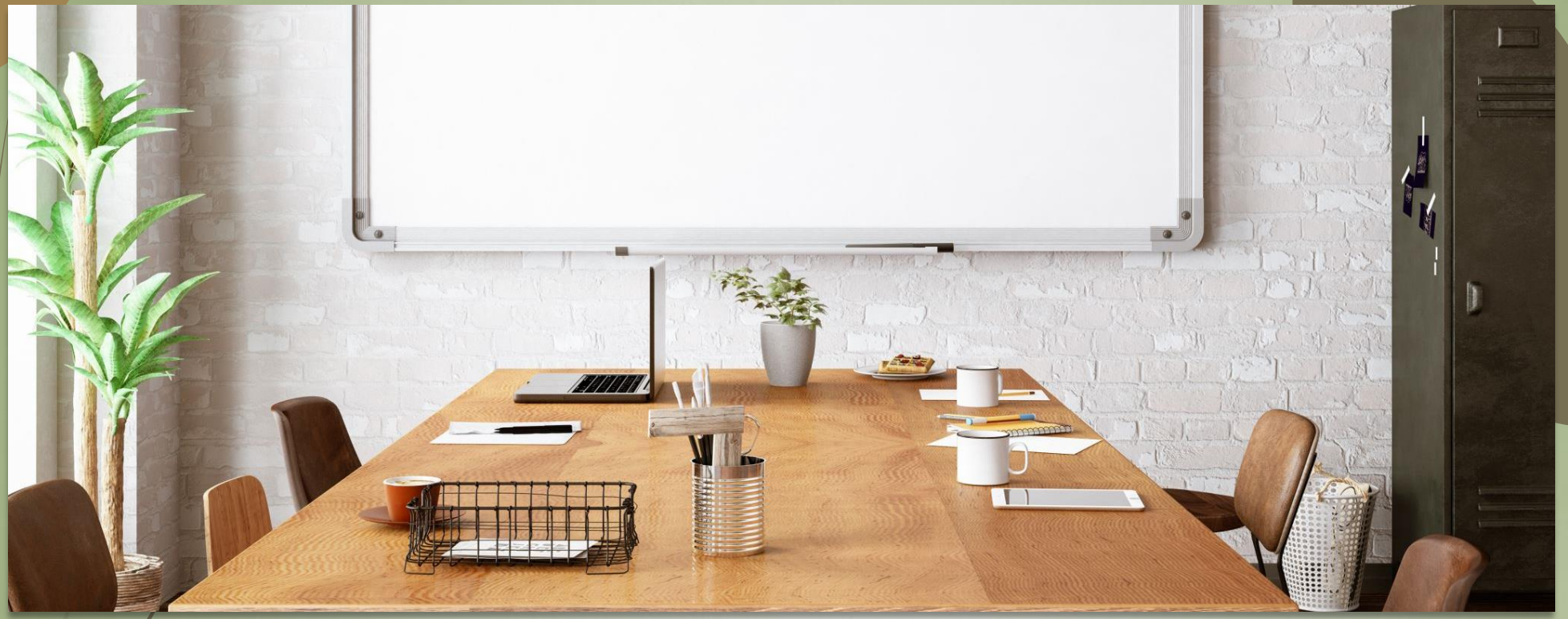


Presentation title

Daniel Locke

PROJECTS

I. GameCo Regional Sales
II. Medical Staffing Plan: Influenza Season
III. Product Analysis: Rockbuster
IV. Customer Purchase Analysis: Instacart
V. Analysis for Anti-Money Laundering: Pig E. Bank

MARKET ANALYSIS REGIONAL SALES, GameCo

Analyzing Global sales for a fictitious company

Objective

GameCo wants to use data to inform the development of new games. As such, you'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.

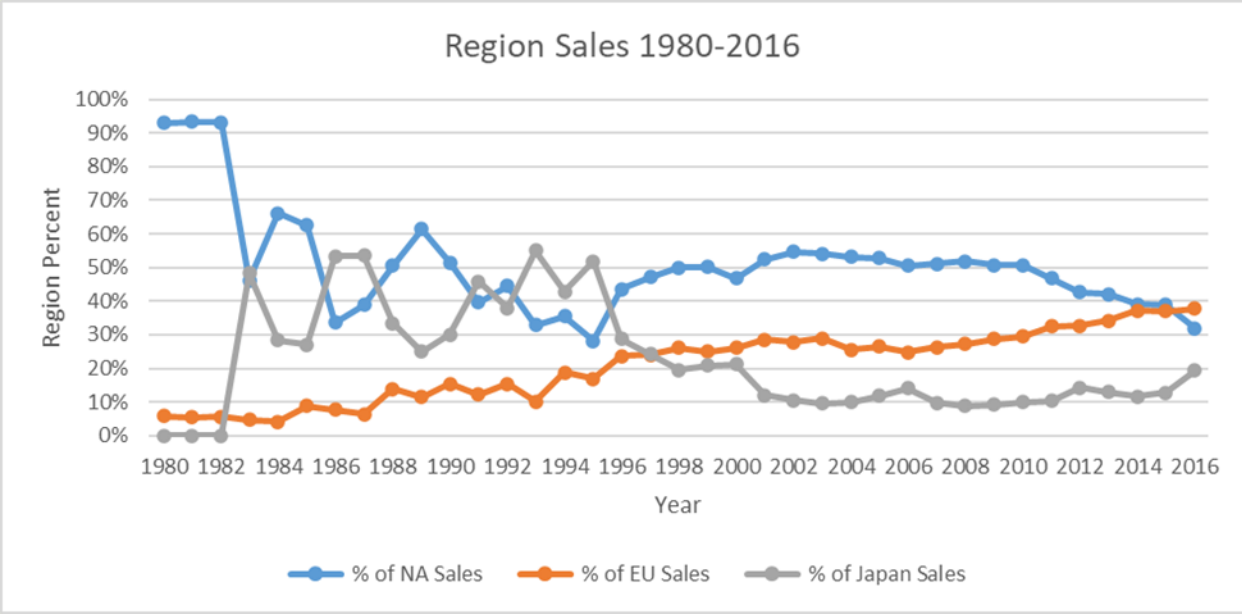
Data

Historical sales of video games (for games that sold more than 100,000 copies) spanning different platforms, genres, and publishing studios. This data was drawn from the website VGChartz

