**Job assistance by Code basics:**

**Application tracking system (ATS):**

1. Resume Building
2. Latest Job Updates By discord server
3. Mock-Up Interview Module
4. Online credibility
5. Interview Questions

**Business project that will cover in this project:**

1. Excel -2
2. Power BI - 1
3. SQL – 1
4. Python -1
5. Practice -2
6. Virtual Internship 1

**The soft skills I will acquire by the end of the Boot Camp:**

1. Domain and function Knowledge
2. online Credibility
3. Problem-solving Mindset
4. Using the internet as the second Brain
5. Stakeholder management
6. Project Management

**Commitments from our side to have a good career:**

1. Work on exercises and practice problems
2. Don’t get distracted and leave in between – Keep away your Phone 20 m distance and no social media
3. Genuinely help others Focus on Non-tech skills too

**Excel:**

**Basics of Excel:**

* Zoom in and Out – Ctrl + Scroll in and out
* For copying of cell memory (Ctrl + C) and pasting in another cell memory (Ctrl + V)
* Windows + D shortcut will directly go to the desktop.
* For Column heads normally we use fill colour option to stand out from the Excel and make the text as Bold (Ctrl + B).
* Generally, the Number alignment in cell will be right aligned and text will be left aligned by default, if we don’t the like the alignment we can change.
* In Home tab under Number section, we can format the cells Ex: percentage.
* In conditional formatting we format the cell by some condition based on our requirement.
* Auto sum option will give the total of your selected cells.
* Sort & Filter option will sort the data based on ascending and descending and filter will be applied to the column.

**Introduction to formulas:**

Whenever you use type (=) then it indicates to use on formula.

For conditional check we use many formulas one of them is IF

IF syntax: =IF (logical test, [value if true], [value if false]) Ex: IF (D2> 100, “Yes”, “No”)

SUMIF syntax: = SUMIF (range, criteria, [Sum range]) Ex: SUMIF (B2:12, G2, D2:D12)

In SUMIF range will be the some if condition block where we have condition to sum and criteria means by what cell value or item to be sum and last sum of range.

**Locking referenced cells in a formula helps prevent dynamic changes when the formula is copied or moved to another cell (Formula Behaviour):** In whenever we apply the formula and apply it other cells the cell range may be changed by automatically if we don’t want to change the cells and rows reference from the formula, we use the dollar ($) symbol. $ symbol will lock the column and row or both at a time.

F4 key will used to fix the rows and column.

Introduction to tables:   
In selecting particular range of cells and rows we can make in table in insert section where we can make a table and we rename the table under name of at the top left and we can make use many options with tables and we can use **Slicer option** where we can filter the by slinge click.

**Table Benefits:**

Named ranges allow calculations using descriptive names rather than cell references.

Formulas automatically applied to the new rows

Additional filter such as Slicer, Auto totals, filters

Headers are visible when you scroll down: Useful for big tables

**Introduction to charts:**

The Recommended charts feature in Excel Suggests the most appropriate charts based on the selected data.

Converting data into tables allows the automatic addition of new data to the charts.

The Unique () function extracts unique values from a range or array.

Other essential features:

When we to show the percentage then we can select the cells and click percentage symbol Number section and we increase of decrease the decimal point in that section itself.

We can also format the dates and in many ways using option format cells.

Wrap test option is used when we had hidden text in any cell and that is missing and if we use this wrap text the hidden text will be come into next line.

Merge and centre option combine multiple cells into one cell and centres the content horizontally within the merged cells.

Format painter quickly copies and applies the cell formatting to the other cells or ranges in your spreadsheet.

Format cell dialog box help you to customize cell appearance and behaviour in your spread sheet.

Excel provide several security features to help protect your spreadsheet from unauthorized access or modification.

**Chapter 3: Excel Basics: cleaning and combining Data**

Cleaning Bad data:

When we encounter with the **missing values** in the cells and **leading space** in the **cells**

How to identify the missing cells in the column. Select a column and **Hold (CTRL + Shift + Down arrow).**  If the data is not available in the cells you can use **(CTRL + F) find or (CTRL + H) replace**

**Leading Space** is the Space that you have before the text and **trailing Space** is the space that you have after the text.

**For Leading Space,** we can use TRIM function select a column = **TRIM (table name[column]**) while automatically trims the data in column. If = **TRIM (@[column]) @** is for when you a applied a trim formula with in table column then @ is also used instead of table name. And we don’t need formula to be copied so that we can use paste only values **(CTRL + Alt + V)**

**Removing the duplicate values** this can be done two ways one is by selecting the column and apply conditional formatting and select Highlight selecting the rules and then select duplicate values. Second one by selecting the column range and then go to **data tab** and click the Remove Duplicates option.

