Data Analytics Portfolio

Nathan Evans

Greetings! I'm Nathan Evans, a passionate data analyst with a knack for turning raw data into compelling narratives. I thrive on uncovering insights that drive strategic decision-making.

Soft Skills

- **Communication:** Clear, concise, and engaging.
- **Problem Solving:** Analytical approach to challenges.
- **Teamwork:** Collaborative mindset, adept and experienced at working in diverse teams.
- **Adaptability:** Flexible and quick to learn new concepts.

Tools & Technologies

- Excel
- Tableau
- Python
  - Jupyter Notebook
  - Pandas
  - Numpy
  - Seaborn
  - Matplotlib
  - Plotly
  - Folium
- PostgreSQL
  - pgAdmin4
- AWS Certified Cloud Practitioner

Analytical Skills

- Translating Business Requirements
- Cleaning & Preparing Data
- Data Wrangling
- Data Merging
- Data Integration
- Forecasting
- Regression
- Clustering
- Descriptive Analysis
- Geospatial Analysis
- Data Visualization
- Presenting Results & Storytelling

Projects & Tool Utilization

GameCo Sales Trends

Influenza Staffing Solutions

Rockbuster Business Intelligence

Instacart Insights

Exploring World Happiness

GameCo Project Overview

In my capacity as an analyst at GameCo, this project aims to conduct a detailed analysis of video game sales data. The primary objective is to uncover market trends, explore popular genres, analyze regional variations in sales, and examine sales fluctuations over time. The insights gleaned from this analysis will play a crucial role in shaping the sales strategies for future game releases.
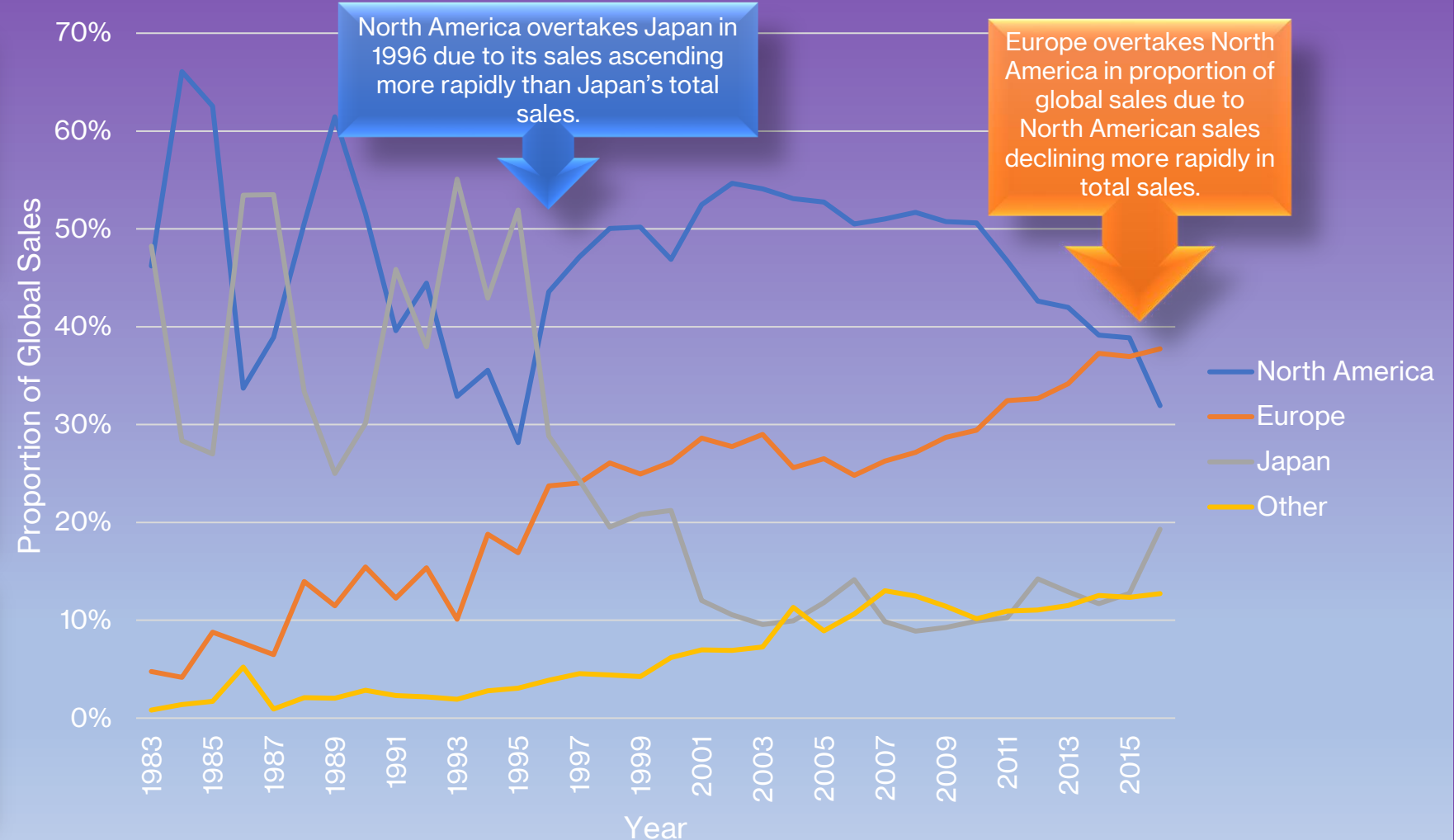
Project Data:
Original Data Set
Project Brief
Presentation
Reflections

Tools:
Excel
PowerPoint
Pivot Tables
Visualization Tools

Figure 2. Proportion of Global Sales by Region Over Time (1983-2016)

North America overtakes Japan in 1996 due to its sales ascending more rapidly than Japan's total sales.

Europe overtakes North America in proportion of global sales due to North American sales declining more rapidly in total sales.

Legend:
— North America
— Europe
— Japan
— Other

Y-axis: Proportion of Global Sales (0% – 70%)
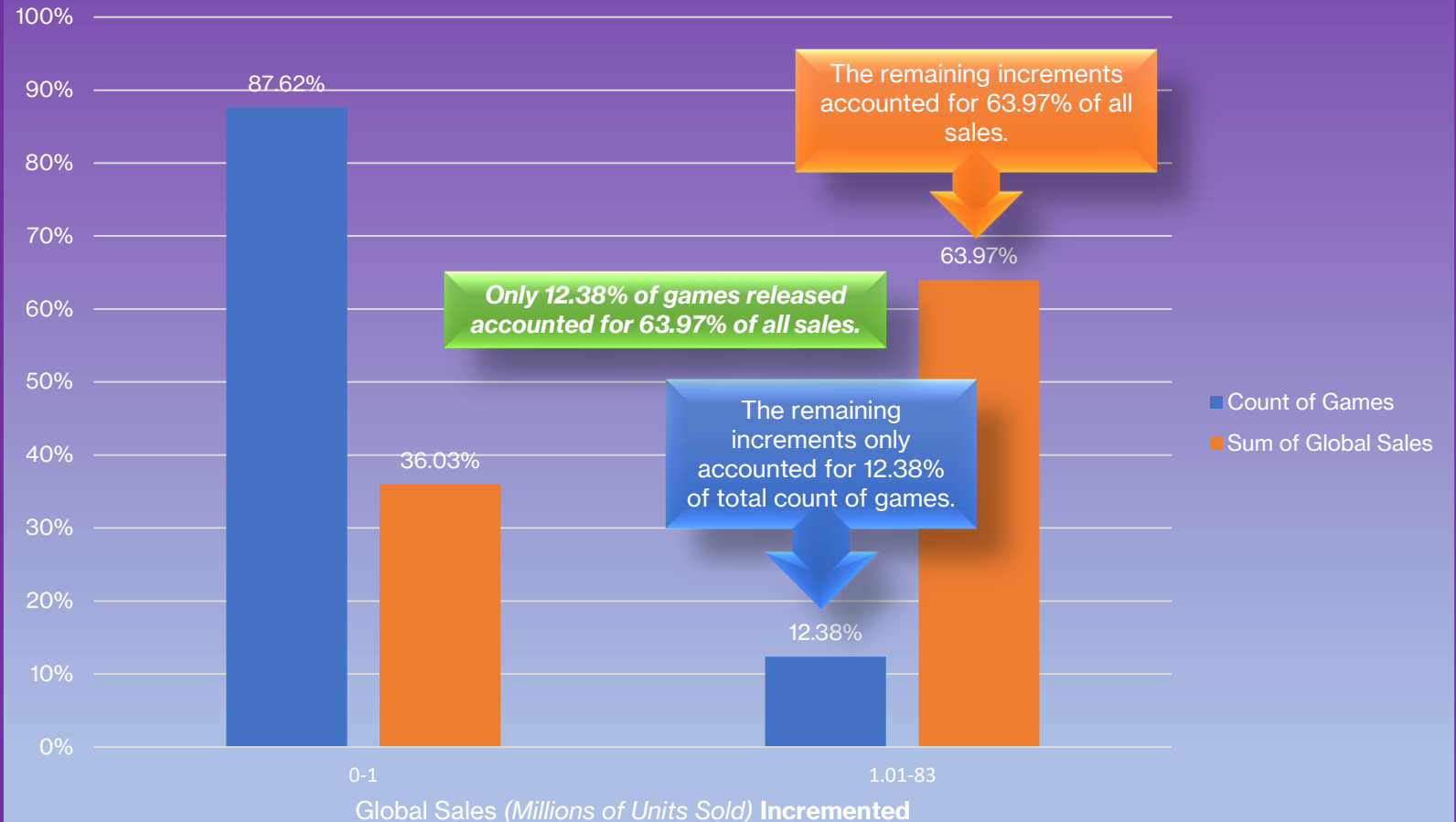X-axis: Year (1983–2015)

