Data Analytics Portfolio

Muhammad Abdulkayyum, Muttaki
04/05/2023

About me

- I am Muhammad Abdulkayyum, a data analyst with programming experience in software development. I have 13 years of experience in software application development with PL/SQL, Python and C++. Along with coding skills, I participated in Analysis, design, development and testing.

- I have always been fascinated by the power of data to drive business decisions, and I am eager to apply my technical expertise to data analytics. I carried practical experience in data analysis and visualisation techniques in Tableau, Python and Excel in Data Analytics certification.
- As I transition into data analytics, I am excited to build upon my existing skills and learn new ones. I want to become a (senior) data analyst in Munich and advance my profession by getting a permanent job.
- I am attaching some of the Data Analytics projects from the CareetFoundry certification course.

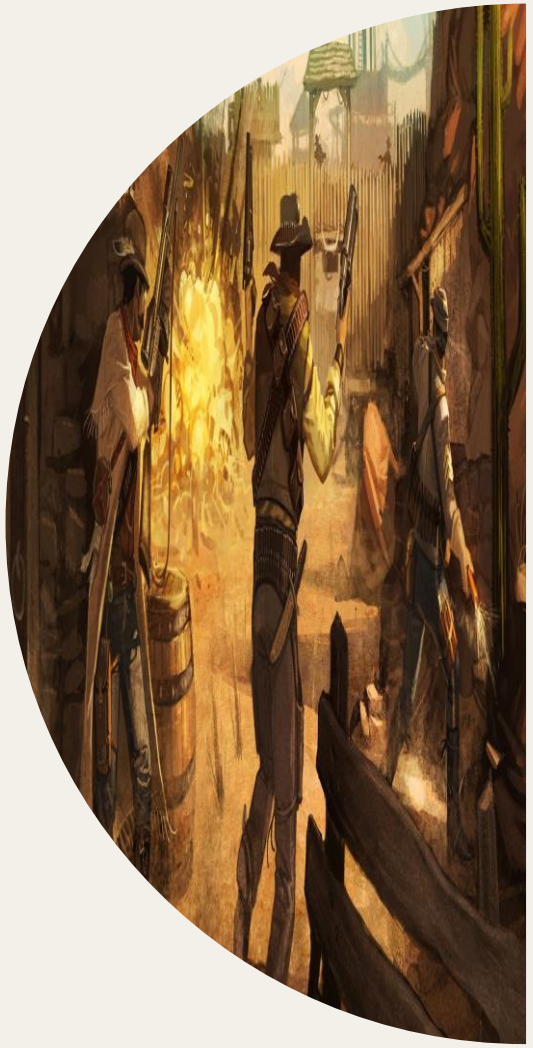
Projects

+ Analyzing global video game sales
+ Preparing for the flu season in the US
+ Answering business questions for an online video rental company
+ Marketing strategy for online grocery stores
+ Anti-money laundering projects at the global bank

Analysing global video game sales

Project duration: 1 month

- About the project
- Approach to the problem
- Stakeholders requirements
- Conclusion with visualisation



Analysing global video game sales

Goal

GameCo is a video game company which wants to use data to inform the development of new games to customers. As an analyst, I've been asked to perform a descriptive analysis of a video game data set to foster a better understanding of how GameCo's new games might fare in the market.

Data set

Data set that covers historical video game sales (for games that sold more than 100,000 copies) spanning different platforms, genres, and publishing studios.

- Source: e VGChartz
- Download Data Set

Skills

- Excel
- Grouping Data
- Summarising the data
- Descriptive Analysis
- Visualising results in Excel
- Presenting results

Approach to the problem

1. Understanding the data

- Understanding the data and classifying the data set as Quantitative or Qualitative.

2. Data cleaning

- Identifying the source of data, data bias and gaps.
- Identify dirty data and analytics practice methods of cleaning data.

3. Analytics

- Experiment with the fundamentals of descriptive analysis by exploratory data analysis (EDA).
- identify the Symmetrical and Skewed Distribution.

