



DATA ANALYST PORTFOLIO

Dominic Bracy

Projects



01

Python &
Tableau
Project on
Global
Youtube
statistics

The Instacart logo, featuring a green leaf icon and the word "instacart" in white lowercase letters on a dark green background.

instacart

02

Python based
project on
consumer
Behavior
trends and
marketing
analysis



03

Business
Analysis of
online video
rental service
using SQL.



04

Developing a
National
Medical staffing
distribution
model using
historical
trends.



05

Comprehensive
analysis of the
global gaming
console
industry in the
marketing
context.

Global YouTube Statistics

Project Goal

The goal of this data analysis project is to comprehensively assess and analyze global YouTube statistics for the year 2023, aiming to identify trends, patterns, and key drivers that can inform strategic decisions and content creation strategies for stakeholders in the digital media industry.

Objective

Conduct a preliminary investigation and examination of the data with the aim of extracting significant insights, subsequently proposing refined segmentation strategies aligned with specified criteria

Skills Applied

- Data Wrangling
- Deriving Variables
- Data Consistency Checks
- Tableau Presentation
- Aggregating Data



The YouTube logo, featuring the word "YouTube" in a bold, sans-serif font, with the "You" in black and "Tube" in white on a red rounded rectangle background. The logo is set against a red background with a halftone pattern and a large white play button icon.

Data Analysis Procedure



DATA CLEANING

Data Cleaning

- Handle missing values
- Remove duplicates
- Handle mixed data types
- Deriving New Columns



pandas

Derive New Columns

- Generate new columns utilizing conditional logic.
- Establish flags and allocate them to new columns.
- Generate summary columns containing descriptive statistics.



seaborn

Visualizations

- Develop regression analysis plots, perform cluster analysis, and examine correlations to visually represent various variables and investigate their interrelationships.



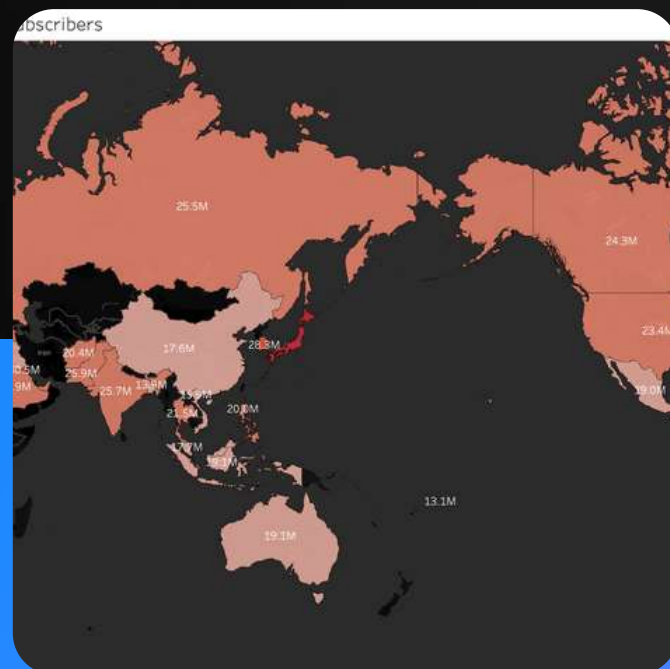
tableau

Present Results

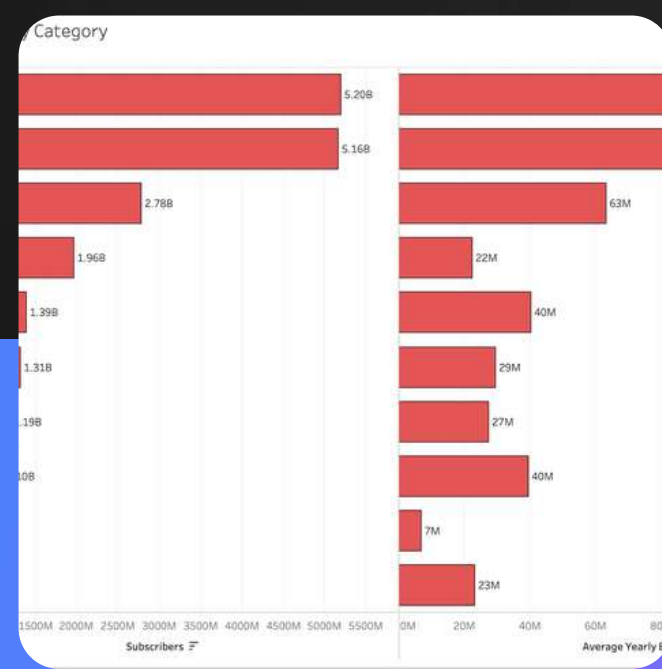
- Generate visual representations and worksheets using Tableau. Compile a conclusive presentation encompassing analytical results and suggestions for Youtube statistics within a Tableau storyboard.



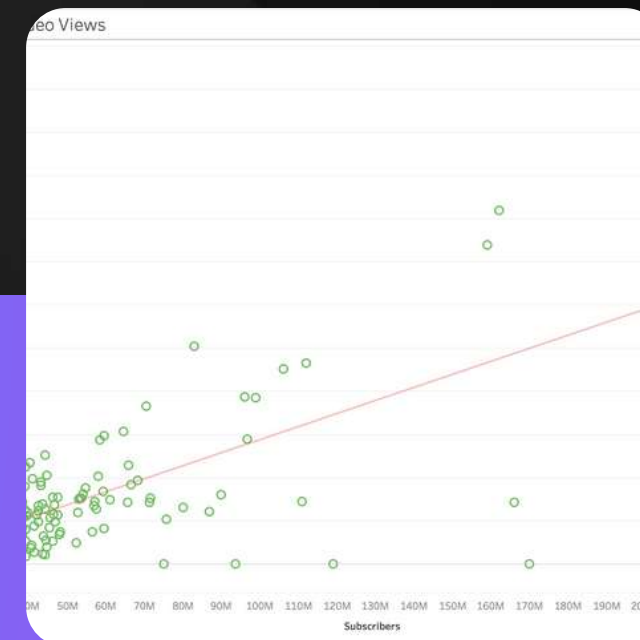
Analysis



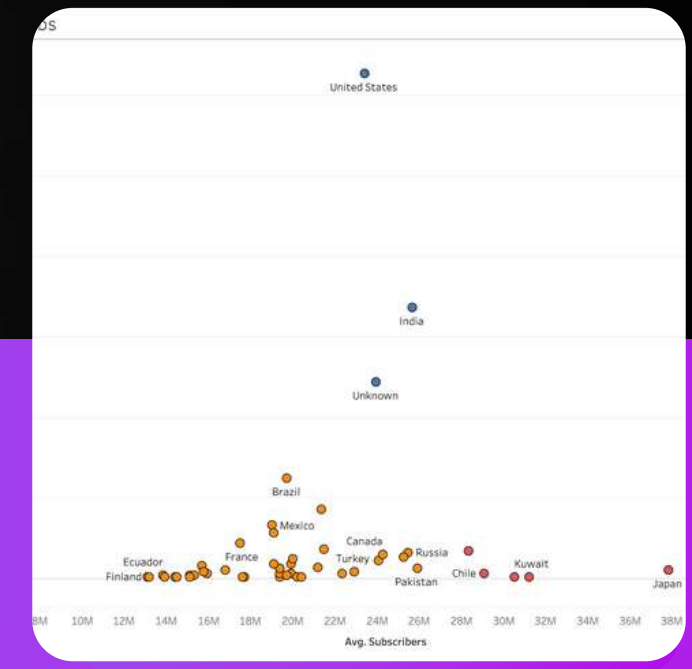
This choropleth map indicates that the US boasts 313 YouTubers, each averaging around 23 million subscribers. In contrast, Cuba, with fewer YouTubers, boasts an average of 46.3 million subscribers owing to a greater share of the subscriber base. This observation suggests that while the path to YouTube stardom may be more challenging in smaller countries, it will lead to a more engaged and potentially lucrative following.



Music & Entertainment is Profitable: The music & entertainment category is the highest revenue-generating category on YouTube. New content creators would be wise to enter this market to maximize earnings.



In the figure, the regression chart makes it clear that there is an interplay between variables, providing insights into their mutual influence. Notably, there exists a positive correlation between subscribers and video views, suggesting that a larger subscriber base is associated with a higher number of video views.



The clustering analysis underscores a compelling phenomenon: smaller nations with fewer YouTubers, such as Cuba, El Salvador, and Barbados, exhibit higher average subscriber counts than their larger counterparts..

Recommendations



Targeting Non-US Audience: As demonstrated in our clustering analysis, smaller countries with fewer YouTubers tend to yield higher average subscriber counts. This observation prompts us to consider untapped potential and opportunities that may await exploration in these regions.



Subscribers & Views Correlation: The positive linear regression relationship between subscribers & video views suggests that channels with more subscribers tend to also have higher view counts. This indicates an engaged and active audience base.



Diversify content: While Music & Entertainment is a high-revenue category, creators and businesses should also consider diversifying content to cater to different audience interests and potentially expand revenue streams.

Challenges & Solution



During the analysis of extensive datasets, it was observed that the performance of the Jupyter Notebook experienced significant slowdown, resulting in an incomplete generation of essential outputs.

