Research Data Analysis with Power BI

Vijay Krishnan   *S Bharanidharan*   *G Krishnamoorthy*

Abstract

Power BI has taken the world of business intelligence, data visualization and analytics by storm. Power BI is an online service that enables searching data, transforming it, visualizing it, and sharing the developed reports and dashboards with other users in the same or different department/organizations or even with the general public. As of February 2017, more than 200,000 organizations across 205 countries are using Power BI. Power BI is having a free option that has adequate features and functionality, it has become a serious contender for use as a business intelligence platform in small and medium organizations. One of the innovative features of Power BI is its Quick Insights feature (Michael Hart, 2017) that is built on a growing set of advanced analytical algorithms. After uploading a dataset to Power BI, a click of a button can be used to invoke this feature that automatically builds many reports based on its analysis of the data, without any human intervention being required. This also helps reduce human errors, in calculations and statistical techniques, which lead to un-verifiable research. Accepting even Excel spreadsheets as input, Power BI is easy to use and ripe for adoption as a platform for Research Data Analysis. In this paper, an attempt has been made to show how easily a dataset of research data can be transformed by Power BI into a set of analytical reports and dashboards, and which can be shared with ease.

Keywords: Research Data Analysis, Power BI, Anna University

Flow of Power BI

Power BI, from Microsoft, is a suite of business analytics tools that is used to analyze data and share insights in the form of reports and dashboards. User data in various forms – spreadsheets, text files, databases, etc. form the input for Power BI. Datasets are formed by transforming the data provided by the users. Data transformations are decided by the users. This step is used to remove errors and redundant data, correct formatting, and prepare data for further analysis by organizing them into suitable normalized forms, and so on. Based on the report and dashboard being developed, filtering the data to only include the relevant bits enables one to focus on only the data that matters.

Once a dataset is ready, reports can be created from them by adding from a choice of multiple visualization elements. Visualization elements in Power BI ranges from showing a single number to a gradient coloured map. These visuals help present data in a way that provides context and insights. Filters can be applied on the reports so that relevant data is surfaced for users interested in analyzing the data. Such reports can be built manually or by

using the Quick Insights feature, which uses various algorithms to analyze the data and returns a set of reports that it builds automatically.

Once reports are published, elements of the report or the whole report can be included in dashboards. Power BI dashboards show a 360 degree view of the data by enabling users to keep their most important metrics in one place. They also allow users to interact with the reports for filtering or querying the data, even allowing natural language queries. It is however limited to a single page, thereby surfacing only the relevant portions of the data to make it easy for users to draw insights. It is further possible to constantly update the report and dashboard data, in real time, and make it available on all devices like PCs and smartphones.

Getting Started with Power BI

Power BI Desktop for windows can be downloaded freely from the website https:// powerbi.microsoft.com/en-us/get-started/ . The Power BI Desktop for Windows can be used to create datasets and reports on their computers. This is required for those users that want to work offline, save their report files and later publish it to the online service. It is also meant for data modeling purposes. Users can also Sign Up for the Power BI service at the same url. Online service is available for Quick Insight feature and also where reports and dashboards can be shared.

Frequently asked questions on software requirements, academic pricing, etc. is answered at https://powerbi.microsoft.com/en-us/ documentation/powerbi-frequently-asked-questions/ (Maggie Sparkman, 2017). The best place to get started with learning Power BI is https:// powerbi.microsoft.com/en-us/guided-learning/ and it introduces users to all the relevant topics of this powerful analytical and visualization platform. Once a user is signed up for the Power BI online service, datasets, reports and dashboards can be developed with ease.

Presently one can sign up to use Power BI, only with an organization email address. It does not allow use of email addresses from public or commercial mail service providers like Gmail, Hotmail, Yahoo, etc. Emails from registered domains of colleges, universities and research institutions can be used to sign up for the service. For example, ProfessorEmail@university.edu or ResearcherEmail@researchinstitute.ac.in and such email addresses can be used. If an organization is using Office 365 or Microsoft Azure, their users can login to that and if Power BI service is listed, can access it from there. For users that do not have an organization email, and who are okay with not sharing reports or dashboards with other users in their organization, https://powerbi.microsoft.com/ en-us/documentation/powerbi-admin-signing-up-for-power-bi-with-a-new-office-365-trial/ is an option (Adam Saxton, 2017).

