

|  |  |
| --- | --- |
| **Student Name/ID Number:** | Darren Farrell Andrian |
| **Academic Year:** | 2023-2024 |
| **Unit Assessor:** | Ei Thandar Khaing |
| **Project Title:** | Summative assignment |
| **Issue Date:** |  |
| **Submission Date:** | 19/1/2024 |

|  |
| --- |
| **Learner declaration** |
| I certify that the work submitted for this assignment is my own and research sources are fully acknowledged.    Student signature: Date: 19/1/2024 |

**Contents**

1. Project Initiation and Requirement Gathering
   1. Problem Statement

HelloGreen, a company grappling with employee attrition, has engaged 4V Data Solutions to address the challenge comprehensively. The problem at hand involves understanding the factors contributing to attrition and devising strategies to retain valuable employees.

* 1. Objectives

The primary objective is to identify and analyze key factors influencing employee attrition within HelloGreen. The project aims to develop a data analytics dashboard that provides insights into crucial elements such as distance from home, age, marital status, and more. This dashboard will enable informed decision-making to enhance employee retention strategies.

* 1. Scope

The project scope encompasses the development and implementation of a robust business intelligence solution, with a focus on creating a data analytics dashboard. The scope extends to the exploration of relationships between various factors contributing to attrition.

* 1. Constraints

The project must operate within specified time and budget constraints. The limitations in accessing certain data sources or the unavailability of specific data points may also pose constraints.

* 1. Assumptions and dependencies

Assumptions include the availability of necessary data, cooperation from relevant stakeholders, and adherence to project timelines. Dependencies may include external data sources, technology dependencies, and collaboration with HelloGreen's internal teams.

* 1. User Requirements

User requirements involve understanding the expectations and needs of Mr. Lim and other stakeholders at HelloGreen. This includes the desired features, functionality, and visual representations within the data analytics dashboard.

* 1. System Requirements

Software

* Power Bi Desktop
* HR dataset

Hardware

* Memory: 2 GB or more.
* Graphics Card: Intel HD or equivalent.
* CPU: Dual Core CPU or more.
* OS: Windows 10 / 11, MAC.

1. Data Analytics, BI and Data Science
   1. What data analytics is (using examples)

Data analytics involves the examination, interpretation, and visualization of data to extract meaningful insights and support decision-making. It encompasses various techniques and approaches to analyze and interpret data patterns. Examples of data analytics include:

* + Descriptive Analytics: Summarizing historical data to understand past trends.
  + Predictive Analytics: Forecasting future trends based on historical data.
  + Prescriptive Analytics: Recommending actions based on analysis.
  1. Tools and techniques associated with it.
* Tools
  + - Python and R: Programming languages for statistical analysis.
    - SQL: Query language for managing and analyzing structured data.
    - Tableau and Power BI: Visualization tools for creating interactive dashboards.
* Technique
* Regression Analysis: Predictive modeling to understand relationships between variables.
* Cluster Analysis: Grouping data points with similar characteristics.
* Time Series Analysis: Analyzing data points collected over time.
  1. What business intelligence is (using examples)

Business Intelligence (BI) involves the use of technologies, processes, and tools to transform raw data into actionable business insights. It focuses on reporting, querying, and data visualization to support business decision-making. Examples of BI include:

* Executive Dashboards: Visual summaries of key performance indicators (KPIs).
* Ad-Hoc Reporting: On-demand reporting for specific business queries.
* OLAP (Online Analytical Processing): Analyzing multidimensional data.
  1. Tools and techniques associated with it.
* Tools
* Microsoft Power BI: For creating interactive reports and dashboards.
* Tableau: Visualization tool for exploring and understanding data.
* SAP BusinessObjects: Suite of BI tools for reporting and analysis.
* Techniques
* Drill-Down Analysis: Going from summarized information to detailed data.
* Data Mining: Extracting patterns and knowledge from large datasets.
* ETL (Extract, Transform, Load): Process for data integration.
  1. What data science is (using examples)

