

UNIFIED MENTOR
INTERNSHIP PROJECT

| | |
|---------------------------------|--|
| PROJECT TITLE | GREEN DESTINATION |
| TOOLS USED | PYTHON, TABLEAU, EXCEL, STATSISTICS |
| TECHNOLOGIES USED | DATA ANALYST, DATA SCIENTIST |
| PROJECT DIFFICULTY LEVEL | INTERMEDIATE LEVEL |

ABOUT THE DATASET

The dataset provided for this project contains 35 columns and 1,471 rows, detailing the employees of Green Destination, a well-known travel agency. This dataset includes important factors such as age, years at the company, and income, which are crucial in identifying patterns and trends related to employee attrition—whether an employee is likely to leave the company or not. The goal is to analyse these variables and determine the key factors influencing attrition. The dataset will be cleaned using Python, focusing on removing inconsistencies and ensuring data quality, and data visualization will be done using Matplotlib, a popular Python library, to uncover trends and insights. This project will offer a comprehensive understanding of employee behaviour and attrition patterns at Green Destination.

Some Data Cleaning Steps

1. Identifying and Handling Null Values
2. Removing Duplicates
3. Dropping Unneeded Columns
4. Data Type Conversion if needed
5. Outlier Detection and Treatment
6. Standardizing Categorical Values

STEP 1 : IMPORT LIBRARIES

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

STEP 2 : IMPORT THE DATASET

```
data= pd.read_csv('greendestination (1) (1).csv')
```

```
data.head(2)
```

STEP 3 : DATA CLEANING

```
# sum of null values
```

```
missing_data = data.isnull().sum()
```

```
print(data.isnull().sum())
```

```
# drop the duplicates
```

```
data.drop_duplicates(inplace=True)
```

```
# Dropping unnecessary columns
```

```
columns_to_drop = [ 'Over18', 'EmployeeCount', 'StandardHours']
```

```
# Dropping the columns
```

```
data = data.drop(columns=columns_to_drop)
```

```
# Displaying the remaining columns
```

```
print(data.columns)
```

STEP 4: DESCRIBE THE DATA SET

```
# show the names of all columns included
```

```
data.columns
```

```
Index(['Age', 'Attrition', 'BusinessTravel', 'DailyRate', 'Department',  
       'DistanceFromHome', 'Education', 'EducationField', 'EmployeeNumber',  
       'EnvironmentSatisfaction', 'Gender', 'HourlyRate', 'JobInvolvement',  
       'JobLevel', 'JobRole', 'JobSatisfaction', 'MaritalStatus',  
       'MonthlyIncome', 'MonthlyRate', 'NumCompaniesWorked', 'OverTime',  
       'PercentSalaryHike', 'PerformanceRating', 'RelationshipSatisfaction',
```

```
'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',  
'WorkLifeBalance', 'YearsAtCompany', 'YearsInCurrentRole',  
'YearsSinceLastPromotion', 'YearsWithCurrManager'],  
dtype='object')
```

```
# show start ,end , sep
```

```
data.index
```

```
# show number of rows and columns
```

```
data.shape
```

data.info()

Data columns (total 32 columns):

| # | Column | Non-Null Count | Dtype |
|-----|--------------------------|----------------|--------|
| --- | ----- | ----- | ----- |
| 0 | Age | 1470 non-null | int64 |
| 1 | Attrition | 1470 non-null | object |
| 2 | BusinessTravel | 1470 non-null | object |
| 3 | DailyRate | 1470 non-null | int64 |
| 4 | Department | 1470 non-null | object |
| 5 | DistanceFromHome | 1470 non-null | int64 |
| 6 | Education | 1470 non-null | int64 |
| 7 | EducationField | 1470 non-null | object |
| 8 | EmployeeNumber | 1470 non-null | int64 |
| 9 | EnvironmentSatisfaction | 1470 non-null | int64 |
| 10 | Gender | 1470 non-null | object |
| 11 | HourlyRate | 1470 non-null | int64 |
| 12 | JobInvolvement | 1470 non-null | int64 |
| 13 | JobLevel | 1470 non-null | int64 |
| 14 | JobRole | 1470 non-null | object |
| 15 | JobSatisfaction | 1470 non-null | int64 |
| 16 | MaritalStatus | 1470 non-null | object |
| 17 | MonthlyIncome | 1470 non-null | int64 |
| 18 | MonthlyRate | 1470 non-null | int64 |
| 19 | NumCompaniesWorked | 1470 non-null | int64 |
| 20 | OverTime | 1470 non-null | object |
| 21 | PercentSalaryHike | 1470 non-null | int64 |
| 22 | PerformanceRating | 1470 non-null | int64 |
| 23 | RelationshipSatisfaction | 1470 non-null | int64 |
| 24 | StockOptionLevel | 1470 non-null | int64 |
| 25 | TotalWorkingYears | 1470 non-null | int64 |
| 26 | TrainingTimesLastYear | 1470 non-null | int64 |
| 27 | WorkLifeBalance | 1470 non-null | int64 |
| 28 | YearsAtCompany | 1470 non-null | int64 |
| 29 | YearsInCurrentRole | 1470 non-null | int64 |
| 30 | YearsSinceLastPromotion | 1470 non-null | int64 |
| 31 | YearsWithCurrManager | 1470 non-null | int64 |

dtypes: int64(24), object(8)

memory usage: 367.6+ KB

`data.head()`

| | Age | Attrition | Department | DistanceFromHome | EmployeeNumber | EnvironmentSatisfaction | Gender | JobSatisfaction | MaritalStatus | MonthlyIncome |
|---|-----|-----------|------------------------|------------------|----------------|-------------------------|--------|-----------------|---------------|---------------|
| 0 | 41 | Yes | Sales | 1 | 1 | 2 | Female | 4 | Single | 5993 |
| 1 | 49 | No | Research & Development | 8 | 2 | 3 | Male | 2 | Married | 5130 |
| 2 | 37 | Yes | Research & Development | 2 | 4 | 4 | Male | 3 | Single | 2090 |
| 3 | 33 | No | Research & Development | 3 | 5 | 4 | Female | 3 | Married | 2909 |
| 4 | 27 | No | Research & Development | 2 | 7 | 1 | Male | 2 | Married | 3468 |

This is how the table looks like

After data cleaning, the dataset now consists of 1,470 rows and 31 columns. Unnecessary columns have been dropped, and the data is indexed using a range index starting at 0, with a total of 1,470 rows and a step size of 1. The remaining columns, selected based on their relevance to identifying the reasons behind employee attrition, are listed under `data.columns.name`. We will now analyze the conclusions drawn through diagrams and visualizations.

