

R&D Documentation on PowerBI

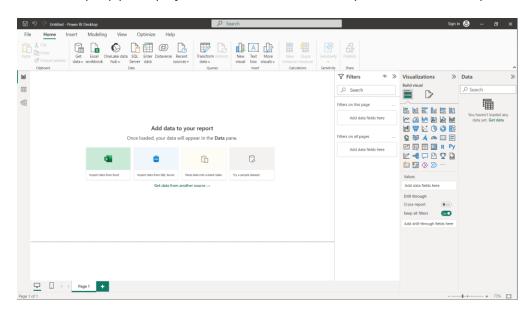
1	What is Microsoft power BI	3
2	Common uses of power BI	3 - 5
3	Data sources	5 - 7
4	Methodology	8
5	Steps involved in project	8 - 17
6	Data visualization & chart designing	17 - 19
7	Challenges faced	19 - 20
8	Conclusion	20

By

ANIRUDH SHARMA

Microsoft Power BI is a business intelligence (BI) platform that provides nontechnical business users with tools for aggregating, analyzing, visualizing and sharing data. Power BI's user interface is intuitive for users familiar with Excel, and its deep integration with other Microsoft products makes it a versatile self-service tool that requires little upfront training.

Users can download an application for Windows 10, called Power BI Desktop, and native mobile apps for Windows, Android and iOS devices. There is also Power BI Report Server for companies that must maintain their data and reports on premises. That version of Power BI requires a special version of the desktop app -- aptly called Power BI Desktop for Power BI Report Server.



66

Common uses of Power BI

"

Microsoft Power BI is used to find insights within an organization's data. Power BI can help connect disparate data sets, transform and clean the data into a data model and create charts or graphs to provide visuals of the data. All of this can be shared with other Power BI users within the organization.

The data models created from Power BI can be used in several ways for organizations, including the following:

telling stories through charts and data visualizations;

examining "what if" scenarios within the data; and

creating reports that can answer questions in real time and help with forecasting to make sure departments meet business metrics.

Power BI can also provide executive dashboards for administrators or managers, giving management more insight into how departments are doing.

In the modern era, data-driven decisions are pivotal for the success of businesses. As data sources and volumes continue to grow, there's an increasing need for tools that can transform this sea of data into actionable insights. This is where Power BI steps in.

Components: Power BI comprises several elements:

- **Power BI Desktop:** A Windows desktop application for PCs and desktops, primarily used for designing and publishing reports.
- **Power BI Service:** Often referred to as "Power BI online", it's a cloud-based platform where users can share, collaborate, and publish their reports and dashboards.
- **Power BI Mobile:** An app for iOS and Android devices, allowing users to access and interact with their reports and dashboards on the go.
- **Power BI Gateway:** A bridge that provides quick and secure data transfer between the onpremises data sources and Power BI Service.
- **Power BI Embedded:** An Azure service that allows developers to embed reports and visuals into applications.

Data Transformation and Modeling: One of Power BI's strengths is its ability to connect to a wide range of data sources, from simple Excel sheets to complex cloud-based databases. Once connected, users can perform data transformation using the Power Query Editor and model data to define relationships, hierarchies, and metrics.

DAX (Data Analysis Expressions): This is Power BI's formula language. It allows users to create custom calculations, aggregations, and measures, thus amplifying the analytical power of the tool.

Visualization: With a simple drag-and-drop interface, users can build a range of visuals from basic bar charts to complex custom visuals. These visuals can then be combined into reports and dashboards, providing a cohesive storytelling medium.

Collaboration: Power BI makes sharing insights a breeze. Users can share reports and dashboards with others, collaborate with colleagues in real-time, and even set up automated data refresh schedules.

Flexibility and Scalability: Whether you're a small business with minimal data sources or a large enterprise with vast amounts of data spread across various databases, Power BI can cater to your needs. With its integration capabilities with other Microsoft products and various third-party tools, it offers both flexibility and scalability.

