U-3 (BI)

* Concept of BI –

- In BI, process of collecting & analyzing data with help of diff modern BI tools.
- capable to provide interactive data visualisations
- helps business orgs to extract required data & optimize org for continuous growth.

* Benefits of Reporting

- Improve workflow
- Uses real time & historical data
- operation optimization
- cost optimization
- Improve procurement process
- performance monitoring of employee
- customer analysis & prediction
- Resource management
- Business forecasting

* Building Reports with Relⁿ & Multidimen data Models.

- Relⁿ Data Model –
  - widely used in all orgs + s/w as a backend
  - work efficiends for data storage processing
  - primary & simplest model having all capabilities → process & store data.
  - works on ACID → atomicity, consistency, isolation & durability
  - Following termonlogies used to represent RDM –

1) Tables → represent rel^n in RDM (Rows & Column)
   ↳ Records → rows & attributes → columns

2) Tuple → • single row of a table ie single record of a rel^n

3) Rel^n instance → • finite set of tuples present in rel^n as: particular instance • don't have duplicates tuples.

4) Relation schema – describe relation name i.e. table name along with its attribute with name.

5) Relation key – identify row in table (rel^n) • consist of one or more attributes used to identify rows uniquely.

6) Attribute domain – every attribute has predefined value scope.

* Multidimensional Data Model –

- Rel^n DM is a 2D dataset ie columns of row tools.
- If we have more info about same row wrt another dimension then we need to add more dim in relation → becomes 3D data. similarly, we can go adding no. of dimensions → multidim DM
- Data is powerful → valuable asset for any org.
- complexity ↑ difficult to extract data ↑ in order to find actionable insights
- multidim data model → reps data in cube form (Data Cube).
- data cube enables data to be modeled & viewed in multiple dim.
  ↳ dimensions & facts
- dimensions → views/entities in an org stored as records.

* Types of Report -

1) List -

- basic forms of report.
- base of reporting about data & its insights.
- by extending basic list next level analysis of data is done & represent in form of graph, map, table, etc.
- eg → student dataset → student academic performance.
  To report this data → list all student as per exam score time distinction, 1st class, 2nd class, pass & fail.

2) Cross tabs -

- cross tabs = extended version of simple table.
- represent single categorical variable, tables or frequency tables to rep us.
- describe rel^n b/w 2 categorical variable → cross tabs used.
- cross tab → categories of one variable → rows of table.
- cross tab → 2nd " " → columns of table.
- cells of table contain no. of times that particular comb^n occured.
- edges / boundaries → summarized / grouped observation of catgories.
- statistical tool for categorical data.
- categorical data = values that are mutually excluse to each other.
- table → detail data in grid structure.
  cross tab → grouped data in grid structure.

3) Statistices -

1) Descriptive statictiss -

- main objective - demonstrate huge portion of collected data.
- brief summary of gathered data using descriptive charts.
- illustrates univariate measures of tendency.
- measures of dispersion like variance & standard deviation.

2) **Inferential statistics -**

- Main objective - provide more detailed & effective statistics data analy
- involved making broader & deper deduction & interpretations usually on interaction b/w variables cause & effect rel^n & scope of
- In data analysis are analysis of variance (ANOVA), T-test, Sample T- test, z-test, linear regression & multiple regression.

3) **Psychometric Tests -**

- Analyse attributes & performance of employed survey to ensure survey data to reliable & valid.
- eg → cronbaci's Alpha.

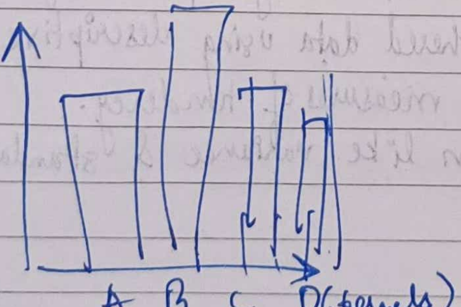
Types of statistical reporting data -

1) Categorical data → result of relative freq. statistics.
2) Ordinal data → Best represented using freq. tables.
   → Data have scales & ordered acc to preference.
3) Interval data → Averaging & standard deviation type of data.
4) Radio data → converted to normal data using alggarithms square roots.

