Welcome to world EXCEL

In Data Industry

1 — Data Analyst

2 — Business Analyst

3 — Data Scientist

Outside Data Industry

1 — Financial Analyst

2 — Demand Planner

3 — Supply Chain Analyst

4 — Accountant

5 — Operations Analyst

Excel Industry Use Cases

1.) Generate Management Reports

2.) Tracking Tool

3.) Project Management Tool

4.) As a Database

Module 0 - EXCEL BASICS

Takeaways

- Excel building blocks include sheets, cells, rows, columns, the ribbon menu, and the formula bar.
- An Excel workbook is saved with the extension **'.xlsx'.**
- The most commonly used operations in Excel are **filtering, sorting,** and **conditional formatting.**

Takeaways

- For writing formula in a cell start with **“=”** sign
- Use **Format painter** tool, to copy paste formats
- Formulas covered - **SUM(), AVERAGE(), IF(), SUMIF()**

Takeaways

- Locking referenced cells in a formula **helps prevent dynamic changes** when the formula is copied or moved to another cell.
- **To lock a column or row reference,** put a **'$'** in front of it. You can also use the **F4 key** to do the same.
- You have the **flexibility to lock** either rows or columns or both.

Takeaways

- **Named Ranges** allow calculations using descriptive names rather than cell references.
- Formulas will be automatically applied to **new rows.**
- Additional features such as **slicers, automatic totals,** and **filters** are available.
- **Table headers remain visible** when scrolling down, which is useful for large tables.

Takeaways

- 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 charts.
- The **UNIQUE()** function extracts unique values from a range or array.

Cell Formatting

Security

Takeaways

- **Merge & Center** option combines multiple cells into one cell and centers the content horizontally within the merged cell.
- **Format Painter** quickly copies and applies cell formatting to other cells or ranges in your spreadsheet.
- **Format Cells** dialog box helps you to customize cell appearance and behavior in your spreadsheet.
- **Wrap text** lets you display the cell contents over multiple lines instead of one.
- Excel provides several **security features** to help protect your spreadsheet from unauthorized access or modification.

Module 1- Data Cleaning - v + x lookup INDEX & MATCH

XLOOKUP

vs

VLOOKUP

S.No.	Name	Math	Language	History	Physics
1	Alex	45	62	100	99
2	Aron	30	90	66	72
3	Aubrey	99	70	51	80
4	Calista	82	52	74	66
5	Chase	92	96	90	58
6	Claris	38	43	52	97
7	Curt	59	76	72	50
8	Delaine	50	37	30	31
9	Enoch	34	35	51	36
10	Florene	45	77	32	87

Aron

Subject	Marks
Language	90
Math	30
History	66
Physics	72

Takeaways

- **Data cleaning** is an important step in the data analysis process because it ensures that the data is accurate and reliable.
- The **TRIM()** function eliminates extra spaces in the text, including those at the start and end.
- **Conditional Formatting** helps to identify and highlight duplicate data in a range of cells.
- The **"Text to Columns"** feature helps to split a single cell containing text into multiple cells based on a specified delimiter, such as a comma, semicolon, or space.

Takeaways

- **VLOOKUP** function is used to look up a value in a table by searching for a corresponding value in the leftmost column of another table.
- Syntax of the **VLOOKUP()** function is: **VLOOKUP (lookup_value, table_array, col_index_num, [range_lookup])**
- **Some limitations of VLookup:**
  - It searches only the first column of a reference table for a matching value.
  - Adding a new column to the reference table can cause errors by shifting column indexes and changing the referenced columns.

