Project 2.1: Data Cleanup

Sarah Alomran

Step 1: Business and Data Understanding

Key Decisions:

1. What decisions needs to be made?

Pawdacity would like to expand and open a 14th store, the decision needs to be made is which city should Pawdacity open its new store in.

2. What data is needed to inform those decisions?

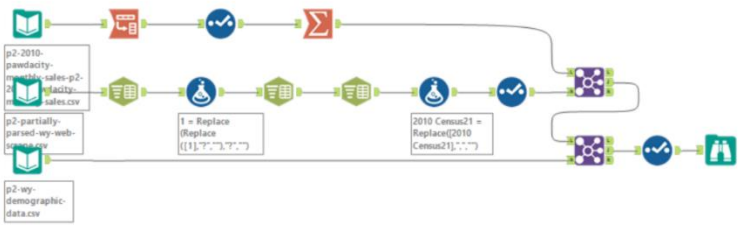
We need to know monthly sales data of each store, population and demographic data of the cities where Pawdacity has operations.

Step 2: Building the Training Set

By building a clean training set given the data provided, we get the following result.

Column	Sum	Average
Census Population	213,862	19442
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

Alterys workflow:



Step 3: Dealing with Outliers

I exported the cleaned and joined data into an excel file, the red highlighted cells are the outliers since they are above the upper fence.
Gillette and Rock Springs values are close enough to the upper fence Hence Cheyenne should be the outlier city.

City	2010 Census Population	Total Pawdacity Sales	Households with Under 18	Land Area	Popuation Density	Total Families
Buffalo	4585	185328	746	3115.508	1.55	1819.5
Casper	35316	317736	7788	3894.309	11.16	8756.32
Cheyenne	59466	917892	7158	1500.178	20.34	14612.64
Cody	9520	218376	1403	2998.957	1.82	3515.62
Douglas	6120	208008	832	1829.465	1.46	1744.08
Evanston	12359	283824	1486	999.4971	4.95	2712.64
Gillette	29087	543132	4052	2748.853	5.8	7189.43
Powell	6314	233928	1251	2673.575	1.62	3134.18
Riverton	10615	303264	2680	4796.86	2.34	5556.49
Rock Springs	23036	253584	4022	6620.202	2.78	7572.18
Sheridan	17444	308232	2646	1893.977	8.98	6039.71
Outliers						
Median	12359	283824	2646	2748.853	2.78	5556.49
1st Quartile	6314	218376	1251	1829.465	1.62	2712.64
3rd Quatile	29087	317736	4052	3894.309	8.98	7572.18
IQR	22773	99360	2801	2064.844	7.36	4859.54
Upper Fence	63246.5	466776	8253.5	6991.575	20.02	14861.49
Lower Fence	-27845.5	69336	-2950.5	-1267.8	-9.42	-4576.67