# Machine Learning

1. Market trend Preddiction
2. Time Series Data
3. Unsupervised Learning
4. The tree representing how close the data points are to each other
5. A distance metric
6. . k-nearest neighbour is same as k-means
7. i. Single-link

ii. Complete-link

iii. Average-link

8. i. Clustering analysis is negatively affected by multicollinearity of features

ii. Clustering analysis is negatively affected by heteroscedasticity

9. 2
10. Given sales data from a large number of products in a supermarket, estimate future sales for each of these products.
11. Option C
12. Option A
13. They can cluster different customer types into one group based on different factors, such as purchasing patterns. The factors analysed through clustering can have a big impact on sales and customer satisfaction, making it an invaluable tool to boost revenue, cut costs, or sometimes even both.
14. Graph-based clustering performance can easily be improved by applying ICA blind source separation during the graph Laplacian embedding step. Applying unsupervised feature learning to input data using either RICA or SFT, improves clustering performance.

# STATISTICS

1. Total Variation = Residual Variation + Regression Variation
2. Binomial
3. 2
4. Type-I error
5. Level of confidence
6. Increase
7. Hypothesis
8. Minimize errors
9. 0
10. In finance, Bayes' Theorem can be used to rate the risk of lending money to potential borrowers. The theorem is also called Bayes' Rule or Bayes' Law and is the foundation of the field of Bayesian statistics.
11. In finance, Z-scores are measures of an observation's variability and can be used by traders to help determine market volatility. The Z-score is also sometimes known as the Altman Z-score.
12. A *t* test is a statistical test that is used to compare the means of two groups. It is often used in hypothesis testing to determine whether a process or treatment actually has an effect on the population of interest, or whether two groups are different from one another.
13. In statistics, a percentile is a term that describes how a score compares to other scores from the same set. While there is no universal definition of percentile, it is commonly expressed as the percentage of values in a set of data scores that fall below a given value.
14. Analysis of Variance (ANOVA) is a statistical formula used to compare variances across the means (or average) of different groups. A range of scenarios use it to determine if there is any difference between the means of different groups.
15. ANOVA is helpful for testing three or more variables. It is similar to multiple two-sample t-tests. However, it results in fewer type I errors and is appropriate for a range of issues. ANOVA groups differences by comparing the means of each group and includes spreading out the variance into diverse sources.