DATA ANALYTICS PORTFOLIO

Ankita Bangalore Shashidhar

PROJECTS

TOOLS

01 GameCo

Global market analysis of video game sales.

Excel

02 Preparing for Influenza Season

Staff deployment planning for influenza season.

Excel, +ableau

03 Rockbuster

Launching Rockbuster Stealth online movie service

PostgreSQL, +ableau, Github

04 Instacart

Market segmentation analysis to uncover sales.

Python, Github

05 PigE.Bank Analysing customer attrition.

Excel, Github

06 World Happiness Analysis

Python, +ableau

01 GAME CO. MARKET ANALYSIS

PROJECT OVERVIEW

Background: Game Co. is a leading global gaming company with a strong presence in key markets, including North America, Europe, Japan, and other regions. The company offers a wide variety of games across multiple genres, available for both purchase and rental, catering to a diverse audience of gamers worldwide.

Objective: The primary objective of this project was to perform an in-depth analysis of historical game sales and rental data to identify critical trends and patterns. By extracting actionable insights from this comprehensive dataset, the aim was to develop predictive models that could forecast market reception for upcoming game releases. These insights are intended to guide strategic decision-making, optimizing the timing and positioning of future games in different markets.

LINKS : Project Brief Final Presentation Project Reflection

Technical Summary

DATA

- **Data source**: VGChartz.
- **File**: Excel - CSV
- **Period of data**: 1980-2016
- **Regions**: North America, Europe, Japan, and others
- **Information**: title, platforms, year, genre, publisher

SKILLS APPLIED

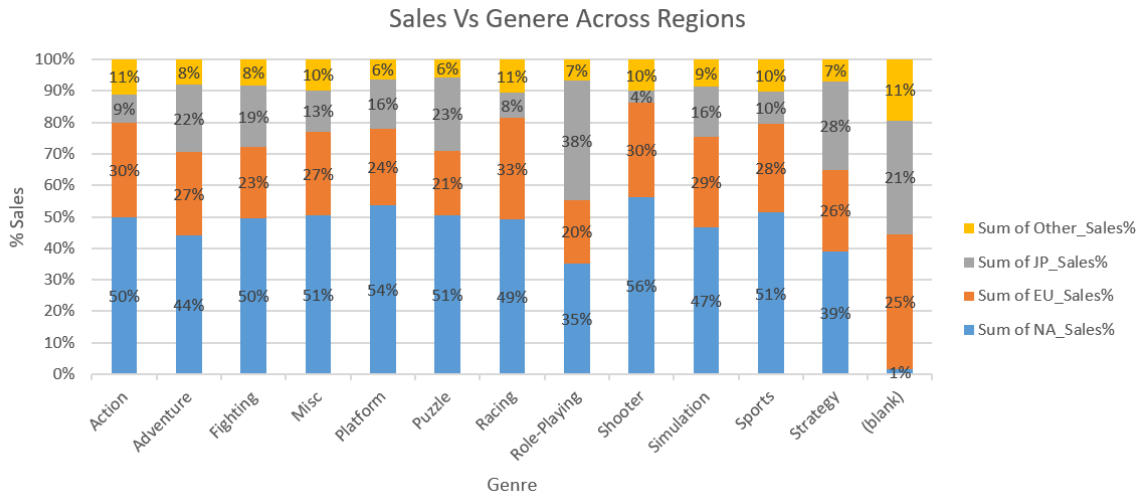
For this project, I used Excel.

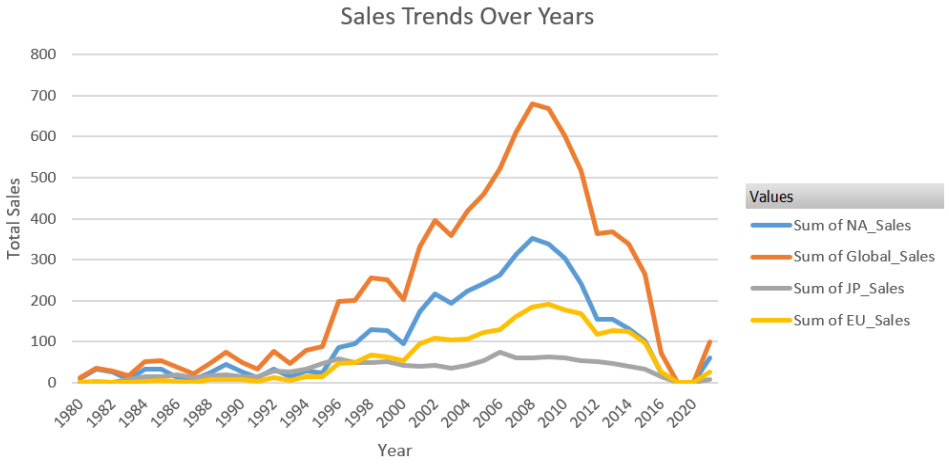
- Improving data quality
- Data grouping and summarizing
- Descriptive analysis
- Pivot table
- Visualization results in Excel
- Presenting results

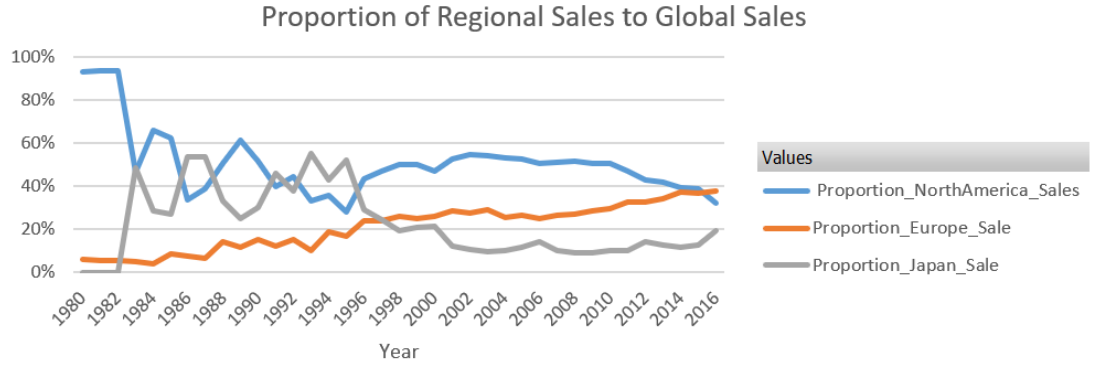
ANALYSIS

The final analysis highlights popular game genres and regional sales variations since 1980, combining detailed regional analysis of top genres and platforms with comprehensive sales history and market share data.

