Question number 2

Explain the following in data analysis

1. **Data validation**

Data validation is a feature in excel which is used to control what users can enter into a cell. It allows you to dictate specific rules, it also allows users to display a custom message if users try to enter invalid data. For examples a user cannot type a number into a cell intended for names and vice visa where the user cant type text into a cell intended for numbers.

1. **Big Data**

Big Data refers to massive and complex datasets that are difficult to handle using traditional data processing methods due to their volume.it can also be refers to extremely large and diverse collections of structured, unstructured and semi structured data that continues to grow exponentially over time these datasets are huge and complex in volume, velocity, and variety.

Big data is used in machine learning, predictive modeling, and other advanced analytics to solve business problems and make informed decisions. Big data collects terabytes and sometimes even petabytes of raw data from many sources that must be received, processed and transformed into the format that business users and analysts need to start analyzing it.

1. **Data Consolidation**

Data consolidation is the process of combining data from multiple sources, cleaning and verifying it by removing errors and storing it in a single location such as data warehouse or database. Data is produced from various sources and in multiple formats. The data consolidation process makes it easier to unify that data. ETL is one of the most widely used data management technique for consolidating data. It is a process in which data is extracted from a source system and loaded into a target system after transformation (including data cleansing, aggregation, sorting etc.)

**d. Data modeling**

Data modeling is the process of creating a simplified diagram of a software system and the data elements it contains, using text and symbols to represent the data and how it flows .Data models provide a blueprint for designing a new database or re-engineering a legacy application and it helps an organization use its data effectively to meet business needs for information.

1. A data model can be thought of as a flow chart that illustrates of as a flow chart that illustrates data entities, their attributes and the relationships between entities. It enables data management and analytics teams to document data requirements for applications and identify errors in development plans before any code is written

**e.** **Functions in excel**

Functions are predefined formulas that perform calculations by using specifics values, called arguments, in a particular order, or structure. Function can be used to perform simple or complex calculations. Functions used for quickly finding the sum, the average, count, maximum value, and minimum value for a range of cells. In order to use function correctly you will need to understand the different parts of a function and how to create arguments to calculate values and cell references. In order to work correctly, a function must be written a specify ways which is called the syntax. The basic syntax for a function is an equals sign (=), the function name (sum, for example), and one or more arguments. Arguments contain the information you want to calculate.

**f. Data wrangling/cleaning**

**D**ata wrangling is the process of removing errors and combining complex data sets to make them more accessible and easier to analyze. Due to the rapid expansion of the amount of data and data sources available today storing and organizing large quantities of data for analysis is becoming increasingly necessary. Data wrangling can be defined as the process of cleaning, organizing and transforming raw data into the desired format for analysts to use for prompt decision making also known as data cleaning, data wrangling enables businesses to tackle more complex data in less time , produce more accurate results and make better decisions.

g. Conditional Formatting

Conditional formatting is a feature in spreadsheet programs like Microsoft Excel that lets you automatically change the appearance of cells based on certain conditions. For example, you could use conditional formatting to highlight cells that contain high or low values, Draw attention to cells that meet specific text criteria Use color scales or data bars to visualize trends in your data

Conditional formatting makes it easy to highlight certain values or make particular cells easy to identify. This change the cell range based on a conditional formatting to highlight a cell that contains certain condition.

QUESTION 3a

WRITE OUT 30 EXCELS FUNCTIONS AND STATE EACH OF THEM USES WITH PRACTICAL EXAMPLES.

**SUM (range)**: Adds values in a range. Example: =SUM (A1:A10) sums cells A1 to A10.

**AVERAGE (range)**: Calculates the average of a range. Example: =AVERAGE

(B2:B15) finds the average of cells B2 to B15.

**COUNT (range)**: Counts the number of cells containing numbers in a range. Example: =COUNT (C1:C20) counts cells C1 to C20 with numbers.

**COUNTBLANK (range)**: Counts the number of empty cells in a range. Example: =COUNTBLANK (D3:D8) counts empty cells D3 to D8.

**SQRT (number)**: Calculates the square root of a number. Example: =SQRT (25) returns 5.

**PI ()**: Returns the value of pi (3.14159...). Example: =PI ()\*4^2 calculates the area of a circle with radius 4.

**SIN (angle)**: Calculates the sine of an angle in radians. Example: =SIN (PI()/2) returns 1 (sine of 90 degrees).

**IF (logical test value if true, value if false)**: Checks a condition and returns a value based on the outcome. Example: =IF (A1>10, "High", "Low") displays "High" if A1 is greater than 10, otherwise "Low".

**AND (logical1, logical2,...)**: Returns TRUE if all conditions are TRUE, FALSE otherwise. Example: =AND (A1>5, B1<10) returns TRUE only if A1 is greater than 5 and B1 is less than 10.

**OR (logical1, logical2,...)**: Returns TRUE if any condition is TRUE, FALSE otherwise. Example: =OR (C1="Apple", C1="Orange") returns TRUE if C1 contains "Apple" or "Orange".

**CONCATENATE (text1, text2,...)**: Joins multiple text strings into one. Example: =CONCATENATE ("Hello ", A1) combines "Hello" with the content of cell A1.

**LEN (text)**: Returns the length of a text string. Example: =LEN ("This is a text") returns 15 (including spaces).

**LEFT (text, num\_chars)**: Extracts a specified number of characters from the left side of a text string. Example: =LEFT (A1, 3) extracts the first 3 characters from cell A1.

