Stefan Rieß

Data Analytics Portfolio

Github
Tableau
Linkedin

About Me

I have a background in technical Customer support, where I have been responsible for identifying issues and bugs within a data set and Programs.

Over the last month, I have acquired additional skills such as Excel, PowerPoint, Tableau, SQL, and Python as data analysis tools displayed in this Portfolio.

Featured Projects

GameCo Inc.

Descriptive Analysis of game Sales across the Globe

Rockbuster Stealth

Geographical Analysis of Movie Rental to assist with Competetive Launch of Streaming Services.

Pig E. Bank

Predictive Analysis on Customer retention and possible money Laundering. Creation of Decision tree

US Influenza Season

Exploratory Analysis, profiling, and forecasting of the upcoming Influenza Season in the United States

Instacart

Exploratory Analysis, profiling Customer data for a vast Grocery selling Service
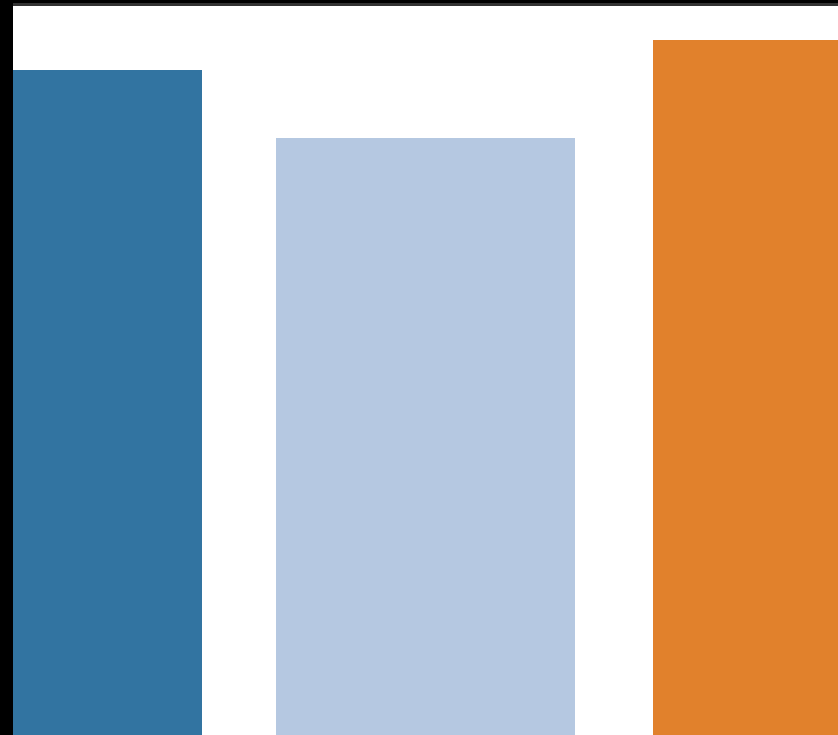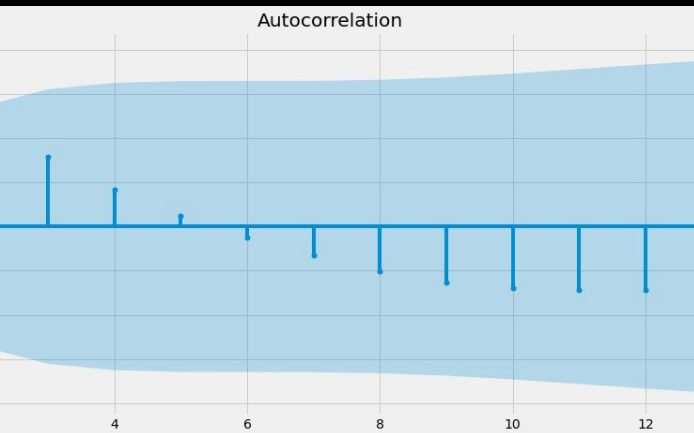
Evolution of Covid 19

Correlation Matrix and Heatmaps Choropleth Maps-Supervised Machine Learning : Regression Unsupervised Machine Learning: Clustering-Analyzing Time Series Data
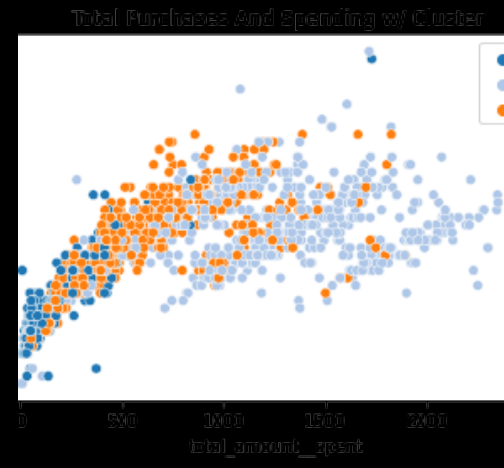
Coding Training

• During my Data Analytics complete Course, I also used different Datasets to create other Visualisations and to gain additional techniques

Game Co

Currently, the Executive assumes that sales over the different Geographic Regions have stayed the same over time.
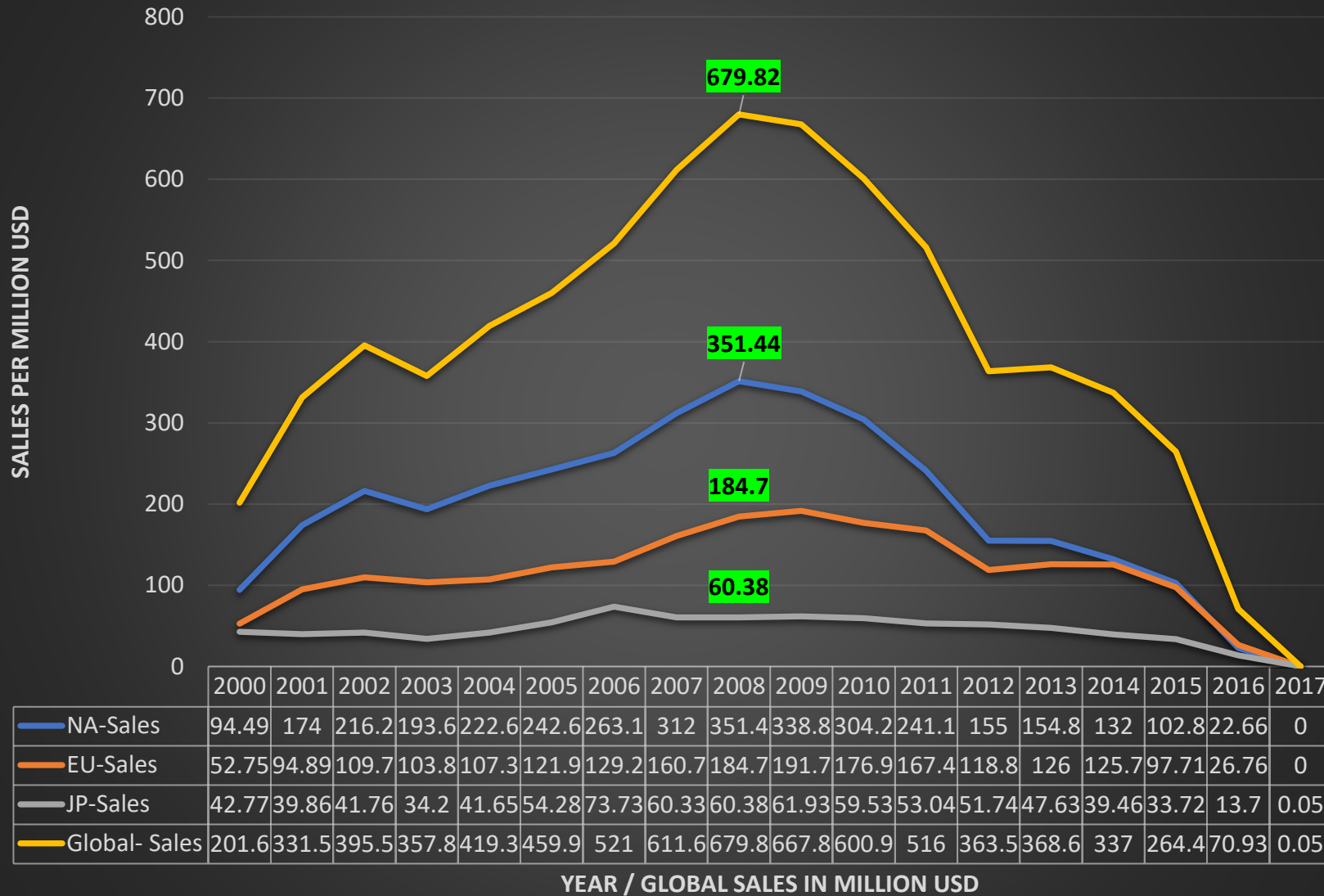
Data:

