Data Analyst Portfolio

Atika AhmedSudi

Projects

Game Co

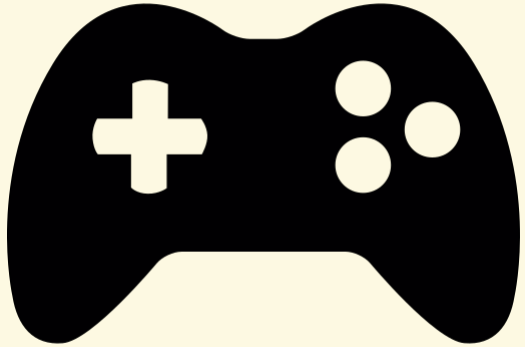
Preparing for Influenza

Rockbuster Stealth

Instacart

Pig E. Bank

Medical Cost Personal



Game Co

Overview

- **Objectives**: GameCo analyst conducted a descriptive analysis of the data set to understand the market demand for new games, aiming to inform game development and market success.

- **Primary goals**: GameCo executives were seeking insights from data on popularity, primary competitors, fluctuations, and sales figures across different geographical regions. They were interested in discrepancies among game genres, publishers' primary competitors, and fluctuations in popularity over time.

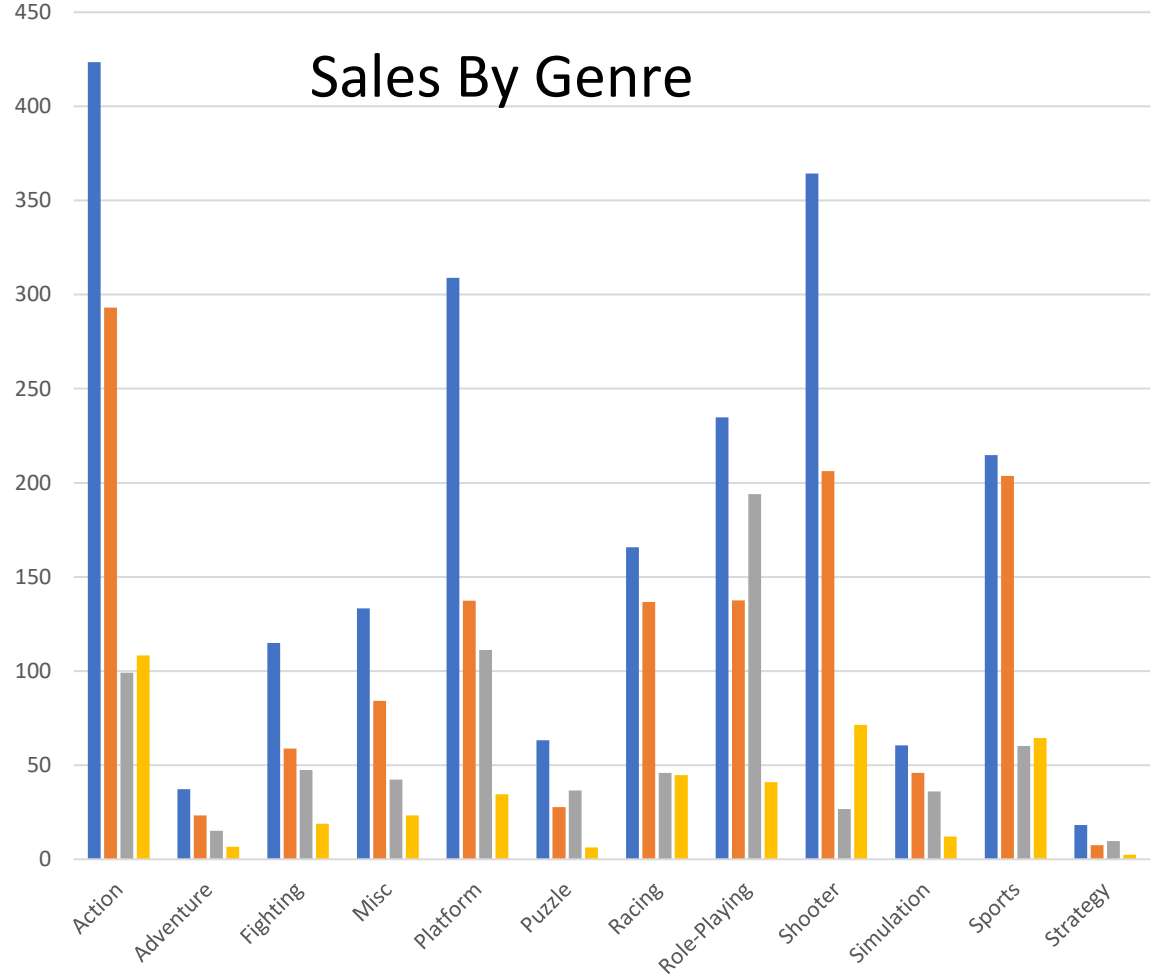
- **Data Set**: From 1980 through 2016, the data collection from VGChart tracks historical sales of video games with over 10,000 copies across platforms, genres, and publishing studios.

