Kelompok 13:

- Vyos Devin (2501974776)
- Haziq Ibrahim Omar (2502021342)
- Muhammad Dzul Qarrnayn Destra Nurmanto (2502027485)
- Alexandre Makarios(2502020440)

# TMDB MOVIE

Dataset TMDB 5000 is data that contains a title film, homepage, id, original title, overview, popularity, production companies, production countries, release date, spoken languages, status, tagline, vote average there is have 14 variable and 4803 object

## TECHNIQUES

in the proses of analysing this data set we use a total of 4 data analysis techniques which is:
1. classification
2. data visualization
3. data cleaning
4. prediction(regression)

## DATA CLEANING

In this data set we found 2 NA value in category runtime, before clean this 2 NA value we have 4803 object after the cleaning we have 4801 object
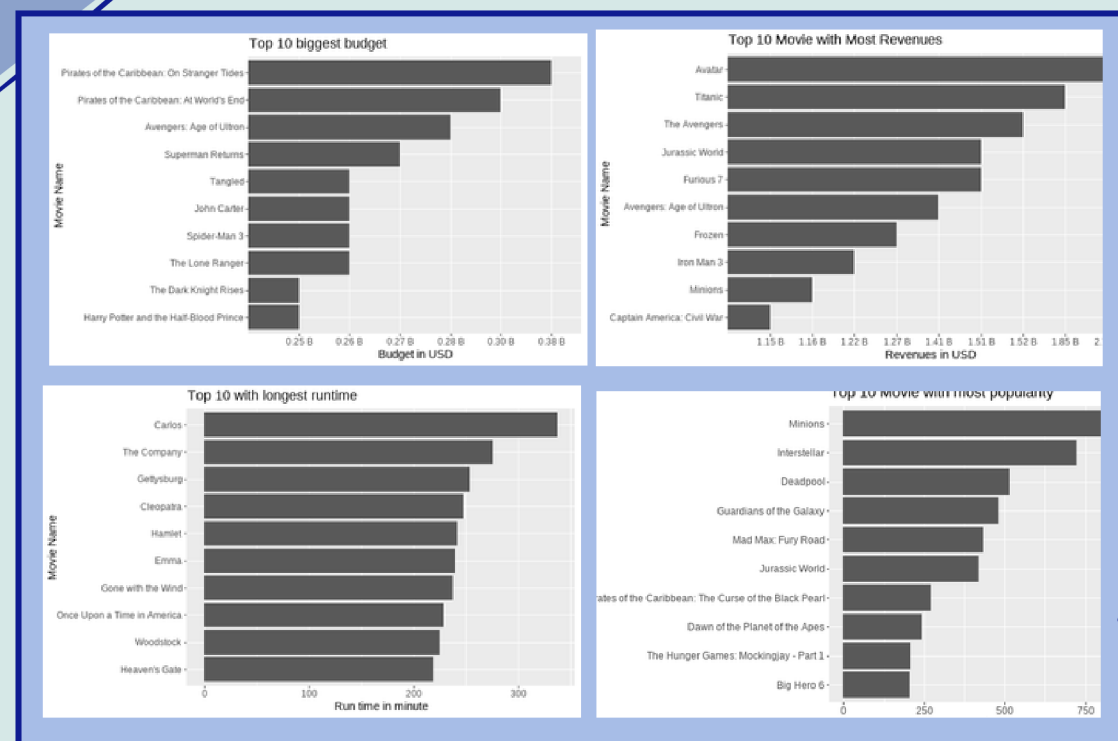
## DATA VISUALIZATION

we use data visualization technique to better show the data that we choose from the classification proses

## CLASSIFICATION

in this data set we used the classification technique to choose only the top 10 of a certain variable that we will analyze for example we pick the top most grossing film of all time

## PREDICTION

we use regression technique to predict what kind of data that will be shown form outside of the data frame for example we used in on budget and revenue to predict how much budget and revenue a data outside of the data frame will get



## CORRELATION



- from the correlation plot we know that vote count and popularity (0.78) we can conclude that popularity and vote count balance because if popularity increase vote count will icnrease too

vote count and revenue have big correlation value(0,78) because if the film get big vote count it will impact to revenue if the film get more vote count we can say the film is recommend for watch

## RESULT

Insights gained from data analysis are summarised below:
- avatar is the most of revenues with 2.79B USD, follow titanic with 1.85B USD, and The Avenger with 1.52B USD
- Pirates of the Caribbean: On Stranger Tides is the most budget make the film with 0.38B USD
- Carlos movie is the most runtime with 3 Hours and 38 Minute runtime

## REGRESSION

from regression method we predict that if the budget were 405.089980 the predicted profit will be about 2.064607829 that is because in the data frame there are a huge gap between the first and the second top grossing film

## CONCLUSION

from this data we can say that the most amount of grossing film is avatar with 2.79B in revenue and the most amount of budget spend in one movie is pirate of caribbean with 250 million and we predicted from the model if a movie spend about 400 million in budget if will gross about 2B and its because budget and revenue have a huge correlation value almost over 0,73