## 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]:
                                                                              Date
                                                                                           Jo
                         Age
                              Gender
                                       Ethnicity
                                                  Job Title
                                                                            Survey
               Name
                                                            Department
                     Bracket
                                                                                    Satisfactio
                                                                        Completed
                John
                                                   Product
                                                                Product
                                                                          2024-10-
         0
                       25-34
                              Female
                                          Asian
                                                                                       Disagre
                Doe
                                                   Manager
                                                           Development
                Jane
                                         Middle
                                                 Operations
                                                                          2024-10-
          1
                       18-24
                               Female
                                                                  Sales
                                                                                          Agre
               Smith
                                                   Manager
                                         Eastern
                                                                                07
              Carlos
                                                                          2024-10-
          2
                       45-54
                              Female
                                          Indian
                                                              Consulting
                                                                                        Neutra
               Reyes
                                                  Designer
                                                                          2024-10-
               Emily
                                                       UX
                       35-44
         3
                                Male
                                      Caucasian
                                                                    HR
                                                                                        Neutra
              Zhang
                                                  Designer
                                                                                07
                                                                          2024-10-
             Michael
                                                                Product
                       18-24
                              Female Caucasian
                                                                                          Agre
             Johnson
                                                  Designer Development
In []:
         # Identifying Data type and null/missing values
In [4]:
         Data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15 entries, 0 to 14
Data columns (total 18 columns):
    Column
                                    Non-Null Count Dtype
____
                                     _____
                                    15 non-null
0
    Name
                                                  object
                                    15 non-null
1
   Age Bracket
                                                    object
                                    15 non-null
15 non-null
15 non-null
   Gender
                                                    object
3
   Ethnicity
                                                    object
    Job Title
                                                   object
5
   Department
                                   15 non-null
                                                    object
6
   Date Survey Completed
                                   15 non-null
                                                    object
7
                                   15 non-null
    Job Satisfaction
                                                    object
                                   15 non-null
8
   Work-Life Balance
                                                    object
    Management Support
                                   15 non-null
                                                    object
                                   15 non-null
10 Team Collaboration
                                                    object
11 Workload Fairness
                                    15 non-null
                                                    object
12 Career Development Opportunities 15 non-null
                                                    object
13 Workplace Inclusivity
                                  15 non-null
                                                    object
                                   15 non-null
14 Company Communication
                                                    object
                                   15 non-null
15 Compensation Satisfaction
                                                    object
16 Job Security
                                    15 non-null
                                                    object
                                    15 non-null
17 Overall Engagement
                                                    object
dtypes: object(18)
memory usage: 2.2+ KB
```

## One Hot encoding - Converting categorical variables into continues variables

```
In [5]: Data['Work-Life Balance'] = Data['Work-Life Balance'].map({ 'Strongly Dis
                                                                       'Disagree':'2
                                                                       'Neutral':'3'
                                                                       'Agree':'4',
                                                                       'Strongly Agr
        Data['Management Support'] = Data['Management Support'].map({ 'Strongly D
                                                                         'Disagree':
                                                                         'Neutral':'
                                                                         'Agree':'4'
                                                                         'Strongly A
In [6]:
        Data['Overall Engagement'] = Data['Overall Engagement'].map({'Strongly Di
                                                                        'Disagree':'
                                                                        'Neutral':'3
                                                                        'Agree': '4',
                                                                        'Strongly Ag
        Data['Job Security'] = Data['Job Security'].map({'Strongly Disagree':'1',
                                                            'Disagree':'2',
                                                            'Neutral':'3',
                                                            'Agree':'4',
                                                            'Strongly Agree':'5'})
```

```
'Disagree':'
                                                                                 'Neutral':'
                                                                                 'Agree':'4'
                                                                                 'Strongly A
          Data['Workload Fairness'] = Data['Workload Fairness'].map({'Strongly Disa
                                                                             'Disagree':'2'
                                                                             'Neutral':'3',
                                                                             'Agree':'4',
                                                                             'Strongly Agre
          Data['Workplace Inclusivity'] = Data['Workplace Inclusivity'].map({'Stron
 In [8]:
                                                                                       'Neutr
                                                                                       'Agree
                                                                                       'Stron
          Data['Company Communication'] = Data['Company Communication'].map({'Stron
                                                                                       'Disag
                                                                                      'Neutr
                                                                                       'Agree
                                                                                       'Stron
 In [9]:
          Data['Compensation Satisfaction'] = Data['Compensation Satisfaction'].map
          Data['Career Development Opportunities'] = Data['Career Development Opportunities']
In [10]:
          Data['Job Satisfaction'] = Data['Job Satisfaction'].map({'Strongly Disagr
                                                                           'Disagree':'2',
                                                                           'Neutral':'3',
                                                                           'Agree':'4',
                                                                           'Strongly Agree
In [11]:
          Data.head(5)
Out [11]:
                                                                              Date
                                                                                           Jo
                         Age
                                                                            Survey
               Name
                               Gender
                                       Ethnicity
                                                  Job Title
                                                            Department
                                                                                    Satisfactio
                      Bracket
                                                                        Completed
                                                                          2024-10-
                John
                                                   Product
                                                                Product
          0
                        25-34
                               Female
                                           Asian
                 Doe
                                                   Manager
                                                           Development
                                                                                05
                Jane
                                          Middle
                                                 Operations
                                                                          2024-10-
                        18-24
                                                                  Sales
                               Female
                Smith
                                         Eastern
                                                   Manager
                                                                                07
                                                                          2024-10-
               Carlos
                                                       UX
          2
                        45-54
                               Female
                                          Indian
                                                              Consulting
               Reyes
                                                   Designer
                                                                                08
                                                                          2024-10-
                Emily
                                                        UX
          3
                                                                    HR
                        35-44
                                 Male
                                       Caucasian
                                                   Designer
               Zhang
                                                                                07
                                                                          2024-10-
              Michael
                                                                Product
                        18-24
                               Female Caucasian
                                                           Development
              Johnson
                                                   Designer
                                                                                07
```

Data['Team Collaboration'] = Data['Team Collaboration'].map({'Strongly Di

In [7]:

```
In [ ]: # Q1 - Calculating Descriptive statistics of Job satification and Over al
In [12]: job satisfaction stats = Data['Job Satisfaction'].astype(int).agg(['mean']
         overall engagement stats = Data['Overall Engagement'].astype(int).agg(['m
         print("Job Satisfaction Statistics:")
         print(job_satisfaction_stats)
         print("\nOverall Engagement Statistics:")
         print(overall_engagement_stats)
         Job Satisfaction Statistics:
         mean
                 3.000000
                   3.000000
         median
         std
                   1.309307
         Name: Job Satisfaction, dtype: float64
         Overall Engagement Statistics:
                   3.400000
         mean
         median
                   3.000000
                   1.298351
         std
         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'
         HO = "There is no significant difference in job satisfaction between IT a
         H1 = "There is a significant difference in job satisfaction between IT an
         t statistic, p value = stats.ttest ind(it job satisfaction, hr job satisf
         print("T-statistic:", t_statistic)
         print("P-value:", p_value)
         alpha = 0.05
         if p_value < alpha:</pre>
             print("There is a significant difference in job satisfaction between
         else:
             print("There is no significant difference in job satisfaction between
         T-statistic: -1.7320508075688774
         P-value: 0.33333333333333336
         There is no significant difference in job satisfaction between IT and HR
         departments and accept null hypothesis.
 In [ ]: ## Correlation analysis between Work lifenbalance 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</pre>
```

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