**HR data analyses using python - by Katari Pavan **

Importing *Libraries*

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

Importing data set into IDE

```
Data = pd.read_csv('employee_experience_survey_data.csv')
```

```
Data.head(5)
```

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

Next steps: Generate code with Data | View recommended plots | New interactive sheet

Identifying Data types and Null or Missing values

```
Data.info()
```

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

*One hat encoding - Converting categorical variables into Continues variables *

```
Data['Work-Life Balance'] = Data['Work-Life Balance'].map({ 'Strongly Disagree'
                                                           'Disagree':'2',
                                                            'Neutral':'3',
                                                             'Agree':'4',
                                                                    'Strongly
Data['Management Support'] = Data['Management Support'].map({'Strongly Disagree
                                                                    'Neutral':
                                                                    'Agree':'4
                                                                    'Strongly
```

```
Data['Overall Engagement'] = Data['Overall Engagement'].map({'Strongly Disagree
                                                             'Neutral':
                                                             'Agree':'4
                                                             'Strongly
Data['Job Security'] = Data['Job Security'].map({'Strongly Disagree':'1','Disag
                                                             'Neutral':
                                                             'Agree':'4
                                                             'Strongly


Data['Team Collaboration'] = Data['Team Collaboration'].map({'Strongly Disagree
                                                             'Neutral':
                                                             'Agree':'4
                                                             'Strongly
Data['Workload Fairness'] = Data['Workload Fairness'].map({'Strongly Disagree':
                                                    'Disagree':'2',
                                                    'Neutral':'3',
                                                    'Agree':'4',
                                                    'Strongly Agree':'5'


Data['Workplace Inclusivity'] = Data['Workplace Inclusivity'].map({'Strongly Di
                                                           'Disagree':'
                                                           'Neutral':'3
                                                           'Agree':'4',
                                                           'Strongly Ag
Data['Company Communication'] = Data['Company Communication'].map({'Strongly Di
                                                           'Disagree':'
                                                           'Neutral':'3
                                                           'Agree':'4',
                                                           'Strongly Ag


Data['Compensation Satisfaction'] = Data['Compensation Satisfaction'].map({'Str
                                                                 'Dis
                                                                 'Neu
                                                                 'Agr
                                                                 'Str
Data['Career Development Opportunities'] = Data['Career Development Opportuniti









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

```
Data.head(5)
```

Department	Date Survey Completed	Job Satisfaction	Work-Life Balance	Management Support	Team Collaboration	Wor Fai
Product Development	2024-10-05	2	5	3	3	
Sales	2024-10-07	4	1	5	3	
Consulting	2024-10-08	3	1	4	2	
HR	2024-10-07	3	4	4	1	
Product Development	2024-10-07	4	5	2	3	

Next steps: Generate code with Data | View recommended plots | New interactive sheet

Exploratory Data Analysis

