DR. PADMASINI VENUGOPAL

DATA ANALYST PORTFOLIO

PROJECTS

- GAMECO

- PREPARING FOR THE NEXT INFLUENZA SEASON 2018 IN THE USA

- ROCKBUSTER STEALTH

- INSTACART

- PIG E. BANK

- DAIRY PRODUCTS CONSUMPTION IN INDIA (2019-2022)

GAMECO VIDEO GAMES SALES

PROJECT 1: GAMECO BUDGET PLANNING

OBJECTIVE
- To develop new games with the data.
- To better understand the games in the market.

KEY QUESTIONS
- Are any specific games more popular than others?
- Which publishers are the main competitors?
- Have any games decreased or increased in popularity over time?
- How have their sales varied between geographic regions over time?

DATA AND PROJECT
- Data set that covers historical sales of video games

DATA: SOURCE

LINK: SOURCE

PROJECT BRIEF: SOURCE

TECHNIQUES APPLIED
- Data sorting
- Data filtering
- Grouping and summarising the data
- Deriving new variables
- Joining the data with VLOOKUP
- Visualizing in Excel
- Presenting results and recommendations

LIMITATIONS IN DATA
- It tracks the total number of units of games sold (not financial figures) from 1980 to 2016.
- The numbers represent units sold in millions. When you see the number "1.2," for instance, this represents a total of 1.2 million units sold.
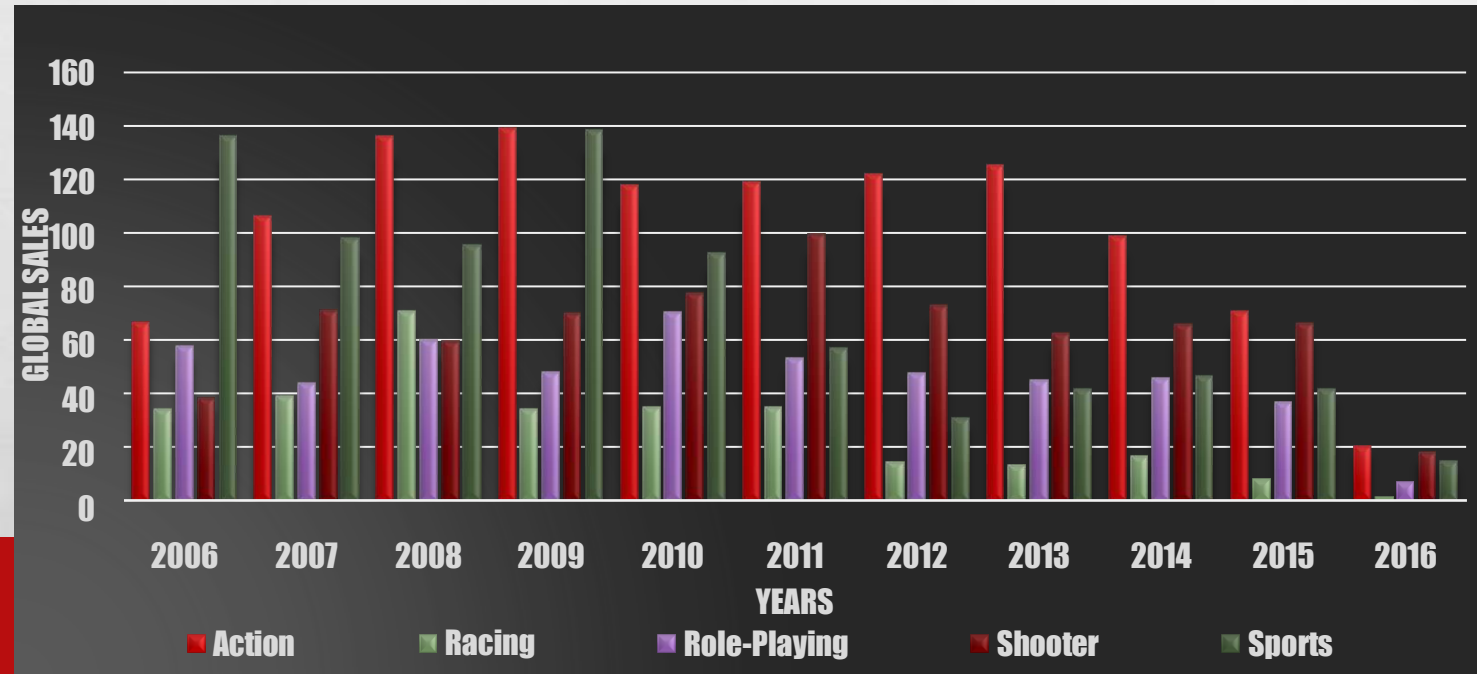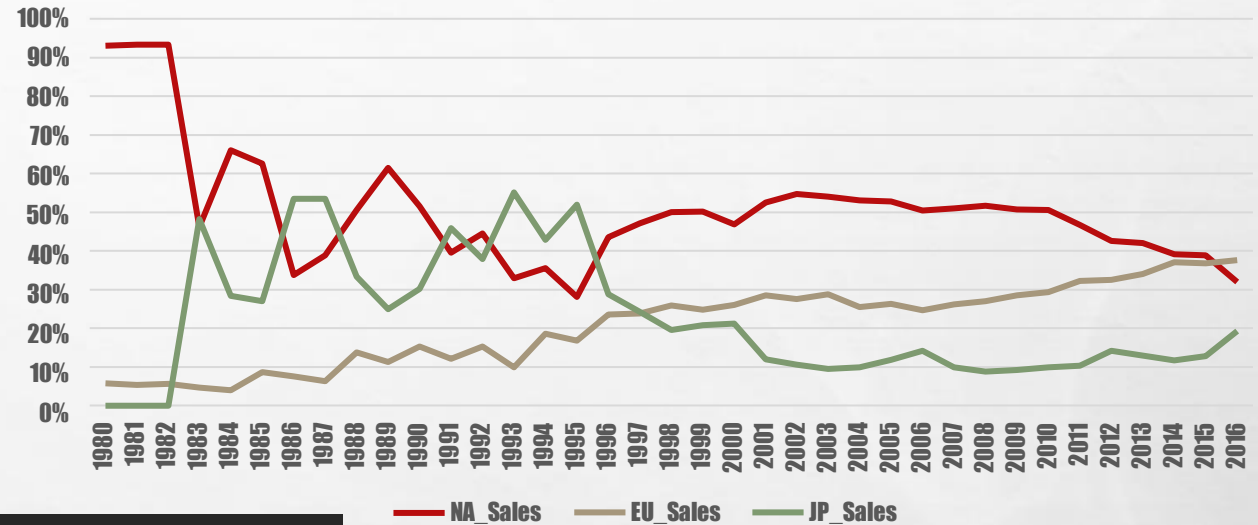
TOOLS USED:

ANALYSIS

PERCENTAGE OF SALES GLOBALLY

- The NA has the largest sales globally for shooter games.

- Eu sales have slowly peaked since 1994 and in the last 2 years, it has crossed the NA American sales globally.

- Since the take 10 years the top revenue-generating game genres are action, racing, role-playing, shooter, and sports.

- After 2010, there seems to be a drop in all the top-selling games. But still, the percentage of global sales has been secured by the increased sales of other games.

DELIVERABLES AND RECOMMENDATIONS

RESULTS/DELIVERABLES

- The project insights and analysis have been presented to the stakeholders.
  Link: source

RECOMMENDATIONS

- More focus should be given to why the high-selling genre "action" has reduced sales now and concentrate on the other high-selling genres.

- Focus on the recent 6% hike in sales in Japan last year may help us to find which games have started to boom.

- Create new games based on the regions and the newer trends.

- Marketing campaigns and promotions can be done to bring back interest in youngsters towards the video games will gradually help in improvising sales.
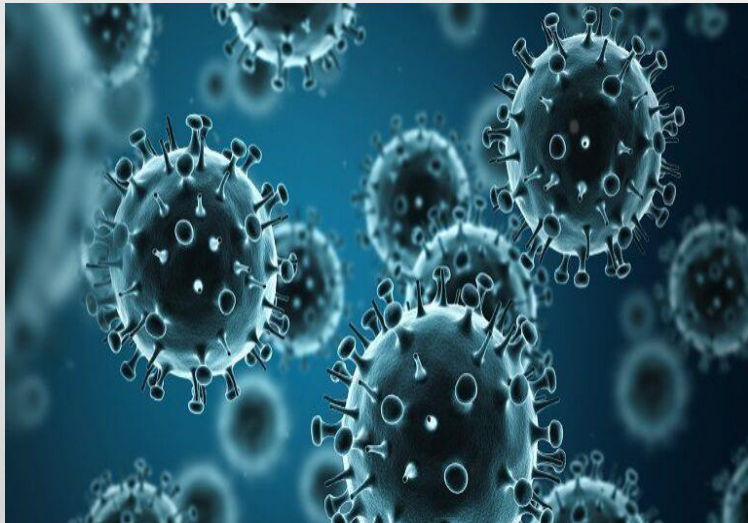
PROJECT 2: PREPARING FOR THE NEXT INFLUENZA SEASON 2018 IN THE USA

OBJECTIVE

- To analyze influenza trends in the US in order to support a medical staffing agency for the deployment of temporary healthcare personnel for the upcoming influenza season.

