A green letter and black text

Description automatically generated

**Analysis on HR Analytics Data based on various features**

**NAME:**

**INSTRUCTOR:**

Data Visualization **INFO 5709**

**Term Project**

1. **Introduction**

It is important to understand the employee behavior to improve the performance of the organizations. Some of the aspects of the EMPLOYEE’S work environment, remunerations and positions affect their level of satisfaction and hence commitment to the employers’ organization. In this project, the focus is on the correlation between business travel needs, employees’ remuneration structure, and the job positions held within organizations in a bid to analyze causes of high turnover and inequalities in the organizations.

Using employee data, this work explores how gender, department, salary slabs, and attrition are related. It aims to answer critical questions like:

1. Is there a high probability of attrition among employees who travel more often?

2. Is there a meaningful gender and department wage differential among employees?

3. How do positions one occupant affect travelling needs and pay structures?

This project utilizes Python for data pre-processing and Tableau for interactive data visualization in order to help the HR teams by making the relevant data-driven decisions with the goal of achieving workforce equity, satisfaction, and retention.

1. **Dataset Description**

The dataset contains many variables describing samples of employees: demographic, organizational status, pay, and other characteristics relevant for studying patterns that impact employee behavior and turnover.

**Source**: [Kaggle HR Analytics Dataset](https://www.kaggle.com/datasets/anshika2301/hr-analytics-dataset/data)

**Size**:

* **Rows**: 1,480 (representing employees)
* **Columns**: 38 (attributes detailing employee information)

**Key Attributes**

**1. Demographics**

* **Age**: Employee age in years.
* **Gender**: Male or Female.
* **Education**: Educational level (1 = Below College, 2 = College, 3 = Bachelor’s, 4 = Master’s, 5 = Doctorate).
* **Education Field**: Field of study (e.g., Life Sciences, Medical, Marketing).
* **Marital Status**: Employee's marital status (e.g., Single, Married, Divorced).

**2. Work-Related Attributes**

* **Department**: Division in the organization (e.g., Sales, Research & Development, Human Resources).
* **Job Role**: Specific job title (e.g., Data Scientist, Sales Executive).
* **Business Travel**: Frequency of business travel required for the job (Travel\_Rarely, Travel\_Frequently, Non-Travel).
* **Years at Company**: Number of years the employee has worked in the organization.
* **Years with Current Manager**: Duration of working under the same manager.

**3. Compensation and Benefits**

* **Monthly Income**: Employee’s monthly salary in dollars.
* **Salary Slab**: Categorized income groups (Low, Medium, High).
* **Stock Option Level**: Availability of stock options (0, 1, 2, 3).

**4. Attrition Indicators**

* **Attrition**: Whether the employee left the organization (Yes/No).
* **Overtime**: Indicates if the employee works overtime (Yes/No).
* **Job Involvement**: Measure of job engagement (Scale: 1-4).
* **Job Satisfaction**: Satisfaction level of the job (Scale: 1-4).
* **Environment Satisfaction**: Satisfaction with the work environment (Scale: 1-4).

**5. Performance Metrics**

* **Performance Rating**: Performance appraisal score (Scale: 1-4).
* **Training Times Last Year**: Number of training programs attended in the past year.
* **Percent Salary Hike**: Percentage increase in salary over the last hike.

**6. Additional Attributes**

* **Distance from Home**: Distance between employee’s residence and workplace (in miles).
* **Work-Life Balance**: Perceived work-life balance (Scale: 1-4).
* **Total Working Years**: Cumulative years of work experience.

1. **Data Preparation and EDA**

Tool Used: Python and Tableau

Using Python libraries, we can clean and preprocess the dataset. After removing 7 duplicate entries we got 1473 records, and the missing values have been handled based on Mean of the column.

A screenshot of a computer

Description automatically generated

To get better view on data, few visualizations on dataset.

A screen shot of a graph

Description automatically generated

The following histogram shows the spread of employees’ ages where a significant number of employees are within the ages of 30-40 years. The age distribution is also presented by Kernel Density Estimate curve (KDE) that lays on the top of histogram which is more smoothed than histogram.

A screenshot of a computer screen

Description automatically generated

The presented bar chart presents the distribution of the employees by departments, and it can be identified that most of them work in the Research & Development department, and the Sales departments are second, while the Human Resources department has the least number of employees. This will show the composition of the workforce by departments as well as give some information on the sizes of the departments.A screenshot of a computer screen

Description automatically generated

The boxplot illustrates the distribution of employee ages across the three departments: research and development wing, a sales department and human resource department. The departments: designing, outing, human resources and IT all fall within the age bracket of 20 to 60 years Though, the median age of the employees in the Sales department is slightly lower than the other departments meaning that this department has a youthful work force.A screen shot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

1. **Hypothesis**
2. **Employees working overtime are more likely to experience attrition.**

The Bar plot is used to compare the overtime and the turnover rate of the employees. The attrition rate is significantly higher for those employees who work overtime (127 employees) while compared with those who do not work overtime (289 employees). This shares the cross-sectional analysis view that employees working overtime will be inclined to turnover hence the importance of workload and work-life interventions.

