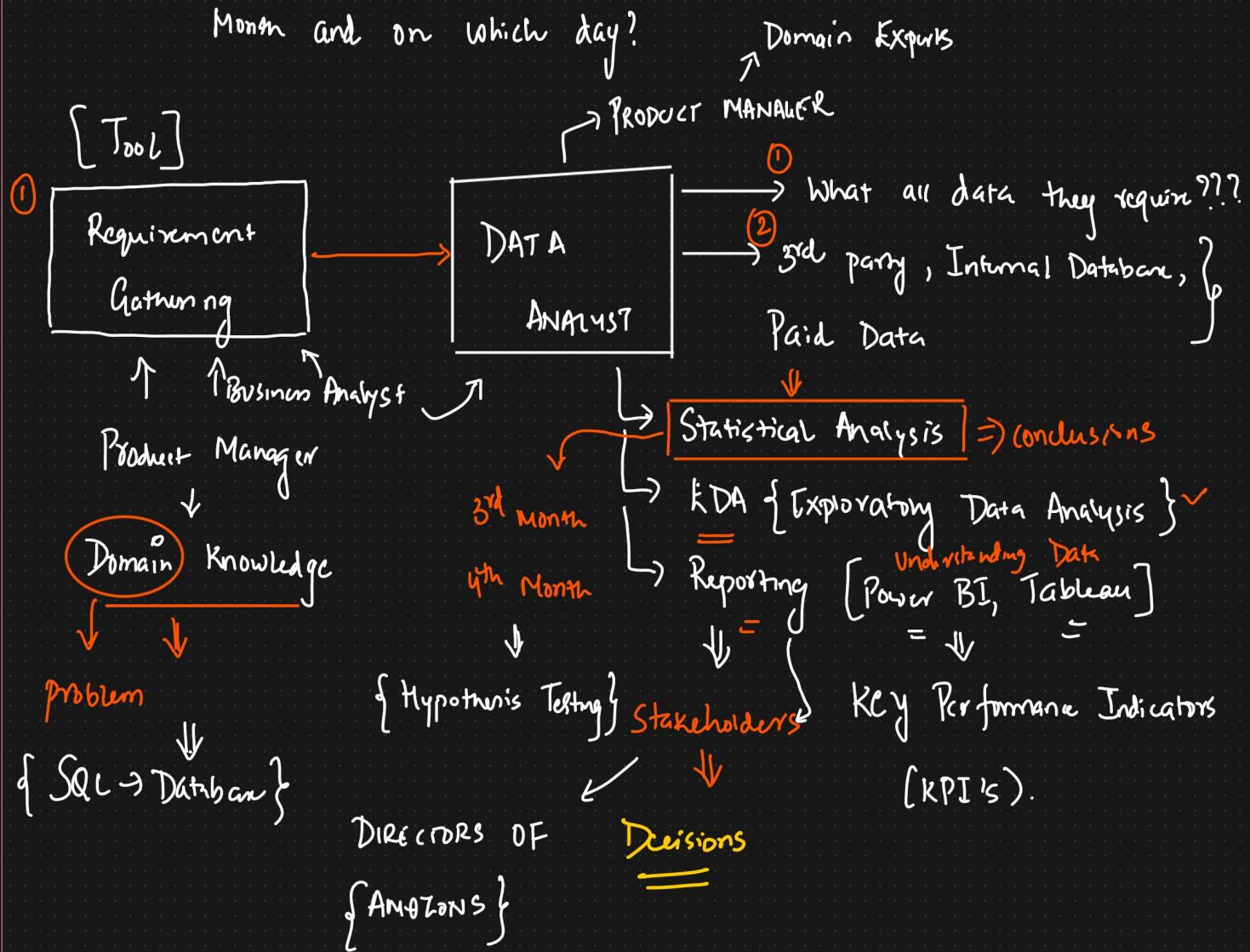


Statistics

Data Analyst → Amazon wants to decide when is the Next Big Billion Sale?



