In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
df = pd.read_csv('HR-Employee-Attrition.csv')
df
```

Out[2]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education
0	41	Yes	Travel_Rarely	1102	Sales	1	2
1	49	No	Travel_Frequently	279	Research & Development	8	1
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2
3	33	No	Travel_Frequently	1392	Research & Development	3	4
4	27	No	Travel_Rarely	591	Research & Development	2	1
...	...	...	...	...	...	...	...
1465	36	No	Travel_Frequently	884	Research & Development	23	2
1466	39	No	Travel_Rarely	613	Research & Development	6	1
1467	27	No	Travel_Rarely	155	Research & Development	4	3
1468	49	No	Travel_Frequently	1023	Sales	2	3
1469	34	No	Travel_Rarely	628	Research & Development	8	3

1470 rows × 35 columns

In [3]:
```
print(df.head())
```

```
   Age Attrition     BusinessTravel  DailyRate              Department  \
0   41       Yes      Travel_Rarely       1102                   Sales
1   49        No  Travel_Frequently        279  Research & Development
2   37       Yes      Travel_Rarely       1373  Research & Development
3   33        No  Travel_Frequently       1392  Research & Development
4   27        No      Travel_Rarely        591  Research & Development

   DistanceFromHome  Education EducationField  EmployeeCount  EmployeeNumb
er  \
0                 1          2  Life Sciences              1
1
1                 8          1  Life Sciences              1
2
2                 2          2          Other              1
4
3                 3          4  Life Sciences              1
5
4                 2          1        Medical              1
7

   ...  RelationshipSatisfaction StandardHours  StockOptionLevel  \
0  ...                         1            80                 0
1  ...                         4            80                 1
2  ...                         2            80                 0
3  ...                         3            80                 0
4  ...                         4            80                 1

   TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance  YearsAtCompan
y  \
0                  8                      0               1
6
1                 10                      3               3              1
0
2                  7                      3               3
0
3                  8                      3               3
8
4                  6                      3               3
2

  YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                  4                        0                    5
1                  7                        1                    7
2                  0                        0                    0
3                  7                        3                    0
4                  2                        2                    2

[5 rows x 35 columns]
```

In [4]:
```
print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 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   EmployeeCount             1470 non-null   int64
 9   EmployeeNumber            1470 non-null   int64
 10  EnvironmentSatisfaction   1470 non-null   int64
 11  Gender                    1470 non-null   object
 12  HourlyRate                1470 non-null   int64
 13  JobInvolvement            1470 non-null   int64
 14  JobLevel                  1470 non-null   int64
 15  JobRole                   1470 non-null   object
 16  JobSatisfaction           1470 non-null   int64
 17  MaritalStatus             1470 non-null   object
 18  MonthlyIncome             1470 non-null   int64
 19  MonthlyRate               1470 non-null   int64
 20  NumCompaniesWorked        1470 non-null   int64
 21  Over18                    1470 non-null   object
 22  OverTime                  1470 non-null   object
 23  PercentSalaryHike         1470 non-null   int64
 24  PerformanceRating         1470 non-null   int64
 25  RelationshipSatisfaction  1470 non-null   int64
 26  StandardHours             1470 non-null   int64
 27  StockOptionLevel          1470 non-null   int64
 28  TotalWorkingYears         1470 non-null   int64
 29  TrainingTimesLastYear     1470 non-null   int64
 30  WorkLifeBalance           1470 non-null   int64
 31  YearsAtCompany            1470 non-null   int64
 32  YearsInCurrentRole        1470 non-null   int64
 33  YearsSinceLastPromotion   1470 non-null   int64
 34  YearsWithCurrManager      1470 non-null   int64
dtypes: int64(26), object(9)
memory usage: 402.1+ KB
None
```

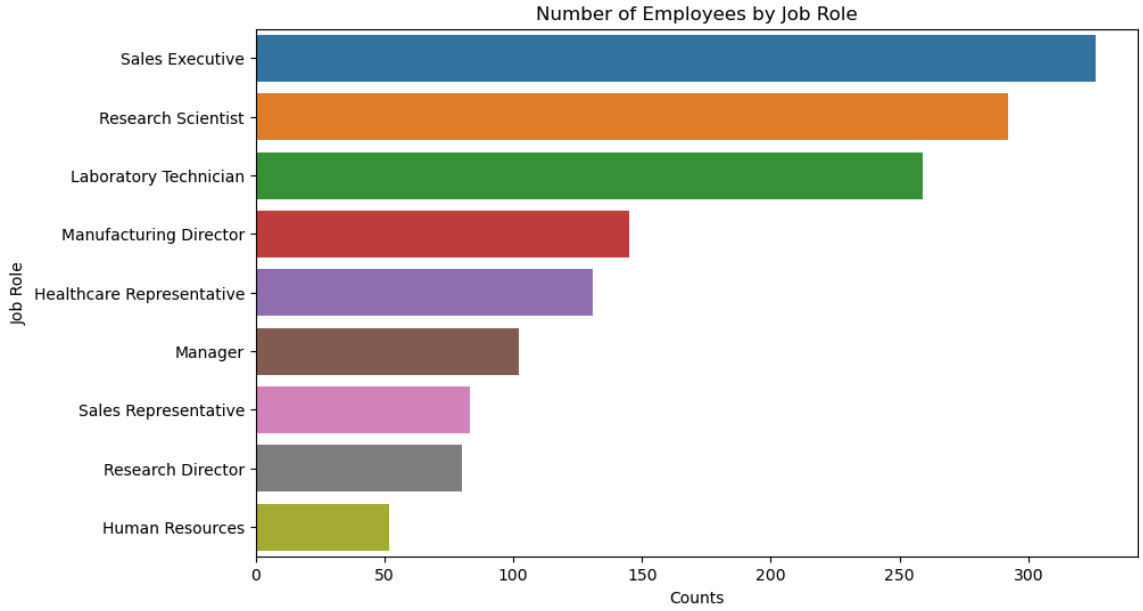
In [5]:
```
print(df.describe())
```

