FMSR assignment: TTR sales revenue improvement

TTR (The Tea Room) is a chain of tea rooms with stores in various cities in The Netherlands. The company is looking for ways to increase its revenue. They hired you as an analyst to write a clear and concise advisory report for TTR on how to do so. TTR requests you to base your advice on a thorough statistical analysis of the company's sales data.

TTR provides you a sample of sales data from their stores in a CSV file which you can read into Python to do your analysis. The variables in the dataset are:

- transaction_id: Unique ID representing an individual transaction
- transaction_date: Date of the transaction (MM/DD/YY)
- transaction_time: Timestamp of the transaction (HH:MM:SS)
- transaction_qty: Quantity of items sold
- store_id: Unique ID of the tea room where the sale took place
- store_location: Location of the tea room where the sale took place
- unit_price: Sales price of the product sold
- product_category: Description of the product category
- product_type: Description of the type of product

The report consists of three parts. In the first part, TTR asks you to report on three compulsory questions. In the second part, TTR asks you to report the examination of at least 3 other questions. In the third part, you write at least 3 points of advice on how to increase revenue. This advice is based on the findings in part 1 and part 2.

Question 1 (compulsory):
The quantity of items sold varies per transaction. TTR would like to know what the probability is that a transaction contains more than 2 items sold. Report and explain your analysis and conclusion.

Question 2 (compulsory):
TTR would like to obtain insight in how the unit price varies for the product types that they sell. Conduct appropriate analysis, report relevant descriptive statistics and use visualizations when reporting your findings. Include skewness in your interpretation.

Question 3 (compulsory):
TTR would like to obtain insight in the average revenue per transaction. Because you are working with sample data, they would like to also report the interval in which the real (population) average revenue per transaction is likely to fall. Report and explain your analysis and conclusion.

Question 4 through 6 (select yourself):

TTR would like receive three points of advice on increasing revenues, based on three other questions that you examine. To provide sound advice, you report for each question the results of a descriptive analysis and inferential analysis. I.e., based on the results of the descriptive analysis, you make an assumption (you formulate hypotheses) which you test using inferential statistical analysis. Your points of advice are based on those results.

TTR proposes some topics (A through E) to investigate. Topics have multiple questions variants. You are asked to select and examine (at minimum) 3 questions covering (at minimum) 3 different topics:

	Topic	**Question**	***Descriptive analysis***	***Inferential analysis***
A	Product types or categories	A1	How well do different product categories sell in terms of revenue?	Are there product categories that sell significantly worse than others?
		A2	How well do different product types sell in terms of revenue?	Are there product types that sell significantly worse than others?
B	Product types per product category	B1	How well do the different product types sell, in terms of quantity, per product category?	Are there product types that sell significantly worse than others in the same category?
		B2	How well do the different product types sell, in terms of revenue, per product category?	Are there product types that sell significantly worse than others in the same category?
C	Tea room locations	C1	How do tea room locations compare in terms of sales revenue?	Are there differences between locations?
		C2	How do tea room locations compare in terms of sales quantity?	Are there differences between locations?
D	Product categories compared across tea rooms	D1	How well do different product categories sell in terms of revenue in different tea rooms?	Are there product categories that sell significantly worse in some tea rooms than in others?
		D2	How well do different product categories sell in terms of quantity in different tea rooms?	Are there product categories that sell significantly worse in some tea rooms than in others?
E	Evolvement of sales during time	E1	How do sales revenues during the week evolve?	Are there significant differences between days?
		E2	How do sales revenues during the day evolve?	Are there significant differences during the day?

How to go about the examination of these questions?

To illustrate how you could examine these questions, TTR gives you the following example related to the first suggested question: The dataset contains information on sales quantities of different product categories. In the *descriptive analysis* you could report how much revenue is generated per category. Perhaps you observe that one product category generates less revenue than other categories. In your *inferential analysis* you may then want to examine if the sales quantity of this one product category is indeed significantly lower than the sales quantities of the other categories. You could use the outcome to formulate advice about considering whether this product category should be retained.

Instructions for the report

TTR gives you the following instructions to follow when writing the report:

1. Report on the examination of the three compulsory questions explained above (part 1 of your report).
2. Report on the examination of (at least) 3 other questions (part 2 of your report).
    - Report per question the results of the descriptive analysis and inferential statistical analysis.
        - In the descriptive analysis, think about reporting appropriate measures of central tendency and variation and the distribution (skewness) of the numerical data or frequencies of values of categorical data. Use visualizations like histograms or frequency tables for ease of reading.
        - In the inferential analysis, report the step-y-step plan for hypothesis testing including the results.
3. Write at least 3 points of advice for TTR on how to increase their revenue (part 3 of your report).
    - To analyse revenue, generate a new variable that contains the revenue per transaction and include this variable in your analysis. You could use the following code: data["revenue"] = data["transaction_qty"] * data["unit_price"].
4. As data preparation and cleaning is key for reliable analysis, examine, and report on, missing data.
    - Delete observations that have missing data. To check for missing data and to delete observations with missing data, you could, for instance, use the functions isna() and drop.na() from the Pandas package. But feel free to use other suitable functions or methods if you like.
5. For each question, report and explain your analysis, i.e., which test did you use and why, how did you implement it and why.
6. Add a readable Python script as an annex to the report. Make sure that the analysis of every question is accompanied by a working Python script.

Submit:

1. A word or PDF file of the advisory report
2. A .py or .ipynb file with readable code used for your analysis