Statistics ÷ Defn

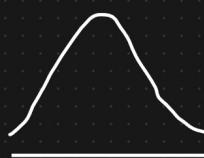
Statistics is the science of collecting, organizing and analyzing data.

Data ÷ "facts or pieces of information"

Eg: Ages of students in a classroom

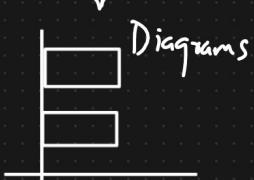
Histograms

{ 24, 30, 21, 34, 20 } \Rightarrow Mean, Median = Analysis



Eg: Weights of students in a classroom

{ 75, 45, 50, 55, 60 } \Rightarrow



2 types

Statistics

Descriptive Stats

Inferential Stats

④ It consists of organizing

and summarizing of
data



pdf

④ It consists of a technique

to form some conclusions

↓ ↓ ↓

population

Country A
West Bengal

1000

Sample EXIT Poll
of People

Aaj Tak

Histograms

BAR CHART

Box Plot

Pie
Chart



CRYPTO



CANDLESTICK

Eg: P value, Z test, t-test, Chi Square test

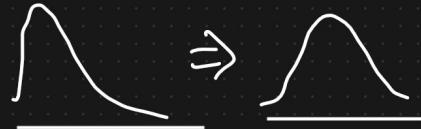
Anova,

Interviews Amazon

What is the average size of All Sharks In the world ??

Eg) Let's say there are 20 classrooms in a university and you have collected the age of students in one classroom.

Ages { 21, 20, 18, 34, 17, 22, 24, 25, 26, 23, 22 }



Descriptive Stats : What is the average age of the students in the classroom ??

{ 1, 3, 4, ^{↑ outlier} 100 }

Minimum Age, Outliers, Distribution

Maximum Age

Sample
||

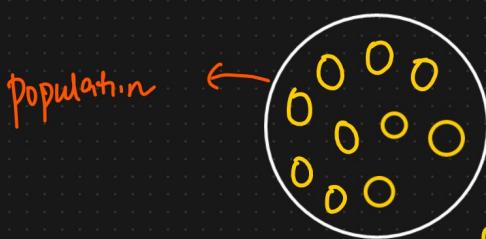
Brace over Sample and Maximum Age

Inferential question : Are the average age of the student in the classroom equal to the average ages of the entire university



↓
Population data.

(N)
Population And (n)
Sample



1 Million
= State A
Sample → Avg height
Population
Sample Size = 1000

{ Average height of all the people in the state ??

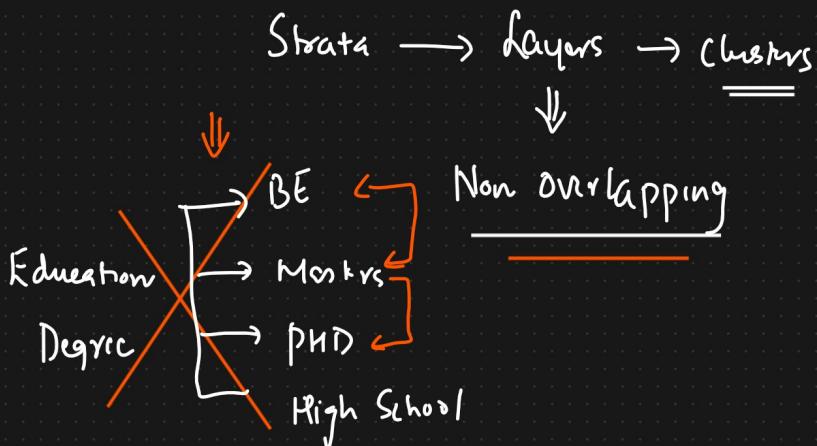
Sampling Techniques

① Simple Random Sampling \div Every member of the population (N)

has an equal chance of being selected for your sample (n)



