# **DATA INTERPRETATION**

## **DOMAIN ANALYSIS:**

**1) EmpNumber:** Unique identifier for each employee in the dataset.

**2) Age:** Age of the employee, providing insight into workforce demographics and potential correlations with attrition.

**3) Gender:** Gender of the employee, which may impact workplace dynamics and attrition patterns.

**4) EducationBackground:** The educational background of the employee, influencing skillset and career trajectory.

**5) MaritalStatus:** Marital status of the employee, potentially affecting work-life balance and job satisfaction.

**6) EmpDepartment:** Department in which the employee works, indicating job role and organizational structure.

**7) EmpJobRole:** Specific job role of the employee within their department, reflecting responsibilities and career path.

**8) BusinessTravelFrequency:** Frequency of business travel for the employee, impacting lifestyle and job satisfaction.

**9) DistanceFromHome:** Distance of employee's residence from the workplace, influencing commuting stress and retention.

**10) EmpEducationLevel:** Level of education attained by the employee, reflecting qualifications and potential for advancement.

**11) EmpEnvironmentSatisfaction:** Employee satisfaction with the work environment, affecting morale and turnover.

**12) EmpHourlyRate:** Hourly wage of the employee, a factor in compensation satisfaction and retention.

**13) EmpJobInvolvement:** Level of involvement and engagement in the job role, affecting performance and attrition risk.

**14) EmpJobLevel:** Level of hierarchy within the organization, indicating seniority and career progression.

**15) EmpJobSatisfaction:** Satisfaction level with the job role, impacting employee morale and retention.

**16) NumCompaniesWorked:** Number of companies the employee has previously worked for, indicating job stability and turnover risk.

**17) OverTime:** Whether the employee works overtime, influencing work-life balance and burnout.

**18) EmpLastSalaryHikePercent:** Percentage of the employee's last salary hike, affecting compensation satisfaction and retention.

**19) EmpRelationshipSatisfaction:** Satisfaction with relationships at work, influencing job satisfaction and likelihood of turnover.

**20) TotalWorkExperienceInYears:** Total work experience of the employee, influencing skill level and career trajectory.

**21) TrainingTimesLastYear:** Number of training sessions attended by the employee last year, indicating investment in skill development and career growth.

**22) EmpWorkLifeBalance:** Employee's perceived balance between work and personal life, affecting job satisfaction and retention.

**23) ExperienceYearsAtThisCompany:** Years of experience at the current company, indicating loyalty and potential for promotion.

**24) ExperienceYearsInCurrentRole:** Years of experience in the current job role, influencing expertise and potential for advancement.

**25) YearsSinceLastPromotion:** Time since the employee's last promotion, impacting career progression and job satisfaction.

**26) YearsWithCurrManager**: Years of tenure with the current manager, affecting job satisfaction and retention.

**27) Attrition:**   indicates whether the employee has left the company or not.

28) PerformanceRating:  Target variable for the given problem.  this is the performance rating assigned to the employee, influencing career development and potential for retention.

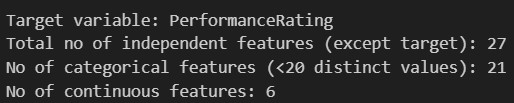
# **SUMMARY OF IMPORTANT ASPECTS OF THE MODEL**