This page will feature a selection of example graphs from the extensive report and visualizations.







Conclusion

- Europe sales have increased steadily over the years and surpassed North America in market share post-2015.
- North American market sales have declined over the years.
- While Action, Shooter, and Sports dominate in North America and Europe, In Japan Role-Playing games is the most preferred genre.

Recommendation

- Marketing strategies to match regional sales trends and shifts in market preferences post-2015 is preferred.
- Advertising campaigns that align with regional tastes and current trends to boost engagement and market share is encouraged.
- In North America and Europe increase focus on Shooter and Sports genres, while in Japan increase focus on Role Playing games.
- With increase in European Sales, Game Co. should allocate more marketing budget for Europe.

LINKS : Project Brief Final Presentation Project Reflection

02 PREPARING FOR INFLUENZA SEASON

PROJECT OVERVIEW

BACKGROUND: During the flu season, U.S. healthcare facilities often see a significant increase in patient volume. Hospitals and clinics need additional staff to adequately treat these extra patients. The medical staffing agency provides this temporary staff.

OBJECTIVE: This project aimed to develop a staffing schedule that optimally addresses these demands. The goal was to enhance flu season preparedness by efficiently managing the staffing needs of clinics and hospitals. This initiative was motivated by the need to support healthcare providers during peak periods when vulnerable populations are at greater risk of serious flu-related complications. The goal is to determine when to send staff and how many to each state.

LINKS : Project Brief Interim Report Tableau Storyboard Presentation Recording

Technical Summary

DATA

- Census Population Dataset

Source: US Census Bureau Contents: Population information from the US by country, time, age and gender for 2009-2017.

- Influenza Death Dataset

Source: CDC Contents: Information about influenza deaths by age groups in the US by state and time for between 2009-2017.

SKILLS APPLIED

- Translating business requirements into analytical questions
- Sourcing relevant datasets
- Data integration and cleaning
  Statistical hypothesis testing
- Visual analysis in Tableau
- Forecasting
- Storytelling in Tableau
- Presenting results to an audience

Research Hypothesis

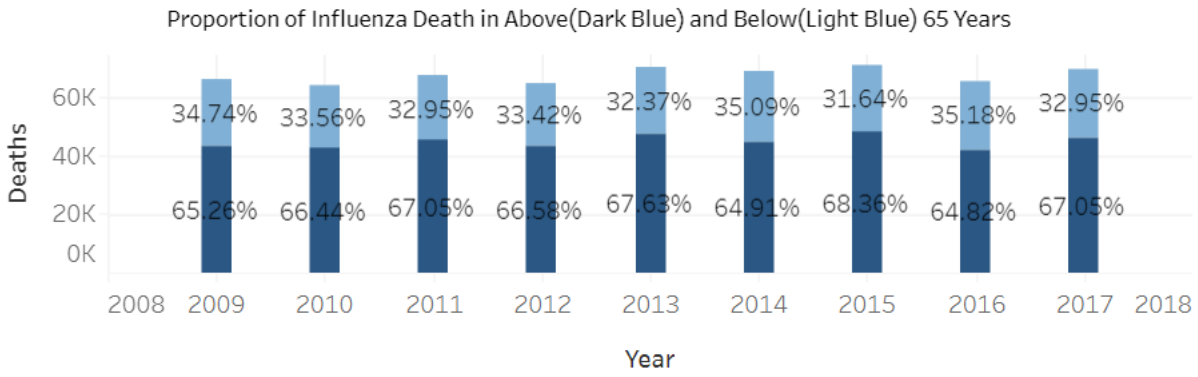
If the age of patient is more than 65 years, then the death rate due to influenza is more.

