High Level Design (HLD)

House Price Data Analysis

Revision Number: 1.0
Last date of revision: 29/07/2023

Subhajit Ghosh

Document Version Control

Date Issued	Version	Description	Author
29th July 2021	1.0	First Version of Complete HLD	Subhajit Ghosh

Document Version Control

Contents

Abstract

This High-Level Design Document presents a comprehensive overview of the approach, methodology, and key components of the HR attrition project. The objective of this project is to conduct an in-depth analysis of HR attrition data to unveil valuable insights that can aid HR managers in understanding and effectively mitigating attrition rates, enhancing employee satisfaction, and retaining valuable talent within the organization.

With attrition being a significant challenge faced by many corporate environments, the need for data-driven HR analytics becomes imperative in formulating strategic measures that contribute to a well-functioning and economically viable company. This document serves as a comprehensive guide for all stakeholders involved in the project, providing a clear roadmap towards achieving its goals and ensuring alignment with organizational objectives.

1 Introduction

1.1 Why this High-Level Design Document?

The creation of this High-Level Design Document is of paramount significance as it serves as the foundational cornerstone that underpins the entire HR attrition project. In a world where data-driven decision-making has emerged as a driving force behind organizational success, this document takes center stage in imparting structure, clarity, and coherence to the project's pursuit of tackling attrition challenges effectively.

1. Clarity of Purpose: This document defines the project's specific goals, avoiding ambiguity and misalignment among stakeholders. It fosters enthusiasm and commitment among team members by showcasing the value their efforts bring to the organization.

2. Alignment with Organizational Goals: The project strategically aligns with the organization's broader objectives of talent management, employee engagement, and productivity enhancement.

3. Stakeholder Communication: The document facilitates effective communication by conveying the project's objectives, scope, and methodologies to relevant stakeholders.

4. Risk Mitigation: Addressing potential challenges and risks proactively enables the development of mitigation strategies, ensuring smooth progress.

5. Resource Allocation: The document optimizes resource allocation by providing a clear blueprint of the project's timeline, deliverables, and milestones.

6. Project Governance: It establishes project governance by defining roles, responsibilities, and reporting structures, promoting accountability and transparency.

7. Scalability and Replicability: The document embraces scalable and replicable methodologies to tackle future HR challenges and promote sustainable data-driven decision-making.

1.2 Scope

The scope of the HR attrition project is extensive and encompasses a multifaceted exploration of various aspects related to employee attrition within the organization. It delves into the comprehensive analysis of historical HR data to gain deep insights into attrition trends, patterns, and underlying factors affecting employee turnover. By exploring a wide range of data points, the project aims to provide a holistic view of attrition that will empower HR managers and decision-makers to devise tailored strategies for talent retention and organizational growth.

Specifically, the scope of the project includes, but is not limited to:

a) Attrition Rates Across Departments: The project seeks to identify and compare attrition rates across different departments within the organization. This analysis allows for the detection of patterns and potential disparities in attrition rates among departments, shedding light on departments that may require more focused attention in terms of employee retention efforts.

b) Employee Demographics: The project aims to explore employee demographics, such as age groups, gender, tenure, and educational background, to discern whether certain demographic segments are more susceptible to attrition. This insight can assist in tailoring retention strategies to address the unique needs and preferences of different employee groups.

c) Key Factors Influencing Attrition: A pivotal aspect of the scope involves identifying the critical factors contributing to attrition. These factors may include job satisfaction levels, work-life balance, compensation and benefits, career growth opportunities, and managerial support. By understanding the primary drivers of attrition, the project endeavors to equip HR managers with actionable information to create a more conducive and fulfilling work environment.

d) Employee Satisfaction Levels: The project will assess employee satisfaction levels through survey data or other feedback mechanisms. Understanding employee satisfaction is essential in gauging overall engagement and motivation, which significantly impacts attrition rates. Consequently, HR managers can target areas of improvement to enhance overall employee experience and loyalty.

e) Historical Attrition Trends: Analyzing historical attrition trends over specific periods allows for the identification of long-term patterns and seasonal fluctuations. This information can be valuable in predicting attrition trends in the future and designing preemptive measures to mitigate potential attrition spikes.

f) Data Sources and Timelines: The project scope defines the data sources to be utilized, which may include HR databases, exit interviews, employee surveys, performance records, and other relevant sources. Additionally, the document outlines the timelines and intervals for data collection and analysis, ensuring a well-organized and efficient execution of the project.

g) Ethical Considerations: The scope also addresses ethical considerations related to data privacy and confidentiality. Ensuring compliance with data protection regulations and safeguarding the anonymity of individual employees is a critical aspect of the project's ethical framework.

