

Data Vista Project Report

Title: DataVista: Sales Data Analysis and Visualization

Project Statement and Outcomes:

- ✓ Develop a data analysis and visualization platform for sales data.
- ✓ Provide insights into sales performance, trends, and patterns.
- ✓ Include features for interactive visualization and exploration of sales datasets.
- ✓ Implement functionalities for comparing sales across products, regions, and time periods.

Project Undertaken under: Infosys SpringBoard Summer Internship Program 4.0

Project Mentor: Sujata Zope

Team Members:

S.no	Name	Registered Email ID
1.	Debjani Guha Biswas	mayadb2266@gmail.com
2.	Sundram Kumar Ray	sundramraycse@gmail.com
3.	Chandrakant Mohanty	0211csbs005@niet.co.in
4.	Pentapati Prasanthi	pentapatiprasanthi15@gmail.com
5.	Yalavarthi Srija	srijayalavarthi254@gmail.com
6.	Pooja Karkhile	poojakarkhile91@gmail.com
7.	Vuyyuru Sai Ram Reddy	sairamreddy1511@gmail.com
8.	Jannat Sheik	shaikhjannat1105@gmail.com
9.	Thummala Keertisree	tkeerthisree1@gmail.com
10.	Varsha Mekala	varshamekala2003@gmail.com
11.	Nandini Dighe	nandinidighe18@gmail.com
12.	Kolli Satya Surya Teja	satyasuryateja9@gmail.com

Abstract:

The Data Vista project, conducted under the Infosys Springboard Summer Internship Program, focuses on the comprehensive analysis and visualization of sales data. The primary objective of this project is to leverage advanced data analytics techniques to derive actionable insights from sales data, enhancing strategic decision-making processes. This project involves collecting and processing sales data, performing detailed exploratory data analysis (EDA), and employing visualization tools to present the findings in an intuitive and accessible manner.

Through the application of various statistical and machine learning methods, this analysis identifies key sales trends, patterns, and anomalies. The project emphasizes the importance of data cleaning and preprocessing to ensure the accuracy and reliability of the results. By utilizing visualization platforms such as Tableau and Power BI, complex data sets are transformed into clear, interactive dashboards and reports.

The outcomes of this project aim to provide stakeholders with a deeper understanding of sales dynamics, including factors influencing sales performance and customer behavior. This information is crucial for optimizing sales strategies, improving customer targeting, and ultimately driving revenue growth. The Data Vista project serves as a testament to the value of data-driven approaches in business operations, showcasing the significant impact of effective data analysis and visualization.

Keywords: Sales Data Analysis, Data Visualization, Exploratory Data Analysis, Machine Learning, Tableau, Power BI, Infosys Springboard.

Project Timeline and their descriptions:

Week 1: Project Initialization and Planning

Understand project requirements and objectives.
Conduct a competency gap analysis.
Create a competency plan and define necessary actions.
Understand the sequence of tasks.
Set up the development environment with required tools.

Week 2: System Design and Initial Implementation

Design database tables (schema) for storing sales data.
Implement database tables and test data entries.
Implement basic frontend modules (e.g., basic data display).
Implemented Data segmentation

❖ Data references

<https://excelbianalytics.com/wp/downloads-18-sample-csv-files-data-sets-for-testing-sales/>
<https://data.world/dataman-udit/us-regional-sales-data>
<https://drive.google.com/file/d/1wFRMr2vP8q6KugL1qjU9uJutzrL28e1e/view?usp=sharing>

Total number of rows in dataset = 1000000

Total number of columns in dataset = 28

List of columns in the dataset:

'FirstName', 'LastName', 'customerID', 'productID', 'OrderID', 'Category', 'Sub_Category',
'ProductName', 'Quantity', 'Segment', 'Region', 'Country', 'City', 'State', 'PostalCode', 'latitude',
'longitude', 'OrderDate', 'ShipDate', 'ShipMode', 'ShipStatus', 'Gender', 'ItemCost', 'ShippingCost',
'MakingCost', 'TotalCost', 'Margin', 'SellingPrice],

Key Roles of Python in Preprocessing Sales Data

Python plays a crucial role in preprocessing sales datasets, providing a robust and flexible environment for data cleaning, transformation, and preparation for analysis. The preprocessing stage is essential for ensuring data quality and making it suitable for analytical tasks such as reporting, machine learning, and business intelligence.

1. Data Cleaning:

- **Handling Missing Values:** Python can identify and address missing values in the dataset, which might involve imputing missing data with statistical measures (mean, median, mode) or removing rows/columns with excessive missing data.
- **Removing Duplicates:** Python efficiently detects and eliminates duplicate records, which can skew analysis results.
- **Data Type Conversion:** Ensuring that all columns have the correct data type (e.g., dates, numerical values) is crucial for accurate analysis. Python libraries make it easy to convert data types as needed.

