

Project on Use of Microsoft Power BI tool in India Power

REPORT OF MAJOR PROJECT SUBMITTED FOR PARTIAL FULFILLMENT OF
THE REQUIREMENT FOR THE DEGREE OF MASTER OF COMPUTER
APPLICATION

ABHIRAJ PAL

REGISTRATION NO.-151170510002 OF 2015-2016

UNIVERSITY ROLL NO.-11701015002

UNDER THE SUPERVISION OF

Mr. Arghya Roy

Chief Manager, DT

India Power Corporation Ltd.

And

Mr. Sandip Saha

Chief Manager, IT

India Power Corporation Ltd.

AT

INDIA POWER CORPORATION LTD.

PLOT X1-2&3, BLOCK EP, SECTOR-V, SALT LAKE,

KOLKATA – 700091

MAY, 2018

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Mr. Arghya Roy, chief manager, DT, India Power Corporation Ltd, and Mr. Sandip Saha, chief manager, IT, India Power Corporation Ltd, whose roles as project guide was invaluable for the project. I am extremely thankful for their keen interest in advising me.

I am also highly thankful to Mr. Sanjeev Sinha, President-IT & DT, for his unconditional help and inspiration.

Date:.....

.....

Reg No.: 151170510002 of 2015-2016

Roll No.: 11701015002

MCA 6th Semester, RCCIIT

TABLE OF CONTENTS

TOPICS

1. Introduction
2. Technology used
3. Current landscape of BI
4. Gartner's Report
5. Feasibility Study
6. Reports created with description
7. Aims and objectives
8. Functionalities
9. Future scope of work
10. Structured
Architecture
11. Challenges
12. Conclusion

Introduction

Power BI is a powerful cloud based and On-premises suit of business analytics tools developed by Microsoft to make it easy to combine data from multiple sources, analyze and visualize information and share insights. It also provides self-service business intelligence capabilities, where end users can create reports and dashboards by themselves, without having to depend on information technology staff or database administrators. Power BI allows us to share dashboards and reports with the right people.

It has flexibility as it operates with all leading operating systems like Windows, IOS and Android. Power BI provides cloud-based BI services, known as "Power BI Services", along with a desktop based interface, called "Power BI Desktop". It offers data warehouse capabilities including data preparation, data discovery and interactive dashboards

Technology Used

Power BI Desktop version is used on windows 7 platform to create real time reports and efficient dashboards after connecting the data from excel to power BI.

Current landscape of BI

Feature	Tableau	Power BI
Data Visualization	Tableau provides strong data visualization and is one of the main data visualization tool in the market.	Power BI provides a strong backend data manipulation feature with access to simple visualizations.
Size of Dataset	Tableau can connect much larger datasets as compared to Power BI.	Power BI has a limit of 1GB data in free version.
Data Sources	Tableau covers a vast range of data sources to connect with for data visualization. In Tableau, you select the dataset first and visualizations are used on the fly.	<p>Power BI covers most of the data sources available in Tableau. It is closely integrated with Office 365, hence provides connectivity to SharePoint.</p> <p>Power BI online version also supports direct visualization on Search Engine, though, only Bing is supported at this point.</p>
Costing	Tableau is expensive as compared to Power BI but still under budget for small and medium enterprise.	Power BI provides a free version with 1GB limit on dataset. Power BI Pro is also a cheaper solution when compared with any other BI tool.
Implementation	Tableau provides different implementation types as per organizational needs panning from few hours to few weeks.	Power BI uses cloud storage and includes simple implementation process.

Gartner recognizes Microsoft as a leader in Analytics and BI platforms for 11 consecutive years

Today we're thrilled to announce that for the 11th consecutive year, Microsoft has been positioned as a leader in [Gartner's 2018 Magic Quadrant for Analytics and Business Intelligence Platforms](#).* And for the third year, Microsoft is positioned furthest to the right for **completeness of vision** within the Leaders quadrant.

Figure 1. Magic Quadrant for Analytics and Business Intelligence Platforms



Source: Gartner (February 2018)

This is an exciting milestone, and it is Microsoft's perspective that this underscores our global leadership and its relentless commitment to innovation, agile execution, and visionary roadmap to deliver high-value to customers and partners in a constantly changing market.

Power BI has helped transform hundreds of thousands of businesses around the world by providing self-service analytics at scale. Here are top reasons why customers have chosen Power BI as their analytic platform solution:

- **High value** | Users can start free with Power BI Desktop. Organizations can then choose to license users individually at a low monthly cost or extend business intelligence broadly with Power BI Premium—purchasing only the capacity that’s needed. Since Power BI integrates with their existing tools, it’s instantly familiar and quick to transition.
- **Comprehensive analytics solution** | Comprehensive capabilities allow users to model, analyze, and author to deliver business insights. Organizations enjoy a flexible architecture that provides strong integration with Microsoft and 3rd party products and technologies for streamlined deployment.
- **Centralized management** | Organizations can deploy in minutes and distribute BI content with just a few clicks. They can take advantage of the agility of self-service analytics with IT governance from Power BI.
- **Global scale** | Organizations have the flexibility to deploy their BI solution where they reside with global presence. Power BI provides guaranteed reliability and performance from a trusted company.
- **Security, governance, and compliance** | Organizations can rely on a BI platform that helps comply with stringent industry standards and certifications, plus keep their data secure and control how it is accessed and used.

We’re honored and proud to see the positive impact that Power BI continuously delivers to customers and to hear their feedback:

- *“We use Power BI in Office 365 to turn data into insights, so we can **ensure our place in the future of the luxury car industry.**” — Andrew Palmer, Chief Executive Officer, Aston Martin*
- *“We’re using Power BI to tell a better, more effective story to our partners. And we know it’s working because **we’re seeing more subscriptions, followers, and retweets, and higher partner engagement numbers.**” — Jessica Raymond, Director, Marketing, BitTitan*
- *“With Office 365 E5, we’re delivering an intelligent workplace, where employees gain self-service access to interactive, graphical reports and dashboards through Microsoft Power BI. Our employees use this service to create their own reports on demand. And for Henkel, **there’s huge benefit in supporting efficient, widespread access to business information, such as increased agility in a market that’s moving faster than ever.** The possibility to quickly assess the impact of an advertising campaign might serve as a good example: **It’s easy to use Power BI to analyze real-time sales**”*

data and fine-tune our marketing activities to drive sales.” - Markus Petrak,
Corporate Director, Henkel

You can see the new AI-driven analytics and data storytelling capabilities driving these results in action watching the latest episode of Microsoft Mechanics.

