### Happiness Index Project Final Report

**Motivation**

Global “happiness scores” are headline material, but the real policy value lies in knowing which levers actually move national well-being. Our project blends the World Happiness Report with basic labour-market data to build a model that classifies each country’s life-satisfaction level (High, Medium, Low) from a concise set of socioeconomic indicators.  By turning the happiness debate into a reproducible machine-learning exercise, we give governments and NGOs a clearer sense of which factors deserve investment.

**Data & Analysis Pipeline**

* **Dataset**

135 countries

Target: Happiness score.

Features:

* GDP per capita
* Social support
* Healthy-life expectancy
* Freedom of choice
* Generosity
* Perceived corruption
* Unemployment rate
* **Pre-processing**
  + Verified ranges and units.
  + Mean-imputed the missing entries.
  + StandardScaler applied to all continuous predictors.
* **Exploration**
  + Descriptive statistics and distribution plots.
  + Pearson/Spearman correlation matrix.
* **Feature Engineering**
  + Created happiness class thresholds: Low ≤ 4.8, Medium 4.8-5.8, High > 5.8.
* **Modelling**
  + **RandomForestRegressor :** Captures non-linear relationships and variable interactions without strong parametric assumptions, making it ideal for mixed socioeconomic data.
  + **RandomForestClassifier:** Ensemble of bootstrapped trees that votes on the most likely class.

**Key Findings**

* Economic strength dominates GDP per capita shows the highest positive correlation with happiness (r = 0.81).
* Health matters almost as much healthy-life expectancy (r = 0.80) rivals GDP as a predictor.
* Community counts social support (r = 0.75) and freedom of choice (r = 0.55) are strong secondary drivers.
* Generosity is surprisingly weak correlation is mild (r = 0.16); perceptions of corruption sit mid-pack (r = 0.45).
* Joblessness erodes well-being unemployment rate is the only factor with a clear negative link (r = -0.20).

**Limitations & Future Work**

* Single-year snapshot: Results capture correlation, not causation, and ignore year-to-year dynamics.
* Omitted variables: Education quality, environmental metrics, inequality, and cultural factors could refine the picture.
* Next steps: Expand to a 10-year panel, test causal paths with fixed-effects models or instrumental variables.