Executive Summary

Absenteeism Analysis Report

This report analyzes absenteeism data within the organization to identify trends and potential causes. Understanding absenteeism patterns can improve employee well-being, increase productivity, and reduce associated costs.

The report will provide valuable insights into absenteeism patterns within the organization, paving the way for developing targeted strategies to improve employee well-being, boost productivity, and minimize the negative effects of absenteeism.

Introduction

1.1 Background and Context

Employee absenteeism, particularly unplanned absences, can significantly impact an organization's performance. It disrupts workflows, increases the workload on present employees, and can lead to missed deadlines and decreased customer satisfaction. Analyzing absenteeism data allows us to identify patterns and potential underlying causes, ultimately helping us develop strategies to reduce its impact.

1.2 Objectives

This absenteeism analysis aims to:

- Calculate short-term and long-term absenteeism rates.
- Identify departments or employee groups with higher absenteeism.
- Explore potential correlations between absenteeism and factors like demographics or absence reasons.
- Gain insights into potential causes of absenteeism within the organization.

Data and Methodology

The analysis utilizes data from the following tables:

- **Absenteeism_AWK:** Contains employee absence details, including hours absent and absence reasons.
- **Employee_Demographic:** Provides demographic information for each employee.
- **Reason_Absence:** This table contain details about specific reasons for absence (e.g., illness, personal leave).

All these tables were created based on the original data source stratified from:

Martiniano, Andrea and Ferreira, Ricardo. (2018). Absenteeism at work. UCI Machine Learning Repository. https://doi.org/10.24432/C5X882.

2.2 Data Cleaning and Preparation (Optional)

This part of the project was performed using SQL to clean data or prepare necessary steps. For example, I am handling missing values, standardizing data formats, or addressing inconsistencies. Attached is a PDF file describing the SQL process used.

2.3 Analytical Techniques

The analysis will employ the following techniques:

- **Descriptive statistics:** To calculate absenteeism rates (overall, by reason, etc.).
- **Comparative analysis:** To identify groups with higher or lower absenteeism rates.
- **Correlation analysis:** To explore possible relationships between absenteeism and absence reasons or employee demographics. This part will be developed in a future version of the report.

2.4 Analytical Questions

Questions	Analysis type				Suitable visualization type(s)					Chart Type	Data requirements	
	Describe	Diagnose	Predict	Prescribe	Trend	Compare	Pattern	Rank	Proportion		Dimensions	Measures
How does the reason for absence (ICD category) affect absenteeism time?	X					X				Bar Chart	Reason for Absence	Absence duration in hours
Is there a correlation between distance from work and absence duration?		X					X			Scatter Plot	Home to work distance in km	Absence duration in hours
Does age impact absenteeism, considering factors like		X					X		X	Line Chart/Scatter Plot	Age Group	Absence_duration in hours

service time?											
Is there a correlation between body mass index (BMI) and absence duration?		X					X			Scatter Plot	Body Mass Index Category	Absence duration in hours
Is there a correlation between number of children or pets and absence duration?		X					X			Scatter Plot	Children Numer	Absence duration in hours
Is there a correlation between social habits (smoking/drinking) and absence duration? (Future version)		X					X			Scatter Plot	Social_smoker (Yes/No) OR Social_drinker (Yes/No)	Absence duration in hours
Does the day of the week (Day of the week), months or season (Seasons) influence absenteeism patterns?		X			X		X			Bar Chart or Line Chart	Day of the week, months, or season	Average Absence duration
Is there a correlation between transportation expense and absence duration?		X					X			Line Chart/Bar Chart	Month name	Total Absence per month

How is absenteeism distributed across different absence categories? (e.g., Illness, Personal Leave)	X								X	Pie Chart	Absence Description	Count of Absences
What is the monthly trend of absenteeism, which days and seasons occur the most?	X				X							

