① Data cleaning Methods / techniques.

- Remove duplicates
- Handle missing data
- manual data cleaning
- use data cleansing tools.
- convert data types
- Remove unnecessary values
- Use a clear format.
- Remove unwanted outliers.

Data cleaning tools.

- OpenRefine
- Microsoft Excel.
- WinPure.

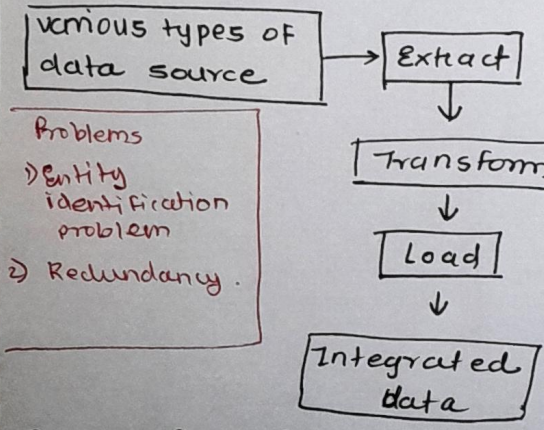
Advantages.

- more accurate model.
- improved decision-making.
- Better data quality.

② Data Transformation.

- Data smoothing
- Attribute construction.
- Normalization ~~attribute raw~~
- Data Aggregation
- Data Discretization.
- Concept Hierarchy generation for nominal data.

③ Data Integration.



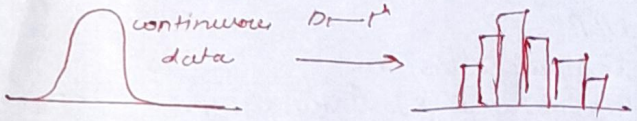
Process of combining, sharing, or synchronizing data from multiple sources to provide users with a unified view.

④ Data Discretization.

- also known binning.
- process of converting continuous data into discrete categories.
- used in datascience & datamining.
- the range of continuous variable gets divided into intervals.
- each data bin gets data points.

Uses

1) Data reduction 2) Computational efficiency
3) ML 4) Data mining.
5) Data analysis.



Discretization →
1) supervised → uses class information
2) unsupervised → does not use class information.

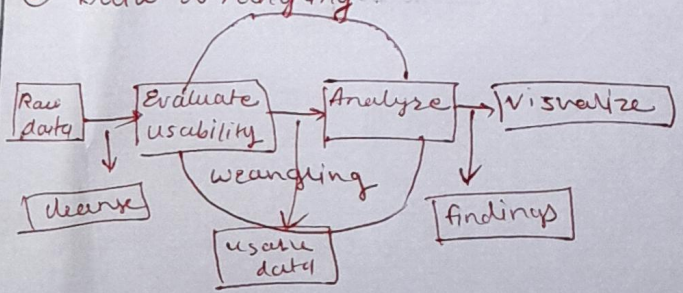
Top down → also called splitting.

Bottom up → also called merging.

⑤ Data Reduction.

① Dimensionality Reduction → Step wise forward selection, Step wise backward selection
② Numerosity Reduction
③ Data cube aggregation
④ Data compression → lossy, lossless.
⑤ Discretization.

⑥ Data Wrangling.



Methods

① Discovering ② structuring ③ Cleaning
④ enriching ⑤ validating ⑥ Publishing.

- the process of cleaning, organizing and transforming raw data into the desired format for analysis and better decision making.

(7) Data science use in AI.

① Data collection & Preprocessing
② Machine learning & Deep learning Models
③ Predictive Analysis
④ NLP
⑤ computer vision
⑥ Big Data & AI Integration
⑦ Model training
⑧ AI business Intelligence & Decision making
⑨ Statistical Analysis.

(8) Datascience Application

① In Search Engine.
② In Transport (Driverless cars)
③ In Finance (Stock market.)
④ E-commerce
⑤ In Health Care
⑥ Image Recognition

(9) Data Science life cycle

① Business understanding
② Data exploration
③ Data visualization
④ Predictive Modeling
⑤ Data cleaning.
⑥ Feature & Engineering
⑦ Data mining.