```
                Age    DailyRate  DistanceFromHome    Education  EmployeeCo
unt  \
count  1470.000000  1470.000000       1470.000000  1470.000000         147
0.0
mean     36.923810   802.485714          9.192517     2.912925
1.0
std       9.135373   403.509100          8.106864     1.024165
0.0
min      18.000000   102.000000          1.000000     1.000000
1.0
25%      30.000000   465.000000          2.000000     2.000000
1.0
50%      36.000000   802.000000          7.000000     3.000000
1.0
75%      43.000000  1157.000000         14.000000     4.000000
1.0
max      60.000000  1499.000000         29.000000     5.000000
1.0

       EmployeeNumber  EnvironmentSatisfaction   HourlyRate  JobInvolvemen
t  \
count     1470.000000              1470.000000  1470.000000     1470.00000
0
mean      1024.865306                 2.721769    65.891156        2.72993
2
std        602.024335                 1.093082    20.329428        0.71156
1
min          1.000000                 1.000000    30.000000        1.00000
0
25%        491.250000                 2.000000    48.000000        2.00000
0
50%       1020.500000                 3.000000    66.000000        3.00000
0
75%       1555.750000                 4.000000    83.750000        3.00000
0
max       2068.000000                 4.000000   100.000000        4.00000
0

          JobLevel  ...  RelationshipSatisfaction  StandardHours  \
count  1470.000000  ...               1470.000000         1470.0
mean      2.063946  ...                  2.712245           80.0
std       1.106940  ...                  1.081209            0.0
min       1.000000  ...                  1.000000           80.0
25%       1.000000  ...                  2.000000           80.0
50%       2.000000  ...                  3.000000           80.0
75%       3.000000  ...                  4.000000           80.0
max       5.000000  ...                  4.000000           80.0

       StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
count       1470.000000        1470.000000            1470.000000
mean           0.793878          11.279592               2.799320
std            0.852077           7.780782               1.289271
min            0.000000           0.000000               0.000000
25%            0.000000           6.000000               2.000000
50%            1.000000          10.000000               3.000000
75%            1.000000          15.000000               3.000000
max            3.000000          40.000000               6.000000

