

**Tutorial 2: Data Mining Methodologies**

1. Projects without data mining often rely on:
  - A. Advanced machine learning algorithms
  - B. Manual analysis or simple statistics**
  - C. Big Data pipelines
  - D. Automated predictive models
2. Which of the following is a limitation of projects without data mining?
  - A. Deep pattern recognition
  - B. Accurate real-time predictions
  - C. Slow decision-making**
  - D. Automated data cleaning
3. A small retail store that restocks products by calculating sales averages is an example of:
  - A. Projects with data mining
  - B. Projects without data mining**
  - C. Predictive analytics project
  - D. Market basket analysis
4. What type of patterns are most likely missed in projects without data mining?
  - A. Hidden patterns and correlations**
  - B. Sales averages
  - C. Customer names
  - D. Data entry errors
5. You are working on a project where the business objective is to increase sales revenue. You are about to start the Deployment stage of the CRISP-DM process model. Which task must be completed before the Deployment stage?
  - A. plan deployment
  - B. review project**
  - C. review process
  - D. produce final report
6. A manufacturer has a business goal of reducing product returns due to poor product quality which is known to occur as a result of random mechanical malfunctioning. Which data mining goal is consistent with the business goal?
  - A. Conduct market basket analysis to determine which products to promote.
  - B. Identify which customers are likely to default on their accounts.
  - C. Classify customers into segments for a target marketing campaign.
  - D. Use predictive maintenance to schedule machine repairs prior to failure.**
7. Dividing a database into 3 parts; a training data set, validation data set and testing data set is known as:
  - A. Data Understanding
  - B. Data Partitioning
  - C. Association Analysis
  - D. Predictive Modeling**
8. What is the correct order for the 6 CRISP DM phases?
  - A. Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, Deployment
  - B. Data Understanding, Data Preparation, Business Understanding, Modelling, Evaluation, Deployment
  - C. Data Understanding, Business Understanding, Data Preparation, Modelling, Evaluation, Deployment
  - D. Business Understanding, Data Preparation, Data Understanding, Modelling, Evaluation, Deployment

9. Which of these is NOT part of the CRISP DM Data Understanding phase?
- A. Collecting relevant data.
  - B. Finding and identifying any problems within the data sets.
  - C. Cleaning and addressing any problems with the data sets.
  - D. These are all part of the data understanding phase
10. What is SAS Enterprise Miner used for:
- A. ONLY for Market Basket Analysis
  - B. ONLY for Predictive Modeling
  - C. Creating Accurate Descriptive and Predictive Models**
  - D. None of the above
11. What is Predictive Modelling?
- A. The process of using decision trees to predict certain outcomes.
  - B. Is the process of developing clusters in order to segregate data and discover the relevant categories of data.
  - C. The process of discovering association rules between variables in a dataset.
  - D. None of the above.**
12. You are working on a project where the business objective is to increase sales revenue. With which CRISP-DM process model stage would you start?
- A. Data Preparation
  - B. Data Understanding
  - C. Evaluation**
  - D. Business Understanding
13. You notice missing values and outliers in your dataset. Which CRISP-DM phase should address this issue?
- A. Business Understanding
  - B. Data Understanding
  - C. Data Preparation**
  - D. Modeling
14. A company wants to predict customer churn. Which CRISP-DM phase involves translating this business goal into a data mining problem?
- A. Data Understanding
  - B. Business Understanding**
  - C. Evaluation
  - D. Deployment
15. Which of the following is **not** a characteristic of CRISP-DM?
- A. Flexible and iterative
  - B. Industry-neutral
  - C. Specific to only one type of data mining algorithm**
  - D. Emphasizes both business and data understanding