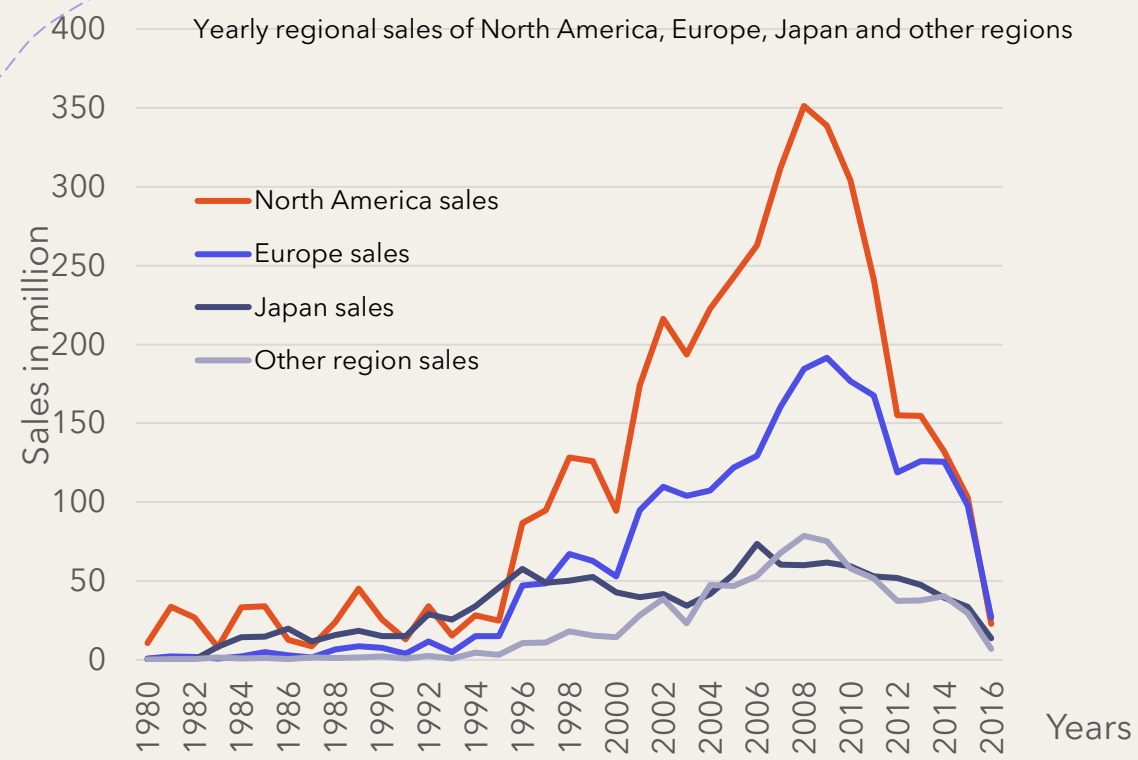
4. Storytelling

- Develop data visualisations for stakeholders.
- Develop a narrative about data insights for stakeholders.

