VICTOR MBOGU

Data Analytics Portfolio

About Me

I'm a skilled Data Analyst with academic background in Statistics and an Azure cloud computing background. I have a passion to work with data. My experience can help me analyze and provide insight that aligns with your company values and overall mission. My skills include knowledge of SQL, Excel, Python, Tableau and PowerPoint. These tools are being used for my cleaning, analysis and data visualisation. All projects are featured in this portfolio.

Featured Projects

01 **GameCo, Inc.**
Examining worldwide video game sales

02 **Influenza Season**
Planning for Flu season in the United States

03 **Rockbuster Stealth**
Marketing strategy for an online grocery store

04 **Instacart**
Creating customer segmentation based on transactional data

05 **Pig E. Bank**
Predictive analysis on customer retention, using decision tree.

06 **Fifa20 Gaming**
Finding connections between variables

01

GameCo, Inc. Project

GameCo, Inc. Overview

OBJECTIVE
Use historical data to perform a descriptive analysis for a game company

ASSUMPTION
Game sales for the various geographic regions have stayed consistent over time.
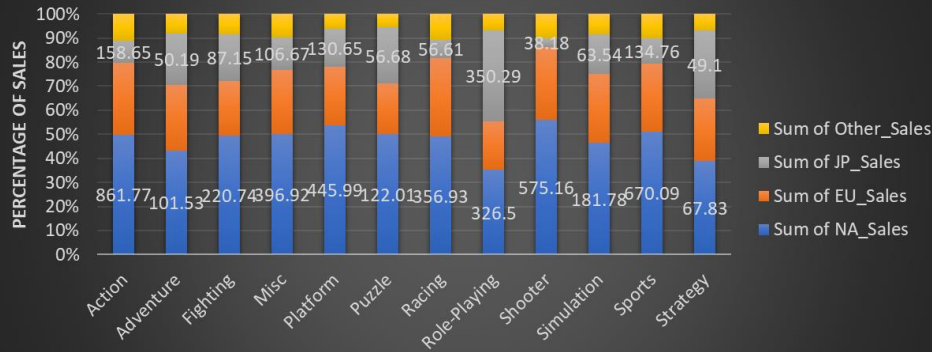
TOOLS
Excel
PowerPoint

SKILLS
Grouping and Summarising Data, Descriptive Analysis, Visualising & Presenting Results.

GameCo.Inc, Analysis

-- The stacked chart provides information about the various sales by Genre and region.
-- Action games were more sold in the whole region.
-- North America has the highest sales for action games while Japan has the lowest sales across the 3 regions.
-- We can also note that European sales are getting closer to the North American sales.
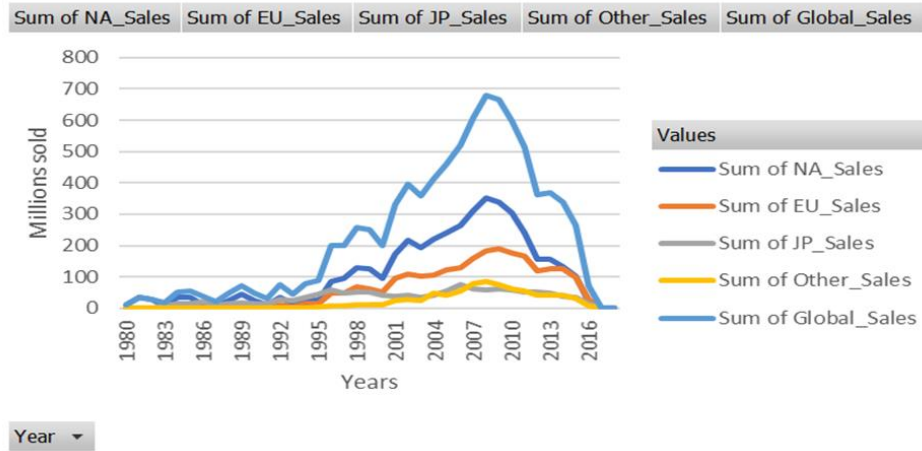
-- The line chart shows the yearly sales of games across all regions. It shows that, there have been an increase in sales from 2008 to 2009 and which includes North American having the highest distribution of sales across all regions.
-- Although there was a sudden decrease from 2008 from all regions and also Japan has the lowest sales of all regions

Conclusion & Recommendation

-- Across all regions games sales haven't remained the same over the years, there have been upward and downward trends in the sales of games in the past few years. While north American and Japan sales have been fluctuating, Europe maintains a upward trend over the years

-- Since Europe has been consistent over the years, GameCo can take advantage of it's trend and increase their percentage in sales, this also makes maximum return in replace of other regions with lower sales

-- Note, creating awareness using different marketing strategies in the regions with lower sales will improve the sales across regions.

