

|  |  |
| --- | --- |
| **Student Name/ID Number:** | Muhammad Kemal/ BDSE-0922-084 |
| **Academic Year:** | 2023-2024 |
| **Unit Assessor:** | Ei Thandar Khaing |
| **Project Title:** | Project Report |
| **Issue Date:** | 06-01-2024 |
| **Submission Date:** | 22-01-2024 |

|  |
| --- |
| **Learner declaration** |
| I certify that the work submitted for this assignment is my own and research sources are fully acknowledged.        Student signature:  Date: 22-01-2024 |

**Contents**

1. Project Initiation and Requirement Gathering
   1. Problem Statement

HelloGreen Company aimed to understand the factors that contribute to employee turnover. The project manager, Mr. Lim, recognized the importance of identifying the elements that influence employees and how to retain the best ones. To achieve this, a data analytics dashboard was created to uncover important factors, such as distance from home, age, marital status, and more, that affect employee turnover.

* 1. Objectives

The Objectives of this project include:

1. Identify and analyze factors contributing to employee turnover within HelloGreen Company.
2. Develop insights into how various elements, such as distance from home, age, marital status, and more, impact employee retention.
3. Provide actionable recommendations for retaining key employees and improving overall employee satisfaction.
4. Create a data analytics dashboard to visualize and interpret the findings effectively.
   1. Scope

The following scope of this project are:

1. Analyzing employee turnover data to identify patterns and trends.
2. Exploring factors such as distance from home, age, marital status, and other relevant parameters.
3. Implementing a data analytics dashboard to present the insights in an accessible and user-friendly manner.
4. Recommendations will be focused on strategies to enhance employee retention.
5. The project will be limited to factors within the control of HelloGreen Company, excluding external economic factors.
   1. Constraints

There is contrains in this project including:

1. The project relies heavily on the availability and accuracy of employee data. Incomplete or inaccurate data may impact the reliability of the findings.
2. Data privacy and confidentiality must be strictly maintained in adherence to company policies and legal regulations.
3. The project operates within a defined timeframe, and the analysis and recommendations must be delivered within this schedule.
   1. Assumptions and dependencies

In the pursuit of understanding employee turnover at HelloGreen Company, certain assumptions are made and dependencies identified to provide clarity on the context and expectations of the project.

1. Assumptions:

It is assumed that the employee data provided for analysis is accurate and up-to-date. Any discrepancies may impact the reliability of the insights.

The factors chosen for analysis, including distance from home, age, marital status, etc., are assumed to be relevant contributors to employee turnover. Adjustments may be made based on emerging insights.

1. Dependencies:

The project is dependent on the availability and accessibility of relevant employee turnover data. Delays or restrictions in accessing this data may impact the project timeline.

The effectiveness of the project depends on active participation and cooperation from employees during data collection, surveys, and feedback sessions.

The project assumes that necessary approvals from the management have been obtained, and any changes in the project scope or direction will be communicated and approved accordingly.

* 1. User Requirements

Understanding the user needs was critical to the success of this project. In the context of HelloGreen Company's initiative to understand employee turnover, the following user needs have been identified:

1. Data Access:

Users need a system that provides easy and secure access to relevant employee turnover data. The system must ensure data privacy and comply with legal regulations.

1. User-Friendly Dashboards:

Users want data analytics dashboards that are intuitive and easy to use, enabling easy navigation and interpretation of key insights related to employee turnover factors.

1. Customizable Data Display:

Dashboards should allow users to customize views based on different parameters such as department, tenure, or certain demographic factors. This ensures that insights are tailored to the individual needs of the user.

1. In-depth Analysis of Factors:

Users need an in-depth analysis of the various factors that influence employee turnover, including distance from home, age, marital status, and other relevant parameters. The system should provide in-depth insights into each factor.

* 1. System Requirements

To implement a data analytics dashboard to understand and address employee turnover at HelloGreen Company, the following hardware specifications are required such as, a desktop environment with Windows 10/11 or Macintosh operating system, minimum 16 GB RAM, Intel Core i7 (10th generation) or AMD Ryzen 7 3xxx processor, 256 GB Solid State Drive (SSD) storage, dedicated graphics card (e.g. NVIDIA GeForce GTX 1660 or AMD Radeon RX 560), installation of advanced data analysis tools such as Power BI or Tableau, integration of a robust backup system and redundancy measures, and use of Microsoft Office tools for project documentation, preferably with the latest version for enhanced features and compatibility. This specification aims to provide a high-performance hardware setup capable of efficiently handling the complex computations involved in data analysis, ensuring optimal functionality and responsiveness of data analysis dashboards.

