

# Data Table Schema

## London 2012 Datasets

### UK\_international\_visits

Tracking of international visitors to the UK considering duration of stay and estimated spend extrapolated from survey results. Covering from 2002 - 2019.

549,853 rows & 11 columns. Size: ~55MB. Source: [Visit Britain](#)

| Field           | Type    | Description  |
|-----------------|---------|--|
| <b>year</b>     | INTEGER | Year   |
| <b>quarter</b>  | STRING  | Quarter  |
| <b>market</b>   | STRING  | Origin country of tourist  |
| <b>dur_stay</b> | STRING  | The range of nights this group is associated to                                      |
| <b>mode</b>     | STRING  | Method of travel (e.g. "Air")  |
| <b>purpose</b>  | STRING  | The reason for travel  |
| <b>area</b>     | STRING  | Region where this group recorded staying   |
| <b>visits</b>   | FLOAT   | The total number of visits made by persons in this group, in thousands               |
| <b>spend</b>    | FLOAT   | The estimated total amount spent by all persons in this group, in millions of Pounds |
| <b>nights</b>   | FLOAT   | The total number of nights spent in the area by this group, in thousands             |
| <b>sample</b>   | FLOAT   | The number of people surveyed to create these spend and nights estimates             |

### london\_earnings\_by\_borough

Average earnings by week or hour based on the borough of residence, covering from 2002 - 2019. Data from the Annual Survey of Hours and Earnings.

6,768 rows & 7 columns. Size: ~318KB. Source: [Office for National Statistics](#)

| Field            | Type    | Description   |
|------------------|---------|---|
| <b>year</b>      | INTEGER | Year of observation   |
| <b>area_code</b> | STRING  | Unique code corresponding to the area where persons reside  |
| <b>area</b>      | STRING  | Name of the area where persons making the earnings reside   |
| <b>identity</b>  | STRING  | What group of people are covered by this entry. Broken into: "Part-Time", "Full-Time", "Male", or "Female". |
| <b>pay_type</b>  | STRING  | "Hourly" or "Weekly". Corresponding to the wage rate over the time period                                   |

|                   |       |  |
|-------------------|-------|--|
| <b>pay</b>        | FLOAT | Income per pay_type                        |
| <b>confidence</b> | FLOAT | Coefficient of variation for the pay entry |

## london\_economic\_activity

Workforce participation segmented by gender and borough, covering from 2005 - 2019. Created from the Annual Population Survey.

1,470 rows & 8 columns. Size: ~97KB. Source: [Office for National Statistics](#)

| Field                      | Type    | Description  |
|----------------------------|---------|--|
| <b>year</b>                | INTEGER | Year   |
| <b>area_code</b>           | STRING  | Unique code corresponding to the area where persons reside                           |
| <b>area</b>                | STRING  | Name of the area where persons making the earnings reside                            |
| <b>gender</b>              | STRING  | Gender of those covered in the entry. Either "M", or "F"                             |
| <b>economically_active</b> | STRING  | Number of persons in the entry's group who are in the workforce                      |
| <b>working_age</b>         | FLOAT   | Income per pay_type  |
| <b>percent</b>             | FLOAT   | Percentage of economically active working aged persons. Given as a number out of 100 |
| <b>confidence</b>          | FLOAT   | 95% confidence interval of the percent figure  |

## london\_infrastructure\_spending

Breakdown of direct expenditures for the London games, including the change in cost estimates over development time.

59 rows & 6 columns. Size: ~4KB. Source: [The Guardian](#)

| Field                      | Type    | Description   |
|----------------------------|---------|---|
| <b>category</b>            | STRING  | The overarching category under which the expenditure falls. Note: the spending from the "Public Sector Funding Package" category will partially overlap with the other major categories |
| <b>breakdown</b>           | STRING  | More detailed expenditure specifics   |
| <b>baseline_budget</b>     | INTEGER | The initial budget for the games in millions of Pounds  |
| <b>31/12/2011_forecast</b> | INTEGER | Updated forecast made on 31/12/2011 for the total expected cost in millions of Pounds   |
| <b>31/03/2012_forecast</b> | INTEGER | Updated forecast made on 31/03/2012 for the total expected cost in millions of Pounds   |
| <b>variance</b>            | INTEGER | Change from 31/12/2011 forecast to 31/03/2012 forecast, in millions of Pounds   |

## london\_sports\_participation

Breakdown of the amount of exercise people in London get by number of exercising sessions. Given at the borough level and covers from 2008 - 2016.

