

## Udacity--Project-Investigate-TMDB-Movies-Dataset [REPORT]

### Project Overview

In this project, we have to analyze a dataset and then communicate our findings about it. We will use the Python libraries NumPy, pandas, and Matplotlib to make your analysis easier.

What do I need to install? You will need an installation of Python, plus the following libraries:

pandas

NumPy

Matplotlib

### Why this Project?

In this project, we have to go through the data analysis process and see how everything fits together. I have also use the Python libraries NumPy, pandas, and Matplotlib, which make writing data analysis code in Python a lot easier!

Questions I posed:

- 1- Movies which had most and least profit
- 2- 2.Movies with largest and lowest budgets
- 3- 3.Movies with most and least earned revenue
- 4- 4.Movies with longest and shortest runtime values
- 5- 5.Average runtime of all the movies
- 6- 6.In which year we had most no. of profitable movies
- 7- 7.Successful genres (with respect to the profitable movies)
- 8- 8.Most frequent cast (with respect to the profitable movies)

- 9- 9.Average budget (with respect to the profitable movies)
- 10- 10.Average revenue (with respect to the profitable movies)
- 11- 11.Average duration of the movie (with respect to the profitable movies)

What did I do for investigating these questions?

I cleaned the data then found the correlations between the dependent and independent columns

Data Wrangling :

- 1- Remove unused columns.
- 2- Remove duplicate rows.
- 3- Change release date to date format.
- 4- Replace zeros in runtime column with NaN and drop them.
- 5- Replace zeros in budget and revenue columns with NaN and dropping them.
- 6- Change the data types of budget and revenue and runtime columns.

What I have learn?

After completing the project, I have learned following :

- 1-Know all the steps involved in a typical data analysis process
- 2-Be comfortable posing questions that can be answered with a given dataset and then answering those questions
- 3-Know how to investigate problems in a dataset and wrangle the data into a format you can use
- 4-Have practice communicating the results of your analysis
- 5-Be able to use vectorized operations in NumPy and pandas to speed up your data analysis code
- 6-Be familiar with pandas' Series and DataFrame objects, which let you access your data more conveniently
- 7-Know how to use Matplotlib to produce plots showing your findings