AAE 875: Fundamentals of OOP and Data Analytics

Week 4, part 1 and 2: IO files and Data Visualization

Before you begin any work make sure to set your input and output working directories in Eclipse. When you are ready tackle the following exercises.

Exercise 1: Output a dictionary containing coefficient estimates derived from a regression that estimates the causal relationship between price and market competition. As you can see in the figure below I have already obtained the competition estimates for you. Print the output to a .txt file. Then output to a .csv file. What is the data type of your output in both cases?

```
In [ ]:  competition_coef = {
             "0.1 quantile": 0.124,
             "0.3 quantile": 0.234,
             "0.5 quantile": 0.423,
             "0.7 quantile": 0.287,
             "0.9 quantile": 0.675
         }
```

Exercise 2: Your next task is to write code to input the .txt and .csv files resulted from your solution to Exercise 1. What data type would these files represent in Python? Finally create a dictionary using the input data, add the key-value pair "mean": 0.345, and output it again in .txt and .csv format.

Exercise 3: Now that you have learned how to write code to input and output files let's explore the topic of data visualization. A popular dataset for learning data exploration is the dataset containing various information about the passengers of the Titanic. The titanic.csv dataset can be found on the course website. Write code to import it in Eclipse. What is the number of rows, column names and data types of columns?

First, visualize this data using the matplotlib.pyplot package. Plot the number of males and females and the number of people who survived by class type. Style your plots using string formatting and line properties. Add a legend and title to each plot.

Second, visualize the data and plot the same variables using the seaborn package. Before you start to write code, what are the datasets available in the seaborn package? Import the titanic dataset using the seaborn package. Is the number of rows, column names and data types of columns the same? If not, what is the difference?

Finally, which package (mathplotlib, seaborn) is more useful to visualize the titanic dataset?