Project Report

Qualification Name	Professional Diploma in Data Science
Module Name	WSQ -Data Modelling and Visualization (SF)

Student name		Assessor name	
Koh Chun Yao			
Date issued	**Completion date**		**Submitted on**

Project title	Project Milestone 1 and 2

Learner declaration
I certify that the work submitted for this assignment is my own and research sources are fully acknowledged.

Student signature: Koh Chun Yao   Date: 25 Feb 2026

Index

Document Version History

Version Number	Effective Date of release	Details	Author
1.0	24 June 2023	Initial Creation	Kishan
2.0	15 November 2024	Format Change	Ismayil Siyad

1. Project Background

This project focuses on applying data analytics skills to a real-world HR dataset covering employee attrition and performance from 2018 to 2019.

The dataset comprises multiple related tables including employee demographics, education levels, job involvement, performance ratings, satisfaction scores, and work-life balance indicators.

Using Microsoft Power BI as the primary tool, the project involves end-to-end analytics work — from importing and cleaning raw data, to building a structured data model, and finally designing interactive dashboards that surface meaningful HR insights.

The project is structured around five key activities:

- data transformation,
- data modelling and
- three rounds of visualization covering an Overview page, a Demographics page, and a Performance Tracker with Attrition analysis.

Throughout, DAX formulas were used to create <u>calculated measures</u> such as attrition rate, active employee count, and total headcount. The final deliverable is a multi-page Power BI report that enables HR decision-makers to explore employee trends, identify attrition risk factors, and track individual performance — ultimately supporting more informed and data-driven workforce strategies.

2. Project Objective

1. To import, inspect, and clean HR data from multiple source tables, ensuring data quality and consistency before analysis.

2. To build a well-structured Power BI data model using a star schema, establishing correct relationships and hierarchies between fact and dimension tables.

3. To develop DAX measures that calculate key HR metrics such as total employees, active/inactive headcount, and attrition rate, enabling dynamic and accurate reporting.

4. To design interactive and visually informative dashboards across multiple report pages (Overview, Demographics, Performance Tracker, and Attrition) that allow HR stakeholders to explore workforce trends at a glance.

5. To derive actionable insights from the data visualizations, such as identifying attrition patterns by department, tenure, travel frequency, and job role, to support evidence-based HR decision-making

3. Project Specifications

Technical Tools Used:

• **Microsoft Power BI Desktop** — Primary tool for data import, transformation, modelling, and visualization. Used to build all report pages and interactive dashboards.

• **Power Query** (M Language) — Used within Power BI for data cleansing tasks, such as removing nulls, correcting data types, and applying transformations like creating the AgeBin column.

• **DAX** (Data Analysis Expressions) — Used to create calculated measures (e.g., %Attrition, ActiveEmployees, InactiveEmployees, TotalEmployees) stored in a dedicated Measure table.

• Microsoft Excel — Source format for the raw HR dataset, containing multiple sheets for each data table (e.g., Education, Job Involvement, Performance Rating, etc.).

First, all seven source tables were loaded into Power BI and inspected for **data quality** issues. Power Query was used to fix incorrect data types and handle missing values. **Calculated columns (e.g., AgeBin) and DAX measures** were then created.

Next, **table relationships** were established in the Model View to form a star schema, with the Employee Data table as the <u>central</u> fact table.

Hierarchies were set up for **drill-down** functionality. **Visualizations** were built incrementally across four report pages — starting with high-level KPIs on the Overview page, then demographic breakdowns, followed by **per-employee** performance tracking with slicers, and finally attrition analysis charts.
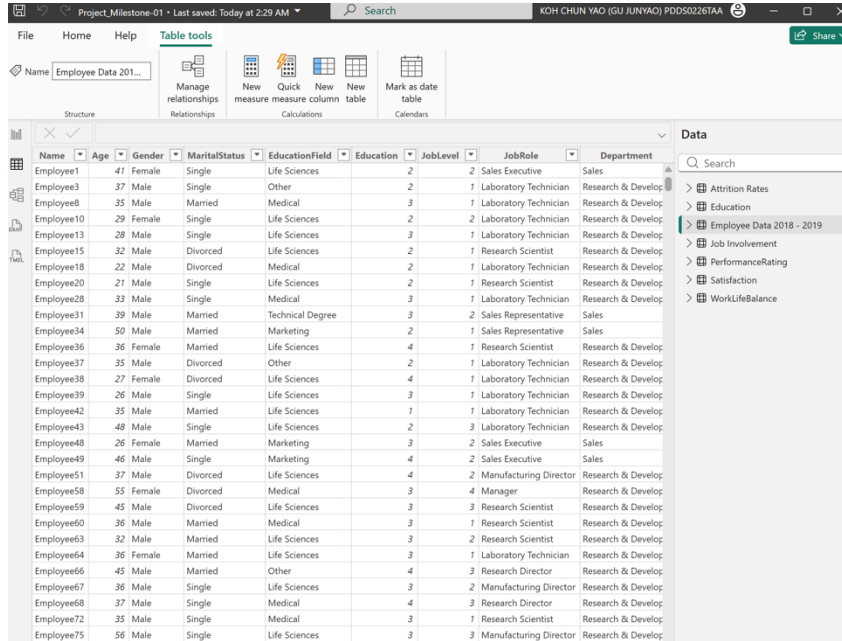
Each page was designed with clarity and usability in mind, using appropriate chart types for the data being presented.