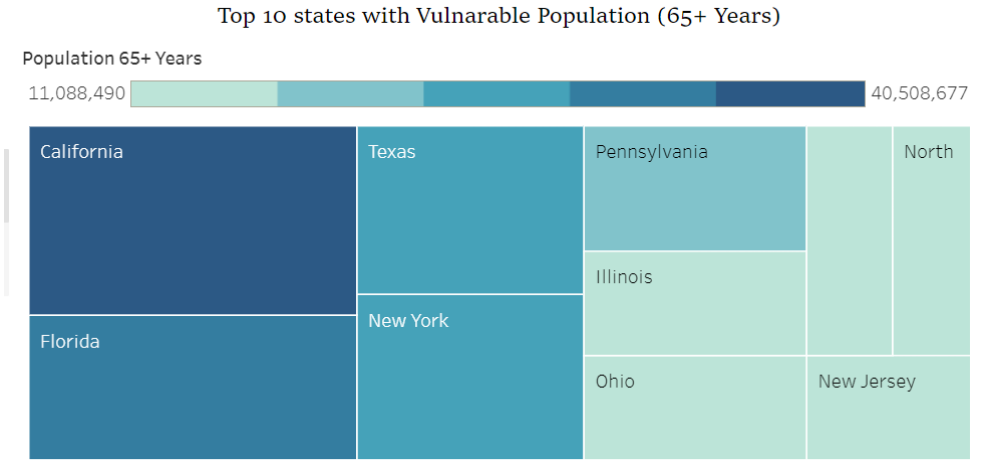
Preparing for Influenza Season

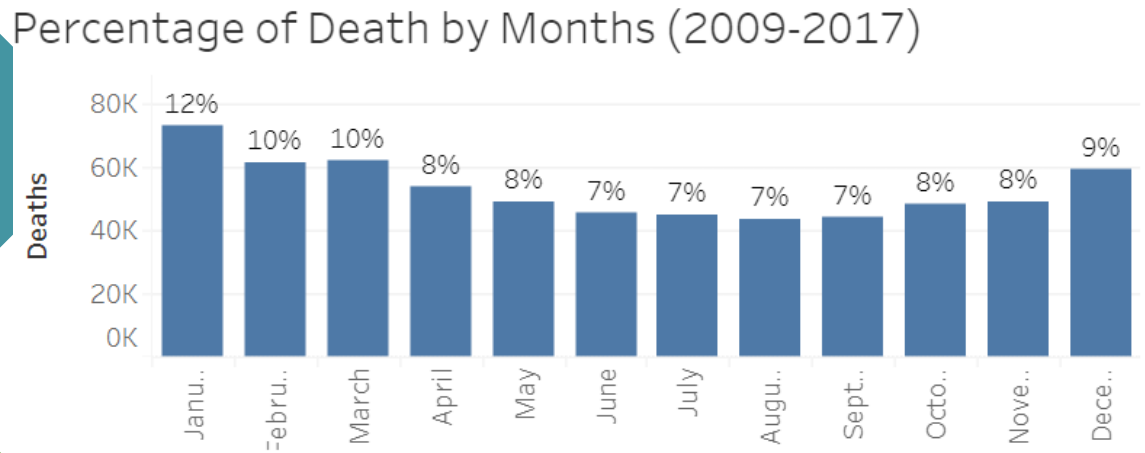
ANALYSIS

In this analysis, I forecasted staffing needs for a medical staffing agency to optimize resource allocation during peak flu periods. Staffing allocation was based on the vulnerable population of 65+ as this was the age group experiencing the most severe disease.

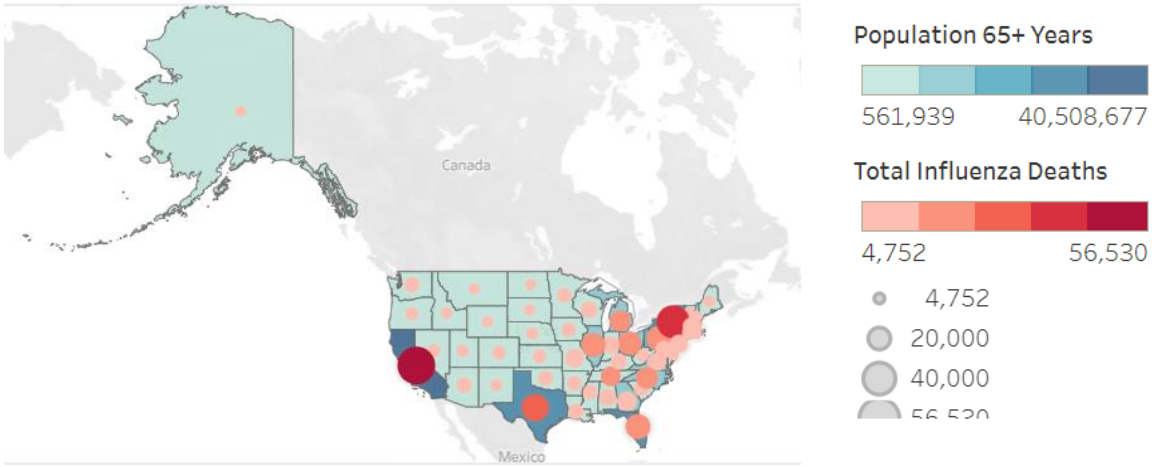
This page features a selection of example graphs from the extensive report and visualizations.







States with high Mortality and Vulnarable Population



Conclusion

- Influenza is mostly seen during winter season. Death rate increases from September and starts declining from March. 40% of Influenza Deaths occur during December to March.
- The population over 65 years are vulnerable with 66% of the total mortality.
- States with a larger vulnerable population have higher mortality rate and hence require more medical staff.
- California, New York, Texas, Pennsylvania and Florida are top 5 states with high mortality and larger vulnerable population

Recommendation

- Be prepared for high demand in staff during December to March and moderate demand during September, October, November and April.
- Main focus should be on states with higher mortality and larger vulnarable population.
- Further analysis can be done on Vaccination rate, Infection rate and other vulnarable groups (children under 5 years, pregnent women, individuals with pre-medical conditions)
- Additionaly a survey can be done to evaluate the preparedness for the influenza season.