4) **Chart -**

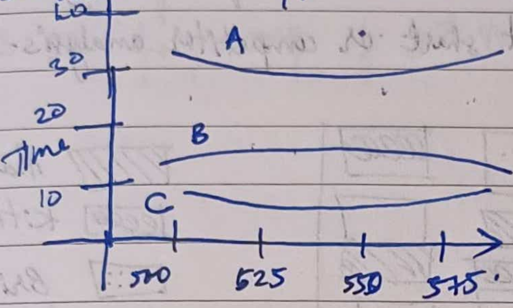
1) Bar chart:

- compare data b/w diff groups & help to track changes in data overtime
- Barecharts most useful when these are big cranees ie to show now one group compares to another group.
- Diff types of bar chart - vertical, horizontal, stacked, grouped, bar chart.



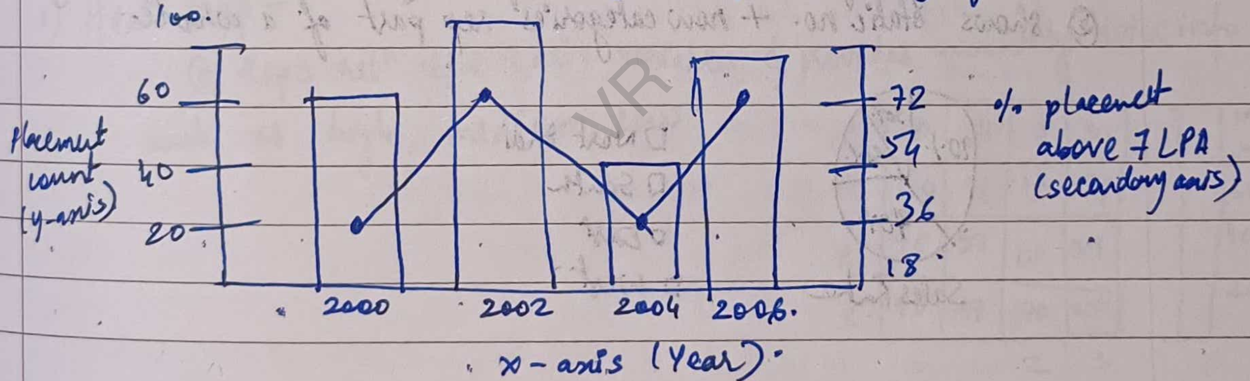
2) Line chart -

- Represent trends or progress of respective variables overtime.
- suitable when i/p data is continous.



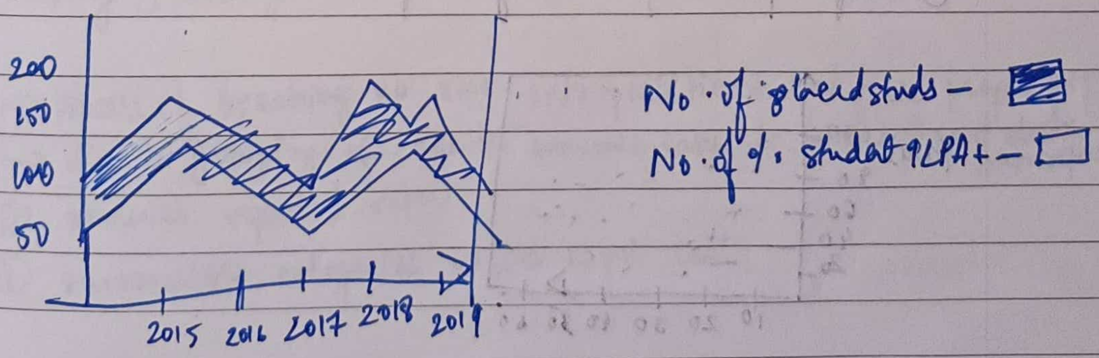
3) Dual axis chart -

- plotted using one x-axis & 2y-axis.
- 3data variables → 1 → continous set of data.
  → 2 → grouping by category.