STEP 5: ANALYSING THE SITUATION

Count the number of employees for each Attrition category (Yes/No)

```
attrition_counts = data['Attrition'].value_counts()
```

Create a pie chart for Attrition

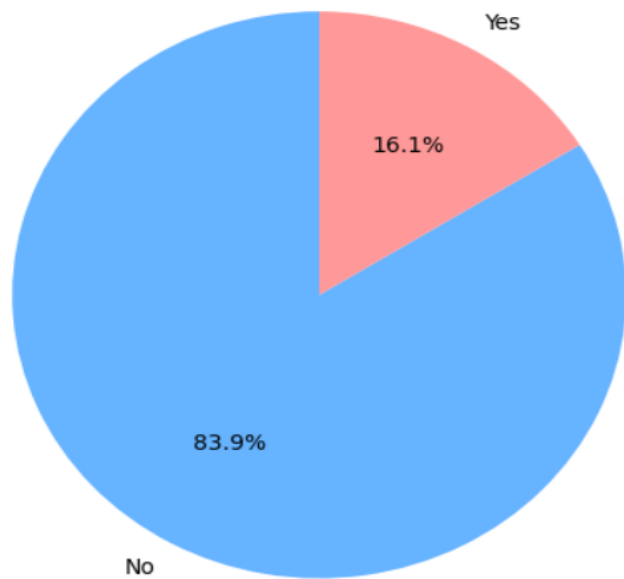
```
plt.figure(figsize=(6, 6))
```

```
plt.pie(attrition_counts, labels=attrition_counts.index, autopct='%1.1f%%',  
startangle=90, colors=['#66b3ff', '#ff9999'])
```

```
plt.title('Employee Attrition Distribution')
```

```
plt.show()
```

Employee Attrition Distribution



16.1% employees are leaving the company , this can be based on many factors like age , years worked in company , past promotion , job satisfaction , monthly income etc.

Scatter plot of Age vs YearsAtCompany colored by Attrition

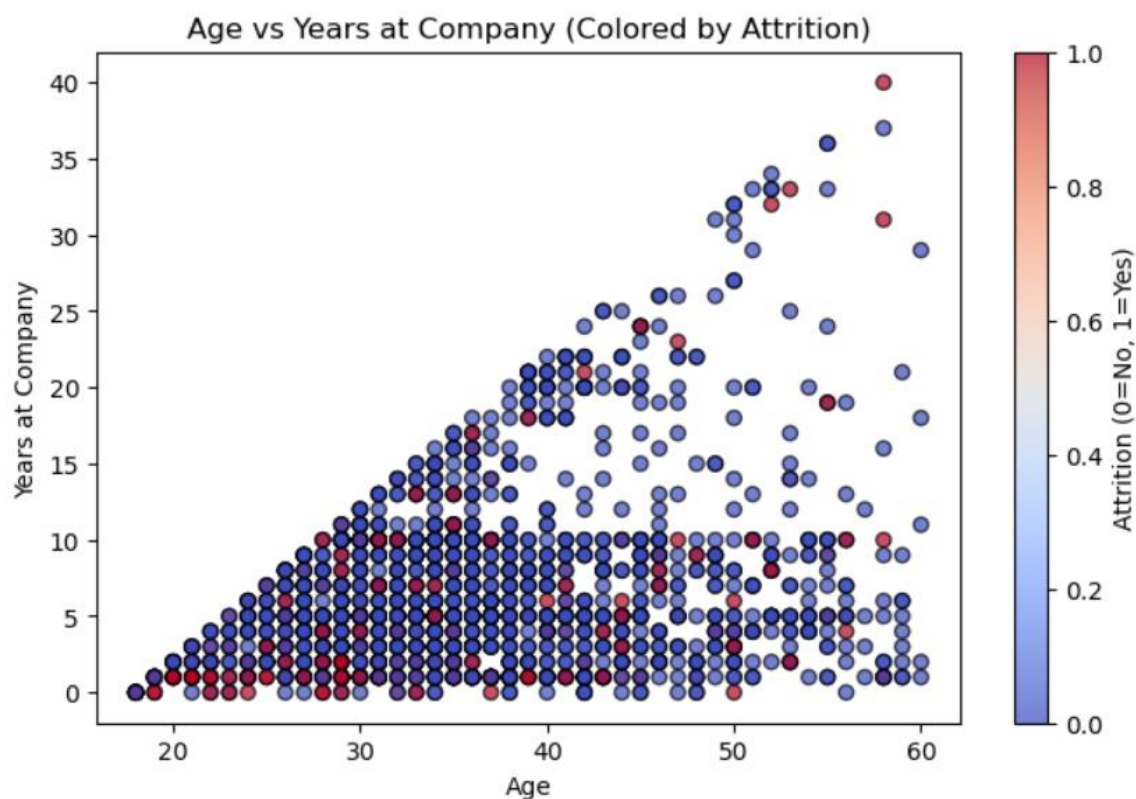
```
plt.figure(figsize=(8, 5))  
  
plt.scatter(data['Age'], data['YearsAtCompany'], c=data['Attrition'],  
            cmap='coolwarm', edgecolor='k', alpha=0.7)  
  
plt.xlabel('Age')  
plt.ylabel('Years at Company')  
plt.title('Age vs Years at Company (Colored by Attrition)')  
plt.colorbar(label='Attrition (0=No, 1=Yes)')  
plt.show()
```

Diagonal Pattern: Younger employees have fewer years at the company, while older employees (40-60) generally have more years of service.

Attrition Highlight: Red dots show higher attrition, mostly among younger employees with short tenure. Blue dots indicate lower attrition, especially for long-serving employees. Age & Tenure

Impact: Older employees with more years at the company (10-15+ years) show lower attrition rates. Attrition decreases as age and tenure increase.

Younger employees and those with fewer years at the company are more prone to leaving (higher attrition rates). Older employees and those with longer tenures are less likely to leave, indicating a possible trend of loyalty or stability among this group.



```
# Create the scatter plot
```

```
plt.figure(figsize=(10, 6))
```

```
plt.scatter(age, monthly_income, color='blue', alpha=0.6)
```

```
plt.title('Scatter Plot of Age vs Monthly Income', fontsize=16)
```

```
plt.xlabel('Age', fontsize=14)
```

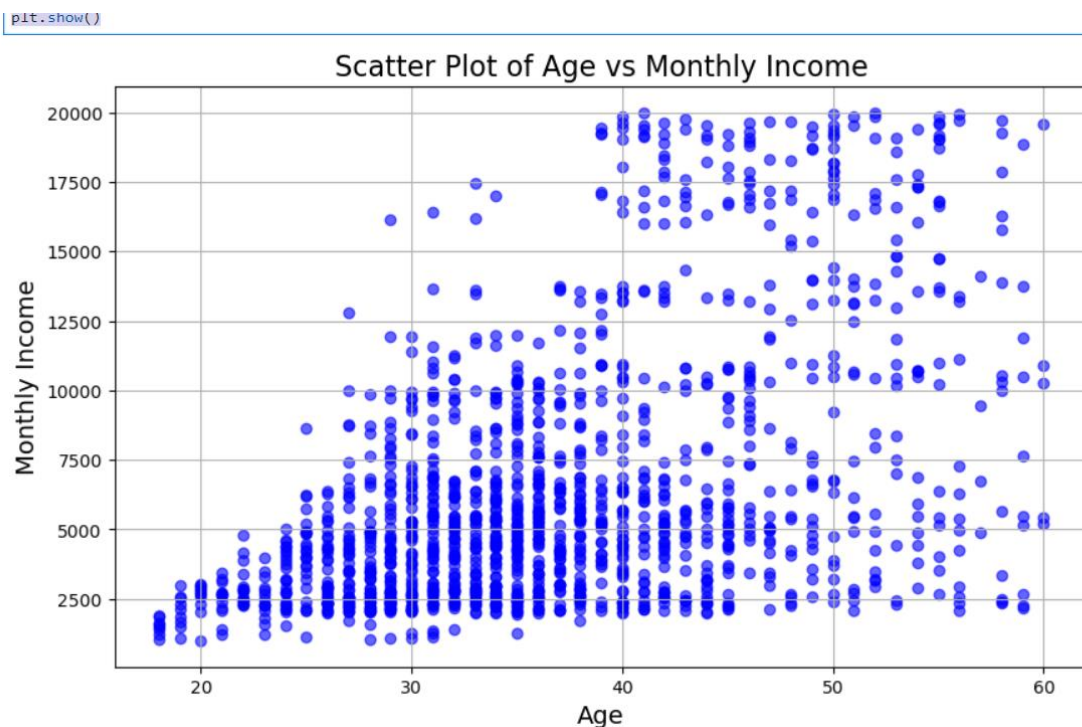


```
plt.ylabel('Monthly Income', fontsize=14)
```

```
plt.grid(True)
```

```
# Show the plot
```

```
plt.show()
```



The scatter plot in your image shows the relationship between **Age** (on the x-axis) and **Monthly Income** (on the y-axis). Here's what it suggests:

1. General Trend:

Monthly income generally increases with age, especially between ages 20 and 30, where the rise is more pronounced.