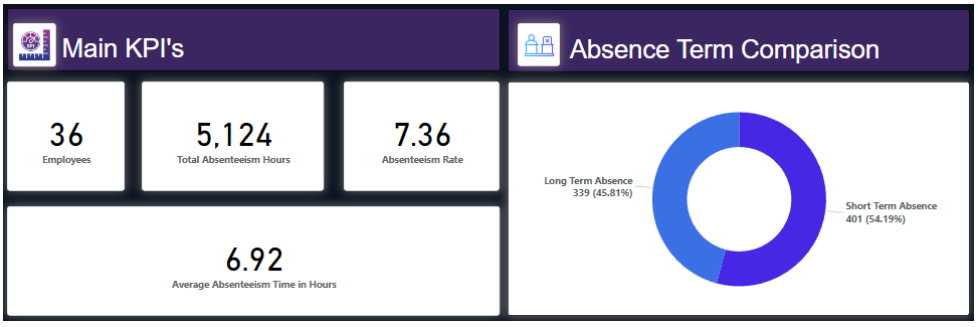
3.1 Absenteeism Analysis

Overall Absenteeism:

The absenteeism rate of 7.36% indicates that employees are absent for an average of over seven days per year (assuming a standard 2,080 working hours annually). This absenteeism rate falls within a moderate range, but further investigation can reveal if there's room for improvement.

Employee Impact:

- The average absenteeism time of 6.92 hours suggests that the absences are relatively short when employees are absent. It could be positive as it minimizes disruption compared to more extended absences.



- **Analysis of Top 10 Absence Reasons:**

This table shows the top 10 most frequent reasons for employee absences, categorized by season (Summer, Fall, Winter, Spring) and total absences throughout the year.

Key Observations:

- **Overall Ranking:**
  - Diseases of the musculoskeletal system and connective tissue are the leading causes of absenteeism yearly (842 absences).
  - Injury, poisoning, and external causes follow closely (729 absences).
  - Medical consultations rank third (424 absences).

- **Seasonal Variations:**

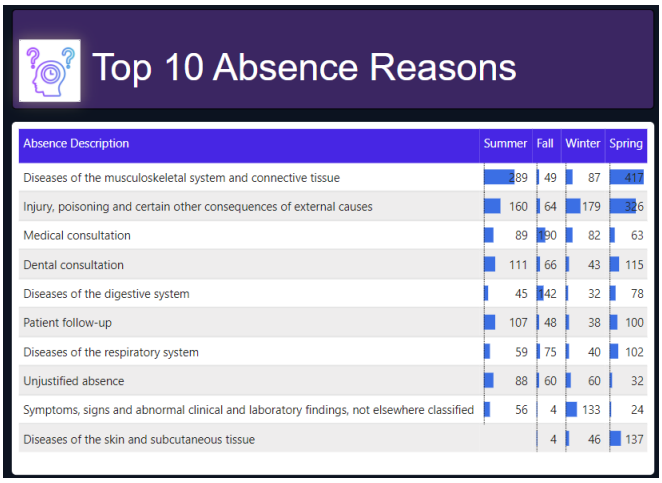
  - While some reasons are prevalent year-round (musculoskeletal injury), others exhibit seasonal variations:

  - **Respiratory System Diseases:** Significantly in Winter with (40) compared to other seasons, suggesting a potential impact of cold and flu season.

  - **Skin Conditions are significantly** lower in Winter (46) and Fall (4) than in Spring (137), possibly due to less sun exposure.

  - **Dental consultations:** Show a slight increase in Summer (111) and Spring (115) compared to other seasons.

- **Unjustified Absences:** Relatively consistent throughout the year, suggesting limited seasonal influence.