2. Data Transformation:

- **Date and Time Manipulation:** Python can parse and manipulate date and time data, allowing for the extraction of useful features such as month, year, day of the week, and fiscal quarters, which are often important in sales analysis.
- **Normalization and Scaling:** To prepare data for machine learning models, Python can normalize or scale numerical features to ensure they have comparable ranges, improving model performance and convergence.
- **Feature Engineering:** Python enables the creation of new features that can provide additional insights. For instance, calculating total sales from quantity and price per unit or creating categorical features from numerical data.

3. Outlier Detection and Handling:

- Outliers can significantly impact statistical analyses and machine learning models. Python provides tools to detect and handle outliers through statistical methods or visualization techniques, ensuring that the dataset is representative of typical sales patterns.

4. Data Integration:

- Sales data often comes from multiple sources (e.g., CRM systems, e-commerce platforms, inventory databases). Python can merge and integrate these disparate datasets, ensuring consistency and completeness.

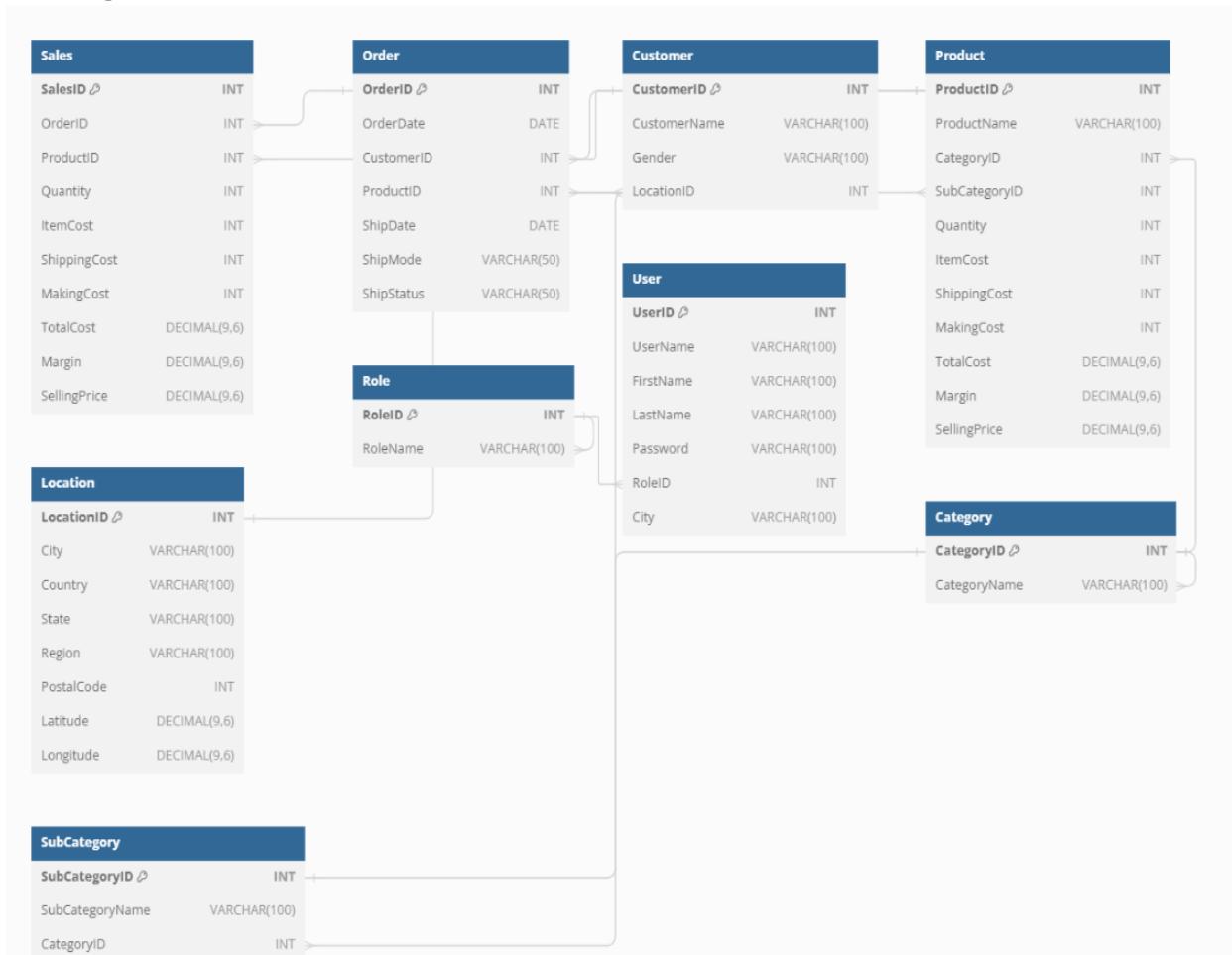
5. Data Aggregation and Summarization:

- Python allows for the aggregation and summarization of data, such as calculating total sales by region, product line, or sales representative. This helps in generating high-level insights and identifying trends and patterns.

