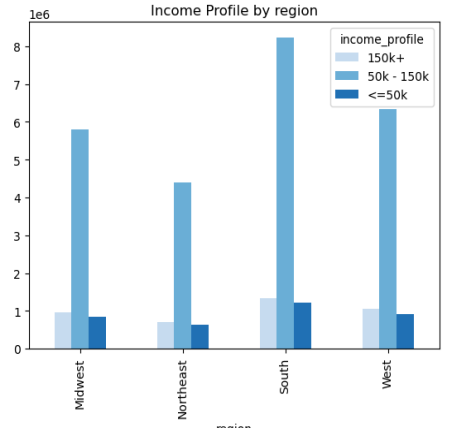




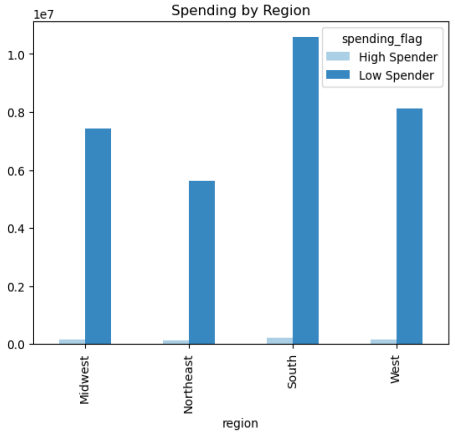
- 10 am to 15pm busiest hours of the day & 3 to 4 am is the time when order count is least.
- The departments with highest frequency of product orders - Produce, Dairy eggs, Snacks, Beverages, Frozen.
- The loyal customers are mostly the frequent customers followed by regular customers .New customers are least frequent.
- Customers will spend very similarly to other customers regardless of their income profile, with a few exceptions.
- At or after 40 years old, income tends to get higher for that age profile.
- Families certainly spend more on produce than single people.

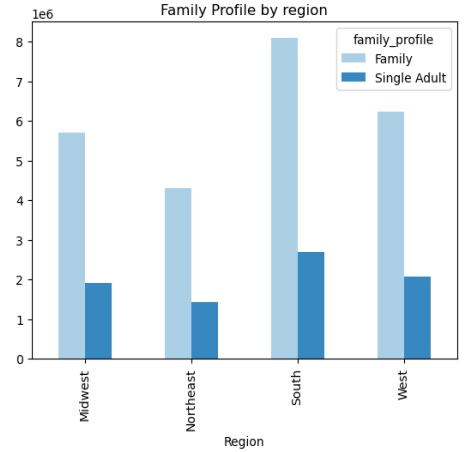
Instacart Analysis by Region





- Number of orders is highest in the region South followed by West, Midwest and Northeast. The South is the region that leads many categories.





Recommendations

- Ad campaigns should ads during busiest day and time .
- Can also improved by placing target ads at times when there are the fewest orders – Tuesday and Wednesday, and in general, between the hours of 11pm and 6am.
- People tend to spend money on pricier items in the mornings, so ads for more expensive items can be placed at this time.
- Instacart should focus on advertising higher priced items to help increase the revenue for higher priced items.
- Advertise to re-order common items will help increase orders and loyalty & should think for some special offer to new customers to increase their order count.
- A significant portion of the advertising budget should be targeted towards family items.
- **For Complete Insights please visit :** GitHub & Project Presentation

Thank you!

