```
!pip install pycaret
```

```
Requirement already satisfied: orjson<4.0.0,>=3.8.0 in /usr/local/lib/python3.10/dist-packages (from plotly-resampler>=0.8.3.1->pycaret) (3.10.9)
Requirement already satisfied: tsdownsample>=0.1.3 in /usr/local/lib/python3.10/dist-packages (from plotly-resampler>=0.8.3.1->pycaret) (0.1.3)
Requirement already satisfied: Cython!=0.29.18,!=0.29.31,>=0.29 in /usr/local/lib/python3.10/dist-packages (from pmdarima>=2.0.4->pycaret) (3.0.11)
Requirement already satisfied: urllib3 in /usr/local/lib/python3.10/dist-packages (from pmdarima>=2.0.4->pycaret) (2.2.3)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests>=2.27.1->pycaret) (3.4.0)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests>=2.27.1->pycaret) (3.10)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests>=2.27.1->pycaret) (2024.8.30)
Requirement already satisfied: Flask<3.1,>=1.0.4 in /usr/local/lib/python3.10/dist-packages (from dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret)
Requirement already satisfied: Werkzeug<3.1 in /usr/local/lib/python3.10/dist-packages (from dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (3.0.
Requirement already satisfied: dash-html-components==2.0.0 in /usr/local/lib/python3.10/dist-packages (from dash>=2.9.0->plotly-resampler>=0.8.3.1-
Requirement already satisfied: dash-core-components==2.0.0 in /usr/local/lib/python3.10/dist-packages (from dash>=2.9.0->plotly-resampler>=0.8.3.1-
Requirement already satisfied: dash-table==5.0.0 in /usr/local/lib/python3.10/dist-packages (from dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret)
Requirement already satisfied: typing-extensions>=4.1.1 in /usr/local/lib/python3.10/dist-packages (from dash>=2.9.0->plotly-resampler>=0.8.3.1->py
Requirement already satisfied: retrying in /usr/local/lib/python3.10/dist-packages (from dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (1.3.4)
Requirement already satisfied: nest-asyncio in /usr/local/lib/python3.10/dist-packages (from dash>=2.9.0->plotly-resampler>=0.8.3.1->pycaret) (1.6.
Requirement already satisfied: jupyter-client in /usr/local/lib/python3.10/dist-packages (from ipykernel>=4.5.1->ipywidgets>=7.6.5->pycaret) (6.1.1
Requirement already satisfied: tornado>=4.2 in /usr/local/lib/python3.10/dist-packages (from ipykernel>=4.5.1->ipywidgets>=7.6.5->pycaret) (6.3.3)
Requirement already satisfied: parso<0.9.0,>=0.8.3 in /usr/local/lib/python3.10/dist-packages (from jedi>=0.16->ipython>=5.5.0->pycaret) (0.8.4)
Requirement already satisfied: attrs>=22.2.0 in /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6->nbformat>=4.2.0->pycaret) (24.2.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6->nbformat>=4.2
Requirement already satisfied: referencing>=0.28.4 in /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6->nbformat>=4.2.0->pycaret) (0.3
Requirement already satisfied: rpds-py>=0.7.1 in /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6->nbformat>=4.2.0->pycaret) (0.20.0)
Requirement already satisfied: platformdirs>=2.5 in /usr/local/lib/python3.10/dist-packages (from jupyter-core!=5.0.*,>=4.12->nbformat>=4.2.0->pyca
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages (from patsy>=0.5.1->category-encoders>=2.4.0->pycaret) (1.16.0)
Requirement already satisfied: ptyprocess>=0.5 in /usr/local/lib/python3.10/dist-packages (from pexpect>4.3->ipython>=5.5.0->pycaret) (0.7.0)
Requirement already satisfied: wcwidth in /usr/local/lib/python3.10/dist-packages (from prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.0.0->ipython>=5.5.
Requirement already satisfied: notebook>=4.4.1 in /usr/local/lib/python3.10/dist-packages (from widgetsnbextension~=3.6.0->ipywidgets>=7.6.5->pycar
Requirement already satisfied: itsdangerous>=2.0 in /usr/local/lib/python3.10/dist-packages (from Flask<3.1,>=1.0.4->dash>=2.9.0->plotly-resampler>
Requirement already satisfied: click>=8.0 in /usr/local/lib/python3.10/dist-packages (from Flask<3.1,>=1.0.4->dash>=2.9.0->plotly-resampler>=0.8.3.