1. Data Analytics, BI and Data Science
   1. What data analytics is

Data analytics involves the meticulous analysis of factual information, such as measurements or statistics, to comprehend, discuss, or calculate. It requires a detailed examination to understand the nature and essential features of digital information. In the realm of data analysis, a data analyst plays a crucial role in studying and determining the relationships within factual information for reasoning, discussion, or calculation. This process involves extracting meaning from data through a rigorous analysis to obtain valuable insights. Data analytics is vital for optimizing business performance, as it enables companies to identify more efficient ways of operating, make better decisions, and analyze customer trends and satisfaction.

2.2 Tools and techniques associated with it.

Data Analytics Steps :

1. Define Data Requirements:

Identify the specific data needed to address the research question or solve the problem at hand. This includes outlining variables, pinpointing data sources, and considering any constraints or limitations that may apply.

1. Collect Data:

Gather data from various sources such as databases, surveys, APIs, and web scraping. It is imperative to ensure the accuracy and quality of the collected data, maintaining a high standard throughout the collection process.

1. Organize and Analyze:

Once the data is collected, it needs to be organized and prepared for analysis. This involves cleaning and transforming the data, addressing missing values, and structuring it in a way that facilitates subsequent analysis.

1. Cleanse Data:

Identify and rectify errors, inconsistencies, or duplicates within the data to enhance accuracy and reliability, ensuring that the dataset is free from discrepancies.

1. Conduct Data Analysis:

Apply various analytical techniques, such as statistical analysis, data mining, machine learning, or other relevant methods, after cleaning the data. This step aims to gain insights and effectively address the research question based on the nature of the data and the analysis goals.

1. Visualize and Share Findings:

Following the analysis, present the results visually using charts, graphs, or other visual aids. This visual representation facilitates clear communication of insights and findings to stakeholders or decision-makers, empowering them to make informed decisions grounded in the analyzed data.

2.3 What business intelligence is Definition of Business Intelligence:

Business intelligence (BI) is a technology-based process for analyzing data and delivering actionable information that helps executives, managers and workers make informed business decisions. As part of the BI process, organizations collect data from internal IT systems and external sources, prepare it for analysis, run queries against the data and create data visualizations, dashboards and BI reports to make the analysis results available to business users for operational decision-making.

Evolution of Business Intelligence:

The evolution of Business Intelligence (BI) has traversed several transformative phases, beginning with manual reporting and progressing to the current era of sophisticated analytics. Initially reliant on static reports, BI advanced with Decision Support Systems, introducing more interactive interfaces. The advent of Online Analytical Processing (OLAP) brought multidimensional analysis, while data warehousing centralized data storage for comprehensive reporting. The integration of data mining enhanced pattern discovery, leading to the era of Enterprise BI with end-to-end solutions. The recent shift towards Self-Service BI democratizes access, allowing non-technical users to create analyses. Advanced analytics and AI integration, real-time analytics, and cloud-based solutions represent the latest frontiers, reflecting BI's continual adaptation to meet the dynamic demands of the modern business landscape.

2.4 Tools and techniques associated with it.

Microsoft Excel, acknowledged as an early business intelligence (BI) tool, empowered users to independently analyze data, facilitating work even when sourced from the IT department or historical data. Initially developed within Excel, Power BI has expanded beyond its origins, comprising the Power BI Service, a web-based platform for visualization and in-depth analysis, and the Power BI Desktop, a free tool for business analysts offering data modeling, transformation, and report creation. As a unified, scalable platform, Power BI caters to both self-service and enterprise BI, seamlessly connecting to diverse data sources with end-to-end data protection. It integrates effectively with Azure and Microsoft 365, offering a library of 500+ connectors. Another notable BI tool, Tableau, revolutionizes data usage through interactive dashboards and visualizations, allowing users to uncover insights in large datasets. Similarly, QlikView serves as a guided analytics solution, enabling the rapid development and delivery of interactive applications and dashboards.

2.5 What data science is

Data science is the study of data to extract meaningful insights for business. Data science is a multidisciplinary approach that combines principles and practices from the fields of mathematics, statistics, artificial intelligence, and computer engineering to analyze large amounts of data. This analysis helps data scientists to ask and answer questions such as what happened, why it happened, what will happen, and what can be done with the results.