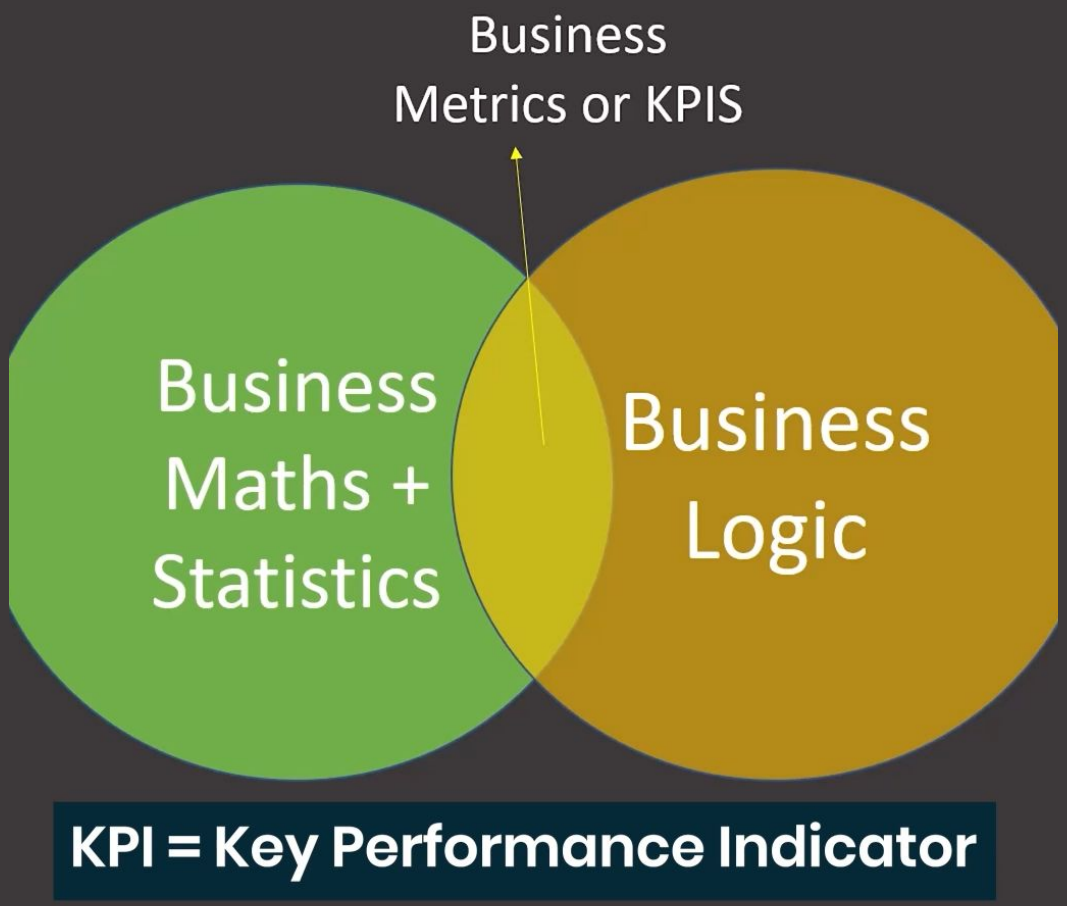
Takeaways

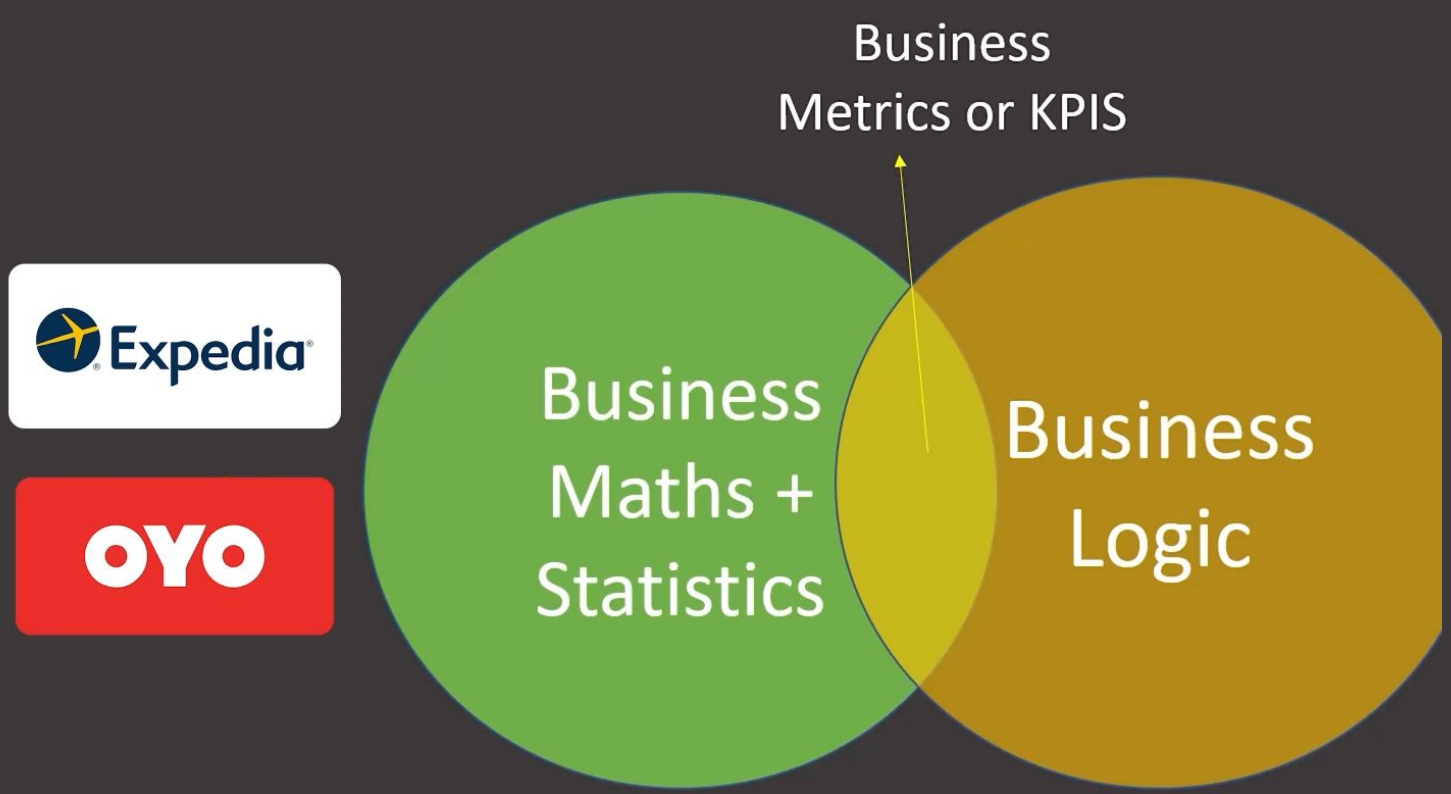
- **INDEX()** and **MATCH()** are two powerful Excel functions that are often used together to search for and retrieve data from a table or range.
- The syntax of the **INDEX()** function is: **INDEX(array, row_num, [column_num]).**
- The syntax of the **MATCH()** function is: **MATCH(lookup_value, lookup_array, [match_type]).**

Takeaways

- The syntax of the **XLOOKUP()** function is: **XLOOKUP(lookup_value, lookup_array, return_array, [if_not_found], [match_mode]).**
- Unlike **VLOOKUP,** which only searches in the leftmost column of a table, **XLOOKUP** is capable of searching in any column.