Requirement already satisfied: pyzmq<25,>=17 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidget
Requirement already satisfied: argon2-cffi in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets>
Requirement already satisfied: nbconvert>=5 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets
Requirement already satisfied: Send2Trash>=1.8.0 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywi
Requirement already satisfied: terminado>=0.8.3 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywid
Requirement already satisfied: prometheus-client in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywi
Requirement already satisfied: nbclassic>=0.4.7 in /usr/local/lib/python3.10/dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywid
Requirement already satisfied: notebook-shim>=0.2.3 in /usr/local/lib/python3.10/dist-packages (from nbclassic>=0.4.7->notebook>=4.4.1->widgetsnbex
Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipyw
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3
Requirement already satisfied: bleach in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ip
Requirement already satisfied: defusedxml in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0
Requirement already satisfied: entrypoints>=0.2.2 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextensio
Requirement already satisfied: jupyterlab-pygments in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextensi
Requirement already satisfied: mistune<2,>=0.8.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension
Requirement already satisfied: nbclient>=0.5.0 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=
Requirement already satisfied: pandocfilters>=1.4.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextens
Requirement already satisfied: tinycss2 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->
Requirement already satisfied: argon2-cffi-bindings in /usr/local/lib/python3.10/dist-packages (from argon2-cffi->notebook>=4.4.1->widgetsnbextensi
Requirement already satisfied: jupyter-server<3,>=1.8 in /usr/local/lib/python3.10/dist-packages (from notebook-shim>=0.2.3->nbclassic>=0.4.7->note
Requirement already satisfied: cffi>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from argon2-cffi-bindings->argon2-cffi->notebook>=4.4.1->wid
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/dist-packages (from beautifulsoup4->nbconvert>=5->notebook>=4.4.1->widget
Requirement already satisfied: webencodings in /usr/local/lib/python3.10/dist-packages (from bleach->nbconvert>=5->notebook>=4.4.1->widgetsnbextens
Requirement already satisfied: pycparser in /usr/local/lib/python3.10/dist-packages (from cffi>=1.0.1->argon2-cffi-bindings->argon2-cffi->notebook>
Requirement already satisfied: anyio<4,>=3.1.0 in /usr/local/lib/python3.10/dist-packages (from jupyter-server<3,>=1.8->notebook-shim>=0.2.3->nbcla