66

Data sources

"

Power BI provides connectivity to a wide range of data sources. These data sources are continually expanding as Microsoft updates Power BI to accommodate a broader set of connectors and direct query capabilities. Below are categories of data sources you can connect to using Power BI:

Files:

- Excel
- CSV
- XML
- JSON
- Text
- SharePoint Folder
- And more
- Databases:
- SQL Server
- MySQL
- PostgreSQL
- Oracle
- IBM Db2
- Sybase
- Amazon Redshift
- Snowflake
- Azure SQL Database
- Azure Data Lake Storage
- Google BigQuery
- And more

Power Platform:

- PowerApps (Common Data Service)
- Power Automate
- Power BI datasets
- Power BI dataflows

Azure:

- Azure SQL Data Warehouse
- Azure Blob Storage
- Azure Table Storage
- Azure HDInsight
- Azure Cosmos DB
- Azure Data Explorer
- And others
- Online Services:
- SharePoint Online List
- Dynamics 365
- Microsoft Exchange Online
- Google Analytics
- Salesforce
- GitHub
- Adobe Analytics
- And many more

Other:

- Web (via web scraping or APIs)
- ODBC
- R script and Python script
- OData feed
- SharePoint List (on-premises)
- Active Directory
- Hadoop (HDFS)
- Spark
- Advanced & Specialized Data Connectors:
- PDF
- Web By Example
- SAP HANA
- SAP Business Warehouse Server
- Teradata
- And more

Third-Party Service Connectors: Microsoft partners with several third-party vendors, which means there are a multitude of additional connectors available, whether it's for CRM systems, ERPs, marketing platforms, or niche database systems.

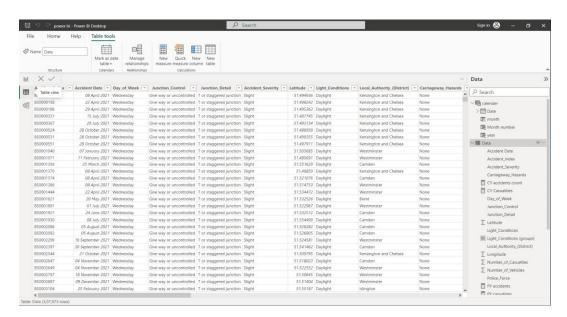
Streaming Data:

- Power BI Streaming datasets
- Azure Stream Analytics

Custom Connectors: If the available connectors don't meet your needs, Power BI provides an SDK where you can build your custom connector.

Power BI's ability to connect to such a vast array of data sources, both on-premises and cloud-based, is one of its significant advantages. It allows organizations to consolidate and visualize data from disparate systems, offering a comprehensive view for analytics and decision-making. Always keep the Power BI documentation handy or visit the official Power BI website to stay updated on the latest supported data sources.

Here I've taken random data about **road accidents** and made data in Excel doc. This is the data.





Methodology:

. 99

For the data I've taken, I've followed these methodologies:

- How to connect to raw data/flat file
- Data cleaning in power query
- Data processing
- Time intelligence function/ calendar date table in power BI
- Data modelling (Relationship between multiple tables)
- YTD and YOY growth calculations using DAX
- KPI and advanced KPI generations
- Creating custom columns and measures in the report
- Importing images
- Creating different charts and generating insights
- Export the report to users

Steps in my project:

- Requirement gathering
- Raw data overview
- Connecting data with power BI
- Data cleaning
- Data processing
- > Data modelling
- > Background design in power point
- Data visualization/ charts design
- Report/ dashboard building
- > insights

Requirement gathering:

Requirement gathering is a critical phase in any project, including those that involve Power BI. In the context of Power BI, requirement gathering refers to the process of understanding and documenting the needs and expectations of stakeholders for a Power BI report or dashboard. This involves understanding what data sources are available, what insights the business wants to extract, how those insights should be presented, and who the intended audience is.

Here's a breakdown of what requirement gathering might entail for a Power BI project:

Business Objectives: Understand the main business goals or questions that the Power BI report or dashboard aims to address. This helps in aligning the project with strategic goals.

Data Sources: Identify where the data is coming from:

- What are the primary and secondary data sources?
- Are they on-premises, in the cloud, or a mix of both?
- Do you have access to these sources, and are there any permissions or security considerations?

Key Metrics & KPIs: Understand the primary metrics or Key Performance Indicators (KPIs) the business wants to track. This could range from sales figures, customer churn rates, operational efficiencies, and more.

Visual Preferences:

Do stakeholders have a preference for specific types of visuals or charts?