Module 2- Business Math and Statistics

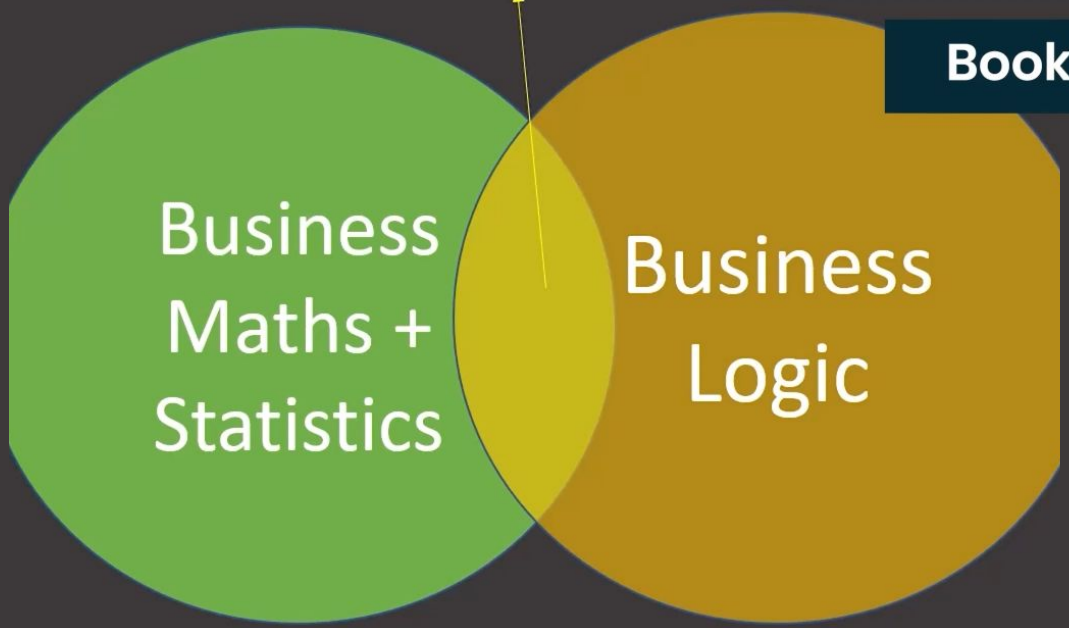

Business
Metrics or KPIS
Business
Maths +
Statistics
Business
Logic
KPI = Key Performance Indicator