```
Data.describe()
```

	Job Satisfaction	Work-Life Balance	Management Support	Team Collaboration	Workload Fairness	Ca Develop Opportuni
count	15.000000	15.000000	15.000000	15.000000	15.000000	15.0(
mean	3.000000	3.333333	2.800000	2.866667	3.266667	3.0(
std	1.309307	1.676163	1.521278	1.407463	1.533747	1.4
min	1.000000	1.000000	1.000000	1.000000	1.000000	1.0(
25%	2.000000	1.500000	1.000000	1.500000	2.500000	2.0(
50%	3.000000	4.000000	3.000000	3.000000	3.000000	4.0(
75%	4.000000	5.000000	4.000000	4.000000	5.000000	4.0(
max	5.000000	5.000000	5.000000	5.000000	5.000000	5.0(

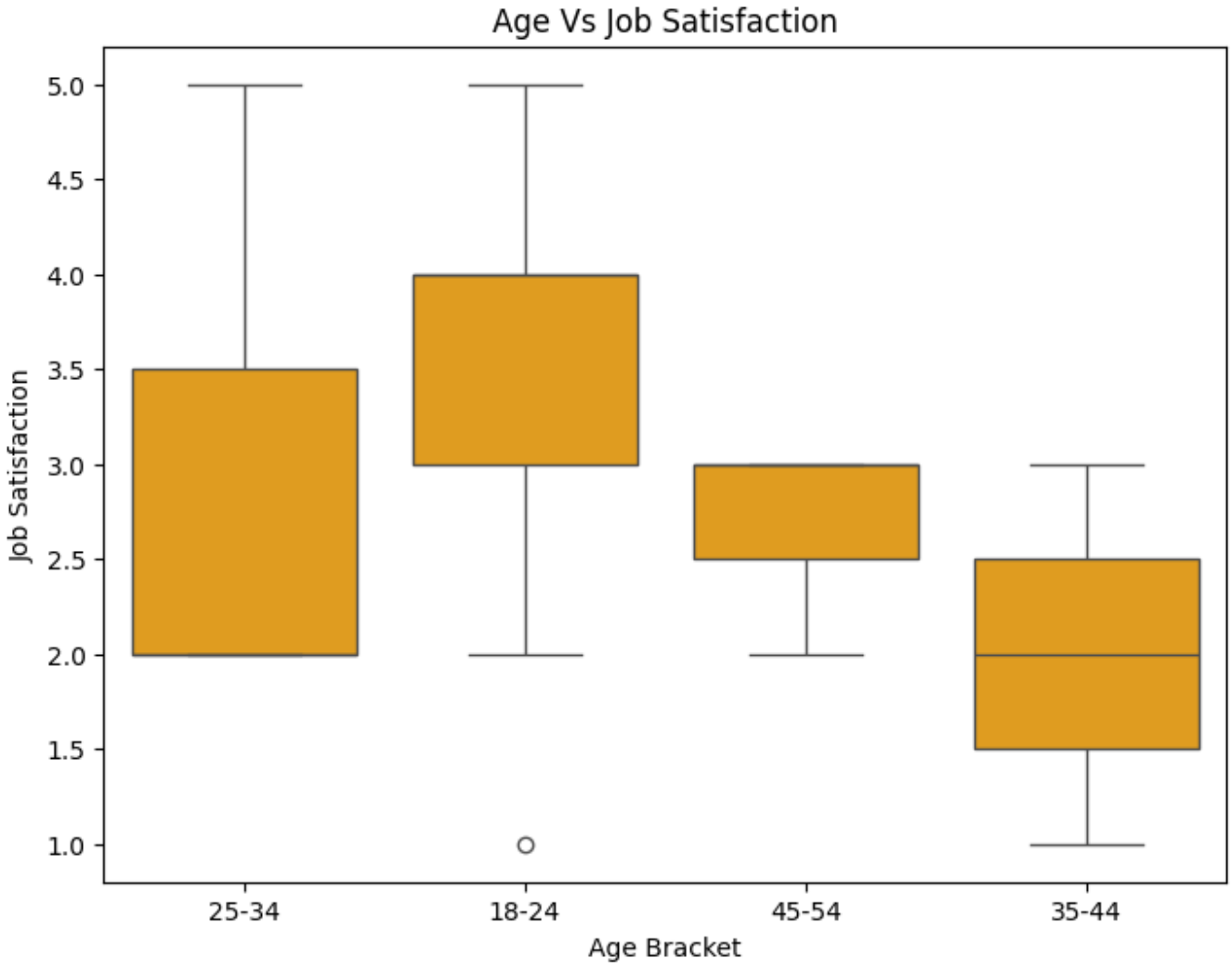
Descriptive statistics of Job satisfaction and Overall Engagement

```
job_satisfaction_stats = Data['Job Satisfaction'].astype(int).agg(['mean', 'mec
overall_engagement_stats = Data['Overall Engagement'].astype(int).agg(['mean',