Data science enables businesses to discover new patterns and relationships that have the potential to transform organizations. It can reveal low-cost changes to resource management for maximum impact on profit margins. For example, an e-commerce company, using data science, discovered that there were too many customer queries being generated after hours. Investigation revealed that customers would be more likely to buy if they were responded to immediately, rather than during business hours the next day. By implementing customer service 24 hours a day and 7 days a week, the business would increase its revenue by 30%.

* 1. Tools and techniques associated with it.
     1. Techniques for data science include:

1. The initial phase of this procedure involves acquiring data, which could encompass the tracking of user interactions across various websites.
2. Following that, the second stage encompasses extraction, cleansing, or annotation, aiming to minimize interference and remove extraneous data.
3. Subsequently, the third stage involves integration, aggregation, and representation, wherein the data is structured to facilitate effective data mining.
4. The fourth stage entails analysis and modeling, where the application of machine learning and data mining techniques occurs.
5. Ultimately, the fifth and final step is interpretation, involving a thorough analysis of the conclusive results obtained from the process.
   * 1. Tools for data science:
   1. Python:

Python stands as a versatile and widely-used programming language in the field of data science. Its extensive ecosystem includes libraries such as NumPy, Pandas, Scikit-learn, and TensorFlow, making it a powerful tool for data analysis, machine learning, and statistical modeling.

* 1. R:

R is a specialized programming language for statistical computing and graphics. Widely embraced in academia and industry, R provides an array of packages for statistical analysis, data visualization, and machine learning, making it a preferred tool for statisticians and data scientists.

* 1. Jupyter Notebooks:

Jupyter Notebooks serve as an interactive computing environment that allows the creation and sharing of documents containing live code, equations, visualizations, and narrative text. Widely employed for data exploration, analysis, and collaborative work, Jupyter Notebooks support multiple programming languages, including Python and R.

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

Data analytics and business intelligence often have a retrospective focus, analyzing historical and current data to understand past performance and inform present decisions. In contrast, data science incorporates both historical and future-oriented perspectives, utilizing predictive modeling and machine learning to make future predictions.

* + 1. Business Intelligence (BI):

Data analytics typically involves statistical analysis and visualization techniques. Business intelligence relies on reporting, querying, and dashboard tools. Data science employs a broader range of techniques, including advanced statistical modeling, machine learning, and algorithm development.

* + 1. Data Science:

While all three disciplines deal with data, data analytics and business intelligence often work with structured data from traditional databases. Data science, on the other hand, deals with both structured and unstructured data, often including big data sources and various data formats.

1. Data Types

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

* + 1. Unstructured Data:

Unstructured data lacks a defined format or structure. In the context of this project, examples of unstructured data may encompass comments or feedback from employees stored in free-text form. For instance, if HelloGreen collects exit surveys or interviews from departing employees, the resulting unstructured data would consist of the open-ended responses and comments provided by the employees.

* + 1. Semi-Structured Data:

Semi-structured data possesses some structure but is not fully organized. Within the project scenario, information such as meeting notes or memos containing a mix of text, tables, and various elements falls under the category of semi-structured data. For instance, if HelloGreen maintains documents like meeting notes pertaining to employee performance or evaluations, the content may include both textual information and tabular data, making it semi-structured.

* + 1. Structured Data:

Structured data is well-organized with a specific format and schema. In the project context, structured data could consist of official information sourced from HelloGreen's HR database. For example, an employee table within the database may include columns such as age, marital status, distance to work, and more, presenting the data in a clearly organized and structured table or database format.

3.2 Why data analytics is important?

Data analytics holds paramount importance for organizations as it facilitates informed decision-making by extracting valuable insights from data. Through the analysis of historical and real-time data, businesses can identify trends, optimize operations, enhance customer experiences, and innovate in product development. The ability to measure performance, manage risks, and ensure compliance with regulations further underscores the significance of data analytics. Ultimately, leveraging data analytics not only provides a competitive advantage but also enables organizations to adapt to market changes, anticipate customer needs, and strategically plan for the future, making it an indispensable tool in the contemporary business landscape.

