

Data Wrangling Steps

As far as gathering and organizing the data for this project, my steps were not too complicated.

In order to prepare to wrangle the data I selected four sources:

- Income by job title from bls.gov
- Income by location from bls.gov
- Income by experience from kaggle
- Income by education level from smartasset.com

The wrangling part of things came next and went as follows:

1. I went through each source and delete unnecessary columns. For example there was a column in the location source having to do with state code.
2. I searched for extreme outliers and deleted them.
3. I made sure each source had a common numerical based value. In this case it was average salary.
4. I reformatted average salary for each source to be in a US currency based format.
5. I went through the income by location source and combine rows so that the dataset represented average income by county.
6. The final step was just reviewing steps 1-5 and making sure no mistakes were made.

There was only 1 outlier and it was very extreme so I removed that row of data. There were no missing values.