- **VG Sales**

Tools:

- Excel
- Word
- Powerpoint

Skills obtained:

- Grouping and Summarising Data
- Descriptive Analysis
- Visualisation
- Presentation

Trendline of Sales in the Gaming Industry

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
NA-Sales	94.49	174	216.2	193.6	222.6	242.6	263.1	312	351.4	338.8	304.2	241.1	155	154.8	132	102.8	22.66	0
EU-Sales	52.75	94.89	109.7	103.8	107.3	121.9	129.2	160.7	184.7	191.7	176.9	167.4	118.8	126	125.7	97.71	26.76	0
JP-Sales	42.77	39.86	41.76	34.2	41.65	54.28	73.73	60.33	60.38	61.93	59.53	53.04	51.74	47.63	39.46	33.72	13.7	0.05
Global- Sales	201.6	331.5	395.5	357.8	419.3	459.9	521	611.6	679.8	667.8	600.9	516	363.5	368.6	337	264.4	70.93	0.05

YEAR / GLOBAL SALES IN MILLION USD

It is visible that Sales, since they have been at an all-time high in 2008 and 2009, constantly have dropped from roughly 680 million Units sold to currently 70 million.

It shows that this should be addressed to avoid problems in the future as we don't know yet what will influence the gaming Market in the future. An adjustment in Marketing Budget should be considered at this point.

The proportion of Sales by Genre in $Millions
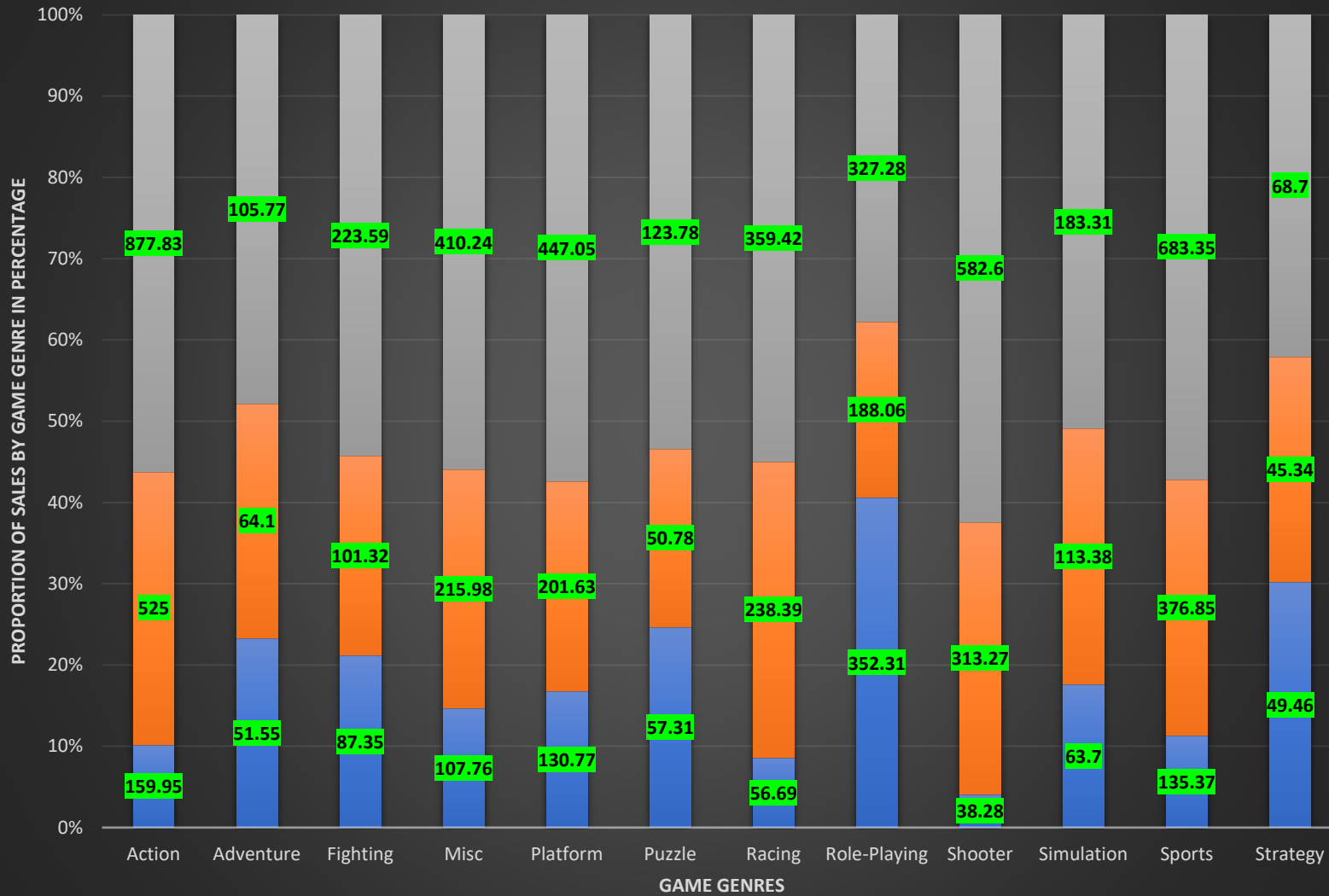
Here we have a view of the Sales of Computer games by genre for the Major selling Regions of North America, Europe, and Japan.

Based on the analysis, we have prepared a Document and Presentation with Recommendations for the Stakeholders.

- NA-Sales
- EU-Sales
- JP-Sales

Recommended Actions for the Games CO Project

Marketing should be focused on Shooter/Action, and Sports Games for North America and Europe as these tend to have a high potential for Sales and Revenue.

Marketing should be shifted towards the Asian Region for Adventure Games as these sold there with a very high Proportion.

Final
Presentation
Game Co

US Influenza Season

Objective:

Create a plan for medical staffing in the upcoming Influenza Season.