1,290 rows & 6 columns. Size: ~60KB. Source: [Sport England](#)

| Field                       | Type    | Description   |
|-----------------------------|---------|---|
| <b>year</b>                 | INTEGER | Year  |
| <b>area_code</b>            | STRING  | Unique code corresponding to the area where persons reside  |
| <b>area</b>                 | STRING  | Name of the area where persons in the entry reside  |
| <b>sports_participation</b> | STRING  | Number of sessions per week that persons in this category participated in at least moderately intense exercise. Options are "one+", "three+", and "zero". Note that all persons counted in "three+" would also be counted in "one+". This is measured annually from <b>October to October</b> |
| <b>percentage</b>           | FLOAT   | Percentage of persons in the area who fall under the sports participation frequency. Percent given in decimal format  |
| <b>population</b>           | INTEGER | Population of the corresponding area  |

## london\_taxpayer\_income

Breakdown of the mean and median annual income of taxpayers based on what borough of London they reside in. Covering from 1999 - 2018. Data is from the Survey of Personal Incomes.

846 rows & 6 columns. Size: ~39KB. Source: [HM Revenue & Customs](#)

| Field                | Type    | Description  |
|----------------------|---------|--|
| <b>year</b>          | INTEGER | Year   |
| <b>area_code</b>     | STRING  | Unique code corresponding to the area where persons reside               |
| <b>area</b>          | STRING  | Name of the area where persons in the entry reside                       |
| <b>population</b>    | INTEGER | Population of taxpayers in the corresponding area                        |
| <b>mean_income</b>   | INTEGER | Average income of individuals from this area during the given year range |
| <b>median_income</b> | INTEGER | Median Income of individuals from this area during the given year range  |

## london\_ticket\_sales

Sales figures for select Olympic events, outlining how many tickets were sold and at what price for each event. These sales do not include group tickets (such as those purchased by the families of athletes).

938 rows & 5 columns. Size: ~29KB. Source: [London Organising Committee of the Olympic & Paralympic Games](#)

| Field                         | Type    | Description   |
|-------------------------------|---------|---|
| <b>session</b>                | STRING  | The unique code given to an Olympic event session   |
| <b>price</b>                  | STRING  | The price of the ticket category. Either a number representing the Pound price, or "Special", representing a non-specific reduced ticket price (e.g. "pay your age" for children) |
| <b>number_sold</b>            | STRING  | The number of public tickets sold at this price point for this event  |
| <b>total_public_sold</b>      | INTEGER | The total number of public tickets sold for the specific event session  |
| <b>percent_sold_to_public</b> | INTEGER | The percent of tickets sold for the specific event session that were not group tickets. Given in a number out of 100  |

## london\_tickets\_for\_sale

Outlines all tickets that were available for sale at the Olympic games and what their pricing was. This dataset does not indicate sales figures, only the sales offerings.

1,128 rows & 15 columns. Size: ~129KB. Source: [The Guardian](#)

| Field                 | Type    | Description  |
|-----------------------|---------|--|
| <b>date</b>           | DATE    | Date of the event. YYYY-MM-DD format   |
| <b>time</b>           | STRING  | Time range of when the event took place. HH:MM-HH:MM format  |
| <b>gender</b>         | STRING  | Gender of the event. Either "M" or "W"   |
| <b>session</b>        | STRING  | The name of the Olympic event  |
| <b>medal_ceremony</b> | STRING  | "M" if this was an event with a medal ceremony, empty otherwise  |
| <b>code</b>           | STRING  | The unique code given to an Olympic event session  |
| <b>AA</b>             | INTEGER | The price of AA grade tickets for the event, in Pounds. Note that events without this pricing option are listed as "-" |
| <b>A</b>              | INTEGER | The price of A grade tickets for the event, in Pounds. Note that events without this pricing option are listed as "-"  |