Methodology

Power BI supports many forms and formats of data that it can analyze. We decided to use an Excel spreadsheet for this purpose. It had data regarding Anna University research productivity – viz. information about conference papers and research articles authored by individuals from Anna University. The data had earlier been obtained from

Scopus in a year wise manner, and had been aggregated for the desired time period along with addition of various calculated fields. Data covering various topics of interest were organized into different sheets. The core data had these fields –

❖ Author Count

❖ Page Count

❖ Subject

❖ Authors

❖ Title

❖ Year

❖ Source title

❖ Cited by

❖ Affiliations

❖ Authors with affiliations

❖ Abstract

❖ Author Keywords

❖ Index Keywords

❖ Publisher

❖ Conference name

❖ Conference date

❖ Conference location

❖ Language of Original Document

❖ Document Type

From the core data in one spreadsheet, desired data was extracted into separate spreadsheets.

❖ Country Wise spreadsheet had two columns – Country and Records.

❖ Subject Wise spreadsheet had two columns – Subject and Records.

❖ Year Wise spreadsheet had two columns – Year and Records.

❖ Quantum of Literature spreadsheet had two columns – Institution and Records.

❖ Single and Multiple author spreadsheet had two columns – Author Count and Records

Figure 1 shows one of the spreadsheets with information regarding research articles.

Author Count	Page Count	Subject	Authors	Title
3	2	Arts and Humanities	Pillai S.O., Natarajan S., Palanisamy P.K.	Free volume of sodium chloride solutions in water-dimethy formamide mixtures from ultra
3	4	Computer Science	Murugeshan M.S., Lakshmi K., Mukherjee S.	Exploiting negative categories and wikipedia structures for document classification
2	31	Engineering	Sivakumar G.D., Shahabudeen P.	Algorithms for the design of a multi-stage adaptive kanban system
3	6	Engineering	Natarajan S., Ramadass N., Rao Y.V.R.	State-based dynamic multi-alphabet arithmetic coding for image compression
3	6	Engineering	Arthanareeswaran G., Thanikaivelan P., Raajenthiren M.	Preparation and characterization of poly (methyl methacrylate) and sulfonated poly (ether
3	4	Computer Science	Rajavel R., Somasundaram T.S., Govindarajan K.	Dynamic Load Balancer Algorithm for the Computational Grid Environment
3	4	Computer Science	Veni S., Kadhar Nawaz G.M., Praba P.	Performance analysis of network traffic behavior in conventional network over MPLS
5	0	Engineering	F. shabir M., Authars S., Ganesan S., Karthik R., Madhan S.K.	
3	5	Physics and Astronomy	Saravanan L., Pandurangan A., Jayavel R.	Synthesis and optical properties of Eu+ doped CdS nanostructures
7	7	Pharmacology Toxicolo	Rajesh V.A., Kandasamy R., Park S.J., Boyapalle S., Mohapatra S., Heller	Isatin down-regulates expression of atrial natriuretic peptide receptor A and inhibits airwa
2	14	Materials Science	Bhavani P., Sangeetha D.	SPSEBS/H3PO4 composite electrolyte membranes for application in PEMFC and DMFC
3	5	Pharmacology Toxicolo	Rajesh V.A., Kamadasan T.A., Anand Vijayakumar P.R.A.	Influence of literacy, occupation and hereditary on cognitive functioning in Diabetic individ
2	7	Computer Science	Johnson S., Shanmugam V.	Feature subset selection for hot method prediction using genetic algorithm wrapped with s
2	5	Engineering	Jasper J., Victoire T.A.A.	Variable neighborhood search guided differential evolution for dynamic load dispatch with
4	6	Engineering	Newlin N.R., Jenifer. A., Rohini. G., Jamuna. V.	Digital simulation of single and interleaved soft switching boost converter for PV
2	6	Social Sciences	Vikkraman P., Ebinezar D.C.	A study on effective people management in IT industry with special reference to software c
2	8	Engineering	Sudhahar R., Jamuna V.	A review of pioneer fiction realization of embedded core FPGA based control of single phas
5	0	Mathematics	Ganesh R., Sakthivel R., Ren Y., Anthoni S.M., Mahmudov N	Controllability of neutral fractional functional equations with impulses and infinite delay

