

# Data Table Schema

## job\_listings

Public job listings and associated company information, from 2014 – 2019. Refer to *onet\_code\_legend.csv* and *SOC\_code\_legend.csv*, respectively, for more detailed ONET (2010) and SOC (2010) occupational code descriptions.

~2.7 million rows & 17 columns. Size: ~740MB. Source: non-public.

Field	Type	Description
<b>title</b>	STRING	Job title scraped from the job post
<b>company_name</b>	STRING	Name of the company posting the job
<b>city</b>	STRING	City location for the job
<b>region</b>	STRING	Regional location for the job
<b>zip</b>	STRING	Zip location for the job
<b>created</b>	STRING	The first time the job listing was scraped
<b>last_checked</b>	STRING	The most recent time the job was seen
<b>last_updated</b>	STRING	The most recent time the job was modified
<b>delete_date</b>	STRING	The most recent time the job was taken down
<b>onet_occupation_code</b>	STRING	ONET classification of the job. Refer to <i>onet_code_legend.csv</i> for more detailed ONET occupation code descriptions. Unclassified jobs are denoted by “9999-99.00”
<b>soc_occupation_code</b>	STRING	SOC code mappings (from ONET) for integration with employment_by_occupation dataset. Refer to <i>SOC_code_legend.csv</i> for more detailed SOC occupation code mappings. Unclassified jobs are denoted by “9999” and unmapped jobs are denoted by “9998”
<b>url</b>	STRING	The unique URL for the specific job
<b>stock_exchange_country</b>	STRING	Country of the stock exchange that the ticker symbol is traded on
<b>stock_exchange_name</b>	STRING	Stock exchange symbol that ticker is traded on
<b>stock_ticker</b>	STRING	Ticker symbol
<b>ticker_end_date</b>	STRING	The first day of listing of the stock ticker
<b>ticker_start_date</b>	STRING	The last day of listing of the stock ticker

## employment\_by\_occupation

UK employment data by occupation, status, and sex. Data is estimated from the Labour Force Survey in Quarter 2 (Apr - Jun) of 2011 – 2018. Refer to *SOC\_code\_legend.csv* for more detailed SOC (2010) occupational code mappings.

29,952 rows & 10 columns. Size: ~2MB. Source: [Office for National Statistics](#)

Field	Type	Description
<b>year</b>	INTEGER	Year of observation

<b>soc_occupation_code</b>	INTEGER	Standard Occupational Classification (SOC) codes follow the format of {main}{sub}{minor}{unit}. Refer to <i>SOC_code_legend.csv</i> for more detailed SOC occupation code mappings.
<b>soc_major</b>	INTEGER	Major is the most general level of classification
<b>soc_minor</b>	INTEGER	Minor is the intermediate level of classification
<b>soc_unit</b>	INTEGER	Unit is the most specific level of classification
<b>description</b>	STRING	Job description, corresponding to the SOC classification
<b>sex</b>	STRING	The sex of the employee, 'male' or 'female'
<b>emp_type</b>	STRING	Employment type, 'employees' or 'self-employed'
<b>fp_time</b>	STRING	Full or part-time, 'full' or 'part'
<b>value</b>	INTEGER	Number of people in employment

## labor\_market\_statistics

UK labor market statistics from the Labour Force Survey from 1992 – 2019. Refer to *labor\_market\_stats\_legend.csv* for indicator codes and units.

324 rows & 254 columns. Size: ~0.5MB. Source: [Office for National Statistics](#)

Field	Type	Description
<b>year</b>	INTEGER	Year of observation
<b>month</b>	STRING	Time period of observation, values include year, quarter, or month
<b>Various Indicators [252 features]</b>	STRING	Various countrywide market data aggregated over the specified time period. Refer to <i>labor_market_stats_legend.csv</i> for indicator codes and units

## immigrant\_statistics

UK citizenship applications by country of nationality from 2007 – 2018. Refer to *immigrant\_legend.csv* for more information.

47 rows & 263 columns. Size: ~0.1MB. Source: [Office for National Statistics](#)

Field	Type	Description
<b>year</b>	INTEGER	Year of observation
<b>quarter</b>	INTEGER	Quarter of observation
<b>Various Countries [261 features]</b>	STRING	Various country-level statistics on citizenship applications to the UK. They follow the format of (geographical_region)-(country_of_nationality) where each use of "-" indicates one more level of specificity. Codes ending without "-" are aggregate values by geographical region. Refer to <i>immigrant_legend.csv</i> for more information

## LSE\_historical\_data

Historical London Stock Exchange (LSE) company data from 2013 – 2019. Market cap data is reported monthly whereas P/E, EPS, and P/S data is annually since 2014 in most cases.

159,400 rows & 11 columns. Size: 15MB. Source: [London Stock Exchange](#) and [Financial Times](#)

Field	Type	Description
<b>year</b>	INTEGER	Year of observation
<b>month</b>	INTEGER	Month of observation
<b>ticker</b>	STRING	Ticker symbol for the company
<b>list_date</b>	STRING	The date when the company was listed on the LSE
<b>company_name</b>	STRING	The name of the company
<b>icb_industry</b>	STRING	Industry Classification Benchmark (ICB) is a four-tier structure, with industry, supersector, sector, and subsector. Industry is the most general classification level
<b>icb_supersector</b>	STRING	ICB supersector, second-most general categorization.
<b>country_of_incorporation</b>	STRING	The country in which the company was incorporated
<b>market_cap</b>	FLOAT	Total dollar market value of a company's outstanding shares, reported monthly (unit: <i>millions of £s</i> )
<b>eps</b>	FLOAT	Earnings per share, bi-annually (unit: <i>millions of £s</i> )
<b>p/e</b>	FLOAT	Price to earnings ratio, bi-annually (unit: <i>millions of £s</i> )
<b>p/s</b>	FLOAT	Price to sales ratio, bi-annually (unit: <i>millions of £s</i> )

## UK\_bill\_data

UK Public General Acts, by content and size of bill. All UK legislation passed from 1996 – 2019.

890 rows & 7 columns. Size: ~10MB. Source: [UK Legislation](#)

Field	Type	Description
<b>act_name</b>	STRING	Name of legislation passed by the UK government
<b>date_of_act</b>	STRING	Date when legislation was passed
<b>amendment</b>	BOOLEAN	Whether or not this legislation later amended
<b>characters_in_act</b>	INTEGER	The length of the legislation's text in characters
<b>words_in_act</b>	INTEGER	The length of the legislation's text in words
<b>link_to_act</b>	STRING	The public link for more information about the act
<b>act_text</b>	STRING	The full text of legislation