GameCo Project Analysis

Conducted a thorough analysis of the dataset to uncover key business insights.

Created visual aids to present the analysis findings, enabling clear communication of data-driven insights to stakeholders for a renewed business understanding.

Figure 6. Count of Games & Sum of Global Sales (1983-2016)

The remaining increments accounted for 63.97% of all sales.

Only 12.38% of games released accounted for 63.97% of all sales.

The remaining increments only accounted for 12.38% of total count of games.

- Count of Games
- Sum of Global Sales

87.62%
36.03%
63.97%
12.38%

Global Sales *(Millions of Units Sold)* **Incremented**

0-1
1.01-83

Process | *Exploratory Data Analysis* | *Data Cleaning & Preparation* | *Grouping & Summarization of Data* | *Initial Analysis* | *Visualization* | *Stakeholder Reporting*

GameCo Project Recommendations

North America:

Focus on *PS4* and *Xbox One* platforms and *Action, Shooter* and **Sports** genres. Focus on games published by the top 10 publishers for North America.

Europe:

Focus on *PS4* and *Xbox One* platforms and *Action, Shooter* and **Sports** genres. Shift PC platform marketing primarily to Europe as it has the most PC sales. Europe has overtaken North America as the top region in proportional sales; therefore, shift marketing towards this region. Focus on games published by the top 10 publishers for Europe.
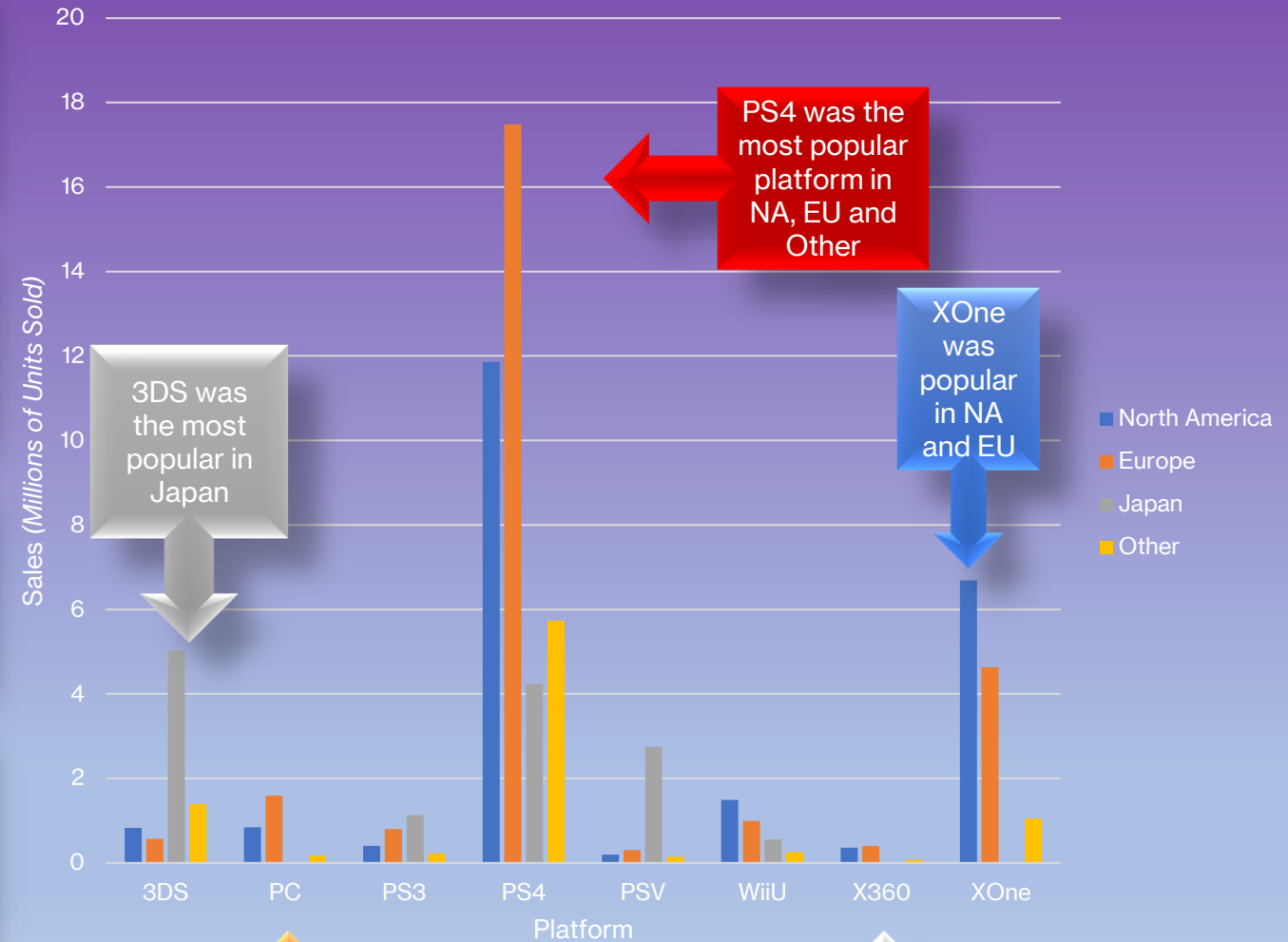
Japan:

Focus on *3DS*, *PS4* and *PSV* platforms and *Role-playing, Action* and *Misc.* genres. **PC and Xbox** is **nearly zero,** reduce marketing for these platforms. Focus on games published by the top 10 publishers for Japan.

Other Regions:

Focus on *PS4, 3DS* and *Xbox One* platforms and *Action, Shooter* and **Sports** genres. Focus on games published by the top 10 publishers for Other. This region's proportional sales have been steadily increasing over time. Consider gathering more data to differentiate these regions as they become a higher proportion of global sales in the future.

Figure 7. Platform Sales Per Region (2016)

PS4 was the most popular platform in NA, EU and Other

XOne was popular in NA and EU

3DS was the most popular in Japan

PC sold best in Europe

X360 and XOne sold poorly in Japan

Legend:
- North America
- Europe
- Japan
- Other

Y-axis: Sales (Millions of Units Sold)
X-axis: Platform (3DS, PC, PS3, PS4, PSV, WiiU, X360, XOne)

Influenza Project Overview

Measure Values

24,517   218,071

In response to the upcoming flu season in the U.S., the aim is to contribute to strategic influenza season planning. This involves pinpointing regions experiencing escalated demand for supplementary staffing on a national scale. The overarching objective is to analyze influenza patterns, leveraging these insights to proactively strategize and optimally allocate staffing resources across the nation.