4. Project Tasks

4A. Activity 1 :Data Transformation

Please include screenshots 📷 and explanations in the provided space below.

1. Screenshots of Tables

2. Data Cleaning

Each imported table was inspected in Power Query for data quality issues. Key checks performed:

• Verified all column data types were correctly assigned (e.g., numeric IDs, text labels).

• Ensured the "date" column in the Attrition Rates table was set to DateTime format.

• Checked for missing or null values across all tables and removed or flagged affected rows. • Confirmed no duplicate records existed in the Employee Data table. • Validated that lookup values (e.g., Education level codes, Satisfaction IDs) matched across related tables before establishing relationships.

3. Data Transformations Applied
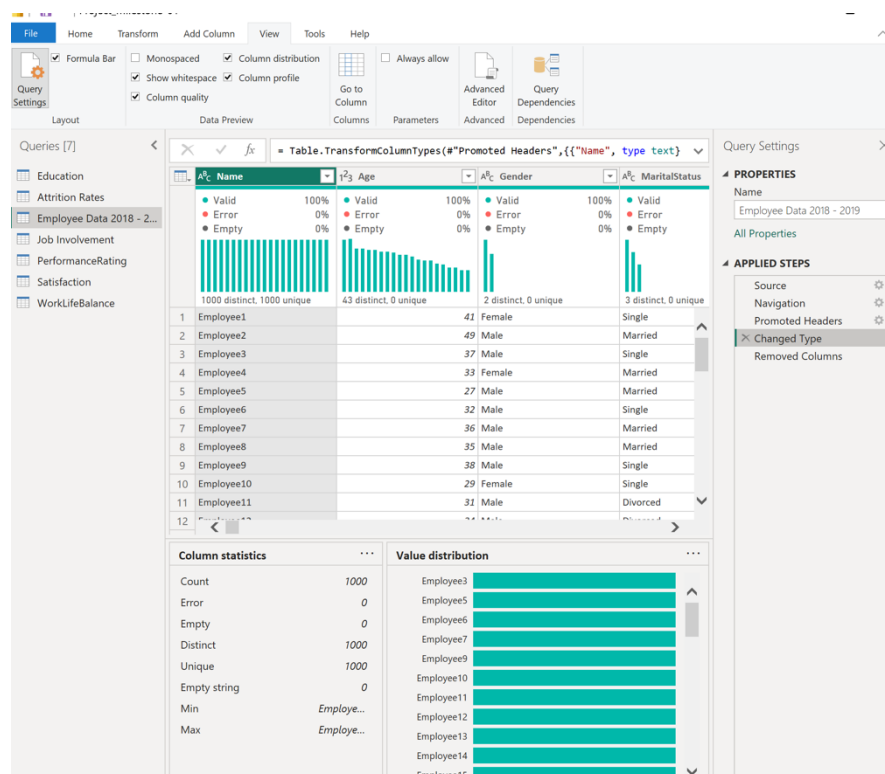
The following transformations were applied in Power Query and the Power BI data model: • AgeBin column (Employee Data 2018–2019): A calculated column was created to group employees into age brackets (e.g., Under 25, 25–34, 35–44, 45–54, 55+). This enables age-based demographic analysis in charts.

• Date format fix (Attrition Rates): The "date" column was converted from a text or numeric format to a proper DateTime type, allowing it to be used in time-based bar charts and trend lines. • Column renaming and type enforcement were applied

across tables to ensure consistency and compatibility with the data model relationships.

