Name: Wassem Messiha

Student #: 501113671

Supervisor Name:

Date: September 27, 2021

Abstract

The outbreak of COVID-19 has hit the economy severely the past year, not only causing higher unemployment rates and lower income but also causing people to practice social distancing, and national and international travel restrictions. Reducing capacity in most businesses caused the complete shutdown of many industries. Speaking of the industries that were heavily impacted by COVID-19 pandemic, I would like to bring the light on the film industry, focusing on 2020 and the first two quarters of 2021. Due to the fear we have all developed during the pandemic and the new regulations implemented, people around the world were forced to stay home. This led to cinemas and movie theaters having to close, and big movies having to push back their release date for a year and others even more due to coronavirus uncertainty. The worldwide box office estimated revenue has reportedly reduced from 44.5 billion U.S. dollars to 16.3 billion in 2020 alone. In this analysis project I will be using Classification and Regression analysis to focus on the revenue trends of the box office in previous years to get an insight about the top grossing months, top grossing movies based on actors and distributors, how are the trends for genres correlated with revenue, and the correlation between budgets and revenues of the movies. The previous areas of focus along with other attributes in the dataset will also allow me to build a Predictive Analytics model to predict the box office revenue in the next 3 months as the film industry starts to go back to normal with new releases announced and as audiences continue to have a strong appetite to enjoy the theatrical experience. I will be using the boxofficemojo dataset found on Kaggle, it has 26 attributes and 3243 observations for the last 30 years. First, I will start with Exploratory Data Analysis in order to understand the data description, the type of each attribute, and finding missing values and outliers. I will also do revenue classification based on movies genres and distributors. Then I will perform a correlation analysis on the dataset to see

which attributes are highly correlated with the revenue. And finally, I will use Linear Regression analysis to predict the movies revenue for the fourth quarter of 2021.