

**A
INTERNSHIP REPORT
on
“PRODUCT SALES DASHBOARDS”**

A report submitted in partial fulfillment of the requirements for the Award of Degree of

BACHELOR OF ENGINEERING

in

INFORMATION TECHNOLOGY

by

Ms. Suprabha Prakash Potinde

Under Supervision of

Mr. Vijay Shirsath, CEO

Softercrowd Technologies, Nashik.

(Duration: 20th Dec, 2024 to 23th Jan, 2025)



DEPARTMENT OF INFORMATION TECHNOLOGY

JAWAHAR EDUCATION SOCIETY'S

INSTITUTE OF TECHNOLOGY, MANAGEMENT & RESEARCH, NASHIK

(Approved by AICTE, New Delhi, DTE, Government of Maharashtra, Affiliated to Savitribai Phule Pune University, Pune)

2024 – 2025

DEPARTMENT OF INFORMATION TECHNOLOGY

JAWAHAR EDUCATION SOCIETY'S

INSTITUTE OF TECHNOLOGY, MANAGEMENT & RESEARCH, NASHIK



CERTIFICATE

This is to certify that the “**Internship report**” submitted by **Ms. Suprabha Prakash Potinde** is work done by her and submitted during 2024 – 2025 academic year, in partial fulfillment of the requirements for the award of the degree of **BACHELOR OF ENGINEERING IN INFORMATION TECHNOLOGY**, at **Jawahar Institute of Technology and Research, Nashik**.

Internship Coordinator
(Ms. D. D. Survase)

Head of Department
(Prof. S. B. Patil)

Internship Completion Certificate



Changes for the Better
www.softcrowdtechnologies.com

Date: 27th January, 2025

Internship Completion Letter

*This is to ensure that Miss. Suprabha Prakash Potinde has done her temporary position for a **1 Month** as a **Data Science Trainee Intern** at SoftCrowd Technologies, Nashik, Maharashtra from **20/12/2024 to 23/01/2025**.*

During the temporary position she showed great abilities with a self-persuaded to learn new things. Her presentation surpassed assumptions and had the option to finish the tasks effectively on schedule.

We want her, to enjoy all that life has to offer for future undertakings.



Sincerely,

Mr. Vijay Shirsath
(Director)
SoftCrowd Technologies

Regn. Office : 2nd Floor, Kaveri Sankul, Wakilwadi Corner, Ashok Stambh, Nashik-422001.
E-mail : softcrowdtech@gmail.com | Mob. : 7447444874

ACKNOWLEDGEMENT

First I would like to thank **Mr. Vijay Shirsath , CEO, Head of Softcrowd Technologies** for giving me the opportunity to do an internship within the organization.

I also would like all the people that worked along with me of **Softcrowd Technologies** with their patience and openness they created an enjoyable working environment.

It is indeed with a great sense of pleasure and immense sense of gratitude that I acknowledge the help of these individuals.

I am highly indebted to Director Principal **Dr. M.V Bhatkar**, for the facilities provided to accomplish this internship.

I would like to thank my Head of the Department **Prof. S. B. Patil** for his constructive criticism throughout my internship.

I would like to thank **Ms. D. D . Survase** internship coordinator Department of Information Technology for their support and advices to get and complete internship

I am extremely great full to my department staff members and friends who helped me in successful completion of this internship.

SUPRABHA PRAKASH POTINDE

ABSTRACT

Programmers and opportunities:

The Institute combines pioneering research with top class education. An innovative curriculum allows the student flexibility in selecting courses and projects. Students, even at the undergraduate level, get to participate in on-going research and technology development - an opportunity unprecedented in India. As a result, a vibrant undergraduate programmer co-exists with a strong postgraduate programmer.

Methodology:

This section can help stakeholders understand the process behind the data analysis and visualization. Clearly state the main objectives of the report. What questions are you trying to answer? What insights are you aiming to provide? Describe any transformations applied to the data, including merging tables, pivoting/unpivoting, and creating calculated columns. Discuss the types of visualizations chosen (e.g., bar charts, line graphs, maps) and the rationale behind these choices. Describe the layout and design principles used to create an intuitive and user-friendly dashboard. This methodology section provides a comprehensive overview of the processes involved in creating the Power BI report. It ensures transparency and helps stakeholders understand the rigor behind the data analysis and visualization efforts.

Key parts of the report:

In existed system, users have to go that particular car and user should reserve it. Here, time consuming is more and there is no guarantee that car will come after waiting for long time. We will not able know the fair details of the distance of our journey

Benefits to the company / institution through your report:

The Institute combines pioneering research with top class education. An innovative curriculum allows the student flexibility in selecting courses and projects. Students, even at the undergraduate level, get to participate in on-going research and technology development - an opportunity unprecedented in India

INDEX

INTERNSHIP COMPLETION CERTIFICATE

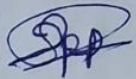
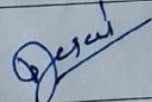

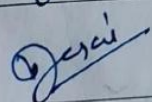

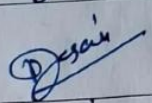

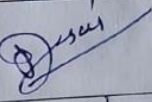
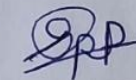
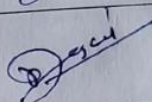

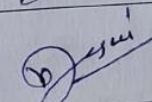

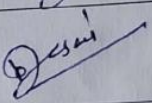
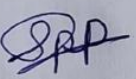
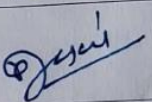

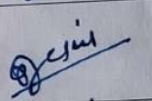

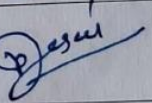
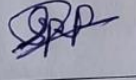
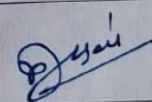
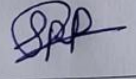
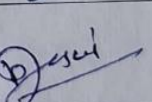
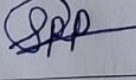
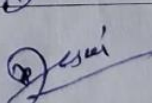
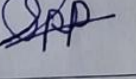
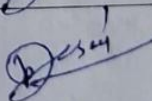
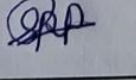
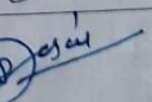
ACKNOWLEDGMENT	iv
ABSTRACT	v
LEARNING/INTERNSHIP OBJECTIVES	ix
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: ANALYSIS	2
CHAPTER 3: SYSTEM REQUIRMENTS	5
3.1 Functional Requirements.....	5
3.2 Non- Functional Requirements	5
3.3 System Requirements.....	5
CHAPTER 4: TECHNOLOGY	8
4.1 Stacked Bar Chart	9
4.2 Stacked Column Chart	9
4.3 Clustered Bar Chart.....	10
4.4 Clustered Column Chart.....	11
4.5 Line Chart.....	11
4.6 Area Chart.....	12
4.7 Line Stacked Chart.....	13
4.8 Ribbon Chart.....	13
4.9 Funnel Chart.....	14
4.10 Scatter Chart.....	15
4.11 Gauge Chart.....	15
4.12 Pie Chart.....	16
4.13 Waterfall Chart.....	16
4.14 Tree-map Chart.....	17
4.15 Slicer	17
4.16 Maps.....	18

CHAPTER 5: TASKS COMPLETION	12
CHAPTER 6: SCREENSHOTS	23
CHAPTER 7: CONCLUSION	41
CHAPTER 8: REFERENCES	42

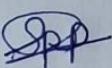
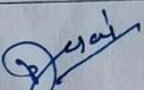

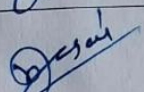
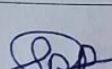
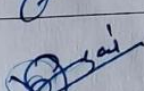
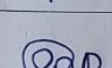
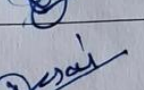
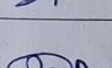
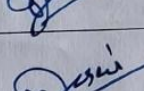
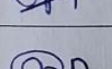
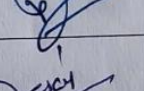
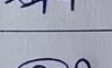
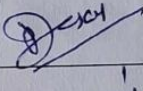
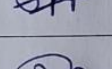
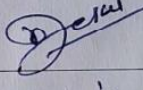
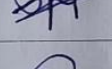
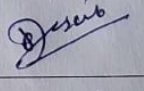
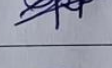
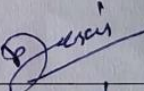
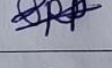
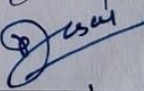
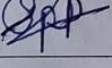
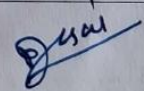
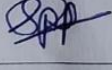
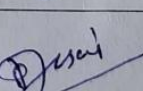
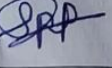
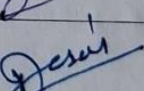
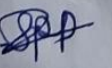
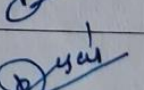

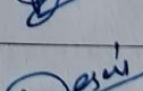
Learning Objectives/Internship Objectives

- Internships are generally thought of to be reserved for college students looking to gain experience in a particular field. However, a wide array of people can benefit from
- Training Internships in order to receive real world experience and develop their skills.
- An objective for this position should emphasize the skills you already possess in the area and your interest in learning more
- Internships are utilized in a number of different career fields, including architecture, engineering, healthcare, economics, advertising and many more.
- Some internship is used to allow individuals to perform scientific research while others are specifically designed to allow people to gain first-hand experience working.
- Utilizing internships is a great way to build your resume and develop skills that can be emphasized in your resume for future jobs. When you are applying for a Training Internship, make sure to highlight any special skills or talents that can make you stand apart from the rest of the applicants so that you have an improved chance of landing the position.

Daily Attendance Record

Date	Signature of Student	Description of Work Carried Out	Signature of Supervisor / Guide
20/12/24		Introduction to Excel Interface, Data Types in Excel, Applying Calculations.	
23/12/24		Relative and Absolute Cell Reference Calculation Order Preference, Cell Errors.	
24/12/24		Functions and Modifying Spreadsheets	
24/12/24		Formatting the spreadsheet, Conditional Formatting, Illustrations	
26/12/24		Charts :- Column Chart, Different chart types, Sorting	
27/12/24		Filters, Subtotals and custom Sort.	
28/12/24		Format as Table, Duplicate Rows Treatment.	
31/12/24		DSUM I, DSUM II and other Functions.	
31/12/24		Subtotal Functions and Pivot Tables.	
31/12/24		Grouping and Modifying Pivot Tables, Freeze Panes, IF Functions	
01/01/25		Introduction to Power BI, Basic visuals : Bar, Pie, Donut, Funnel charts.	
02/01/25		Advanced charts : Ribbon, Line, Area, combo, Scatter and Maps.	
03/01/25		Formatting Graphs, Power View, Dashboard Creation (Orders Dataset)	
06/01/25		Introduction to Databases, Installation of MS SQL server, ACID properties.	
07/01/25		DDL operations :- Create, Alter, Drop, Truncate; SQL Constraints.	

Daily Attendance Record

Date	Signature of Student	Description of Work Carried Out	Signature of Supervisor / Guide
08/01/25		DML Operations: Insert, update, Delete;	
09/01/25		Group By, Having, Aggregate Functions SQL Joins (outer inner, Left, Right, full)	
10/01/25		Subqueries (Pet, vet, owner Data)	
11/01/25		Introduction to Python, Basic Operators Data Types, Variables.	
11/01/25		Control statements, Functions, Object-Oriented Programming.	
11/11/25		File Handling, Exception Handling Practice on Loops & OOP concepts.	
13/01/25		Introduction to 1D & 2D Arrays, Arrays Indexing, Arithmetic & Logical operations.	
14/01/25		Pandas :- Series, Data Frames, merging Data Frames, Data input & output.	
15/01/25		matplotlib :- Data Visualization, plotting, Grids, Labels, Colors.	
16/01/25		Build an ATM System in Python.	
17/01/25		Data Analysis on cars Dataset.	
20/01/25		Introduction to Machine Learning Predictive models, Regression Basics.	
21/01/25		Simple Linear Regression, multiple Linear Regression.	
22/01/25		classification K-nearest Neighbours, SVC.	
22/01/25		PCA and clustering: PCA on Random Dataset, K-Means Clustering.	
23/01/25		Decision trees, Random Forest. Wine Quality & IRIS Dataset Analysis.	

1. INTRODUCTION

In today's data-driven business environment, organizations are increasingly relying on advanced analytics tools to gain insights into their sales performance. A Product Sales Dashboard in Power BI serves as a powerful tool for visualizing and analyzing sales data, enabling businesses to make informed decisions and drive growth.

A product sales dashboard is a visual representation of sales data that allows businesses to monitor and analyze their product performance over time. It consolidates various data points into a single view, making it easier for stakeholders to understand sales trends, customer behavior and overall business performance.

The ability to visualize and interpret sales data effectively can be a game-changer for organizations aiming to enhance their market position and optimize their operations. This is where Power BI, Microsoft's robust business intelligence tool, comes into play, offering a comprehensive solution for creating dynamic product sales dashboards.

Sales dashboards serve as a critical component of a company's analytics strategy. They provide a centralized platform for monitoring sales performance, enabling businesses to track progress against targets, identify areas for improvement, and respond swiftly to market changes. A well-designed sales dashboard can transform complex data into actionable insights, allowing teams to focus on what matters most—driving revenue and improving customer satisfaction.

In a competitive marketplace, the ability to make data-driven decisions is paramount. Product sales dashboards in Power BI empower organizations to analyze historical sales data, understand customer preferences, and evaluate the effectiveness of marketing strategies.

