جامعة حمدان بن محمد الذكية
Hamdan Bin Mohammed Smart University

General Education Department

Project Fall 2024

Course	**Introduction to Programming - PROG101**		
Assessment Method	Individual Project		
Date of Assessment		**Duration / Deadline(s)**	10 Days
Maximum Mark	**100**	**Percentage of Final Grade**	**20%**
Student ID		**Student Name**	

Instructions to Students

1. **All work must be genuine**
2. **Copying from any source is not allowed.**
3. **Each student must submit the project deliverables individually through the VLE (Virtual Learning Environment).**

Academic Honesty Statement

- **Students are expected to uphold integrity by avoiding all forms of academic dishonesty, as outlined in HBMSU procedures and guidelines, as well as instructions from HBMSU personnel.**
- **Any student found guilty of academic dishonesty may face disciplinary actions as specified in the Student and Academic Regulations.**

Question (Section) No.	Deliverable Report (PDF), & Python Code in PDF	Presentation in PPT	Total	Percentage
Marks Allocated	**70**	**30**	**100**	**20%**
Awarded Marks				

Learner's Name: _____

جامعة حمدان بن محمد الذكية
Hamdan Bin Mohammed Smart University

PROJECT INTRODUCTION

The purpose of this data analysis project is to explore factors influencing global happiness by analyzing the World Happiness dataset. This dataset, sourced from open data platforms, includes metrics such as Country, Region, Happiness Score, Economy, Health, Freedom, and others related to well-being.

The aim is to uncover meaningful insights into how various socioeconomic indicators correlate with happiness levels across different countries and regions. By analyzing relationships between factors like economic stability, freedom, and health, this project seeks to provide a data-driven understanding of what contributes to national happiness. This can aid policymakers, researchers, and social scientists in identifying trends and areas for improvement to enhance citizens' quality of life.

ANALYSIS OBJECTIVES

DATA ACQUISITION AND CLEANING

Statistical Analysis Question

1. *What is the average Happiness Score across all countries?*

Conditional Filtering Questions

2. *How many countries have an Economy score above 1.0 and a Freedom score above 0.5?*
3. *Which countries have a Health score below 0.1 but a Happiness Score below 5?*

Grouping Data Question

4. *What is the average Happiness Score per region?*

Sorting Data Question

5. *What are the top 5 happiest countries in descending order of their Happiness Score?*

Combination Questions

6. *In the region with the highest average Happiness Score, what are the top 5 countries by Happiness Score?*
7. *List countries with above-average Economy and Health scores, sorted by Freedom in descending order.*

Visualization Questions

8. *How does the average Happiness Score vary across regions?*
   A bar chart for this comparison can provide a visual reference for regional happiness trends.

9. *Is there a relationship between Economy and Happiness Score?*
   A scatter plot showing this relationship can reveal the correlation between economic strength and happiness.

10. *What is the proportion of countries in each region?*
    A pie chart of regional distribution can highlight data representation across regions and support region-specific decisions.

Learner's Name: _____

Data Cleaning

Data cleaning is essential for the World Happiness dataset to ensure accurate and reliable analysis. This process involves identifying and handling any missing values, duplicates, or irrelevant data that could skew the results or produce misleading insights. For instance, missing values in columns like Happiness Score, Economy, or Health could result in inaccurate averages or correlations, leading to incorrect conclusions. Cleaning also involves removing any unnecessary columns that don't contribute to our analysis, streamlining the dataset for focused exploration. Furthermore, outliers or inconsistencies in numerical values need to be managed, as these could distort statistical calculations, impacting decisions based on economic or health-related factors. Overall, data cleaning ensures that the dataset accurately reflects meaningful patterns and relationships, making the insights more reliable and actionable for understanding global happiness trends.

Load data

```
[ ] import pandas as pd

    # Load dataset
    df = pd.read_csv('World Happiness.csv')
    df.head()
```

	Country	Year	GDP	Happiness Rank	Happiness Score	Economy	Family	Health	Freedom	Generosity	Corruption	Dystopia	Job Satisfaction	Region
0	Norway	2008	7.370	1	7.537	1.616463	1.533524	0.796667	0.635423	0.362012	0.315964	2.277027	94.6	Western Europe
1	Denmark	2009	7.540	2	7.522	1.482383	1.551122	0.792566	0.626007	0.355280	0.400770	2.313707	93.5	Western Europe
2	Iceland	2010	7.647	3	7.504	1.480633	1.610574	0.833552	0.627163	0.475540	0.153527	2.322715	94.5	Western Europe
3	Switzerland	2011	7.620	4	7.494	1.564980	1.516912	0.858131	0.620071	0.290549	0.367007	2.276716	93.7	Western Europe
4	Finland	2012	7.705	5	7.469	1.443572	1.540247	0.809158	0.617951	0.245483	0.382612	2.430182	91.2	Western Europe

Clean data

```
# Drop rows with significant missing values
df.dropna(inplace=True)

# Convert relevant columns to appropriate data types if necessary
df['Year'] = df['Year'].astype(int)

# Check for duplicates and remove them
df.drop_duplicates(inplace=True)

# Display sample cleaned dataset
df.head()
```