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
```

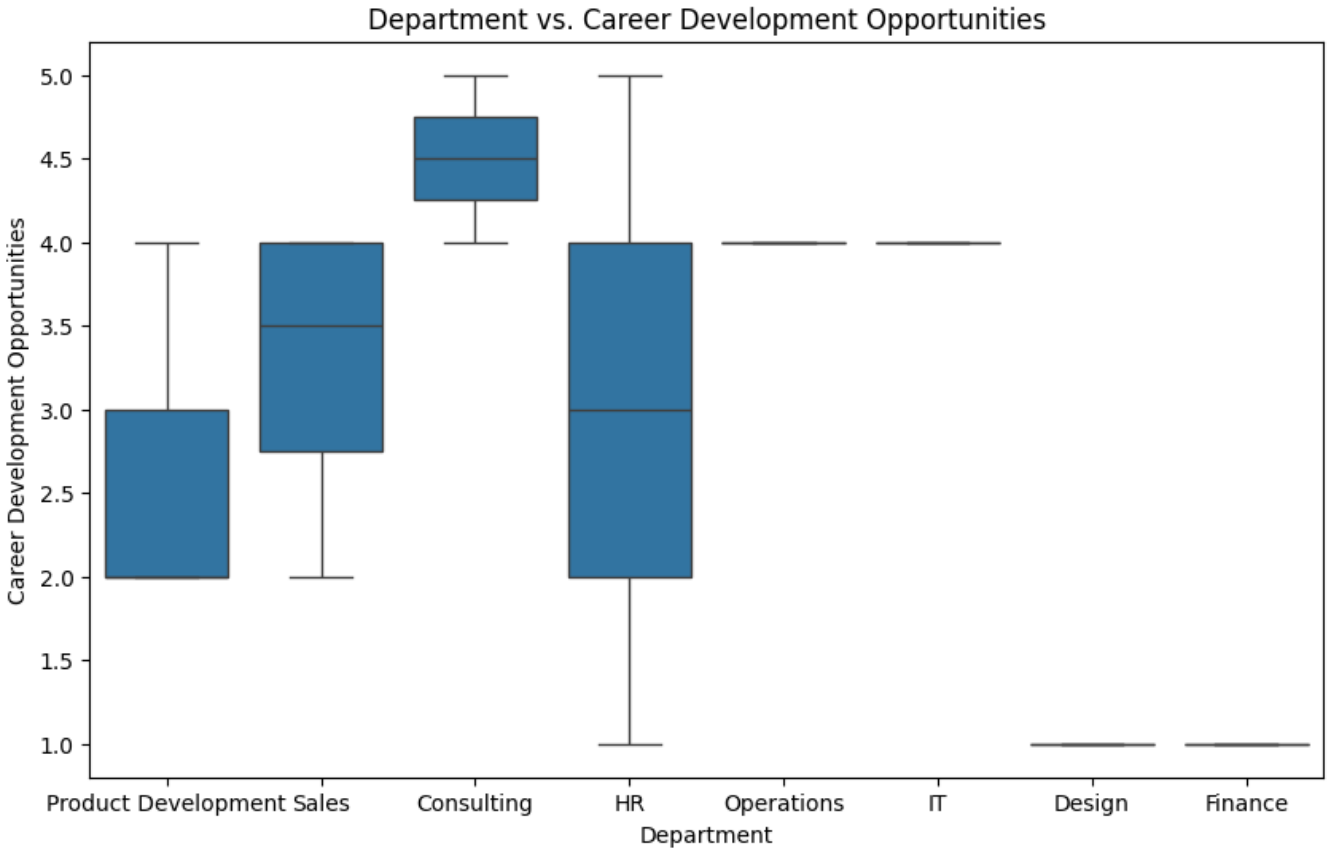
Bi variant analysis