Tools

I. Excel
II. Summarizing
III. Grouping Data
IV. Visualizing Data in Excel
V. Presenting Results

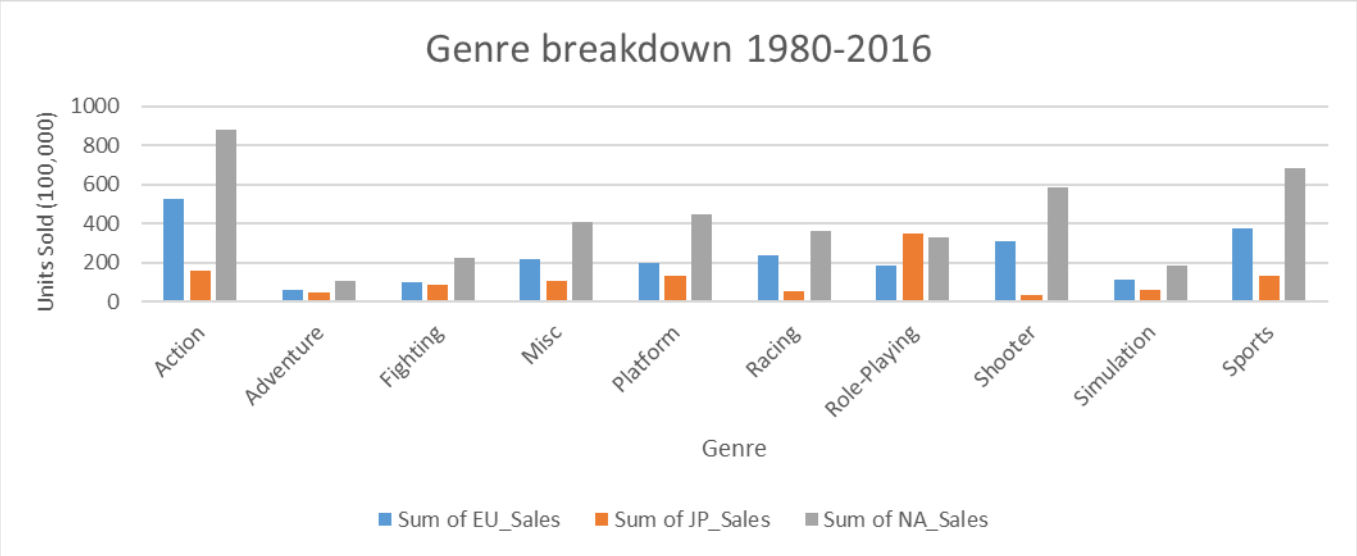
REGIONAL SALES 1980-2016



i. Action, Sports & Shooter were the top 3 genres over the time period for NA & EU.

ii. Role Playing, Action and Sports were the top 3 for JP.

BREAKDOWN OF UNITS SOLD BY GENRE IN EACH REGION 1980-2016

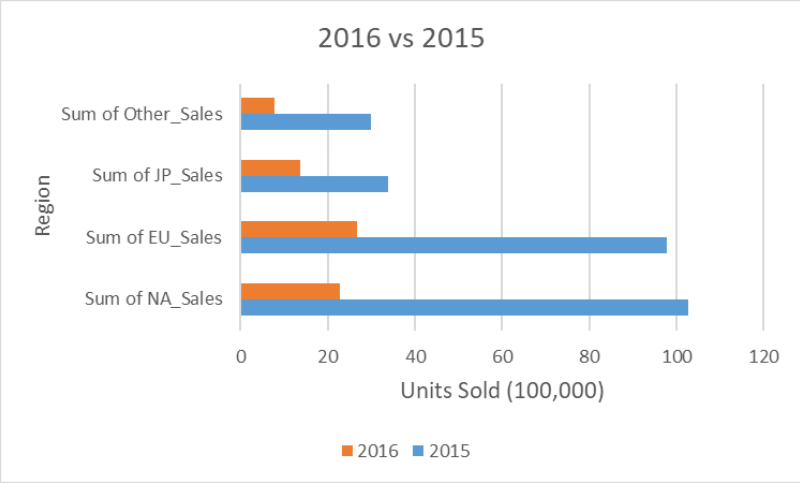


I. NA Sales peaked in 1980-1982, then dropped significantly. NA averaged 50% of Global Sales. They remained a top competitor in the years 1996-2015.

II. EU did not see double digit growth till 1988. The first 8 years averaged 8% of Global Sales. During the whole timeframe they averaged 25% of Global Sales.

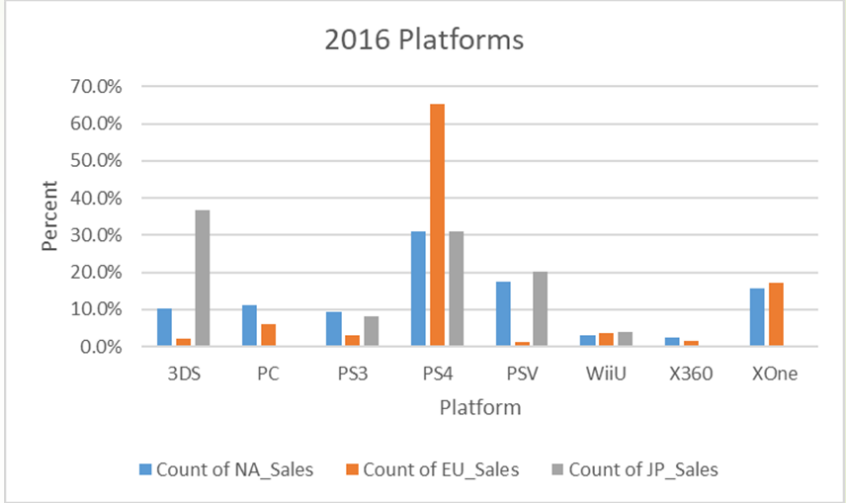
III. JP does not get on the leaderboard till 1983.They peak in 1993 at 55% and then have a couple of good years before seeing a downward trend. They averaged 19% of Global Sales

ANALYSIS



A dramatic drop in units sold compared to the prior year. Before Marketing $ are spent, worth investigating why.

A deeper look into World events, demographics, and the games that were released is warranted.



Based on the data PS4 is leading the platform in 2016 for both EU & NA.

JP's 3DS is their top performing followed by PS4.

