



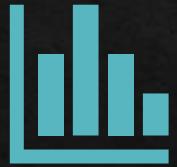
Mnguni Zulu  
Data Analyst  
[Case Studies]

# Video Game Sales



## Tools

Excel was the main tool used in the cleaning, analysis and visualization of data



## Data

- 1980 – 2016
- 16,000 + rows
- 3 Countries



## Recommendations

- Visualisations
- Power Point Presentation

# Introduction

- ❖ Goals: The goal of this project was to perform descriptive analysis of historical data of video game sales across numerous platforms and generate insights which could be used for recommendations for GameCo (client).
- ❖ The dataset was sourced from VGCHartz and was limited to the years 1980 to 2016. The dataset had about 16,000 rows
- ❖ The cleaning, analysis and visual analysis of data was performed in Excel.

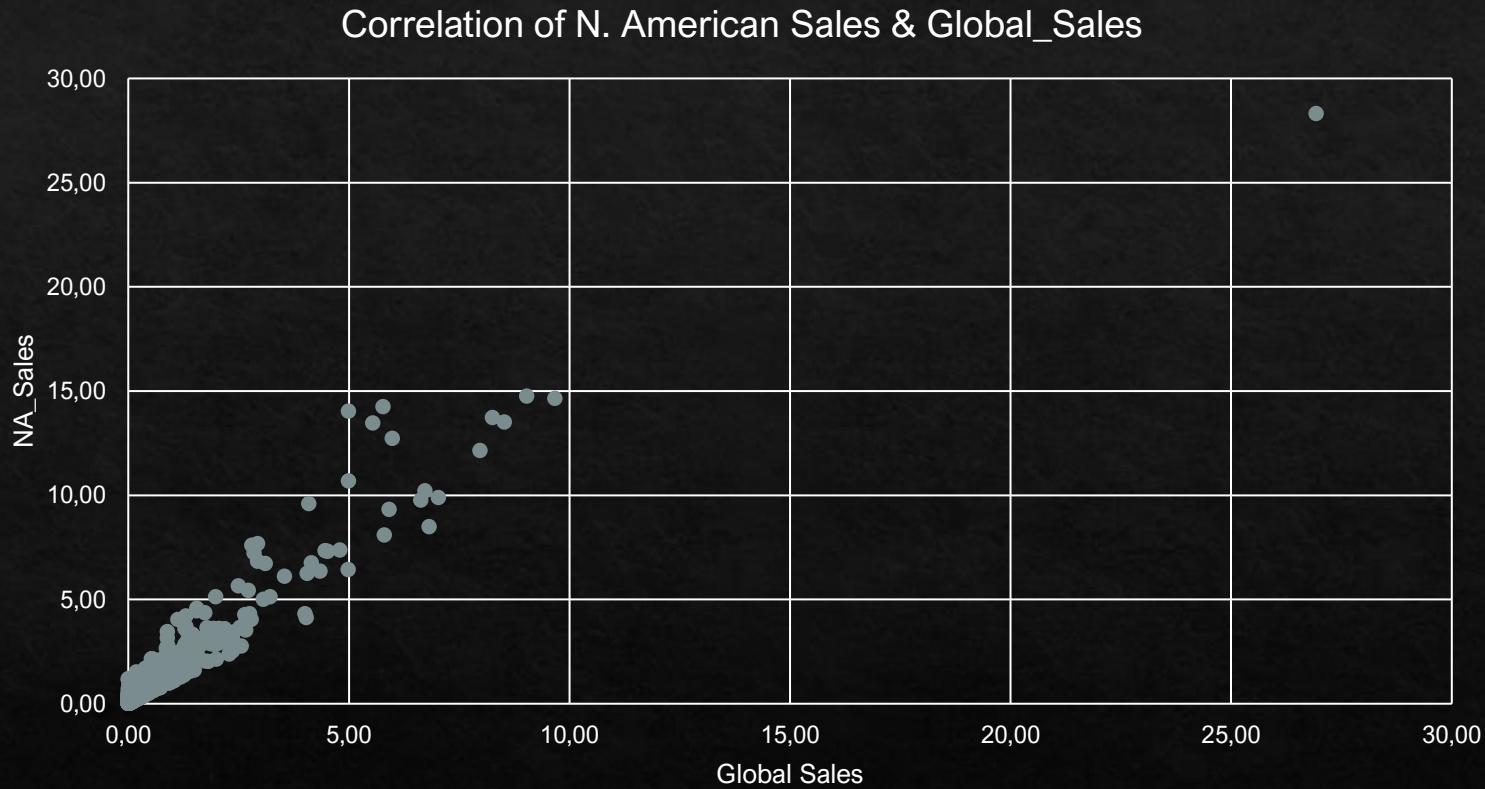
# Analysis

- ❖ The dataset had missing data, irrelevant data, inconsistencies and more, which had to be identified and dealt with in excel.
- ❖ Filters were used to identify inconsistencies, the Replace function was used to correct inconsistencies and missing values, where applicable.
- ❖ Pivot tables were used to create new groupings and reveal invaluable insights for video games per country, genre and platform

