

STEAM DATASET ANALYSIS

PROBLEM STATEMENT AND OBJECTIVES:

The objective is to develop an interactive dashboard for analyzing Steam data. This includes cleaning and preprocessing the dataset, addressing missing values, and conducting feature engineering to ensure analysis readiness. Subsequently, the dashboard should be able to unveil meaningful insights. The design emphasizes user-friendliness, visual appeal, and interactivity, allowing stakeholders to explore data dynamically and discover meaningful relationships. Ultimately, the dashboard serves to engage stakeholders by providing actionable insights through intuitive visualizations and interactive features.

DATA PREPROCESSING:

1. Handling Null values

A. For certain columns (Name, About the game, Reviews, Website, Support URL, Support email, tags, Screenshots, Movies, Notes, Developers, Publishers, Categories, and Genres) where filling null values using mean, median, or mode would mislead the data due to the categorical nature of the values, we've replaced null values with "unknown".

B. For columns with more than 90% null values ('Score rank', 'Metacritic URL'), we've removed these columns from the dataset as we can't get much insight from these

2. Feature Engineering

A. Genres:

- Since Genres have multiple combinations of values, we've extracted the first individual genre from the set in each row and created a new column named '**Genre**' which holds a single genre value for each entry.

B. Price:

- We created a new column named **price_category** based on the **price** column. If the price is 0, the category is set as "free". If the price is less than or equal to 10, the category is set as "below 10\$". If the price is above 10, the category is set as "over 10\$".

C. Estimated Owners:

- The **Estimated Owners** column had a range of estimated customers in the form of a string (e.g., '0-1000'). we've extracted the lower and upper bounds of this range, found the average, and created a new column named **Owners_Avg** containing the integer value of the average number of estimated owners.

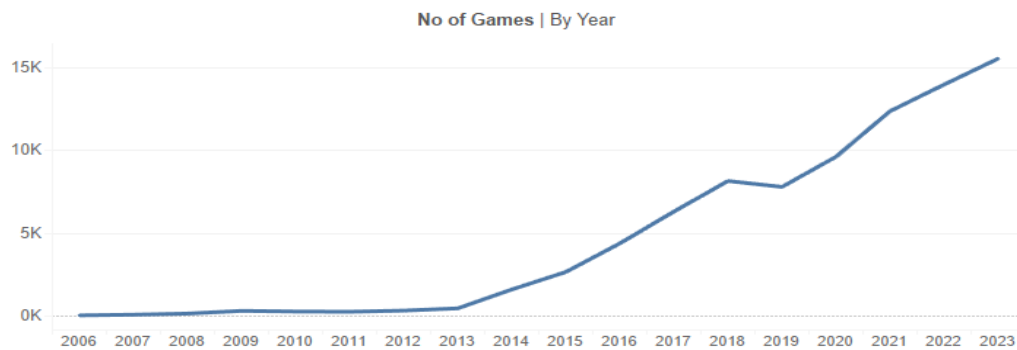
- These preprocessing steps helped us in preparing the dataset for further analysis or modeling.

Analytics and Dashboard:

1. KEY INSIGHTS

-While going through the data, lots of meaningful insights were discovered, some key points among them are

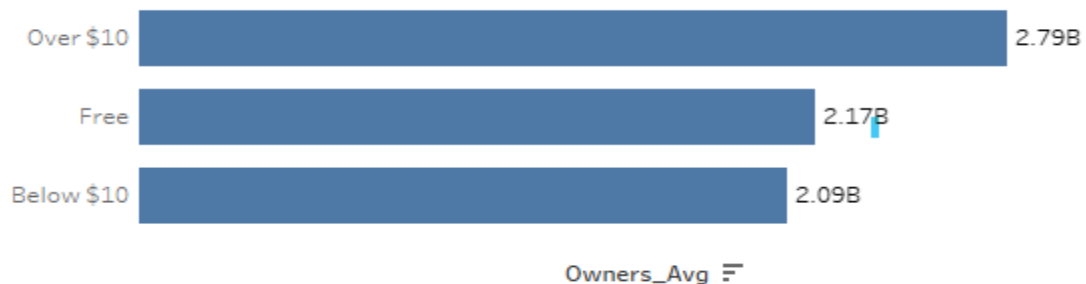
A. Trend of games over the years :