KEY QUESTIONS

- Determine the timing and spatial distribution of medical personnel
- Determine whether influenza occurs seasonally or throughout the entire year
- Prioritize states with large Vulnerable populations

DATA AND PROJECT

- The influenza death data is collected from the CDC LINK: SOURCE THE US CENSUS DATA FROM THE US CENSUS BUREAU LINK: SOURCE PROJECT BRIEF LINK: SOURCE

TECHNIQUES APPLIED

- Data cleaning
- Data quality and integrity checks.
- Data integration & transformation
- Data grouping and summarizing
- Statistical analysis
- Hypothesis testing
- Forecasting
- Visualizations using tableau

LIMITATIONS

- 82% of influenza mortality data entries were suppressed.
- It is manually collected data that may have typographical errors on it.
- If I had details about the yearly flu vaccination shots that would have helped me to analyze whether the vaccination shots helped in decreasing death rates.
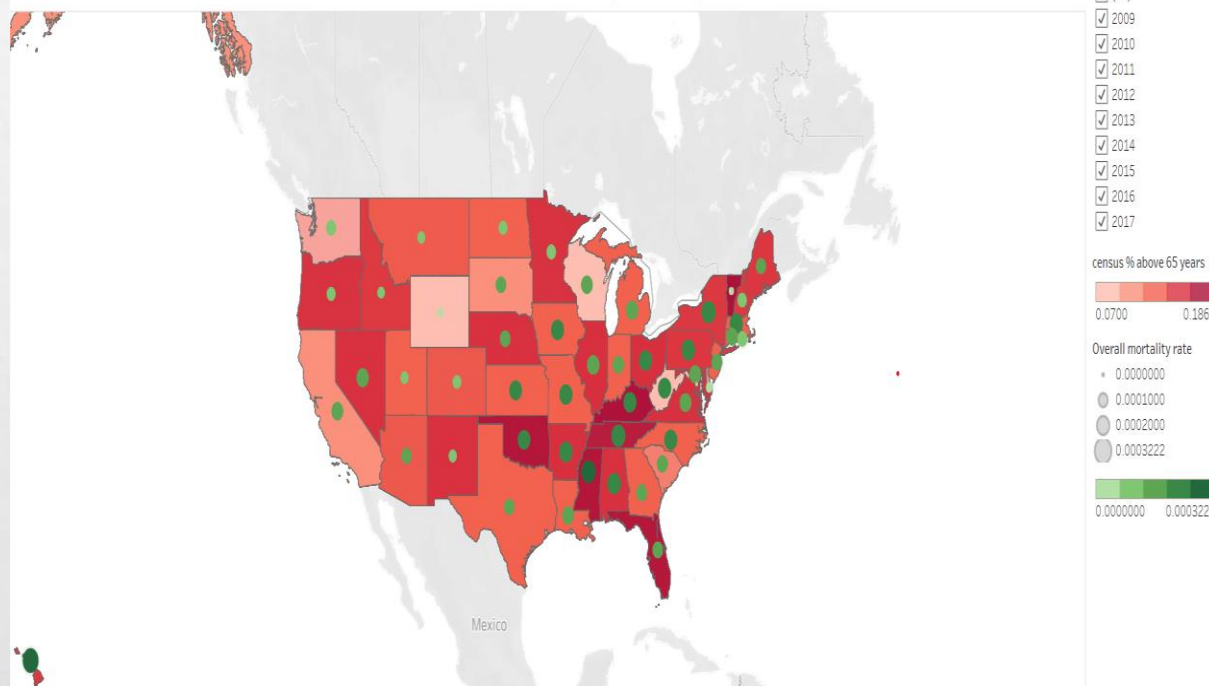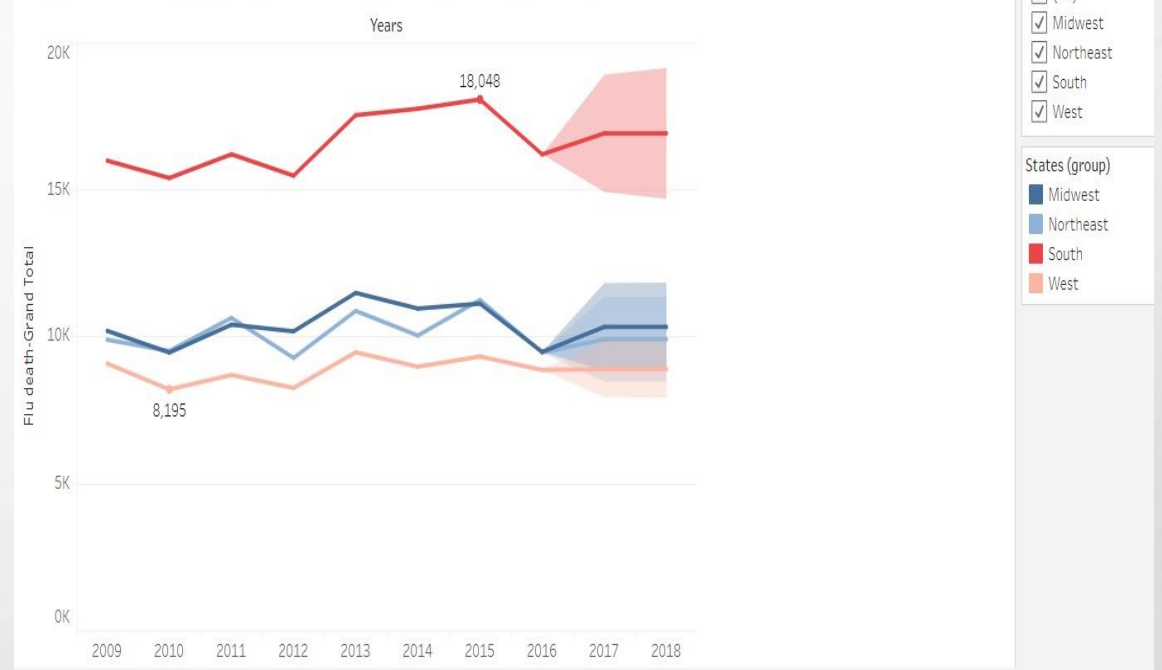
TOOLS USED :

ANALYSIS

Relationship between % of Census data over 65 years and Total Mortality Rate(2009-2017)

TOTAL FLU DEATH RATES(2009-2017) & FORECASTING FOR 2018

Hawaii state has the highest mortality rate followed by it is New York, Tennessee, and Mississippi this shows that the state with a more vulnerable population has the most death rates.
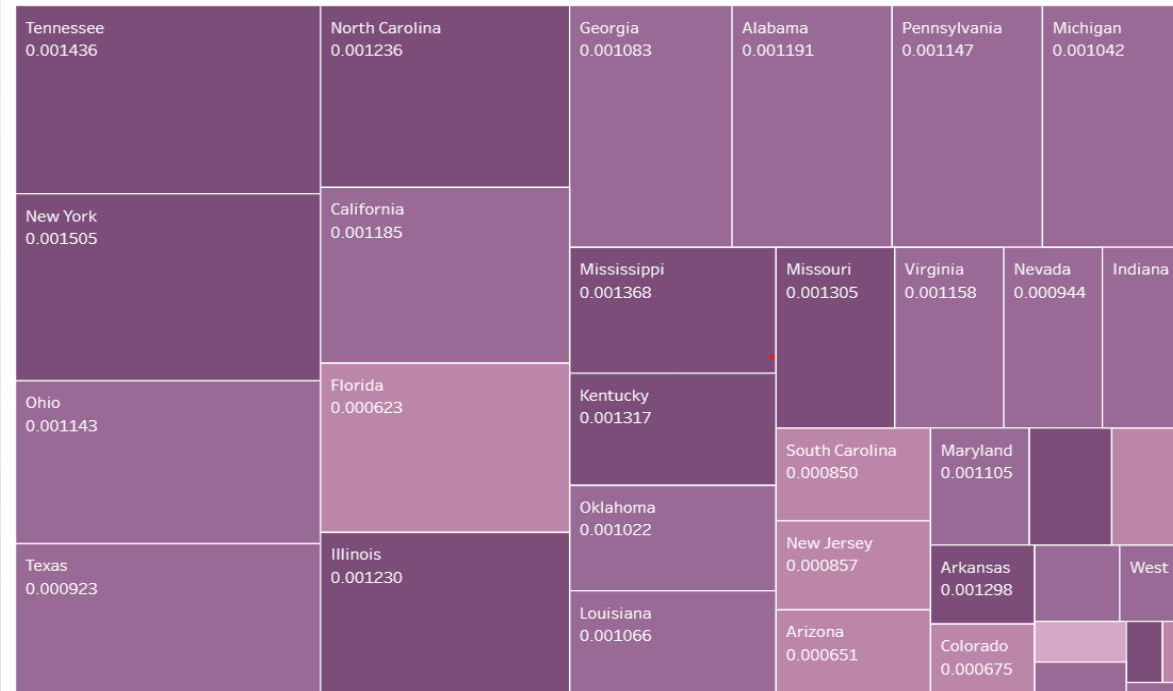
States from the southern region are going to have high flu deaths next year so it's best to send the faculties there in advance
Next is the West (California) will have a high death rate.

