

Project: Import the dataset “Real Estate Texas.csv,” containing data related to real estate sales in Texas. The variables in the dataset are:

- **city:** city
- **year:** reference year
- **month:** reference month
- **sales:** total number of sales
- **volume:** total sales value in millions of dollars
- **median_price:** median sales price in dollars
- **listings:** total number of active listings
- **months_inventory:** amount of time needed to sell all current listings at the current sales pace, expressed in months.

Task:

1. **Identify the variable types contained in the dataset.** (Pay attention to variables that imply time and how they are treated!)
2. **Calculate measures of central tendency, variability, and shape** for all the variables where it makes sense to do so. For other variables, create a frequency distribution. Provide brief commentary on your findings.
3. **Which variable has the highest variability?** How did you determine this? And which one is the most skewed?
4. **Divide one of the quantitative variables into classes** (you choose which one and how), construct the frequency distribution, create the corresponding bar chart, and finally calculate the Gini index.
5. **"Guess" the Gini index for the variable "city."**
6. **What is the probability that a randomly selected row from this dataset reports the city "Beaumont"?** What is the probability that it reports the month of July? And the probability that it reports December 2012?
7. **There is a column with the median price.** Create another column that indicates the mean price, using the other variables at your disposal.
8. **Try to create another column that provides an idea of the "effectiveness" of the sales listings.** Can you make any observations based on this?
9. **Try to create some summary statistics (or simply mean and standard deviation) of some variables of your choice, conditional on the city, year, and month.**