       WorkLifeBalance  YearsAtCompany  YearsInCurrentRole  \
count      1470.000000     1470.000000         1470.000000
mean          2.761224        7.008163            4.229252
```

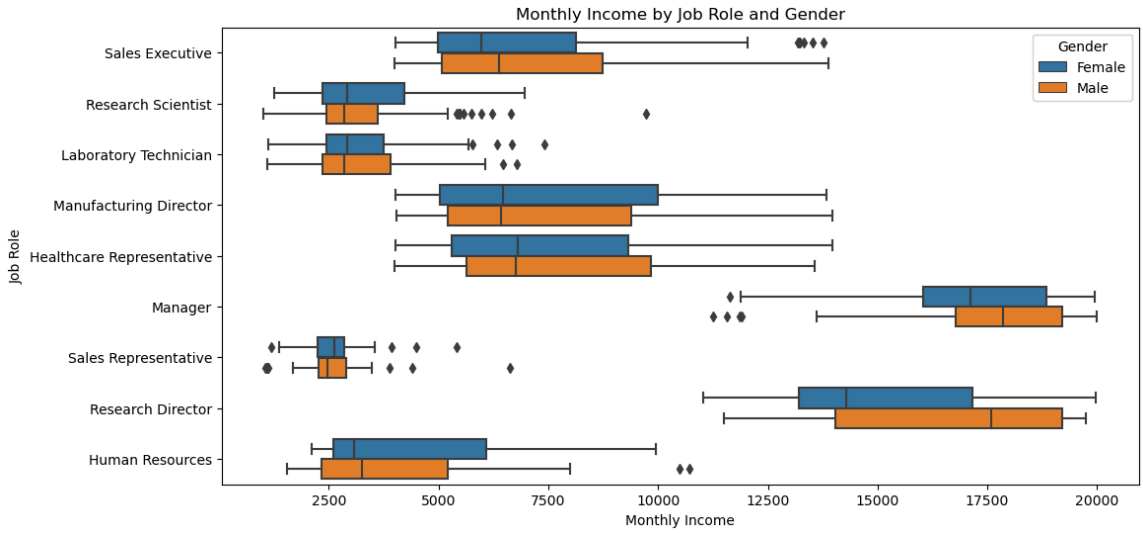
```
std           0.706476        6.126525           3.623137
min           1.000000        0.000000           0.000000
25%           2.000000        3.000000           2.000000
50%           3.000000        5.000000           3.000000
75%           3.000000        9.000000           7.000000
max           4.000000       40.000000          18.000000

       YearsSinceLastPromotion  YearsWithCurrManager
count              1470.000000           1470.000000
mean                  2.187755              4.123129
std                   3.222430              3.568136
min                   0.000000              0.000000
25%                   0.000000              2.000000
50%                   1.000000              3.000000
75%                   3.000000              7.000000
max                  15.000000             17.000000

