MIS0855: Data Science

Assignment 1: Analyze a Data Set Using Tableau

Task:

Use Tableau to analyze and reveal various relationships within a data set.

Scenario:

Earlier in the course you worked with a data set containing fuel economy data for 2015 model year cars. Now you’re going to work with that same data set in Tableau to answer a series of questions.

Deliverable:

1. Create a Tableau workbook based on the Excel data file “2015 Car Fuel Econ.xlsx.”
2. Create eight data visualizations, one for each question listed on the next page. Each visualization should be on a separate worksheet within the workbook and it should provide the answer to the complete question. Label each worksheet in the Tableau workbook as “Question 1,” “Question 2,” and so on.
3. Using the visualization you created in 2 above, answer each question using the Deliverable Worksheet at the end of this document.
4. You should use the chart type specified by the question.

Tips:

* You should construct your visualizations so that they appropriately display the data.
* You should create a graphic for all questions EXCEPT question 2 – for that one create a text table.
* Keep in mind principles that we’ve discussed in class, such as good use of colors, legends, scale, and aspect ratio. Don’t be afraid to play around with fonts and colors. Also remember, simplicity is good.
* Make sure you are aware of when to use sums versus averages.
* Take the hints seriously – they will help you!

You should **upload** the Deliverable Worksheet and your Tableau Workbook to Canvas by the start of class on the due date.

Grading:

For each question, your work will be evaluated using the following criteria:

|  |  |  |
| --- | --- | --- |
| Criteria | Weight | Description |
| Answer | 50% | Is the answer to the question correct? |
| Visualization Method | 20% | Is the right visualization chosen as specified in the instructions? |
| Visualization Design | 30% | Are the axes and data points labeled properly and readable? |

Questions:

Question 1 – Use a treemap:

Which car manufacturer has the greatest number of carlines?  
(consider variations as separate models-use count function – i.e., BMW 228i versus BMW 228i xDrive)

Question 2 – Use a single data (text) table:

1. Which car manufacturer has the highest **average** fuel economy for city driving?
2. Which car manufacturer has the highest **average** fuel economy for highway driving?

Question 3 – Use a bar chart:

Which carline has the greatest **average** difference between its highway fuel economy and its city fuel economy?  
(HINT: Create a calculated field for the difference and then create a graphic using that value.)

Question 4 – Use a line chart:

What is the relationship between **average** (combined) fuel economy and more powerful engines (i.e., higher engine displacement)? (i.e., when one variable is high/low, what is the value of the other one?)

Question 5 – Use a bar chart:

1. Which transmission type has the highest overall **average** fuel economy?
2. Which transmission type has the worst overall **average** fuel economy?

Question 6 – Use a bar chart:

Which manufacturers have, on **average**, more powerful engines in their four-wheel drive small SUVs than their two-wheel drive small SUVs? (HINT: Use a filter! List all of them and ignore the ones that do not have one of the versions)

Question 7 – Use a scatter plot (only one graphic for a, b and c):

1. What is the relationship between city fuel economy (MPG) and highway fuel economy (MPG)? (i.e., when one variable is high/low, what is the value of the other one?)
2. Using the same graphic, identify the single carline with the lowest city AND highway fuel economy?
3. Which manufacturer makes that car?  
   (HINT: Use labels!)

Question 8 – Use a scatter plot (only one graphic for a, b and c):

1. What is the relationship between engine displacement and overall (combined) CO2 emissions? (i.e., when one variable is high/low, what is the value of the other one?)
2. Which carlines get better city mileage than highway mileage?
3. What can you say about their CO2 emissions of those cars in (b), compared to the rest of the group?

(HINTS: Use a calculated field to create a categorical variable to differentiate between the two groups. Use the format of the calculated field from the in-class activity as a guide to create this one. Then use that new variable to color code your data points. Finally, put labels containing the model names on your data points.)

Deliverable Worksheet Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Use this worksheet to complete the assignment. The questions are on page 2.

|  |  |
| --- | --- |
| Question | Your Answer |
| 1 |  |
| 2a |  |
| 2b |  |
| 3 |  |
| 4 |  |
| 5a |  |
| 5b |  |
| 6 |  |
| 7a |  |
| 7b |  |
| 7c |  |
| 8a |  |
| 8b |  |
| 8c |  |