**Mixing to columns** **need to separate** because of bad formatting data there is an option called Text to columns in Data tab where u need to select the column and then click the option there are two option one delimited of separation of text and one fixed with separation. Delimiter separator is need have kind value which need to separate the two texts like (dot, comma, semicolon etc) and fixed with band where we need select the data at point need to separated and value need to separated, we need to select both points. For these two methods we need to have some column because the separated content is overlaping the other column.

**Combing data from two tables VLOOK UP formula:**

The VLOOKUP works up when we want combine the data of two tables and then we can use this =VLOOKUP (lookup value, table array, col index number, range lookup)

LOOKUP value needs to the first column in other reference table then only the formula works.

Table array the reference table which we need to get data form that table.

Col index number which column number data we need to have it (col index number start from 1,2…...)

Range lookup value is true or false when we need have the exact match then we need to have False

**Definition (VLOOKUP):** It is used to look up a value in a table by searching for a corresponding value in the left most column of another table.

**Some limitations:**

* **It searches only the first column of reference table for matching value.**
* **Adding a new column to the reference table can cause errors by shifting the column indexes and changing the referenced columns.**

**Combing data from two tables (INDEXMATCH):**

We can over VLOOKUP limitations with index match.

Index function is used to find and index from a given table.

=INDEX (array (table name), row no, col number,)

=MATCH (lookup value (“dumplings”, lookup array (table or column in a table food[items], exact match 0,1)

Using Match function based on column name finding the column index number ex = MATCH(“PRICE”, FOOD[#HEADERS]) will give index number

When we are using indexmatch and we are not finding some values then we can given =IFNA((INDEXFMATCH FUNCTION VALUE, “VAULE which u to fill if no match happened”))

**Combing data from two tables (XLOOKUP):** XLOOKUP is a function in Microsoft Excel that is designed to perform flexible and powerful lookups in a worksheet. It was introduced as a replacement for older functions like VLOOKUP, HLOOKUP, and LOOKUP, providing improvements in functionality and ease of use. XLOOKUP allows you to search a range or array, find the corresponding value, and return that value to a specified cell.

The syntax of XLOOKUP function is XLOOKUP (look up value, look up array, return array, “if not found, match mode)

Lookup value is which we want look a value which is in one table and will be available in another table as well.

Return array means which we want the values column to be filled with in an another table.

Lookup array we can select same match column will be in both Table.

Unlike VLOOKUP, which only searches in the leftmost column of a table, XLOOKUP is capable of searching any column.

**Control + G** short cut we can use to fill blank cells to some value at a time select the references and select blank and type the value in the cell and Press Control and enter it will automatically fill the value in the all selected Blank cells.

=INDEX(products,MATCH(products[@[product\_name]],products[product\_name],0),MATCH(products[[#Headers],[product\_name]],products[#Headers],0))

=INDEX(products,MATCH(products[@[product\_name]],products[product\_name],0),MATCH(products[[#Headers],[product\_name]],products[#Headers],0))

**Chapter 4: Applying Business Math & Statistics:**

Business knowledge will you the better understanding on the KPI key performance indicators with combination of Domain Knowledge and Tools Experiences and Business math and Statics will help being a better Data Analyst.

Learned about different examples how we need to deal with business requirements different currency rates are changed to single currency unit and learned how to amylase the categorized data.

Business Knowledge + Business Math & Statistics = KPI [key performance indicators]

IF, SUM, SUMIF, COUNT, COUNTIF, AVERAGE

**Null Values Can be handled in different ways based on the business situation:**

* Delete the data
* Replace with the mean or median value.

**Business Matrix:**

**Profit and loss Statement:**  Every company will have the profit and loss where the revenue is more than the budget then will get profit and less than budget will go into loss.

The Profit loss % will calculate is based on budget example profit/ Budget.

Market Share will be equal to revenue / total revenue.

Mostly communicate with business language with the client it will make more effective way of communication. Ex: If we can say one third of total revenue instead of 35 %.

Targets will be set at the start of year in any company and if value you see in (11)

Will be considered as negative value in finical documents.

* Revenue – Budget = Profit/Loss
* Budget is the baseline for Calculating the profit/ loss Percentage.
* **Market Value:** Total revenue of the all the movies.
* **Market Share:** The Revenue Contribution of a particular industry to the overall market value.
* **Target:** Estimated revenue for a particular time period.
* Use conditional formatting to Highlight values that are above or below a certain threshold.

**Commonly used Statistic Concepts (Mean, Median, Mode):**

* With 20% statistics, you can cover 80% of your work.
* **Mean** – Synonym of Average.
* **Median** – Middle value of a dataset when it is ordered in ascending order.
* If the dataset has even number of values, the median is the average of two middle values.
* **Mode** – Most frequent occurring data value.

