Data Cleanup

REVIEW

HISTORY

Meets Specifications

Hello Udacian!
Congratulations!!! You made it. 🎉 😀
You should be proud of your work, I know I am proud of it.
Your first submission was really outstanding and you have improved your work very well.  :+1:
I am sure you have enjoyed working on this project, I can see the hard work in your project.
I really hope the lectures and this project in particular have effectively taught you the fundamentals of this course.
Well Done 👏
It was a pleasure reviewing your project.
I wish you all the best!
Happy Learning 😇

Business and Data Understanding

✓

The section is written clearly and is concise. The section is written in less than 250 words.

Good job!
All answers are written clearly and is concise, and not exceeding more than 250 words.

✓

All the following questions have been accurately answered:

1. What decisions need to be made?
2. What data is needed to inform those decisions?

Good job!
Your answers are correct for :
What decisions need to be made?
Correct! To decide the newest store location for Pawdacity depending on the highest sales value in a year.
What data is needed to inform those decisions?
✅
So data that we need in order to inform this decision, like "sales, city, 2010 census Population, Age groups, Total families, households under 18 and some geographical data like Land area" etc.

Building the Training Set

✓

The averages for each column is correct in the training set

Good job!
The averages for each column is correct in the training set, your calculations are correct 👏🏼

✓

Outliers have been analyzed for each field in the training set.

The outliers are accurately identified.

The decision to keep, remove, or impute each outlier is well justified.

After pointing out all the possible outlier cities only one city should be decided upon to remove.

Good job!
You have correctly identified Cheyenne & Gillette as an outlier. You have correctly used the Interquartile Range (IQR) method to calculate the Upper and Lower Fences.
You correctly justified the decision to keep, remove, or impute each outlier as well :+1:
Suggestion
You can check the lesson for Data Issues section where you will find the video for "what is an outlier", once you go through this lesson you will understand it clearly.
And you can also plot the Scatterplots of Population-related variables versus Pawdacity Total Sales and then figure out the outliers, check the greatest distance from the linear trend (that could be the outlier) .

- For IQR method :
  Calculate the Interquartile Range: IQR = Q3 - Q1

Add the product of 1.5 & IQR to Q3 to get the upper fence
Subtract product of 1.5 & IQR from Q1 to get the lower fence

DOWNLOAD PROJECT

RETURN TO PATH

Rate this review

START