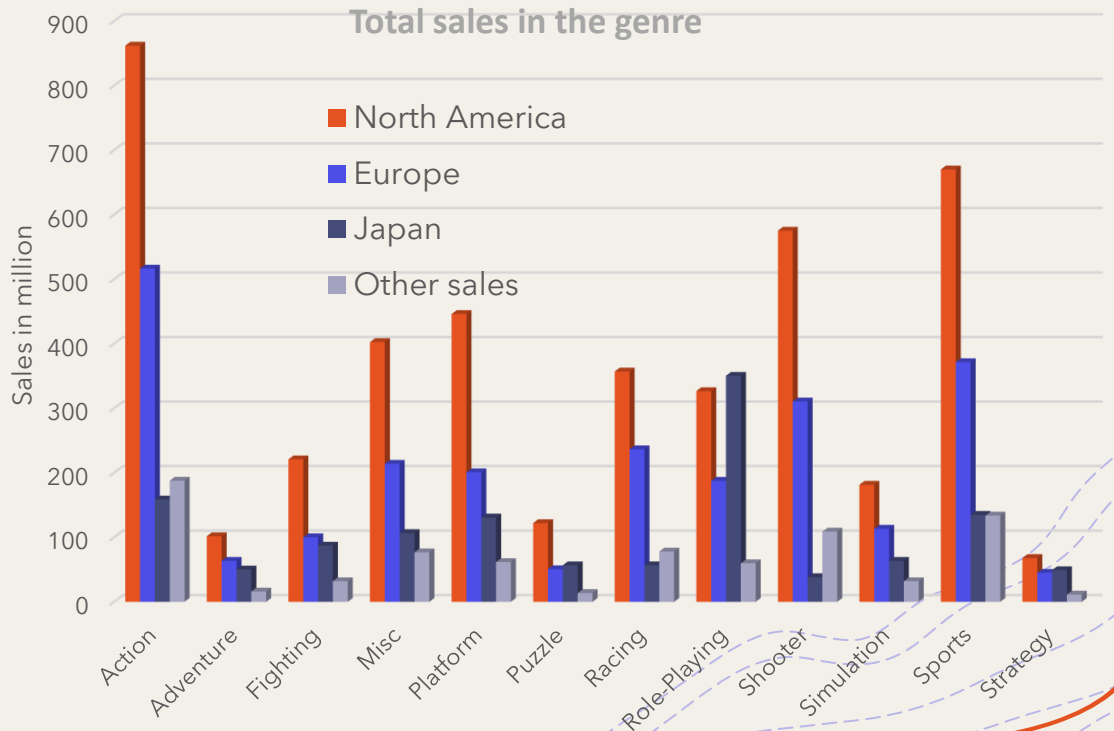
Final presentation to stakeholders

Using pivot tables and visualisation tools some of the insights from the data.

Sales figures never stayed the same worldwide over time.



Action, Sports and shooter games are the most popular games in North America, Europe and Japan.



Conclusion

1. GameCo administrative assumed sales across regions is consistent over the year, and the distribution of marketing resources equally would generate more revenue. But the sales timeline graph across areas has different profits.

2. Marketing resource distribution is prioritised depending on the publication, genre and platform the games are played in each region.

3. Shipping dates could be the possibility for the lower sale values in other regions. Online access/streaming of the games could generate faster revenue.

4. The lower-sale genre games like puzzle, adventure, and strategy games require more marketing.

GameCo.

- **Resource Links**
  - Reports
  - Exploratory datasheet
  - Storyboard

Preparing for the flu season in the US

Project duration: 1 month

- About the project
- Approach to the problem
- Stakeholders requirements
- Conclusion with visualisation

Preparing for the flu season in the US

Goal

To help a medical staffing agency that provides temporary workers to clinics and hospitals as needed. The analysis will help plan for influenza season when additional staff are in high demand. The final results will examine trends in influenza and how they can be used to plan for staffing needs across the country proactively.

Skills

- Excel
- Translating business requirements
- Data cleaning
- Data integration & transformation
- Statistical hypothesis testing
- Visual Analysis & Forecasting
- Storytelling in Tableau
- Presenting results

Data set

- Influenza deaths by geography, time, age, and gender
  - Source: CDC
  - Download Data Set
- Population data by geography
  - Source: US Census Bure
  - Download Data Set
- Counts of influenza laboratory test results by state (survey)
  - Source: CDC (Fluview)
  - Download Influenza Visits Data Set
  - Download Data Set
- Survey of flu shot rates in children
  - Source: CDC
  - Download Data Set

Stakeholders requirements

- Provide information to support a staffing plan, detailing what data can help inform medical personnel's timing and spatial distribution throughout the United States.

- Determine whether influenza occurs seasonally or throughout the entire year. If seasonal, does it start and end simultaneously (month) in every state?

- Prioritise states with large vulnerable populations. Consider categorising each state as low-, medium-, or high-need based on its vulnerable population count.

- Assess data limitations that may prevent you from conducting your desired analyses.

Approach to the problem

1. Understanding the data

- Translate business requirements into data questions by Clarifying, Adjoining, Funneling and Elevating methods.
- Formulate a research hypothesis and distinguish between relevant and irrelevant data for a given data project.

2. Data preparation

- Preparing data for analysis by profiling and cleaning the data.
- Identifying key columns and integrating them using VLOOKUP().
- Applying statistical methods—such as standard deviation, variance, and correlation—to analyse your data.

3. Hypothesis testing

- Conducting an inferential analysis via hypothesis testing.
- Interpreting the results of a hypothesis test.