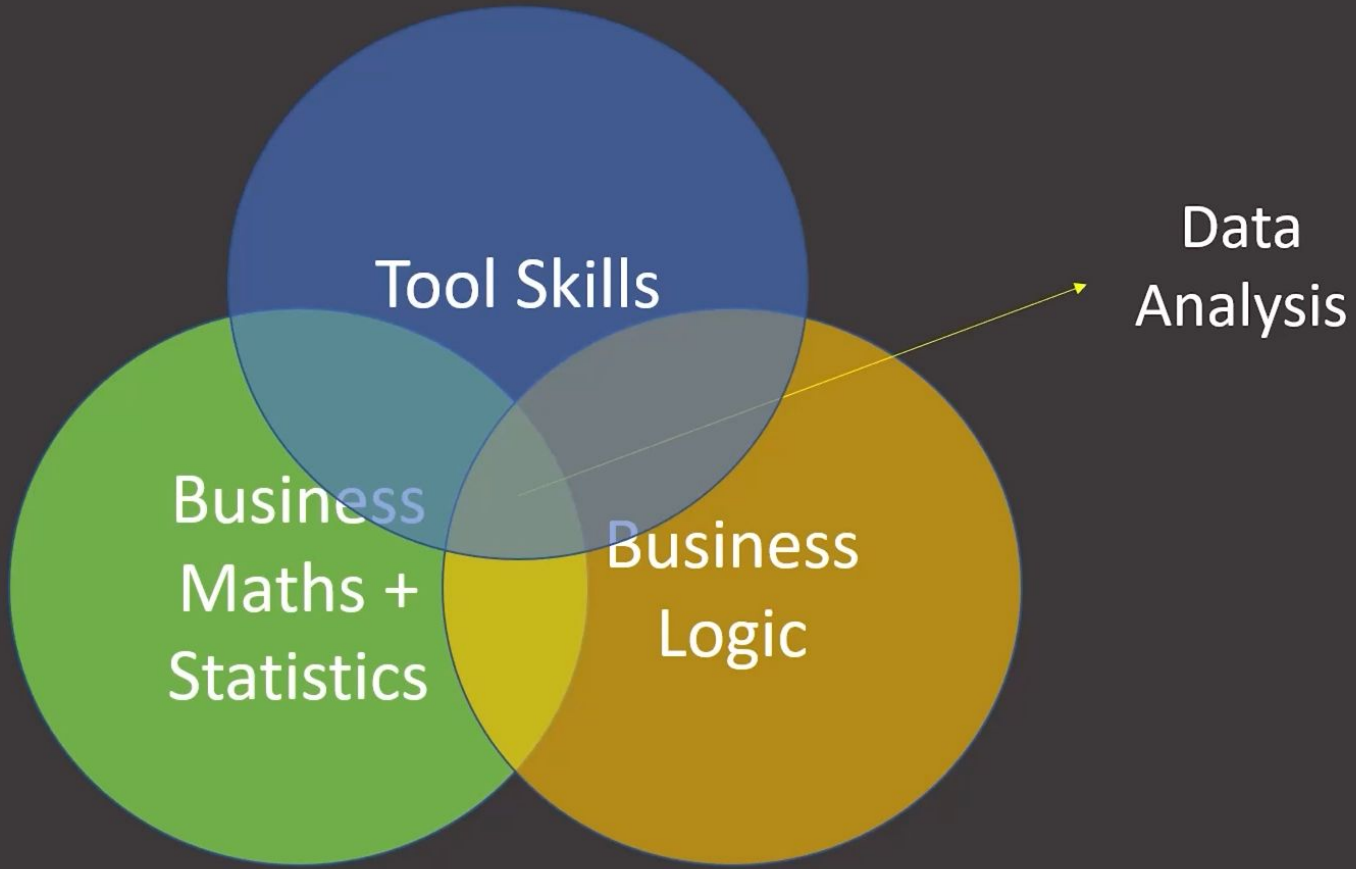

Business
Metrics or KPIS
Business
Maths +
Statistics
Business
Logic
Expedia
OYO

Working of Hotel Business

What is Occupancy Rate

Booking Capacity




Tool Skills
Business
Maths +
Statistics
Business
Logic
Data
Analysis

Takeaways

- Business Knowledge + Business Math & Statistics = **KPIs (Key Performance Indicators)**
- Formulas covered: **IF, SUM, SUMIF, COUNT, COUNTIF, AVERAGE**
- **Null values** can be handled in different ways based on the business situation:
  - Delete that data
  - Replace with the mean or median value