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

The above horizontal bar plot focuses on the relationship between overtime, attrition and average monthly income. The findings suggest that employees with larger average monthly incomes remain with the organization regardless of whether they work overtime, while lower paid employees are more likely to leave. This means that even in organizations with many people working on extra hours, pay is a moderate force in retention, underlining the need for organizations to offer reasonable wages.**A screenshot of a computer

Description automatically generated**

This is a box and whisker plot of attrition, overtime and average monthly income in a scatter plot with count of employees. It points out that those employees who work for more than 40 hours per week and have lower average monthly wages are the ones who are likely to turnover. It continues the trend of demonstrating that overtime does cause attrition, but competitive compensation can help to reduce its impact with practical information on how retention may be accomplished.

1. **Employees with salary slabs varies by Gender and Department with influence of age and job roles.**

The scatter plot below represents the flow of the employees based on the gender and departments with further drill down based on the age of the employee and their job title in a way providing a filter on the salary slabs. It demonstrates how the salary disparity exists through gender and departments and how many variations are seen on higher salary brackets in Research & Development and Sales departments, of course.**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

The above categorized and filterable stacked bar chart represents the distribution of employees based on their salary slabs, Gender, Department, Age & Job Roles. This tells that, in terms of salary, the male employees in Research & Development and Sales departments have higher number of employees who earn in the higher salary bands such as 10k-15k and 15k+ respectively, while female employees are distributed more or less equally in all the low and middle salary bands thus pointing to the possibility of any gender pay gap in various departments.

**A screenshot of a computer

Description automatically generated**

The following highlighted table demonstrates the number of employees, grouped by the salary slabs, gender and department (Human & Resource, Research & Development, and Sales). The darkness ranges from the number of employees in each category and show that the male employees more are in higher paying bracket (15000 and above) in Research & Development and female employees are more in lower paying bracket (5000 and below) across departments. This has raised some questions whether there is unfair remuneration between the gender in the organization.

1. **Employees with frequent or no business travel are divided, and their travel requirements may influence with Salary Slab and Job Role.**

This pie chart visualizes the distribution of employees based on their business travel frequency, segmented into categories: non-travel, occasional travel, and frequent travel. The biggest share is of workers who travel none or less frequently – 1,038 of them; 278 workers travel frequently and 150 of them do not travel at all. This means that majority of the employees have little or no travelling needs, and more frequent travelling is less typical, which might have been affected by the salary slab and the job positions.**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

This stacked bar chart represents the count of employees based on their business travel frequency: Non-Travel, Travel\_Rarely, Travel\_Frequently. From the chart, it is evident that majority of the employee’s travel rarely (1,038), with the second largest group employees that travel frequently (278) and the smallest group being those employees that do not travel at all (150). This visualization concerns the travel distribution among employees and can be further distinguished by salary slab and job role.

**A screenshot of a computer

Description automatically generated**

This box-and-whisker plot represents the distribution of employees and the count of employees travelling for business and the frequency at which they travel (Non-Travel, Travel\_Rarely, and Travel\_Frequently). The plot also evaluated the distribution and the median of the employees for each of the four categories and depicted that ‘’Travel\_Rarely’’ category has the highest density of employees whereas ‘’non-travel’’ and ‘’Travel\_Frequently’’ categories have least density of employees. This visualization provides an understanding of differences in travel needs and their impact on workforce distribution.

**Dashboard in Tableau**

This dashboard combines multiple visualizations to present insights into employee behavior and organizational dynamics:

1. Pie Chart (Top Left): Shows the distribution of the employees according to the business travel status: Non-Travel, Travel\_Rarely, Travel\_Frequently where more than half of the employees are categorized as Travel\_Rarely.

**A screenshot of a computer

Description automatically generated**

2. Bar Chart (Bottom Left): Shows that overtime has a positive correlation with turnover, which means that those employees who work overtime leave the organization.

3. Scatter Plot (Right): Analyzes the salary distribution in different slabs and finds out that many differences exist, where most of the males are present in the higher salary slab in groups like Research & Development and Sales.

Age, department, and job position work as filters for interactivity so that users can go to the individual categories for further analysis. By using this dashboard, it is easy to get an overview of trends in travel, salary, and overtime and how this information influences future HR strategies.

**Conclusion**

During the analysis of the data set, we identified key patterns about employees and business processes in the organization. Workers, especially those who have worked extra hours, are more likely to quit the organization quickly demonstrating why workload is a critical aspect and why employees must be paid fairly. The study also revealed that ongoing competitive pay structures decrease the effects of overtime as a cause of attrition. This means that the gender pay gap was evident; males dominated the higher salary classes more so in technical positions while females were dominant in the lower and middle salary categories; therefore, the need for gender sensitive remuneration policies. Employee travel behavior analysis identified that majority of employees travel less frequently while few sections of employees travel frequently, and they belong to higher rank of job designation and higher salary bracket. New data showed that Research & Development became the largest department that employed people of all salaries. Future studies could involve the use of tools including attrition risk prediction models, sentiment of employee satisfaction and results from multi-year analysis to effectively champion equity, retention and productivity.

**Reference**

*HR Analytics Dataset*. (2023, October 27). Kaggle. <https://www.kaggle.com/datasets/anshika2301/hr-analytics-dataset/data>