2. ANALYSIS

Introduction of Industry/ Organization, Types of Product and Services, History, Number of Employees

Softcrowd Technologies is a company established in April 2017 . It has industry-specific software expertise in technology, healthcare, financial, media, manufacturing and many other sectors. This company specializes in offering Web Designing, Web Application Development, Mobile App Development, Software Development, Digital Marketing, Software Testing, Quality Assurance Services and many more.

Softcrowd Technologies aspire to be the global sourcing choice of the world market and revolutionizes the way service processes function. To reach out to the common people across the globe and making Information Technology a tool for the “MASS” along with the tool for the “CLASS”. Creating innovative IT solutions and provide IT-enabled services to delight customers worldwide and build Relationships based on Trust, Values and Professionalism.

Softcrowd Technologies has industry-specific software expertise in Technology, Financial, Healthcare, Media, Manufacturing, and many other sectors. The company specializes in offering Web Designing, Web Application Development, Mobile Application Development, Software Development, Digital Marketing, Software Testing, Quality Assurance services, and many more. We are a team of committed innovative, clientsensitive and experienced software professionals who always strive to deliver customized, cost- effective and long-term software solutions that complement our client's objective and result in a satisfied customer.

At Softcrowd Technologies, “Strive with Technology” to provide the most effective and affordable service that fulfills our customer’s needs and budget. They provide customized websites and software solutions that suit customer’s company objectives.

SERVICES PROVIDED BY SOFTCROWD TECHNOLOGIES

WEB DEVELOPMENT

Softcrowd Technologies is a web design studio that offers corporate web design and custom web design. Our website designers offer cheap website design that is user-friendly at the same time easily customizable. We offer both corporate web design as well as custom web design. web design solutions include Complete Site Design from Scratch, Redesign of your existing website, Error correction in your website, New Sections in your current website, Complete revamp of your present website with old/new content and Daily site content maintenance

MOBILE APP DEVELOPMENT

Android is the Latest technology. Most Companies Prefer using Android development because it is cheap and efficient with respect to speed and execution. Many Mobile Apps are developed in android because it provides a wide range of functionality.

Software on Android Application is used widely because it is available on every mobile, iPad and tablets. Android application is developed in cheap charges due to free availability with respect to other tools.

SOFTWARE DEVELOPMENT

Software development is a process by which standalone or individual software is created using a specific programming language. It involves writing a series of interrelated programming code, which provides the functionality of the developed software. Software development may also be called application development and software design.

Web hosting

Web hosting is a service that allows organizations and individuals to post a website or web page onto the Internet. A web host, or web hosting service provider, is a business that provides the technologies and services needed for the website or webpage to be viewed in the Internet. Websites are hosted, or stored, on special computers called servers. When Internet users want to view your website, all they need to do is type your website address or domain into their browser. Their computer will then connect to your server and your webpages will be delivered to them through the browser.

Bulk SMS

Bulk sms is part of digital Marketing solution. Bulk SMS is easier way to communicate thousands of people just in second. You are able to send SMS in bulk with Single click from Desktop and Laptop by using our online Web portal with secure Username and Password.

Social Media marketing

Whether it's creating your digital product or just designing it, whether it's engaging with your audiences or just marketing to them, we do it all.

Social media is the collective of online communications channels dedicated to community-based input, interaction, content-sharing and collaboration. Websites and applications dedicated to forums, micro blogging, social networking, social bookmarking, social curation, and wikis are among the different types of social media.

ERP

The need of the hour for manufacturing industries to overcome aggressive business competition does not rest only with techno-savvy production line equipment's and gadgets, but also depends on smart usage of technology to leverage their true potential.

Products of Softcrowd Technology

- Online Examination System
- ERP
- Magic Bricks Application
- Food Delivery Application
- Online Service Provider Application
- Home Automation System
- E-commerce Application

3. SYSTEM REQUIREMENTS SPECIFICATIONS

3.1 Functional Requirements

- **Data Sources:** List the data sources that will be integrated (e.g., SQL databases, Excel files, cloud services). Specify any APIs or data connectors required.
- **Data Processing:** Describe the data cleaning, transformation, and modelling processes. Specify any DAX calculations or measures that need to be implemented.
- **Dashboard Features:**
 - **Visualizations:** List the types of visualizations required (e.g., bar charts, line graphs, pie charts, maps).
 - **Interactivity:** Specify features such as filters, slicers, drill-throughs, and tooltips.
 - **User Roles:** Define different user roles and their access levels (e.g., admin, viewer).
- **Export and Sharing:** Describe the requirements for exporting reports (e.g., PDF, Excel). Specify sharing options (e.g., sharing via Power BI Service, embedding in other applications).

3.2 Non-Functional Requirements

- **Performance:** Specify performance requirements, such as load times for dashboards and data refresh rates.
- **Usability:** Define usability requirements, including user interface design principles and accessibility standards.
- **Security:** Outline security requirements, including data encryption, user authentication, and row-level security (RLS).
- **Scalability:** Describe the expected scalability of the dashboards to accommodate growing data volumes and user numbers.
- **Compliance:** Specify any compliance requirements (e.g., GDPR, HIPAA) that the dashboards must adhere to.

3.3 System Requirements

- **Hardware Requirements:**
 - **Client-Side:** Specify the minimum and recommended hardware specifications for users accessing the dashboards (e.g., CPU, RAM, storage).
 - **Server-Side:** If applicable, outline the server requirements for hosting data sources or Power BI Report Server.

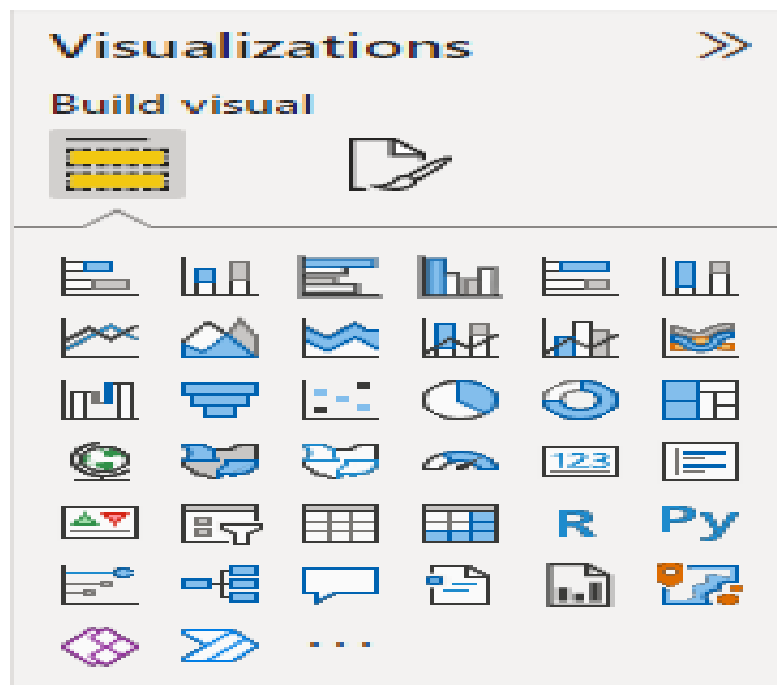
- **Software Requirements:**
 - **Power BI Desktop:** Specify the version and any prerequisites for Power BI Desktop.
 - **Power BI Service:** Outline any requirements for using Power BI Service (e.g., licensing).
 - **Operating Systems:** List supported operating systems for client machines (e.g., Windows, macOS).
- **Network Requirements:** Specify network bandwidth requirements for optimal performance, especially for cloud-based data sources.

4. TECHNOLOGY

A Power BI report is a multi-perspective view into a semantic model, with visuals that represent findings and insights from that semantic model. A report can have a single visual or many pages full of visuals. Depending on your job role, you might be someone who designs reports, or you might be a business user who consumes reports. This article is for business users. Power BI bases a report on a single semantic model. Report designers create the visuals in a report to represent pieces of information. Presenting data as visuals makes it easier to grasp the meaning of what you're seeing. And, the visuals aren't static. They update as the underlying data changes. The designers do the pre-work, adding visuals that they believe are interesting. But you get to explore and dig deeper. You interact with the visuals and filters as you dig into the data to discover insights and find answers. The extent of what you can do with a report depends on your role and the permissions the report designer assigns to you.

1. Visualization Pane

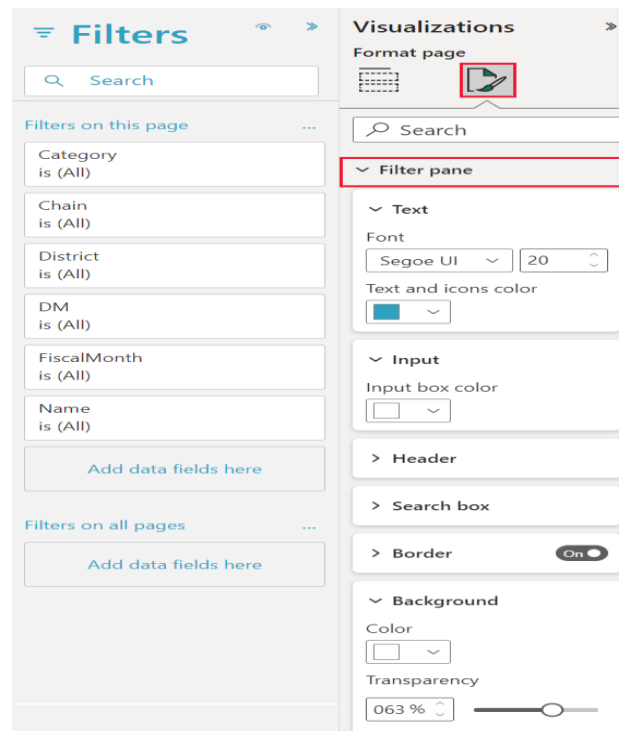
The small icons show the different types of visualizations you can create. If you start building a visualization by selecting fields without selecting a visualization type first, Power BI picks the visualization type for you. You can keep Power BI's selection, or change the type by selecting a different icon.



2. Filter Pane

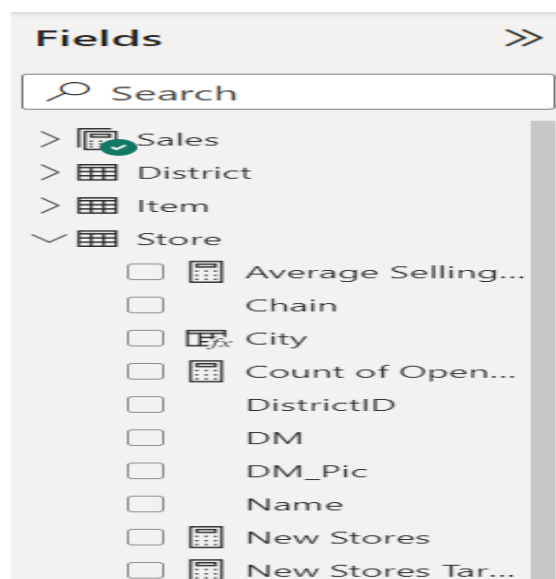
Use the Filters pane to view, set, and modify persistent filters to your reports at the page, report, drill-through, and visual-level. Yes, you can do ad-hoc filtering on report pages and visuals by

selecting elements of the visuals or by using tools like slicers. Filtering in the Filters pane has the advantage that the state of the filters is saved with the report.



3. Field Pane

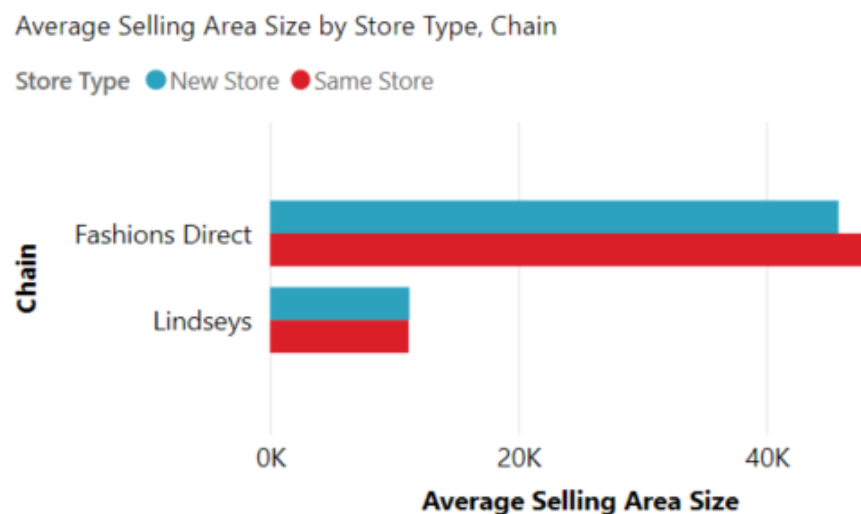
The Fields pane displays the tables, folders, and fields in your data that are available for you to use to create visualizations. Drag a field onto the page to start a new visualization. You can also drag a field onto an existing visualization to add the field to that visualization. When you add a checkmark next to a field, Power BI adds that field to the active or new visualization. It also decides which bucket to place that field into.



4.1 Stacked Bar Chart

A stacked bar chart is a type of chart that represents the total of different categories stacked on top of one another for each item in a category. It allows for easy comparison of the total values across categories while also highlighting the contributions of individual sub-groups.

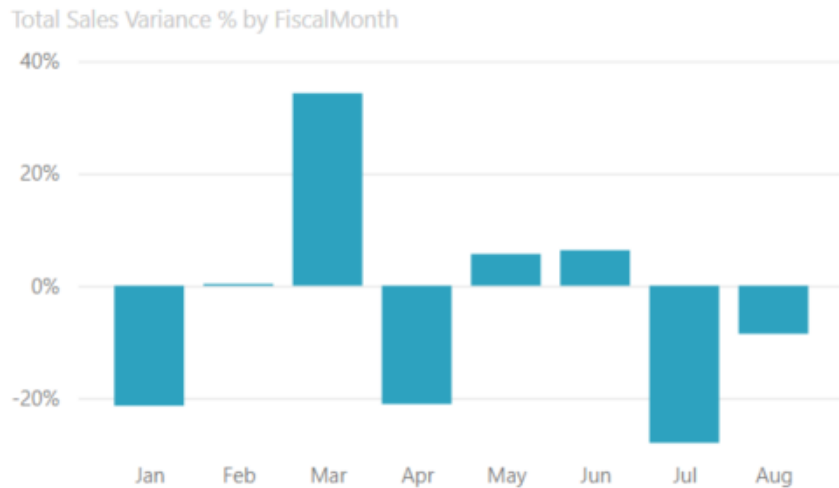
- **Categories on X-axis:** Each bar represents a different product category: Sports, Clothing, Toys, Electronics, Home & Kitchen, Books, and Beauty.
- **Categories on Y-axis:** The height of each segment within the bars represents the sum of sales or quantities sold, indicating performance or distribution in each category.



4.2 Stacked Column Chart

Used to show the total of different categories while comparing the sub-categories within each major category. Each column represents a main category (e.g., product types like Sports, Clothing, Toys, etc.). Each section of a stacked column represents a sub-category or division within the product type, indicating the distribution of sales data.

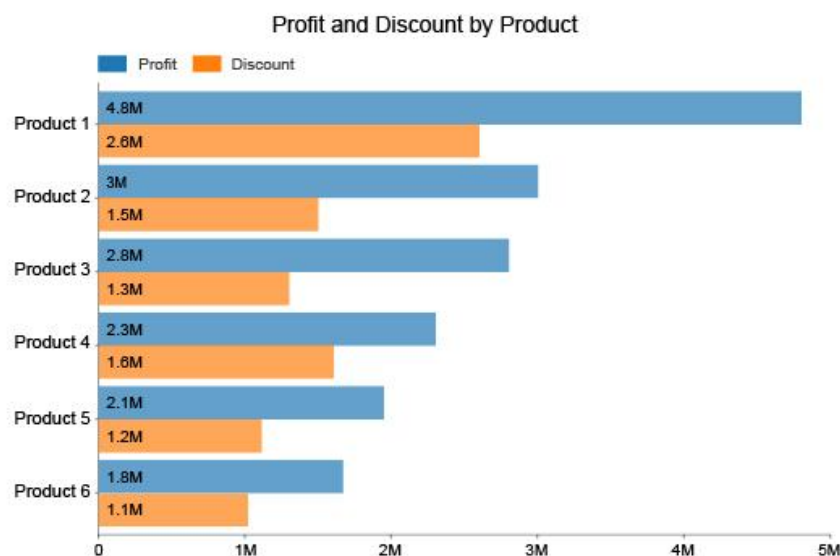
- The **X-axis** typically displays product categories.
- The **Y-axis** indicates the numeric value (like sales quantities or totals).



4.3 Clustered Bar Chart

A clustered bar chart is a graphical representation that uses bars to display and compare multiple sets of data across categories. Clustered bar charts are an effective tool for visualizing multi-dimensional data, making them valuable for business insights, performance tracking, and decision-making processes.

- **Multiple Data Series:** Represents different data series for each category, making comparisons straightforward.
- **Grouped by Category:** Categories are often grouped together on the x-axis, with separate colored bars for each series.
- **Orientation:** Bars can be horizontal or vertical, with vertical being more common for clustered bar charts.

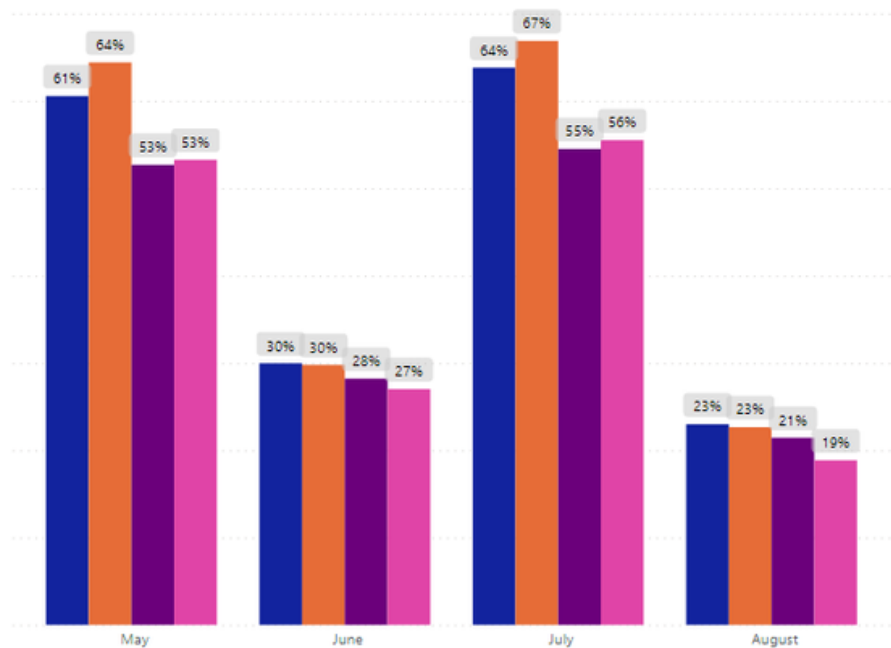


4.4 Clustered Column Chart

A clustered column chart is commonly used for comparing values across different categories.

By understanding and utilizing clustered column charts effectively, businesses can gain valuable insights into performance trends and comparisons. Display multiple series of data for easy comparison within categories.

- **Horizontal Axis (X):** Displays different categories (e.g., product types, regions).
- **Vertical Axis (Y):** Represents the value of the data (e.g., quantity sold, total sales).

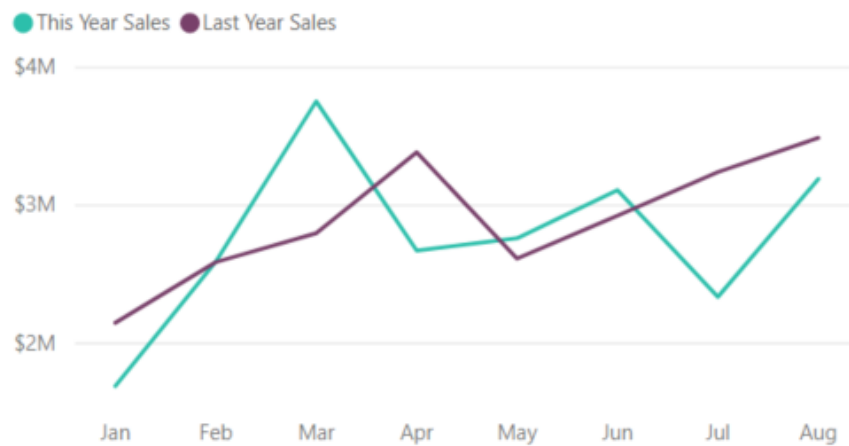


4.5 Line Chart

The line chart is visually clear with color coding based on product category, facilitating easy identification of trends and data points. The line indicates fluctuating sales across the different categories, with notable values recorded for each category

- **Y-Axis:** Represents the sales figures, ranging from approximately 3600 to 4200.
- **X-Axis:** Displays product categories, including:

This Year Sales and Last Year Sales by FiscalMonth

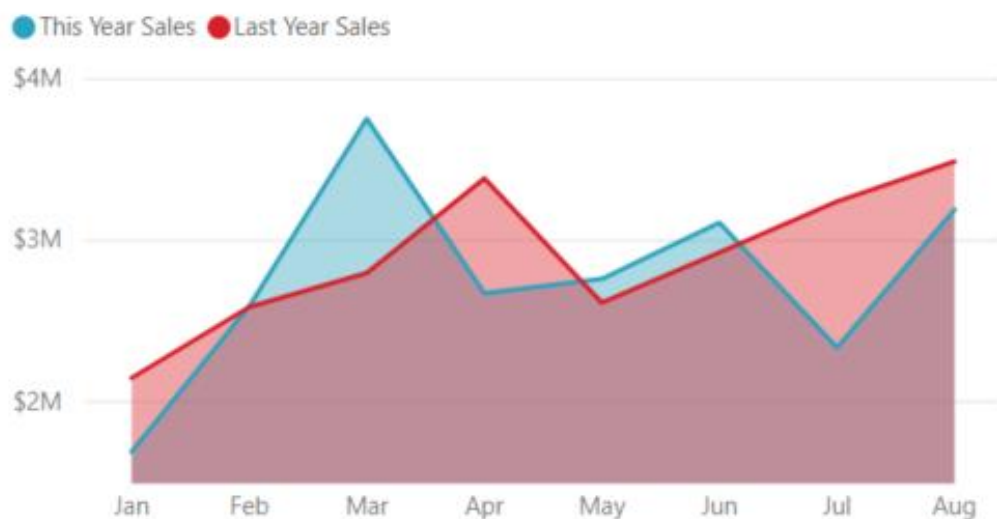


4.6 Area Chart

An area chart is a graphical representation of data that uses filled areas to show the magnitude of values over a continuous interval. Certain categories may show higher sales than others (e.g., Clothing tends to have higher peaks)Fluctuations in sales can be visualized, allowing for quick analysis of which product categories are more popular or performing better.

- **X-Axis:** Represents different product categories, such as Sports, Clothing, Toys, Electronics, Home & Kitchen, Books, and Beauty.
- **Y-Axis:** Displays the corresponding sales figures, ranging from a minimum of around 3600 to a maximum of approximately 4200.

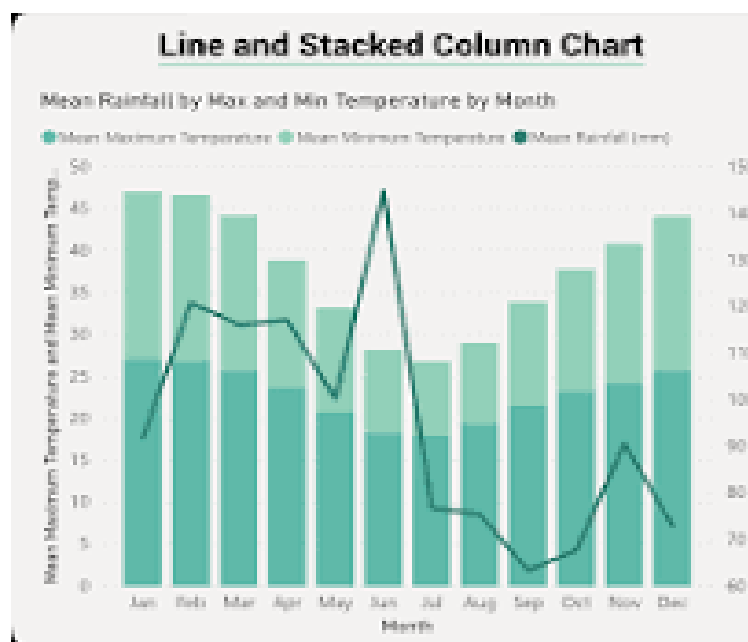
This Year Sales and Last Year Sales by FiscalMonth



4.7 Line Stacked Column Chart

A line stacked column chart combines two types of visualizations: stacked columns and a line chart. It is ideal for displaying both individual category values (in stacked columns) and the overall trend (in a line). The chart shows individual contributions of categories while also illustrating the trend over time or between different categories.

- **Categorical Data:** Utilizes categorical data for a clear comparison of parts to a whole.
- **Trend Analysis:** The line chart overlays the columns, making it easy to observe trends across different time periods or categories.



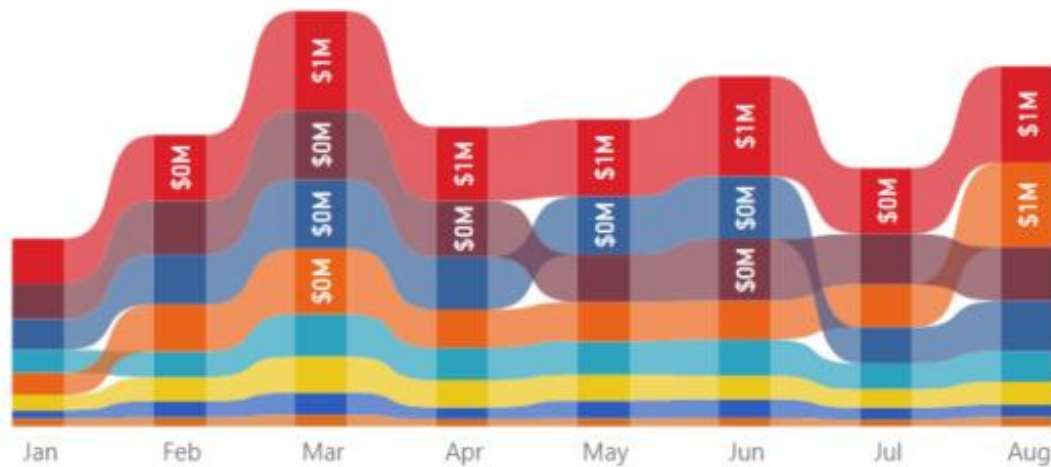
4.8 Ribbon Chart

A ribbon chart is a data visualization tool that allows users to see variations in data across multiple categories and how these categories relate to each other over time or across a dimension. Ribbon charts are effective in providing a complex overview of data dynamics, allowing businesses to visualize relationships and trends over time efficiently.