[8 rows x 26 columns]
```

In [6]:
```
plt.figure(figsize=(8, 4))
sns.histplot(df['Age'], bins=30, kde=True)
plt.title('Age Distribution of Employees')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
```

In [7]:
```
plt.figure(figsize=(10, 6))
sns.countplot(y='JobRole', data=df, order = df['JobRole'].value_counts().in
plt.title('Number of Employees by Job Role')
plt.xlabel('Counts')
plt.ylabel('Job Role')
plt.show()
```

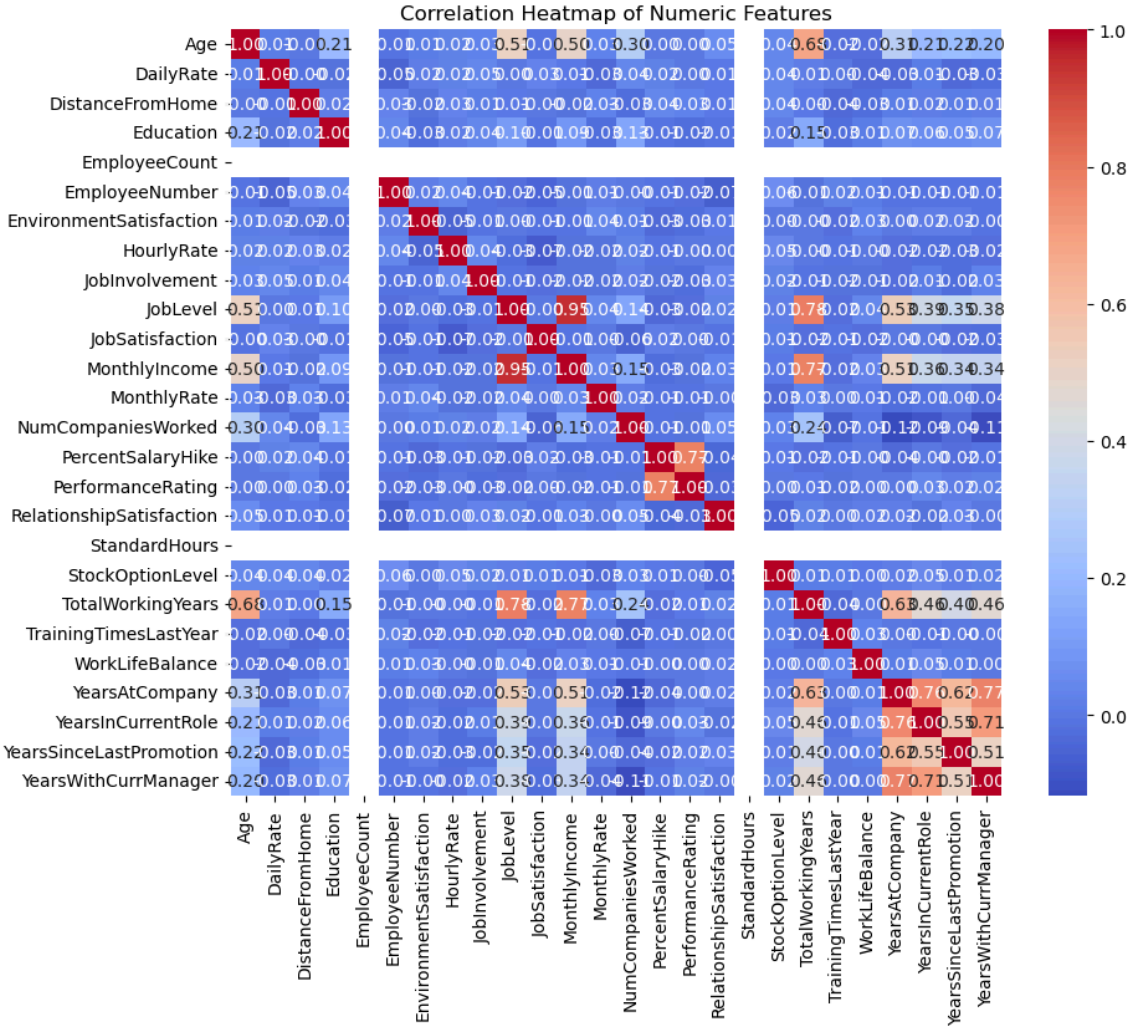


In [8]:
```
plt.figure(figsize=(12, 6))
sns.boxplot(x='MonthlyIncome', y='JobRole', hue='Gender', data=df)
plt.title('Monthly Income by Job Role and Gender')
plt.xlabel('Monthly Income')
