CA-C24L: Data Analytics Lab

Part- A: Spreadsheet (Excel)

1. CONDITIONAL FORMATTING, IF, COUNTIF, SUMIF, AVERAGE, CONCAT
2. INDEX, MATCH, UNIQUE, IFS, COUNTIFS, SUMIFS, AVERAGEIFS
3. VLOOKUP, HLOOKUP, XLOOKUP, COUNT, COUNTA
4. LEFT, MID, RIGHT, LEN, SUBSTITUTE, SEARCH, ISNUMBER
5. TODAY, NOW, YEAR, MONTH, NETWORKDAYS, EOMONTH
6. OFFSET, CHOOSE, LET, MAX, SORT, SORTBY, RANK
7. FILTER, FRQUENCY, SEQUENCE, RANDARRAY, IFERROR
8. PIVOT TABLES, WHAT-IF ANALYSIS, DATA VALIDATION, SUBTOTALS WITH RANGES
9. Develop an interactive dashboard for the Financial Sample Excel workbook

(https://learn.microsoft.com/en-us/power-bi/create-reports/sample-financial-download) or Sample-Superstore Excel data

Part- B: Data Analysis using Python

Note: Download the sample data file from the open sources (Kaggle.com, etc.,) or from prescribed study materials to apply & practice all these methods using Python.

1. Probability
   a) Calculating the simple probabilities
   b) Applications of Probability distributions to real life problems
2. Test of significance
   a) T-Test: one sample, two independent samples and paired
   b) ANOVA & Chi-Square Test
3. Correlation and Regression analysis
   a) Scattered diagram, calculating of correlation coefficient
   b) Linear regression: fitting, testing model adequacy and prediction (simple and multiple)
   c) Fitting of logistic regression

Part- C: Power BI

1. Introduction to Power BI- Get Started with Power BI - Sign up for Power BI - Overview: Power BI data sources - Connect to a SaaS solution - Upload a local CSV file - Connect to Excel data that can be refreshed - Create a Report with Visualizations
2. Using visualizations - Create a new report - Create and arrange visualizations – Format a visualization - Use text, map, and gauge visualizations and save a report - Use a slicer to filter visualizations - Sort, copy, and paste visualizations
3. Modify and Print a Report - Rename and delete report pages - Add a filter to a page or report Set visualization interactions - Send a report to PowerPoint
4. Create a Dashboard - Create and manage dashboards - Pin a report tile to a dashboard - Pin a live report page to a dashboard - Pin a tile from another dashboard - Pin an Excel element to a dashboard - Add a tile to a dashboard

Part- A: Spreadsheet (Excel)

Note:

- Download the sample data file from the open sources (Kaggle.com, etc.,) to apply & practice all these functions.
- **Syntax depends on your Excel language settings**:
  - ✓ If you're using **English (US)** Excel, you should use **commas (,)** as argument separators, not semicolons (;).
  - ✓ If you're using **another locale**, like many European versions, the semicolon is correct.
  - ✓ **Optional improvement**: Make the labels consistent (capitalization and spacing).

1. CONDITIONAL FORMATTING, IF, COUNTIF, SUMIF, AVERAGE, CONCAT

Definition: Conditional Formatting is a feature in Excel that allows you to automatically apply formatting (like colors, bold text, or borders) to cells based on their values or formulas.
Example: Select Salary column -> HomeTab -> Conditional Formatting -> Highlight All -> Greater than -> option -> Enter the value -> Click Ok

Definition: The **IF function** checks whether a condition is true or false, and returns one value if true and another if false.
Syntax: =IF(logical_test, value_if_true, value_if_false)
Example: =IF(ColumnNumber>50000, "Good", "Average")
This checks if the value in cell A1 is greater than 50. If it is, it returns “Good”; otherwise, it returns “Average”.

Definition: The **COUNTIF function** counts the number of cells that meet a specified condition in a range.
Syntax: =COUNTIF(range, criteria)
Example: =COUNTIF(B2:B10, ">100")

Definition: The **SUMIF function** adds up all the values in a range that meet a specified condition.
Syntax: =SUMIF(range, criteria, [sum_range])
Example: =SUMIF(A2:A10, ">50", B2:B10)

Definition: The **AVERAGE function** calculates the arithmetic mean of a group of numbers.
Syntax: =AVERAGE(number1, [number2], ...)
Example: =AVERAGE(B2:B10)

Definition: The **CONCAT function** joins (or concatenates) two or more strings or cell values into one continuous text.
Syntax: =CONCAT(text1, [text2], ...)
Example: =CONCAT(A2, " ", B2)

2. INDEX, MATCH, UNIQUE, IFS, COUNTIFS, SUMIFS, AVERAGEIFS

Definition: Returns the value of a cell at a specific row and column within a range or array.
Syntax: =INDEX(array, row_num, [column_num])
Example: =INDEX(A2:C5, 2, 3)
Returns the value from **2nd row** and **3rd column** of the range **A2:C5**.