```
plt.figure(figsize=(8, 6))
sns.boxplot(x='Age Bracket', y='Job Satisfaction', data=Data,color='orange')
plt.title('Age Vs Job Satisfaction')
plt.xlabel('Age Bracket')
plt.ylabel('Job Satisfaction')
plt.show()
```



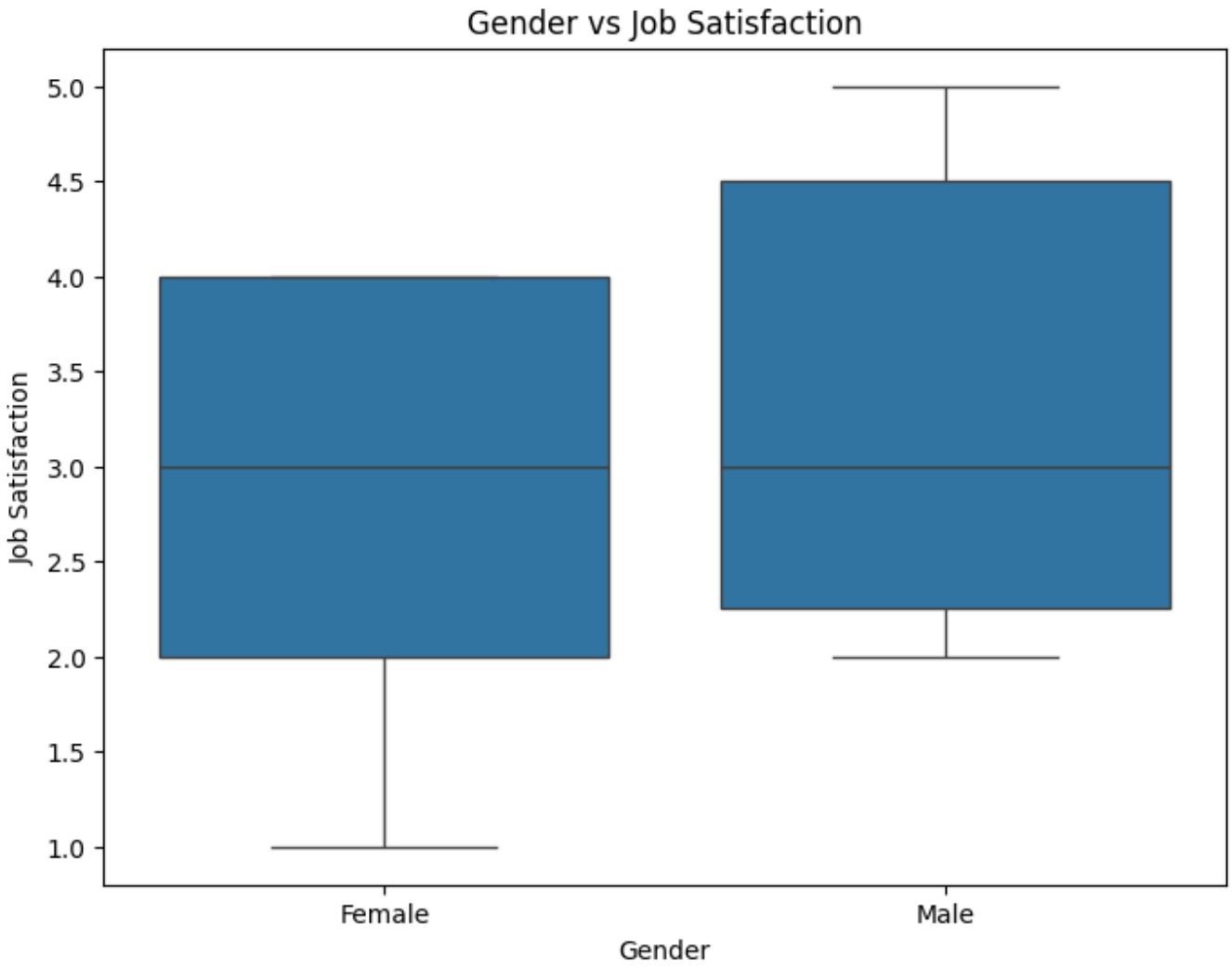
Interpretation : Job satsification for Age group people between 25-34 is Netural and low in age group of 35-44