RECOMMENDATIONS

Sales and Genres fluctuate year over year, there are many variables to consider as to why this happens: demographics, world events, prior year marketing dollars, on-hand inventory levels to name a few. Reviewing the history for the last 36 years has shown the uncertainty of the numbers. Focusing on 2016 alone is a good foundation on which to base the current trend and use that year to build upon for 2017.

The Adventure & Action Genre's can be combined into one Genre, giving a larger audience to target; by combining the two, you will get the best result for the money spent; because these two genres are so similar. Then focus on Sports, Shooter, and Role-Playing.

The top Platform in 2016 is PS4 followed by Nintendo 3DS. Action, Sports Shooter, and Role-Playing are the top 4 Genres across all 3 Regions. Based on the data, it would be worth investing a considerable amount of marketing dollars into Action, Sports & Shooters in the EU and NA Regions. The JP Region should focus on Shooter games first before moving onto the Action Genre.

MEDICAL STAFFING PLAN

Influenza Season

Objective

Plan for the influenza season, a time when additional staff are in high demand. The results will examine trends in influenza and how they can be used to proactively plan for staffing needs across the country.

Data

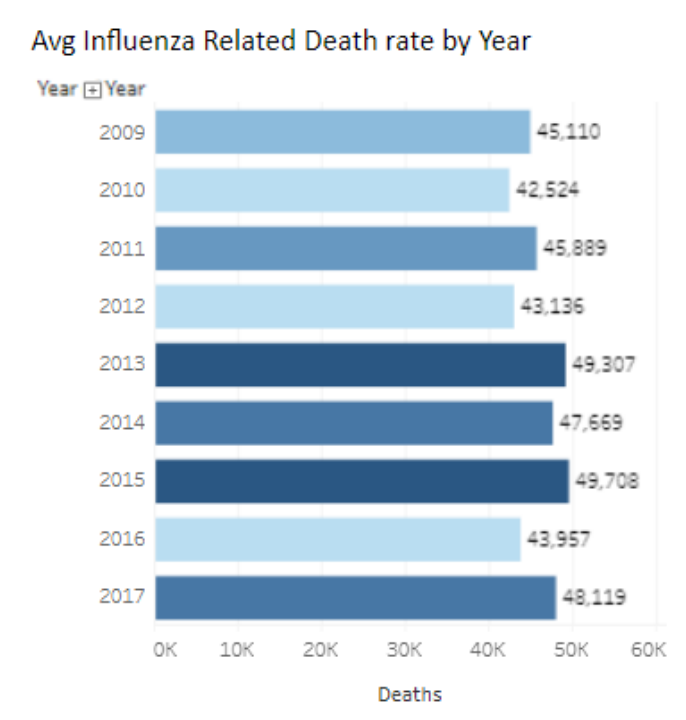
Population data 2009-2017
US Census Bureau

Influenza Deaths by Geography, Time, Age & Gender (CDC)

Tools

I. Excel
II. Summarizing
III. Cleaning Data
IV. Data Transformation
V. Storytelling in Tableau
VI. Forecasting
VII. Presenting Results

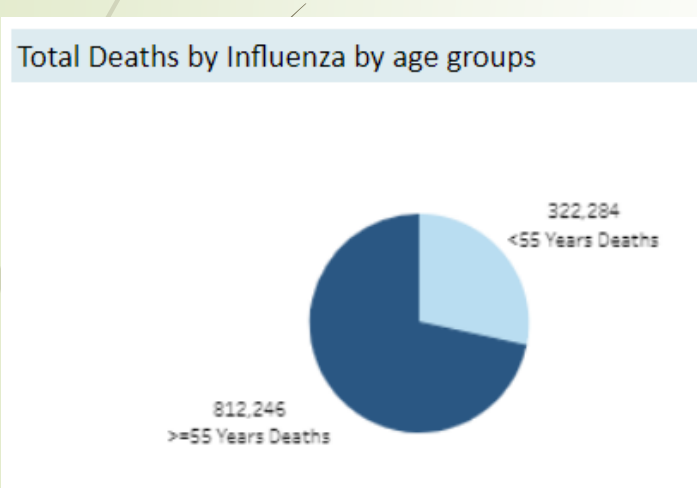
ANALYSIS

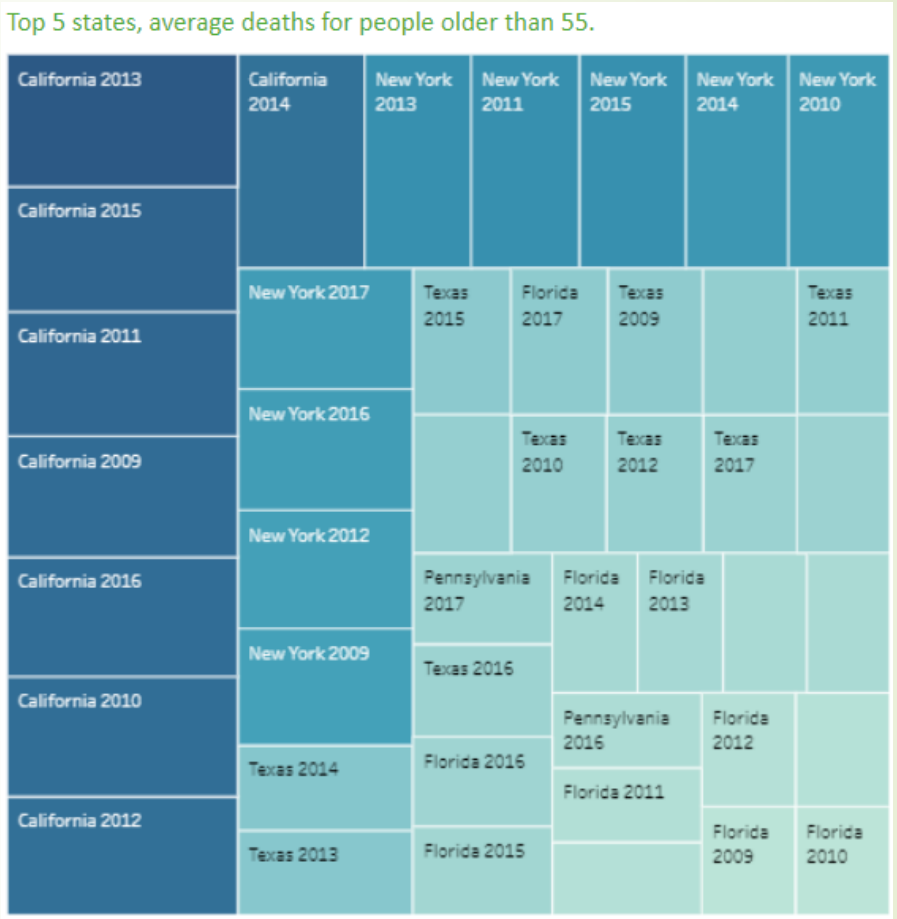


The CDC works with its various partners across the states, public health, and other local groups to collect the data. Unfortunately, Influenza deaths are not always directly related to having the flu. For some, there may also be underlying conditions such as cancer, heart issues, or other failing organs.