ANALYSIS

% OF FLU DEATH RATE 65+ BY STATES

State	Rate
Tennessee	0.001436
North Carolina	0.001236
Georgia	0.001083
Alabama	0.001191
Pennsylvania	0.001147
Michigan	0.001042
New York	0.001505
California	0.001185
Mississippi	0.001368
Missouri	0.001305
Virginia	0.001158
Nevada	0.000944
Indiana	
Ohio	0.001143
Florida	0.000623
Kentucky	0.001317
South Carolina	0.000850
Maryland	0.001105
Oklahoma	0.001022
New Jersey	0.000857
Texas	0.000923
Illinois	0.001230
Arkansas	0.001298
West	
Louisiana	0.001066
Arizona	0.000651
Colorado	0.000675

Relationship between the Census above 65 years and the overall mortality rate(2009-2017)

Overall Mortality rate = 3.00425e-11*Census- >65 years + 9.93126e-05
R-Squared: 0.162149
P-value: < 0.0001

Despite California and Texas having the highest 65+ population count the states that have the highest flu mortality rate are New York, Tennessee, and Hawaii.
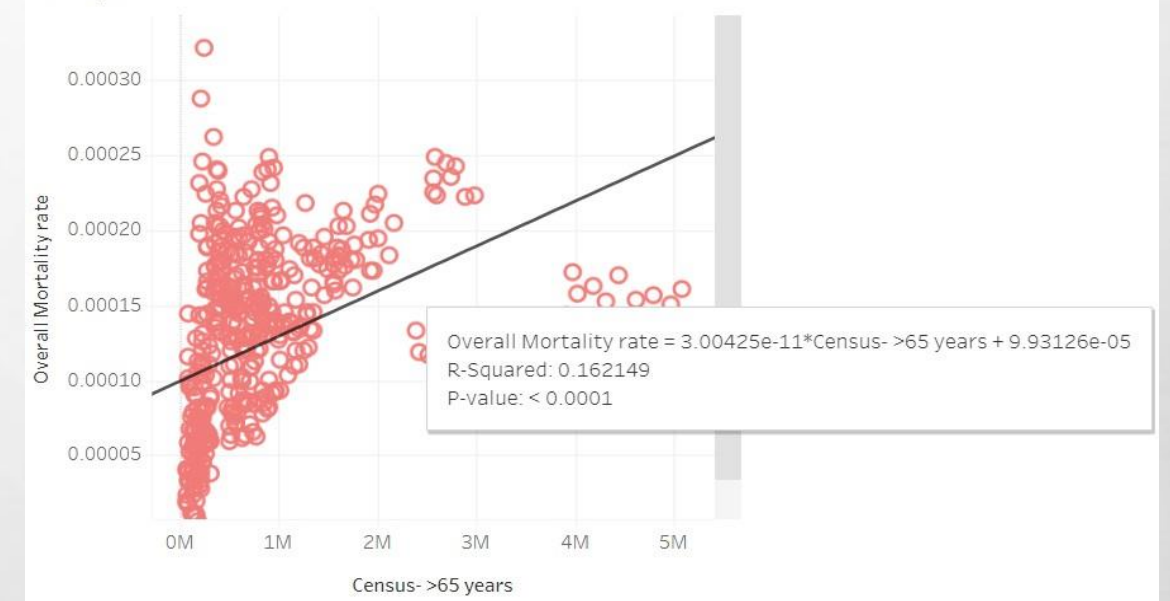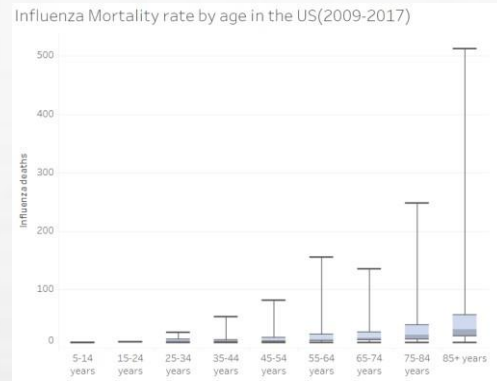
The scatter plot is moderately bound to the trend line and shows a positive correlation this shows that the above 65 years pop. has the highest influenza mortality rate and they are a vulnerable population.
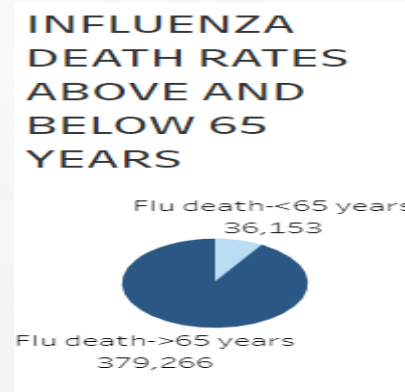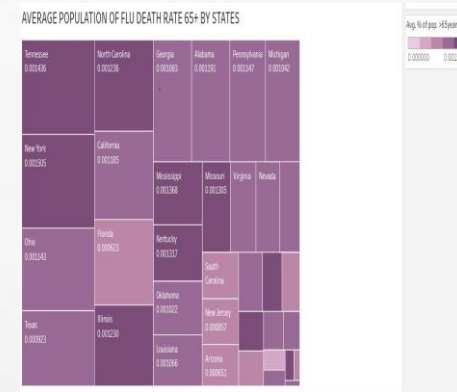
VISUALIZATIONS PERFORMED WITH TABLEAU
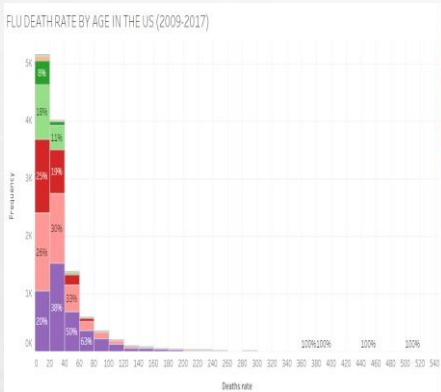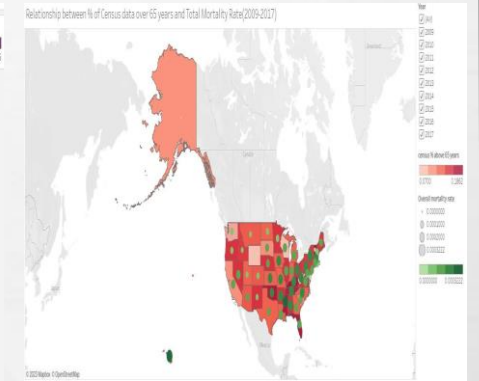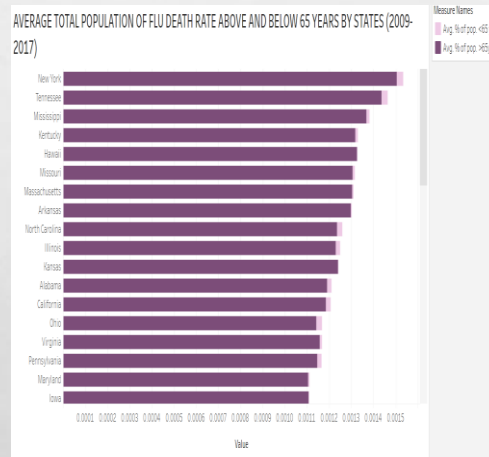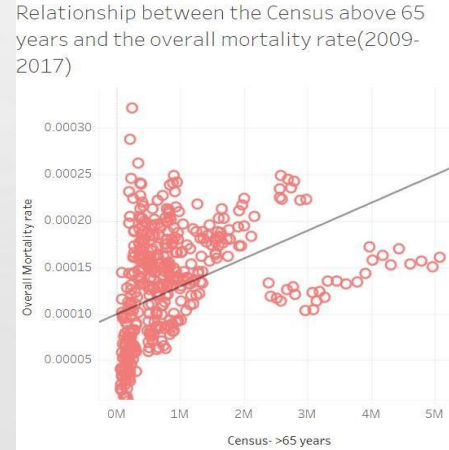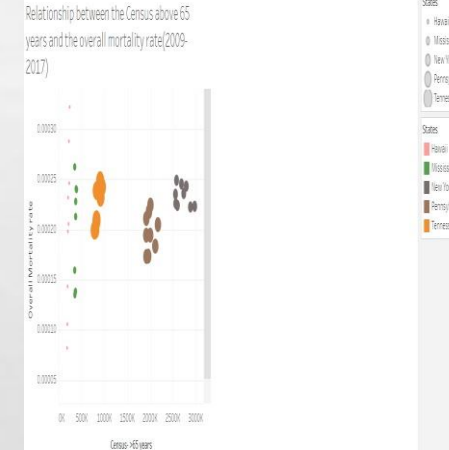

STACKED BAR CHART

BOX PLOT

PIE CHART

TREE MAPS

MAPS

HORIZONTAL BAR CHART

SCATTERPLOT

BUBBLE CHART

DELIVERABLES AND RECOMMENDATION

RESULTS/DELIVERABLES

The presentation is done in Tableau through a [Tableau dashboard](#)

RECOMMENDATIONS

- To deploy more staff to the highest mortality states to prevent more flu deaths.
- Should prepare prior to the season- so sending adequate staff in late November to all the states(especially to those 5 states) will help us to be prepared for the season.
- To create health awareness by publishing more about flu spread and its prevention in newspapers and social media platforms.