Data science involves the extraction of knowledge and insights from structured and unstructured data through a combination of various techniques. It integrates statistics, machine learning, and domain expertise to solve complex problems. Examples of data science include:

* Recommendation Systems: Analyzing user preferences to suggest products.
* Natural Language Processing (NLP): Understanding and generating human language.
* Predictive Maintenance: Forecasting equipment failures to optimize maintenance.
  1. Tools and techniques associated with it.
     1. Tools
* Python and R: Core programming languages for data science.
* Jupyter Notebooks: Interactive computing environments for data analysis.
* TensorFlow and PyTorch: Frameworks for machine learning and deep learning.
  + 1. Technique
* Machine Learning Algorithms: Supervised and unsupervised learning techniques.
* Big Data Analytics: Processing and analyzing large volumes of data.
  1. Compare and contrast data analytics, business intelligence, and data science.
* Data Analytics:

Encompasses a broad range of techniques for analyzing and interpreting data, including descriptive, predictive, and prescriptive analytics. Focuses on understanding patterns, trends, and relationships within data.

* Business Intelligence:

Primarily concentrates on transforming raw data into actionable insights, emphasizing reporting, querying, and data visualization. Provides historical and current data analysis to support decision-making.

* Data Science:

Involves advanced statistical and machine learning techniques for predictive modeling and extracting actionable insights. Aims to derive insights from both historical and real-time data, focusing on complex and large datasets.

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

* Unstructured data  
  In the context of HelloGreen's employee attrition project, unstructured data refers to information that lacks a predefined data model or is not organized in a specific manner. Examples of unstructured data in this scenario could include employee feedback from surveys, comments, and open-ended responses. These data points may not follow a rigid format and may contain diverse information related to employee sentiments and experiences within the organization.
* Semi-structured data  
  Semi-structured data within the project scenario involves information that has some level of structure but does not fit neatly into a traditional relational database. Examples of semi-structured data could include employee profiles with varying fields, such as job titles, departments, and project involvements. While some structure may exist, the data may have flexible elements, allowing for the addition of new fields or attributes as needed.
* Structured data  
  Structured data in the project context pertains to well-organized and formatted information that fits into a conventional relational database. Examples include structured employee records containing standardized fields such as name, age, department, and performance ratings. Structured data allows for easy querying, sorting, and analysis, making it suitable for integration into the data analytics dashboard.  
  1. Why data analytics is important?
* Informed Decision-Making  
  In the case of HelloGreen, leveraging analytics helps in identifying patterns and trends related to employee attrition, enabling management to make strategic decisions to enhance employee retention.
* Predictive Analysis for Proactive Measures  
  For HelloGreen, this means predicting possible attrition scenarios based on historical data, enabling the implementation of proactive measures to retain key employees and mitigate risks.
* Improved Strategic Planning  
  By analyzing data, HelloGreen gains a deeper understanding of the factors contributing to employee attrition. This insight facilitates the development of targeted strategies and interventions, improving overall strategic planning for talent retention.
* Enhanced Operational Efficiency

In the context of employee attrition, this could involve optimizing HR processes to identify and address factors influencing attrition more effectively.

* Competitive Advantage

This means that being better equipped to create a work environment that attracts and retains top talent, giving the company a competitive edge in the market.

1. Data Analytics tools and techniques
   1. How data analytics tools can contribute to effective decision-making

* Insights Generation

Data analytics tools enable the exploration and analysis of vast datasets, extracting valuable insights that might not be apparent through traditional methods. By uncovering patterns, trends, and correlations, these tools empower decision-makers to make informed choices based on a deeper understanding of the data.

* Real-Time Monitoring

Many data analytics tools, including Power BI Desktop, offer real-time monitoring capabilities. This feature allows decision-makers to track key performance indicators (KPIs) and critical metrics, providing timely information for proactive decision-making. Real-time insights are particularly valuable in dynamic environments where quick responses are essential.