Definition : Returns the **position** of a value in a range.
Syntax: =MATCH(lookup_value, lookup_array, [match_type])
match_type: Use 0 for exact match.
Example: =MATCH(75, B2:B6, 0)
Finds the position of **75** in the range B2:B6.

Definition: Returns a list of **unique values** from a range or array.
Syntax: =UNIQUE(array)
Example: =UNIQUE(A2:A10)
Extracts unique names from the list in A2:A10.

Definition: Checks multiple conditions and returns a value for the **first TRUE condition**.
Syntax: =IFS(condition1, value1, condition2, value2, ...)
Example: =IFS(A2>=90, "A", A2>=75, "B", A2>=60, "C", A2<60, "Fail")
Grades a score in A2 based on conditions.

Definition: Counts how many times multiple criteria are met across one or more ranges.
Syntax: =COUNTIFS(range1, criteria1, range2, criteria2, ...)
Example: =COUNTIFS(A2:A10, "IT", B2:B10, ">70")
Counts how many students are from the "IT" department and have marks > 70.

Definition: Sums values based on multiple conditions.
Syntax: =SUMIFS(sum_range, criteria_range1, criteria1, ...)
Example: =SUMIFS(C2:C10, A2:A10, "HR", B2:B10, ">30000")
Adds salaries in C2:C10 where department is HR and salary > 30,000.

Definition: Calculates the **average** of values that meet **multiple criteria**.
Syntax: =AVERAGEIFS(average_range, criteria_range1, criteria1, ...)
Example: =AVERAGEIFS(C2:C10, A2:A10, "Science", B2:B10, ">75")
Averages the marks (C2:C10) of students in "Science" who scored more than 75.

3. VLOOKUP, HLOOKUP, XLOOKUP, COUNT, COUNTA

Definition: Searches for a value **vertically** in the first column of a range and returns a value from the same row in another column.
Syntax : =VLOOKUP(lookup_value, table_array, col_index_num, [range_lookup])

- lookup_value: The value you want to find.
- table_array: The range to search in.
- col_index_num: Column number to return value from.
- range_lookup: Optional. Use FALSE for exact match.

Example : =VLOOKUP("Alice", A2:C6, 3, FALSE)

Definition: Searches for a value horizontally in the first row of a range and returns a value from the same column in another row.
Syntax: =HLOOKUP(lookup_value, table_array, row_index_num, [range_lookup])
Example: =HLOOKUP("Marks", A1:E3, 2, FALSE)

Definition: Searches for a value in a range and returns a corresponding value from another range. It works vertically or horizontally and is more flexible than **VLOOKUP or HLOOKUP**(Modern Replacement for VLOOKUP/HLOOKUP)
Syntax : =XLOOKUP(lookup_value, lookup_array, return_array, [if_not_found], [match_mode], [search_mode])
Example: =XLOOKUP("David", A2:A6, C2:C6)

Definition- Count: Counts numeric cells only.
Syntax: =COUNT(range)
Example: =COUNT(A2:A10) counts numbers in A2 to A10.

Definition-Counta: Counts all non-empty cells.
Syntax: =COUNTA(range)
Example: =COUNTA(A2:A10) counts everything filled in A2 to A10.

4. LEFT, MID, RIGHT, LEN, SUBSTITUTE, SEARCH, ISNUMBER

LEFT: Gets a set number of characters from the beginning of text.
Syntax: =LEFT(text, [num_chars])
Example: =LEFT("Excel",2) returns “Ex”.

MID: Gets characters from the middle of text.
Syntax: =MID(text, start_num, num_chars)
Example: =MID("Excel",2,3) returns “xce”

RIGHT: Gets a set number of characters from the end of text.
Syntax: =RIGHT(text, [num_chars])
Example: =RIGHT("Excel",3) returns “cel”

LEN: Shows number of characters in text.
Syntax: =LEN(text)
Example: =LEN("Excel") returns 5.

SUBSTITUTE: Replaces one part of the text with another.
Syntax: =SUBSTITUTE(text, old_text, new_text, [instance_num])
Example: =SUBSTITUTE("Hello World", "World", "Excel") returns “Hello Excel”.

