Assignment 2 - 30 Marks

In this assignment, you have a sales dataset that contains data from different countries and cities. Using python, you are required to explore the data, find the size of the data, visualize the different columns and find the relation between these columns such as cost, profit and revenue. However, before exploring these relations you need to make sure the dataset is free of the following:

- Inconsistent column names
- Missing data
- Outliers
- Duplicate rows (full duplication)
- Untidy
- Need to process columns
- Column types can signal unexpected data value

You are required to explain what you found and how you solved it. Visualize the data before and after the cleaning process.

Grading Criteria

	Unsatisfactory (33%)	**Fair (75%)**	**Good (100%)**
Delivery 5 Points	• Completed less than 50% of the requirements. • Does not comply with requirements (does something other than requirements).	• Completed between 50-64% of the requirements. • Delivered on time, and in correct format.	• Completed between 65-89% of the requirements.
Problems Answered Correctly and Appropriately **5 Points**	• No answer or wrong • Non-informative answer • Answer is ambiguous or not relevant	• Short answer. • Partially informative answer. • Answer is partially clear.	• The answer is just right. • Informative answer • Answer is sufficiently clear.
Concept Comprehension **10 points**	• The student failed to demonstrate a clear understanding of the assignment concepts	• The student demonstrated a clear understanding of some of the assignment concepts	• The student demonstrated a clear understanding of the assignment concepts
Analytics Skills **10 Points**	• The student did not provide analytical insights for the attributes in the data.	• The provided partial analysis of the data and the attributes.	• The student provided insightful analytics for the attributes and underlying data.