Figure 1: Research Data spreadsheet with info about Articles – Author Count, Page Count, Subject, Authors, Title are the columns shown.

Building a Dataset with our Input Data

It is necessary to login to Power BI or use the Power BI Desktop to build a dataset using an Excel spreadsheet. Online service option was chosen and logged in to the account, where the Navigation Bar had an option for Get Data (Michele Hart, 2017).

On invoking Get Data, it gives various options to get data ranging from content packs to online services. Below steps were followed to complete this phase -

❖ The clear option, in our case, to "Import or Connect to Data" had Files as one option, and Database as another.

❖ On selecting File, it further prompted us with options for Local File, One Drive – Business, One Drive – Personal, and SharePoint – Team Sites. In our case, since we had an Excel workbook, it was a Local File that we selected.

❖ At the last step, it gave options whether to Import or Upload. Upload is simply used to view the data, and not to create reports. So, we selected the option to Import the data. (David Iseminger, 2017)

Creating Reports

After completing the steps for Get Data, on the navigation pane, we saw the name of our workbook, FINAL RAWDATA – Copy being listed under the Datasets option. On the workspace, in the middle of the screen, was a blank/white canvas and on the right, it had Visualization, Filter and Field options which could be used to create a report.

On clicking the ellipsis "…" button next to the dataset name, it gave other options including Quick Insights. We used the Quick Insights option and it

generated some reports for us automatically. Figure 2 shows one of the Quick Insights report that was generated automatically.
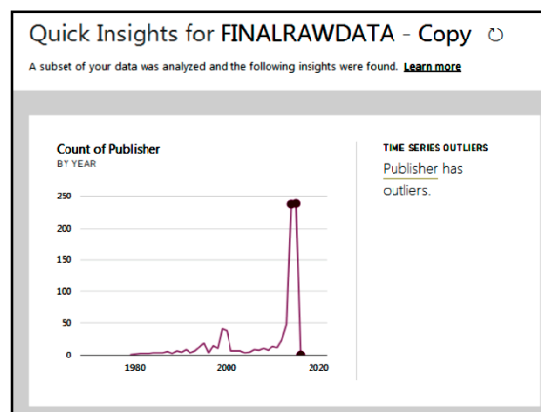

Figure 2: One of the Quick Insights report that was generated automatically for the dataset

With reports being automatically generated using Quick Insights, we used the View Insights option next to the dataset name to see the various reports. It had created a lot of reports for the dataset we created and these were the ones at the top –

❖ Count of Publisher by Year

❖ Page Count by Language by Original Document

❖ Count of Publisher by Language by Original Document

❖ Count of Publisher and Page Count

❖ Page Count by Subject

These reports created by Quick Insights usually compared one or two fields where there was a clear trend in the data. We needed a few specific reports according to our requirement, and we could create them from the scratch by clicking on the dataset name, and dragging the required fields to the canvas (Michele Hart, 2017). Choice of visualization was

changed if the dragged fields were shown as a table. Once the report met our requirements, we saved it and gave it a name. It appeared in the list of Reports on the navigation pane.

Creating a Dashboard

On each report and View Insights listing, the contents (visualizations) were shown in rectangles, called Tiles. These tiles had a Pin icon. Selecting that gave the option to pin that content to an existing or new dashboard. For the first visualization, we selected the new option and gave a name - AnnaUniversityResearch. For the other visualizations, we selected the existing dashboard option and added it to AnnaUniversityResearch.

Figure 3 shows a partial view of the dashboard that we were able to create. All this was achieved in 15-20 minutes, without any coding, and shows how easy it is to perform Research Data analysis using Power BI.