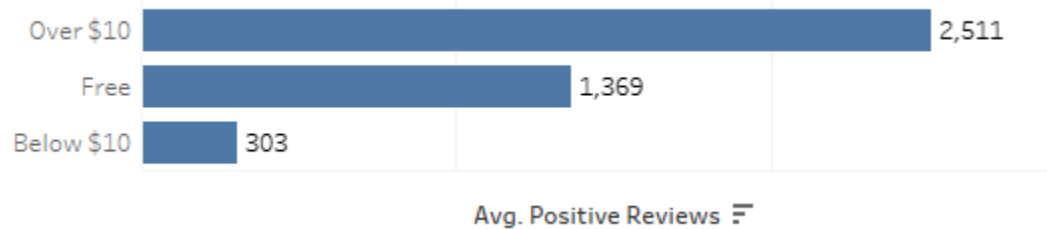
- We can see that each year, the no. of games released is increasing by a lot when compared to previous years

-We have to stand out in the market to get people to like our game

B. Price Category vs No of People Who bought the game

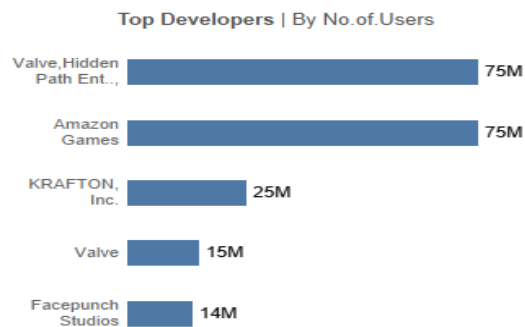


- From the above, we can see that more people prefer games over \$10 than Free games or Low-cost games (Less than \$10)
- Why? Look at the below graph:



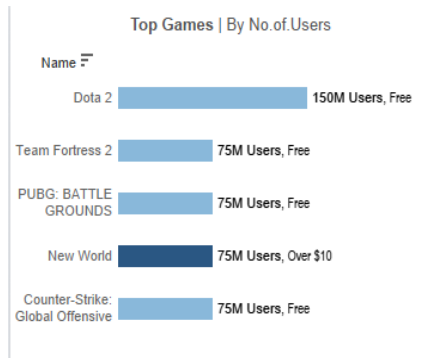
- Here, we can see that games priced over \$10 have a higher average Positive review, which means that these games are of a better Quality
- ***This shows that people are willing to spend more for Quality Games***

C. Top Developers and their strategy



- The Top Developers whose games are bought by lots of people are 'Valve, Hidden Path Entertainment', And 'Amazon Games' which are bought by a significantly higher number of people compared to other Developers in the top 5.
- Both have launched only one game but it is used by a lot of people
- ***This shows that the quality of the game is more important than number of games***

D. Top Games and their similarity



- We can see that **all games in top 5**(Based on no. of. users) are **Action games**
- It is also noted that **4/5** of them are **Free Real-time Multiplayer Games**

2. DASHBOARD

- Although we can extract a lot of valuable insights from data, it is important to note that we cannot display all of them on our dashboard
- Dashboard is meant to provide stakeholders with an overall understanding of the dataset. Therefore, it is important to keep the dashboard simple, consistent, and engaging.
- Here's Our Dashboard:



- The dashboard provides an overall view of our data
- It shows the trend of key metrics and the distribution of price categories
- It is simple and easily interpretable
- There are not too many visualizations, making it easier for stakeholders to grasp the information
- There are various filters available, making it more engaging
- Each item on the dashboard is clickable for interactivity
- It covers all the key metrics that need to be monitored
- Good spacing enhances our visualization's interpretability for easy data insights.

CONCLUSION:

1. **Data Readiness:** The Steam dataset has been cleaned and pre-processed, setting the stage for meaningful analysis.
2. **Insight Highlights:** Analysis reveals trends such as rising game releases annually and correlations like higher-priced games garnering better reviews.
3. **Dashboard Implementation:** The interactive dashboard offers stakeholders an easy-to-use platform to explore Steam data insights effectively.
4. **Future Focus:** Continued analysis can delve deeper into specific aspects of the dataset, while updates to the dashboard will maintain its relevance and usefulness to stakeholders.