After 30, while the income continues to rise for some, the relationship becomes more varied, with a wider range of incomes for the same age group.

2. Income Distribution:

For younger employees (around ages 20-30), the income is mostly clustered between ₹2,500 and ₹10,000.

For older employees (ages 30-50), the income becomes more spread out, with some earning above ₹15,000, but there are also individuals with incomes lower than ₹5,000.

3. Outliers:

- There are a few points above ₹15,000 and ₹20,000, indicating higher earners, likely in senior roles or with specific skills, regardless of age.

In summary, while age correlates with income, particularly early in the career, there is significant variation as age increases.

```
import matplotlib.pyplot as plt
import seaborn as sns

# Box plot for Monthly Income and Attrition
plt.figure(figsize=(8,6))
sns.boxplot(x='Attrition', y='MonthlyIncome', data=data)

plt.title('Monthly Income vs Attrition')
plt.xlabel('Attrition (0=No, 1=Yes)')
plt.ylabel('Monthly Income')
plt.show()
```

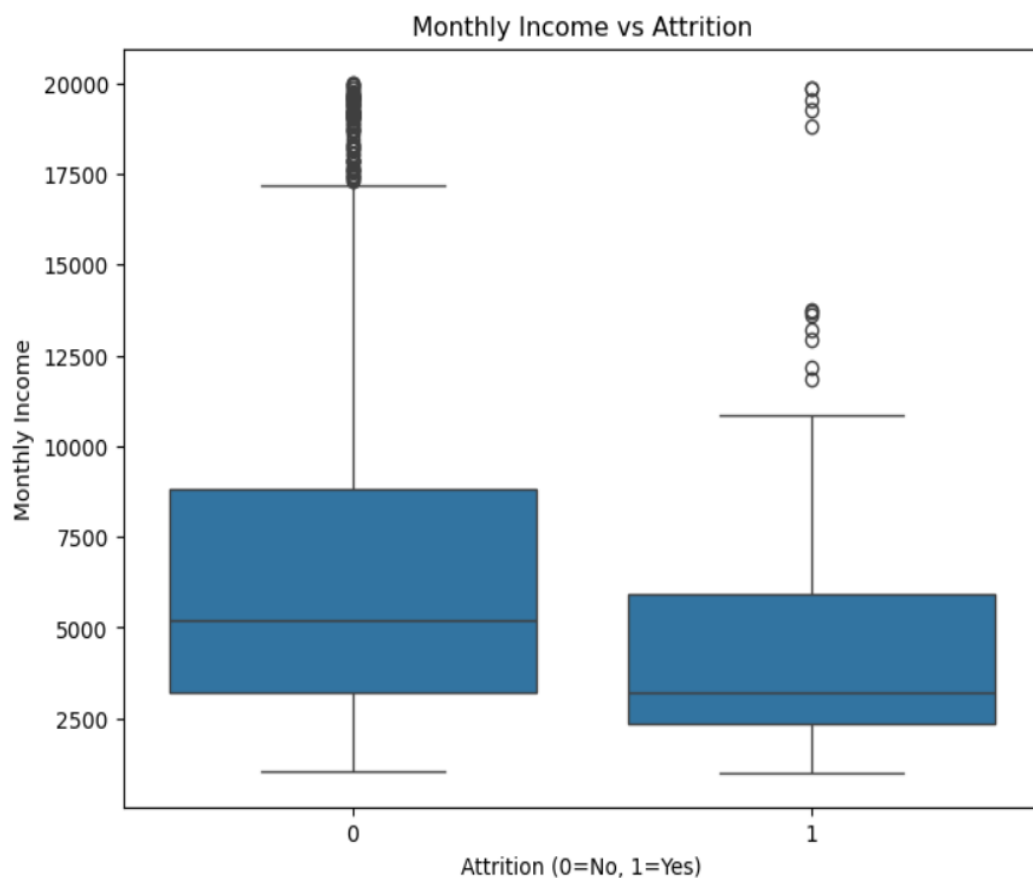
This box plot shows the relationship between "Monthly Income" and "Attrition" (whether an employee has left the company, with 0 meaning 'No' and 1 meaning 'Yes'). Here's what the plot illustrates:

Attrition = 0 (No): Employees who did not leave the company tend to have higher monthly incomes, with a median income around 5000-6000. The box is larger, indicating more variation in the monthly incomes of these employees, with some extreme outliers having incomes around 20,000.

Attrition = 1 (Yes): Employees who left the company generally have lower monthly incomes, with a median income below 5000. There are a few outliers earning much more than the typical leaver, but most employees who left had lower incomes compared to those who stayed.

Conclusion:

The graph suggests that employees with lower monthly incomes are more likely to leave the company, whereas those with higher incomes tend to stay.



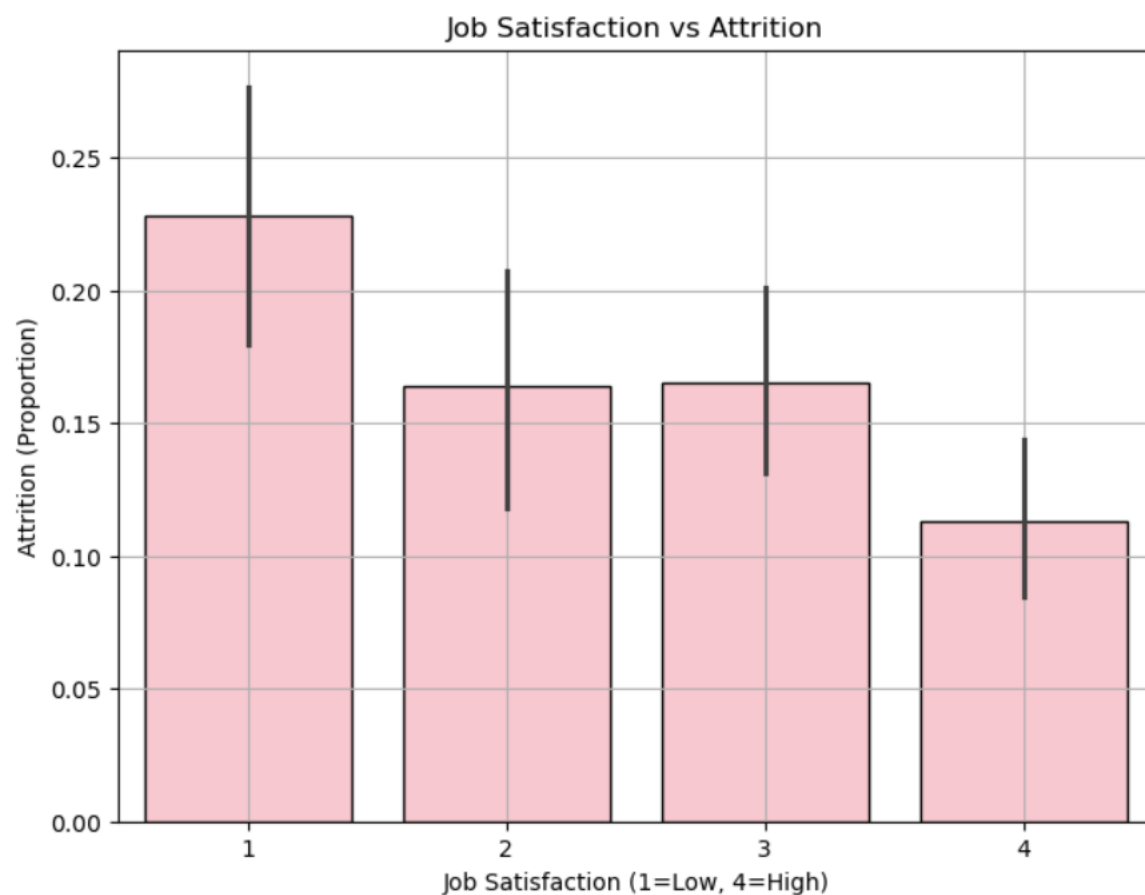
Create a bar plot to show the relationship between Job Satisfaction and Attrition

```
plt.figure(figsize=(8, 6))
```

```
sns.barplot(x='JobSatisfaction', y='Attrition', data=data, color =  
'pink',edgecolor='black')
```

```
# Set plot labels and title
plt.title('Job Satisfaction vs Attrition')
plt.xlabel('Job Satisfaction (1=Low, 4=High)')
plt.ylabel('Attrition (Proportion)')
plt.grid(True)

# Show the plot
plt.show()
```