-- To expand the financial plan for promoting, designate the percentage rate to reflect the sales - 38% or more to EU, 32% or less to North America and 19% or more to Japan.

02

Influenza Season Project

Influenza Season Overview

OBJECTIVE
Use historical data to help plan for staffing demand in the country

ASSUMPTION
Vulnerable populations suffer the most-severe impacts from the flu and are the most likely to end up in the hospital
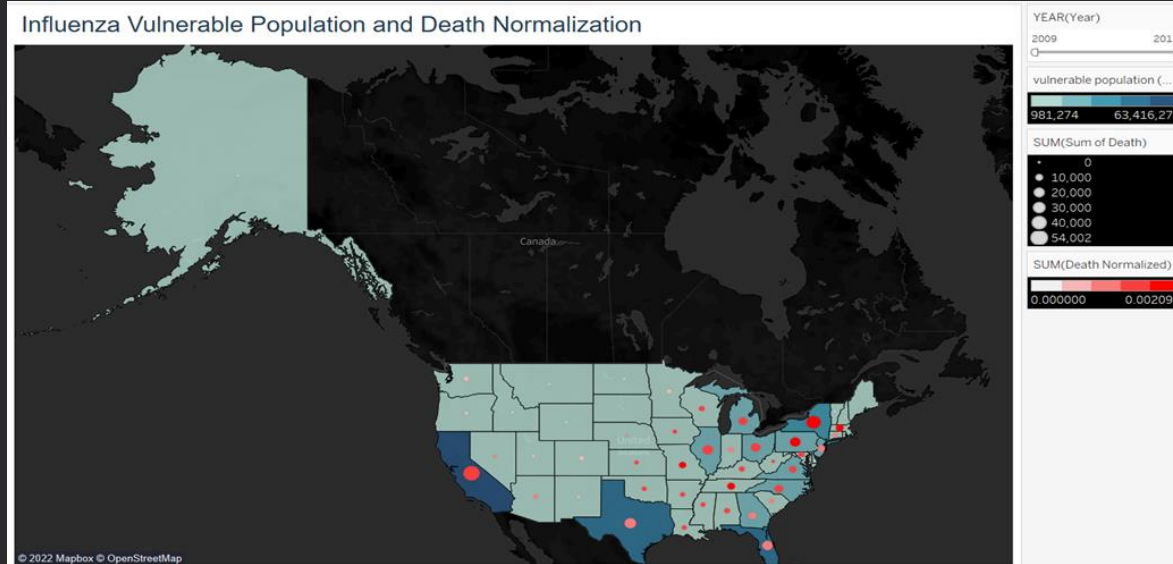
TOOLS
Excel
Tableau

SKILLS
Data profiling, Data cleaning, Integration, & Transformation.
Statistical Hypothesis Testing, Visual Analysis, Forecasting, & Storytelling

Data Set: Influenza Deaths,   Population Census Data   Interim-Report: PDF

Influenza Season Analysis

-- The combination map shows the vulnerable population (under 5 and over 65 years) and the influenza death normalization, which represents the percentage of death over population.

-- Looking at the map we can see that the number of deaths increased, and various state with high mortality rate. Note hovering around the map shows their total number of death

