Q1) (B) The various methods for handling the problem of missing values in data tuples include:

(i) Ignoring the tuple: This is usually done when the value is missing. It is especially poor when the percentage of missing values per attribute varies considerably.

(ii) Manually filling in the missing value: This approach is time consuming & may not be reasonal task for large data sets with many missing values.

(iii) Using a global constant to fill in the missing value: Replacing all missing attribute values by the same constant.

(iv) Using the attribute mean for quantitative (numeric) values or attribute mode for categorical (nominal) values.

(v) Using the most probable value to fill in the missing value.

18DCE115

a) (A) Boys : 66, 66, 67, 67, 68, 68, 68, 68, 68, 69, 69, 69, 70, 71, 72, 72, 72, 73, 73, 74

$$\text{Median} = \frac{69+69}{2} = 69$$

For the left part of 69

Median = 68 = Q1

For the right part of 69

$$\text{Median} = \frac{72+72}{2} = 72 = Q3$$

∴ IQR for boys = 72 − 68 = 4

Girls : 61, 61, 62, 62, 63, 63, 65, 65, 65, 66, 66, 66, 67, 68, 68, 69, 69, 69

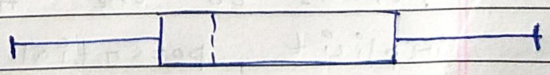
$$\text{Median} = \frac{65+66}{2} = 65.5$$

$Q_1 = 63$

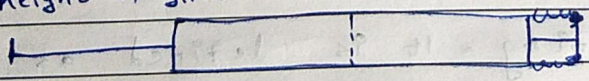
$Q_2 = 68$

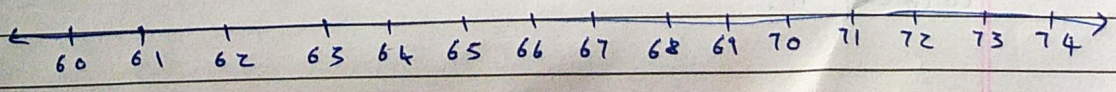
∴ IQR = 68 − 63 = 5

Height of boys



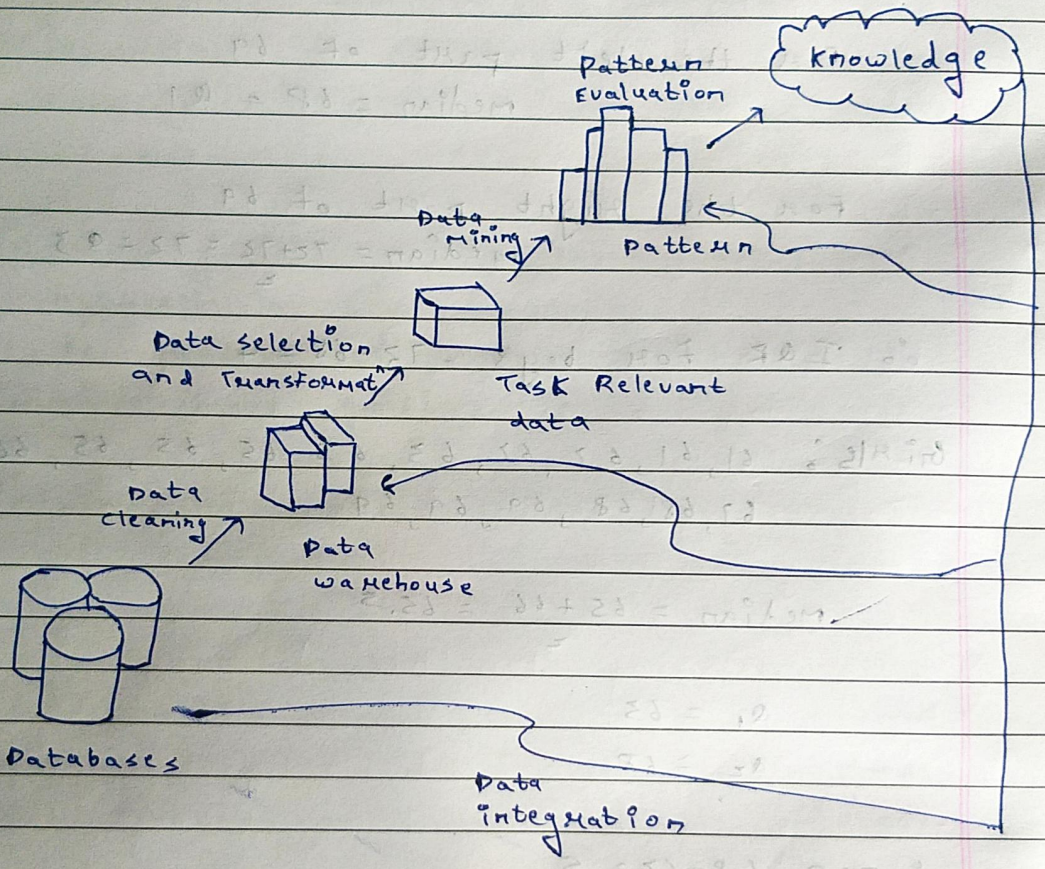
Height of girls





Q3) KDD is acronym for knowledge discovery in databases.

Steps involved in KDD process:



Data mining is also known as KDD which refers to the non-trivial extraction of implicit, potentially useful info. from data stored in databases.

1) Data cleaning - It is defined as removal of noisy & irrelavant data from collection.

2) Data integration : It is defined as heterogenous data from multiple sources combined in common source (data warehouse)

3) Data selection : It is defined as the process where data relevant to the analysis is decided & retrieved from data collection.

4) Data Transformation : It is defined as process of transfering data into appropriate from req. mining process. It has two ~~typ~~ steps : Data mining & code generation

5) Data mining : It is defind as clever techniques to extract potential from potential use.

6) Pattern Evaluation : It is defined as identifying strictly patters representing knowledge base on given measures.

7) Knowledge representation : It is techniques which uses tools to present data mining resources, generates reports, table, etc.

04) Issues in data Mining :

1) Mining Methodology & user interaction.
- Handling noise & incomple data
- Incorporation of background knowledge
- Mining different kinds of knowledge in databases.
- Pattern evaluation
- Expression & visualisation of data mining results.

2) Performance & scalability
- Efficiency & scalability of data Mining algorithms
- Parallel, distributed and incremental Mining methods

3) issues relating to the diversity of data types :
- Handling relational & complex types of data
- Mining information from heterogeneous databases.

4) Issues related to application & social impacts
- Application of discovered knowledge.
- protection of data security, integrity & privacy
5) - integration of discovered knowledge & existing knowledge.