- **Data Relationship:** Illustrates the flow and connection between different categories over a timeline or in a categorical hierarchy
- **Visual Relationship:** Uses colored ribbons that connect various segments or categories, showing entire trends and making it easy to identify changes and relationships.

This Year Sales by FiscalMonth and Category

Category ● 010-Womens ● 020-Mens ● 030-Kids ● 040-Juniors ● 060-Intimate ● 070-Hosiery

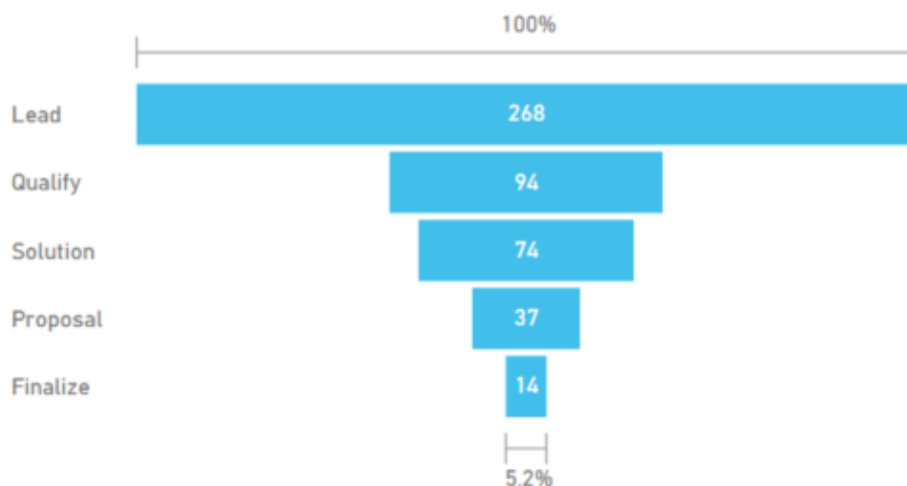


4.9 Funnel Chart

A funnel chart is a visualization tool commonly used in sales, marketing, or analytics to represent the progressive reduction of data as it passes through a defined process. Here's a breakdown based on the provided dashboard image. A funnel chart effectively summarizes complex data sets by visualizing transitions through stages, making it invaluable for performance analysis in sales and marketing.

- **Visual Representation:** Funnel charts visually depict stages in a process. Each segment of the funnel represents a stage where data is filtered down.
- **Stage-by-Stage Analysis:** They provide insights into how many items pass from one stage to the next, helping identify drop-off points.

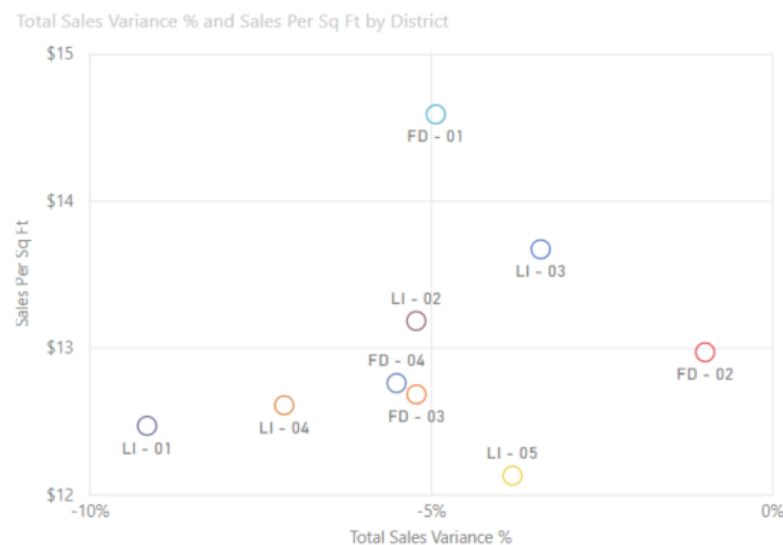
Opportunity Count by Sales Stage



4.10 Scatter Chart

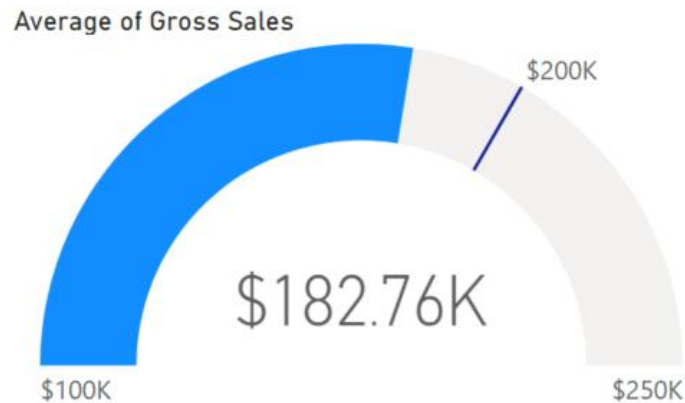
A scatter chart is a data visualization tool used to represent values for typically two different variables for a set of data. A scatter chart is an effective way to visualize complex relationships between variables, helping businesses make informed decisions based on data insights. Understanding the visual representation of data can drive strategic decisions and identify areas for growth.

- **Data Representation:** Each point on the scatter chart represents an individual data point. The chart can show trends, correlations, and distributions of the data points.
- **Interpretation:** Patterns may indicate relationships or correlations between variables. Clusters of points can highlight groupings or categorizations of data. Outliers can be identified as points that deviate significantly from the overall pattern.



4.11 Gauge Chart

A gauge chart visually represents quantitative values within a defined range, typically used to indicate performance measures, such as sales figures or other KPIs (Key Performance Indicators). The gauge displays a numeric value (1.71K) which likely represents a key metric (e.g., sales, quantity sold) against a defined scale (0 to 6K). It facilitates easy visualization of progress towards a goal or target.



4.11 Pie Chart

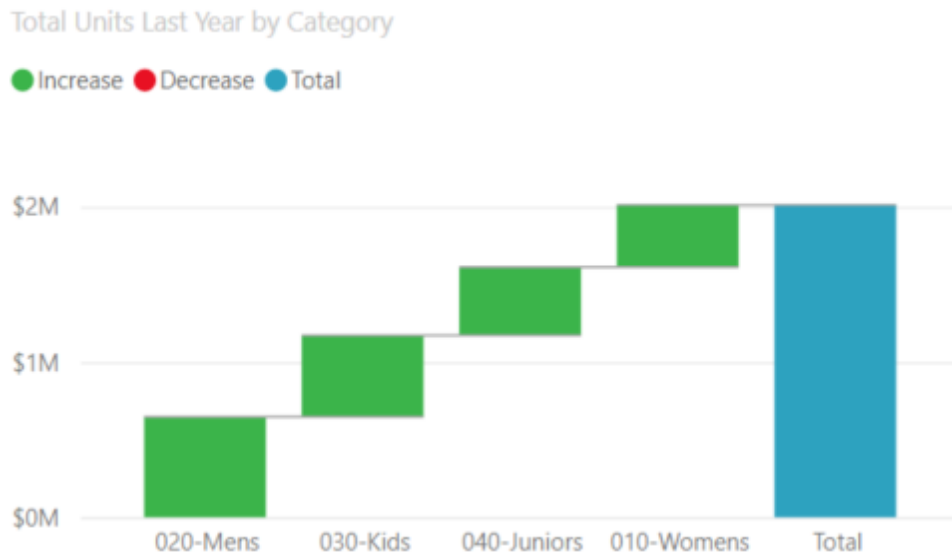
The pie chart illustrates the distribution of sales across different regions. This pie chart and the accompanying data provide a comprehensive overview of the sales performance across different regions, useful for strategic business decision-making.

This Year Sales by Chain



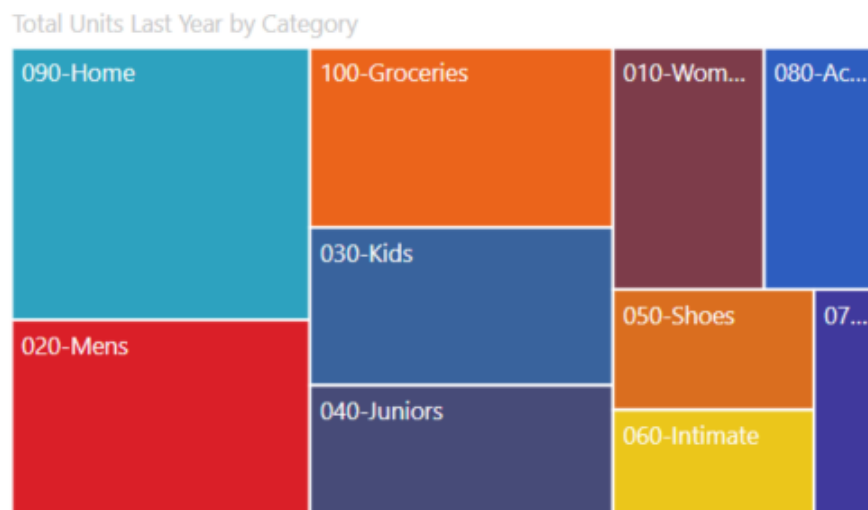
4.12 Waterfall Chart

A waterfall chart shows a running total as values are added or subtracted. It's useful for understanding how an initial value (for example, net income) is affected by a series of positive and negative changes. The columns are color coded so you can quickly tell increases and decreases. The initial and the final value columns often start on the horizontal axis, while the intermediate values are floating columns. Because of this "look", waterfall charts are also called bridge charts.



4.13 Tree-map

Tree-maps are charts of colored rectangles, with size representing value. They can be hierarchical, with rectangles nested within the main rectangles. The space inside each rectangle is allocated based on the value being measured. And the rectangles are arranged in size from top left (largest) to bottom right (smallest).

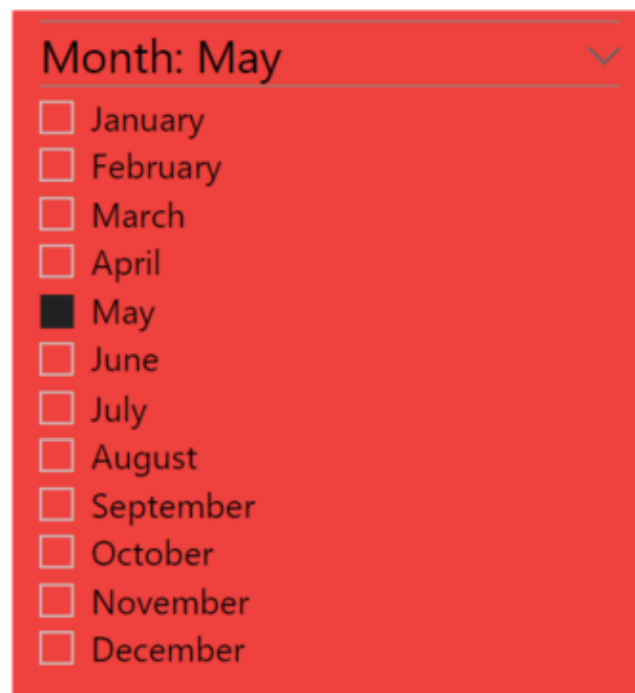


4.14 Slicer Chart

The slicer chart allows users to filter and analyze various aspects of product sales data efficiently. The slicer chart is a valuable tool in the Product Sales Insights Dashboard, facilitating comprehensive analysis and decision-making based on visualized data metrics. Users can adjust the slicers to focus on specific regions, product categories, and order statuses.

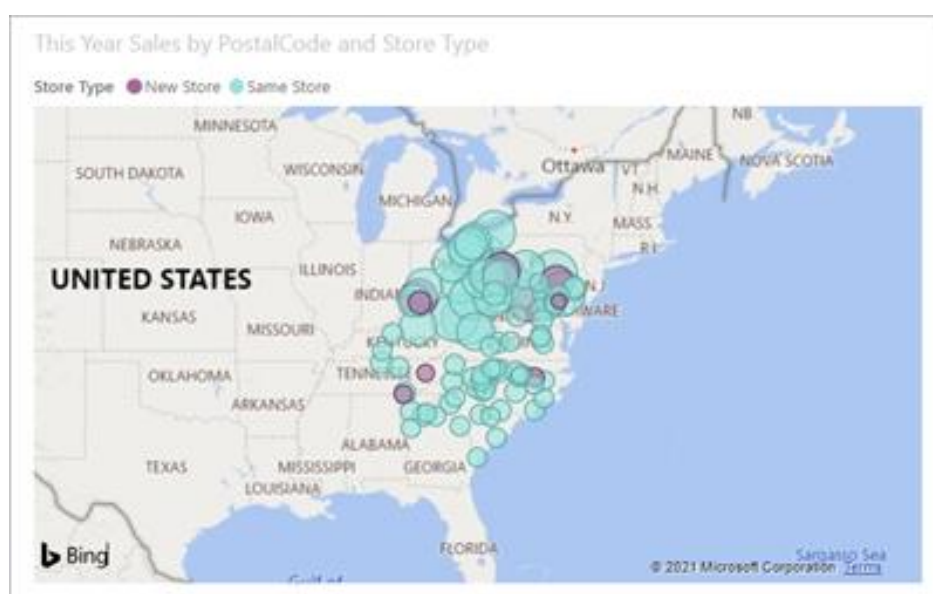
- **Visual Representation:** Different types of charts are used (e.g., pie charts, bar graphs) to visualize sales quantities and financial metrics.