JobSatisfaction: A scale from 1 (low satisfaction) to 4 (high satisfaction).

Attrition: 0 means the employee stayed, and 1 means the employee left the company.

The chart shows the proportion of employees who left (Attrition = 1) for each level of Job Satisfaction.

```
# Set the figure size
```

```
# Find the top 10 employees by MonthlyIncome
```

```
top_10 = data.nlargest(10, 'MonthlyIncome')
```

```
# Find the bottom 10 employees by MonthlyIncome
```

```
bottom_10 = data.nsmallest(10, 'MonthlyIncome')
```

```
# Plotting the combined bar chart
```

```
plt.figure(figsize=(12, 6))
```

```
# Top 10
```

```
plt.bar(top_10['EmployeeNumber'].astype(str), top_10['MonthlyIncome'],  
color='skyblue', label='Top 10 Employees')
```

```
# Bottom 10
```

```
plt.bar(bottom_10['EmployeeNumber'].astype(str),  
bottom_10['MonthlyIncome'], color='salmon', label='Bottom 10 Employees')
```

```
plt.title('Top 10 and Bottom 10 Employees by Monthly Income')
```

```
plt.xlabel('Employee Number')
```

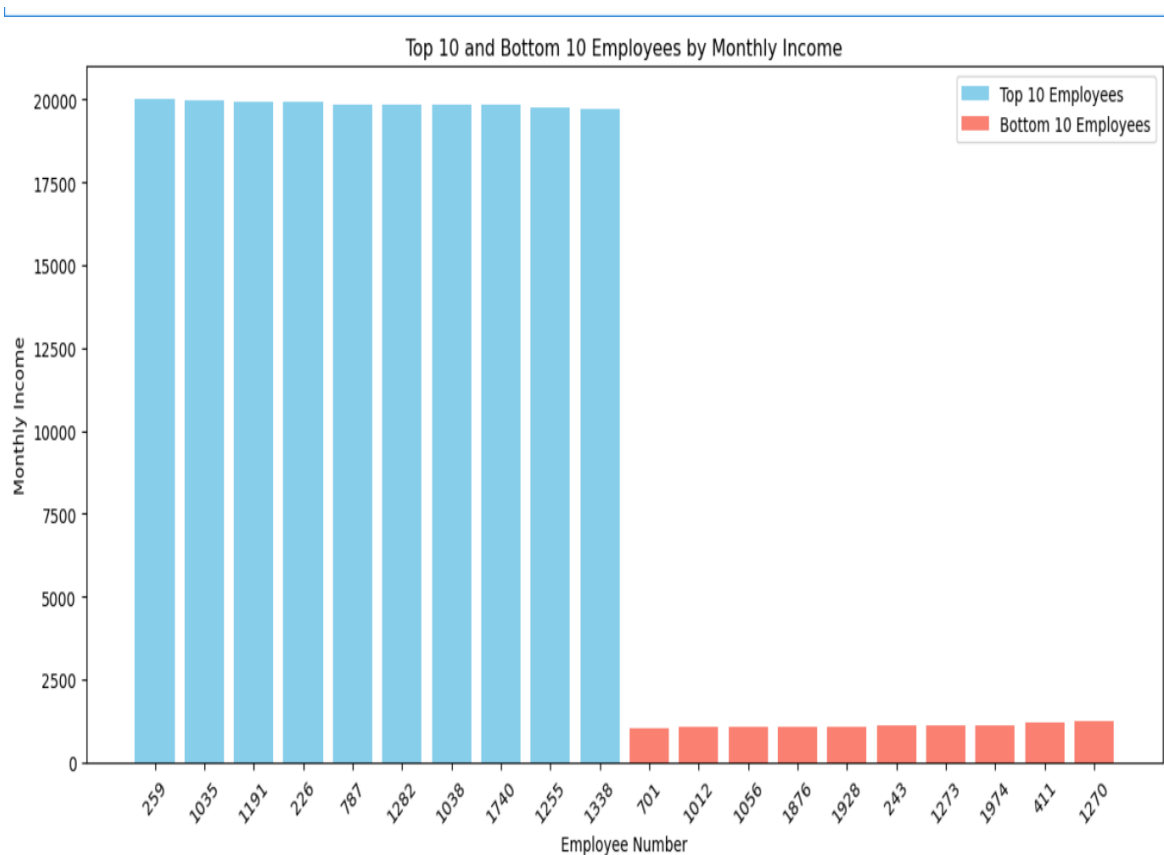
```
plt.ylabel('Monthly Income')
```

```
plt.xticks(rotation=45)
```

```
plt.legend()
```

```
plt.tight_layout()
```

```
plt.show()
```