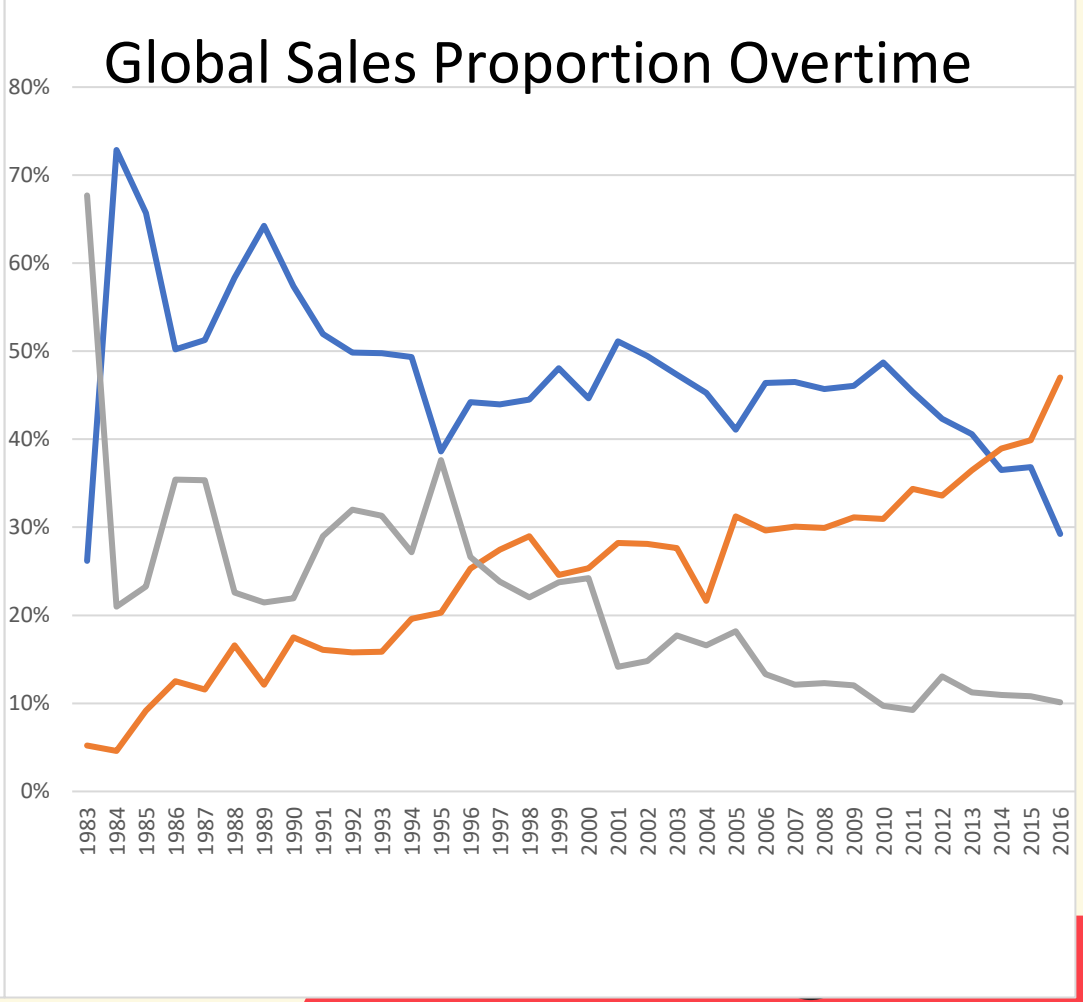
- **Methods**: Data cleansing, data integrity, quality, pivot tables, data grouping, summarising, calculated fields, and descriptive analysis were all developed techniques.

Analysis

- **Data analysis**: In this study, I used Microsoft Excel to retrieve, organize, and refine data, create a pivot table, and create visual representations. Data were categorized, examined, and identified biases. Descriptive analysis techniques were used to investigate fluctuations in video game sales. Prioritizing marketing budgets based on sales trends was suggested, with Europe, North America, and Japan being the top three regions.
- **Recommendation:** To ensure success in the upcoming year, research the decline in game sales and determine customer base locations. Phase out miscellaneous and racing games, launch local-targeted marketing campaigns and develop culturally relevant games for specific regions.