```
plt.figure(figsize=(10, 6))
sns.boxplot(x='Department', y='Career Development Opportunities', data=Data)
plt.title('Department vs. Career Development Opportunities')
plt.xlabel('Department')
plt.ylabel('Career Development Opportunities')
plt.show()
```



Interpretation : It is identified that career development opportunites are very low for Design and finance department and good for Consulting department

```
plt.figure(figsize=(8, 6))
sns.boxplot(x='Gender', y='Job Satisfaction', data=Data)
plt.title('Gender vs Job Satisfaction')
plt.show()
```

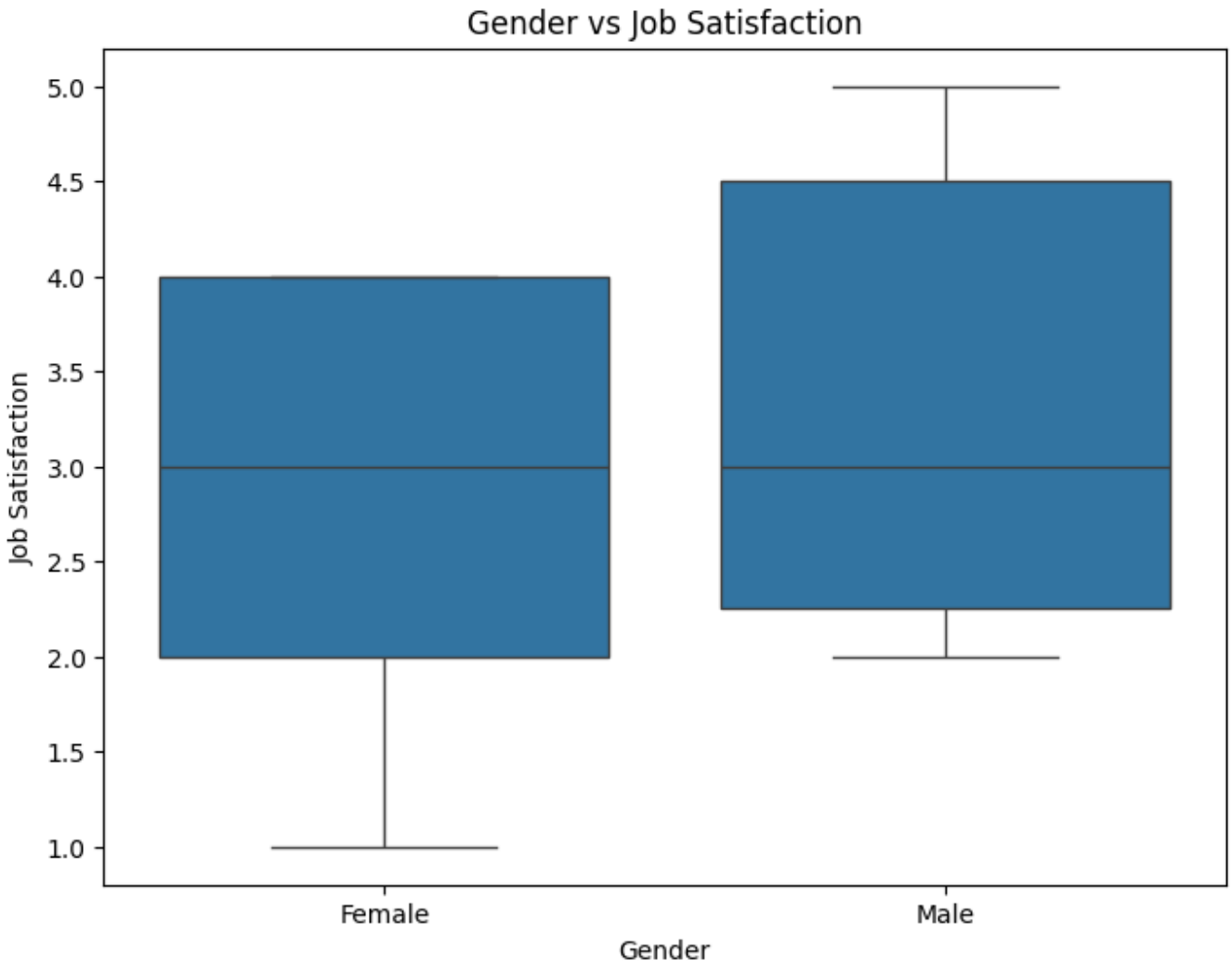


Interpretation : It is resulted that Job satsification in Female is slightly low as compared to Male