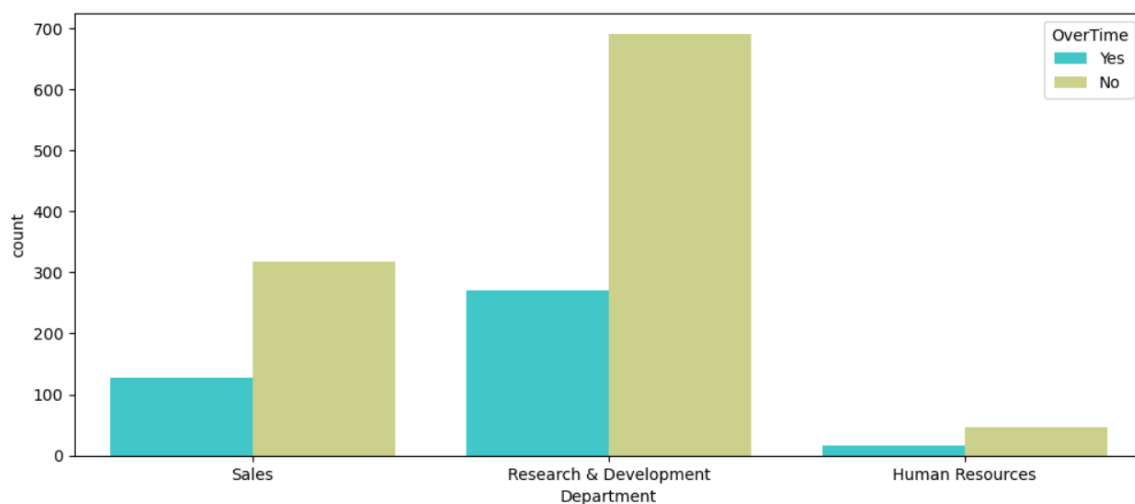


The diagram above highlights the top 10 and bottom 10 employees by monthly income. It clearly illustrates a significant income disparity between these two groups. This inequality is concerning because as the wage gap widens, employees with lower earnings are more likely to experience higher levels of attrition. Those in the bottom income bracket may leave the company due to inadequate compensation, feeling undervalued and underpaid. This growing disparity can negatively impact employee retention and overall company morale.

```
plt.figure(figsize=(12,5)) # width =12, height=5
```

```
sns.countplot(x=data['Department'], hue=data['OverTime'], palette='rainbow')
```

```
plt.show()
```



```
# Calculate the percentage of employees with PercentSalaryHike greater than 15%
```

```
total_employees = len(data)
```

```
high_hike_employees = len(data[data['PercentSalaryHike'] > 15])
```

```
percentage_high_hike = (high_hike_employees / total_employees) * 100
```

```
# Plotting the percentage
```

```
labels = ['> 15%', '<= 15%']
```

```
sizes = [percentage_high_hike, 100 - percentage_high_hike]
```

```
colors = ['#ff9999', '#66b3ff']
```

```
explode = (0.1, 0) # explode 1st slice
```

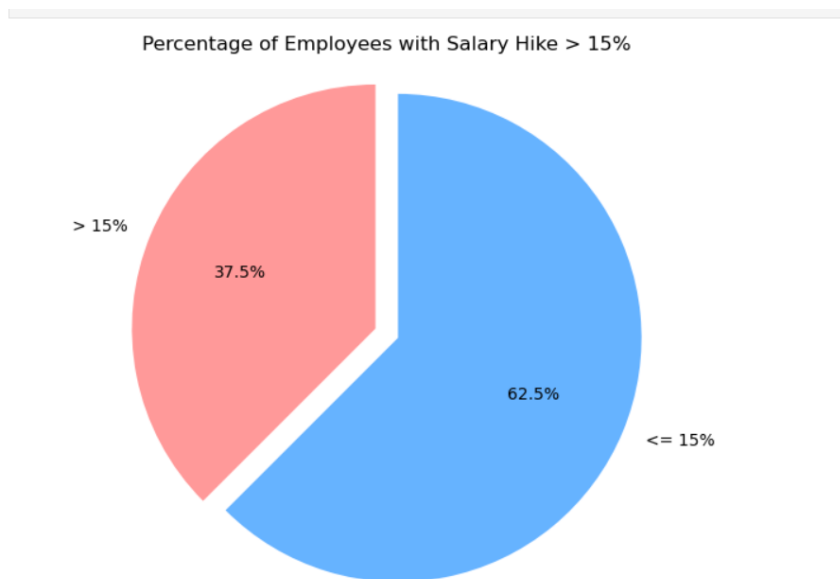
```
plt.figure(figsize=(8, 6))
```

```
plt.pie(sizes, explode=explode, labels=labels, colors=colors, autopct='%1.1f%%',  
startangle=90)
```

```
plt.axis('equal') # Equal aspect ratio ensures that pie chart is a circle.
```

```
plt.title('Percentage of Employees with Salary Hike > 15%')
```

```
plt.show()
```



The pie chart illustrates that only 37.5% of employees received a salary hike of more than 15%, while the majority did not see an increase of this magnitude. This limited wage growth can lead to dissatisfaction, especially for those who feel their efforts are not adequately rewarded. The disparity in salary hikes may create a sense of inequality within the organization, causing frustration among employees who perceive their compensation as insufficient. Over time, this frustration can result in higher turnover rates, as employees seek better financial incentives and recognition for their contributions in other companies. Addressing this imbalance through fair and consistent salary adjustments could help improve employee retention and morale.

Filtering employees who have worked more than 8 years and have not been promoted in the last x years

years_threshold = 2 # Set this to the desired number of years since the last promotion

long_term_employees = data[(data['YearsAtCompany'] > 8) & (data['YearsSinceLastPromotion'] > years_threshold)]

Calculate the total number of employees

total_employees = len(data)

Calculate the number of long-term employees without recent promotion


```
long_term_count = len(long_term_employees)
```

```
# Calculate the percentage
```

```
percentage_long_term = (long_term_count / total_employees) * 100
```

```
# Data for plotting
```

```
labels = ['> 8 Years & No Recent Promotion', '<= 8 Years or Recently Promoted']
```

```
sizes = [percentage_long_term, 100 - percentage_long_term]
```

```
colors = ['pink', 'orange']
```

```
explode = (0.1, 0) # explode 1st slice
```

```
# Plotting the percentage
```

```
plt.figure(figsize=(8, 6))
```

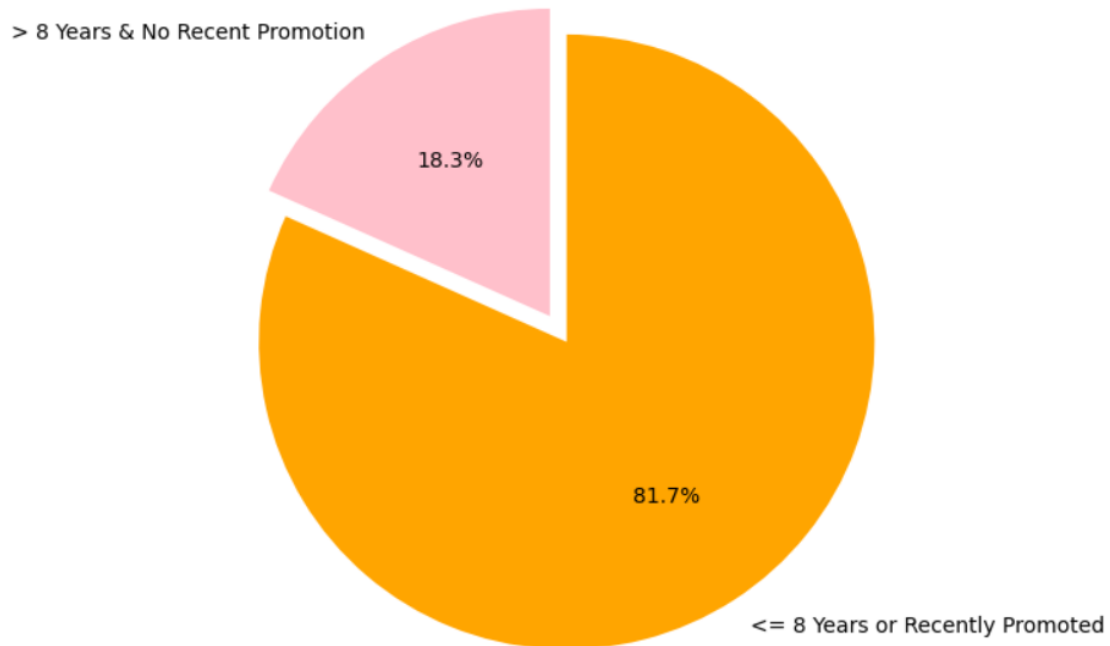
```
plt.pie(sizes, explode=explode, labels=labels, colors=colors, autopct='%1.1f%%',  
startangle=90)
```

```
plt.axis('equal') # Equal aspect ratio ensures that pie chart is a circle.
```

```
plt.title('Percentage of Employees Who Have Worked More Than 8 Years and  
No Recent Promotion')
```