Visualization





Preparing for the influenza

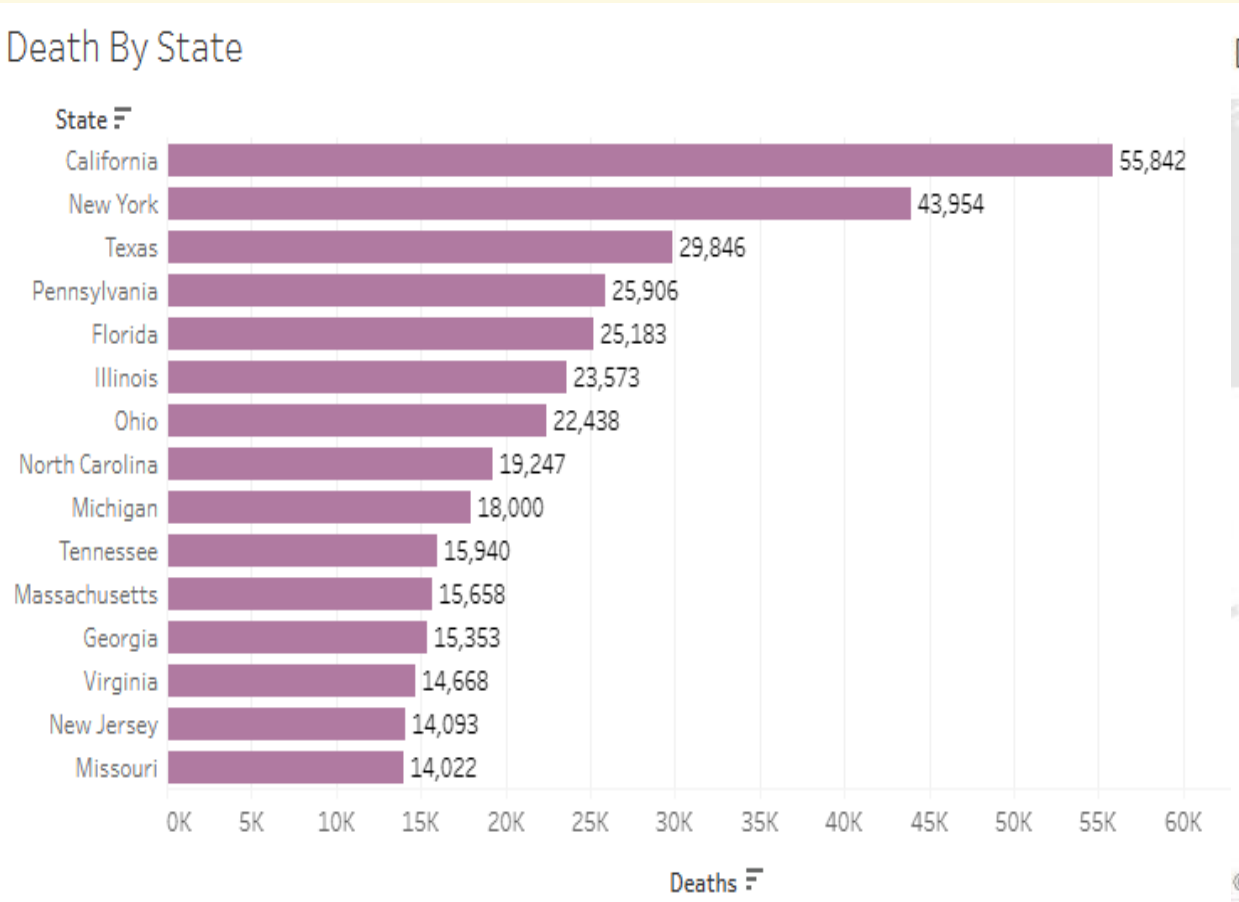
Overview

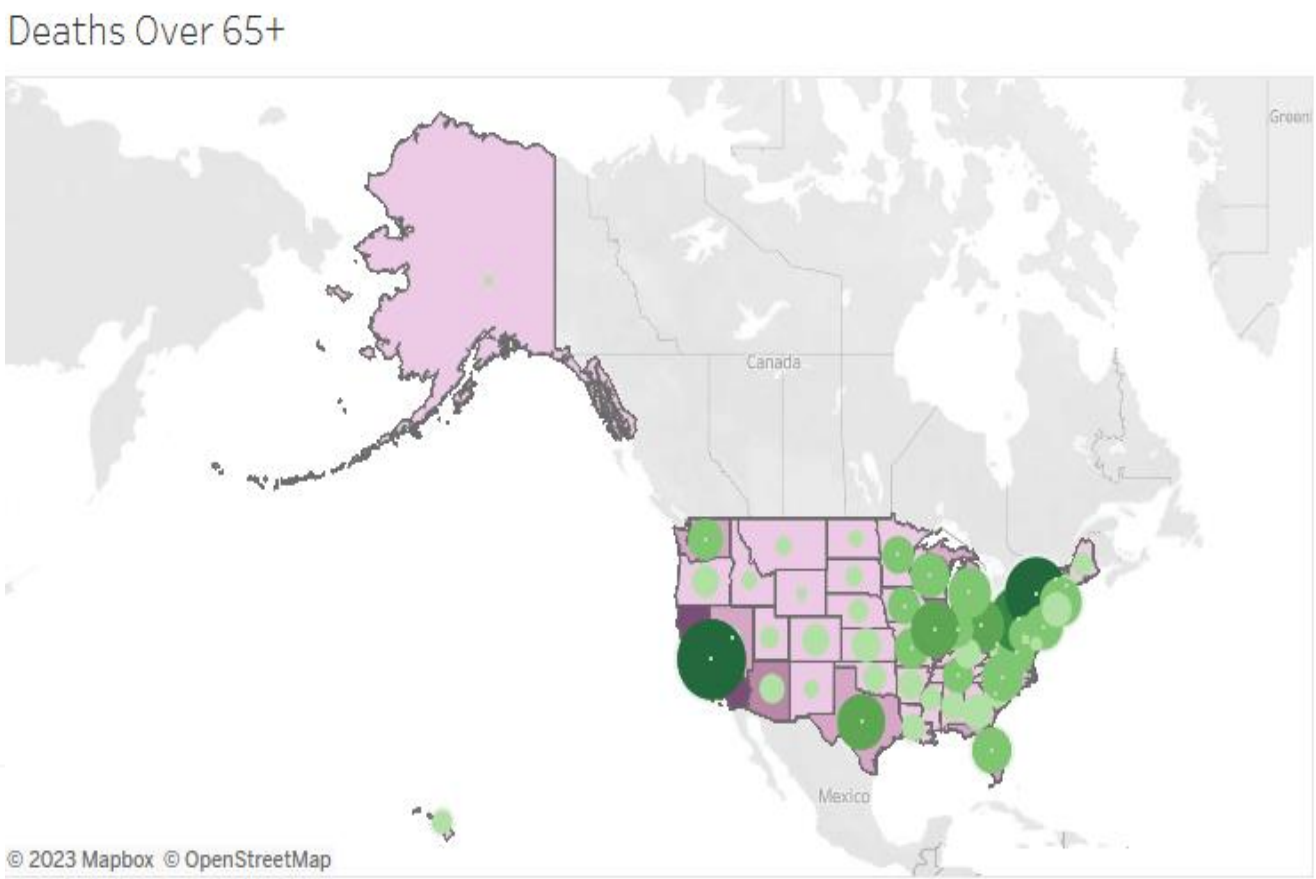
- **Objectives**: The influenza season in the US affects vulnerable populations, increasing demand for healthcare personnel. A medical staffing firm supplies temporary workers, determining dispatch timing and number for each state.
- **Primary goal**: Medical staffing service analyses influenza trends to plan for a high-demand workforce during flu season, ensuring temporary worker availability in clinics and hospitals.
- **Data Set**: Influenza deaths Sourced from the CDC, Population data Sourced from the US Census Bureau
- **Methods**: Techniques gained data cleaning, data transformation, integration, data profiling, integrity, research project design, and visualisation tools such as PivotTables, and VLOOKUP.

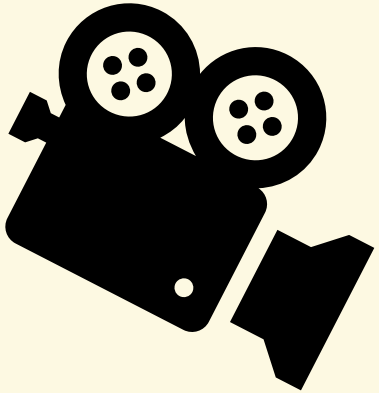
Analysis

- **Data Analysis**: Excel and Tableau were used for data profiling, integrity concerns, and data consolidation in a research project. Statistical studies were conducted to compute variance, and standard deviation, and analysed relationships. Results were summarized in a comprehensive report.
- **Recommendations**: To combat influenza during January, February, March, and December, an augmented healthcare workforce is needed in states with higher population densities, such as Florida, Ohio, Tennessee, North Carolina, and Alabama. Prioritization should be given to treating individuals aged 65 and above.