LINKS : Project Brief Interim Report Tableau Storyboard Presentation Recording

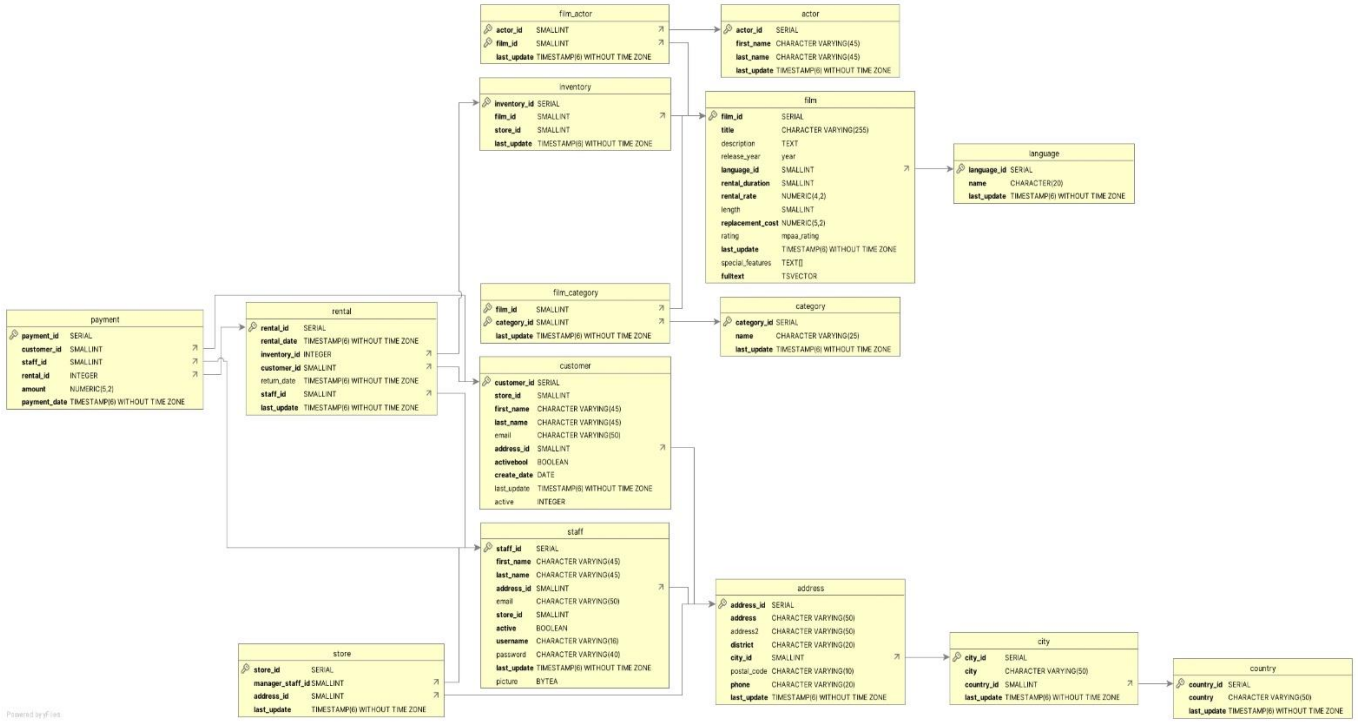
03 Rockbuster Stealth

PROJECT OVERVIEW

BACKGROUND: Rockbuster Stealth LLC, a global movie rental company with traditional physical stores, faces increasing competition from digital streaming platforms such as Netflix and Amazon Prime. In response, Rockbuster is considering a transition to an online video rental model.

OBJECTIVE: This project aims to use data analytics to address key business questions regarding sales, revenue, customer behaviour and inform Rockbuster's strategy for transitioning to an online service model in 2020.

Rockbuster Entity Relationship Diagram



LINKS : Project Brief Data Dictionary Final Presentation Technical File Tableau Storyboard Github Repository

Technical Summary

DATA

- Rockbuster Stealth Data Set

Contents: Movie information, customer data, rental records, financial data, and inventory data

SKILLS APPLIED

- Relational databases
- SQL
- Creating a data dictionary
- Database querying
- Data filtering
- Data cleaning and summarizing
- Joining tables
- Subqueries
- Common table expressions
- Presentation

Rockbuster Stealth

ANALYSIS

In this analysis, Rockbuster's data was managed within a PostgreSQL database, enhancing the data integrity with data cleaning and validation. With SQL uncovered key insights about top-performing movies and strategic customer locations, which I visualized using Tableau. The comprehensive analysis is documented in a detailed report with SQL code in Excel and supported by an ERD and data dictionary.

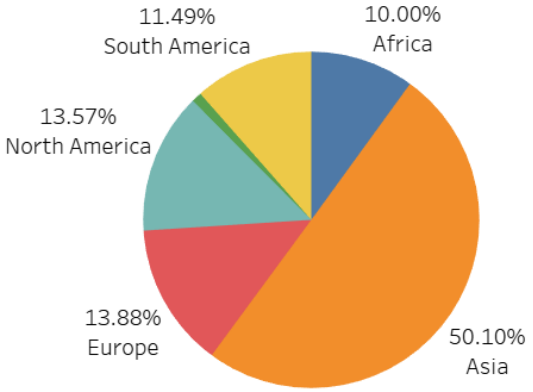
This page features a selection of example graphs from the extensive report and

