Project 2.1: Data Cleanup

Step 1: Business and Data Understanding

Provide an explanation of the key decisions that need to be made. (250 word limit)

Key Decisions:

Answer these questions

1. What decisions needs to be made?
   Ans. Pawdacity is a leading pet store chain in Wyoming, In this we need to recommend a city of on where to open a newest 14th store.

2. What data is needed to inform those decisions?
   Ans. To inform these decisions we required some datasets like City, Population Density, Household with under 18, land area, 2010 Census Population, Total Sales in Pawdacity and Total Families.

Step 2: Building the Training Set

Build your training set given the data provided to you. Your column sums of your dataset should match the sums in the table below.

In addition provide the averages on your data set here to help reviewers check your work. You should round up to two decimal places, ex: 1.24

Column	Sum	Average
Census Population	*213,862*	*19,442*
Total Pawdacity Sales	*3,773,304*	*343027.65*
Households with Under 18	*34,064*	*3,096.7*
Land Area	*33,071*	*3,006.49*
Population Density	*63*	*5.72*
Total Families	*62,653*	*5695.72*

Step 3: Dealing with Outliers

Answer these questions

Are there any cities that are outliers in the training set? Which outlier have you chosen to remove or impute? Because this dataset is a small data set (11 cities), **you should only remove or impute one outlier**. Please explain your reasoning.

Ans. From the result of the last Join Tool and Summarize Tool we predict that Gillette and Cheyenne seems to be the outliers as their sales data is higher than the expected or others.

Record #	STATE	CITY	Total Sales	NAME
1	WY	Buffalo	185328	Pawdacity
2	WY	Casper	317736	Pawdacity
3	WY	Cheyenne	917892	Pawdacity
4	WY	Cody	218376	Pawdacity
5	WY	Douglas	208008	Pawdacity
6	WY	Evanston	283824	Pawdacity
7	WY	Gillette	543132	Pawdacity
8	WY	Powell	233928	Pawdacity
9	WY	Riverton	303264	Pawdacity
10	WY	Rock Springs	253584	Pawdacity
11	WY	Sheridan	308232	Pawdacity

Alteryx WorkFlow