➢ Absence Term Comparison:

Short-term absences are slightly more prevalent than long-term absences within your organization. Here's a breakdown:

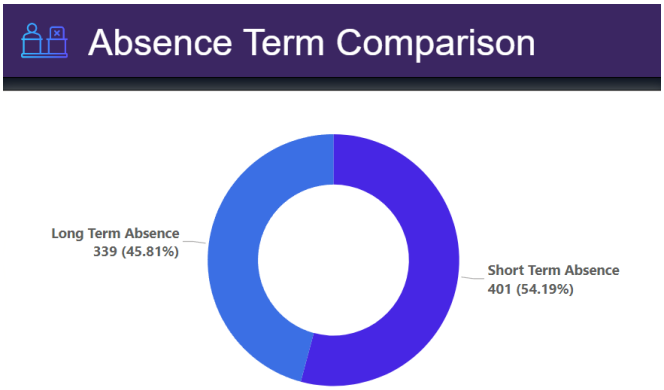
- **Short-Term Absence:** Represents 54.19% (401 absences) of the total absences.
- **Long-Term Absence:** Represents 45.81% (339 absences) of the total absences.

This data suggests that most employees experience shorter absences when they are away from work. However, a significant portion (almost half) still experience extended absences. Here are some additional insights:

- **Threshold Definition:** Short-term absence is less than or equal to 3 hours, while long-term absence is greater than 3 hours.
- **Potential Cost Impact:** While short-term absences seem less disruptive individually, their higher frequency could significantly impact productivity and potentially higher costs associated with covering absent employees.

Further Analysis for Deeper Understanding:

- **Investigate Reasons:** Analyze the reasons behind short-term and long-term absences to identify potential areas for intervention. For example, short-term absences might be due to minor illnesses, while long-term absences could be related to chronic health conditions or personal leave.



➢ **Analysis of Monthly Absence Trends and Season Proportion:**

Overall Observations:

- Absenteeism hours fluctuate throughout the year, with a significant peak in March (765 hours) and July (734) compared to other months.
- Generally, absenteeism seems higher in the first half of the year (January-June) compared to the latter half (July-December).

<u>Potential Explanations for Variations:</u>

- **Seasonal Influences:** Higher absenteeism in the first half might correlate with cold and flu season (though further analysis is needed to confirm this).
- **School Holidays:** If the organization has many employees with school-aged children, absenteeism might increase during school breaks (e.g., March break) due to childcare needs.
- **Year-End Trends:** The lower absenteeism in the latter half could be due to employees using up their allowance earlier or being less likely to take time off around holidays and year-end deadlines.

Further Considerations:

- **Year-over-Year Comparison:** If data from previous years is available, comparing monthly trends across those years can reveal if the observed pattern is consistent or unique to the current year.
- **Reason Analysis:** Investigate the most frequent reasons for absence each month to see if they correlate with the observed trends (e.g., high illness rates in March).

Regarding with Season Proportion:

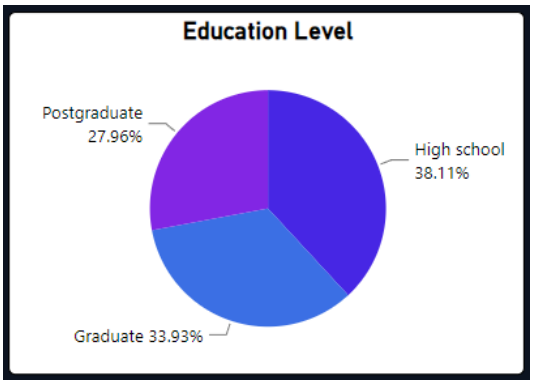
- **Summer:** Contributes 27.97% to the total absences.
- **Fall:** Follows closely with 21.74% of absences.
- **Winter:** Has the lowest proportion of absences at 18.15%.
- **Spring:** Represents the highest proportion of absences at 32.14%.

Key Takeaways:

Absenteeism is significantly higher in spring compared to other seasons. It might be holiday-related disruptions or other factors.

