CS 2704 – Final Project Proposal

Hypothesis: GDP per capita and Happiness Score correlation

1. Group Members

- Malek Karoui
- Tusshar Gopaul
- Tishant Doorgapersand
- Fady Haytham Elgohary

2. Chosen Dataset

- **Dataset Name**: World Happiness Report 2019
- **Source**: Kaggle World Happiness Report Dataset

This dataset includes data for **156 countries**, with variables such as **GDP per Capita**, **Social Support**, **Life Expectancy**, and **Freedom to Make Life Choices**, all of which contribute to the calculation of the **Happiness Score**.

3. GitHub Repository

- **Repository URL**: https://github.com/Malek-karoui20/Python-

4. Hypothesis

- **Statement**:
  "We hypothesize that **GDP per Capita** has a significant positive correlation with **Happiness Score**."
- **Relevance**:
  This hypothesis is important because it explores the relationship between wealth and happiness. Understanding this correlation can provide insights into how economic factors affect the well-being of a nation's population. If supported, the hypothesis could suggest that improving the economic situation of a country might contribute to increasing its citizens' happiness levels.

5. Plan for Testing the Hypothesis

- **Descriptive Analysis**:
  - **Summary Statistics**:
    We will calculate the **mean**, **median**, **standard deviation**, and other summary statistics for **GDP per Capita** and **Happiness Score**. This will provide insight into the central tendencies and spread of the data.
  - **Visualizations**:
    - A **scatter plot** will be created to visualize the relationship between **GDP per Capita** (X-axis) and **Happiness Score** (Y-axis).
    - A **correlation heatmap** will help visualize the correlation between various variables in the dataset, including **GDP per Capita** and **Happiness Score**.
- **Predictive Analysis**:
  - We will perform **Linear Regression** to predict the **Happiness Score** based on **GDP per Capita**. The linear regression model will allow us to examine the predictive power of GDP in determining happiness levels.
- **Statistical Tests**:
  - **Pearson correlation** will be computed to assess the strength and direction of the linear relationship between **GDP per Capita** and **Happiness Score**. A **p-value** will be used to test the hypothesis and determine statistical significance.

6. Data Snippet

Here is a preview of the dataset showing key variables:

# Overall rank	Country or region	# Score	# GDP per capita	# Social support	# Healthy life expe...
1 – 156	156 unique values	2.85 – 7.77	0 – 1.68	0 – 1.62	0 – 1.14
1	Finland	7.769	1.340	1.587	0.986
2	Denmark	7.600	1.383	1.573	0.996
3	Norway	7.554	1.488	1.582	1.028
4	Iceland	7.494	1.380	1.624	1.026
5	Netherlands	7.488	1.396	1.522	0.999
6	Switzerland	7.480	1.452	1.526	1.052
7	Sweden	7.343	1.387	1.487	1.009
8	New Zealand	7.307	1.303	1.557	1.026
9	Canada	7.278	1.365	1.505	1.039
10	Austria	7.246	1.376	1.475	1.016

7. Expected Output

- **Anticipated Results**:
  We expect to find a **positive correlation** between **GDP per Capita** and **Happiness Score**, indicating that wealthier countries tend to have higher happiness scores. We anticipate that **GDP per Capita** will be a significant predictor of the **Happiness Score** in our regression model.
- **Conclusion**:
  If our hypothesis is supported, it could suggest that policies aimed at increasing economic growth might contribute to improving national well-being. On the other hand, if the correlation is weak or negative, it may suggest that other factors, such as social support or freedom of choice, play a more significant role in determining happiness.