Visualization







Rockbuster

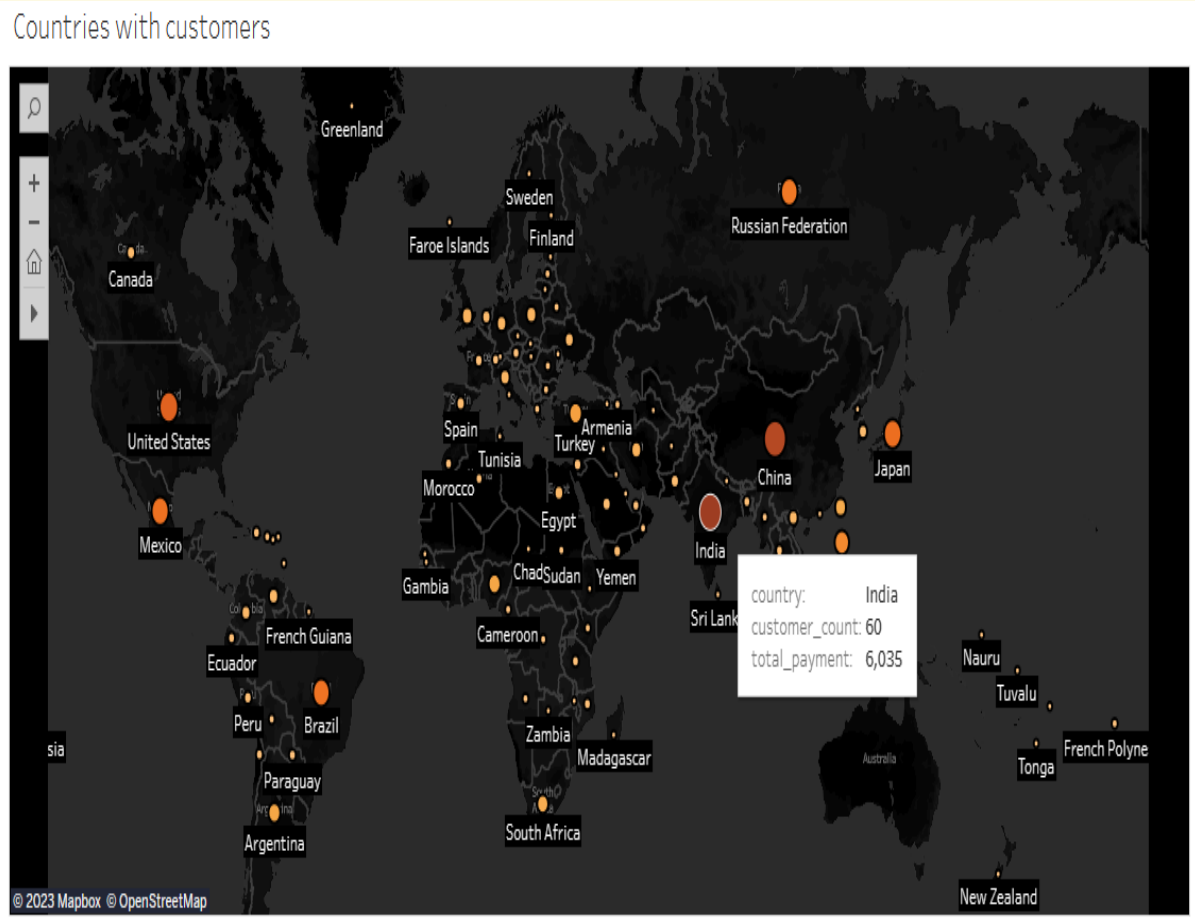
Overview

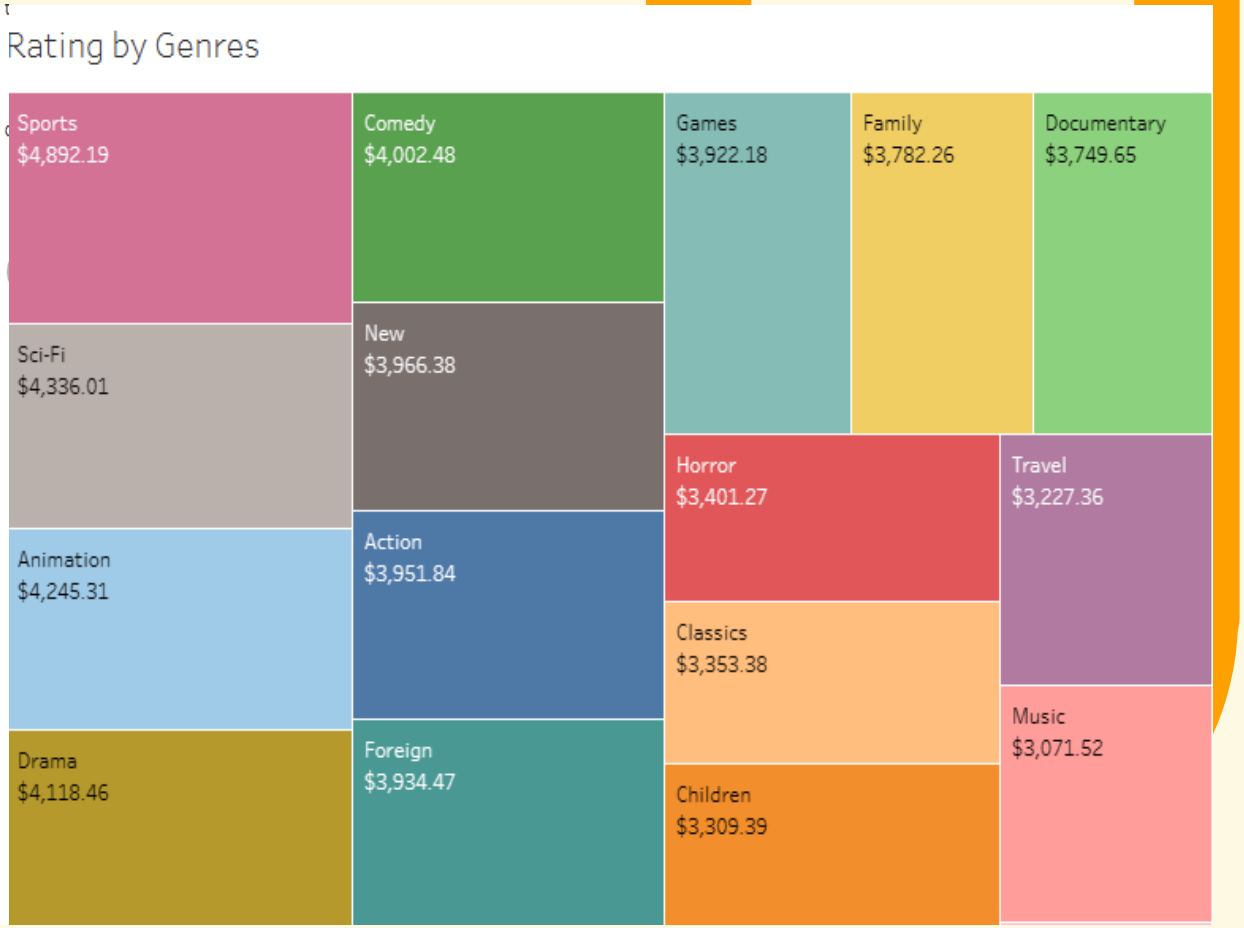
- **Objective:** Rockbuster Stealth is shifting from physical storefronts to online streaming services in order to boost revenue, rental duration, customer base, and sales across regions. For the analysis, a data analyst is hired.
- **Primary Goals:** The Rockbuster Stealth Management Board seeks data-driven insights for their 2020 company plan, focusing on revenue, rental durations, customer locations, high lifetime value, and sales performance across regions.
- **Data set:** Data set from Rockbuster Stealth LLC. All rentals' history data is included.
- **Methods:** Data cleaning involves relational databases, the creation of a data dictionary, the creation of an ERD, the filtering of data, the joining of tables, and the execution of subqueries.