SEARCH: Finds the position where text first appears in another text.
Syntax: =SEARCH(find_text, within_text, [start_num])
Example: =SEARCH("c", "Excel") returns 3.

ISNUMBER: Checks if something is a number.
Syntax: =ISNUMBER(value)
Example: =ISNUMBER(A2) returns TRUE if A2 is a number.

5. TODAY, NOW, YEAR, MONTH, NETWORKDAYS, EOMONTH

TODAY: Returns the current date (as per your system clock).
Syntax: =TODAY()
Example: If today's date is August 17, 2025,
=TODAY() → 17-Aug-2025

NOW : Returns the current date and time.
Syntax: =NOW()
Example: If now is August 17, 2025, 4:09PM,
=NOW() → 17-Aug-2025 16:09

YEAR: Returns the year component from a date.
Syntax: =YEAR(serial_number)
Example: For the date 17-Aug-2025 in cell A1:
=YEAR(A1) → 2025

MONTH: Returns the month (as a number 1–12) from a date.
Syntax: =MONTH(serial_number)
Example: For the date 17-Aug-2025 in cell A1:
=MONTH(A1) → 8

NETWORKDAYS : Returns the number of whole working days between two dates (excluding weekends and optionally specified holidays).
Syntax: =NETWORKDAYS(start_date, end_date, [holidays])
Example: From August 1, 2025 (in A1) to August 17, 2025 (in A2), excluding Aug 15, 2025 (holiday in B1):
=NETWORKDAYS(A1, A2, B1) → 12

6. EOMONTH: Returns the serial number for the last day of the month, n months before or after a specified date.
Syntax: =EOMONTH(start_date, months)
Example: For date 17-Aug-2025 in A1, last day of current month:
=EOMONTH(A1,0) → 31-Aug-2025
For two months ahead:
=EOMONTH(A1,2) → 31-Oct-2025

6. OFFSET, CHOOSE, LET, MAX, SORT, SORTBY, RANK

OFFSET: Returns a reference to a range that is a specified number of rows and columns from a cell or range of cells. Useful for creating dynamic ranges.

Syntax: OFFSET(reference, rows, cols, [height], [width])

- reference: Starting cell or range
- rows: Number of rows to offset
- cols: Number of columns to offset
- height: (Optional) Number of rows to return
- width: (Optional) Number of columns to return

Example: =OFFSET(A1, 3, 1)

Returns reference 3 rows down and 1 column to the right of cell A1.

CHOOSE: Returns a value from a list using a given position or index.

Syntax: CHOOSE(index_num, value1, [value2], ...)

- index_num: Position to return (1–254)
- value1, value2, ...: Values to choose from

Example: =CHOOSE(2, "Red", "Blue", "Green")

Returns "Blue" as it is the second value.

LET: Allows you to assign names to calculation results inside a formula, making formulas easier to read and faster.

Syntax: LET(name1, name_value1, [name2], [name_value2],…, calculation)

- name1: Variable name
- name_value1: Value or calculation
- calculation: Expression using the defined names

Example: =LET(x, 2, y, 5, x*y)

Returns 10 (2 multiplied by 5).

4.MAX : Returns the largest (maximum) numeric value in a range of cells, ignoring text.

Syntax: MAX(number1, [number2], ...)

- number1, number2, ...: Numbers or ranges

Example: =MAX(C2:C7)

Returns the highest value in the range C2 to C7.

5.SORT : Sorts the contents of an array or range by columns or rows, in ascending or descending order.

Syntax:

SORT(array, [sort_index], [sort_order], [by_col])

- array: Range to sort
- sort_index: (Optional) Which column/row to sort by (default: 1)
- sort_order: (Optional) 1 for ascending (default), -1 for descending
- by_col: (Optional) FALSE for rows (default), TRUE for columns

Example: =SORT(A2:B10, 2, -1)

Sorts the range A2:B10 by the second column in descending order.

6.SORTBY: Sorts a range or array based on values in one or more corresponding arrays.
Syntax: SORTBY(array, by_array1, [sort_order1], ...)

- array: Range to sort
- by_array1: Array to sort by
- sort_order1: (Optional) 1 for ascending (default), -1 for descending

Example: =SORTBY(A2:B10, B2:B10, 1)
Sorts A2:B10 by the values in B2:B10 in ascending order.

7.RANK : Returns the rank of a number in a list of numbers: its size relative to other values in the list.
Syntax: RANK(number, ref, [order])

- number: The number to rank
- ref: Array of numbers
- order: (Optional) 0 for descending (default), 1 for ascending

Example: =RANK(C2, C2:C10, 0)
Returns the rank of the value in C2 among values C2 to C10, with 1 as highest.