Data:

- Influenza Deaths
- Influenza Hospital Visits
- Laboratory Data Sets
- Flu Shot Survey
- Population Data

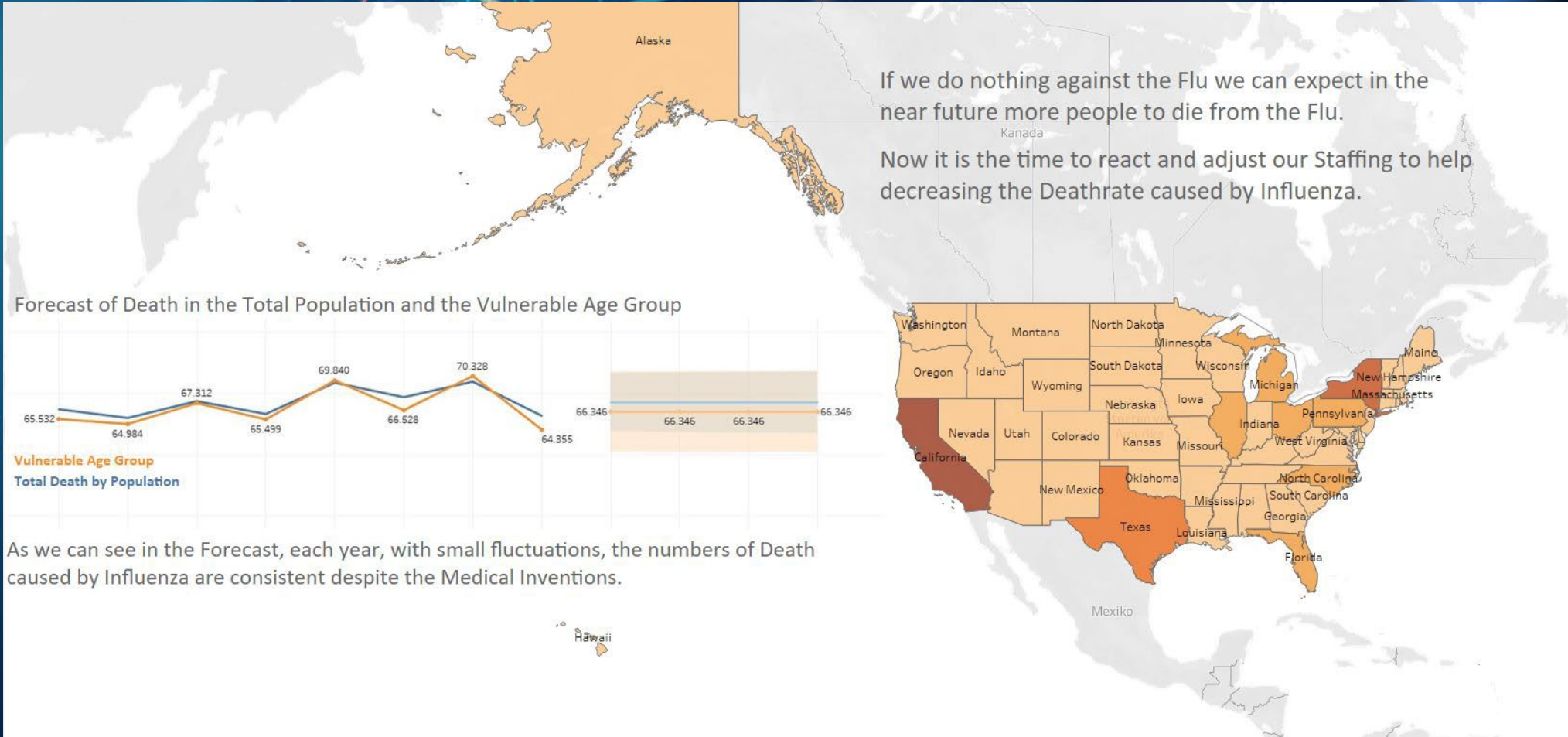**Tools:**

- Excel
- Tableau

Skills obtained:

- Profiling Data
- Cleaning, Transformation, and Integration of Data
- Statistical Hypothesis
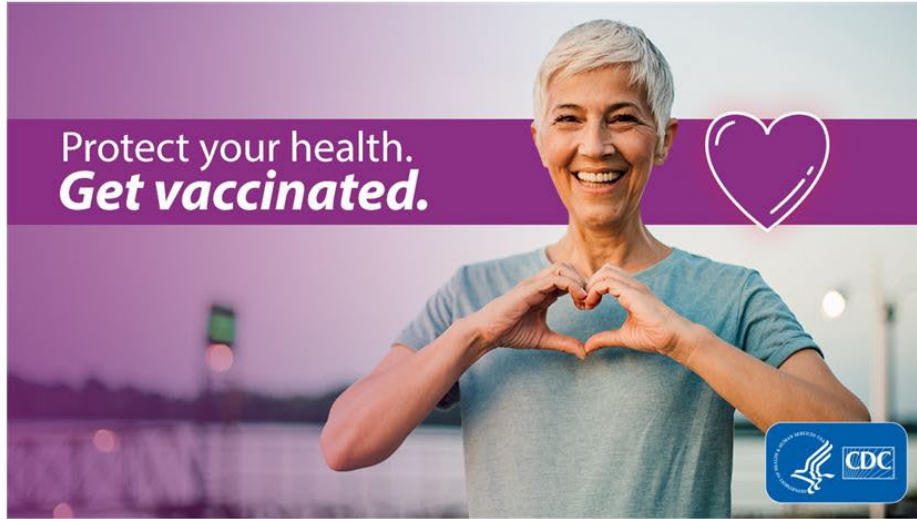- Visual Analysis
- Forecasting
- Storytelling with Data

US Influenza Season Analysis Forecast

Conclusion and Recommendation for the US Influenza Season Project

Preparation for the 2018 Influenza Season

Recommendation

In order to tackle the Deathrate and lower it down we should send additional Medical Personal to the Top Affected States we have identified in our Analysis.

This will help the Overall Deathrate to be lowered but expecially also the Vulnerable Age Group.

With this in effect , ongoing Surveys to adress Changes in Workload, we can expect lower Deathrate overall.

Increasing the Budget spent on Marketing for Vaccinations, this will help aswell for more people getting Vaccinated and lowering the Deathrate as a Result.

Protect your health.
Get vaccinated.

CDC

Conclusion:

We have identified the following States which are the most affected by the Flu and should receive more staffing in order to decrease the Death-Rate caused by the Flu.

California
New York
Texas
Pennsylvania
Florida

Rockbuster Stealth LLC

Objective:

Rockbuster Stealth LLC is a fictive Movie Rental Company that wants to go into Market Streaming. To find the best entrance, an Analysis of current sales/rentals was requested, with Key questions asked by the Stakeholders.

Data:

- Rockbuster Data Set
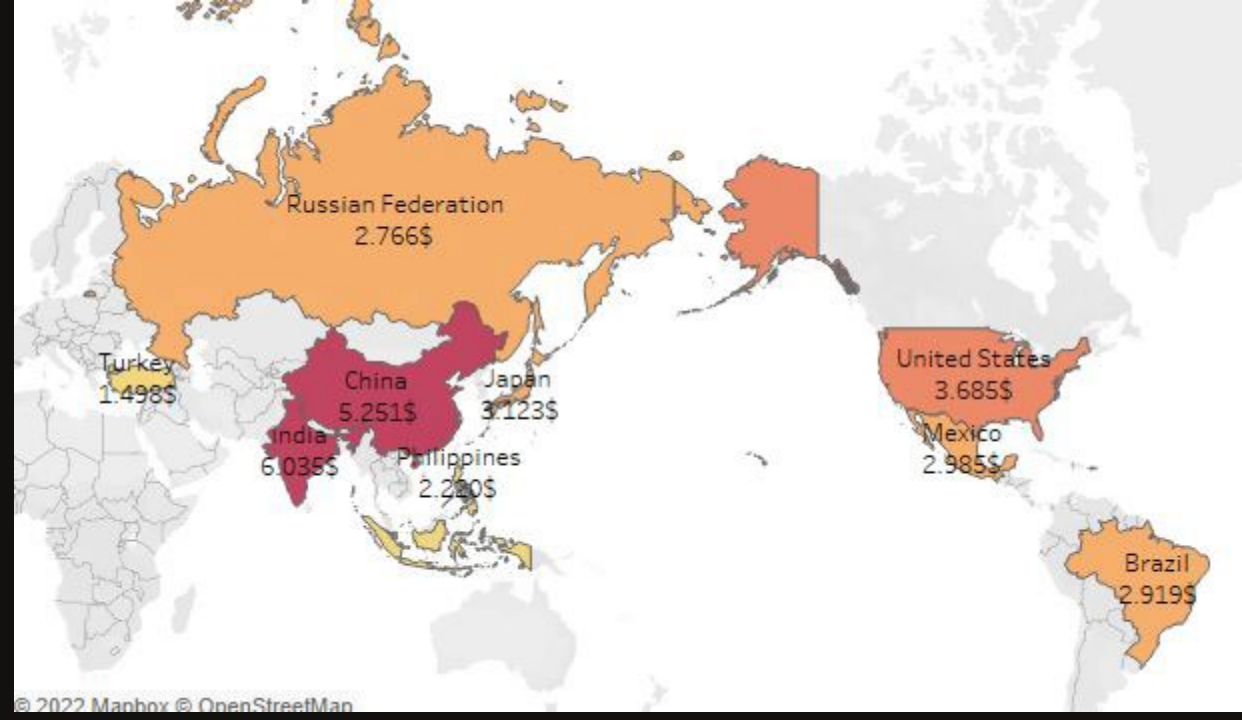- Data Dictionary

Skills obtained:

- Database Querying
- Relational Databases
- Cleaning, Filtering, and Summarising Data
- Common Table Expressions
- Table Joins and Subqueries in SQL

Tools:

- PostgreSQL
- Tableau

Which countries contributed the most revenue gain?

Here I have displayed the Top 10 countries with the most revenue gain for Rockbuster Stealth LLC.

It is visible that the most revenue comes from India and China, but also a cultural total different Customerbase such as Russia and the United States.

Sales Variation between different Regions for Rockbuster Stealth LLC

To gain a better Overview of the wholesome Situation, I created a map to display how the Sales Vary between Regions.

In that case, I could see that Rockbuster Stealth has an Overall Worldwide Customerbase with Spikes in Asian/Eastern Regions and North and South America.

Conclusion and Recommondation for the Rockbuster Stealth LLC Project

As a conclusion and recommendation in this Project, we suggested increasing Marketing in countries with a very high Customerbase, such as India, China, Japan, and Russia, as the current Movies Registered for Rockbuster aim for People in these [i.E. Asian Bollywood Movies]

We also recommended implementing a Loyalty System for recurring Rentals and Customers based on our findings.

The whole presentation can be looked up under Tableau via the Following Link.

Rockbuster Stealth LLC Data Analysis

Instacart Grocery Basket Analysis

Objective:

Instacart already has very good sales, but they want to uncover more
information about their sales patterns. Your task is to perform an initial data and exploratory
analysis of some of their data in order to derive insights and suggest strategies for better
segmentation based on the provided criteria.

Data

[Instacart Customers Data Set](#)

[Data Dictionary](#)

Citation (required in your final report): "The Instacart Online Grocery Shopping Dataset 2017", Accessed from https://www.instacart.com/datasets/grocery-shopping-2017 on <date>.

Tools:

- Jupyter Notebooks
- Python
- Excel

Skills obtained:

- Data Wrangling/Consistency Checks

- Combining and Exporting Data

- Deriving New Variables

- Grouping and Aggregating Data

- Visualization with Python

- Reporting with Excel

Instacart Grocery Basket Analysis
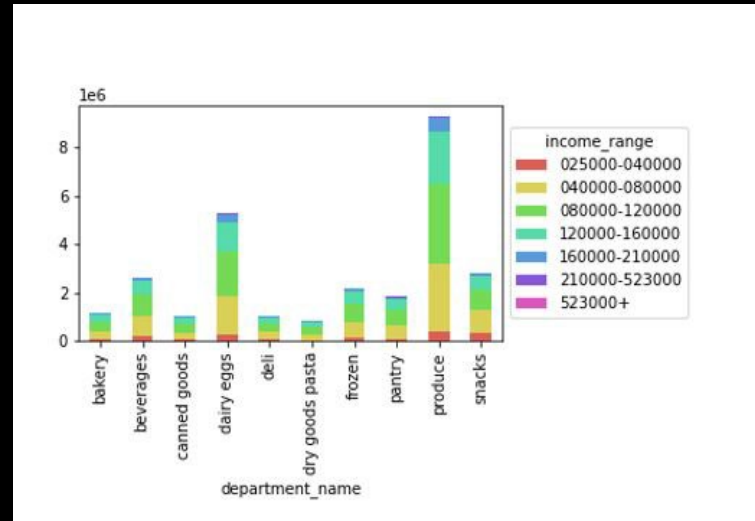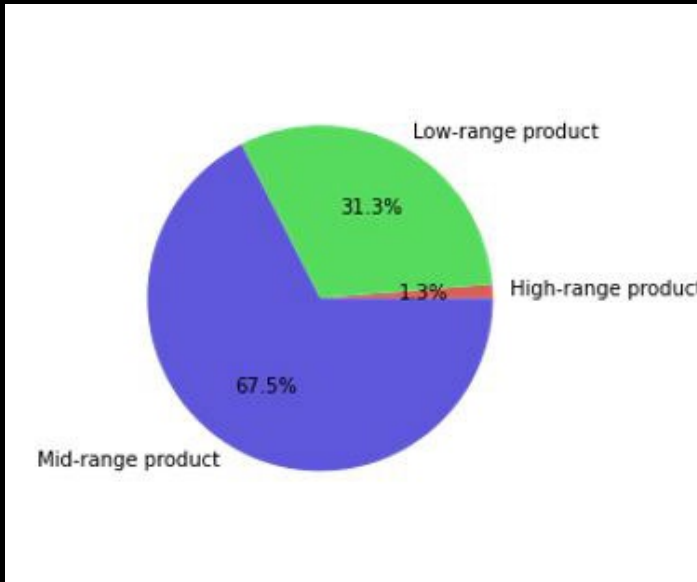
Key Questions:

The sales team needs to know the busiest days of the week and hours of the day.
They also want to know whether there are particular times of the day when people spend.
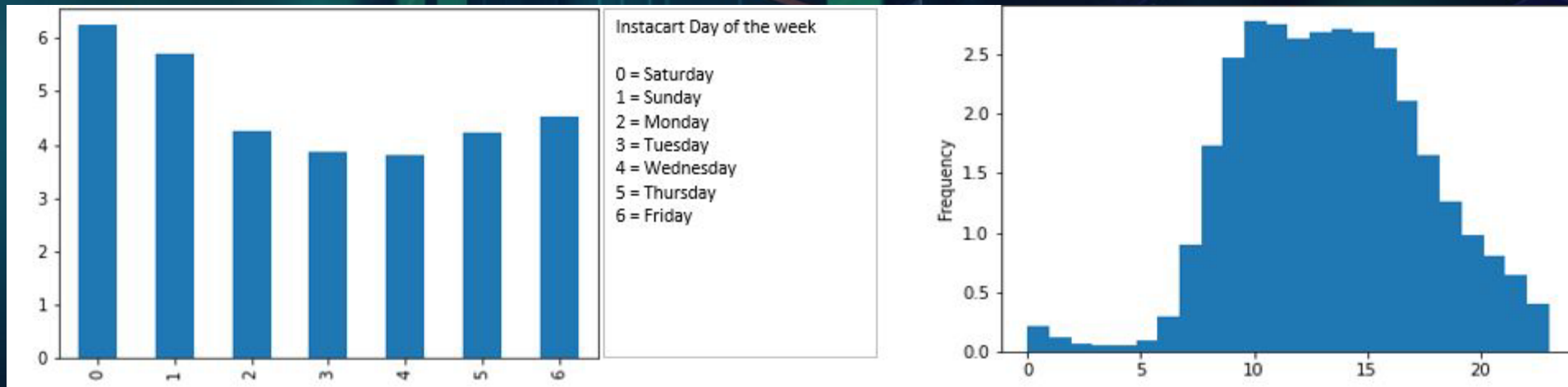
I have created several Visualisations for this Project to Display my findings to Instacarts Management, such as the frequency of Customers or the most sold products based on the Customer's income.
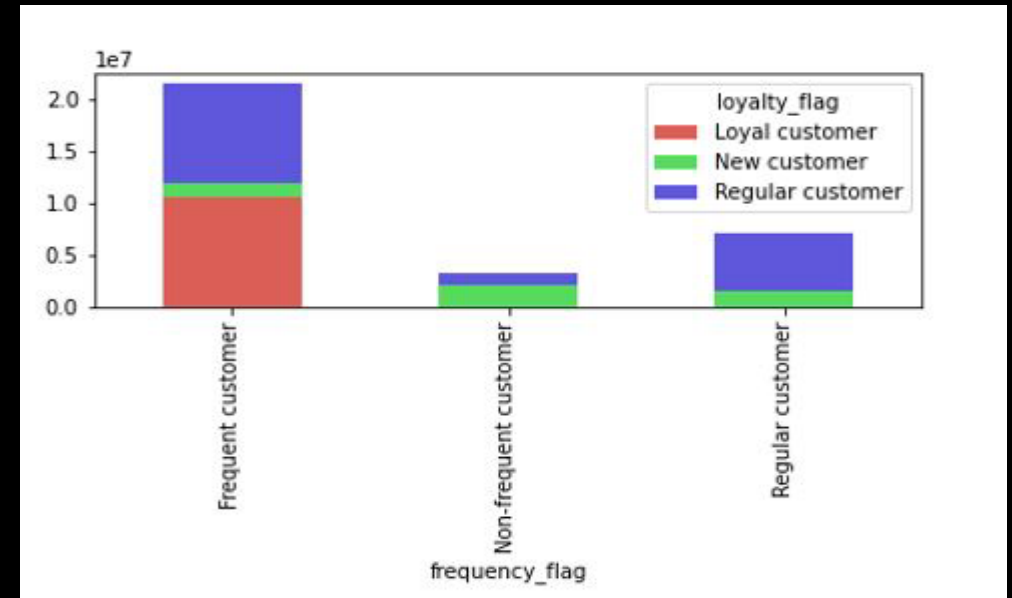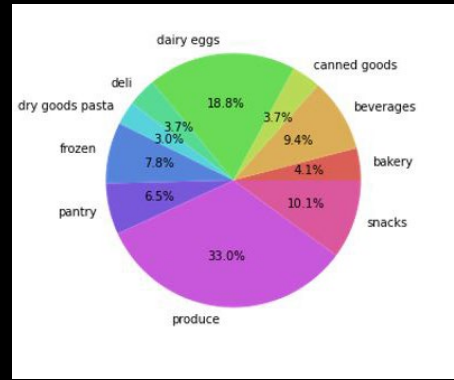
We could identify a pattern in income and spending habits.

Summary and Recommendations for the Instacart Grocery Basket Analysis

Based on the Key Questions provided in the Project Brief, we came to the following conclusion regarding Instacarts Grocery Analysis.

As we can see in our Analysis, the busiest days for Instacart are around the weekend days since that are usually the days people have time for shopping.

To increase business in the middle of the Week, marketing campaigns should be shifted from Tuesday to Thursday in the morning from midnight to 7 am.

We categorized the selling goods into 3 Categories—low-range, Mid-range, and High-range products.

If products from the High-Range categories can be shifted with pricing into the Mid-Range, sales revenue could be increased.

To gain more revenue, marketing should be adjusted towards the leading products.

The highest frequency of departments can use as well t adjust marketing.

The more massive part of Instacarts Customers is loyal and regular customers with the same frequency of Instacart usage.

The whole Analysis, including all Calculations, Queries, and Visualisations, can be reviewed here.

Pig E. Bank Money Analysis

Objective:

Predictive Analysis on Customer retention and a possible reason for Customers leaving the Bank.

Creation of Decision tree

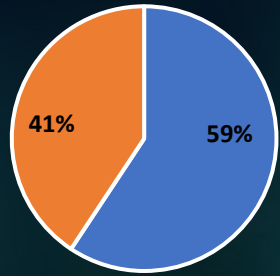## Data:

- **Pig E. Bank Clients**

Tools:

- Excel
- Powerpoint
- Github

Skills obtained:

- Data Ethics
- Data Mining and cleaning
- Predictive Analysis
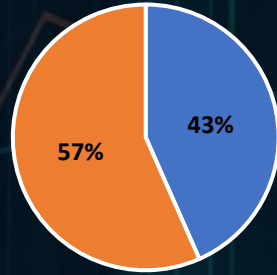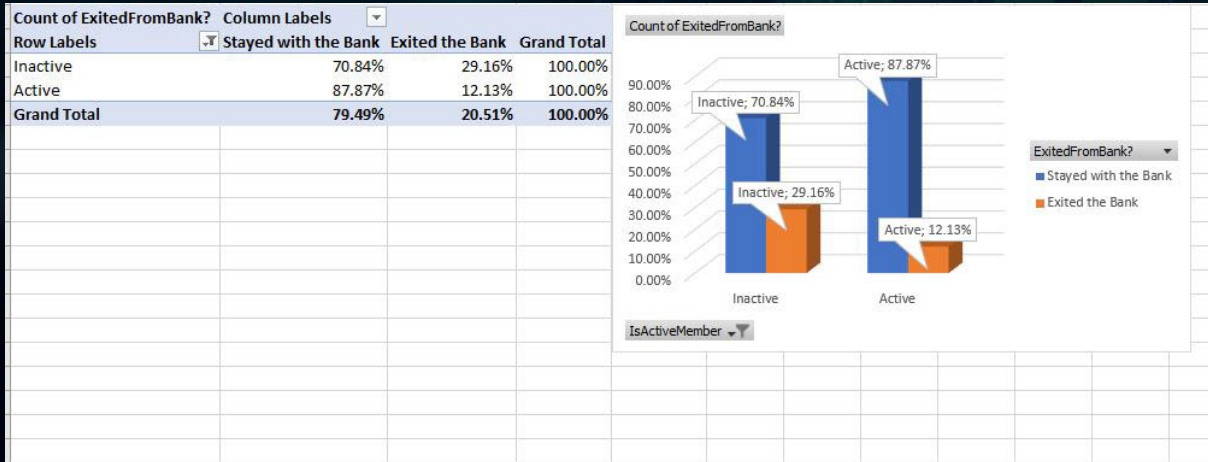- Time-Series Analysis and Forecasting

Exited Bank by Gender

41% 59%

☐ Female ☐ Male

Stayed with Bank by Gender

57% 43%

☐ Female ☐ Male

Here we can see that most customers leaving the bank are female, with 59 Percent who left being Women compared to Man with 41 Percent.

| Count of ExitedFromBank?	Column Labels		
Row Labels	Stayed with the Bank	Exited the Bank	Grand Total
Inactive	70.84%	29.16%	100.00%
Active	87.87%	12.13%	100.00%
Grand Total	79.49%	20.51%	100.00%

I also compared the Activity Status of the Pig. E's Customers in another Pivot Table.
It was visible that customers considered inactive are more likely to leave Pig. E Bank.

