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.

1. What data is needed to inform those decisions?

We will need county-by-county historical sales data, population and demographics data of counties/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* |  |
| *Total Pawdacity Sales* | *3,773,304* |  |
| *Households with Under 18* | *34,064* |  |
| *Land Area* | *33,071* |  |
| *Population Density* | *63* |  |
| *Total Families* | *62,653* |  |

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.