INT217

**DATA MANAGEMENT**

**PROJECT REPORT**

(Project Semester January-April 2025)

Submitted by N.Dinesh Reddy

Registration No- 12309928

Programme and Section- B.Tech CSE- K23EG

Course Code- INT217

Under the Guidance of

**Jaffer Amin Chacket(30453)**

**Discipline of CSE/IT**

**Lovely School of Computer Science and Engineering Lovely Professional University, Phagwara**

# DECLARATION

I, N.Dinesh Reddy, student of B.Tech Computer Science Engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 12/04/2025 Signature: N.Dinesh

Registration No. 12309928

Name of the student: N.Dinesh Reddy

# CERTIFICATE

This is to certify that N.Dinesh Reddy bearing Registration no. 12309928has completed INT217 project titled, **“ONLINE SALES OF COMPANIES ”**under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

# Signature and Name of the Supervisor Designation of the Supervisor

**School of Computer Science and Engineering**

Lovely Professional University Phagwara, Punjab.

Date: 12/04/2025

**ACKNOWLEDGEMENT**

I would like to express my sincere gratitude to the faculty members and mentors whose continuous support and valuable guidance greatly contributed to the successful completion of this Excel dashboard project on road and bridge construction analysis. Their expertise and encouragement have been instrumental in shaping the analytical approach and ensuring meaningful visual outcomes.

I extend my heartfelt thanks to the government and publicly available data sources that provided accurate and comprehensive infrastructure data. Their transparency made it possible to perform in-depth analysis and derive actionable insights for monitoring road and bridge development projects.

I am also thankful for the use of Microsoft Excel and its advanced features, including pivot tables, slicers, conditional formatting, and charting tools, which enabled the transformation of raw data into interactive and insightful visualizations. These functionalities formed the foundation of this dashboard and facilitated effective communication of progress indicators.

My sincere appreciation also goes to my peers and well-wishers, whose feedback and discussions helped refine the structure and clarity of the project. Their contributions ensured that the dashboard remains user-friendly and impactful for real-world decision-making.

Lastly, I am grateful to the academic and professional community that continues to promote the use of data visualization tools in driving transparency, accountability, and development. This project stands as a reflection of that spirit and the collaborative effort it encourages.