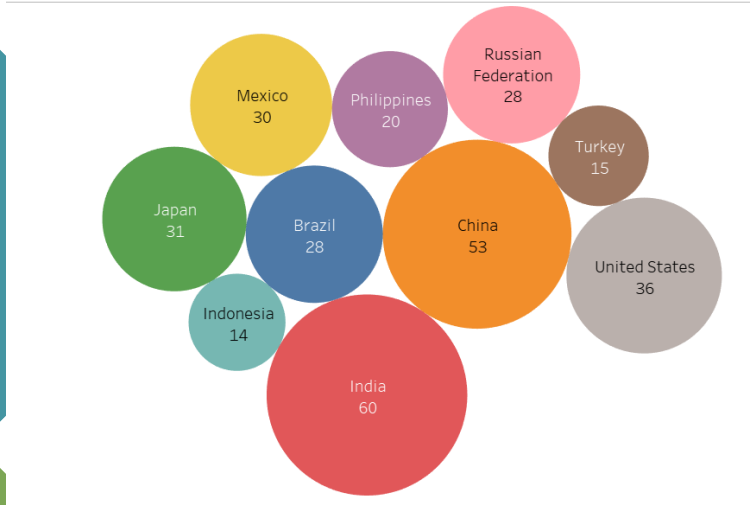
Highest Revenue: **India**

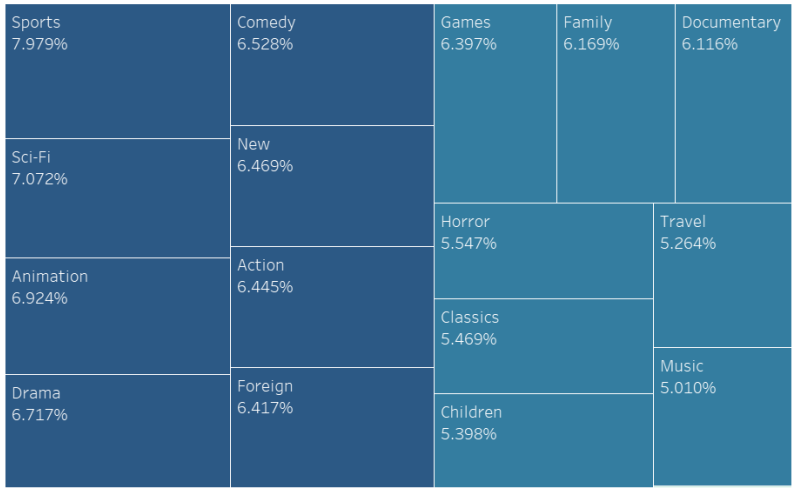
Highest Grossing Category: **Sports**

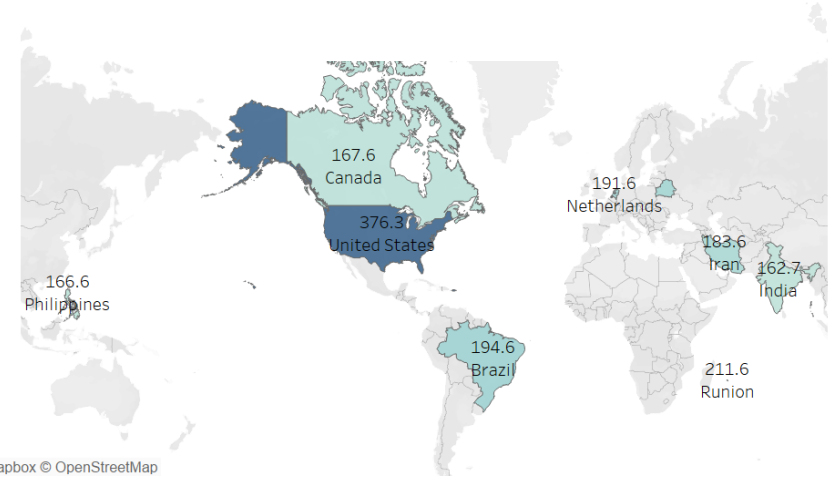
High LifeTimeValue Customer Location : **USA**

Highest Revenue Region : **Asia**









Conclusion

- Telegraph Voyage contributed the most revenue gain while Duffel Apocalypse contributed the least.
- The average rental duration is 5 days.
- Rockbuster customers are located all around the world.
- Customers with high lifetime value are located in United States followed by Reunion.
- Asia contributes highest with nearly 50% of the total revenue followed by Europe.
- The most popular movie category is PG-13.
- Our database includes 599 customers across 108 countries.
- There are 1000 movies distributed across 20 genres.
- Rental rates are almost evenly distributed between $0.99, $2.99, and $4.99, with average being $2.98.

Recommedation

- **Geographic Market Expansion:** Top Regions: Asia (50.10%), Europe (13.88%) and North America (13.57%) are the strongest revenue drivers. Loyalty programs in these regions can solidify customer retention and could yield substantial returns. Underperforming Regions: Consider strategies to boost revenue in regions like Oceania and Africa, which currently have lower revenue contributions.
- **Focus on Top -Performing Genres:** Since PG -13 and NC -17 rated movies are the highest revenue generators, Rockbuster should promote movies with these ratings. Additionally, exploring why G -rated movies generate less revenue could lead to strategies to boost their performance.
- **Targeted Marketing by Rating:** High revenue generating genres like Sports, Sci -Fi, Animation, Drama, and Comedy should be expanded and promoted.
- **Enhance Customer Experience:** Focus on high lifetime value customers in countries like Reunion, the United States, and Brazil to deepen customer relationships and retention.

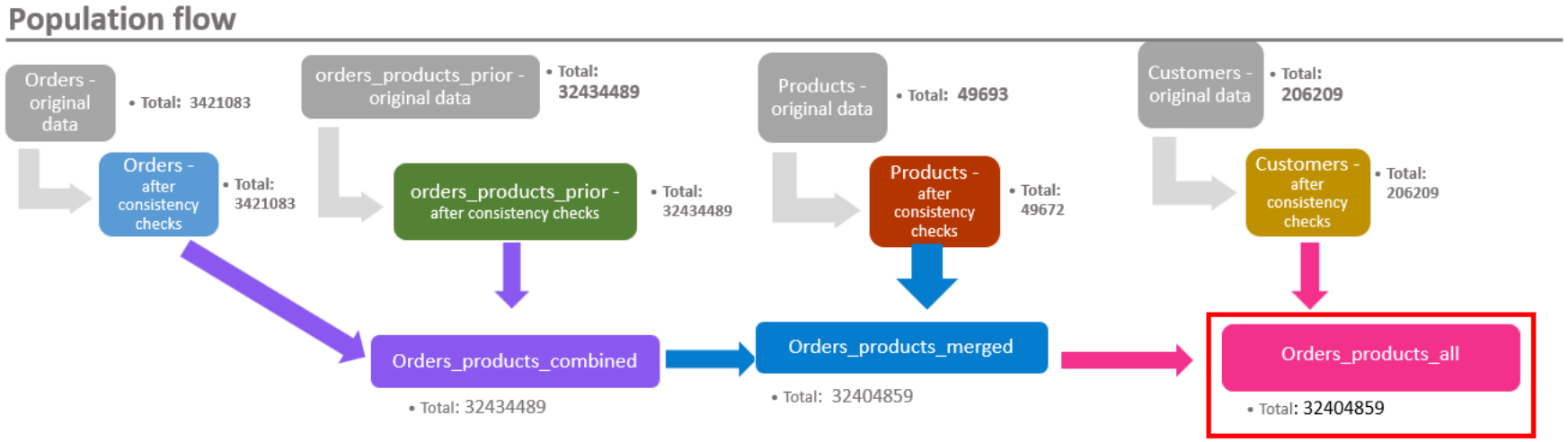
LINKS : Project Brief Data Dictionary Final Presentation Technical File Tableau Storyboard Github Repository

04 Instacrt

PROJECT OVERVIEW

BACKGROUND: Instacart, a leading online grocery store, enables customers to order groceries through an app. The company aimed to boost sales by adopting a more strategic approach to targeting customers and improving segmentation through the analysis of historical data.

OBJECTIVE: The aim of this project was to conduct a detailed analysis of Instacart's sales data to enhance their marketing strategy, focusing on more precise customer targeting and effective segmentation to drive sales growth.



