Data Table Schema

clickstreams\_with\_purchase

Information about the series of pages that users clicked on which resulted in a purchase on the Wayfair platform. *Note: in order to keep the dataset size manageable, the provided data is a 25% unbiased sample of the raw data. If using click count metrics, remember to multiply quantities by 4 to approximate the actual data.*

*~11 million rows & 6 columns.* Size: ~85MB zipped, ~850MB unzipped. Source: [Wayfair](https://www.wayfair.com/).

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| customer\_id | INTEGER | Unique ID for each customer |
| date | STRING | The date recorded at time of click |
| time | STRING | The time recorded at time of click |
| page\_viewed | STRING | The name of the page viewed on Wayfair’s website |
| product\_id | INTEGER | Unique ID for the product concerned |
| ordered | INTEGER | 1 if ordered, else 0 |

clickstreams\_without\_purchase

Information about the series of pages that users clicked on which did not result in a purchase on the Wayfair platform. *Note: in order to keep the dataset size manageable, the provided data is a 1.25% unbiased sample of the raw data. If using click count metrics, remember to multiply quantities by 80 to approximate the actual data.*

*~11 million rows & 5 columns.* Size: ~90MB zipped, ~850MB unzipped.Source: [Wayfair](https://www.wayfair.com/).

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| customer\_id | INTEGER | Unique ID for each customer |
| date | STRING | The date recorded at time of click |
| time | STRING | The time recorded at time of click |
| page\_viewed | STRING | The name of the page viewed on Wayfair’s website |
| product\_id | INTEGER | Unique ID for the product concerned |

orders

Information about Wayfair online orders during a single week in July 2018.

*~140,000 rows & 9 columns.* Size: ~12MB. Source: [Wayfair](https://www.wayfair.com/).

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| order\_product\_id | INTEGER | Unique ID for each order-product combination |
| order\_id | INTEGER | Unique ID for each order |
| wayfair\_product\_id | STRING | Unique ID for the product being ordered |
| wayfair\_website\_id | STRING | Wayfair vendor identifier |
| order\_date | STRING | Date the order was made |
| order\_time | STRING | The time (24hr) the order was made |
| customer\_id | INTEGER | Unique ID for the customer making the order |
| customer\_zip | INTEGER | The zip code the order was made from |
| quantity\_ordered | INTEGER | How many of order\_product\_id was ordered |

products

Information about Wayfair products sold online.

*~500,000 rows & 13 columns.* Size: ~70MB zipped, ~200MB unzipped. Source: [Wayfair](https://www.wayfair.com/).

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| website\_name | STRING | Name of Wayfair vendor website |
| product\_id | INTEGER | Unique ID for product in question |
| product\_name | STRING | Name of product in question |
| category\_name | STRING | Name of category given to product |
| description | STRING | Product description |
| manufacturer\_name | STRING | Name of the manufacturer of the product |
| weight | FLOAT | The weight of the product in pounds |
| dimension\_1 | FLOAT | First product dimension in inches |
| dimension\_2 | FLOAT | Second product dimension in inches |
| dimension\_3 | FLOAT | Third product dimension in inches |
| num\_reviews | FLOAT | Number of reviews on the product |
| onsite\_price | FLOAT | The price Wayfair listed for the product on their website |
| on\_promotion | INTEGER | 1 if the product is on promotion else 0 |

comp\_products

Information about products of Wayfair’s competitors that are sold online.

*~1.25 million rows & 6 columns.* Size: ~200MB zipped, ~500MB unzipped. Source: [Wayfair](https://www.wayfair.com/).

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| comp\_website\_name | STRING | Name of competitor website |
| comp\_product\_id | INTEGER | Unique ID for each competitor product |
| comp\_product\_name | STRING | Name of the product |
| category\_name | STRING | How Wayfair categorizes this product |
| comp\_description | STRING | Product description |
| comp\_price | FLOAT | The price listed on the competitor’s website |

rental\_prices

Data about the rental prices of various property types across U.S. zip codes in 2018.

*19,782 rows & 17 columns.* Size: ~2MB. Source: [Zillow](https://www.zillow.com/).