Solution - Subsequent investigation and consultation with my tutor revealed that my computer settings were impeding the updating of Anaconda and Python libraries. As a corrective measure, I had to undertake a reinstallation of all the necessary tools to resume and progress with my analysis.



Visualization & Exploration - Analyzing the YouTube dataset presented a challenge, particularly in the generation of comprehensible visual representations. This challenge was notably evident in the creation of choropleth and cluster charts, which tend to appear cluttered.

Solution - To offset the issue, I employed the interactive visualization platform Tableau. This choice streamlined the visualization process, allowing for a more nuanced exploration of various facets within the dataset.



The YouTube dataset presented a challenge in that understanding user behavior with a large dataset required advanced modeling. Subscribers were counted multiple times in a few countries.

Solution - To address this intricacy, a strategic solution was implemented, leveraging clustering methodologies to meticulously scrutinize the relationships among key variables and their impact on each distinct cluster.

Instacart

Project Goal

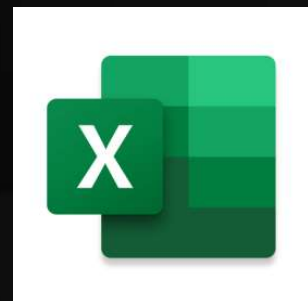
Analyze customer profiles, purchasing behavior, and departmental sales patterns to obtain valuable insights and provide targeted marketing campaign recommendations to the marketing and sales teams.

Objective

Conduct an initial data exploration and analysis to extract valuable insights and propose improved segmentation strategies based on the given criteria.

Skills Applied

- Data Wrangling
- Deriving Variables
- Data Consistency Checks
- Grouping Data
- Aggregating Data



Data Analysis Procedure



Data Cleaning

- Handle missing values
- Remove duplicates
- Handle mixed data types
 - Grouping Data
- Merging data sets



Derive New Columns

- Generate new columns utilizing conditional logic.
- Establish flags and allocate them to new columns.
- Generate summary columns containing descriptive statistics.



Visualizations

- Construct histograms, bar charts, line charts, and scatterplots to visualize different variables and explore relationships between them.

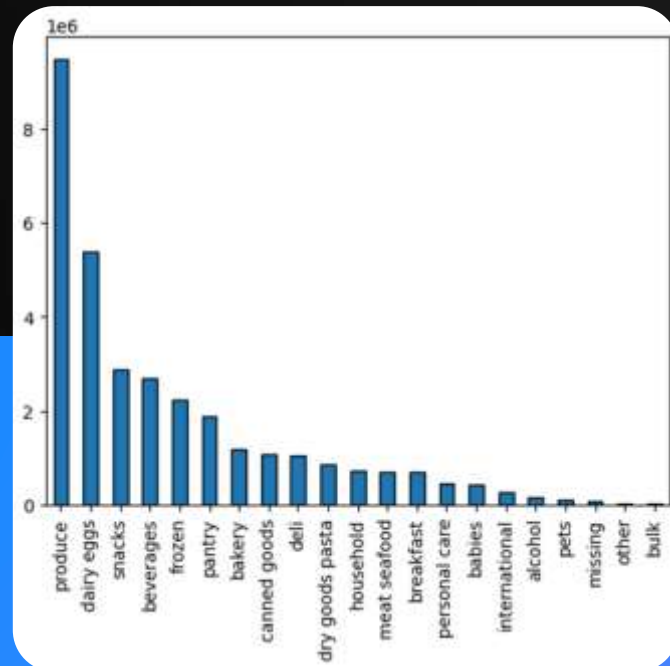


Excel Reporting

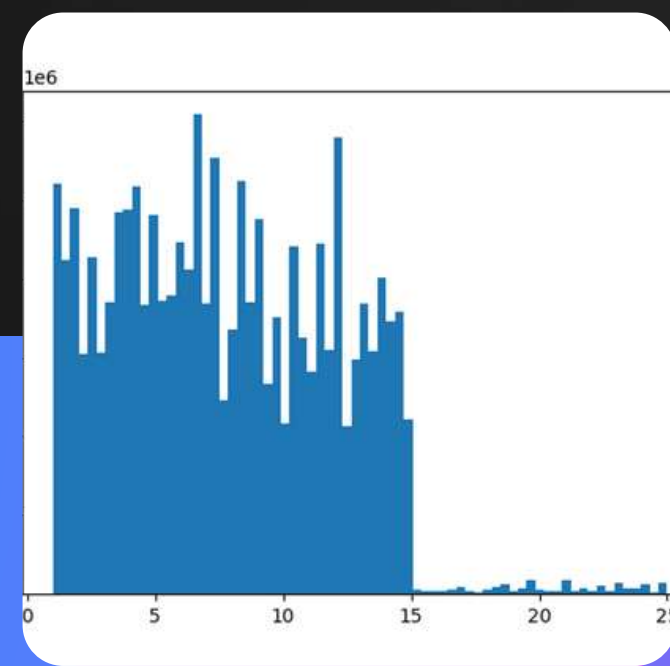
- Summarize analysis findings and elucidate the discovered connections within the data.
- Generate a comprehensive report outlining the analysis methodology, results, visualizations, and recommendations.