Finally, and most importantly, we remain committed and inspired by the great feedback and engagement we see from our community around Power BI. Since July 2015, our engineers have worked hard to listen and incorporate feedback from more than 200,000 active community participants who have submitted over 6,000 ideas. This level of input has directly influenced the evolution of our product. Also, with more than **250 showcased solutions**, our partner network brings a wealth of experience across a wide number of industries, technologies and geographies.

**This graphic was published by Gartner, Inc. as part of a larger research document and should be evaluated in the context of the entire document. The Gartner document is available upon request from Microsoft. Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner’s research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.*

Feasibility Study

Power BI is developed by Microsoft in September 2013 and it was first released to the general public on July 24, 2015.

Power BI (Business intelligence) is a cloud based data analysis, from wide range of data sources, which can be used for data analysis and reporting. Power BI is user friendly and simple that power users and business analysts can work with it. Power BI is a powerful tool that can be used by BI developers in enterprise systems for modeling scenarios and complex data mash-up.

Benefits

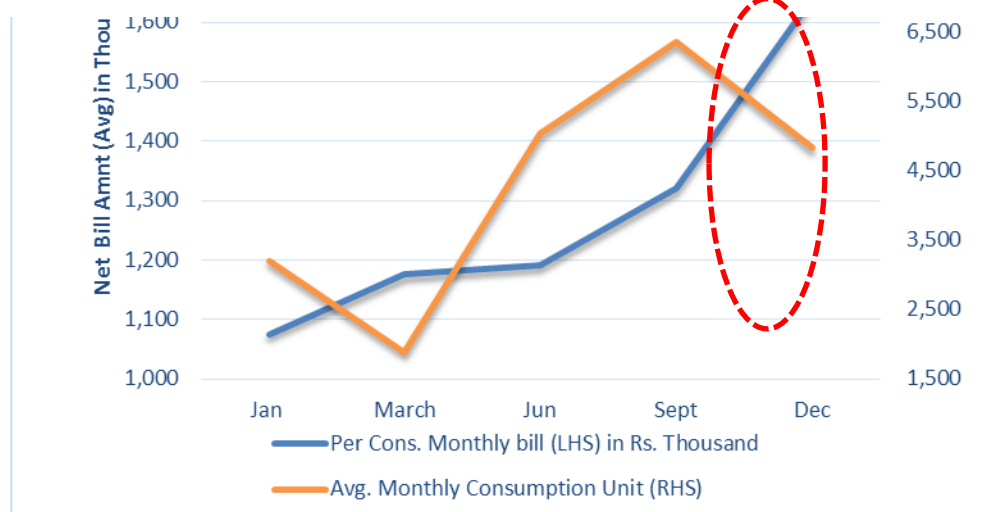
- Real time dashboard updates.
- Live connection to data sources, secure, on premises and in the cloud.
- Uses natural language intuitive data exploration.
- Integrated with familiar Microsoft products.
- **Reports created with description**

Important KPIs on which reports are made

1. % of Net Bill AMT paid within due date – month wise
2. % of Count bill paid within due date –Route wise
3. Average monthly Unit consumed and Average Net Bill AMT – month wise and
 - a. Unit consumed > 500 (arbitrary, can be changed)
 - b. Unit Consumed < 500
4. Unit consumed – Route Wise. Also plot SD. Do it for last four months
5. Top 100 customers – % of Net Bill AMT and Unit Consumed wise for last 12 months.
 - a. Can we identify count of new customer in this 100 list in any month?
6. Meter Replacement V/s Unit consumed – Month wise
 - a. Perform a suitable Statistical Hypothesis Test to conclude whether changing meter has actually caused in lower units consumption?
7. % Meter Status Data – Month wise and Route Wise- We just want see the status = locked and / or reading not taken
8. % Meter Status Data V/s SD of unit consumed – Month wise