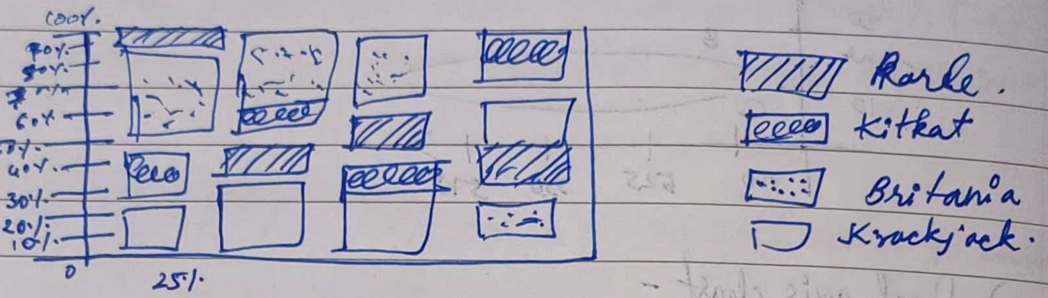
4) Area charts -

- actually line chart but fills up space b/w x-axis x-axis & lineplot.
- Helps analyse both individual & overall contribution against total contrib



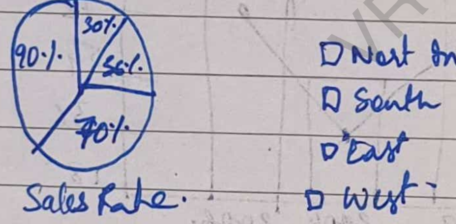
5) Mekko Chart -

- also known as marimekko chart.
- used to compare measures, values, values, quantities & shown data distribution
- used to show growth, market share or competitor analysis.



6) Pie Chart -

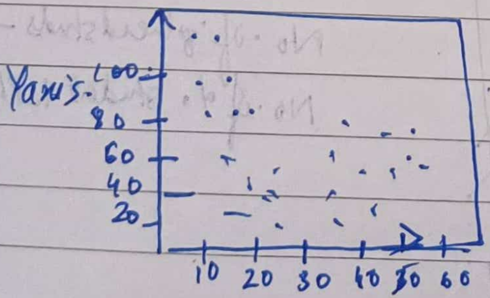
① Percentage of data distribution of any variable among categories.
② shows static no. + how categories rep part of a whole.



Sales Rate.

7) Scatter plot chart -

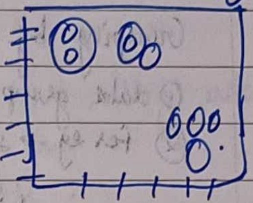
① shows rel^n ~~distribution pattern~~ b/w 2 variables.
② helps reveal data distribution pattern.
③ useful for to find insights from data like outliers, pattern & similarities.



4) Bubbled Chart –

① similar to scatter plot → reps data distribution among 2 var^n.

② additionaly in bubble chart 3rd data variable shows size. as per frequency of 3rd variance.

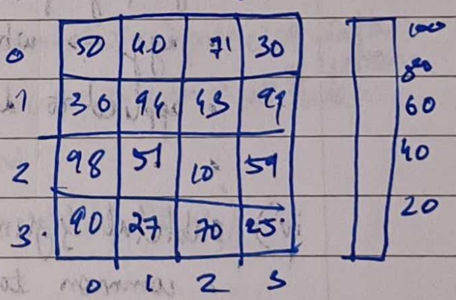


5) Map –

- Identify insights & make proper decisions.
- Heat map, point map, flow map, statistical map.
- Maps can be further divided into 2D, 3D, static, dynamic.
- often used in combination w/time, pt, bubble & dim.

1) Heat Map –

① Reps rel^n b/w 2 data variables & provides quantity wise info such as high, medium, low.



	0	1	2	3
0	50	40	71	30
1	30	94	43	99
2	98	51	10	59
3	90	27	70	25

* Operations on Data Reports –

DATA Grouping & Sorting –

① Grouping → process of breaking up rows into partitions that share properties.
→ due to grouping of doc it becomes easy to read relevant data from report.

