Looking at HR Data for a company SQL and Power BI Project

Why this project?

I was motivated to see what the use case of data analysis when it comes to HR data and how I can potentially find key insights and patterns by looking at the row data and find out within a year how many employees were absent, how reasons of being absent and their overall health status.

Key Takeaways

Dataset Details.

This Dataset was Used from a popular data analysis/scientists' community Kaggle. It has about 740 rows which are employees. This dataset consists of 3 tables with details like Absenteeism/reasons/compensation. We will be using some relevant columns for our data analysis which in this case only focus on absenteeism reasons and compensation.

Questions from HR

1: Hr wants to find out the average absentee hours from last year?

2: Hr requested to find out the top 3 reasons why employees were absent from work.

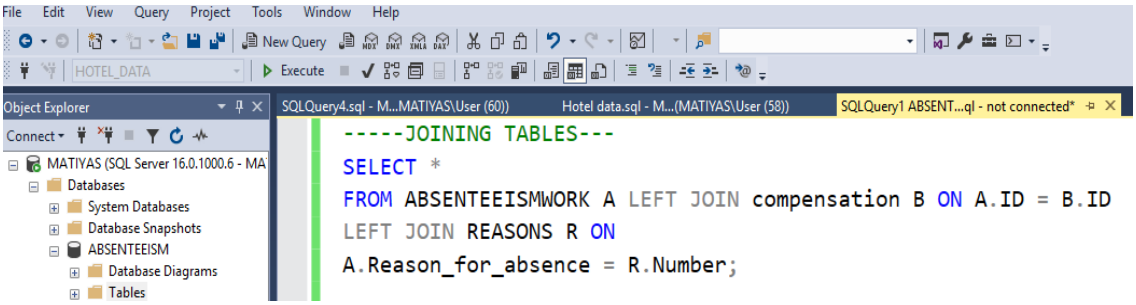
3: How many employees actively smoke? Was that the reason for being absent?

4: Looking at the season when those employees took leave on what Season how many and the main reason?

5: Hr. also wanted to give a raise for employees that don't smoke and were not absent for more than two days that year as top employee performance?

Analysis / Cleaning process

I begin by cleaning by data set. I begin by first Joining my tables in SQL which are Absenteeism, Compensation and Reasons. Using the left join I would bring two of my tables as one and begin looking at my data.



Upon joining I will be looking at making some changes of column fields from the absenteeism data which are the Body mass index and month of absence.

Boday Mass Index -Healthy, Over and underweight Accordingly

Month Of absence – Winter Fall Summer and spring.

Using Case When Statement I will be writing a query for better visualization for my Power BI Dashboard later

I will be also cleaning and removing Unknown data since they are missing key information for my analysis which I will either make as Null or 0 moving forward.

```sql
SELECT A.ID,
R.REASON,
MONTH_OF_ABSENCE,
BODY_MASS_INDEX,
CASE WHEN Body_mass_index < 18.5 THEN 'UNDERWEIGHT'
     WHEN Body_mass_indeX BETWEEN 18.5 AND 25 THEN 'HEALTHY WEIGHT'
     WHEN Body_mass_index BETWEEN 25 AND 30 THEN 'OVERWEIGHT'
     WHEN Body_mass_index > 18.5 THEN 'UNDERWEIGHT'
     ELSE 'UNKNOWN' END AS BMI_CATAGORY,

CASE WHEN MONTH_OF_ABSENCE IN (12,1,2) THEN 'WINTER'
     WHEN MONTH_OF_ABSENCE IN (3,4,5) THEN 'SPRING'
     WHEN MONTH_OF_ABSENCE IN (6,7,8) THEN 'SUMMER'
     WHEN MONTH_OF_ABSENCE IN (9,10,11) THEN 'FALL'
     ELSE 'UNKNOWN' END AS SEASON_NAMES,
     MONTH_OF_ABSENCE,
     DAY_OF_THE_WEEK,
     TRANSPORTATION_EXPENSE,
     Education,
     Son,
     Social_drinker,
     Social_smoker,
     PET,
     Disciplinary_failure,
     Age,
     Work_load_Average_day,
     Absenteeism_time_in_hours
FROM ABSENTEEISMWORK A LEFT JOIN compensation B ON A.ID = B.ID
LEFT JOIN REASONS R ON
A.Reason_for_absence = R.Number;
```

Optimized my query so I can do my analysis furthermore on my power bi dashboard. I will be also attaching .SQL file to further dive into it by only using SQL and I potentially looked at some insights before my power Bi dashboard.

First Question asked was what is the average late time?

(SELECT AVG (ABSENTEEISM_TIME_IN_HOURS) AS average_late_time

FROM ABSENTEEISMWORK A LEFT JOIN compensation B ON A.ID = B.ID

LEFT JOIN REASONS R ON

A.Reason_for_absence = R.Number)

Top 3 reasons why employees are absent from work?

Looking at our Power Bi dashboard for this observation the top three reasons why employees are absent from work are medical consultation which take up 149 employees followed by the dental consultation which are 112 employees the last one is physiotherapy.

What is the season with the highest absenteeism rate?

Spring has one of the highest absenteeism rates where employees go take days off to go to their doctors office and also get other personal matters sorted out.

Based out off all the given Hr data provided by Human Resources for a particular year employees were taking off ours for probably reasoning. And we have seen. We have also managed to identify and give salary raise by 3$ more hourly for 686 employees that are nonsmokers, which is 92.7% of our employee demographics. The other trends we looked at is the division of total hours that were allocated to off time for all 740 employees we have also managed to divide these hours by seasons reasons and employee typs.