DA Session 11 – Correlation & Causation Assignment

Name: Sarthak Sahu
Course: Data Analytics
Platform: PW Skills

Q1. Define correlation. Why is correlation important in data analytics?

Answer:
Correlation is a statistical measure that describes the strength and direction of the relationship between two variables. It is important in data analytics because it helps identify patterns, supports prediction and trend analysis, assists in decision-making, and is widely used in business and academic analysis.

Q2. Explain the different types of correlation based on direction with examples.

Answer:
Positive correlation occurs when both variables increase or decrease together, for example study hours and exam scores. Negative correlation occurs when one variable increases while the other decreases, such as mobile usage and exam scores. Zero correlation means no relationship exists between variables, for example shoe size and exam performance.

Q3. Differentiate between linear and non-linear correlation with suitable examples.

Answer:
Linear correlation represents a straight-line relationship between variables, such as study hours and exam score. Non-linear correlation represents a curved relationship, such as stress level and performance, where moderate stress improves performance but excessive stress reduces it.

Q4. What is Pearson's correlation coefficient? Explain how its values are interpreted.

Answer:
Pearson's correlation coefficient measures the strength and direction of a linear relationship between two continuous variables. Its value ranges from -1 to +1, where +1 indicates perfect positive correlation, -1 indicates perfect negative correlation, and 0 indicates no correlation.

Q5. Why is it risky to assume causation from correlation?

Answer:
It is risky because correlation does not prove cause-and-effect. A third variable may influence both variables, the relationship may be coincidental, or the direction of influence may be unclear. For example, ice cream sales and drowning cases both increase in summer, but ice cream does not cause drowning.

Q6. Calculate Pearson's correlation between Study Hours and Exam Score using Excel. Interpret the result.

Answer:
Using Excel's CORREL() function, a strong positive correlation is obtained between study hours and

exam score. This indicates that students who spend more time studying generally achieve higher exam scores.

Q7. Identify whether Mobile Usage and Exam Score have a positive or negative correlation.

Answer:
Mobile usage and exam score show a negative correlation. This means that increased mobile usage is associated with lower exam scores.

Q8. Does high attendance cause better exam scores? Justify using data.

Answer:
The data shows a positive correlation between attendance and exam scores. However, this does not confirm causation, as other factors such as study habits, motivation, and health can also affect exam performance.

Q9. Create a scatter plot for Attendance vs Exam Score and comment on the nature of correlation.

Answer:
The scatter plot shows an upward trend, indicating a strong positive linear correlation between attendance and exam scores.

Q10. Identify a possible confounding variable in the relationship between Study Hours and Exam Score.

Answer:
Sleep duration is a possible confounding variable. Students who get adequate sleep can study more effectively and perform better in exams, influencing both study hours and exam scores.