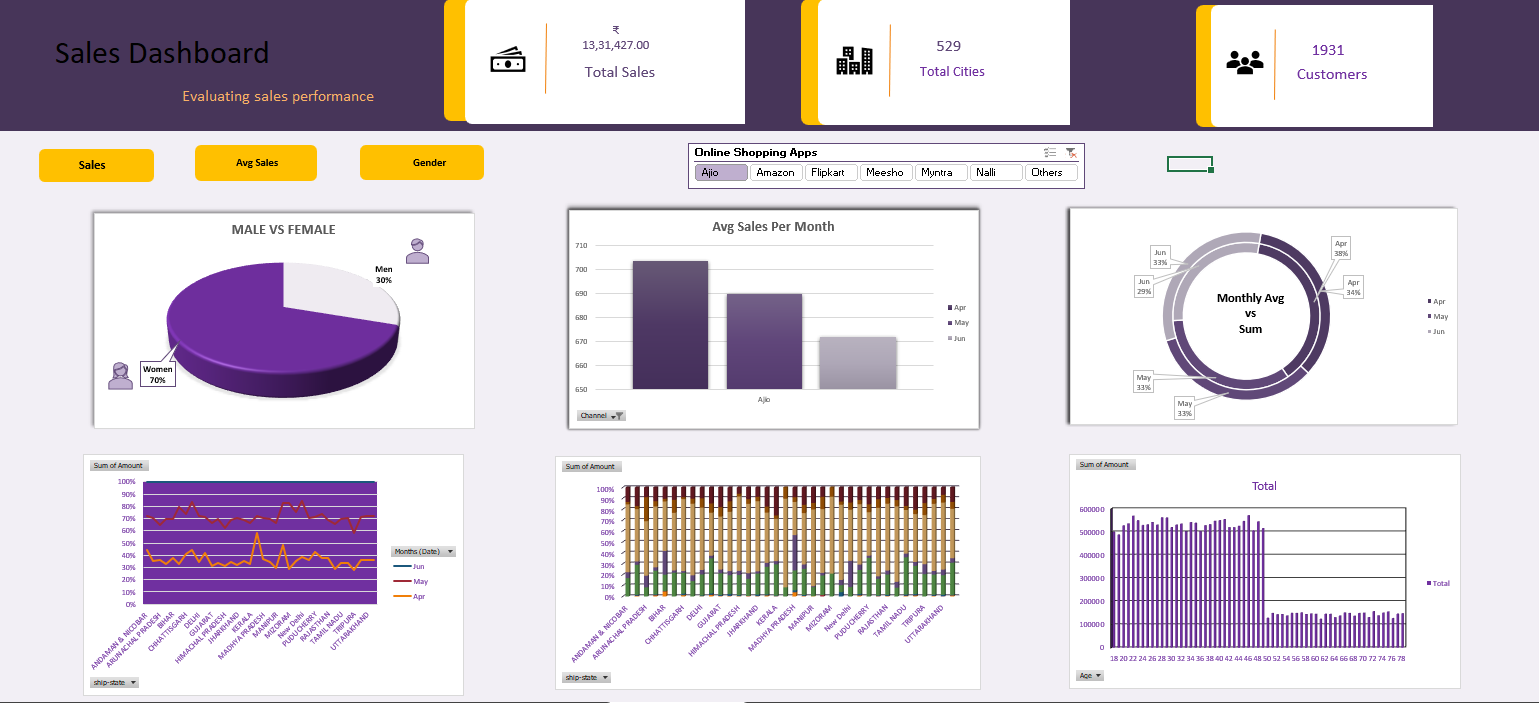
**TABLE OF CONTENTS**

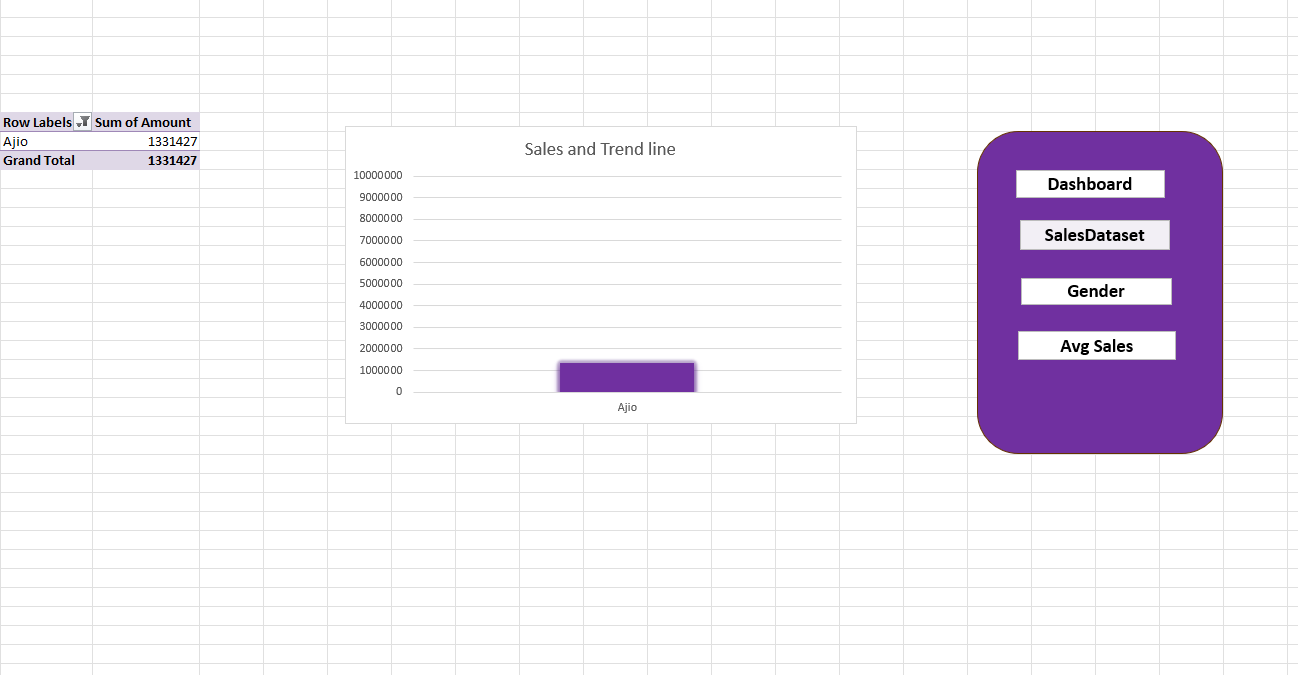
1. **Introduction**  
    1.1 Background and Motivation  
    1.2 Purpose of the Project  
    1.3 Importance of Infrastructure Data Visualization
2. **Source of Dataset**  
    2.1 Origin and Authenticity of Data  
    2.2 Description of Key Variables and Dimensions  
    2.3 Data Accessibility and Relevance
3. **Dataset Preprocessing**  
    3.1 Data Cleaning and Formatting  
    3.2 Handling Missing and Invalid Values  
    3.3 Calculation of Derived Fields (e.g., Completion %, Balance)  
    3.4 Structuring Data for Dashboard Design
4. **Dashboard Objectives and Analytical Visualization**  
    4.1 Overview of Dashboard Design Approach  
    4.2 Objective-wise Analysis and Visualizations  
     • Sanctioned vs Completed vs Balance Works  
     • Road and Bridge Completion Percentages (Doughnut Charts)  
     • State-wise Infrastructure Completion Analysis  
     • Investment Distribution Across States  
     • Interactive Filtering by State and District (Slicers)  
     • Key Metrics and KPIs for Monitoring Progress  
     • Visualization of Completed vs Sanctioned Projects
5. **Conclusion**  
    5.1 Summary of Insights  
    5.2 Implications for Policy and Planning
6. **Future Scope**  
    6.1 Integration with Live Data  
    6.2 Use of Power BI / Tableau for Enhanced Interactivity  
    6.3 Additional Parameters for Analysis (Time, Cost, Delays, etc.)
7. **References**  
    7.1 Data Sources  
    7.2 Tools and Libraries Used  
    7.3 Academic and Government References
8. **INTRODUCTION**