6. Data Validation and Quality Assurance:

- Python can automate the validation of data against business rules and consistency checks, ensuring that the dataset meets predefined quality standards. This includes checking for logical inconsistencies, such as negative sales amounts or invalid dates.

ER Diagram:



Week 3: Data Analysis Design and Implementation

Designed Dashboards for additional data analysis features (e.g., aggregations, filters).

Design additional database tables for specific data analysis requirements.

Implement database tables and test data entries.

Enhance frontend modules for better data representation and visualization

Week 4: Integration and Testing Phase I

Integrate data retrieval and basic analysis modules.

Test integration and ensure smooth data flow.

Implement frontend modules for displaying basic analysis results.

Execute test cases for data retrieval, basic analysis, and frontend functionalities.

Key Concepts of Row-Level Security in Power BI

Row-level security (RLS) in Power BI is a feature that enables you to control access to data at the row level within your reports and dashboards. This means you can restrict data visibility for users based on their roles, ensuring that they can only see the data that is relevant to them. RLS is crucial for maintaining data security, confidentiality, and compliance with organizational policies.

1. **Roles:** RLS is implemented through the creation of roles within Power BI Desktop. Each role defines a set of rules that specify which rows of data are accessible to users assigned to that role.
2. **DAX Filters:** Data Analysis Expressions (DAX) are used to create row-level filters. These filters are applied to tables in your data model to restrict data based on specific criteria.
3. **User Assignment:** Once roles are created and filters are defined, users or groups are assigned to these roles in the Power BI Service. When a user views a report, the RLS filters are applied, and they can only see the data permitted by their role.

How to Implement Row-Level Security in Power BI

1. Define Roles in Power BI Desktop:

- Open your Power BI Desktop file.
- Go to the "Modeling" tab and select "Manage Roles."

- Create a new role and define DAX filters to restrict data access. For example, to restrict data to a specific region:

[Region] = "North America"

- Save the role definitions.

2. Test Roles:

- Use the "View as Roles" feature under the "Modeling" tab in Power BI Desktop to test how the data looks for a user assigned to a specific role.

3. Publish to Power BI Service:

- Publish your Power BI Desktop file to the Power BI Service.

4. Assign Users to Roles:

- In the Power BI Service, navigate to the dataset you have published.
- Select "Security" to open the RLS settings for the dataset.
- Assign users or groups to the roles you defined in Power BI Desktop.