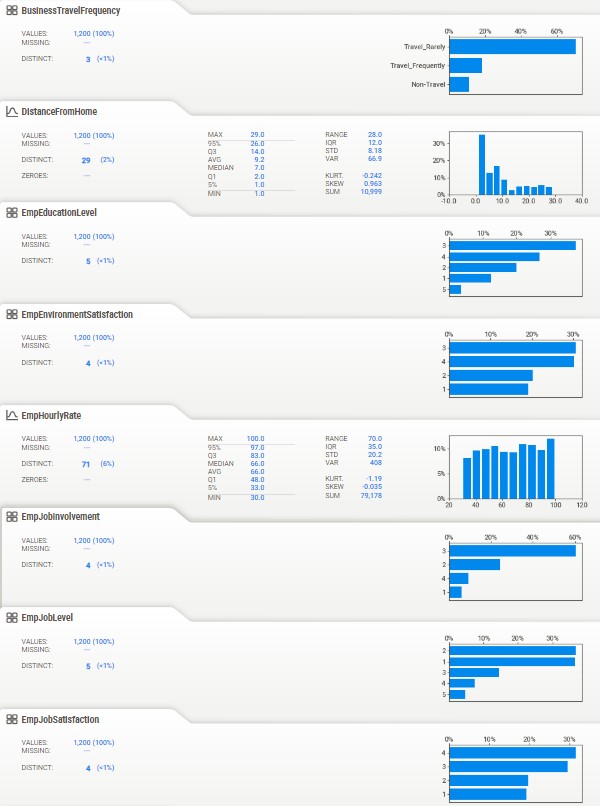
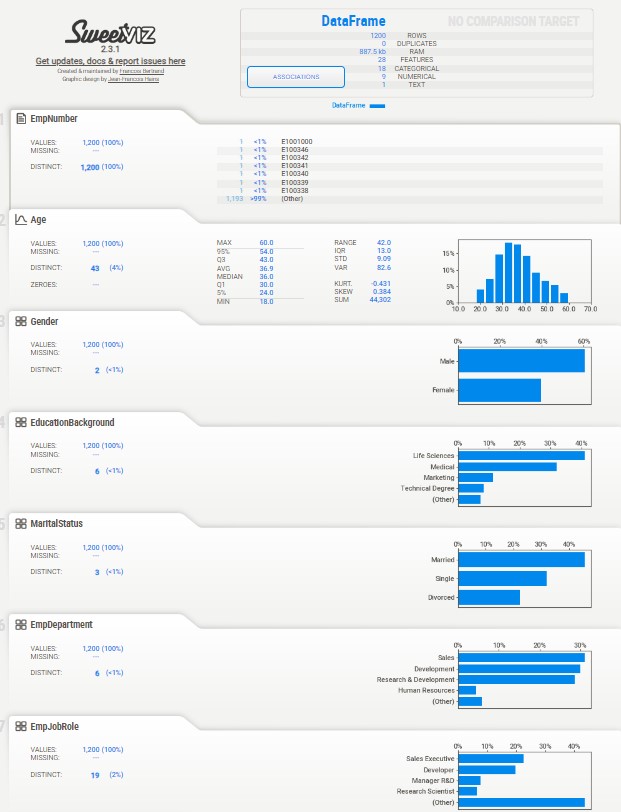
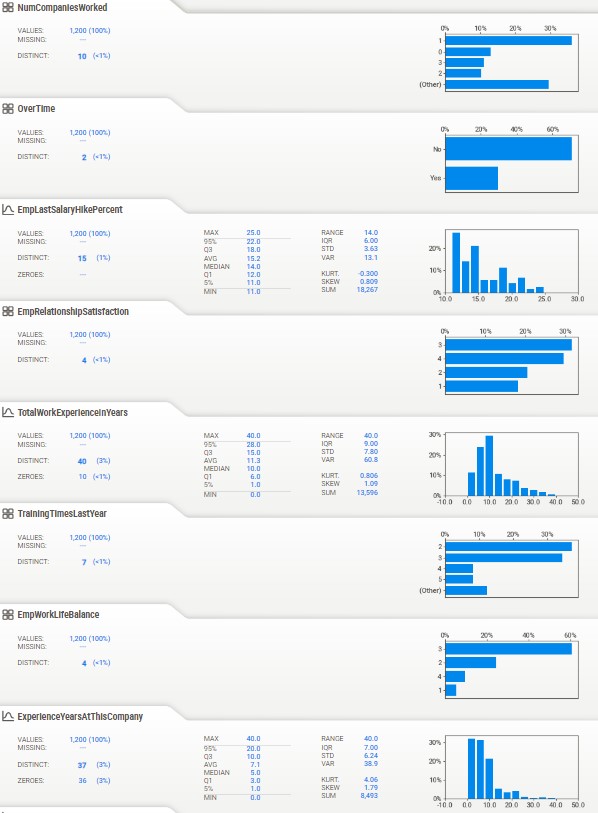
## **DATA ANALYSIS**

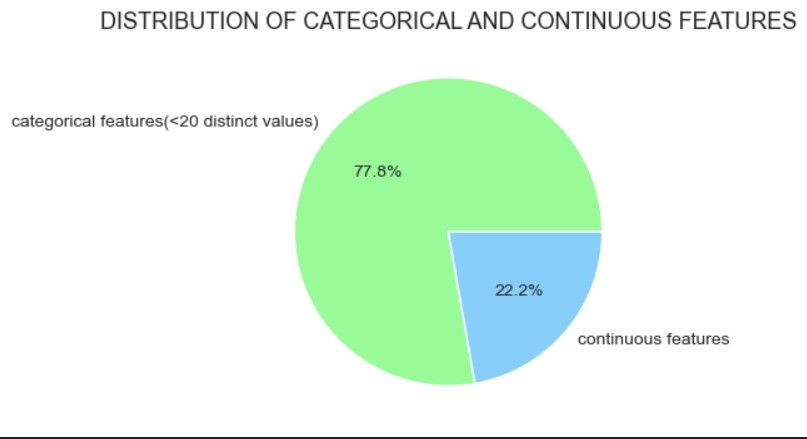
### **VISUALIZATION**

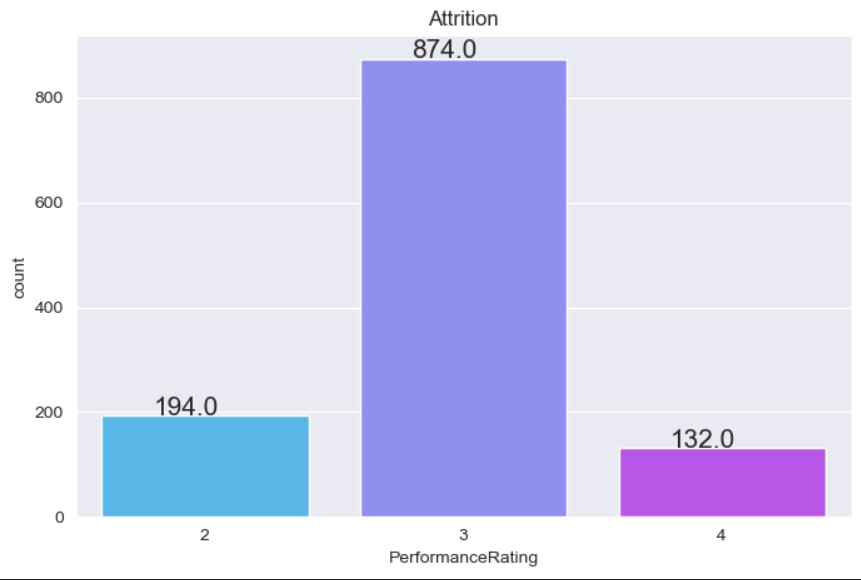
Used visualizations such as **histograms, count plots, scatter plots, pie charts, box plots, line plots, cat plots, pair plot and heatmaps** to effectively communicate analysis findings. Visual representations of data can make complex patterns and relationships more accessible and understandable.