# Conclusion

- ❖ Recommendations were made to GameCo with regard to trends in historical data analysis.
- ❖ Fastest growing markets, were identified as well as which genres were the largest part of that growth and also which platforms they were played on.
- ❖ This was communicated in the form of a Powerpoint presentation with various charts to represent the data and findings in a visual manner.

# Visualisation Example



The visualisations shows that sales in North America and global sales are strongly positively correlated, highlighting the importance of a game's success in this market.

# Upcoming Influenza Season



## Tools

- Excel for data manipulation
- Excel for statistical analysis
- Tableau for visual analysis



## Data

- US Census
- Influenza Visits
- Influenza Laboratory Tests



## Recommendations

- Visualisations in Tableau
- Video presentation of results
- Tableau dashboard

# Introduction

- ❖ Goal: The aim objective was to use historical data to uncover trends which would allow a medical staffing agency to predict the required deployment of first staff across the USA.
- ❖ The dataset were sourced from publicly available information generated by official US agencies.
- ❖ A hypothesis linking age to risk of hospitalisation due to Influenza like illness was created and tested using regression in Excel

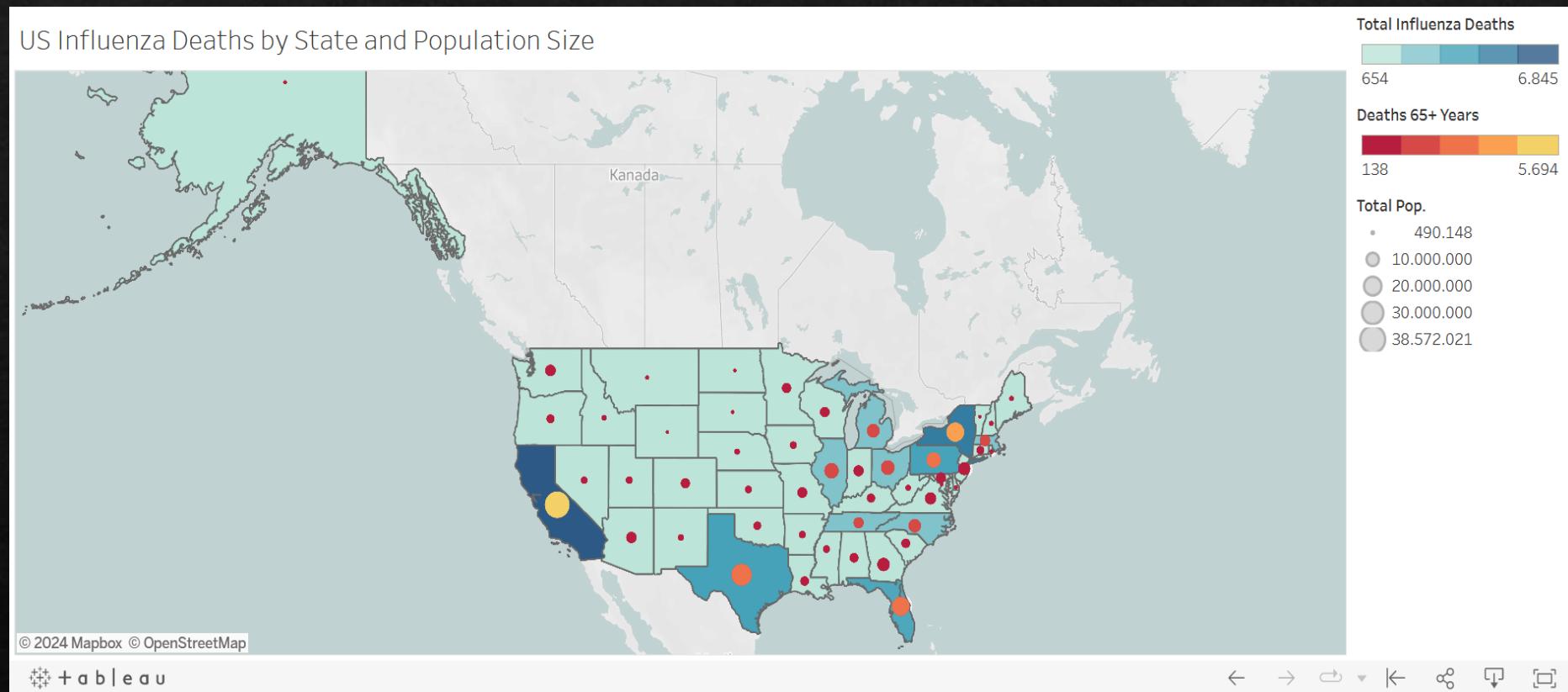
# Analysis

- ❖ Decisions had to be made regarding which datasets to accept or reject for the purpose of the analysis. To decide this, I had to understand the business requirements and develop several hypotheses which would perhaps be tested at the later stage. These were formulated with the help of stakeholder input.
- ❖ I created a plan and schedule for how the analysis project would proceed, and how and when the stakeholders would be able to provide input and when and how results would be communicated.
- ❖ The dataset had missing values, inconsistencies which had to be resolved.