ROCKBUSTER STEALTH

PROJECT 3: ROCKBUSTER STEALTH

BACKGROUND & OBJECTIVE

- Rockbuster Stealth is a movie rental company

- The management board expect data-driven answers that they can use for their 2020 company strategy.

KEY QUESTIONS

- Which movies contributed the most/least to revenue gain?
- What was the average rental duration for all videos?
- Which countries are Rockbuster customers based in?
- Where are customers with a high lifetime value based?
- Do sales figures vary between geographic regions?

DATA AND PROJECT

The data set contains information about Rockbuster's film inventory, customers, and payments, among other things

DATA: DATASET
BRIEF: SOURCE
DATA DICTIONARY: SOURCE

TECHNIQUES APPLIED

- Database querying
- Data filtering
- Data cleaning
- Summarizing
- Joining tables
- Subqueries
- Common table expressions

LIMITATIONS

Data has personal details of stores, customers, payments, inventory, etc.
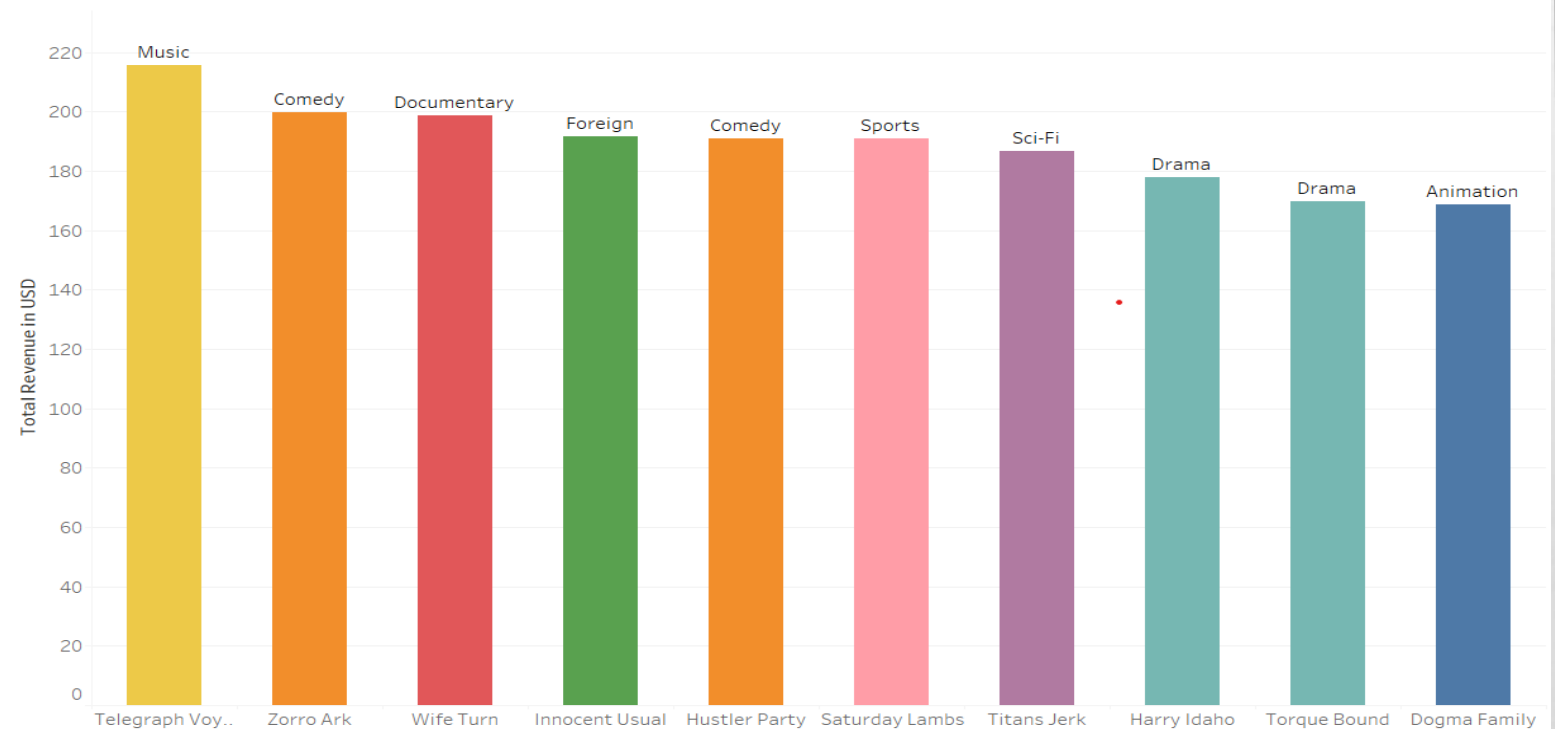
MOVIE TIME

TOOLS USED:

ANALYSIS

DATA OVERVIEW

- Rental rate - $0.99 -- $4.99

- Total no of movies – 1000

- No of active users – 584

- Average film length -- 115 minutes

- Movie ratings -- g (178), pg (194), pg - (223), r (195), nc-17 (210)

- Rental duration -- 3-7 days

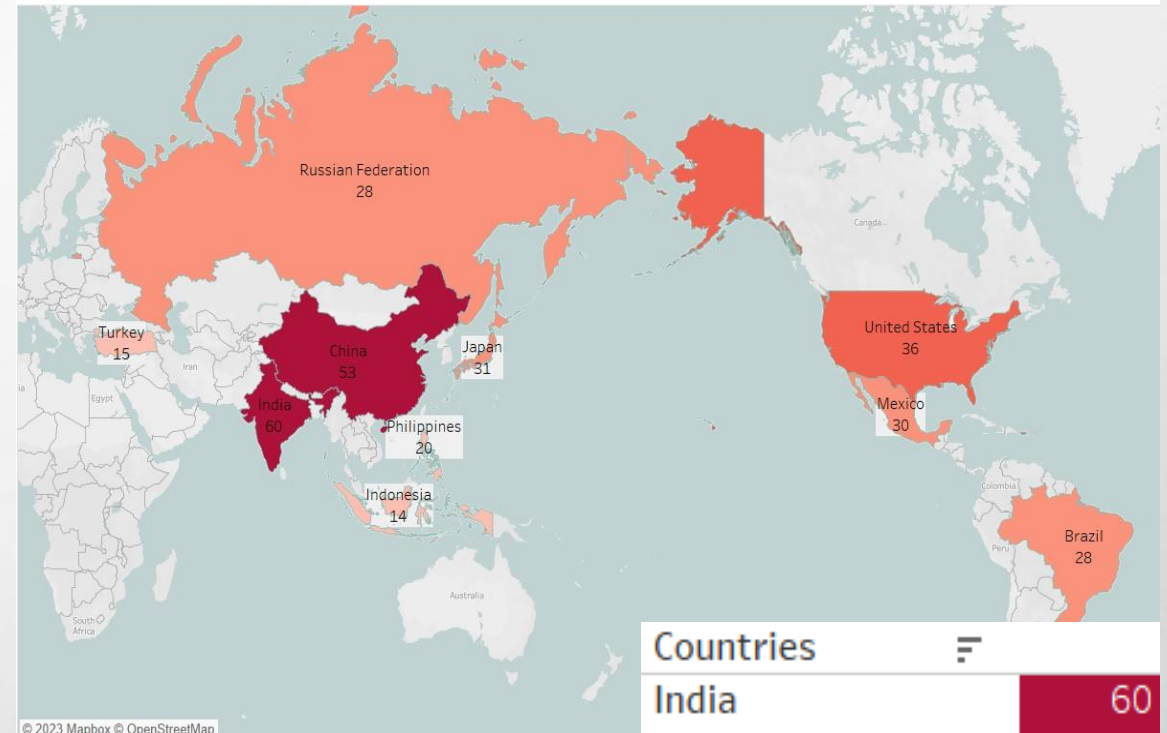Top 10 movies contributing higher for revenue based on revenue

The top movies that contribute the most revenue belong to music, comedy, and documentary genres. The name of the movies is Telegraphic Voyage, Zorro Ark, and WifeTurn.

