

|  |  |
| --- | --- |
| **Student Name/ID Number:** | Syukur Sidiq Nur Alam/ BDSE-0922-082 |
| **Academic Year:** | 2023-2024 |
| **Unit Assessor:** | Ei Thandar Khaing |
| **Project Title:** | Project Report |
| **Issue Date:** | 06 Jan 2024 |
| **Submission Date:** | 21 Jan 2024 |

|  |
| --- |
| **Learner declaration** |
| I certify that the work submitted for this assignment is my own and research sources are fully acknowledged.  Student signature: Date: 21 Jan 2024 |

**Contents**

# Project Initiation and Requirement Gathering

* 1. **Problem Statement**

HelloGreen, a company, is having trouble figuring out what causes staff attrition. The project manager, Mr. Lim, uses a data-driven strategy to find factors that impact workers and keep the best ones on board.

* 1. **Objectives**

The research aims to identify and analyze the elements that contribute to employee attrition.

- Create a data analytics dashboard-based corporate intelligence system.

- Assess how HelloGreen's problems can be addressed via data analytics.

- Evaluate the dashboard's usefulness in aiding in decision-making.

* 1. **Scope**

The project scope includes:

* Identifying business processes.
* Analyzing data sets.
* Implementing business intelligence solutions.
* Creating a data analytics dashboard.
  1. **Constraints**

The project has a limited timeframe for completion and limited human and technological resources.

* 1. **Assumptions and dependencies**

In developing the data analytics dashboard for HelloGreen's staff attrition analysis, certain assumptions and dependencies are integral to the project's success. Firstly, we presume the accuracy and currency of the provided data, as this forms the basis for generating credible insights. Additionally, our work relies on the existence of essential datasets encompassing employee demographics and attrition statistics. These assumptions underpin the reliability and efficacy of the ensuing analysis and visualization efforts.

* 1. **User Requirements**

Expectations from stakeholders include:

• The creation of an interactive dashboard that shows variables influencing attrition.

• The delivery of an extensive report detailing the vital function of data analytics.

• Offering a user-friendly interface adapted to the requirements of stakeholders.

* 1. **System Requirements**

Hardware :

* Intel(R) Core(TM) i7-8550U CPU
* RAM : 8GB
* HDD : 1TB

Software :

* Power BI
* Microsoft Excell
* Edge Microsoft
* Windows 10

# Data Analytics, BI and Data Science

* 1. **What data analytics is (using examples)**

Data analytics is the discipline that involves examining raw data to draw conclusions and insights from the information. Its application aids businesses in optimizing performance, operating more efficiently, maximizing profits, and making strategically guided decisions.

In various industries, data analysis plays a crucial role, as business leaders rely on data to inform their decision-making processes. For instance, a sneaker manufacturer might analyze sales data to decide which designs to continue producing and which to discontinue. Similarly, a healthcare administrator could examine inventory data to determine the necessary orders for medical supplies.  
Example:

Shoe Manufacturing: Sneaker manufacturers utilize data analysis by examining sales data. This analysis helps determine which designs are performing well in the market, thus allowing the company to make decisions on which designs to continue producing and which to discontinue. This ensures alignment of production with consumer preferences, ultimately improving overall business performance.

* 1. **Tools and techniques associated with it.**

In the realm of data analytics, analysts rely on various tools and techniques to process information and derive meaningful insights.

**Tools**

Data analysts benefit from reporting and explaining conclusions with the assistance of tools like Tableau and Power BI. Both are robust applications for data visualization and analysis, allowing users to compile information, conduct data analytics, and communicate results effectively through dashboards and reports. Additionally, emerging tools such as SAS, an analytics platform specializing in data mining, and Apache Spark contribute to the evolving landscape of tools available to data analysts.

**Techniques**

Data analysts leverage a range of analytical methods to process data and extract valuable information. Some widely used techniques include:

Regression Analysis: Involves scrutinizing the relationship between dependent variables to understand how a change in one variable may influence another, aiding in quantifying and identifying relationships within datasets.

Factor Analysis: This technique aims to condense large datasets into smaller, more manageable forms. By doing so, analysts seek to reveal hidden trends that might be challenging to discern in the original dataset, facilitating a more focused analysis.

Cohort Analysis: This process entails breaking down a dataset into groups of similar data, often based on customer demographics. Such segmentation enables analysts and other users to delve deeper into specific subsets of data, gaining nuanced insights.