#### **UNIVARIATE ANALYSIS:**

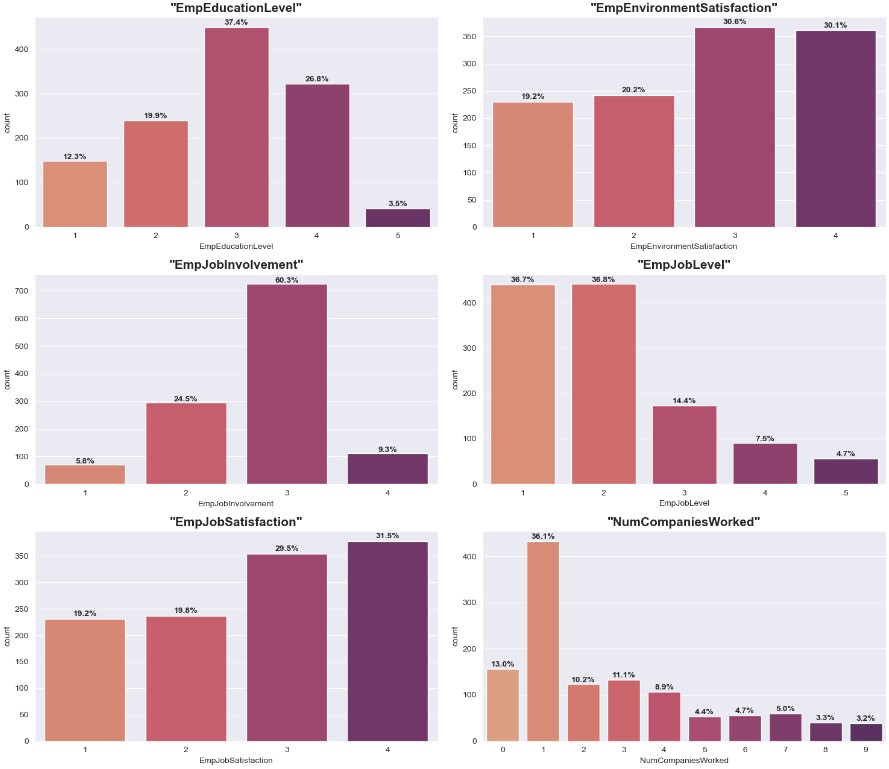


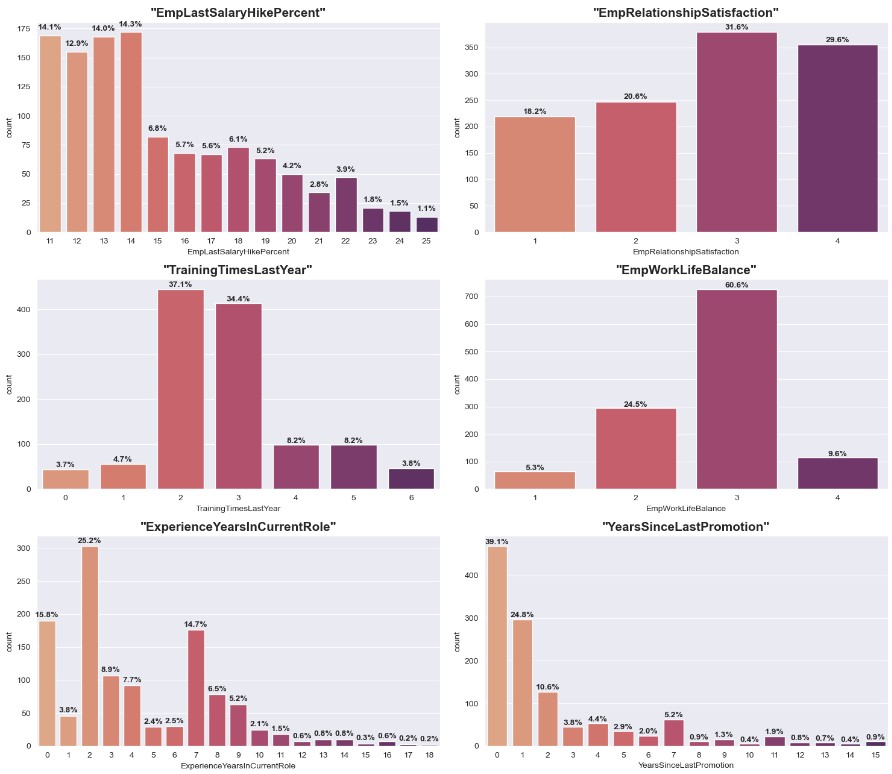






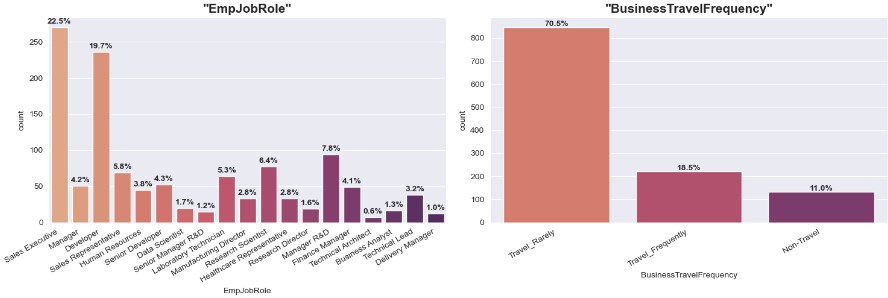
**Discrete features:**

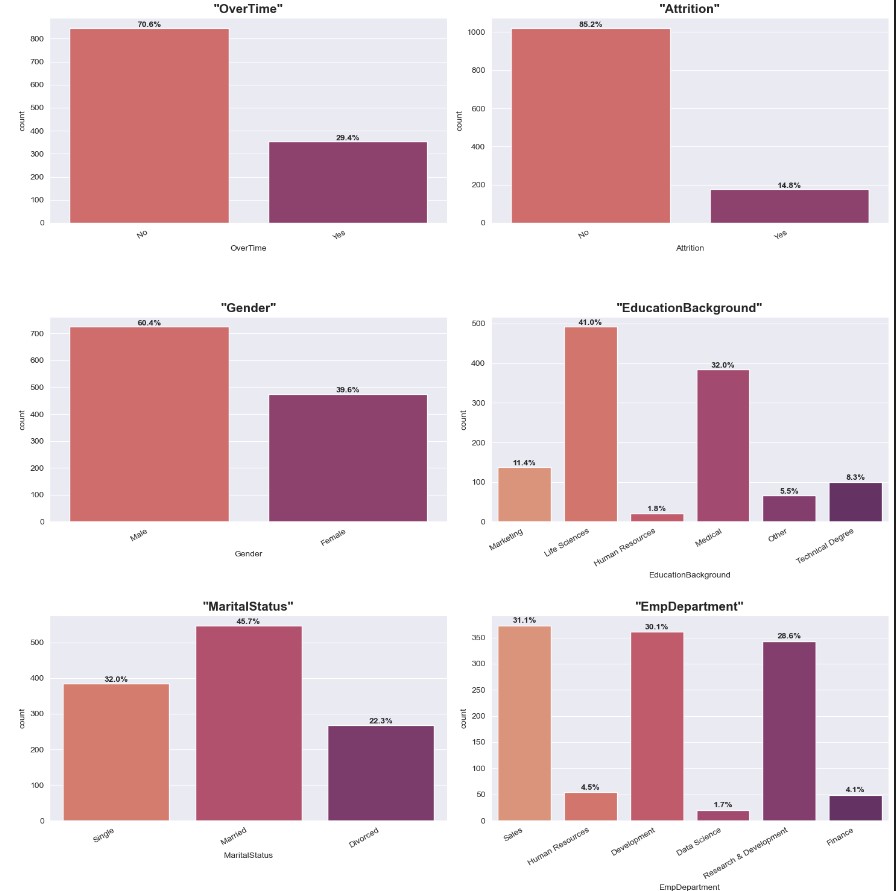




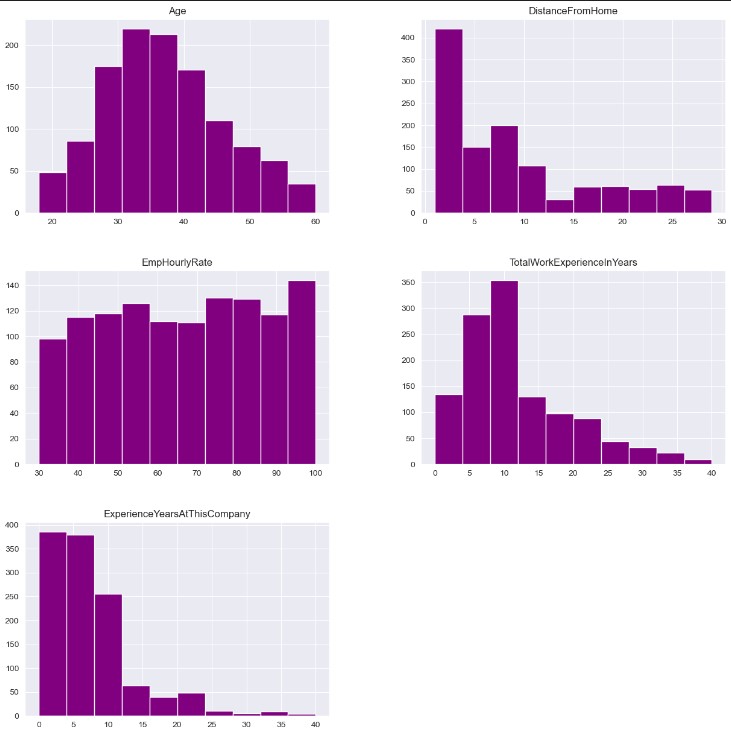


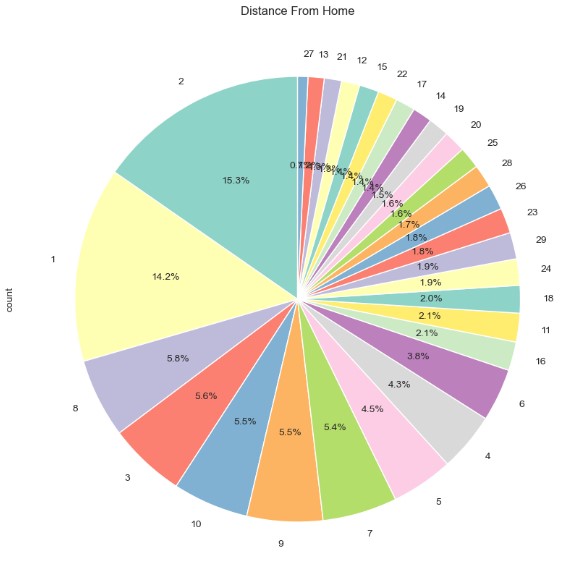
**Binary and nominal features:**





**Continuous features:**



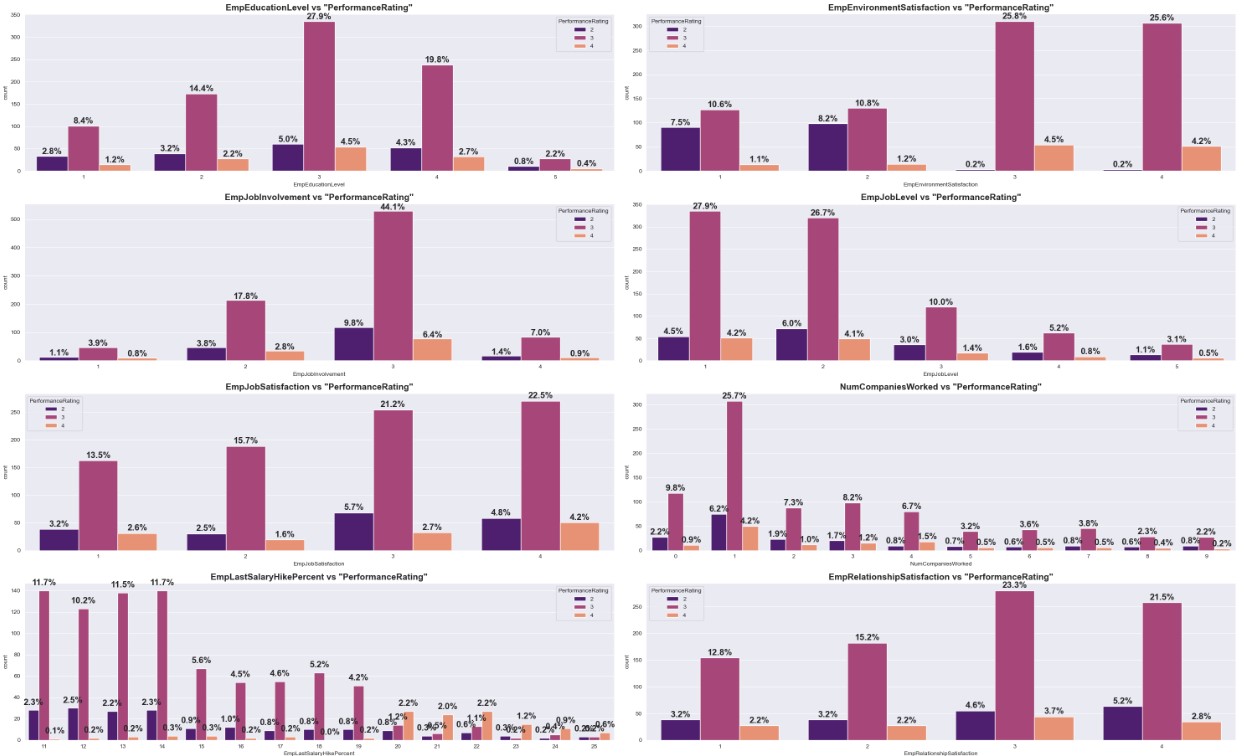


