

Step 1: Business and Data Understanding

Q1. What decision needs to be made?

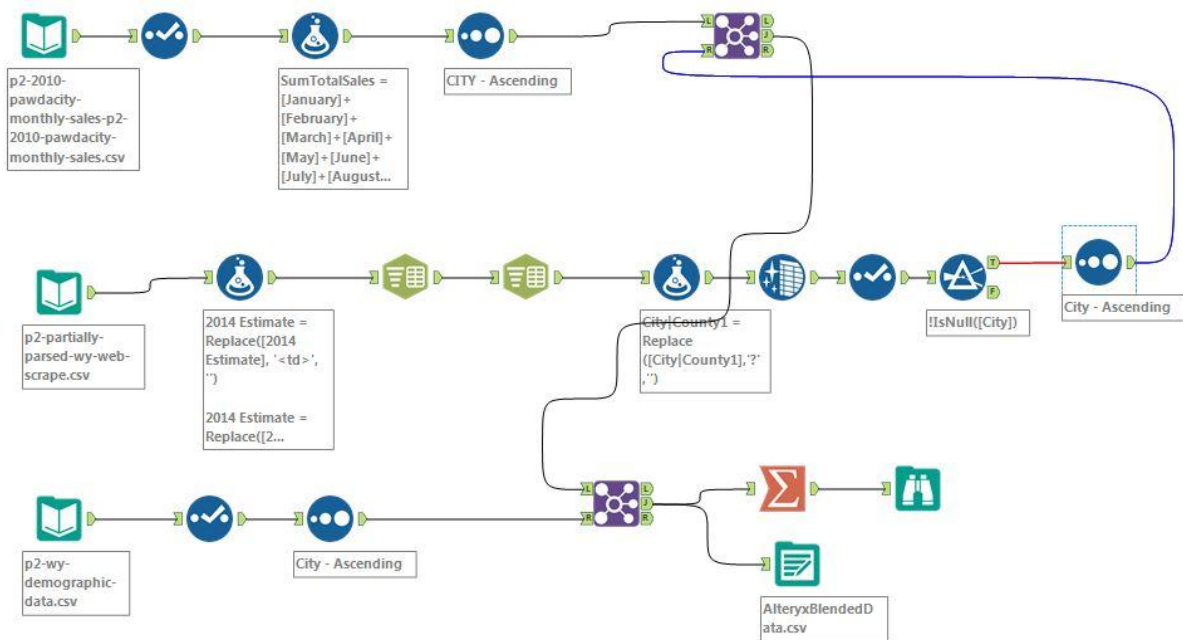
Ans 1. The decision that needs to be made is to give a recommendation for a new city for opening a new Pawdacity store .

Q2. What data is needed to inform those decisions?

Ans 2. Sales data for previous years for already functioning Pawdacity stores segregated based on various demographics like population density, land area etc. would be required.

Step 2: Building the Training Set

The Alteryx workflow has been included in the zip file. Given below are two screenshots, first one is of the Alteryx workflow and the second one is of the values calculated from the data.



Record #	Census Population	Total Pawdacity Sales	Households with Under 18	Land Area	Population Density	Total Families
1	213862	3773304	34064	33071.380389	62.8	62652.79

Step 3: Dealing with Outliers

An excel sheet with all the calculations for computing outliers has been included in the zip file. The city 'Cheyenne' should be imputed (as it is a very small data set with just 11 cities) as it has outliers in two attributes – 'Population Density' and 'Total Pawdacity Sales'.