

GRAPH ANALYSIS USING MDIPL DATA

Synopsis

Submitted in partial fulfilment of the requirement for the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE & ENGINEERING

By

Avinav (CSE-19/27)

Sahil (CSE-19/71)

Under the supervision of

Rohit Arya



Department of Computer Science & Engineering

Satyug Darshan Institute of Engineering & Technology

Faridabad-121002

MARCH-2023

CERTIFICATE

This is to certify that the project titled “ **GRAPH ANALYSIS USING MDIPL DATA** ” is the bona fide work carried out by Avinav and Sahil student of B.Tech (CSE) of Satyug Darshan Institute of Engineering and Technology, Bhopani (Faridabad) affiliated to **JC BOSE University, Faridabad, Haryana(India)** in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology (Computer Science and Engineering)** and that the project has not formed the basis for the award previously of any other degree, diploma, fellowship or any other similar title.

Signature of the Guide

ACKNOWLEDGEMENT

I would like to take this opportunity to thanks everybody who guided me and extended possible help for the successful completion of the project “GRAPH ANALYSIS USING MDIPL DATA” .This work would not have been a reality without the permission of _____, Department of Computer Science.

I am indebted to _____, Project Guide, to whom I owe the success of this project, for his whole hearted corporation and guidance, which helped me in overcoming the teething problems of an industrial environment and carrying on the project work smoothly.

I am thankful to all of my friends who involved in the development of project and also indebted to my friends whose moral support always inspired me to come out with the best.

Last but not the least, my project would not have seen the light of the day without the blessings of my parents which was a constant help in the times of need.

Avinav (19020004027)

Sahil(19020004071)

ABSTRACT

Nowadays, Data Analytics is playing a crucial role in many sectors including financial organizations, manufacturing, etc. Analyzing data and presenting data as dashboards is one of the recent strategies. Dashboards help in saving time of the analyst working in any organization, typically by highlighting overall statistics in graphs or summaries in tables. The dashboard makes it easy to understand the scenario and support the decision-making process effectively. We propose the dashboard mechanism to monitor the entire Administrative work of Clients and Customers and their leading project that will help in reducing man-hours of any organization. The basic information which will include in our project will be the analysis of Project, their Customers, Suppliers and total Bill for specific Project and Customer. The proposed system not only helps to showcase the statistical data in the form of graphs but also gives the facility to drill down the summary for in-depth analysis. It further helps in reducing man-hours and makes the work to be done efficiently and effectively so as to answer the queries of the organization from anywhere.

Keywords - Data Analytics, Dashboard, Google Spreadsheet, Decision Making, Organization , Statistical Data

TABLE OF CONTENT

S.No.	Particulars
1.	Introduction to the Project
	1.1. Title
	1.2. Objective/Scope
2.	Literature Review
3.	Technology to be Used
	3.1. Python
	3.2. Power BI
	3.3. Libraries and Packages Included
4.	Implementation Methodology
5.	Future Enhancements
6.	References

INTRODUCTION TO PROJECT

This project aims to create a dashboard that analyzes and visualizes data from MDIPL using various Python libraries. The MDIPL dataset is a comprehensive collection of data from external APIs that provide information about a variety of factors that impact business operations, such as sales, project, customers and their ordering behaviour. The dashboard will enable business stakeholders to monitor Key Performance Indicators (KPIs) in real-time and make data-driven decisions based on the insights gained from the data analysis.

The project involves the collection and processing of data from the MDIPL dataset, including Data Cleaning, Data Preprocessing, and Data Integration. The Data is then visualized using various Python libraries, such as Matplotlib, Seaborn, Patchworklib, Plotly and many more to create Interactive Charts, Graphs, and other visualizations that enable stakeholders to quickly identify Trends and Patterns in the Data. The dashboard will be developed using the Dash framework, a popular open-source platform for building web-based applications in Python.

The dashboard will include features such as Filtering, Sorting, and Searching, allowing users to drill down into specific data points and gain deeper insights into the Data. Additionally, the dashboard will be designed to be Scalable and Responsive, capable of handling large amounts of data and accommodating a growing user base.

Overall, this project represents a valuable contribution to the field of Data Analytics, Business Analytics and Dashboard Development, providing a powerful tool for businesses to monitor their KPIs in real-time and make data-driven decisions based on the insights gained from the analysis of the MDIPL dataset.

OBJECTIVE

The objective of creating a dashboard with real-time data

- Is to provide a comprehensive and interactive tool for business stakeholders to monitor Key Performance Indicators (KPIs) in real-time and make data-driven decisions based on the insights gained from the data analysis.
- The dashboard will enable users to visualize and analyze real-time data from various sources. Users can gain a more comprehensive understanding of their business operations and adapt quickly to changing market conditions.
- The dashboard will be designed to be user-friendly and visually appealing, with a range of interactive visualizations, such as charts, tables, and maps, that enable users to explore the data and gain deeper insights.
- To provide a powerful tool for businesses to monitor their KPIs in real-time, enabling them to make data-driven decisions.
- By providing a flexible and customizable solution, the dashboard can be tailored to meet the specific needs of different businesses, making it a valuable asset for businesses across different industries.
- Providing a powerful tool for businesses to monitor their KPIs in real-time and make data-driven decisions based on the insights gained from the analysis.

