**EMPLOYEE DATA ANALYSIS**

**Data Cleaning:**

* First, I opened the dataset and checked for blanks in all the columns but there was none.
* Then I checked for duplicates in the employee ID and realize that there are 172 duplicate records and these duplicated employee IDs had unique employee names. I proceeded to use conditional formatting to sort the duplicate records by colour. For each duplicate record. I changed the last digit of the employee ID ensuring that there were no more duplicates.
* While cleaning, I noticed a discrepancy on the job title. I had 110 records of Sr. Manger in place of Sr. Manager so I used the Find and Replace function to replace Sr. Manger with Sr. Manager.
* I cross-checked the datatype of each column to ensure that they were correct. I converted the Annual Salary datatype to Currency, the Bonus % datatype to Percentage, the Hire date and Exit Date datatypes to Date etc.
* I added a new column named ‘Hire Month’ which was extracted from the hire date
* I added another column and named it ‘Age Group’ and this segregates the employees into age groups of Young adults, Adults, Middle-Aged Adults and Elderly. I used this formula: =IF(AND(Age>=20, Age<=30), "Young Adult", IF(AND(Age>=31, Age<=40), "Adult", IF(AND(Age>=41, Age<=50), "Middle Aged Adult", "Elderly"))).
* I added another column called ‘Duration of Employment’ which I calculated using this formula =IF(AND(Hire Date<>"", Exit Date<>""), Exit Date - Hire Date, IF(Hire Date <>"", DATE(2024,4,16)- Hire Date, ""))

**Data Transformation and Visualization in Power BI:**

* First, I imported the cleaned Excel file into Power BI using the Get Data option and clicked on Transform Data
* Then I checked that the datatype of each column is correct and adjusted those that needed adjustments
* I calculated the average age of employees by using the average filter on Power BI and I visualized that with a card
* Then I visualized the number of employees by department using the Clustered bar chart visual
* I showed the most frequent job title with a card visual using the Top N filter in Power BI (this can also be done on Excel using this formula =INDEX(C2:C1001, MODE(MATCH(C2:C1001,C2:C1001,0)))
* I visualized the department with the highest number of employees using the card visual
* I calculated the average annual salary by using the average filter option on the visualization pane and visualized that using a card visual.
* I used a pie chart to show the gender distribution among employees
* I used a line chart to show the trend in hires over the years
* I showed the age group across different genders using the stacked column chart
* I calculated the average highest average bonus percentage for each department and visualized that using a stacked bar chart
* I used a clustered column chart to show the top 5 employees who have stayed the longest in the organization. I used different shades of one colour to differentiate between the employees since the values are quite close i.e the darker the shade, the higher the employment duration.
* On the second page on Power Bi titled employee demographics, I visualized the total number of exited employees, total number of active employees, number of employees against the months they were hired, the month where employees resigned the most and the average annual salary for each job title across all gender with different visuals and I used a slicer to filter these records by age group

**Answer to questions from the dataset:**

1. What is the average age of employees in the dataset?

The average age of employees in the dataset is 44.38

1. How many employees are there in each department?

IT – 241 employees, Engineering – 158 employees, Sales – 140, Human Resources – 125 employees, Finance – 120 employees, Marketing – 120 employees, Accounting – 96 employees

1. Which job title appears most frequently in the dataset?

The job title with the most occurrence is Director

1. What is the distribution of gender among employees?

The gender distribution is 518 female employees and 482 male employees

1. Which department has the highest average bonus percentage?

The department with the highest average bonus percentage is Marketing with 0.12%

1. Are there any missing values in the exit date column, if so, how many?

Yes, there are 915 missing values in the Exit Date column

1. Are there any duplicate employee IDs?

There are 172 duplicate employee IDs

1. What is the average annual salary?

The average annual salary is $113.22K

1. Which department has the highest number of employee?

The IT department has the highest number of employees with 241

**Recommendations for the company:**

1. Recognize and celebrate the top 5 employees with the highest duration of employment
2. Since most employees exit the company in May, hire more employees in April to ensure business continuity
3. About 33% of the total number of employees (331) are elderly and might leave the company soon for retirement. The company should hire more adults, young adults and milled-aged adults