High Level Design
HR Analytics – Absenteeism

Revision No.: 1.2

Last Date of Revision: 24/01/2024

Submitted By: -
Aditya Sahoo

Document Control

Date	Version	Description	Author
18/01/2024	1.0	Introduction, General Description	Aditya Sahoo
20/01/2024	1.1	Architecture, KPI	Aditya Sahoo
24/01/2024	1.2	Final Revision	Aditya Sahoo

Table Of Contents

1. Introduction

1.1 Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions before coding and can be used as a reference manual for how the modules interact at a high level.

1.2 Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly technical terms which should be understandable to the administrators of the system

2. General Description

2.1 Product Perspective & Problem Statement

HR is not just about hiring people it is an ocean of its own. The HR department goes through a constant journey of finding, selecting, onboarding and monitoring the right talent. You are required to use the analytics concept to provide smooth monitoring of the workforce for the HR department. This data set is suitable for identifying pockets of absence in the organization.

These pockets may require intervention. 'Absent Hour' will be used as a dependent variable. In addition, age and length of service may also be associated with absence – but how? This is for you to find out. The data set can also be used as an exercise set to predict absence using decision trees or linear models.

This data set is quite straightforward. It is large but still manageable in software like SPSS or Excel. You may have to code a number of nominal variables into number values before you can do your analysis but on top of that, the data itself doesn't pose much of a challenge. Note: The data does need to be cleaned. Everyone under 18 or above 65 may be removed from the data set. This HR data set focuses on absence at work. The data set contains 740 rows and 21 columns of data.

The data set contains a number of employee IDs. Each row represents a certain quantity of absence – meaning that one employee can have multiple rows. Information on employees includes the number of children, workload, distance from work, transportation expense, education, height, weight, BMI, and absenteeism time in hours. Other information includes the season, month of absence, day of absence, and day of the week. This data set can help you find predictors of absence.

Potential analyses could be to see if there is an association between BMI and absence, as well as season, workload, distance from work and the other factors in the data set. The challenge of this data set is mostly in structuring the data. An individual employee has multiple records. These need to be combined prior to analysis. This data set also enables you to do longitudinal research.

2.2 Tool Used

Microsoft Excel is a powerful spreadsheet program developed by Microsoft as part of the Microsoft Office suite. It is widely used for various purposes, including data analysis, calculation, charting, and financial modeling. Excel provides a grid interface where users can organize, manipulate, and analyze data in tabular form.

Excel is widely used in business, finance, education, and various other fields due to its versatility and user-friendly interface. Its capabilities make it a valuable tool for tasks ranging from simple data entry to complex financial modeling and analysis.

3. Architecture

Microsoft Excel is a spreadsheet program that is part of the Microsoft Office suite of productivity software. It's widely used for tasks such as data analysis, visualization, and tabular calculations. The architecture of Microsoft Excel involves several key components and functionalities:

1. User Interface (UI):

- Ribbon: The ribbon interface is a prominent feature, providing a set of tabs, each containing related commands and functions. Users can navigate through the ribbon to access various tools for formatting, data manipulation, and more.
- Cells and Worksheets: The basic unit of data in Excel is a cell, arranged in a grid on worksheets. Worksheets are organized within workbooks, allowing users to manage and analyze data in a structured manner.

2. Calculation Engine:

- Formulas and Functions: Excel has a powerful calculation engine that supports a wide range of formulas and functions. Users can create custom formulas or use built-in functions for mathematical calculations, statistical analysis, and more.
- Recalculation: Excel automatically recalculates formulas and updates the results when changes are made to the data. This ensures that users always have up-to-date calculations.

3. Data Storage:

- Workbooks: Workbooks are the files that contain Excel data. Each workbook can have multiple worksheets, and users can create, save, and open workbooks to organize their data.
- File Format: Excel workbooks are typically saved in proprietary file formats such as .xlsx or .xlsb. These formats preserve various elements, including formatting, formulas, and charts.

4. Data Analysis and Visualization:

- Charts and Graphs: Excel provides a variety of chart types for visualizing data. Users can create charts to represent trends, patterns, and comparisons within their datasets.
- PivotTables and Pivot Charts: PivotTables allow users to summarize and analyze large datasets interactively. Pivot Charts provide visual representations of PivotTable data.

5. Integration and Compatibility:

- Integration with Other Office Applications: Excel seamlessly integrates with other Microsoft Office applications, allowing users to embed Excel tables or charts in Word documents, PowerPoint presentations, etc.
- Data Import and Export: Excel supports importing data from various sources, such as databases, text files, and online sources. Users can also export Excel data to different formats.

6. Macro and VBA (Visual Basic for Applications):

- VBA Programming: Excel includes a powerful scripting language called VBA. Users can create macros using VBA to automate repetitive tasks and extend Excel's functionality.
- Macro Recorder: The Macro Recorder allows users to record a series of actions in Excel and then play them back as a macro. This is a convenient way to automate tasks without writing VBA code.

7. Security and Collaboration:

- Worksheet Protection: Excel allows users to protect worksheets and workbooks with passwords, controlling access to sensitive data.
- Collaboration Features: Excel supports real-time collaboration, allowing multiple users to work on the same workbook simultaneously. Changes made by one user are updated for others in real-time.

Understanding the architecture of Microsoft Excel helps users leverage its features efficiently for data management, analysis, and reporting. The combination of a user-friendly interface and powerful calculation capabilities makes Excel a versatile tool for various tasks.

4. KPI

- Total Absent Hours by Gender
- Average Absent Hours by Gender
- Department wise average absent hours
- Division wise average length of service
- Division wise average absent hours
- Average absent hours by age group
- Total absent hours by age group
- Average absent hours by business unit
- Average absent hours by length of service
- Top 6 job titles by total absent hours
- Top 5 cities by total absent hours