The Relationship between Sum of Death and Sum of Total Population
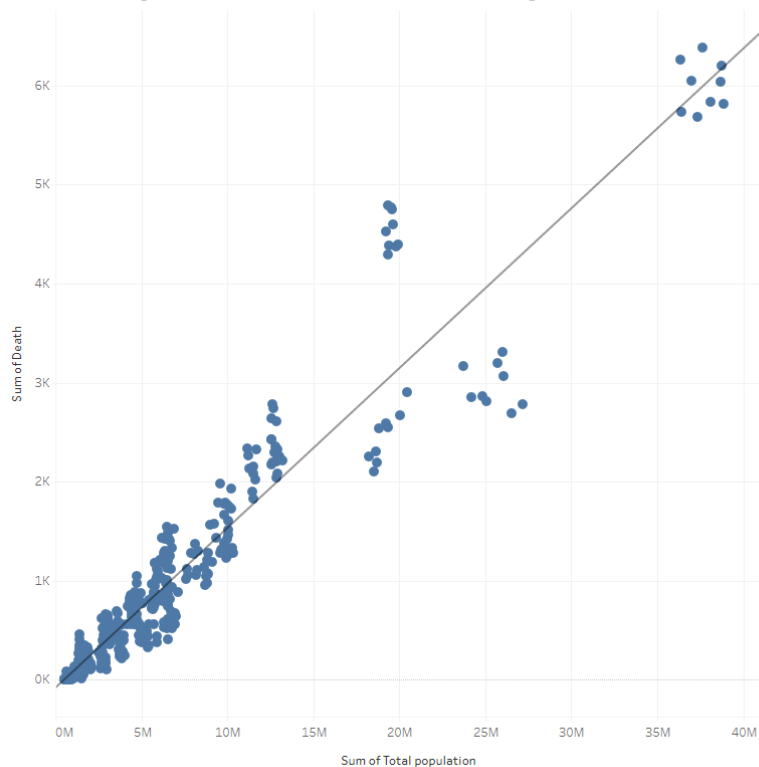
-- Scatterplot showing the relationship between sum of death and sum of population with a trend line, showing a strong corelation.

-- The data values are tightly clustered around the trend line and a few extreme values. This makes sense considering that we have a positive correlation

Conclusion and Recommendation

In preparation of the influenza season, more additional staffs should be allocated to high risks areas such as:

*California
*New York
*Texas
* Pennsylvania
*Florida
*Illinois

Elderly people above 50 years of age should be vaccinated to prevent high rate of mortality

Click in the icon below for final report

03

Rockbuster LLC Project

Rockbuster Stealth LLC Overview

OBJECTIVE

To use relational database to analyze
and provide data-driven answer to more complex business questions

ASSUMPTION

Which countries has the most
existing customers?
What was the average rental
duration?

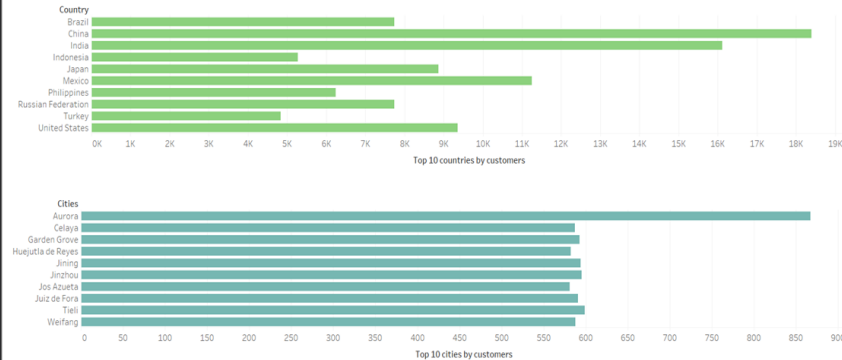## TOOLS

Excel
PostgreSQL
Tableau
DbVisualizer

SKILLS

Relational Database, Data Querying,
Filtering, Cleaning, Summarising Data,
Table Joins & Subqueries,
Common Table Expressions
.

Data Set: Rockbuster

Rockbuster Stealth LLC Analysis

Top countries and cities with high rate of customers

Top 10 countries by customers

Top 10 cities by customers

Top 5 customers

Coun..	City	First Na..	Last Name	
Turkey	Sivas	Gabriel	Harder	$108.75
Mexico	Atlixco	Sara	Perry	$128.70
	Celaya	Sergio	Stanfield	$102.76
Japan	Kurashiki	Rebecca	Scott	$89.76
Brazil	So Leopoldo	Francisco	Skidmore	$93.79

-- India, China, United States, Japan, and Mexico have high rates of revenue and claim top 5 customer count. Other countries had lower counts

-- While we can see the top 5 paying customer are majorly in Mexico, Turkey, Japan and Brazil