- **Region Filter:** Categories: Asia, Australia, Europe, North America, South America
- **Order Status:** Options: Cancelled, Completed, Pending, Returned

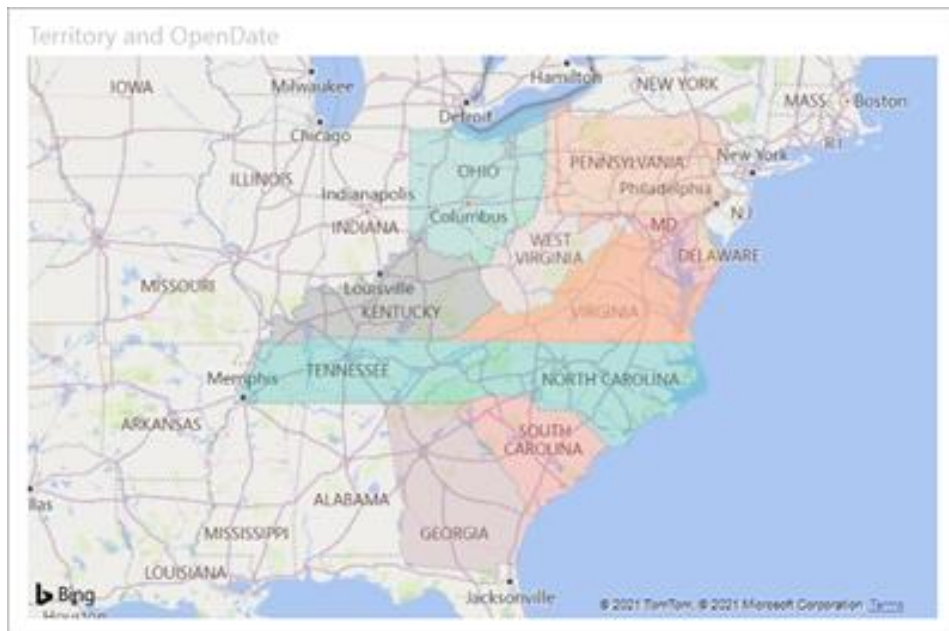


4.15 Maps

Basic Maps: Use a basic map to associate both categorical and quantitative information with spatial locations.



Filled map: A filled map uses shading or tinting or patterns to display how a value differs in proportion across a geography or region. Quickly display these relative differences with shading that ranges from light (less-frequent/lower) to dark (more-frequent/more).



5. SCREENSHOTS

First Dashboard :

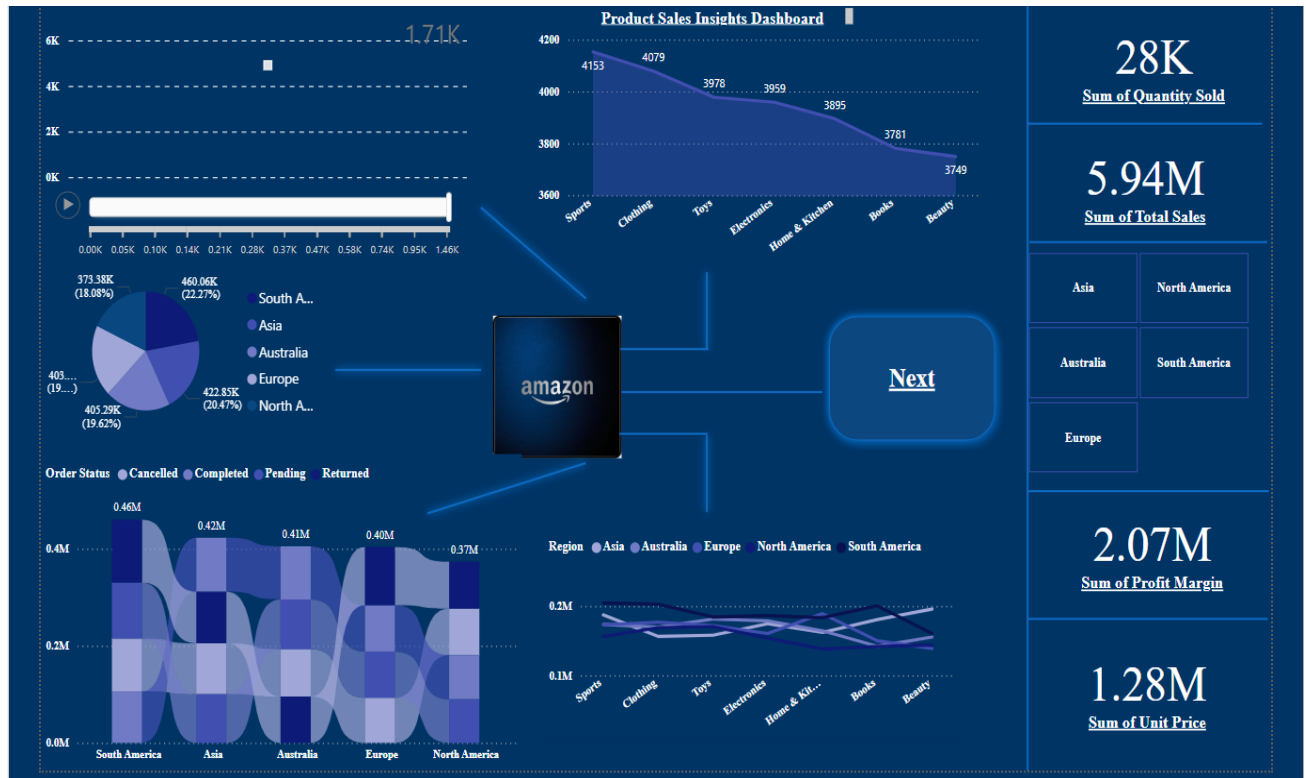


Fig. 1

First dashboard using Scatter chart, Pie chart, Ribbon chart, Line chart, Area chart , Cards, Buttons, Logos

Second Dashboard :

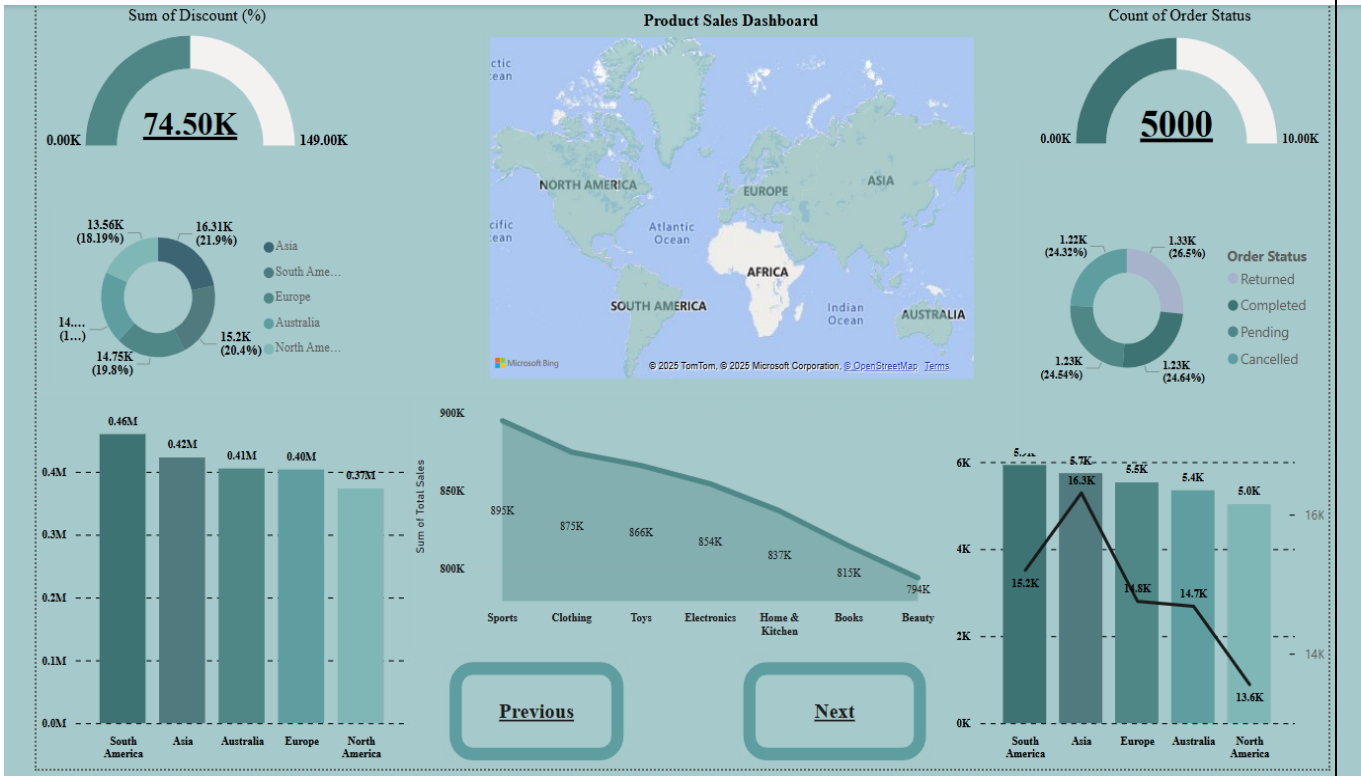


Fig. 2

Second dashboard using Gauge, Donut chart, Stacked column chart, Stacked area chart, Buttons, Line and stacked column chart, Maps

Third Dashboard :

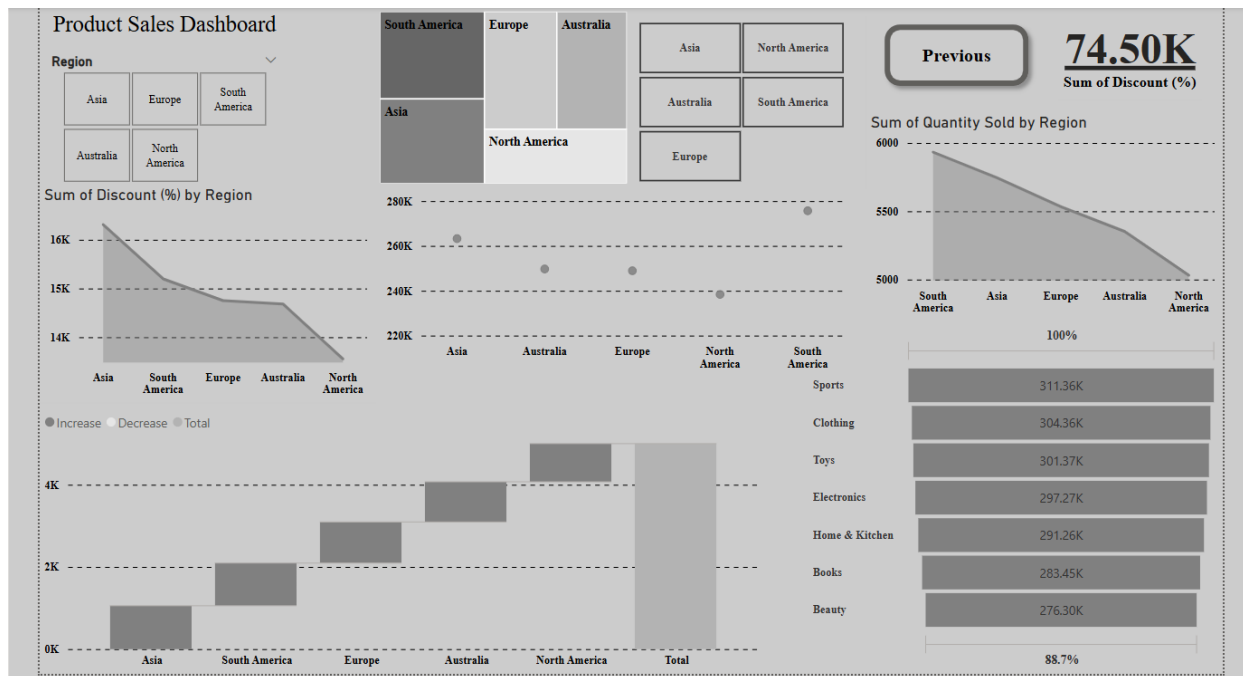


Fig. 3

Third dashboard using Slicer, Treemap, Waterfall chart, Funnel chart, Stacked area chart, Slicer, Button, Card

6. TASK COMPLETION

EXCEL

Introduction to Excel Interface, Data Types in Excel, Applying Calculations Relative and Absolute Cell Reference, Calculation Order Preference, Cell Errors