Graph Of the above kpis are given below with description

Per Consumer Bill v/s Average Monthly Units consumed

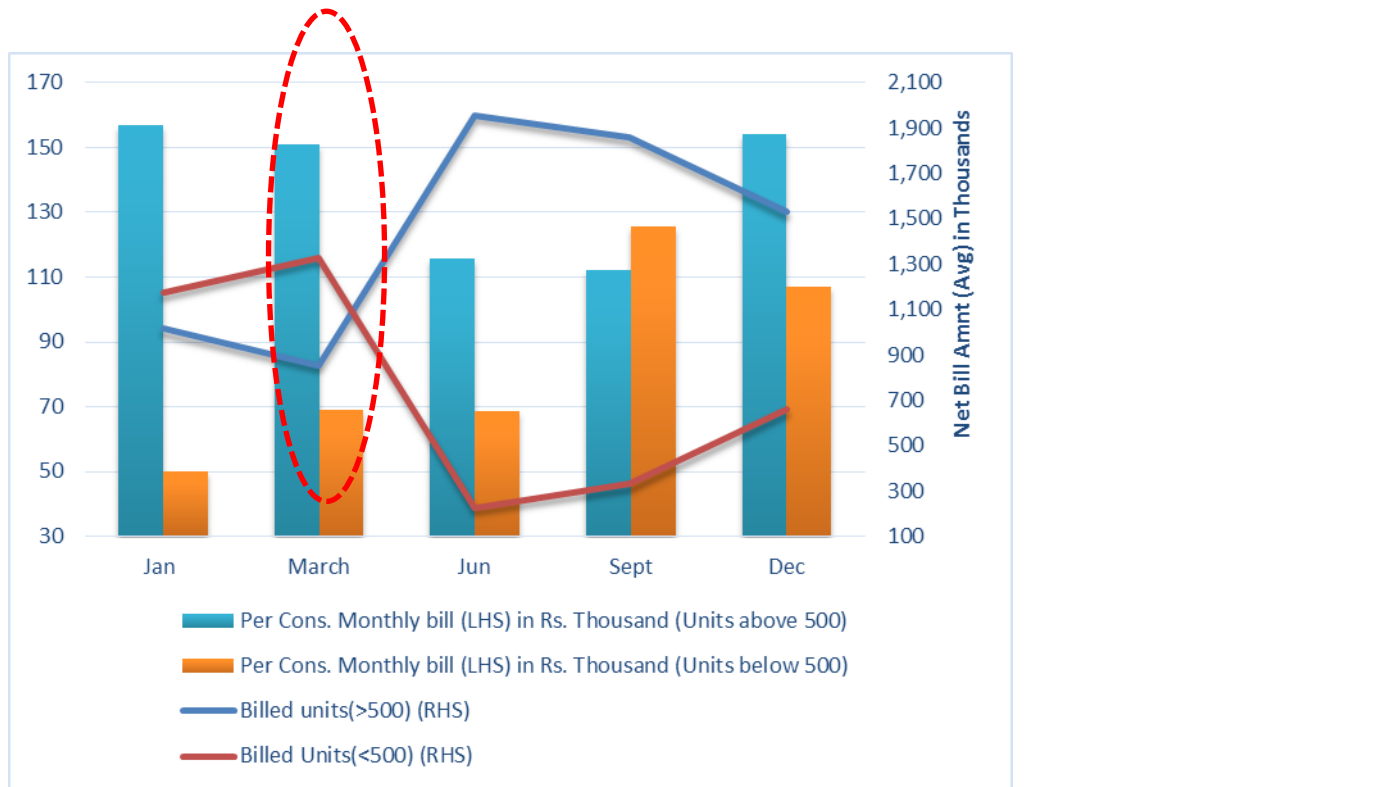


Ideally,
Consumer
bill amount

and electricity

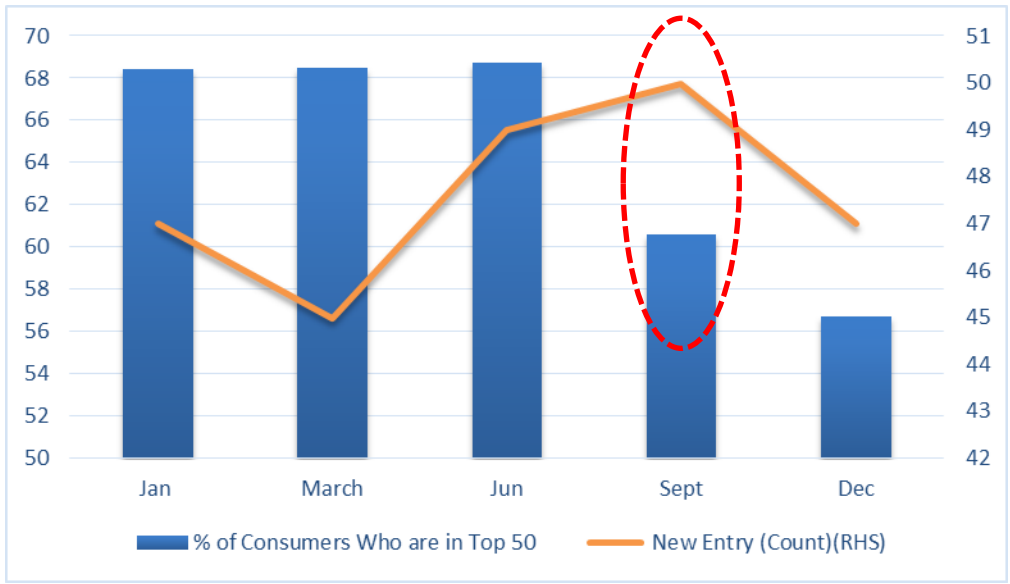
consumption (measured in unit) should move in tandem. However, in Dec 2017 we have seen the reversal in trend.

Per consumer monthly bill amount and unit consumption (in units) with filter of 500 units



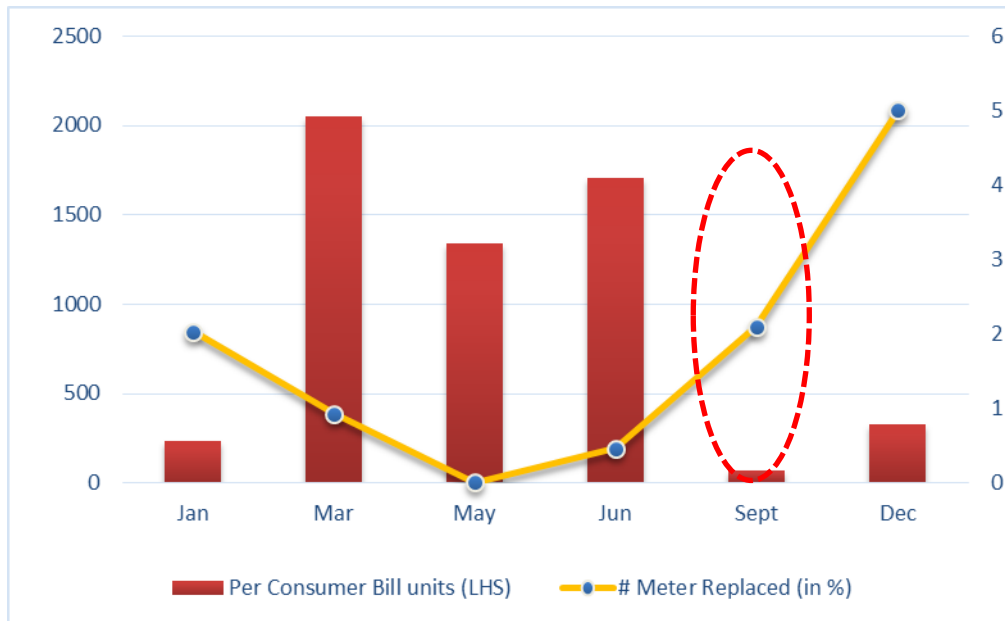
From this graph, it appears that from June number of high value customers (billed unit more than 500 units) have increased. However, that increase eventually doesn't translate into monetary benefit immediately. That got reflected from September billing. What needs to be pondered is the stark difference in number of customers whose consumption is less than 500 units and above 500 units. In fact, from the trend it seems that again the numbers are leading to a *misnomer*.