ANALYSIS

The Flu season affects different ages across each state, the frontline staff should be assigned in areas that have people 55 years and older.





RECOMMENDATIONS

Deploy the extra medical staff to the following states first, California, New York, Texas, Florida, and Pennsylvania. Although most states will remain flat in 2018 compared to 2017, Florida will be the exception, they will see a slight increase.

Tableau Final Project

PRODUCT ANALYSIS: ROCKBUSTER

Influenza Season

Objective

Rockbuster Stealth LLC is a movie rental company that used to have stores around the world. Facing stiff competition from streaming services such as Netflix and Amazon Prime. The Rockbuster Stealth management team is planning to use its existing movie licenses to launch an online video rental service in order to stay competitive.

Data

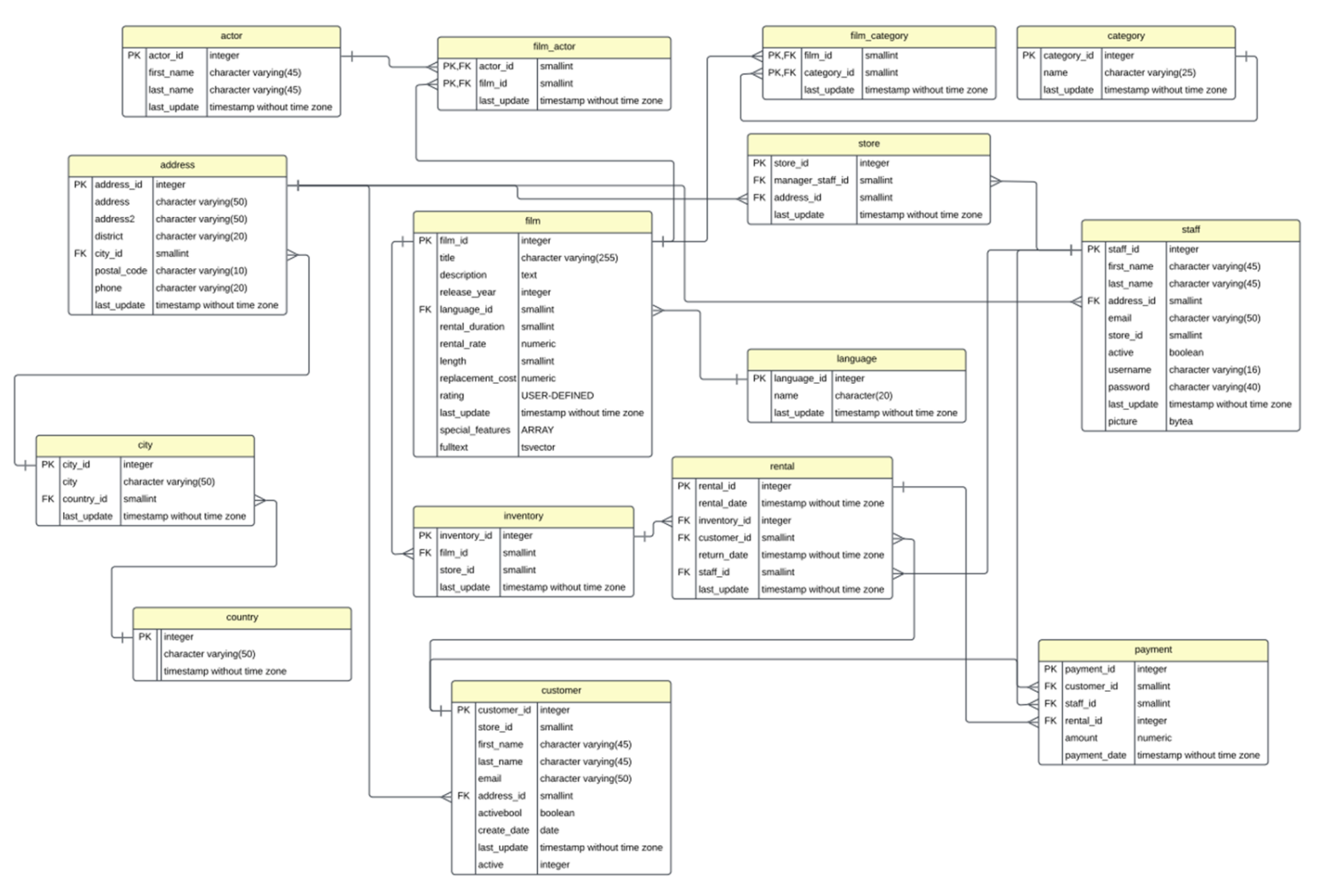
Data Dictionary

Internal Rockbuster data was loaded into a relational database including data on customers, film inventory, rentals & and payments.