Rockbuster revenue by genre

Sum Of Revenue
$47.89 — 5K

Genre	Revenue
Sports	$4,892.19
Comedy	$4,002.48
Games	$3,922.18
Family	$3,782.26
Documentary	$3,749.65
Sci-Fi	$4,336.01
New	$3,966.38
Horror	$3,401.27
Travel	$3,227.36
Animation	$4,245.31
Action	$3,951.84
Classics	$3,353.38
Music	$3,071.52
Drama	$4,118.46
Foreign	$3,934.47
Children	$3,309.39

-- Sports, Comedy, Sci-fi, Animation, and Drama, are clearly the top genres in Rockbuster company.
These genres generate more revenues for Rockbuster

Top 10 movies by revenue

Title	Genre	Rating	Revenue
Dogma Family	Animation	G	$168.72
Harry Idaho	Drama	PG-13	$177.73
Hustler Party	Comedy	NC-17	$190.78
Innocent Usual	Foreign	PG-13	$191.74
Saturday Lambs	Sports	G	$190.74
Telegraph Voyage	Music	PG	$215.75
Titans Jerk	Sci-Fi	PG	$186.73
Torque Bound	Drama	G	$169.76
Wife Turn	Documentary	NC-17	$198.73
Zorro Ark	Comedy	NC-17	$199.72

Bottom 10 movies by revenue

Title	Genre	Rating	Revenue
Cruelty Unforgiven	Classics	G	$6.94
Duffel Apocalypse	Documentary	G	$5.94
Freedom Cleopatra	Comedy	PG-13	$5.95
Lights Deer	Classics	R	$7.93
Oklahoma Jumanji	New	PG	$5.94
Rebel Airport	Music	G	$6.93
Stallion Sundance	Sci-Fi	PG-13	$7.94
Texas Watch	Horror	NC-17	$5.94
Treatment Jekyll	Drama	PG	$6.94
Young Language	Documentary	G	$6.93

-- We can clearly see the top 10 and bottom 10 movies, this will enable Rockbuster to know the kind of movies most customers demand and generate revenue.

Conclusion and Recommendation

-- Advertisement and awareness should be carried out in the lower regions where there are lower customers such as Faroe islands, Togo and Madagascar.
-- Promotion should be given to the top 5 customers; this can enable them to tell people about offered discounts.

-- The minimum rental duration should increase, so more customers could easily have access to the film, especially customers from the lower regions.
-- Since Rock buster's movies are based online, monitoring of customers will help account for what services should be recommended towards the betterment of sales and increase in customers in the lower regions.

Data Dictionary:

Final Presentation:

04

Instacart Project

Instacart Overview

OBJECTIVE

To ensure Instacart targets the right customer profiles with the appropriate products.

ASSUMPTION

What are the busiest days of the week and hours of the day?
Are there particular times of the day when people spend the most money?

TOOLS

Python
Excel

SKILLS

Data Wrangling, Merging, Deriving Variables, Grouping & Aggregating Data, Visualising with Python, Population flows
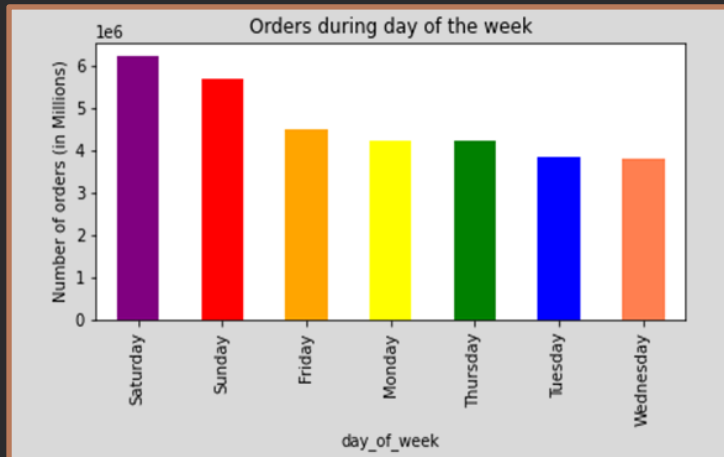
Data Set: Orders and Products   Departments   Customers   Final Report:

-- The Pie chart shows the Clearly, mid-range-cost items dominate the Instacart catalogue. The low-range products account for only 31.25% of the whole product portfolio.

