DATA ANALYTICS

Case studies Portfolio
By: Azeez Babarinde

GAMECO

Analyzing Video Game Popularity

PROJECT

Company

GameCo is a fictional new video game company that wants to use data to inform the development of new games.

Objective

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 used was drawn from VGChartz (Video Game Charts), an industry research firm. The video game sales record is from 1980-2016.

OVERVIEW

Key questions

How have the sales figures varied between geographic regions over time?

What is the regional percentage of global sales over the years?

What is the proportion of all time sales by genre in each region?

Skills/Tools

- Excel
- Grouping data
- Summarizing data
- Descriptive analysis
- Visualizing results in Excel
- Presenting results

ANALYSIS

Steps

Clean the dataset in Excel

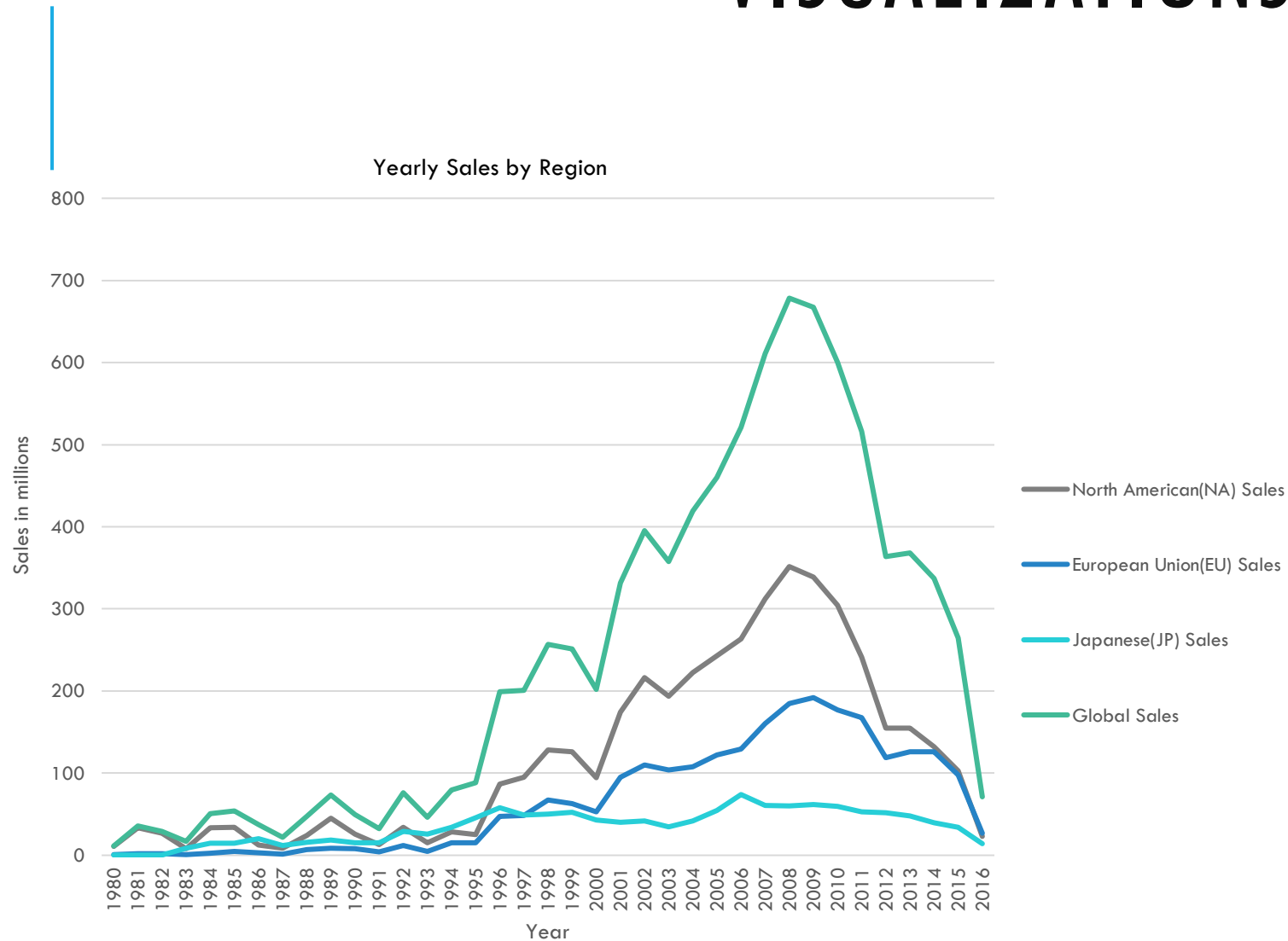Use pivot tables to glean insights

Group data so similar items can be analyzed together

Filter data to look at subcategories

Respond to hypothetical challenges facing GameCo, and choose an appropriate analysis

