

Data Collection and Preprocessing Phase

Date	15 May 2024
Team ID	740065
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.
Data Preprocessing Code Screenshots	
Loading Data	<pre>df=pd.read_csv('2019.csv')</pre>

Describe data in dataset

```
df.describe().T
```

		count	mean	std	min	25%	50%	75%	max
Overall rank	156.0	78.500000	45.177428	1.000	39.750000	78.5000	117.25000	156.000	
Happiness Score	156.0	5.407096	1.113120	2.853	4.54450	5.3795	6.18450	7.769	
GDP per capita	156.0	0.905147	0.398389	0.000	0.6075	0.9600	1.23250	1.684	
Social support	156.0	1.208814	0.299191	0.000	1.05575	1.2715	1.45250	1.624	
Healthy life expectancy	156.0	0.725244	0.242124	0.000	0.54775	0.7880	0.88175	1.141	
Freedom to make life choices	156.0	0.392571	0.142089	0.000	0.30800	0.4170	0.50725	0.631	
Generosity	156.0	0.164846	0.095254	0.000	0.10875	0.1775	0.24825	0.566	
Perceptions of corruption	156.0	0.119603	0.094538	0.000	0.04700	0.0875	0.14125	0.453	

pairplot