- Are there corporate branding or color schemes that need to be adhered to?
- User Interactivity: Understand if users want to interact with the report:
- Will there be a need for slicers or filters?

Do they want drill-down or drill-through capabilities?

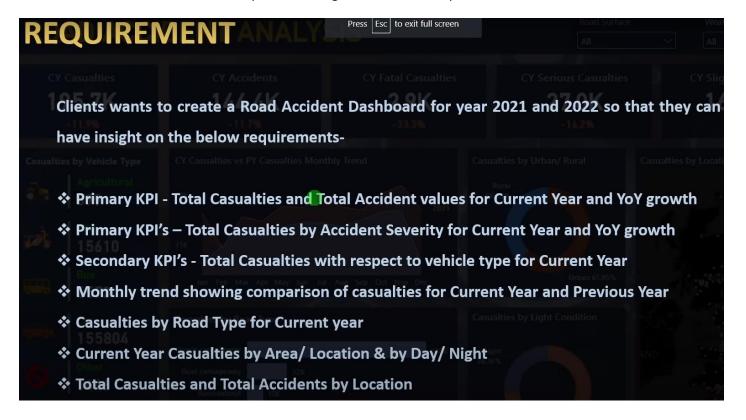
Frequency of Data Refresh: Understand how often the data needs to be updated. Is it real-time, daily, weekly, or monthly? This impacts how you'd set up data refreshes in Power BI.

Security & Sharing:

- Who are the intended viewers of the report or dashboard?
- Are there any row-level security requirements?
- How will the report be shared through Power BI service, embedded in an app, or some other means?

Integration with Other Tools: Determine if there's a need to integrate the Power BI reports or dashboards with other tools or platforms, like SharePoint, Teams, or custom web applications.

Feedback Loop: Establish a mechanism for stakeholders to provide feedback on initial designs or iterations. This ensures the final product aligns with their expectations.



Raw data overview:

In Power BI, the term "raw data" usually refers to the original, unprocessed data that you source from various data repositories before any transformation, cleaning, or modeling has been applied to it. Understanding your raw data is crucial for several reasons, including ensuring data accuracy, understanding the data's structure, and planning the necessary transformations to derive meaningful insights.

Here's an overview of working with raw data in Power BI:

Importing Raw Data:

- Power BI supports a wide variety of data sources, including databases, online services, Excel files, CSVs, and many others.
- When connecting to these data sources, you're often accessing the raw, untouched data.

Previewing Raw Data:

- Once you connect to a data source, Power BI allows you to preview the data before loading it into the model.
- This preview gives you a glimpse of the data's structure, the types of fields available, and any potential issues or inconsistencies.

Understanding Data Quality:

- Raw data can often contain missing values, duplicates, or inaccuracies.
- By examining the raw data in Power BI's query editor, you can identify and rectify these quality issues.

Planning Transformations:

Based on the raw data's structure and quality, you can determine the necessary transformations, such as:

- Removing or replacing null values.
- Splitting columns.
- Changing data types.
- Adding calculated columns or measures.
- Power BI's Power Query Editor provides an intuitive interface to perform these transformations.

Data Modeling:

Once you've assessed the raw data and made initial transformations, you'll proceed to model the data, which might include:

- Defining relationships between tables.
- Setting up hierarchies.
- Creating calculated columns and measures using DAX.

Documentation:

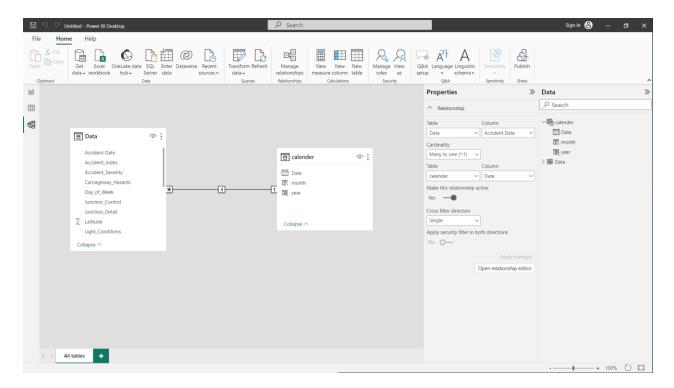
It's good practice to document the state and structure of the raw data, including any anomalies or quirks you notice. This will help in future data refreshes or if other team members work on the project.