|              |         |   |
|--------------|---------|---|
| <b>B</b>     | INTEGER | The price of B grade tickets for the event, in Pounds. Note that events without this pricing option are listed as “–” |
| <b>C</b>     | INTEGER | The price of C grade tickets for the event, in Pounds. Note that events without this pricing option are listed as “–” |
| <b>D</b>     | INTEGER | The price of D grade tickets for the event, in Pounds. Note that events without this pricing option are listed as “–” |
| <b>E</b>     | INTEGER | The price of E grade tickets for the event, in Pounds. Note that events without this pricing option are listed as “–” |
| <b>limit</b> | INTEGER | The maximum number of tickets an individual could purchase for the event  |
| <b>venue</b> | STRING  | Where the event took place  |
| <b>event</b> | STRING  | The name of the overarching event type  |

## UK\_inflation

UK inflation data using 2018 as a base year, covering back to 1751.

268 rows & 3 columns. Size: ~4KB. Source: [Kate Rose Morley](#)

| Field             | Type    | Description                         |
|-------------------|---------|-------------------------------------|
| <b>year</b>       | INTEGER | Year                                |
| <b>inflation</b>  | FLOAT   | Rate of inflation in decimal format |
| <b>multiplier</b> | FLOAT   | Inflation multiplier                |

## london\_underground\_activity

Tracking of entry and exit figures from London's various underground subway stations based on the type of day. Covering from 2007 - 2017.

2,953 rows & 11 columns. Size: ~207KB. Source: [Transport for London](#)

| Field                 | Type    | Description  |
|-----------------------|---------|--|
| <b>year</b>           | INTEGER | Year   |
| <b>nlc</b>            | STRING  | Unique code corresponding to the station                                   |
| <b>station</b>        | STRING  | Name of the underground station  |
| <b>area</b>           | STRING  | The name of the area where the station is located                          |
| <b>weekday_entry</b>  | INTEGER | The number of times people entered this station on an average weekday      |
| <b>weekday_exit</b>   | INTEGER | The number of times people exited from this station on an average weekday  |
| <b>saturday_entry</b> | INTEGER | The number of times people entered this station on an average Saturday     |
| <b>saturday_exit</b>  | INTEGER | The number of times people exited from this station on an average Saturday |

|                              |         |   |
|------------------------------|---------|---|
| <b>sunday_entry</b>          | INTEGER | The number of times people entered this station on an average Sunday          |
| <b>sunday_exit</b>           | INTEGER | The number of times people exited from this station on an average Sunday      |
| <b>annual_entry_and_exit</b> | FLOAT   | The annual number of entries and exits at this station this year. In millions |

## london\_underground\_station\_info

Information about London's underground stations. It outlines what local authority the station belongs to, in what zone(s) it operates and when the station first opened.

269 rows & 5 columns. Size: ~14KB. Source: [Transport for London](#)

| Field                  | Type   | Description  |
|------------------------|--------|--|
| <b>station</b>         | STRING | Name of the underground station  |
| <b>line</b>            | STRING | Which line(s) the station is on. If multiple, they are separated by ampersands |
| <b>local_authority</b> | STRING | Name of the underground station  |
| <b>zone</b>            | STRING | The zone(s) the station is in. If multiple, they are separated by ampersands   |
| <b>opened</b>          | DATE   | The date the underground station opened. In the format of YYYY-MM-DD           |

# Vancouver 2010 Datasets

## vancouver\_tourism\_indicators

High-level indicators relating to tourism in Vancouver, including revenue, GDP, employment, salaries, and regional estimates. Covering from 2000 - 2015.

16 rows & 14 columns. Size: ~2KB. Source: [British Columbia](#)

| Field                             | Type  | Description  |
|-----------------------------------|-------|--|
| <b>year</b>                       | FLOAT | Year   |
| <b>revenue</b>                    | FLOAT | Revenue, in millions of dollars  |
| <b>GDP</b>                        | FLOAT | Gross domestic product, in millions of CAD   |
| <b>real_GDP</b>                   | FLOAT | Real gross domestic product, in millions of CAD (base year: 2007)  |
| <b>implicit_price_index</b>       | FLOAT | Implicit price index, where 2007 = 100   |
| <b>employment</b>                 | FLOAT | Employment, in thousands of people   |
| <b>wages_and_salaries</b>         | FLOAT | Wages & salaries, in millions of CAD   |
| <b>consumption_taxes</b>          | FLOAT | Sales tax, in millions of CAD  |
| <b>employment_XYZ [6 regions]</b> | FLOAT | Employment by region. XYZ stands for the region code (VCI - Vancouver, Coast & Mountains, VI - Vancouver Island, TO - Thompson Okanagan, NBC - Northern British Columbia, BCR - British Columbia Rockies, CCC - Cariboo Chilcotin Coast) |

