

KESHAV DEWAN

DATA ANALYTICS PORTFOLIO



ABOUT ME

I am an **Engineer** with a professional experience in the **Real Estate Industry**.

In my **previous roles**, I have worked in the **Research & Consulting** field where I provided **Data Insights** to the clients through MS-Excel & PowerPoint, which eventually led me to study about the Data Analytics Field.

In Data Analytics, I have been lucky to work on certain projects that helped me in developing my skills of **Data Cleaning, Wrangling**, conduct **Descriptive Analysis** and further work on **SQL, Python** and **Tableau** to create meaningful **visualizations** and in turn providing important **insights** to the client.

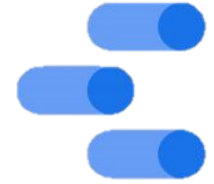
ANALYTICS TOOLS



Tableau



PowerBI



Google Data Studio



PostgreSQL



SQL



Python



Excel



PowerPoint



Anaconda



PORTFOLIO PROJECTS

GameCo

Descriptive analysis on global video game sales to inform development into new games

RockBuster LLC

Movie rental company planning to launch an online video streaming service

Pig E. Bank

Provide analytical support and assess client risk, provide data insights against money laundering, using big data concepts

Influenza Season

Helping a medical staffing agency for proper workforce allocation during Influenza Season in the US

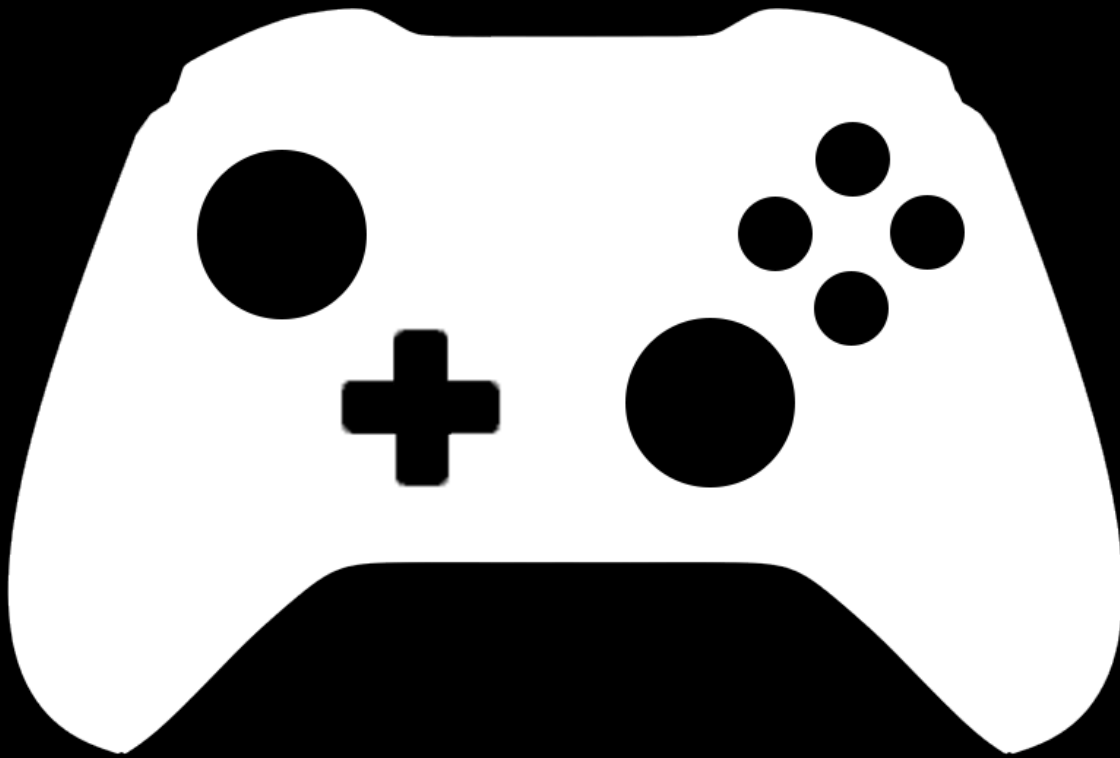
Instacart

Conduct an initial data and exploratory analysis for an online grocery store in order to derive insights and suggest strategies

World Happiness Index

Build an interactive dashboard that will visually showcase well-curated results of an advanced exploratory analysis conducted in Python



GAMECO



GameCo, a video game company, wants to use data for the development of new games.

