stat 3:

MCT:
--> purpose: to explain a variable's performance with a single value..

rejected aggregations --> sum() , count(), max(), min()

accepted aggregations --> mean, mode, median.

mean--> is a average.
3 means :
1. Arithmetic mean --> simple average
formula --> sum(x)/n
x --> all values a variable (list)
n --> number of elements.

```
def amean(x):
    tot = sum(x)
    n = len(x)
    am = tot/n
    return am

x = [10,20,30]
amean(x)
# 20
```

2. Geometric mean
-->
formula --> nth root of (product of all elements)

```
def gmean(x):
    m = 1
    for v in x:
        m *= v
    n = len(x)
    gm = m**(1/n)
    return gm

gmean(x)
# am > gm
```

3. Harmonic mean(scientific mean)
formula -->
hm = n/(sum of [all receprocals of each element])

```
def hmean(x):
```

```
        xi = [1/v for v in x]
        s = sum(xi)
        n = len(x)
        hm = n/s
        return hm

    hmean(x)
    # am> gm > hm
    # hm < gm < am
```

#----------

when to use what mean.

am --> where fraction is neglectable (not important)
  air temperature , person weight, class time in hh:mm:ss

gm --> where fraction is not neglectable (important)
  body temperature, gold weight

hm --> where sub fractions also important (not roundable).
    rocket launching (missile), dimond carrot calculation

for hm 2 situations-->
    1. sub fraction is also important. --> scientific applications used.
    2. outliers existed in data.
        --> wrong interpretation

#----------

mct --> mean, mode, median
mean --> avg--> most values are close to mean
mode -->highly repeated value --> most of occurance
median --> central value --> 50% above to median , 50% below to median
---

competition --> Comparative analytics.
    -->ex:
        1. which Stock is good for invensting based on returns.
        2. who is best player among Sachin and gangooly.

2. who is best player among Sachin and gangooly. (based on runs)

Mean-->

case1

	mean
Sachin	70
gangooly	30

--> Sachin is best

case2

	mean
Sachin	70
gangooly	70

--> can not take decision

case 3:

	mean
Sachin	70
gangooly	69

--> can not take decision

if two means are equal or close together, we can not conclude about performance

#-------------------------------------

solution : "Mode"

mode ---> highly repeated value.

applicants
male
female
male
male
male
female
----

n(male) --> 4
n(female) --> 2

mode --> male
interpretation --> most of applicants are males.

problem:

sachin_runs

0
90
91
99
100
110
0
---

number of matches --> 7

n(0) --> 2
n(90) ->1
n(91) -> 1
n(99) -> 1
n(100) --> 1
n(110) --> 1

mode(suchin_runs) --> 0
interpretation --> most of Sachin runs are 0 --> is it correct ? --> no

solution: ->

transform runs into "Labels" , ex: above avg , below avg

sachin_runs

0
90
91
99
100
110
0
20
30
120
110
100
:

0
9

(200 matches)
mean(Sachin) --> 94

sachin_runs	avg_status
0	below
90	below
91	below
99	above
100	above
110	above
0	below
20	:
30	:
120	above
110	above
100	above
:	:
0	:
9	:

---

n(above) --> 130
n(below) --> 70

mode(avg_status) --> 'above'
--> in most cases, Sachin gives 'above average score' .

comparative analytics.

case 1:

	mean	mode
Sachin	70	above
gangooly	70	below

Sachin is best.

case 2:

	mean	mode
Sachin	70	above
gangooly	70	above

We can not take decision.

#----------

if two modes are equal , then we can conclude decision.

solution : median

#----------

x --> 10, 90, 30, 20, 100, 40, 100, 30, 35, 45

sort(x)-->
10, 20, 30, 30, 35, 40, 45, 90, 100, 100
20, 30, 30, 35, 40, 45, 90, 100,
30, 30, 35, 40, 45, 90,
30, 35, 40, 45
35, 40 -->
avg(35,40) --> 37.5

interpretation --> 50% of x is above to 37.5,
50% of x is below to 37.5

case1

	mean	mode	median
Sachin	70	above	60
gangooly	70	above	40

--> Sachin is best

case2

	mean	mode	median

Sachin	70	above	60
gangooly	70	above	61

we can not ---> take decision
if two medians are equal or close together, we can not conclude decision..

solution --> quartiles.

3 quartiles.

Quartile 1 --> 75% of data is above to Q1, 25% of data is below to Q1

Quartile 2-->median --> 50% above , 50% below

Quartile 3 --> 25% above to Q3, 75% below to Q3.

where to look ? (at q1 or q3)

--> number of absents --> low score is priority

k1	k2
2	10

--> score of marks --> high score is priority.

k1	k2
90	20

if low value is priority , then choose q1
if high value is priority , then cho0se q3

case1

	mean	mode	median	Quartile3
Sachin	70	above	60	65
gangooly	70	above	61	74

Gangooly is best.

case2

	mean	mode	median	Quartile3
Sachin	70	above	60	75
gangooly	70	above	61	74

can not take decision.

if two q1 or q3 are equal or close together, we can not conclude decision.

#----------
in above case,
entire mct could not able to help us to take decision.

all mct explains --> positivity of a variable (goodness)

solution --> we need to look their negativity .

#----------
The negativity of a variable is explained by "Dispersion Techniques".

Disperison --> variability (spread)
3 dispersions.
1. range
2. variance
3. standard deviation.