plt.ylabel('Job Role')
plt.legend(title='Gender')
plt.show()
```

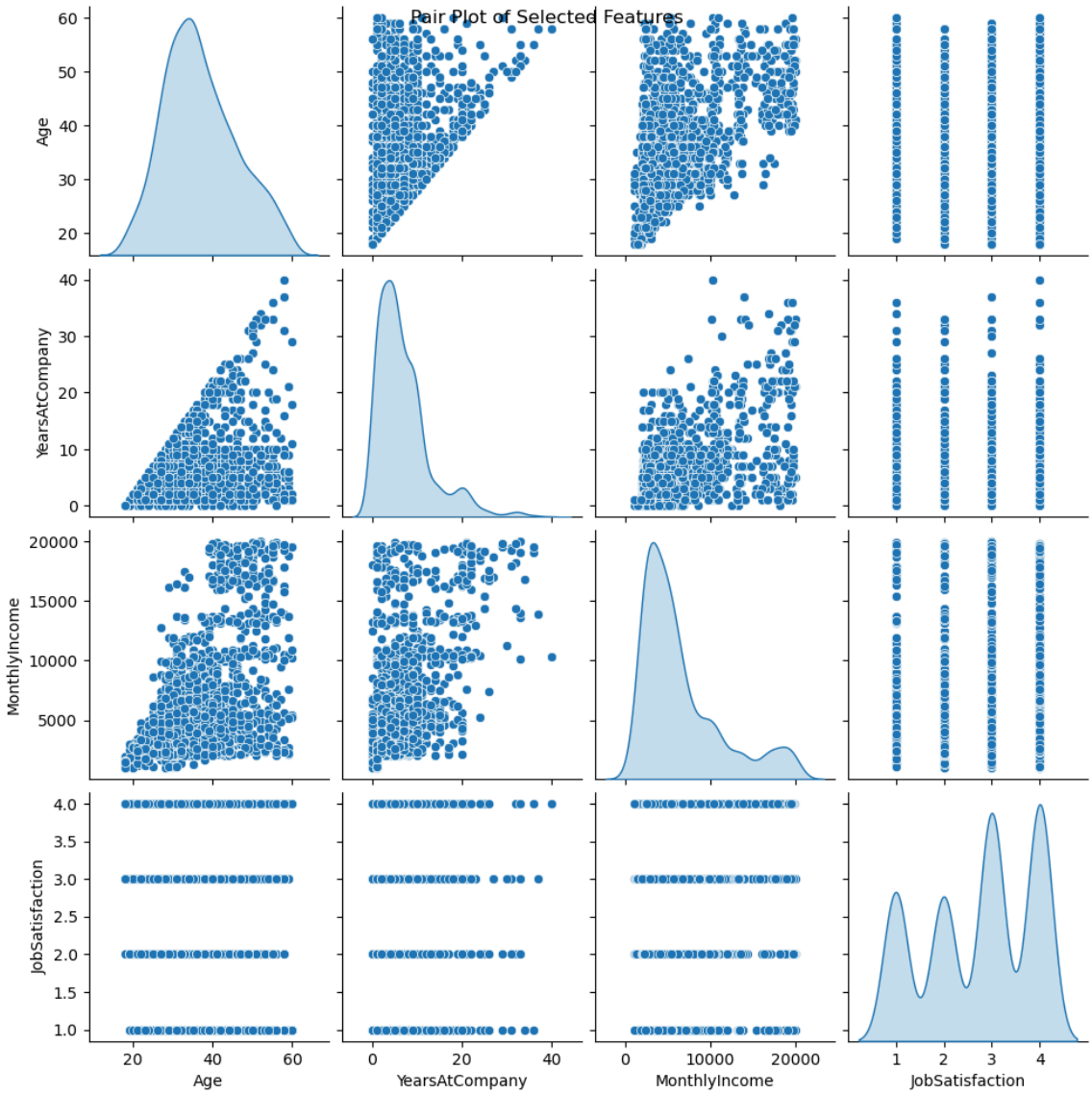


In [9]:
```
plt.figure(figsize=(10, 8))
corr = df.corr()
sns.heatmap(corr, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Heatmap of Numeric Features')
plt.show()
```

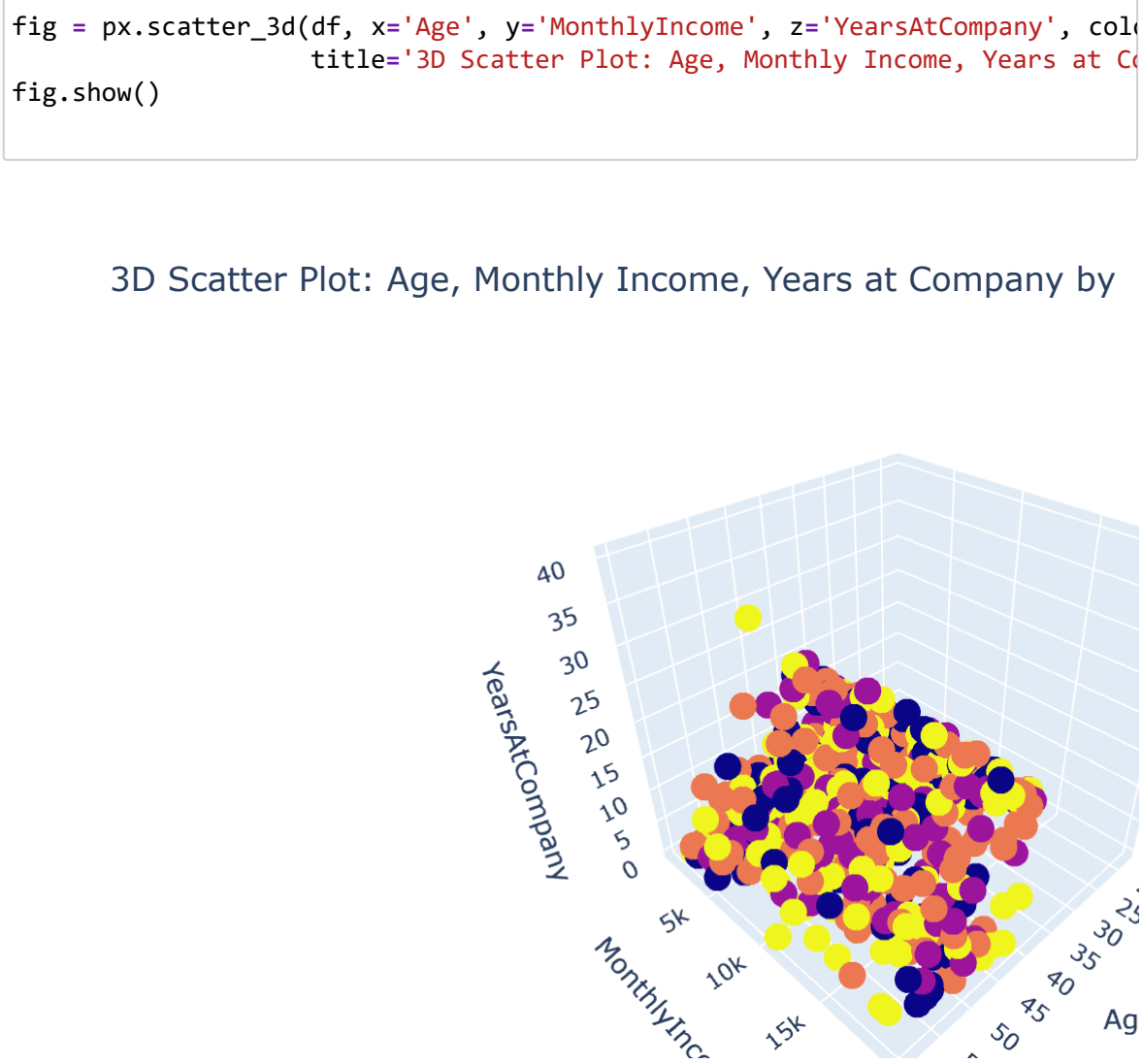
```
C:\Users\Aditya Kudva\AppData\Local\Temp\ipykernel_26404\2903477720.py:2:
FutureWarning: The default value of numeric_only in DataFrame.corr is depr
ecated. In a future version, it will default to False. Select only valid c
olumns or specify the value of numeric_only to silence this warning.
  corr = df.corr()
```