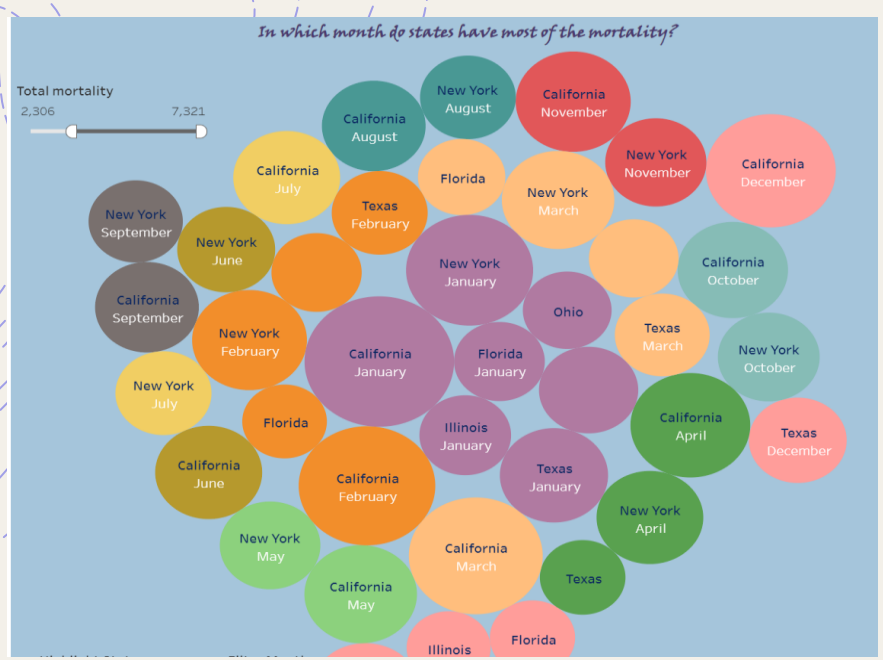
4. Storytelling

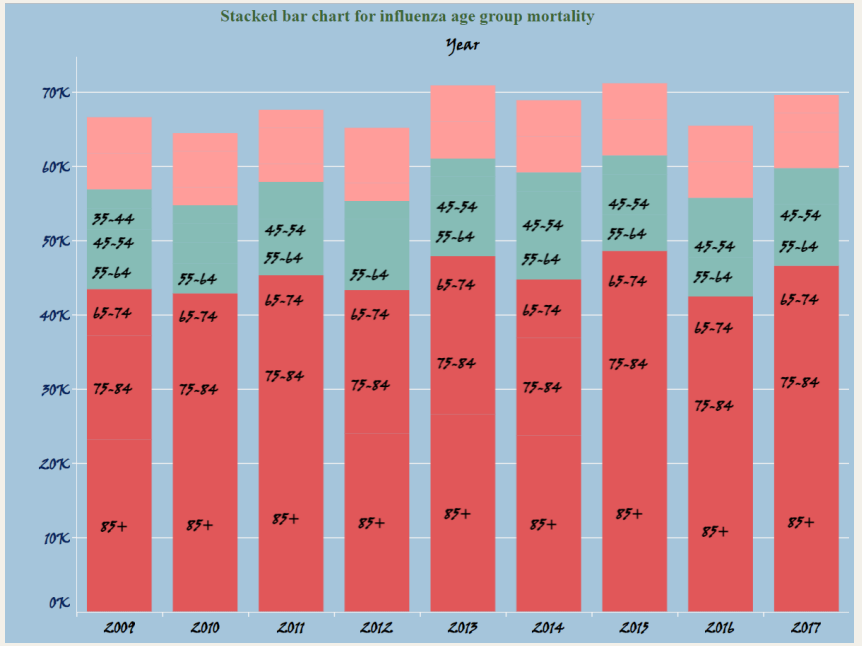
- Consolidating Analytical Insights.
- Preparing visualisations using Tableau