Analysis

- **Data analysis**: Data cleaning entails maintaining databases, building dictionaries and ERDs, filtering and connecting tables, and running SQL subqueries to provide profiles and summary statistics.
- **Recommendation**: Rockbuster's film collection should be expanded to include international languages, with a focus on Chinese dialects and Indian films. Marketing activities should be aimed at India, China, the United States, Japan, and Mexico. Sports, science fiction, animation, drama, and comedy films generate revenue.

Visualization





Instacart
pandas
jupyter
matplotlib

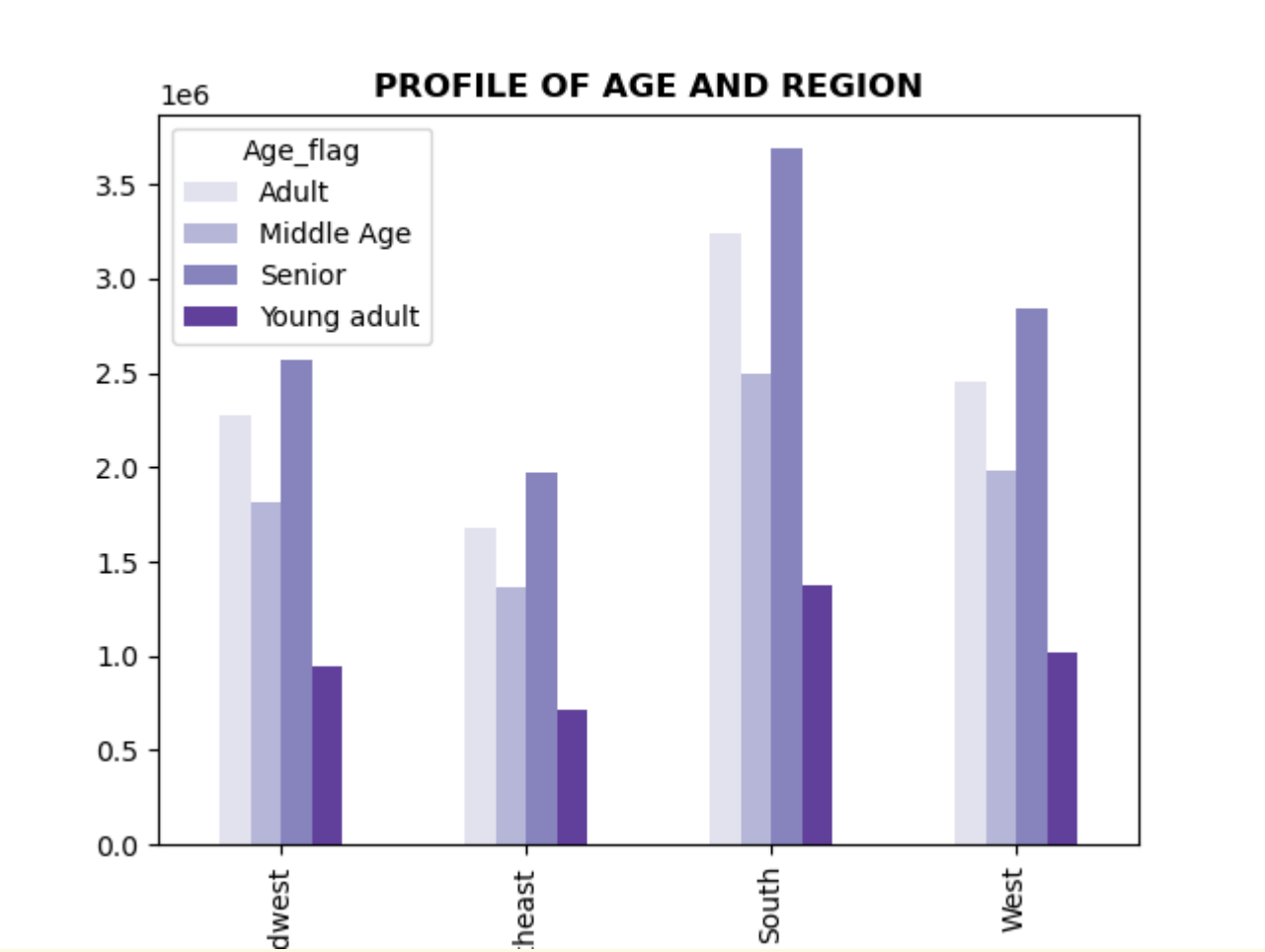
Instacart

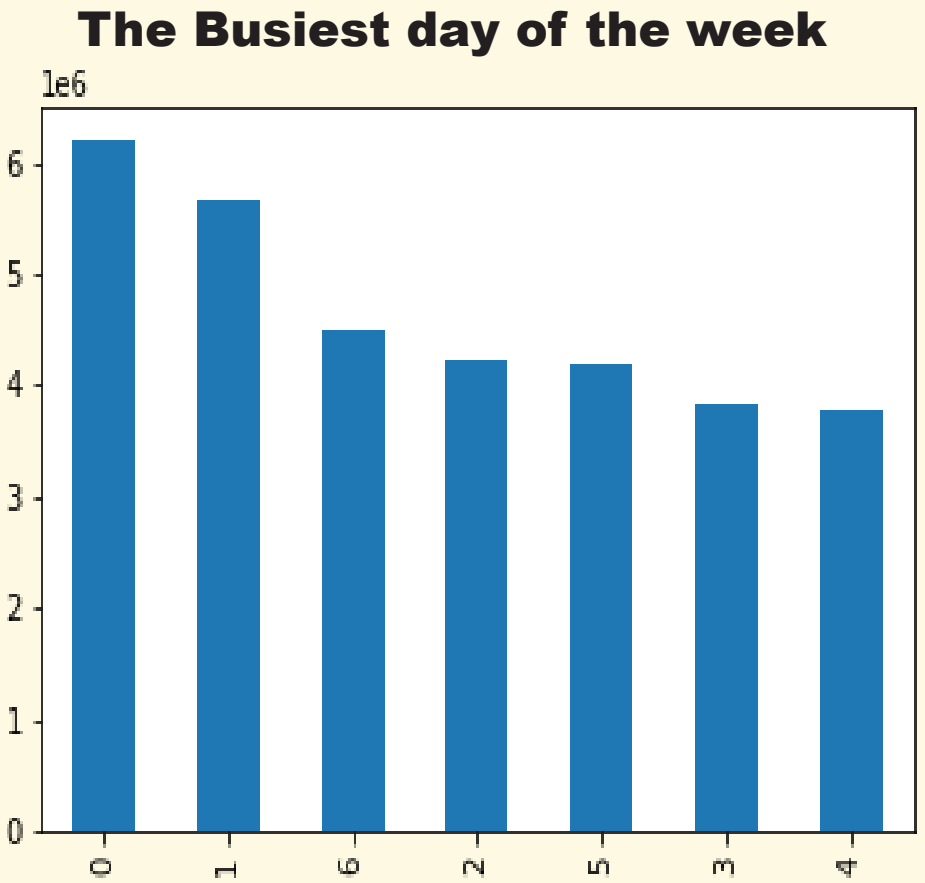
- **Objective**: Analyse Instacart's sales patterns, derive insights, and propose segmentation strategies based on provided criteria for enhanced customer experience.
- **Primary Goals**: Sales team schedules ads based on busy days, hours, popular products, and expensive time periods. Instacart simplifies pricing, understands customer behaviours, and customises strategies based on order price, frequency, and product types.
- **Data set**: Instacart's 2017 Online Grocery Shopping Dataset includes Data Dictionary.
- **Methods**: Wrangling, merging, exporting, grouping, and aggregating variables, Python visualisation, and Excel reports were all part of the data cleaning process.

Analysis

- **Data analysis**: Anaconda and Jupyter were used for project directory setup, with Python libraries imported. Data wrangling, subset extraction, consistency testing, merging outcomes examined, conditional logic applied, and summary columns created. Customer group behaviour was investigated, and data links were discovered.
- **Recommendation**: Schedule ads before and after 10 am -3 pm, Monday-Thursday to increase sales. Consider non-peak hours between 7-8 am and after 4 pm. Incentivize high-range products with coupons, discount vouchers for departments with lower popularity, and establish loyalty rewards. Make use of patterns from loyal customers and tailor rewards programs for the Northeast. Conduct research to understand the reasons behind lower popularity in the region.