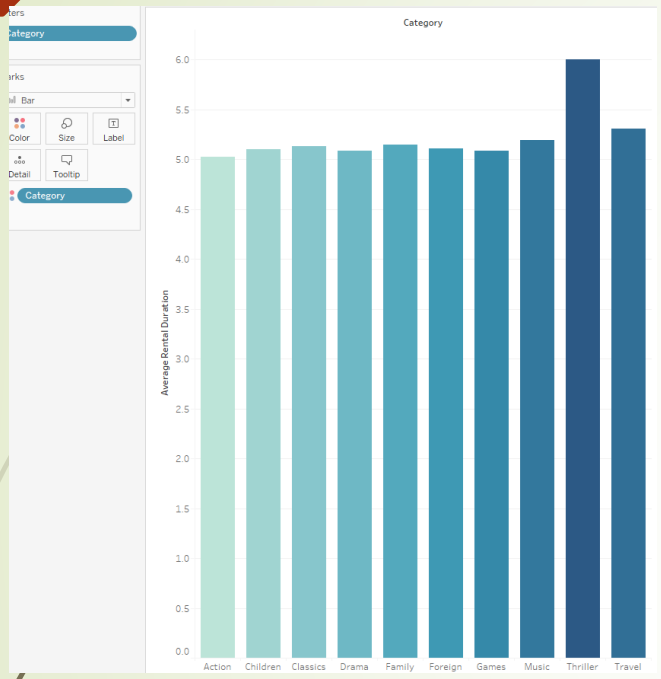
Tools

I. Relational databases
II. SQL
III. Database querying
IV. Filtering, Cleaning and summarizing
V. Joining tables
VI. Subqueries

ENTITY RELATIONSHIP DIAGRAM



ANALYSIS



- Min rental 3 days
- Max Rental 7 days
- Avg rental 5 days
- Highest Rental THRILLER

```
Query  Query History
1  SELECT category,
2  NAME genre
3  FROM category
4  LIMIT 16
```

Data Output | Messages | Notifications

	category category	genre character varying (25)
1	(1,Action,"2006-02-15 09:46:27")	Action
2	(2,Animation,"2006-02-15 09:46:27")	Animation
3	(3,Children,"2006-02-15 09:46:27")	Children
4	(4,Classics,"2006-02-15 09:46:27")	Classics
5	(5,Comedy,"2006-02-15 09:46:27")	Comedy
6	(6,Documentary,"2006-02-15 09:46:27")	Documentary
7	(7,Drama,"2006-02-15 09:46:27")	Drama
8	(8,Family,"2006-02-15 09:46:27")	Family
9	(9,Foreign,"2006-02-15 09:46:27")	Foreign
10	(10,Games,"2006-02-15 09:46:27")	Games
11	(11,Horror,"2006-02-15 09:46:27")	Horror
12	(12,Music,"2006-02-15 09:46:27")	Music
13	(13,New,"2006-02-15 09:46:27")	New
14	(14,Sci-Fi,"2006-02-15 09:46:27")	Sci-Fi
15	(15,Sports,"2006-02-15 09:46:27")	Sports
16	(16,Travel,"2006-02-15 09:46:27")	Travel

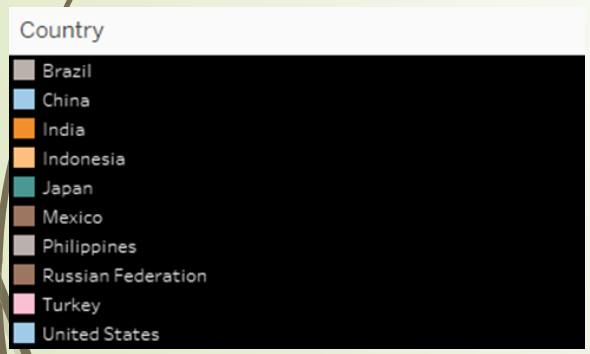
category	rental_count	average_rental_duration
Thriller	0	6.00
Travel	203	5.31
Music	179	5.19
Family	235	5.14
Classics	204	5.13
Foreign	252	5.11
Children	212	5.10
Drama	229	5.09
Games	208	5.09
Action	228	5.03

ANALYSIS


```
Query  Query History
1  SELECT A. Customer_id,
2         A. first_name AS "customer first name",
3         A. last_name AS "customer last name",
4         D. country,
5         C. city,
6  SUM (E.amount) AS "total amount paid"
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     WHERE city IN ('Cianjur', 'Acua', 'So Leopoldo', 'Iwaki', 'Ambattur', 'Ahanwei', 'Teboksary', 'Tianjir')
13     GROUP BY A. customer_id, first_name, last_name, country, city
14     ORDER BY "total amount paid" DESC
15     LIMIT 5
```

Data Output | Messages | Notifications

	country character varying (50)	number OF customers bigint
1	India	60
2	China	53
3	United States	36
4	Japan	31
5	Mexico	30
6	Brazil	28
7	Russian Federation	28
8	Philippines	20
9	Turkey	15
10	Indonesia	14



The size of the bubble shows revenue earnings in comparison to other countries.

RECOMMENDATIONS

I. Continue to build the foundation for the top 5 countries: India, China, United States, Japan & and Mexico.

II. A few of the lifetime customers live outside the top 5 countries. Investigate why that is and then strategize how you can build the area up.

III. Build the inventory with the higher revenue movie categories, i.e., Sports, Sci-Fi, and drama.

IV. Invest in Global Marketing where and when should the money be invested

CUSTOMER PURCHASE ANALYSIS

Objective

Instacart is an online grocery store that operates through an app. Instacart already has very good sales, but they want to uncover more information about their sales patterns.

Perform an initial data and exploratory analysis of some of their data in order to derive insights and suggest strategies for better segmentation based on the provided criteria.

Data

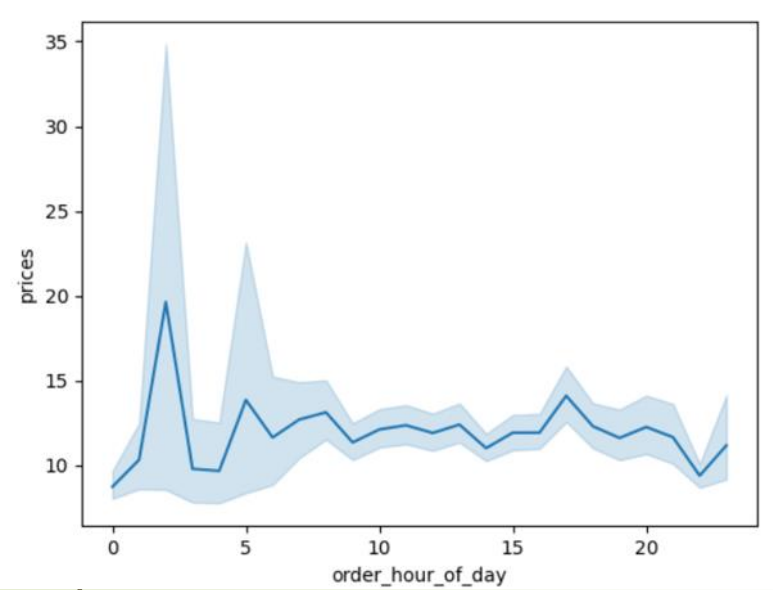
“The Instacart Online Grocery Shopping Dataset 2017” via Kaggle on 28/10/23