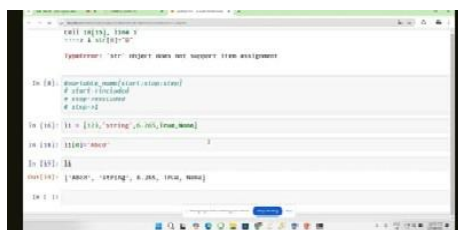
Function and Modifying Spreadsheets

1. Formatting the Spreadsheets, Conditional Formatting, Illustrations

Student Data						
Student	s1	s2	s3	s4	percentage	
DBMS	55	40	42	57	0.268867925	
SE	30	51	38	45	0.212264151	
PA	57	49	32	29	0.136792453	
CG	53	56	45	35	0.16509434	
M3	50	52	41	46	0.216981132	
Total	245	248	198	212		
avg	49	49.6	39.6	42.4		
min	30	40	32	29		
max	57	56	45	57		
count	5	5	5	5		
gst	18%	44.1	44.64	35.64	38.16	

2.Filters, Sub-Totals and Custom Sort

Subtotal		
Buyers	Products	Price
Saurabh	A	15
Pratibha	A	36
Rushi	A	10
A Total	0	61
Siddhant	B	18
Akshada	B	65
B Total	0	83
Amruta	C	74
Rahul	C	52
C Total	0	126
Grand Total	0	270



3.Format as Table, Duplicate Rows Treatment

ID	NAME
1	A
2	B
3	C
4	D
5	E
2	B
6	T
4	D

ID	NAME
1	A
2	B
3	C
4	D
5	E

4.DSUM I.DSUM II, and Other functions

ID	PRODUCT	COUNTRY	PRICE
1	S	q	12
2	A	w	52
3	U	e	14
4	R	t	16
5	A	y	74
7	H	i	16
8	S	m	15
9	A	p	21
10	U	e	20
	unhide		259
	hide		240

DSUM		
PRODUCT	PRICE	
S		27

DSUM		
PRODUCT	COUNTRY	PRICE
S	m	15

Column1	DAVG	
A		49

Column1	Column2	Column3
A	p	21

DCOUNT		
PRODUCT	PRICE	
A		3

DCOUNT		
PRODUCT	COUNTRY	PRICE
U	e	2

DMIN		
PRODUCT	PRICE	
A		21

DMIN		
PRODUCT	COUNTRY	PRICE
U	e	14

DMAX		
PRODUCT	PRICE	
A		74

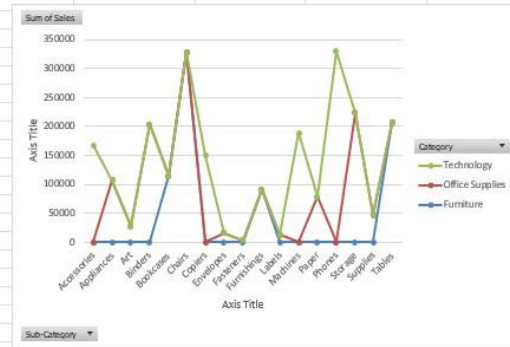
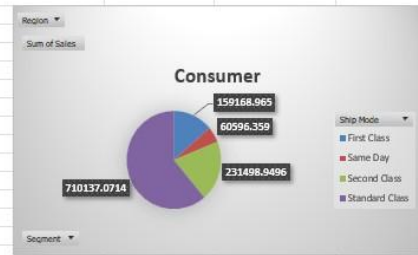
DMAX		
PRODUCT	COUNTRY	PRICE
U	e	20

5.Pivot Table

Grouping and Modifying Pivot tables, Freeze panes

Sum of Sales	Column Labels				
Row Labels	Central	East	South	West	Grand Total
Consumer	252031.434	350908.167	195580.971	362880.773	1161401.345
Corporate	157995.8128	200409.347	121885.9325	225855.2745	706146.3668
Home Office	91212.644	127463.726	74255.0015	136721.777	429653.1485
Grand Total	501239.8908	678781.24	391721.905	725457.8245	2297200.86

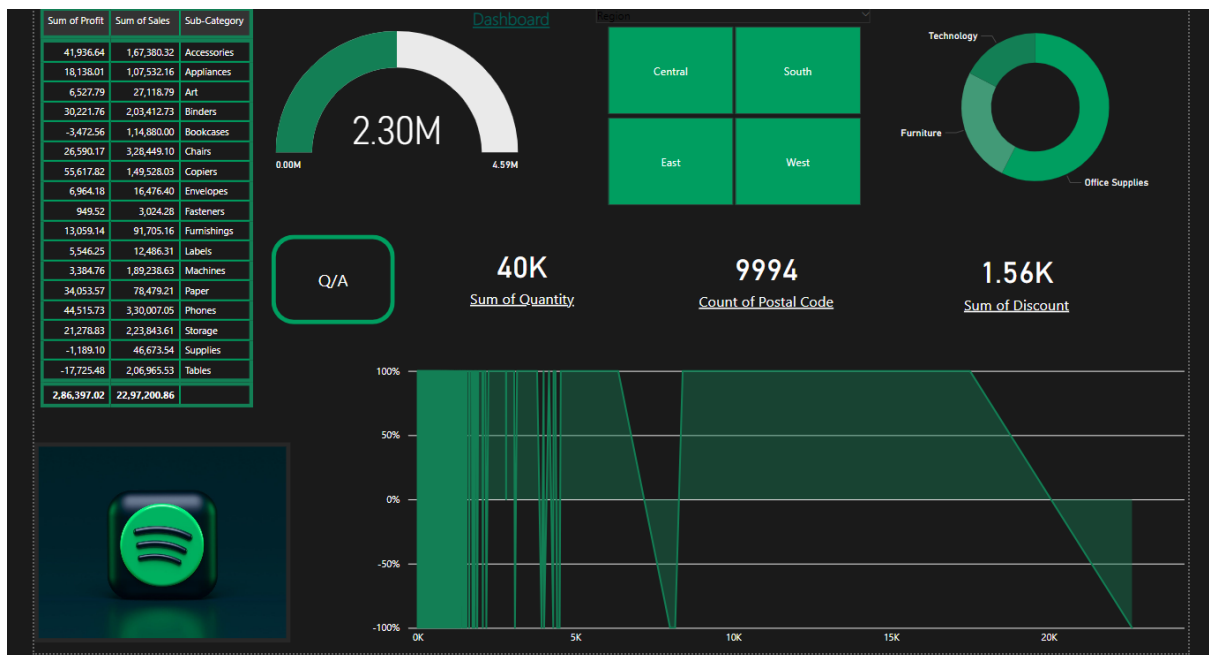
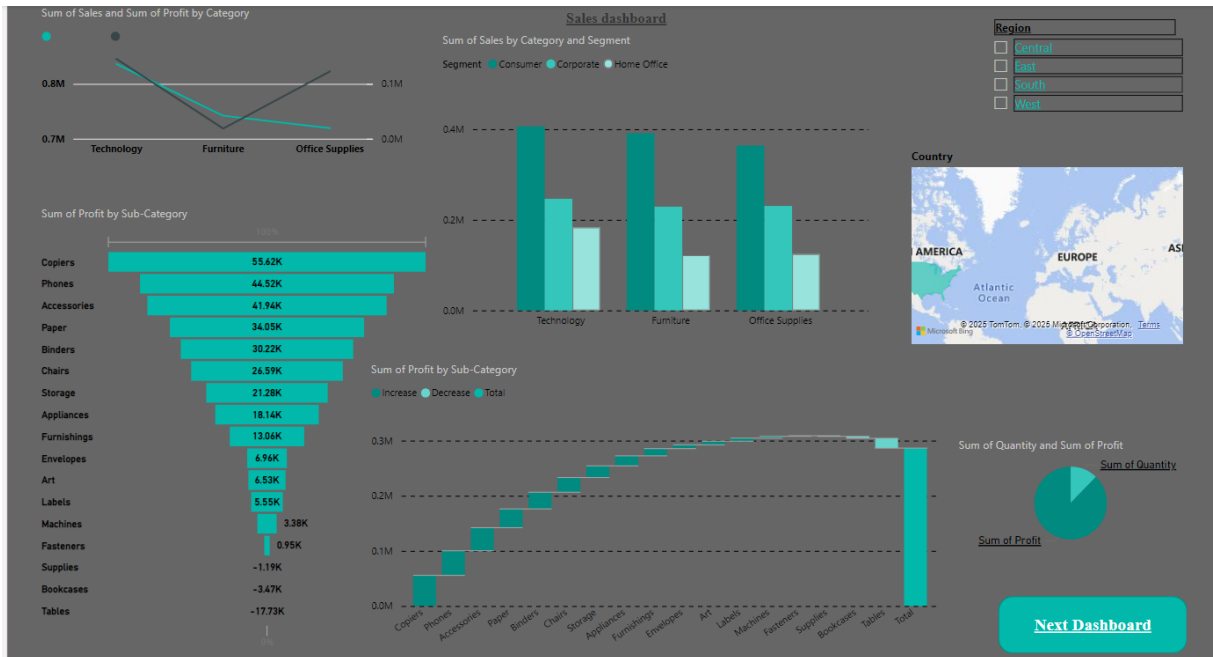
Region	(All)			
Sum of Sales	Column Labels			
Row Labels	Consumer	Corporate	Home Office	Grand Total
First Class	159168.965	105858.4699	86400.988	351428.4229
Same Day	60596.359	45121.323	22645.443	128363.125
Second Class	231498.9496	146126.0388	81568.581	459193.5694
Standard Class	710137.0714	409040.5351	239038.1365	1358215.743
Grand Total	1161401.345	706146.3668	429653.1485	2297200.86
Sum of Sales	Column Labels			
Row Labels	Furniture	Office Supplies	Technology	Grand Total
Accessories			167380.318	167380.318
Appliances		107532.161		107532.161
Art		27118.792		27118.792
Binders		203412.733		203412.733
Bookcases	114879.9963			114879.9963
Chairs	328449.103			328449.103
Copiers			149528.03	149528.03
Envelopes		16476.402		16476.402
Fasteners		3024.28		3024.28
Furnishings	91705.164			91705.164
Labels		12486.312		12486.312
Machines			189238.631	189238.631
Paper		78479.206		78479.206
Phones			330007.054	330007.054
Storage		223843.608		223843.608
Supplies		46673.538		46673.538
Tables	206965.532			206965.532
Grand Total	741999.7953	719047.032	836154.033	2297200.86

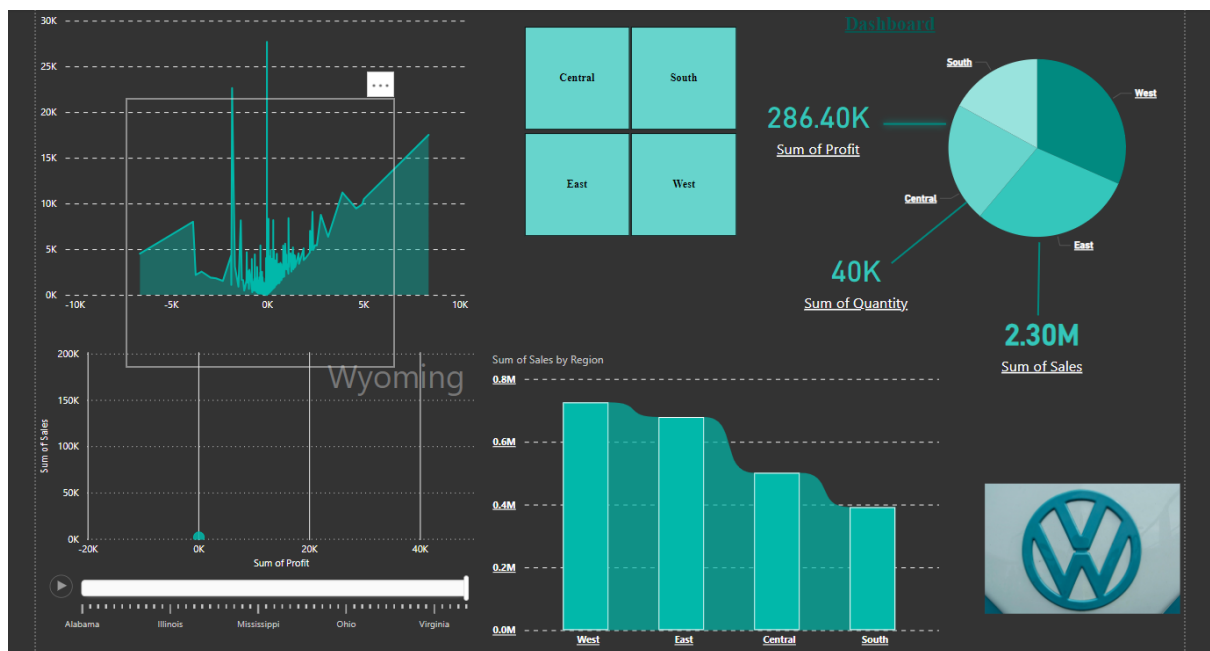


POWER BI

Introduction to Power-BI, Basic Visuals: Bar, Pie, Donut , Funnel Charts. Advances chart Ribbon, Line, Area, Combo, Scatter and Maps all charts, Button, Drill through , All action buttons, Shapes. Formatting, Power view, Dashborad creation