```
plt.figure(figsize=(8,6), dpi=100 )
sns.boxplot(x='Gender', y='Work-Life Balance', data=Data, color='lightgreen')
plt.title('Gender Vs work life balance')
plt.ylabel('Work life balance')
plt.show()
```

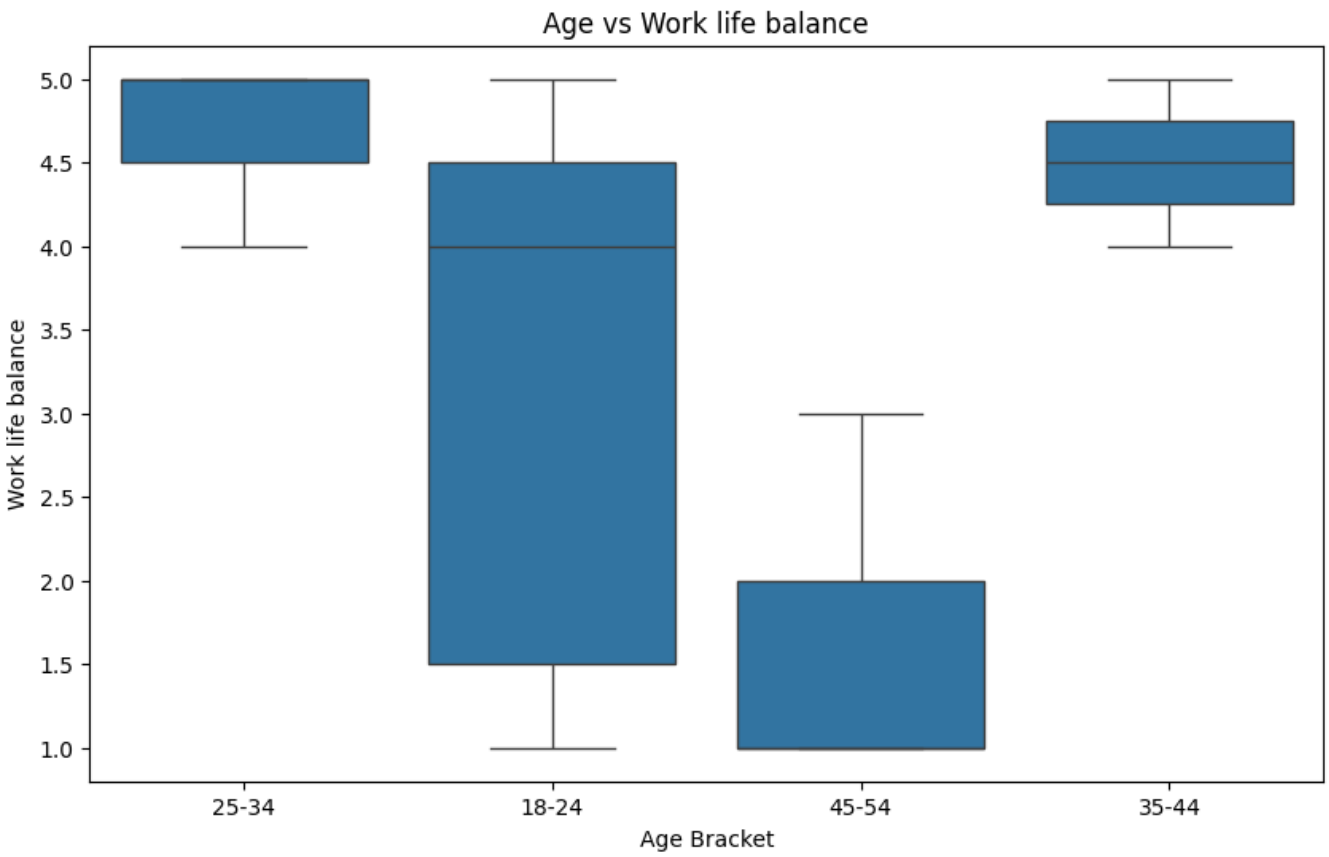
Interpretation : Work life balance for male is good as compared to female