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| zipcode | STRING | Zip code |
| city | STRING | City name |
| state | STRING | State name |
| metro | STRING | Name of metropolitan area |
| county | STRING | County name |
| month | INTEGER | The month the data were collected in |
| 1\_br\_total | FLOAT | The median total price (USD) of a 1 bedroom rental unit per month |
| 2\_br\_total | FLOAT | The median total price (USD) of a 2 bedroom rental unit per month |
| 3\_br\_total | FLOAT | The median total price (USD) of a 3 bedroom rental unit per month |
| 4\_br\_total | FLOAT | The median total price (USD) of a 4 bedroom rental unit per month |
| 5+\_br\_total | FLOAT | The median total price (USD) of a 5 or more bedroom rental unit per month |
| 1\_br\_psqft | FLOAT | The median price per square foot per month (USD) of a 1 bedroom rental unit |
| 2\_br\_psqft | FLOAT | The median price per square foot per month (USD) of a 2 bedroom rental unit |
| 3\_br\_psqft | FLOAT | The median price per square foot per month (USD) of a 3 bedroom rental unit |
| 4\_br\_psqft | FLOAT | The median price per square foot per month (USD) of a 4 bedroom rental unit |
| 5+\_br\_psqft | FLOAT | The median price per square foot per month (USD) of a 5 or more bedroom rental unit |
| avg\_house\_size | FLOAT | The estimated average size of a rental unit based on the available prices per square foot and total prices |

sale\_prices

Data about the sale prices of various property types across U.S. zip codes in 2018.

*42,735 rows & 17 columns.* Size: ~4MB. Source: [Zillow](https://www.zillow.com/).

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| zipcode | STRING | Zip code |
| city | STRING | City name |
| state | STRING | State name |
| metro | STRING | Name of metropolitan area |
| county | STRING | County name |
| month | INTEGER | The month the data were collected in |
| 1\_br\_total | FLOAT | The median total price (USD) of a 1-bedroom unit per month |
| 2\_br\_total | FLOAT | The median total price (USD) of a 2-bedroom unit per month |
| 3\_br\_total | FLOAT | The median total price (USD) of a 3-bedroom unit per month |
| 4\_br\_total | FLOAT | The median total price (USD) of a 4-bedroom unit per month |
| 5+\_br\_total | FLOAT | The median total price (USD) of a 5 or more bedroom unit per month |
| 1\_br\_psqft | FLOAT | The median price per square foot per month (USD) of a 1-bedroom unit |
| 2\_br\_psqft | FLOAT | The median price per square foot per month (USD) of a 2-bedroom unit |
| 3\_br\_psqft | FLOAT | The median price per square foot per month (USD) of a 3-bedroom unit |
| 4\_br\_psqft | FLOAT | The median price per square foot per month (USD) of a 4-bedroom unit |
| 5+\_br\_psqft | FLOAT | The median price per square foot per month (USD) of a 5 or more bedroom unit |
| avg\_house\_size | FLOAT | The estimated average size of a rental unit based on the available prices per square foot and total prices |

taxes

Data containing details about the tax returns filed across U.S. zip codes in 2016.

*~180,000 rows & 54 columns.* Size: ~30MB. Source: [Internal Revenue Service](https://www.irs.gov/).

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| state\_fips | STRING | Federal Information Processing Code for the given state |
| state | STRING | 2-letter state abbreviation code |
| zipcode | INTEGER | 5-digit zip code |
| agi\_stub | INTEGER | Adjusted gross income size:  1 = $1 under $25,000 2 = $25,000 under $50,000 3 = $50,000 under $75,000 4 = $75,000 under $100,000 5 = $100,000 under $200,000 6 = $200,000 or more |
| num\_returns | INTEGER | Number of tax returns |
| num\_single | INTEGER | Number of filing status single returns |
| num\_joint | INTEGER | Number of filing status joint returns |
| num\_hoh | INTEGER | Number of filing status head of household returns |
| num\_dependents | INTEGER | Total number of dependents |
| num\_elderly | INTEGER | Number of elderly returns |
| Parameters (44 total) | INTEGER | Detailed zipcode-wide metrics about tax returns |

taxes\_dictionary

Detailed description of each of the Parameters in the ***taxes*** data table.

*44 rows & 3 columns*. Size: ~0.1MB. Source: [Internal Revenue Service](https://www.irs.gov/).