Project Data:
Original Data Set 1
Original Data Set 2
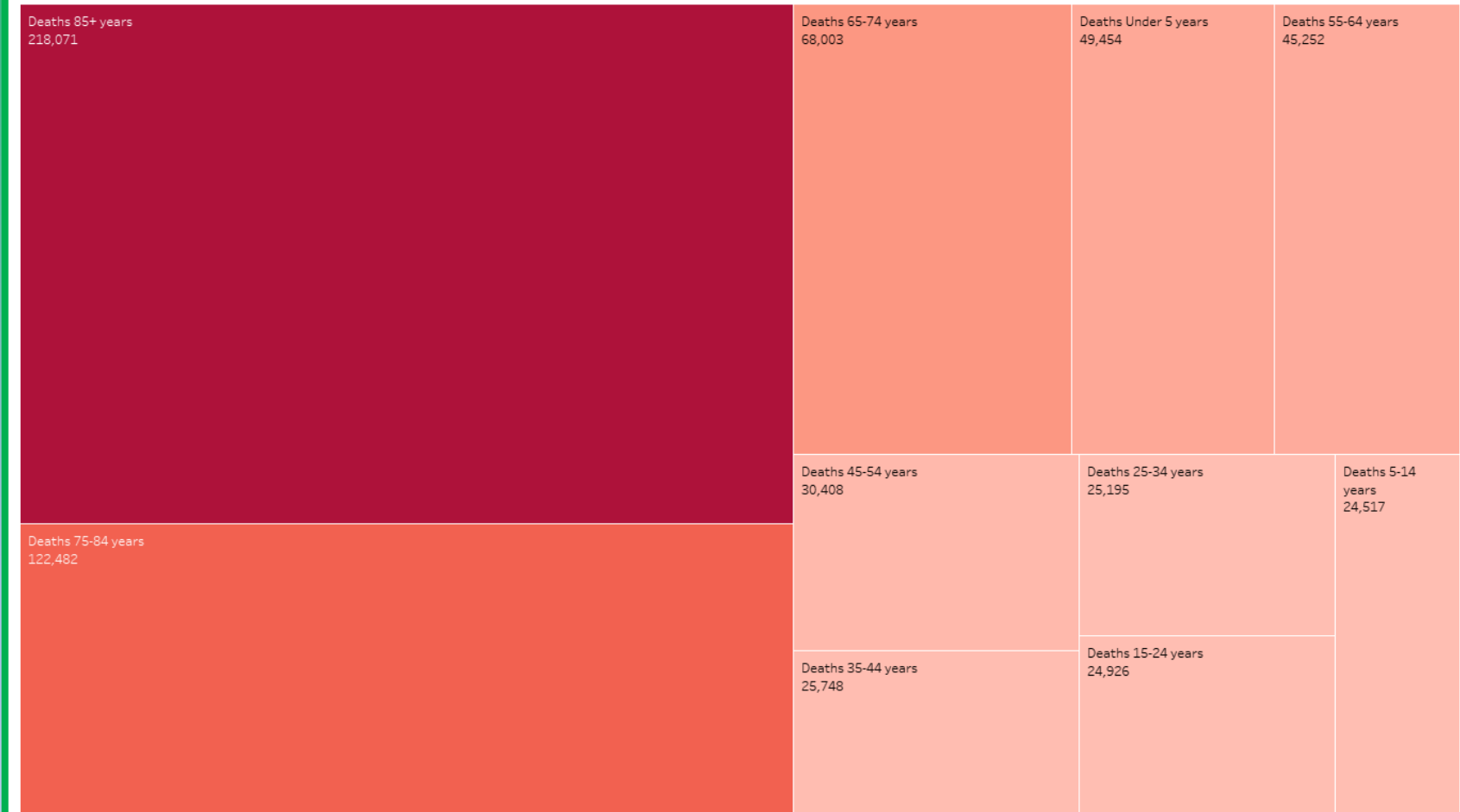Project Brief
Tableau Presentation
Video Presentation

Tools:
Excel
PowerPoint
Tableau

Influenza Deaths by Age Group Treemap (2009-2017) United States

Deaths 85+ years
218,071

Deaths 65-74 years
68,003

Deaths Under 5 years
49,454

Deaths 55-64 years
45,252

Deaths 45-54 years
30,408

Deaths 25-34 years
25,195

Deaths 5-14 years
24,517

Deaths 75-84 years
122,482

Deaths 35-44 years
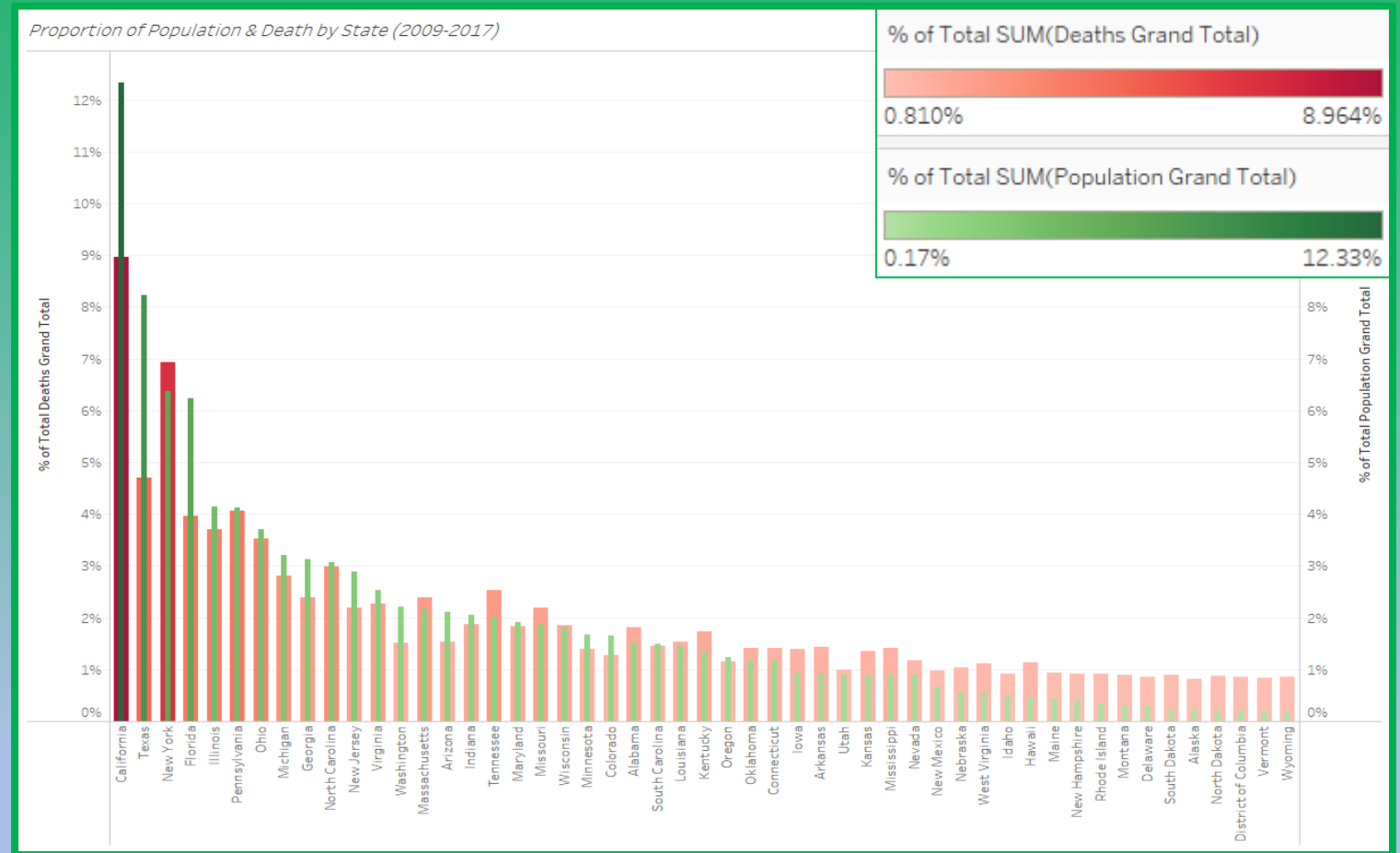25,748

Deaths 15-24 years
24,926

Influenza Project Analysis

Utilized advanced statistical techniques within Excel to perform comprehensive data cleaning, aggregation, and analysis.

Executed inferential analysis by employing hypothesis testing methodologies on pivotal variables.

Extracted strategic insights by leveraging sophisticated visual analysis tools within Tableau, addressing key business inquiries.