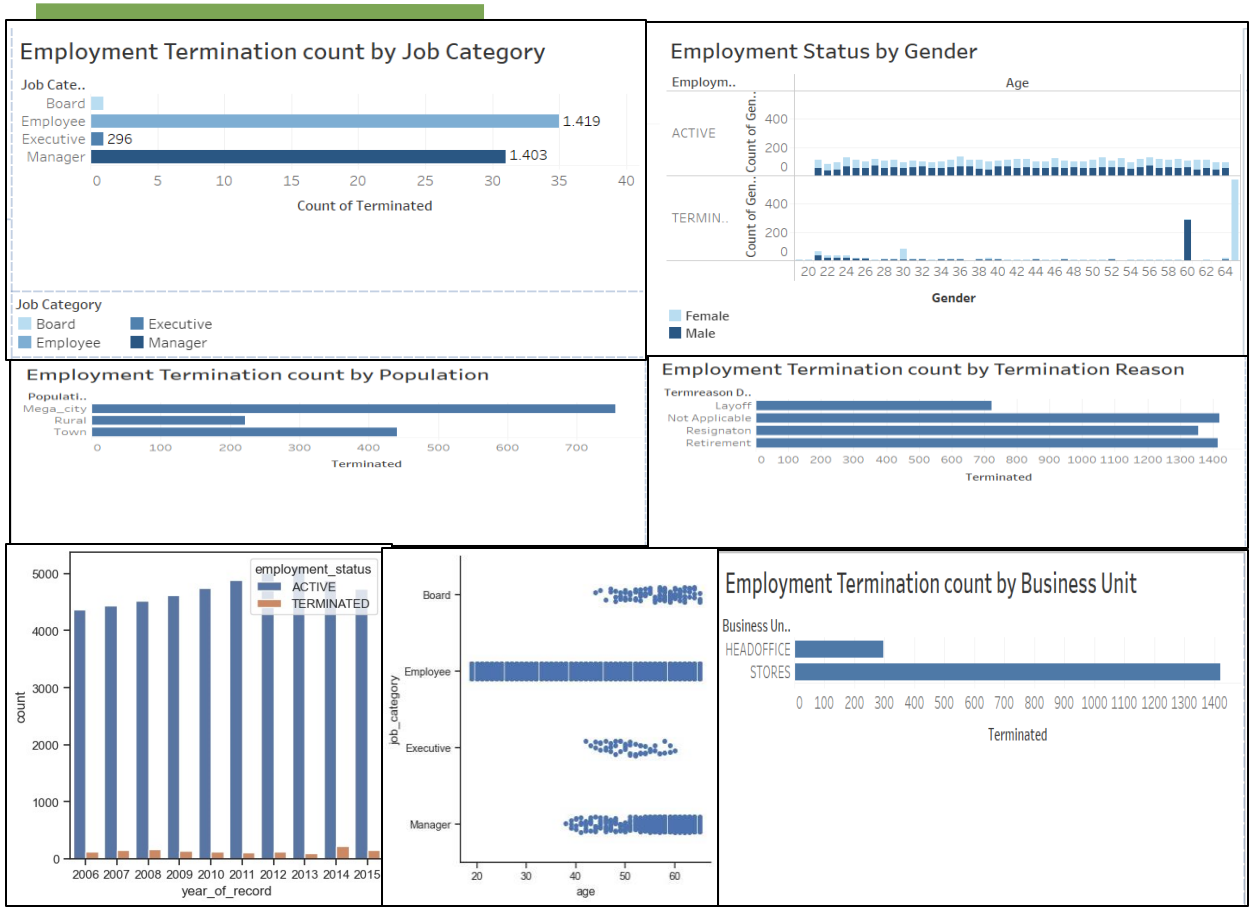
05. Employee Attrition Analysis

Overview

- **Objective:** Perform an advanced analysis on a fictitious dataset 'Employee Attrition' of a Canadian Company in order to derive insights where each of 10 years (2006-2015) it shows employees that are active and those that are terminated.
- **Link To:** Project Brief
- **Skills used :** Data Sourcing, Data Consistency Check, Data merging & Wrangling, Deriving Variables, Grouping & Aggregating Data using Python libraries , Exploratory Visual Analysis, Geospatial Analysis ,Machine Learning, Time-Series Analysis , Creating visualization & Reporting in Tableau.
- **Link To :** Dataset
- **Tools** 

