Project: Investigate the TMDb Movie Dataset.

- A statement of the question(s) you posed

In today's fast-paced technologically driven society, various means of entertainment are available and are easily accessible electronically. Movies are one such entertainment measure and with the advancement of technology, data related to each movie can now be obtained easily.

Here, I will be investigating the 2017 TMDb movie dataset sourced from Kaggle, with about 10,000 movies.

Below are questions we will try to answer based on the 2017 TMDb movies dataset.

1. Top/Popular Movies based on budget?
2. What movie made the highest profit based on budget?
3. What kinds of properties are associated with movies that have high revenues?

- A description of what you did to investigate those questions

Import Packages
Load the dataset
Data Wrangling
Exploratory Data Analysis (EDA)
Conclusion.

In the Data Wrangling Phase i did:
I load the data,

Checked for null value
Checked for duplicated rows and columns
Checked for datatypes.

- Summary statistics and plots communicating your final results

The summary statistics and plot can be seen in the jupyter notbook attached to this document.

References:

https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.apply.html

https://stackoverflow.com/questions/35376387/extract-int-from-string-in-pandas

https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.append.html

https://www.geeksforgeeks.org/bar-plot-in-matplotlib/