2 General Description

2.1 Product Perspective & Problem Statement

Human Resources (HR) is a vast and dynamic department responsible for finding, selecting, onboarding, and monitoring talent. To streamline workforce monitoring, analytics concepts play a key role. Job classifications, which consider job families and pay grades, are crucial, especially when introducing new roles into the existing job structure.

Jobs have distinct features like education level, experience, impact, supervision, and budget responsibilities. Analyzing these factors enables job analysts to group jobs based on pay scales and benefits. Linear Discriminant Analysis (LDA) is suggested by Sundmark for characterizing different job classes based on their features.

By applying LDA to job classification data, HR can efficiently classify newly created jobs within the existing structure, providing valuable guidelines for organizational functions. This data-driven approach empowers HR to make informed decisions and optimize talent management practices. The seamless monitoring of the workforce ensures the right talent is retained, contributing to the overall success of the organization.

2.2 Tools used

Here, the tools and technologies employed in the project are listed. This includes data analytics and visualization tools such as Power BI, Python, or any other relevant software used for data processing and generating meaningful visualizations.
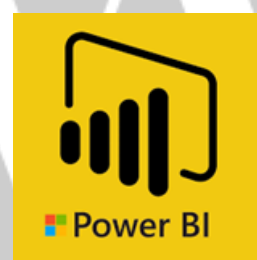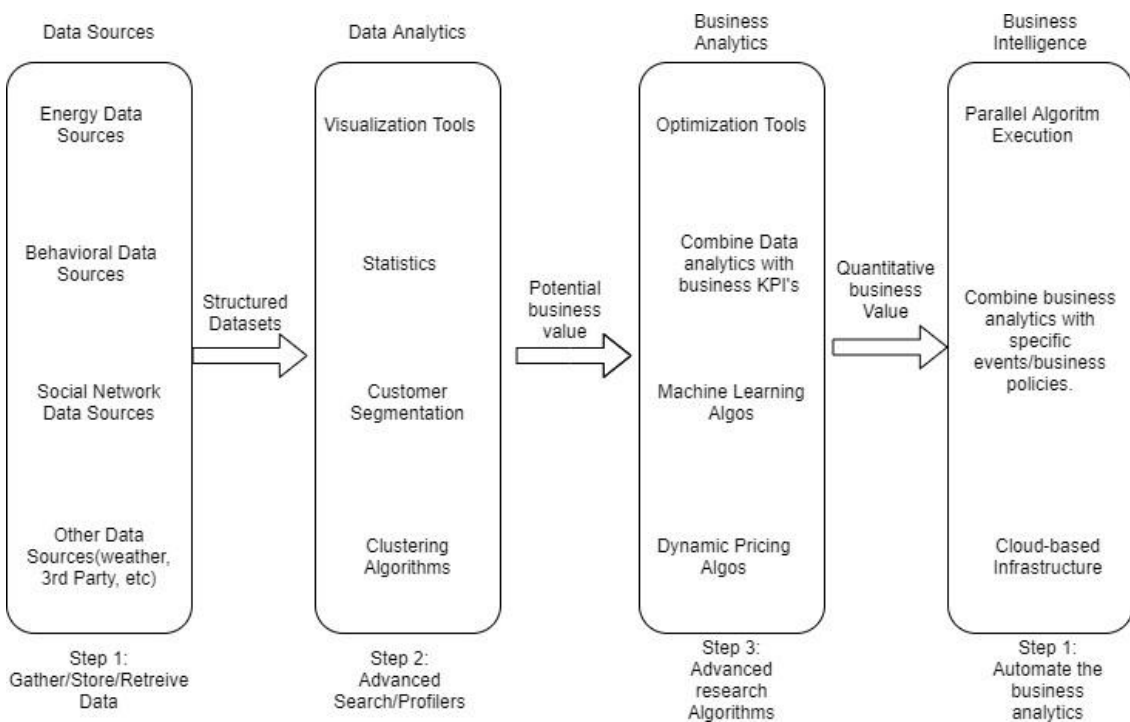
3 Design Details
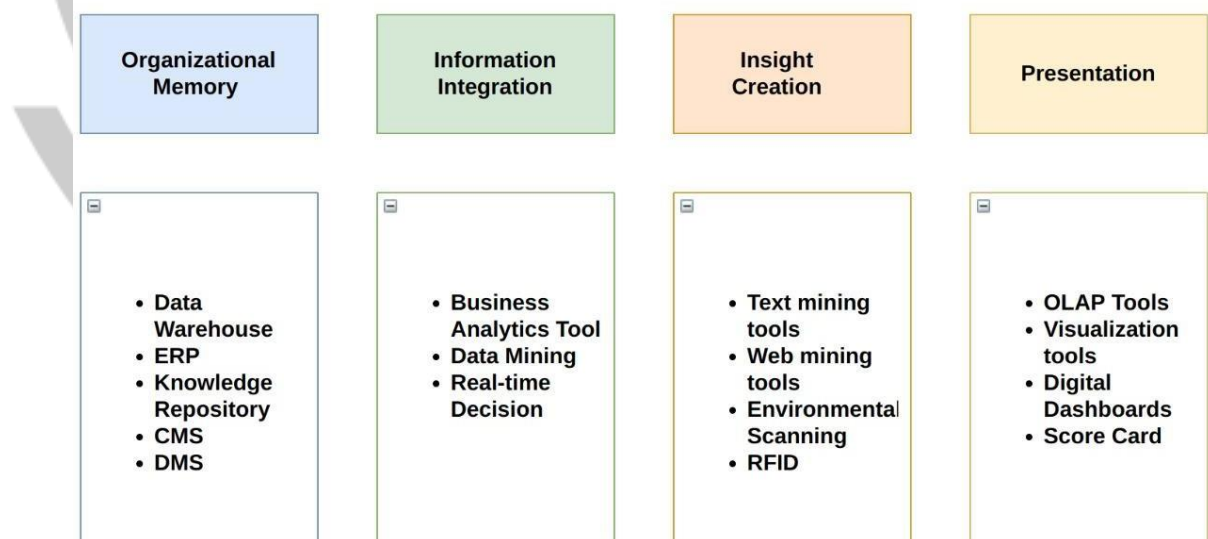
3.1 Functional Architecture

Figure 1: Functional Architecture of Business Intelligence

3.2 Optimization

Your data strategy drives performance

- Minimize the number of fields
- Minimize the number of records