Example Scenario

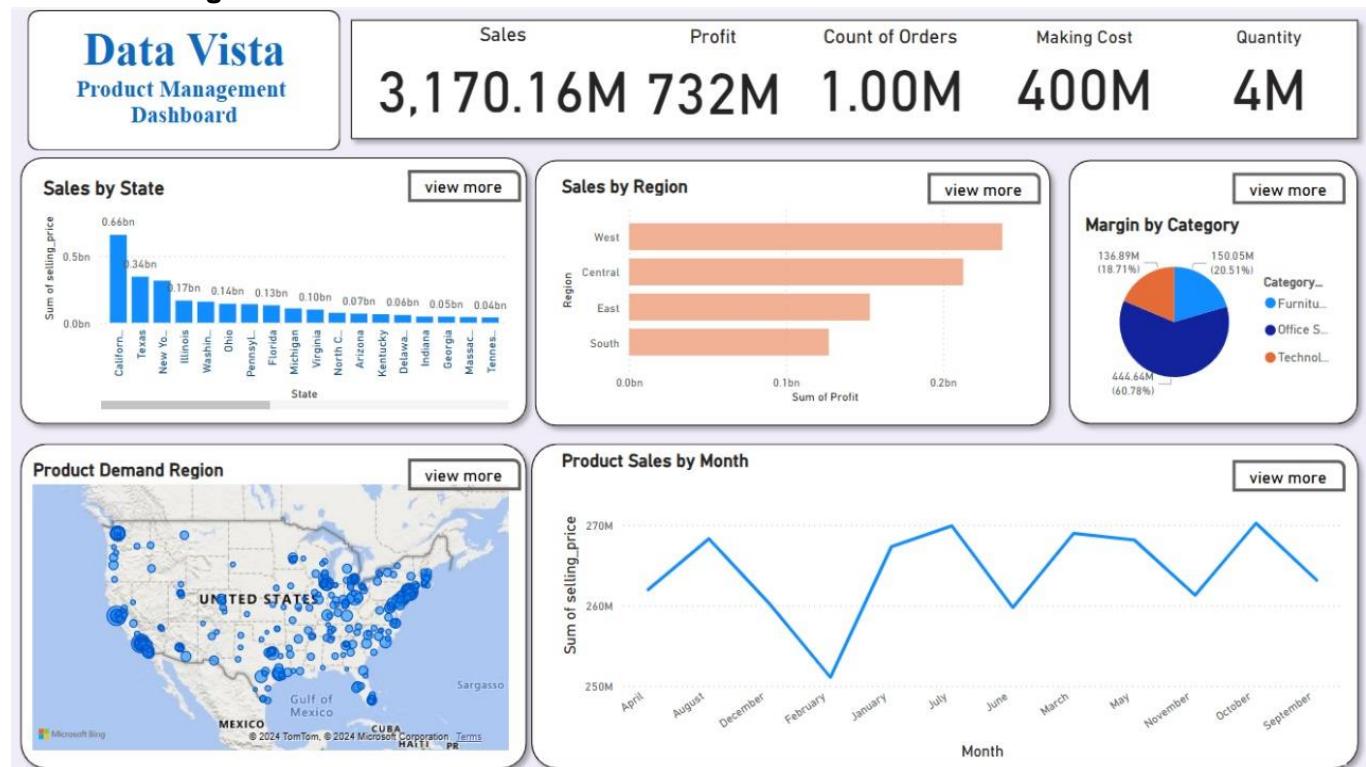
Imagine you have a sales dataset with data from different regions. You want each regional manager to only see the sales data for their respective region. You can create a role for each region and apply a DAX filter that limits data to that region. Assign each manager to their respective role, ensuring they only see the relevant data.

Benefits of Row-Level Security

- **Data Security:** Ensures sensitive information is only accessible to authorized users.
- **Compliance:** Helps meet regulatory requirements by enforcing strict data access controls.
- **Personalized Insights:** Users see only the data relevant to them, enhancing the relevance and accuracy of insights

- In summary, row-level security in Power BI is a powerful feature that allows for precise control over data visibility, ensuring users access only the data they are authorized to see. This is essential for data security, compliance, and providing targeted insights to users.

Product Management Dashboard



OrderID	OrderDate	CategoryName	Sum of Quantity	Sum of shipping_cost	Sum of margin	Sum of selling_price
6	Tuesday, July 02, 2024	Furniture	9	90	1,306.80	5,662.80
15	Tuesday, June 04, 2024	Furniture	3	30	558.72	2,421.12
19	Sunday, December 22, 2024	Furniture	1	10	42.96	186.16
21	Friday, April 12, 2024	Furniture	2	20	227.76	986.96
22	Sunday, October 27, 2024	Furniture	6	60	1,123.92	4,870.32
23	Saturday, May 04, 2024	Furniture	3	30	745.56	3,230.76
Total			3735767	37357670	721.60M, 551.60	3,126,935,723.60

CategoryName	SubCategoryName	ProductID	ProductName	Region
Office Supplies	Envelopes	1000464	#10- 4 1/8 x 9 1/2 Recycled Envelopes	Central
Office Supplies	Envelopes	1000464	#10- 4 1/8 x 9 1/2 Recycled Envelopes	Central
Office Supplies	Envelopes	1000464	#10- 4 1/8 x 9 1/2 Recycled Envelopes	Central
Office Supplies	Envelopes	1000464	#10- 4 1/8 x 9 1/2 Recycled Envelopes	East
Office Supplies	Envelopes	1000464	#10- 4 1/8 x 9 1/2 Recycled Envelopes	East
Office Supplies	Envelopes	1000464	#10- 4 1/8 x 9 1/2 Recycled Envelopes	East
Office Supplies	Envelopes	1000464	#10- 4 1/8 x 9 1/2 Recycled Envelopes	East
Office Supplies	Envelopes	1000464	#10- 4 1/8 x 9 1/2 Recycled Envelopes	East
Office Supplies	Envelopes	1000464	#10- 4 1/8 x 9 1/2 Recycled Envelopes	West
Office Supplies	Envelopes	1000464	#10- 4 1/8 x 9 1/2 Recycled Envelopes	West
Office Supplies	Envelopes	1000911	#10- 4 1/8 x 9 1/2 Security-Tint Envelopes	Central
Office Supplies	Envelopes	1000911	#10- 4 1/8 x 9 1/2 Security-Tint Envelopes	Central
Office Supplies	Envelopes	1000911	#10- 4 1/8 x 9 1/2 Security-Tint Envelopes	South
Office Supplies	Envelopes	1000911	#10- 4 1/8 x 9 1/2 Security-Tint Envelopes	South
Office Supplies	Envelopes	1000911	#10- 4 1/8 x 9 1/2 Security-Tint Envelopes	West
Office Supplies	Envelopes	1000911	#10- 4 1/8 x 9 1/2 Security-Tint Envelopes	West