	Country	Year	GDP	Happiness Rank	Happiness Score	Economy	Family	Health	Freedom	Generosity	Corruption	Dystopia	Job Satisfaction	Region
0	Norway	2008	7.370	1	7.537	1.616463	1.533524	0.796667	0.635423	0.362012	0.315964	2.277027	94.6	Western Europe
1	Denmark	2009	7.540	2	7.522	1.482383	1.551122	0.792566	0.626007	0.355280	0.400770	2.313707	93.5	Western Europe
2	Iceland	2010	7.647	3	7.504	1.480633	1.610574	0.833552	0.627163	0.475540	0.153527	2.322715	94.5	Western Europe
3	Switzerland	2011	7.620	4	7.494	1.564980	1.516912	0.858131	0.620071	0.290549	0.367007	2.276716	93.7	Western Europe
4	Finland	2012	7.705	5	7.469	1.443572	1.540247	0.809158	0.617951	0.245483	0.382612	2.430182	91.2	Western Europe

Data cleaning ensures accurate and reliable results. Steps include handling missing values, removing duplicates, and adjusting formats where necessary.

DATA AND EXPLORATORY ANALYSIS

Learner's Name: _____

Statistical Analysis Question

1. *What is the average Happiness Score across all countries?*

```
[3]  avg_happiness = df['Happiness Score'].mean()
     print("Average Happiness Score:", avg_happiness)

Average Happiness Score: 5.357874172417219
```

Conditional Filtering Questions

2. *How many countries have an Economy score above 1.0 and a Freedom score above 0.5?*

```
[ ]  high_scores = df[(df['Economy'] > 1.0) & (df['Freedom'] > 0.5)]
     print("Number of countries:", len(high_scores))

Number of countries: 33
```

3. *Which countries have a Health score below 0.1 but a Happiness Score below 5?*

```
low_happiness_health = df[(df['Health'] < 0.1) & (df['Happiness Score'] < 5)]
low_happiness_health[['Country', 'Happiness Score', 'Health']]
```

	Country	Happiness Score	Health
103	Sierra Leone	4.709	0.005565
125	Ivory Coast	4.180	0.048642
134	Chad	3.936	0.041135
136	Lesotho	3.808	0.000000
137	Angola	3.795	0.049869
152	Central African Republic	2.693	0.018773

Grouping Data Question

4. *What is the average Happiness Score per region?*

```
avg_happiness_by_region = df.groupby('Region')['Happiness Score'].mean()
print(avg_happiness_by_region)

Region
Africa            4.254581
Asia-Pacific      5.358326
Eastern Europe    5.498952
Europe            4.096000
Latin America     5.957818
North America     7.154500
Western Europe    6.880474
Name: Happiness Score, dtype: float64
```

Sorting Data Question

Learner's Name: _____

5. What are the top 5 happiest countries in descending order of their Happiness Score?

```
[ ] top_happy_countries = df.sort_values(by='Happiness Score', ascending=False).head(5)
    top_happy_countries[['Country', 'Happiness Score']]
```

	Country	Happiness Score
0	Norway	7.537
1	Denmark	7.522
2	Iceland	7.504
3	Switzerland	7.494
4	Finland	7.469

Combination Questions

6. In the region with the highest average Happiness Score, what are the top 5 countries by Happiness Score?

```
# Find the region with the highest average Happiness Score
top_region = avg_happiness_by_region.idxmax()
top_countries_in_region = df[df['Region'] == top_region].sort_values(by='Happiness Score', ascending=False).head(5)
top_countries_in_region[['Country', 'Happiness Score']]
```

	Country	Happiness Score
6	Canada	7.316
13	United States	6.993

7. List countries with above-average Economy and Health scores, sorted by Freedom in descending order.

```
avg_economy = df['Economy'].mean()
avg_health = df['Health'].mean()
top_countries = df[(df['Economy'] > avg_economy) & (df['Health'] > avg_health)].sort_values(by='Freedom', ascending=False)
top_countries[['Country', 'Economy', 'Health', 'Freedom']].head(6)
```

	Country	Economy	Health	Freedom
0	Norway	1.616463	0.796667	0.635423
2	Iceland	1.480633	0.833552	0.627163
1	Denmark	1.482383	0.792566	0.626007
3	Switzerland	1.564980	0.858131	0.620071
4	Finland	1.443572	0.809158	0.617951
7	New Zealand	1.405706	0.816760	0.614062

DATA ANALYSIS - VISUALIZATION

Visualization Questions

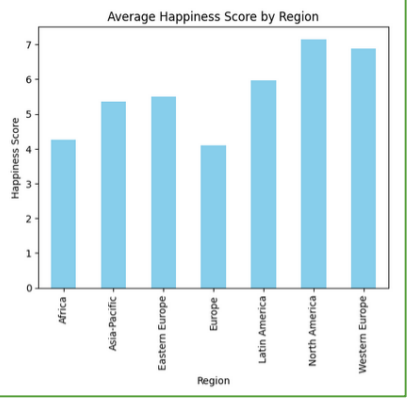
8. How does the average Happiness Score vary across regions?

