

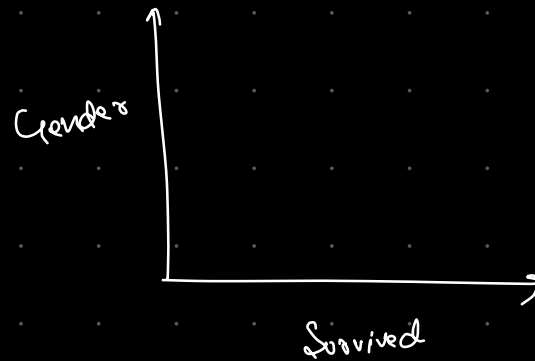
Class 5 \Rightarrow Data Science Library

- \Rightarrow 30 min matplotlib, + stats (MM, central tendencies)
- \Rightarrow Hypothesis testing
- \Rightarrow Sampling / S no. summary / Distribution
- \Rightarrow Inference ~~statistics~~
- \Rightarrow missing value + outliers.
- \Rightarrow Maths on ML

Seaborn

titanic

titanic['Gender']
titanic['Survived']



⇒ Python statistical Graphic library

⇒ Built on top of matplotlib

It incorporates matplotlib
with pandas

⇒ Its plotting function work on dataframe
& arrays creating entire datasets.

Data Science

data \Rightarrow Base of analytics

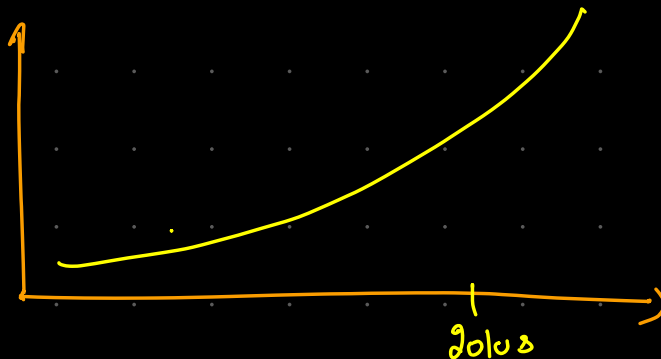
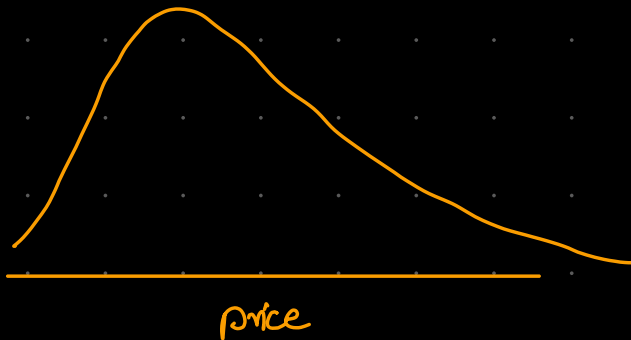


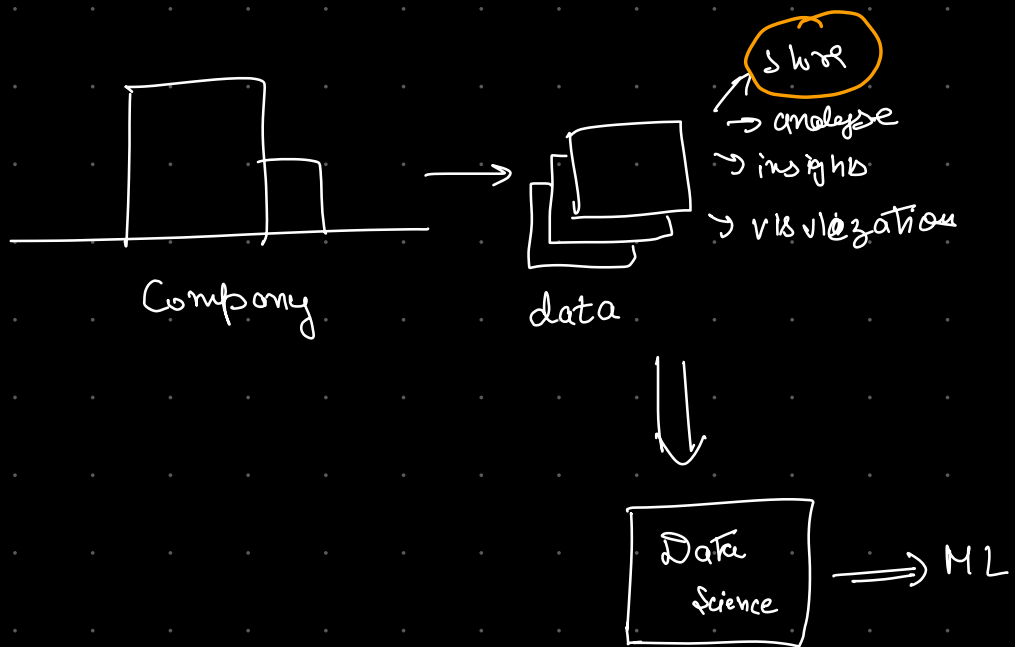
Science

\Downarrow
maths

1. Stats
2. Prob
3. Calculus

\Rightarrow It's a concept / technique
used to analyze & derive
insights from our data





Tools to work on data

- python
- sql
- tableau

⇒ Statistics

⇒ Predictive analysis

⇒ Advance Predictive Analysis

Statistics

Statistics is the science of conducting studies to

Collect

organize

summarize

analyze

and draw conclusion
out of your data.