Refreshing Raw Data:

- Depending on how frequently the raw data changes or updates, you might need to set up scheduled refreshes in Power BI. This ensures that your reports and dashboards reflect the most recent data.
- Always monitor these refreshes for potential issues, especially if the structure of the raw data source changes.

Connecting data with power BI

Connecting data to Power BI is the fundamental step to start any analytics project within the platform. Power BI provides multiple ways to connect to different data sources, be it databases, files, online services, or streaming data.



Data cleaning & processing:

Data cleaning and processing are essential steps in the data preparation process, ensuring that data is accurate, consistent, and ready for analysis. "Garbage in, garbage out" is a common saying in data science, emphasizing the importance of quality input data for quality outputs. Here's an overview of data cleaning and processing:

Data Cleaning:

Handling Missing Values:

- Identify and address missing data.
- Options include deleting rows/columns with missing data, imputing missing values using statistical methods (mean, median, mode), or using prediction algorithms.
- Outliers Detection & Treatment:
- Detect outliers using visualization tools (box plots) or statistical measures (Z-score, IQR).
- Depending on the context, outliers can be capped, transformed, or left untouched if they represent genuine observations.

Duplicate Removal:

Identify and remove duplicate records. This often happens during data collection or merging processes.

Standardization:

Ensure that data is consistent, e.g., converting all text to lowercase or ensuring date formats are consistent.

Error Correction:

Identify and correct inaccuracies in data, which can be due to human error during data entry or system errors.

Data Processing:

Feature Engineering:

Create new features based on existing data to provide more insights or to better represent underlying patterns.

E.g., Extracting the month or day from a date column.

Transformation:

- Apply transformations to make data more suitable for analysis. Common transformations include normalization (scaling data between 0 and 1) and standardization (scaling data to have a mean of 0 and a standard deviation of 1).
- Other transformations include logarithmic, square root, and box-cox transformations, especially for addressing skewed data.

Encoding:

Convert categorical data into a format that can be used by algorithms. Common methods include one-hot encoding (create new columns for each category) and ordinal encoding (assigning unique integers to categories).

Aggregation:

Summarize or group data for analysis. E.g., you might aggregate daily sales data to a monthly or yearly level.

Data Splitting:

In machine learning, split data into training, validation, and test sets to evaluate model performance accurately.

Data Integration:

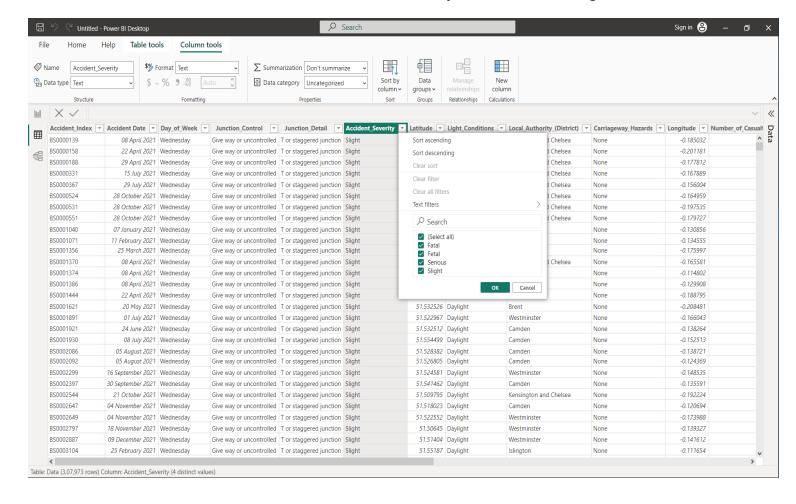
Combine data from different sources, ensuring consistency and accuracy. Techniques include data concatenation and merging.