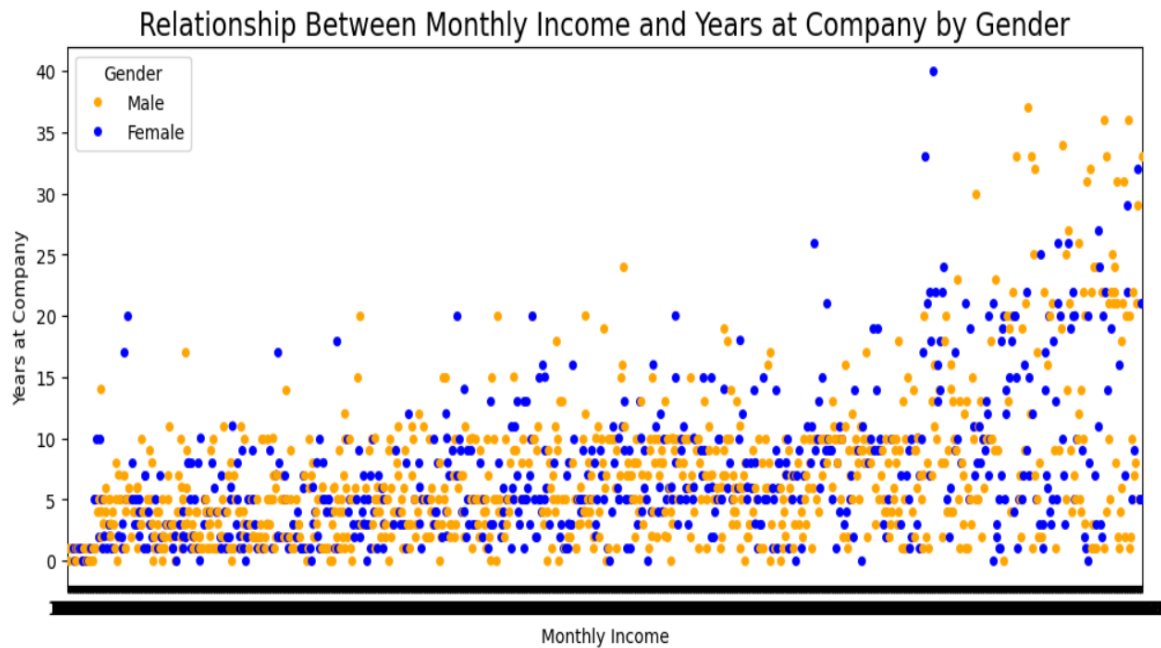
```
plt.show()
```

Percentage of Employees Who Have Worked More Than 8 Years and No Recent Promotion



The pie chart illustrates the distribution of employees based on their duration of service and recent promotion status. It shows that 18.3% of employees have worked for more than 8 years without a recent promotion, while 81.7% are either employees who have been with the company for 8 years or less, or have recently received a promotion.

```
plt.figure(figsize=(12, 5))  
  
sns.stripplot(data=data, x='MonthlyIncome', y='YearsAtCompany',  
hue='Gender', palette=['orange', 'blue'], jitter=True)  
  
plt.title("Relationship Between Monthly Income and Years at Company by  
Gender", fontsize=16)  
  
plt.xlabel("Monthly Income")  
  
plt.ylabel("Years at Company")  
  
plt.legend(title='Gender')  
  
plt.show()
```

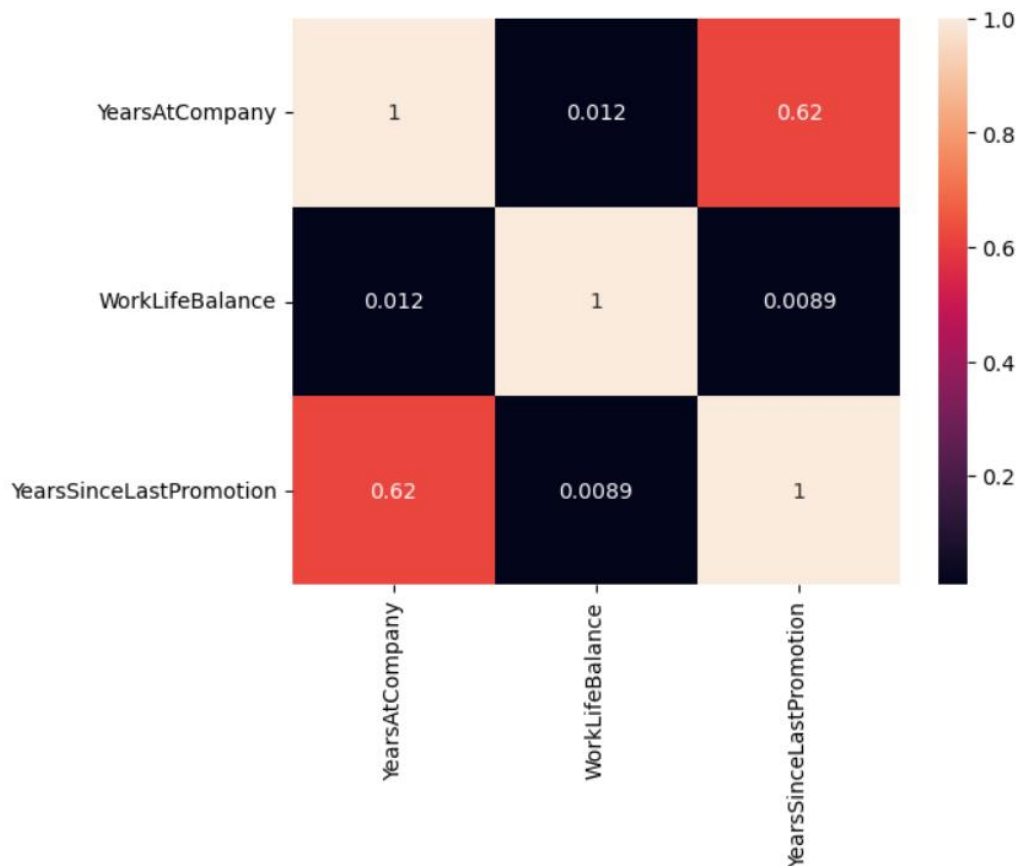


Insights: Income Distribution: Both Male and Female employees show a wide range of monthly incomes, with many clustered at lower income levels, though some earn significantly more. Tenure Variation: Employees with lower incomes tend to have shorter tenures (closer to the bottom of the plot), while those with longer tenures tend to earn higher incomes. Gender Representation: The distribution across genders is relatively even, though there may be slight variations in income and tenure patterns by gender.

```
corr =
data[['YearsAtCompany','WorkLifeBalance','YearsSinceLastPromotion']].corr()
corr
```

| | YearsAtCompany | WorkLifeBalance | YearsSinceLastPromotion |
|-------------------------|----------------|-----------------|-------------------------|
| YearsAtCompany | 1.000000 | 0.012089 | 0.618409 |
| WorkLifeBalance | 0.012089 | 1.000000 | 0.008941 |
| YearsSinceLastPromotion | 0.618409 | 0.008941 | 1.000000 |

```
sns.heatmap(corr,annot=True)
plt.show()
```