Created Columns and Measures

A dedicated "**Measure**" table was created in Power BI to house all DAX measures. The following measures and calculated columns were defined: **Calculated Column** — AgeBin: Categorises employee age into bins using a DAX SWITCH/IF formula (e.g., "Under 25", "25–34", "35–44", "45–54", "55+").

DAX Measures: • TotalEmployees = COUNTROWS('Employee Data 2018 - 2019') • ActiveEmployees = CALCULATE([TotalEmployees], 'Employee Data 2018 - 2019'[Attrition] = "No") • InactiveEmployees = CALCULATE([TotalEmployees], 'Employee Data 2018 - 2019'[Attrition] = "Yes")

• %Attrition = DIVIDE([InactiveEmployees], [TotalEmployees], 0) formatted as a percentage. These measures power the KPI cards and allow dynamic filtering across all report pages.

4B Activity 2 : Data Modeling
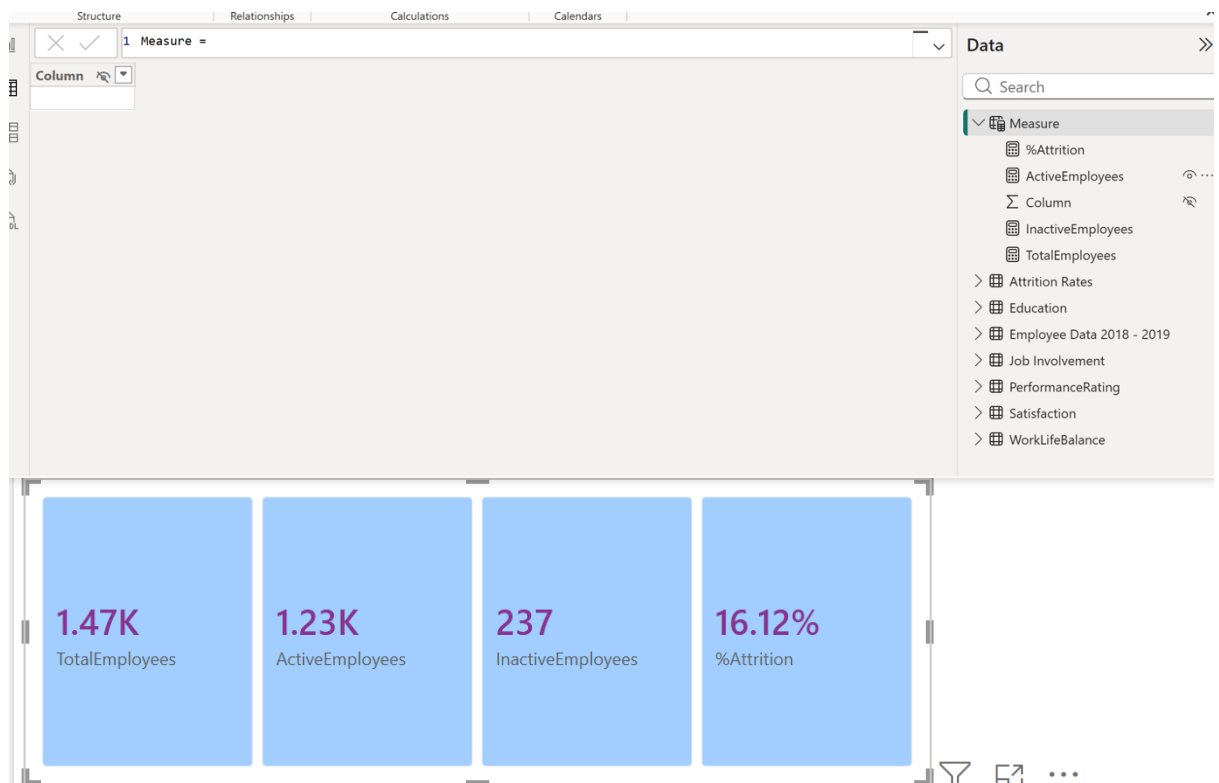
Please include screenshots 📷 and explanations in the provided space below.