Proportion of Population & Death by State (2009-2017)

% of Total SUM(Deaths Grand Total)
0.810% — 8.964%

% of Total SUM(Population Grand Total)
0.17% — 12.33%

Process | Exploratory Data Analysis | Data Cleaning & Profiling | Data Transformation | Hypothesis Testing | Visualization | Tableau Spatial & Temporal Analysis | Dashboard & Storyboard Creation | Stakeholder Reporting

Influenza Project Recommendations

If only utilizing total deaths in our decision, small population states with less resources and high mortality rates will not receive enough help as the larger population states would take the lion's share of medical staffing resources.
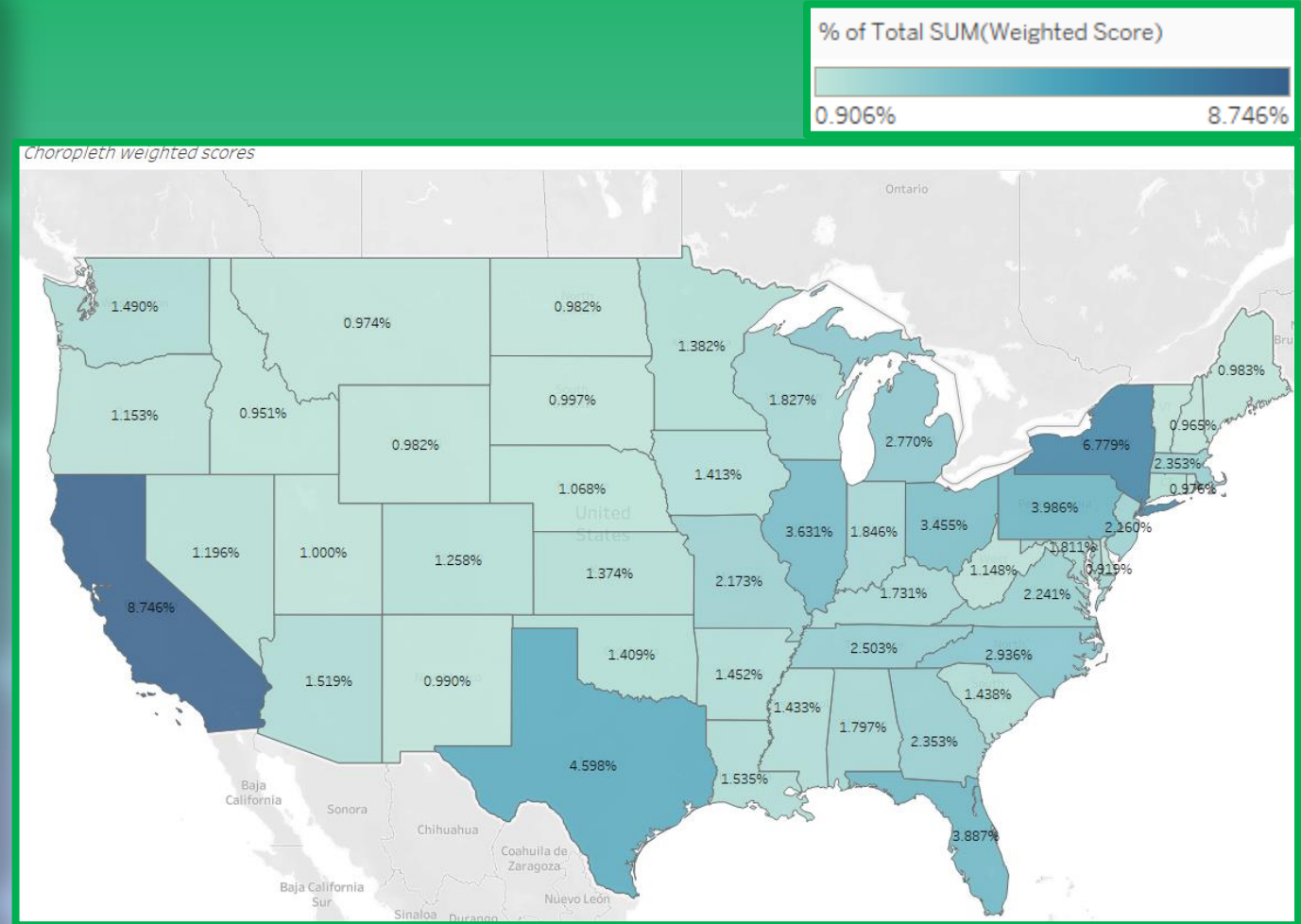
If only utilizing mortality in our decision, populations with higher total deaths would not receive the necessary help required, and the states with high mortality would receive too much help.

Utilizing a weighted score, an informed decision on how much medical staffing resources can be sent to each state that takes into account both total deaths & mortality.

Weighted Score:
*(0.5 * [Death Count Per Capita]) + (0.5 * [Deaths Grand Total])*

If we use a **50/50 weight** for **Mortality** and **Total Deaths**, it means that equal weight is given to the importance of both factors when determining the level of assistance needed for each state.

In this case, both the severity of the influenza outbreak (as reflected by per capita deaths/mortality) and the practical need for resources in larger populations (as reflected by total death count) are considered equally important in the medical staffing resource allocation decision.

% of Total SUM(Weighted Score)

0.906% — 8.746%

Choropleth weighted scores

1.490% 0.974% 0.982% 1.382% 0.983%
1.153% 0.951% 0.997% 1.827% 0.965%
0.982% 2.770% 6.779% 2.353%
1.068% 1.413% 0.376%
3.631% 1.846% 3.455% 3.986% 2.160%
1.196% 1.000% 1.258% 1.374% 1.811% 0.915%
8.746% 2.173% 1.731% 1.148% 2.241%
1.409% 2.503% 2.936%
1.519% 0.990% 1.452% 1.438%
1.433% 1.797% 2.353%
4.598% 1.535%
3.887%

Rockbuster Project Overview

As a data analyst hired by the Business Intelligence department, the objective is to harness Rockbuster's existing movie licenses and leverage data-driven insights to formulate a competitive edge in the online video rental market.

Project Data:
Original Database File
Project Brief
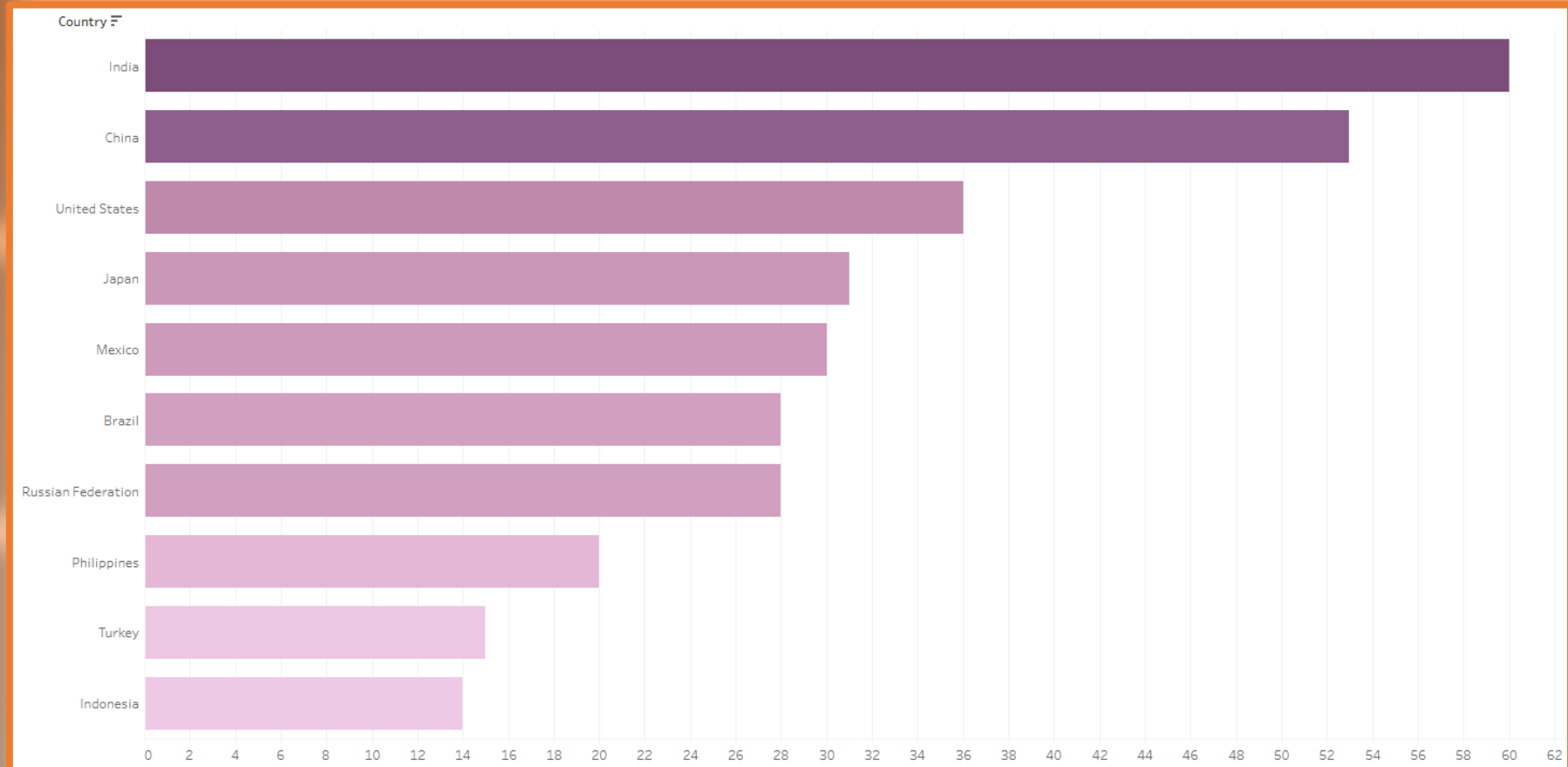SQL Queries
Data Dictionary
Presentation
Data Sheet