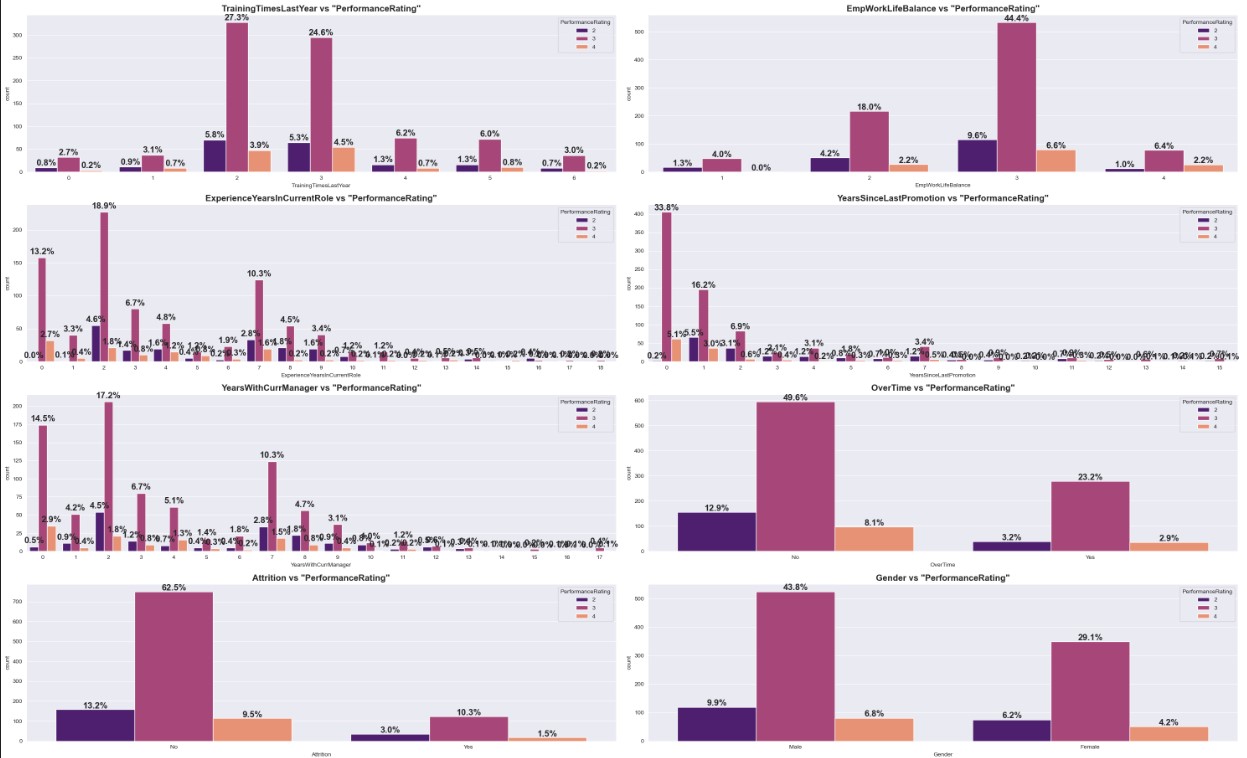
**Insights from univariate analysis:**

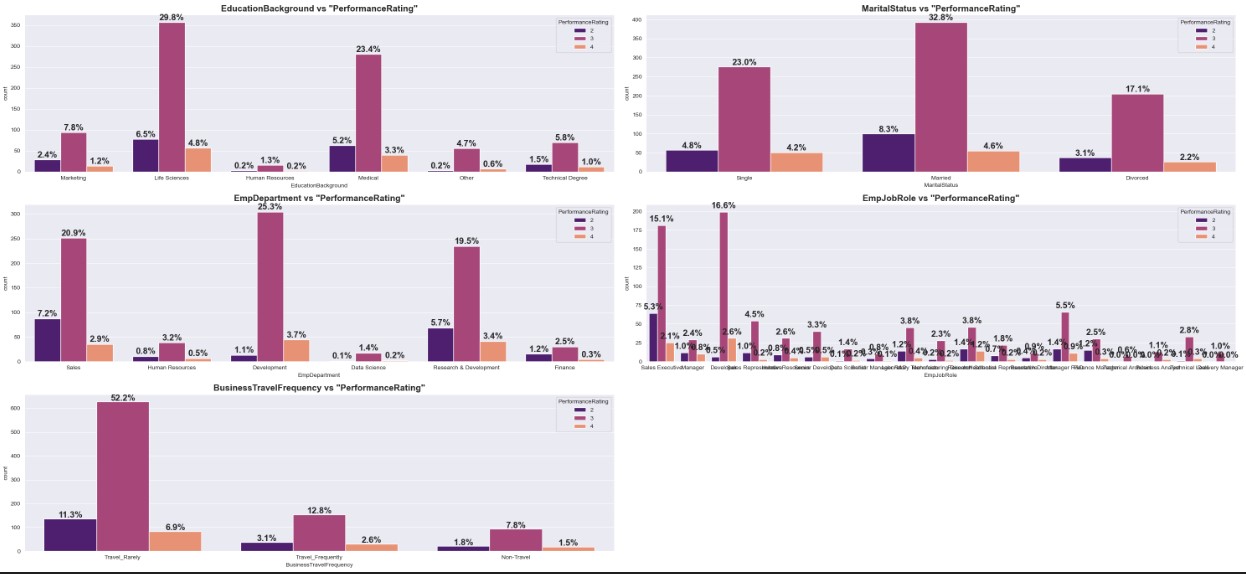
* Employee between the age group 30-35 is the majority.
* 70% of the people travel rarely, 20% travel frequently rest do not travel.
* Majority of the employees belong to research and development.
* Almost 35% of the people are nearer to the office i.e. the distance from their home is lesser than or equal to 10.
* More than 60% of the people have educational qualification of 2 and 4.
* Majority (40%) of the people are from life science field and 30% are from medical field.
* 60% of the people are almost satisfied with environment condition of the office with more than 3 and 4 ratings.
* Gender count: 60% male 40% female.
* Almost 40% of the people have partial involvement in job and 20% have good involvement.
* More than 45% employees seem to be satisfied with their job.
* 50% of the people are married, 30% single and the rest are divorced.
* 40% of the employee have worked experience less than 10 year
* 15% of the people have worked for less than 1 company which implies they are freshers.
* 30% of the people have worked for more than 5 companies.
* 80% of the people have average work rating.