Hospitality

Tech

business

e-commerce

agriculture

Two areas of Statistics

Descriptive Inferential

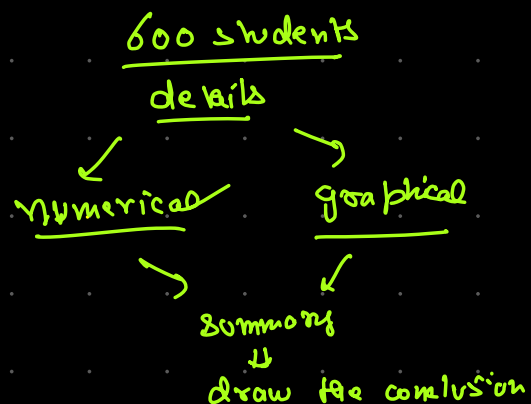
⇒ Descriptive Statistics

1. helps to understand raw data & drive the data based on our business need.
2. Uses statistical & graphical approaches to look for pattern in our dataset & summarize the data, to present information

Numerical ⇒ mean, median, mode, s.d., var

Graphical ⇒ hist plot, boxplot, graphs.

GNI PER CAPITA ANALYSIS USING WORLD BANK 2011 DATA	
Mean	11616.7
Standard Error	1361.2
Median	4200
Mode	430
Standard Deviation	17591.1
	3094480
Sample Variance	35.5
Kurtosis	4.99
Skewness	2.27
Range	88680
Minimum	190
Maximum	88870
Sum	1939990
Count	167

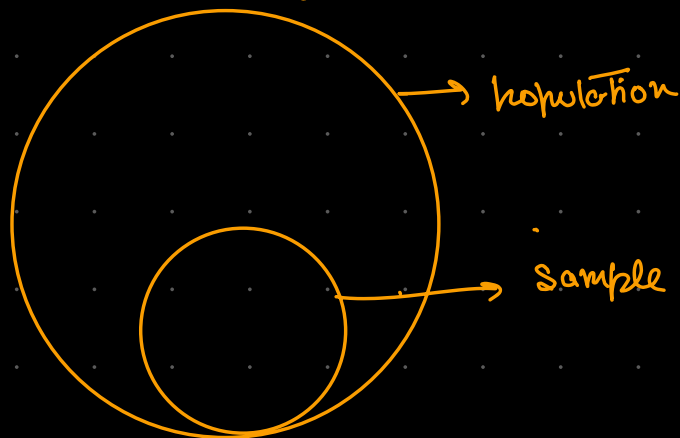


Inferential statistics

Inference

⇓
to infer

⇒ use sample data to make estimates, decision or prediction or generalization about larger set of data.



exit polls

POLL OF EXIT POLLS					
	SEATS	MAJORITY	PROJECTED RESULTS		
GUJARAT	182	92	BJP 131	Cong + NCP 41	AAP 7
HIMACHAL PRADESH	68	35	BJP 36	Cong 29	AAP 0
DELHI MCD	250	126	AAP 155	BJP 84	Cong 7

Source: Regional and National Channels

Health Warning: Aggregate of exit polls, these polls are not always accurate. Delhi MCD wards 'matched' to ensure a degree of comparability between 2017& 2022 elections

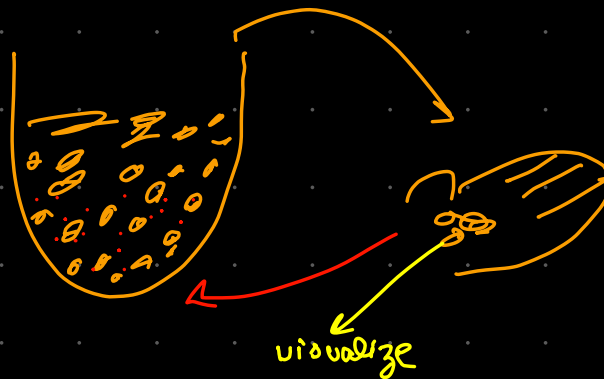
#PollofExitPolls

NDTV

With inferential stats, we are trying to reach conclusion beyond our immediate data

sample data → population conclusion.

Grain



Blood Test

