

Feasibility of Combining 2020 GDP and Happiness Index Data for Cross-Country Wellbeing Analysis

Author: Rohan Chawla

This proposal evaluates the feasibility of combining the World GDP 2020 and World Happiness Index 2020 datasets to examine the relationship between economic prosperity and well-being, focusing on G20 countries. The G20 represents over 80% of global GDP and 60% of the population, making it a valuable subset for high-impact, policy-relevant analysis. Both datasets, sourced from Kaggle, reflect global data for 2020 and are well-suited for integration.

Data Quality and Cleaning

The GDP dataset includes GDP per capita, total GDP, and population figures for 184 countries. Although complete, the GDP values contain formatting symbols that require cleaning. The Happiness dataset offers Ladder Scores and contributing dimensions like social support, freedom, and life expectancy for 153 countries. It is well-structured with minimal missing data. Restricting the analysis to G20 countries improves relevance and reduces data noise from smaller economies. The datasets share the country name as a key attribute, though minor discrepancies (e.g., "United States" vs. "United States of America") may require standardization.

Compatibility and Wrangling Techniques

Both datasets are from 2020, use country-level granularity, and are stored in CSV format, ensuring high compatibility. GDP per capita values will be cleaned and converted to floats. Country names will be standardized using strip and replace functions. A left join will merge records on the country column, followed by filtering for G20 entries. Validation will address mismatched or missing data. Aggregation methods, such as calculating average GDP and happiness scores by region or development category, may be used to uncover broader patterns. The OECD Better Life Index will be a backup if merging is infeasible due to data issues. It includes comparable metrics and can be integrated with the same wrangling methods.

In conclusion, focusing on the G20 ensures analytical relevance and data quality. The merged dataset will support insights into how economic strength aligns with citizen well-being among the world's leading economies.

References

- International Monetary Fund. (2020). The G-20's role in the post-pandemic recovery. <https://www.imf.org/en/News/Articles/2020/11/19/na111920-the-g20s-role-in-the-post-pandemic-recovery>
- Helliwell, J. F., Layard, R., & Sachs, J. D. (2020). World Happiness Report 2020. Sustainable Development Solutions Network. <https://worldhappiness.report/ed/2020/>