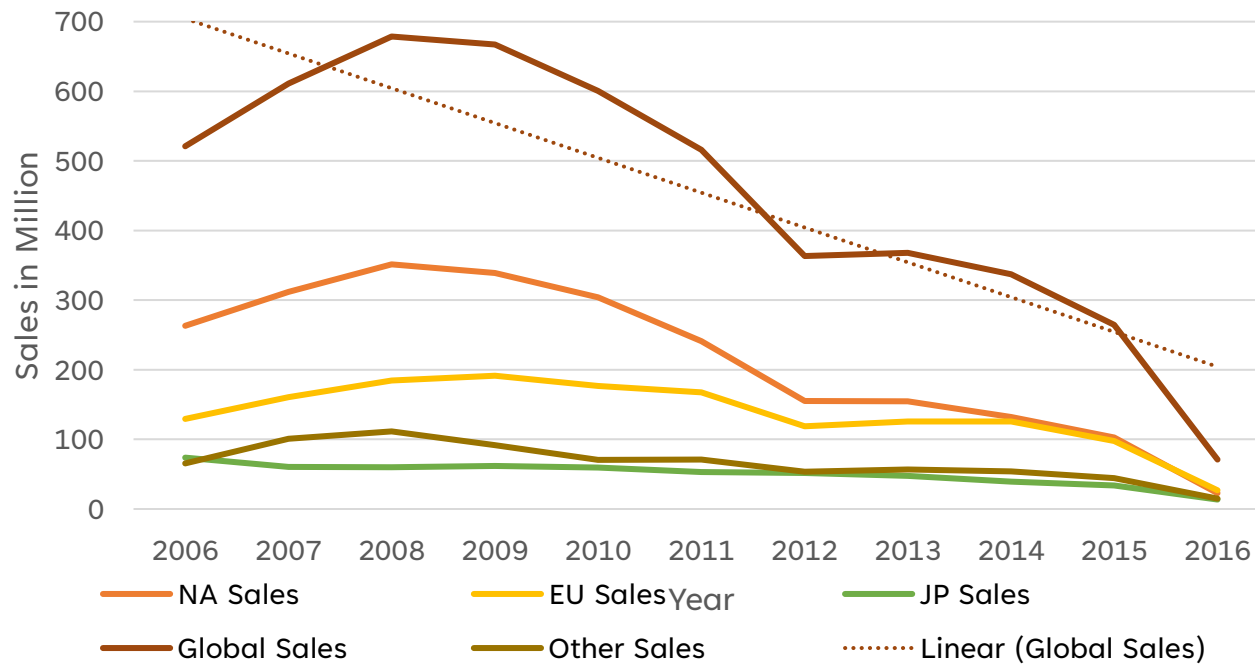
The executives at GameCo are interested in knowing the most popular games available, any games that have witnessed a decrease in popularity, varying sales across regions.

PROJECT OVERVIEW

OBJECTIVE	Perform analysis of video game data set to foster a better understanding of how GameCo's new games might fare in the market.
DATASET	Data was drawn from the website VGChartz VG Sales Data Set
TOOLS	 
SKILLS	Grouping data, Summarizing data, Descriptive analysis, Visualizing results in Excel

GAMECO ANALYSIS

Regional Sales (Million Units)



NA – North America

JP – Japan

EU – European Union

2012 witnessed a huge **drop** in **global video games sales** of **approx. 53%** as compared to the **peak year of 2008**

North America (32%) has lost the **top spot of global sales** contribution from **European Union (38%)**

Japan has witnessed **maximum volatility** in terms of sales

Action games were **most preferred** followed by **Shooting and Sports Genre**

RECOMMENDATIONS

Marketing budget recommendation

Continue focus on North America and European Union region

Genre recommendation

Focus on genre centric games w.r.t the regions.

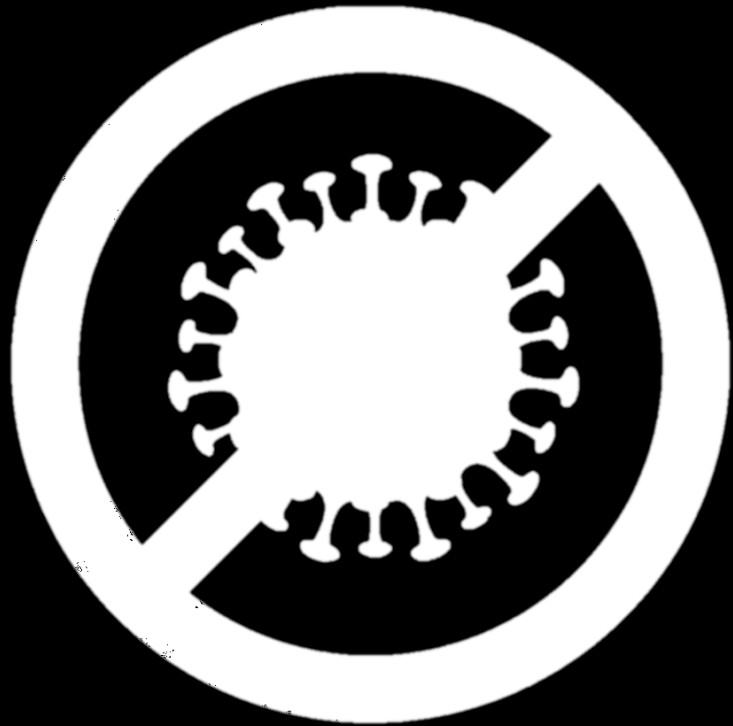
Changing Technology

Start exploring the change in gaming technology – online games, mobile games, cloud gaming etc.