Final Dashboards





SQL

Introduction to Databases ,Installation of MS SQL SERVER ,ACID properties

DDL operations: Create,Alter,Drop,Truncate;SQL constraints

The screenshot displays the SQL Server Enterprise Manager interface. At the top, two query windows are open: 'SQLQuery1.sql - DE...rabah Potinde (55)' and 'SQLQuery2.sql - DE...rabah Potinde (59)'. The active window shows the following SQL code:

```
create table pet1(pet_id int, name varchar(30), kind varchar(30),gender varchar(30),age int,owner_id int);
select *from pet1;

insert into pet1 values(101,'Blackie','Dog','Male',11,5168),(102,'Roomba','Cat','Male',09,5508);
insert into pet1 values(103,'Simba','Cat','Male',01,3086),(104,'Keller','Parrot','Female',02,7908),(105,'Cuddles','Dog','Male',13,4378)

-->WHERE
--> 1) pet who are dog
select name from pet1 where kind='Dog';
--> 2) pet with age greater than 5
-->select name,age, kind from pet where (age)>=5;
select* from pet1 where age >= 5;

-->GROUP BY
-->3) COUNT NO OF CATS
select kind,count(*) as no_of_cats
from pet1 where kind ='Cat'
group by kind;
-->4)CALCULATE AVG AGE OF EACH KIND
select kind,avg(age) as avg_age
from pet1
group by kind;
-->5)MIN AND MAX AGE OF MALES AND FEMALES
select gender,min(age) as min_age,max(age) as max_age
```

Below the query window, the 'Results' tab is active, showing a grid of 16 rows and 6 columns. The columns are labeled: pet_id, name, kind, gender, age, and owner_id. The data is as follows:

	pet_id	name	kind	gender	age	owner_id
1	101	Blackie	Dog	Male	11	5168
2	102	Roomba	Cat	Male	9	5508
3	103	Simba	Cat	Male	1	3086
4	104	Keller	Parrot	Female	2	7908
5	105	Cuddles	Dog	Male	13	4378
6	106	Vuitton	Parrot	Female	11	7581
7	107	Priya	Cat	Female	7	7343
8	108	Simba	Cat	Male	0	2700
9	109	Cookie	Cat	Female	8	7606
10	110	Heisenberg	Dog	Male	3	1319
11	111	Stowe	Cat	Female	15	1132
12	112	Scout	Dog	Female	2	7846
13	113	Lily	Dog	Female	3	7846
14	114	Danger	Dog	Male	9	9037
15	115	Scooter	Dog	Male	9	9850
16	116	Bandit	Parrot	Male	11	6102

```

---right join
select student.name, Studentcourse.course_id, student.roll_no,
Studentcourse.course from student_final student
right join Studentcourse
on student.roll_no=Studentcourse.roll_no;

---Full Outer Join
select student.name, Studentcourse.course_id, student.roll_no,
Studentcourse.roll_no from student_final student
full join Studentcourse
on Studentcourse.roll_no=student.roll_no;

---Cross Join
select * from student_final cross join Studentcourse;

---self join
CREATE TABLE self (
    emp_id INT,
    emp_name VARCHAR(50),
    salary INT,
    manager_id INT
);

```

98 %

Results Messages

	roll_no	name	city	age	course_id	roll_no	course
1	1	Harsh	Delhi	18	1	1	Data Science
2	2	Pratik	Bihar	19	1	1	Data Science
3	3	Riyanka	Bihar	20	1	1	Data Science
4	4	Deep	Bihar	18	1	1	Data Science
5	5	Saptarshi	Bihar	19	1	1	Data Science

Query executed successfully.

```

---Give Name & ID of pets who have been treated by Vet
select pet_id, name from pets
where pet_id in(select pet_id from Vet);

---Display name,kind and age of pets
---whose age is equal to average age of all pets
select name,kind,age from pets
where age in (select avg(age) from pets);

select name,kind,age from pets
where age=(select avg(age) from pets);

---Give name and occupation of owners who have a pet treated for
---ear or skin infection
select name,occupation from Owners
where owner_id in(select owner_id from pets
where pet_id in (select pet_id from vet
where disease='Ear Infection' or disease='Skin Infection'));

```

98 %

Results Messages

	name	occupation
1	Hasan Ansari	chef
2	Tanishka Ahire	teacher
3	Stanley Desouza	lawyer
4	Gagan Sharma	businessman


```

enter the number/a
number is even

In [1]: a = int(input("enter a= "))
       b = int(input("enter b= "))
       c = int(input("enter c= "))
       if a > b and a > c:
           print("a is largest")
       if b > a and b > c:
           print("b is largest")
       if c > a and c > b:
           print("c is largest")

Enter a= 45
Enter b= 65
Enter c= 34
b is largest

In [10]: number = int(input("Enter the number?"))
         if number==10:
             print("number is equals to 10")
         elif number==50:
             print("number is equal to 50")
         elif number==100:
             print("number is equal to 100")
         else:
             print("number is not equal to 10, 50 or 100")

Enter the number/a
number is not equal to 10, 50 or 100

```

```

In [8]: #variable_name[start:stop:step]
       # start->included
       # stop->excluded
       # step->1

In [16]: l1 = [123,"string",6.265,True,None]

In [18]: l1[0]="Abcd"

In [19]: l1
Out[19]: ['Abcd', 'string', 6.265, True, None]

In [ ]:

```

BREAK	CONTINUE	PASS
<pre> s = "python" for i in s: if i == 'h': break print(i) </pre> <p>p y t</p>	<pre> s = "python" for i in s: if i == 'h': continue print(i) </pre> <p>p y t o n</p>	<pre> s = "python" for i in s: if i == 'h': pass print(i) </pre> <p>p y t h o n</p>

```
def student(name,age,*args):
    print("Name: ", name)
    print("Age: ", age)

    for i in args:
        print("printing")
        print(i)

student("Raj",20,10,20,30,40)
```

```
Name: Raj
Age: 20
printing
10
printing
20
printing
30
printing
40
```

```
def student(name,age,**kwargs):
    print("Name: ", name)
    print("Age: ", age)

    for key,value in kwargs.items():
        print(key,":",value)

student("Raj",20,maths=90,science=100,history=80)
```

```
Name: Raj
Age: 20
maths : 90
science : 100
history : 80
```

```
In [5]: try:
        a = int(input("Enter a:"))
        b = int(input("Enter b:"))
        c = a/b
        print(c)
    except:
        print("Can't divide with zero")

Enter a:12
Enter b:2
6.0
```

```
In [6]: try:
        a = int(input("Enter a:"))
        b = int(input("Enter b:"))
        c = a/b
        print("a/b = %d"%c)
        # Using Exception with except statement. If we print(Exception) it will return exception class
    except :
        print("can't divide by zero")
    else:
        print("Hi I am else block")

Enter a:12
Enter b:4
a/b = 3
Hi I am else block
```

```
] : f = open("abc.txt","w")

f.write("hellow good afternoon everybody")
f.close()
f = open("abc.txt", "r")
print(f.read())
f.close()

hellow good afternoon everybody
```

```
] : #Write to an Existing File

f = open("abc.txt", "w")
f.write("Now the file has more content!")
f.close()

#open and read the file after the appending:
f = open("abc.txt", "r")
print(f.read())

Now the file has more content!
```

Introduction to 1D,2D Arrays, Array indexing ,Arithmetic and Logical operations

Pandas: Series Dataframe, Merging dataframes , Data input and output Matplotlib: Data visualization, Ploting ,Grids, Labells , Colors

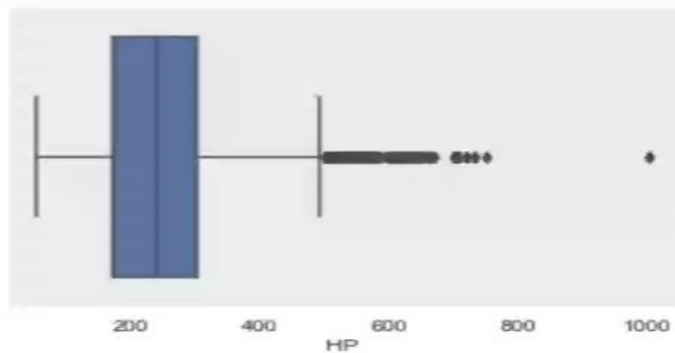
```
In [1]: import numpy as np
# as-> Alias
```

```
In [6]: a = np.array(10) #0-Dimension -> Scalar
b = np.array([1,2,3]) #1-D -> Vector
c = np.array([[1,2],[3,4]]) #2-D -> Metrix
d = np.array([[[1, 2, 3], [4, 5, 6]], [[1, 2, 3], [4, 5, 6]]])
#3-D -> Tensor
```

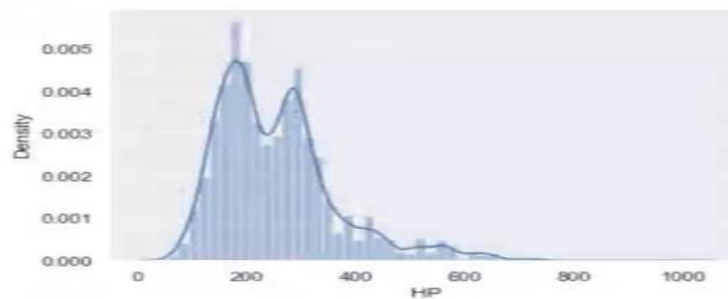
```
In [9]: print(a)
print("-"*25)
print(b)
print("-"*25)
print(c)
print("-"*25)
print(d)
```

```
10
-----
[1 2 3]
-----
[[1 2]
 [3 4]]
```

```
## Plot a boxplot for 'HP' columns in dataset
sns.boxplot(x=df['HP'])
plt.show()
```



```
#plotting distplot for variable HP
sns.distplot(df['HP'])
plt.show()
```

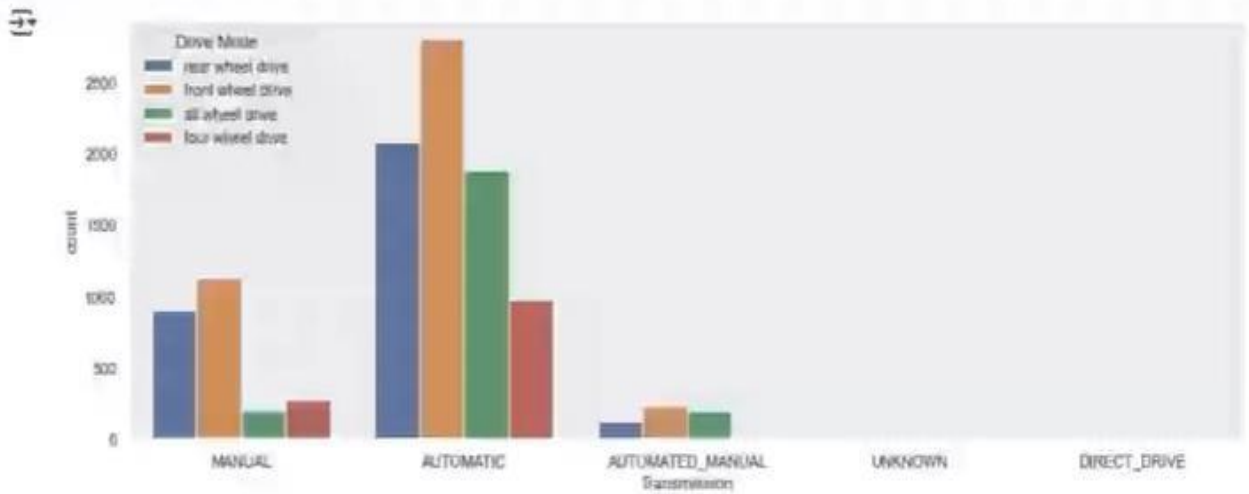


```

plt.figure(figsize=(15,5))
df = df.dropna()
# plot countplot on transmission and drive mode
sns.countplot(x='Transmission',hue='Drive Mode',data=df)
plt.show()

# 'Cylinders', y='Price'

```



Data Cleaning using Pandas

```
#create d1
d1 = pd.DataFrame( {'Temperature' : [1, np.nan, 3, 2, 3] ,
                    'Humidity' : [22, np.nan, 2, np.nan, 20]})
```

t the dataframe d1

```
#print d1
print(d1)
```

	Temperature	Humidity
0	1.0	22.0
1	NaN	NaN
2	3.0	2.0
3	2.0	NaN
4	3.0	20.0

```
# imputing mean
d1.fillna(d1.mean())
```

	Temperature	Humidity
0	1.00	22.000000
1	2.25	14.666667
2	3.00	2.000000
3	2.00	14.666667
4	3.00	20.000000

```
#drop E
df.drop(["E"], axis=1, inplace=True)
df = df.drop(["E"], axis=1)
```

check if column E is removed by printing head of df

```
#df head
df.head()
```

	A	B	C	D
0	9	15	64	28
1	93	29	8	73
2	40	36	16	11
3	88	62	33	72
4	49	51	54	77

Machine Learning

1.Feature Engineering

```
: data[mask].shape
: (7, 3)

: mask1 = (data['Height']>=Lower_limit) & (data['Height']<=Upper_limit)
: new_df = data[mask1]

: new_df
```

	Gender	Height	Weight
0	Male	73.847017	241.893563
1	Male	68.781904	162.310473
2	Male	74.110105	212.740856
3	Male	71.730978	220.042470
4	Male	69.881796	206.349801

