

TASK 1: HR DATA ANALYSIS

MITALI NITIN KUBAL

[mitali.kubal2001@gmail.com](mailto:mitali.kubal2001@gmail.com)

Greetings,

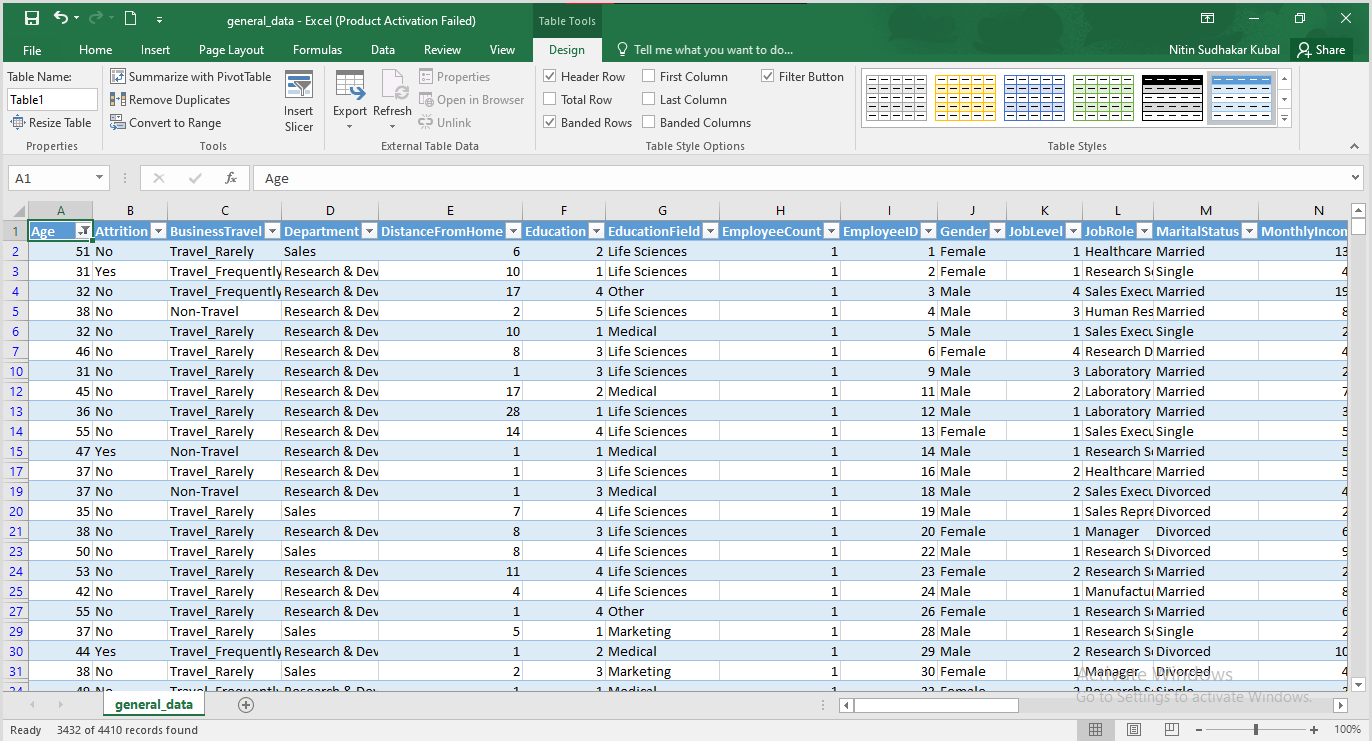
I am excited to embark on a virtual internship assigned by Psyliq, focused on data analysis using hr datasets. Leveraging advanced Excel functions & Power BI, I aim to gain valuable insights and deepen my understanding of the dataset. As a student pursuing MSc in Statistics, this internship offers a unique opportunity to apply my skills in a practical setting, enhancing my proficiency in analyzing complex data.

The dataset provided encompasses various facets of hr data. Based on these datasets I have to answer the assessment questions. Through these tasks, I aspire not only to meet the internship requirements but also to contribute meaningfully to the exploration of data analysis dynamics through evidence-based insights.