Complete Report link: 



INFLUENZA SEASON



The United States has an influenza season where more people than usual suffer from the flu.

Some people, particularly those in vulnerable populations, develop serious complications and end up in the hospital.

Hospitals and clinics need additional staff to adequately treat these extra patients. The medical staffing agency provides this temporary staff.

PROJECT OVERVIEW

OBJECTIVE — Helping the staffing agency in proper workforce allocation with the available data. 'Fight the Flu'

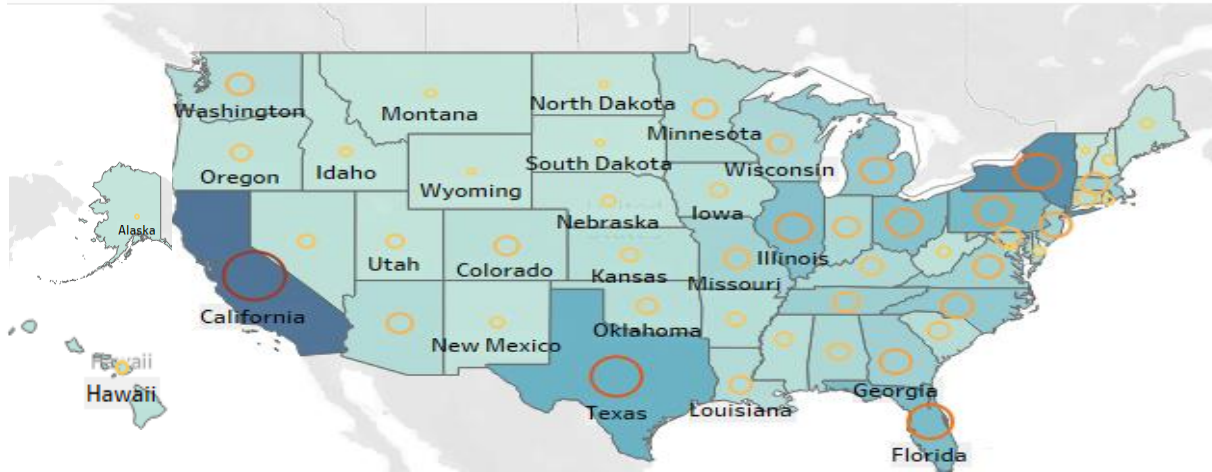
DATASET — Data was drawn from different websites: [CDC](#), [US Census Bureau](#)
Influenza deaths by geography, time, age, and gender – [CDC](#)
Population data by geography – [US Census Bureau](#)

TOOLS —  

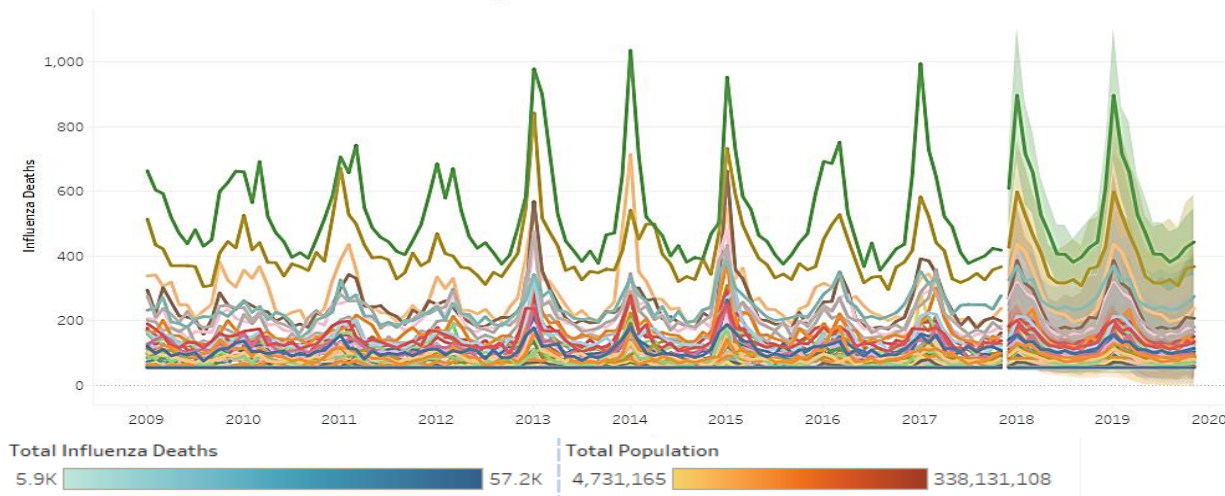
SKILLS — Translating business requirements, Data cleaning, Data integration, Data transformation, Statistical hypothesis testing Visual analysis, Forecasting, Storytelling in Tableau, Presenting results to an audience

INFLUENZA SEASON ANALYSIS

Total Influenza Death & Population Data (2009-2017)