**RIGHT (text, num\_chars)**: Extracts a specified number of characters from the right side of a text string. Example: =RIGHT (B1, 4) extracts the last 4 characters from cell B1.

**UPPER (text)**: Converts all characters in a text string to uppercase. Example: =UPPER ("hello*")*returns "HELLO".

**LOWER (text)**: Converts all characters in a text string to lowercase. Example: =LOWER ("WORLD") returns "world".

**TODAY ()**: Returns the current date. Example: =TODAY () shows the current date in your system format.

**NOW ()**: Returns the current date and time. Example: =NOW () shows the current date and time in your system format.

**YEAR (date)**: Extracts the year from a date. Example: =YEAR (A1) extracts the year from the date in cell A1.

**MONTH (date)**: Extracts the month from a date. Example: =MONTH (B2) extracts the month from the date in cell B2.

**DAY (date)**: Extracts the day from a date. Example: =DAY (C3) extracts the day from the date in cell C3.

Question 3b

Write out 45 Excel short cut keys and state their uses

1. **Ctrl + N:** Create a new workbook.
2. **Ctrl + O:** Open an existing workbook.
3. **Ctrl + S:** Save the current workbook.
4. **Ctrl + W:** Close the current workbook.
5. **Ctrl + F4:** Close Excel.
6. **Ctrl + Tab:** Switch to the next open workbook.
7. **Ctrl + Page Up:** Switch to the previous worksheet.
8. **Ctrl + Page Down:** Switch to the next worksheet.
9. **Ctrl + delete:** clear the contents of the selected cells
10. **Enter:** Edit the content of the selected cell.
11. **F2:** Edit the content of the selected cell.
12. **Esc:** Exit edit mode for the selected cell.
13. **Tab:** Move one cell to the right.
14. **Shift + Tab:** Move one cell to the left.
15. **Up arrow:** Move one cell up.
16. **Down arrow:** Move one cell down.
17. **Home:** Move to the beginning of the current row.
18. **End:** Move to the end of the current row (last used cell).
19. **Ctrl + Home:** Move to cell A1 (top-left corner).
20. **Ctrl + End:** Move to the last used cell on the worksheet.
21. **Pg Up:** Scroll up one page.
22. **Pg Dn:** Scroll down one page.
23. **Ctrl + E:** open the “sort” dialog box to sort data
24. **Ctrl + B:** Bold selected text.
25. **Ctrl + I:** Italicize selected text.
26. **Ctrl + U:** Underline selected text.
27. **Ctrl + Shift + $:** Apply currency format.
28. **Ctrl + Shift + %:** Apply percentage format.
29. **Ctrl + 1:** Open the format cells dialog box.
30. **Alt + H + B + Borders:** Add a border to the selected cells.
31. **Alt + H + A + Fill Colors:** Fill the selected cells with color.
32. **Ctrl + shift++:** Add a new blank worksheet.
33. **Ctrl + C:** Copy selected cells.
34. **Ctrl + V:** Paste copied cells.
35. **Ctrl + X:** Cut selected cells.
36. **Ctrl + Z:** Undo the last action.
37. **Ctrl + Y:** Redo the last undone action.
38. **Delete:** Delete the content of the selected cells.
39. **Backspace:** Delete the character to the left in the cell.
40. **Ctrl + D:** Fill down the contents of the above cell into the selected cells.
41. **Ctrl + R:** Fill right the contents of the leftmost cell into the selected cells.
42. **Ctrl + space:** select the entire current column.
43. **=:** Enter the formula editing mode in the selected cell.
44. **F4:** Absolute cell reference toggle (e.g., A1 vs. $A$1).
45. **F2:** Edit the formula in the formula bar.

**Question 4b**

**State the benefit of Tables in data analysis**

Tables are fundamental tools in data analysis, offering several key benefits:

1. **Organization and Clarity:** They present data in a structured format with rows and columns, making it easy to read and understand, especially for large datasets.
2. **Efficient Comparison:** Tables allow for side-by-side comparisons between different data points or categories. You can easily spot trends or identify outliers by looking across rows or down columns.
3. **Preserving Precision:** Unlike charts, tables maintain all the exact values, which is crucial for detailed analysis where specific numbers matter.
4. **Data Exploration:** Tables are great for initial data exploration. You can quickly scan through rows and columns to get a sense of the data distribution and identify potential areas for further investigation.
5. **Integration with Other Tools:** Data tables can be easily exported and imported into various data analysis software and tools, facilitating further manipulation and analysis.

**Question 4c**

**What is Pivot Table**

A pivot table is a powerful tool used to summarize and analyze data in spreadsheets. It takes a large dataset and reorganizes it into a more concise and informative way, allowing you to see trends and patterns that might not be readily apparent in the raw data.

**Source data:** Imagine you have a large table with lots of rows and columns containing information, like sales figures for different products across various regions.

**Rearrangement:** The pivot table takes this data and lets you group it by different categories, such as region or product category. You can also choose how to summarize the data within each group. Common summaries include sums, averages, and counts.

**Insights:** By rearranging and summarizing the data, the pivot table provides insights you can't easily get from the original table. For example, you can quickly see which region has the highest sales or which product category is performing the best.