Customer Data fabricated by Career Foundry for the purpose of this project

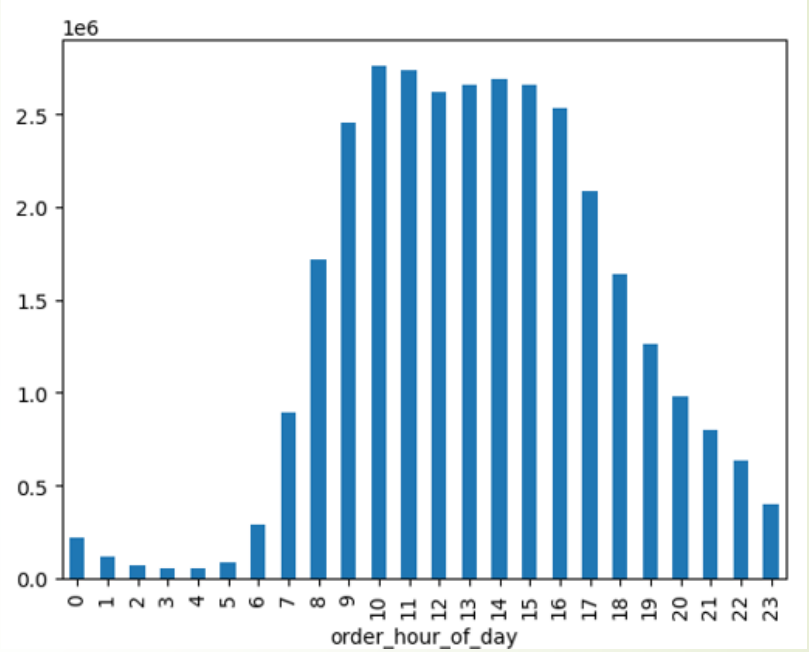
Tools

I. Python
II. Data wrangling/merging
III. Deriving variables
IV. Grouping data
V. Aggregating data
VI. Reporting in Excel
VII. PII procedures

ANALYSIS

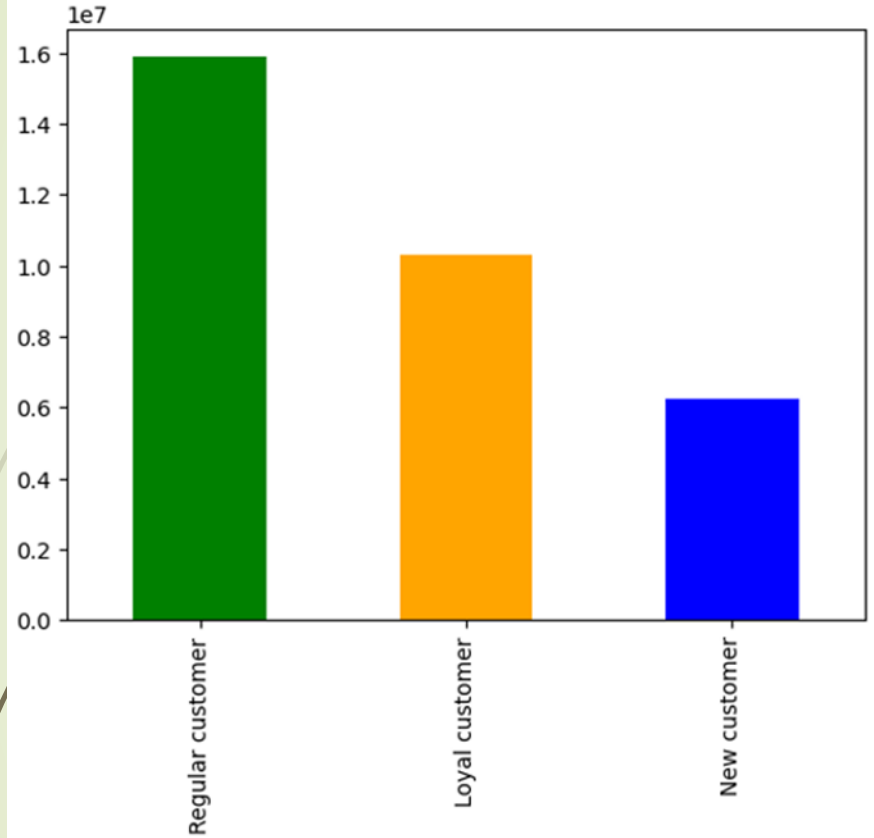


The busiest hour and most expensive is around 3 am. 5 am there is a small spike but throughout the rest of the day, it's consistent on the price.

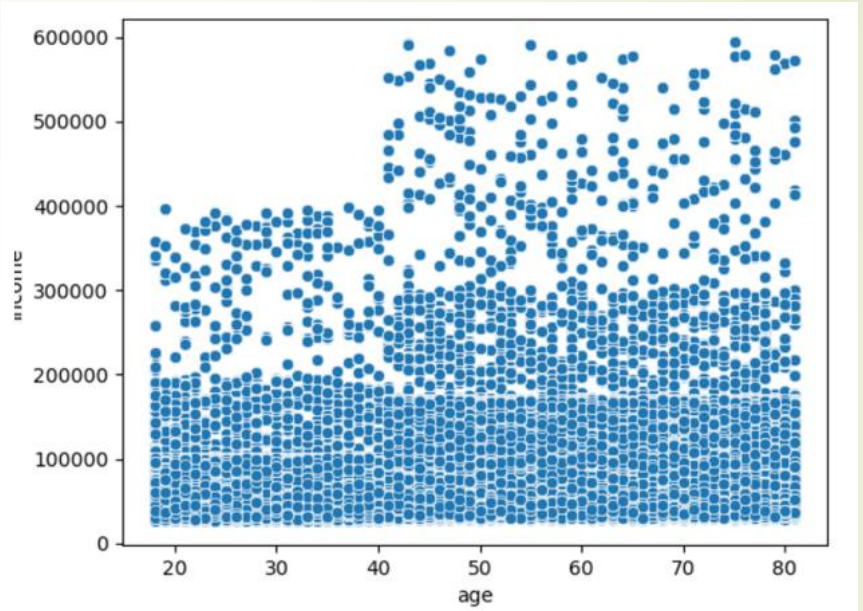


The busiest hours of the day when orders are placed.