- ❖ Data from different sources also had to be joined to create one dataset to subject to the necessary data manipulations and analysis.
- ❖ Temporal and spatial analysis were conducted using Tableau

# Conclusion

- ❖ After the statistical analysis found that there was a strong link between individuals in age groups 65+ and influenza mortality, a visual analysis was performed.
- ❖ The visual analysis was performed in Tableau
- ❖ The states with the largest at risk population were identified, as well the times of the year that they are most vulnerable.
- ❖ An interactive dashboard and storyboard were created in Tableau to present the data findings and communicate recommendations
- ❖ Video Presentation Link:  
<https://vimeo.com/881139281/f10b3261f6?share=copy>
- ❖ Tableau Link:  
[https://public.tableau.com/views/2\\_9\\_PreparingfortheUpcomingInfluenzaSeason\\_Revised/Story1?:language=de-DE&:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/views/2_9_PreparingfortheUpcomingInfluenzaSeason_Revised/Story1?:language=de-DE&:display_count=n&:origin=viz_share_link)

# Visualisation Example



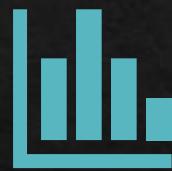
This is a spatial analysis highlights the link between general population sizes, populations of 65+ years age group and influenza deaths

# Rockbuster Stealth



## Tools

- PostgreSQL database
- SQL for queries
- Database Visualiser



## Data

- Relational Database
- 2 Fact Tables
- 13 Dimension Tables



## Objectives

- Important Business Questions answered
- Key marketing questions answered
- Powerpoint of key data and visualisations

# Introduction

- ❖ The objective was to answer key business questions relating to marketing and operations for a fictional global movie rental business
- ❖ The business has been struggling since the rise of streaming services like Netflix
- ❖ I used PostgreSQL for the relational database management software which is SQL compliant.