Analysis based on Top 50 consumers (as sorted by Bill Amount)



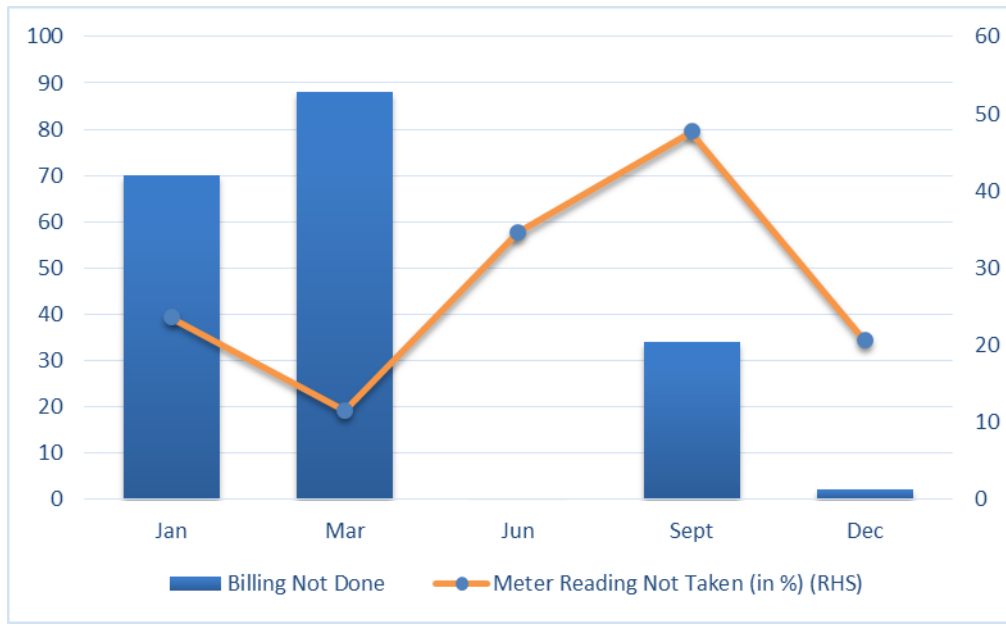
From Sept, we see a decline in number of customers who are in Top 50 in terms of Paying Bill (Amount). Also we see a decline in new customers who are high value in nature.

Monthly meter replacement rate and Per Consumer Monthly bill units



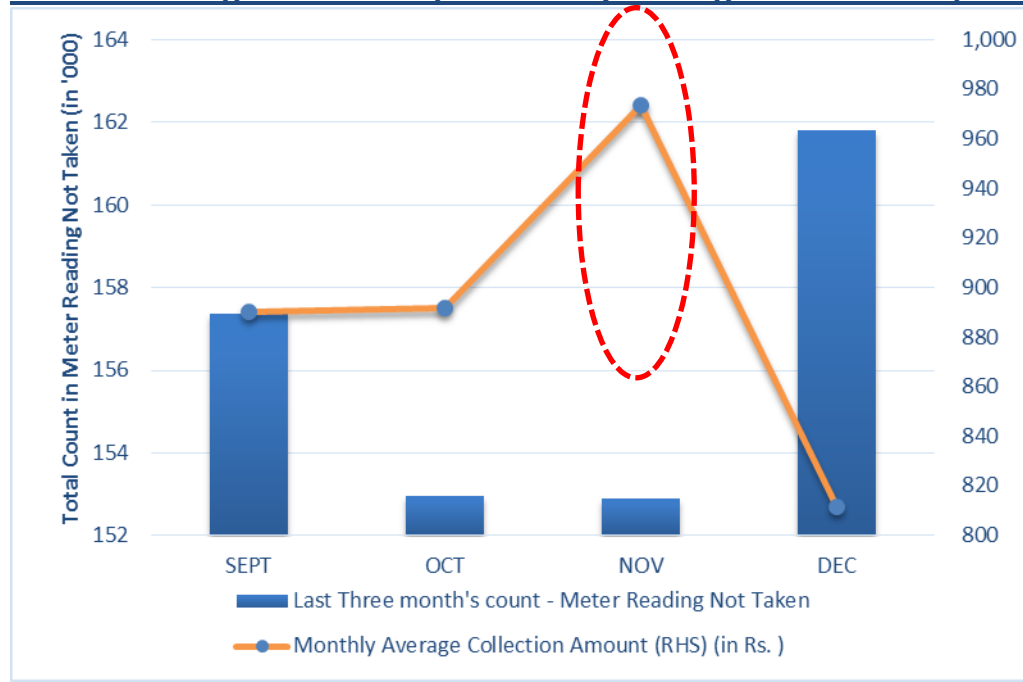
From June, we can see that there is a steady trend in number of meter being replaced. This has actually resulted in lesser amount of monthly consumption. Does it indicate that Meters were being "tampered" before June?

Monthly Incomplete Billing Status and Meter Reading Not Taken (Monthly)



From September, we see a healthy trend in meter reading activities. We see a more number of customers are being billed (though in Dec an abrupt drop is unexplainable) and more number of meters are read.

Meter Reading Not Taken v/s Monthly Average Bill Amount (in Rs.)



This graph plots two parameters viz. in consecutive previous three month's meter reading status (i.e. reading not taken) and Monthly Average collection amount. We see a drastic drop in collection and corresponding jumps in "Meter reading not taken" count. Please note that December bar graph represents the cumulative numbers of Sept, Oct and Nov counts.