ANALYSIS



The chart shows the breakdown of the type of customer buying.



Ages 40-60 have more spending $ than that of the younger generation. So, there is a correlation as the older you get the more earnings you make. spending

RECOMMENDATIONS

I. Based on the histogram the busiest hours for placing orders are from 9 am-4 pm. And specific days that draw the highest orders placed are Friday & and Saturday.. It would be beneficial to run a promotional campaign during these hours/days to continue to build on the momentum. That said, it is also worth noting the slower times and what is the reason holding people back.

II. The price range that is spent is $2-$15 meaning items purchased are reasonable and less expensive and as such cater to a wide variety of people and ages. This gives the sales team a large avenue to develop campaigns targeting various products.

III. Just like in restaurants, retail is able to look at high-selling items (dairy, frozen items, produce & snacks) and build a marketing/promotional campaign around them.

IV. The regular customer builds a consistent pattern of shopping via Instacart. And is the strongest customer base for the business. The loyal customer is the 2nd strongest and although they are loyal, they still shop elsewhere, so devising a marketing campaign that targets their needs is how you switch them to a regular customer. Loyalty is a fickle relationship, if you disappoint the Loyalty customer (a one-off situation), you are liable to lose their business, and they may never come back.

V. It is recommended to target specific demographics (age groups, and what they are spending week after week. Build a campaign around them and tie another campaign into turning loyal customers into regular customers.

ANTI-MONEY LAUNDERING

GLOBAL PIG E. BANK (fictional)

Objective

Suggest ways of controlling for bias and communicating concerns to stakeholders.

Data

Fictional client data made by Career Foundry for the purpose of this task

Tools

I. Big data
II. Data ethics
III. Data mining
IV. Predictive analysis
V. Time series analysis and forecasting
I. Differentiating between regression and classification models
I. Using GitHub

ANALYSIS

Row_number	Column removed	Unneccesary
Customer_ID	No Changes	-
Last_name	column removed	unneccesary and PII
Credit Score	3 missing values: changed to N/A	remove missing data
Country	FR changed to France: 244 replacements DE changed to Germany: 23 replacements ES changed to Spain: 118 replacements	Changed for consistency
Gender	F changed to Female: 19 replacements M changed to Male: 49 replacements Null changed to N/A: 1 replacement	Changed for consistency
Age	11 ages of 2: changed to N/A 1 NULL chnged to N/A	2 yrs old not correct & Null changed to N/A for consistentcy
Tenure	No Changes	-
Balance	No Changes	-
NumOfProducts	No Changes	-
HasCrCard?	No Changes	-
IsActiveMember	No Changes	-
Estimated Salary	1 Blank & 1 Null changed to N/A	Changed for consistency
ExitedFromBank?	No Changes	-

The variables that play an active role in which shows NETFLIX should expand its library.

Objective

This dataset provides a comprehensive view of movies and shows available on Netflix. It captures a range of information, offering insights into the type of content, its popularity, and other significant aspects.

This dataset is invaluable for tasks like trend analysis in entertainment, predictive modeling for viewer preferences, and comparative studies of genres and release years. It allows for an in-depth understanding of the factors that contribute to the success and popularity of movies and shows on Netflix.

Data

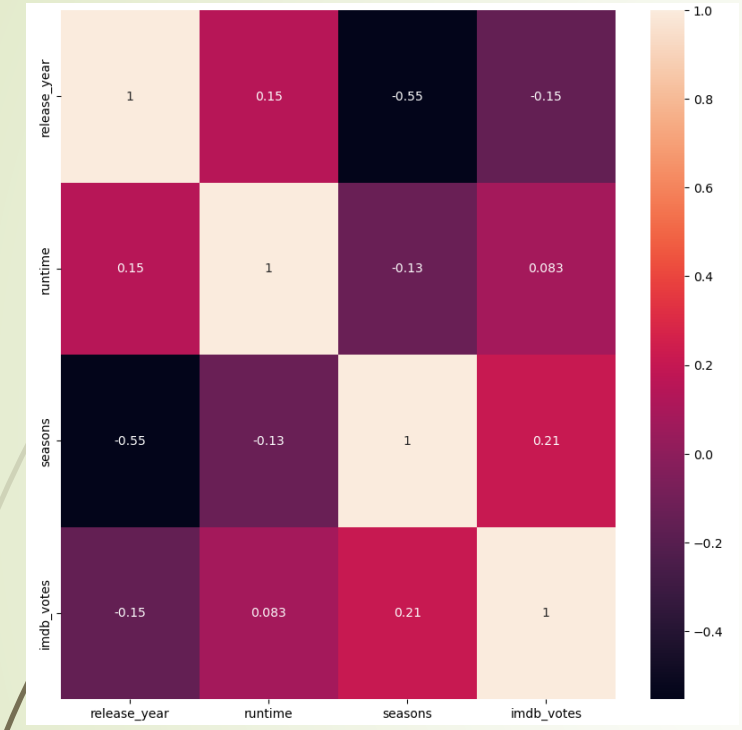
https://www.kaggle.com/datasets/maso0dahmed/netflix-movies-and-shows/data

Netflix Movies & Shows (primary focus was TV shows only)

