

## 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?  
We need to decide in which city to build a new store.
2. What data is needed to inform those decisions?  
We will need city-by-city historical sales data of pet stores and pawdacity stores, as well as population and demographics data of cities.

## 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	19442
Total Pawdacity Sales	3,773,304	343027.64
Households with Under 18	34,064	3096.72
Land Area	33,071	3006.49
Population Density	63	5.71
Total Families	62,653	5695.71

## Step 3: Dealing with Outliers

*Answer these questions*

Are there any 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.

I calculated the interquartile range (IQR, 75<sup>th</sup> percentile - 25<sup>th</sup> percentile) and looked for anything that was above 75<sup>th</sup> pct + 1.5\*IQR or below 25<sup>th</sup> pct - 1.5 \* IQR. Cheyenne was an outlier in a few categories: total sales, census population, population density, total families, and under 18 households. A few other cities were an outlier in one category, but Cheyenne is the outlier I would remove.

Outlier analysis:

Casper

Households with Under 18 above 1.5\*IQR: 6747.0 value: 7788

Cheyenne

Total Pawdacity Sales above 1.5\*IQR: 399816.0 value: 917892

Cheyenne

2010 Census above 1.5\*IQR: 44206.0 value: 59466

Cheyenne

Population Density above 1.5\*IQR: 13.06 value: 20.34

Cheyenne

Total Families above 1.5\*IQR: 11838.2 value: 14612.64

Cheyenne

Households with Under 18 above 1.5\*IQR: 6747.0 value: 7158

Gillette

Total Pawdacity Sales above 1.5\*IQR: 399816.0 value: 543132

Rock Springs

Land Area above 1.5\*IQR: 5148.095526 value: 6620.201916