

HR Data analysis using python - by katari Pavan

```
In [ ]: # Importing libraries
```

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib as plt
from scipy import stats
```

```
In [ ]: # Importing Data set into IDE (jupyter)
```

```
In [2]: Data = pd.read_csv('employee_experience_survey_data.csv')
```

```
In [3]: Data.head(5)
```

Out[3]:

	Name	Age Bracket	Gender	Ethnicity	Job Title	Department	Date Survey Completed	Job Satisfaction
0	John Doe	25-34	Female	Asian	Product Manager	Product Development	2024-10-05	Disagree
1	Jane Smith	18-24	Female	Middle Eastern	Operations Manager	Sales	2024-10-07	Agree
2	Carlos Reyes	45-54	Female	Indian	UX Designer	Consulting	2024-10-08	Neutral
3	Emily Zhang	35-44	Male	Caucasian	UX Designer	HR	2024-10-07	Neutral
4	Michael Johnson	18-24	Female	Caucasian	UX Designer	Product Development	2024-10-07	Agree

```
In [ ]: # Identifying Data type and null/missing values
```

```
In [4]: Data.info()
```



```
In [7]: Data['Team Collaboration'] = Data['Team Collaboration'].map({'Strongly Disagree': '1',
                                                                    'Disagree': '2',
                                                                    'Neutral': '3',
                                                                    'Agree': '4',
                                                                    'Strongly Agree': '5'})

Data['Workload Fairness'] = Data['Workload Fairness'].map({'Strongly Disagree': '1',
                                                            'Disagree': '2',
                                                            'Neutral': '3',
                                                            'Agree': '4',
                                                            'Strongly Agree': '5'})
```

```
In [8]: Data['Workplace Inclusivity'] = Data['Workplace Inclusivity'].map({'Strongly Disagree': '1',
                                                                              'Disagree': '2',
                                                                              'Neutral': '3',
                                                                              'Agree': '4',
                                                                              'Strongly Agree': '5'})

Data['Company Communication'] = Data['Company Communication'].map({'Strongly Disagree': '1',
                                                                      'Disagree': '2',
                                                                      'Neutral': '3',
                                                                      'Agree': '4',
                                                                      'Strongly Agree': '5'})
```

```
In [9]: Data['Compensation Satisfaction'] = Data['Compensation Satisfaction'].map({'Strongly Disagree': '1',
                                                                                     'Disagree': '2',
                                                                                     'Neutral': '3',
                                                                                     'Agree': '4',
                                                                                     'Strongly Agree': '5'})

Data['Career Development Opportunities'] = Data['Career Development Opportunities'].map({'Strongly Disagree': '1',
                                                                                          'Disagree': '2',
                                                                                          'Neutral': '3',
                                                                                          'Agree': '4',
                                                                                          'Strongly Agree': '5'})
```

```
In [10]: Data['Job Satisfaction'] = Data['Job Satisfaction'].map({'Strongly Disagree': '1',
                                                                    'Disagree': '2',
                                                                    'Neutral': '3',
                                                                    'Agree': '4',
                                                                    'Strongly Agree': '5'})
```

```
In [11]: Data.head(5)
```

Out[11]:

	Name	Age Bracket	Gender	Ethnicity	Job Title	Department	Date Survey Completed	Job Satisfaction
0	John Doe	25-34	Female	Asian	Product Manager	Product Development	2024-10-05	4
1	Jane Smith	18-24	Female	Middle Eastern	Operations Manager	Sales	2024-10-07	3
2	Carlos Reyes	45-54	Female	Indian	UX Designer	Consulting	2024-10-08	5
3	Emily Zhang	35-44	Male	Caucasian	UX Designer	HR	2024-10-07	4
4	Michael Johnson	18-24	Female	Caucasian	UX Designer	Product Development	2024-10-07	3

```
In [ ]: # Q1 - Calculating Descriptive statistics of Job satisfaction and Overall Engagement
```

```
In [12]: job_satisfaction_stats = Data['Job Satisfaction'].astype(int).agg(['mean', 'median', 'std'])
overall_engagement_stats = Data['Overall Engagement'].astype(int).agg(['mean', 'median', 'std'])

print("Job Satisfaction Statistics:")
print(job_satisfaction_stats)

print("\nOverall Engagement Statistics:")
print(overall_engagement_stats)

Job Satisfaction Statistics:
mean      3.000000
median    3.000000
std       1.309307
Name: Job Satisfaction, dtype: float64

Overall Engagement Statistics:
mean      3.400000
median    3.000000
std       1.298351
Name: Overall Engagement, dtype: float64
```

```
In [ ]: # Hypothesis testing
```

```
In [13]: it_job_satisfaction = Data[Data['Department'] == 'IT']['Job Satisfaction']
hr_job_satisfaction = Data[Data['Department'] == 'HR']['Job Satisfaction']

H0 = "There is no significant difference in job satisfaction between IT and HR departments"
H1 = "There is a significant difference in job satisfaction between IT and HR departments"

t_statistic, p_value = stats.ttest_ind(it_job_satisfaction, hr_job_satisfaction)

print("T-statistic:", t_statistic)
print("P-value:", p_value)

alpha = 0.05

if p_value < alpha:
    print("There is a significant difference in job satisfaction between IT and HR departments")
else:
    print("There is no significant difference in job satisfaction between IT and HR departments")

T-statistic: -1.7320508075688774
P-value: 0.33333333333333326
There is no significant difference in job satisfaction between IT and HR departments and accept null hypothesis.
```

```
In [ ]: ## Correlation analysis between Work life balance and Overall engagement
## WLB - Work-Life Balance
## OE - Overall Engagement
## corr - correlation
```

```
In [14]: WLB_OE_corr = Data['Work-Life Balance'].astype(int).corr(Data['Overall En  
WLB_OE_corr
```

```
if WLB_OE_corr > 0.5:  
    print("There is a strong positive correlation between Work-Life Balan  
elif WLB_OE_corr > 0:  
    print("There is a moderate positive correlation between Work-Life Bal  
elif WLB_OE_corr < -0.5:  
    print("There is a strong negative correlation between Work-Life Balan  
elif WLB_OE_corr < 0:  
    print("There is a moderate negative correlation between Work-Life Bal  
else:  
    print("There is little to no correlation between Work-Life Balance an
```

There is a moderate negative correlation between Work-Life Balance and Overall Engagement.