Hej alle data managere. I denne uge skal I genbesøge Power BI, og udarbejde et nyt dashboard med nedslag i datasammenhænge I finder interessante. Dataene er en anonymiseret udgave af Novo nordisk HR-data, og er baseret på en "virkelig" arbejdsopgave med at optimere et dashboard der kigger på medarbejder performance og fravær.

Da opgaven er en fra det "virkelige liv" bærer den også præg lige, som værende ret åben. Ligesom i de forrige uger på akademiet, vil vi opfordre jer til at danne jer et generelt overblik over opgaven og herefter lave et nedslag i noget I finder interessant. Husk ikke lade jer afskrække af at afgrænse jeres analytiske fokus i dataene - det er ofte når man skærer elementer fra, at man finder de mere "snævre" og sigende sammenhænge i dataene.

Jeres optimerede dashboards skal afleveres som en præsentation på 10 - 15 min hvor I gennemgår jeres tilgang til opgaveløsningen, prioritering af interesseområder samt en demonstration af jeres grafiske repræsentation af dataene i Power BI. Vi vil endnu engang opfordre jer til at sparre med hinanden i jeres kreative/analytiske tilgang til opgaveløsningen. For at understøtte denne sparring vil vi også afholde endnu et sparringsseminar på tværs af Aarhus & Ballerup om onsdagen.

I skal også være opmærksomme på at denne opgave er mere avanceret end den sidste i løste og Specialisterne forventer ikke at I kan nå løse alle dele eller forslag i opgaven. Prioritér derfor jeres tid og løs de dele i finder mest interessant og forsøg at kom helt i bund i en datasammenhæng inden i går videre til en ny.

Fordi opgaven er udviklet i samarbejde med Novo nordisk, er den stillet på engelsk, hvis nogen af jer har udfordringer med dette skal i blot kontakte en teamcoach. Alternativt kan I gøre brug af AI til en oversættelse af opgaven.

OBS: Datasættet skal åbnes i desktop udgaven af Power BI!

HR Absence Tool for People & Organization (P&O) in Novo Nordisk

Novo Nordisk's P&O department has hired you to improve their absence dashboard. The attached file contains the current dashboard, and all the data used for it.

Ultimate Calendar Table:
For this assignment you should rely on a table named Calendar Table, you can see how it was created using the advance dax editor in power query. This table includes a wide range of date-related columns that enhance the functionality and flexibility of time-based analyses.

The table starts with a specified start date (January 1, 2022) and dynamically calculates the end date based on the current date. It includes columns for calendar months, fiscal months, quarters, years, and various offsets that allow for easy filtering and comparison of data across different time periods. Additionally, the table provides flags for year-to-date (YTD), month-to-date (MTD), and quarter-to-date (QTD) calculations, which are useful for tracking performance metrics over these periods. The inclusion of weekday and weekend indicators, as well as week sequence numbers, further enhances the ability to perform detailed time-based analyses.

This table is particularly valuable for creating dynamic and interactive dashboards in Power BI, enabling users to visualize and analyze data with greater precision and insight. It is highly recommended to use this table or time measures for filtering.

The original database schema diagram with multiple interconnected tables. Each table contains various fields, and the relationships between the tables are indicated by lines connecting them. Here are the tables and their fields:

1. **OriginalReductionSchema**
    - Calculation Month
    - Reduction Month
2. **Employee Access Table**
    - Organizational Node Id
    - Employee Id
3. **Employee**
    - Job Group
    - Job Code
    - Employee Number
    - Employee Name (First)
    - Employee Name (Last)
    - Company Code
    - Start Date
4. **Absence Monthly**
    - 202301
    - 202302
    - 202303
    - Absence in Days, Hours
    - Absence in Days, Memory_Hours
    - Absence in Hours, Memory_Days
    - Total Absences in Days, Hours
5. **Measure Table**
    - Measure Value
6. **Type of Abs Category**

    - CategoryId
    - CategoryName
7. **Absence Options**
    - OptionId
    - OptionName
8. **Hierarchy**
    - Level_1
    - Level_2
    - Level_3
    - Level_4
9. **Ultimate Calendar Table**
    - CalendarMonth
    - CalendarQuarter
    - CalendarYear
10. **Absence Period**

- 202301
- 202302
- 202303
- Period in Days, Hours

The diagram visually represents the structure of a database and how different pieces of data are related to each other through various keys and relationships. This is crucial for your work to find a better way to do this as this model is heavy and makes the file extremely heavy and slow.

