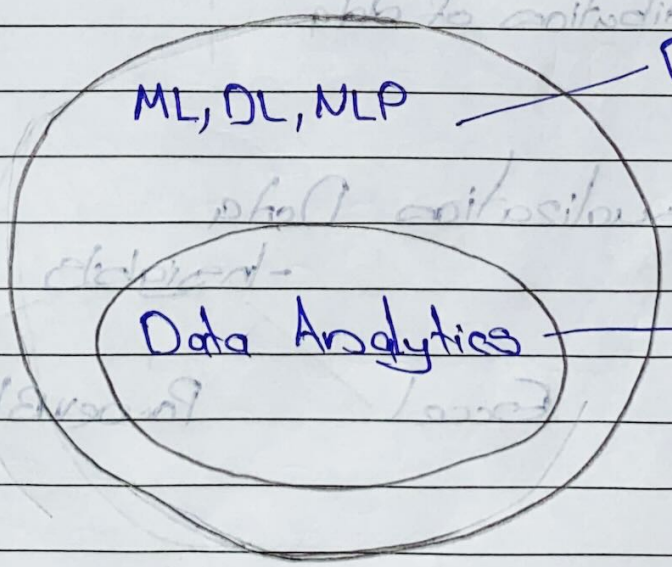
Data Analytics Project Life Cycle

⊙ Data Analytics vs Data Science vs Data Engineer

- Data Analytics
  → Insights → Generate Revenue
  (SQL, stats, BI tools: Excel, Power BI, Python)

- Data Science
  → Prediction of new set of data points
  (ML, DL, NLP)

- Data Engineering
  → Big Data → Data pipeline



⊙ Parts:

① Identify the Problem
② Data Requirement / Collection
3. Data Preparation → SQL
4. Exploring the Data
   - Outliers (Extreme data)
   - Null / Missing

Statistics

Descriptive stats	Inferential Stats
• Description of the data	• Conclusion → Hypothesis testing (Null & alternative hypothesis)
• Measure of Dispersion • Measure of Central tendency • Central limit theorem • Distribution of data	

EDA = Exploratory Data Analysis

5. Visualization Data
   - Insights

Excel — PowerBI — Tableau

- SQL
- Statistics — Descriptive, Inferential
- Python framework
  - Pandas
  - Numpy
  - Matplotlib
  - Seaborn
- Tools — Excel, Power BI

Excel

Data types

1. Textual (left aligned)
2. Numeric (Right aligned)
3. Boolean (True / False)
4. Date

Cell Referencing

Cell Referencing
- Absolute Referencing
  - Fixed ($ Prior to column & row)
- Relative Referencing
  - changed (Default)

⊙ Logical Operators (Function)

- AND (both should be satisfied)
- OR (Any one of them should be satisfied)
- NOT NOT 1 → 0

⊙ Function

= CONCATENATE	= COUNT (All only Num type)
= UPPER	= COUNTA (All except blank value)
= SUM	= COUNTBLANK (All BLANK)
= SUMIF	= COUNTIF (All records according to the condition will written inside the "")
= AVERAGE	
= MEDIAN	
= MIN	
= MAX	
= LEN	
= EXACT (	

⊙ If (B2<C2, "Profitable", "failure")

where it is sort of If else statement in the Excel