Outlier Removal using Std Dev

```
: standard_deviation = data['Height'].std()
: mean = data['Height'].mean()

: standard_deviation,mean
: (3.8475281287732324, 66.36755975482124)

: Upper_limit = mean + (3*standard_deviation)
: Lower_limit = mean - (3*standard_deviation)
: print(Upper_limit)
: print(Lower_limit)

: 77.91014411714094
: 54.82497539250154

: mask = (data['Height']<Lower_limit) | (data['Height']>Upper_limit)
: data[mask]
```

	Gender	Height	Weight
994	Male	78.066867	255.090835
1317	Male	78.462053	227.342565
2014	Male	78.996742	209.969099
3285	Male	78.528210	253.889004

Outlier Removal using Z-score

```
: def z_score(df,column):
:     x = df[column].values #data['Height']
:     mean = df[column].mean()
:     standard_deviation = df[column].std()
:     return (x-mean) / standard_deviation

: data['z_score'] = z_score(data,'Height')

: data.head()
```

	Gender	Height	Weight	z_score
0	Male	73.847017	241.893563	1.943964
1	Male	68.781904	162.310473	0.627505
2	Male	74.110105	212.740856	2.012343
3	Male	71.730978	220.042470	1.393991
4	Male	69.881796	206.349801	0.913375

```
: mask = (data['z_score']<-3) | (data['z_score']>3)
: data[mask]
```

Outlier Removal using IQR

```
: q1 = np.quantile(data['Height'],0.25)
: q3 = np.quantile(data['Height'],0.75)
: IQR = q3-q1

: Upper_limitIQR = q3 + (1.5*IQR)
: Lower_limitIQR = q1 - (1.5*IQR)
: print(Upper_limit)
: print(Lower_limit)

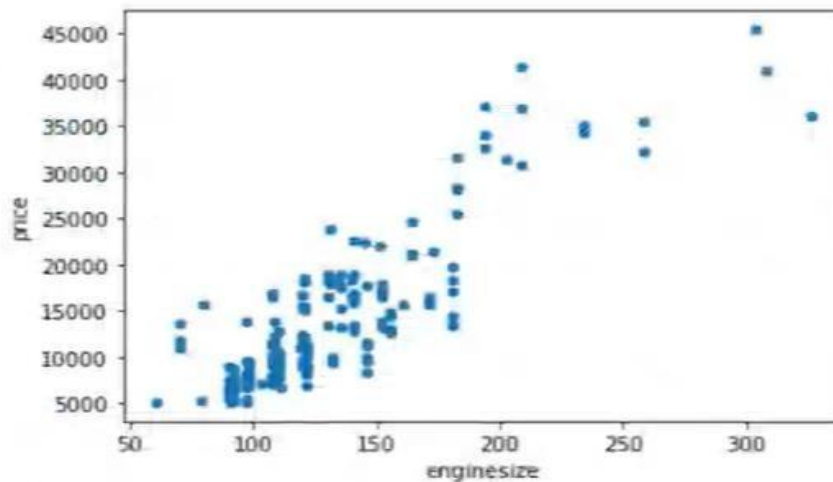
: 77.91014411714094
: 54.82497539250154

: mask2 = (data['Height']>=Lower_limitIQR) & (data['Height']<=Upper_limitIQR)
: new_df = data[mask2]

: new_df.shape
: (9992, 3)
```

2.Linear Regression

```
: # check the distribution of data by plotting as scatter
new_df.plot(x='engine_size', y='price', kind='scatter')→
: <AxesSubplot:xlabel='engine_size', ylabel='price'>
```



```
: # import ML related packages of sklearn
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score

: # split the data as train and test
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=0)

: #shapes of splitted data - sizes of the X_train and Y_train should be same. Also, tests.
print("X_train:",x_train.shape)
print("X_test:",x_test.shape)
print("Y_train:",y_train.shape)
print("Y_test:",y_test.shape)

X_train: (164,)
X_test: (41,)
Y_train: (164,)
Y_test: (41,)

: # create a linear regression model
model = LinearRegression()

: # train the model using training data
# since we have only one independent variable, you should use 'values.reshape(-1,1)'. Otherwise, x_train is enough.
model.fit(x_train.values.reshape(-1,1), y_train)

: LinearRegression()
```



```
# print y intercept / bias
model.intercept_

-7613.3709394898

# make predictions using test data
y_pred = model.predict(x_test.values.reshape(-1,1))

# Let's calculate the metrics
# MSE
mse = mean_squared_error(y_test, y_pred)
print("MSE --> ", mse)

MSE --> 16835544.03813768

# RMSE
import math
rmse = math.sqrt(mse)
print("RMSE --> ", rmse)

RMSE --> 4103.113944084137

# MAE
mae = mean_absolute_error(y_test, y_pred)
print("MAE --> ", mae)

MAE --> 3195.031239500043

# R2
r2 = r2_score(y_test, y_pred)
print("R2 --> ", r2)
```

3. Logistic Regression

```
import pandas as pd
from matplotlib import pyplot as plt
%matplotlib inline
```

```
df = pd.read_csv("insurance_data.csv")
df.head()
```

	age	bought_insurance
0	22	0
1	25	0
2	47	1
3	52	0
4	46	1

```
plt.scatter(df.age, df.bought_insurance, marker='+', color='red')
```

```
<matplotlib.collections.PathCollection at 0x20a8cb15d30>
```



```
import math
def sigmoid(x):
    return 1 / (1 + math.exp(-x))

def prediction_function(age):
    z = 0.042 * age - 1.53 # 0.04150133 ~ 0.042 and -1.52726963 ~ -1.53
    y = sigmoid(z)
    return y
```

```
age = 35
prediction_function(age)
```

```
0.4850044983805899
```

0.485 is less than 0.5 which means person with 35 age will *not* buy insurance

```
age = 43
prediction_function(age)
```

```
0.568565299077705
```

0.485 is more than 0.5 which means person with 43 will buy the insurance

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(df[['age']], df.bought_insurance, train_size=0.8)

X_test

age
4 46
8 62
28 23
17 58
24 50
26 54

from sklearn.linear_model import LogisticRegression
model = LogisticRegression()

model.fit(X_train, y_train)

C:\ProgramData\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: Default
to 'lbfgs' in 0.22. Specify a solver to silence this warning.
FutureWarning)

LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
intercept_scaling=1, max_iter=100, multi_class='warn',
n_jobs=None, penalty='l2', random_state=None, solver='warn',
tol=0.0001, verbose=0, warm_start=False)
```



```

from sklearn.datasets import load_digits
import pandas as pd

dataset = load_digits()
dataset.keys()

dict_keys(['data', 'target', 'frame', 'feature_names', 'target_names', 'images', 'DESCR'])

dataset.data.shape

(1797, 64)

dataset.data[0]

array([ 0.,  0.,  5., 13.,  9.,  1.,  0.,  0.,  0.,  0., 13., 15., 10.,
        15.,  5.,  0.,  0.,  3., 15.,  2.,  0., 11.,  8.,  0.,  0.,  4.,
        12.,  0.,  0.,  8.,  8.,  0.,  0.,  5.,  8.,  0.,  0.,  9.,  8.,
        0.,  0.,  4., 11.,  0.,  1., 12.,  7.,  0.,  0.,  2., 14.,  5.,
        10., 12.,  0.,  0.,  0.,  0.,  6., 13., 10.,  0.,  0.,  0.])

dataset.data[0].reshape(8,8)

array([[ 0.,  0.,  5., 13.,  9.,  1.,  0.,  0.],
       [ 0.,  0., 13., 15., 10., 15.,  5.,  0.],
       [ 0.,  3., 15.,  2.,  0., 11.,  8.,  0.],
       [ 0.,  4., 12.,  0.,  0.,  8.,  8.,  0.],
       [ 0.,  5.,  8.,  0.,  0.,  9.,  8.,  0.],
       [ 0.,  4., 11.,  0.,  1., 12.,  7.,  0.],
       [ 0.,  2., 14.,  5., 10., 12.,  0.,  0.],
       [ 0.,  0.,  6., 13., 10.,  0.,  0.,  0.]])

```

Use components such that 95% of variance is retained

```

from sklearn.decomposition import PCA

pca = PCA(0.95)
X_pca = pca.fit_transform(X)
X_pca.shape

(1797, 29)

pca.explained_variance_ratio_

array([0.14890594, 0.13618771, 0.11794594, 0.08409979, 0.05782415,
       0.0491691 , 0.04315987, 0.03661373, 0.03353248, 0.03078806,
       0.02372341, 0.02272697, 0.01821863, 0.01773855, 0.01467101,
       0.01409716, 0.01318589, 0.01248138, 0.01017718, 0.00905617,
       0.00889538, 0.00797123, 0.00767493, 0.00722904, 0.00695889,
       0.00596081, 0.00575615, 0.00515158, 0.0048954 ])

```

pca.n_components_

29

PCA created 29 components out of 64 original columns

X_pca

```

array([[ -1.25946645,  21.27488348, -9.46305462, ...,  3.67072108,
        -0.9436689 , -1.13250195],
       [  7.9576113 , -20.76869896,  4.43996004, ...,  2.18261819,

```

Let's now select only two components

```

pca = PCA(n_components=2)
X_pca = pca.fit_transform(X)
X_pca.shape

(1797, 2)

X_pca

array([[ -1.25946639,  21.27487891],
       [  7.95760922, -20.76869518],
       [  6.99192341, -9.95598163],
       ...,
       [ 10.80128435, -6.96025523],
       [ -4.87210315, 12.42395926],
       [ -0.34438701,  6.36554335]])

```

```

pca.explained_variance_ratio_

array([0.14890594, 0.13618771])

```

You can see that both combined retains $0.14+0.13=0.27$ or 27% of important feature information

```

X_train_pca, X_test_pca, y_train, y_test = train_test_split(X_pca, y, test_size=0.2, random_state=30)

model = LogisticRegression(max_iter=1000)
model.fit(X_train_pca, y_train)
model.score(X_test_pca, y_test)

```

5.SVM

```
# categorize wine quality
# 2-6.5 are "bad" quality, and wines that are between 6.5 - 8 are "good"
bins = (2,6.5,8) # range type for good bad
group_names = ['bad','good']
categories = pd.cut(df['quality'], bins, labels = group_names)
df['quality'] = categories

#after categorize
df['quality'].value_counts()

bad    1382
good    217
Name: quality, dtype: int64

#barplot of quality vs alcohol.
# more alcohol, better red wine
sns.barplot(x='quality', y='alcohol', data=df)
plt.show()

<Axes: xlabel='quality', ylabel='alcohol'>
```

```

> more alcohol, better red wine
sns.barpilot(x='quality', y='alcohol', data=df)
plt.show()

<Axes: xlabel='quality', ylabel='alcohol'>

X = df.drop(['quality'], axis = 1)
y = df['quality']

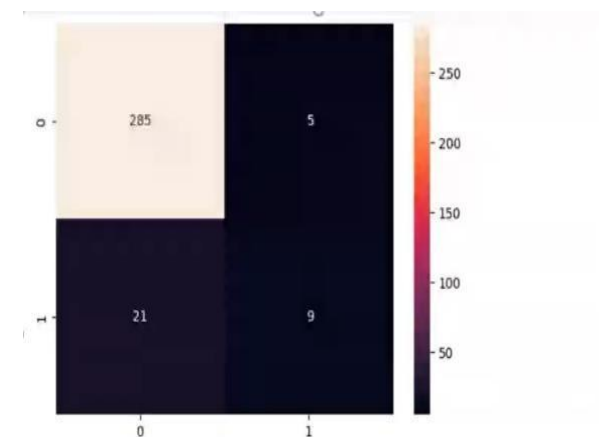
#Now quality is categorical i.e. good or bad . So to convert it into class 1 and 0 following step is taken.
from sklearn.preprocessing import LabelEncoder
labelencoder_y = LabelEncoder()
y = labelencoder_y.fit_transform(y)

y
array([0, 0, 0, ..., 0, 0, 0])

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 0)

X_test

```

[illegible]

```
from sklearn import metrics
print("Accuracy:", metrics.accuracy_score(y_test, predictions))
```

Accuracy: 0.91875

```
new_pred = list(result.predict([[11,1,0.100,0.99,4,0.99,1,2,0.1,1,0.50,9]]))
new_pred
```

7. CONCLUSION

In conclusion, Power BI dashboards are an essential component of modern data analytics and business intelligence strategies. They empower organizations to visualize data effectively, derive insights, and make data-driven decisions. By leveraging the capabilities of Power BI, businesses can enhance their analytical capabilities, improve operational efficiency, and ultimately drive better outcomes. As organizations continue to embrace data-driven cultures, the role of Power BI dashboards will only become more significant in facilitating informed decision-making and strategic planning.

By leveraging the capabilities of Power BI, businesses can enhance their analytical capabilities, improve operational efficiency, and ultimately drive better outcomes. As data continues to play a critical role in decision-making, mastering Power BI dashboards will be increasingly valuable for professionals across various industries.

8. REFERENCE

1. <https://learn.microsoft.com/en-us/power-bi/create-reports/sample-datasets>
2. <https://www.microsoft.com/en-us/power-platform/products/power-bi/topics/data-visualization>
3. <https://learn.microsoft.com/en-us/power-bi/consumer/end-user-dashboards>
4. <https://www.blackbox.ai/chat/kz9XweV>
5. Data Visualization with Microsoft Power BI: How to Design Savvy Dashboards
6. Microsoft Power BI Cookbook