```
plt.figure(figsize=(8, 6))  # Now calling figure() from the pyplot module
sns.boxplot(x='Gender', y='Job Satisfaction', data=Data)  # Now calling boxplot
plt.title('Gender vs Job Satisfaction')
plt.show()
```



Interpretation : It is find that there is a slight difference in job satisfaction between genders, with males reporting slightly higher levels of satisfaction on average.

```
plt.figure(figsize=(10, 6), dpi= 100)
sns.boxplot(x='Age Bracket', y='Work-Life Balance', data=Data)
plt.title('Age vs Work life balance')
plt.ylabel('Work life balance')
plt.show()
```



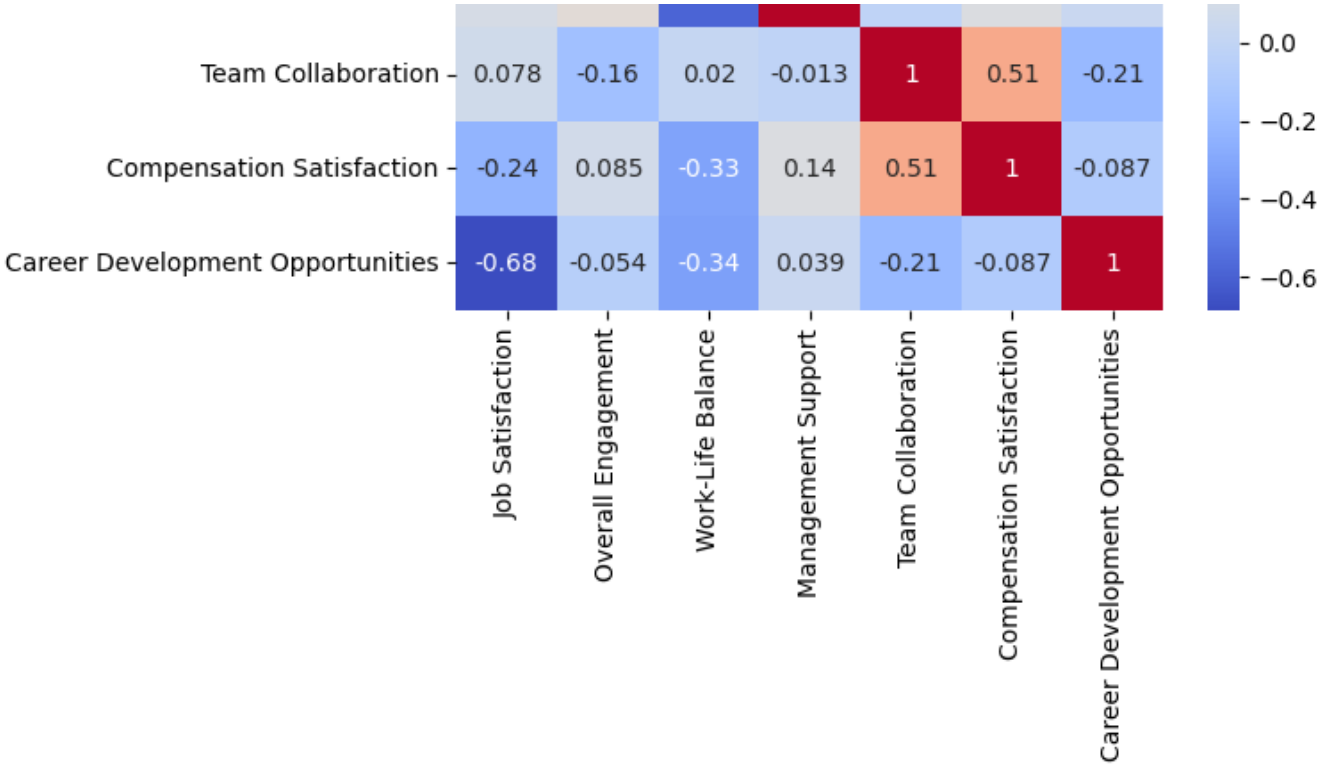
Interpretation : It identified that there is significant difference in work-life balance across the age groups. Work life balance of employees of age group between 25-34 and 35-44 are good.

Correlation analysis