Count of ExitedFromBank?	Column Labels		
Row Labels	Stayed with Bank	Exited from Bank	Grand Total
18-27	96.88%	3.13%	100.00%
28-37	91.34%	8.66%	100.00%
38-47	72.73%	27.27%	100.00%
48-57	52.38%	47.62%	100.00%
58-67	48.08%	51.92%	100.00%
68-77	90.91%	9.09%	100.00%
78-87	100.00%	0.00%	100.00%
Grand Total	79.49%	20.51%	100.00%

To identify the most common reasons customers are leaving Pig. E Bank, I have compared the Age of their customers with the status of going or staying with the Bank.

I used a pivot table where I created 10 Year Age Groups to display the Data. Here we can see that most of the customers leaving the Bank are between 48 to 67 of Age.
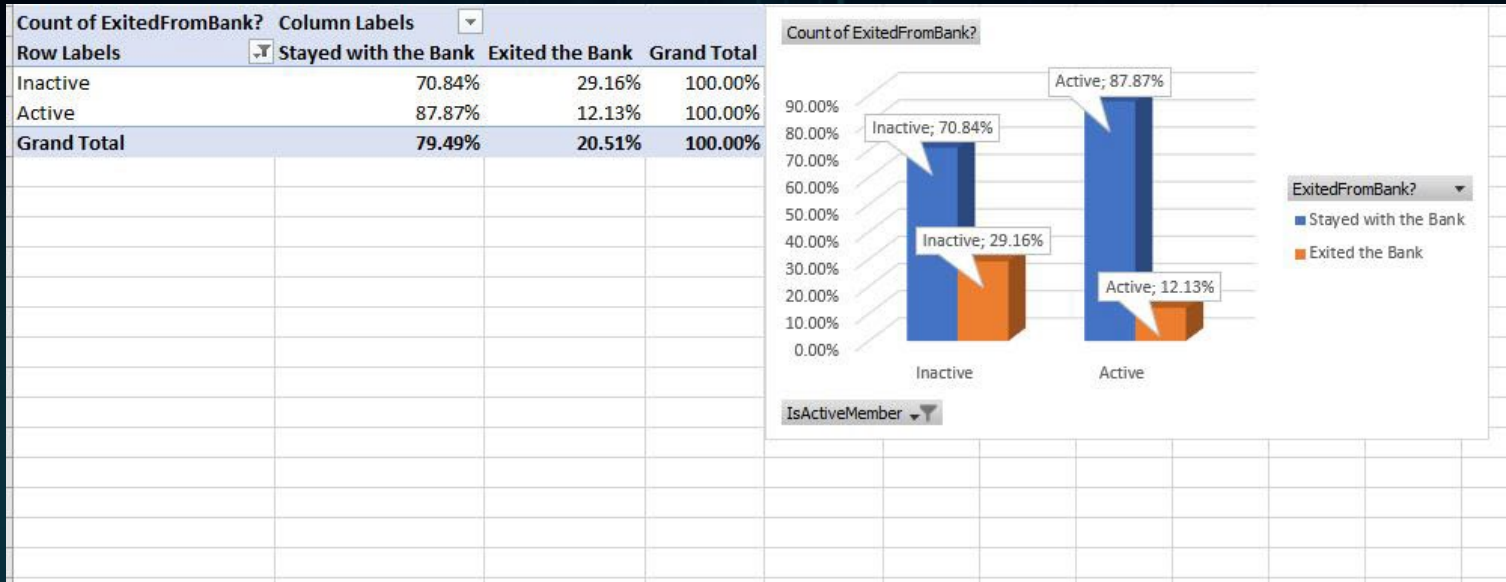
I also compared the Activity Status of Pig. E's Customers in another Pivot Table.

It was visible that customers considered inactive are more likely to leave Pig. E Bank.

Summary and Recommendations Pig E. Bank Analysis

In summary, I can see that several main factors play a Role if and why customers are leaving Pig E. Bank. Mainly I would recommend focusing on Gender, Activity, and Age.
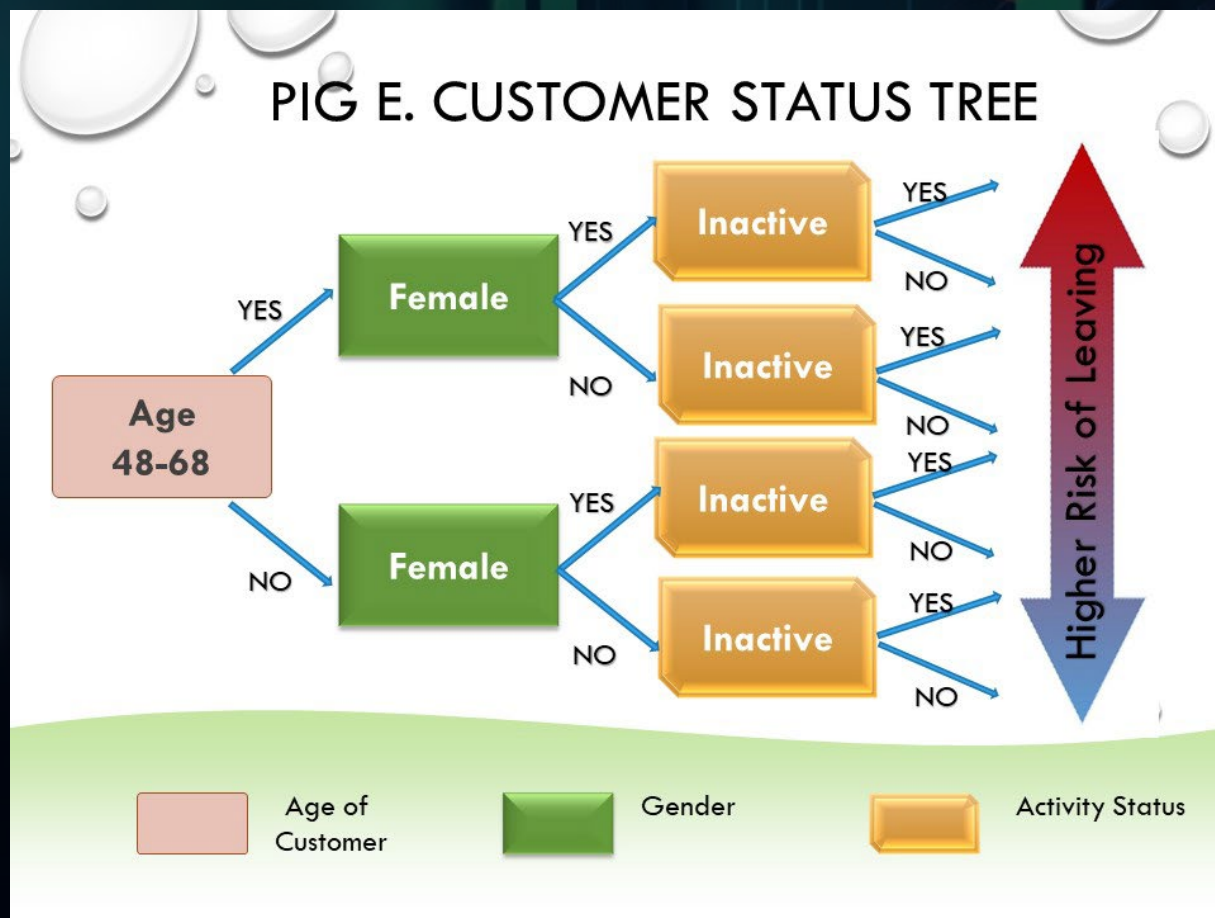
I have created a Decision Tree based on the Customer Status of Pig E. Bank's probability of Leaving the Bank.
The Charts have shown that customers, often women between 48 and 68 Years of Age, are more likely to Leave Pig E. Bank.

