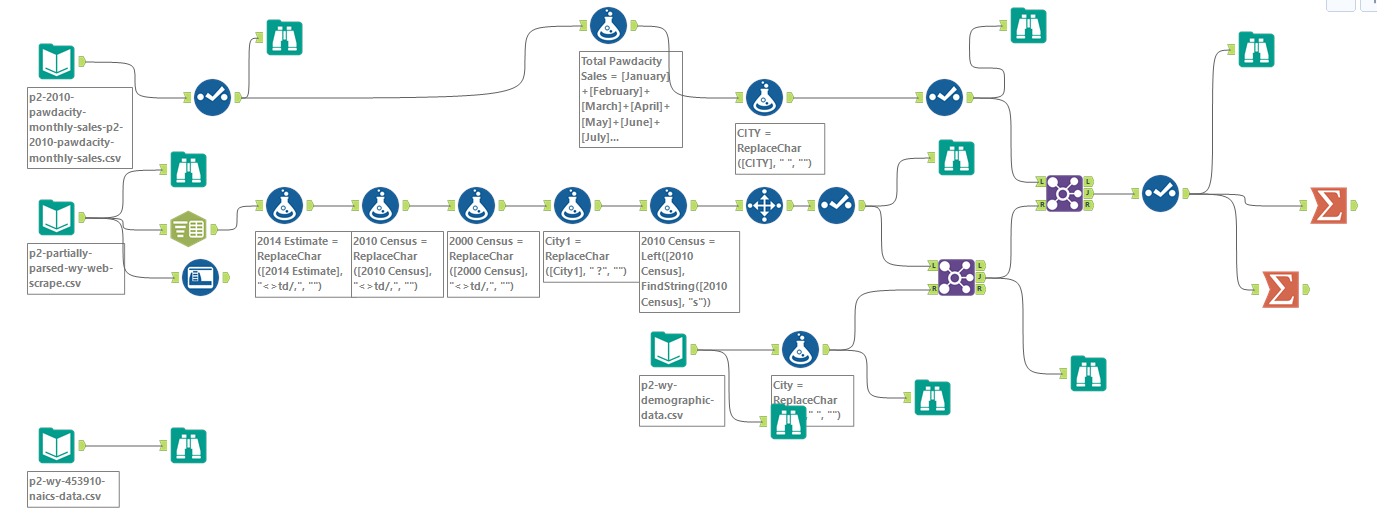
Project 2.1: Data Cleanup

## Step 1: Business and Data Understanding

***Pawdacity is a leading pet store chain in Wyoming with 13 stores in the state. They are looking to open a 14th store and would like recommendation the city for newest store based on predicted yearly sales. I picked monthly sales data for year 2010, demographic data for each city and county in Wyoming, population numbers and the sales of all competitor stores to make the analysis.***

## Step 2: Building the Training Set

|  |  |  |
| --- | --- | --- |
| **Column** | **Sum** | **Average** |
| *Census Population* | *213,862* | *19,422* |
| *Total Pawdacity Sales* | *3,773,304* | *343027.64* |
| *Households with Under 18* | *34,064* | *3096.73* |
| *Land Area* | *33,071* | *3006.49* |
| *Population Density* | *63* | *5.71* |
| *Total Families* | *62,653* | *5695.71* |



## Step 3: Dealing with Outliers

***After we pull the final data into Excel, we performed the IQR analysis to find the outliers.***

***City Cheyenne is the outlier in the training set. Since both value for Total sales and population density is outside of upper fence. This is a small data set with only 11 cities, I will choose to remove this city.***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| CITY | 2010 Census | Total Pawdacity Sales | Households with Under 18 | Land Area | Population Density | Total Families |
| Buffalo | 4585 | 185328 | 746 | 3115.5075 | 1.55 | 1819.5 |
| Casper | 35316 | 317736 | 7788 | 3894.3091 | 11.16 | 8756.32 |
| **Cheyenne** | 59466 | **917892** | 7158 | 1500.1784 | **20.34** | 14612.64 |
| Cody | 9520 | 218376 | 1403 | 2998.95696 | 1.82 | 3515.62 |
| Douglas | 6120 | 208008 | 832 | 1829.4651 | 1.46 | 1744.08 |
| Evanston | 12359 | 283824 | 1486 | 999.4971 | 4.95 | 2712.64 |
| Gillette | 29087 | **543132** | 4052 | 2748.8529 | 5.8 | 7189.43 |
| Powell | 6314 | 233928 | 1251 | 2673.57455 | 1.62 | 3134.18 |
| Riverton | 10615 | 303264 | 2680 | 4796.859815 | 2.34 | 5556.49 |
| RockSprings | 23036 | 253584 | 4022 | 6620.201916 | 2.78 | 7572.18 |
| Sheridan | 17444 | 308232 | 2646 | 1893.977048 | 8.98 | 6039.71 |
|  |  |  |  |  |  |  |
| **Q1** | 6314 | 218376 | 1251 | 1829.4651 | 1.62 | 2712.64 |
| **Q3** | 29087 | 317736 | 4052 | 3894.3091 | 8.98 | 7572.18 |
| **IQR:** | 22773 | 99360 | 2801 | 2064.844 | 7.36 | 4859.54 |
| **Upper:** | 63246.5 | 466776 | 8253.5 | 6991.5751 | 20.02 | 14861.49 |
| **Lower:** | -27845.5 | 69336 | -2950.5 | -1267.8009 | -9.42 | -4576.67 |