ANALYSIS PERFORMED

Europe
$12,803.85

South/Latin America
$10,941.00

Africa
$4,782.71

Middle east
$3,528.73

North America
$4,245.01

Asia & Pacific
$24,002.16

Top 10 countries with high customers

Russian Federation
28

Turkey
15

China
53

Japan
31

India
60

Philippines
20

United States
36

Mexico
30

Indonesia
14

Brazil
28

© 2023 Mapbox © OpenStreetMap

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

Most of Rockbuster's total revenue comes from Asia & Pacific, Europe, and South/Latin America.

RECOMMENDATIONS

✓ It's better to remove a few movies which has the least revenue.

✓ Special offers can be introduced to people who hire the movies for more than 5 days.

✓ Should increase the marketing strategies to our major markets Asia & Pacific, Europe, and South/Latin America.

✓ The top 5 countries should be taken into account and we can introduce more movies in their preferred language which can create more interest among people and will improvise our sales.

✓ The high-paid customers should be rewarded for their loyalty and we should also include the referral codes system which helps in increasing the customer numbers into improve South.

DELIVERABLES

(CLICK BELOW)
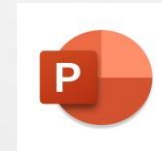
PROJECT 4- INSTACART BASKET ANALYSIS

BACKGROUND AND OBJECTIVE

- Instacart, is an online grocery store that operates through an app.
- Perform an initial data and exploratory analysis of some of their data in order to derive insights and suggest strategies.

KEY QUESTIONS

- What are the busiest days of the week?
- When do people spend the most money?
- Are there certain types of products that are more popular?
- Customer's loyalty status?
- Are there differences in ordering habits based on a customer's region?
- Is there a connection between age and family status in terms of ordering habits?
- What differences can you find in the ordering habits of different customer profiles?

DATA AND PROJECT

- It's an open-source data set from Instacart.

DATA: CUSTOMERS
DEPARTMENTS
ORDERS
PRODUCTS
DATA DICTIONARY:
SOURCE
PROJECT BRIEF: SOURCE

DATA LIMITATIONS

- Data only contains Records from 2017

TECHNIQUES APPLIED

- Data cleaning: wrangling and subsetting
- Data consistency check
- Combining and exporting data
- Deriving new variables
- Grouping data and aggregating variables
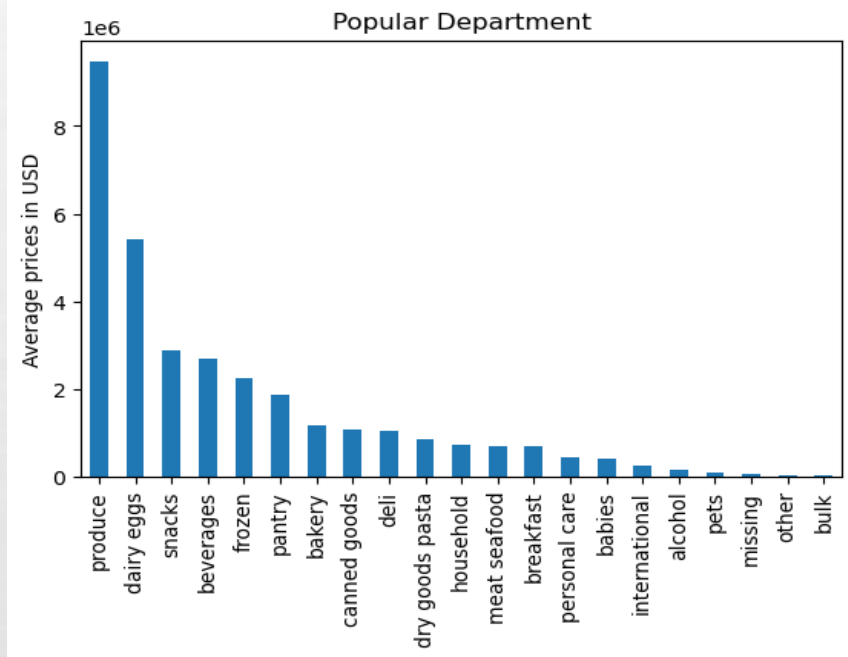- Python visualization and excel report

TOOLS USED:

DEPARTMENT ANALYSIS

Popular Department

Department Profiles by Region
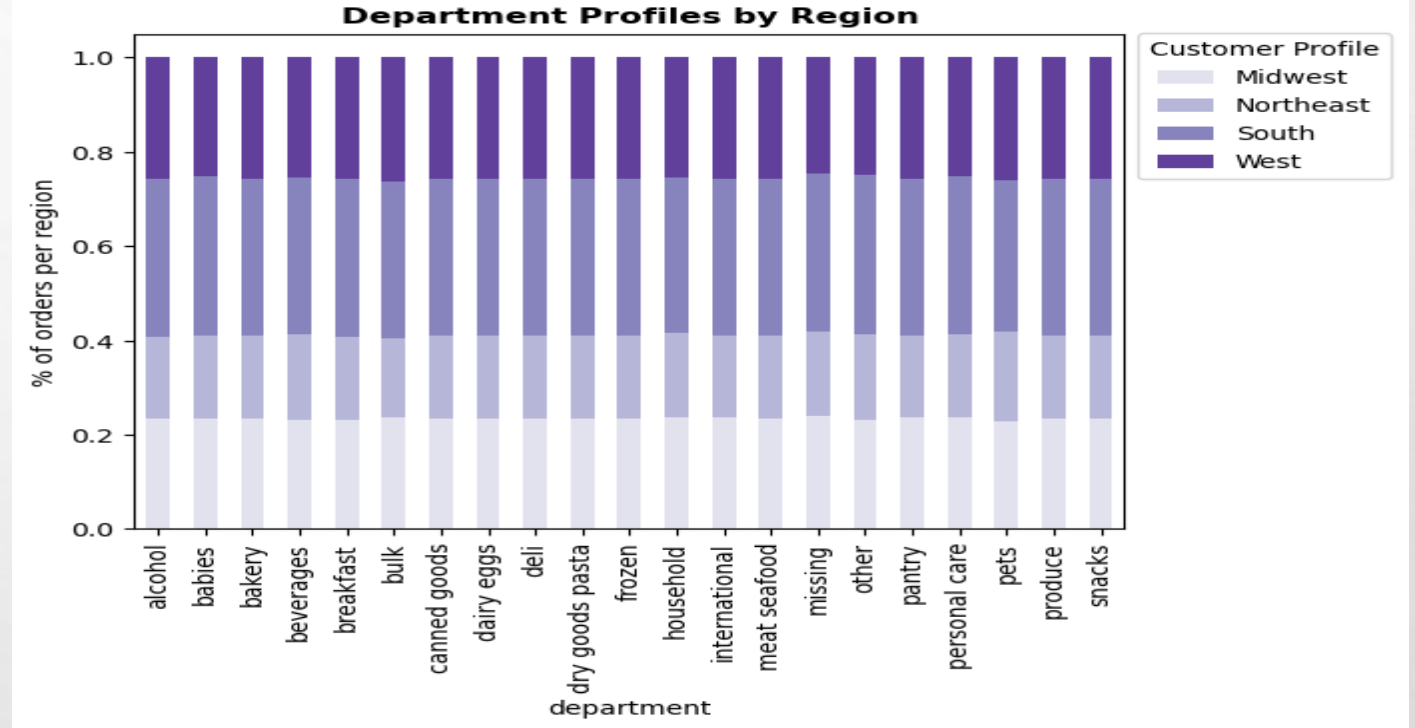
The most popular department with high purchase frequency is "produce, dairy eggs, snacks".
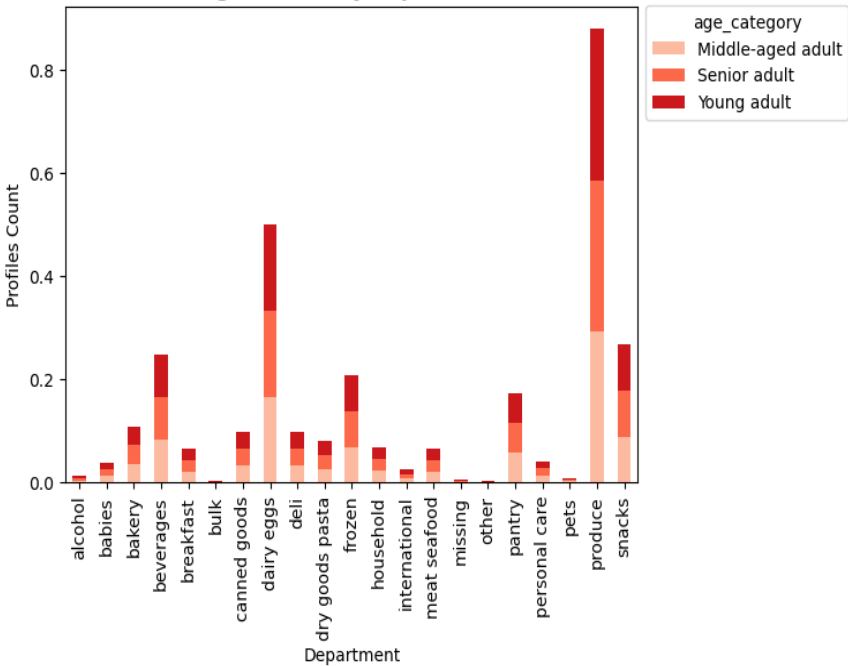The least sold department is "Pets, other, bulk".