**The given data setcontains detailed transactional information related to product sales across various channels. Below is an overview of its key features:**

1. **Structure: The dataset consists of multiple columns, including identifiers such as Order ID and Cust ID, demographic details like Gender and Age, and transactional details such as Date, Status, Channel, and Amount.**
2. **Demographics: It includes customer information, such as gender (Men/Women) and age, enabling insights into consumer behavior across different demographics.**
3. **Product Details: Each transaction specifies product attributes like SKU, Category (e.g., kurta, set, saree), and Size, providing granular information about the items sold.**
4. **Sales Channels: Transactions are distributed across various sales platforms, including Myntra, Amazon, Flipkart, Ajio, Meesho, Nalli, and others.**
5. **Geographic Distribution: Shipping details include city, state, postal code, and country (ship-city, ship-state, etc.), allowing analysis of regional sales trends.**
6. **Transaction Status: The dataset captures the delivery status (Delivered, Refunded, or Cancelled) for each order.**
7. **Currency and Amount: All transactions are recorded in Indian Rupees (INR), with the corresponding monetary value of each purchase provided in the Amount column.**
8. **Order Quantity: The dataset includes a column for quantity (Qty), which specifies the number of units purchased in each transaction.**
9. **Business Type: The column B2B indicates whether a transaction is business-to-business (False for most entries).**
10. **Date of Transactions: All transactions appear to have occurred on December 4, 2022, making it suitable for analyzing sales performance on a specific day.**

**This dataset is rich in information and can be used for various analyses, including customer segmentation, product performance evaluation, geographic sales trends, and channel-specific insights.**