LITERATURE REVIEW

In this research we conducted a state of the art review about Data Visualization and identified a set of distinct Data Visualization techniques. A Dynamic Dashboard for Analyzing MDIPL data developed using the Dash Framework as the base for development.

Due to the massive amount of information that exists and is created every day, the number of existing artifacts and the needs of all users and consumers of knowledge out there, it is difficult to find ways to present the information in an accessible way or make sense out of it. Our proposal is to develop a Dashboard where users will be able to interact with the information, based on an initial set of hints, charts, tables and reports, produced by the Dashboard. This Dashboard will be free and easy to use by users. No matter how great the technology, a Dashboard's success as a medium of communication, is a product of design, a result of a display that speaks clearly and immediately. Regarding the implemented Dashboard, it is functional and can be used for the basic aim for which it was designed.

From this research work, a scientific article was published in the CISTI'2018 - 13th Iberian Conference on Information Systems and Technologies, that was held between the 13th and 16th of June 2018, at Cáceres, Spain. This article was presented at the Conference by Renato Toasa (see Appendix H).

TECHNOLOGIES TO BE USED

PYTHON:

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. Python is a general-purpose interpreted, interactive, object-oriented, and high-level programming language. It was created by Guido van Rossum during 1985- 1990. Like Perl, Python source code is also available under the GNU General Public License.



Python has been around since 1990. It is one of the best programming languages widely used in data analytics. It is easy to use, fast, and manipulates data seamlessly. It supports various data analytics activities such as data collection, analysis, modelling, and visualization. Python offers the best features in the industry today. Reasons for this are that it increases productivity, quality of coding, and saves significant amounts of time for programmers. Python has been popular for a number of reasons, including its simple syntax that mimics natural language, its versatility, its affordability (open-source, free-to-use), its active support community, a dynamic resource for other programmers to seek support and guidance from, and its extensive archive of modules and libraries. Python also provides seamless integration with other data science technologies such as TensorFlow and SQL. Python is popularly used by statisticians, engineers, and scientists to perform data analytics. Here are some of the reasons why Data Analytics using Python has become popular:

1. Python is easy to learn and understand and has a simple syntax.
2. The programming language is scalable and flexible.
3. It has a vast collection of libraries for numerical computation and data manipulation.
4. Python provides libraries for graphics and data visualization to build plots.
5. It has broad community support to help solve many kinds of queries.

POWER BI:

Power BI is an interactive Data Visualization software product developed by Microsoft with a primary focus on Business Intelligence. It is part of the Microsoft Power Platform. Power BI is a collection of software services, apps, and connectors that work together to turn unrelated sources of data into coherent, visually immersive, and interactive insights. Data may be input by reading directly from a database, webpage, or structured files such as spreadsheets, CSV, XML, and JSON.



Power BI is a collection of software services, apps, and connectors that work together to turn your unrelated sources of data into coherent, visually immersive, and interactive insights. Your data might be an Excel spreadsheet, or a collection of cloud-based and on-premises hybrid data warehouses. Power BI lets you easily connect to your data sources, visualize and discover what's important, and share that with anyone or everyone you want. Power BI consists of several elements that all work together, starting with these three basics:

- A Windows desktop application called *Power BI Desktop*.
- An online software as a service (SaaS) service called the *Power BI service*.
- Power BI Mobile apps for Windows, iOS, and Android devices.

These three elements—Power BI Desktop, the service, and the mobile apps—are designed to let you create, share, and consume business insights in the way that serves you and your role most effectively.

Beyond those three, Power BI also features two other elements:

- Power BI Report Builder,
- Power BI Report Server.

One common workflow in Power BI begins by connecting to data sources in Power BI Desktop and building a report. You then publish that report from Power BI Desktop to the Power BI service, and share it so business users in the Power BI service and on mobile devices can view and interact with the report.

This workflow is common, and shows how the three main Power BI elements complement one another.

LIBRARIES AND PACKAGES INCLUDED

1. Requests

Requests enable you to send HTTP/1.1 requests and include headers, form data, multipart files, and parameters using basic Python dictionaries. Similarly, it also enables you to retrieve the answer data.

2. Dash

From exploring data to monitoring your experiments, Dash is like the front end to the analytical Python backend. This productive Python framework is ideal for data visualization apps particularly suited for every Python user. The ease we experience is a result of extensive and exhaustive effort.