② Stratified Sampling

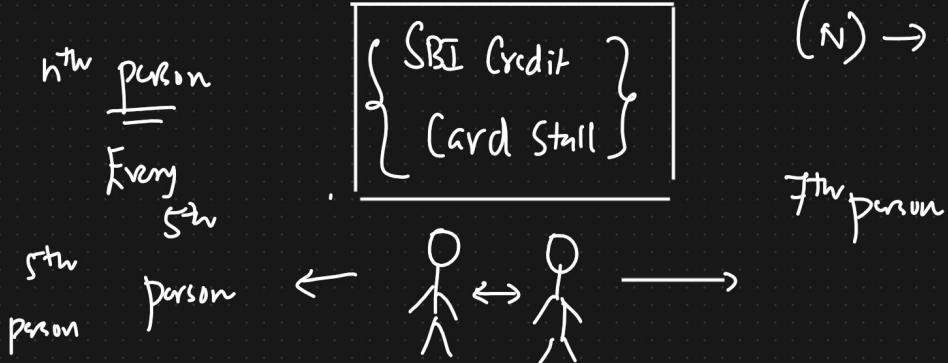


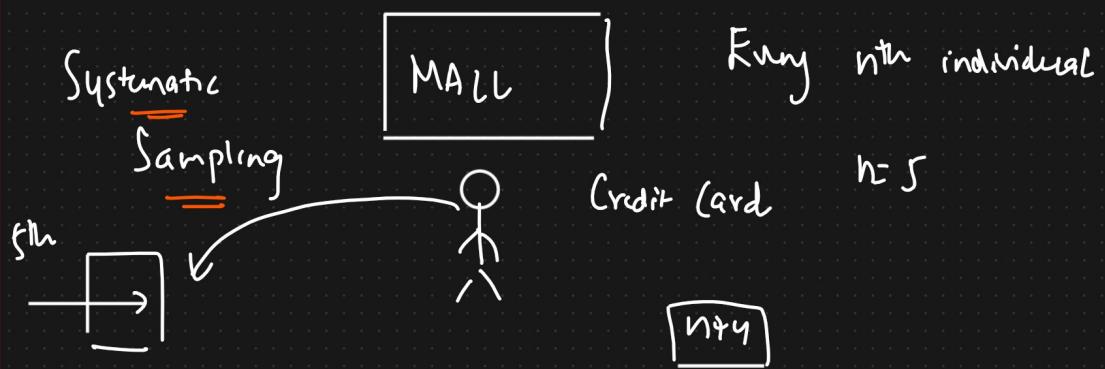
Blood Groups



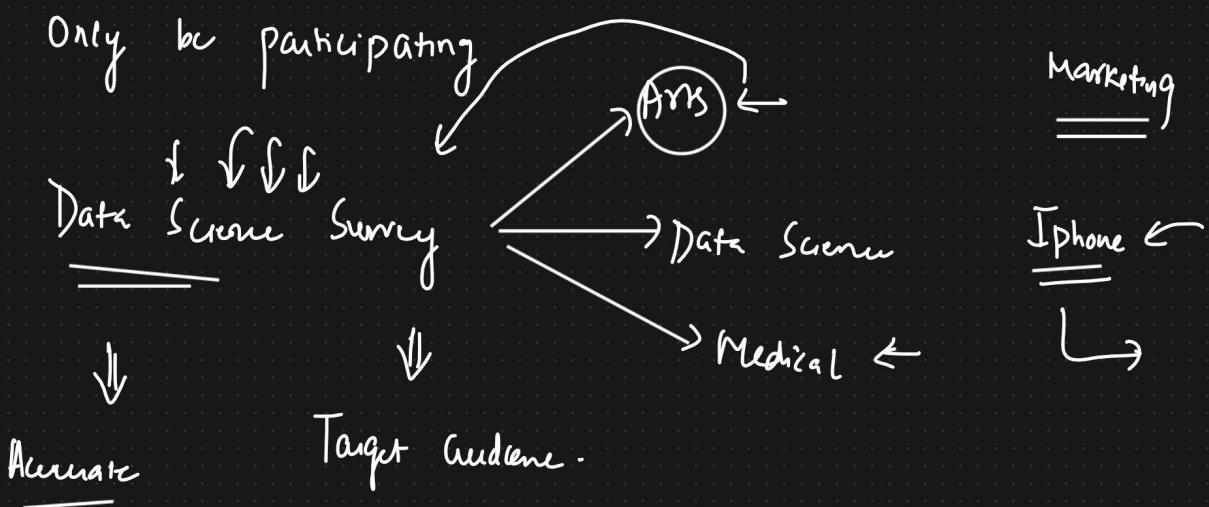
③ Systematic Sampling \div

(n) \rightarrow select every n^{th} individual





④ Convenience Sampling : Only those people who are interested will



Teams → Calls

- ① Calls Credit Card → Airport → Systematic }
Random }
- ② Exit Polls → Stratified + Random Sampling
- ③ Survey Regarding New Technology

Variable

A variable is a property that can take on any value

Eg: $Age = 24$

$Age = 25$

$Age = 26$

Age = {24, 25, 28, 29, 30} \Rightarrow Collection

=

Two kinds of variable

① Quantitative Variable \rightarrow Measured Numerically {Add, Subtract, \times & \div }

② Qualitative Variables \rightarrow Categorical Variable

L Eg: Gender $\begin{cases} \rightarrow \text{Male} \\ \rightarrow \text{Female} \end{cases}$ {Based on some characteristics we can group Categorical variables}

Type of flowers $\begin{cases} \rightarrow \text{Rose} \\ \rightarrow \text{daisy} \\ \rightarrow \text{Sunflower} \end{cases}$

Types of Movies $\begin{cases} \rightarrow \text{Action} \\ \rightarrow \text{Comedy} \\ \rightarrow \text{Action Comedy} \end{cases}$

Quantitative Variable

|

↓

Discrete Variable

Continuous Variable

Eg: Whole numbers

Eg: Height = 172.5cm, 150.1cm

160.62cm

Eg: No. of Bank Account

$$\{1, 2, 3, 4, 0, 5, \dots\}$$

~~2x~~ ~~3x~~

~~3x~~

Rainfall: 1.35cm, 1.25, 1.3589cm

Weight, Temperature.

Eg: No. of Children in a family

$$\{0, 1, 2, 3, 4, 5, \dots\}$$

~~2x~~ children

Whole Number
Numerical

~~2x~~

Assessment

$$\left. \begin{array}{l} \text{Widow} \\ \text{Married} \\ \text{Not Married} \end{array} \right\} \xrightarrow{\text{Categorical variable}} \begin{array}{l} 1 \\ 2 \end{array} \quad \begin{array}{l} 100 \text{ or } 5 \text{ km} \\ = \end{array}$$

17 years 5 months 24 day

Age?: 20 seconds.

- ① What kind of variable is Marital Status? [Categorical]
- ② What kind of " is Ganga River length? [Continuous]
- ③ What kind of " is Movie duration? [continuous]
- ④ What kind of variable IQ is? [continuous]

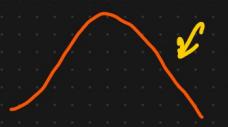
$$\{[90-100] \quad [100-120] \quad [120-150]\} \rightarrow \text{Categorical variable}$$

Histograms {Construct a histogram}

Discrete \rightarrow 10 unique

$$\left\{ \begin{array}{l} \text{Ages} = \{10, 12, 14, 18, 24, 26, 30, 35, 36, 37, 40, 41, 42, 43, 50, 51\} \\ \text{Ages} = \{65, 68, 78, 90, 95, 100\} \end{array} \right. \rightarrow \text{continuous value}$$

Probability density function



- ① Sort the Numbers {Ascending} frequency

- ② Bins \rightarrow No. of groups

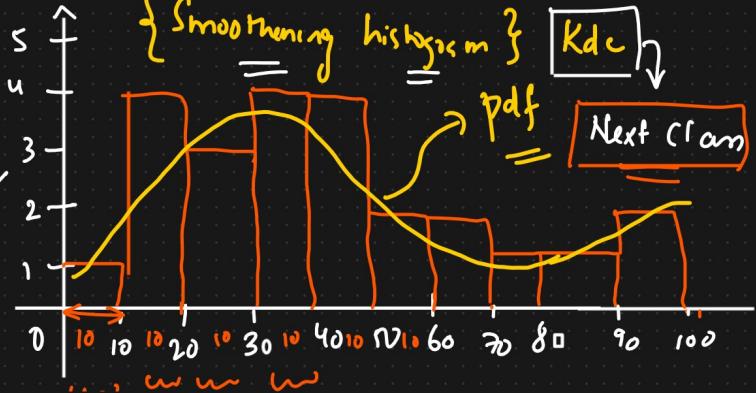
- ③ Bins size \rightarrow Size of Bins

$$[0-100]$$

$$\text{Bins} = 10$$

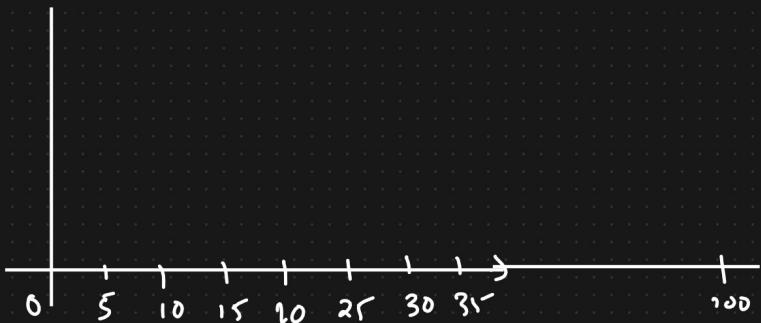
$$\begin{matrix} 0-10 \\ \hline 10-20 \\ \hline \end{matrix}$$