LINKS : Project Brief Final Report GitHub Repository

Technical Summary

Data

Data Sets:
Customers: Analyzed for purchasing patterns and loyalty.
Orders: Studied to determine busy times and spending habits.
Products: Categorized to understand popularity and sales impact.
Departments: Analyzed for sales volume per department.
Data Citations: "The Instacart Online Grocery Shopping Dataset 2017", Accessed via Kaggle. "Customers Data Set", Provided by CareerFoundry

Skills Applied

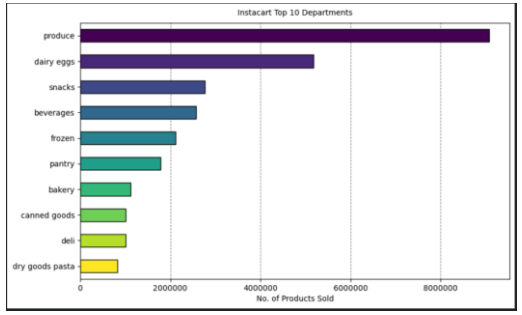
Python Programming:
- Utilized in Jupyter Notebook for all coding and analysis tasks.
- Data wrangling, subsetting, filtering, and summarizing with Pandas.
- Data merging and consistency checks to ensure accuracy.
- Deriving new variables and complex data transformations.
- Grouping and aggregating data for detailed insights.

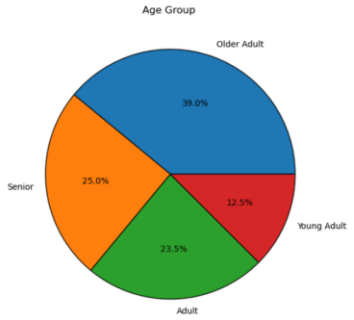
Data Visualization:
- Advanced visualizations created using Matplotlib and Seaborn.
- Presentation of analytical results through clear and effective charts.

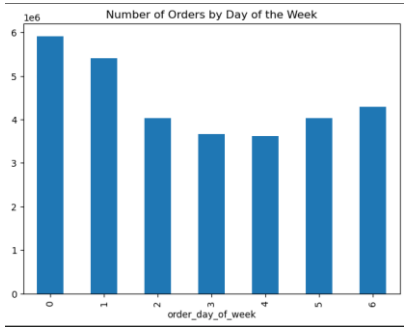
ANALYSIS

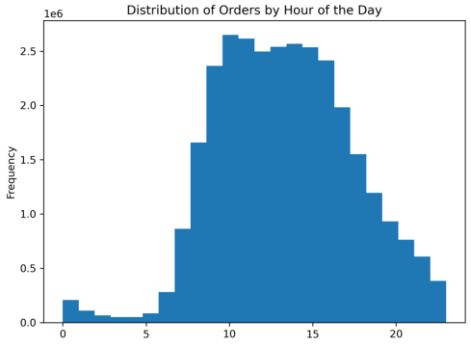
Using pandas cleaned and merged multiple datasets and performed data wrangling, aggregation, and column derivation. Removed outliers to refine the dataset. The comprehensive analysis provided actionable insights for crafting targeted marketing campaigns.

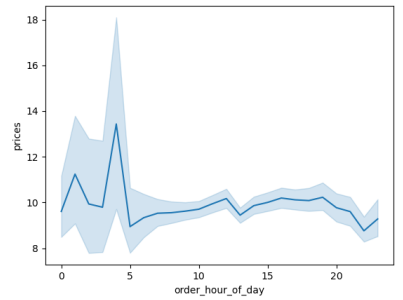
This page features a selection of example graphs from the extensive report and visualizations.