Analyzing the data using natural language

At the top of the dashboard presentation, "Ask a question about your data" can be clicked and a question can be entered in natural language. We asked the question – Average page count by subject. A report got generated automatically and shown. It could be pinned to the dashboard, if needed. This is the Q&A feature of Power BI, that helps with quick analysis of our research data. Figure 4 shows the result of our Q&A.
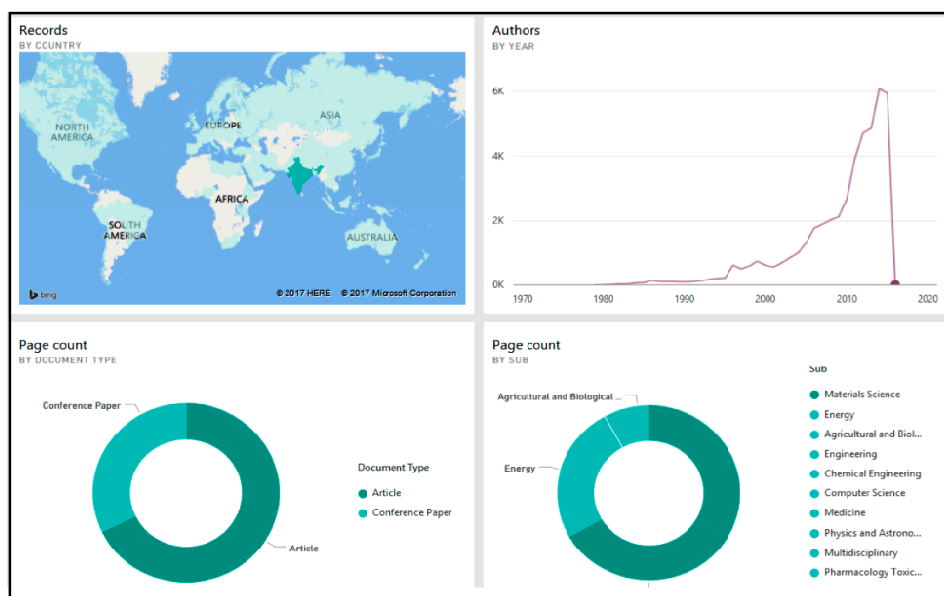

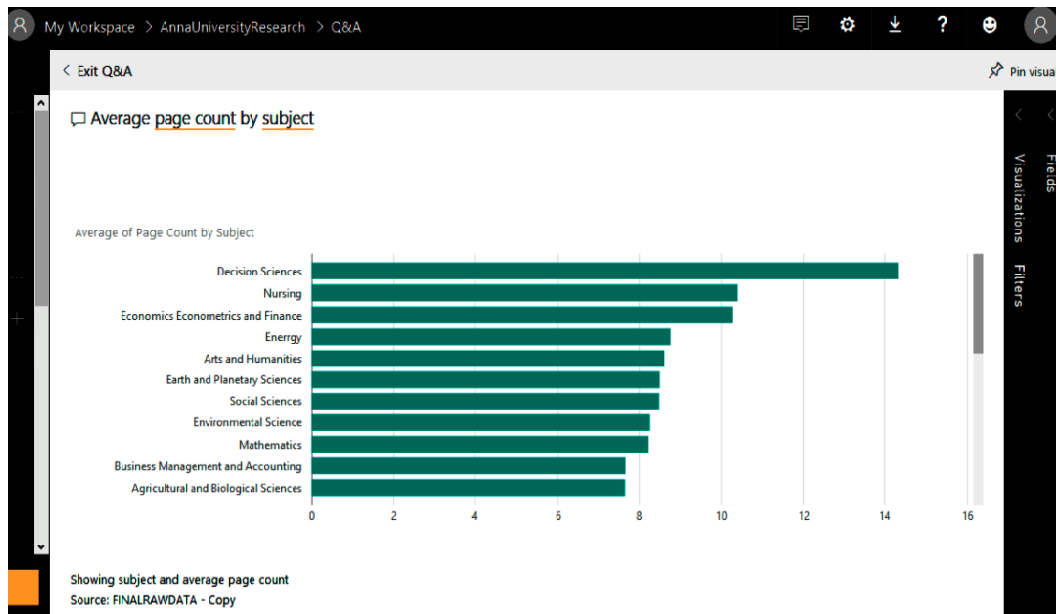**Figure 3: A partial view of the dashboard Anna University Research.**

Figure 4: Q&A feature allows users to ask questions in natural language to query the data, via the dashboard.