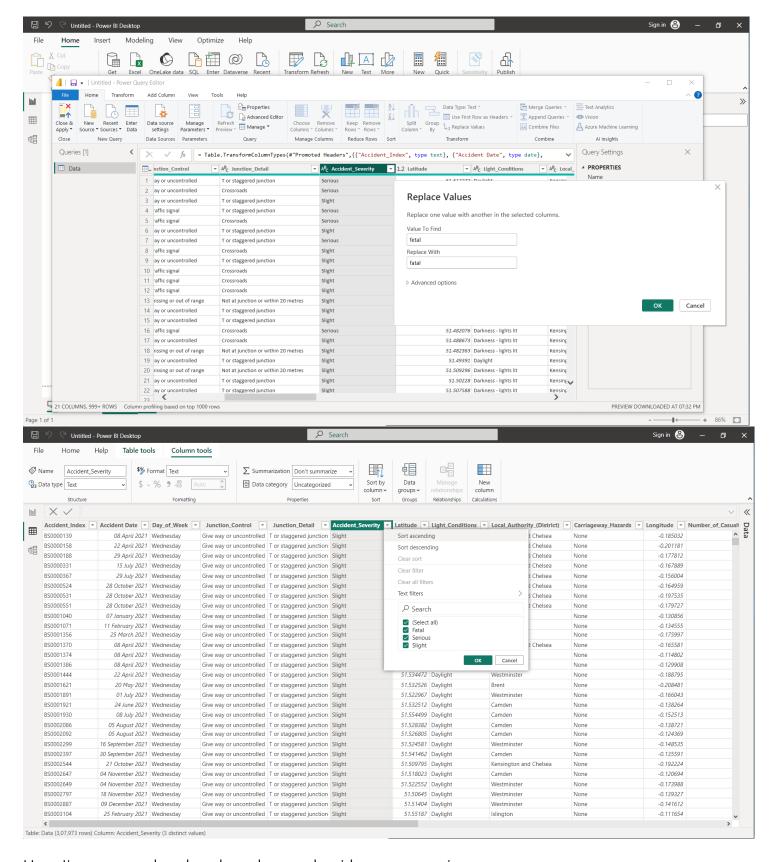
Reshaping:

Pivot or melt datasets to get them into the desired shape for analysis.

Data Indexing & Sorting:

Index data for faster retrieval or sort it for better readability and understanding.





Here I've removed and replace the words with same meaning

Fetel -> fetal

Data modelling:

Data modeling in Power BI refers to the process of organizing, relating, and enhancing your imported data in ways that best serve your analysis and reporting needs. Power BI's data model capabilities are built on top of the Vert iPAQ engine (the same engine that powers Microsoft Analysis Services Tabular). This allows you to create efficient and flexible data models that can handle and relate vast amounts of data.

Defining Hierarchies:

- Hierarchies allow you to drill down into data, such as going from year to month to day.
- In Power BI, you can define hierarchies that let users explore data at different levels.

Handling Date and Time:

Dates can be tricky in data modeling.

It's often beneficial to create a separate date table in your model, ensuring you can leverage time-based functions and hierarchies effectively.

Categorizing Data:

Power BI allows you to define categories for columns, such as marking a column as containing geographical data (like city names or postal codes). This enhances the tool's automatic detection and visualization capabilities.

Row-Level Security (RLS):

If you need to control data access at a granular level, you can implement RLS. With RLS, you can define rules that control data visibility based on user roles.

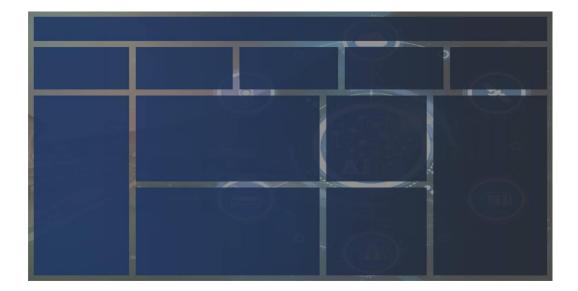
Optimizing Performance:

- Efficient data modeling can also improve report performance. Considerations include:
- Reducing the number of unnecessary columns or tables.
- Minimizing the use of calculated columns where measures would be more efficient.
- Using summary tables for large datasets.

Background design in PowerPoint:

We must create a background to insert certain columns in the power Bi while doing our project.

It will create a frame like structure so that we can easily understand and insert the visualizations.



66

Data visualization / chart designing:

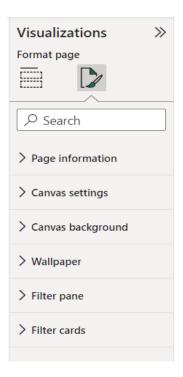
99

Data visualization is a critical component of Power BI, allowing users to convey information clearly and efficiently through graphical means. Power BI offers a broad range of visualization options, catering to various data representation needs.