The Full Analysis and Presentation can be reviewed here.

Covid-19 Evolution 2019-2022

Objective:

Visualization of the current ongoing Pandemic. Display the Evaluation from the Start until now. How did different Countries change their Tactics in living and fighting against Covid-19 and what is the result.
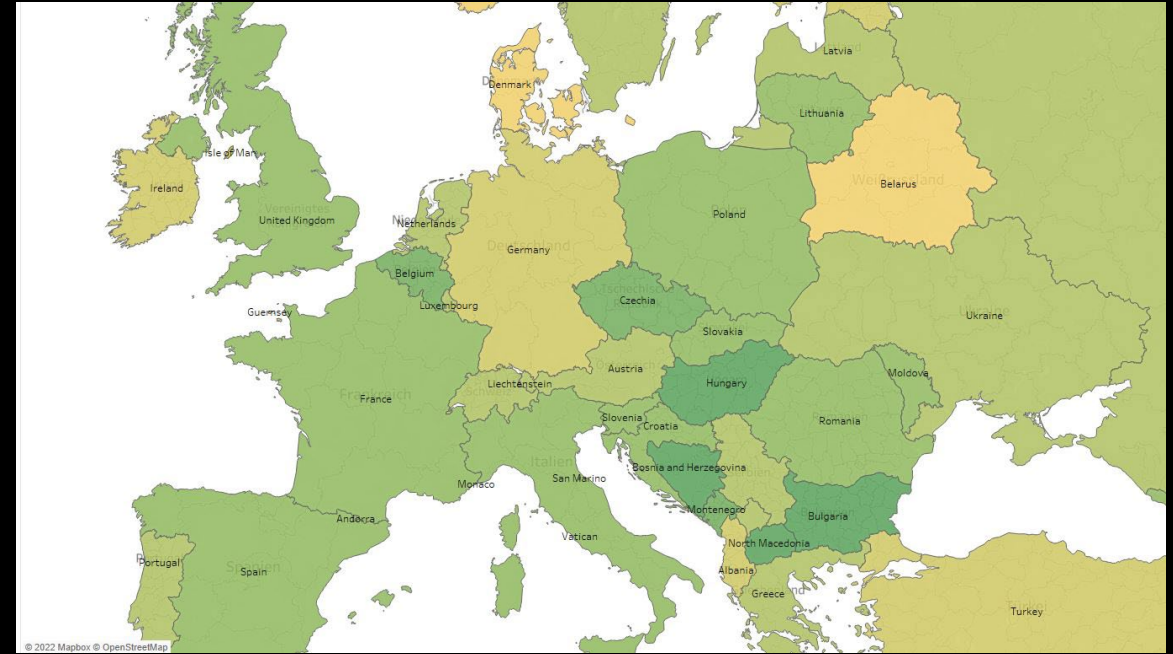
Data:

- Our World in Data

Tools:

- Python
- Tableau
- Microsoft Office Tools
- Machine Learning

Skills obtained:

- Consistency checks
- Correlation Matrix and Heatmaps
- Choropleth Maps-Supervised Machine Learning : Regression
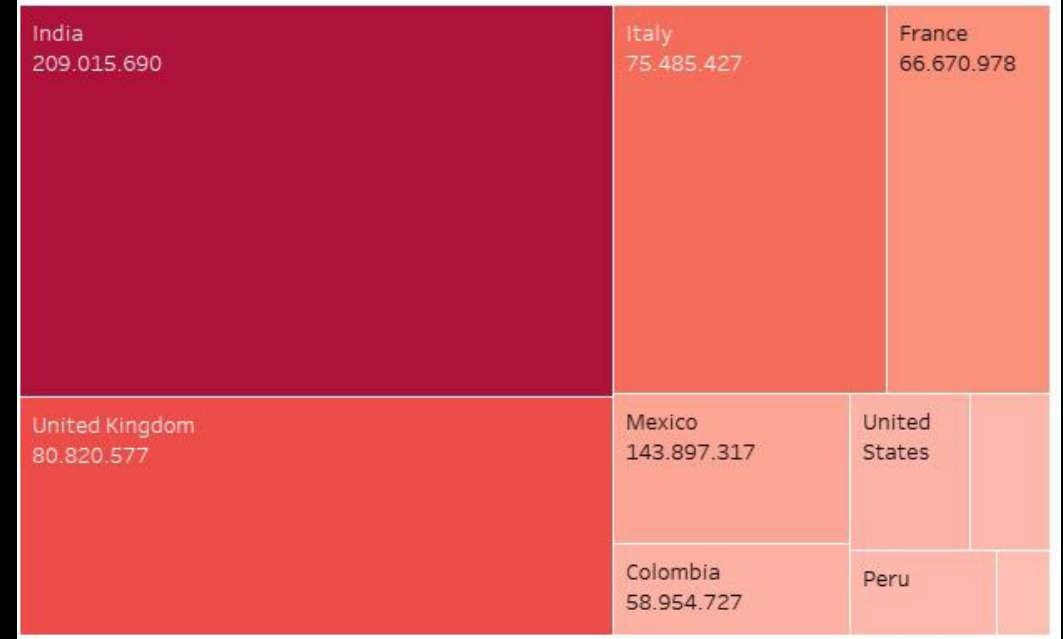- Unsupervised Machine Learning: Clustering-Analyzing Time Series Data

Covid-19 Evolution

- In our Analysis we wanted to have a look on the Covid-19 Disease and the Evolution since the start in 2019. What happened during the ongoing crisis and compare how different Countries adapted changes in the Virus, developed Vaccines and review Restrictions and their Effects on the Evolution such as the 0 Covid Strategie in China or the relatively easy , less Restriction Tactic in Germany in order to gain Herd Immunity.

Top Countries affected by Covid-19 with Population Density

| India 209.015.690 | Italy 75.485.427 | France 66.670.978 |
| United Kingdom 80.820.577 | Mexico 143.897.317 | United States |
| | Colombia 58.954.727 | Peru |

Covid-19 Evolution

- I created several Visualisations to show the Evolving part of the Covid 19 Disease and how it affects each country differently.

