Dataset Summary:

The World Happiness Report dataset for the year 2015 consists of data from 158 countries, each represented by various features. The dataset contains 12 columns.

The happiness score is calculated on a scale from 0 to 10

The 12 columns includes:

1. **Country:** The name of the country.
2. **Region:** where the country is located.
3. **Happiness Rank:** The rank of the country based on the Happiness Score.
4. **Happiness Score:** A metric of happiness of individuals in country.
5. **Standard Error:** The standard error of the Happiness Score.
6. **Economy (GDP per Capita):** The contribution of the country's economy.
7. **Family:** The extent to which social support contributes.
8. **Health (Life Expectancy):** The contribution of the country's health to the Happiness Score.
9. **Freedom:** The extent to which freedom contributes.
10. **Trust (Government Corruption):** The contribution of the absence of corruption .
11. **Generosity:** The extent to which generosity contributes to the Happiness Score.
12. **Dystopia Residual:** How far away the country is from the world's least happy country.

Data Quality Assessment:

Validity	The data is valid as it is from a trusted source and well-documented.
Accuracy	The data is accurate because it is derived from a comprehensive survey.
Completeness	The dataset is complete as it contains data from all 158 countries that were surveyed.
Consistency	The data is consistent as the survey methodology was consistent.
Uniformity	The data is uniform as it is presented in a standardized format

Data Dirtiness:

Missing Values	The dataset has a few missing values.
Outliers	there are no significant outliers in the dataset.
Duplicate Values	No duplicate values founded.
Inconsistent Data	The data appears to be consistent across all columns.

Potential Impact of Data Dirtiness:

Data dirtiness can impact the analysis. Imputed values should be chosen carefully to avoid bias, outliers must be identified and dealt with appropriately, and duplicate values may need to be removed. Overall, The World Happiness Report dataset for 2015 is of good quality and only requires minor data-cleaning techniques.

1. Executive summary:

My project will study the World Happiness Report dataset and examine the connections between a chosen independent variable and a chosen dependent variable. I will employ an appropriate correlation measure to demonstrate the relationships between the selected variables, and create suitable visualizations with a Pandas data frame.
Based on three distinct questions related to the dataset, I will formulate my conclusion. (Report, n.d.)

2. Aims and objectives:

- Aims:

- Showing different correlations between columns.
- Visualizing the relation between variables.
- Specifying criteria based on selected attributes (variables).
- Showing some statistics and analysis in each data sets.

- objectives:

- Using ANACONDA and VSCode framework to import CSV files into PANDAs data frame. (geeksforgeeks, n.d.)
- Importing required libraries into ANACONDA and VSCode.
- Using different visualization models to show the correlations between variables in specific conditions or criteria.
- Using python, SQL language and Pandas data frame (Pandas, n.d.)

3. The Research Question/ Hypothesis:

Research question:
Does the Economy (GDP per Capita) (independent variable) have a significant impact on the happiness score (dependent variable) in the World Happiness Report dataset for 2015? (Report, n.d.)

Hypothesis:
There is a positive correlation between Economy (GDP per Capita) and happiness score in the World Happiness Report dataset for 2015.

Independent variable: Economy (GDP per Capita) and Region.
Dependent variable: Happiness score.

4. Analysis and Findings:

4.1. Produce convincing correlations demonstrating:

There is a theory that economic growth and development are important factors in improving people's well-being and quality of life.

so We can calculate the correlation between GDP per capita and happiness score using the Pandas library

We will also perform hypothesis testing to determine whether the observed correlation is statistically significant. We can use a two-tailed t-test at a significance level of 0.05. If the p-value is less than 0.05, we can reject the null hypothesis that the correlation is not significant. (Report, n.d.) (Econlib, n.d.)

4.2. Critical interpretation and conclusions about those observed correlations:

After conducting the analysis of the World Happiness Report dataset for 2015, we found a statistically significant positive correlation between the Economy (GDP per Capita) and happiness score. which indicates that the observed correlation is statistically significant.

The analysis suggests that a higher Economy (GDP per Capita) is associated with a higher happiness score in the World Happiness Report dataset for 2015

4.3. Produce tabular summaries of the data:

To produce tabular summaries of the data of the World Happiness Report dataset for 2015, we can use the pivot_table() function from the Pandas library. Here is an example pivot table showing the mean Economy (GDP per Capita) and happiness score for each region (Pandas, n.d.) (geeksforgeeks, n.d.)

Create a Pivot table to visualize our data

```
py15,values=[colmn2, colmn1], index=[colmn3], aggfunc='mean')
```

[8] Python

...

