

CEBD1260

Final Project - Online GDP Prediction Tool

Summary

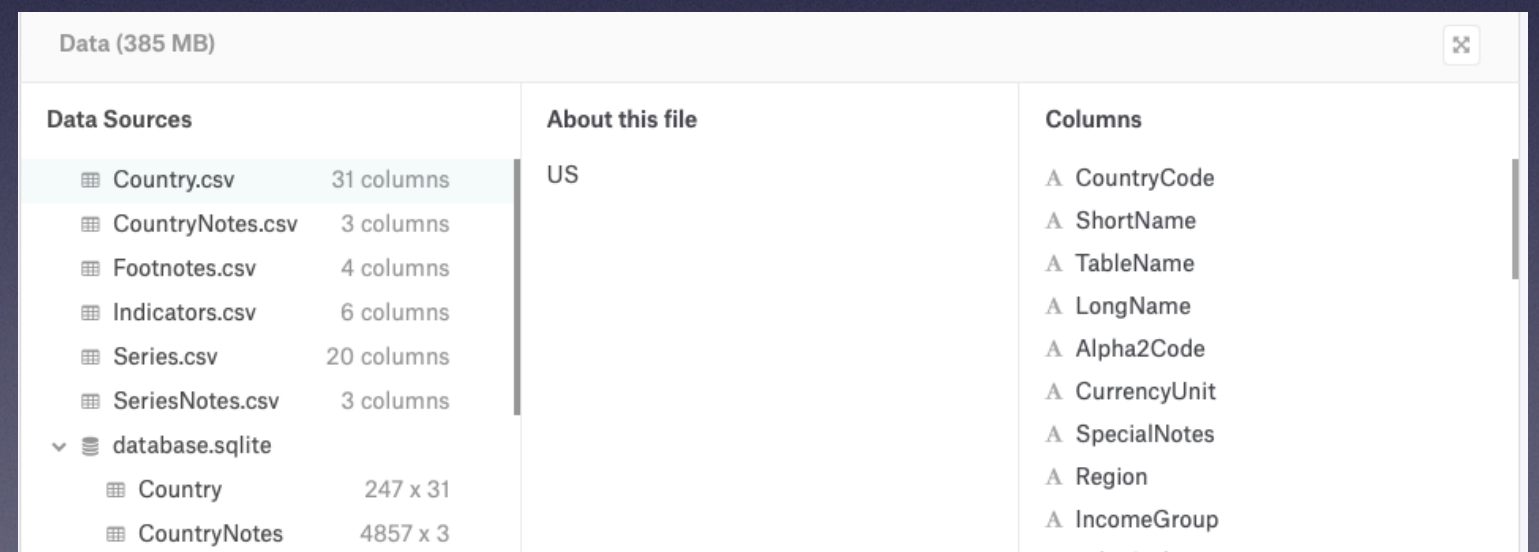
- Problem definition
- Data set
- Approach
- Results
- Q&A

Data Set

- The world development indicators from the world bank contain over thousand annual indicators of economic development from hundreds of countries.

- 385mb

- Semi-structured



The screenshot shows a data viewer interface with three main sections: Data Sources, About this file, and Columns.

Data Sources	About this file	Columns
<ul style="list-style-type: none">Country.csv 31 columnsCountryNotes.csv 3 columnsFootnotes.csv 4 columnsIndicators.csv 6 columnsSeries.csv 20 columnsSeriesNotes.csv 3 columnsdatabase.sqlite<ul style="list-style-type: none">Country 247 x 31CountryNotes 4857 x 3	US	<ul style="list-style-type: none">CountryCodeShortNameTableNameLongNameAlpha2CodeCurrencyUnitSpecialNotesRegionIncomeGroup

Problem Definition

- Can we predict the GDP of a country given several economic factors?
- Exploratory analysis (China, fertility rate)
- Who would use this? - NGO & Governmental Agencies

Approach

- Previous data was similar (WDI's - query through Google BigQuery)
- Cleaning - target specific indicators by code e.g.
- Modelling - LinearRegression Vs. RandomForest?
- Deploying - Choosing columns + herokuapp

Approach - Challenges

- All values being in one column (semi-structured)
- .str contains to find IndicatorCode
- Create several data frames + aggregate
- Dealing with NaN's by filling with mean

Results

- <https://gdp-prediction-tool.herokuapp.com>

Questions