**Variance and Standard Deviation:**

* **Variance** is a measure of how spread out a distribution is. It is calculated as the **average of the squared differences from the mean.**
* The Smaller the variance, the **less spread out** the data is. Conversely, the larger the variance, the **more spread out** the data is.
* **Standard deviation** is measure of the amount of variation or dispersion of a set of values. It is calculated as the **square root** of the variance.
* The smaller the standard deviation, the closer the data points are to the mean. Conversely, the larger the standard deviation, the more spread out the data points are.
* The Stock market’s volatility is the best value case for variance and standard deviation.

**Correlation:**

Correlation is a statical measure that shows the degree to which two variables are related.

A correlation coefficient can range from -1 to 1.

* -1 (perfect Negative correlation)
* 0 (no correlation)
* 1(perfect positive correlation)

**PMT function:**  
In Excel, the **PMT function** is used to calculate the periodic payment for a loan or investment, given a fixed interest rate, a number of periods, and a present value or loan amount. The PMT function syntax is as follows:

excel Copy code

=PMT (rate, nper, pv, [fv], [type])

rate: The interest rate for each period.

nper: The total number of payment periods.

pv: The present value, or the total amount of the loan or investment.

[fv] (optional): The future value, or a cash balance you want to attain after the last payment. If omitted, it is assumed to be 0.

[type] (optional): The timing of the payment. Use 0 if payments are due at the end of the period, and use 1 if payments are due at the beginning of the period. If omitted, it is assumed to be 0.

Here's an example:

excel Copy code

=PMT (5%/12, 5\*12, -10000)

In this example:

The interest rate is 5% per year, so we divide it by 12 to get the monthly rate (5%/12).

The loan term is 5 years, so the total number of payment periods is 5 multiplied by 12 (5\*12).

The present value (loan amount) is $10,000, represented as a negative value (-10000).

This formula will calculate the monthly payment necessary to pay off a $10,000 loan over 5 years at an annual interest rate of 5%.

Remember to be consistent with the units of the rate and the number of periods. If the rate is an **annual rate**, the number of periods should be the total number of payments over the loan term in years. If the rate is a **monthly rate**, the number of periods should be the total number of payments in months.

**Total cost of a loan:**

The total cost of a loan, also known as the **total repayment amount**, can be calculated by adding up all the payments made over the life of the loan. The formula for calculating the total cost of a loan depends on the type of loan and the payment frequency.

For a simple loan with fixed monthly payments, you can use the following formula:

**Total Cost = Monthly Payment** **× Number of Payments**

If you have the monthly payment amount (PMT) calculated using the PMT function in Excel or another method, and you know the total number of payments (n), you can find the total cost.

**For example**, if your monthly payment is $500 and the loan term is 5 years (60 monthly payments), the total cost would be:

Total Cost= $500 **×** 60 = $30,000

This formula assumes that the interest rate remains constant over the life of the loan, and there are no additional fees or costs associated with the loan.

If your loan has **variable interest rates** or if there are fees associated with the loan, the calculation becomes more complex. In such cases, you may need to use a financial calculator, spreadsheet software, or specialized financial formulas to determine the total cost accurately. Additionally, it's essential to consider any upfront fees, closing costs, and other charges that may be part of the loan agreement.

**Total interest amount:**

The total interest paid on a loan can be calculated by subtracting the original loan amount (or present value) from the total amount repaid over the life of the loan. The formula for calculating the total interest is:

**Total Interest = Total Repayment Amount − Original Loan Amount**

Here's a breakdown of the components:

**Total Repayment Amount:** This is the total amount you will pay over the life of the loan, including both the principal (original loan amount) and the interest.

**Original Loan Amount (Present Value):** This is the initial amount of the loan before interest.

The formula ensures that you're capturing the total interest paid by subtracting the original loan amount from the total amount repaid.

**Chapter 5:**

**Power query:**

Power Query is a **data connectivity and preparation tool** that is part of Microsoft Power BI, Excel, and other Microsoft products. It allows users to connect to various data sources, transform and shape the data according to their needs, and load it into the desired destination for analysis and reporting.

Here are some key features and reasons why Power Query is considered a game-changer:

**Data Integration and Connectivity:** Power Query supports a wide range of data sources, including databases, Excel files, text files, web pages, and many more. This enables users to bring in data from various sources into a unified environment for analysis.

**Data Transformation and Cleaning:** Power Query provides a user-friendly interface for transforming and cleaning data. Users can easily filter, sort, pivot, unpivot, group, and perform various transformations without the need for complex coding. This helps in preparing the data for analysis and reporting.

**Data Loading and Refreshing:** Power Query enables users to load the transformed data into their desired destination, such as Power BI, Excel, or other supported platforms. Moreover, it supports automatic refreshing of the data, ensuring that the analysis is always based on the latest available data.

**Repeatability and Consistency:** Power Query allows users to create reusable and shareable queries. This promotes consistency across analyses and makes it easier to repeat the data preparation process, reducing the risk of errors and saving time.