- Optimize extracts to speed up future queries by materializing calculations, removing columns and the use of accelerated views

Reduce the marks (data points) in your view

- Practice guided analytics. There's no need to fit everything you plan to show in a single view. Compile related views and connect them with action filters to travel from overview to highly-granular views at the speed of thought.
- Remove unneeded dimensions from the detail shelf.
- Explore. Try displaying your data in different types of views.

Limit your filters by number and type

- Reduce the number of filters in use. Excessive filters on a view will create a more complex query, which takes longer to return results. Double-check your filters and remove any that aren't necessary.

- Use an include filter. Exclude filters load the entire domain of a dimension, while include filters do not. An include filter runs much faster than an exclude filter, especially for dimensions with many members.

- Use a continuous date filter. Continuous date filters (relative and range-of-date filters) can take advantage of the indexing properties in your database and are faster than discrete date filters.

- Use Boolean or numeric filters. Computers process integers and Booleans (t/f) much faster than strings.

- Use parameters and action filters. These reduce the query load (and work across data sources).

Optimize and materialize your calculations

- Perform calculations in the database
- Reduce the number of nested calculations.

- Reduce the granularity of LOD or table calculations in the view. The more granular the calculation, the longer it takes.
  - LODs - Look at the number of unique dimension members in the calculation.
  - Table Calculations - the more marks in the view, the longer it will take to calculate.