Monte Carlo Simulations: These simulations model the probability of different outcomes and find applications in risk mitigation and loss prevention. By incorporating multiple values and variables, Monte Carlo simulations offer enhanced forecasting capabilities compared to other data analytics approaches.

Time Series Analysis: Involves tracking data over time to establish the relationship between the value of a data point and its occurrence. Frequently used to identify cyclical trends or project financial forecasts, time series analysis provides valuable insights into temporal patterns within datasets.

* 1. **What business intelligence is**

Business Intelligence (BI) constitutes a collection of strategies and technologies employed by enterprises to scrutinize business information, converting it into actionable insights that guide both strategic and tactical decision-making. BI tools operate by accessing and analyzing datasets, presenting their analytical findings through reports, summaries, dashboards, graphs, charts, and maps. This comprehensive approach furnishes users with detailed intelligence concerning the business's current state.

As an illustration, consider a company aiming to enhance its supply chain management. Leveraging BI capabilities becomes crucial for identifying delays and variabilities in the shipping process. The company can utilize BI tools to pinpoint which products frequently encounter delays and discern the modes of transportation most commonly associated with these delays. In this manner, BI empowers organizations to gain valuable insights, enabling them to make informed decisions for improving operational efficiency and overall business performance.

* 1. **Tools and techniques associated with it.**

Microsoft Excel, an early BI tool, empowered users to independently analyze data, bridging the gap between IT and end-users. Power BI, initially an Excel extension, has evolved into a standalone platform with components like the web-based Power BI Service for visualization and analysis, and the Power BI Desktop for data modeling and reporting.

BI Tools:

Microsoft Power BI: A scalable platform for self-service and enterprise BI. It connects to various data sources, provides end-to-end data protection, and offers 500+ connectors for on-premises and cloud data sources.

Tableau: A visual analytics platform transforming data use. Users create interactive dashboards, connect to diverse data sources, and visualize data effortlessly, uncovering insights within large datasets.

QlikView: A classic guided analytics solution, enabling the rapid development and delivery of interactive applications and dashboards for insightful data analysis.

* 1. **What data science is**

Data science is a multidisciplinary field that involves the use of scientific methods, processes, algorithms, and systems to extract insights and knowledge from structured and unstructured data. It combines expertise from various domains, including statistics, mathematics, computer science, and domain-specific knowledge, to analyze and interpret complex data sets. The goal of data science is to uncover patterns, trends, and valuable information that can inform decision-making, prediction, and optimization in diverse fields such as business, healthcare, finance, and technology. Data scientists often employ techniques such as machine learning, data mining, statistical analysis, and data visualization to derive meaningful insights from data.

* 1. **Tools and techniques associated with it.**

The process begins with obtaining data, potentially by recording user behavior across various websites. Subsequently, the second step involves extraction, cleaning, or annotation, aiming to reduce noise and eliminate unnecessary data. Moving on to the third step, integration, aggregation, and representation come into play, organizing the data in a manner conducive to supporting data mining. The fourth step encompasses analysis and modeling, incorporating machine learning and data mining techniques. Ultimately, in the fifth step, interpretation occurs, where the final results are thoroughly analyzed and understood.  
  
Data Science Tools:

a. Microsoft Azure Machine Learning:

Microsoft Azure Machine Learning empowers data scientists and developers to swiftly build, deploy, and manage high-quality models. It accelerates time to value with industry-leading machine learning operations (MLOps), open-source interoperability, and integrated tools. This AI learning platform is designed for responsible AI applications, ensuring robust and ethical machine learning practices.

b. Rstudio:

Rstudio serves as the integrated development environment (IDE) for the R programming language. Offering an excellent user interface, it simplifies code writing, data visualization, and interactive application development. Widely utilized by data scientists, RStudio's extensive package selection and features make it a go-to tool for statistical analysis, data visualization, and machine learning tasks.

c. Python:

Python stands out as a widely used programming language in the realm of Data Science. Equipped with built-in mathematical libraries and functions, Python facilitates easy calculation of mathematical problems and data analysis. Its rich ecosystem includes tools like NumPy, Pandas, SciPy, and scikit-learn, making it a versatile choice for data processing, analysis, and machine learning tasks.