Tools:
Excel
PostgreSQL
pgAdmin4
Tableau

Figure 1. Top 10 Countries by Rockbuster Customer Count
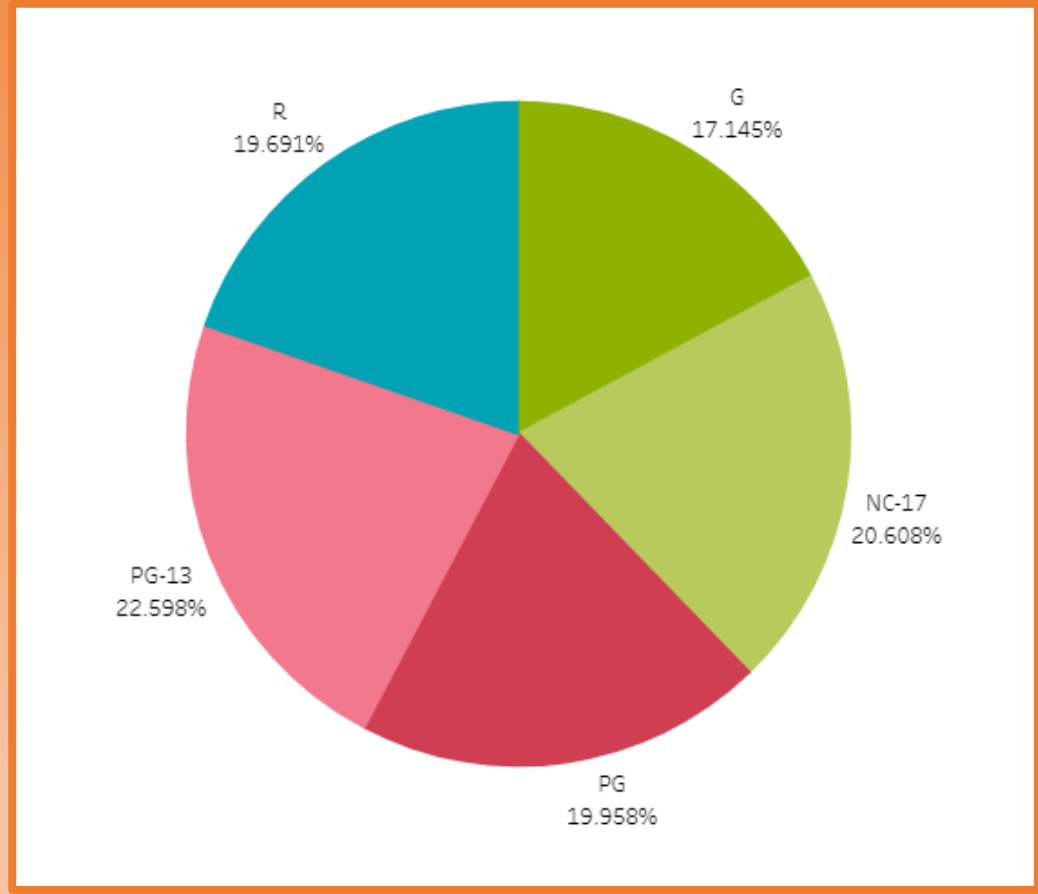
Rockbuster Project Analysis

Collated data from diverse warehouse tables to directly target key business objectives, extracting valuable insights.

Utilized Tableau Public to create visuals based on SQL-generated reports.

Created clear and concise documentation, encompassing a data dictionary and analysis methodology, ensuring transparency and reproducibility in data interpretation.

Figure 2. Percentage of Rockbuster Revenue by MPAA Rating

R
19.691%

G
17.145%

NC-17
20.608%

PG-13
22.598%

PG
19.958%

| Process | Data Dictionary Creation | Exploratory Data Analysis | Cleaning & Summarizing Data in SQL | User Profile Grouping | Deriving Insights using Nested Queries & Common Table Expressions | Visualizing Insights in Tableau | Stakeholder Reporting |

Rockbuster Project Recommendations

```
1   --calculate revenue and rental count by genre
2   WITH RevenueByGenre_CTE AS (
3       --calculate revenue by genre
4       SELECT category.name AS genre, SUM(payment.amount) AS revenue
5       FROM film_category
6       INNER JOIN category ON film_category.category_id = category.category_id
7       INNER JOIN film ON film_category.film_id = film.film_id
8       INNER JOIN inventory ON film.film_id = inventory.inventory_id
9       INNER JOIN rental ON inventory.inventory_id = rental.inventory_id
10      INNER JOIN payment ON rental.rental_id = payment.rental_id
11      GROUP BY category.name
12  ),
13  RentalsByGenre_CTE AS (
14      --calculate rental count by genre
15      SELECT category.name AS genre, COUNT(*) AS rental_count
16      FROM film_category
17      INNER JOIN category ON film_category.category_id = category.category_id
18      INNER JOIN film ON film_category.film_id = film.film_id
19      INNER JOIN inventory ON film.film_id = inventory.inventory_id
20      INNER JOIN rental ON inventory.inventory_id = rental.inventory_id
21      GROUP BY category.name
22  )
23  --join revenue and rental count by genre
24  SELECT RevenueByGenre_CTE.genre, RevenueByGenre_CTE.revenue, RentalsByGenre_CTE.
25  FROM RevenueByGenre_CTE
26  INNER JOIN RentalsByGenre_CTE ON RevenueByGenre_CTE.genre = RentalsByGenre_CTE.
27  ORDER BY RevenueByGenre_CTE.genre;
```

- Focus on obtaining and maintaining Licenses for **Sports**, **Foreign**, **Documentary**, and **Family Genres** as they are the top performing in both **Revenue** and Number of **Rentals.**

- Focus on **PG-13** films as it is the highest performing **MPAA Rating**; however, it is wise to not ignore other ratings as they also perform well, and it is sensible to have a variety of content.

- Focus resources towards **Top 10 Nations by Revenue** and consider shifting resources away from Bottom 10 Nations by Revenue.

- Keep a close eye on nations which have a growing **Customer Count**, as this is closely correlated with revenue.

- Set **prices** close to the average of **$3** per rental.

- Develop a Perks or **Rewards System** for **High Lifetime Value Customers** as they generate much more revenue than the average customer.

Data Output | Messages | Notifications

genre character varying (25)	revenue numeric	rental_count bigint
Action	895.92	228
Animation	944.93	226
Children	826.08	212
Classics	795.17	204
Comedy	774.19	192
Documentary	911.85	239
Drama	742.01	229
Family	987.85	235
Foreign	1008.72	252
Games	862.04	208
Horror	740.25	194
Music	593.36	179
New	812.99	214
Sci-Fi	820.13	201
Sports	976.65	259
Travel	667.20	203

Total rows: 16 of 16   Query complete 00:00:00.051

Instacart Project Overview

As an analyst for Instacart, an online grocery store operating through an app, the goal is to conduct an initial data and exploratory analysis. This project aims to derive insights and recommend strategies for better customer segmentation based on provided criteria.