Region wise Influenza Death



Influenza Deaths were higher among the elderly population (age 65+ years)

Smaller states with less population like Wyoming, Alaska had the maximum Elderly Population to Death ratio (more than 2%)

It could be seen from the line graph that maximum deaths peaked during the winter season i.e., Nov-March

With the help of line graph, we had also prepared a future forecast (up till Dec-2020) of the elderly deaths

RECOMMENDATIONS

Age

Allocate more staff to states with higher number of elderly population

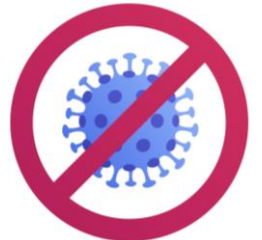
Season

Staff allocation during the winter months i.e., Nov-March

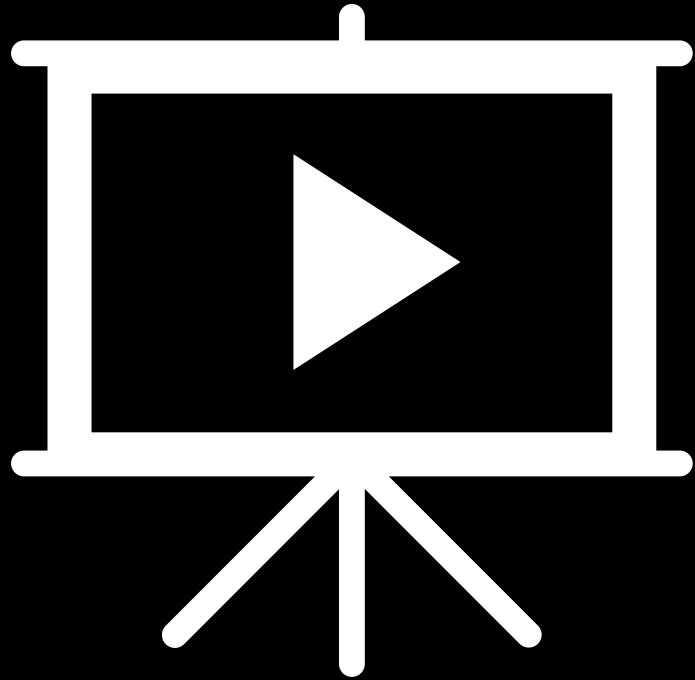
Health Infrastructure

Improve the number of staffs in states with comparatively poor health infrastructure

Complete Report link: 



ROCKBUSTER LLC



Rockbuster Stealth LLC is a movie rental company that used to have stores around the world. Facing stiff competition from streaming services such as Netflix and Amazon Prime, the Rockbuster Stealth management team is planning to use it's existing movie licenses to launch an online video rental service in order to stay competitive.

PROJECT OVERVIEW

OBJECTIVE — Launch an online video rental service in order to stay competitive against online streaming services

DATASET — Rockbuster LLC

TOOLS —



SKILLS —

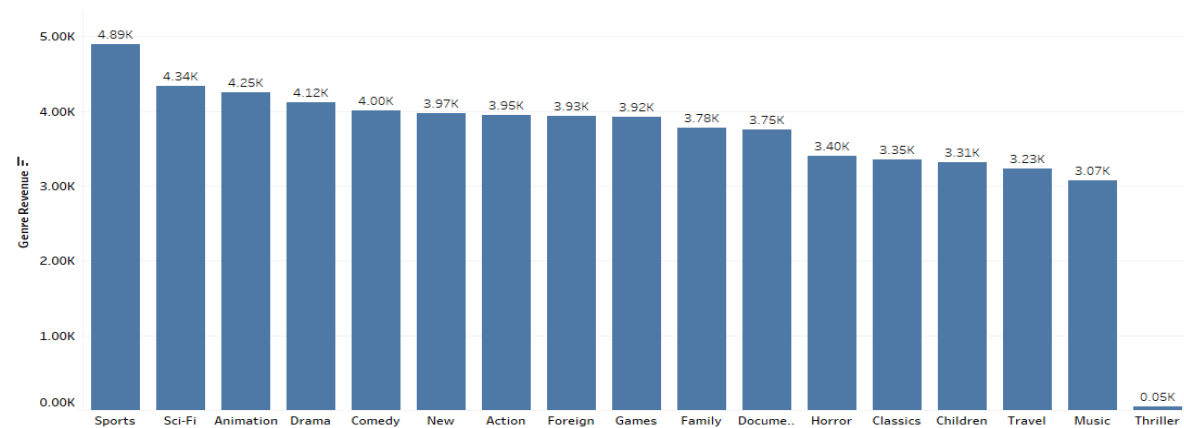
Relational databases, SQL, Database querying, Filtering, Cleaning and summarizing, Joining tables, Subqueries, Common table expressions (CTE)