- This Map shows which countries have been most affected by Covid-19 by their Population Density.
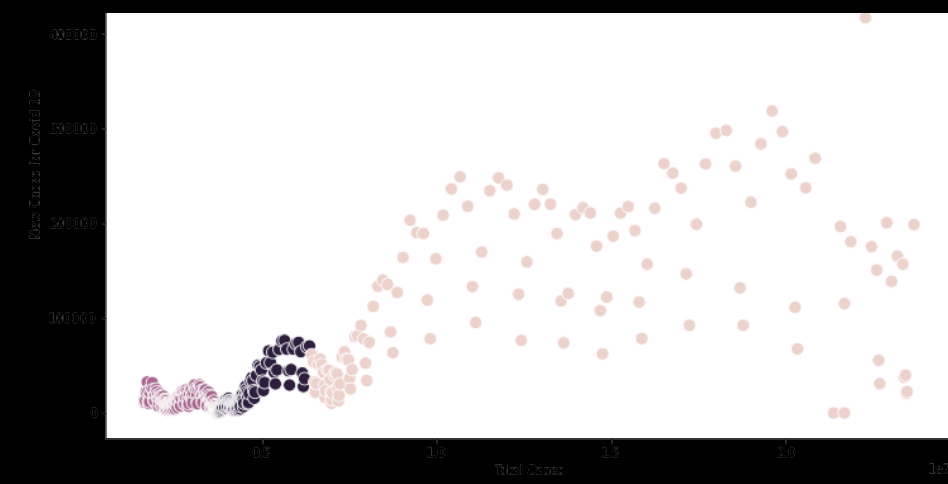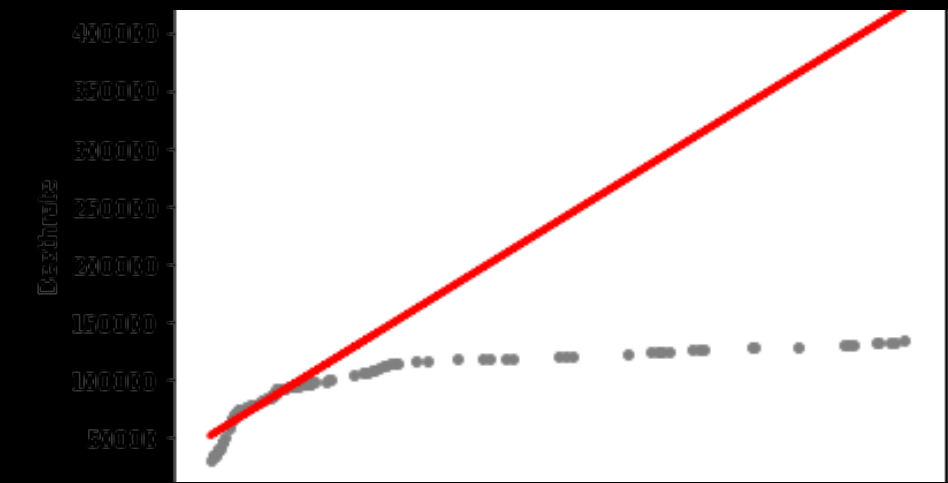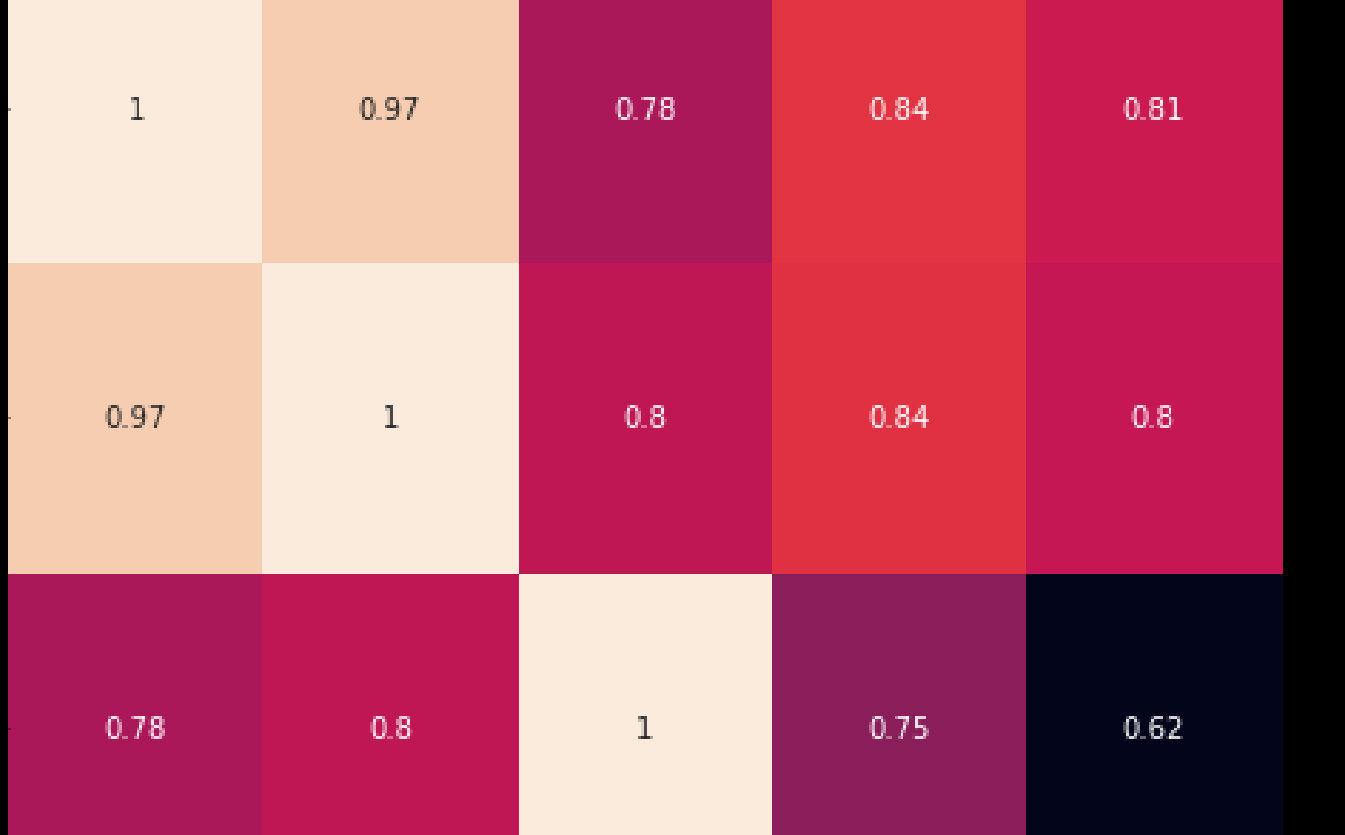
- I created Heatmaps that shows the Correlation

- For Analysis i created Trendlines and Clustering Models

- I used different techniques to analyze and present my results and findings.
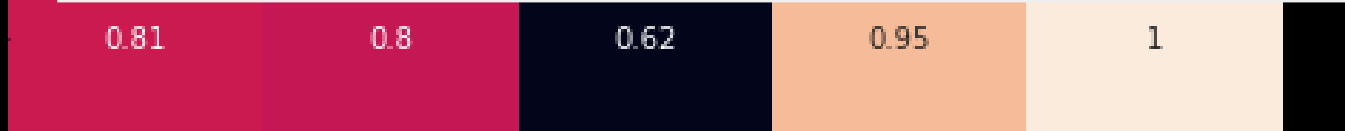
Summary of Covid-19 Evolution Analysis

- Since the Covid-19 Disease started to spread around the Globe in 2019, we can see the Evolution of the Disease in Every Country. From China beginning with a 0 Covid Tolerance, bringing a Lockdown to whole Cities with Millions of People, to Countries that denied the Existence of Covid contemplating a Conspiracy to other countries who have been entirely "relaxed."

- Overall, the World concluded that only together this can be defeated as even North Korea recently started to confirm and report Cases.

- However, worldwide, it has been shown, despite the differences in Politics, the efficiency in Vaccinations that it is possible to Live with Covid.

- The whole presentation can be reviewed under Tableau and Github

Stefan Rieß

Data Analytics Portfolio

Questions? Feel free to contact me!

Github
Tableau
Linkedin