1. Data Analytics tools and techniques
   1. How data analytics tools can contribute to effective decision-making Data analytics tools play a crucial role in enhancing effective decision-making through various analytic approaches:

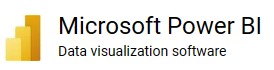
Data analytics tools play a pivotal role in augmenting effective decision-making through a multitude of analytic approaches. Firstly, these tools enable descriptive analytics, offering a comprehensive overview of historical data patterns and trends. This retrospective analysis provides decision-makers with valuable insights into past performance, facilitating a better understanding of what has transpired.

Secondly, predictive analytics, a key facet of data analytics tools, utilizes statistical algorithms and machine learning models to forecast future outcomes. By extrapolating patterns from historical data, decision-makers can anticipate trends and make proactive decisions to address potential challenges or capitalize on emerging opportunities.

Furthermore, prescriptive analytics goes beyond prediction by recommending specific actions to optimize outcomes. Data analytics tools employing prescriptive analytics help decision-makers by suggesting the most effective courses of action based on the analyzed data, allowing for strategic planning and resource optimization.

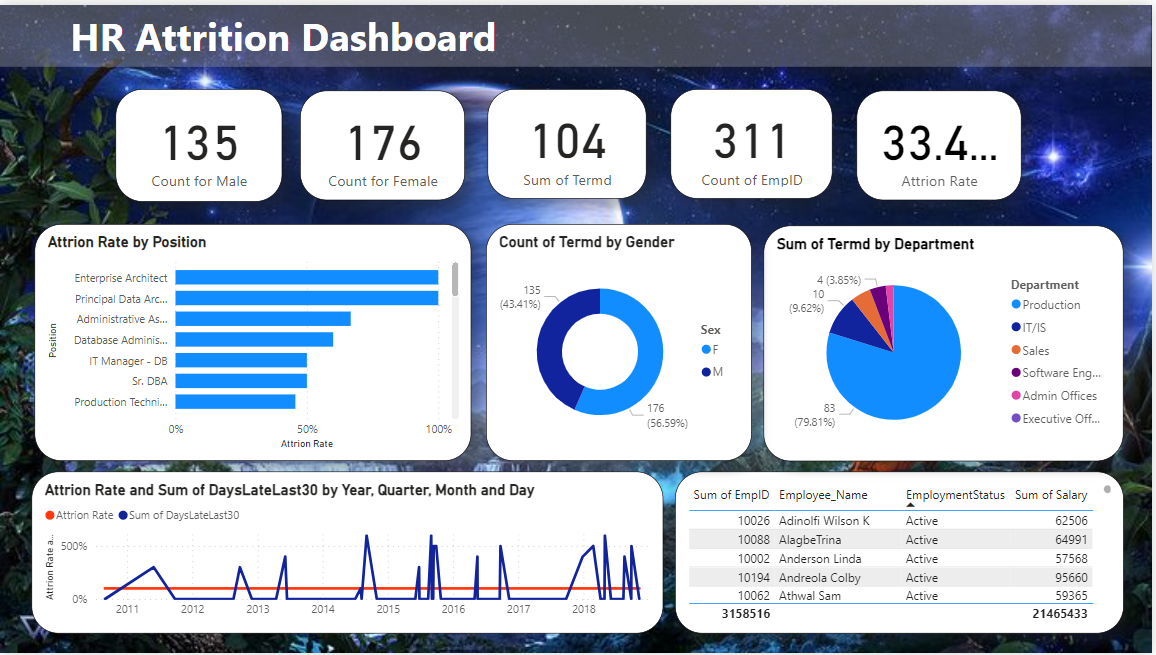
* 1. Select appropriate data analytics tool.
     1. Microsoft Power BI

Selecting Microsoft Power BI as a data analytics tool provides a user-friendly interface, seamless data integration, and cloud-based collaboration. Its intuitive design caters to both technical and non-technical users, allowing for the creation of compelling visualizations. Power BI's robust data modeling features, DAX language, and integration with Microsoft 365 and Azure enhance its capabilities, making it a comprehensive and accessible solution for effective decision-making.



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

In the customization phase, the design of the data analytics platform is tailored to meet specific needs, emphasizing key performance indicators (KPIs) and user-friendly features. This involves incorporating relevant KPIs that align with organizational goals and objectives, providing a focused view of performance metrics. Visualizations are carefully crafted to convey complex information intuitively, enhancing comprehension. User-friendly features are integrated to ensure accessibility for a diverse user base, enabling seamless interaction with the platform. Customization in design aims to optimize the user experience, making it efficient and tailored to the unique requirements of the organization.