Requirement already satisfied: websocket-client in /usr/local/lib/python3.10/dist-packages (from jupyter-server<3,>=1.8->notebook-shim>=0.2.3->nbcl
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.10/dist-packages (from anyio<4,>=3.1.0->jupyter-server<3,>=1.8->notebook-shim
Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/dist-packages (from anyio<4,>=3.1.0->jupyter-server<3,>=1.8->notebook-sh
```

```
import pandas as pd
import pycaret
import numpy as np
import scipy
import sklearn
from pycaret.classification import setup
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split, GridSearchCV, cross_val_score
from sklearn.metrics import classification_report, confusion_matrix, roc_curve, auc
import matplotlib.pyplot as plt


from google.colab import drive
drive.mount("/content/drive")


df = pd.read_csv("/content/drive/MyDrive/df_final.csv")
#df = pd.read_csv("df_final.csv")
```

Exploratory Analysis (EDA)

```
df.head()
```

	Intuition_Encoded	Age	Income	Employment_Status	High_Expectation	Industry_Experience	Highest_Degree_Ordinal	Hours_Learning_Weekly	Months_Prog
0	3.0	35.0	2.0	1.0	1.0	1.0	4.0	2.0	
1	3.0	27.0	2.0	1.0	1.0	1.0	4.0	10.0	
2	3.0	24.0	2.0	1.0	1.0	1.0	6.0	5.0	
3	3.0	44.0	2.0	1.0	1.0	1.0	3.0	8.0	
4	3.0	21.0	2.0	1.0	1.0	1.0	4.0	42.0	

5 rows × 29 columns

```
print(df.isnull().sum())
```

```
Intuition_Encoded                          0
Age                                        0
Income                                     0
Employment_Status                          0
High_Expectation                           0
Industry_Experience                        0
Highest_Degree_Ordinal                     0
Hours_Learning_Weekly                      0
Months_Programming                         0
Money_Spent                                0
Count_Learning_Methods                     0
Count_Online_Resources                     0
In-person Events                           0
Listen_Podcasts                            0
Youtube_Channels                           0
Months_Finding_New_Job                     0
Laid_Off_Potential                         0
Replacable_Job_Potential                   0
Study_Field_Computer-related               0
Study_Field_Not applicable                 0
Study_Field_Other Science & Engineering    0
Study_Field_Others                         0
Field_Working_Education                    0
Field_Working_Others                       0
Field_Working_Self-employed                0
Field_Working_Software development and IT  0
Field_Working_unemployed                   0
Job_Status_Expecation(Objective2)          0
Job_Status_Income(Objective1)              0
dtype: int64
```

```
print(df.describe())
```

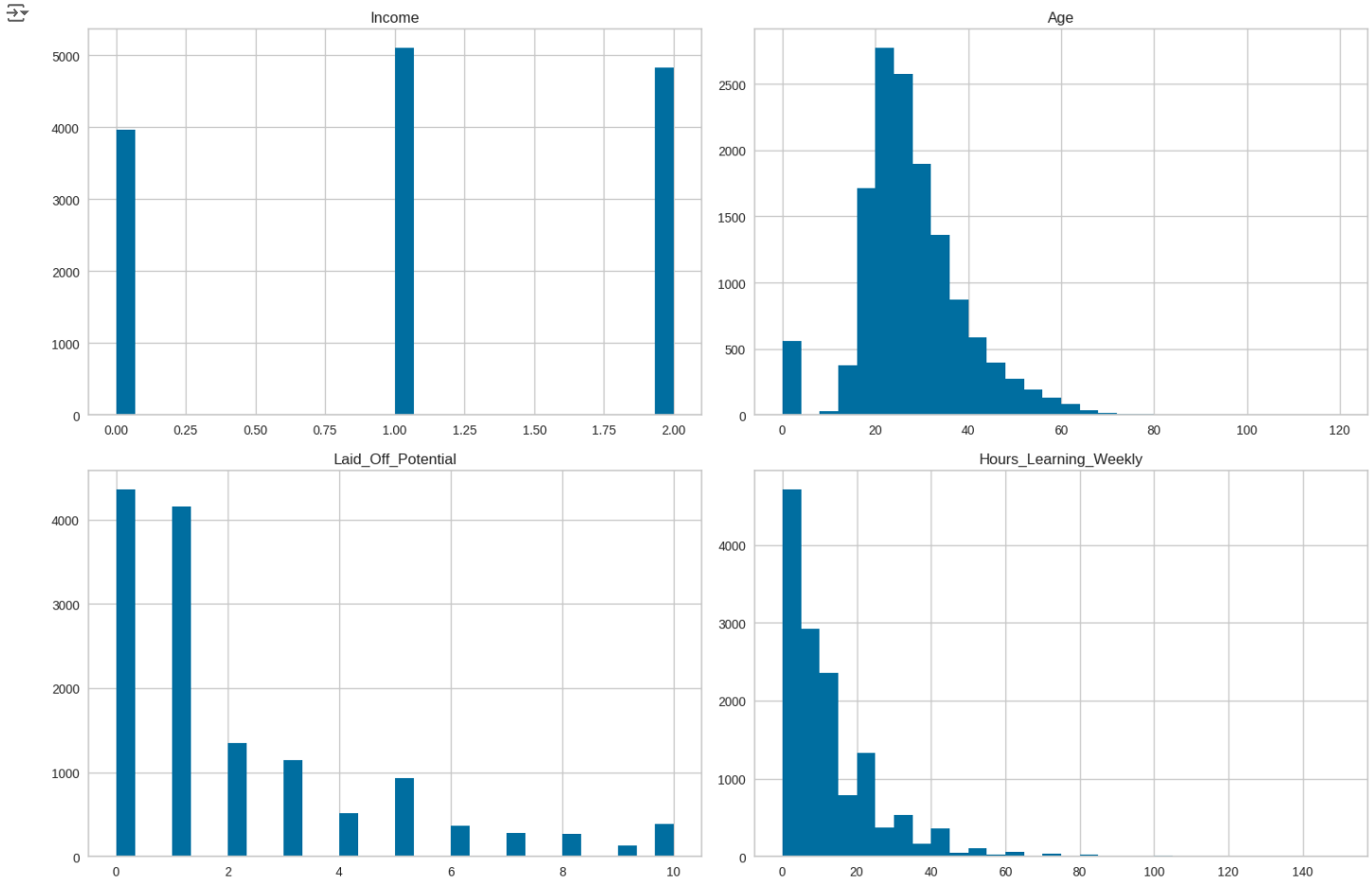
```
std               13.090505          107.847497  2.170295e+04  ...