② removes repeated report

③ summarises categories varies as per call.

② Grouping based on selected category & conditional parameters.

③ consolidated data at a higher level of granuality.

Grouping by single attributed -

① data grouped based on single attributes as column.

② for eg → sales report → group data by product category.

- It summaries sales info each category/ product separetly
- grouping operation will consolidate data & provide aggregated values as Total sales, revenue, avg quantity sold.

ii) Grouping by mutiple attributes:

- data in 1st grouped by primary attribute.
- within each group then it is further subdivided based on secondary attribute

iii) Grouping with aggregation -

- grouping often involves aggregation ie calc summary value for each group
- agg → when sum, aggreg ate, average, count, max, min.
- applied to decive meaningful insights.
- .

iv) subtotal & grand total -

- common to include subtotal & grand total of row/ cols to provide an overview of grouped data.
- subtotal → aggregate values will each group.
- grand total → overall avg value each group & entire dataset.

→ Data G widely supported various data analysis & reporting tools

→ allows users to summarice.

- condense large amt of data into manageable.
- enabling better analysis & decision meaning

sorting -

Sorting -

→ essential operation when working of reports.
→ arrange in order based on one or more attribute or column.
→ ascending, descending & sorting by multiple Attr.
- primary attribute → 1st sorting criteria.
- secondary → & so on.

→ sorting operations available in s/w reporting tools.

FILTERING REPORTS -

- allows one to selectively display or exclude data based on specified criteria.
- Helps focus on specific subsets of data meet conditions.

i) single condition filtering
ii) multi condn → • allows for more complex + precise data selection.
iii) Range based → range of values for a particular column or attribute.
iv) Text based → Filter text value on patterns.
v) Interactive filtering (dynamic apply or modify filters)

Adding Calculations to Reports -

SUM	sum (select cells / range).
AVG	Avg (select / cols / range).
COUNT	Count (select cells / cols / range).
MAX	Max (———)
MIN	MIN (———)
Standard deviation	STD (———
variance	VAR (———).

VR

* Conditional Formatting -

- feature in spreadsheets s/w & data visualization tools tools that allows you to apply formatting to cells.
- visually highlights or emphasize certain certain data based on prefined criteria.
- CF enables us to indentify pattens, trends, outliers, expections in their data.

* Overview of CF -

- Formatting Rules → CF rules → cond^n + cors. formatting instructions
  → evalutes data cell & defines how cell is formatted
- Types of conditions → numerical, text based, date based, formula based cond^n
- Formatting option → font color, bg color, font style.
- Rules priorisation → mutiple rules → same range of data
  ↳ diff rules → conflict.
- Interactive & dynamic formating → reflected in real time & formatting adjusts

* DRILL DOWN -

① moving higher level summary to detained level of data
② users → delve deeper into data → expanding or revealing additional.
③ eg → sales report initally displays to sales revenue by year.
④ access next level of detail → ie quaterly sales revenue.
⑤ each drill down → provide more granular r data uncovering more specific insights.

DRILL UP-

* · moves from lower level detail to higher level sumary
· summarize & aggregate data · -> higher level of hiearachy.
· exe -> derived down to daily sales rev, drill up to monthly sales rev.
· Drill up allows users to rollup data & collapsing details & focusing on higher level summaries.

* PDF (Portable Doc Format)-

· finalised reports meant for suitable cases used in distributing to others.
· presentations, exec summaries, documentation.
· reports that have consistent layout or formatting across diff devices.
· focus on visual presentation rather than data manipulation.

* CSV (comma seperated values):

· stores tabular data data in plain text format, with values seperated by commas
· lightweight -> can be used for data ~~insurance~~ in charge b/w diff sys.

* XLS (excel spreadsheets).

· XLS may contain mutiple worksheets, formulae, charts, macros-
· suitable for complex data analysis & reporting scenarios.

* XLM. (Extensible Markup Language)-
· uses tags to define structure of data.
· often used to exchange data between diff application because they are are machine-readable & easily customized-
· complex to understand for people who arenot familiar with format.