Text

Description automatically generated

BC2402 ANALYTICS I:  
VISUAL & PREDICTIVE TECHNIQUES

AY21/22 SEM 1 | SEMINAR 2, TEAM 7

DATA DICTIONARY

PISA\_PREDICTORS.csv → PISA\_results\_cart.csv → GDP\_INPUT\_EDU.csv → GDP\_INPUT\_ALL.csv → GDP\_INPUT\_ALL\_P.csv

* Starting with publicly available data from the world bank, we generate PISA\_PREDICTORS.csv. Each subsequent dataset generated and used for subsequent models is built upon the last.
* Owing to this, many data tables will be repeated in this data dictionary. This is intentional, to avoid repetition while showing the data gathering process.
* We also include the csvs of the publicly available data, to allow the R script to run properly.

| File | Variable Name | Variable |
| --- | --- | --- |
| PISA\_PREDICTORS.csv  PISA\_results\_cart.csv  GDP\_INPUT\_EDU.csv  GDP\_INPUT\_ALL.csv | PISA\_math | PISA: Mean performance on the mathematics scale |
| PISA\_sci | PISA: Mean performance on the science scale |
| PISA\_read | PISA: Mean performance on the reading scale |
| Pre\_pri\_enrolment\_num | Enrolment in pre-primary education, both sexes (number) |
| Pre\_pri\_sch\_expectancy | School life expectancy, pre-primary, both sexes (years) |
| Pre\_pri\_enrolment\_ratio | Gross enrolment ratio, pre-primary, both sexes (%) |
| Pri\_enrolment\_ratio | Gross enrolment ratio, primary, both sexes (%) |
| Pre\_pri\_pupil\_teacher\_ratio | Pupil-trained teacher ratio in pre-primary education (headcount basis) |
| Pri\_enrolment\_num | Enrolment in primary education, both sexes (number) |
| Pri\_sch\_expectancy | School life expectancy, primary, both sexes (years) |
| Pri\_completion | Completion rate, primary education, both sexes (%) |
| Sec\_enrolment\_num | Enrolment in secondary education, both sexes (number) |
| Sec\_enrolment\_ratio | Gross enrolment ratio, secondary, both sexes (%) |
| Pri\_sec\_sch\_expectancy | School life expectancy, primary and secondary, both sexes (years) |
| Sec\_sch\_expectancy | School life expectancy, secondary, both sexes (years) |
| Lower\_sec\_completion | Completion rate, lower secondary education, both sexes (%) |
| Upper\_sec\_completion | Completion rate, upper secondary education, both sexes (%) |
| Ter\_enrolment\_num | Enrolment in tertiary education, all programmes, both sexes (number) |
| STEM\_ter\_grad\_percent | Percentage of graduates from Science, Technology, Engineering and Mathematics programmes in tertiary education, both sexes (%) |
| Literacy\_rate | Adult literacy rate, population 15+ years, both sexes (%) |
| Illiterate\_pop | Adult illiterate population, 15+ years, both sexes (number) |
| Gov\_ex\_percent\_GDP | Government expenditure on education as % of GDP |
| Gov\_ex | Government expenditure on education, constant US$ (millions) |
| Population | Population, total |
| Gov\_ex\_capita | Government expenditure on education, constant US$ (millions) per capita |
| Gov\_ex\_3 | Government expenditure on education, constant US$ (millions) lagged by 3 years |
| Log\_gov\_ex\_3 | Log (Government expenditure on education lagged by 3 years) |
| GDP\_INPUT\_EDU.csv  GDP\_INPUT\_ALL.csv | GDP\_1 | Gross Domestic Product (US$) lagged by 1 year |
| GDP\_growth\_1 | Gross Domestic Product growth (annual %) lagged by 1 year |
| GDP\_capita\_1 | Gross Domestic Product per capita (US$) lagged by 1 year |
| GDP\_capita\_growth\_1 | Gross Domestic Product per capita growth (annual %) lagged by 1 year |
| worldbank/macroecons.csv  GDP\_INPUT\_ALL.csv | Life\_expectancy | Life expectancy at birth |
| Gross\_capital\_formation | Gross capital formation as a percentage of GDP |
| Trade | Value of trade as a percentage of GDP |
| Govt\_expenditure\_total | Total government expenditure as a percentage of GDP |