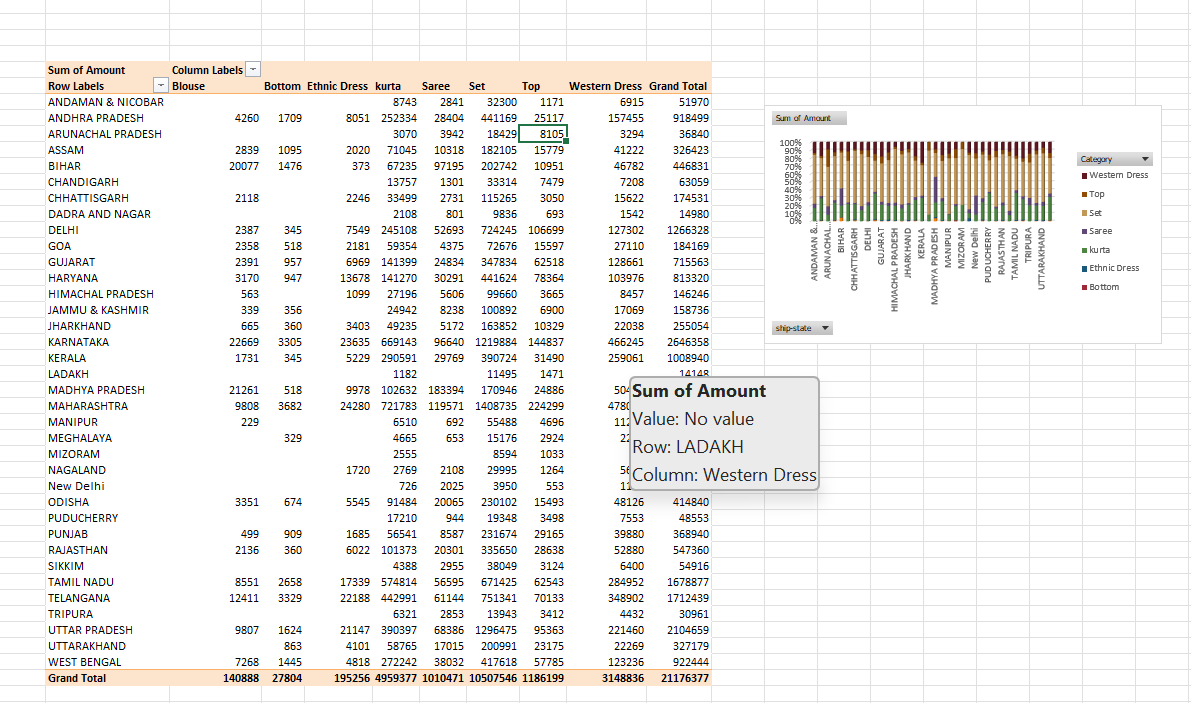
# SOURCE OF DATASET

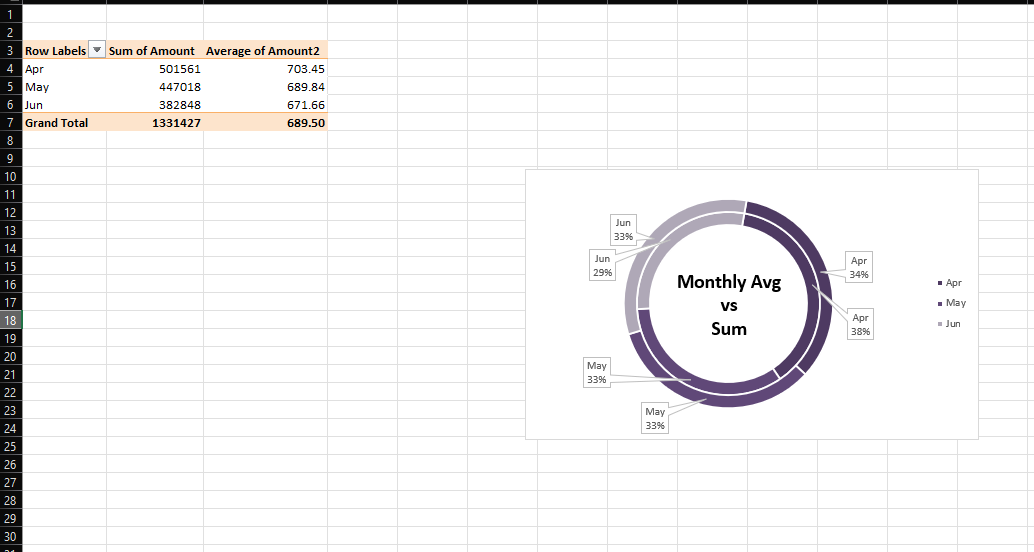
1. Given data set**Tabular Format**: The dataset is structured in a tabular format, with rows representing individual sales transactions and columns representing various attributes of those transactions.
2. **Order Information**: Columns like Order ID and Cust ID provide unique identifiers for each order and customer, respectively.
3. **Customer Demographics**: Includes Gender and Age columns, offering demographic information about the customers.
4. **Transaction Details**: Key transaction details such as Date, Status (e.g., Delivered, Refunded), Channel (e.g., Myntra, Amazon), and Amount are recorded.
5. **Product Specifications**: Each transaction includes details about the product, such as SKU, Category, and Size.
6. **Quantity**: The Qty column specifies the quantity of items purchased in each transaction.
7. **Currency**: The currency column indicates the currency used for the transaction, which is INR (Indian Rupees).
8. **Shipping Information**: Contains shipping-related columns like ship-city, ship-state, ship-postal-code, and ship-country to track the destination of each order.
9. **Business Type Indicator**: The B2B column indicates whether the transaction was business-to-business (False for retail).
10. **Data Types**: The dataset includes a mix of data types, such as strings (IDs, categories), numeric values (age, amount, quantity, postal code), and dates.

