**Data Analytics Capstone Topic Approval Form**

**Student Name:** Laurie Narcisse

**Student ID:** 011392441

**Capstone Project Name:** Netflix Time Series Analysis

**Project Topic**: Forecasting Viewer Engagement Trends for Movies and TV Shows on Netflix

**This project does not involve human subjects research and is exempt from WGU IRB review.**

**Research Question:** How can we effectively forecast viewer engagement trends for movies and TV shows on Netflix using time series analysis?

**Hypothesis**: **Null hypothesis**-. There is no significant relationship between historical viewer engagement data and future viewer engagement trends for movies and TV shows on Netflix. **Alternate Hypothesis**-. Historical viewer engagement data can be used to accurately forecast future viewer engagement trends for movies and TV shows on Netflix.

**Context:***.*Understanding viewer engagement trends is crucial for content creators, producers, and the Netflix platform to optimize content selection, scheduling, and promotion strategies. By analyzing historical viewer engagement data through time series analysis, we aim to develop accurate forecasts of future trends. This analysis will provide valuable insights into audience behavior and preferences, enabling stakeholders to make data-driven decisions to enhance viewer engagement and platform performance.

**Data:** The dataset used for this analysis is sourced from Kaggle's Netflix Movies and TV Shows Dataset [1], which contains detailed information on movies and TV shows available on Netflix. The dataset includes attributes such as viewership counts, ratings, duration watched, release dates, and other relevant time series data. Each entry in the dataset represents a specific title on Netflix, allowing for comprehensive analysis of temporal patterns and trends. The dataset is located at <https://www.kaggle.com/datasets/ashfakyeafi/netflix-movies-and-shows-dataset?resource=download>

**Data Gathering:** The data will be collected from the Kaggle dataset [1], which provides historical viewer engagement metrics for movies and TV shows on Netflix. The dataset covers a specific time period and includes time series information, enabling us to analyze trends over time.

**Data Analytics Tools and Techniques**: The primary data-analysis technique for this project will be time series analysis, specifically autoregressive integrated moving average (ARIMA) modeling. ARIMA models are well-suited for forecasting time series data by capturing temporal dependencies and patterns in the data [2].

**Justification of Tools/Techniques:** ARIMA modeling is appropriate for this analysis because it can capture both the trend and seasonality present in time series data, allowing for accurate forecasting of future viewer engagement trends. By fitting ARIMA models to historical viewer engagement data, we can generate forecasts and assess their predictive performance [2].

The planned tools to use include Python and Jupyter Notebooks.

Jupyter Notebooks will serve as the primary platform for conducting essential tasks such as data cleaning, analysis, visualization, and time series forecasting to predict future fluctuations in the stock price of the associated tech companies. These tasks will be accomplished using the Python programming language, leveraging its extensive array of packages and libraries specifically designed to streamline data analysis and model development processes.

**Project Outcomes**: Key anticipated outcomes and deliverables include:

* Development of ARIMA models to forecast viewer engagement trends for movies and TV shows on Netflix.
* Identification of significant factors influencing viewer engagement trends, such as seasonality, release dates, and promotional activities.
* Evaluation of forecast accuracy and model performance through statistical measures and validation techniques.
* Insights into temporal patterns and trends in viewer engagement, facilitating strategic decision-making for content selection and promotion.
* A final report summarizing findings, conclusions, and actionable insights for stakeholders in the entertainment industry.

**Projected Project End Date**: 3/31/2024

**Sources**:

[1] Kaggle. (n.d.). Netflix Movies and TV Shows Dataset. Retrieved from https://www.kaggle.com/datasets/ashfakyeafi/netflix-movies-and-shows-dataset?resource=download

[2] Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice (2nd ed.). OTexts. <https://otexts.com/fpp2/>

[3] Mitta, S. N. (2023, June 26). Time Series Analysis of Netflix Stocks with Pandas. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2023/06/time-series-analysis-of-netflix-stocks-with-pandas/>

[4] Wei, W. W. S. (2006). Time series analysis: univariate and multivariate methods (2nd ed.). Pearson Education.

[5] Pierre, S. (2022, October 12). A guide to time series analysis in Python. Built In. https://builtin.com/data-science/time-series-python

[6] Bramhani, G., Indla, V. D. S., Munnuru, B., & Bhuvaneswari, M. (2024). Crunching the Data: A Machine learning approach to Netflix stock price analysis. ResearchGate. <https://doi.org/10.5281/zenodo.10633595>

[7] Wilson, G. T. (2016). Time Series Analysis: Forecasting and Control, 5th Edition, by George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel and Greta M. Ljung, 2015. Published by John Wiley and Sons Inc., Hoboken, New Jersey, pp. 712. ISBN: 978-1-118-67502-1. Journal of Time Series Analysis, 37(5), 709–711. <https://doi.org/10.1111/jtsa.12194>

**Course Instructor Signature/Date:**

The research is exempt from an IRB Review.

An IRB approval is in place (provide proof in appendix B).

Course Instructor’s Approval Status: Approved

Date: Click here to enter a date.

Reviewed by:

Comments: Click here to enter text.