| worldbank/gdp.csv | GDP | Gross Domestic Product (US$) |
| GDP\_growth | Gross Domestic Product growth (annual %) |
| GDP\_capita | Gross Domestic Product per capita (US$) |
| GDP\_capita\_growth | Gross Domestic Product per capita growth (annual %) |
| GDP\_INPUT\_ALL\_P.csv | Life\_expectancy\_change | % change in Life expectancy at birth from the previous year |
| Gross\_capital\_formation\_change | % change in Gross capital formation from the previous year |
| Trade\_change | % change in Value of trade from the previous year |
| Govt\_expenditure\_total\_change | % change in Total government expenditure from the previous year |
| Pri\_sec\_sch\_expectancy\_change | % change of school life expectancy, primary and secondary, both sexes (years) from previous year |
| Gov\_ex\_change | % change of government expenditure on education from previous year |
| Gov\_ex\_percent\_GDP\_change | % change of government expenditure on education as % of GDP from previous year |
| Population\_change | % change of total population from previous year |
| Pre\_pri\_enrolment\_num\_change | % change in enrolment in pre-primary education, both sexes from previous year |
| Pri\_enrolment\_num\_change | % change in enrolment in primary education, both sexes from previous year |
| Sec\_enrolment\_num\_change | % change in enrolment in secondary education, both sexes from previous year |
| Ter\_enrolment\_num\_change | % change of enrolment in tertiary education, all programmes, both sexes (number) from previous year |
| Pre\_pri\_enrolment\_ratio\_change | % change of gross enrolment ratio, pre-primary, both sexes from previous year |
| Pri\_enrolment\_ratio\_change | % change of gross enrolment ratio, primary, both sexes from previous year |
| Sec\_enrolment\_ratio\_change | % change of gross enrolment ratio, secondary, both sexes from previous year |
| GDP\_growth\_positive | Categorical variable that represents if GDP growth for the next year is positive or negative |
| PISA\_math\_change | % change of PISA score on the mathematics scale from previous year |
| PISA\_sci\_change | % change of PISA score on the science scale from previous year |
| PISA\_read\_change | % change of PISA score on the reading scale from previous year |
| PISA\_math\_change\_3 | % change of PISA score on the mathematics scale from previous year lagged by 3 years |
| PISA\_sci\_change\_3 | % change of PISA score on the science scale from previous year lagged by 3 years |
| PISA\_read\_change\_3 | % change of PISA score on the reading scale from previous year lagged by 3 years |
| PISA\_math\_new\_3 | Predicted PISA score on the mathematics scale based on our model, and actual PISA scores where applicable, lagged by 3 years |
| PISA\_sci\_new\_3 | Predicted PISA score on the science scale based on our model, and actual PISA scores where applicable, lagged by 3 years |
| PISA\_read\_new\_3 | Predicted PISA score on the reading scale based on our model, and actual PISA scores where applicable, lagged by 3 years |
| RSCRIPT\_Machine\_Learning\_BC2406S2T7.R | PISA\_pred\_math | Predicted PISA score on the mathematics scale based on our model |
| PISA\_pred\_sci | Predicted PISA score on the science scale based on our model |
| PISA\_pred\_read | Predicted PISA score on the reading scale based on our model |
| PISA\_results\_cart.csv  GDP\_INPUT\_ALL.csv | PISA\_math\_new | Predicted PISA score on the mathematics scale based on our model, and actual PISA scores where applicable |
| PISA\_sci\_new | Predicted PISA score on the science scale based on our model, and actual PISA scores where applicable |
| PISA\_read\_new | Predicted PISA score on the reading scale based on our model, and actual PISA scores where applicable |