```
correlation_matrix = Data[['Job Satisfaction', 'Overall Engagement', 'Work-Life
print(correlation_matrix)

sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Employee Satsifaction Survey')
```

```
plt.show()
```

	Job Satisfaction	Overall Engagement	Work-Life Balance	Management Support	Team Collaboration	Compensation Satisfaction	Career Development Opportunities
Job Satisfaction	1.000000	-0.042018	0.097642	0.107583	0.077522	-0.239732	-0.683074
Overall Engagement	-0.042018	1.000000	-0.361041	0.188051	-0.164169	0.084614	-0.053576
Work-Life Balance	0.097642	-0.361041	1.000000	-0.588256	0.020185	-0.327709	-0.335953
Management Support	0.107583	0.188051	-0.588256	1.000000	-0.013344	0.144430	0.039193
Team Collaboration	0.077522	-0.164169	0.020185	-0.013344	1.000000	0.512930	-0.207105
Compensation Satisfaction	-0.239732	0.084614	-0.327709	0.144430	0.512930	1.000000	-0.087336
Career Development Opportunities	-0.683074	-0.053576	-0.335953	0.039193	-0.207105	-0.087336	1.000000



Hypothesis testing (Paired T- Test)

```
it_job_satisfaction = Data[Data['Department'] == 'IT']['Job Satisfaction'].astyp
hr_job_satisfaction = Data[Data['Department'] == 'HR']['Job Satisfaction'].astyp

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

alpha = 0.05

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

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

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

```
T-statistic: -1.7320508075688774
P-value: 0.33333333333333326
There is no significant difference in job satisfaction between IT and HR de
```

Calculating correlation between Work life balance and Overall engagement

```
## Correlation analysis between Work lifenbalance and Overall engagement
## WLB - Work-Life Balance
## OE - Overall Engagement
## corr - correlation

WLB_OE_corr = Data['Work-Life Balance'].astype(int).corr(Data['Overall Engageme
WLB_OE_corr

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

```
There is a moderate negative correlation between Work-Life Balance and Over
```

Key findings :

It is identified that career development opportunites are very low for Design and finance department and good for consulting department

There is strong positive correletion between Team collabration and compensation satsification.

There is strong negative coorelation between job satisfaction and carrer development opportunities.

Those employees who are satisfied with their jobs are more likely to be engaged and less likely to feel they have limited development opportunities.

It identified that there is significant difference in work-life balance across the age groups. Work life balance of employees of age group between 25-34 and 35-44 are good.

Suggestions :

1.Provide career development opportunites by recognizing and appreciating there contribution.

2.Provide management support in work life balance

3.Improve areas like overall engagement, job satisfaction.