

# DATA ANALYTICS

जब हमारे पास २ या ३ लाइन डाटा होता है , तब मनुष्य पूरा डाटा याद ही रख लेता है, पर अगर डाटा उससे अधिक है तो मनुष्य डाटा नहीं याद रख सकता और इसलिए वो केवल insights (सूत्र) याद रख पाता है।

डाटा एनालिसिस वो टेक्निक्स हैं जिससे हम इकट्ठा किये हुए डाटा के बारे में इनसाइट्स निकलने में प्रयोग करते हैं।

# DATA STORAGE

डाटा को प्रायः हम कंप्यूटर फाइल [जैसे की txt, csv या xlsx फाइल ] या डेटाबेस सिस्टम [जैसे MySQL , PostGres या Mongo ] में सेव करते हैं। आप जो भी सॉफ्टवेयर उसे करते हैं जैसे Tally , Facebook , Whatsapp या TikTok , ये सभी डाटा को डेटाबेस सिस्टम में रखते हैं। जैसे की Facebook Mysql का प्रयोग करता है। जो डाटा हम अक्सर देखना या चेक करना चाहते हैं , उसे एक्सेल के माध्यम से xlsx फाइलो में सेव करते हैं।

हम इन टुटोरिअल्स में यह मान कर चलेंगे की डाटा पहले से ही एक्सेल फाइल में या डेटाबेस में रखा हुआ है और अब हमे उसपर कुछ इनसाइट्स निकलने को कहा गया है।

Roll Number	Name	Marks(/100)
1	ABC	57
2	DEF	88
3	GHI	76
4	JKL	93
5	MNO	64
6	PQR	78
7	STU	95
8	VWX	74
9	YZA	91
10	BCD	75

Roll Number	Paper Section	Marks
1	1	27
1	2	30
2	1	44
2	2	44
3	1	26
3	2	50
4	1	44
4	2	49
5	1	30
5	2	34
6	1	39
6	2	39
7	1	49
7	2	46
8	1	38
8	2	36
9	1	46
9	2	45
10	1	37
10	2	38

Roll Number	Taught By
1	1
1	2
2	2
3	2
4	1
5	2
5	1
6	1
7	1
7	2
8	1
9	2
10	2

# Data Analysis Tools

भिन्न भिन्न प्रकार के टूल्स हैं जिनका प्रयोग Data Analytics में किया जा सकता है :

1. Python / Pandas : पाइथन लैंग्वेज की pandas लाइब्रेरी का हम इन टुटोरिअल्स में मुख्यतः प्रयोग करेंगे।

2. R : पाइथन के अलावा R भी एक बहुत प्रसिद्ध लैंग्वेज है जो डाटा एनालिटिक्स के लिए उपयोगी है। कई लोग पाइथन का प्रयोग करते हैं बाकी लोग R का। अगर आपको दोनों में से एक आती है तो आपका सारा काम हो जाएगा। दोनों एक साथ सीखने की ज़रूरत नहीं है।

3. SQL : SQL किसी भी डेटाबेस सिस्टम में स्टोर किये हुए डाटा को मैनेज और एनालाइज करने के काम आती है। आप database में सेव किये हुए डाटा पर एनालिटिक्स SQL से भी कर सकते हैं। इन टुटोरिअल्स में हम SQL का प्रयोग देखेंगे पर सीधा फाइल में सेव किये हुए डाटा पर, डेटाबेस सिस्टम पर नहीं। यहाँ सीखी हुई SQL आप किसी भी डेटाबेस सिस्टम या बिग डाटा सिस्टम पर चला पाएंगे।

4. अन्य : कई लोग डाटा एनालाइज करने के लिए माइक्रोसॉफ्ट Excel , SPSS , Matlab , Julia लैंग्वेज जैसे अन्य टूल भी प्रयोग करते हैं। हम इनके बारे में इस टुटोरिअल में नहीं सीखेंगे।

# Tabular Data

डाटा को ज़्यादातर tables के फॉर्म में रखा जाता है। 1 डेटाबेस में कई टेबल होते हैं। आज हम जो डाटा देख रहे हैं उसे हमने 1 excel फाइल के 3 tabs में 3 टेबल रखे हैं : StudentOverallData , StudentSectionWiseMarks और StudentTeacher . टेबल में 1 row में 1 entity होती है और 1 attribute का डाटा 1 कॉलम में होता है. 1 entity क्या होनी चाहिए हो यह डेटाबेस प्लान करने वाले के ऊपर निर्भर होता है।

ध्यान दें की कुछ डाटा टेबल के रूप में स्टोर करना मुश्किल होता है बूत इस टुटोरियल में हम ऐसे डाटा पर काम नहीं करेंगे ।

## Table

Roll Number	Name	Marks(/100)
1	ABC	57
2	DEF	88
3	GHI	76
4	JKL	93
5	MNO	64
6	PQR	78
7	STU	95
8	VWX	74
9	YZA	91
10	BCD	75

# Data Analytics Operations

डाटा एनालिटिक्स में luckily कुछ तरह के operations ही बार बार प्रयोग करके insights निकाले जाते हैं। आपको जब कोई डाटा मिला तो आपको इन्ही ऑपरेशन्स को अलग अलग तरह से जोड़कर उससे काम insights निकालनी होती हैं।