## vancouver\_employment\_by\_industry

British Columbia employment by North American Industry Classification System (NAICS) industry. Please refer to *vancouver\_employment\_by\_industry\_legend.csv* for the corresponding industry names. Covering from 1987 - 2019.

4654 rows & 3 columns. Size: ~116KB. Source: [British Columbia](#)

| Field                | Type    | Description  |
|----------------------|---------|--|
| <b>year</b>          | INTEGER | Year   |
| <b>industry_code</b> | STRING  | NAICS industry code. Please refer to <i>vancouver_employment_by_industry_legend.csv</i> for the corresponding industry names |
| <b>value</b>         | FLOAT   | Number of people employed in the industry  |

## vancouver\_business\_size

Business locations by employee size, by region and industry. Covering from 1999 - 2019.

15,909 rows & 10 columns. Size: ~882KB. Source: [British Columbia](#)

| Field            | Type   | Description   |
|------------------|--------|---|
| year             | FLOAT  | Year  |
| region_code      | STRING | Region code. Refer to <i>region_codes.csv</i> to find the names and relationships between regions   |
| business_type    | STRING | NASICS2 description of the industry   |
| unknown          | FLOAT  | Businesses that do not maintain an employee payroll or did not have employees in the past 12 months |
| one_to_4         | FLOAT  | Businesses with one to four employees   |
| five_to_9        | FLOAT  | Businesses with five to nine employees  |
| ten_to_19        | FLOAT  | Businesses with ten to nineteen employees   |
| twenty_to_49     | FLOAT  | Businesses with twenty to forty-nine employees  |
| fifty_to_199     | FLOAT  | Businesses with fifty to one-hundred ninety-nine employees  |
| two_hundred_plus | FLOAT  | Businesses with more than two hundred employees   |

## vancouver\_small\_business\_by\_region

Small business counts segmented by region. Covering from 2007 - 2018.

85 rows & 3 columns. Size: ~1KB. Source: [British Columbia](#)

| Field       | Type   | Description   |
|-------------|--------|---|
| year        | FLOAT  | Year  |
| region_code | STRING | Region code(s). <b>Note:</b> multiple region code may be noted. Refer to <i>region_codes.csv</i> to find the names and relationships between regions. |
| value       | FLOAT  | Number of small businesses  |

## vancouver\_visits

Monthly overnight visitors to Vancouver. Covering from 1994 - 2018.

301 rows & 3 columns. Size: ~1KB. Source: [Link](#)

| Field | Type  | Description        |
|-------|-------|--------------------|
| year  | FLOAT | Year               |
| month | FLOAT | Month              |
| value | FLOAT | Number of visitors |



## vancouver\_room\_revenues\_2000\_2010

Room revenues, property counts, and room counts by municipal jurisdictions. Covering from 2000 - 2010.

58,212 rows & 7 columns. Size: ~1MB. Source: [British Columbia](#)

| Field       | Type   | Description  |
|-------------|--------|--|
| year        | FLOAT  | Year   |
| month       | FLOAT  | Month  |
| region_code | STRING | Region code. Refer to <i>region_codes.csv</i> to find the names and relationships between regions  |
| acc_type    | STRING | Accommodation types. 'TOT' = all types, 'FL' = fishing lodge, 'VR' = vacation rental, 'HO' = hotels, with a breakdown by number of rooms ('H1' = 251+ rooms, 'H2' = 151 - 250 rooms, 'H3' = 76 - 150 rooms, 'H4' = 1- 75 rooms), 'MO' = motels, 'MI' = all other accommodation types |
| revenue     | FLOAT  | Monthly room revenues (\$)   |
| properties  | FLOAT  | Monthly property counts  |
| rooms       | FLOAT  | Monthly room counts  |

## vancouver\_room\_revenues\_2010\_2019

Room revenues by municipal jurisdictions. Covering from 2010 - 2019. **Note:** the region codes were manually mapped to best match up with the 2000 - 2010 data, but discrepancies may arise.