A bar chart for this comparison can provide a visual reference for regional happiness trends.

Learner's Name: _____

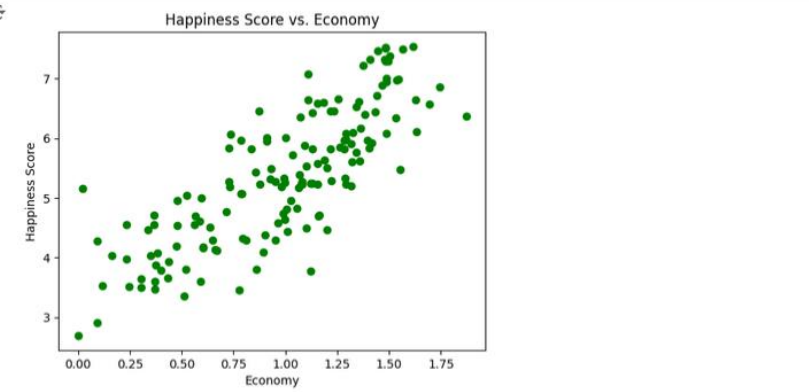
```
[ ] import matplotlib.pyplot as plt

    avg_happiness_by_region.plot(kind='bar', color='skyblue')
    plt.title('Average Happiness Score by Region')
    plt.xlabel('Region')
    plt.ylabel('Happiness Score')
    plt.show()
```



9. *Is there a relationship between Economy and Happiness Score?*
   A scatter plot showing this relationship can reveal the correlation between economic strength and happiness.

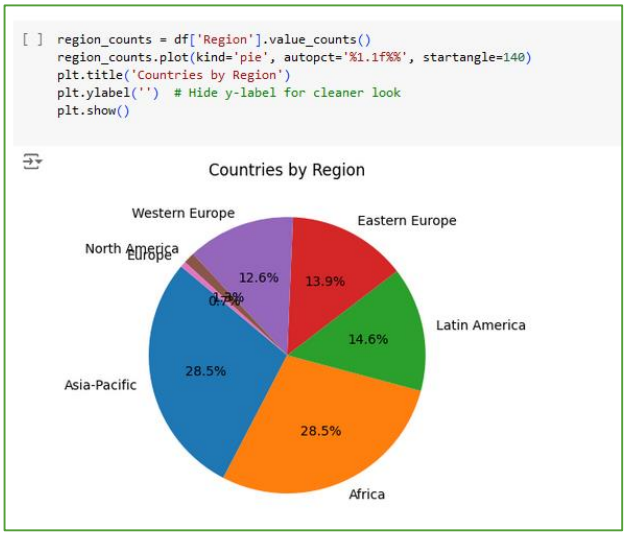
```
[ ] plt.scatter(df['Economy'], df['Happiness Score'], color='green')
    plt.title('Happiness Score vs. Economy')
    plt.xlabel('Economy')
    plt.ylabel('Happiness Score')
    plt.show()
```



10. *What is the proportion of countries in each region?*
    A pie chart of regional distribution can highlight data representation across regions and support region-specific decisions.

Learner's Name: _____

```
[ ] region_counts = df['Region'].value_counts()
    region_counts.plot(kind='pie', autopct='%1.1f%%', startangle=140)
    plt.title('Countries by Region')
    plt.ylabel('')  # Hide y-label for cleaner look
    plt.show()
```



EXECUTIVE SUMMARY

This project explores global happiness metrics by analyzing the World Happiness dataset, which includes key indicators such as Happiness Score, Economy, Health, Freedom, and Social Support across different countries and regions. The aim is to uncover insights into the factors that contribute most significantly to national happiness, allowing policymakers and researchers to understand potential areas for societal improvements. Through a series of ten analysis questions, we performed statistical assessments, conditional filtering, grouping, and sorting to explore relationships between happiness and various socio-economic indicators. Our findings reveal that economic stability and health are consistently associated with higher happiness scores, but the role of freedom and social support varies by region. For instance, countries with high scores in freedom and economy tend to have higher overall happiness, suggesting that these are core drivers of well-being.

Visualizations using bar charts, scatter plots, and pie charts helped us illustrate key findings. The data shows a clear correlation between Economy and Happiness Score, and regions like the Americas and Europe have higher average happiness scores compared to other areas. Additionally, a significant proportion of countries in these regions demonstrate a balanced performance in health and economic indicators. The analysis concludes that improving citizens' quality of life requires a holistic approach that not only strengthens the economy but also promotes personal freedom and public health. Policymakers can leverage these insights to prioritize initiatives that foster economic growth, social support, and health services, which may ultimately contribute to a more satisfied and thriving population.

REFERENCES

1. Data.gov. (n.d.). USA Government Open Data Portal. Retrieved from https://www.data.gov
2. Kaggle. (n.d.). Dataset Repository. Retrieved from https://www.kaggle.com/datasets
3. Happiness Research Institute. (Various years). World Happiness Data. Accessed from various sources including Kaggle and national data repositories for global well-being analysis.

Learner's Name: _____