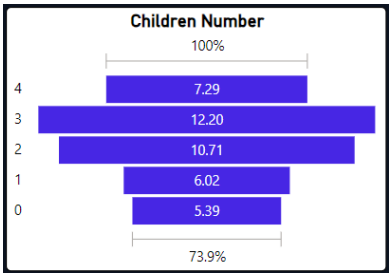
➢ Average Absence & Employee Characteristics

1) **Education Level:** Employees with a high school diploma have the highest total absence time 38.11% .

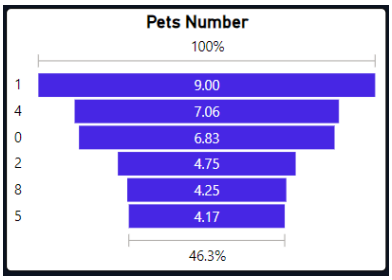


2) **The funnel chart titled "Children Number"** represents the distribution of employees based on the number of children they have, considering data for 73.9% of the total employee population.

Key Observation: The chart reveals that the highest percentage of employees (12.20%) have three children. A significant portion of employees with represented data (73.9%) have children, with a focus on those having three children.

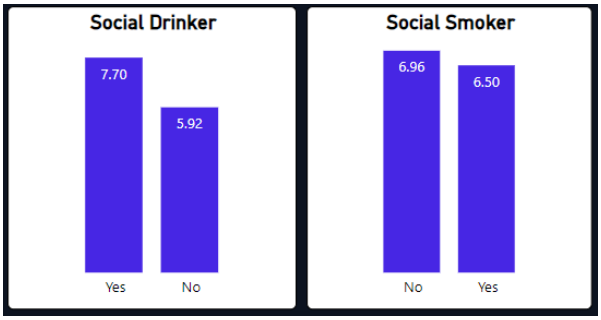


3) There is a similar situation with **the absence of employees who have pets**. The highest average is 9.0% for employees who have only one pet. The total percentage this group of employees represents concerning absences is 46.3 out of 100%.



4) The bar chart titled "**Social Drinker**" compares employees who identify as social drinkers (Yes) and those who do not (No).

- A higher percentage of employees (7.70%) identify as social drinkers than those who don't (5.92%).
- Regarding the Social Smoker chart, the data shows a very close split between employees who are social smokers (6.50%) and those who are not (6.96%).

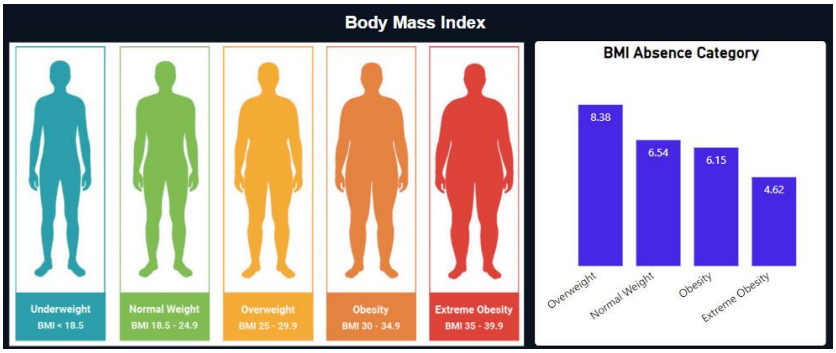


5) The "**Body Mass Index (BMI) Absence Category**" bar chart presents the distribution of employee absences across four BMI categories: Overweight, Normal Weight, Obesity, and Extreme Obesity. Here's a breakdown of the information it conveys:

- **Employee Absence Distribution:**
  - Overweight employees (8.38%) have the highest percentage of absences.
  - The remaining categories follow a decreasing trend: Normal Weight (6.54%), Obesity (6.15%), and Extreme Obesity (4.62%).