The ordering habits across all departments remain consistent for all four regions.
The south region represents the highest number of orders.

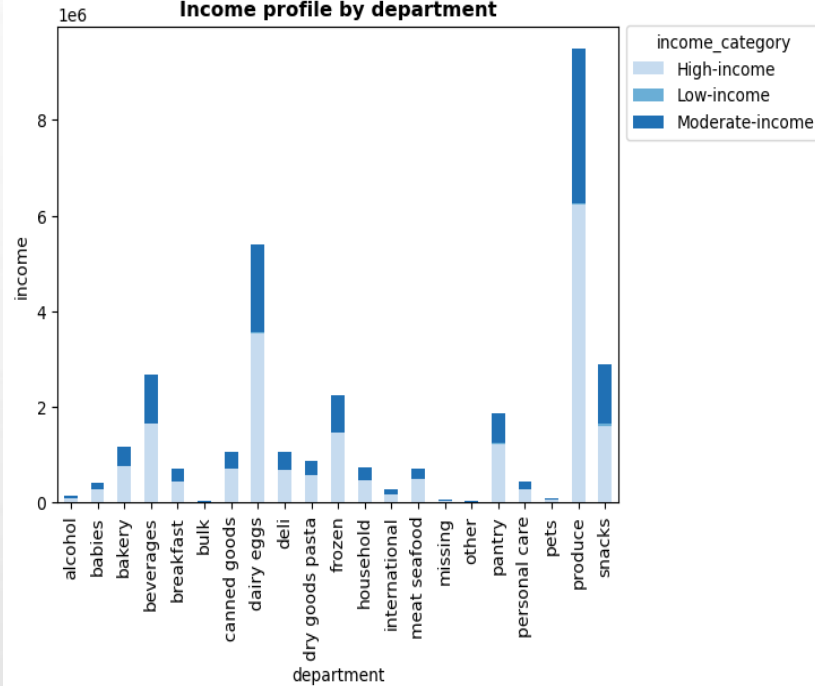# CUSTOMER BEHAVIOURAL ANALYSIS

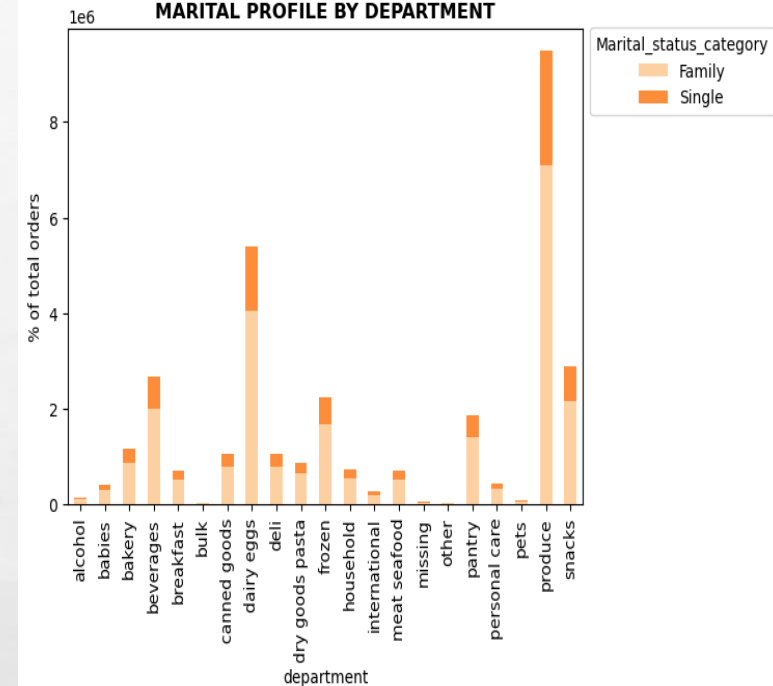
The top 5 departments based on profile counts are produce, dairy eggs, snacks, beverages, and frozen.

Produce, dairy eggs, and snacks are the top three departments across all income profiles and the least three departments are pets, missing, and others.

From the stacked bar chart, Families and singles mostly prefer Produce, dairy, and snacks department products.

RECOMMENDATIONS

- Create more marketing campaigns for young age groups by introducing trendy pay-later options.

- Instacart must enhance the marketing for the least sold products such as pets, babies, and alcohol.

- Family-oriented promotions by developing marketing campaigns that cater to the needs of families.

- For young mothers promoting the baby's products to them are advised.

DELIVERABLES
(CLICK BELOW)

PIG E. BANK

PROJECT 5 – PIG E. BANK

PROJECT OBJECTIVE

Providing analytical support to its anti-money-laundering compliance department and helping build and optimize models that assist the bank in running its compliance program more efficiently.

DATA AND PROJECT

PROJECT BRIEF – SOURCE
DATA SET - DATA

LIMITATIONS

Customers' demographic data are limited.
The data is limited to the ATMs within the 100-meter radius of the border.

TECHNIQUES APPLIED

- Data cleaning, sorting, and checking for integrity.
- Big data management
- Data ethics & data mining
- Predictive analysis
- Time series analysis and forecasting

TOOLS USED:

PREDICTIVE ANALYSIS

Descriptive Statistics Combined

ExitedFromBank	Count of ExitedFromBank?	Average of Estimated Sa	Average of I	Average of NumOfI	Average of HasCrCard?	Average of Balance	Average of Tenure
Stayed	787	98943.39	0.56	1.54	0.71	74830.87	5.16
Left	204	97155.20	0.30	1.46	0.71	90239.22	4.72
Grand Total	991	98574.54	0.51	1.52	0.71	78002.72	5.07

Descriptive Statistics - Left

| ExitedFromBank	Count of ExitedFromBank?	Average of Credit Score	Average of E	Average of NumOfI	Average of HasCrCard?	Average of IsActiveMember
Left	204	636.51	90239.22	1.46	0.71	0.30
Grand Total	204	636.51	90239.22	1.46	0.71	0.30

Based on Countries

Country	Average of Age	Average of Credit Score	Average of E	Average of IsActive	Average of Estimated Sa	Average of NumOfProducts	Average of HasCrCard?
France	39.09	645.86	60180.81	0.50	97160.96	1.53	0.72
Germany	39.02	648.37	119206.47	0.50	102782.95	1.52	0.73
Spain	37.58	653.67	69991.51	0.53	96998.13	1.51	0.67
Grand Total	38.68	648.51	78002.72	0.51	98574.54	1.52	0.71

Based on Gender

Gender	Count of Gender	Average of Balance	Average of A	Average of Credit S	Average of Estimated Sa	Average of IsActiveMember
Female	462	80670.84	38.96	646.57	98623.73	0.48
Male	528	75815.84	38.48	650.50	98573.61	0.53
N/A	1	0.00	24.00	497.00	76390.01	0.00
Grand Total	991	78002.72	38.68	648.51	98574.54	0.51

Decision Tree

The descriptive statistics were calculated for all the variables for those who stayed in the back as well as who left.
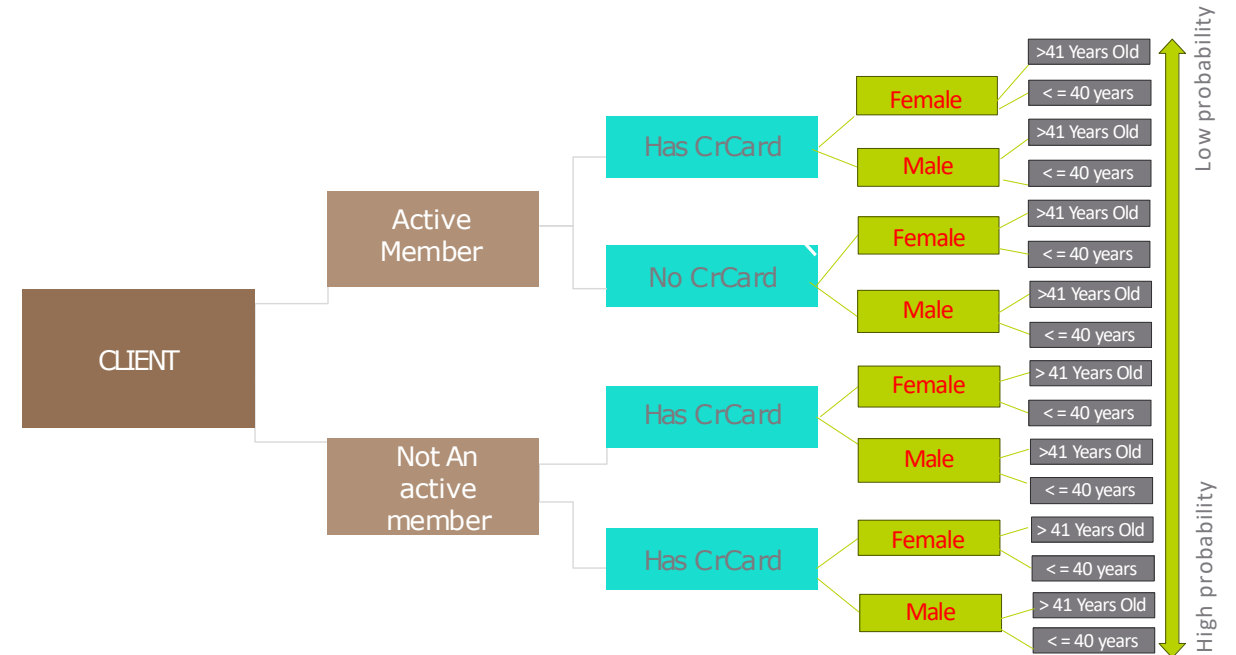
Nonactive female customers above 41 years category have a high probability of leaving the bank.