Low-range product
31.25%
1.29%
High-range
67.46%
Mid-range product

Demographic breakdown of customers

Customer profile	Young parent	Single adult
Age	less or equals30	lesser or equals 30
Income	$400000 or less	$400000 or less
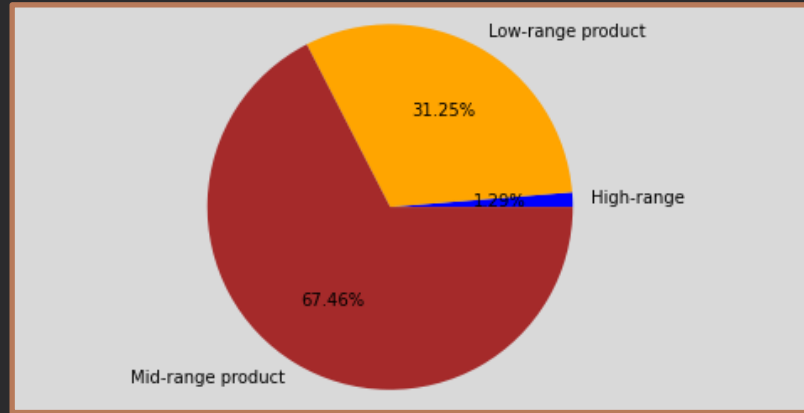
-- The table shows the classification of customer profile between Age <30, Top 5 department & Income.
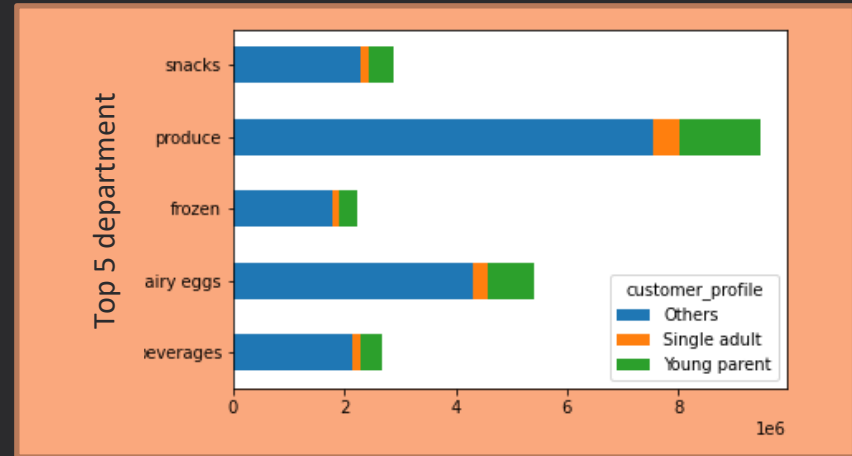
-- Between these groups, customers order more from the produce department and it comprises of young parents, single adult use less from the produce department.

Conclusion and Recommendation

-- Instacart could find a way to generate more customers for the days with low customers by creating awareness, more about the products. Also the Instacart needs to look at the time frame to know low demand of orders and find a way to boost more customers with the low order.

-- Customers spend more money during afternoon and low sales are in the evening, these periods of day give Instacart a clue on how to improve or increase other sales during the hours from 5pm to 11 pm

-- From the departments below we can clearly see that the produce department has the highest frequency in sales and generates more income among all departments. This gives Instacart clue on how they can improve other departments regarding to sales and price.

-- The parents uses more of Instacart product, while the non parent are the least customers to use the Instacart product. 66.29% frequently use Instacart products

05

PIG E. Bank Project

PIG E. Bank Overview

OBJECTIVE
Finding the leading indicators why a customer may quit the bank, in order to boost customer retention.