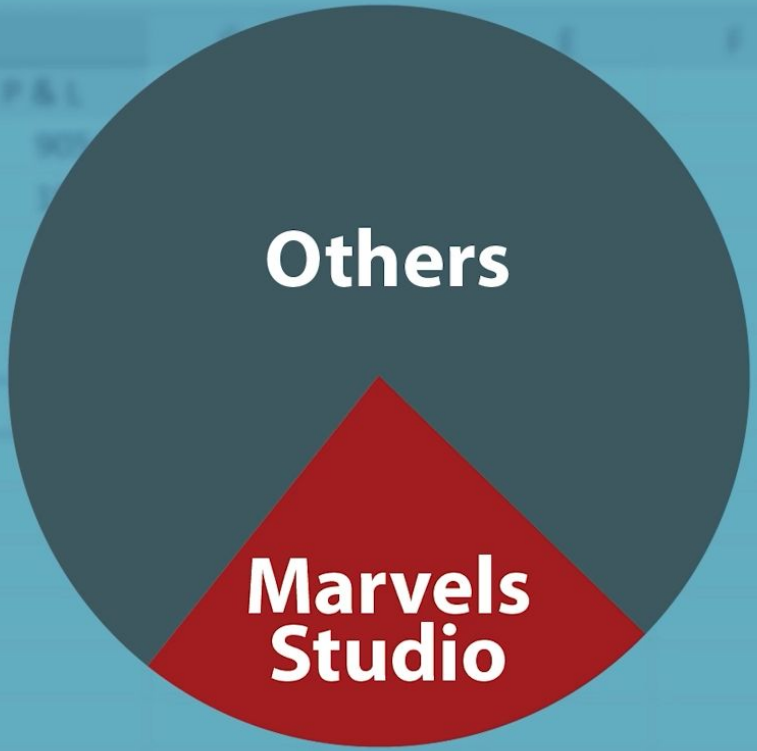
Business Metrics

MARVEL


MARVEL STUDIOS
AVENGERS
ENDGAME
APRIL 26


ANT-MAN AND THE WASP
QUANTUMANIA
FEBRUARY 17


MARVEL STUDIOS
THOR


Others
Marvels Studio

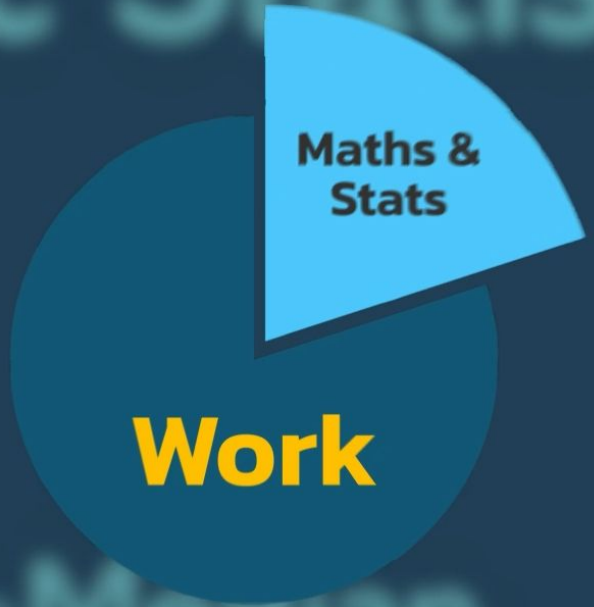
Basic Statistics

Mean

Median

Mode

Basic Statistics



Median

Mode

Name	Monthly Income ($)
Rob	5000
Rafiq	6000
Nina	4000
Sofia	7500
Mohan	8000
Tao	7000

Average	6250

Name	Monthly Income ($)
Rob	5000
Rafiq	6000
Nina	4000
Sofia	7500
Mohan	8000
Tao	7000

Average	6250

Name	Monthly Income ($)
Rob	5000
Rafiq	6000
Nina	4000
Sofia	7500
Mohan	8000
Tao	7000
Elon Musk	10 million

Name	Monthly Income ($)
Rob	5000
Rafiq	6000
Nina	4000
Sofia	7500
Mohan	8000
Tao	7000

Average	6250

Name	Monthly Income ($)
Rob	5000
Rafiq	6000
Nina	4000
Sofia	7500
Mohan	8000
Tao	7000
Elon Musk	10 million

Name	Monthly Income ($)
Rob	5000
Rafiq	6000
Nina	4000
Sofia	7500
Mohan	8000
Tao	7000

Average	6250

Name	Monthly Income ($)
Rob	5000
Rafiq	6000
Nina	4000
Sofia	7500
Mohan	8000
Tao	7000
Elon Musk	10 million

Nina	Rob	Rafiq	Tao	Sofia	Mohan	Elon Musk
4000	5000	6000	7000	7500	8000	10 million

Median = 7000

Nina	Rob	Rafiq	Tao	Prem	Sofia	Mohan	Elon Musk
4000	5000	6000	7000	8000	7500	8000	10 million

Median = 7500

What is Mode?