- **For Complete Insights please visit :** GitHub & Tableau Storyboard

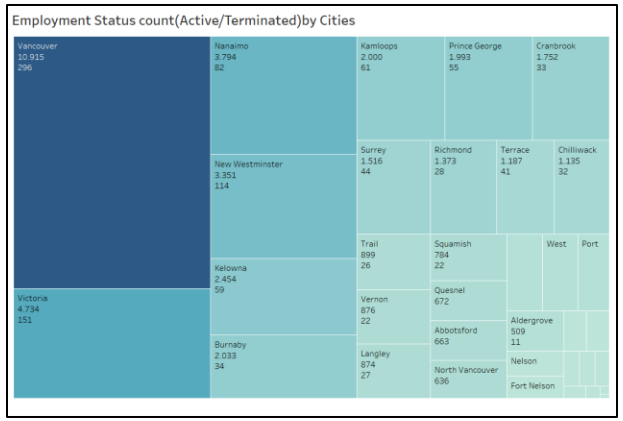
Analysis

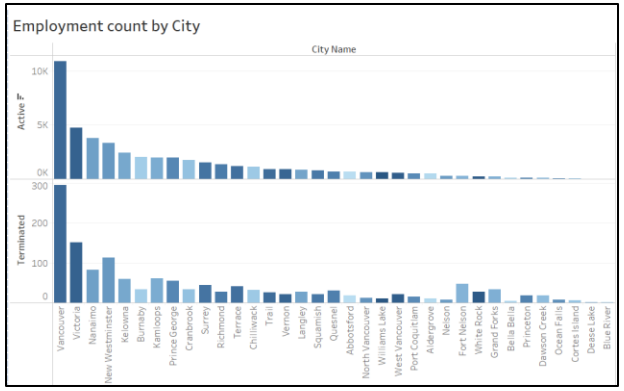


Observations:

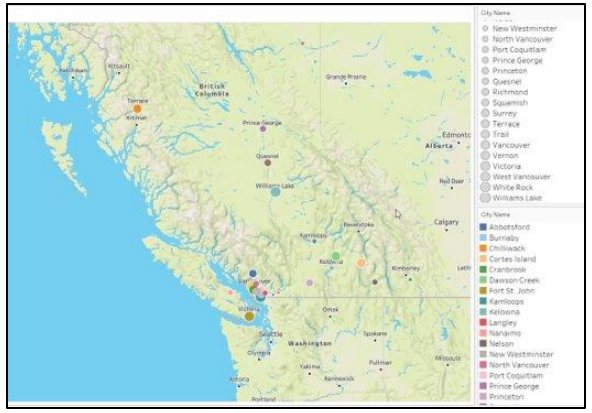
- Very few have stopped employment when they were at executive level. Nobody was terminated if they were at the executive level, or higher.
- Maximum number of "Employees" are the ones with the employee job title.
- The largest age peak of above 60 years old. This would be people who are retiring from the work force.
- No major difference between male and female termination, however female are more in both.
- Maximum number of employees are from the mega city in which the ratio of active to terminated status of the employee is higher.
- As expected, layoffs are involuntary, whereas resignations and retirements are voluntary.
- 2014 had an unusually high number of employment terminations.
- Cases of termination are higher in stores whereas cases of termination are lower in the head office.

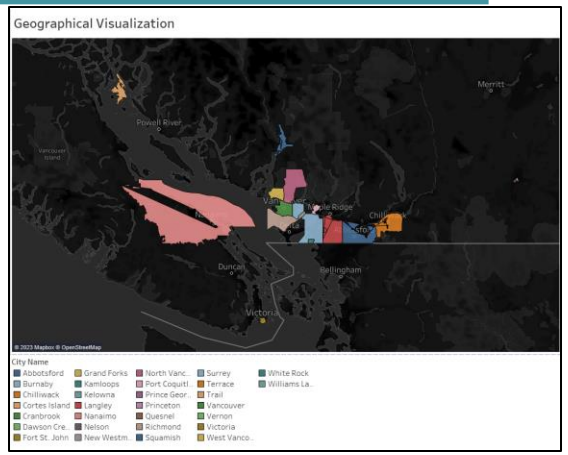
Geospatial Analysis





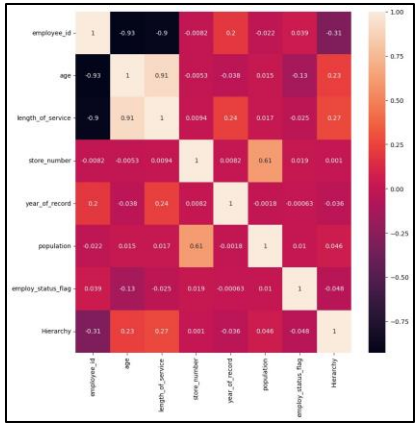
- In the above charts, we see that Vancouver and Victoria have the highest share of terminated employees. This skewness might be because the highest concentration of employees in the dataset is in these 2 cities.