Sharing Reports and Dashboards

Data visualization and analytics being a tool in the management process, it is necessary to share the reports and dashboards in quick time for appropriate action. Power BI enables various options to share reports and dashboards (ajayan, 2017). One can publish the results to a website or blog, share it with other users within the organization, and share it with user groups, etc. with confidence that those sharing rules will be applied effectively.

Conclusion

Our experience with Power BI has shown that it is a radical approach to simplifying the business intelligence and data analytics space, whereby individuals and organizations can easily provide data, build reports or have them automatically created, aggregate them in dashboards, and share it with minimal investment of time and effort. When this service is delivered by an organization of the repute of Microsoft, with independent verification by Gartner which has compared the competition, it is evident that Power BI is a unique opportunity for research institutions and professionals to fulfil their data analysis needs.

References

1. Miguel Martinez, 2017, "Gartner positions Microsoft as a leader in BI and Analytics Platforms for ten consecutive years", Accessed online at https://powerbi.microsoft.com/en-us/blog/gartner-positions-microsoft-as-a-leader-in-bi-and-analytics-platforms-for-ten-consecutive-years/

2. Michael Hart, 2017, "Quick Insights with Power BI", Accessed online at https://powerbi.microsoft.com/en-us/documentation/powerbi-service-auto-insights/

3. Microsoft Documentation, 2017, "What is Power BI", Accessed online at https://powerbi.microsoft.com/en-us/what-is-power-bi/

4. Microsoft, 2017, "Choose how to get started", Accessed online at https://powerbi.microsoft.com/en-us/get-started/

5. Maggie Sparkman, 2017, "Frequently asked questions about Power BIFrequently asked questions about Power BI", Accessed online at https://powerbi.microsoft.com/en-us/documentation/powerbi-frequently-asked-questions/

6. Microsoft, 2017, "Guided Learning", Accessed online at https://powerbi.microsoft.com/en-us/guided-learning/

7. Adam Saxton, 2017, "Signing up for Power BI with a new Office 365 Trial", Accessed online at https://powerbi.microsoft.com/en-us/documentation/powerbi-admin-signing-up-for-power-bi-with-a-new-office-365-trial/

8. Michele Hart, 2017, "Power BI - basic concepts for Power BI service", Accessed online at https://powerbi.microsoft.com/en-us/documentation/powerbi-service-basic-concepts/

9. David Iseminger, 2017, "Get data from files", Accessed online at https://powerbi.microsoft.com/en-us/documentation/powerbi-service-get-data-from-files/

10. David Iseminger, 2017, "Get data from Excel workbook files", Accessed online at https://powerbi.microsoft.com/en-us/documentation/powerbi-service-excel-workbook-files/

11. Michele Hart, 2017, "Create a new Power BI report", Accessed online at https://powerbi.microsoft.com/en-us/documentation/powerbi-service-create-a-new-report/

12. Ajayan, 2017, "How should I collaborate on & share dashboards and reports?", Accessed online at https://powerbi.microsoft.com/en-us/documentation/powerbi-service-how-should-i-share-my-dashboard/

About Authors

Mr. Vijay Krishnan, Research Scholar, Anna University, Chennai, India
Email: vijayreports@live.com

Dr. S. Bharanidharan, Computer Scientist, Anna University, Chennai, India
Email: s_bharani@annauniv.edu

Dr. G. Krishnamoorthy, Director, University Library. Anna University, Chennai, India
Email:krishnamoorthyg(at)annauniv.edu