Predictive Analytics for Business Nanodegree

Project: Create an Analytical Dataset

Laila Hussain Alqawain

_____

Step 1: Business and Data Understanding

Key Decisions:

1. What decisions needs to be made?

   The decision to need to be made is that we should decide which city should new store be located in.

2. What data is needed to inform those decisions?

   We need to know store locations of company, total of all months for 2010, population data and demographic data.

Step 2: Building the Training Set

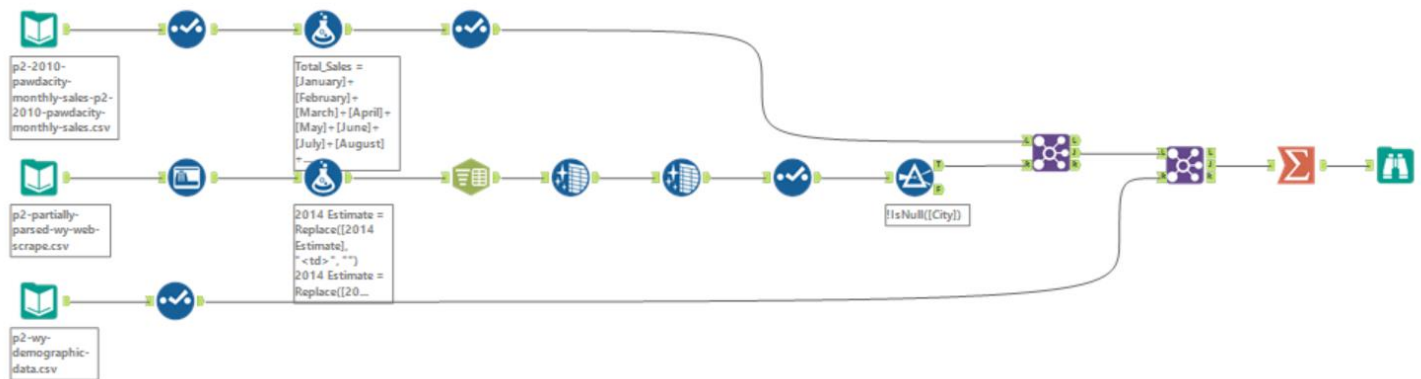I reached the same results which provided.

Sum_2010 Census	Sum_Total_Sales	Sum_Total_Sales2	Sum_2010 Census2	Sum_Land Area
213862	3773304	3773304	213862	33071.380389

Sum_Land Area	Sum_Households with Under 18	Sum_Population Density
33071.380389	34064	62.8

The outputs are **11** records.

Record	City	Total_Sales	2010 Census	County	Land Area	Households with Under 18	Population Density	Total Families
1	Buffalo	185328	4585	Johnson	3115.5075	746	1.55	1819.5
2	Casper	317736	35316	Natrona	3894.3091	7788	11.16	8756.32
3	Cheyenne	917892	59466	Laramie	1500.1784	7158	20.34	14612.64
4	Cody	218376	9520	Park	2998.95696	1403	1.82	3515.62
5	Douglas	208008	6120	Converse	1829.4651	832	1.46	1744.08
6	Evanston	283824	12359	Uinta	999.4971	1486	4.95	2712.64
7	Gillette	543132	29087	Campbell	2748.8529	4052	5.8	7189.43
8	Powell	233928	6314	Park	2673.57455	1251	1.62	3134.18
9	Riverton	303264	10615	Fremont	4796.859815	2680	2.34	5556.49
10	Rock Springs	253584	23036	Sweetwater	6620.201916	4022	2.78	7572.18
11	Sheridan	308232	17444	Sheridan	1893.977048	2646	8.98	6039.71

This is the workflow

Step 3: Dealing with Outliers

A s we see, there are some outliers in **Cheyenne** city, so I'll remove it.

City	Total_Sales	2010 Census	County	Land Area	Households with Under 18	Population Density	Total Families
Buffalo	185328	4585	Johnson	3115.5075	746	1.55	1819.5
Casper	317736	35316	Natrona	3894.3091	7788	11.16	8756.32
Cheyenne	917892	59466	Laramie	1500.1784	7158	20.34	14612.64
Cody	218376	9520	Park	2998.95696	1403	1.82	3515.62
Douglas	208008	6120	Converse	1829.4651	832	1.46	1744.08
Evanston	283824	12359	Uinta	999.4971	1486	4.95	2712.64
Gillette	543132	29087	Campbell	2748.8529	4052	5.8	7189.43
Powell	233928	6314	Park	2673.57455	1251	1.62	3134.18
Riverton	303264	10615	Fremont	4796.859815	2680	2.34	5556.49
Rock Springs	253584	23036	Sweetwater	6620.201916	4022	2.78	7572.18
Sheridan	308232	17444	Sheridan	1893.977048	2646	8.98	6039.71
Q1	**226152**	**7917**		**1861.721074**	**1327**	**1.72**	**2923.41**
Q3	**312984**	**26061.5**		**3504.9083**	**4037**	**7.39**	**7380.805**
IQR	**86832**	**18144.5**		**1643.187226**	**2710**	**5.67**	**4457.395**
Upper Fence	**443232**	**53278.25**		**5969.689139**	**8102**	**15.895**	**14066.8975**
Lower Fence	**95904**	**-19299.75**		**-603.059765**	**-2738**	**-6.785**	**-3762.6825**