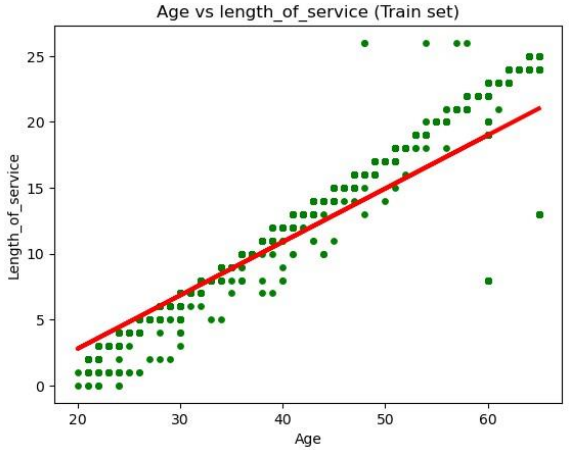


Analysis by Swati Gaikwad

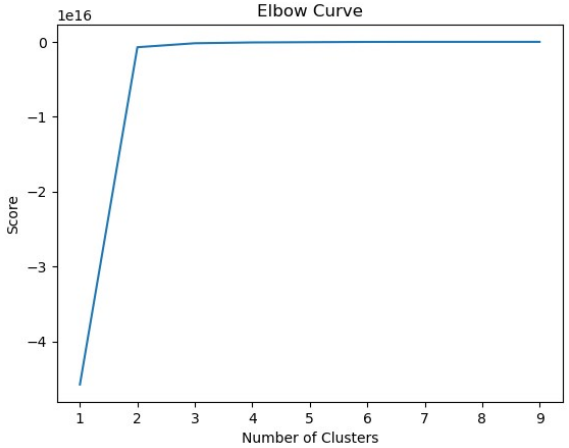
Advanced Analysis

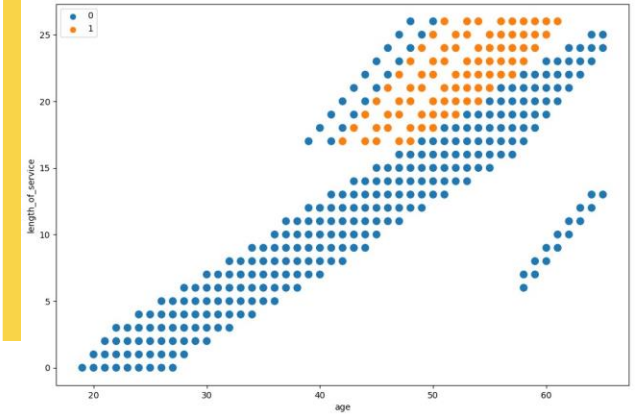


- The correlation heatmap shows a strong positive correlation between length of Service and Age .
- The linear regression analysis also identifies a strong correlation between length of Service and Age, indicating a trend.



- Based on the elbow curve, the optimal number of clusters is 2. After value of 2 on the x-axis the curve flattens out.
- Employees with longer lengths of service are less likely to leave.





Analysis by Swati Gaikwad

Summary

In order to perform a regression analysis, I have determined length of service as a potential dependent variable, which represents the length of service for each employee . Based on that I analyzed other factors as independent variables like :

- Age: It may have an impact on the length of service or other aspects of employment.
- Gender: I explored if there are any gender-based differences in the length of service or other employment-related factors.
- Job Category: It indicated differences in job responsibilities, career progression, or other factors that may impact the length of service.
- Business unit: Different units may have distinct work environments, opportunities for growth, or other factors that could affect the length of service.

Limitations of the case study:

- There were not enough data points to yield a highly significant result. The dataset does not have many features, but the approach can be used as preliminary work for more complicated employee churn models.
- The data contained a limited number of variables upon which to conduct the analysis.

Recommendations:

- Continuous Monitoring: Regularly monitor employee attrition rates and conduct periodic analysis to identify any emerging trends or issues. This allows for timely interventions and adjustments to retention strategies.
- Address Termination Peaks: Identify and address the periods with higher termination counts, such as in December 2014 and December 2015. Investigate the reasons behind these peaks and take corrective actions, if necessary. It could involve improving working conditions, addressing workload concerns, or ensuring appropriate support for employees during busy seasons.

Next steps:

- Gather more data points or information for these variables.
- Analyze the impact of additional variables on the Employee Attrition count.

Thank you!