ASSUMPTION
Who are the customers leaving the bank and why?
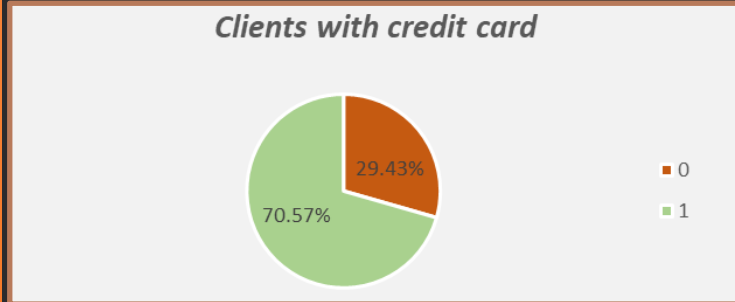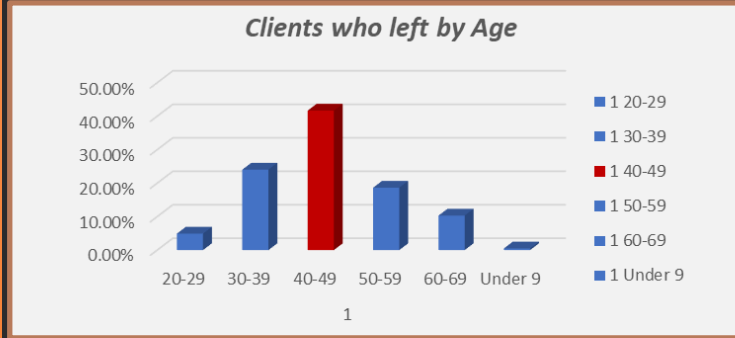
TOOLS
Microsoft Excel

SKILLS
Big Data, Data Ethics
Data Mining and Cleaning
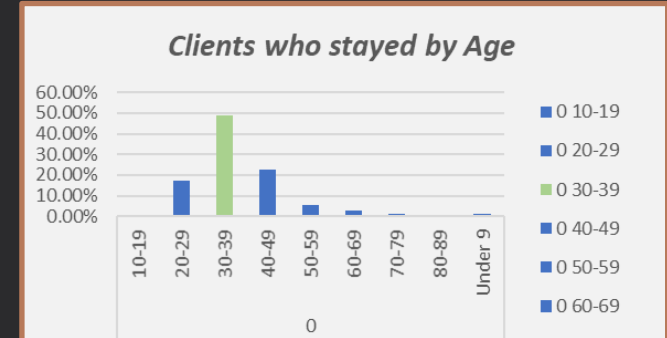Predictive Analysis
Time-Series Analysis and Forecasting

.

Instacart Analysis

-- Clients aged 40-50 are more likely to leave.

-- It's obvious clients within the age bracket 30 to 39 have a higher tendency to stay, followed by the age bracket 40 to 49.
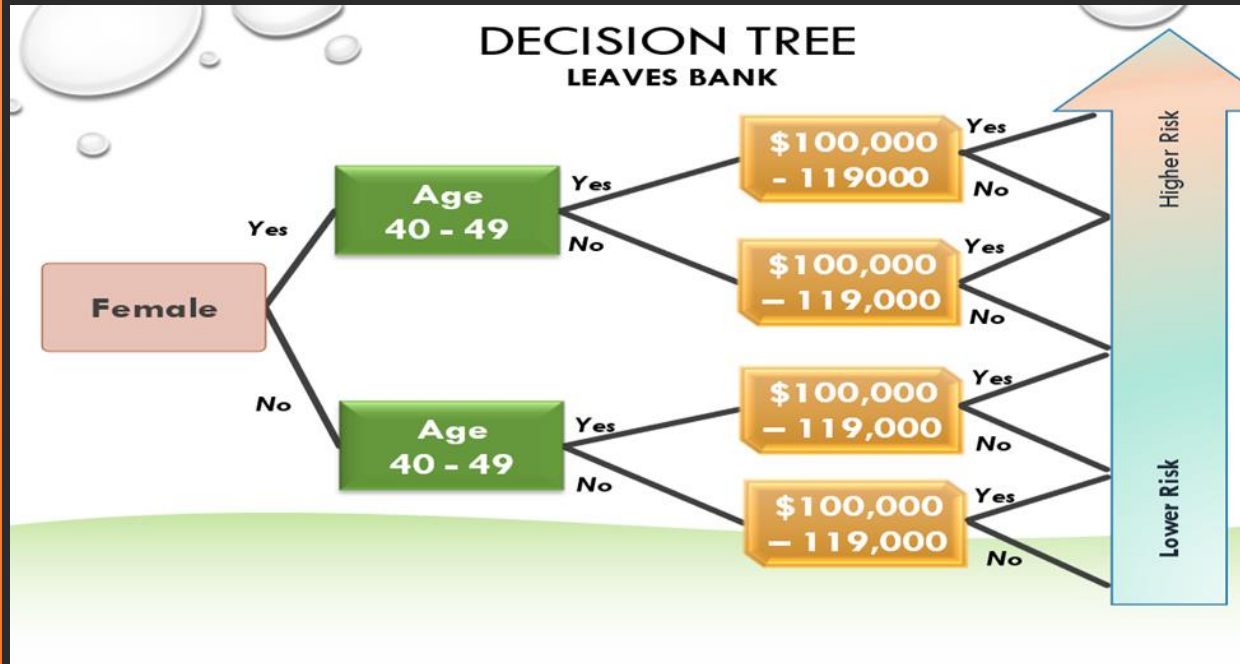