- Where possible, use MIN or MAX instead of AVG. AVG requires more processing than MIN or MAX. Often rows will be duplicated and display the same result with MIN, MAX, or AVG.

- Make groups with calculations. Like include filters, calculated groups load only named members of the domain, whereas Tableau's group function loads the entire domain.

- <u>Use Booleans or numeric calculations instead of string calculations</u>. Computers can process integers and Booleans (t/f) much faster than strings. Boolean>Int>Float>Date>DateTime>String

4 KPIs

Dashboards will be implemented to display and indicate certain KPIs and relevant indicators for the disease.

As and when, the system starts to capture the historical/periodic data for a user, the dashboards will be included to display charts over time with progress on various indicators or factors

4.1 KPIs (Key Performance Indicators)

The key indicators displaying a summary of the HR attrition dataset and its relationship with different metrics are as follows:

1. Tiles for Quick Analysis: The first set of tiles provide a quick overview of the total number of employees in different categories. The first tile represents the total number of employees in the company. The second tile represents the total number of employees classified by family, such as Research & Development, Sales, and HR. The third tile represents the distribution of employees based on their education level.
2. Column and Bar Chart Visualization: The column chart visualizes the relationship between PG (Performance Grade) and contact level. This chart showcases the distribution of performance grades across various contact levels, helping identify any patterns or correlations.
3. The bar chart displays the relationship between job family and job description. It illustrates the distribution of job descriptions within each job family, providing insights into the diversity of roles within different departments.
4. Chart Visuals: The first chart represents the relationship between PG (Performance Grade) and the total number of job families. This visualization offers an understanding of how performance grades are distributed across different job families, shedding light on areas where performance may be particularly strong or weak.
5. The second chart showcases the relationship between PG and the total number of jobs. By analyzing this chart, HR managers can gain insights into the performance distribution among different job roles and identify any trends or discrepancies.
6. Education and Grade-Based Analysis Visuals: The first set of tiles presents an analysis based on employee ID and their education level. This visualization helps in understanding the educational background of employees and its potential impact on attrition.

7. The second set of tiles showcases the top 10 pay grades and their respective family descriptions. By examining this visualization, HR managers can identify the highest-paying job families and the corresponding descriptions, aiding in understanding the relationship between pay grades and job roles.
8. These key indicators and visualizat

5 Deployment

Deployment plays a crucial role in the HR attrition project, as it involves effectively sharing the insights and findings with relevant stakeholders. Leveraging various tools like Jupyter Lab, Power BI, and Novypro ensures a seamless and impactful dissemination of the project's outcomes.

1. Jupyter Lab: Jupyter Lab serves as a powerful platform for conducting data analysis, exploration, and model building. After analyzing the HR attrition dataset using Python or other data analysis libraries in Jupyter Lab, the team can create interactive notebooks showcasing the step-by-step analysis and methodologies employed. These notebooks can include data visualizations, statistical analyses, and machine learning models utilized to derive insights. The advantage of using Jupyter Lab is its flexibility, allowing stakeholders to review and reproduce the analysis if required, fostering transparency and credibility.

2. Power BI: Power BI is an invaluable tool for data visualization and reporting. After completing the analysis in Jupyter Lab, the team can integrate the results into a dynamic and interactive dashboard using Power BI. The dashboard can include key metrics, KPIs, and graphical representations of the HR attrition insights. Power BI allows stakeholders to explore the data in real-time, filter information based on specific criteria, and gain a comprehensive understanding of the findings. This visually engaging and user-friendly dashboard is a powerful medium to communicate complex data insights effectively.

3. Novypro for Report Sharing: Novypro offers a secure and efficient platform for sharing reports and insights with various stakeholders. The team can upload the Jupyter Lab notebooks, Power BI dashboards, and any other relevant project documentation to Novypro. With access controls, the team can manage who can view, comment, or edit the shared materials, ensuring data privacy and confidentiality. Stakeholders, such as HR managers, executives, and decision-makers, can access the reports from any location and device, facilitating seamless collaboration and knowledge dissemination.