**Advanced Capabilities:** Power Query includes advanced functionalities such as the ability to create custom functions using the M language, reference other queries, and handle complex data scenarios. This makes it a powerful tool for users with varying levels of expertise.

**Query Folding:** Power Query supports query folding, which means that it pushes certain data transformation operations back to the source database if possible. This can significantly improve performance by reducing the amount of data transferred over the network.

**User-Friendly Interface:** Power Query features a user-friendly and intuitive interface, making it accessible to a wide range of users, including those without extensive programming skills. This empowers business users to independently manage and transform their data.

Overall, Power Query simplifies and streamlines the process of data preparation, making it a game-changer for analysts, business users, and data professionals who need to work with diverse data sources and perform complex transformations for analysis and reporting.

Powery is simple to use when u click on data tab and if u see Table/ range then that option is called power query. Here we can do whether we did in excel without much bothering about the excel formulas.

* Powery query simplifies data cleaning and transformations compared to using excel formulas.
* Power query uses **M language** internally for all the steps performed using the UI controls.
* It is advisable to meaningful names to the transformations steps in power query

**(your future self will thank you for this!)**

* In power query, you can perform various types joins between tables based on your specific requirements.
* Take some time to explore all the options in power query and familiarize yourself with them
* To Quickly check the quality status of columns, use the **view** option.
* Unique values are values that appear only once in the data.
* Distinct values are values that appear at least once in the data.

**Custom Column in power query:**

A condition column allows you to add a column to your table based on a specific condition or set of conditions.

In Power Query, you can create a custom column to derive new values based on existing columns or perform calculations. Here's how you can add a custom column in Power Query:

* **Load Your Data into Power Query:**

1. Open Excel or Power BI.
2. Import your data into Power Query by going to the "Data" tab and selecting "Get Data" or "Power Query" (depending on your version).

* **Access Power Query Editor:**

1. In Excel, go to the "Home" tab, and click on "Transform Data."
2. In Power BI, go to the "Home" tab, and click on "Transform Data."

* **Add a Custom Column:**

1. Once you are in the Power Query Editor, select the table for which you want to add a custom column.
2. Go to the "Add Column" tab in the ribbon.

* **Choose a Transformation:**

1. Select "Custom Column" from the menu. This opens the "Add Custom Column" dialog.

* Enter Your Formula:

1. In the "New column name" field, enter a name for your custom column.
2. In the "Custom column formula" field, enter your formula using the M language. The M language is used in Power Query for creating custom formulas. It's similar to Excel formulas but has some differences.
3. For example, if you want to concatenate two columns "Column1" and "Column2," the formula might look like this:

= [Column1] & [Column2]

* **Click OK:**

1. After entering your formula, click "OK" to create the custom column.

* Once you've added your custom column, click "Close & Apply" to apply the changes and **Apply Changes:**

1. close the Power Query Editor.

Now, your custom column should be part of your data table. The new column will be calculated based on the formula you provided.

Remember to review and adjust your formula based on the specific requirements of your data and the transformations you want to apply. If you are familiar with Excel formulas, creating custom columns in Power Query should feel relatively intuitive.

**Custom column in power query:**

A Custom column allows you to add a new column to your table using a formula that you define.

This formula can reference other columns in your table and perform calculations or manipulations on their values.

**In Power Query Editor:**

* **Load Your Data:**

1. Open Excel or Power BI.
2. Import your data into Power Query.

* **Access Power Query Editor:**

1. In Excel, go to the "Data" tab and click "Transform Data."
2. In Power BI, go to the "Home" tab and click "Transform Data."

* Select the Table:

1. In Power Query Editor, select the table for which you want to add a custom column.

* Add Custom Column:

1. Go to the "Add Column" tab and click on "Custom Column."

* Define the Custom Column:

1. In the "New column name" field, enter a name for your custom column.
2. In the "Custom column formula" field, enter your formula using the M language. For example, if you want to concatenate two columns, the formula might look like this:

[Column1] & [Column2]

* Click OK:

1. Click "OK" to create the custom column.

* **Apply Changes:**

1. Click "Close & Apply" to apply the changes and close the Power Query Editor.

Your custom column is now part of your data. The formula you entered will be applied to each row, creating values based on the specified logic.

**Example:**

Let's say you have a table with columns "FirstName" and "LastName," and you want to create a custom column called "FullName" by concatenating these columns.

1. Select your table.
2. Go to the "Add Column" tab and click "Custom Column."
3. Enter "FullName" as the new column name.
4. In the formula field, enter: [FirstName] & " " & [LastName]
5. Click "OK" and then "Close & Apply."

Now, you'll have a new column called "Full Name" that combines the first and last names in each row.

Remember to adjust the formula based on your specific requirements and the columns you are working with.

**Pivot tables:**

* As a data analyst it may be difficult to survive without pivot tables.
* You can insert a pivot table in the same worksheet or in a new worksheet.
* pivot tables make it easy to summarize data.
* Pivot tables allow you to slice and dice data in any format you want.

