data-mining-homework-2

This README acts as the log addressed in the assignment.

Files

hw2

The main file. Does the homework.

util

A collection of general functions that are used across the project.

Diagram

Contains the code for the pyplot graphs.

Math Utils

Takes care of the math envoved in the linear regression.

dataOperations

Handles Step 4 with approach one and approach two.

constants

Creates the constants for the dataOperations file to use (DOE ratings).

analysis

Analyzes the data to help create the pyplot graphs.

Log

Resolving Cases

1. Created a dictionary to hold the indexes of each of the columns in our dataset. Similarly, we created two lists to store categorical and continuous data.
2. Wrote the functions in diagram.py that will graph each of the graphs required.

3. Wrote the function calls in hw2.py to get the graphs to display as a pdf.

Hiccups

None.

Steps

Step 1

Created the frequency diagram. Frequency diagrams are drawn for each of the categorical attributes.

Step 2

Created the pie chart. Pie charts are drawn for each of the categorical attributes.

Step 3

Created the dot chart. Dot charts are drawn for each of the continuous attributes.

Step 4

Approach 1: We used a dictionary in our file constants.py to convert continuous attribute to categorical attribute.

Approach 2: We created 5 equal width bins to denote the subranges of the MPG values.

Step 5

Created the histogram. Histograms are drawn for each of the continuous attributes.

Step 6

Created the scatter plot. Scatter plots are drawn for each of the continuous attributes vs MPG.

Step 7

Drew the linear regression line for each of the scatter plots drawn.
Did this by using the formulas given in the lecture notes. Needed to created a new file to do this.

Step 8

Part 1: Created a graph with multiple box plots. The y axis is the MPG and the x axis is the car years. Each of the car year displays its own box plots.

Part 2: Created a multiple frequency diagram. X and Y axis are the same as in part 1, but the colors of the frequency diagram bars represent where the car was made.