#### **BIVARIATE ANALYSIS:**

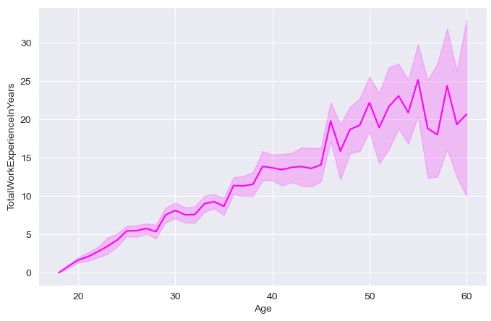
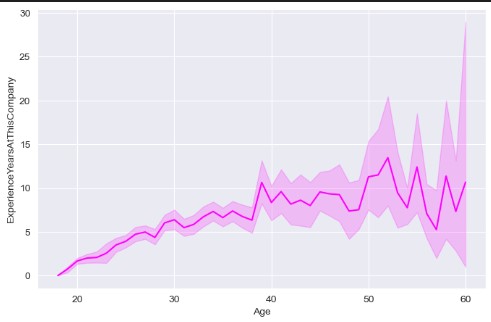
**Categorical features Vs target variable:**

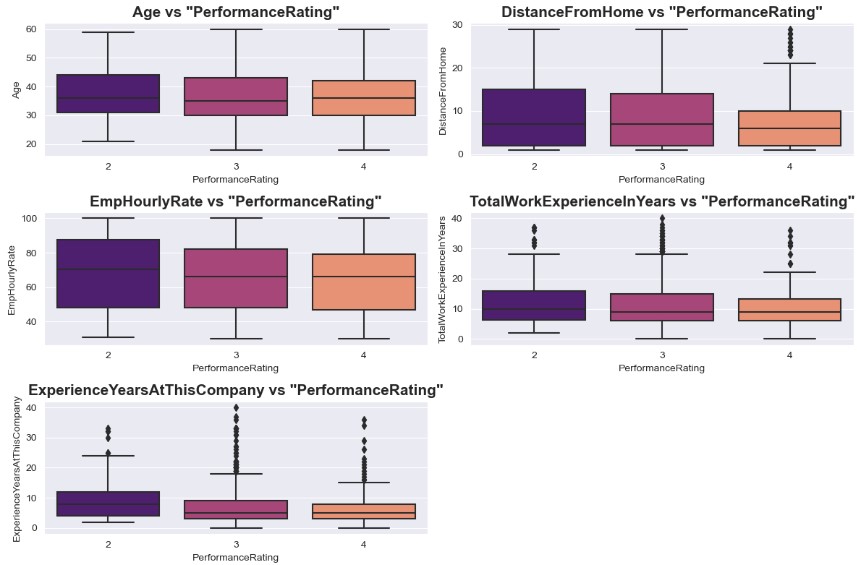


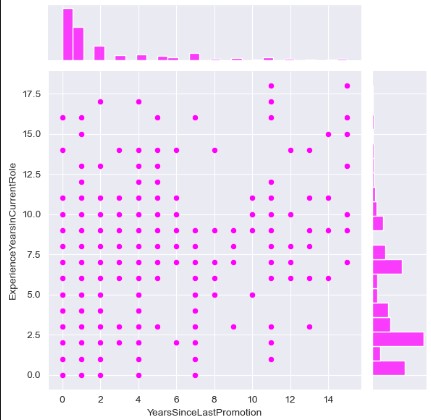
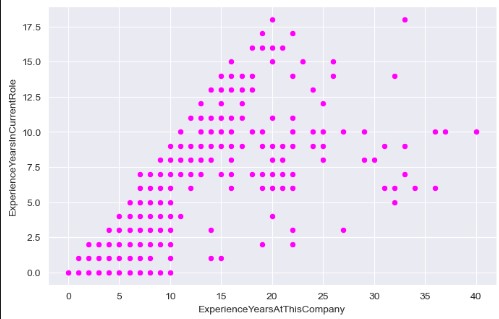


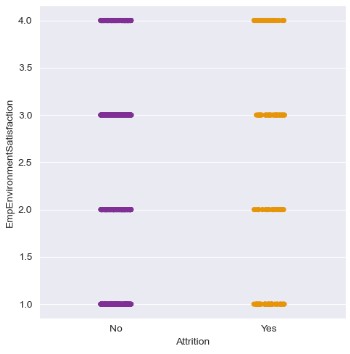


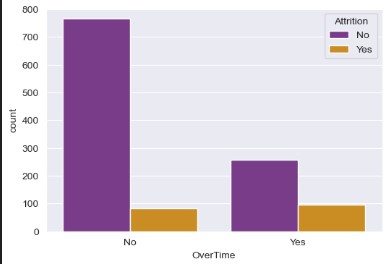
**Continuous features Vs target variable:**









