**Basic Data Dictionary**

**Zillow Home Value Index (ZHVI):**

Region – Appears to be a primary key index for identifying region.

Region Name – Name for a given region (associated with a unique ID – “Region”)

Region Type – Factor, indicating the type of region for a given record, Levels as follows:

* Country – country level granularity of estimates
* State – state level granularity of estimates
* MSA – metropolitan area granularity of estimates

Regionid – Appears to be an id associated with regions, possibly from the original Zillow dataset. Data is not complete for this field. May want to omit from analysis.

Indicator – Factor corresponding to which level of the index we’re working with.

Indicator Name – User friendly name for indicator (factor level)

Indicator Unit – Scale type for measure of home value – “Value” column (i.e. U.S. Dollars)

Scale – appears to correspond to frequency

Frequency – Frequency of the value measurement

Date – Date of the measurement.

Value – The typical home value for a given region. This value is calculated using the Zillow’s proprietary “Zestimate.” Value is a seasonally adjusted and smoothed measure. Value can be sliced and diced by the factors listed above.

[ZHVI User Guide](https://www.zillow.com/research/zhvi-user-guide/)

[ZHVI Methodology](https://www.zillow.com/research/zhvi-methodology-2019-deep-26226/)

**Zillow Observed rent index (ZORI):**

Region – Appears to be a primary key index for identifying region.

Region Name – Name for a given region (associated with a unique ID – “Region”)

Region Type – Factor, indicating the type of region for a given record, Levels as follows:

* Country – country level granularity of estimates
* State – state level granularity of estimates
* MSA – metropolitan area granularity of estimates

Regionid – Appears to be an id associated with regions, possibly from the original Zillow dataset. Data is not complete for this field. May want to omit from analysis.

Indicator – Factor corresponding to which level of the index we’re working with.

Indicator Name – User friendly name for indicator (factor level)

Indicator Unit – Scale type for measure of home value – “Value” column (i.e. U.S. Dollars)

Scale – appears to correspond to frequency

Frequency – Frequency of the value measurement

Date – Date of the measurement.

Value – The typical observed market rate rent for a given region for a given region. Weighted to rental housing stock (what’s available). Dollar value is given by taking the mean of listed rents that fall into the 40th to 60th percentile range for a given region, again weighted to reflect rental housing stock.

[Zillow Observed Rent Index (ZORI) Methodology](https://www.zillow.com/research/methodology-zori-repeat-rent-27092/)

**Regional Price Parities by State and Metro Areas (Transformed Regional Price Parities by MSA.csv)**

[**https://www.bea.gov/data/prices-inflation/regional-price-parities-state-and-metro-area**](https://www.bea.gov/data/prices-inflation/regional-price-parities-state-and-metro-area)

Regional price parities (RPPs) measure the differences in price levels across states and metropolitan areas for a given year and are expressed as a percentage of the overall national price level.

The transformed version has been cleaned up for PowerBI and structure has been changed to be more conducive to time series analysis.

GeoName – synonymous with Metropolitan Statistical Area (MSA)

LineCode – indication for factor level (most likely irrelevant for us but left it in case)

Description – factor indicating which type of expenses a particular price parity measurement applies to

Year – the year the price parity was recorded

Percentage of Mean cost US – price parity expressed as a percentage of the overall national price level.

**Per Capita Personal Consumption Expenditures by State and Major Product(Transformed Per Capita Personal Consumption Expenditures by State.csv)**

[**https://www.bea.gov/news/2021/personal-consumption-expenditures-state-2020**](https://www.bea.gov/news/2021/personal-consumption-expenditures-state-2020)

Income that people get from wages, proprietors' income, dividends, interest, rents, and government benefits. A person's income is counted in the county, metropolitan statistical area, or other area where they live, even if they work elsewhere.

The transformed version has been cleaned up for PowerBI and structure has been changed to be more conducive to time series analysis.

GeoName – US aggregate or state where the measure was recorded. (We might consider multiplying US aggregate Regional Price Parities to determine dollar cost of certain services in each MSA).

LineCode - indication for factor level (most likely irrelevant for us but left it in case)

Description – factor indicating which type of expense a particular dollar amount applies to

Year – the year the measure was recorded

US Dollars – US dollar amount for a particular consumption expenditure

**Per Capita Personal Income by MSA (Transformed Per Capita Disposable Personal Income by MSA.csv)**

[**https://www.bea.gov/data/income-saving/personal-income-county-metro-and-other-areas**](https://www.bea.gov/data/income-saving/personal-income-county-metro-and-other-areas)

Income that people get from wages, proprietors' income, dividends, interest, rents, and government benefits. A person's income is counted in the county, metropolitan statistical area, or other area where they live, even if they work elsewhere.

The transformed version has been cleaned up for PowerBI and structure has been changed to be more conducive to time series analysis.

GeoName – synonymous with Metropolitan Statistical Area (MSA)

LineCode – indication for factor level (most likely irrelevant for us but left it in case)

Description – factor indicating what is being measured (note population measurements are contained in this file as well)

Year – the year the measurement was recorded

US Dollars – Measurement in US dollars (for income only, NOTE: population is contained in this file as well)