This heatmap shows the correlation matrix among three variables: YearsAtCompany, WorkLifeBalance, and YearsSinceLastPromotion. Each cell represents the correlation coefficient between two variables, with values ranging from -1 to 1:

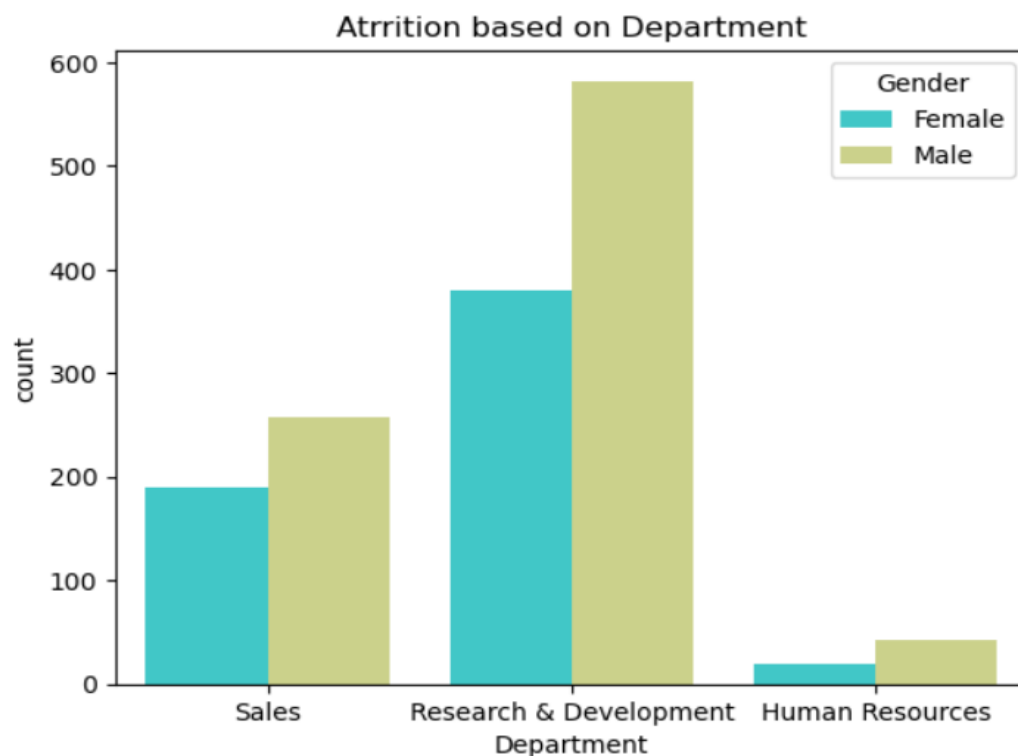
- YearsAtCompany and YearsSinceLastPromotion have a relatively strong positive correlation (0.62), suggesting that employees who have been with the company longer are more likely to have had more time since their last promotion.
- WorkLifeBalance has very low correlation values with both YearsAtCompany (0.012) and YearsSinceLastPromotion (0.0089), indicating no significant relationship between WorkLifeBalance and the other two variables.

The colour gradient represents the strength of correlation, with darker colours indicating lower correlation and lighter colours showing higher correlation.

```
sns.countplot(x=data['Department'],hue=data['Gender'],palette='rainbow')
plt.title('Attrition based on Department')
```

```
plt.show()
```

```
plt.show()
```



From the chart, we can observe:

In the Sales department, there is a slightly higher attrition rate among males compared to females.

In the Research & Development department, the attrition rate for males is significantly higher than for females.

The attrition in the Human Resources department is quite low for both genders, with a slightly higher count for males.

Overall, Research & Development shows the highest attrition, particularly among male employees.

EDA

```
def count_per_cat(col):
```

```
    vc = data[col].value_counts()
```

```
    plt.figure(figsize=(12,5)) # width= 12, height = 5
```

```

plt.subplot(1,2,1) # row=1,cols= 2, chart_num = 1

ax =
sns.countplot(x=data[col],order=data[col].value_counts().sort_values(ascending=False).index)

ax.bar_label(ax.containers[0])

plt.xticks(rotation=75)

plt.title(f'Countplot for {col}')

plt.subplot(1,2,2) # row=1,cols= 2, chart_num = 2

plt.pie(vc.values,labels=vc.index,autopct='%.2f%%')

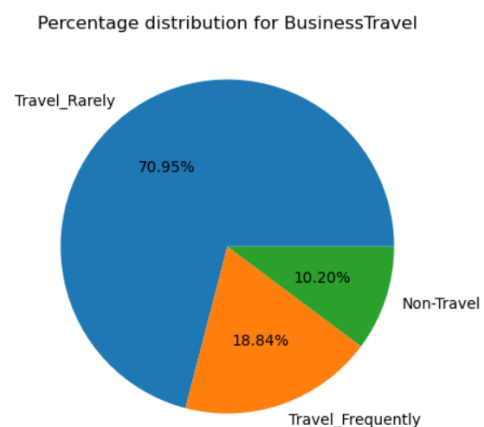
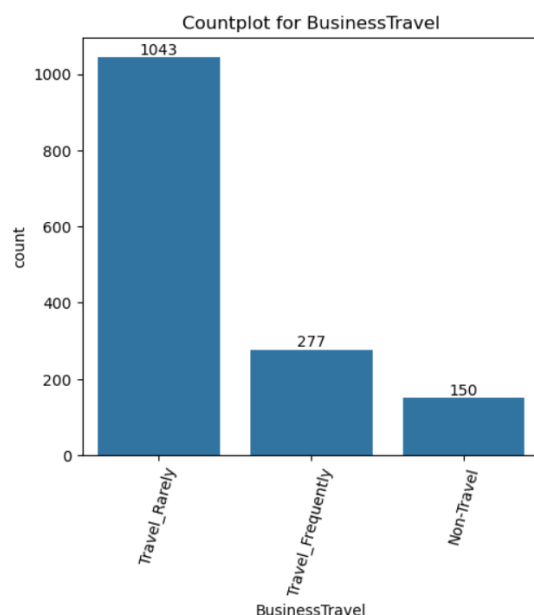
plt.title(f'Percentage distribution for {col}')

plt.show()

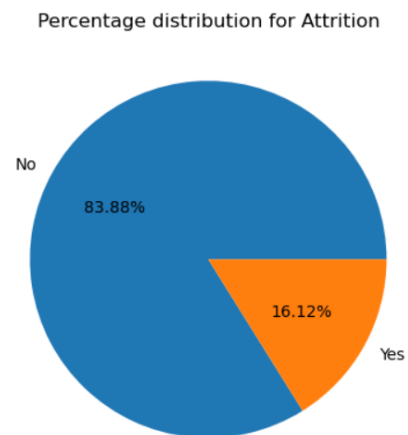
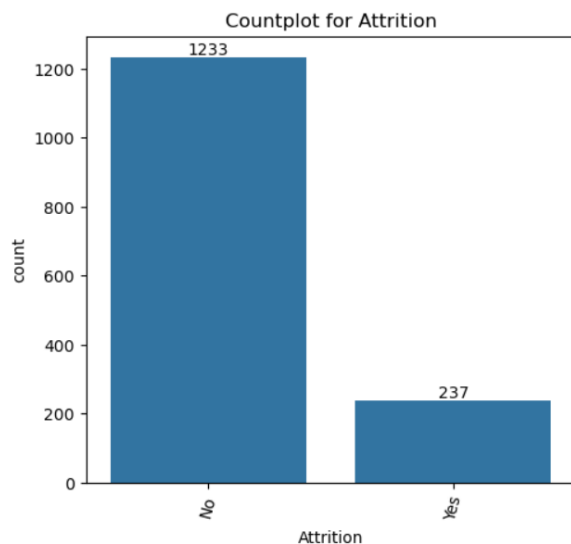
return vc