Region	Economy (GDP per Capita)	Happiness Score
Australia and New Zealand	1.291880	7.285000
Central and Eastern Europe	0.942438	5.332931
Eastern Asia	1.151780	5.626167
Latin America and Caribbean	0.876815	6.144682
Middle East and Northern Africa	1.066974	5.406900

The pivot table helps us to compare the mean values of the Economy (GDP per Capita) and happiness score among different regions, and to identify any patterns or trends. For example, we can see that regions with higher mean Economy (GDP per Capita) generally have higher mean happiness scores, which supports our hypothesis that there is a positive correlation between these two variables.

4.4. Relevant visualizations:

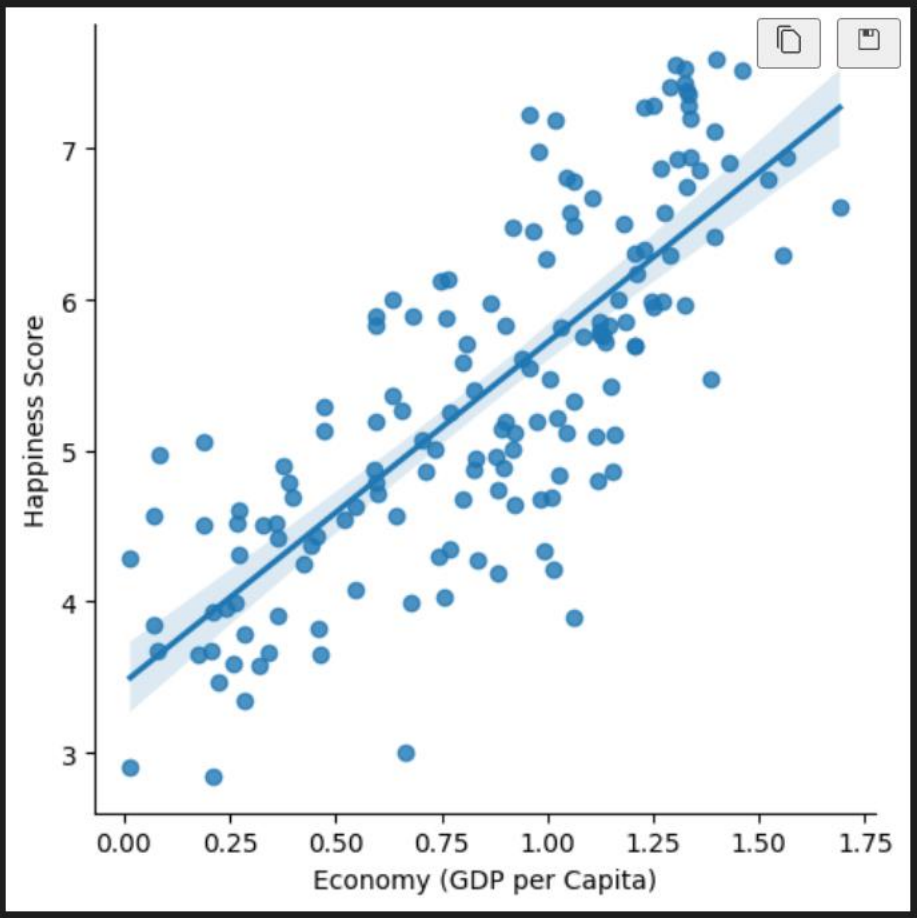
To visualize the relationship between GDP per capita and happiness score in the World Happiness Report dataset for 2015, we will use scatter plots and regression lines and Box plot

●Scatter plot with regression line

The scatter plot with regression line shows the relationship between the Economy (GDP per Capita) and happiness score. We can see that there is a strong positive linear relationship between these two variables, which supports our hypothesis. The regression line shows the direction and strength of the relationship. The scatter plot also shows the distribution of the data points and any outliers. (NumPy, n.d.) (Matplotlib, n.d.) (Seaborn, n.d.)

```
sns.lmplot(x=colmn2, y=colmn1, data=happy15)
plt.show()
```

