ITU Artificial Intelligence & Data Engineering Department
YZV 211E Intro to Data Science & Engineering, Fall 2022
Homework #2
Due October 28, 2022 11:59pm

Problem Definition

Along with analyzing the data with statistical tools, visual analysis can help data analysts to see trends, relations and aggregations within the datasets more easily. It is also a good way to present your findings in a visually appealing way to your peers or supervisors.

In this homework, we will be working with the survey data that you have participated. We will be asking questions to data and try to get an answer by creating plots. First of all, **you need to clean the survey data as it includes NaN and invalid values**. You may perform the cleaning manually, since there is not much data. **However**, **formatting for the "Neighborhood" column should be fixed with Python codes in relevant cell within the notebook.** Also, you should explain your cleaning process in several words as it is described in the notebook. **After achieving the clean data, answer 5 questions with creating plots and explaining what have you understand from the plots**.

Requirements

- Use both **matplotlib** and **seaborn** for plots.
- **You should fully clean the data before moving on the plotting section and explain your process in several words.** This should not take your much time. **Consider the cleaning instructions and expected data types specified in the notebook. If there is not an instruction for specific case, you can clean the data with your own way** You might check your cleaning process by checking the data types after (df.dtypes)
- **Make explanations on the plots** you have generated to show what you have understand from the plot. **What does the graph tells you about the data?**
- Be aware that your plots should have **meaningful title, labels and legends (if necessary).**
- Graphs should be visually appealing such that it can show how much effort you put in to those graphs.
- Python 3.9.5 is recommended. Required libraries and preferred versions are listed below (**Versions are not selected specifically. Therefore, you may work with other versions as long as they are close to the latest version. Old versions are not recommended.**). Openpyxl is needed to read excel file.
  - matplotlib 3.5.2
  - pandas 1.3.4
  - numpy 1.22.3
  - basemap 1.3.4
  - openpyxl 3.0.10

Homework Deliverables

- CSV file which was cleaned by you
- Jupyter Notebook including plots and explanations

Homework Rules

- You must submit your solution as a Jupyter (iPython) Notebook.
- Solution must include code and your analysis/comments. If the analysis is not self-explanatory and does not include adequate comments, you might get penalized.
- Do not share any code or text that can be submitted as a part of an assignment with your friends (discussing ideas is okay).
- Do not copy-paste code from internet sources. Do not post homework assignment problems to the internet to ask for help. If you need help, e-mail, come and talk to us.
- You may discuss the problems at an abstract level with your classmates, but you should not **share or copy code** from your classmates or the Internet. You should submit your **own and individual** homework.
- Only electronic submissions through Ninova will be accepted no later than the deadline.
- Academic dishonesty, including cheating, plagiarism, and direct copying, is unacceptable.
- If you have any questions about the homework, you can send an e-mail to Erhan Biçer (bicer21@itu.edu.tr).
- Note: The **submitted solutions WILL BE CHECKED WITH THE PLAGIARISM TOOLS!**