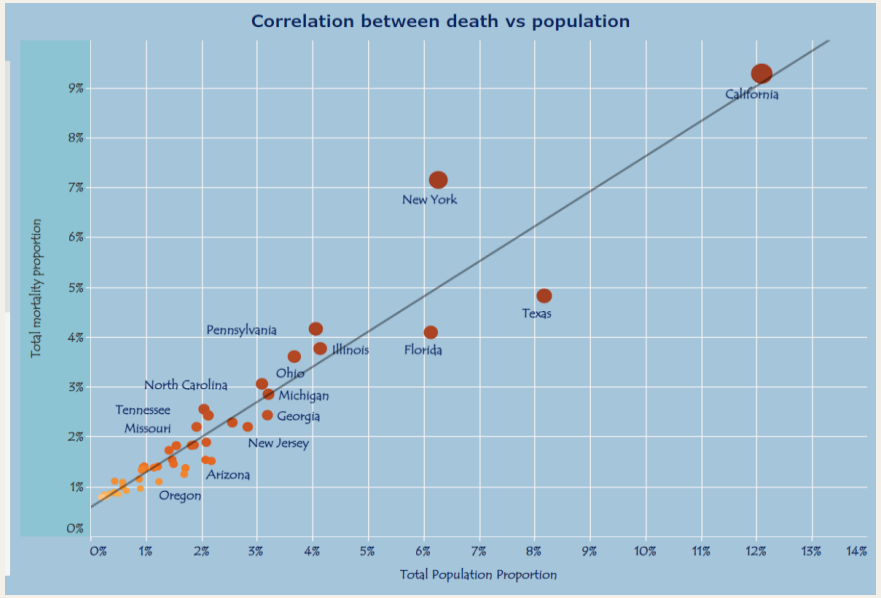
Visualisation using Tableau

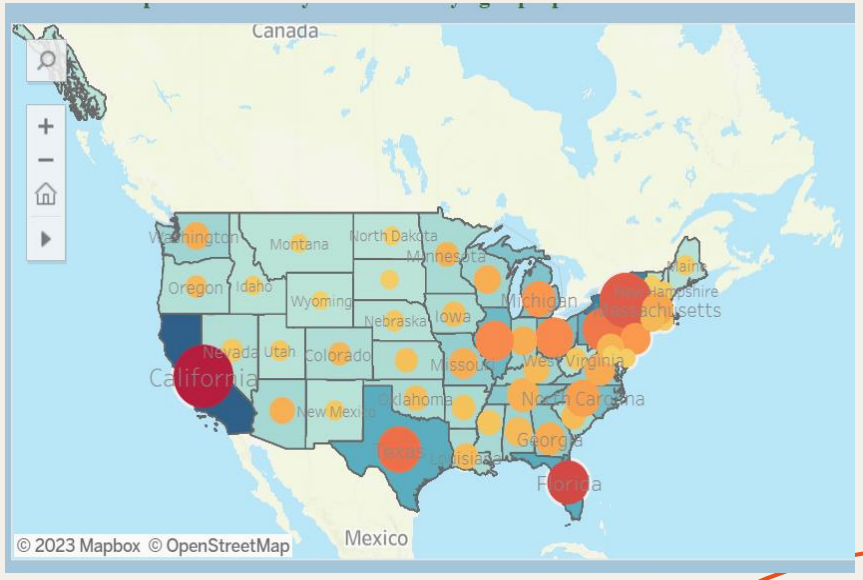
- Identifying vulnerable age groups affected highly by flu using composition and comparison charts in Tableau.
- Statistical Visualizations using Histograms & Box Plots
- Tableau temporal visualisations & forecasting were used to find the most affected region and to allocate sufficient medical staff demand.
- Creating visualisation to find regional flu-affected patients through spatial analyses.

Some of the Tableau visualisation & links









- **Resource Links**
  - Reports
  - Exploratory datasheet
  - Storyboard(Public tableau link)

Conclusion

1. Medical staff distribution

VULNERABLE STATE: California, New York, Pennsylvania and Texas the most.

HIGHLY POPULATED STATES: California, New York, Pennsylvania and Texas

The spread of the influenza virus is *positively correlated with the population.*

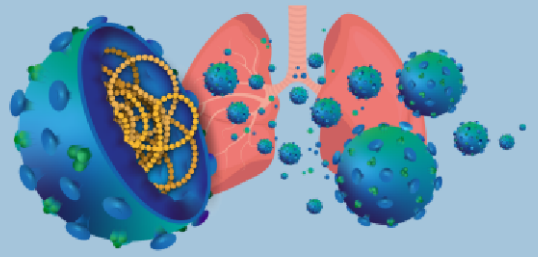
2. Medical Staff reservation

VULNERABLE MONTHS: Dec, Jan, Feb and March

ii. SENTIMENT ANALYSIS: Patient's reviews describe negativity for the flu vaccination waiting period. Flu vaccination staff can be reserved to reduce patient waiting time during Sept, Oct, and November,

VULNERABLE AGE GROUP: 65 and above 65

The vulnerable patient count is Positively correlated with the state population.