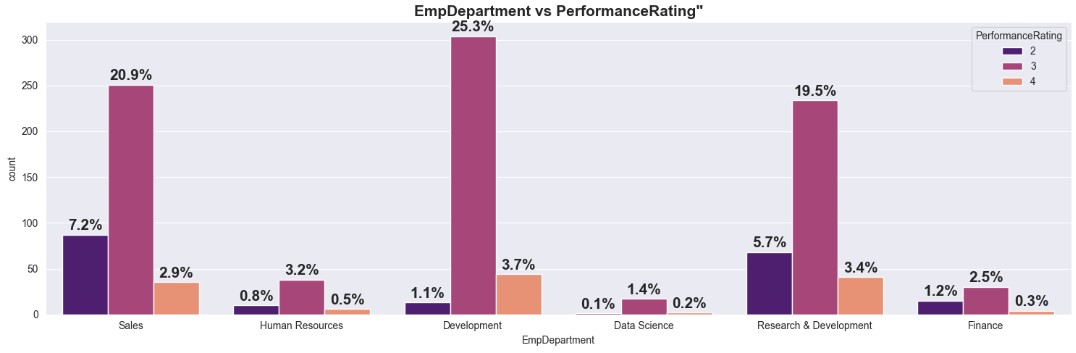


**Insights from bivariate analysis:**

* One common insight from all the above graphs seems to be that the "3" category of Performance Rating dominates in all the categorical input features.
* This represents the previous insights as similar to that of univariate analysis and when comparing the categories of each input feature, most of the employees in the dominating category have moderate performance rating of 3.
* For instance, in the **Employees education level**, the dominating category is 3rd level. Among them, most of the employees have shown moderate performance rating than the rest of the categories.
* However, the trend is different in few features.
* Even though the **employees have work environment satisfaction of 4th level, they show less performance rating** compared to the other categories. This goes in a similar way to that of employee’s job satisfaction.
* Since **the target variable is imbalanced**, this visualization may not reveal the true trend of the data distribution.
* Age does not significantly affect the performance rating as shown in the boxplot.
* Employees with travel **distance of more than 12 miles** are more probably low performing employees. Greater the distance from Home, lower the performance rating of the employees.
* **Hourly rate** of employees shows **significant different trend** with respect to the performance.
* **Total work experience** as well as **Experience at this company** shows similar trend as that of hourly rate.
* **Total work experience:**  Less than 15 years experienced employees have high performance rating compared to employees with >15 years of experience.
* **Experience years at this company**: experience of >8 years have poor performance rating whereas employees with **< 8 years of experience at this company** show high performance rating.
* This work experience trends could be due to their increasing age along with experience, **but due to the presence of many outliers**, these features might not capture the underlying pattern of the data.
* People who travel more are more expected to leave the job.
* People who do not do overtime less likely to leave the job.
* People from job roles of sales executive, developer, sales representative, laboratory technician, research scientist and manager of R&D are more likely to leave the job.

#### **DEPARTMENT WISE PERFORMANCE ANALYSIS:**

The analysis assesses the performance of employees across different departments to identify any disparities or trends.



**Insights:**

* Development department dominates the other departments in number.
* In terms of performance rating, employees who belong to development department perform well with the dominating 25% of rating 2.
* Hence departments such as **development, sales and Research & development** constitute the high performing employees.

### **FEATURE SELECTION/ FEATURE ENGINEERING:**

Feature selection techniques determine the relative importance of different features in predicting employee performance.

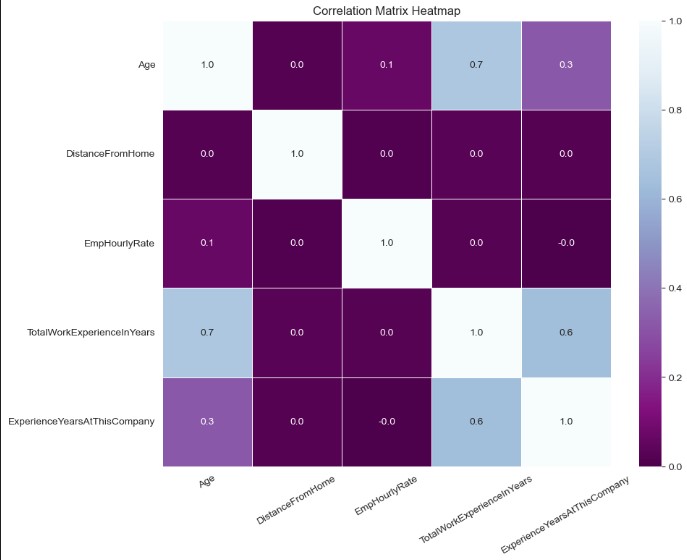
These techniques:

1. Ensure that visual analysis results are rigorously validated through appropriate methodologies and techniques.
2. Continuously monitor and evaluate analysis outcomes to ensure consistency and validity over time.

#### **THE TECHNIQUES USED IN THIS PROJECT ARE:**

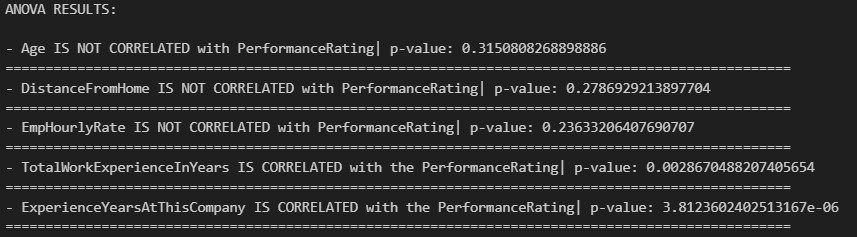
#### **Correlation analysis:**

Pearson’s correlation coefficient is calculated to find out the correlation coefficients between two continuous variables.



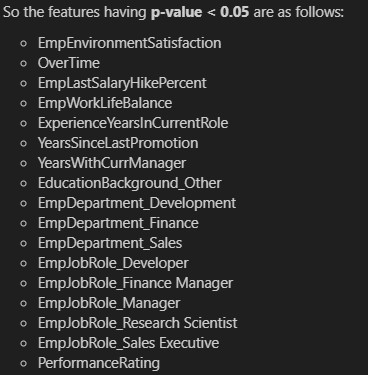
#### **ANOVA Test:**

ANOVA test is performed between categorical and continuous variable to find out the statistical significance of the features.



#### **CHI SQUARE Test:**

Chi square test is performed between two categorical features to assess the feature relevance for the predictions.



#### **INSIGHTS:**