A PivotTable in Excel is a powerful tool that allows you to summarize and analyse large amounts of data in a dynamic and interactive way. It helps you organize, reorganize, and visualize your data, making it easier to draw insights and make informed decisions. Here's a brief introduction to PivotTables in Excel:

**What is a PivotTable?**

A PivotTable is a **data processing tool** in Excel that allows you to quickly summarize and analyse large datasets. It enables you to transform raw data into a more structured and organized format, making it easier **to identify trends, patterns, and relationships within the data**.

**How to Create a PivotTable:**

* **Prepare Your Data:**

1. Ensure your data is organized in a tabular format with headers for each column.
2. Include meaningful labels for clarity.

* **Select Your Data:**

1. Highlight the range of cells containing your data.

* **Insert a PivotTable:**

1. Go to the "Insert" tab on the Excel ribbon.
2. Click on "PivotTable" and select the range of cells you highlighted.

* **Design Your PivotTable:**

1. In the PivotTable Field List, you have four areas: **Rows, Columns, Values, and Filters**.
2. Drag and drop fields from your dataset into these areas to define how you want to summarize and arrange your data.

**Key Components of a PivotTable:**

* **Rows and Columns:**

1. Determine the organization of your data. For example, you might want to see sales data by product categories (rows) and years (columns).

* **Values:**

1. Specify the data you want to analyse, such as sum, count, average, etc. For example, you might want to see the total sales or average prices.

* **Filters:**

1. Apply filters to focus on specific subsets of your data. This could be helpful when you want to analyse a particular region or time period.

**Interactivity and Dynamism:**

* **PivotTables offer a dynamic environment:**

1. You can easily rearrange and reorganize your data by dragging and dropping fields.
2. Filter and drill down into specific details for deeper analysis.
3. Quickly update your PivotTable as you modify the underlying data.

**Conclusion:**

PivotTables are a versatile tool for data analysis in Excel, providing a user-friendly way to organize and analyse large datasets. With their dynamic nature, you can gain valuable insights and present your data in a more digestible format. As you become more familiar with PivotTables, you'll discover their ability to streamline complex data analysis tasks in Excel.

**Power Pivot:**

* Power Pivot is data modelling tool for Microsoft Excel. It allows users to create relationships between tables and perform **advanced calculations** with large amount of data.
* To create measures, we can use the DAX (Data analysis Expression) language.
* **Primary key:** It is a column or combination of columns that uniquely identifies each row in table.
* **Foreign Key:** It is a column or combination of columns that refer to the primary of another table and establishes a link between the two tables.

Power Pivot is an add-in for Microsoft Excel that allows you to import, manipulate, analyse, and visualize large datasets quickly and easily. Power Pivot is particularly useful when working with multiple related tables of data. One of the powerful features of **Power Pivot is its ability to create PivotTables based on multiple tables**, providing more advanced data analysis capabilities.

Here's a basic guide on how to create a Power Pivot table in Excel:

**Step 1: Enable Power Pivot**

* Open Excel and go to the "File" tab.
* Click on "Options" to open the Excel Options dialog box.
* In the Excel Options dialog box, click on "Add-ins" in the left sidebar.
* In the "Manage" box at the bottom, select "COM Add-ins" from the dropdown menu and click on "Go..."
* Check the "Microsoft Office Power Pivot" box and click "OK."

**Step 2: Load Data into Power Pivot**

* After enabling Power Pivot, you'll see a new tab in the Excel ribbon called "Power Pivot." Click on it.
* Click on the "Add to Data Model" button to import data into the Power Pivot Data Model.

**Step 3: Create Relationships**

* In the Power Pivot tab, click on "Diagram View."
* Drag and drop tables from the Tables list to the Diagram view.
* Create relationships between tables by dragging fields from one table to the corresponding field in another.

**Step 4: Create a PivotTable**

* Go back to the Excel sheet and click anywhere in your dataset.
* Go to the "Insert" tab and click on "PivotTable."
* In the Create PivotTable dialog box, ensure that you select "Use this workbook's Data Model."
* Choose where you want to place the PivotTable and click "OK."

**Step 5: Build the PivotTable**

* In the PivotTable Field List, you'll see fields from all the tables in the Data Model.
* Drag and drop fields into the Rows and Values areas to build your PivotTable.

**Step 6: Use Power Pivot Features**

* You can use DAX (Data Analysis Expressions) formulas to create calculated columns and measures.
* Explore the PivotTable and chart features to analyse and visualize your data effectively.

Power Pivot provides advanced features for data analysis, and this is just a basic overview. If you're new to Power Pivot, it's recommended to explore tutorials and documentation to fully leverage its capabilities.

Use the SUMX([table],[Expression])function to add up the values in a table based on specified Expression.