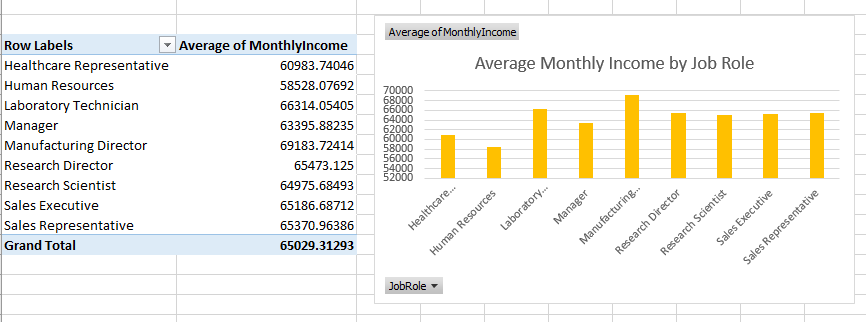
**Aim:**

To analyze the given datasets and perform the tasks provided.

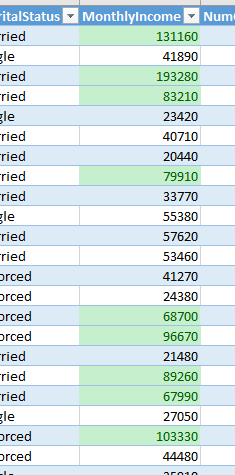
1. Using Excel, how would you filter the dataset to only show employees aged 30 and above?



1. Create a pivot table to summarize the average Monthly Income by Job Role.

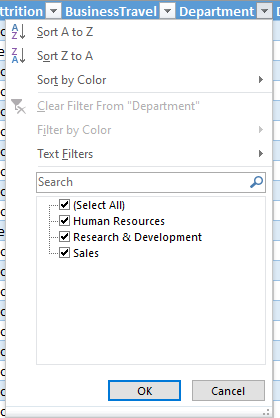


1. Apply conditional formatting to highlight employees with Monthly Income above the company's average income.

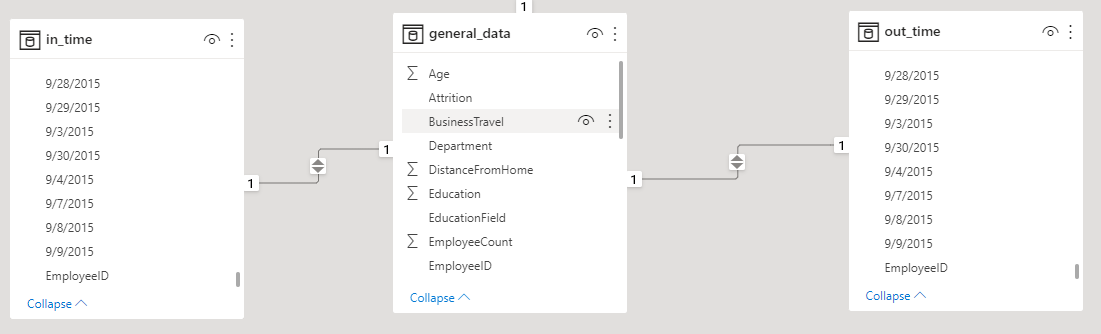


1. Create a bar chart in Excel to visualize the distribution of employee ages.
2. Identify and clean any missing or inconsistent data in the "Department" column.

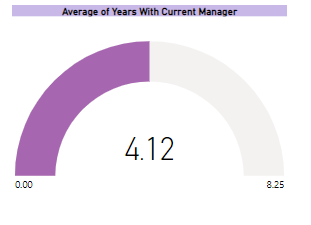
No missing values or inconsistent data in the “Department” column.