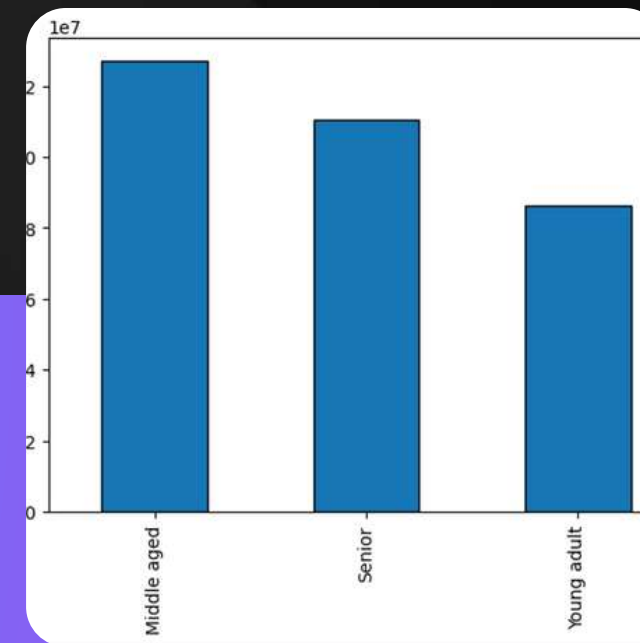
Analysis



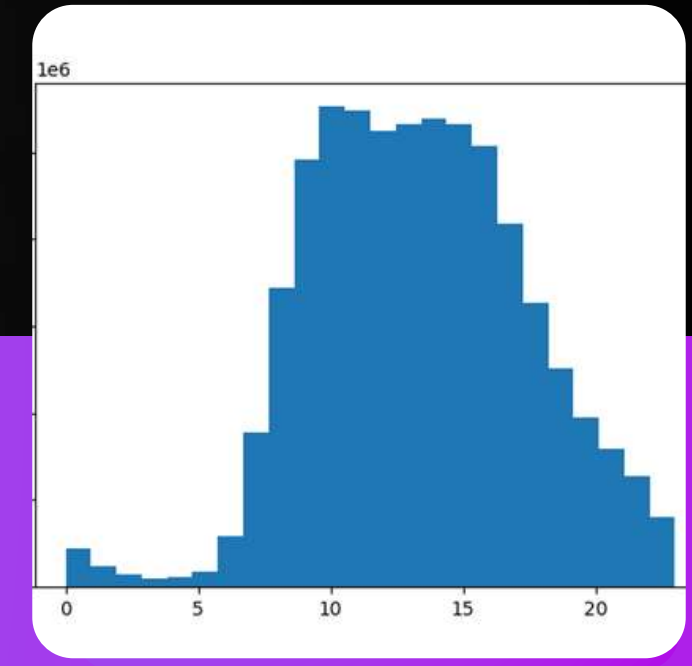
Here we can see the distribution of sales by department. Produce represents the largest department by quite a large distance. Dairy and egg products had the second highest sales with snack and beverages next.



Here we can see the distribution of pricing (under \$25). The vast majority of goods are priced under \$15. With the most common price being approximately \$7.



We have defined three age groups; middle aged is the largest of those groups with people aged between 35 and 60. Elderly people is the second largest group and is over 60s and the under 35s are the smallest group.



Here we can see the distribution of sales by hour throughout the day. The sales follow a clear pattern with a normal distribution. The sales peak between 10am and 3pm before tapering off into the evening. Unsurprisingly the lowest sales are in the early morning hours.

Recommendations



Instacart should prioritize high-performing items like "produce," "dairy/eggs," "snacks," and "beverages" while also seeking to generate interest in "pets," "other," and bulk products. Conducting thorough customer research would help target their needs better.



Instacart can boost revenue from loyal customers by implementing loyalty programs and rewards. Special discounts and incentives will foster deeper connections, enhance satisfaction, and encourage continued patronage, strengthening overall customer loyalty.



High income customers have a higher average price paid than both of our other income groups. We can continue to maximize that market by looking into various ways to reward our high income earners. Maybe a friend spends a certain amount and they get a gift card or credit towards an order.