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

In the realm of data analytics, the application of exploratory techniques is instrumental in unveiling patterns, trends, and actionable insights within datasets. This involves employing various statistical and visual methods to delve into the data's inherent structure and relationships. Exploratory data analytics allows for a preliminary understanding of the dataset, aiding in the identification of outliers, distributions, and potential correlations between variables. Through techniques such as data visualization, summary statistics, and data profiling, analysts can uncover valuable insights that serve as a foundation for more in-depth analysis and informed decision-making. The application of exploratory data analytics is a crucial initial step in the data analysis process, enabling a comprehensive exploration of the data landscape and setting the stage for subsequent analytical endeavors.

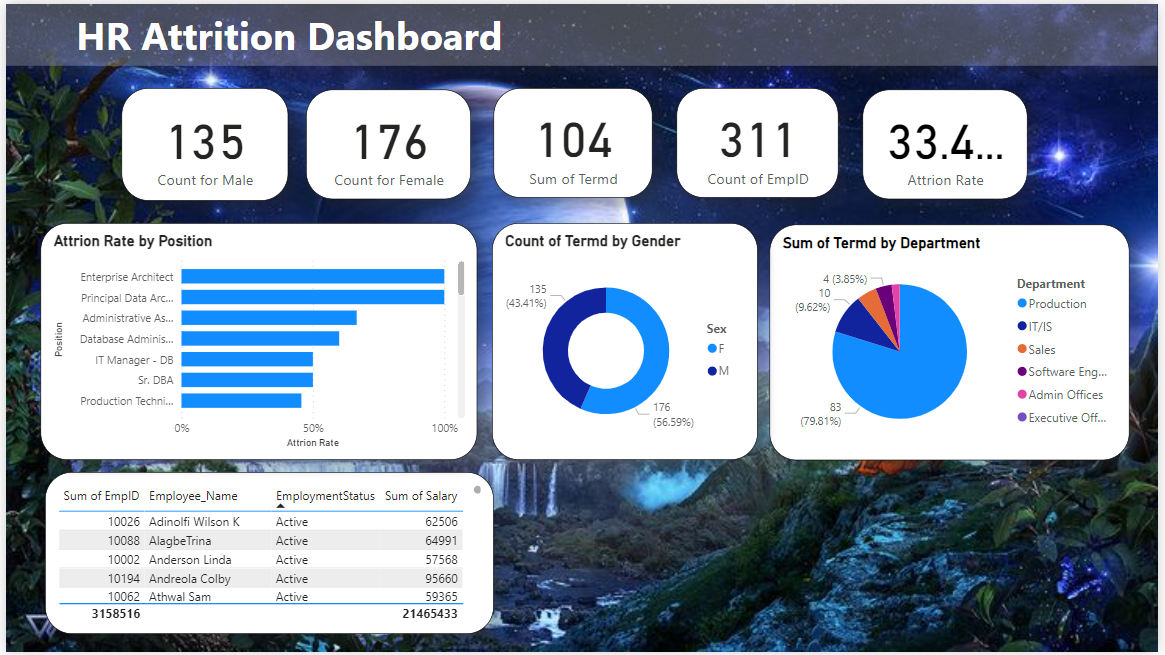
The HR Attrition Dashboard in data analytics is a comprehensive tool designed for monitoring and analyzing employee turnover within an organization. This dashboard encompasses features such as visualizations of employee demographics, attrition trends over time, and a breakdown of the reasons for employee departure. It provides key performance indicators related to employee performance, tenure analysis, and insights from exit interviews. Additionally, the dashboard includes metrics on the cost of attrition, employee satisfaction scores, and predictive analytics for forecasting potential attrition. With user-friendly interactive features, HR professionals can customize their view, enabling data-driven decision-making, proactive retention strategies, and a holistic understanding of attrition dynamics within the organization.

1. Evaluation
   1. Critical review of the design

The feedback received on the Power BI HR Dashboard was generally positive, with specific appreciation for the excellent visual design. The use of various elements such as donut charts, graphs, tables, and cards was commended. In particular, the display of cards with pictures was highly favored.

However, there were a few areas suggested for improvement. One of the key points raised was the need for alignment in the color theme. The combination of dark green and yellow was found to be less than optimal for clear readability. The text color was also mentioned as being difficult to discern. Therefore, it would be beneficial to reconsider these aspects to enhance the overall user experience of the dashboard.

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



After

