



GRACE SKELLEY

DATA ANALYST PORTFOLIO



Hi I'm Grace Skelley

Welcome to my portfolio!



"I am a professional business intelligence analyst with experience in the Medical and Healthcare industry.

My goal is to have a sustainable impact on the organization's success with the use of my extensive research, compelling visualizations, and project coordination/management abilities. I aim to make a difference by leveraging my accumulated analytical skill sets in discovering innovative solutions, through a data-driven approach fueled by creativity, and critical thinking.

My professional experience in business operations in multiple departments – big and small organization settings has deepened my understanding of the importance of well-organized, competent workflow procedures influenced by data-rich decisions. Therefore, my desire and determination to bring positive changes are what pushed me to evolve my career map."

A handwritten signature in black ink that reads "Grace Skelley". The signature is fluid and cursive, with a distinct "G" at the beginning.

[VIEW MY PROJECTS HERE](#)

PROJECT EXCERPTS

- **INSTACART**
Python-based project on Consumer Behavior Trends and Marketing Analysis
- **ROCKBUSTER STEALTH LLC**
International Business Analysis of Online Video Rental Services through SQL
- **PREPARING FOR INFLUENZA SEASON IN THE US**
A National Medical Staffing Distribution based on Historical Trends
- **GAMECO**
Global Marketing Analysis of the Gaming Console Industry
- **PIG E. BANK**
A Predictive Analysis of customer retention risk for Global Finance Service Company
- **THE WORLD'S LARGEST PUBLIC COMPANIES**
Python-based project on Global Industry Performances



EXCERPT

INSTACART

A successful online grocery store that operates through an app. Executives are curious to learn about different customer segments and understand their purchasing behavioral patterns for applicable marketing campaigns.

NOTE: Instacart is a real company that's made its data publicly available online. However, the content of this project has been fabricated for the purposes of this Achievement.

OBJECTIVE

Produce an exploratory analysis of consumer behavior and sales patterns in sustaining information that will benefit sales and marketing departments in developing tactical promotional efforts.

PROJECT & DATA

- Project Brief
- Customer Data Set, Customer Orders, Products, and Department | Modified Open Source from [Instacart](#)
- Data Dictionary | Link provided by CareerFoundry

LIMITATIONS

- Data only contains records from 2017.
- Customer demographics are limited to income, age, family size, and marital status.

TECHNIQUES APPLIED

- Data Cleaning: Wrangling and Subsetting
- Data Consistency Checks
- Combining and Exporting Data
- Deriving New Variables
- Grouping Data and Aggregating Variables
- Python Visualization and Excel Report

TOOLS



DATA METHODOLOGY



ORGANIZING DATA

Tracking data irregularities like mixed-type variables, missing and/or duplicate values for ratification. Methodically noting steps trailing the transformation of the dataframe will allow other users to identify procedural actions taken.

DERIVING VARIABLES

Using conditional logic in the form of if-statements, user-defined functions, the loc() function, and for-loops in deriving new columns.

AGGREGATING DATA

Creating flags and placing new columns for a summary of descriptive analysis with groupby() function.

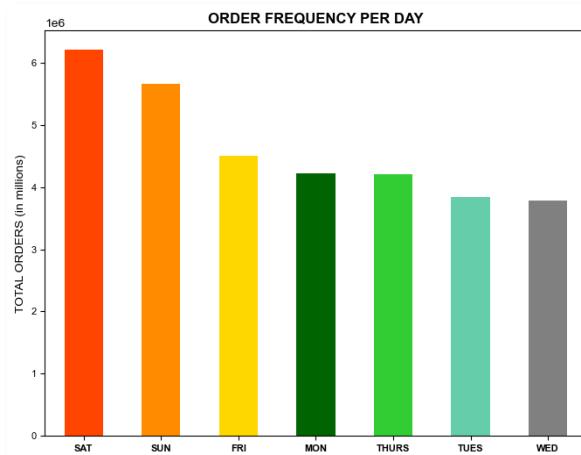
PYTHON VISUALIZATION & EXCEL REPORT

A merged dataframe is utilized to generate a compelling visualization addressing the stakeholders' questions.

BUSINESS ANALYSIS



Saturday, Sunday, and Friday are the **busiest** days for shoppers.

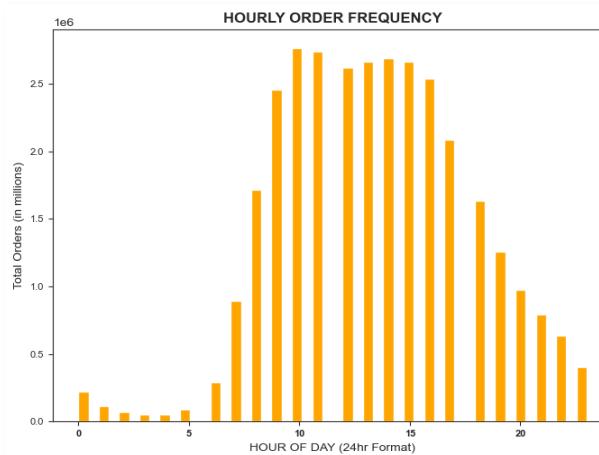


```
bar_chart = final_df['order_day_of_week'].value_counts().plot.bar(color = ['orangered', 'darkgreen', 'blue', 'red', 'purple', 'brown', 'teal'], figsize = (8,6))

# Adding Chart Labels
sns.set_style('ticks')
bar_chart.set_xticklabels(['SAT', 'SUN', 'FRI', 'MON', 'THURS', 'TUES', 'WED'])
plt.ylabel('TOTAL ORDERS (in millions)')
plt.xticks(fontsize = 8, fontweight = 'bold')
plt.yticks(fontsize = 8)
plt.title('ORDER FREQUENCY PER DAY', fontweight = 'bold')
plt.show()
```

FIG. 4a

Purchasing frequency **increases** between 9 am through 3 pm.



```
hist2 = final_df['order_hour_of_day'].plot.hist(bins = 50, color = 'orange', figsize = (8,6))

# Adding Chart Labels
sns.set_style('ticks')
plt.ylabel('Total Orders (in millions)')
plt.xlabel('HOUR OF DAY (24hr Format)')
plt.xticks(fontsize = 8, fontweight = 'bold')
plt.yticks(fontsize = 8)
plt.title('HOURLY ORDER FREQUENCY', fontweight = 'bold')
plt.show()
```

FIG. 4b

There are **21 departments** that offer **multiple price** ranges.

department	HIGH	LOW	MID
alcohol	-	33,046	111,581
babies	-	121,484	288,908
bakery	-	274,986	845,842
beverages	-	814,697	1,757,204
breakfast	-	209,185	461,665
bulk	-	1,181	32,270
canned goods	-	281,711	730,363
dairy eggs	4,877	1,370,908	3,801,397
deli	-	299,220	704,614
dry goods pasta	-	284,346	537,790
frozen	-	647,617	1,474,114
household	-	245,070	454,787
international	-	73,203	182,788
meat seafood	392,855	-	281,926
missing	-	18,516	46,252
other	-	15,259	19,152
pantry	221	504,521	1,277,963
personal care	-	123,874	300,432
pets	-	28,165	64,895
produce	-	2,585,708	6,493,565
snacks	-	1,742,143	1,024,263

#16 Department Price Range and Check Output

```
dept_prc_rnge = pd.crosstab(df_cstmr_prfl['department'], df_cstmr_prfl['price_range'])

dept_prc_rnge
```

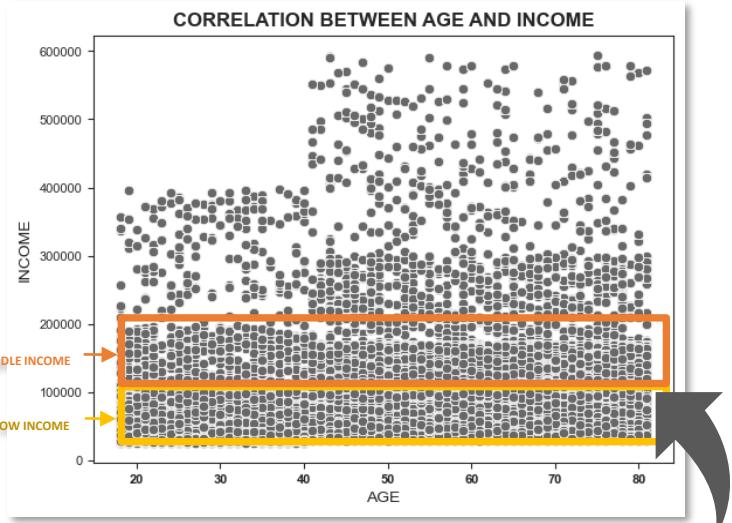
FIG. 4c

VIEW FULL REPORTS BELOW

CONSUMER SEGMENT ANALYSIS



The customers' income ranges between **low to mid-range** describing probable spending power.

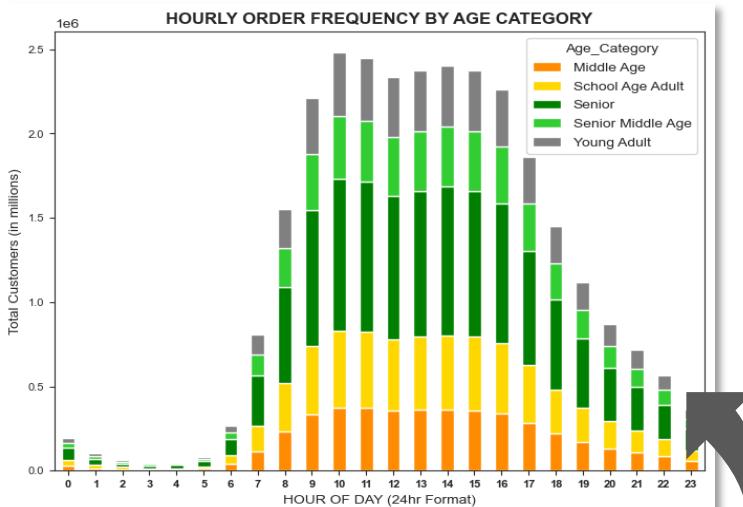


```
# Creating Scatterplot2 for spending power by age group
scatterplot2 = sns.scatterplot(x = 'Age', y = 'Income', data = final_df)

# Adding Chart Labels
sns.set_style('ticks')
plt.ylabel('INCOME')
plt.xlabel('AGE')
plt.xticks(fontsize = 8, fontweight = 'bold')
plt.yticks(fontsize = 8)
plt.title('CORRELATION BETWEEN AGE AND INCOME', fontweight = 'bold')
plt.show()
```

FIG. 4d

Seniors appear to be **frequent shoppers**.

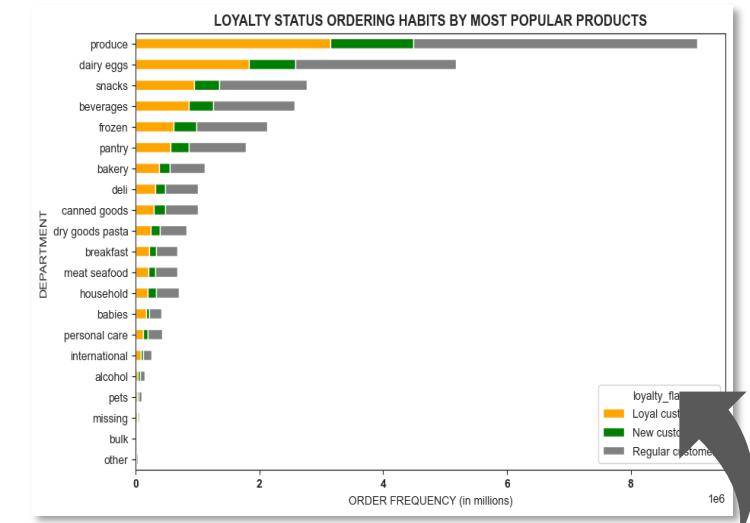


```
#2a Create a Stacked Bar for Age Category Profile of Shopping habits (Order Hour of Day)
stckbr_age_shppngtm_hbts = Age_shppngtm.plot(kind = 'bar', stacked = True, color = ['darkgreen', 'green', 'yellow', 'orange', 'grey'])

# Adding Chart Labels
sns.set_style('ticks')
plt.ylabel('Total Customers (in millions)')
plt.xlabel('HOUR OF DAY (24hr Format)')
plt.xticks(fontsize = 8, fontweight = 'bold')
plt.yticks(fontsize = 8)
plt.title('HOURLY ORDER FREQUENCY BY AGE CATEGORY', fontweight = 'bold')
plt.show()
```

FIG. 4e

Produce, dairy eggs, and snacks are the most **profitable** goods.



```
#11a Create a Bar Chart for department distribution by Loyalty Status
br_dept_lylty = dept_lylty.sort_values(by = 'Loyal customer', ascending = True).plot(kind = 'bar')

# Adding Chart Labels
sns.set_style('ticks')
plt.ylabel('DEPARTMENT')
plt.xlabel('ORDER FREQUENCY (in millions)')
plt.xticks(fontsize = 10, fontweight = 'bold')
plt.yticks(fontsize = 10)
plt.title('LOYALTY STATUS ORDERING HABITS BY MOST POPULAR PRODUCTS', fontweight = 'bold')
plt.show()
```

FIG. 4f

[VIEW FULL REPORTS BELOW](#)

RECOMMENDATIONS

CONSUMER TREND

ADVERTISEMENT

Product endorsements of the least popular products like international, pantry, etc. should be promoted on weekends (Fridays, Saturdays, and Sundays) between 9 am and 3 pm, leveraging fast-moving goods such as dairy eggs, beverages, and beverages so on.

COMPETITOR ANALYSIS

Looking into competitors' pricing schemes and conducting surveys on consumers' household expenditures can disclose key information for effective pricing tactics.

CUSTOMER SEGMENT

Offering special rates and rewards through a point system tailored to members' interests maintains loyal buyers and converts non-frequent buyers' status to a higher level.

RESEARCH & DEVELOPMENT

Expand the services to new target audiences that may be interested in the premium brands and an exclusive offer on executive packages can spark new interests and eventually extend business in that area.



[VIEW FULL REPORTS BELOW](#)



Python Codes



Final Excel Report



EXCERPT

ROCKBUSTER STEALTH LLC

A fictitious rental movie company that had storefronts across the world, is facing tough competition from online streaming services like Netflix and Amazon Prime. To remain competitive, the management plans to launch an online rental service.

OBJECTIVE

Provide insights on the current business standings to support the company's 2020 marketing campaign for a new product launch.

PROJECT & DATA

- Project Brief
- This mock dataset is provided by CareerFoundry and can be downloaded [here](#).

LIMITATIONS

- Data covers internal records of stores, customers, payments, inventory, films, and more.

TECHNIQUES APPLIED

- Relational Databases
- Entity Relationship Diagram (ERD)
- Data Dictionary
- Database Querying
- Filtering Data
- Data Cleaning and Summarizing
- Joining Tables
- Subqueries
- Common Table Expressions

TOOLS



DATA METHODOLOGY



ASSESSING DATABASE

Initial review and crafting of an Entity Relationship Diagram (ERD) through DbVisualizer describes the structure of the database for analysis and capturing it through a data dictionary for user accessibility.

DATA CLEANING

Performing CRUD functions ensures a clean and consistent format for filtering as well as summarizing data for output accuracy.

DATA EXTRACTION & SUMMARIZATION

Extracting records using Joining Tables, Subqueries, and Common Table Expressions (CTEs) commands allow an extensive comprehension of the current business standings.

DATA VISUALIZATION & STORYTELLING

SQL results are converted to CSV files to generate visualization with ease for the final presentation and tableau storyboard.

BUSINESS PROFILE OVERVIEW



The business is accessible internationally and presently **servicing 108 countries**.

OFFERING WIDE SELECTIONS in ENGLISH:

- 5 Film ratings
- 16 Genres

AUSTRALIA & CANADA BRANCHES ARE FULLY OPERATING:

- 2 Staff
- 4581 Inventory
- 16,044 Movie Rentals

[VIEW FULL REPORTS BELOW](#)



Data Dictionary



SQL Codes



PowerPoint Presentation



Interactive Tableau Storyboard

BUSINESS ANALYSIS

22% of the viewers are likely to engage in **PG-13 movies** while the **sports film genre** has the dominant rental rate.

```

1 -- RENTAL COUNT BY FILM GENRE
2
3 SELECT 1 -- TOTAL REVENUE PER FILM RATING
4
5 FROM film A
6
7 INNER JOIN film_category B ON A.film_id = B.film_id
8 INNER JOIN payment C ON B.inventory_id = C.inventory_id
9 INNER JOIN rental D ON C.rental_id = D.rental_id
10 INNER JOIN film_category E ON A.film_id = E.film_id
11 INNER JOIN category F ON E.category_id = F.category_id
12 GROUP BY F.name;
13
14 GROUP BY F.name;
  
```

FIG. 3a

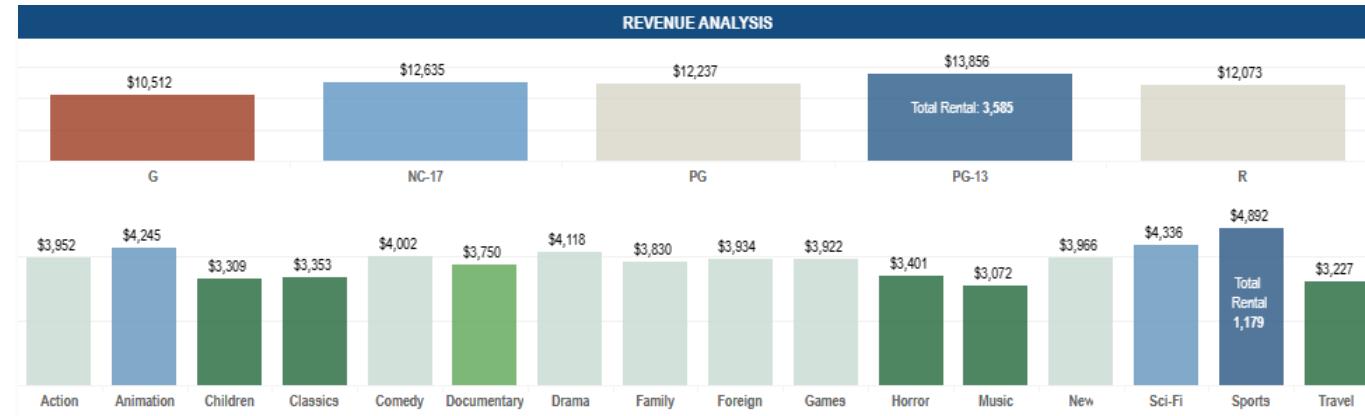


FIG. 3b

Asia takes up 45% of the customer base with **China** and **India** generating the **highest revenue** globally.

```

1 -- REGIONAL SALES REVENUE
2
3 SELECT D.country
4     ,COUNT(DISTINCT A.customer_id) AS customer_count
5     ,SUM(E.amount) AS total_revenue
6
7 FROM customer A
8 INNER JOIN address B ON A.address_id = B.address_id
9 INNER JOIN city C ON B.city_id = C.city_id
10 INNER JOIN country D ON C.country_id = D.country_id
11 INNER JOIN payment E ON A.customer_id = E.customer_id
12 GROUP BY D.country
13 ORDER BY total_revenue DESC;
  
```

FIG. 3c



FIG. 3d

[VIEW FULL REPORTS BELOW](#)



Data Dictionary



SQL Codes



PowerPoint Presentation



Interactive Tableau Storyboard

RECOMMENDATIONS

ONLINE VIDEO RENTAL SERVICES

TARGET AUDIENCE

Determine the micro and macro environment in China, India, and its surrounding areas adapting services that meet cultural demands.

REWARDS PROGRAM

Incentivizing top-paying customers establishes strong client rapport and opens up opportunities to offer new services that best fit their interests.

PRODUCT DEVELOPMENT

Fostering new services that conform to viewers' specific preferences can encourage old and new viewers to optimize services.

BUSINESS DEVELOPMENT

Assess high-performing countries and understand consumer demands to replicate successful operations in other potential markets.



[VIEW FULL REPORTS BELOW](#)



Data Dictionary



SQL Codes



PowerPoint Presentation



Interactive Tableau Storyboard

EXCERPT

PREPARING FOR INFLUENZA SEASON IN THE U.S.

A medical staffing agency aims to allocate temporary workforces in the most impacted areas to mitigate the influenza outbreak for the coming season. With limited resources to hire new employees, they are determined to support the healthcare system across all 50 states.

OBJECTIVE

Analyze historical influenza trends in the U.S. to assist the medical staffing agency in the deployment of temporary healthcare personnel for the upcoming season.

PROJECT & ALTERED DATA

- Project Brief
- Influenza deaths | *Source from CDC*
- U.S. Population | *Source from U.S. Census Bureau*
- Influenza Visits & Influenza Lab Tests | *Source from CDC (Fluvview)*
- Children's Flu shot rates | *Source from CDC*

LIMITATIONS

- Influenza deaths data comprises 82% of death counts below 10 that are suppressed for confidentiality.
- Death records identify a single underlying cause of deaths (influenza-initiated may not be counted).
- Data sets are dated from 2009 to 2017.
- Information on hospital sizes and staffing capacity is unknown.

TECHNIQUES APPLIED

- Designing a Data Research Project
- Data Profiling and Integrity
- Data Cleaning
- Data Transformation and Integration
 - Excel: Pivot Tables and VLOOKUP
- Statistical Analysis and Hypothesis Testing
- Data Visualization and Storytelling (Tableau)

TOOLS



DATA METHODOLOGY



DESIGNING DATA RESEARCH PROJECT

Interpreting business requirements to data questions leads to a research hypothesis that serves as a guideline for the analysis. A project management plan is prepared to keep track of progress.

DATA PREPARATION

Exploring the datasets for information relevancy, integrity, completeness, etc. will help produce valuable insights. Then, transforming and integrating multiple data discloses the influenza case developments that will warrant the planning phase.

STATISTICAL ANALYSIS & HYPOTHESIS TESTING

Performing statistical methods that detect critical age populace targeted by influenza will model the next steps in preparation for the staff distribution. The relations of multiple variables are then confirmed through t-testing.

DATA VISUALIZATION & STORYTELLING

Data results are utilized for a compelling tableau presentation disclosing influenza trends and vulnerable age populations.

STATISTICAL ANALYSIS & HYPOTHESIS TESTING



ANALYTICAL TOOL APPLICATION

According to statistics influenza **mortality rate is much higher** in adults **over 65 years old**, hence susceptible to developing severe cases than in younger age groups.

Research Hypothesis:	If the mortality rate is high among adults over 65, there's an increased demand for medical workers.
Dependent Variable:	Influenza death
Independent Variable:	US Population
Null Hypothesis:	H_0 : Influenza Mortality Rate \leq Adults over 65 years
Alternative Hypothesis:	H_A : Influenza Mortality Rate $>$ Adults over 65 years
P-Value:	1.50984960014892E-130

FIG. 2a

	H_0 : Influenza Mortality Rate $<$ Adults over 65 years	H_A : Influenza Mortality Rate $>$ Adults over 65 years
Mean	0.21%	0.63%
Mean Difference		
Standard Deviation	0.00219507	0.002149116
t-Test: Two-Sample Assuming Unequal Variances		
Mean	0.002145587	0.006296545
Variance	4.81833E-06	4.6187E-06
Observations	455	455
Hypothesized Mean Difference	0	
df	908	
t Stat	-28.82280638	
P(T<=t) one-tail	1.53E-130	
t Critical one-tail	1.646533511	
P(T<=t) two-tail	3.0583E-130	
t Critical two-tail	1.962580045	

FIG. 2b

The **normalized death rate** shows the correlation by age group and illustrates the **magnitude** of influenza deaths.

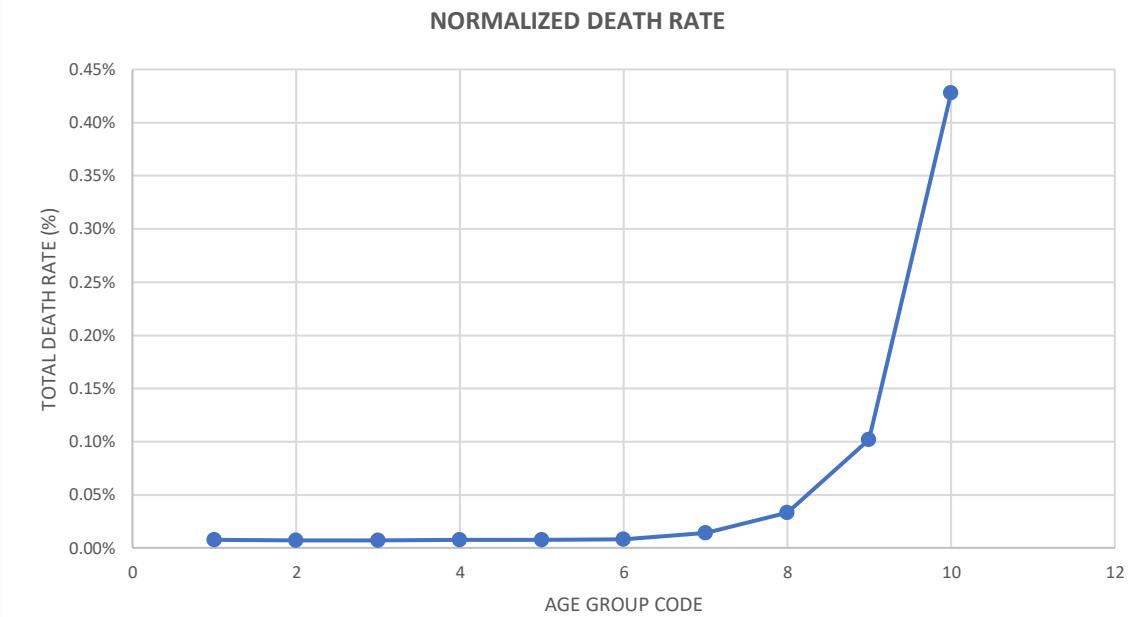


FIG. 2c

[VIEW FULL REPORTS BELOW](#)



Project Management Plan



Data Sourcing



Interim Report



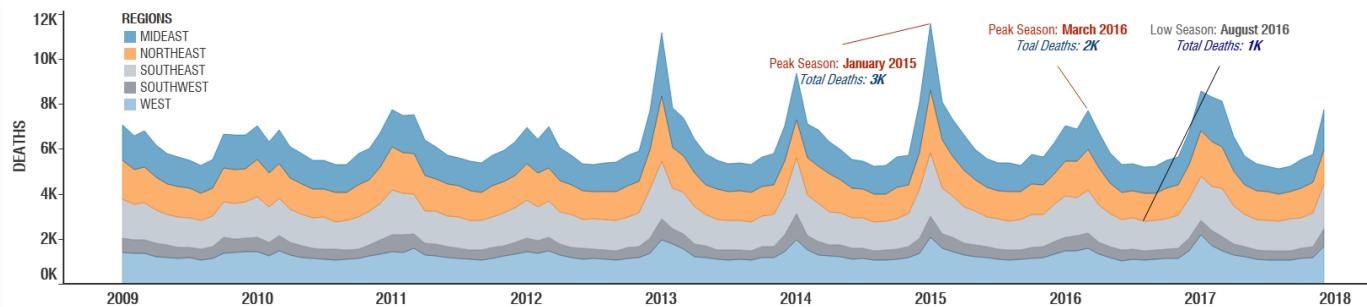
Interactive Tableau Storyboard

TIME SERIES ANALYSIS

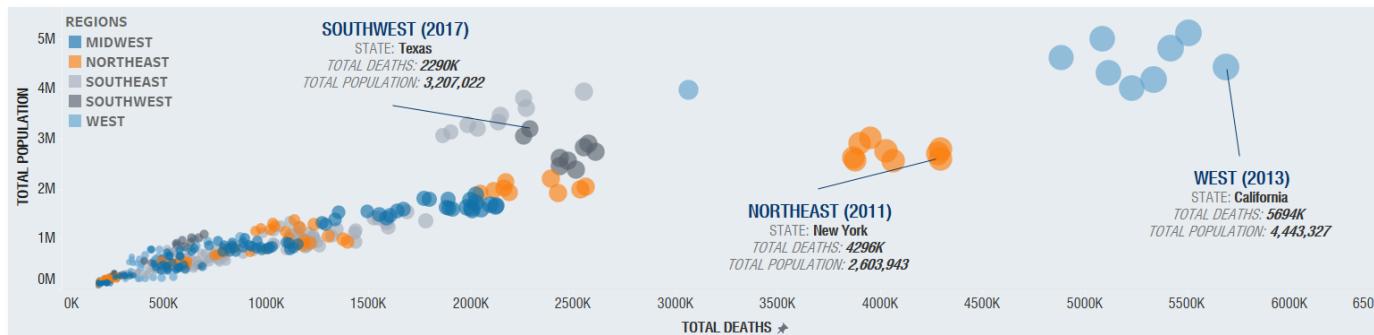


VISUALIZATION TOOL APPLICATION

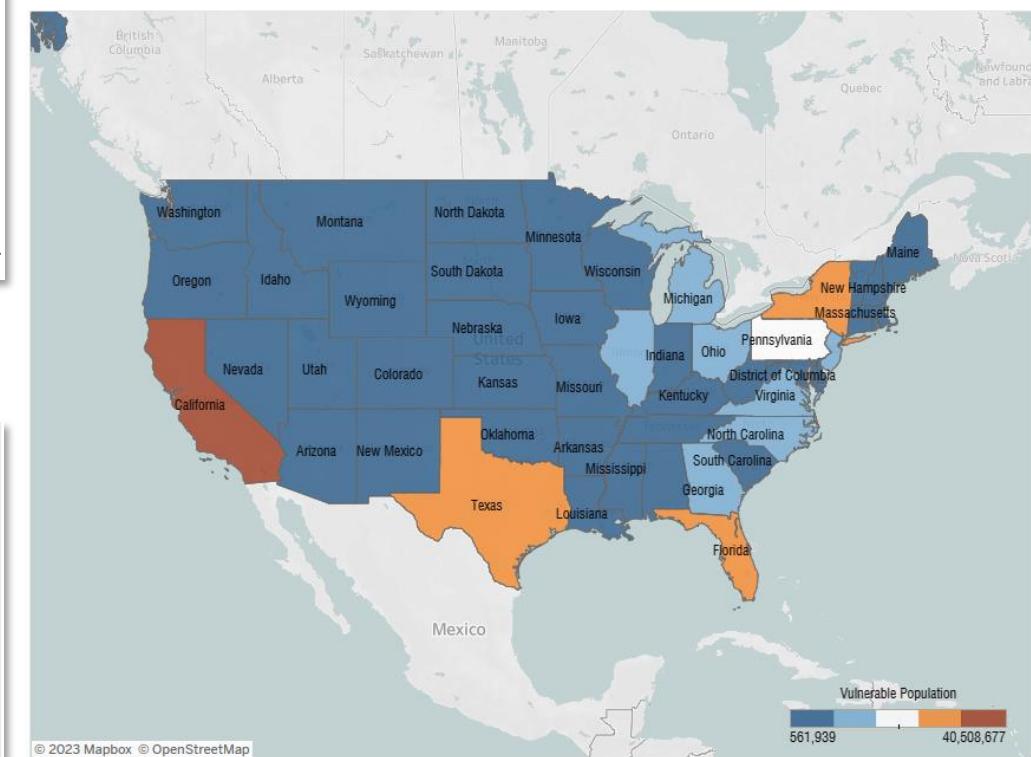
Influenza transpires all year round, however, it has penetrating effects in **colder climates**.



The **death ratio** is proportional to population size signifying a substantial impact on regions with **higher concentrations** of vulnerable age groups.



California, New York, Texas, and Florida are states with a strong density level of vulnerable populations with **significant death rates**.



[VIEW FULL REPORTS BELOW](#)



Project Management Plan



Data Sourcing



Interim Report



Interactive Tableau Storyboard

RECOMMENDATIONS

NATIONAL MEDICAL STAFFING

STAFFING DISTRIBUTION

Allocate 32% of medical personnel to high-priority level regions - California, New York, Texas, and Florida arranging the deployment close to the colder weather.

RESEARCH ANALYSIS

Investigating the hospitals' and clinics' capacities can reinforce actionable numbers to efficiently designate temporary healthcare workers at state and county levels.



[VIEW FULL REPORTS BELOW](#)

MONITORING PROGRAM

Yearly surveillance of influenza cases should be carefully measured to aid in planning medical staffing for future events.

PUBLIC AWARENESS CAMPAIGN

Educating the community on the benefits of vaccination is the best defensive measure to mitigate the spread of influenza.



Project Management Plan



Data Sourcing



Interim Report



Interactive Tableau Storyboard



EXCERPT

GAMECO

A fictional newcomer to the video game industry deems that sales have been the same over time and is interested to understand the current marketing developments.

OBJECTIVE

Perform a 10-year descriptive analysis to gain insights into the current video game developments for marketing and sales teams' 2017 budget planning.

PROJECT & DATA

- [Project Brief](#)
- This mock 1980 to 2016 video game sales data is available [here](#).
- Data sourced from [VGChartz](#).
- VGChartz's data collection methodology is available [here](#).

LIMITATIONS

- Tracks the total number of units sold (*not financial figures*) to the retail stores.
- 2016 is the latest year logged with partial records.

TECHNIQUES APPLIED

- Data Integrity, Quality and Consistency Assessment
- Data Cleaning
- Pivot Tables
 - Data Grouping & Summarizing
 - Calculated Fields
- Descriptive Analysis
- Excel Visualization Results
- PowerPoint Presentation

TOOLS



DATA METHODOLOGY



DATA CLEANING

Preparing data by removal of duplicate and irrelevant values, imputing missing figures with mean values. Normalizing text formats for accessibility (i.e. corrections on typos and names with special characters, and so on).

GROUP & SUMMARIZING DATA

Using excel pivot tables to group and summarize data with categorical filters enhances customize views.

DESCRIPTIVE ANALYSIS

An application of basic exploratory data analysis measures central tendency, distribution, and outliers.

EXCEL VISUALIZATION CHARTS

Data findings are converted to charts illustrating geographical trends by segment. Suitable graphs are used for the final presentation report.

INITIAL ANALYSIS



ANALYTICAL TOOLS APPLICATION

A **declining pattern** in overall sales **refutes the executives' current business understanding.**

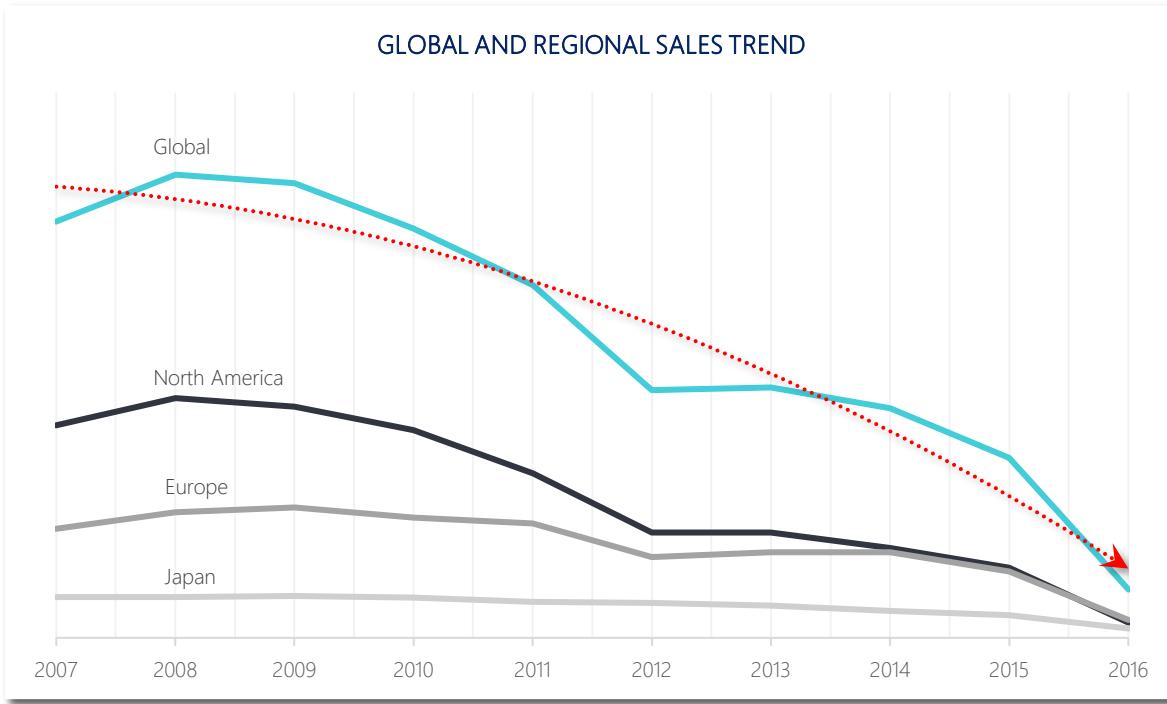


FIG. 1a

- **North America** is the strongest sales performing region suffers the sharpest drop.
- **Europe** 2nd largest market follows similar trend with all other regions.
- **Japan** a single-country market shows minimal impact of the declining development.

The sales ratio molds the **market share trend** demonstrating **distinctive** results of each **regional business growth**.

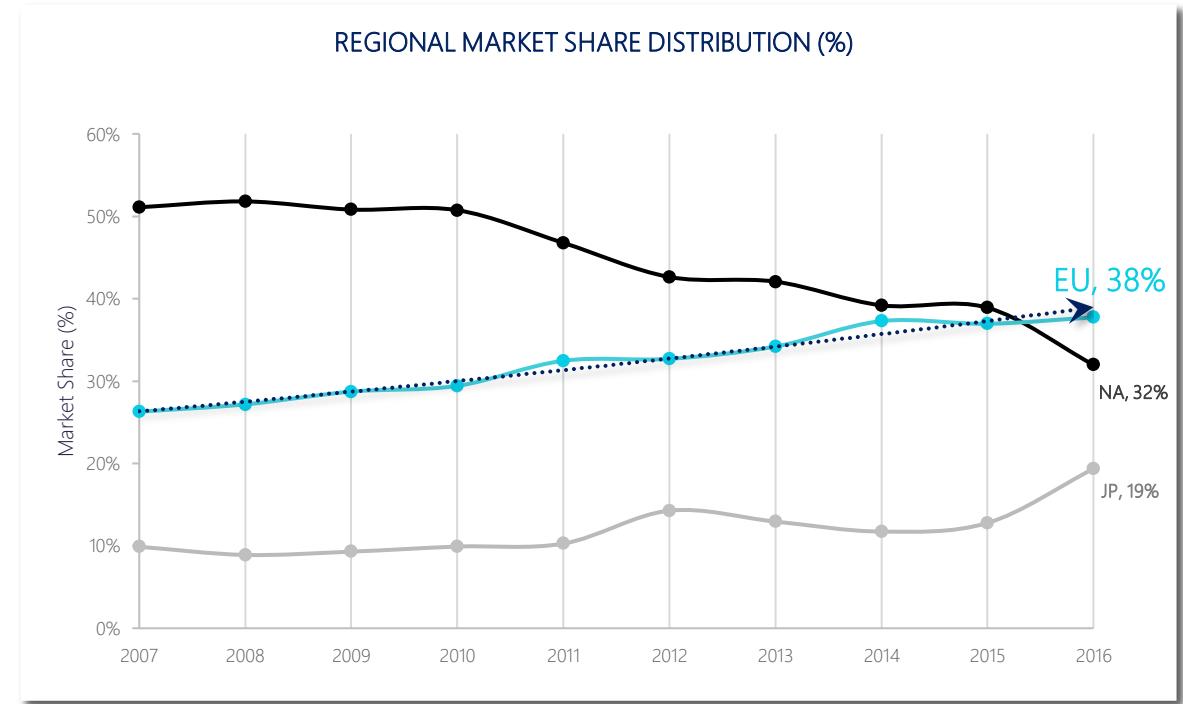


FIG. 1b

- **North America** plummets to the 2nd leading profitable region.
- **Europe** continuously **climbs** taking over the highest market share.
- Japan's unstable growths illustrates positive growth from 2014.

[VIEW THE FULL REPORT BELOW](#)



PowerPoint Presentation

MARKET TREND ANALYSIS



ANALYTICAL TOOL APPLICATION

Genre popularity changes over time suggesting **Action**, **Sports**, **Shooter**, and **Role-Playing** are likely to maintain **video game players' interests** while other categories eventually diminish globally.

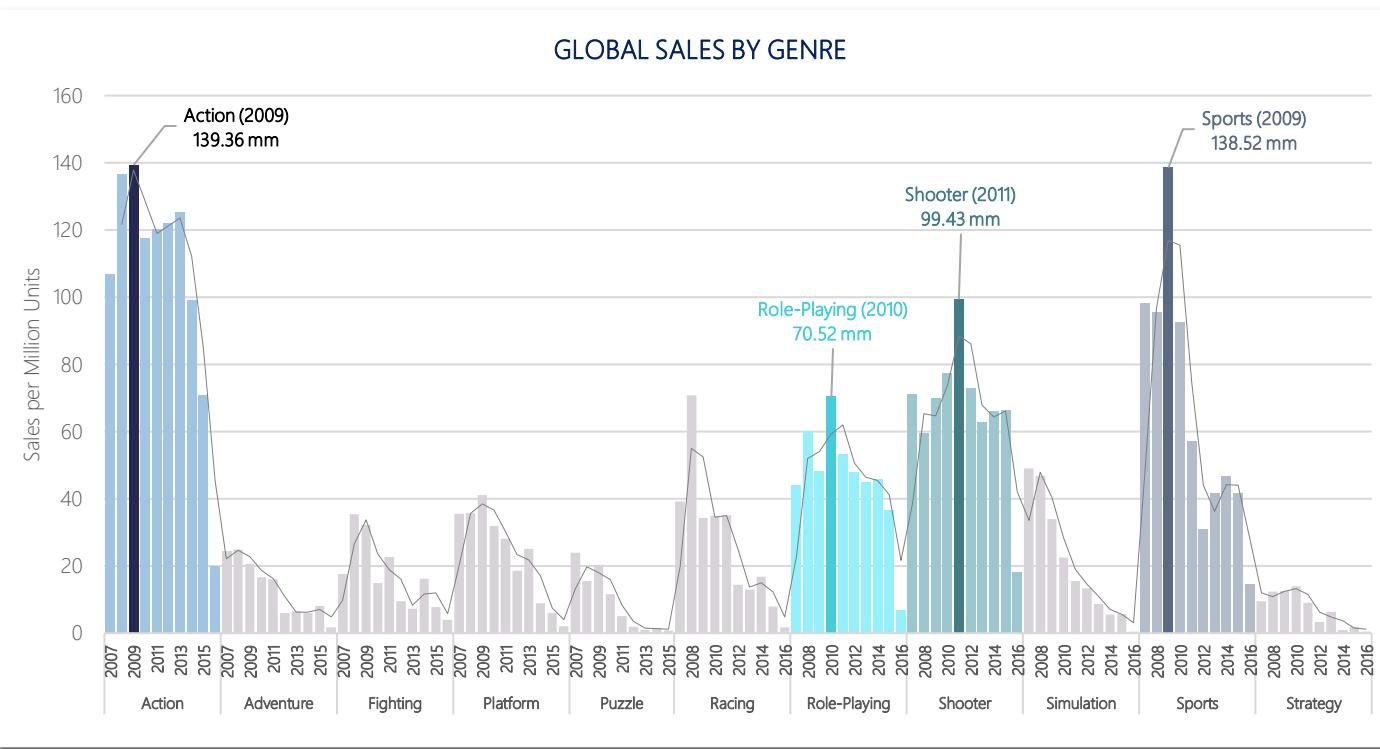


FIG. 1c

[VIEW THE FULL REPORT BELOW](#)

Identifying **regional consumer developments** observes **opportunities** leveraging on high-demand game categories adapted to each target market's interests.

TOP-PERFORMING GENRE makes up more than 50% of sales by region.			
RANK	EUROPE	NORTH AMERICA	JAPAN
1	Action	Action	Role-Playing
2	Shooter	Shooter	Action
3	Sports	Sports	

FIG. 1d

TOP-SELLING PLATFORM makes up above 69% of sales in each region.			
RANK	EUROPE	NORTH AMERICA	JAPAN
1	PS3	X360	DS
2	X360	WII	3DS
3	WII	PS3	PS3
4	DS	DS	PSP

FIG. 1e

RECOMMENDATIONS

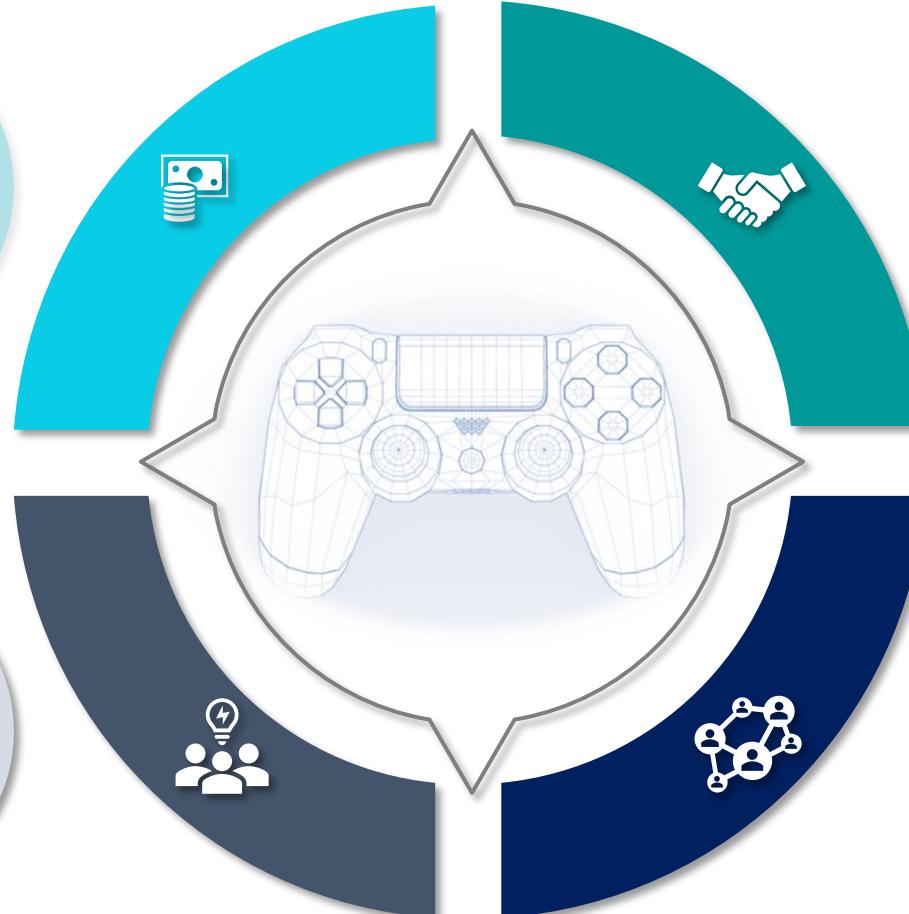
GLOBAL MARKET

BUDGET DISTRIBUTION

Investing 38% of the resources to European market support its the year-to-year stable growth.

TARGET MARKET

Implement consumer behavior reports in diagnosing buying and spending habits to disclose detailed facts on product preferences including purchasing power trends.



SALES DEVELOPMENT

Explore sales channels (*online vs. retail*) and acquire current platform trends used for promoting and selling products.

SCHEMATIC TRADING

Conduct an in-depth marketing analysis on publishers' recent product developments to discover the profitable trends and partnering with new enterprises.

[VIEW THE FULL REPORT BELOW](#)



PowerPoint Presentation

EXCERPT

PIG E. BANK

A theoretically well-known global bank seeks analytical support for its anti-money laundering compliance division and is determined to improve services to increase customer retention.

OBJECTIVE

Contribute to the development and optimization of models to enhance the efficiency of the bank's compliance program in identifying client loss risk factors and flagging suspicious transaction behavior.

PROJECT & DATA

- Project Brief
- Client Dataset | provided by CareerFoundry

LIMITATIONS

- Customer demographics are limited to gender, age, and country with records of their account balance, estimated salary, membership status, etc.

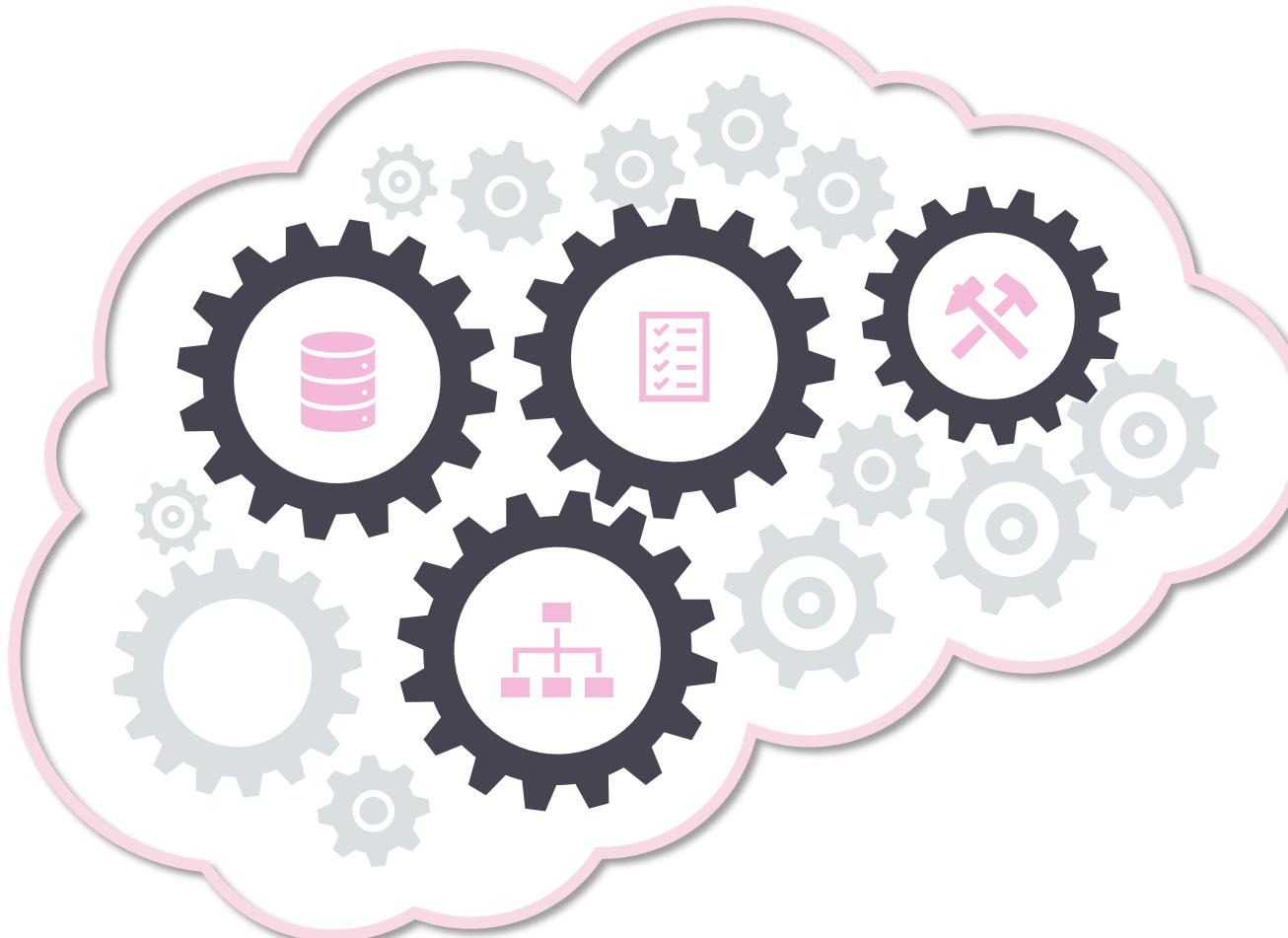
TECHNIQUES APPLIED

- Big Data Management
- Data Ethics
- Data Mining
- Predictive Analysis
- Time Series Analysis and Forecasting

TOOLS



DATA METHODOLOGY



BIG DATA

Defining the data characteristics (structured and unstructured) and recognizing its applications and limitations. Identifying software tools appropriate for handling big data.

DATA ETHICS

Identifying ethical dilemmas in managing big data in compliance with security and privacy laws.

DATA MINING

Performing data cleaning and descriptive statistics using pivot tables that will be utilized in generating a decision tree modeling for testing outcomes of the analysis.

PREDICTIVE, TIME SERIES ANALYSES AND FORECASTING

Optimizing linear regression models through testing correct predictive prototypes in classifying risk factors that may have contributed to customer loss with the application of different scenarios.

CUSTOMER RETENTION ANALYSIS



ANALYTICAL TOOLS APPLICATION

There is a **notable percentage** of customers that have **closed** bank accounts.

EXIT STATUS DISTRIBUTION

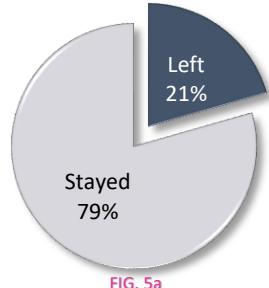


FIG. 5a

Non-active customers are likely to **exit**.

TOTAL AVERAGE EXIT STATUS BY MEMBERSHIP STATUS

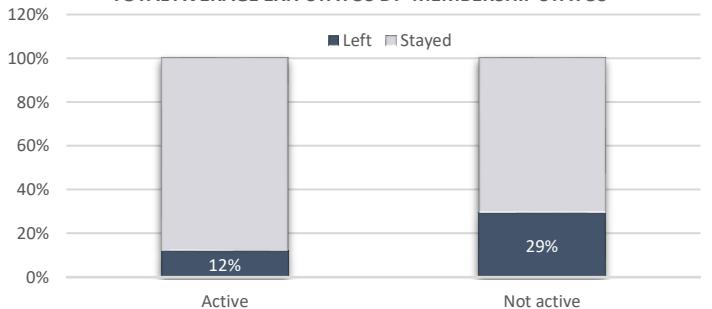


FIG. 5b

18% of the clients with **1-2** products have **left**.

OVERALL DISTRIBUTION

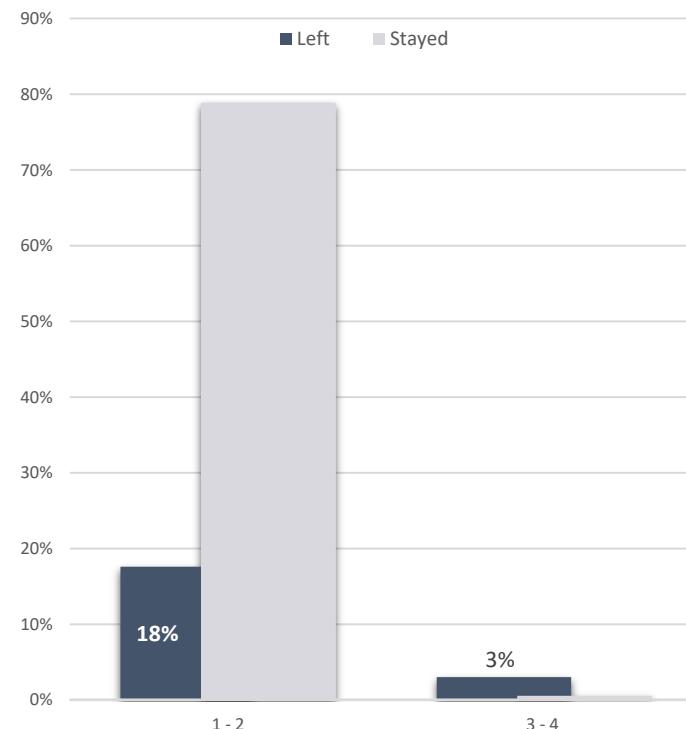


FIG. 5c

Clients residing in **Germany** have a significant proportion of **withdrawing** services.

TOTAL AVERAGE EXIT STATUS BY COUNTRY

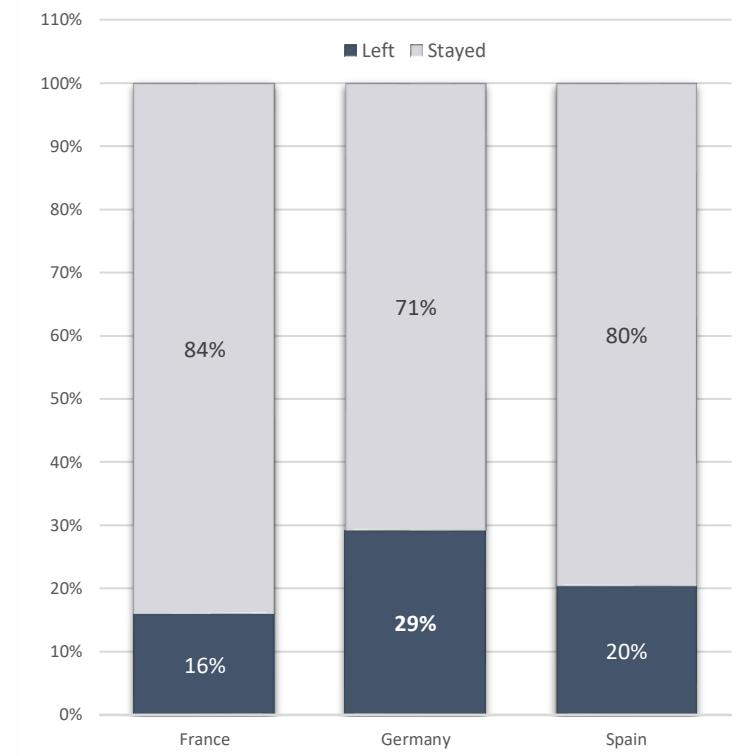


FIG. 5d

[VIEW THE FULL REPORT BELOW](#)



Final Report

PREDICTIVE MODEL ANALYSIS



ANALYTICAL TOOLS APPLICATION

The **factors** that have a **substantial impact** on customer attrition are **non-active members** that had **one to two products**. Additionally, the residential location, specifically the **German** market, poses a **higher risk** of customers terminating their relationship with the bank, even when demographic information is not taken into account.

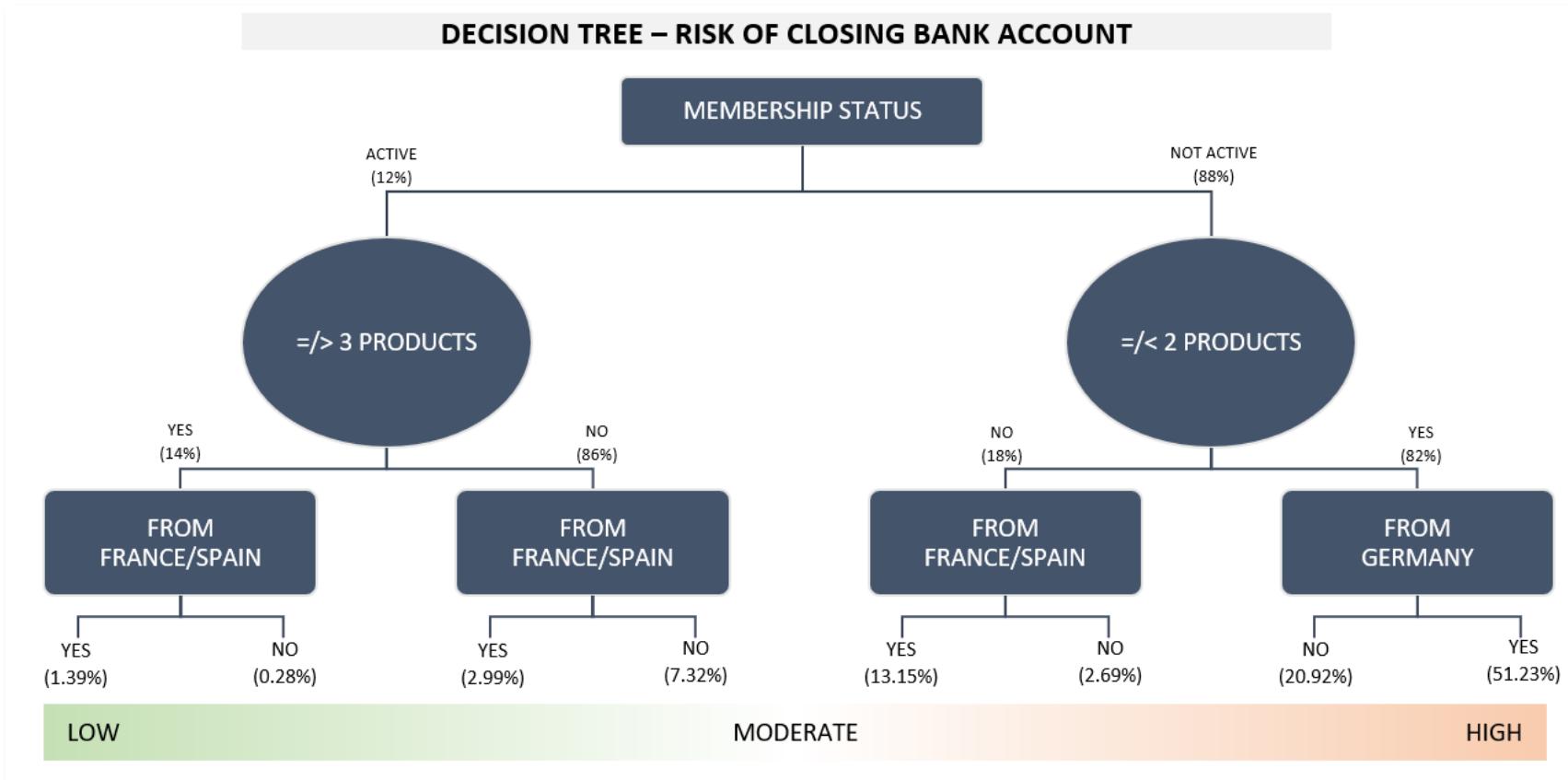


FIG. 5e
VIEW THE FULL REPORT BELOW



Final Report

RECOMMENDATIONS

CUSTOMER PRESERVATION

BUSINESS DEVELOPMENT

Implement banking strategies aimed at driving product sales within the target markets of France and Spain, targeting potential clients and business communities.

PRODUCT EVALUATION

Examine the products and services provided to various demographic groups to enhance and develop innovative solutions aimed at converting non-active members.

CUSTOMER BEHAVIOR ANALYSIS

Undertaking a survey to gain insights into the specific needs of customers, including those who are planning to terminate their accounts, with the aim of addressing their concerns and improving customer retention rates.

RESEARCH & DEVELOPMENT

Perform a marketing research analysis on the German market to delve deeper into the factors contributing to the elevated customer attrition risk.



[VIEW THE FULL REPORT BELOW](#)



Final Report

EXCERPT

THE WORLD'S LARGEST PUBLIC COMPANIES

The Forbes Global 2000 releases the annual ranking of the top 2000 public companies in the world, distributed by Forbes magazine. "The Global 2000" annual ranking is determined by a weighted evaluation of four key criteria: sales, profit, assets, and market value.

OBJECTIVE

Evaluate the performance of top-performing industries from 2008 to 2022 to identify successful operational strategies using key performance indicators (KPIs) and gain insights into the impact of COVID-19 on business operations.

PROJECT & DATA

- Project Brief
- Datasets: [2008](#), [2009](#), [2010](#), [2011](#), [2012](#), [2013](#), [2014](#), [2015](#), [2016](#), [2017](#),
[2018](#), [2019](#), [2020](#), [2021](#), [2022](#) | Open Source from [data.world](#).
- Click [here](#) for Forbes' Global 2000 methodology

LIMITATIONS

- Data contains records from 2008 – 2022.
- The global largest public companies are limited to the top 2000 performers from various industry types.

TECHNIQUES APPLIED

- Data Sourcing & Data Cleaning: Wrangling, Subsetting, and Consistency Checks
- Data manipulation: Deriving New Variables, Aggregating, and Grouping Data
- Exploratory Visual Analysis: Linear Regression, Geospatial
- Linear Regression (Unsupervised Machine Learning Model)
- K-means Clustering (Supervised Machine Learning Model)
- Time Series Analysis
- Data Dashboard (Tableau Storyboard)



TOOLS



DATA METHODOLOGY



DATA SOURCING & DATA CLEANING

Source datasets that align with project objectives for advanced analytics, apply necessary preparatory methods to optimize diagnostic outcomes, and define questions to explore the data content.



DATA MANIPULATION & EXPLORATORY ANALYSIS

Utilize Excel and Python functions to aggregate and group data, perform basic statistical analysis, create new columns, and optimize the process through exploratory visual analysis using techniques such as geospatial analysis, scatterplots, histograms, and more.



ADVANCED ANALYTICAL TECHNIQUES

Employing a combination of techniques, such as supervised and unsupervised machine learning models, along with time series analysis, valuable insights can be conveyed regarding potential developments in regions and industry types.



DATA DASHBOARD

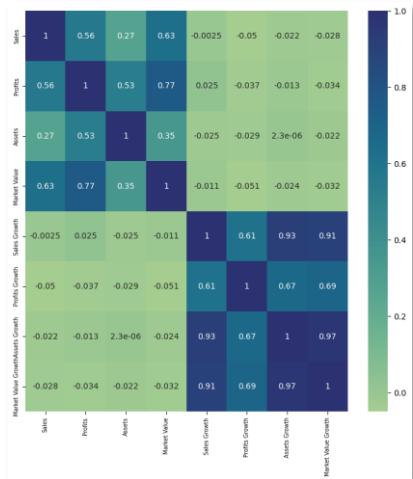
Forming a Tableau storyboard that presents curated significant findings of the analysis in an interactive format.

EXPLORATORY VISUAL ANALYSIS



To analyze the relationships between quantitative variables like sales, profits, assets, and market values, a **correlation heat map** was generated. The heat map helps identify initial connections between these variables.

The analysis revealed a **strong positive** correlation between **sales** and **profits**.



```
# GLOBAL
# Create a subplot with matplotlib

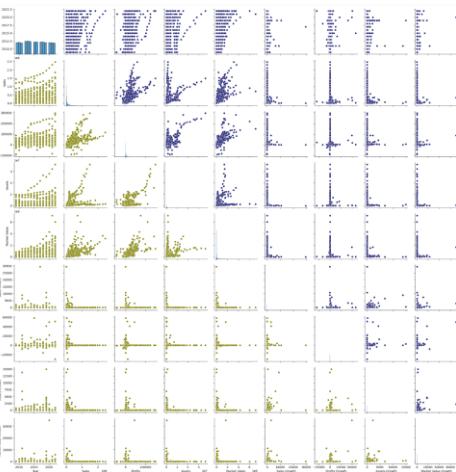
dfgwth_hm1,ax = plt.subplots(figsize = (9,9))
plt.xticks(range(dfgwth_hm1.shape[1]), dfgwth_hm1.columns, fontsize = 7, rotation = 90) # x axis labels
plt.yticks(range(dfgwth_hm1.shape[1]), dfgwth_hm1.columns, fontsize = 7) # y axis labels

# Create the correlation heatmap in seaborn by applying a heatmap onto the correlation matrix
gbl_hm2 = sns.heatmap(usa_hm.corr(), cmap = 'crest', annot = True, ax = ax)
```

FIG. 6a

Pair plots, similar to heat maps, depict relationships and directionality between variables. They help determine the potential effects of one variable on another.

In the pair plot analysis, variables related to **sales** show positive growth trends, indicating that sales could be a **significant predictor** of performance.



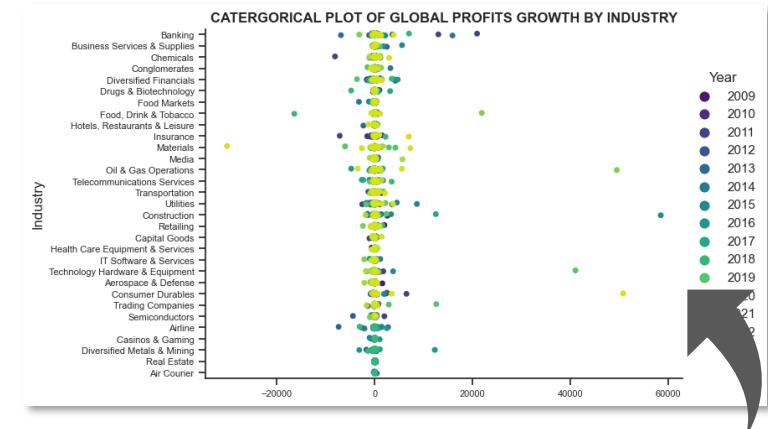
```
# GLOBAL
# Create a first Pair Plots

gbl_pp = sns.pairplot(gblpp)
gbl_pp.map_upper(sns.scatterplot, color = 'midnightblue')
gbl_pp.map_lower(sns.scatterplot, color = 'olive')
```

FIG. 6b

Categorical plots provide valuable insights into non-linear relationships, complementing the information from pair plots.

Analyzing **industry-specific** developments over time reveals an overview of **performance trends**.



```
# Set the figure size
plt.figure(figsize = (50,10))

# Creating a Categorical plot
sns.set(style = 'ticks')
cp_gbl = sns.catplot(x = 'Profits Growth', y = 'Industry', hue = 'Year', palette = 'viridis', aspect = 1)

# Adding annotations to the graph
plt.xlabel('')
plt.yticks(fontsize = 8)
plt.xticks(fontsize = 8)
plt.title('CATEGORICAL PLOT OF GLOBAL PROFITS GROWTH BY INDUSTRY', fontweight = 'bold')
```

FIG. 6c

VIEW FULL REPORTS BELOW

ADVANCED ANALYTICAL TECHNIQUES



The **linear regression** analysis identifies a strong correlation between sales and profits, indicating a trend.

However, the statistical testing yields **poor results**, suggesting that there is not enough evidence to depict the presence of the trends.



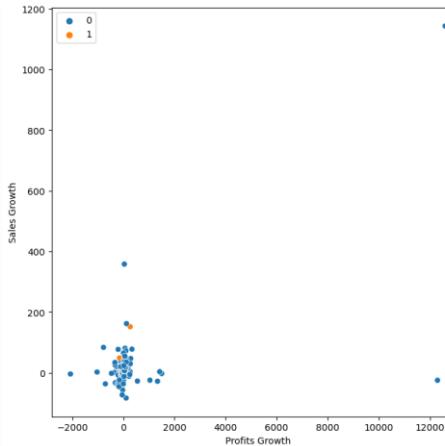
Visualizing the training set results.

```
plot_test = plt
plot_test.scatter(X_train, y_train, color = 'seagreen', s = 15)
plot_test.plot(X_train, y_predicted_train, color = 'red', linewidth = 3)
plot_test.title('Sales vs Profit (Train set)')
plot_test.xlabel('Sales Growth')
plot_test.ylabel('Profit Growth')
plot_test.show()
```

FIG. 6d

Further testing was conducted using the **K-means algorithm**, which groups data points with similar traits into clusters.

However, the scatterplot reveals overlapping data points without clear groupings, indicating an **inadequate fit** to the prediction model.



Plot the clusters

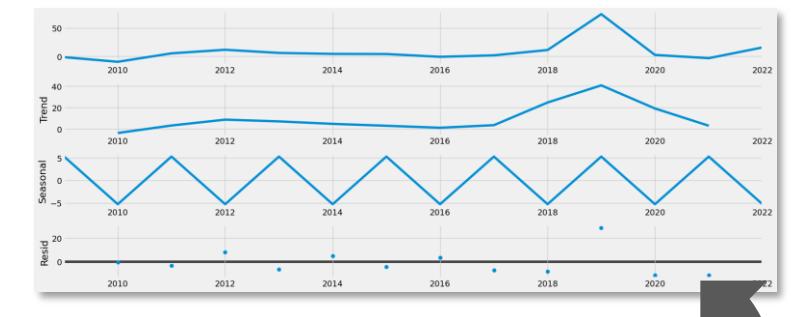
```
plt.figure(figsize = (8,8))
ax1 = sns.scatterplot(x = usa2['Profits Growth'], y = usa2['Sales Growth'], hue = kmeans.labels_, size = s)
# Here, you're subsetting `X` for the x and y arguments to avoid using their labels.
# `hue` takes the value of the attribute `kmeans.labels_`, which is the result of the K-Means clustering.
# `s` represents the size of the points you want to see in the plot.

ax1.grid(False) # This removes the grid from the background.
plt.show()
```

FIG. 6e

The **decomposition** of the time series allows for the assessment of individual components, such as seasonality, to identify possible trends.

It is evident that **sales fluctuate** over time, with a significant decline in 2019, coinciding with the onset of the pandemic.



Decompose the time series using an additive model

```
decomposition1 = sm.tsa.seasonal_decompose(usa2, model = 'additive', period = 2)

from pylab import rcParams # This will define a fixed size for all special charts.
rcParams['figure.figsize'] = 18, 7

# Plot the separate components
decomposition1.plot()
plt.show()
```

FIG. 6f

VIEW FULL REPORTS BELOW

RECOMMENDATIONS

NEXT STEPS

QUALITATIVE RESEARCH

Supplement quantitative analysis with qualitative research to gain a holistic view of a company's prospects. Consider non-financial factors like industry dynamics, competitive landscape, market trends, and overall economic conditions (GDP growth, inflation rates, labor market condition, consumer spending patterns) to understand the broader context in which financial metrics operate.

INVESTOR SENTIMENT ANALYSIS

Analyze stock price movements, market reactions to earnings releases or product launches, and current events.

Provide insights into how investors perceive and react to company developments, which can influence market value and investor sentiment.

COMPARATIVE ANALYSIS

Compare the sales, profit, assets, and market value of the company with its competitors in the same industry or sector. Identify relative strengths and weaknesses, assess market positioning, and spot any significant divergences.

FINANCIAL RATIO ANALYSIS

Analyze the ratios over time and/or compare them to industry benchmarks, investors and analysts can gain insights into a company's financial health, efficiency, and market valuation.



[VIEW FULL REPORTS BELOW](#)



Python

Codes |



+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

+

Interactive Dashboard

THANK YOU



GRACE SKELLEY
DATA ANALYST PORTFOLIO