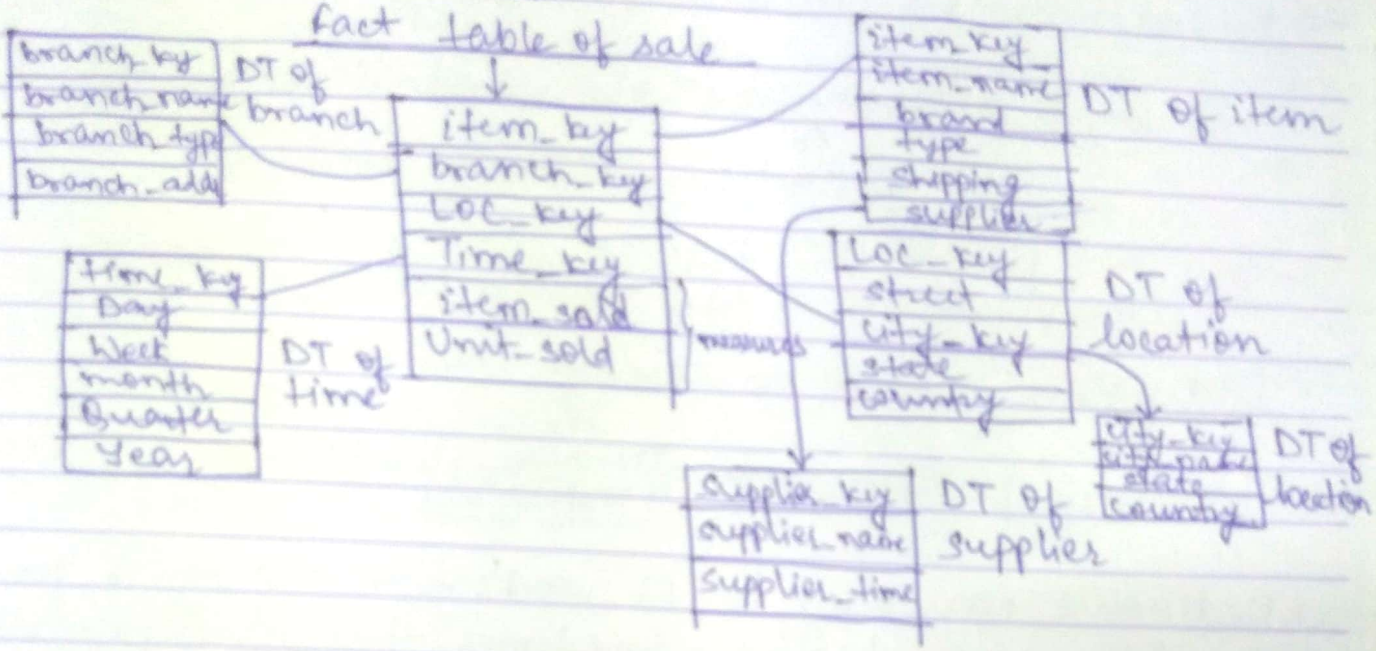
**

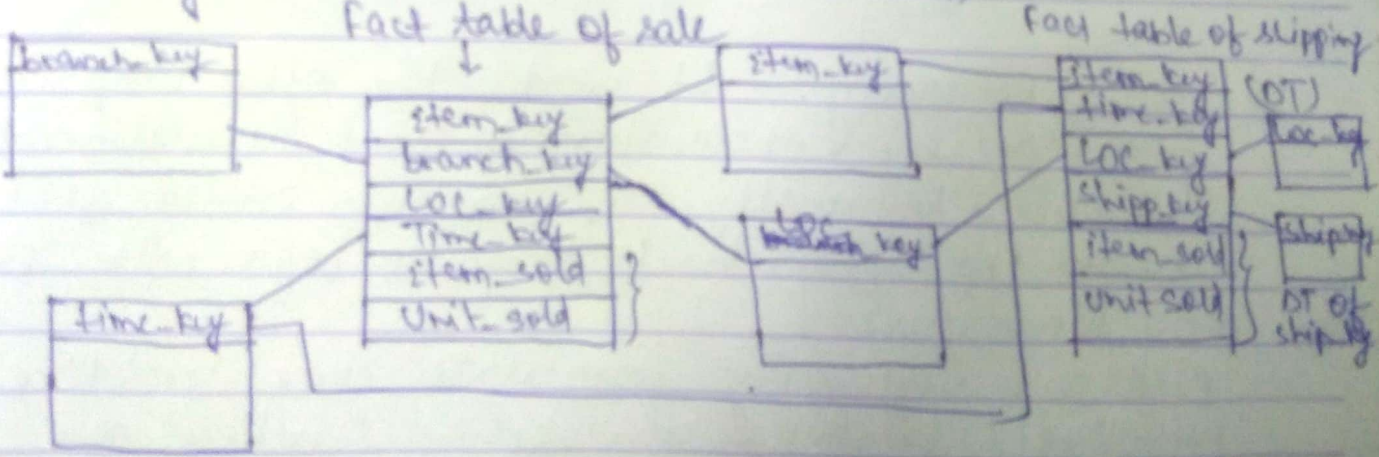
(2) Snowflake Schema :-



→ It is a variant of the star schema, where some dimensional table is normalised. Thereby giving the data input of additional table.