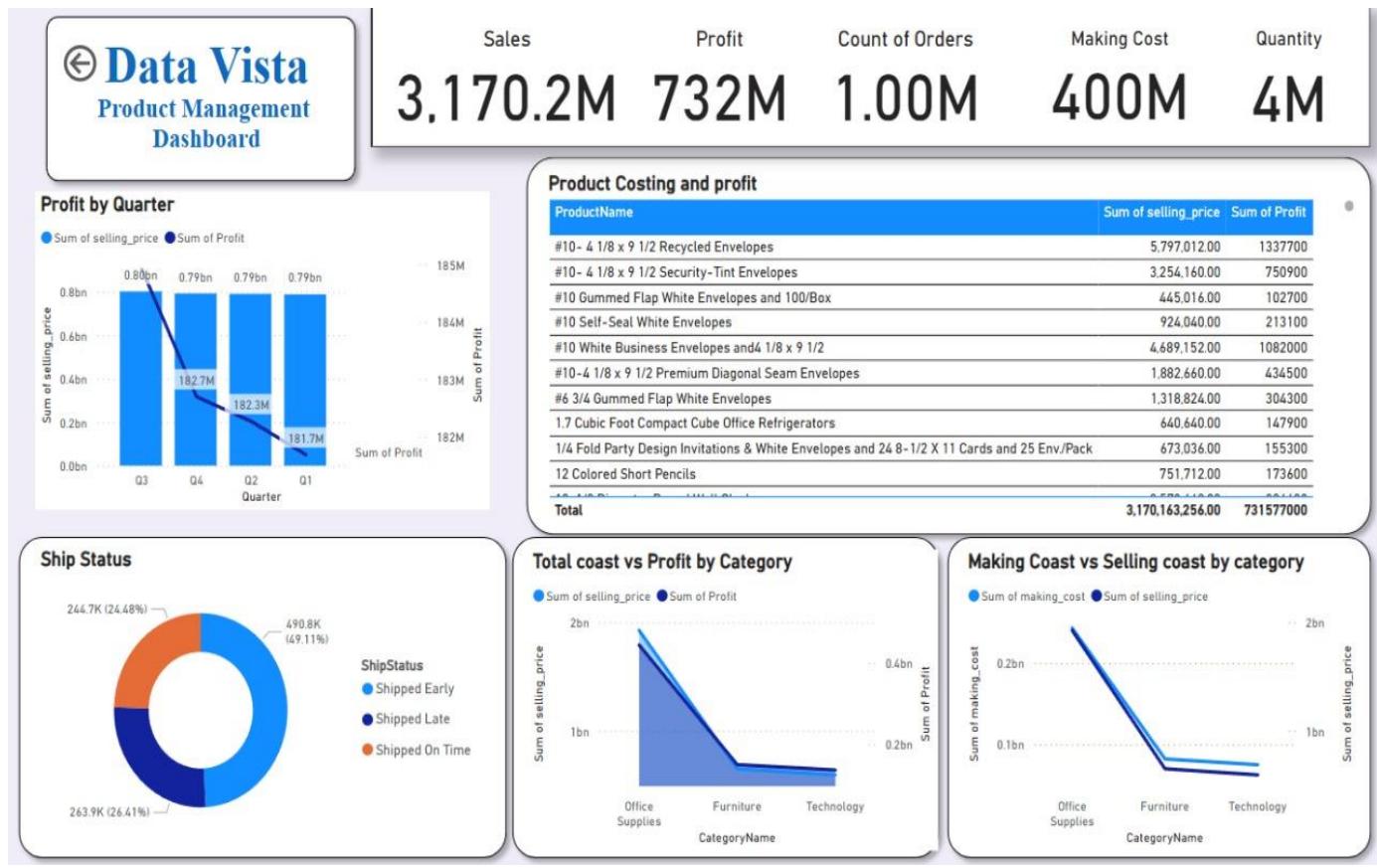
California

Having Highest Sales

Order Date

1/6/2024 12/30/2024

Region	
Central	South
East	West



Week 5: Advanced Analysis Features

Designed dashboards for advanced analysis features (e.g., trend analysis, forecasting).
Test endpoints and used filters on the dashboards.
Implement additional database tables for advanced analysis requirements.
Implement advanced analysis features in the backend.
Enhance frontend modules to incorporate advanced analysis results.