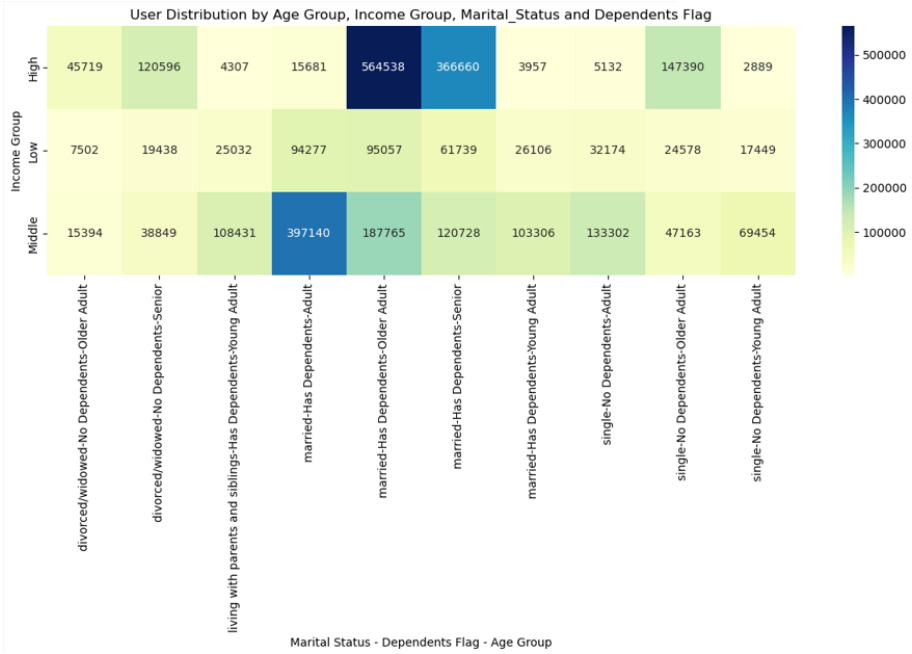


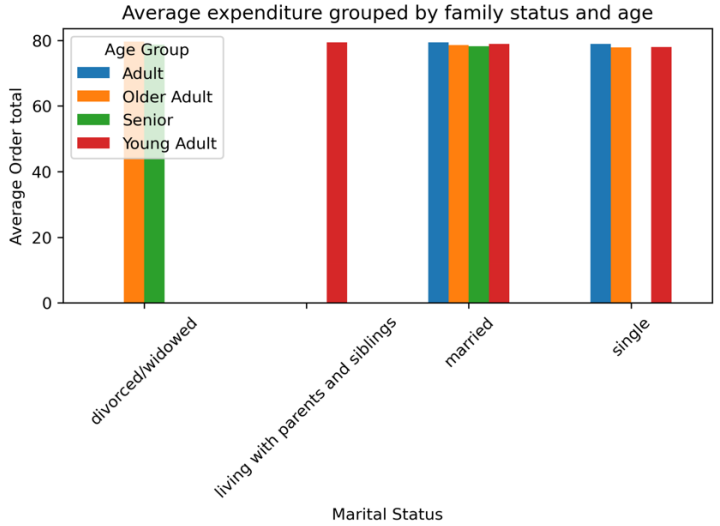


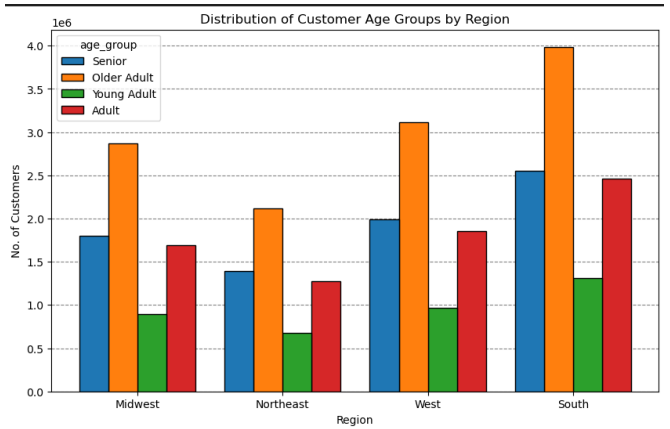












Conclusion

- The busiest days of the week are Saturday and Sunday between the hours of 8 am to 4 pm. The least busy day is Wednesday between the hours of 2 am to 5 am.
- People spend the most money during the hours of 8am to 4pm. 9 am is when people spend the most money.
  The top 5 products are produce, dairy and eggs, snacks, beverages and frozen food.
- Majority of customers purchase mid range products.
- Total Product Orders are more in southern states followed by West, Midwest and East. However average order total is almost same across all region and most customers are low spenders
- 'Older Adult' Age Group Customers, married and have dependents and belong to high income group contribute maximum to sales.
- There is a positive correlation between age and income. As the age increases, income also increases.

Recommendation

Marketing Sales

- Schedule promotional ads and promote early bird discounts particularly on slower days like Tuesday and Wednesday during the times 6 am to 8 am and 5pm to 10pm
- Expand Niche Categories: Increase visibility and appeal of underperforming departments like "Personal Care" and "Pets & Babies" by offering targeted promotions or introducing new products that cater to specific needs, particularly for low-income or niche customer segments.
- Develop Regional Marketing Strategies: Tailor campaigns to regional differences, emphasizing high-income and senior profiles in the South and West. In regions where customer profiles are underrepresented, explore new product offerings or localized promotions to stimulate growth.

Customer Profile

- Aim to convert the decent flow of its new customers to regular customers or even loyal customers through promotional welcome incentives.
- Target High-Income Seniors: Focus marketing efforts on high-income senior profiles across all regions. These customers consistently show high spending in the grocery category, making them a key segment for loyalty programs, premium offerings, and exclusive deals.
- Provide bulk buying options, or subscription services to small families. These strategies could increase purchase frequency and basket size among this consistently reliable customer base.

LINKS : Project Brief Final Report GitHub Repository

05 PIG E. BANK

PROJECT OVERVIEW

BACKGROUND: Pig E. Bank is a global bank dedicated to providing exceptional financial services.

OBJECTIVE: The aim of this project was to perform an in-depth analysis of customer satisfaction data at Pig E. Bank. This analysis targeted the identification of factors leading to customer attrition with the ultimate goal of developing robust strategies to improve customer retention

LINKS : Project Brief Project Report

Technical Summary

Data

Data source: Career Foundry

Skills Applied

- Big data
- Data ethics
- Data mining
- Predictive analysis
- Time series analysis and forecasting

PIG E. BANK

ANALYSIS

Key Factors influencing customers churn

		Loyal	Exited
	Female	43.44%	59.31%
Gender	Male	56.56%	40.69%
	Not Active	43.77%	70.10%
Activity	Active	56.23%	29.90%
	1	46.82%	69.61%
Num. Of	2	52.54%	15.69%
Products	3	0.64%	13.73%
	4	0.00%	0.98%
Credit Card	No Credit Card	29.26%	29.41%
Holders	Has Credit Card	70.74%	70.59%
	France	51.27%	37.75%
Country	Germany	23.03%	36.76%
	Spain	25.70%	25.49%
	10-19	0.38%	0.00%
	20-29	18.45%	5.39%
	30-39	49.11%	24.02%
Age Group	40-49	22.52%	41.67%
	50-59	5.60%	18.63%
	60-69	2.54%	10.29%
	70-79	1.15%	0.00%
	80-90	0.25%	0.00%

In this analysis, the goal was to determine key risk-factors that make a customer more inclined to churn. In order to do this, data was segmented into two groups for analysis: Loyal clients and Exited clients. From here, the exited client's data was then elevated to detect patterns that may contribute to a customer churning. Finally, a model was constructed using a decision tree to systematically predict the determinants of customer churn based on the identified patterns.

Customer Profile

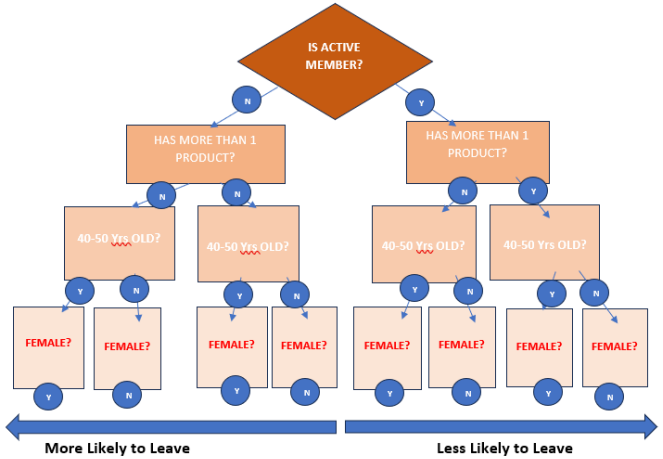
Loyal Clients

	Credit Score	Age	Tenure	Balance	NumOfProducts	HasCrCard?	IsActiveMember	Estimated Salary
Min	376	18	0	$0	1	0	0	371.05
Max	850	82	10	$ 213146.2	4	1	1	199725.39
Mean	648	39	5	$77962.026858	1.52	0.71	0.5	98574.54

Exited Clients

	Credit Score	Age	Tenure	Balance	NumOfProducts	HasCrCard?	IsActiveMember	Estimated Salary
Min	376	18	0	$ 0	1	0	0	371.05
Max	850	82	10	$ 213146.2	4	1	1	199725.39
Mean	648.46	38.97	5.07	$ 77916.642998	1.52	0.71	0.5	98731.58

Decision Tree



Conclusion

- Being an inactive member seems to be a major contributing factor in leaving the bank.
- A higher proportion of the people who left the bank are above age 45.
- Majority of the former customers held only one product.
- More females are among former customers while current customers are predominantly male.