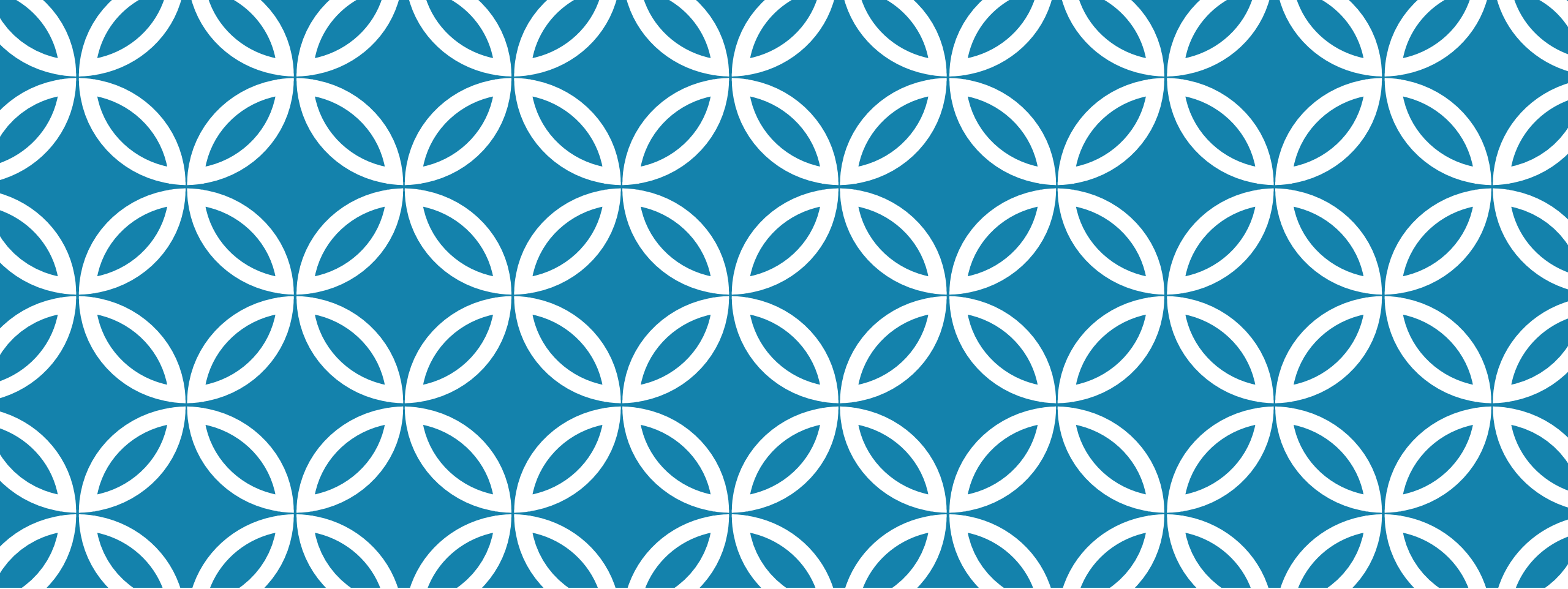Create a PowerPoint presentation that tells a compelling story

VISUALIZATIONS

Yearly Sales by Region

Sales in millions (y-axis: 0, 100, 200, 300, 400, 500, 600, 700, 800)
Year (x-axis: 1980–2016)

Legend:
- North American(NA) Sales
- European Union(EU) Sales
- Japanese(JP) Sales
- Global Sales

The Executives current assumption of the market is that game sales for the various geographic regions have stayed the same over time

Insight

- Our findings tells us that the market have been changing over the years and there are also new business opportunities.
- EU market have been the most consistent region by sales growth from time and have become the leader since 2015

PREPARING FOR FLU SEASON IN THE U.S

Medical Staffing Agency

PROJECT

Company

A fictional medical staffing agency want to plan for the upcoming influenza season by providing temporary workers to clinics and hospitals across the United States on an as-needed basis.

Objective

To determine when to send staff, and how many, to each state.

Data Set

The data used for this analysis is influenza death data from the CDC, and population data from the U.S. Census Bureau.

OVERVIEW

Key Questions

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

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

Skills/Tools

- Excel
- Translating business requirements
- Data cleaning, integration and transformation
- Statistical hypothesis testing
- Visual analysis
- Forecasting
- Storytelling in Tableau
- Presenting results to an audience