Aims & Objectives

The main objective of the Project on “The uses of Microsoft Power BI tool in India Power” is to collect data about the organization from various data sources and connect them to Power BI to create some rich reports and after that create dash-boards. . The purpose of the project is to create and share real time insights for an organization so that business users will find it easy to understand the status of the organization on the basis of time. Real time

Dash boards help management to take accurate decision as and when required. Management also plan for future initiatives on the present trends.

Functionalities

Microsoft Power BI connects to multiple data sources such as excel, .csv, oracle, SAP business warehouse, MySQL, SQL server and many others and creates interactive reports using those data with rich graphical visualizations, publish and share insights on website and blogs.

Power BI maintains a specific cycle:

At first, the report is created in Power BI Desktop.

Then, the created report is published in Power BI service.

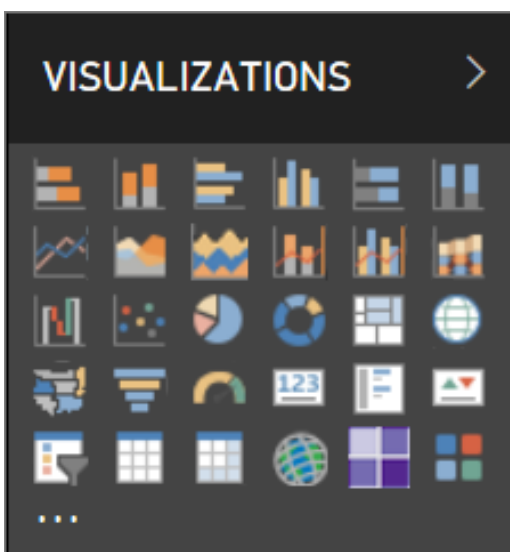
Finally, the users of Power BI can see the reports on the Power BI mobile app.

Some of the main functions of Power BI are:

- **Customizable Dashboards:** Microsoft Power BI tool enables to create dashboards, A Power BI *dashboard* is a single page, often called a canvas, that uses visualizations to tell a story. Because it is limited to one page, a well-designed dashboard contains only the most-important elements of that story.



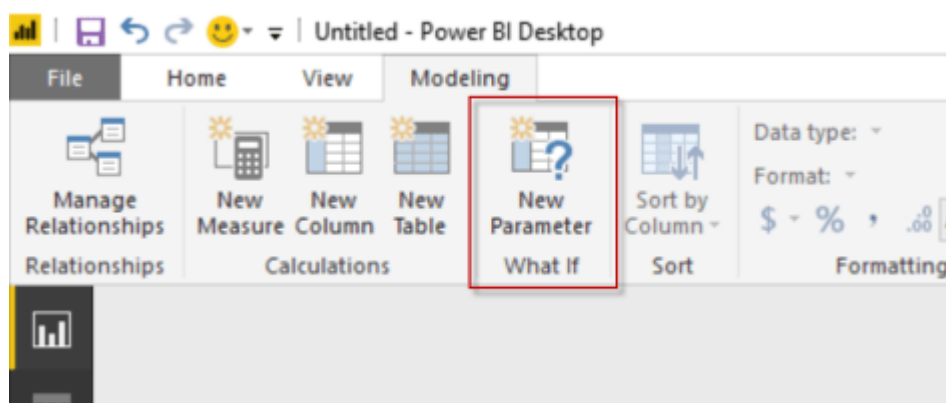
- **Rich graphical Visualizations:** When creating or editing a Power BI report, there are many different types of visuals available for us to use. These visuals display in the **Visualizations** pane. When we download Power BI Desktop or open Power BI service (app.powerbi.com), this set of visuals comes "pre-packaged."



But we are not limited to this set of visuals, selecting the ellipses opens up another source of report visuals: *custom visuals*.

Custom visuals are created by developers, using the custom visuals SDK, to enable business users to see their data in a way that fits the business best.

- **Ad-hoc Reporting:** Power BI allows us to create ad hoc reporting. Ad hoc analysis is a **business intelligence** process designed to answer a single, specific business question. The product of ad hoc analysis is typically a statistical model, analytic report, or other type of data summary. Ad hoc reporting should be business user controlled and not technical user controlled.
- **What-if Analysis :**The What If Parameter feature in Power BI allows us to dynamically perform what-if type analysis by using a slider bar to visualize changes to a calculation. This becomes a very flexible way to perform all kinds of different “what if” type scenarios.



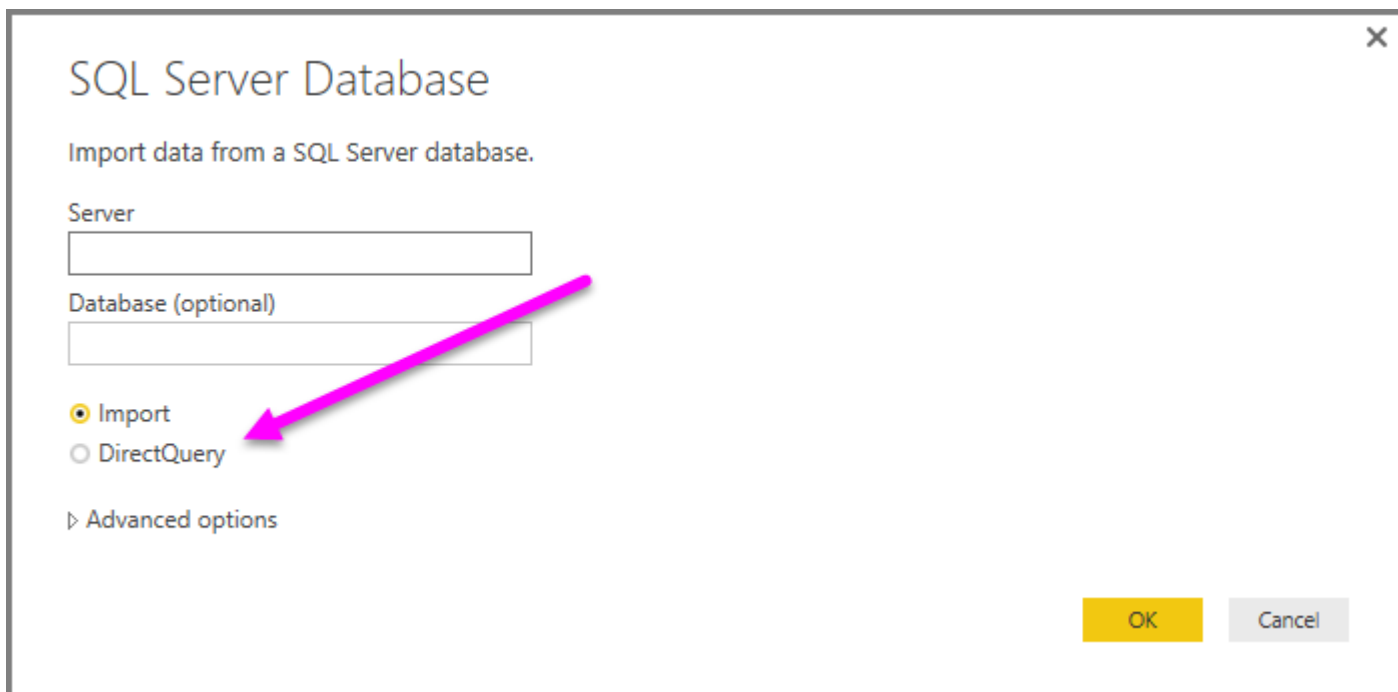
- **Statistical Analysis:** R is a software environment for data mining, statistical analysis, and data presentation. It is widely used in statistical applications for sophisticated modeling and data mining solutions. R scripts can be executed in Power BI with the NAV database as the data source for computations.

