

**TASK**

**Exploratory Data Analysis on the Automobile Data Set**

[](http://www.hyperiondev.com/portal/)

**Introduction**

The main objective here is to carry out preliminary exploration of movie data. The goal is to select the right attributes for use in the computation of the following: “Movie with the Highest Budget”, “Viewing the top genre”, “Selecting Movies by how they are rated, particularly from high to low”, ‘Finding the most popular movie”, “Most profitable movies” and so forth.

**DATA CLEANING**

# SUMMARY OF THE METHODS AND VISUALIZATIONS DONE DURING DATA CLEANING

* **Removing unused columns**
* **Removing duplicated rows if there’s any**
* **Changing release date column into date format**
* **Changing the format of budget and revenue column**
* **All JSON data in the dataset was cleaned so that it can be used**

MISSING DATA

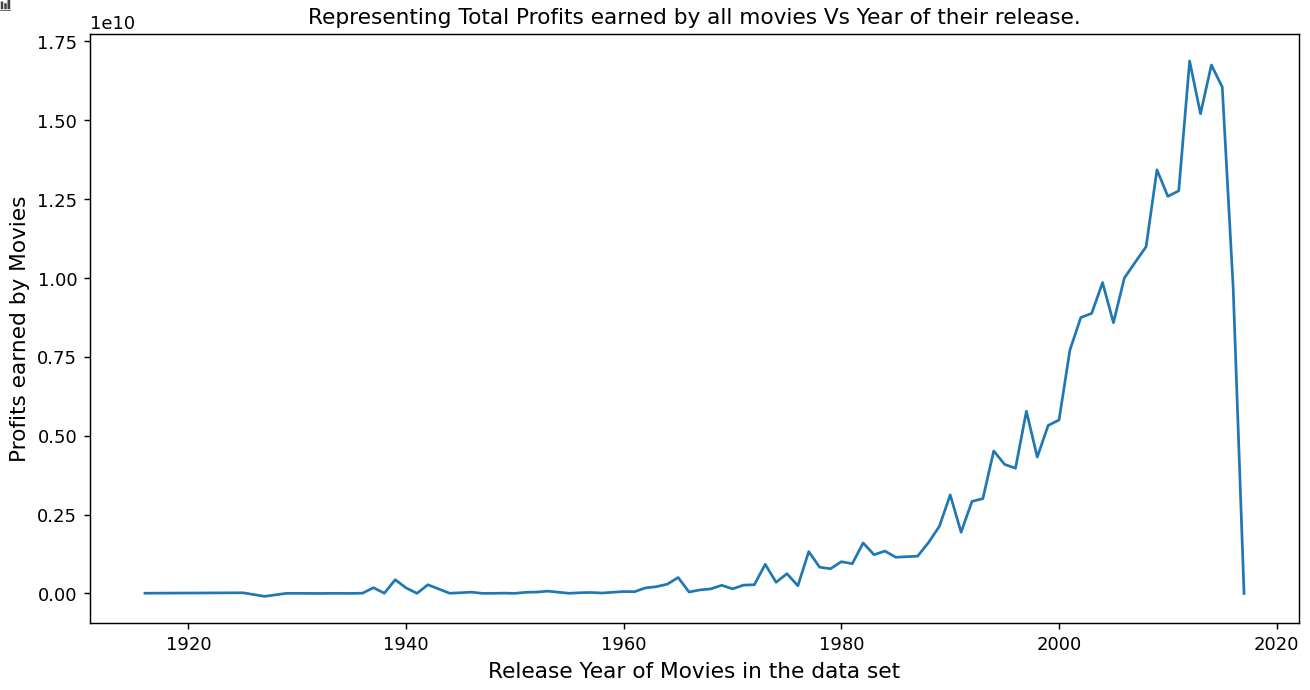
# ANY MISSING DATA? HOW DID YOU HANDLE IT

* It was found that there’s missing data particularly on the budget and revenue entries. These entries were having 0 values suggesting that the data is missing
* Steps taken in order to handle the missing data was to remove the entries with 0 values in the data frame looking at these two columns (‘revenue’ and ‘budget’)

DATA STORIES AND VISUALIZATIONS

# THIS IS THE BULK OF THIS PROJECT. EXTRACT STORIES AND ASSUMPTIONS BASED ON VISUALIZATIONS OF THE DATA

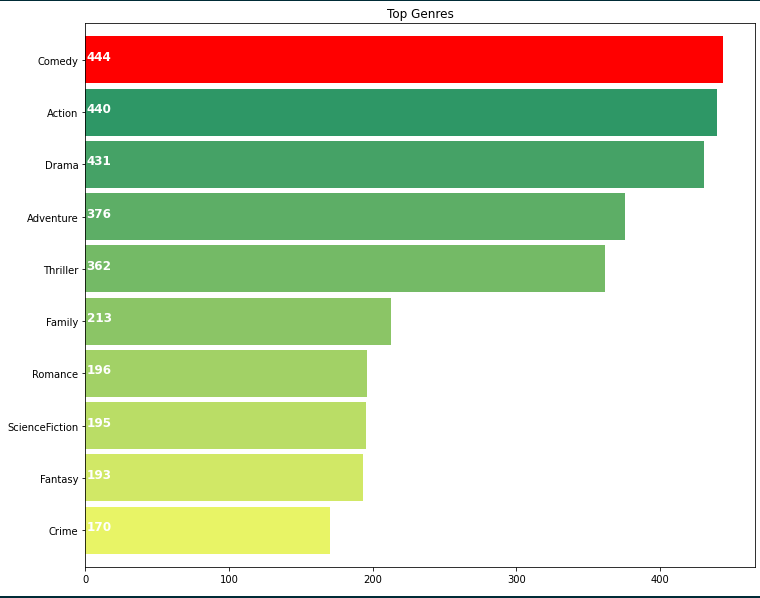
1. Representing total profit earned by all movies VS year of their release



For us to know which year had the most profit in the data frame we had to first clean some of the data such as date and after we had to plot this graph so that we can visually see how the data trends.

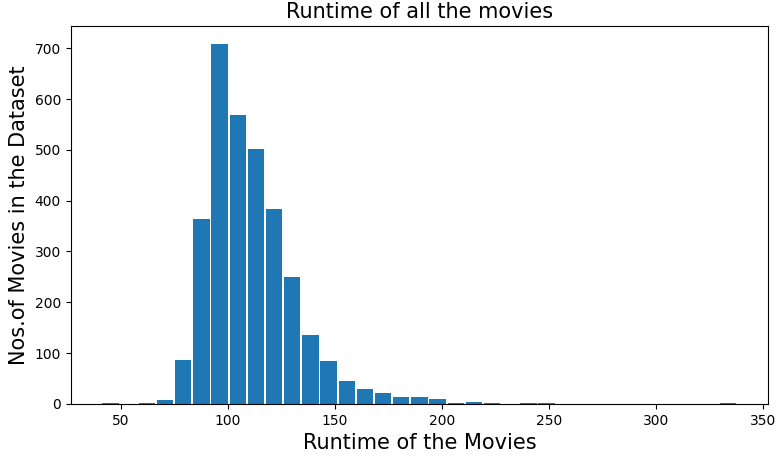
We could finally know from the graph that 2014 was the year with most profit.

1. Comparing Genre



To find out the top genre for all our movies in the data frame we plotted a graph to do the comparisons which help us in making decisions for example on which genre is popular, where to improve, etc. This graph particularly looks at the top 10 genres

1. **Viewing Runtime of all the movies**

****

**THIS REPORT WAS WRITTEN BY : LIVHALANI DAU**