Name	Restaurant Vote
Rob	Mexican
Rafiq	Mexican
Nina	Italian
Sofia	Thai
Mohan	Italian
Tao	Mexican
Bantu	Indian

Mode here is Mexican

Mode means most frequently occurring value in a dataset

Takeaways

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

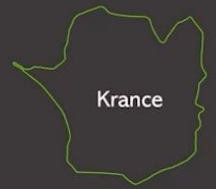
Variance

How far each number is from every other number in a dataset


Juganda
Krance

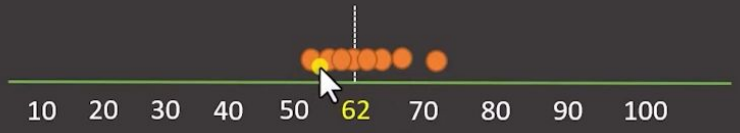
name	yearly income
nishith	71
veeral	62
angelina	66
salma	61
Nitin	54
dhaval	67
venkat	55
Priya	60

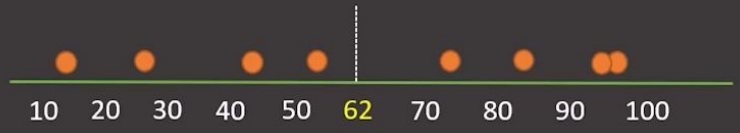
Mean = 62



name	yearly income
mohan	99
rita	14
bhavin	75
michael	84
abdul	44
kiran	54
ahmed	98
gayatri	28

Mean = 62





name	yearly income
nishith	71
veeral	62
angelina	66
salma	61
Nitin	54
dhaval	67
venkat	55
Priya	60

name	yearly income
mohan	99
rita	14
bhavin	75
michael	84
abdul	44
kiran	54
ahmed	98
gayatri	28

why do we square the difference in variance formula?

In the formula for variance, we calculate the average of the squared differences between each data point and the mean. The squared differences are used because they have several useful properties, such as:

1. Squaring the differences ensures that they are all positive. If we didn't square them, then the sum of the differences would always be zero, since some differences would be positive and others negative, and they would cancel each other out. By squaring them, we eliminate the negative signs and ensure that all the differences are positive.
2. Squaring the differences gives more weight to larger deviations from the mean. This is because the squared value of a large deviation is much greater than the squared value of a small deviation. By giving more weight to the larger deviations, we get a better measure of how spread out the data is.
3. Squaring the differences allows us to use calculus to find the minimum variance. Calculus requires us to differentiate the variance formula with respect to the mean, which would not be possible if we did not use squared differences.

$$variance = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\mathbb{N}}$$

$x_i = each\ value\ in\ a\ dataset$

$\bar{x} = average$

$\mathbb{N} = total\ data\ points$

Takeaways

- 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 a 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 use case for variance and standard deviation.


Discount
Unit Price
Profit
Sales
Costs
1
0
-0.02
-0.02
-0.01
0
1
0.2
0.53
0.53
-0.02
0.2
1
0.53
0.19
-0.02
0.53
0.53
1
0.93
-0.01
0.53
0.19
0.93
1
Discount
Unit Price
Profit
Sales
Costs
-1
1

Takeaways

- **Correlation is a statistical 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)**

Module 3- Power Query

Power Query

Why Power Query is the Game Changer

1.Much Quicker than Regular Excel

2.Handle Bigger & Multiple Datasets

3.Bridge to Power BI & Advanced Business Analytics

4.Can Reduce VBA Significantly

Takeaways

- Power Query simplifies data cleaning and transformation compared to using Excel formulas.
- Power Query uses **M-language** internally for all the steps performed using the UI controls.
- It is advisable to give meaningful names to the transformation steps in Power Query. (Your future self will thank you for this!)

Takeaways

- In **Power Query**, you can perform various types of 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.

Takeaways

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

Module 4 - Pivot Tables

Takeaways

- As a Data Analyst, it may be difficult to survive without **Pivot tables.**
- You can insert a Pivot table either 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.

Takeaways

- To create a pivot table, follow these steps:
  Insert > Pivot Table > Select the table/range > Create Pivot Table
- You can use the recommended pivot table for several use cases.
- Make sure to explore all other options available in the pivot table.

Takeaways

- Pivot tables allow you to sort data based on different values.
- You can make pivot tables more presentable by using **formatting** and other options.
- Pivot tables are **very similar to the matrix visualizations** that you will use in BI tools.