Use Direct Query in Power BI Desktop

With **Power BI Desktop**, when we connect to our data source, it is always possible to import a copy of the data into the **Power BI Desktop**. For some data sources, an alternative approach is available: connect directly to the data source using **Direct Query**.

How to Connect using Direct Query

When we use **Get Data** to connect to a data source supported by **Direct Query**, the connection window lets us select how we want to connect.



The differences between selecting **Import** and **Direct Query** are the following:

Import – the selected tables and columns are imported into **Power BI Desktop**. As we create or interact with a visualization, **Power BI Desktop** uses the imported data. We must refresh the data, which imports the full data set again, to see any changes that occurred to the underlying data since the initial import or the most recent refresh.

Direct Query – no data is imported or copied into **Power BI Desktop**. For relational sources, the selected tables and columns appear in the **Fields** list. For multi-dimensional sources like SAP Business Warehouse, the dimensions and measures of the selected cube appear in the **Fields** list. As we create or interact with a

visualization, **Power BI Desktop** queries the underlying data source, which means we are always viewing current data.

Many data modeling and data transformations are available when using **Direct Query**, though with some limitations. When creating or interacting with a visualization, the underlying source must be queried and the time necessary to refresh the visualization is dependent on the performance of the underlying data source. When the data necessary to service the request has recently been requested, Power BI Desktop uses recent data to reduce the time required to display the visualization. Selecting **Refresh** from the **Home** ribbon will ensure all visualizations are refreshed with current data.

The [Power BI and Direct Query](#) article describes **Direct Query** in detail. Also, see the following sections for more information about benefits, limitations, and important considerations when using **Direct Query**.

Benefits of using Direct Query

There are a few benefits to using **Direct Query**:

- **Direct Query** lets you build visualizations over very large datasets, where it otherwise would be unfeasible to first import all of the data with pre-aggregation
- Underlying data changes can require a refresh of data, and for some reports, the need to display current data can require large data transfers, making re-importing data unfeasible. By contrast, **Direct Query** reports always use current data
- The 1 GB dataset limitation does *not* apply to **Direct Query**

Limitations of Direct Query

There are currently a few limitations to using **Direct Query**:

- All tables must come from a single database
- If the **Query Editor** query is overly complex, an error will occur. To remedy the error you must either delete the problematic step in **Query Editor**, or *Import* the data instead of using **Direct Query**. For multi-dimensional sources like SAP Business Warehouse, there is no **Query Editor**
- Relationship filtering is limited to a single direction, rather than both directions (though it is possible to enable cross filtering in both directions for **Direct Query** as a Preview feature). For multi-dimensional sources like SAP Business Warehouse, there are no relationships defined in the model
- Time intelligence capabilities are not available in **Direct Query**. For example, special treatment of date columns (year, quarter, month, day, so on) are not supported in **Direct Query** mode.
- By default, limitations are placed on DAX expressions allowed in measures; see the following paragraph (after this bulleted list) for more information
- There is a 1 million row limit for returning data when using **Direct Query**. This does not affect aggregations or calculations used to create the dataset returned using **Direct Query**, only the rows returned. For example, you can aggregate 10 million rows with your query that runs on the data source, and accurately return the results of that aggregation to Power BI using **Direct Query** as long as the data returned to Power BI is less than 1 million rows. If more than 1 million rows would be returned from **Direct Query**, Power BI returns an error.

1

To ensure that queries sent to the underlying data source have acceptable performance, limitations are imposed on measures by default. Advanced users can choose to bypass this limitation by selecting **File > Options** and then **Settings > Options and settings > Direct Query**, then selecting the option *Allow unrestricted measures in Direct Query mode*. When that option is selected, any DAX expression that is valid for a measure can be used. Users must be aware, however, that some expressions that perform very well when the data is imported may result in very slow queries to the backend source when in Direct Query mode.

Important considerations when using Direct Query

The following three points should be taken into consideration when using **Direct Query**:

- **Performance and load** - All **Direct Query** requests are sent to the source database, so the time required to refresh a visual is dependent on how long that back-end source takes to respond with the results from the query (or queries). The recommended response time (with requested data being returned) for using **Direct Query** for visuals is five seconds or less, with a maximum recommended results response time of 30 seconds. Any longer, and the experience of a user consuming the report becomes unacceptably poor. In addition, once a report is published to the Power BI service, any query that takes longer than a few minutes will timeout, and the user will receive an error.

Load on the source database should also be considered, based on the number of Power BI users who will consume the published report. Using *Row Level Security* (RLS) can have a significant impact as well; a non-RLS dashboard tile shared by multiple users results in a single query to the database, but using RLS on a dashboard tile usually means the refresh of a tile requires one query *per user*, thus significantly increasing load on the source database and potentially impacting performance.