* 1. **Compare and contrast data analytics, business intelligence, and data science.**

|  |  |  |
| --- | --- | --- |
| **Business Intelligence** | **Data Analysts** | **Data Scientists** |
| Grasp the business requirements and effectively convey the findings to the team. | Conduct data analysis to draw conclusions and make predictions for informed business decisions. | Develop and train predictive machine learning models to establish intelligent systems |
| Apply business frameworks to analyze data, such as marketing funnels and cohort analysis. | Engage in the process of cleaning and structuring data, as well as generating reports. | Might create prototypes to explore potential applications of data and then collaborate with engineering teams for implementation. |
| Engage in the tasks of cleaning data and creating reports and dashboards. | Utilize Python and statistical methods for forecasting events. |  |

# Data Types

* 1. **Differentiate between unstructured, semi-structured and structured data within an organization based on project scenario.**

Data within enterprises is widely divided into three types: unstructured, semi-structured, and structured. Each category indicates a distinct amount of structure and uniformity, which affects how data is kept, processed, and analyzed.

• Unstructured Data:

Unstructured data lacks a predefined structure or data model, often presenting itself in the form of free-text, multimedia, or other formats without a formal framework.

Project Scenario Examples:

Employee Feedback Comments: These comments may offer valuable insights into employee sentiment but lack a standardized framework.

• Semi-Structured Data:

Definition: Semi-structured data possesses some level of structure but does not adhere to the tabular format typical of relational databases. It may include tags, elements, or hierarchies.

Project Scenario Examples:

Survey Results: While surveys may feature predefined questions (providing structure), the responses may vary in length and type, resulting in semi-structured data.

• Structured Data:

Structured data is meticulously organized and adheres to a predefined data model, commonly found in relational databases and characterized by rows and columns.

Project Scenario Examples:

Employee Database: This structured dataset likely includes well-defined fields such as age, marital status, and other attributes following a consistent format.

* 1. **Why data analytics is important?**

\*\*3.2 Importance of Data Analytics\*\*

Data analytics is crucial for:

1. Informed Decision-Making:

Enables evidence-based decisions.

2. Operational Efficiency:

Identifies areas for process improvement.

3. Trend Recognition:

Helps anticipate future developments.

4. Customer Insights:

Provides valuable information for tailored strategies.

5. Competitive Advantage:

Offers a faster and more accurate analysis, ensuring a competitive edge.

6. Risk Management:

Identifies and mitigates potential risks.

7. Innovation:

Fosters product and service development aligned with market demands.

8. Personalization:

Enables tailored customer interactions for a better experience.

# Data Analytics tools and techniques

* 1. **How data analytics tools can contribute to effective decision-making**

Data analysis tools empower organizations to rapidly examine extensive datasets, recognize patterns, and form decisions guided by data. Utilizing these tools allows for the generation of pertinent information from the data, offering valuable insights for business decision-making. This includes discerning customer requirements, enhancing operational efficiency, refining marketing strategies, and supporting businesses in making informed, data-driven decisions.

* 1. **Select appropriate data analytics tool.**

In this scenario, the author opted for the Power BI tool due to its capability to seamlessly connect to data sources, facilitate visualization, uncover essential insights, and allow easy sharing with specific individuals or a broader audience.

# Dashboard Development

* 1. **Create an HR attrition dashboard.**

Dashboard in Power BI Service



* 1. **Customize the design including key performance indicators (KPI) and visualizations and user-friendly features.**



# Data Analytics

* 1. **Apply exploratory data analytics techniques to uncover patterns, trends, and actionable insights.**

Based on the data analytic techniques explained above, the author chose the factor analysis technique, namely shrinking large data that includes several other data into smaller data. This technique can be said to be like concluding data. By first carrying out an analysis of attrition, we try to reveal the factors related to attrition. Apart from that, the dashboard design created also follows current trends.

# Evaluation

* 1. **Critical review of the design**

By creating a fresh blue themed dashboard, the dashboard will feel more elegant and contemporary. Considering the user experience, usually users will feel mesmerized if they have a sea blue background.

With various color combinations, the author makes several cards that are appropriate and clear to see. The author also includes data related to attrition so that the data is easier to understand and the company can make decisions easily.

* 1. **Identify what customization has been integrated into the design**

I applied some cards like total employees, total exits, active employees, t and average exits.

I also implemented some static diagrams so that I can see the static Attrition in more detail.