→ In this, the DTs in normalised form.
(simplified form) → remove corrupt data

(3) Galaxy schema :- (or constellation)



→ There are multiple fact tables that share information (Dim. Tables).

* Inductive and Deductive Reasoning :-

Theory → Hypothesis → Observation → Confirmation

Deduction - theory to fact

eg :- All plants are green.
All planets move around the sun. } theory

Induction - fact to theory

Observation → Pattern → Iterative Hypothesis → theory

→ Deductive Reasoning :- It works from the most general to more specific. Sometimes, this is informally called a top-down approach. You might begin with thinking up it a theory of a ~~it~~ topic of interest ~~wit~~ with a narrowdown even further when we collect observations ~~that we can test~~ to address the hypothesis. This ultimately leads us to be ~~ed~~ able to test the hypothesis with specific data, i.e, a confirmation of our original theories.

→ Inductive Reasoning :- It works the other way moving from specific observations to broader visualizations & theories. Informally, we sometimes called this a bottom-~~top~~ approach. In this, we begin with specific obs. & measures — pattern & regularity, formulate some tentative hypothesis that we can explore further and end up finally developing some general conclusion of theory.

eg :- Earth moves around the sun.
Induction
All planets move around the sun.

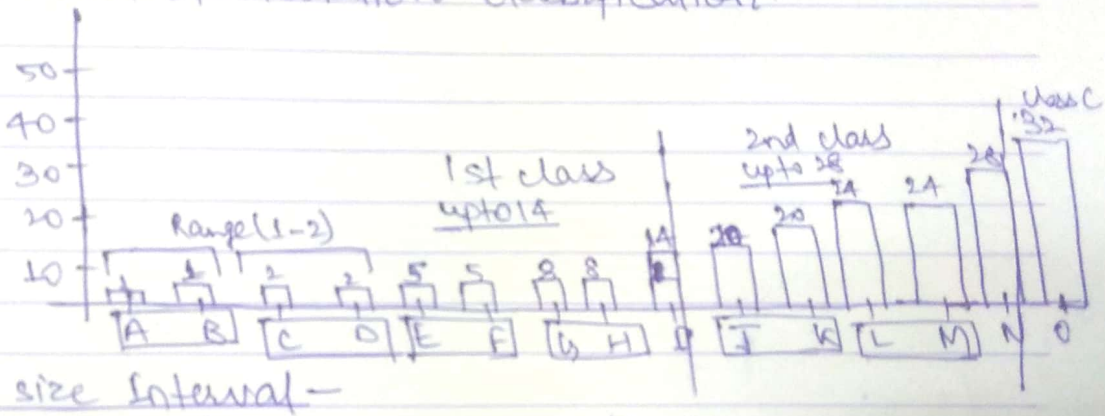
→ Research papers are examples of both.

⇒ Inductive reasoning is wide in nature, is more open-ended and exploratory. Deduction reasoning is more narrow in nature and deal with testing or confirming hypothesis.

* Data classification Method :-

→ Numeric data

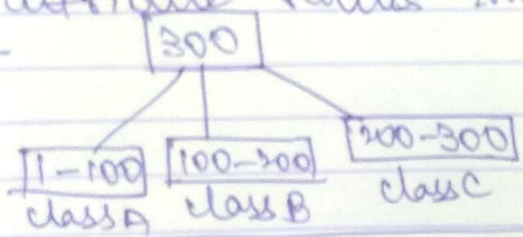
1. Equal size interval — 300 → 1-100, 100-200, 200-300 } 3 classes of equal interval
2. Quantile classification
3. Mean standard deviation classification.



1. Equal size Interval -

This classification scheme divide the range of attribute values into equal size subranges.

eg :-



⇒ No. of observation = total observation / no. of class

2. Quantile classification :-

Each class contain an equal number of features. In this method, we have to pre-define how many classes we want to use, then we rank and order our data classes by placing an equal no. of ~~stats~~ observations in each class. This method classifies data into a certain no. of category equal in size.

3. Mean standard deviation :-

eg :- 96, 104, 126, 134, 140

Mean = 120

Mean deviation (σ) = $\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - x_0)^2}$

$= \sqrt{(96-120)^2 + (104-120)^2 + 6^2 + 14^2 + 20^2/5}$

$= \sqrt{576 + 256 + 36 + 196 + 400/5}$

$= 17.11$