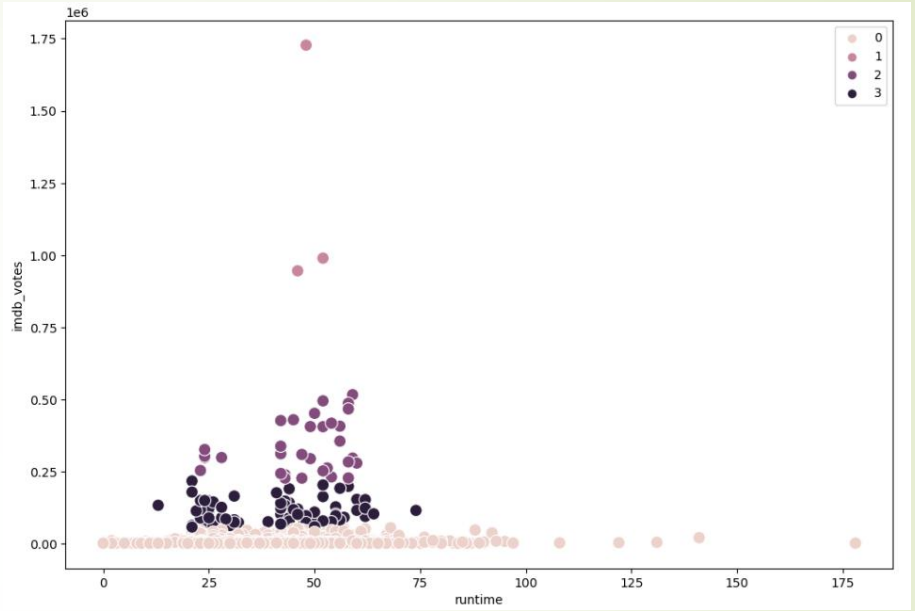
Tools

I. Python
II. Data wrangling/merging
III. Deriving variables
IV. Grouping data
V. Aggregating data
VI. Reporting in Excel
VII. Tableau

ANALYSIS

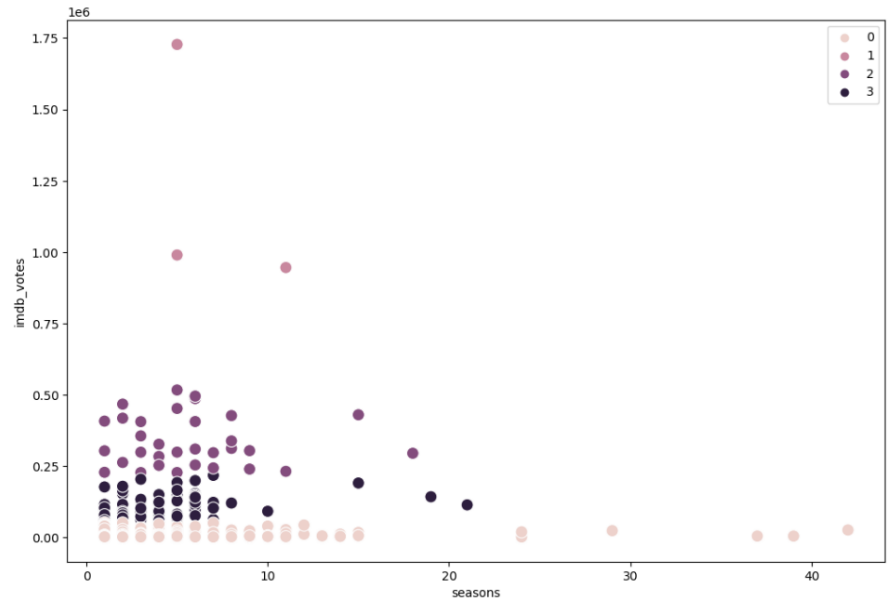


Imdb scores & and Imdb votes have the highest correlation at .25. However, runtime and score are at .22. The votes and the score are too similar, and the scores are a direct result of the # of votes. Too many or too little can have both a positive and negative impact on the result.

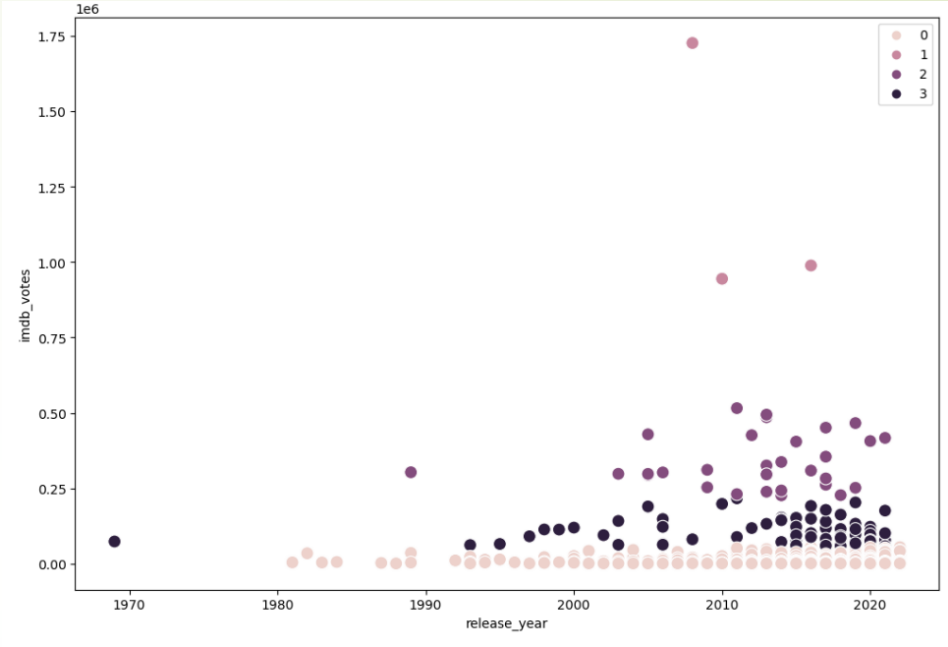


The shorter the runtime the fewer votes. This begs the question as to why. Further analysis is needed to better understand how these two variables correlate if at all.

ANALYSIS



The fewer seasons the more votes are congregating. This information can lead to new questions. How does genre play a role, when was it released?



The number of votes has increased through the years, partly by social media influencers and social media in general. With so many platforms out there and people wanting to have a say in everything, there is no surprise to see the data where it's at.

RECOMMENDATIONS

I. The original dataset looked at movies and shows. Focusing on shows only, more research is warranted on how the number of seasons impacts the voting.

II. The release year range is from 1949 through 2021. When you think back to the status of the world during COVID-19 and how this impacted voting, production countries, scores, and even the release years. Further research is warranted. Release years were all pushed back or put on hold due to the pandemic. Perhaps filtering the data to exclude anything past 2018 would help to generate more reliable data.

    I. During the pandemic with everyone forced to stay inside, more and more people were watching anything and everything. Anything to not be bored.

III. The variables that were looked at are Title, age certification, genres, production countries, IMDB ID, IMDB votes, IMDb scores, type, and release year. Taking into account what drove the increase in votes in the latter years is crucial to understanding how best to expand the library.

Thank you

Daniel82676@gmail.com

Tableau Profile