A recommendation is to find the columns in the data that haven't been used in the file and remove them.

Case description

Your manager has also attached a wishlist for the contents of the dashboard and certain data insights they view as critical for their business operations. However, you quickly realize that this list is quite thorough and will be hard to realize within the given amount of time, you also question your manager's knowledge of the data. To accommodate both the managers' requests and the short timeframe you decide to use their inquiries and guidelines as suggestions and attempt to compromise to fulfill the assignment. Make sure to pick and choose the suggestions you deem most impactful or interesting:

1. **Overview Report**:
    - Total number of employees (headcount)
    - Absence due to illness
    - Year-to-date absence percentage

    - Rolling 12-month absence percentage
    - Subcategories of absence (very short, short, medium, extended, long) for each EVP Area
2. **Yearly Summary**: This page should:
    - Update itself every month
    - Allow directors to compare this year's data with last year's data
    - Show absence percentage, year-to-date absence percentage, and the same measures for last year
    - Use columns for percentages and lines with dots for year-to-date data
3. **Yearly Metrics**: This page should show:
    - Absence
    - Year-to-date absence
    - Rolling 12-month absence
    - Refer to the page named "Correct Selections, AJUE Yearly, PSQIT" for guidance
4. **Employee valuation proposition (EVP) Area Summary**: This page should show a summary of absence by EVP Area.
5. **Top 10 Departments**: This page should show the top 10 departments with the highest absence rates, represented by the column "EVP – 2 Area".

Additional Tasks:

- **Redesign the Many-to-many (MMR) Structure**: Based on discussions with your Product Owner, it has been concluded that the current MMR structure is not suitable. One of your tasks is to redesign the MMR to find a more efficient and optimal way to set the semantic model.
- **Review Metrics**: Your Product Owner has also recommended reviewing the metrics, as they seem too lengthy for their intended purpose.

Steps to Consider for a Better Semantic Model in Power BI:

1. **Data Collection and Preparation**:
    - **High-Quality Data**: Ensure the data you collect is relevant, accurate, and comprehensive. Remove any duplicates or irrelevant information.
    - **Data Cleaning**: Preprocess the data to remove noise and inconsistencies. This includes handling missing values, correcting errors, and normalizing text.
    - **Data Documentation**: Maintain clear documentation of the data sources, preprocessing steps, and any transformations applied.
2. **Data Modeling**:
    - **Define Relationships**: Clearly define relationships between tables. Use primary and foreign keys to establish connections.

- **Star Schema**: Consider using a star schema for simplicity and efficiency. This involves a central fact table connected to multiple dimension tables.
- **Normalization**: Ensure that data is normalized to reduce redundancy and improve data integrity.

3. **Measure Creation**:
    - **DAX (Data Analysis Expressions)**: Use DAX to create calculated columns and measures. This allows for dynamic calculations and aggregations.
    - **Key Metrics**: Identify and create key metrics that are essential for the dashboard, such as total absence hours, absence percentage, and year-to-date absence.
4. **Data Visualization**:
    - **Clear and Intuitive Visuals**: Choose visuals that clearly represent the data and different metrics (eg. bar charts, line charts, and pie charts etc.)
    - **Interactive Elements**: Add interactive elements like slicers and filters to allow users to explore the data.
    - **Consistent Formatting**: Ensure consistent formatting across all visuals for a professional look.
5. **Performance Optimization**:
    - **Efficient Queries**: Optimize queries to ensure fast data retrieval. Use query folding and avoid complex transformations in Power BI.
    - **Data Aggregation**: Aggregate data where possible to reduce the volume of data processed.
    - **Incremental Refresh**: Implement incremental refresh to update only the data that has changed, improving performance.

Things to Avoid:

1. **Overcomplicating the Model**:
    - **Keep It Simple**: Avoid overly complex models with too many relationships and tables. Simplicity improves performance and maintainability.
2. **Ignoring Data Quality**:
    - **Validate Data**: Ensure data quality by validating and cleaning data before importing it into Power BI.
3. **Poor Visualization Choices**:
    - **Avoid Clutter**: Do not overcrowd the dashboard with too many visuals. Focus on key metrics and ensure clarity.
4. **Neglecting Performance**:
    - **Optimize Regularly**: Regularly review and optimize the model to ensure it performs well with large datasets.