```

vc_pizza_cat = count_per_cat('BusinessTravel')



vc_pizza_cat = count_per_cat('Attrition')



STEP 6 SOME STATS MEASURE

```
data['MonthlyIncome'].mean()
```

```
6502.931292517007
```

```
data['MonthlyIncome'].median()
```

```
4919.0
```

```
import pandas as pd
```

```
from scipy.stats import skew
```

```
# Calculate the skewness of the 'MonthlyIncome' column
```

```
monthly_income_skewness = skew(data['MonthlyIncome'])
```

```
# Print the skewness value
```

```
print(f"Skewness of MonthlyIncome: {monthly_income_skewness}")
```

The skewness of a dataset indicates its asymmetry. A **skewness value of 1.368** for the **MonthlyIncome** suggests the following:

- **Positive Skewness:** Since the skewness is greater than 0, it means the data is **positively skewed**, or **right-skewed**.

- **Interpretation:** In this case, a skewness of 1.368 indicates that most employees' monthly incomes are concentrated around lower values, but there are a few employees who earn significantly higher incomes, pulling the tail of the distribution to the right.
- **Practical Impact:** In terms of salary distributions, this often suggests that a small number of employees have much higher salaries compared to the majority, which is common in real-world income data.

Convert 'Attrition' column to binary format (Yes -> 1, No -> 0)

```
data['Attrition'] = data['Attrition'].apply(lambda x: 1 if x == 'Yes' else 0)
```

Now calculate the correlation between 'Age' and 'Attrition'

```
correlation = data['Age'].corr(data['Attrition'])
```

```
print(f"Correlation between Age and Attrition: {correlation}")
```

The correlation value of -0.159 between Age and Attrition suggests a weak negative relationship.

- **Negative Correlation:** The negative sign indicates that as age increases, the likelihood of attrition decreases. In other words, older employees are slightly less likely to leave the company compared to younger employees.
- **Weak Correlation:** The magnitude of the correlation (0.159) is quite small, meaning that age alone is not a strong predictor of whether an employee will leave. Other factors likely play a more significant role in employee attrition.

In summary, while age has a slight impact on attrition (older employees tend to stay longer), the correlation is not strong enough to draw major conclusions without considering other variables

STEP 7 TABLEAU DASHBOARD LINK

https://public.tableau.com/views/projectUNIFIEDMENTOR/Dashboard1?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

The bar chart titled **"Attrition based on overtime and gender"** visualizes the relationship between employee attrition, gender, and whether they worked overtime.

Here's what the chart conveys:

X-Axis (Horizontal): It categorizes employees based on their gender (Female or Male) and whether they worked overtime ("No" or "Yes").

Y-Axis (Vertical): It shows the average age of employees.

The chart contains four groupings for each gender and their overtime status:

Female/No Overtime

Female/Yes Overtime

Male/No Overtime

Male/Yes Overtime

Each group is further broken down by attrition (whether the employee left the company or not) with the labels No and Yes indicating attrition status

Employees who did not work overtime (both males and females) seem to have a higher average age compared to those who did.

For both genders, those who worked overtime and stayed with the company (attrition "No") tend to be younger on average compared to those who did not work overtime.

This visualization helps in understanding how overtime and age vary with gender for employees who left the company (attrition "Yes") versus those who stayed (attrition "No").

The bar chart titled "**Attrition based on avg distance from home and marital status**" visualizes how employee attrition varies by their marital status (Divorced, Married, Single) and the average distance from home.

Here's what the chart shows:

X-Axis (Horizontal): It categorizes employees by their marital status (Divorced, Married, Single) and whether they experienced attrition (labeled "No" for those who stayed and "Yes" for those who left).

Y-Axis (Vertical): It represents the average distance from home for employees.

Colouring: The orange bars represent employees who left the company (attrition "Yes"), while the blue bars represent those who stayed (attrition "No").

For divorced employees, those who left the company (orange) lived farther from home on average compared to those who stayed (blue).

For married employees, the pattern is similar, with those who left having a greater average distance from home than those who stayed.

For single employees, those who left also lived farther from home compared to those who stayed.

The chart highlights a potential relationship between longer commuting distances and higher attrition rates across all marital statuses. Employees who lived farther away from work were more likely to leave, regardless of their marital status.

FINAL CONCLUSIONS

1. Initially, the dataset contained 35 columns and 1,470 rows. After data cleaning, it was refined to 31 columns with a step size of 1.
2. At Green Destinations, a well-known travel agency, 16.1% of employees are experiencing attrition, indicating they are leaving the company.
3. The data reveals that employees with fewer years at the company show a higher attrition rate. Similarly, younger employees, particularly those between the ages of 20 and 30, exhibit higher attrition rates compared to older employees and those with longer tenure. In contrast, employees with more years at the company and older ages tend to have lower attrition rates.
4. The data indicates a positive relationship between age and monthly income, as shown by an upward-sloping scatterplot. Generally, as age increases, monthly income also rises. Most employees in their 20s to 40s have monthly incomes ranging from 2,500 to 10,000. This trend suggests that the majority of employees earn higher incomes as they age, though there are some older employees with relatively low monthly incomes.
5. There is significant income inequality between the top 10 and bottom 10 employees, creating a notable wage disparity within the company. This disparity contributes to a widening wage gap across the organization.
6. The analyses illustrates that only 37.5% of employees received a salary hike of more than 15%, while the majority did not see an increase of this magnitude.
7. Shows 18.3% of employees have worked for more than 8 years without a recent promotion, while 81.7% are either employees who have been with the company for 8 years or less, or have recently received a promotion.
8. A bar plot revealed that as job satisfaction levels increase, both the attrition rate and attrition proportion decrease.
9. Examining the relationship between monthly income and years at the company by gender, we observe that most data points fall within 0 to 10 years of tenure. There are numerous outliers, particularly among employees with longer tenure and higher monthly incomes, who are predominantly male compared to female employees.
10. When analyzing attrition by department, both male and female employees exhibit the highest levels of attrition in the Research and

Development department. In contrast, the Sales and Human Resources departments show the lowest attrition rates for both genders.

Specifically, males in Research and Development have the highest attrition counts and rates.

11. Employees in the Research and Development department tend to work significant overtime compared to those in other departments. In contrast, the Human Resources department has the least amount of overtime among the various departments, including Sales.
12. Work Life Balance has very low correlation values with both Years At Company (0.012)
13. Years At Company and Years Since Last Promotion have a relatively strong positive correlation (0.62) and Years Since Last Promotion (0.0089
14. Employees who did not work overtime (both males and females) seem to have a higher average age compared to those who did.
15. For both genders, those who worked overtime and stayed with the company (attrition "No") tend to be younger on average compared to those who did not work overtime.
- 16.. Employees who lived farther away from work were more likely to leave, regardless of their marital status.
17. The correlation value of -0.159 between Age and Attrition suggests a weak negative relationship.

PROJECT MADE BY **PRIYAL KESWANI** UNDER THE INTERNSHIP WITH UNIFIED MENTOR .

Contact info: 9717929316

Email id: priyalkeswani@gmail.com