3. Pandas

It is an open-source, BSD-licensed library. Pandas enable the provision of easy data structure and quicker data analysis for Python. For operations like data analysis and modeling, Pandas makes it possible to carry these out without needing to switch to more domain-specific language like R.

4. Matplotlib

All the libraries that we have discussed are capable of a gamut of numeric operations, but when it comes to dimensional plotting, Matplotlib steals the show. This open-source library in Python is widely used for publishing quality figures in various hard copy formats and interactive environments across platforms. You can design charts, graphs, pie charts, scatterplots, histograms, error charts, etc., with just a few lines of code.

5. NumPy

When it comes to scientific computing, NumPy is one of the fundamental packages for Python, providing support for large multidimensional arrays and matrices along with a collection of high-level mathematical functions to execute these functions swiftly.

6. Plotly Express

The `plotly.express` module (usually imported as `px`) contains functions that can create entire figures at once, and is referred to as Plotly Express or PX. Plotly Express is a built-in part of the plotly library, and is the recommended starting point for creating most common figures. Every Plotly Express function uses graph objects internally and returns a `plotly.graph_objects.Figure` instance.

7. Seaborn

When it comes to the visualization of statistical models like heat maps, Seaborn is among the reliable sources. This Python library is derived from Matplotlib and is closely integrated with Pandas data structures.

IMPLEMENTATION METHODOLOGY

- **Define the requirements:** Identify the key metrics, data sources, and visualizations that will be included in the dashboard. This will provide a clear direction for the development process.
- **Choose a suitable framework:** There are many Python frameworks available for creating dashboards, such as Dash, Bokeh, and Flask. In our Project we will be using Dash Plotly.
- **Gather and preprocess the data:** Collect and preprocess the data that will be displayed on the dashboard. This may involve retrieving data from APIs, databases, or files and cleaning and transforming the data to prepare it for visualization. We will fetch the data through an external API.
- **Create the visualizations:** Use the chosen framework to create the visualizations for the dashboard. This may include charts, tables, and other graphical elements. The visualizations are visually appealing and must effectively communicate the data.
- **Design the user interface:** Create a user interface for the dashboard that is intuitive and easy to use. This may include navigation menus, filters, and other interactive elements. We will use Power BI desktop to create UI for the dashboard.
- **Integrate the visualizations and user interface:** Integrate the visualizations and user interface into the chosen framework, using HTML, CSS, and JavaScript as per choice.
- **Test the dashboard:** Thoroughly test the dashboard to ensure that it is functioning as expected. Test for usability, performance, and accuracy of data.
- **Deploy the dashboard:** Once the dashboard has been thoroughly tested, deploy it to the production environment. This may involve setting up a web server.
- **Maintain and update the dashboard:** As the data changes and new requirements emerge, maintain and update the dashboard.

FUTURE ENHANCEMENTS

- Improving the ability to share and collaborate on dashboards, or adding support for version control.
- As the amount of data being displayed on the dashboard grows, performance can become an issue. Future enhancements could include improving the performance of data processing and visualization, or adding support for more powerful hardware.
- The integration of AI and ML algorithms can enhance the decision-making capabilities of dashboards by analyzing data patterns, identifying trends, and providing data-driven insights.
- By incorporating predictive analytics algorithms into the dashboard, users can gain insights into future trends and make more informed decisions.
- Ensuring that dashboards are optimized for mobile devices, so users can access data on the go.
- Dashboards could be enhanced to support multiple languages, allowing users to interact with the dashboard in their preferred language.
- Dashboards could be enhanced to support voice commands, providing a hands-free experience for users.
- Improving the performance of real-time data visualization, or adding support for streaming data.
- We can add more interactive components, such as drop-down menus, sliders, and buttons to provide more options for user input.

CONCLUSION

We proposed a dashboard that is useful for representing data in Visual form so as to make available all information at one place, which will save time, as well as represent all data in organized format. Because of central and accurate availability of data the higher authorities can focus on improvement of organization in terms of administration and results.

We conclude that, using the manual system we require manpower to get accurate data and accurate count of data, but using our dashboard we don't require many manpower to get data, we can easily visualize our organizational data. Our contribution is that we develop the dashboard for financial data of an organization so to take the decisions very quickly and accurately using the proposed dashboard to increase Productivity.

Thus, the proposed system is expected to solve the problems of diversity, equity and integration into the current system.

REFERENCES

- <https://www.mygreatlearning.com/blog/open-source-python-libraries/>
- <https://plotly.com/python/creating-and-updating-figures/>
- https://github.com/Coding-with-Adam/Dash-by-Plotly/blob/master/Other/Dash_Introduction/intro.py
- <https://www.kaggle.com/>
- <https://colab.research.google.com/drive/1z003LabPwofbsN87xs36FBqKcMgTswO4?usp=sharing#scrollTo=nearby-underground>