6 ऐसे डाटा एनालिटिक्स ऑपरेशन्स हैं :

1. Select : जहाँ पर आप कुछ given criteria के हिसाब से उपयुक्त डाटा के कुछ हिस्से को बाहर निकालें।
2. Decode : जहाँ आप एक या एक से ज़्यादा data के attributes को उपयोग करके नया attribute बना दें।
3. LOOKUP या JOIN : जहाँ आप दो अलग जगह store किये हुए डाटा को किसी common attributes की मदद से जोड़ दें।
4. GROUP : जब आप किसी attribute का प्रयोग करके डाटा के कई टुकड़े बना लें।
5. AGGREGATE : जब आप किसी गणित की फंक्शन जैसे SUM , VARIANCE , AVERAGE का प्रयोग करके डाटा की कई rows की जगह 1 sankhya स्टोर कर लें [वस्तुतः हम डाटा में जोड़ तोड़ करने के बाद ऐसे इसी नंबर को insight की तरह सेव करते हैं ]
6. GRAPHING : कभी कभी सिर्फ १ नंबर सेव करने के बजाय हम ग्राफ बनाने के माध्यम से ट्रेंड देखना चाहते हैं।

# SELECT

Roll Number	Name	Marks(/100)
1	ABC	57
2	DEF	88
3	GHI	76
4	JKL	93
5	MNO	64
6	PQR	78
7	STU	95
8	VWX	74
9	YZA	91
10	BCD	75

## SELECT

Example: selecting students with Marks more than 90.



Roll No	Name	Marks(/100)
4		93
7		95
9		91

# DECODE

## Decoding

Example : Adding teacher name  
to each row where Teacher = 1  
[KLM], Teacher = 2 [OPQ]

Roll Number	Taught By
1	1
1	2
2	2
3	2
4	1
5	2
5	1
6	1
7	1
7	2
8	1
9	2
10	2



Roll Number	Taught By	Teacher Name
1	1	KLM
1	2	OPQ
2	2	OPQ
3	2	OPQ
4	1	KLM
5	2	OPQ
5	1	KLM
6	1	KLM
7	1	KLM
7	2	OPQ
8	1	KLM
9	2	OPQ
10	2	OPQ

# LOOKUP Or JOIN

## LOOKUP / JOIN

Example - Combining data with  
Student's name and marks with  
data containing their marks in  
each section of the paper. Roll  
Number is the key used to join.

Roll Number	Name	Marks/(100)
1	ABC	57
2	DEF	88
3	GHI	76
4	JKL	93
5	MNO	64
6	PQR	78
7	STU	95
8	VWX	74
9	YZA	91
10	BCD	75

Roll Number	Paper Section	Marks
1	1	27
1	2	30
2	1	44
2	2	44
3	1	26
3	2	50
4	1	44
4	2	49
5	1	30
5	2	34
6	1	39
6	2	39
7	1	49
7	2	46
8	1	38
8	2	36
9	1	46
9	2	45
10	1	37
10	2	38



Roll Number	Name	Marks/(100)	Paper Section	Marks
1	ABC	57	1	27
1	ABC	57	2	30
2	DEF	88	1	44
2	DEF	88	2	44
3	GHI	76	1	26
3	GHI	76	2	50
4	JKL	93	1	44
4	JKL	93	2	49
5	MNO	64	1	30
5	MNO	64	2	34
6	PQR	78	1	39
6	PQR	78	2	39
7	STU	95	1	49
7	STU	95	2	46
8	VWX	74	1	38
8	VWX	74	2	36
9	YZA	91	1	46
9	YZA	91	2	45
10	BCD	75	1	37
10	BCD	75	2	38



# GROUP

## Group

Splitting the data into many parts based on one or more datapoints.  
Example : Separate marks according to paper section.

Roll Number	Paper Section	Marks
1	1	27
1	2	30
2	1	44
2	2	44
3	1	26
3	2	50
4	1	44
4	2	49
5	1	30
5	2	34
6	1	39
6	2	39
7	1	49
7	2	46
8	1	38
8	2	36
9	1	46
9	2	45
10	1	37
10	2	38

GROUP BY Paper Section

Roll Number	Paper Section	Marks
1	1	27
2	1	44
3	1	26
4	1	44
5	1	30
6	1	39
7	1	49
8	1	38
9	1	46
10	1	37

Roll Number	Paper Section	Marks
1	2	30
2	2	44
3	2	50
4	2	49
5	2	34
6	2	39
7	2	46
8	2	36
9	2	45
10	2	38

# AGGREGATE

## Aggregate

Perform an operation on one attribute of data converting it into a single number. This can be Average, Variance, Median, Maximum, Minimum, SUM, PRODUCT etc.

For example, here we find the average marks in exam.

Roll Number	Name	Marks(100)
1	ABC	57
2	DEF	88
3	GHI	76
4	JKL	93
5	MNO	64
6	PQR	78
7	STU	95
8	VWX	74
9	YZA	91
10	BCD	75

AGGREGATE (AVERAGE) ON Marks/100

39.55