Assignment: 3

Subject Name: Introduction to Data Mining and data warehouse

Title: Data Processing

1. What is data quality? Explain five factors of high quality data.
2. Define following terms
   1) data cleaning
   2) noisy data
   3) Data transformation
   4) sampling
3. What is missing values? Explain methods for handle missing values.
4. Explain data smoothing techniques with example.
5. What data reduction? Explain data reduction strategies.
6. What is histogram explain with example.
7. Explain types of sampling with example.
8. Differentiate between stratified sampling and cluster sampling.
9. What is association rule mining? Explain with example and formulas for support and confidence.
10. Write main applications of association rule mining.
11. Explain Apriori algorithm with example.