

## Data Collection and Preprocessing Phase

Date	15 May 2024
Team ID	739845
Project Title	WORLD HAPPINESS REPORT
Maximum Marks	6 Marks

### Data Exploration and Preprocessing Template

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	Description
Data Overview	Basic statistics, dimensions, and structure of the data.
Univariate Analysis	Exploration of individual variables (mean, median, mode, etc.).
Bivariate Analysis	Relationships between two variables (correlation, scatter plots).
Multivariate Analysis	Patterns and relationships involving multiple variables.
Outliers and Anomalies	Identification and treatment of outliers.
<b>Data Preprocessing Code Screenshots</b>	
Loading Data	<pre>df=pd.read_csv('2019.csv')</pre>

	count	mean	std	min	25%	50%	75%	max
Overall rank	156.0	78.50000	45.177428	1.000	30.75000	78.5000	117.25000	156.000
Happiness Score	156.0	5.407096	1.111120	2.853	4.584652	5.3795	6.18403	7.769
GDP per capita	156.0	0.92547	0.388380	0.00	0.67075	0.9689	1.23259	1.884
Scid support	156.0	1.208114	0.399181	0.000	1.05575	1.3715	1.42393	1.654
Healthy life expectancy	156.0	0.735344	0.243104	0.000	0.54575	0.7880	0.88175	1.141
freedom to make life choices	156.0	0.395751	0.142489	0.000	0.30800	0.4170	0.50275	0.631
Generosity	156.0	0.194846	0.095254	0.000	0.10875	0.1775	0.24605	0.566
Perceptions of corruption	156.0	0.110801	0.094538	0.000	0.04700	0.0895	0.14125	0.453

The image shows a Jupyter Notebook interface with a code cell containing the following Python code:

```
sns.pairplot(
    ndf,
    x_vars=["gdp_per_capita", "generosity", "social_support", "freedom_to_make_life_choices"],
    y_vars=["happiness_score", "human_happiness_score"]
)
```

Below the code cell, the output displays four scatter plots arranged horizontally, showing the relationship between the four x-variables and the y-variables. The y-axis for all plots is labeled "Happiness Score" and ranges from 3 to 8. The x-axes are labeled "GDP per capita", "Generosity", "Social support", and "Freedom to make life choices". The legend indicates that the color of the points represents the "Happiness Score", with values 3 (light pink), 4 (pink), 5 (light purple), 6 (purple), and 7 (dark purple).