ROCKBUSTER LLC ANALYSIS

Genre Rentals

Thriller Duration: 6.00 Rate: 2.99	Foreign Duration: 5.12 Rate: 2.95	Classics Duration: 4.94 Rate: 2.64	Action Duration: 4.93 Rate: 2.67	Horror Duration: 4.92 Rate: 3.10
Travel Duration: 5.47 Rate: 3.32	Drama Duration: 5.06 Rate: 3.18	Comedy Duration: 4.80 Rate: 3.28		Sports Duration: 4.74 Rate: 3.07
Music Duration: 5.25 Rate: 3.06	Games Duration: 5.01 Rate: 3.13	Animation Duration: 4.80 Rate: 2.76		
Family Duration: 5.15 Rate: 2.70	Children Duration: 5.00 Rate: 2.69	Sci-Fi Duration: 4.78 Rate: 2.99		

Genre Revenue



Travel & Music Genre have the highest average rental duration and rates

Sports & Foreign Genre has the highest number of films among the data studied

SPORTS & SCI-FI Genre generated the maximum revenue

Travel, Music & Thriller Genre generated the least revenue

Asian region shares the highest revenue among the countries

RECOMMENDATIONS

Ratings

Increase the number of movies with Pg-13 ratings

Genre

Purchase and Promote more number of movies with Sports, Sci-Fi Genres to generate more revenue

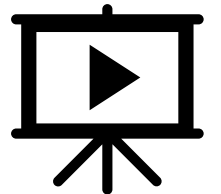
Promotion & Incentive Program

Conduct promotional and incentive programs for countries such as India, China for customer addition and retainership

Complete Report link:



Data Dictionary link:






INSTACART



Instacart, an online grocery store operates through an app. The Instacart stakeholders want to uncover more information about their sales patterns. They are most interested in the variety of customers in their database along with their purchasing behaviors in order to optimize their targeting marketing strategy.

PROJECT OVERVIEW

OBJECTIVE	Perform an initial data and exploratory analysis of some of Instacart's data in order to derive insights and suggest strategies for better segmentation based on the provided criteria.
DATASET	The Instacart Online Grocery Shopping Dataset 2017
TOOLS	  
SKILLS	Python, Data wrangling, Data merging, Deriving variables, Grouping data, Aggregating data, Reporting in Excel, Population flows