Chapter 8:

ETL stands for Extract, transform, load. It refers to the process of extracting data from various sources, transforming the data to fit the desired data model or format, and loading the data into target data store or destination.

ETL stands for Extract, Transform, Load, and it refers to a process in data warehousing and data integration. Here's a brief overview:

**Extract**: This is the first step where data is gathered from various sources. It could be databases, logs, applications, or any other data storage systems.

**Transform**: Once the data is extracted, it often needs to be cleaned, transformed, and converted into a suitable format. This step involves cleaning up inconsistencies, handling missing data, and transforming the data into a structure that suits the target system or database.

**Load**: After the data is transformed, it is loaded into the target system, which could be a data warehouse, a database, or any other data repository. This step involves putting the data in the right tables, columns, and rows.

The ETL process is crucial for ensuring that data from different sources can be effectively analysed and used for reporting and decision-making purposes.

Power query transformations and connects data from different sources such as data bases and spreadsheets, into a format suitable for analysis and reporting.

Power Query is a data connection technology that enables you to discover, connect, combine, and refine data across a wide variety of sources. In Microsoft Excel, Power Query is a feature that helps you import, transform, and combine data from different sources before it's loaded into your workbook. It's part of the Power BI suite of tools and is available in Excel 2010 and later versions.

Here's a basic overview of how you can use Power Query in Excel:

* **Get Data:**

1. Open Excel and go to the "Data" tab.
2. Click on "Get Data" or "Get & Transform Data" (the exact wording may vary depending on your version of Excel).

* **Connect to a Data Source:**

1. Choose the source from which you want to import data. Power Query supports a wide range of sources, including Excel files, databases, web pages, and more.

* **Transform Data:**

1. After connecting to a data source, you can use the Power Query Editor to transform the data. This involves cleaning, filtering, and shaping the data to suit your needs.
2. The Power Query Editor provides a user-friendly interface with various tools for data manipulation.

* **Load Data:**

1. Once you've transformed the data, you can load it into your Excel workbook.
2. You can choose to load the data directly into a worksheet or create a connection that you can refresh later.

* **Refresh Data:**

1. If your source data changes, you can refresh the data in Excel with the latest changes without going through the entire import process again.

Power Query has a powerful formula language called "M," which you can use for more advanced data transformations. Additionally, Power Query can be used not only in Excel but also in Power BI, Power Automate, and other Microsoft products.

In power quey in view section by using column distribution and column quantity we can most of the things in data when we had a unique field that the distinct and unique values should be same.

Data cleaning in power quey involves the process of identifying and correcting inaccuracies, inconsistencies, and errors in data to ensure that it is accurate and usable for analysis and reporting.

Giving proper names to power query steps can make your queries more organized and easier to understand.

Data cleaning in Power Query within Excel involves using various transformation steps to clean, filter, and shape your data before loading it into your workbook. Here are some common data cleaning tasks you can perform using Power Query:

1. **Removing Duplicates:**
   * Use the "Remove duplicates" feature to eliminate duplicate rows from your dataset.
2. **Filtering Rows:**
   * Use filters to exclude or include specific rows based on certain criteria.
3. **Handling Null or Missing Values:**
   * Use the "Fill" or "Replace Values" options to handle null or missing values in your dataset.
4. **Changing Data Types:**
   * Ensure that columns have the correct data type. For example, you might want to change a column containing dates or numbers to the appropriate data type.
5. **Splitting and Merging Columns:**
   * If a column contains combined information, you can split it into separate columns. Conversely, you can merge columns if needed.
6. **Text and Case Transformations:**
   * Convert text to lowercase, uppercase, or proper case using the available transformation options.
7. **Trimming Spaces:**
   * Remove leading or trailing spaces from text values using the "Trim" option.
8. **Handling Errors:**
   * Use error handling options to manage and correct errors in your data.
9. **Creating Custom Columns:**
   * Add new columns to your dataset using custom formulas or expressions.
10. **Conditional Columns:**
    * Create columns with values based on specific conditions using the "Add Conditional Column" feature.
11. **Data Reordering:**
    * Change the order of columns to organize your data more effectively.
12. **Scaling or Normalizing Data:**
    * Apply mathematical operations to scale or normalize numeric values.

Here's a step-by-step example of how to perform some of these tasks in Power Query:

1. **Get Data:**
   * Open Excel, go to the "Data" tab, and select "Get Data" or "Get & Transform Data."
2. **Connect to a Data Source:**
   * Choose your data source and load the data into the Power Query Editor.
3. **Transform Data:**
   * Use the various transformation options in the Power Query Editor to clean and shape your data.
4. **Load Data:**
   * Once you're satisfied with the transformations, load the cleaned data into your Excel workbook.
5. **Refresh Data:**
   * If your source data changes, you can refresh the data in Excel to apply the latest changes.

Remember to preview the changes in the Power Query Editor before loading the data to ensure that your cleaning operations produce the desired results.