ANALYSIS

Steps

Source, profile, and clean the data sets

Integrate the two data sets using Excel

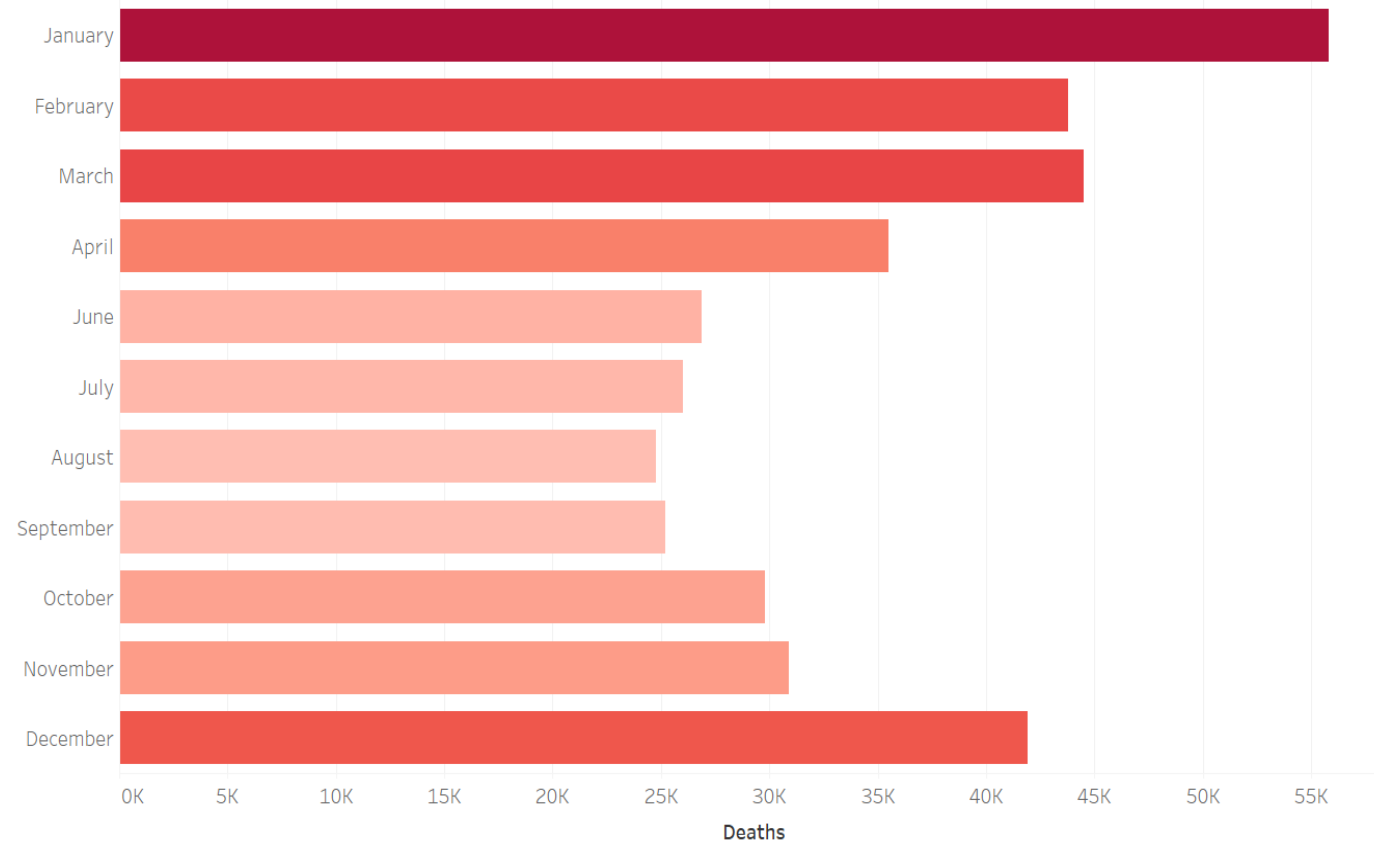Calculate summary statistics and identify any correlations

Create an interim report

Use Tableau visualizations to explore the data

Create a Tableau interactive story board with staffing recommendations

VISUALIZATIONS

When is the Flu Season?

This chart show us the influenza death by month, it tells us that influenza occurs seasonally and it starts and end in different months

Link to Tableau presentation

ROCKBUSTER STEALTH LLC

Answering business questions for an online video rental company

PROJECT

Company

Rockbuster Stealth LLC is a fictional movie rental company that used to have stores around the world.

Objective

I have been hired by Rockbuster Stealth's business intelligence (BI) department to help with the launch strategy for the new online video service.

Data set

Data set that contains information about Rockbuster's film inventory, customers, and payments. It is provided by Careerfoundry.

OVERVIEW

Key questions

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?

Skills/Tools

- Relational databases
- SQL
- Database querying
- Filtering
- Cleaning and summarizing
- Joining tables
- Subqueries
- Common table expressions

ANALYSIS

Steps

Fix consistency issues in the data

Explore the data using SQL queries and answer ad-hoc business questions

Prepare an Excel file with important SQL queries/outputs for technical colleagues