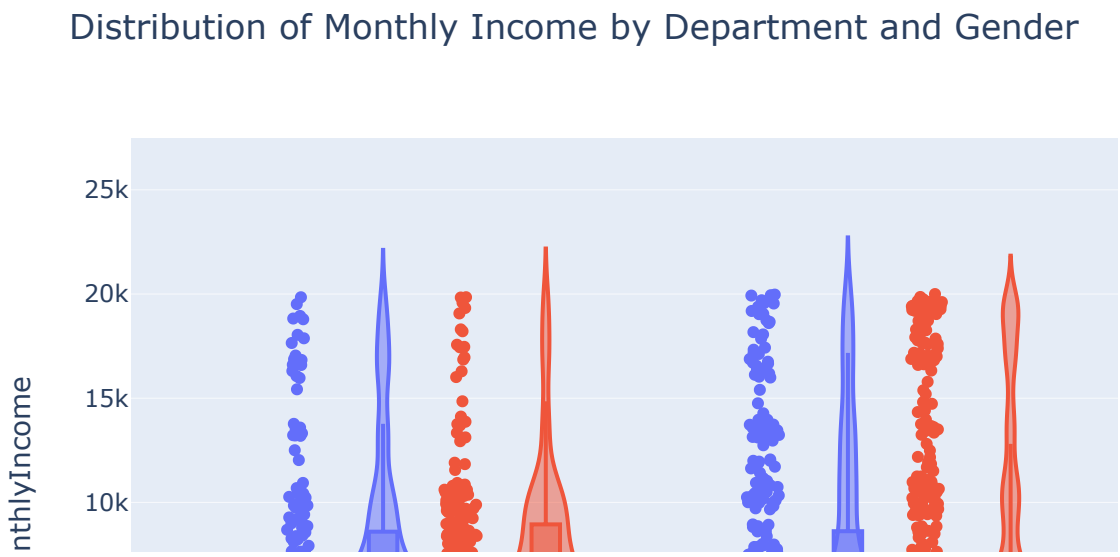
In [10]:
```
# Selecting a subset of features for pairplot to make it more readable
sns.pairplot(df[['Age', 'YearsAtCompany', 'MonthlyIncome', 'JobSatisfaction
plt.suptitle('Pair Plot of Selected Features')
plt.show()
```



In [11]:
```
import plotly.express as px

fig = px.scatter_3d(df, x='Age', y='MonthlyIncome', z='YearsAtCompany', col
                    title='3D Scatter Plot: Age, Monthly Income, Years at C
fig.show()
```



In [12]:
```
fig = px.violin(df, y='MonthlyIncome', x='Department', color='Gender', box='
                title='Distribution of Monthly Income by Department and Gend
fig.show()
```



In [ ]: