TOPIC NAME : _____ DAY : _____

TIME : DATE : / /

① Real for data science Excell is essential,

- many company by default take it that i know excell.

② Why excell is Popular ? 1990 launced and still it is popular why?

⇒ very intutive, everything can bee see in front,

- this is data
- this is column
- this is Row,

@when we are doing some operation, we can understand, which Row and column are being used.

⇒ very human interactive that's why it's Relevent

- very good UI

③ Excel column are shown in Alphabatical order

{A to Z} then {AA - AZ} {BA - BZ}

④ Excel Rows are shown in numbe {1 to 1 millon

"GOOD LUCK"

TOPIC NAME : _____ DAY : _____

TIME : DATE : / /

(#) famous interview question ?,

(#) How many total Rows are in Excel ?,

↳ how many data we can store ?,

⇒ if we Pressed {Cntrl + down Arrow} then we go last row, and that Row Number is.

{ 10 lakh 48 thousand 576 }

=> maximum Roughly 1 million on Excel Apps.

=> in Excel website version it can go up to 5 million

(#) Every Excell cell has unique address

like A2, B12, f16 - - -

⊙ Excell address start with [column Name][Row value]

GOOD LUCK™

TOPIC NAME : _____ DAY : _____

TIME : DATE : / /

④ font tool where it used : Bold, italic, center, letter big and small, font color Heading Higlight.

this all are thing used in Excell dashboard making.

⊙ Excell is calculation kind of tools.

④ mainly we will deal with {data} tab

④ we have Excell, so what is the need of databases?

⊙ Excell is Electronical version of old databook.

⊙ Excell is ment for data crunching (activities or Processes concerned with number or mathmatical calculation) for example in finance, statistics or computing.

⊙ Excell is meant for data analysing, and find out about information from data.

④ Excell is purpose is not storing data:-

TOPIC NAME : _____ DAY : _____

TIME : _____ DATE : / /

① what is the work of databases ?

⊙ database are meant to store data, and to interact with data. millions of users can interact with databases data, like uber, one ata time they handle millons of Ride, everything connect with one databases.

⊙ this kind of flexibility Excell doesn't give, to., in real time data i can extract and use it.

⊙ Excell and databases use cases both are different. that's why we don't use Excell as database.

② the excell file we can store in 2 way?

① .csv

② .xlsx.

③ what is the difference between this two. actuall

④ and when we should store data as .csv and when we should store data as .xlsx

GOOD LUCK

TOPIC NAME : _____ DAY : _____

TIME : DATE : / /

① CSV (Comma-Separated value):

⊙ format: CSV files store data as plain text where each line Represent a row of data and columns are separated by a delimiter {often a comma, but can also be a semicolon or tab}

Advantages:

⊙ Light weight and easy to create and read using simple text editors or Programming languages.

⊙ universal compatibility across different Software and Platforms

⊙ Efficient for larges dataset since they are smaller in size compared to XLSX files.

Disadvantages:

⊙ Lacks formatting options like fonts, colors and multiple sheets.

⊙ Limited support for metadata (such as column types, formulas)

GOOD LUCK™

TOPIC NAME : DAY :
TIME : DATE : / /

② .xlsx (Excel spreadsheet)

format: .xlsx files are binary files that store data in a structured format xml and ZIP compression. They can contain multiple sheets, formatting options, formulas, and charts.

Advantages:

- Support complex data structure, formula and formatting
- Ideal for storing datasets that require detailed formatting and calculation.
- offers features like multiple sheets, cell formatting and chart embedding.

Disadvantage:

- Large file size compared to csv due to additional formatting and data structure information.
- Require specific software (like microsoft Excel or other spreadsheet application) to create and view.

GOOD LUCK

TOPIC NAME : _____ DAY : _____

TIME : DATE : / /

when to use csv:

① Data exchange: use csv when you need to exchange data between different software applications or systems that support csv format

② Large dataset: csv is preferable for large datasets where file size and processing speed are concerns.

③ Simplicity: when the data doesn't require complex formatting or formulas.

when to use xlsx:

① complex data: use xlsx when your data requires complex formatting, multiple sheets, formulas or embedded charts.

② Data analysis: if i need to perform detailed data analysis, xlsx is often more suitable due to it's support for formulas and structured data.

in summary: csv for simplicity, compatibility and efficiency with large dataset, while opt for xlsx when your data requires advanced formatting calculations or multiple sheets.

"GOOD LUCK"

. TOPIC NAME : DAY :

TIME : DATE : / /

Excell formula

Excell formula use on stand with ⇒ =(formula)

often applying formula in one column, when we drag the colum for multiple column the formula automatically adjusted. this is the most underrated feature of excell

① what is aggregession function?

⇒ I have a data, group of numbers., After summarizing the whole group of numbers to come up one number is called aggregession function.

⇒ For Data Select in Row

Fo select the whole column, { cntr + shift + down Arrow }

② To edit any formula, Press. { fn + f2 }

GOOD LUCK™

TOPIC NAME : _____ DAY : _____

TIME : DATE : / /

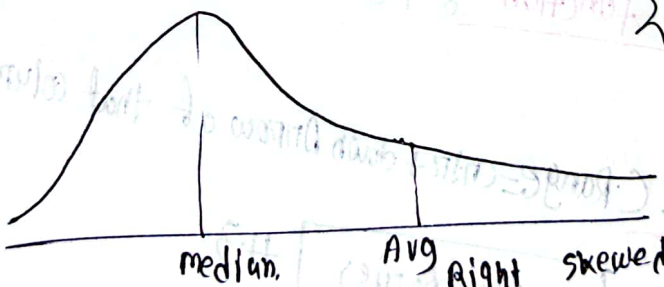
(11) my Avg = 330.031, and median = 225, so is my data is Right skewed or left skewed by knowing this 2 value?

(a) if my Avg > median then the distribution will be Right Skewed.

(a) if my Avg < median then the distribution will be left skeweed.

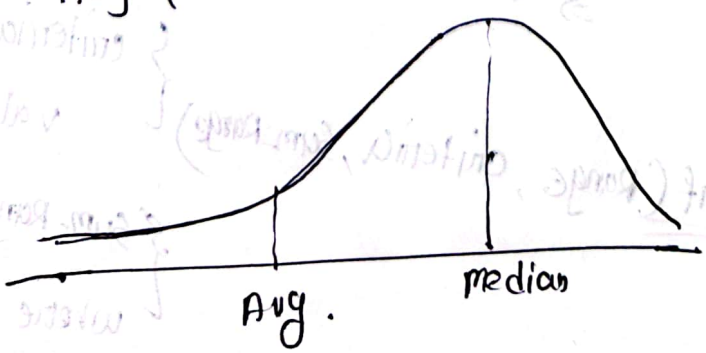
(1) for ~~median~~ & Avg > median /

{median means middle value}



Right skewed.

(1) for Avg < median.



"GOOD LUCK"

TOPIC NAME : DAY :

TIME : DATE : / /

⑪ Conditional Formatting (use for highlight value)

① Select the whole colum by Pressing Cntr + Shift + down Arrow

② the go to home tab and select conditional formatting

③ in the conditional formatting i can highlight Perticular value that are greater or less or equal ,?

⑫ Categories wise sales

④ unique function : return unique value from categories.

=> unique (Range = Cntr + down Arrow of that column)

=> Accessories H3

=> Clothing

=> Election.

{ Range = is basically, that column where we want to do operation }

=> Sumif (Range, criteria, Sum_Range)

{ criteria = basically categories value, H3 .

{ Sum_Rang = the Price column where I want to do sum }

GOOD LUCK™

TOPIC NAME : _____ DAY : _____

TIME : DATE : / /

② Particular year, how much sales has happend?

① first find out all the unique year.

i) unique (year whole colum) → this will give

=> 2007
=> 2008
=> 2006

} unique column

ii) second sumif(year whol column, unique column, Price whole Column)

iii) drag across all thos year cell.

④ if i want to sort the unique year column, i cant sort that column because of formula.

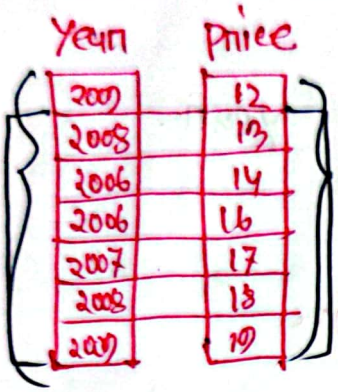
⑤ so what i should do is first copy all the year value from unique column then go past special option, and click past value then sort the value by going Data tab and sort the Range.

GOOD LUCK™

TOPIC NAME : DAY :

TIME : DATE : / /

lock in concept



① Sumif (select year, unique year, Price colum).

② if i don't lock the cell then it will gonna shift cell one down for next operation and so on.

which I don't want i want to lock the cell, ~~that's~~ why I have to use {cntr + f4} ~~after~~ select the Range value.

9a) you want to show a visualization that has one discrete (categorical) [continent] and one continuous variable (Population).

What will be your choice of visualization?

① Bar chart
② Histograme
③ Pie chart
④ Line chart

GOOD LUCK

TOPIC NAME : _____ DAY : _____

TIME : DATE : / /

what is bar chart ?

A bar chart or bar graph is a chart or graph that present categorical data with rectangulars bars with heights or length propotional to the value that they Represent. A vertical bar chart is sometimes called a column chart. use bar chart when the order of categories is not important, and have few categories

what is histograme chart ?

the histograme chart is similar to count plot.

① it create beans

② the count the value that falls into that beam.

③ then creat bar chart with this

histograme is basically combination of { bar chart + count plot }

"GOOD LUCK"

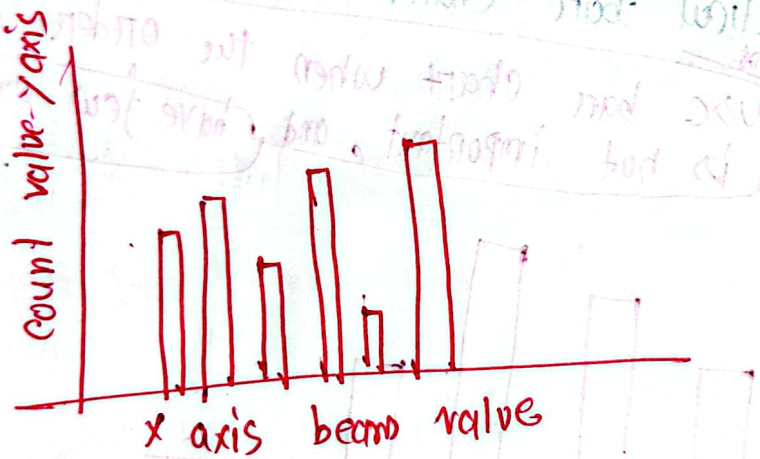
TOPIC NAME : _____ DAY : _____

TIME : DATE : / /

(#) in order to plot histograme how minimum how many point on variables needed?

(44) only one variable.

histograme is count that variable then plot it.



(#) the question answer will not be bare chart because bar chart show count of numerical value. Range.

GOOD LUCK™

TOPIC NAME : _____ DAY : _____

TIME : DATE : / /

why pie chart will not be answer? why not appropiate

- generall we don't prefer pie chart. because
- pie chart is 2 dimensional chart.
- it's preferable if the categorii is 3 or less number.
- in pie chart it's is difficult to analysis height (bar ▯▯) than areas.

why not Line chart will not be the answer?

because Line chart generall use for showing trend.

{trend means over the time?} and when there are many data points and order is important

in that question there were no time. so Line chart will not be answered.

{year} {population}

⇒ here we should use Line chart since it will show trend over the period of time

→ continent is categories

{continent} {population}

there no time → so the answer is bar chart

GOOD LUCK™

TOPIC NAME : DAY :
TIME : DATE : / /

(11) A date variables and numbers of customers in a restaurant?
what will be your choice of visualization?

(a) Bar chart
(b) Histograme
(c) pie chart
(d) Line chart.

(12) here answer will be Line chart, since in question it says (date) over the period how many customer increase or decrease.

visual claten means?

we have to be stay away from practice where visualization chart look like visualy heavey. it will not be easy for your brain to process.

(13) (b)

(14) visualizations purpose is, it should easy for our brain to process, visualization should be simplen. unnessary element should not be added in the

GOOD LUCK

TOPIC NAME : _____ DAY : _____

TIME : DATE : / /

(21) Trend line, how it's get calculated?

↳ using Linear Regression.