Components for 8th chapter exercise:

* Net sales
* Year
* Division
* Country
* Region

##### **Creating Connections Among Tables Using Data Modelling:**

Data model Involves the process of connecting different data sets together by establishing the relationship between them.

In Excel, a data model refers to a collection of tables and their relationships used to analyse and visualize data. The data model is a feature available in Excel that allows you to create relationships between multiple tables, perform more advanced calculations, and create PivotTables and Pivot Charts based on these relationships. Here's a basic guide on how to work with a data model in Excel:

**Creating a Data Model:**

1. **Prepare Your Data:**
   * Organize your data into tables. Each table should have a unique column that can be used to establish relationships with other tables.
2. **Create Tables:**
   * Select the data range.
   * Go to the "Insert" tab and click "Table" or use the shortcut **Ctrl + T**. Ensure the "Create Table" dialog box correctly identifies the data range.
3. **Name Your Tables:**
   * After creating the table, give it a meaningful name in the "Table Tools Design" tab. A good table name should reflect the data it contains.
4. **Repeat for Other Tables:**
   * If you have multiple sets of data, repeat the process to create tables for each.
5. **Define Relationships:**
   * Go to the "Data" tab, and click "Relationships" in the "Connections" group.
   * In the Relationships dialog box, click "New" and define the relationships between tables.

**Using the Data Model:**

1. **Create PivotTables:**
   * Go to the "Insert" tab and click "PivotTable."
   * In the Create PivotTable dialog box, select "Use this workbook's Data Model."
2. **Build Relationships in PivotTable Fields:**
   * Drag fields from different tables to the Rows and Columns areas of the PivotTable to create relationships.
3. **Create Measures (Calculated Fields):**
   * In the PivotTable, you can create calculated fields, also known as measures, using Data Analysis Expressions (DAX) to perform advanced calculations.
4. **Create PivotCharts:**
   * Use the PivotTable to create PivotCharts that represent the data visually.
5. **Refresh Data:**
   * If your source data changes, refresh the PivotTable to update the results.

**Tips for Working with Data Models:**

* **Learn DAX:**
  + Understanding Data Analysis Expressions (DAX) is crucial for creating more complex calculations.
* **Data Cleaning:**
  + Ensure your data is clean and well-organized before creating tables and relationships.
* **Relationships:**
  + Clearly define relationships between tables to ensure accurate results.
* **Performance Considerations:**
  + Large datasets may impact performance. Be mindful of the amount of data in your data model.

The start schema is a data modelling technique that arranges data into a central fact table, surrounded by several dimensions’ tables.

##### **Adding a Date Table Using Power Query:**

**Fiscal year:**

A fiscal year is a period of 12 consecutive months chosen to start at the beginning of an organization's accounting year. Unlike the calendar year, which begins on January 1 and ends on December 31, a fiscal year can start and end at any point during the year.

Organizations often choose a fiscal year that aligns with their business cycle, revenue recognition, or industry norms. The fiscal year is used for financial reporting, budgeting, and planning purposes. The choice of when to start the fiscal year can be influenced by various factors, including seasonality, regulatory requirements, and the nature of the business.

Common fiscal year designations include:

* **Calendar Year (CY):** January 1 to December 31.
* **Fiscal Year (FY):** Any 12-month period that an organization uses for budgeting, financial reporting, and accounting purposes.
* **Academic Year:** Used by educational institutions and often starts in the late summer or early fall and ends in the spring or early summer.
* **Government Fiscal Year:** Many governments use fiscal years that don't align with the calendar year. For example, in the United States, the federal government's fiscal year runs from October 1 to September 30.

To represent fiscal years in a data model or analysis, you may need to create a Fiscal Year dimension or attribute. This dimension typically includes information such as the fiscal year's start date, end date, and any additional attributes like quarters or months.

In some cases, organizations might have variations of fiscal years, such as a fiscal year that aligns with a retail season or a manufacturing cycle. Understanding and using fiscal years correctly is important for accurate financial analysis, forecasting, and planning.

Top of Form

**Calculate function in data model:**

In the context of a data model, especially within tools like Microsoft Excel's Power Pivot or Power BI, the **CALCULATE** function is a powerful DAX (Data Analysis Expressions) function used for manipulating the context in which a calculation is performed. It allows you to change filters, modify context transitions, and combine or nest calculations.

Here's a general structure of the **CALCULATE** function in DAX:

DAXCopy code

CALCULATE(<expression>, <filter1>, <filter2>, ...)

**<expression>** is the measure or expression that you want to evaluate.

**<filter1>, <filter2>, ...** are optional filters that you want to apply.

For example, if you have a measure called **TotalSales** and you want to calculate the total sales for a specific product category, you might use **CALCULATE** like this:

DAXCopy code

CALCULATE([TotalSales], 'Product'[Category] = "Electronics")

In this example, **CALCULATE** is used to modify the context of the **TotalSales** measure, applying an additional filter where only the "Electronics" category is considered.