RockBuster



Project Goal

Assist the business intelligence department in developing a launch strategy for the new online video service. Perform thorough data analysis and provide valuable recommendations to optimize the launch approach.

Objective

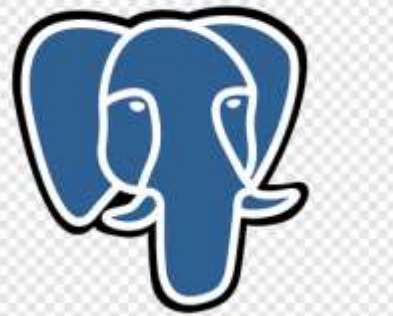
Perform an exploratory analysis of the data, focusing on identifying patterns related to metrics such as the number of customers, average rental cost and other relevant factors.

Skills Applied

- Subqueries
- Filtering Data
- Joining Tables
- Creating a data dictionary
- CTE's



Data Analysis Procedure



PostgreSQL

Understanding Database

- Set up a PostgreSQL database environment.
- Analyze database keys and indexes.
 - Extract an Entity-Relationship Diagram (ERD) and initiate a data dictionary.



Perform Data cleaning in SQL

- Generate new columns utilizing Perform basic CRUD operations and SQL commands.
- Organize, sort, and filter data.
- Identify and clean dirty data.
- Generate a profile of summary statistics (Exploratory Data Analysis – EDA)



Answer Business Questions

- Utilize SQL to join tables, write subqueries, and apply common table expressions (CTEs).
- Address complex business questions using CTEs, subqueries, and joins.

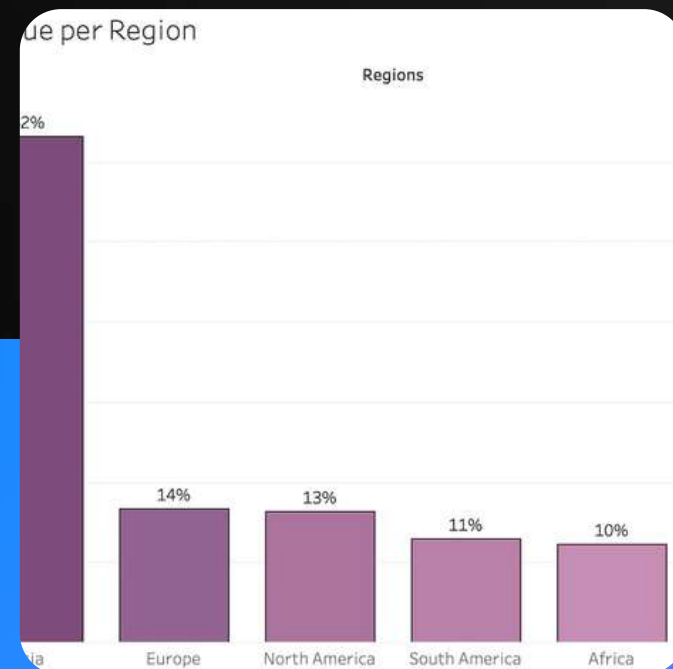


Present Results

- Create visualizations and worksheets in Tableau.
- Prepare a final presentation with analysis results and recommendations for Rockbuster Stealth in Tableau storyboard.



Analysis



Vertical Bar Chart Analysis

Our vertical bar chart illustrates the revenue distribution among six different regions, with the Asia Market leading at 52%. Europe and North America follow as distant second and third, contributing 14% and 13% to the total revenue respectively.



Top 10 Movie analysis

The chart showcasing our top 10 movies by revenue reveals "Telegraph Voyage" leading with \$216 million, followed closely by "Zorro Ark" with \$200 million.



Revenue Per Customer Across Country

Examining revenue per customer in each country unveils an intriguing observation. Despite its relatively lower total revenue, Runion exhibits a higher spending per customer compared to all other countries.



Tree Map Graph Insights

Our tree map graph indicates customer preferences, with a notable inclination towards sports and sci-fi genres. Sports and sci-fi movies contribute significantly to our revenue, accounting for the highest shares of the revenue at 1.9% and 10.6% respectively.

Recommendations



Establish a program to reward our most loyal customers with the highest lifetime value. A few recommendations are discounts and referral programs. Continue to push out and capitalize on our most like movie genres like sports & Thriller titles.



Identify possible reasons why asia is so dominate in the market share. Is marketing an issue or not enough stores available in each region. If marketing is an issue we need to focus our efforts on rolling out promotions in our weak selling regions and potentially look to stengthen our visibility in terms of opening new stores.



Prioritize titles with the most contributions to our revenue and consider scaling back the titles that have been poor amongst our viewers.

Adjust Rental prices for short term and long term rentals to encourage our customers to rent with us more frequently.