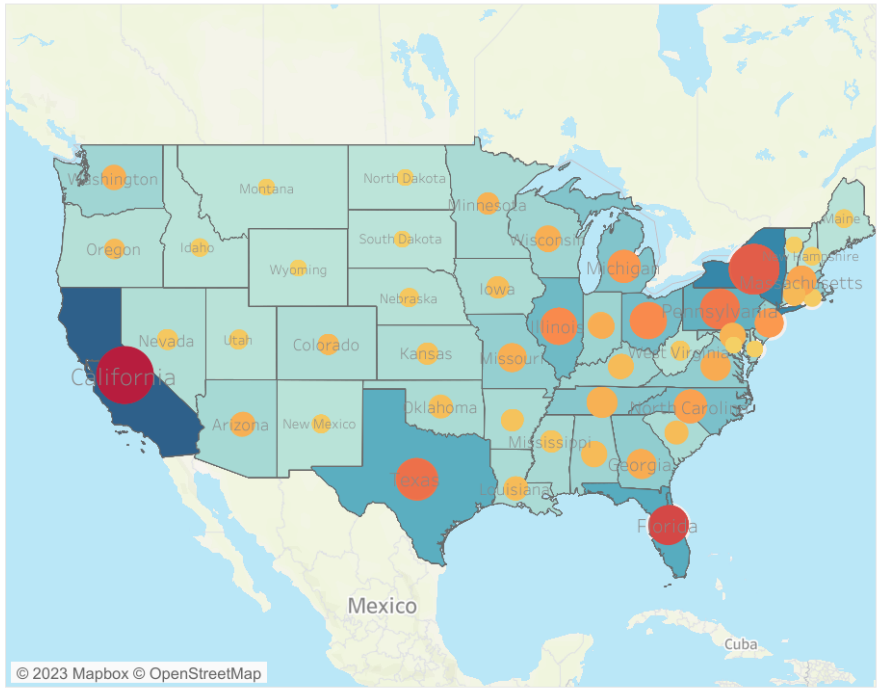
The Influenza Virus affects the respiratory system.

State	
California	9.28%
New York	7.15%
Texas	4.82%
Pennsylvania	4.16%
Florida	4.09%
Illinois	3.77%
Ohio	3.60%
North Carolina	3.06%
Michigan	2.85%
Tennessee	2.55%
Georgia	2.43%
Massachusetts	2.43%
Virginia	2.28%
Missouri	2.20%

Total mortality

4,752 — 56,530

Dual-axis map: Total mortality and the elderly aged people's death in each US state



4. Further Analysis

i. Predict what percentage of additional staff required in 2018

ii. Predict in which month medical staff should be reserved in 2018

iii. Find the elderly aged people's proportion in the state with the respective total population of the state.

Answering business questions for an online video rental company

Project duration: 1 month

- About the project
- Approach to the problem
- Stakeholders requirements
- Conclusion with visualisation

Answering business questions for an online video rental company

Goal

Rockbuster Stealth LLC is a movie rental company that used to have stores worldwide. Rockbuster Stealth's business intelligence (BI) department asked to help with the launch strategy for the new online video service. The BI department helps other departments with data-related queries, from inventory to customer insights.

Data set

Data set contains information about Rockbuster's film inventory, customers, and payments, among other things.

- Download the Rockbuster data set

Skills

- Relational database: PostgreSQL
- SQL
- Database querying using ERD
- Filtering
- Cleaning and summarising the data
- Joining tables
- Subqueries
- Common table expression(CTE)

Stakeholders requirements

- Which movies contributed the most/least to revenue gain?
- What was the average rental duration for all videos?
- Which countries are Rockbuster customers based in?
- Where are customers with a high lifetime value based?
- Do sales figures vary between geographic regions?