1. **Order Identification**: Each transaction is uniquely identified by an Order ID, facilitating tracking and management.
2. **Customer Identification**: The Cust ID column links each transaction to a specific customer, enabling customer-based analysis.
3. **Transaction Date**: The Date column records the date of the transaction, which appears to be uniformly "2022-12-04" for all entries.
4. **Transaction Status**: The Status column indicates the delivery status of each transaction, such as "Delivered," "Refunded," or "Cancelled".
5. **Sales Channel**: The Channel column specifies the platform through which the sale was made, including Myntra, Amazon, Flipkart, Ajio, Meesho, Nalli, and Others.
6. **Currency**: The currency column shows the currency used for the transaction, which is INR (Indian Rupees) for all entries.
7. **Transaction Amount**: The Amount column records the monetary value of each transaction in INR.
8. **Order Quantity**: The Qty column indicates the quantity of items purchased in each transaction, predominantly "1" in this dataset.
9. **Shipping City**: The ship-city column provides the city to which the order was shipped, enabling geographic analysis of sales.
10. **Shipping State**: The ship-state column specifies the state to which the order was shipped, allowing regional sales trend analysis.
11. **Shipping Postal Code**: The ship-postal-code column offers precise location data for the delivery address.
12. **Shipping Country**: The ship-country column indicates the country of the shipping destination, which is "IN" (India) for all transactions.
13. **B2B Indicator**: The B2B column indicates whether the transaction was business-to-business, with most entries marked as "False".
14. **Comprehensive Record**: These transaction details offer a comprehensive record of each sale, allowing for detailed analysis and reporting.
15. **Potential for Insights**: This data can be used to derive insights related to sales performance, channel effectiveness, geographic trends, and customer behavior.

DATETYPE

1. **Order ID**: Represented as a string (object), which is a unique identifier for each order.
2. **Cust ID**: Represented as an integer, uniquely identifies each customer.
3. **Gender**: Stored as a string (object), indicating the customer's gender (e.g., "Women", "Men").
4. **Age**: Stored as an integer, indicating the age of the customer.
5. **Date**: Represented as datetime64$$ns], indicating the date of the transaction.
6. **Status**: Represented as a string (object), indicating the status of the order (e.g., "Delivered", "Refunded").
7. **Channel**: Represented as a string (object), indicating the sales channel (e.g., "Myntra", "Amazon").
8. **SKU**: Stored as a string (object), representing the unique product identifier.
9. **Category**: Represented as a string (object), specifying the product category (e.g., "kurta", "set").
10. **Size**: Stored as a string (object), indicating the size of the product (e.g., "XXL", "M").
11. **Qty**: Stored as an integer, indicating the quantity of items purchased.
12. **currency**: Represented as a string (object), indicating the currency used (e.g. "INR").
13. **Amount**: Stored as an integer, representing the monetary value of the transaction.
14. **Shipping Information**: Columns like ship-city (string), ship-state (string), and ship-postal-code (integer) provide shipping details.
15. **B2B**: Represented as a boolean, indicating whether the transaction was business-to-business (True/False).