Recommendation

- Increase customer engagement through loyalty programs, personalized offers, and regular communication.
- Develop tailored financial products and services that cater specifically to the needs of people above 45.
- Encourage product diversification among customers , with special focus on users with none or only one product.
- Understand the specific needs and preferences of female customers.

LINKS : Project Brief Project Report

06 WORLD HAPPINESS ANALYSIS

PROJECT OVERVIEW

BACKGROUND: The analysis of world happiness is rooted in understanding the well-being and quality of life across different nations. Several factors influence happiness, including social support, health, income, freedom, trust in government, and perceptions of corruption. It is interesting to explore relationships between these factors and happiness levels, aiming to derive insights that could guide policies, interventions, and investments.

OBJECTIVE: The objective of this project is to quantify and compare Happiness Levels Globally, analyse contributing factors and identify key trends and patterns and to explore how GDP of a country relate to happiness..

LINKS : Github Repository Tableau Storyboard

Technical Summary

DATA

Source:
https://www.kaggle.com/datasets/sazidthe1/global-happiness-scores-and-factors

This dataset provides valuable insights in understanding the dynamics of happiness and well-being worldwide. This dataset includes key metrics related to global happiness across all the regions for 9 years from 2015-2023.

SKILLS APPLIED

- Defining business questions and objective.
- Sourcing open data
- Data integration, cleaning and transformation.
- Exploring data, data wrangling and consistency checks.
- Graphical Visualization in python.
- Regression Analysis.
- Cluster Analysis through K-means algorithm.
- Sourcing and Analysis Time series data.
- Creating data dashboards and Visual analysis in Tableau.

Research Hypothesis

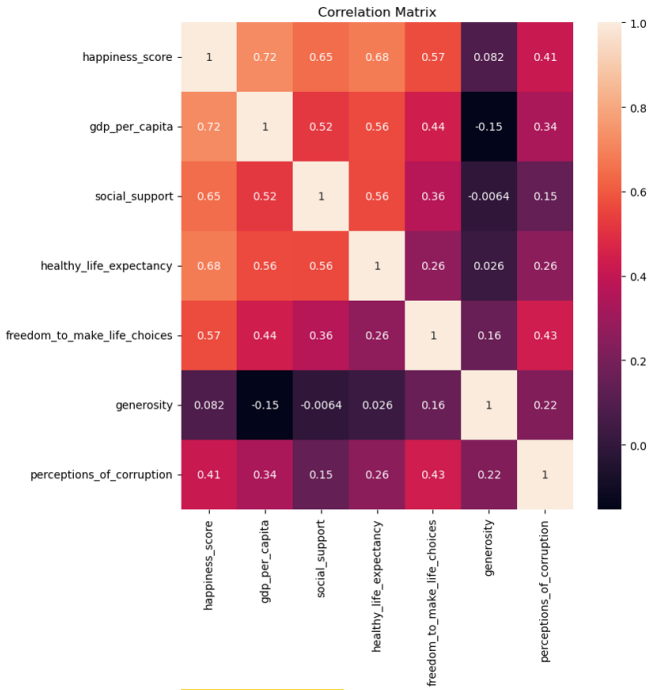
If the GDP per capita of a country increases, then Happiness score also increases.

ANALYSIS

As part of exploratory analysis, performed Linear Regression to discover that, the relation between GDP per capita and Happiness score is not linear. To explore data further, I conducted Cluster Analysis using k-means algorithm which showed how Happiness score varies widely based on GDP per Capita.

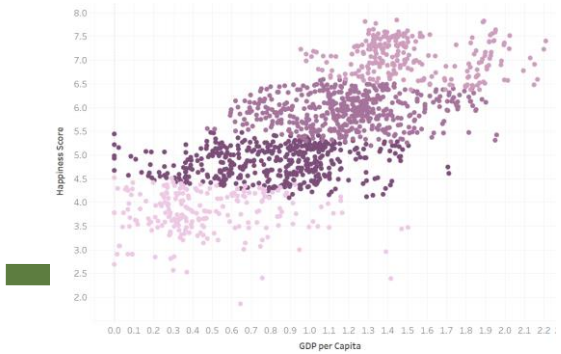
This page features a selection of example graphs from the extensive report and visualizations.

Correlation Matrix



Linear Regression

Cluster Analysis



Conclusion

- Happiness of a country is influenced by different factors including GDP per Capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption.
- We performed exploratory analysis to discover how Happiness score and factors influencing Happiness score are related.
- The Linear Regression showed that there is no linear relation between Happiness score and GDP per Capita. The Cluster Analysis, shows how Happiness score varies widely based on GDP per capita.

Recommendation

- Analyze Subnational Data: If available, explore happiness scores and influencing factors at the state, province, or city level to uncover localized trends.
- Compare Regions: Identify disparities between regions or countries with similar economic conditions to isolate cultural or governance-related drivers of happiness.
- Forecast Future Trends: Use predictive modelling to estimate future happiness scores based on current data and trends in key factors.
- Expand Data Sources: Include other datasets, such as education levels, employment rates, environmental quality, and crime statistics, to enrich the analysis.

LINKS : Github Repository Tableau Storyboard

THANK YOU

Ankita Bangalore Shashidhar