Clients who left by Age

-- It's obvious that more client with credit card left and this shows their percentage rate. 29.43% with credit card stayed.

Clients who stayed by Age

-- Clients who came from France has 51% of inactive members that tends to stay while 37% of inactive members from Germany tends to leave the bank

Clients with credit card

Inactive member with country who left and stayed
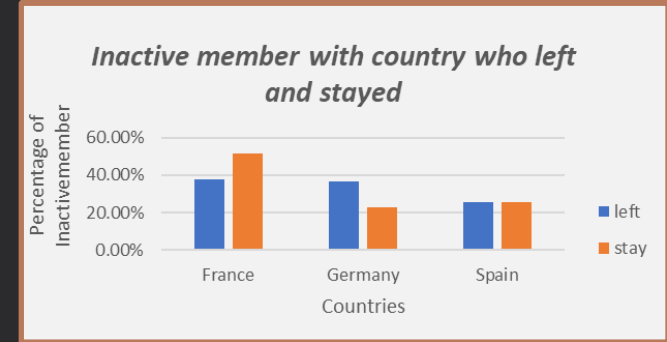
-- These higher-risk characteristics are broken down further by gender and then by the higher-risk income levels. These reveal a marginally greater leave rate, but it is still significant.

DECISION TREE
LEAVES BANK

Female

Yes — Age 40 - 49
No — Age 40 - 49

$100,000 – 119000 — Yes / No
$100,000 – 119,000 — Yes / No
$100,000 – 119,000 — Yes / No
$100,000 – 119,000 — Yes / No

Higher Risk

Lower Risk

-- Exiting customers are clearly more likely to be female, which corresponds to the higher male ratio among those who stayed

-- Germany has the biggest percentage of its overall population that has left, followed by France and Spain, respectively.

06

Fifa2020 Project

Fifa 2020 Analysis

-- The satterplot shows different variable using the pair plot
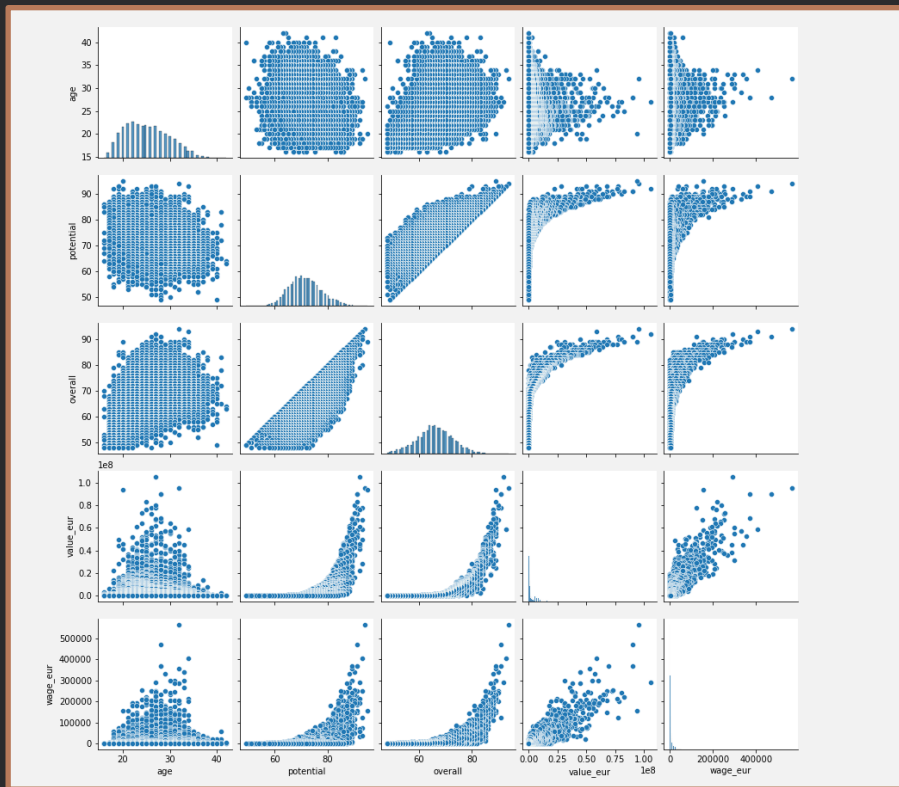
-- Regression Modell -