Influenza Season

Project Goal

Develop a strategy to allocate additional medical staff to US hospitals during the upcoming influenza season, considering the optimal timing and specific locations that would require the most support.

Objective

Identify states requiring additional medical staff for the 2018 influenza season. Examine characteristics and locations of vulnerable populations.

Skills Applied

- Data Quality Checks
- Data Transformation
- Data Integration
- Tableau storytelling
- Statistical hypothesis
- Data Grouping & summarizing



INFLUE

Data Analysis Procedure



Designing Research Project

- Formulating research questions and hypotheses to address business requirements.
- Sourcing relevant data for the project.
- Creating a project management plan for effective execution.



Perform Data cleaning

- Verifying integrity and ensuring data quality.
- Cleaning the data by removing duplicates and handling missing values.
- Transforming the data and integrating multiple datasets.
- Deriving new variables from the existing data.



Statistical Testing

- Computing variance and data spread (variance, standard deviation, identifying outliers) for the two variables under consideration.
- Calculating the correlation between the two variables.
- Performing a two-sample t-test to test the hypothesis.

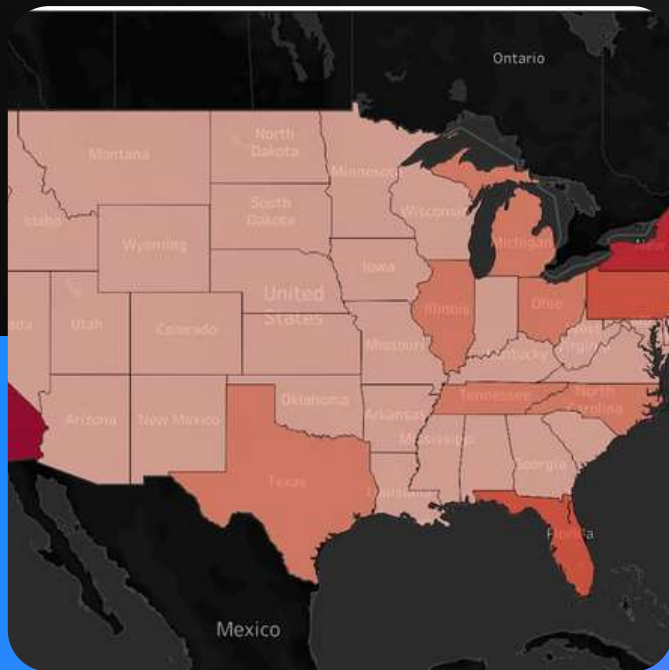


Present Results

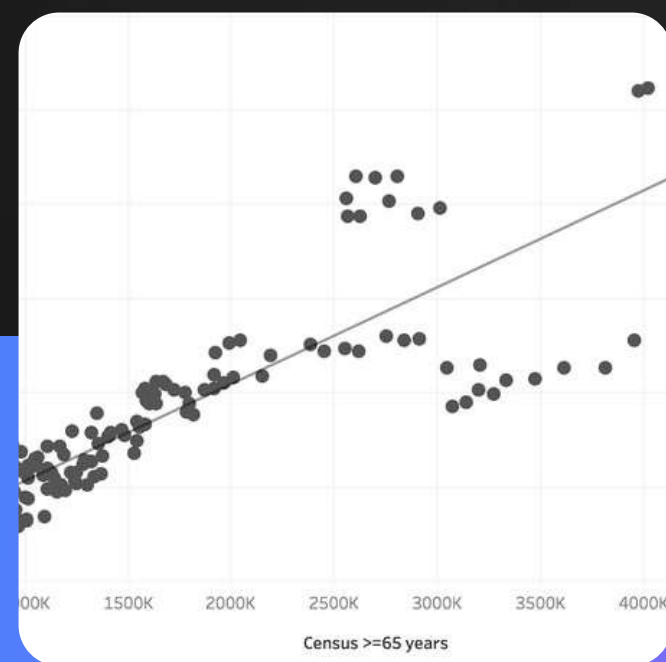
- Generating composition and comparison charts.
- Creating temporal visualizations and conducting forecasting.
- Utilizing statistical visualizations such as box plots, histograms, bubble charts, and scatter plots.