Here's a breakdown of how you might use **CALCULATE**:

**Modify Filter Context:** You can use **CALCULATE** to change the filter context for a calculation. For instance, you might want to calculate a measure within a specific date range or for a certain category.

DAXCopy code

CALCULATE([TotalSales], 'Date'[Year] = 2023)

**Use Multiple Filters:** You can apply multiple filters within **CALCULATE** to further refine your calculations.

DAXCopy code

CALCULATE([TotalSales], 'Date'[Year] = 2023, 'Product'[Category] = "Electronics")

**Nested Calculations:** You can nest **CALCULATE** functions to create more complex calculations.

DAXCopy code

CALCULATE( CALCULATE([TotalSales], 'Date'[Year] = 2023),

'Product'[Category] = "Electronics")

Keep in mind that the effectiveness of **CALCULATE** depends on your data model and the specific requirements of your analysis. It's a fundamental function in DAX and is crucial for creating dynamic and context-aware calculations in Power Pivot and Power BI.

**User Empathetic Report Design:**

Designing user Empathetic Reports requires giving significant consideration to user **readability** and **time** **to Action**.

* **User Readability**
* **Time to action**

A **P&L (Profit and Loss statement)** is a finical report that provides an overview of a company’s financial performance over a period of time, typically a month, quarter or a year.

**P&L statement**s include several critical metrices such as net sales, Cost of Goods Sold (COGS), Gross Margin, Gross Margin %, etc. These matrices evaluate a company’s Finical performance profitability and pricing tactics.

**FORMAT Function:**

In Microsoft Excel, the **FORMAT** function is used to format the value of a cell or a formula result as text, using a specified format code. This function is particularly useful for changing the display of dates, numbers, or other values.

The syntax of the **FORMAT** function is as follows:

**=FORMAT (value, format\_text)**

**value**: This is the value that you want to format.

**format\_text**: This is the format code that specifies how you want the value to be displayed.

Here are a few examples:

**Format a Date:**

**=FORMAT (value, format\_text)**

This formula formats the date in cell A1 to display in the "dd-mmm-yyyy" format.

**Format a Number:**

**=FORMAT (B1, "$#,##0.00")**

This formula formats the number in cell B1 as currency with two decimal places.

**Format a Percentage:**

**=FORMAT (C1, "0.00%")**

This formula formats the value in cell C1 as a percentage with two decimal places.

Remember to replace A1, B1, C1, and the format\_text with the actual cell references and format codes that match your specific data and formatting needs.

Keep in mind that the **FORMAT** function is available in Excel 2013 and later versions. If you are using an earlier version, you might need to use other formatting functions or custom formatting options available in those versions.

**MONTH function:**

In Microsoft Excel, the MONTH function is used to extract the month from a given date. The syntax of the MONTH function is as follows:

=MONTH(serial\_number)

serial\_number: This is the date from which you want to extract the month.

Here's a simple example:

Assuming cell A1 contains a date, you can use the MONTH function like this:

=MONTH(A1)

This formula will return the month as a number (1 for January, 2 for February, and so on) corresponding to the date in cell A1.

If you want the result to be displayed as the name of the month, you can combine the MONTH function with the TEXT function:

=TEXT(A1, "mmmm")

This formula will return the full name of the month for the date in cell A1.

Remember to replace A1 with the actual cell reference containing your date.

Here's a breakdown of the functions used:

MONTH: Extracts the month as a number from a date.

TEXT: Converts a value to text using a specified format. In this case, "mmmm" is the format code for the full month name.

Feel free to customize the formulas based on your specific needs and the format in which you want to display the month.

**ROUNDUP Function:**

The **ROUNDUP** function in Excel is used to round a number away from zero, towards the next highest number. The syntax of the **ROUNDUP** function is as follows:

=ROUNDUP(number, num\_digits)

**number**: This is the number you want to round up.

**num\_digits**: This is the number of digits to which you want to round.

Here's a simple example:

Assuming cell A1 contains a decimal number, you can use the **ROUNDUP** function like this:

=ROUNDUP(A1, 0)

This formula will round the number in cell A1 to the nearest whole number, rounding up if the decimal part is 0.5 or greater.

If you want to round to a specific number of decimal places, you can adjust the **num\_digits** argument. For example, to round to two decimal places:

=ROUNDUP(A1, 2)

This formula will round the number in cell A1 to two decimal places, rounding up if the third decimal place is 0.5 or greater.

Here are a few key points to keep in mind:

If **num\_digits** is positive, the number is rounded to that many decimal places to the right of the decimal point.

If **num\_digits** is 0, the number is rounded to the nearest integer.

If **num\_digits** is negative, the number is rounded to the left of the decimal point.

Feel free to adjust the function based on your specific rounding requirements.