7. FILTER, FRQUENCY, SEQUENCE, RANDARRAY, IFERROR

1. FILTER: Returns only the rows or columns that meet supplied criteria from a given range.
Syntax: =FILTER(array, include, [if_empty])
Example: =FILTER(A2:C10, B2:B10="Sales")
Returns rows from A2:C10 where column B is "Sales".

2. FREQUENCY: Calculates how often values occur within a range of values, and returns a vertical array of numbers.
Syntax: =FREQUENCY(data_array, bins_array)
Example: =FREQUENCY(A2:A20, D2:D5)
Counts how many values in A2:A20 fall into the bins specified in D2:D5.

3. SEQUENCE: Generates a sequence of numbers in an array, useful for creating number lists or ranges.
Syntax: =SEQUENCE(rows, [columns], [start], [step])
Example: =SEQUENCE(5, 1, 10, 2)
Creates a vertical list: 10, 12, 14, 16, 18.

4. RANDARRAY : Returns an array of random numbers, flexible for array size, min/max values, and integer/decimal choice.
Syntax: =RANDARRAY([rows], [columns], [min], [max], [whole_number])
Example: =RANDARRAY(3, 2, 5, 15, TRUE)
Fills a 3x2 array with random whole numbers between 5 and 15.

5. IFERROR : Returns a value you specify if a formula evaluates to an error; otherwise, returns the formula's result.
Syntax: =IFERROR(value, value_if_error)
Example: =IFERROR(A2/B2, "Error!")
Divides A2 by B2; if B2 is 0 (causing an error), returns "Error!".

8. PIVOT TABLES, WHAT-IF ANALYSIS, DATA VALIDATION, SUBTOTALS WITH RANGES

1. Pivot Tables : A Pivot Table is a powerful Excel tool that allows you to summarize, analyze, and organize large data sets by grouping, counting, or totaling data in an interactive table.

Syntax/How to Use:

- Select your data range (with headers).
- Go to Insert > PivotTable.
- Choose either a new worksheet or existing worksheet for table location, and click OK.
- Drag and drop fields between Rows, Columns, Values, and Filters.

Example: Suppose your table has columns: Date, Product, and Sales.
Creating a Pivot Table lets you view total sales by product or date, and filter by month, product, etc.

2. What-If Analysis : What-If Analysis lets you explore outcomes by changing input values in your spreadsheets. Common tools are Data Tables, Scenario Manager, and Goal Seek.

Syntax (Data Table Example):

- Set up a formula (e.g., =A2*B2 for Price x Quantity).
- Highlight the table area with changing values (e.g., different prices).
- Go to Data > What-If Analysis > Data Table.
- For single input: Enter the row or column input cell.

Example:
You want to see profit if price changes:

Price	Profit
10	=A2*B2
20	……..

After using Data Table, you get profits for each price.

3. Data Validation : Data Validation is an Excel feature to restrict or control the type of data entered into a cell (e.g., only numbers, certain date range, specific list).

Syntax:

- Select cell/range.
- Go to Data > Data Validation.
- Choose criteria (Whole number, List, Date, etc.).

Example: Restrict a cell to values from a drop-down list (say: “Yes”, “No”, “Maybe”).

- Select cell A1.
- Data > Data Validation > Allow: List > Source: Yes,No,Maybe.

4. Subtotals with Ranges :
The SUBTOTAL function calculates aggregate values (sum, average, count, etc.) for a specified range, optionally including hidden/filtered rows.

Syntax: =SUBTOTAL(function_num, range1, [range2], ...)

- function_num: Specifies calculation type (e.g., 9 for SUM, 1 for AVERAGE, 2 for COUNT)
- range1: The data range

Example: =SUBTOTAL(9, B2:B20)

This formula gives the sum of B2:B20, ignoring rows hidden by filter.

Summary Table

Pivot Table	Summarize/analyze large data interactively	N/A (Insert > PivotTable)	Sum sales by product
What-If	Test outcomes by changing inputs	Data > What-If > Data Table	Profit for prices 10, 20
Data Validation	Restrict cell entries (list, numbers, dates, etc.)	Data > Data Validation	Drop-down list "Yes/No"
Subtotal	Aggregate calculation for ranges (sum, avg, etc.), works well with filtered rows	=SUBTOTAL(9, B2:B20)	Sum of visible sales amount
Pivot Table	Summarize/analyze large data interactively	N/A (Insert > PivotTable)	Sum sales by product