-- Using the regression model to run multiple columns to see how each feauture contributes or impacts the wages of the players.
-- While the results show that the slope of each variable and their connections -
Overall = 1645.41308393
Age = -319.40627294
Potential = 403.18212547
-- The above results mean the result of the coefficients of Overall has the most effect with a positive correlation and a higher wage than the Potentials' coefficient of players. The Age's coefficient is negative with no effect on wage.
All analysis was performed in Python.

-- Clustering between the Age and Potentials of players
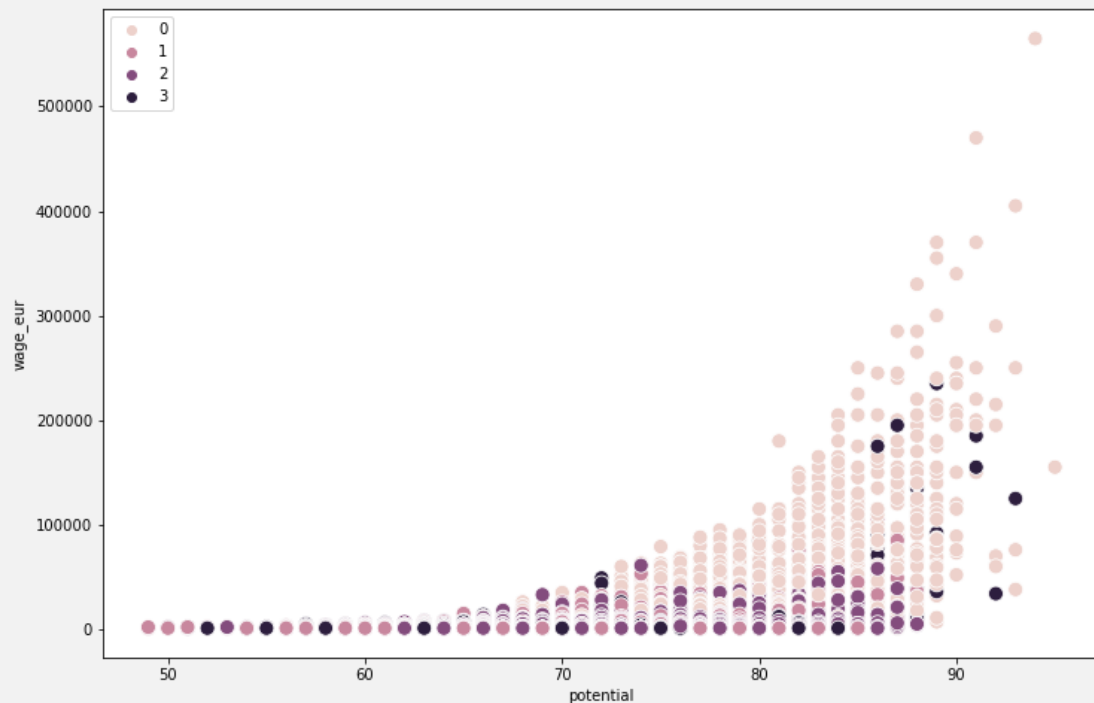
-- There are four clusters in this model, which show that there are different variables that make these players have similarities or have features in common. Number of elements in different clusters are more or less equal.

07

Feature Project Background

ROCKBUSTER STEALTH LLC Overview

OBJECTIVE

Upcoming project

ASSUMPTION

Upcoming project

TOOLS

Upcoming project

SKILLS

Upcoming project
.

Data Set:

THANKS!

Do you have any questions?
victormbogu1@gmail.com
+49 1590 6824479
GitHub Repository Link -

—VICTOR MBOGU