Project Data:
Original Data Set 1
Original Data Set 2
Original Data Set 3
Original Data Set 4
Project Brief
Python Scripts
Final Report

Tools:
Python
Pandas
Numpy
Matplotlib
Seaborn

Total Spending by Hour of the Day and Price Range
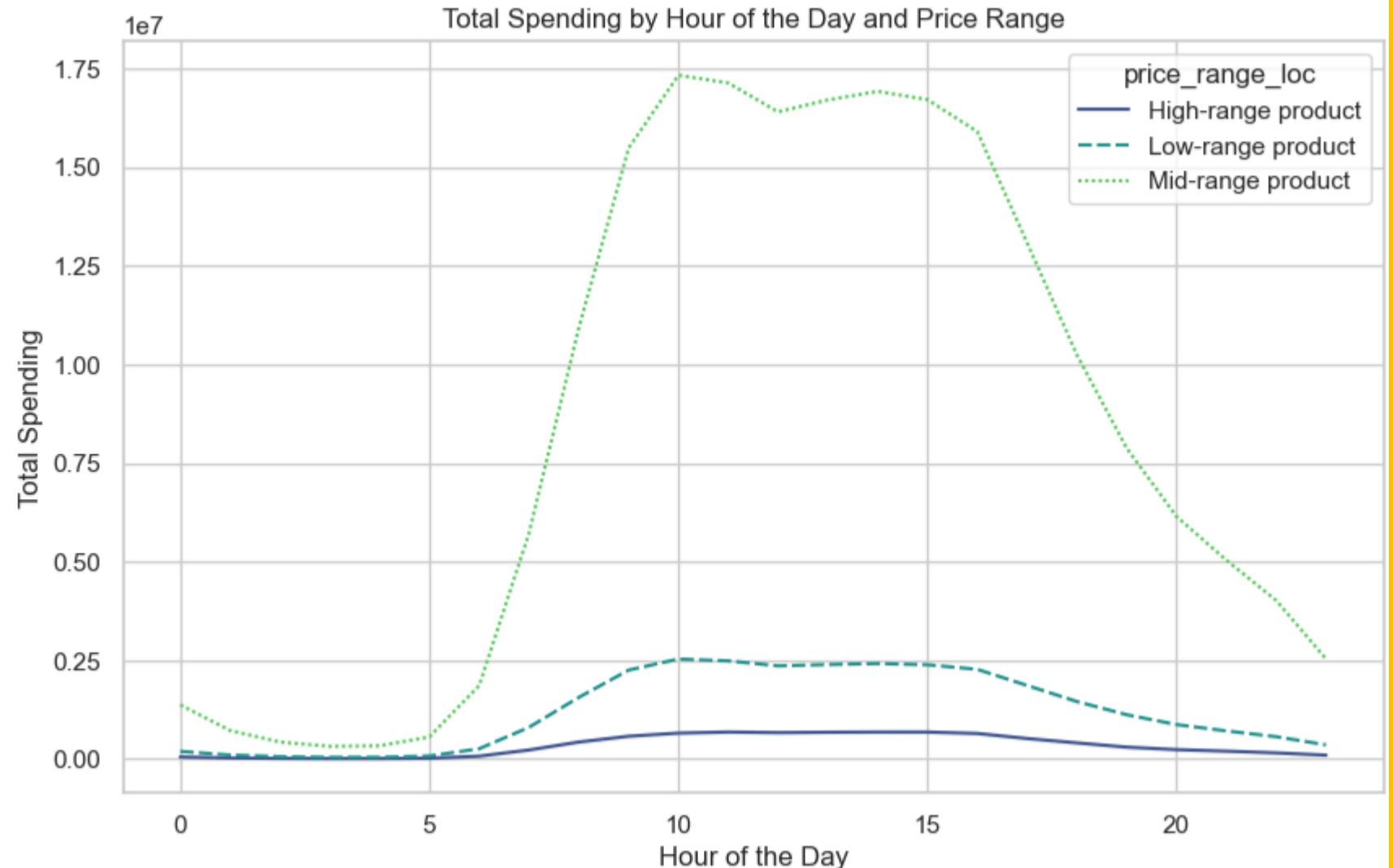
Instacart Project Analysis

Utilized Python to conduct initial analysis, employing sophisticated grouping and aggregating functions to extract crucial business insights.

Employed visual analysis techniques and descriptive statistics to tackle fundamental business queries.

Supplementary insights into the analysis methodology and additional visual representations are available in the Final Report.

Spending Power based on Age and Income
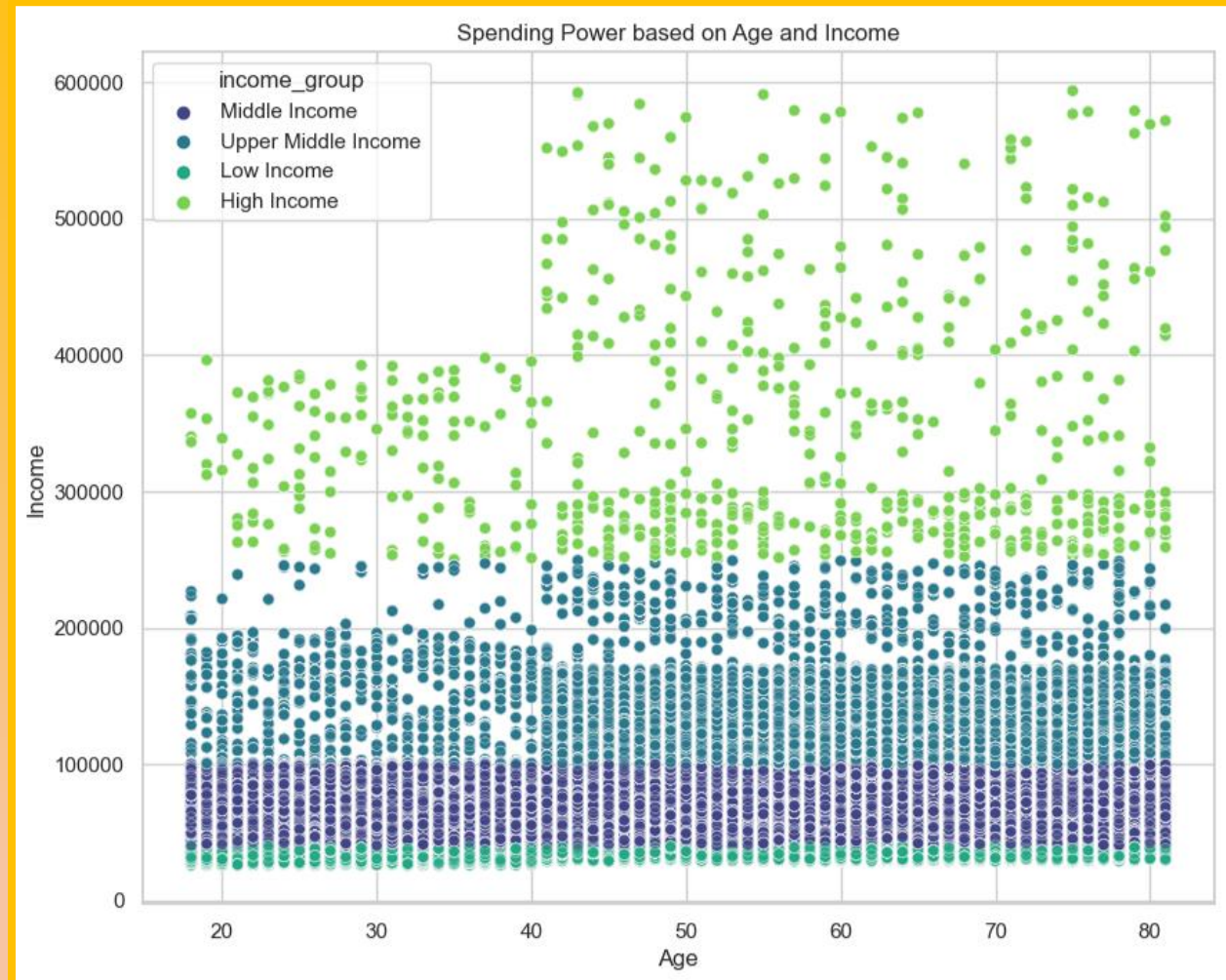
Process → Exploratory Data Analysis → Data Wrangling → Consistency Checks → Concatenation of Tables → Derivation of New Tables → Grouping & Aggregation → Visualization of Insights → Stakeholder Reporting

Instacart Project Recommendations

aligning advertising campaigns with the busiest days and peak order hours, while tailoring promotions to the preferred price range hours, can significantly enhance Instacart's visibility and engagement with its customer base.

The largest segment comprises regular customers, followed by loyal customers, with new customers representing the smallest group.

Given that produce, dairy, and snacks consistently emerge as the most popular departments across all regions, it is advisable for Instacart to focus its advertising efforts on highlighting and promoting products within these categories.

The predominant share of orders is attributed to mid-range products, followed closely by low-range items, while high-range products constitute only a small fraction of total orders.