Create a data dictionary in Word documents

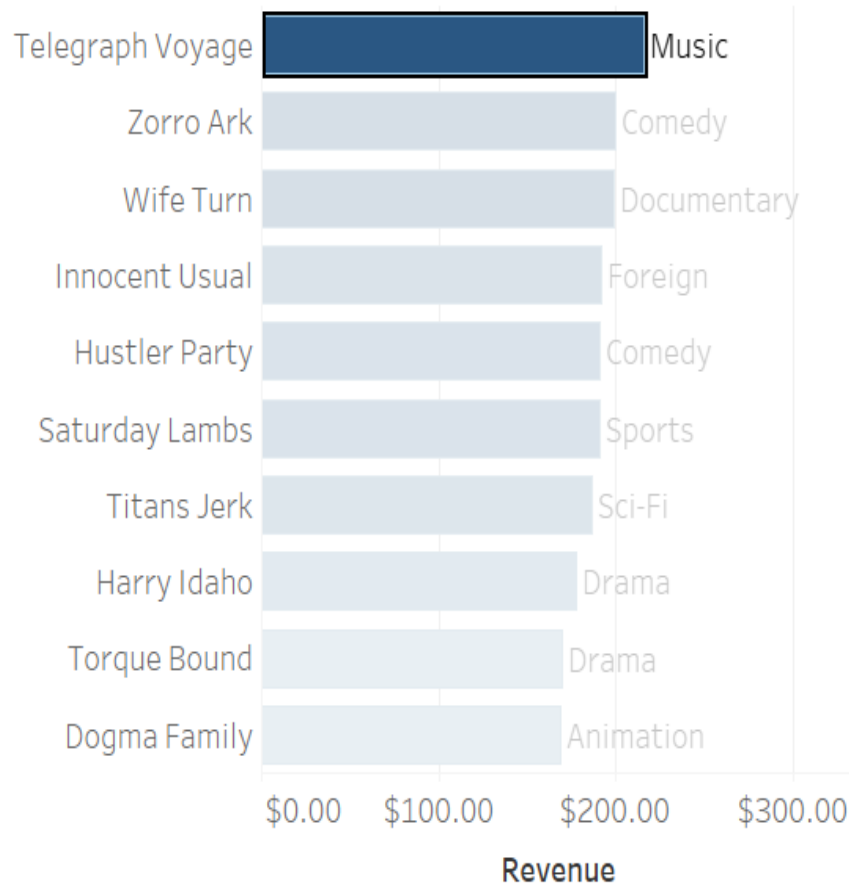Create a presentation of findings

GitHub Repository for the SQL Queries

ANALYSIS: SQL QUERY

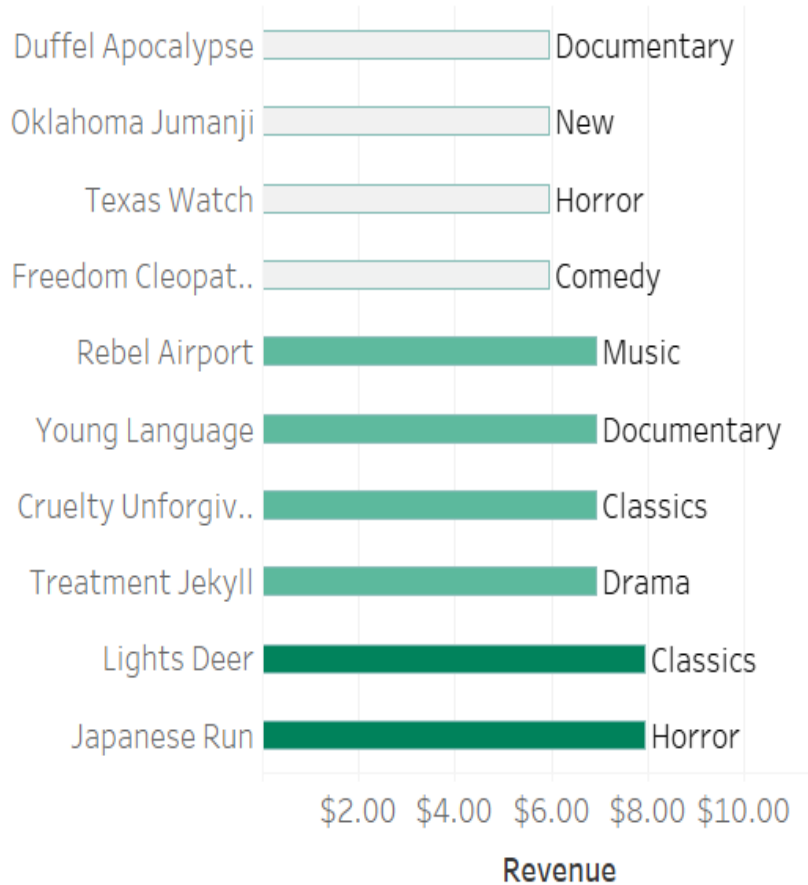**#SQL query showing Which movies contributed the most to revenue gain**

SELECT D.title, F.name AS genre, COUNT(D.film_id), SUM(amount) AS total_payment

FROM payment A

INNER JOIN rental B ON A.rental_id = B.rental_id

INNER JOIN inventory C ON B.inventory_id = C.inventory_id

INNER JOIN film D ON C.film_id = D.film_id

INNER JOIN film_category E ON D.film_id=E.film_id

INNER JOIN category F ON E.category_id=F.category_id

GROUP BY D.film_id, F.name

ORDER BY total_payment DESC

LIMIT 10

VISUALIZATIONS

Highest Earning (2016)

Title	Revenue	Genre
Telegraph Voyage		Music
Zorro Ark		Comedy
Wife Turn		Documentary
Innocent Usual		Foreign
Hustler Party		Comedy
Saturday Lambs		Sports
Titans Jerk		Sci-Fi
Harry Idaho		Drama
Torque Bound		Drama
Dogma Family		Animation

Lowest Earning (2016)

Title	Revenue	Genre
Duffel Apocalypse		Documentary
Oklahoma Jumanji		New
Texas Watch		Horror
Freedom Cleopat..		Comedy
Rebel Airport		Music
Young Language		Documentary
Cruelty Unforgiv..		Classics
Treatment Jekyll		Drama
Lights Deer		Classics
Japanese Run		Horror

The charts showing the lowest and highest earning movies and its genre.

Insight
Rockbuster should launch first the genre of the top earning movies and they should launch first in the China, India and USA because they have the highest number of customer and generates most revenue

Link to Tableau presentation

INSTACART

Marketing strategy for an online grocery store

PROJECT

Company

Instacart, an online grocery store that operates through an app.

Objective

To uncover more information about their sales patterns. I will perform exploratory analysis to derive insights and suggest strategies for better segmentation based on the provided criteria.

Data set

The data set is open source, it contains Customers Data Set and Data Dictionary.

Link to data set

OVERVIEW

Key questions

- what the busiest days of the week and hours of the day?
- What's the distribution among users in regards to their brand loyalty?
- Are there differences in ordering habits based on a customer's loyalty status?
- Are there differences in ordering habits based on a customer's region?
- Is there a connection between age and family status in terms of ordering habits?