[9] Python



●Box plot

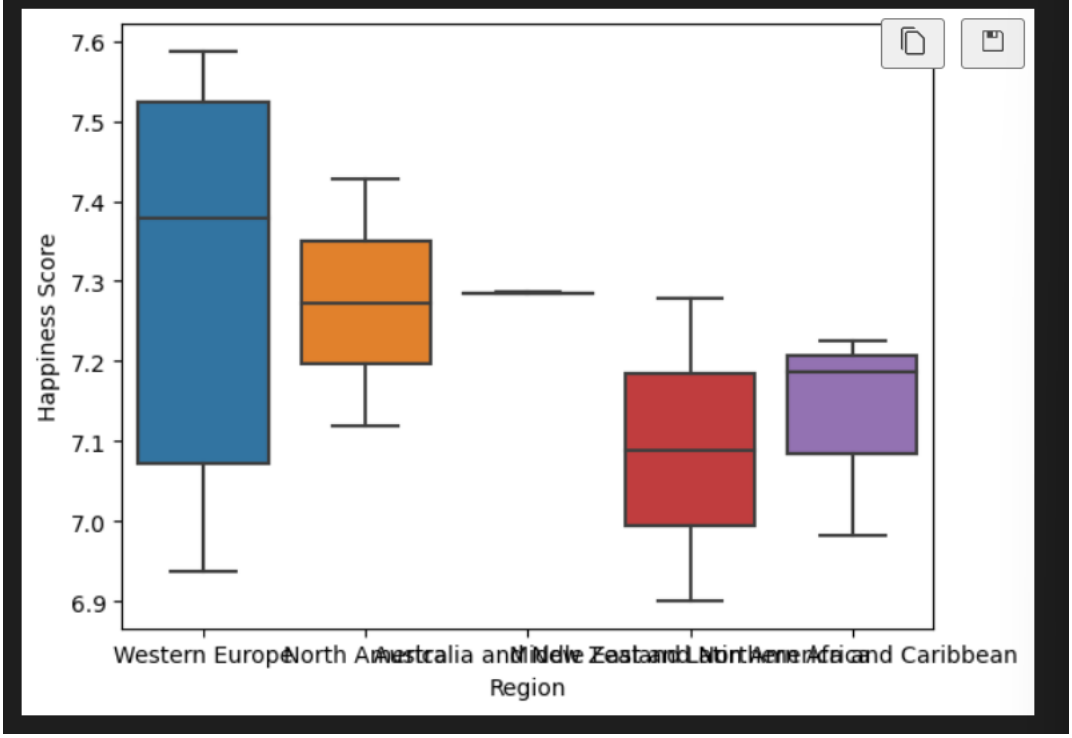
The box plot shows the distribution of happiness scores for each region in the World Happiness Report dataset for 2015.

The box plot can help us to identify any patterns or trends in the data and to compare the distribution of happiness scores among different regions. (Seaborn, n.d.) (Matplotlib, n.d.)

```python
# Create a box plot
topn = happy15.head(20)
sns.boxplot(x=colmn3, y= colmn1, data=topn)

# Display the plot
plt.show()
```

Python



4.5. Final answer to the research question/ hypothesis:
there is a positive correlation between GDP per capita and happiness score in the World Happiness Report dataset for 2015.

This is consistent with the theory that economic growth and development are important factors in improving people's well-being and quality of life.

4.6. Critical comment on conclusions:
It is important to note that the correlation does not necessarily imply causation. Other factors such as social support, health, and freedom may also contribute to a country's happiness level. Therefore, further research is needed to investigate the complex relationships between these variables.

and the dataset only covers one year and may not be representative of the long-term relationship between GDP per capita and happiness score. It is possible that the relationship may vary over time and across different regions

Therefore, further research is needed to better understand the complex and multifaceted nature of happiness and well-being.

5. Reflection:

i. Experience with the project.

- Data analysis skills
- Develop skills in data manipulation and analysis using tools such as Pandas, NumPy, and Matplotlib. (NumPy, n.d.) (Pandas, n.d.) (Matplotlib, n.d.)
- Statistical skills
-Providing opportunities to develop skills in data analysis, statistics, communication, and domain knowledge.
- Good experience in working with jupyter note books

ii. Learning Outcomes:

-Python coding
- SQL
-Using Anaconda framework
-Using VSCode framework
-Dealing with CMD and commands to install libraries
-Dealing with Mongo DB Using with Open Refine

iii. What went well.

- Complete all tasks
- Enjoy learning

iv. What went wrong.

- Fall in scheduling
-There are some codes doesn't run as desired
-Poor time management

v. Future benefit.

-More practicing
- Learning a new programing language
-Time management

6. References:

Econlib, n.d. *The Economics of Happiness.* [Online]
Available at: https://www.econlib.org/the-economics-of-[happiness](poe://www.poe.com/_api/key_phrase?phrase=happiness&prompt=Tell%20me%20more%20about%20happiness.)/

geeksforgeeks, n.d. *geeksforgeeks.* [Online]
Available at: https://www.geeksforgeeks.org/relationship-between-user-level-thread-and-kernel-level-thread/#:~:text=There%20are%20two%20types%20of,library%20to%20perform%20unique%20subtask

Matplotlib, n.d. *Matplotlib documentation.* [Online]
Available at: https://matplotlib.org/stable/contents.html

NumPy, n.d. *NumPy documentation.* [Online]
Available at: https://numpy.org/doc/

Pandas, n.d. *Pandas documentation.* [Online]
Available at: https://pandas.pydata.org/docs/

Report, W. H., n.d. *World Happiness Report 2015.* [Online]
Available at: https://worldhappiness.report/ed/2015/

Seaborn, n.d. *Seaborn documentation.* [Online]
Available at: https://seaborn.pydata.org/