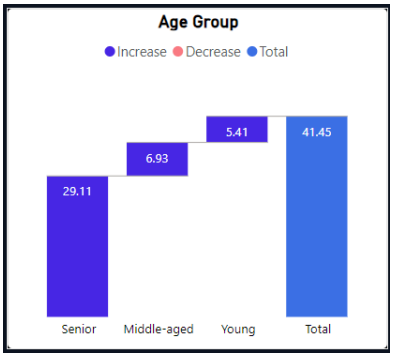
<u>Interpretations:</u>

- This data suggests a potential correlation between overweight status and absenteeism, with overweight employees having the highest absence rate. However, it's important to note the limitations:
- **Causation vs. Correlation:** This chart only shows correlation, not causation. It doesn't necessarily mean being overweight causes absences. Other factors could be influencing both weight and absences.



6) The waterfall chart titled "**Age Group Chart**" depicts the distribution of employees across three age categories: senior, Middle-aged, and Young, representing 41.45% of the employee population.

- Senior employees (29.11%) comprise the most significant absences.
- Middle-aged (6.93%) and Young (5.41%) are the remaining categories that significantly decrease.
- This data suggests a potentially higher concentration of senior employees within the 41.45% of the population represented in the chart.



7) The "**Seniority of Employment**" bar chart presents a comparison between employees categorized by their seniority levels: Senior, Mid-Level, and Junior. Here's a breakdown of the information it reveals:

- Senior employees (7.63%) represent the most significant portion of absences.
- Junior (7.19%) and Mid-Level (5.20%) employees follow with a relatively close distribution.

8) The treemap chart titled "**Distance from Home to Work in KM**" depicts the distribution of employees based on their commute distance, which is categorized as Near (9.71%) and Far (6.13%). Here's a breakdown of the information it conveys:

- A significant proportion of employees (9.71%) live near their workplace and have a major absence average.
- Fewer employees (6.13%) have a long commute (far)

Discussion and Recommendations

4.1 Key Findings:

- ✓ **Musculoskeletal conditions are the leading cause of absenteeism**, followed by injuries and medical consultations. Seasonal variations exist, with **respiratory illnesses peaking in winter** and **skin conditions less prevalent**.

- ✓ **Short-term absences are slightly more frequent than long-term absences.** Absenteeism tends to be **higher in the first half of the year**. Spring has the highest proportion of absences overall.

- ✓ The data suggests a **potential correlation between overweight status and absenteeism**. However, it's crucial to consider limitations and focus on promoting employee wellness for all.

- ✓ The analyzed data set might have a **higher concentration of senior employees**. However, missing data and potential bias require further investigation.

- ✓ The distribution of employees across seniority levels (Senior, Mid-Level, Junior) appears **relatively even** within the data presented.

4.2 Potential Explanations:

- Musculoskeletal conditions might be linked to ergonomic factors in the workplace. Further investigation into specific conditions could reveal areas for improvement.

- Seasonal variations in illnesses could be related to weather patterns and potential spread of viruses during colder months.

- The correlation between overweight status and absenteeism could be due to various health factors. Focusing on employee wellness initiatives that encourage healthy lifestyles is essential.

- The specific data set analyzed could explain the higher concentration of senior employees. Investigating turnover rates across age groups might provide context.

4.3 Recommendations:

- ✓ **Promote ergonomic assessments and workplace safety programs** to potentially reduce musculoskeletal conditions and injuries.

- ✓ **Consider offering flu vaccinations or encouraging healthy practices** during winter to address potential seasonal illness spikes.

- ✓ **Develop and promote employee wellness programs** that encourage healthy habits and disease prevention for all employees, regardless of weight.
- ✓ **Investigate the representativeness of the analyzed data** and explore reasons behind missing data, particularly regarding age and seniority distribution. Consider collecting additional data if needed.

- ✓ **Explore the possibility of flexible work arrangements** to reduce absenteeism related to long commutes or personal needs.

- ✓ **Further, analyze the reasons behind short-term vs. long-term absences** to develop targeted interventions for each category.