Given the positive correlation between spending power and age for customers over 40, Instacart should consider tailoring advertising strategies to specifically target this demographic.

Middle-aged adults (ages 35-64) clearly dominate across all family profiles, comprising the largest customer segment. They represent the highest numbers in every category.

Department Distribution by Region
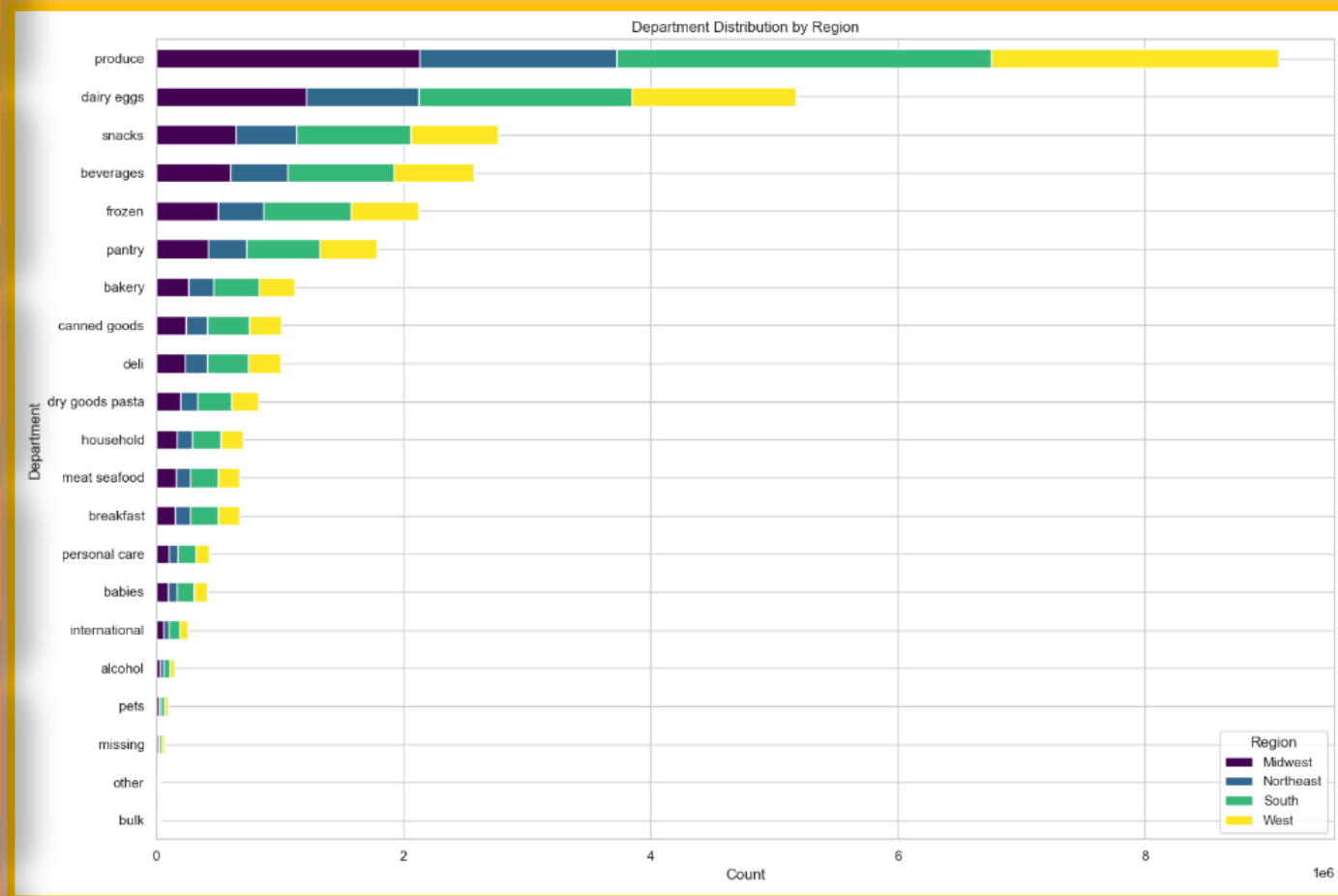
World Happiness Project Overview

Conducted an advanced exploratory analysis of world happiness data using Python. Employing visualizations, geospatial analysis, regression, clustering, and Tableau dashboards and Storyboards to extract valuable insights and trends within countries, regions, and globally.

Project Data:
World Happiness Report 2015-2019 Data
Geospatial JSON Data
Project Brief
Storyboard Plan
Python Scripts
Final Report

Tools:
Python
Pandas
Numpy
Matplotlib
Seaborn
Plotly
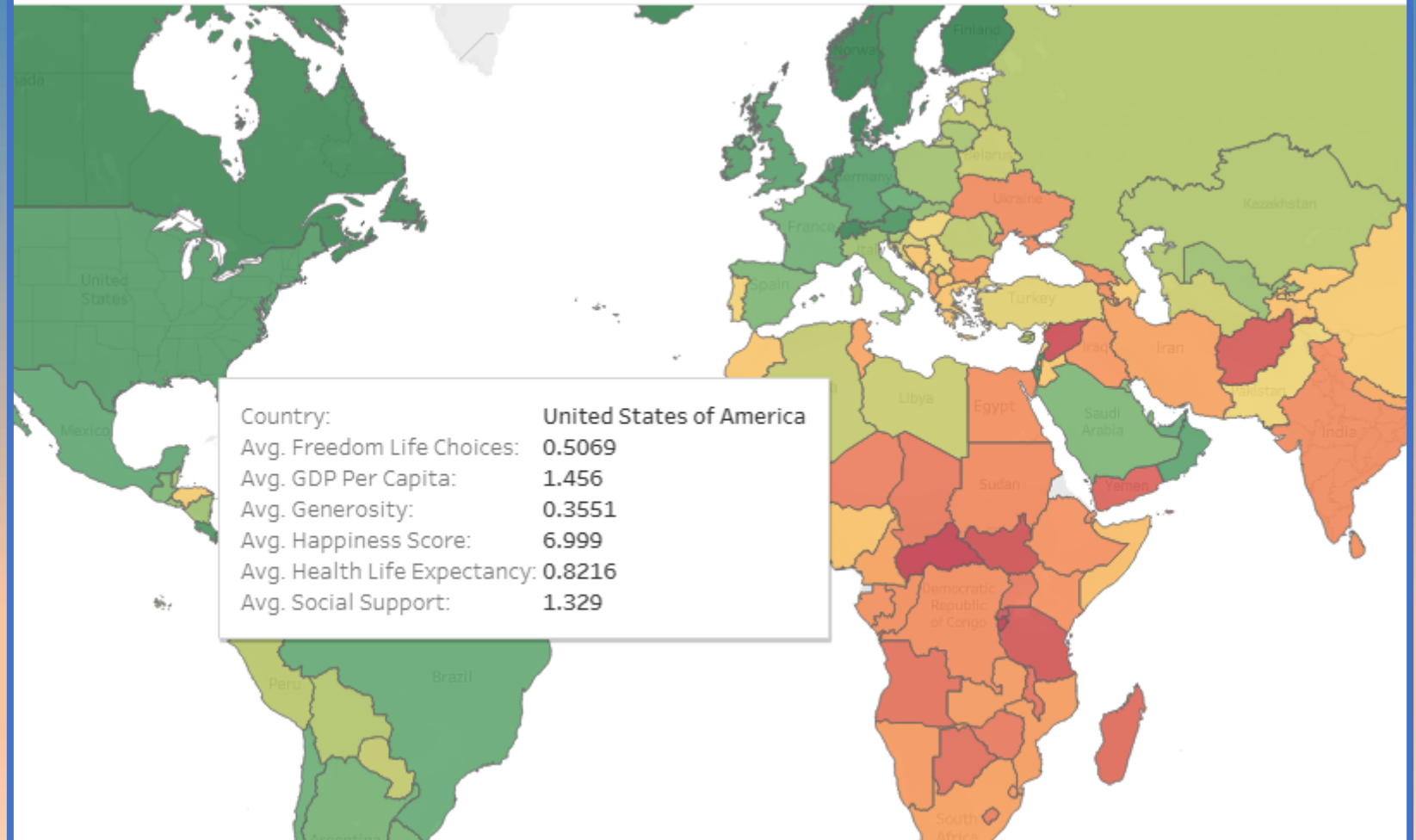Folium

Average Happiness Score by Country (2015-2019)