1. Data Model

A star schema data model was built in Power BI's Model View. The "Employee Data 2018–2019" table serves as the central fact table, linked to six dimension tables: Education, Job Involvement, Performance Rating, Satisfaction, WorkLifeBalance, and Attrition Rates. All relationships are many-to-one (dimension to fact), with single-direction cross-filtering applied to maintain clean and predictable filter propagation across visuals.
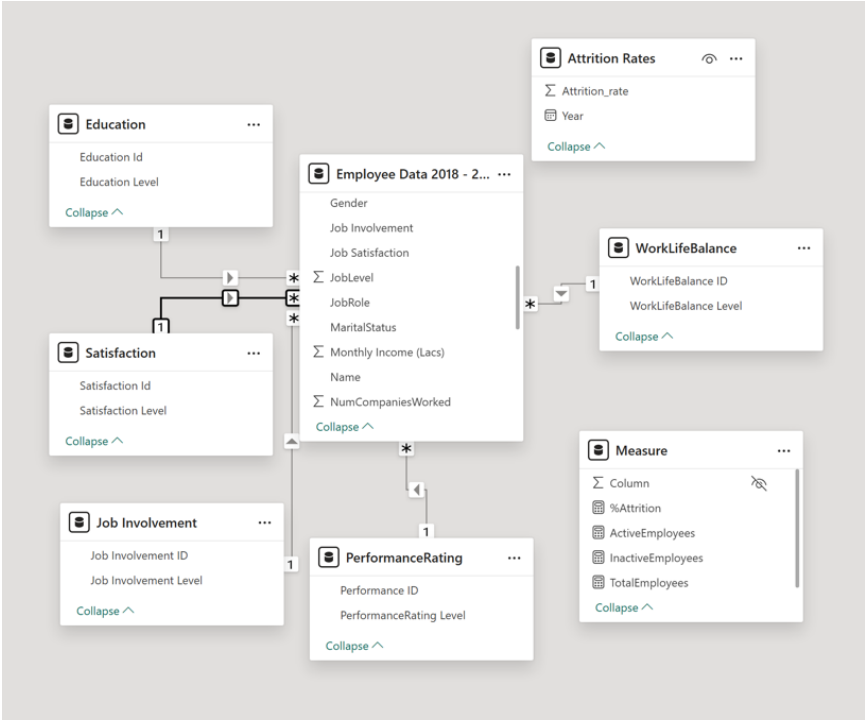
2. Data Relationships

All supporting tables were linked to the Employee Data fact table using many-to-one relationships with single-direction filtering, ensuring a proper star schema structure.
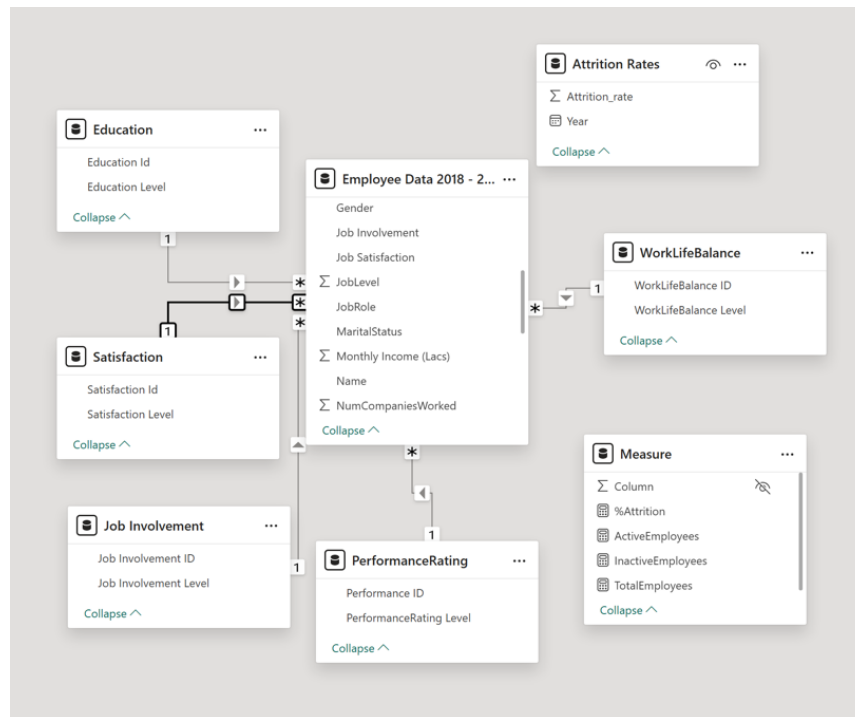
3. Created Hierarchies

A drill-down hierarchy was created within the Employee Data table using "Department" as the top level and "JobRole" as the child level. This hierarchy enables users to interact with visuals (e.g., the stacked bar chart on the Overview page) by drilling from a department-level view down to individual job roles within a department, without needing to change the chart manually.