**Data Analytics Capstone Topic Approval Form**

The purpose of this document is to help you clearly explain your capstone topic, project scope, and timeline. Identify each of the following areas so you will have a complete and realistic overview of your project. Your course instructor cannot approve your project topic without this information*.*

*Note: You must fill out and submit this form. Space within each section will expand as needed.*

*Any costs associated with the development of the data analytics solution will be the responsibility of the student.*

**Student Name:** Kenneth Exchange

**Student ID:** 010994904

**Capstone Project Name:** Relationship of Movie Popularity/Rating with Revenue

**Project Topic:** To identify relationships between movie revenue and movie ratings/reviews

**Research Question:** Do popular movies make more money?

**Hypothesis:** Movies with higher viewer ratings generate more revenue.)

**Context:** Many people base what movies they will pay to watch on view ratings such as those found on IMDB.com. It is highly likely that the film industry bases decision on what movies to produce based on similar metrics. This research will identify if a movie’s viewer ratings are an indicator of the its potential monetary success.

**Data:** In order to answer this question, data related to movie ratings and movie revenue will be needed. This dataset, or datasets, will need to include data related to a large number of movies. This data will also need to cover a given number of years in order to identify overall success of a movie.

*Two datasets have been found related to movie ratings and revenue. Both datasets are available fron openml.org and are public domain. The first dataset is titled “IMDb-Ratings” and includes viewer dating data for 67408 titles. The second dataset is titles “movies” and includes general movie information, including revenue, for 4803 movies.*

The “*IMDb-Ratings” dataset was provided by IMDb and publicly posted on OpenML by* *Elif Ceren Gok . The “movies” dataset is publicly posted on OpenML by* *Sven Eschlbeck.*

**Data Gathering:** The “movies” dataset will be downloaded manually from OpenML and saved locally. “IMDb-Ratings” will be gathered from OpenML using an API.

**Data Analytics Tools and Techniques:** *Identify the appropriate data-analysis technique you will use to analyze the data.* The datasets will be cleaned and merged so that an analysis of the combined dataset can be performed. This analysis may include the use of various plots

**Justification of Tools/Techniques:** *Explain why the data-analysis technique you chose is an appropriate technique to analyze the data collected*. Click here to enter text.

**Application Type, if applicable (select one):**

mobile

web

stand-alone

**Programming/Development Language(s), if applicable:** Python

**Operating System(s)/Platform(s), if applicable:** Windows 11, Jupyter Notebooks

**Database Management System, if applicable:** NA

**Project Outcomes:** *List the key anticipated project outcomes and deliverables in fewer than 500 words*. The outcome of this project is to identify the strength of any correlation between movie viewer ratings and movie revenue as to identify if highly rated movies generate more revenue vs. movies with low ratings.

**Projected Project End Date:** 12/18/2024

**Sources:**   
  
OpenML. (n.d.). https://www.openml.orolsearch?tvpe=data&status=active&id =43113&sort= runs   
  
OpenML. (n.d.-b). https: //www.openml.orq/search?type=data&sort= runs&id=43784&status=active

**Human Subjects or Proprietary Information**

Does your project involve the potential use of human subjects? (Y/N): N

Does your project involve the potential use of proprietary company information? (Y/N): N

**STUDENT SIGNATURE**

**Kenneth Exchange**

**­­­­­­­­­­­­­­­­­­­­­­­­­­­­­\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**By signing and submitting this form, you acknowledge** that any cost associated with the development and execution of your data analytics solution will be your (the student) responsibility.

**TO BE COMPLETED BY AN INSTRUCTOR**

**The capstone topic is approved by an instructor.**

**INSTRUCTOR’S NAME AND SIGNATURE: **

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**INSTRUCTOR APPROVAL DATE: 12/2/2024**

**Project Compliance with IRB (Y/N): Y**