ETL PROJECT REPORT GROUP 10

Contributors:

* Alvin
* Makintha

Data Sources:

* Esports earnings per team/player

<https://www.kaggle.com/jackdaoud/esports-earnings-for-players-teams-by-game?select=highest_earning_teams.csv>

* Historical Esports Data earnings per game

<https://www.kaggle.com/rankirsh/esports-earnings>

Motivation:

We selected this data source to investigate why Esports is becoming more popular in the newer generations. Is this due to a different generation mindset, or is it because becoming a professional esports athlete is a financially rewarding profession.

The purpose of transforming data is to gain a greater understanding from the data. The findings will be presented in the form of graph data-visualisations.

Extract:

The data was in the form of CSV format which can be downloaded from the above data source links

Transform:

Data Transformation Process:

* Importing CSV files to Jupyter Notebook
* Checking for null values/duplicate values
  + If null values found, delete
  + If duplicates found, keep first item
* Data exploration through aggregating data
  + Sort by
  + Group by
  + Reset index
* Transforming DataFrames for data visualisation

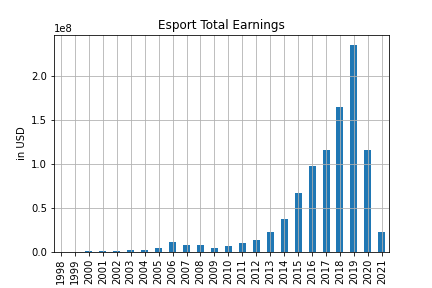
Load:

The final database chosen was a relational type in PostgreSQL. The justifications are as follows:

* Dataframe type
* Columns are fixed
* Some relational data between DataFrames e.g. Country Code
* Time consuming to convert to JSON format for non-relational database

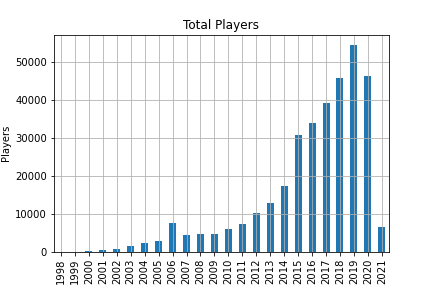
Data Analysis:

Below are the findings from the data exploration. All earnings are USD (unless otherwise stated)



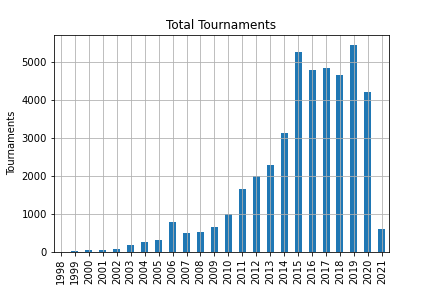
Esport Total Earnings:

* Earnings increasing in the past 20 years
* Peak in 2019 (~$250 000 000)
* Rapid drop in 2020/2021, possible due to COVID



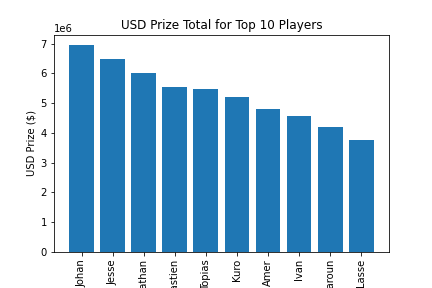
Total Players:

* Trend similar to total earnings
* Increasing for past 20 years
* Peak in 2019
* Drop in 2020 and 2021, possible due to COVID



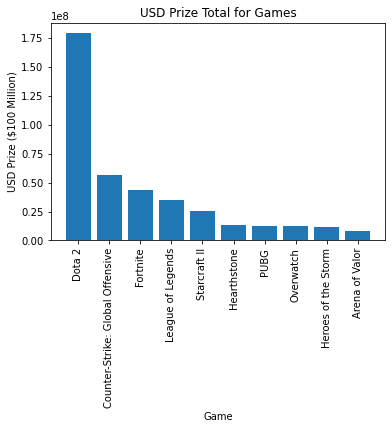
Total Tournaments:

* Trend again similar to previous findings
* Increasing for past 20 years
* Drop in 2020 and 2021, possible due to COVID



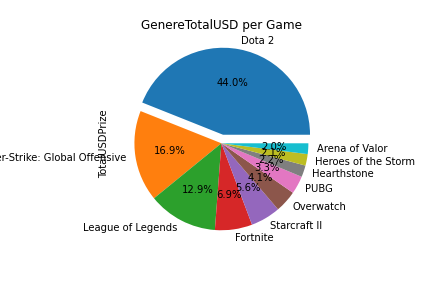
Prize Money for Top 10 Players:

* Top players win almost $7 000 000
* Average earnings for 1000 highest earning players, $397 793.21



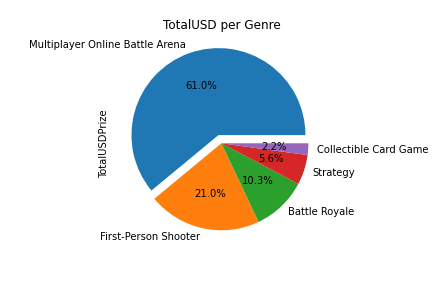
Game Total Prizes:

* Dota 2 is the highest earning game, over triple the amount of the second highest earning game



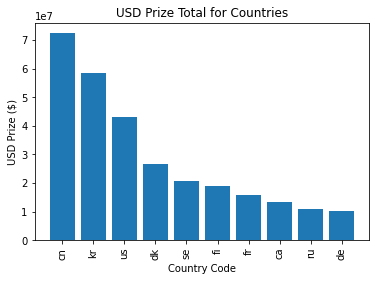
Game Earning Distribution:

* Dota 2 has almost 50% of the total earnings



Genre Earning Distribution:

* Multiplayer Online Battle Arena contributes more than 60% of the total prize earnings
* Collectible Card Game is the lowest contributor



Countries Earnings:

* China (cn) has the highest total prize earnings with over $70 000 000
* Continent of Asia has two of the highest earning countries (China; cn and Korea; kr)

Conclusion

From the data analysis, it can be seen that the trend for total prize earnings is increasing, peaking at ~ $250 000 000. We believe that, based on the previous years, the trend would continue to trend upwards if COVID did not have an effect on earnings. From the graphs, it can be seen that the total prize has a positive correlation with the total players and total tournaments. We also conclude that the highest players earned just under $7 000 000, the highest prize earnings for games comes from Dota 2 and Multiplayer Online Battle Arena genre.

To conclude, it seems that it is possible to make a living off being an Esports athlete, however, it is recommended to play in games and genres with high prize earnings (e.g. Multiplayer Online Battle Arena, Dota 2) or to play in an Asian region.