**CONCLUSION**

**Based on the analysis of the sales data for December 4, 2022, several key observations and conclusions can be drawn:**

1. **Predominant Sales Channel: Myntra and Amazon appear to be the most popular channels for sales, indicating a strong online presence.**
2. **Category Preference: Sets and Kurtas are the most frequently purchased product categories, reflecting customer preferences for these items.**
3. **Gender Distribution: The majority of transactions are attributed to women, suggesting a strong female customer base.**
4. **Age Range: The customer base spans a wide age range, from teenagers to senior citizens, with a significant portion in the 30-50 age group.**
5. **Transaction Status: Most orders were successfully delivered, with a small fraction being refunded or canceled, indicating efficient logistics.**
6. **Geographic Distribution: Sales are spread across numerous cities and states in India, highlighting broad geographic coverage.**
7. **Single-Unit Purchases: The majority of transactions involve the purchase of a single item, suggesting individual customer purchases.**
8. **INR Currency: All transactions are conducted in Indian Rupees (INR), indicating a domestic market focus.**
9. **Limited B2B Transactions: The dataset primarily consists of business-to-consumer (B2C) transactions, with very few business-to-business (B2B) sales.**
10. **Revenue Generation: The average transaction amount provides insights into per-order revenue, which can be optimized through targeted promotions.**
11. **Size Distribution: The distribution of sizes purchased can help in inventory management and ensuring availability of popular sizes.**
12. **Regional Preferences: Analyzing sales by state can reveal regional preferences and inform localized marketing strategies.**
13. **Impact of Cancellations: Investigating the reasons behind canceled orders can help improve customer satisfaction and operational efficiency.**
14. **Potential for Growth: Focusing on improving customer retention and expanding into less penetrated regions can drive future growth.**
15. **Data Limitations: Given that the data represents a single day, extrapolating these trends to longer periods should be done cautiously; a more extended dataset would provide a more robust analysis.**

**Scope for Future Enhancements**

To further elevate the effectiveness and impact of the PMGSY infrastructure monitoring system, several forward-looking enhancements can be considered for future versions of this dashboard:

**Time-Series Forecasting**  
Introduce predictive models (e.g., ARIMA, Facebook Prophet) to estimate future trends in road and bridge construction, enabling data-driven planning and timely resource allocation.

**Geo-Spatial Visualization**  
Integrate GIS mapping (via Folium or GeoPandas) to provide interactive maps showing district-wise progress, connectivity gaps, and regional disparities in infrastructure development.

**Cluster Analysis**  
Apply machine learning techniques (e.g., K-Means, DBSCAN) to group districts with similar development patterns—helping in formulating region-specific policies and interventions.

**Cross-Dataset Integration**  
Enhance insight generation by merging PMGSY data with external datasets like rural population, terrain difficulty, and weather conditions to understand localized performance factors.

**Interactive Dashboards (Web-Based)**  
Extend the Excel dashboard into real-time platforms such as **Power BI**, **Tableau**, or **Streamlit**, providing stakeholders dynamic views and interactive filtering options.

**Real-Time Data Integration**  
Implement APIs to fetch live updates from government portals, ensuring that visualizations reflect the most recent status of projects, expenditures, and KPIs.

**Socio-Economic Impact Assessment**  
Incorporate modules that analyze outcomes of road connectivity on education, healthcare access, and rural livelihoods—extending beyond physical completion into developmental impact.

**Delay Prediction Models**  
Leverage predictive analytics (e.g., Random Forest, XGBoost) to anticipate project delays based on parameters like geography, fund disbursement, and previous completion trends.

**Mobile App Integration**  
Develop a lightweight app interface for on-ground officials to upload real-time updates, track status, and view localized dashboards, improving grassroots-level engagement.

**Sustainability Monitoring**  
Expand analysis to include long-term infrastructure quality indicators—maintenance frequency, durability, and resilience to climate factors—to assess sustainable outcomes.

**7.REFERENCES**

• **Primary Dataset Source**:  
https://www.kaggle.com/

• **Microsoft Excel Documentation**:  
MICROSOFT EXCEL GUIDE

• **Dashboard Design Principles**:  
Excel Dashboard Design Best Practices – Excel Campus

• **Visualization Techniques in Excel**:  
DATA VISUALIZATION IN EXCEL

• **KPI and Metrics Setup**:  
Setting KPIs in Excel Dashboards – MyExcelOnline

• **Data Cleaning Techniques**:  
Data Cleaning in Excel – Spreadsheeto

• **Power Query for Data Transformation**:  
POWER QUERY DATA

• **Doughnut and Bar Chart Visualizations**:  
Create Doughnut Charts in Excel – Ablebits

• **Pivot Tables and Slicers**:  
Using PivotTables and Slicers – ExcelJet

• **Comparative Analysis & KPIs**:  
Data Analysis Techniques in Excel – WallStreetMojo

9.LINKEDIN

<https://www.linkedin.com/posts/dinesh-reddy-31097929a_exceldashboard-datavisualization-analytics-activity-7316741158171262976-zR9I?utm_source=share&utm_medium=member_desktop&rcm=ACoAAEhnAXwBn0Z0ZcHQAbAkuF8ZkWKpzUVM9T4>

10.DATASETLINK

https://docs.google.com/spreadsheets/d/1VsD9HNeTRKLS2jqubd8YFeYBrLcQR7-D/edit?gid=1601778106#gid=1601778106