1. Choosing the Right Visualization:

Understanding your data and the story you want to tell will guide you in selecting the most effective visualization. Some popular visuals in Power BI include:





- Column and Bar Charts: Great for comparing categories or displaying data changes over a period.
- Line Charts: Ideal for showcasing trends over time.
- ❖ Pie and Donut Charts: Useful for representing part-to-whole relationships, but best with limited categories to avoid confusion.
- ❖ Map Visuals: For displaying geographical data.
- Scatter and Bubble Plots: To represent relationships between numeric values or threedimensional data.
- ❖ Tables and Matrices: Useful when you need to display raw numbers or when details are essential.
- Cards: To highlight single values like total sales or average profit.
- ❖ Gauge and KPI: Show progress towards a goal or key performance indicators.
- Waterfall Chart: Breaks down cumulative data into its contributing elements, often used for financial data.
- Ribbon and Tree map: Useful for hierarchical data.

2. Customizing Visualizations:

Power BI offers extensive formatting options:

- Data Colors: Customize the colors of data points for consistency with corporate branding or to enhance readability.
- Title, Background, and Border: Add and format titles, adjust the visual background, or define borders.
- Axes and Labels: Control the appearance and scaling of chart axes, and customize data labels.
- Grids and Bands: Adjust gridlines and create alternating bands of color for better visualization clarity.
- Tooltips: Enhance information displayed when users hover over data points.
- Drillthrough and Drilldown: Allow users to dive deeper into the data directly from a visual.

3. Interactivity:

One of Power BI's strengths is its interactive capabilities:

- Cross-filtering and Highlighting: Selecting a data point in one visual can filter or highlight associated data in other visuals.
- Slicers: Provide a way for viewers to filter data interactively.
- Drillthrough: Allows users to click on a data point and "drill through" to a detailed report based on that point.

4. CUSTOM VISUALS:

If the built-in visuals aren't enough:

- Power BI supports custom visuals from the AppSource or visuals you create.
- ❖ You can use tools like the Power BI Developer toolkit to build your own visuals if you have the technical expertise.

Challenge: The default visualizations might not cater to specific or advanced requirements.

Solution: Explore the Power BI Marketplace for custom visuals. If necessary, develop custom visuals using the Power BI developer tools, but ensure they are optimized and safe.

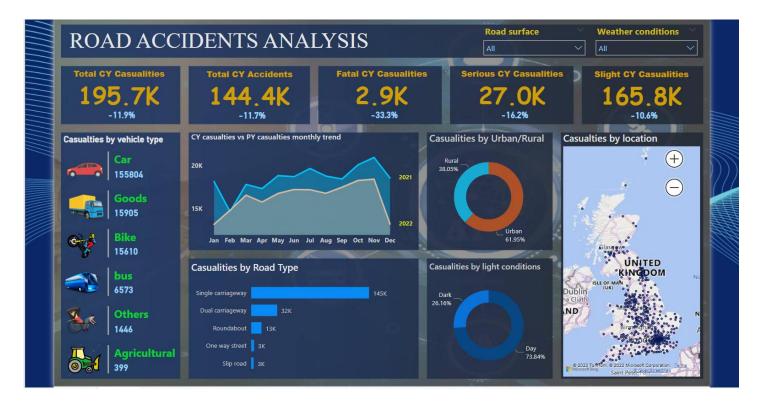
Challenge: Power BI's features and interface are updated frequently, which might disrupt users or workflows.

Solution: Stay informed with the Power BI blog or update documentation. Consider joining the Power BI community to learn from other users and share knowledge.

.....

Challenge: Data sources may evolve, leading to broken connections or outdated reports.

Solution: Implement a robust change management process. Use parameters in P



- **♥** CY current year
- PY previous year

66

conclusion

"

In the above project I've covered the topics:

- ❖ What is Power BI
- Import data in Power BI
- Data Cleaning in Power BI
- Data Modelling in Power BI
- DAX in Power BI
- Measures and Calculations in Power BI
- Charts in Power BI
- Filters and Slicers in Power BI
- Dashboard in Power BI