Conclusion

Absenteeism analysis is crucial for organizations to understand workforce health and productivity. This report has identified key trends, such as musculoskeletal conditions being the leading cause of absences and a potential correlation between overweight status and absenteeism. While some correlations require further exploration, the report highlights the importance of promoting a healthy work environment and employee wellness initiatives. By implementing the recommendations outlined and continuing to analyze absenteeism patterns, your organization can develop targeted strategies to create a healthier and more productive workforce.

Appendix

- Employee Absenteeism Analysis Dashboard with Power BI

- PDF generated from Jupyter notebook attached to the project about data processing, cleaning and transformation using Microsoft SQL Server and Transact-SQL.

Absenteeism at Work Data Modeling, Tranformation, Cleaning and Creation

Tracking employee absences helps businesses understand how much time is lost and why. Since absences can cost millions each year, it's crucial to monitor them and find ways to bring them down. This reduces disruption and saves the company money.

All information used in this notebook is based on source data:

> Martiniano, Andrea and Ferreira, Ricardo. (2018). Absenteeism at work. UCI Machine Learning repository. https://doi.org/10.24432/C5X882.

In [108…
```
-- Note: This database was created manually also the csv file
-- was imported to the database in the same way.

-- Using the database
USE Absenteeism_at_work
```

Commands completed successfully.
Total execution time: 00:00:00.001

In [109…
```
-- Check all the columns from the new imported csv table
SELECT
    TABLE_NAME,
    COLUMN_NAME,
    ORDINAL_POSITION
FROM INFORMATION_SCHEMA.COLUMNS WHERE TABLE_NAME = 'Absenteeism'
ORDER BY ORDINAL_POSITION ASC
```

(21 rows affected)
Total execution time: 00:00:00.097

Out[109]:

TABLE_NAME	COLUMN_NAME	ORDINAL_POSITION
Absenteeism	ID	1
Absenteeism	Reason_for_absence	2
Absenteeism	Month_of_absence	3
Absenteeism	Day_of_the_week	4
Absenteeism	Seasons	5
Absenteeism	Transportation_expense	6
Absenteeism	Distance_from_Residence_to_Work	7
Absenteeism	Service_time	8
Absenteeism	Age	9
Absenteeism	Work_load_Average_day	10
Absenteeism	Hit_target	11
Absenteeism	Disciplinary_failure	12
Absenteeism	Education	13
Absenteeism	Son	14