Skills/ Tools

- Python
- Data wrangling
- Data merging
- Deriving variables
- Grouping data
- Aggregating data
- Reporting in Excel
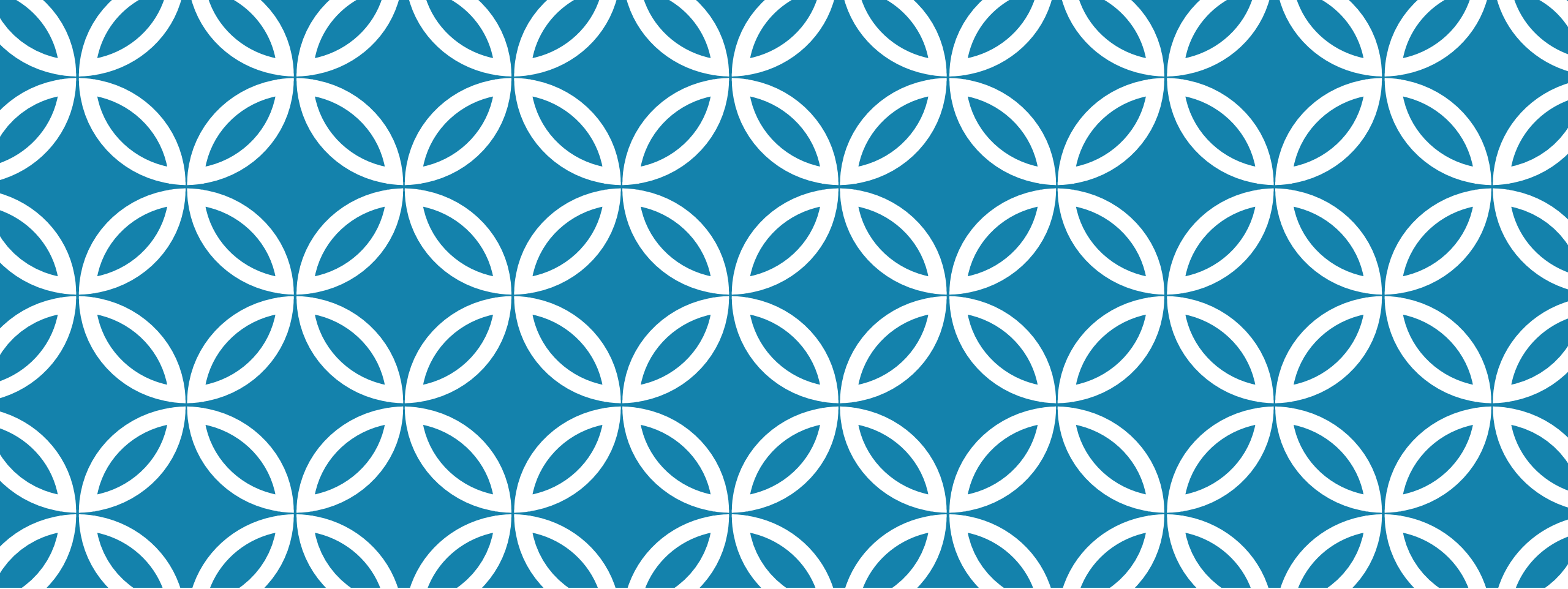- Population flows

ANALYSIS

Steps

- Organize Python project folders to industry standards
- Clean, prepare, and merge the data
- Create customer profiles based on age, marital status, and number of dependents
- Derive new columns and flag the data by different customer profiles
- Create visualizations in Python and in Excel
- Summarize findings, visualizations, and the analysis process in an Excel report

VISUALIZATIONS
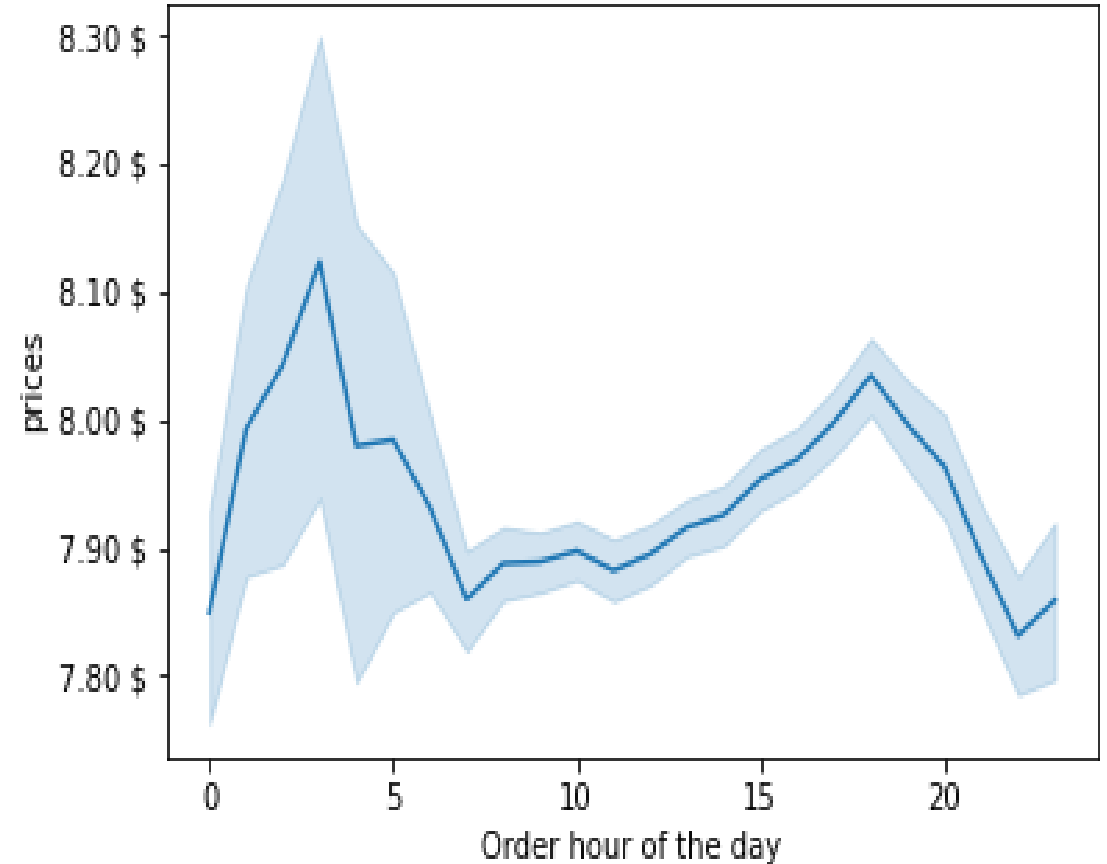
Insight: Instacart should make more staffs available on Saturdays and Sundays, while they should reduce them on Tuesday and Wednesday, the peak hours should also be noted,this will help reduce cost.