INSTACART ANALYSIS

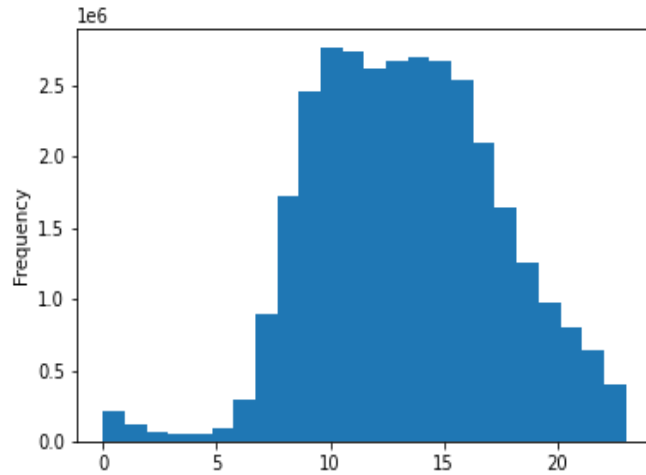


Fig 1

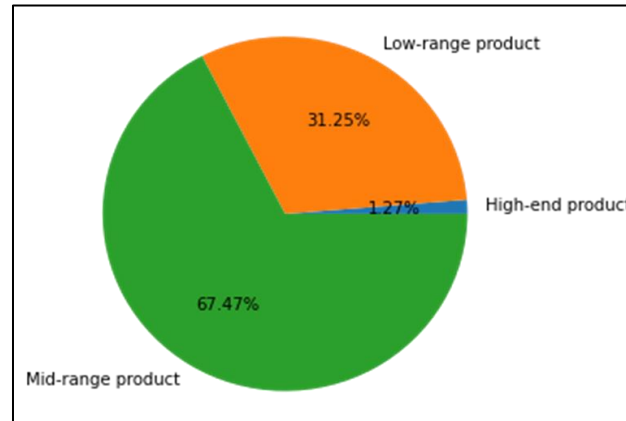


Fig 2

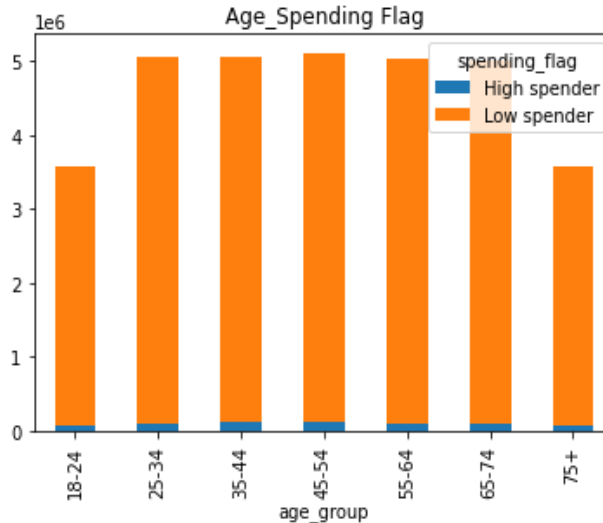


Fig 3

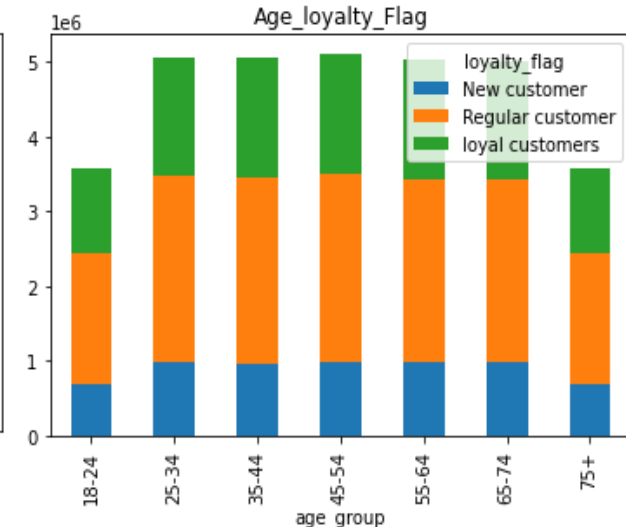


Fig 4

Maximum order execution takes place around 10 am and remains high till the evening time (fig 1)

Mid-range products (68%) are most ordered by the customers (fig 2)

Most of the customers are in the low-spender range (98%)(fig 3)

Maximum customers are in the 25-65 years of age group and are regular customers (fig 4)

Most of the customers order during the weekends

RECOMMENDATIONS

Marketing Time

Increase marketing during the middle of the week and Target customers during day time

Product Range

Maintain focus and bring more products of low-mid range

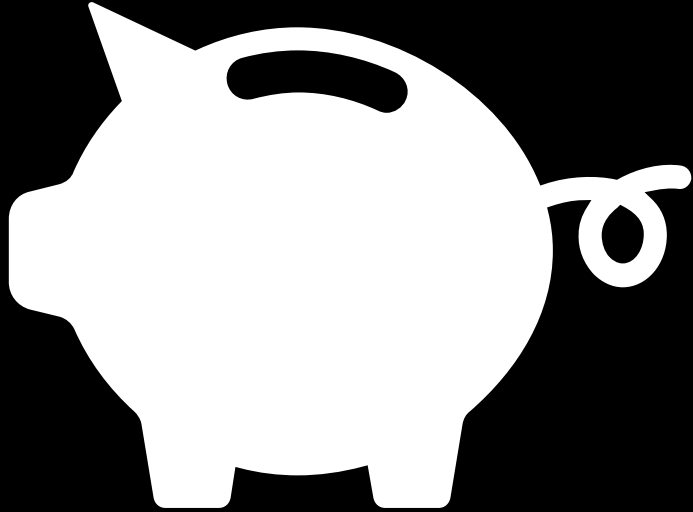
Competition Mapping

Competition mapping should be done especially in regions with low sales

Complete Report link: 



Pig E. Bank



A well known Global Bank requires an analyst to provide analytical support to its anti-money-laundering compliance department.

Work with the bank in order to build and optimize models that assist the bank in running their compliance program more efficiently

PROJECT OVERVIEW

OBJECTIVE — Provide analytical support and assess client risk, provide data insights against money laundering, using big data concepts

DATASET — [Pig E. Bank client data](#)