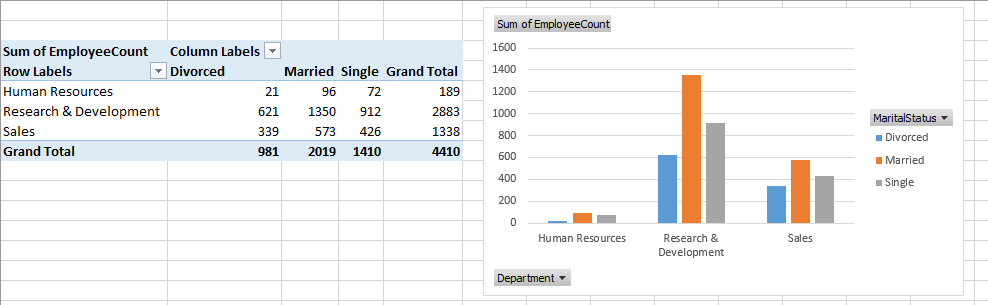
1. In Power BI, establish a relationship between the "EmployeeID" in the employee data and the "EmployeeID" in the time tracking data.



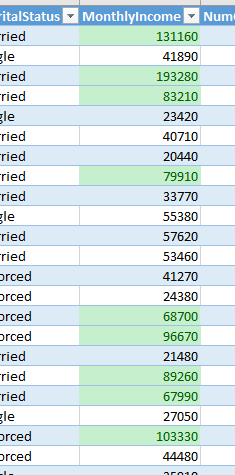
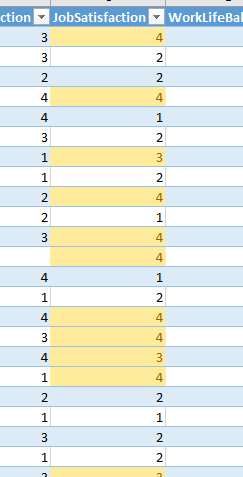
1. Using DAX, create a calculated column that calculates the average years an employee has spent with their current manager.



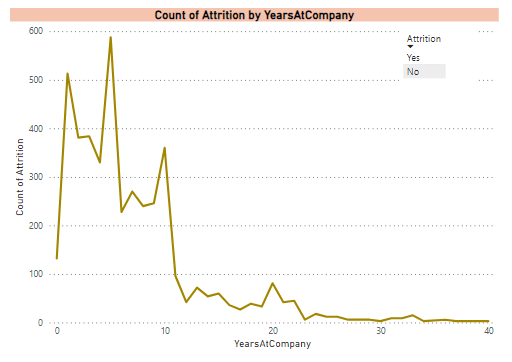
1. Using Excel, create a pivot table that displays the count of employees in each Marital Status category, segmented by Department.



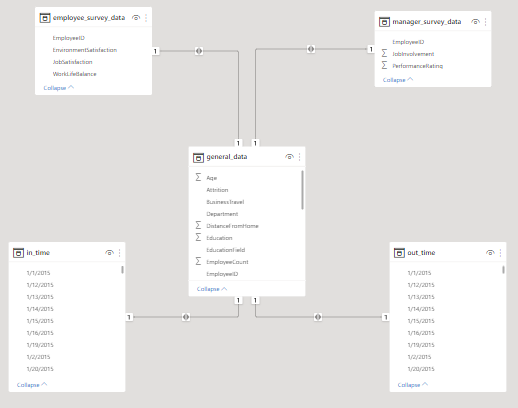
1. Apply conditional formatting to highlight employees with both above-average Monthly Income and above-average Job Satisfaction

1. In Power BI, create a line chart that visualizes the trend of Employee Attrition over the years.



1. Describe how you would create a star schema for this dataset, explaining the benefits of doing so.



* Fact table: general\_data
* Dimension tables: employee\_survey\_data, manager\_survey\_data, in\_time, out\_time
* Primary key: EmployeeID

The star schema offers several benefits for organizing and querying data in a data warehouse or relational database:

* Simplified Querying: Star schemas simplify the querying process by reducing the number of joins required.
* Enhanced Performance: The structure of a star schema often leads to improved query performance. The database engine can efficiently process queries involving the fact table and its related dimensions due to the clear and optimized relationships.
* Improved Readability: Users can easily understand the relationships between the central fact table and dimension tables, making it more accessible for analysts and report developers.
* Scalability: Star schemas are scalable and flexible, allowing for the addition of new dimensions or facts without disrupting the existing structure
* Normalized Data: While fact tables contain numerical data for analysis, dimension tables can be normalized to reduce data redundancy. This normalization ensures consistency and reduces the risk of data anomalies.