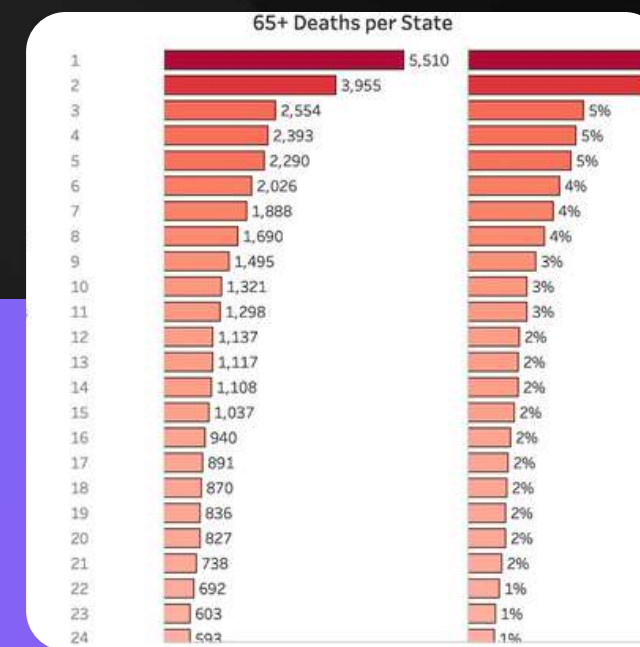
Analysis



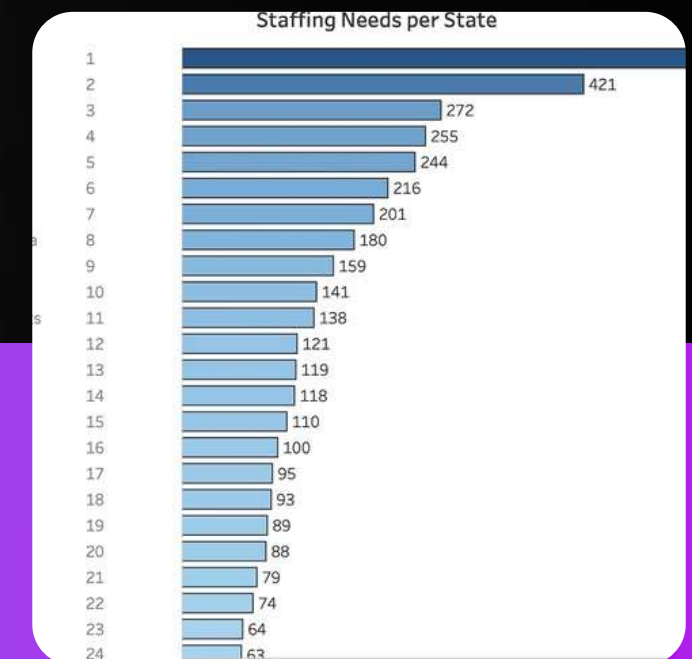
In this spatial map our darker shaded states contain the most deaths of the higher risk age group. California, New York and Texas were our top 3 states.



Here we can see the relationship between the census and deaths of our at risk group which was our 65 and older population.



In this dual bar graph we have the states ranking, the total deaths and % of deaths of each of our state.



In this bar graph we created a parameter with each states % deaths and their ranking adding up to 100%. With that we are able to distribute staff according to the rank and what states need the additional staff.

Recommendations



Using our parameters range we can successfully distribute our staff that are available to the states depending on their rank and percentage.



Ensure that enough staff is available to the states with highest totals of influenza deaths – California, Texas, New York, Pennsylvania, and Florida.



Sending Extra staff to the Top 5 states during our summer months before Influenza season begins from December to March.

GameCo

Project Goal

Gain insights into GameCo's performance in the gaming market

Objective

Examination of global and regional trends in video game sales.

Analyze popular video game genres and platforms and identify regional markets experiencing rapid growth.

Skills Applied

- Data Integration
- Data transformation
- Descriptive analysis
- Data Sorting and filtering
- Data Cleaning



GAMECO

Data Analysis Procedure



Prepping our data

- Removing duplicates.
- Eliminating erroneous data values.
- Imputing missing values.
- Correcting inconsistencies in formatting.



Data Transformation

- Grouping and summarizing data.
- Using pivot tables to gain general data insights.
- Applying filters to examine specific segments.



Analysis

- Analyzing basic statistical features: mean, median, mode.
- Studying data distribution and skewness.
- Identifying outliers.



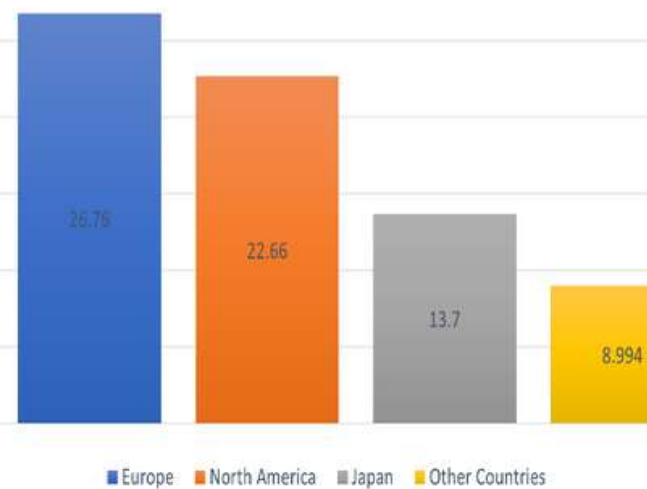
Present Results

- Conducting an analysis of market trends based on key objectives.
 - Creating visualizations.
- Consolidating project deliverables, including reflections and presentations.