Power BI creates queries that are as efficient as possible. Under certain situations however, the generated query may not be efficient enough to avoid refresh that would fail. One example of this situation is when a generated query would retrieve an excessively large number of rows (more than 1 million) from the back-end data source, in which case the following error occurs:

```
The result set of a query to external data source has exceeded the maximum allowed size of '1000000' rows.
```

This situation can occur with a simple chart that includes a very high cardinality column, with the aggregation option set to *Don't Summarize*. The visual needs to only have columns with a cardinality below 1 million, or must have appropriate filters applied.

- **Security** - All users who consume a published report connect to the back-end data source using the credentials entered after publication to the Power BI service. This is the same situation as data that is imported: all users see the same data, irrespective of any security rules defined in the backend source.
- **Supported features** - Not all features in **Power BI Desktop** are supported in **Direct Query** mode, or have some limitations. In addition, there are some

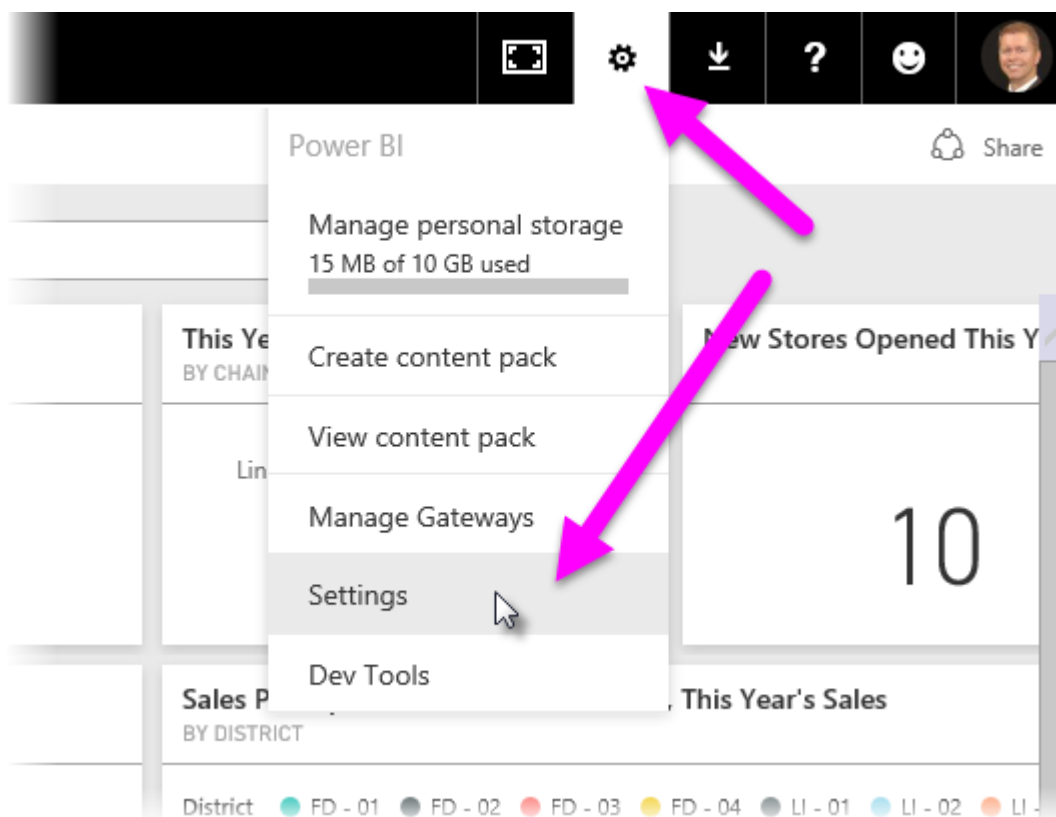
capabilities in the Power BI service (such as *Quick Insights*) that are not available for datasets using **Direct Query**. As such, the limitation of such features when using **Direct Query** should be taken into consideration when determining whether to use **Direct Query**.

Publish to the Power BI service

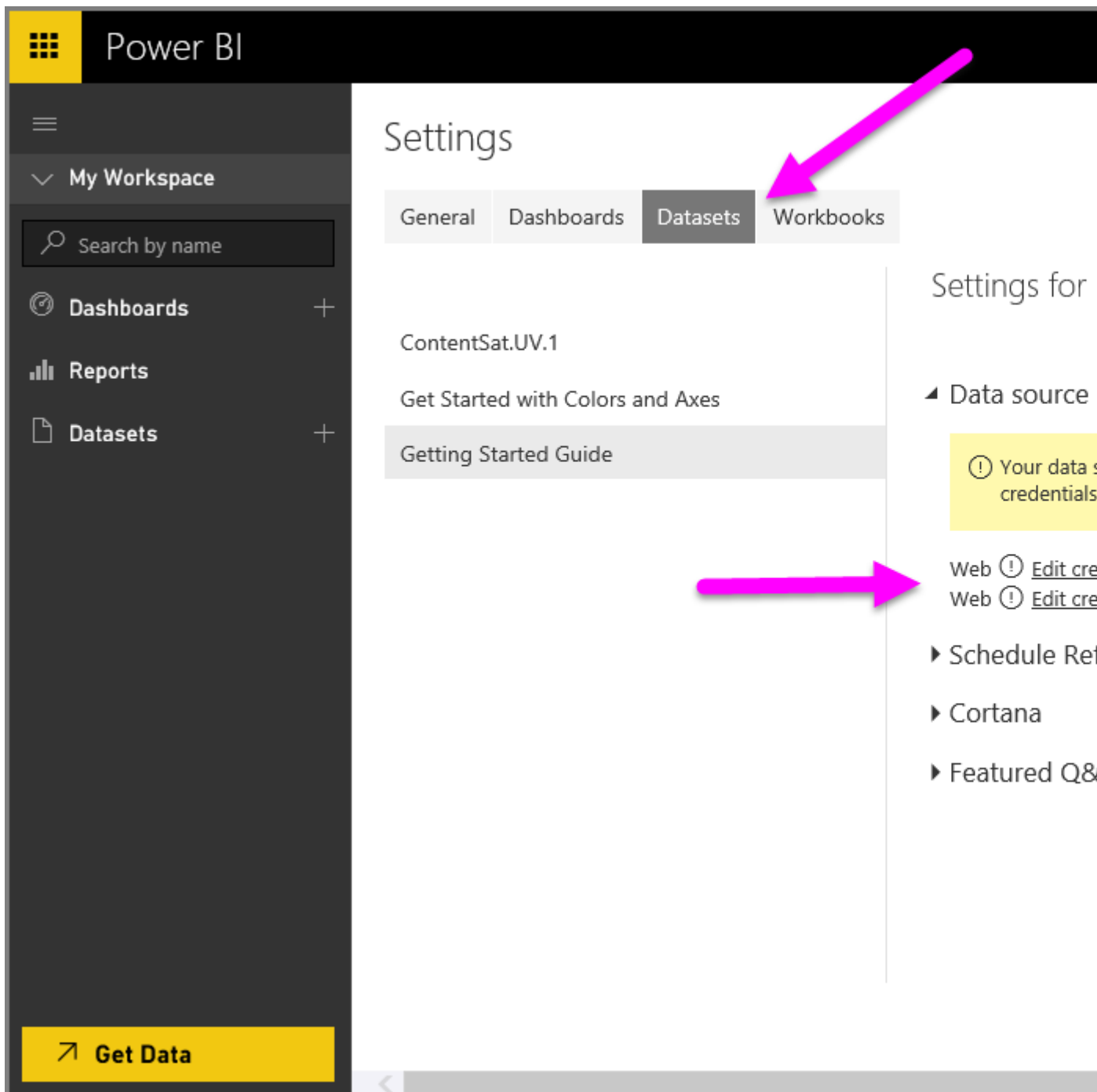
Reports created using **Direct Query** can be published to the Power BI Service.