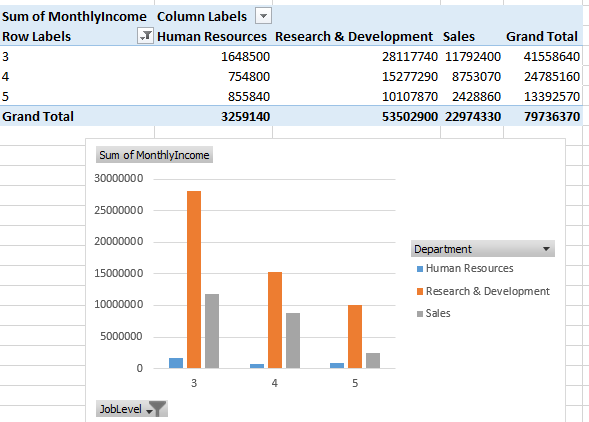
13) Create a hierarchy in Power BI that allows users to drill down from Department to Job Role to further narrow their analysis.

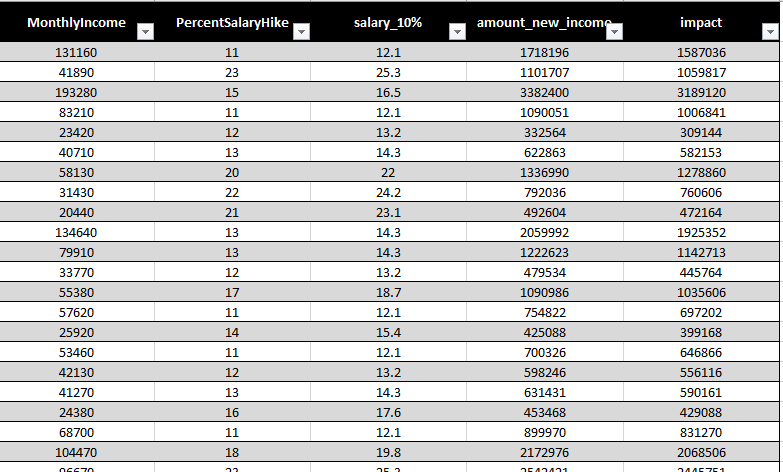
14. How can you set up parameterized queries in Power BI to allow users to filter data based on the Distance from Home column?

* In Power BI, set up parameterized queries by creating a parameter for the Distance from Home column.
* Go to "Home" and select "Manage Parameters" to define a new parameter, let's call it "DistanceParameter."
* Apply this parameter in the Power Query Editor to replace the hardcoded distance filter with the parameter. For example, in your query, use each [DistanceFromHome] = DistanceParameter.
* Now, users can dynamically filter data based on Distance from Home by adjusting the parameter value in the Manage Parameters window, providing a flexible way to interact with the report.

15. In Excel, calculate the total Monthly Income for each Department, considering only the employees with a Job Level greater than or equal to 3.



16. Explain how to perform a What-If analysis in Excel to understand the impact of a 10% increase in Percent Salary Hike on Monthly Income.



17. Verify if the data adheres to a predefined schema. What actions would you take if you find inconsistencies.

* Yes, the Data adheres to a predefined schema.

Actions to take if inconsistencies found are:

* Investigate Sources: Identify the origin of inconsistencies by examining data collection, entry, and processing methods.
* Correct Data: Take corrective actions such as data cleaning, removal of duplicates, and rectifying erroneous values.
* Implement Data Checks: Set up data quality checks and validation procedures to prevent future inconsistencies.
* Enhance Data Governance: Review and improve data governance practices, including access controls and data stewardship programs, to maintain consistency and integrity.