TOOLS —

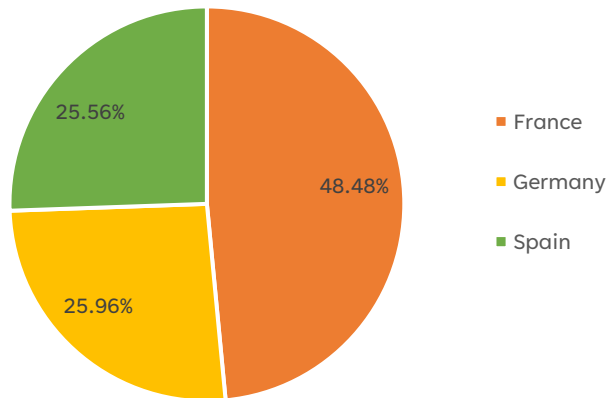


SKILLS —

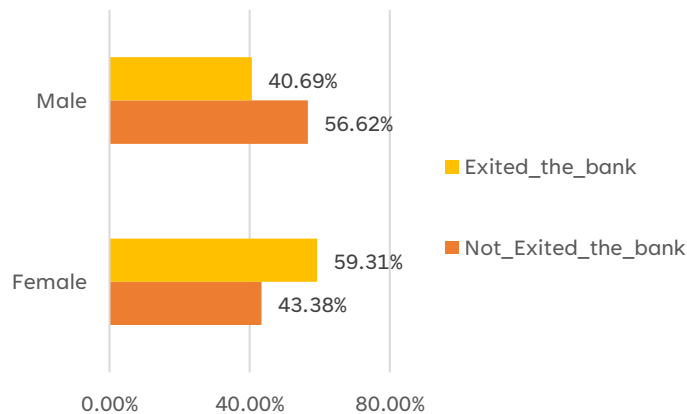
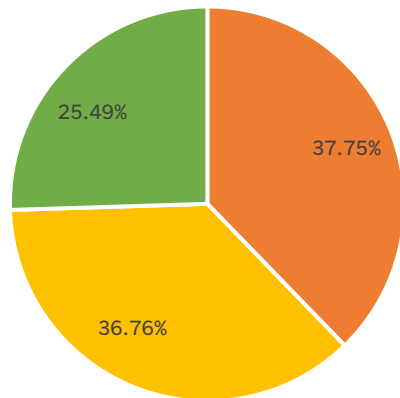
Big data, Data ethics, Data mining,
Predictive analysis, Time series analysis
and forecasting

PIG E. BANK ANALYSIS

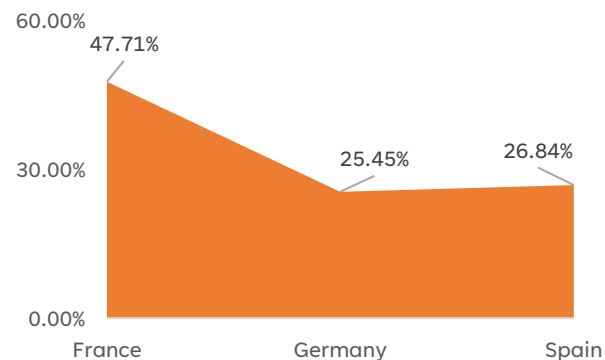
Total Customers (%)



Total Customers Exit (%)



Active Members



France (48%) has the **maximum** number of customers followed by Germany(26%) & Spain (25%)

France (47%) has the **maximum active customers** followed by Spain (27%) & Germany (25%)

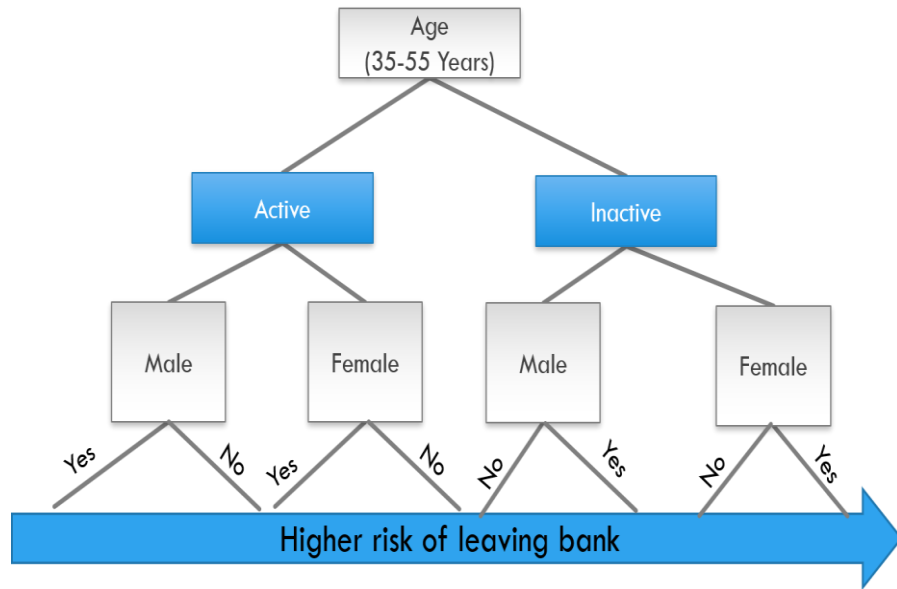
Germany has the maximum customers **to exit ratio of 37%**

Female customers have the exited the bank the most

People from the age group 36-45 (38%) tend to **leave the bank more** as compared to other age groups

RISK ASSESSMENT

Decision Tree in order to assess the reason for customers exit the bank



Complete Report link: 

CONCLUSION

Age

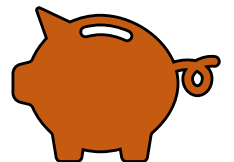
Middle aged group witnessed the maximum exits from the bank

Gender

Female exited the most

Country

Despite of having a 26% of customer share, Germany had the maximum customer to exit ratio



World Happiness Index



The World Happiness Report is published by the United Nations Sustainable Development Solutions Network. The report is published annually and contains articles and rankings of National Happiness. This data is collected from over 150 countries

The study is based on the different Parameters that correlates with various quality of life factors and are categorized into 12 different regions.