Approach to the problem

1.

Understanding the data

- Extracting entity relationship diagram and creating a first draft of a data dictionary.
- Answer some basic business questions using SQL.
- Creating EDA using SQL ordering, limiting, and grouping data

2.

Data Cleaning deriving more columns

- Identifying and namely duplicate, non-uniform, incorrect, and missing data and cleaning the data using a view
- Preparing data for analysis by profiling and cleaning the data.
- Deriving more additional columns using SQL group clause.

3.

Preparing for the business queries

- Creating a flat file for analysis using join queries.
- Writing subqueries to answer complex business questions.
- Using CTE to create complex queries.
- Evaluate the query performance and understand the DPA model.

4.

Storytelling

- Using flat data sheets from analytics queries upload to the tableau.
- Preparing visualisations using Tableau.
- Creating presentations of findings using Excel.

Conclusion

- Analysis of film data
  + Sports, sci-fi and animation are the top genres that made the highest films sales
  + Travel, Music and Thriller are the least purchased movies
  + The top 3 movies are Telegraph Voyage, Zorro Ark, Wife Turn
  + Most negligible revenue-generated films are Texas Watch, Oklahoma Jumanji, Duffe
- Rockbuster customer base in
  + India, China and US are the top 3 countries
  + Total number 108 countries where Rockbuster customers are present.
- Variation in global sales
  + Maximum sales by country is 6034.78, and minimum sale is 47.85. Average sales from all the countries are 567,00

ROCKBUSTER

- **Resource Links**
  - Reports
  - Exploratory datasheet
  - Storyboard

Marketing strategy for online grocery stores

Project duration: 1 month

- About the project
- Approach to the problem
- Stakeholders requirements
- Conclusion with visualisation

Marketing strategy for online grocery stores

Goal

Instacart is an online grocery store that operates through an app. Instacart already has very good sales but wants to uncover more information about its sales patterns. The data analyst's task is to perform an initial data and exploratory analysis of some of their data to derive insights and suggest strategies for better segmentation based on the provided criteria.

Data set

- Customers Data Set
- Data Dictionary

Skills

- Python(pandas, numpy, matplotlib.pyplot, seaborn)
- Jupyter anaconda
- Data wrangling
- Data merging
- Deriving variables
- Grouping data
- Aggregating data
- Reporting in Excel
- Population flows

Stakeholders requirements

- The sales team needs to know the busiest days of the week and hours of the day (i.e., the days and times with the most orders) to schedule ads when there are fewer orders.
- They also want to know whether people spend the most money at particular times of the day, which might inform the type of products they advertise.
- Instacart has many products with different price tags. Marketing and sales want specific price range groupings to help direct their efforts.
- Are there certain types of products that are more popular than others? The marketing and sales teams want to know which departments have the highest frequency of product orders.
- The marketing and sales teams are particularly interested in their system's different types of customers and how their ordering behaviours differ.

Approach to the problem

1. Understanding the data

- Perform basic descriptive statistical operations with pandas to draw early insights into data.
- Applying data wrangling on columns of the data source

2. Data Cleaning

- Finding and addressing mixed type variables in a dataframe.
- Finding and addressing missing values in a dataframe.
- Finding and addressing duplicate values in a dataframe.

3. Data Preparation

- Concatenating and merging data frames using join techniques.
- Deriving new columns from existing datasets to answer business questions.
- Using group and aggregate functions in pandas to answer complex business questions.

4. Storytelling

- Creating basic visualisations using Python libraries such as matplotlib.pyplot, seaborn .
- Consolidate results of Python analysis in a final report for stakeholders.

Conclusion

InstaCart

1. The busiest days for the ad campaign are Saturdays and Sundays.
2. Busy hours for the purchase orders are from 9 AM to 5 PM. During these hours, advertising/marketing is recommended.
3. To increase sales order counts, the high-price range products require an additional ad campaign/marketing.
4. Departments 4 and 16 have received the highest orders. The least purchase orders are received in the department 10,11,2, 21 and 8

- **Resource Links**
  - Reports
  - Exploratory datasheet
  - Storyboard

Thank You

+Muhammad Abdulkayyum

+Muttaki.mak@gmail.com

+github.com