5,280 rows & 5 columns. Size: ~147KB. Source: [British Columbia](#)

| Field       | Type   | Description  |
|-------------|--------|--|
| year        | FLOAT  | Year   |
| month       | FLOAT  | Month  |
| region_code | STRING | Region code(s). <b>Note:</b> multiple region code may be noted. Refer to <i>region_codes.csv</i> to find the names and relationships between regions |
| region_name | STRING | The name of the region, district, or urban center  |
| revenue     | FLOAT  | Monthly room revenue estimate (\$)   |

## region\_codes

Mappings of region code and region names, as well as their relationships and descriptions.

75 rows & 4 columns. Size: ~3KB. Source: [British Columbia](#)

| Field              | Type   | Description   |
|--------------------|--------|---|
| <b>region_code</b> | STRING | The code associated with a region. The first two characters indicate the type of region (DR = Developmental region, RD = Regional district, and UC = urban center). <b>Note:</b> certain RD codes are combined and contain a “+” (e.g. RD51+57) |
| <b>region_name</b> | STRING | The name of the region, district, or urban center   |
| <b>composition</b> | STRING | DRs are made up of RDs, and RDs contain UCs. The composition of DRs and RDs are specified   |
| <b>description</b> | STRING | Brief description of UCs  |

## Rio 2016 Datasets

### brazil\_gdp

Gross domestic product by region. Covering from 2002 - 2017.

2,193 rows & 4 columns. Size: ~100KB. Source: [Link](#)

| Field  | Type  | Description                                   |
|--------|-------|---|
| state  | FLOAT | State   |
| region | FLOAT | Region/city within state                      |
| year   | FLOAT | Year  |
| value  | FLOAT | Gross domestic product, in thousands of Reals |

### brazil\_monthly\_income

Nominal average income, usually received per month from the main job, segmented by occupational groups. Covering from 2012 - 2020.

9,802 rows & 5 columns. Size: ~659KB. Source: [Link](#)

| Field    | Type  | Description   |
|----------|-------|---|
| year     | FLOAT | Year  |
| quarter  | FLOAT | Quarter   |
| job_type | FLOAT | Occupational Group  |
| state    | FLOAT | State   |
| value    | FLOAT | Average income from the main job, usually earned per month (in Reals) |

### brazil\_unemployment

Unemployment dataset, by employment type and state. Covering from 2012 - 2019.

2,674 rows & 5 columns. Size: ~118KB. Source: [Link](#)

| Field    | Type  | Description  |
|----------|-------|--|
| year     | FLOAT | Year   |
| quarter  | FLOAT | Quarter  |
| category | FLOAT | “outside the workforce”, “workforce: unemployed”, or “workforce: employed” |
| state    | FLOAT | State  |
| value    | FLOAT | Number of people   |

## brazil\_tourism\_jobs

Government dataset from IPEA (Institute of Applied Economic Research) with data about the population working in tourism roles. Covering from 2006 - 2018.

33,661 rows & 6 columns. Size: ~2MB. Source: [Link](#)

| Field           | Type   | Description     |
|-----------------|--------|-----------------|
| <b>year</b>     | FLOAT  | Year            |
| <b>month</b>    | FLOAT  | Month           |
| <b>region</b>   | STRING | Region          |
| <b>state</b>    | STRING | State           |
| <b>job_type</b> | STRING | The type of job |
| <b>jobs</b>     | FLOAT  | Number of jobs  |

## brazil\_international\_arrivals

Government dataset from Brazil's Ministry of Tourism detailing the number of international tourists that arrived in Brazil. Covering from 1989 - 2018.

32,761 rows & 6 columns. Size: ~689KB. Source: [Link](#)

| Field          | Type   | Description        |
|----------------|--------|--------------------|
| <b>year</b>    | FLOAT  | Year               |
| <b>month</b>   | FLOAT  | Month              |
| <b>country</b> | STRING | Country of origin  |
| <b>value</b>   | FLOAT  | Number of arrivals |