Based on data modeling design, transforming, cleaning and creating the necessary tables.

```
--> Creating each new table based on the original data: Absenteeism

-- Creating: Absenteeism_AWK
IF OBJECT_ID(N'dbo.Absenteeism_AWK', N'U') IS NOT NULL
    DROP TABLE dbo.Absenteeism_AWK;
SELECT
    id AS IdEmployee,
    Reason_for_absence AS IdAbsence,
    Transportation_expense,
    Service_time,
    Work_load_Average_day AS Work_load_avg_day,
    Hit_target,
    Absenteeism_time_in_hours,
    Disciplinary_failure,
        Day_of_the_week,
        CASE WHEN Day_of_the_week = 1 THEN 'Monday'
             WHEN Day_of_the_week = 2 THEN 'Tuesday'
                 WHEN Day_of_the_week = 3 THEN 'Wednesday'
                 WHEN Day_of_the_week = 4 THEN 'Thursday'
                 WHEN Day_of_the_week = 5 THEN 'Friday'
                 WHEN Day_of_the_week = 6 THEN 'Saturday'
                 WHEN Day_of_the_week = 7 THEN 'Sunday'
    END AS Day_name,
        Month_of_absence,
        CASE WHEN Month_of_absence = 1 THEN 'January'
             WHEN Month_of_absence = 2 THEN 'February'
                 WHEN Month_of_absence = 3 THEN 'March'
                 WHEN Month_of_absence = 4 THEN 'April'
                 WHEN Month_of_absence = 5 THEN 'May'
                 WHEN Month_of_absence = 6 THEN 'June'
                 WHEN Month_of_absence = 7 THEN 'July'
                 WHEN Month_of_absence = 8 THEN 'August'
                 WHEN Month_of_absence = 9 THEN 'September'
                 WHEN Month_of_absence = 10 THEN 'October'
                 WHEN Month_of_absence = 11 THEN 'November'
         -- month with 0 also setting as december
                 WHEN Month_of_absence IN(12,0) THEN 'December'
  END AS Month_name,
  Seasons,
  CASE WHEN Seasons = 1 THEN 'Summer'
       WHEN Seasons = 2 THEN 'Fall'
           WHEN Seasons = 3 THEN 'Winter'
           WHEN Seasons = 4 THEN 'Spring'
   END AS Season_name
INTO Absenteeism_AWK
FROM Absenteeism;

-- UPDATE Month_of_absence WITH 0 BY 12 (December)
UPDATE Absenteeism_AWK SET Month_of_absence = 12 WHERE Month_of_absence = 0


-- Because the dataset is based on a Brazilian company, the year stations are different
-- from North America, so I will update this to have the stations based on
-- Canada where I live.

UPDATE [Absenteeism_AWK]
SET SEASONS = 1, SEASON_NAME = 'Summer'
WHERE Month_of_absence IN (6,7,8)

UPDATE [Absenteeism_AWK]
SET SEASONS = 2, SEASON_NAME = 'Fall'
WHERE Month_of_absence IN (9,10,11)

UPDATE [Absenteeism_AWK]
SET SEASONS = 3, SEASON_NAME = 'Winter'
WHERE Month_of_absence IN (12,1,2)

UPDATE [Absenteeism_AWK]
SET SEASONS = 4, SEASON_NAME = 'Spring'
WHERE Month_of_absence IN (3,4,5)
```

- **Data modeling and DAX formulas used within power bi to achieve calculations, additional tables etc.**

Measures:

Absenteeism Rate:
Calculates the absenteeism rate by dividing the total absenteeism time in hours by 8 hours (which represents one workday), effectively converting the absenteeism time from hours to days, and then taking the average of these values.

Average Absenteeism time in hours:
Calculates the average absenteeism time in hours across all records in the Absenteeism_AWK table.

Long Term Absence:
Calculates the count of rows in the 'Absenteeism_AWK' table where the absenteeism time is considered "long term", defined as being greater than 3 hours.

Short Term Absence:
Calculates the count of rows in the 'Absenteeism_AWK' table where the absenteeism time is considered "short term", defined as being 3 hours or less.

Calculated tables:

Age Category Table:
Generates a new table that categorizes employees based on their maximum age into three age groups: Young, Middle-aged, and Senior.

BMI Category Table:
Generates a new table that categorizes employees based on their maximum BMI into five BMI categories: Underweight, Normal Weight, Overweight, Obesity, and Extreme Obesity.

