

Data Source

1. World Development Indicators (WDI):

Source: data.worldbank.org

The World Development Indicators (WDI) is the primary World Bank collection of development indicators, compiled from officially recognized international sources. It presents the most current and accurate global development data available and includes national, regional, and global estimates.

- **Data Source:** External and open (World Bank-WDI).
- **Trustworthiness:** High, as it's the official record from a consultation with national authorities.
- **Data Collection:** Administrative, via reports gathered by the Bank's country management units and data obtained from official and state sources, with quarterly updates shared through the Internet.
- **Contents:** Countries, indicator's names & codes, and statistical time series data.
- **Limitations:** Quarter updates with potential fluctuations, and minimal errors due to computerization.

2. World Development Indicators *Countries* (WDIC): Another data frame with the same characteristics from the same source with additional complementary and categorical information.

- **Contents:** Countries, Currency Unit, Special Notes, Region Income Group, WB-2 code, National accounts base year, National accounts reference year, SNA price valuation, Lending category, and other groups.

I selected this dataset based on my personal experience as a retail investor and trader in the capital markets, along with my understanding of the macroeconomic variables that influence them. It also aligns with my background in finance from my associate degree. Additionally, I have a strong interest in predictive analysis, and using this dataset to forecast potential investments is something I enjoy.

Data Profile

Consistency checks are realized to ensure there are no duplicates, mixed-type data, missing values, or outliers that are unreasonable given the context of this data set. The following changes were made:

- **Dropped columns from World Development Indicators Countries (WDIC) data frame:**
'Table Name', 'Long Name', '2-alpha code', 'Special Notes', 'WB-2 code', 'National accounts base year', 'National accounts reference year', 'SNA price valuation', 'System of National Accounts', 'Alternative conversion factor', 'PPP survey year', 'Balance of Payments Manual in use', 'External debt Reporting status', 'System of trade', 'Government Accounting concept', 'IMF data dissemination standard', 'Latest population census', 'Latest household survey', 'Source of most recent Income and expenditure data', 'Vital registration complete', 'Latest agricultural census', 'Latest industrial data', 'Latest trade data', 'Latest water withdrawal data' from World Development Indicators Countries (WDIC) data frame
- **Updated data:** Replace 'Turkiye' with 'Turkey'
- Descriptive Statistics in Jupyter code notebook.

Limitations & Ethics

The data, while globally authoritative, may have biases and limitations stemming from the way it is collected, interpreted, and its underlying objectives, particularly when focusing on economic growth

Limitations:

1. **Growth Bias:** Focus on economic growth (e.g., GDP) may overshadow social, environmental, and inequality factors.
2. **Aggregation:** Data at national levels may hide regional disparities within a country.
3. **Timeliness:** Data can be outdated, leading to misinformed decisions in fast-changing environments.
4. **Gaps & Inconsistencies:** Missing data and differing methodologies across countries can make comparisons unreliable.
5. **Official Data Reliance:** Dependence on government-reported data, which can be politically influenced.

Ethical Considerations:

1. **Inequality:** Growth-focused metrics may worsen inequality or harm marginalized populations.
2. **Environmental Neglect:** GDP growth indicators might ignore environmental sustainability.
3. **Western-Centric Models:** Economic policies may push multinational interests over local economies.
4. **Exclusion of Informal Economies:** Informal sectors, critical in emerging markets, may be overlooked.
5. **Data Collection:** Lack of transparency and informed consent in data collection processes can exploit vulnerable communities.

Key Questions

- How do **GDP growth** and **gross savings** reflect potential demand for banking services in the target markets?
- What impact do **inflation (CPI)** and **unemployment rates** have on financial service usage?
- How do **FDI inflows** indicate the attractiveness of the financial sector in the emerging markets?
- Is **GDP growth** driving demand for consumer banking products, or is it concentrated in sectors with less financial service needs?
- What proportion of **gross savings** is held in banks, signaling opportunities for expansion?
- How are banks responding to **GDP growth** and **savings rate** changes?

Privacy and Ethics Questions:

- Are **data collection practices** in emerging markets compliant with local **privacy laws**?
- How can banks avoid **financial exclusion** of marginalized communities in their expansion efforts?