Course	DS 203 : Programming for Data Science
Activity	Type: Exercise Title: E01 – Data Analysis using Pivot Tables
Overview	• Pivot Tables are data summarization tools • Structured data (Tabular data) can be analyzed quickly and effectively by creating Pivot Tables. • Columns containing "categorical data" can be used for row and/or column-based grouping / slicing / dicing of data and calculation of aggregates • Aggregates can be counts, sums or even descriptive statistics like mean, standard deviation, variance, etc.
Expected outcomes of this activity	• Understand the concept of Pivot Tables and their applications • Develop expertise in using Pivot Tables during "exploratory data analysis" "data visualization", "detailed data analysis", and general problem solving involving reasonably sized structured data.

Tools	• At least one of the following
	  ○ Google Sheets
	  ○ Microsoft Excel
	  ○ SmartOffice Sheet
	• Recommended: use at least two of these tools to solve the problems

Effort estimate	• **2 – 3 hours**
Submission type	• **Mandatory submission**
Due date and time	• **9th August, 23:55 Hrs**
Submission instruction	• Submit to the appropriate Moodle submission point • Your final submission should be in the form of a single PDF file. ○ (You may use text editors or the text/image documenting capabilities of tools like Jupyter Notebook to compose your document. Convert it to PDF.) • Your solutions and answers should include explanations / graphs / charts / Tables that are necessary to fully explain the solution(s).
Marks for the exercise	• Credit will be given for timely submission to Moodle • The exercise itself will not carry marks. • **Your understanding and skill – expected to be gained by completing this exercise – will be gauged in a quiz, test, or viva that will be conducted subsequently.**
Suggested references	• Interactive help / online help / product documentation • Articles, blogs and other sources
Note of caution	LLM tools will not help you at the time of your actual evaluation. Use them judiciously, to gain good understanding of the subject. The final submission should be yours, not the one created by an LLM.

This exercise is based on the data file **options-trading-backtest.csv** file. It has been uploaded to the course page on Moodle. Follow the instructions given below to analyze the data and answer the questions posed.

1. Based on the **Entry-Date** column, create the following three derived columns as explained in class: **Day, Month, Year**
2. Review the data columns and create a Table to classify and document the type of each column into one of the following: Discrete (categorical), Continuous (numerical).
    a. Hint: Entry-Weekday is 'discrete' while 'Entry-Price' is 'continuous'.
    b. Include this Table in your final submission.
3. In a separate sheet of the spreadsheet create a pivot table
4. Create a filter in the pivot table to only analyze CE trades. (Hint: Use the column **Instrument-Kind** as the basis for filtering, and select only rows with **CE**)
5. Configure the pivot to create the **average Entry-Price** for **every month of every year**.
    a. (Hint: Use the following columns as row aggregators: **Year** followed by **Month**; and use **Entry-Price** as the column for calculating aggregates, take care to ensure that the average is calculated and not the sum – which is the default.
6. Plot a line graph of the data resulting from the above step and interpret it
    a. Include this line graph and your interpretation in your submission.
7. We would like to understand the **Entry-Price** behaviour across every day of the week. In order to achieve this, use the categorical column **Entry-Weekday** in the **columns** section of the pivot table. Once you configure it this way, each distinct value contained in the **Entry-Weekday** is used to create a column in the pivot table. The **average Entry-Price** is now further sliced into a distinct average value for each day of the week for every month of every year.
    a. Take a snapshot of the image of the pivot at this stage and submit it.
8. Create a plot resulting from the above step and analyze it.
    a. Submit the entire plot, along with your analysis of the plot.
9. Modify the pivot configuration to i) remove the weekday breakup ii) create line plots for monthly **average Entry-Price** and **average Exit-Price**
    a. Analyze the plot and submit the plot and your analysis
10. Modify the pivot table filter to only consider **blank** cells under the **Instrument-Kind** column. (This action will result in considering only the total profit / loss made on a day)
11. Create a plot of **monthly P/L**
    a. Create a plot, analyze it and submit the plot and your analysis
12. Create a plot of **average monthly P/L**

a. Create a plot, analyze it by comparing it with the outcome of the above step and submit the plot and your analysis

13. The following tasks must be done **without using the pivot table** i.e. on the raw data but only after applying a column filter to the column **Instrument-Kind** to **filter out blank cells** from the analysis (i.e. consider only cells that are CE/PE)
    a. Create a Histogram of P/L column, analyze the results and submit
    b. Create a Histogram of P/L-Percentage column, analyze the results and submit
    c. Compare the two histograms and explain the reason for the observed differences.
    d. Why is the P/L-Percentage histogram shaped in the form of Normal Distribution?
14. Coming back to the pivot table, carry out any further analysis that you may wish to, and state the problem and submit the results. Following are some pointers for further analysis:
    a. What are the weekday trends of absolute profits and percentage profits? Is there any pattern? Any anomaly observed? Any explanation?
    b. Is there a seasonal pattern of daily profits observed over the months, across the years?
    c. Series of questions!
        i. How many months have ended with profit, how many months have ended with loss?
        ii. What was the max monthly profit?
        iii. What was the max monthly loss?
        iv. Identify periods when there have been consistent losses. For how many consecutive days has this happened? Which year? Why?
        v. Have there been any months with continuous losses?
    d. Can you add some more column(s) to the date to analyze the trends of cumulative P/L over the days, months and years? What questions will you ask of this new data created? How will it help?

*****