If the data source used does not need the **on-premises data gateway** (**Azure SQL Database**, **Azure SQL Data Warehouse**, or **Red shift**), credentials must be provided before the published report will be displayed in the Power BI Service.

We can provide credentials by selecting the **Settings** gear icon in Power BI, then select **Settings**.



Power BI displays the **Settings** window. From there, we need to select the **Datasets** tab and choose the dataset that uses **Direct Query**, and select **Edit credentials**.



+

Until credentials are supplied, opening a published report or exploring a dataset created with a **Direct Query** connection to such data sources results in an error.

For data sources other than **Azure SQL Database**, **Azure SQL Data Warehouse** and **Red shift** that use Direct Query, an **on-premises data**

gateway must be installed and the data source must be registered to establish a data connection.

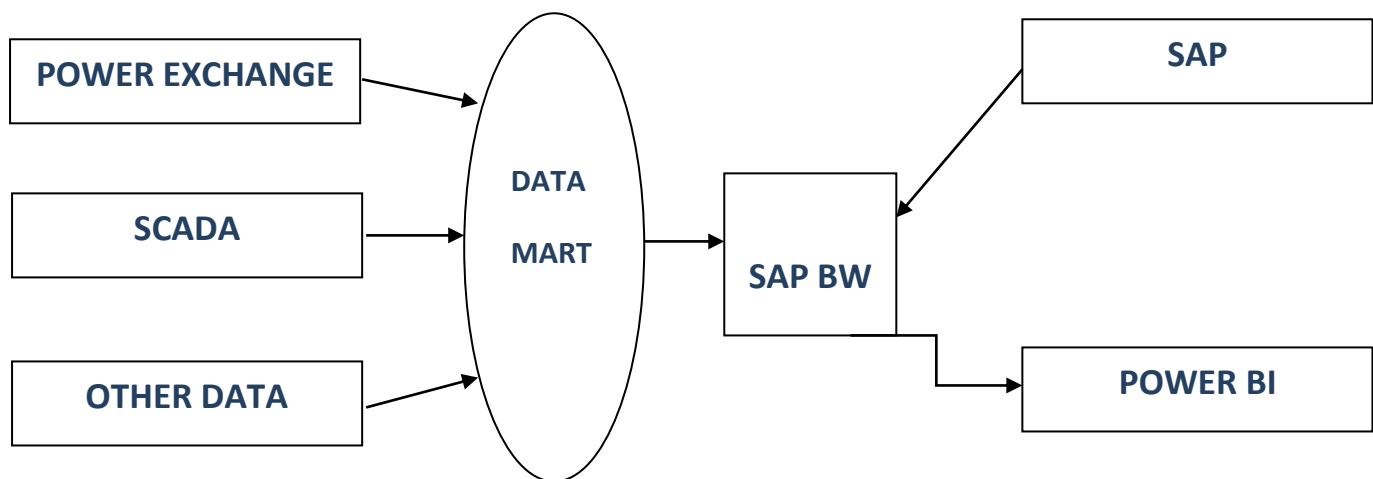
Future Scope

Business Intelligence & Data warehousing is a very good & demanding field in Information Technology industry. Its present as well as future is very bright.

Power BI is a really exciting product. It's fast paced, the Community is growing and really interactive, there are a bunch of ways to learn and interact.

The major benefit I see is that it's an avenue to take the data that we have already been working with and helping business to make that data useful. Power BI is a tool that makes that really straightforward.

Structured Architecture



Challenges

The challenge of this project is to link different data sources with the Power BI tool in a very cost effective module. Connecting the SAP database with Power BI needs huge license cost where the project is not viable.

Conclusion

The choice of BI tool is a function of organization's need, user's maturity and overall budget. Since, India Power Corporation Ltd has just started its journey towards a BI driven data culture and the end users have not yet been evolved fully, it was anonymously decided to go for Power BI as a data visualization tool with SQL server as a backend data mart. The idea is to explore deeper co-relation or discrepancies amongst various data points and visualize interactively through an effective data visualization application. Power BI primarily fulfills this objective. Although, tableau scores much higher on every count (except cost) over Power BI.

Signature of the student

Signature of the project guide(s)