**Data Management and 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 beneath each number will expand as needed.*

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

**Student Name:** Evan Boyd

**Student ID:** 001214997

**Capstone Project Name:** Predicting Life Expectancy from Socio-Econominc Factors

**Project Topic**: Life Expectancy

**Research Question:** *Summarize one question or decision you will answer by collecting and analyzing a set of data.*How does corruption effect life expectancy ?

**Hypothesis:** More corrupt nations will have lower life expectancy.

**Context:** *Explain in 500 words or fewer why the situation or question would benefit from data analysis.*Finding what factors correlate to a long and healthy life could provide many useful insights. If we can isolate disease as being a larger bane to life expectancy than nutrition then efforts can be focused accordingly. If we can find that C02 output correlates to a shorter life we can organize society to reduce that output. Finding a correlation between corruption and life expectancy may push leaders to focus on corruption as a way to raise the standard of living in their nations.

**Data:** *Identify data you will need to collect that are relevant to the situation or question.*Life Expectancy and Socio-Economic data from countries around the world.

*If an existing dataset will be used, describe the dataset. The data set contains 10 years of life expectancy and socio-economic data from 174 countries in 6 geographic regions.*

*Explain who owns the data and why you are allowed to use the data for your capstone project.* The World Bank owns the data. The use of the data in this form falls under their license agreement. Creative Commons Attribution 4.0 International License with additional terms that do not pertain to this use case.

*Note: If you are using restricted information, please have the Third-Party Authorization Form signed by an authorized agent on behalf of the data owner. The data owner’s legal name is required on the form.*

**Data Gathering:** *Describe the data-gathering methodology you will use to collect data.*I will gather my data from kaggle who acquired it via the world bank. The data is in csv format.

**Data Analytics Tools and Techniques**: *Identify the appropriate data-analysis technique you will use to analyze the data.* I will uitilize python and a variety of libraries to preform analysis. I will incorporate graphs, heatmaps, and tabular data to visualize my data to assists with analysis. I’ll implement random forests tree, linear regression, or stochastic gradient descent to build a model from the data that predicts life expectancy. I will use a Z-test to provide a Z-statistic and p-value to help communicate the statistical signifigance of my findings.

**Justification of Tools/Techniques:** *Explain why the data-analysis technique you chose is an appropriate technique to analyze the data collected*. The techniques and tools above are proven methods of approaching exploratory data analysis. Finding the p-value will help me in proving my hypothesis indicating strong(or weak) evidence against the null hypothesis.

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

Mobile

Web

Stand-alone

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

**Operating System(s)/Platform(s), if applicable:** Jupyter notebook

**Database Management System, if applicable:** Click here to enter text.

**Project Outcomes:** *List the key anticipated project outcomes and deliverables in fewer than 500 words*. The project will encompass a jupyer notebook that lays out my data exploration process steps, and my method of model building. I will include descriptions of my thought process and explain the evidence I provide. At the end, my project should provide insight into the effects corruption plays in life expectancy.

**Projected Project End Date:** 6/30/2024

**Sources:** Chavan, S. S. (2023, September 5). *Life expectancy & socio-economic (world bank)*. Kaggle. https://www.kaggle.com/datasets/mjshri23/life-expectancy-and-socio-economic-world-bank/data

**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**

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

**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 FILLED BY A COURSE INSTRUCTOR**

**The capstone topic is approved by a course instructor.**

**COURSE INSTRUCTOR’S NAME AND SIGNATURE: **

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

**COURSE INSTRUCTOR APPROVAL DATE: 6/18/2024**

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