Business Intelligence: An EDA for the new launch of a movie studio

Robert Wong
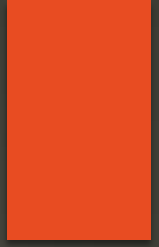**Part-time Data Science Program, Flatiron School**
Date: 17th Feb 2020
Instructor: Eli Thomas

Outline

- Problem Statement
- Methodology
- Data Collection
- Data Analysis and EDA
- Findings
- Conclusion

Problem statement

This project has to do with the types of movies to produce that will intrigue a wide audience.

The issue here is that Microsoft has never created movies before and now needs data to decide on what genres of movies will reach the movie goers.

This endeavor by Microsoft just recently came about through their newly launched movie studio.

The problem is how to be able to predict what genres of movies that will bring box office success.

The project is significant to Microsoft's continuing business expansion due to rapid performance growth in other parts of the business operations such as movie streaming, Xbox video game consoles, AI, etc. The inclusion of movie production as part of Microsoft's business growth will benefit stakeholders, shareholders, customers, and the corporation as whole. For example, more revenue generated can be reinvested into future endeavors for the business operations. Also, increase in performance and profits will create more productivity within the organization. This can benefit the company as well in terms meeting their obligations with the stakeholders. Lastly, if the inclusion of making movies in the business operations brings in an increase in revenue, then the shareholders could also benefit such the rise in stock prices or maybe a dividend payment to stockholders.

Methodology

• Primary: To collect relevant statistical categorical data that presents certain proofs about specific movie genres.

• Secondary: To analyze variable factors of the data under various demographic subcategories.

• Third: To use a few algorithm statistical models to perform data analysis and EDA on the for classification, regression, and clustering.

Data Collection

The data collected was from IMDB through Grouplens' site which provides various amounts of datasets in zipped files. The url is http://files.grouplens.org/datasets/movielens/ml-latest-small-README.html. Under this project's set of instructions data analysis and exploratory data analysis were performed to reach a decision as well as explore the data and build a few statistical algorithmic models to provide decision-making strategies as well as conclusive insights on the categories of movies Microsoft can produce to reach a large audience. Subsequently, I used charts and graphs towards the data visualization along with statistical analysis.

DATA ANALYSIS AND EDA

TEAM CONNECTION

Findings

What types of statistical models can be used to make a prediction of certain types of movies?

How does these models help strategize on new business endeavors?

To what means does EDA along with statistical models provide a solution on decision-making strategies?

Conclusion

Decision Making Strategies

In conclusion of this project on Microsoft's newly launched movie studio, the three questions in slide six were answered from the statistical analysis and EDA performed in jupyter notebooks.

For instance, Principal Component Analysis (PCA), Bayesian Ridge Regression Algorithm, and Decision Tree Regressor were used to provide guidance to the CEO of Microsoft and the organization's team in their decision on what genres of movies to produce.

By utilizing the three statistical models, I was able to implement some strategies to provide an in-depth exploratory data analysis report on what categories of films to produce for their new business venture. Upon answering the three questions for the findings in slide number 6, depending on what type of business operation, I have concluded that Microsoft already a point of entry into the movie industry based upon their success in the gaming industry with its game console Xbox as well as the game development division. So, Microsoft can basically build on that to decide which types of film categories and genres that will serve their audiences in which they did the same for their gaming entertainment business operations.