Visualization





Pig E. Bank

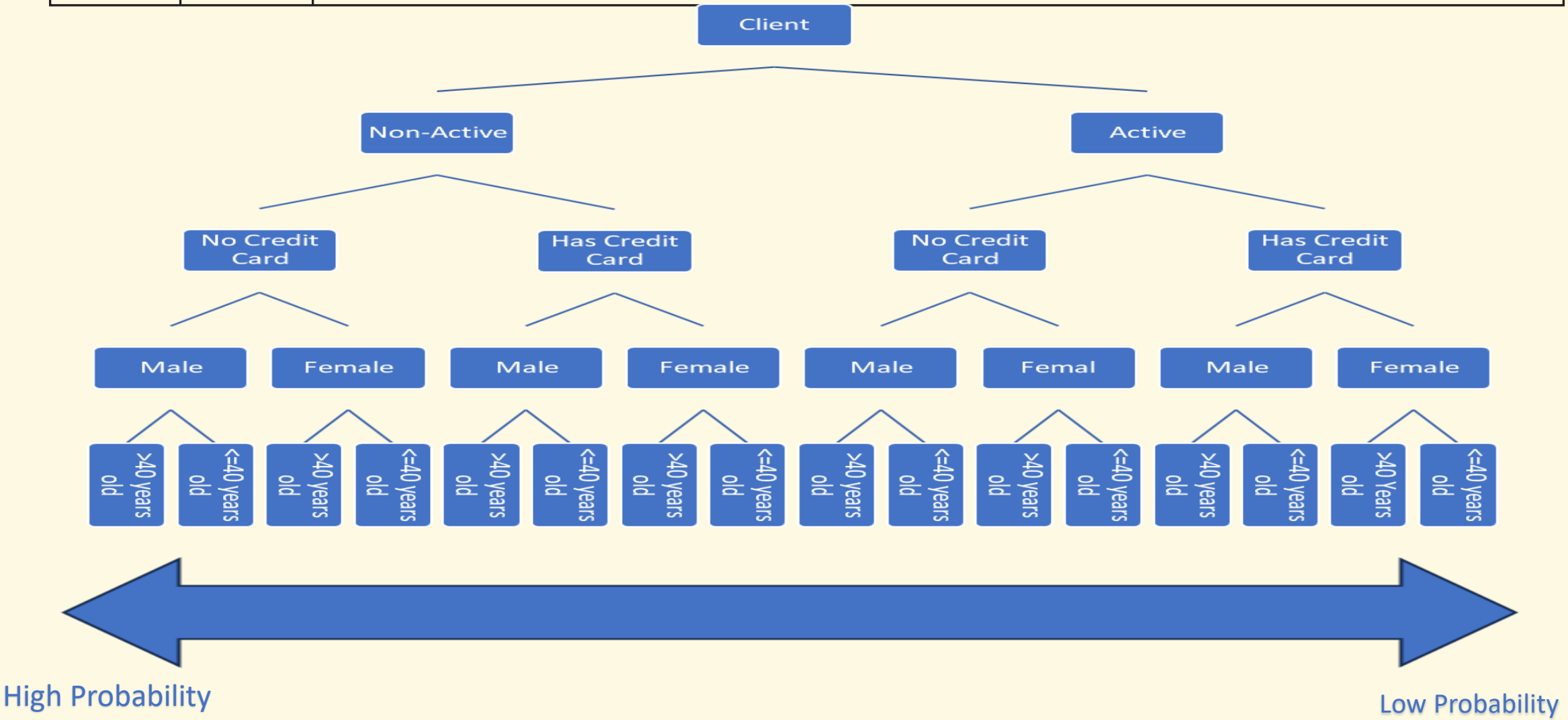
Overview

- **Objective**: The role involves data-related projects evaluating risks, generating metrics, and improving compliance. Challenges include strong technical skills and ethical dilemmas in data handling.
- **Primary Goals**: Decision trees forecast bank client departures using data mining techniques. Data bias and ethics impact data consumption, sharing, and security. Fundamentals like predictive analysis, linear regression, and time-series forecasting are covered. GitHub helps refine abilities and display portfolios for job applications.
- **Methods**: Sorting, filtering, summarising, descriptive analysis, data ethics, and data mining with decision trees were all part of the data analysis.

Analysis

- **Data analysis:** In this project, I explored the structured and unstructured data properties, uses, limitations, and big data computing systems' potential. The analysis addressed workplace prejudices, bias minimization, and ethical concerns, focusing on data mining methods and workplace security.

Decision Tree



Medical Cost Personal

Overview

- **Objective**: To create an engaging dashboard that visually presents carefully selected outcomes from a sophisticated exploratory analysis carried out using Python.

- **Primary Goals**: This analysis examines the relationship between age and BMI variables and medical charges, investigating their impact on costs. A predictive model is constructed to estimate charges for new customers based on their characteristics.

- **Data Set:** Medical Cost Personal Datasets Insurance

- **Methods:** Cleaning, Wrangling, Exploring relationships with Python, Cluster visualization with Tableau, supervised and unsupervised machine learning.

Analysis

- **Data analysis:** The impact of age and BMI on insurance expenses cannot be definitively concluded, as insurance pricing involves multiple factors beyond age and BMI. Further investigation is needed to discern significant components. The dataset's constrained age range limits insights into potential variations and intricacies influencing observed outcomes.

- **Recommendation:** Expanding the medical insurance analysis dataset to include elderly populations and supplementary factors beyond age and BMI is crucial for precision. This holistic understanding of insurance pricing and translating insights into actionable policy recommendations will enhance healthcare coverage and informed decision-making.