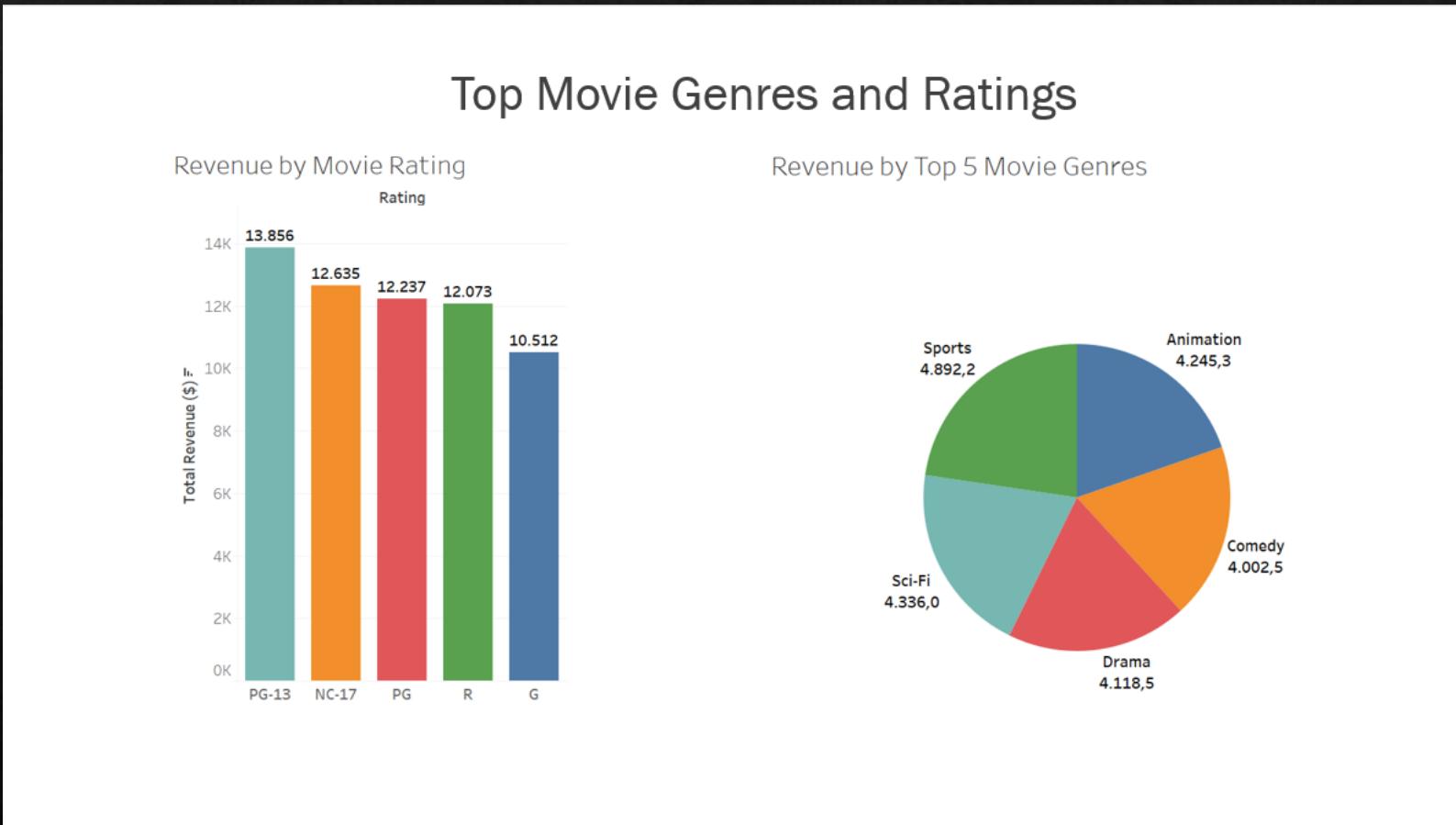
# Analysis

- ❖ I used SQL to create new tables, insert, delete, modify, replace data in tables.
- ❖ I grouped, sorted, filtered and retrieved data from the RDMS to answer specific business questions.
- ❖ I also used SQL to generate summary statistics for numerical data.
- ❖ I utilised subqueries and Common Table Expressions to answer the most complex questions.
- ❖ I utilised keys for different types of joins to generate new tables.

# Conclusion

- ❖ As outputs, I generated a data dictionary and power point presentation presenting the information I had retrieved from the Database.
- ❖ The presentation answered key marketing and business questions and provided recommendations on the best way forward for the Rockbuster Stealth.
- ❖ Important generated tables were exported as csv files and then used to create visualisations in Tableau
- ❖ Github Repo: [https://github.com/MetaMongoose/sql\\_rockbuster](https://github.com/MetaMongoose/sql_rockbuster)

# Visualisation Example



The queries were used to create tables, which in turn were used to create visualisations to present valuable insights into revenue by genre and age rating.

# Instacart



## Tools

- Python for data manipulation
- Compiled in Jupyter Notebook
- Pandas, Numpy, Seaborn and Matplotlib libraries



## Data

- +32 Million rows
- Order, Customer and Product data
- Visualisations



## Recommendations

- Visualisations with Python
- Presentation of Notebooks

# Introduction

- ❖ Instacart is a real online store which makes some of its data publicly available. Customer data was synthesised for the purpose of this project.
- ❖ The objective of this project was to answer complex business questions for the marketing and sales teams at Instacart.
- ❖ Because of the large data size, it was most appropriate to use Python to perform data manipulation like: identifying missing values and inconsistencies. As well as perform merging of data, creation of new calculated variables and visualisations