* Predictive Analytics

Advanced data analytics tools often incorporate predictive analytics techniques. By leveraging historical data, these tools can forecast future trends, enabling organizations to anticipate potential challenges or opportunities. Predictive analytics enhances decision-making by providing a forward-looking perspective based on data-driven models.

* Data Visualization

Effective communication of insights is crucial for decision-making. Data analytics tools facilitate the creation of visually compelling dashboards and reports. These visualizations simplify complex data, making it easier for stakeholders to grasp key information at a glance, fostering more effective and informed decision-making.

* Decision Support Systems

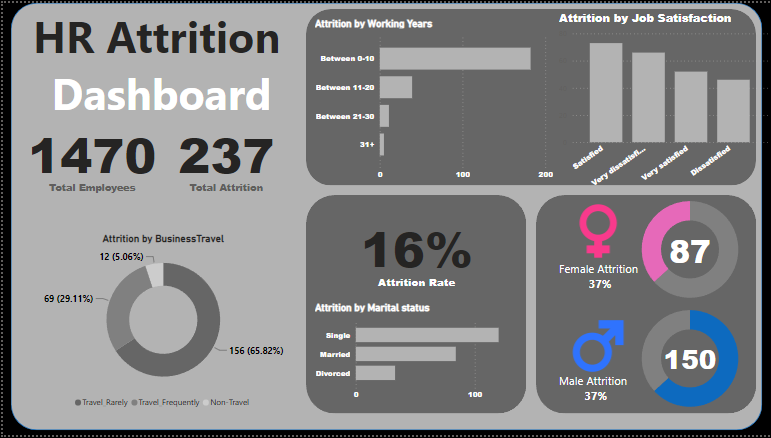
Data analytics tools, when integrated into decision support systems, provide a structured framework for decision-making processes. These systems leverage algorithms and business rules to guide decision-makers, ensuring that choices align with organizational objectives and are based on a thorough analysis of relevant data.

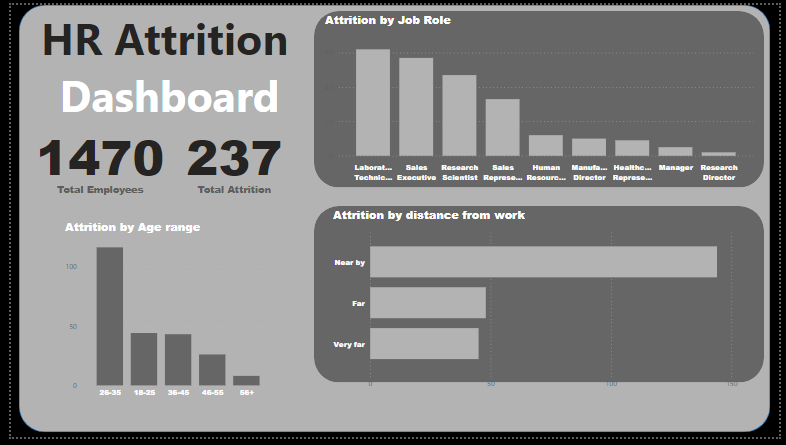
* 1. Select appropriate data analytics tool.

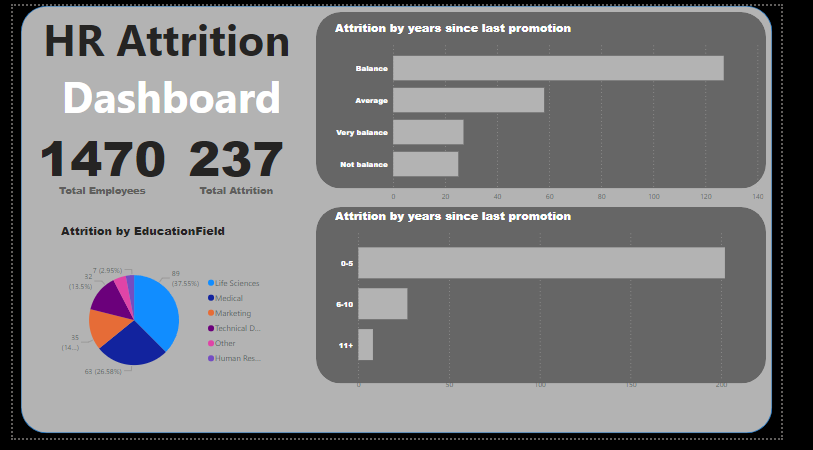
Given the project scenario and the need to create a data analytics dashboard for HelloGreen, Power BI Desktop is a highly suitable choice for several reasons:

* User-Friendly Interface: Power BI Desktop offers an intuitive and user-friendly interface, making it accessible to users with varying levels of technical expertise. This aligns with the diverse audience, including Mr. Lim and other stakeholders, who may not have extensive technical backgrounds.
* Integration Capabilities: Power BI Desktop seamlessly integrates with a variety of data sources, including databases, Excel files, and online services. This flexibility allows for easy integration of HelloGreen's employee data, which may be stored in different formats.
* Robust Data Visualization: Power BI Desktop excels in creating visually appealing and interactive dashboards. The diverse range of visualization options, including charts, graphs, and maps, will enable the representation of attrition-related insights in a compelling manner.
* Real-Time Updates: Power BI Desktop supports real-time data connectivity, which is crucial for monitoring and addressing employee attrition in near real-time. This aligns with the objective of providing timely insights to support decision-making.
* Scalability: Power BI is scalable and can handle large datasets, ensuring that HelloGreen's evolving data needs can be accommodated as the project progresses.

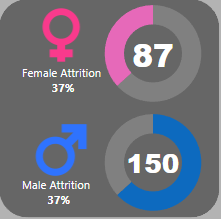
1. Dashboard Development
   1. Create an HR attrition dashboard.

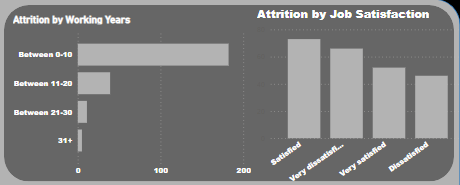


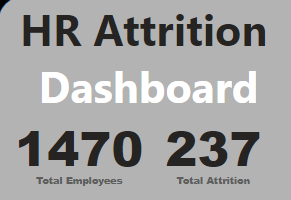




* 1. Customize the design including key performance indicators (KPI) and visualizations and user-friendly features.



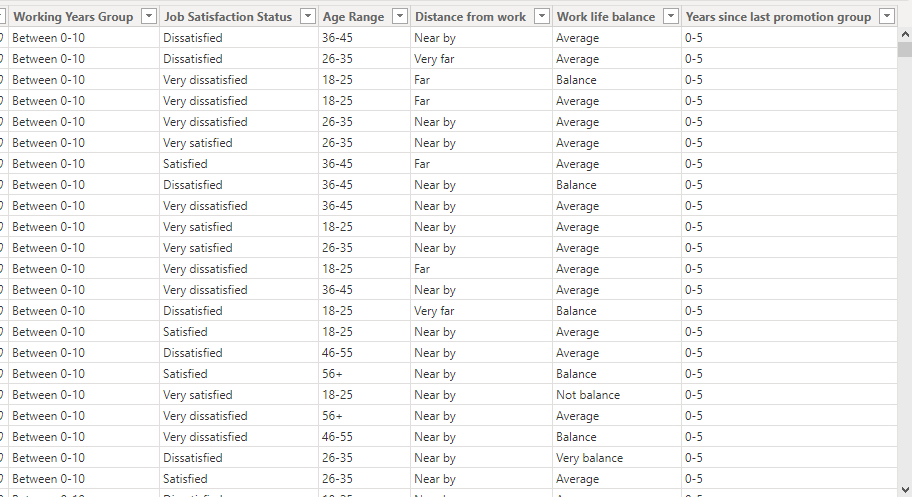




1. Data Analytics
   1. Apply exploratory data analytics techniques to uncover patterns, trends, and actionable insights.