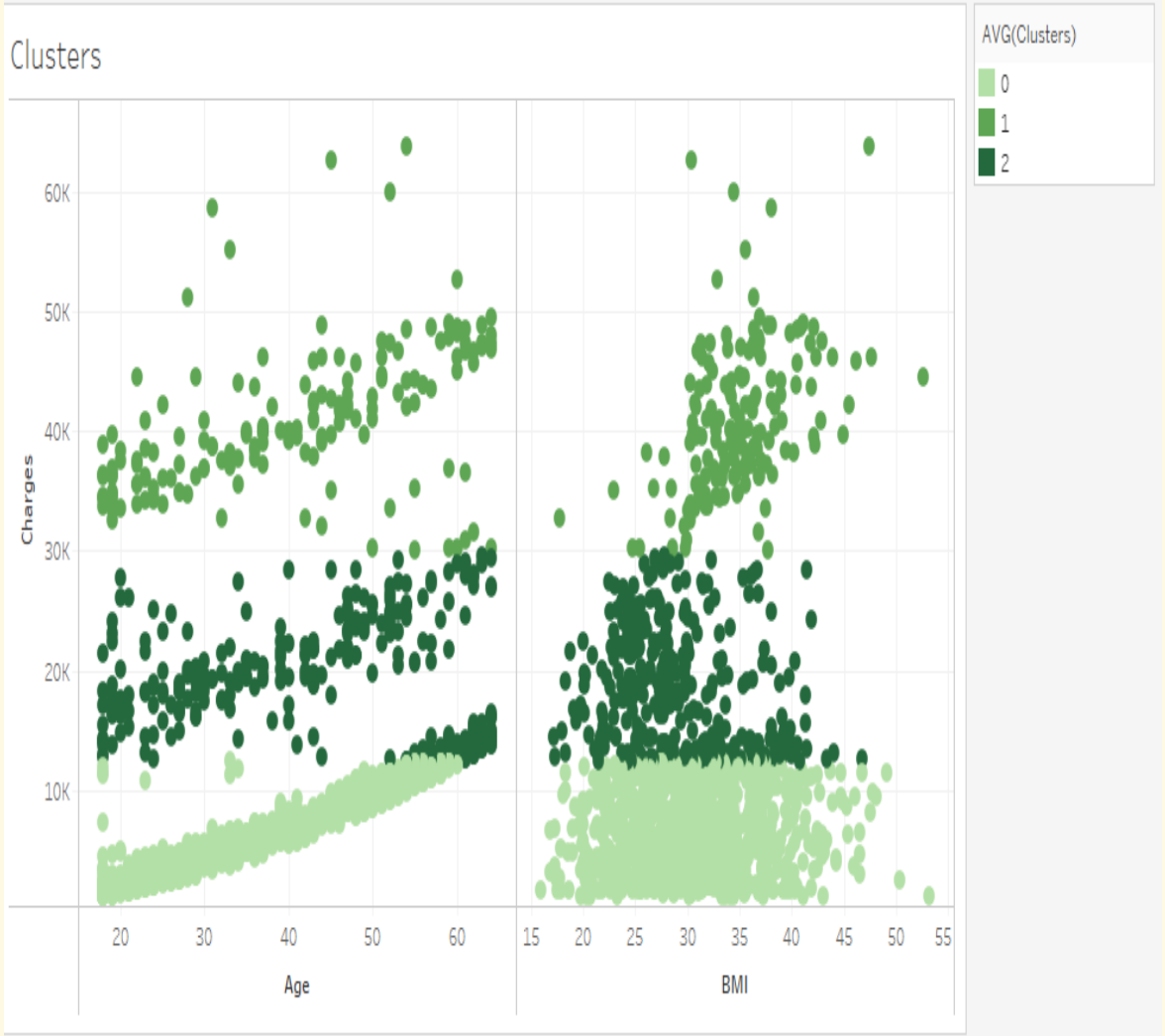
Computation

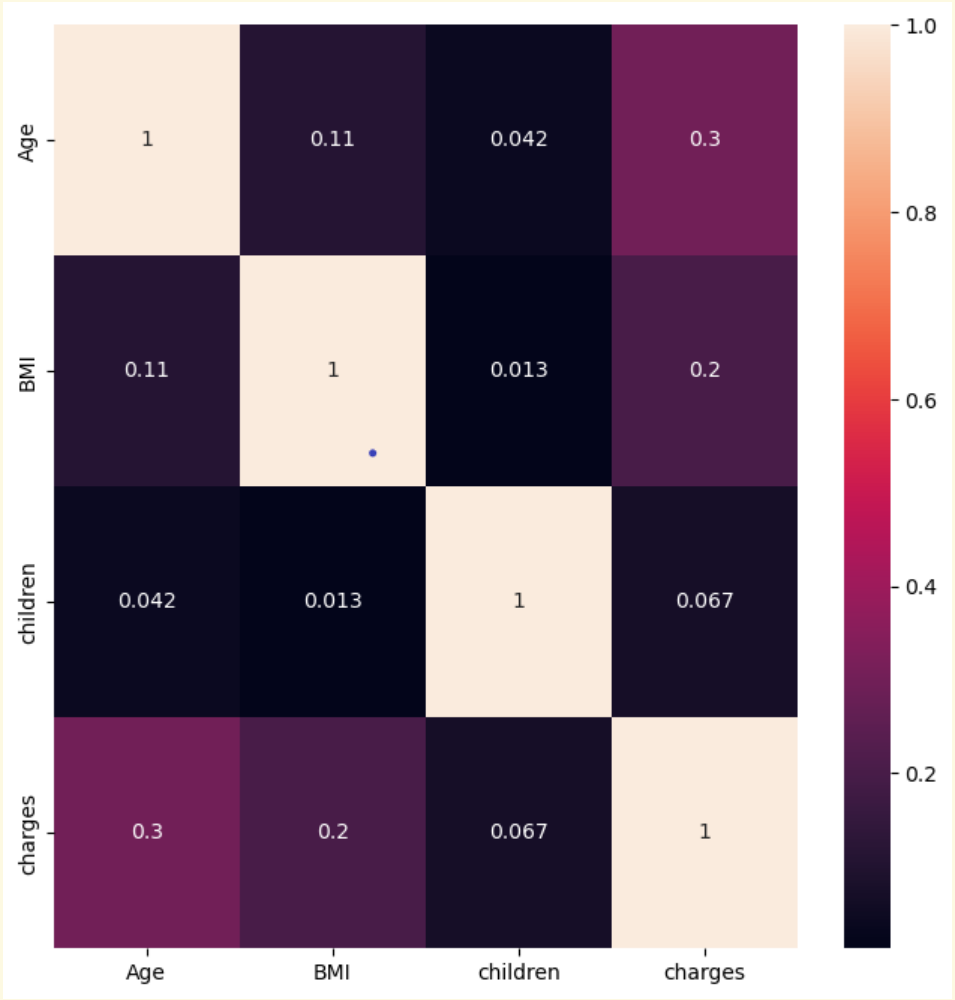
	Age	*charges*
Mean	39.22213912	13279.12149
Variance	197.2432819	146660811
Observations	1337	1337
Hypothesized Mean Difference	0	
df	1336	
t Stat	-39.97542172	
P(T<=t) one-tail	8.741E-231	
t Critical one-tail	1.645994969	
P(T<=t) two-tail	1.7482E-230	
t Critical two-tail	1.961741219	

	BMI	*charges*
Mean	30.66345176	13279.12149
Variance	37.21571482	146660811
Observations	1337	1337
Hypothesized Mean Difference	0	
df	1336	
t Stat	-40.00128491	
P(T<=t) one-tail	5.4573E-231	
t Critical one-tail	1.645994969	
P(T<=t) two-tail	1.0915E-230	
t Critical two-tail	1.961741219	

	Age		BMI		charges	
	mean	median	mean	median	mean	median
cluster						
dark purple	39.828025	41.0	35.023917	34.800	41097.912181	40003.332250
pink	44.448148	47.0	28.703944	27.915	19613.166668	19074.327800
purple	37.567033	38.0	30.492544	30.115	6600.283753	6367.275475

Visualization





Thank you

Atika AhmedSudi

AtikaSudi@gmail.com