

Objective: Exploratory Data Analysis

The objective of this project is to allow you to exercise a very crucial toolkit of the AI/ML flow: data cleaning and data visualization. Being fluent in these skills will be the foundational setup for how you approach data modeling.

You can accept the assignment here <https://classroom.github.com/a/gc2XS9a1>

Make sure you accept the assignment and read through it!

Details

1. We will provide you with dataset(s) that contain 12 months worth of sales data.
2. Your task to answer the following questions:
 - a. Which month had the highest and lowest sales? What do you think the reason for this was?
 - b. At what time of the day would you put out advertisements so that sales increase?
 - c. What products were sold the most?

Expected Output

1. Add project to jupyter notebook, and upload the notebook to github
2. Have 3 visualizations that give a clear indication of an answer to the above questions.

End Date for submissions : 1st Oct.

Grading Criteria

1. The visualizations created- which plot is used(most important), labelling, scaling etc.
2. The use of numpy and pandas to clean and explore the data
3. Insights pulled from the questions asked
4. Software Engineering skills like clean code with understandable variable names, comments etc.

Bonus Points

1. For digging deeper into the visualizations that answer the question.
 - a. Example: If you were given a dataset about shopping trends vs cities and san francisco showed high sales and your guess is that it is because san fran has high population density, then creating visualization that shows the top 25 cities with high population density will show that you have gone the extra mile to understand the insight further.