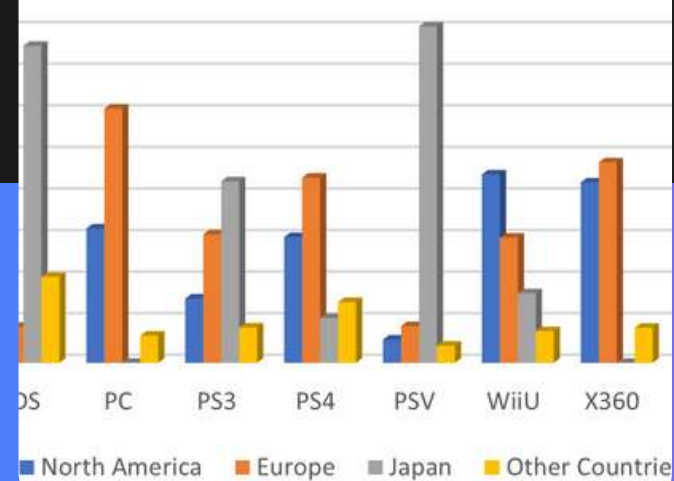
Analysis

2016 Total Sales by Region



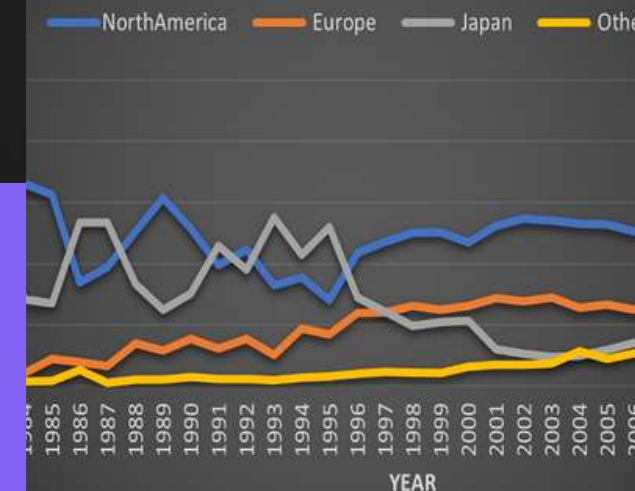
This bar graph shows us the total sales by each of our 4 regions. Europe is our highest selling market with north America a very close second.

Platform Regional Sales Distribution



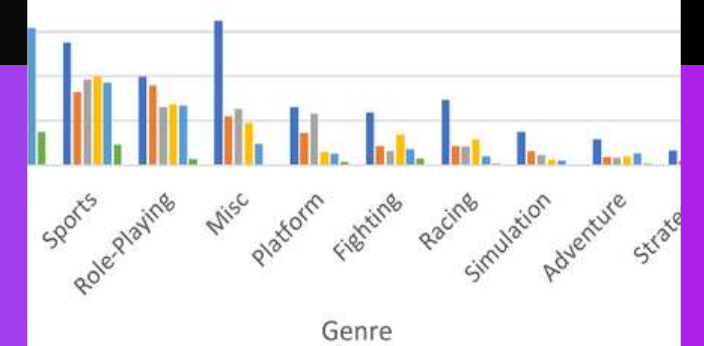
The multi dimension bar graph illustrates our regional sales by gaming platforms. Japan dominates our 3DS and psv platforms.

% of Global sales by Region



The line graph illustrates Europe's continuous rise to attain the top position in regional sales percentage over the years.

Genre Favorites 5 year Outlook North America



The graph details for us the 5 year outlook of north American sales by Genres. There are decline in sales in each of our genres signaling a problem for the US market.

Recommendations



As overall Market Percentage shows constant declines since 2008 we must conduct a heavy analysis to establish the underlying reasons for this Fall and pursue actions to recover.



For 2017 allocate the highest % to Europe. They have been in a steady incline and have surpassed North America. Allocate more of the budget for Action, Shooter, and sports games for North America, Europe and Other countries.



ØFor North America Exclusively we need to conduct an analysis understand the reasons the dominance is no in the Gaming market.



THANK YOU

For watching this presentation

Any Questions?



dbracy3@gmail.com



<https://github.com/dbracy1127>



<https://public.tableau.com/app/profile/dominic.bracy>