Link to report on Github

PIG E. BANK

Anti-money laundering projects at a global bank

PROJECT

Company
Pig E. Bank is a fictional well-known global bank.

Objective
To develop models that detect suspicious account activity associated with money laundering.

Data set
The source of the data used for this analysis is provided by Career Foundry

OVERVIEW

Skills/ Tools

- Big data
- Data ethics
- Data mining
- Predictive analysis
- Time series analysis and forecasting
- Using GitHub

Steps

- Clean and prepare the Excel data, with thorough documentation
- Use descriptive statistics to explore the data
- Identify ethical dilemmas in hypothetical business scenarios
- Create a decision tree model to identify clients who are likely to leave the bank

VISUALIZATIONS
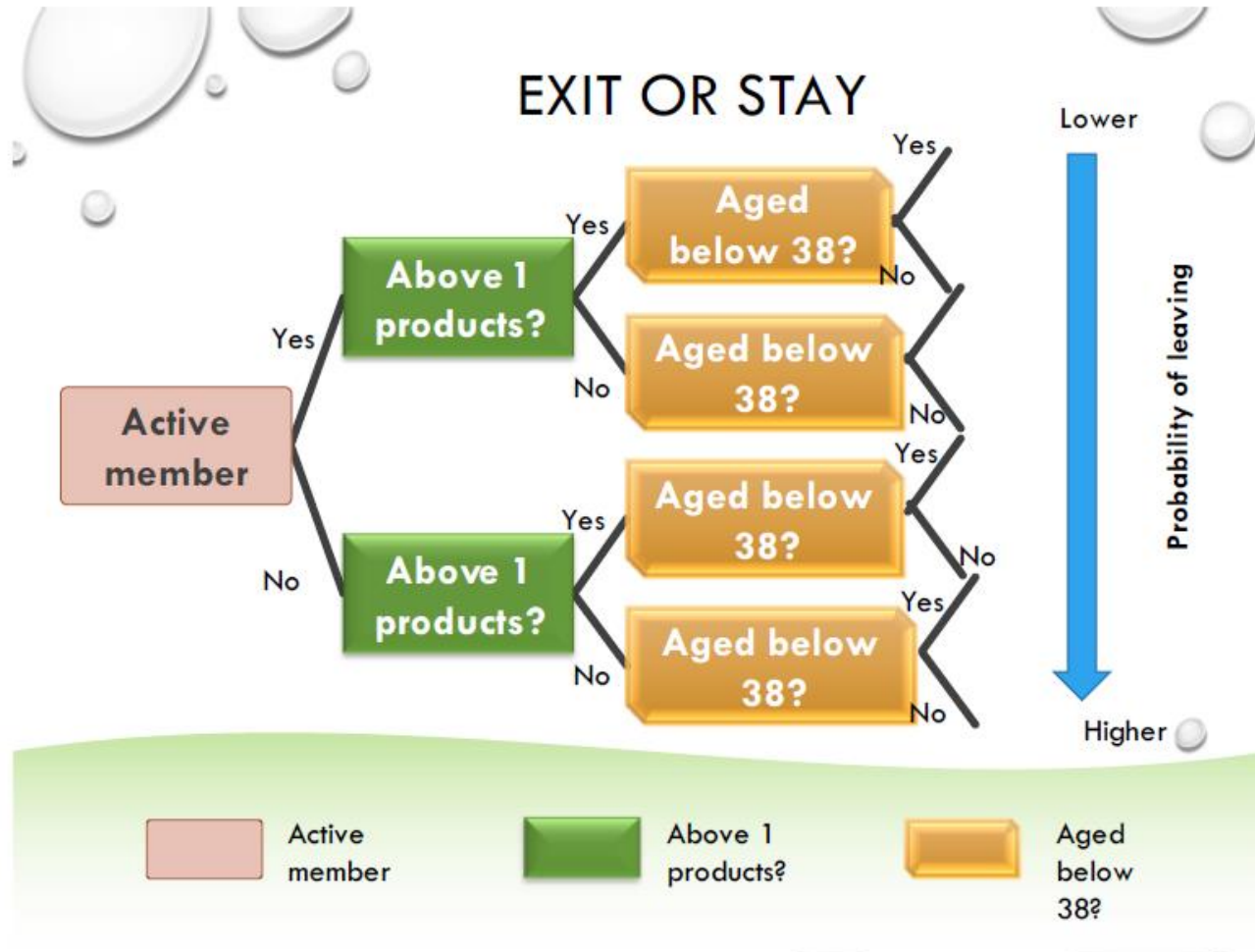
The factors that contributes to clients leaving are:- age, active status, number of products.
This decision tree help to identify those that are more likely to leave .