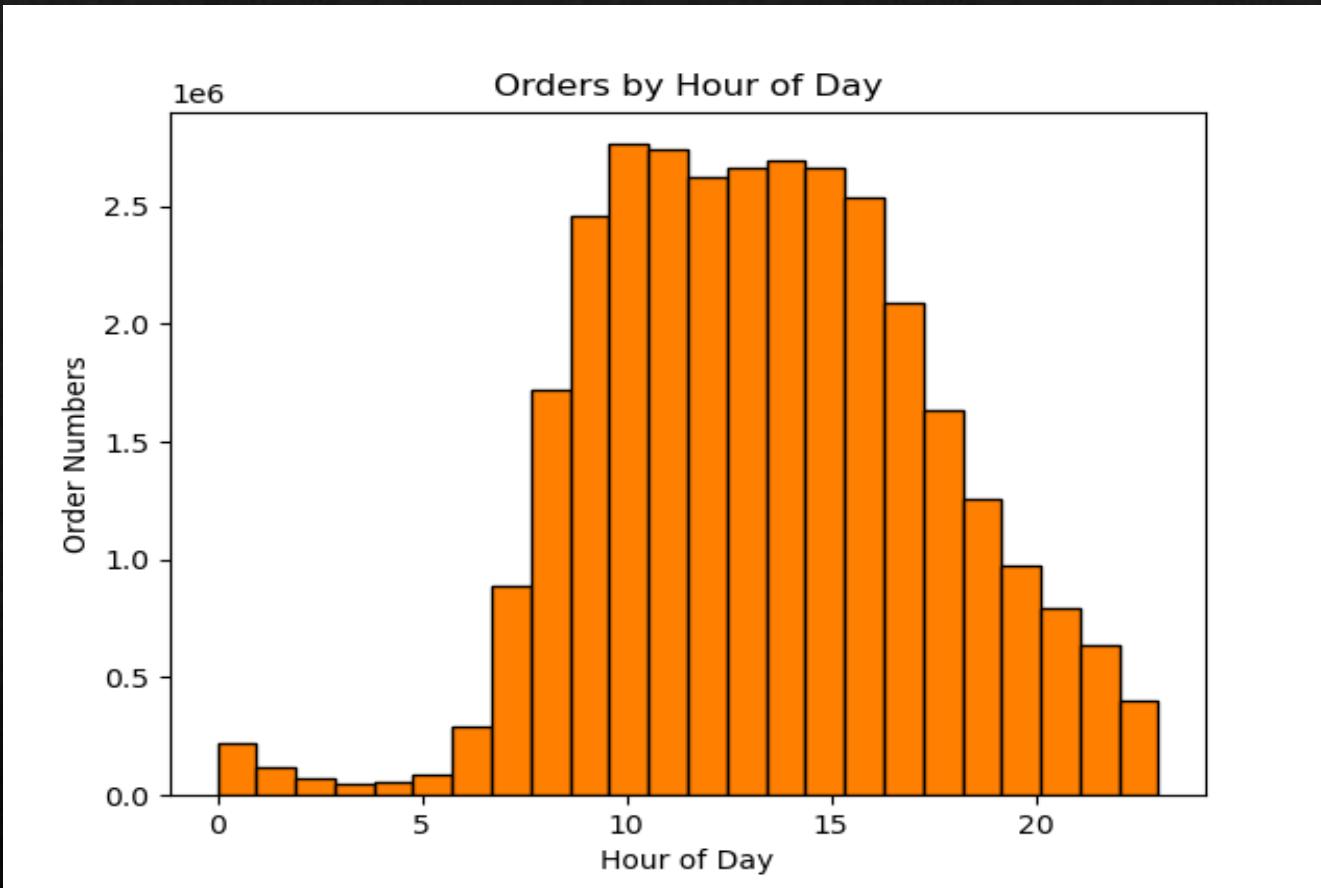
# Analysis

- ❖ Jupyter was used to compile the code into neat and manageable scripts.
- ❖ Use was made of libraries within Python, such as seaborn, pandas, numpy and matplotlib
- ❖ The datasets were manipulated to remove sensitive customer data for privacy reasons. Also other columns found to be irrelevant for the analysis were either deleted all together or omitted from the new versions of the dataframe.
- ❖ Order, customer and product dataframes were joined, with consideration given to the most suitable type of join at each stage.
- ❖ Descriptive statistics were generated from the numerical values and visualisations. Both for the puropse of generating insights.

# Conclusion

- ❖ Visualisations and the recommendations that accompanied them were communicated with help of a report in an excel sheet.
- ❖ A data flow diagram was made to assist in making the steps of the analysis transparent to stakeholders. This flow diagram showed major changes made to the dataframes at different stages, as well as how the datframe changed in size and shape.
- ❖ Recommendations were made in response to the business questions made.
- ❖ The analysis focused on historical data.

# Visualisation Example



This chart was generated to visualise the frequency of orders by hour of day.