Week 6: Integration and Testing Phase II

Integrate advanced analysis modules.
Test integration and ensure accurate results.
Implement frontend modules for displaying advanced analysis results.
Execute test cases for advanced analysis features.

Approach Towards Adding Filters In Power BI, you can achieve the functionality you described using slicers for filtering and drill-through functionality for providing more detailed tabular data upon selection. Here's how you can do it:

1. Create a Slicer:

- In Power BI Desktop, navigate to the "Visualizations" pane.
- Select the slicer visualization (it looks like a dropdown menu) and add it to your report canvas.
- Drag the relevant field (e.g., "Sale Date", "Product", etc.) into the slicer to create a filter.

2. Connect Slicer to Data:

- Ensure that the slicer is connected to your dataset. You can do this by selecting the slicer, then dragging the appropriate field from your data model into the "Fields" well of the slicer.

3. Add Tabular Data Visualization:

- Add a table visualization to your report canvas to display the initial tabular data.
- Drag the necessary fields into the table to display the relevant information (e.g., sales amount, customer names, etc.).

4. Implement Drill-Through:

- Right-click on a data point within the table visualization and select "Drill Through" > "New Drill Through".
- In the "Fields" pane, select the fields you want to include in the detailed tabular view.
- Create a new page in your report and add a table visualization to display the detailed tabular data.
- Ensure that the new table visualization is filtered based on the slicer selection from the previous page.

5. Add "See More" Functionality (Optional):

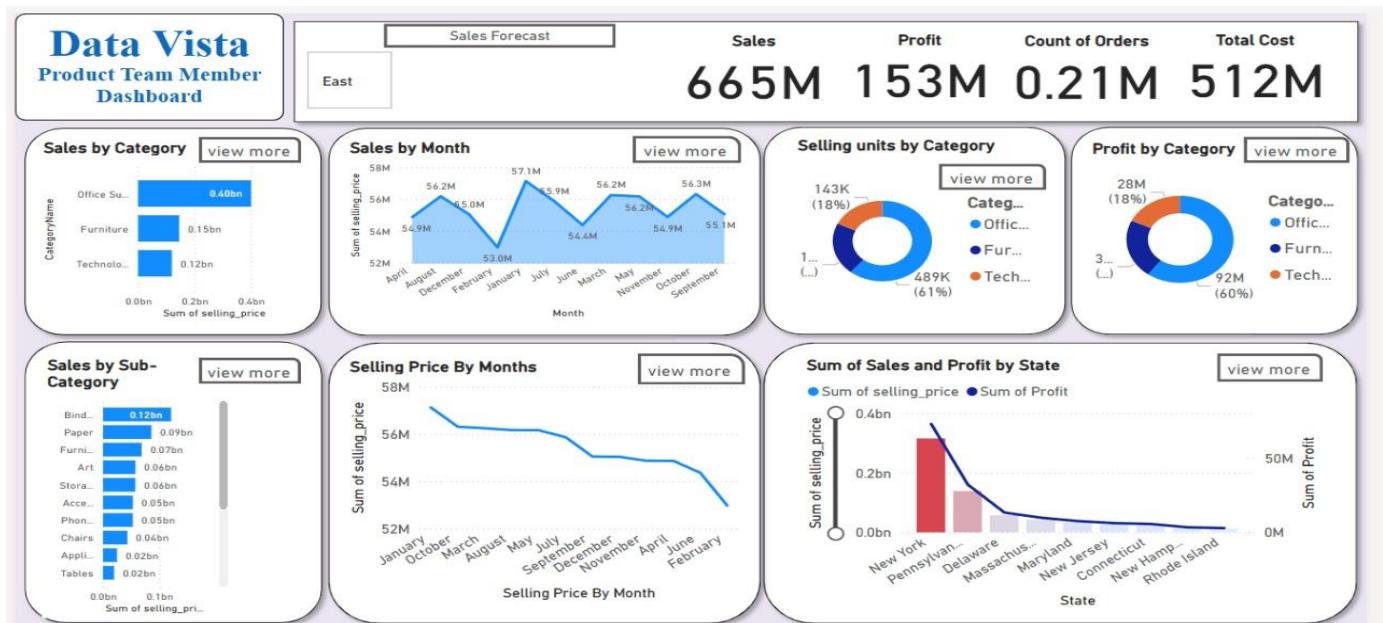
- You can add a button or link to the initial tabular data visualization to provide users with the option to navigate to the detailed tabular view.
- Add an action to the button or link to navigate to the page with the detailed tabular data.