EA SPORTS FIFA PLAYER RATINGS

Open exploratory project

PROJECT

Goal

To build an interactive dashboard that will visually showcase well-curated results of EA Sports FIFA Player rating analysis conducted in Python.

Objective

In this case study, we will be exploring what determines the FIFA player ratings.

Data

EA sports collected FIFA player rating data through a network of EA data reviewers (players, coaches. etc.); these reviewed details were then handled and finalized by the editors.

Link to Dataset

OVERVIEW

Key questions

- Does high player wages lead to high ratings?
- Does player market value affect their ratings?
- Is there correlation between player ages and their ratings ?

Skills/ Tools

- Excel
- Python
- Machine learning
- Exploratory Analysis
- Data sourcing, cleaning, integration and transformation
- Hypothesis testing
- Tableau Presentation

ANALYSIS

Steps

- Source, clean and prepare data
- Create a list of questions to explore
- Answer and research the questions using python visualization, excel and statistical methods
- Derive new column and merge with existing data
- Create a story board in Tableau and answering my research question with it

Link to Tableau Presentation

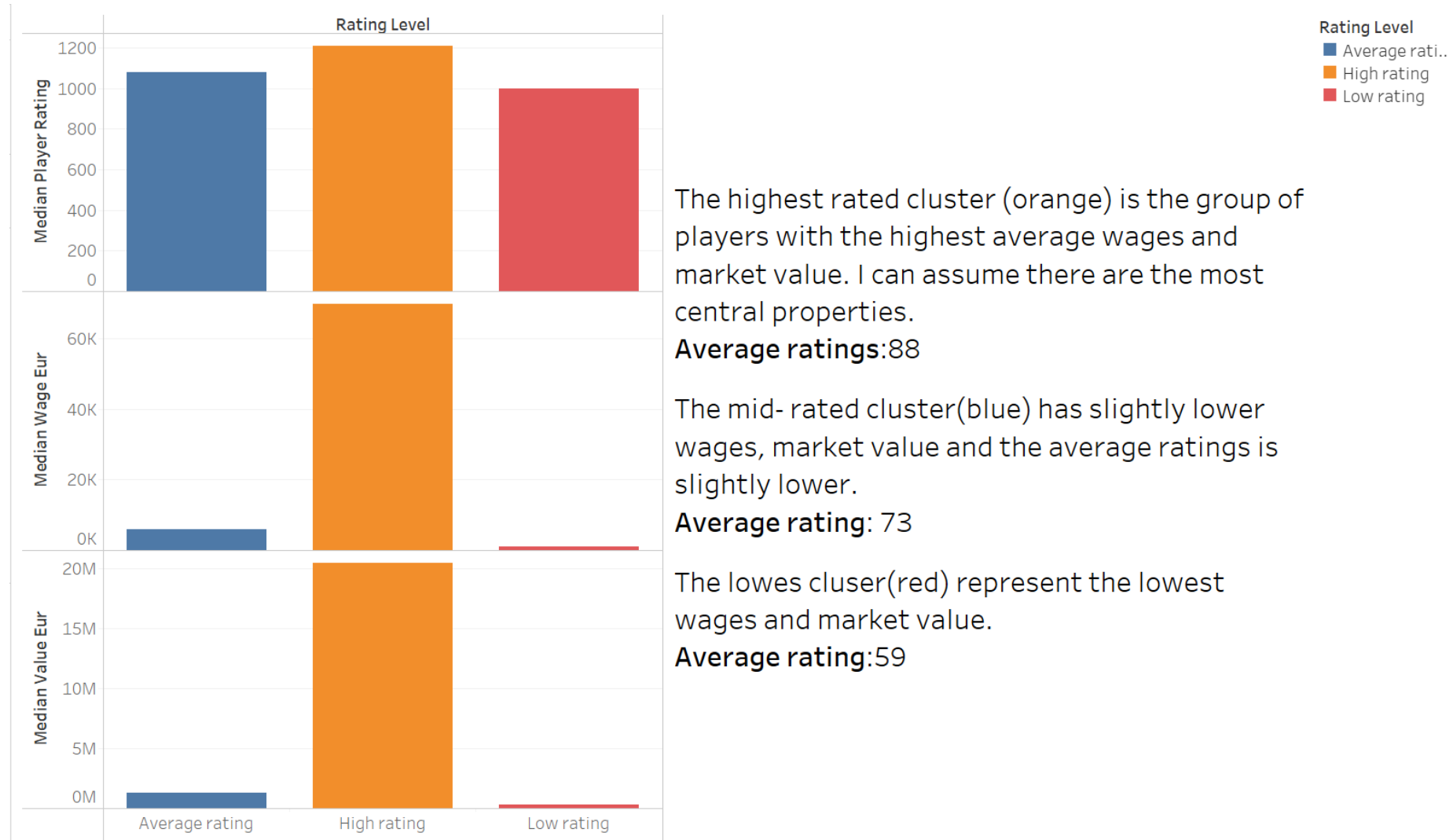Link to GitHub repository

VISUALIZATIONS
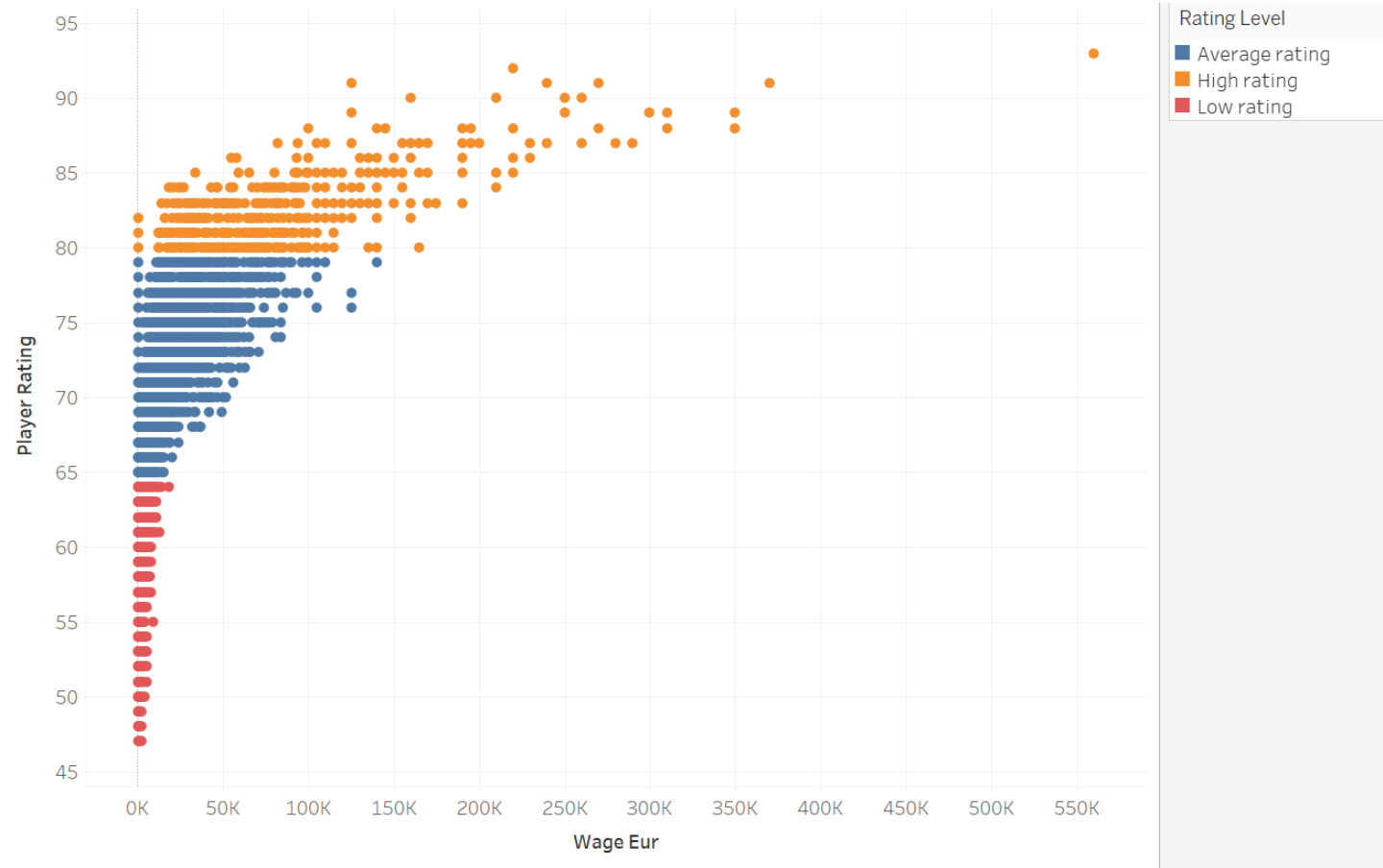
This chart shows that an increase in a player's wages will lead to a rise in their ratings. It also shows them in a group by rating level.

The color code represent
- Blue -Average ratings
- Orange- High ratings
- Red- Low ratings

VISUALIZATIONS

Rating Level
- Average rati..
- High rating
- Low rating

The highest rated cluster (orange) is the group of players with the highest average wages and market value. I can assume there are the most central properties.
Average ratings:88

The mid- rated cluster(blue) has slightly lower wages, market value and the average ratings is slightly lower.
Average rating: 73

The lowes cluser(red) represent the lowest wages and market value.
Average rating:59

INSIGHTS AND CONCLUSION

When choosing the rating for a FIFA player, the following player's data should be considered:

- Players with the highest wages and market value are given higher ratings.
- Players with slightly lower wages and market value have slightly lower ratings.
- Players with the lowest wages and lowest market value have the lowest ratings.