PROJECT OVERVIEW

OBJECTIVE ——— To build an interactive dashboard that will visually showcase well-curated results of an advanced exploratory analysis conducted in Python

DATASET ——— [World Happiness Index](#)

TOOLS ———   

SKILLS ——— Sourcing Open Data, Geographical Visualizations
Filtering, Cleaning and Summarizing Data,
Supervised and Unsupervised Machine Learning:
Clustering, Analyzing Time Series Data, Creating
Data Dashboards

WORLD HAPPINESS INDEX ANALYSIS

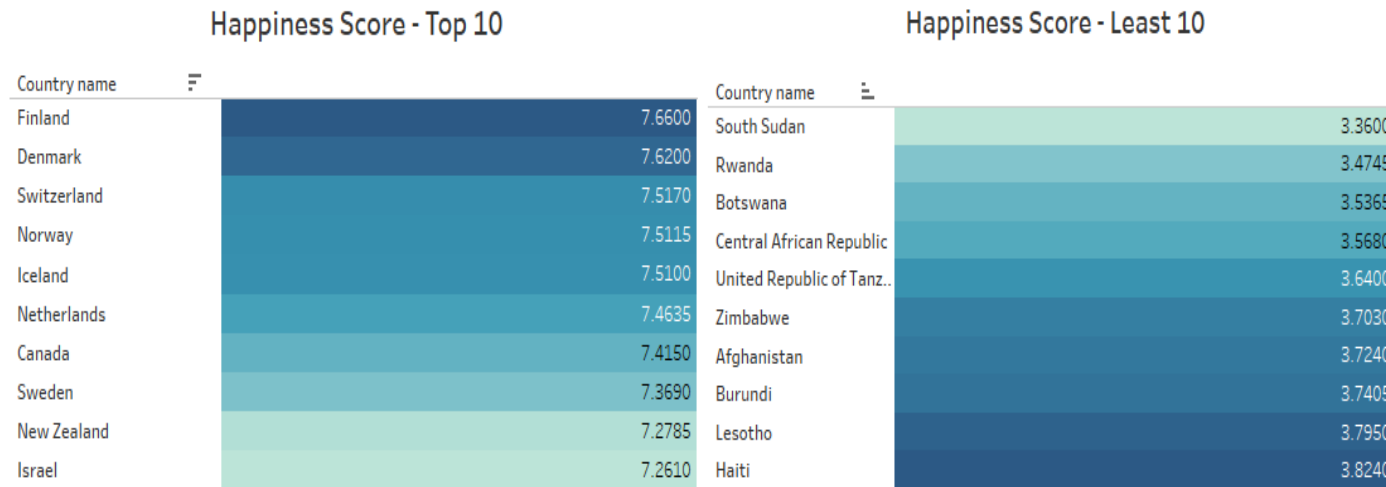


Fig 1

Generosity
 Perceptions of corruption
 Log GDP per capita
 Freedom to make life choices
 Social support
 Healthy life expectancy at birth

Fig 2

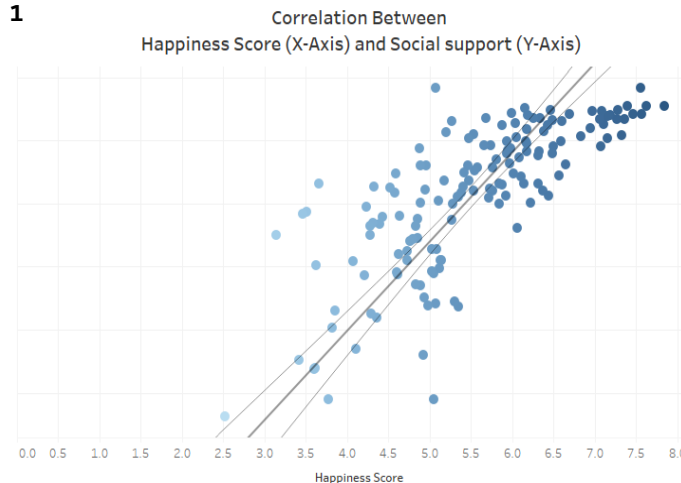


Fig 3

Finland has been the Most Happy Country with a Happiness Score of 7.66 whereas South Sudan has been the least Happy country with a Happiness Score of 3.36(fig 1)

This Happiness Score was calculated on various parameters as shown (fig 2)

With the help of Linear regression we figured out that Happiness Score had a strong correlation with Social Support(fig 3)

This correlation was further seen with different parameters like GDP per capita, Freedom to make life choices

KEY TAKEAWAYS

FINDINGS

Social Security, GDP per Capita, Freedom to make life choices are important factors that play an important role in a country's Happiness Score.

Data Limitations

Under developed and Developing nations have data limitation constraints as there is no proper method to calculate and follow these defined parameters

Going Forward

Spread awareness of the importance of well being among the developing and under-developed nations

Tableau Report link:



Python Analysis link:





FOR FURTHER QUERIES, CONTACT:



[Keshav Dewan](#)



Geneva, Switzerland



keshavdewan1990@gmail.com



[GitHub Profile](#)



[Tableau Profile](#)



[Resume](#)

THANK YOU