6. Test and Publish:

- Test your Power BI report to ensure that the slicer filtering and drill-through functionality work as expected.
- Once you're satisfied with the report, publish it to Power BI Service or share it with your intended audience.

By following these steps, you can create a Power BI report with slicer filtering and drill-through functionality to provide users with more detailed tabular data upon selection. Remember to consider the user experience and ensure that your report is intuitive and easy to navigate.

Product Team Member Dashboard



Data Vista
Product Team Member Dashboard

Sales by Category

Sum of Quantity	Sum of selling price	Sum of Total cost	Margin by Sub-Category
802K	665.1M	511.6M	19.56M

CategoryName	SubCategoryName	ProductName	Sum of Quantity	Sum
Office Supplies	Binders	Zipper Ring Binder Pockets	600	
Office Supplies	Binders	Xtralife ClearVue Slant-D Ring Binder and White and 3	400	
Office Supplies	Storage	X-Rack File for Hanging Folders	700	
Office Supplies	Paper	Xerox Color Copier Paper and 11 x 17 and Ream	500	
Office Supplies	Paper	Xerox 232	500	
Office Supplies	Paper	Xerox 231	400	
Office Supplies	Paper	Xerox 230	700	
Office Supplies	Paper	Xerox 23	700	
Office Supplies	Paper	Xerox 228	700	
Office Supplies	Paper	Xerox 227	800	
Office Supplies	Paper	Xerox 224	200	
Office Supplies	Paper	Xerox 221	500	
Office Supplies	Paper	Xerox 220	600	
Office Supplies	Paper	Xerox 22	700	
Office Supplies	Paper	Xerox 219	500	
Office Supplies	Paper	Xerox 216	600	
Office Supplies	Paper	Xerox 214	500	
Office Supplies	Paper	Xerox 213	700	
Office Supplies	Paper	Xerox 212	300	
Office Supplies	Paper	Xerox 211	1600	
Total			802300	

Office Supplies

Having highest sales

SLICER PANEL

CategoryName
All

Selling_price	
169.00	16,027.96

Clear

Sales Manager Dashboard

Data Vista
Sales Manager Dashboard

Sales and Profit by Quarter

Sales	Profit	Sales by Quarter	Profit by Quarter
3,170.2M	732M	791.69M	182.7M

OrderID	CategoryName	Quarter	ProductName	Sum of selling_price	Sum of Profit
1	Office Supplies	Q1	Message Book and Wirebound and Four 5 1/2 X 4 Forms/Pg. and 200 Dupl. Sets/Book	2,656.16	613
2	Office Supplies	Q1	Avery 508	1,513.20	349
3	Office Supplies	Q1	GBC Standard Plastic Binding Systems Combs	924.56	213
4	Office Supplies	Q2	SAFCO Boltless Steel Shelving	670.80	155
5	Office Supplies	Q2	Avery Hi-Liter EverBold Pen Style Fluorescent Highlighters and 4/Pack	3,375.84	779
6	Furniture	Q3	Global Deluxe High-Back Managers Chair	5,662.80	1307
7	Office Supplies	Q1	Dixon Prang Watercolor Pencils and 10-Color Set with Brush	2,627.04	606
8	Office Supplies	Q3	Alliance Super-Size Bands and Assorted Sizes	1,443.52	333
9	Office Supplies	Q1	Ibico Hi-Tech Manual Binding System	2,887.04	666
10	Office Supplies	Q4	Rogers Handheld Barrel Pencil Sharpener	2,999.36	692
11	Office Supplies	Q1	Southworth 25% Cotton Granite Paper & Envelopes	1,418.56	327
12	Office Supplies	Q2	Xerox 225	1,269.84	293
Total			3,170,163,256.00	731,577,000	

Q3

Having Highest Sales & Profit

SLICER PANEL

Quarter
All

Category Name
All

Clear

Week 7: Visualization Design and Implementation

Designed dashboards for visualization features.

Test endpoints .

Implement additional database tables for visualization requirements.

Implement backend functionalities for data visualization.

Enhance frontend modules for interactive and meaningful visualizations.

Week 8: Final Testing and Deployment

Conduct integration testing for the entire system.

Execute system tests and fix any defects.

Create acceptance test cases.

Perform acceptance testing and address any issues.

Finalize the code in the version control system.

Prepare a deployment package.

Deploy the finalized code to production.

Deliverables:

Hosted Power BI Dashboards

Test cases Report with execution results.

Codebase with peer and external reviews.

Demo video showcasing the data analysis and visualization features.

Final Project Report.