* Data Cleaning and Preprocessing

it's essential to clean and preprocess the data. This involves handling missing values, removing duplicates, and transforming data into a suitable format for analysis.

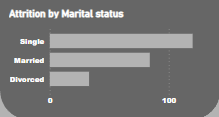
I have added these additional data in the table to better visualize the problem at hand  


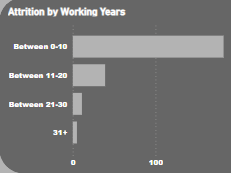
* Data visualization

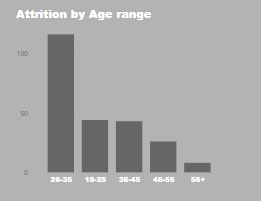
Utilizing data visualization techniques is a powerful way to uncover patterns and trends.  
  
many people that is leaving the company is very dissatisfied with their job



A lot of the people that is leaving the company is young, single people with less than 10 years working experience

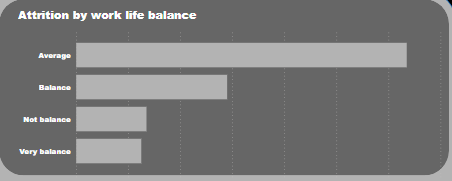






* Correlation Analysis

Examining correlations between different variables can provide insights into relationships. For instance, a lot of the attrition can be seen trough the chart of work life balance, people are having an average work life balance, not good work life balance, this could be a contributing factors of the attrition



we can reduce the attrition by making sure the employee have a good work life balance and reduce the stressful work environment

1. Evaluation
   1. Critical review of the design

* Visual appeal

The visual is very appealing, the chosen color schemes, fonts, and layout contribute to a user-friendly and engaging experience.

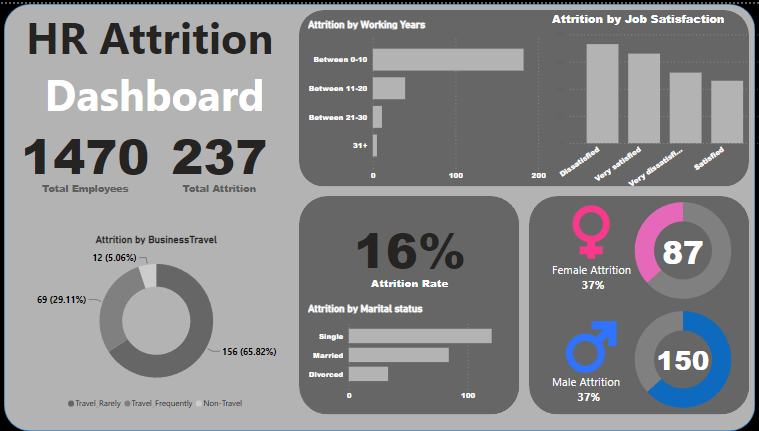
* Information clarity

The visualization is very clear and effectively convey key insights without causing confusion. users can interact with the dashboard seamlessly to explore data and gain insights. chosen metrics and visualizations align with HelloGreen's business objectives related to employee attrition.

* Navigation.

The dashboard supports decision-making within the organization.The navigation is very clear and concise can easily move between different sections and access relevant information without complications.

* 1. Identify what customization has been integrated into the design
* Data Filters: the inclusion of data filters that enable users to focus on specific criteria, such as department, tenure, or performance ratings.



Accessibility Customizations: The dashboard is very accessible. The visualization is very clear and effectively convey key insights without causing confusion. users can interact with the dashboard seamlessly to explore data and gain insights. chosen metrics and visualizations align with HelloGreen's business objectives related to employee attrition.  
  
  
By critically reviewing the design and identifying customizations, we aim to ensure that the data analytics dashboard not only meets the baseline requirements but also aligns with user expectations, enhances usability, and contributes effectively to HelloGreen's decision-making processes.