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.

Student Name: [redacted]

Student ID: [redacted]

Capstone Project Name: Analysis of World Sustainability Dataset

Project Topic: Examination sustainability levels amongst nations with varying degrees of wealth.

Research Question: Do nations with higher GDP's operate in a more sustainable way compared to countries with much lower GDP's?

> **Hypothesis:** The average sustainability score is significantly higher in nations with high GDPs compared to those with low GDPs.

Context: The world's climate is changing quickly and, as a result, a global emphasis on sustainability has increased in popularity. It would be very useful to see which nations' practices result in effective sustainability and how that relates to their GDP. If it's true that wealthier nations are more sustainable, it would challenge the idea that more wealth equals more environmental damage. Furthermore, wealthier nations could potentially support poorer countries in best practices to improve global sustainability.

The connection between wealth and sustainability could also provide a road map for developing countries to become more economically robust and eventually adopt sustainable practices without harming the well-being of their populations. Increasing public awareness about this connection will also likely lead to an increase in political will for effective policy change.

Data: I will be using the following dataset from Kaggle:

- World Sustainability Dataset (2021) – World Bank Data Bank; United Nations Statistics; Our World in Data (https://www.kaggle.com/datasets/truecue/worldsustainabilitydataset?select=WorldSustainabilityDataset.csv)

Data Gathering: I will download the data directly from the Kaggle website and clean the data as needed.

Data Analytics Tools and Techniques:

Tools:

- **Python:** For data preprocessing, visualization, and statistical analysis.
- **Excel:** For preliminary data exploration and visualization.
- **Tableau:** For creating interactive visualizations and dashboards.

Techniques:

- **Correlation Tests:** To determine if there's a significant association between GDP and sustainability scores.
- **T-tests:** To compare the means of sustainability scores between high GDP and low GDP nations.
- **Regression Analysis:** To evaluate the strength and direction of the relationship between GDP and sustainability scores.
- **Confidence Intervals:** To estimate the range within which the true mean sustainability score for high GDP and low GDP countries lies.

Justification of Tools/Techniques:

- **Python:** With Pythons many libraries, it will be useful for data handling, visualization, and statistical tests.
- **Excel:** Excel is user-friendly and good for a quick look at data. I'll be able to easily visualize and summarize data.
- **Tableau:** Tableau is great for visualizing data. Its interactive features will let me see the GDP-sustainability relationship clearly.
- **Correlation Tests:** These tests will show if and how much GDP and sustainability scores are related.
- **T-tests:** These will help me see if sustainability scores really differ between high and low GDP countries.
- **Regression Analysis:** This will help me predict how changes in GDP could affect sustainability scores.
- **Confidence Intervals:** These will give me a range where the true average sustainability score might lie, helping me understand the data's accuracy.

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

Database Management System, if applicable: N/A

Project Outcomes: I aim to understand the link between a nation's GDP and its sustainability practices. I'll provide intuitive visualizations and predictive insights for future trajectories. My goal is to equip stakeholders with clear insights and tools to drive global sustainability.

Projected Project End Date: [redacted], 2023

Sources:

- DataBank | The World Bank. (n.d.). https://databank.worldbank.org/home
- United Nations Statistics Division. (n.d.). SDG Indicators — SDG Indicators. https://unstats.un.org/sdgs/indicators/indicators-list/
- Political regime. (n.d.). Our World in Data. https://ourworldindata.org/grapher/political-regime

---

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

[redacted]

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 SIGNATURE:

Jim Ashe, Ph.D. Mathematics

COURSE INSTRUCTOR APPROVAL DATE:

Wednesday, [redacted] 2023

Project Compliance with IRB (Y/N): Y