DAIRY PRODUCTS CONSUMPTION IN INDIA (2019-2022)

PROJECT 6: DAIRY PRODUCTS CONSUMPTION IN INDIA (2019-2022)

PROJECT BACKGROUND

India is the world's largest milk producer with almost 22% of global production

- To understand the sales pattern to improvise the pricing strategies among different brands.

KEY QUESTIONS

- Which products have the highest sales?
- Sales performance among different brands based on location, land area, and cow population.
- How to manage the inventory for stocks?
- Forecasting the sales for the upcoming year 2023.
- What are the customer preference and buying behaviour?

DATA SOURCE

It's an open-source data set from Kaggle.
DATA SOURCE: LINK
DATASET: SOURCE

LIMITATIONS

- There might be some typography errors.
- As it has time data there can be some lag in the duration.
- The data set contains records from 2019- 2022. If we had more data from past years, it would be better for the analysis.

TECHNIQUES APPLIED

- Exploratory visual analysis
- Created correlation matrices and heatmaps
- Created a choropleth map in Python
- Regression analysis (supervised machine learning)
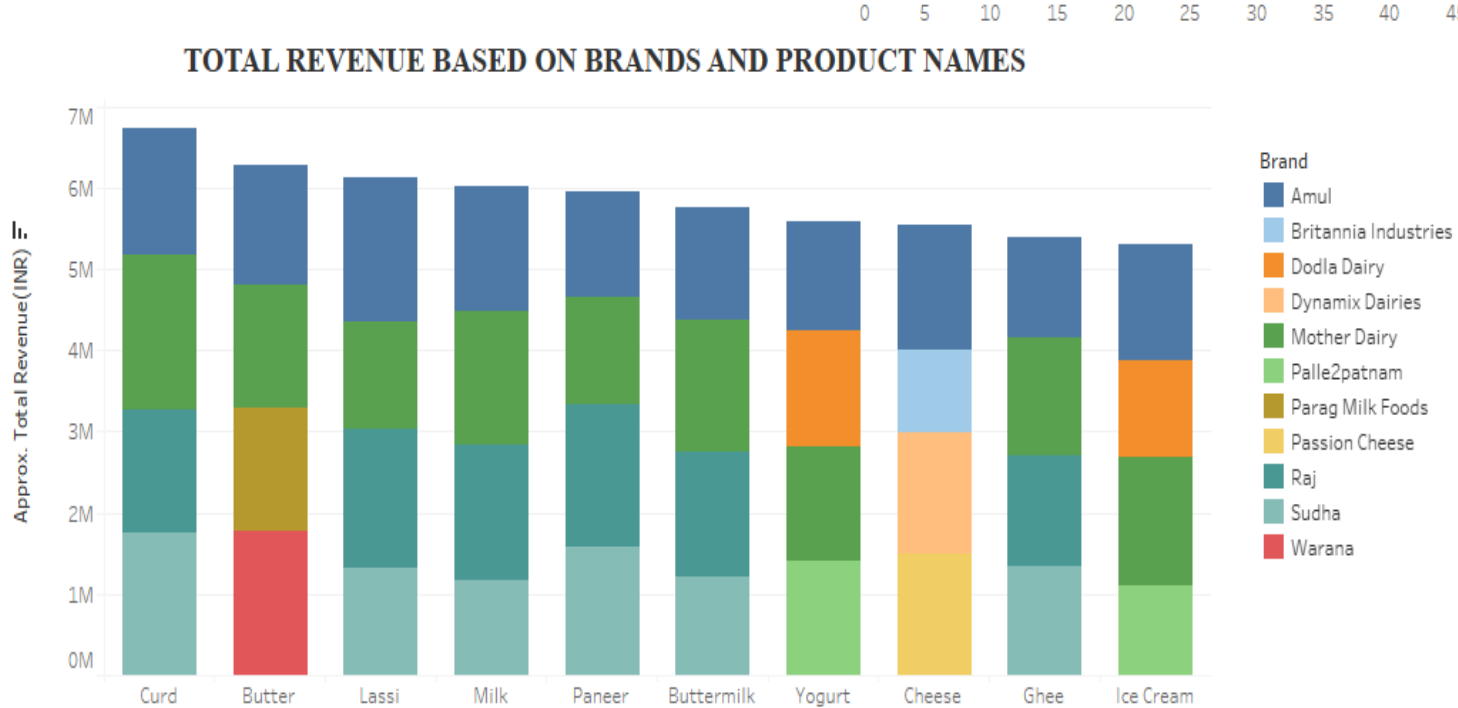- Cluster analysis (unsupervised machine learning)
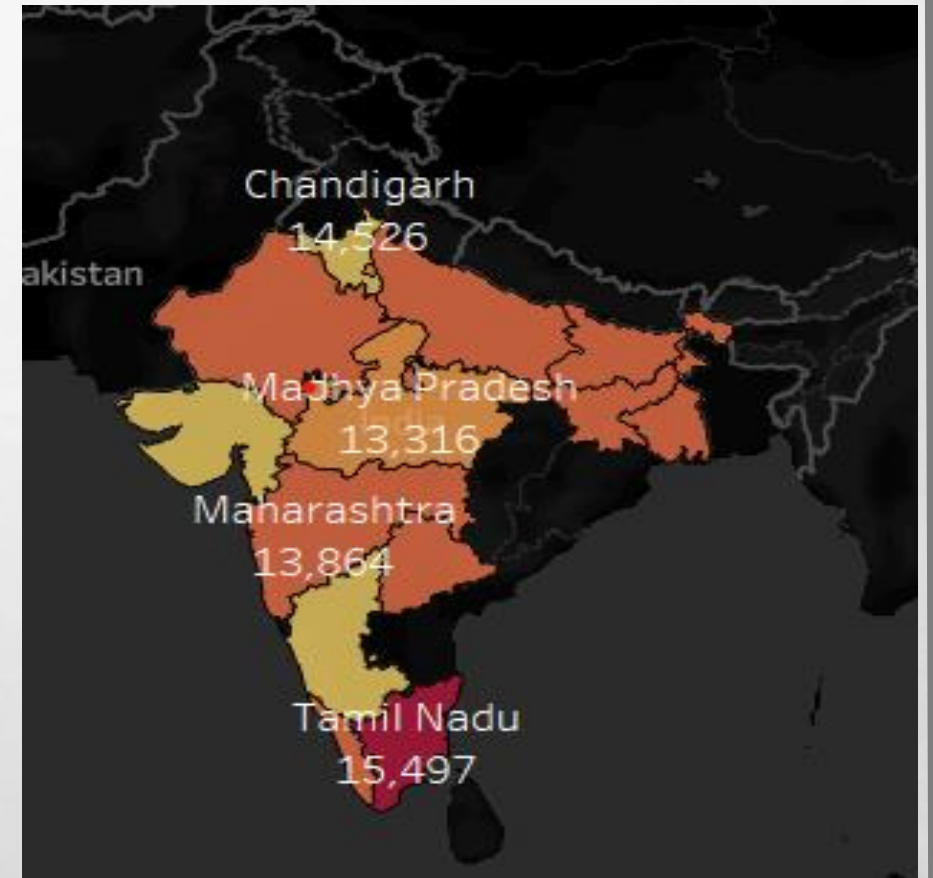- Time series analysis

TOOLS USED:

EXPLORATORY ANALYSIS


TOTAL REVENUE BASED ON BRANDS AND PRODUCT NAMES

THE DAIRY PRODUCTION IN DIFFERENT STATES OF INDIA AND THEIR AVERAGE REVENUE PER DAY

- The most sold product which contributes to the highest sales revenue is Curd followed by it is butter and lassi.
- Amul, Mother Dairy, and Raj are the top three brands with the highest revenue.
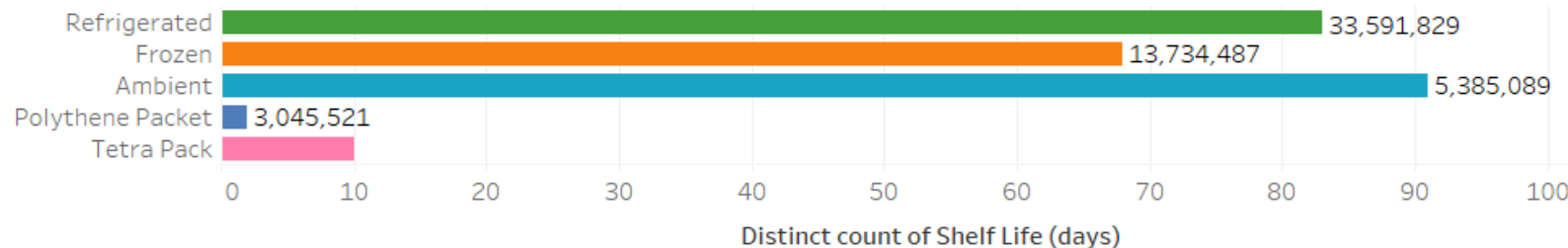
INVENTORY MANAGEMENT

