



SPSS Homework Data Preparation WEEK 4



**Data Science
Academy**

Homework



Disclaimer: Homework is made up for educational purposes.

Answer the following questions.

1. Open "**Customer_dbase.sav**". Find outliers and extreme values. Discard extremes and coarse outliers for variable having many extreme values. For other variables, use the nullify option.
2. Open "**Future-500.xlsx**". Clear '**Dollars**' in **Expenses** column using Derive node.

Tip: use **replace()** function.

3. Derive 3 separate new Flag (True or False) columns (Column names: Low, Moderate and High) from Growth field. Growth intervals for new variables are listed.
 - a. '**Low**' 0.0-0.30
 - b. '**Moderate**' 0.30-0.70
 - c. '**High**' more than 0.70
4. Calculate total profit for each industry type using **Aggregate** node.
5. Generate new variable called 'Margin' equal to Profit / Revenue using Derive node.
6. Exclude data where Margin is less than 0.1.
7. Select IT service from "**Industry**" field using Select node and Audit data for IT Services.
8. Decrease decimal places to zero in "**ID**", "**Inception**", "**Employees**" fields.

Tip: use Format tab of Table node.