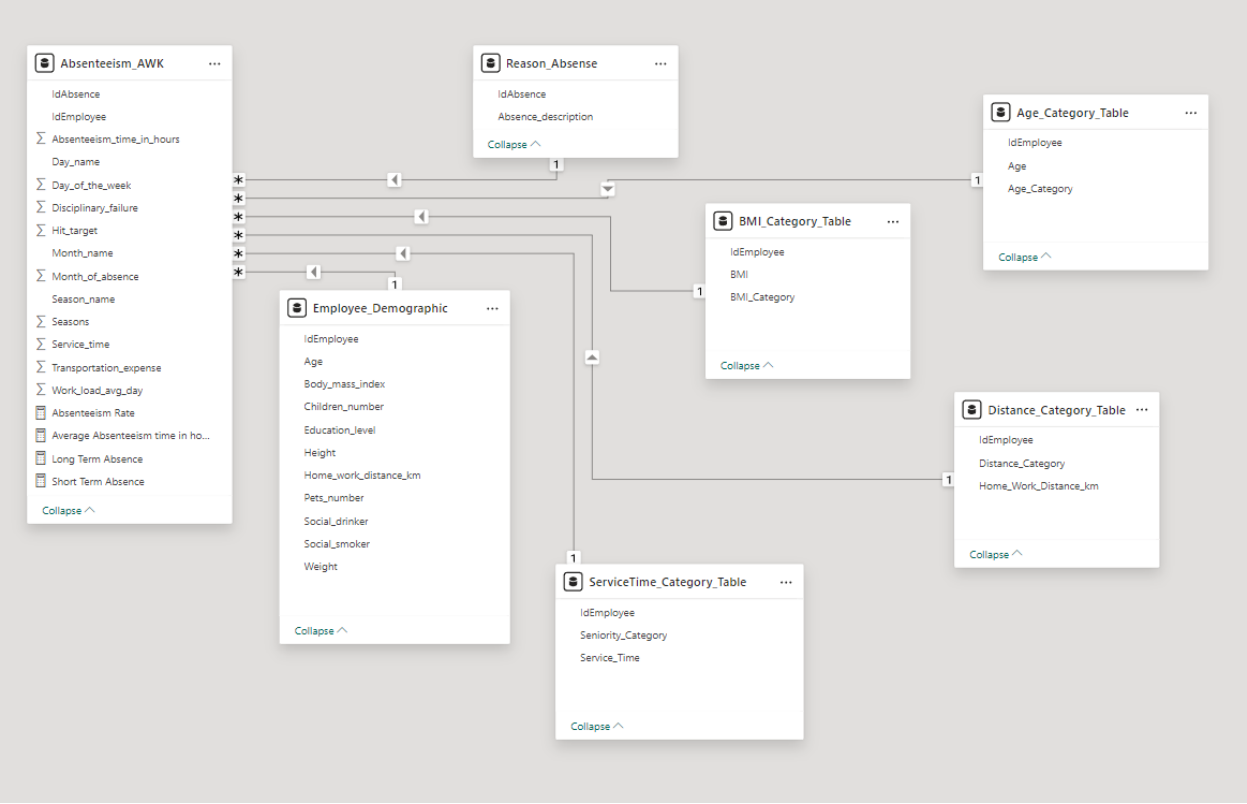
Distance Category Table:
Generates a new table that categorizes employees based on their maximum home-work distance into three distance categories: Very Near, Near, and Far.

ServiceTime Category Table:
This new table aims to categorize employees' service time into three groups: Junior, Mid-level, and Senior.

Physical Tables:

Absenteeism, Employee Demographic, Reason Absence

Data modeling:



References

Martiniano, Andrea and Ferreira, Ricardo. (2018). Absenteeism at work. UCI Machine Learning Repository. https://doi.org/10.24432/C5X882

Inc, T. (2024, March 22). 20 Statistics centered around employee absenteeism [2024]. TeamSense. https://www.teamsense.com/blog/absenteeism-workplace-statistics

Gianluigilopardo. (n.d.). Absenteeism_prediction/absenteeism_notebook.ipynb at master gianluigilopardo/Absenteeism_prediction. GitHub. https://github.com/gianluigilopardo/Absenteeism_prediction/blob/master/absenteeism_notebook.ipynb

Haim, L. S. (2024, March 12). *Visualizing absenteeism at work*. Littal Shemer Haim. https://www.littalics.com/visualizing-absenteeism-at-work/

Icons attribution:

Key performance indicator icons created by HAJICON - Flaticon
Reason icons created by nangicon - Flaticon
Trend icons created by HAJICON - Flaticon
Multitasking icons created by Eucalyp - Flaticon
Notion icons created by meaicon - Flaticon
Absence icons created by afif fudin - Flaticon
Capability icons created by HAJICON - Flaticon
Arrows icons created by Creative Stall Premium - Flaticon