DETERMINING THE SHELF LIFE OF STOCKS
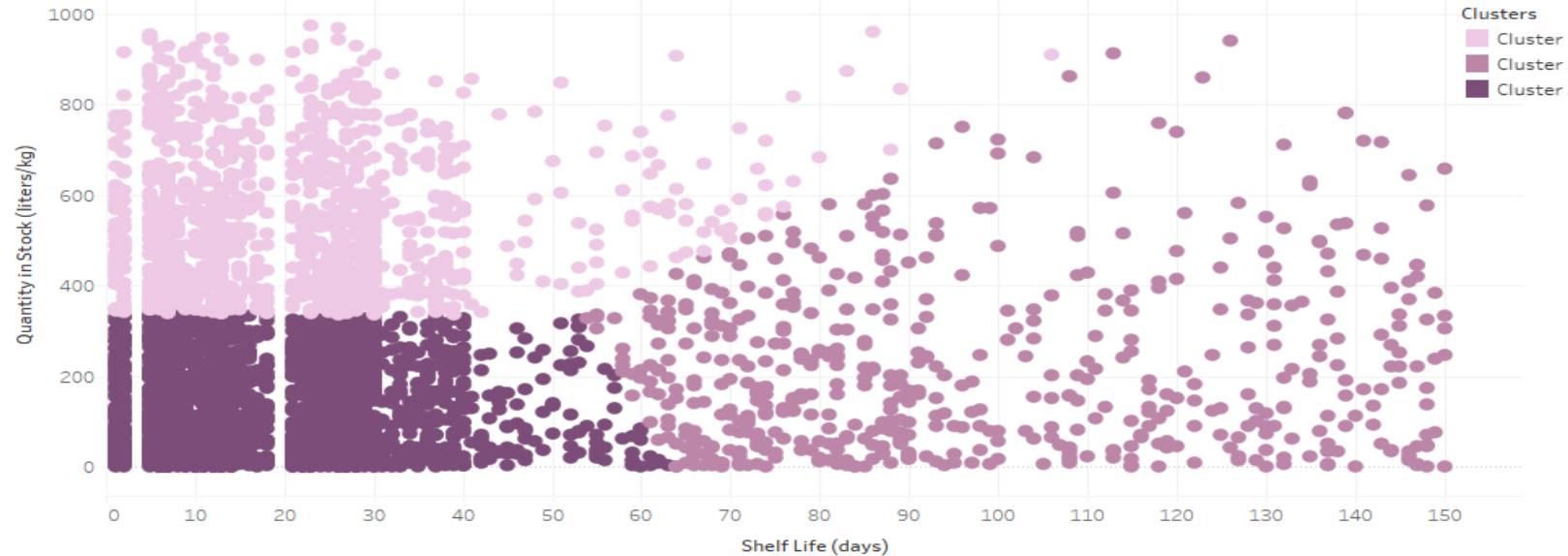
The clustering plot between "shelf life" and "quantities in stock" shows the milk products in stock have a shelf life of 0 to 40 days.

SHELF LIFE BASED ON STORAGE CONDITION

| Storage	Distinct count of Shelf Life (days)
Refrigerated	33,591,829
Frozen	13,734,487
Ambient	5,385,089
Polythene Packet	3,045,521
Tetra Pack	

Based on shelf life, the most durable storage method is Ambient storage for up to 90 days.
Based on sales, the predominant sales are driven by refrigerated, followed by it is frozen and ambient products.

FORECAST FOR 2023

FORECASTING SALES BY BRANDS FOR THE UPCOMING YEAR(2023)
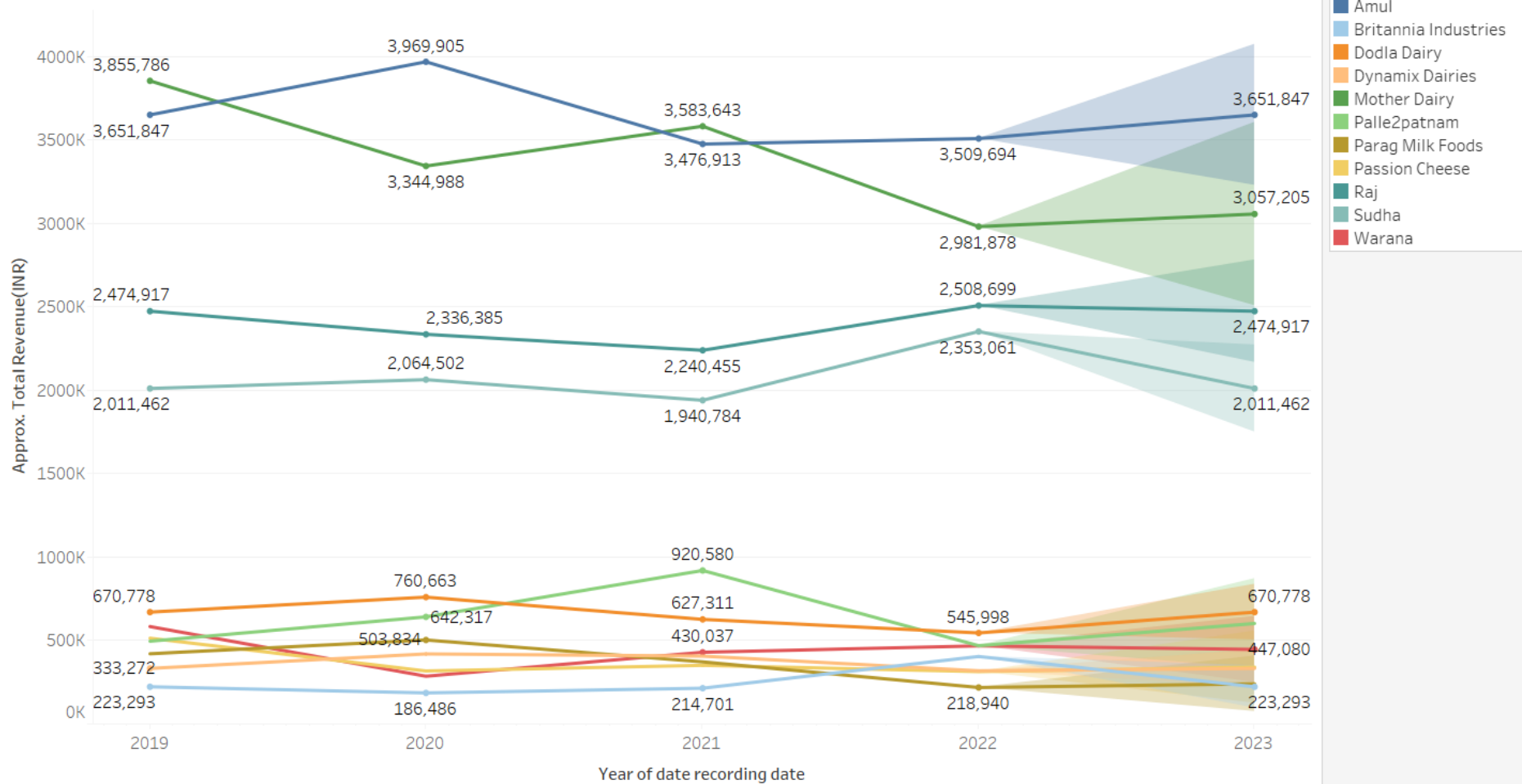
Based on the forecast predictions, there may be a mild increase in sales in most of the dairy brands including the top brands like Amul and Mother Dairy by 2023 in India.

RECOMMENDATIONS

- It is recommended to expand sales into other states and union territories.
- We should create more marketing campaigns for moderately performing brands in order to improve their sales revenue.
- It is important to note that customers tend to prefer moderately priced products, so we should focus on marketing high-priced products by offering promotions for a limited time, which can help boost sales.
- As we live in an online generation, it would be beneficial to promote more online methods for purchase and delivery.
- India has many regional languages, so it would be wise to advertise in the regional language of each area, which can help customers better understand the products. Lastly, it is essential to focus on the states with high sales and create more marketing campaigns in areas where sales are significantly lower.

DELIVERABLES

PYTHON SCRIPTS:

TABLEAU LINK:

FOR FURTHER QUERIES

KINDLY CONTACT

drpadmasini.v@gmail.com