$$\text{Bins size} = \frac{100}{10} = 10$$



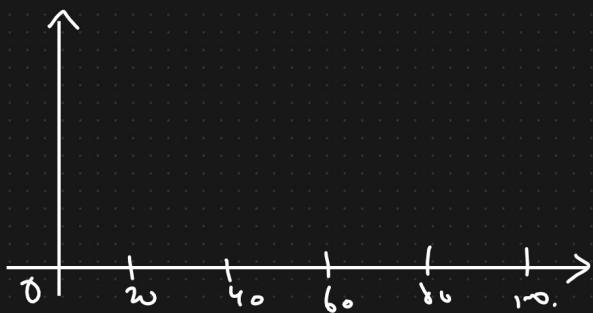
$[0-100]$ $B_{\text{bins}} = 20$

$$\text{Bin size} = \frac{100}{20} = 5$$



$[0-100]$ $B_{\text{bin}} = 5$

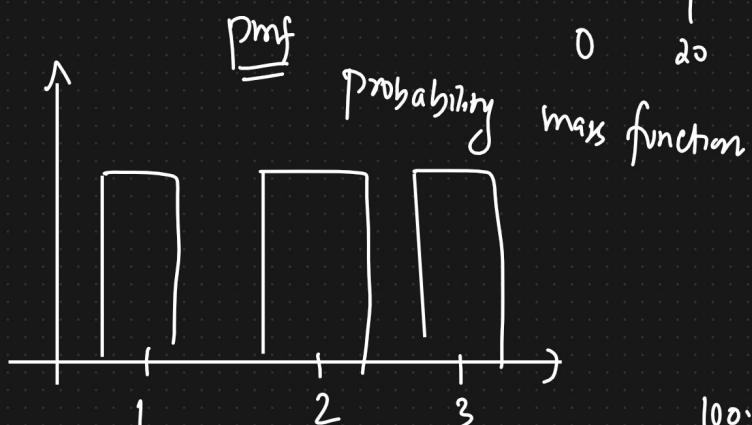
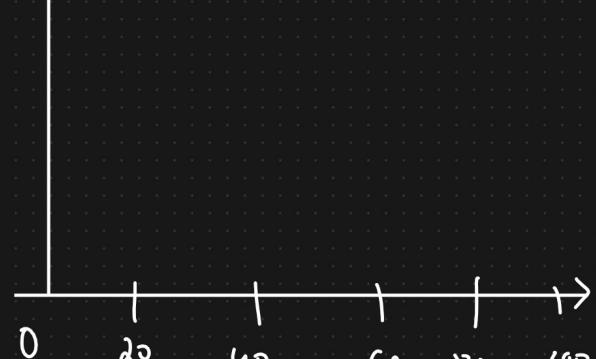
$$\text{Bin size} = \frac{100}{5} = 20$$



Assignment : Data Science $\left[\text{jayantg@inuonai} \right] \rightarrow$

Eg: $10, 13, 18, 22, 27, 32, 38, 40, 45, 51, 56, 57, 88, 90, 92, 94, 99$

$b_{\text{bins}} = 5$
$\text{bin size} = 20$



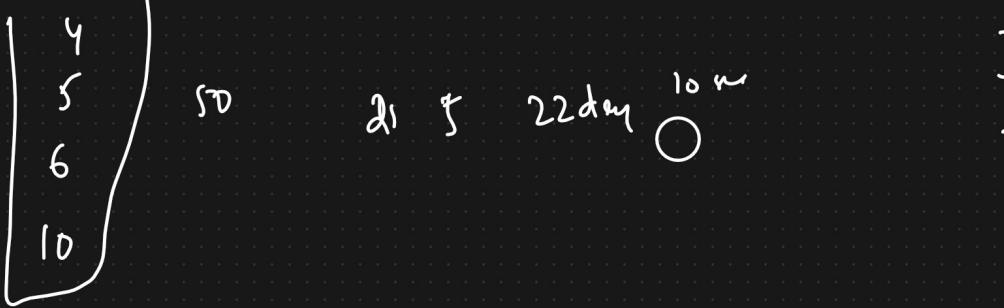
$120.5 \leftarrow$
 $\underline{\underline{\text{Age}}} = \boxed{\underline{\underline{\text{JQ}}}}$

No. of house (Discrete) \rightarrow Whole number

1	2	3
2	2	2

$- 10 \cancel{x} \rightarrow +$

-	-	-
-	-	-

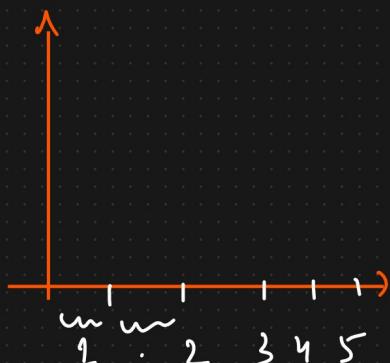


	No. of years	Age of house	No. y
<u>continuous</u>	→ 2	2.5	
<u>=</u>	→ 4	4	
<u>> 25 =></u>	→ 6	28	
<u>=</u>	→ 8	29	
(25 ↓)	→ 10	3.1	
<u>discrete</u>	→ 12	6	
{ whole number }	→ 14	4	
		2	⇒ Continuous X
		0	⇒ Restricted limit
		4	limit
		6	=
		10	
		12	⇒ Discrete

$\{ \overset{\checkmark}{1}, \overset{\checkmark}{2}, \overset{\checkmark}{3}, \overset{\checkmark}{4}, \overset{\checkmark}{5} \}$

bins = 5
0-5

$$\text{bin size} = \frac{5}{5} = 1$$



Assignments : Need to be submitted

✓ [Email: jayant @neuron.ai]