Deployed Application Details

- ❖ **Product Manager:-** https://app.powerbi.com/links/kxDIfu0BdV?ctid=7a91c6a1-80fd-4ef9-a673-ab1937ff5239&pbi_source=linkShare

login:-

productmanager1@mbapun13.onmicrosoft.com

password:-

Product@123

- ❖ **Product Team Members:-** https://app.powerbi.com/links/2EJhYYR7bH?ctid=7a91c6a1-80fd-4ef9-a673-ab1937ff5239&pbi_source=linkShare

login:-

eastproduct@mbapun13.onmicrosoft.com

centralproduct@mbapun13.onmicrosoft.com

southproduct@mbapun13.onmicrosoft.com

westproduct@mbapun13.onmicrosoft.com

passwords:-

East@123
Central@123
South@123
West@1234

- ❖ **Sales Manager:-** https://app.powerbi.com/links/u03wmf2wPn?ctid=7a91c6a1-80fd-4ef9-a673-ab1937ff5239&pbi_source=linkShare

login:-
salesmanager1@mbapun13.onmicrosoft.com
password:-
Salesmanager@12

- ❖ **Sales Team Member :-** https://app.powerbi.com/links/54BQSiSbvq?ctid=7a91c6a1-80fd-4ef9-a673-ab1937ff5239&pbi_source=linkShare

login:-
eastsales@mbapun13.onmicrosoft.com
centralsales@mbapun13.onmicrosoft.com
southsales@mbapun13.onmicrosoft.com
westsales@mbapun13.onmicrosoft.com
passwords:-
East@123
Central@123
South@123
West@1234

Project Outcomes:

The Data Vista Project, part of the Infosys Springboard Summer Internship Program, has provided a remarkable platform to delve into advanced data analytics using Power BI. This project has significantly enhanced our technical proficiency and practical understanding of data visualization and business intelligence.

Key Achievements Using Power BI:

1. Data Filtering and Transformation:

- Utilized Power BI's robust data transformation capabilities to clean, filter, and preprocess large datasets.
- Implemented advanced filtering techniques to refine data views, ensuring the relevance and accuracy of the information displayed.

2. Dashboard Creation:

- Designed interactive and user-friendly dashboards that provide comprehensive insights at a glance.
- Leveraged Power BI's drag-and-drop interface to integrate various data sources, ensuring seamless data connectivity.
- Created custom visuals and utilized Power BI's extensive library of visual elements to enhance the dashboards' interactivity and appeal.

3. Data Plotting and Visualization:

- Employed a wide range of plotting options including bar charts, line graphs, scatter plots, and heat maps to effectively communicate data trends and patterns.
- Used Power BI's built-in AI capabilities to generate predictive analytics and trend forecasts, aiding in proactive decision-making.

4. Advanced Analytics:

- Incorporated DAX (Data Analysis Expressions) to perform complex calculations and derive meaningful insights from the data.
- Implemented real-time data updates and live dashboards, ensuring that the data is always current and actionable.

5. Collaboration and Sharing:

- Facilitated team collaboration by sharing dashboards and reports securely within the organization.
- Enabled role-based access controls to ensure data privacy and compliance with organizational policies.

The successful application of Power BI in the Data Vista Project has led to the following outcomes:

- Enhanced decision-making processes through the provision of clear, data-driven insights.
- Improved efficiency in data analysis workflows, reducing the time required to generate reports and insights.

- Fostered a data-centric culture within the team, emphasizing the importance of data accuracy and integrity.

Acknowledgements:

We would like to extend our deepest gratitude to Infosys for providing this invaluable opportunity. The mentorship and resources made available through the Springboard Summer Internship Program have been instrumental in our success. Special thanks to our project mentors and peers for their constant support and collaboration.

The Data Vista Project has been a cornerstone of our professional development, equipping us with the skills and knowledge to excel in the dynamic field of data analytics. We are excited to apply these learnings in our future careers and continue our journey of exploration and innovation in technology.

Conclusion:

The **Data Vista Project**, undertaken as part of *the Infosys Springboard Summer Internship Program*, has been a valuable and enriching experience. Throughout this project, we have developed a deep understanding of data analytics, honed our technical skills, and applied innovative solutions to real-world problems. The collaborative environment and mentorship provided by Infosys have been instrumental in our learning journey, enabling us to achieve significant milestones and deliver actionable insights.

Our findings have not only met the project objectives but also paved the way for future explorations in data science and analytics. The skills and knowledge gained from this internship have prepared us to tackle complex data challenges and contribute effectively to the ever-evolving field of technology.

We extend our gratitude to Infosys, our mentors, and our peers for their unwavering support and guidance. This experience has been a cornerstone in our professional development, and we look forward to leveraging the lessons learned in our future endeavours.

Thank you ☺