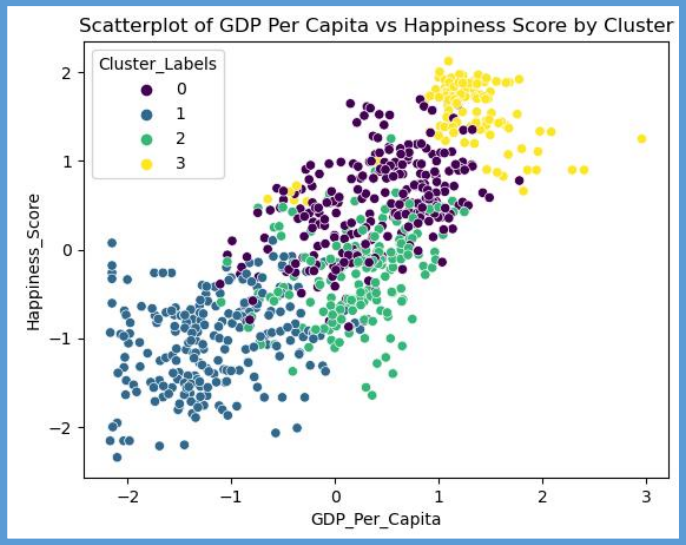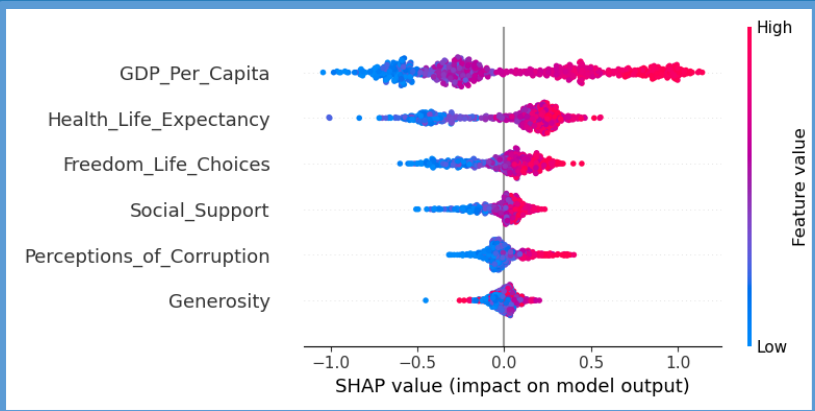| Country:	United States of America
Avg. Freedom Life Choices:	0.5069
Avg. GDP Per Capita:	1.456
Avg. Generosity:	0.3551
Avg. Happiness Score:	6.999
Avg. Health Life Expectancy:	0.8216
Avg. Social Support:	1.329

World Happiness Analysis

Utilized Python for the initial exploratory analysis, incorporating visualizations, geospatial techniques (including JSON file integration), and statistical summaries to gain insights into the world happiness dataset.

Applied regression analysis to discern the influential factors affecting happiness scores, progressing towards a deeper understanding of the dataset's dynamics and geographical variations.

Implemented clustering techniques to unveil patterns and groupings within the data, paving the way for the creation of interactive Tableau dashboards and Storyboards that visually communicate nuanced insights and trends across countries, regions, and the globe.

Scatterplot of GDP Per Capita vs Happiness Score by Cluster

Correlation Matrix of World Happiness Variables

| Process | Exploratory Data Analysis | Data Wrangling | Consistency Checks | Geospatial Analysis | Machine Learning Techniques & Clustering | Visualization of Insights | Data Dashboards | Tableau Storyboard |

World Happiness Recommendations

Conclusion: Based on the analysis encompassing linear regression, clustering, analysis of variable importance, and Shap values, the findings consistently demonstrate that an increase in **GDP per Capita** correlates significantly with heightened levels of happiness.

Recommendation: Advocate for inclusive economic policies aimed at reducing income inequality and creating opportunities for all citizens, fostering economic growth.

Social Support: Strengthen social support systems through community programs and services, promoting emotional and practical assistance to enhance overall well-being
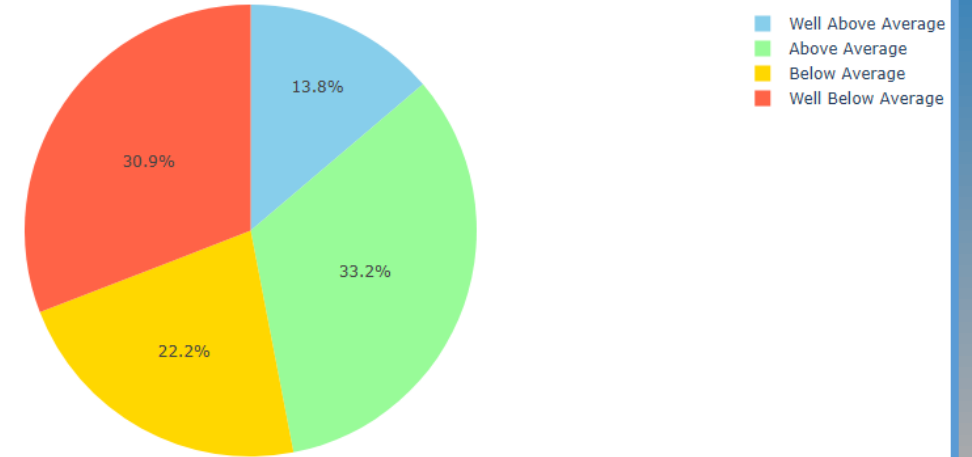
Perceptions of Corruption: Advocate for transparent governance to address perceptions of corruption, fostering trust and confidence among citizens.

Freedom Life Choices: Uphold individual freedoms and choices through policies that respect personal autonomy, allowing citizens to make decisions aligned with their values.
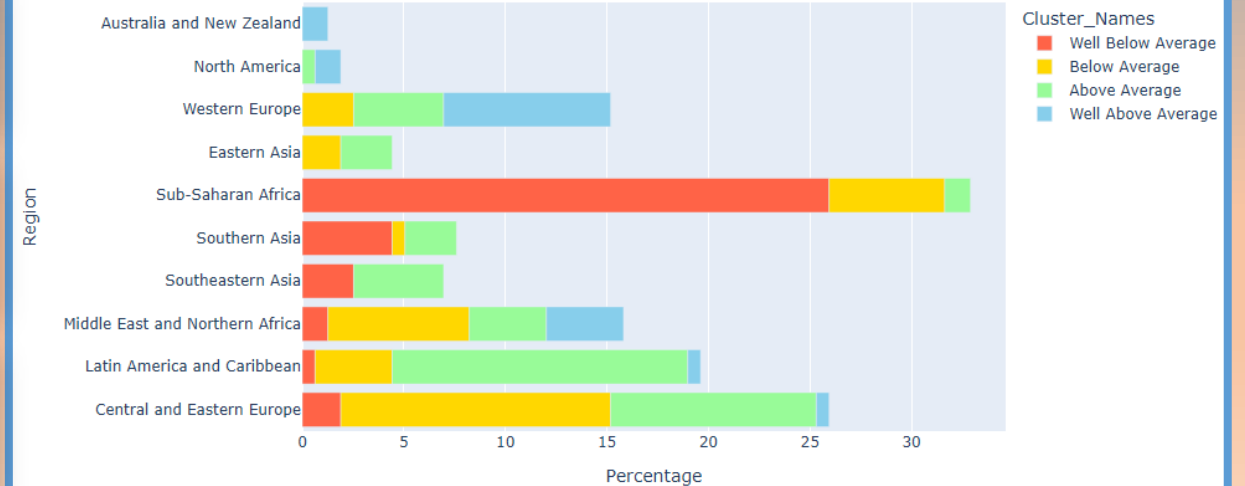
Health Life Expectancy: Ensure access to quality healthcare services for all citizens, investing in healthcare infrastructure and preventive measures to improve health

Generosity: Encourage a culture of generosity by promoting philanthropy and volunteerism, fostering a sense of community and shared responsibility.

Distribution of Clusters by Percentage of Countries

Well Above Average
Above Average
Below Average
Well Below Average

13.8%
30.9%
33.2%
22.2%

Percentage of Countries by Region by Cluster Type 2015-2019

Cluster_Names
- Well Below Average
- Below Average
- Above Average
- Well Above Average

Region:
- Australia and New Zealand
- North America
- Western Europe
- Eastern Asia
- Sub-Saharan Africa
- Southern Asia
- Southeastern Asia
- Middle East and Northern Africa
- Latin America and Caribbean
- Central and Eastern Europe

Percentage: 0, 5, 10, 15, 20, 25, 30

Let's turn data into actionable insights together! 💡