min                0.000000            0.000000  0.000000e+00  ...
25%                3.000000            1.000000  0.000000e+00  ...
50%                7.000000            3.000000  0.000000e+00  ...
75%               15.000000           12.000000  7.000000e+01  ...
max              150.000000         9000.000000  2.500001e+06  ...

       Study_Field_Not applicable  Study_Field_Other Science & Engineering  \
count                13898.000000                             13898.000000
mean                     0.253274                                 0.275004
std                      0.434902                                 0.446532
min                      0.000000                                 0.000000
25%                      0.000000                                 0.000000
50%                      0.000000                                 0.000000
75%                      1.000000                                 1.000000
max                      1.000000                                 1.000000

       Study_Field_Others  Field_Working_Education  Field_Working_Others  \
count        13898.000000             13898.000000          13898.000000
mean             0.287523                 0.054540              0.380990
std              0.452624                 0.227089              0.485648
min              0.000000                 0.000000              0.000000
25%              0.000000                 0.000000              0.000000
```

```
       Job_Status_Income(Objective1)
count                   13898.000000
mean                        1.489495
std                         1.179597
min                         0.000000
25%                         0.000000
50%                         1.000000
75%                         3.000000
max                         3.000000

[8 rows x 29 columns]
```

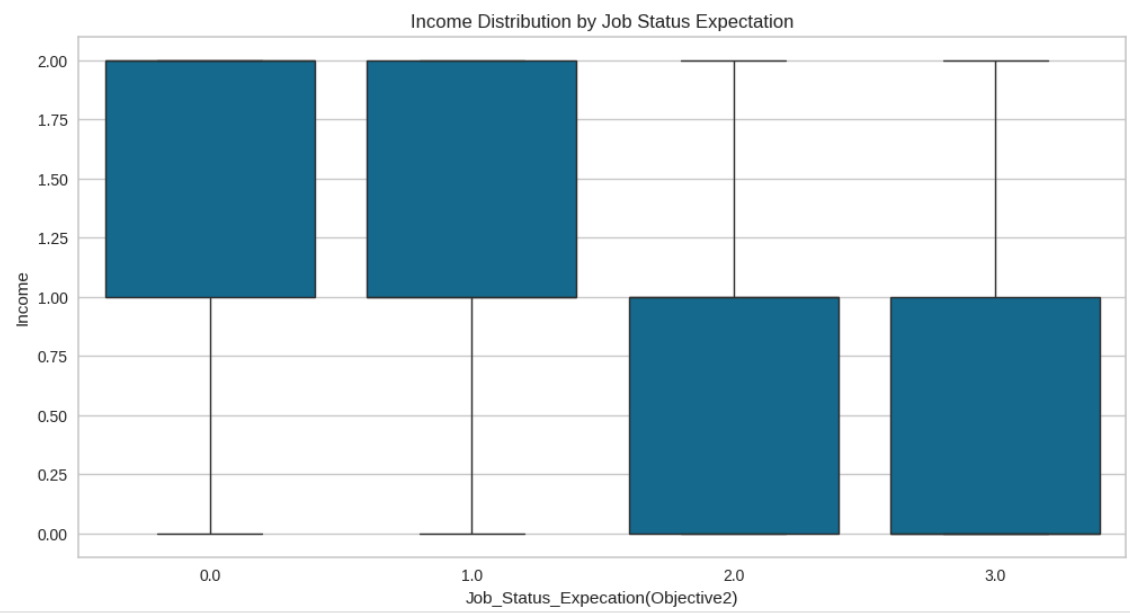
```
print(df["Job_Status_Expecation(Objective2)"].value_counts())
```

```
Job_Status_Expecation(Objective2)
0.0    5555
2.0    3525
3.0    2711
1.0    2107
Name: count, dtype: int64
```

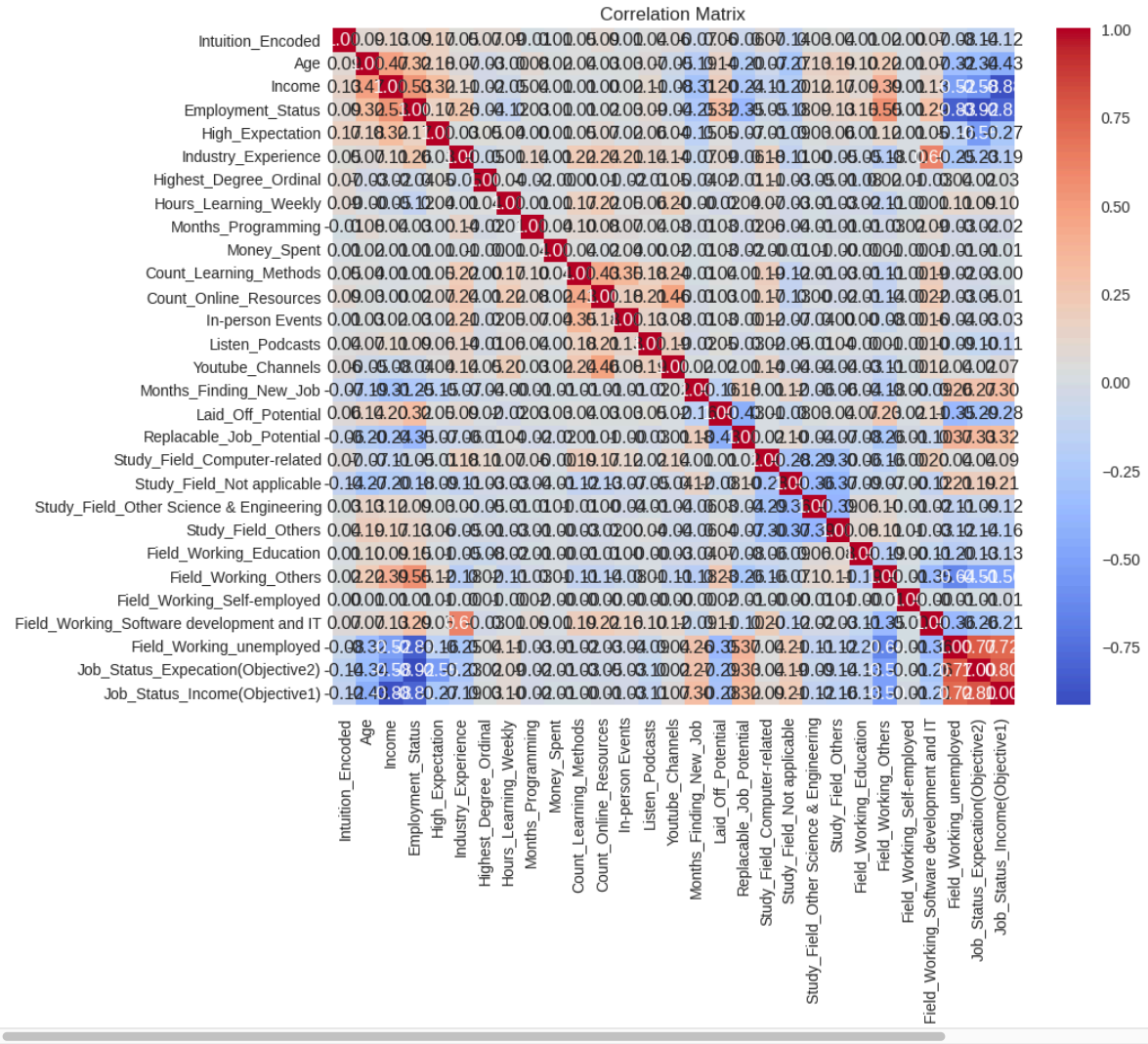
```
import matplotlib.pyplot as plt
import seaborn as sns

numerical_features = ["Income", "Age", "Laid_Off_Potential", "Hours_Learning_Weekly"]
df[numerical_features].hist(bins=30, figsize=(15, 10))
plt.tight_layout()
plt.show()
```



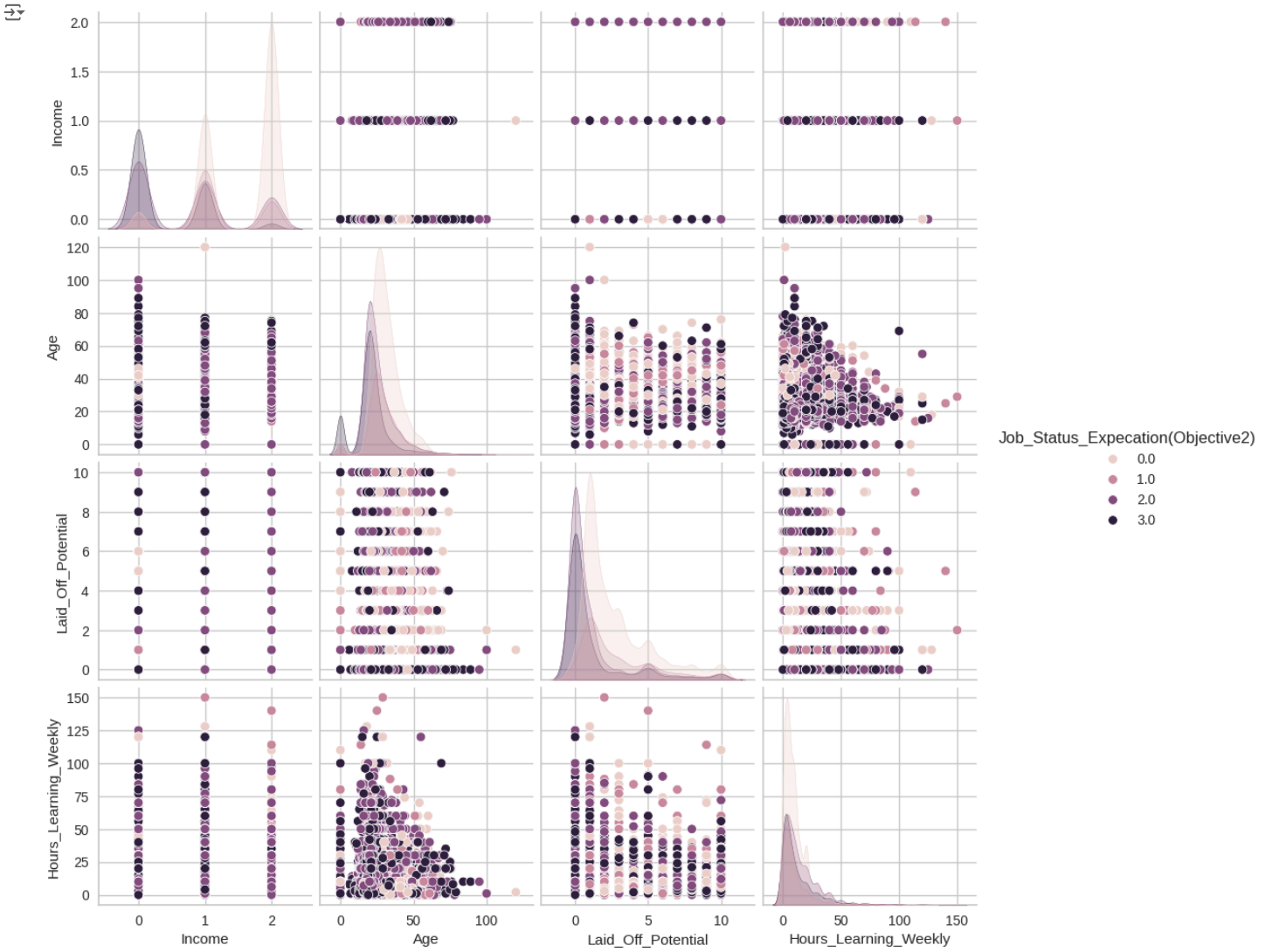

```
plt.figure(figsize=(12, 6))
sns.boxplot(x = "Job_Status_Expecation(Objective2)", y = "Income", data = df)
plt.title("Income Distribution by Job Status Expectation")
plt.show()
```



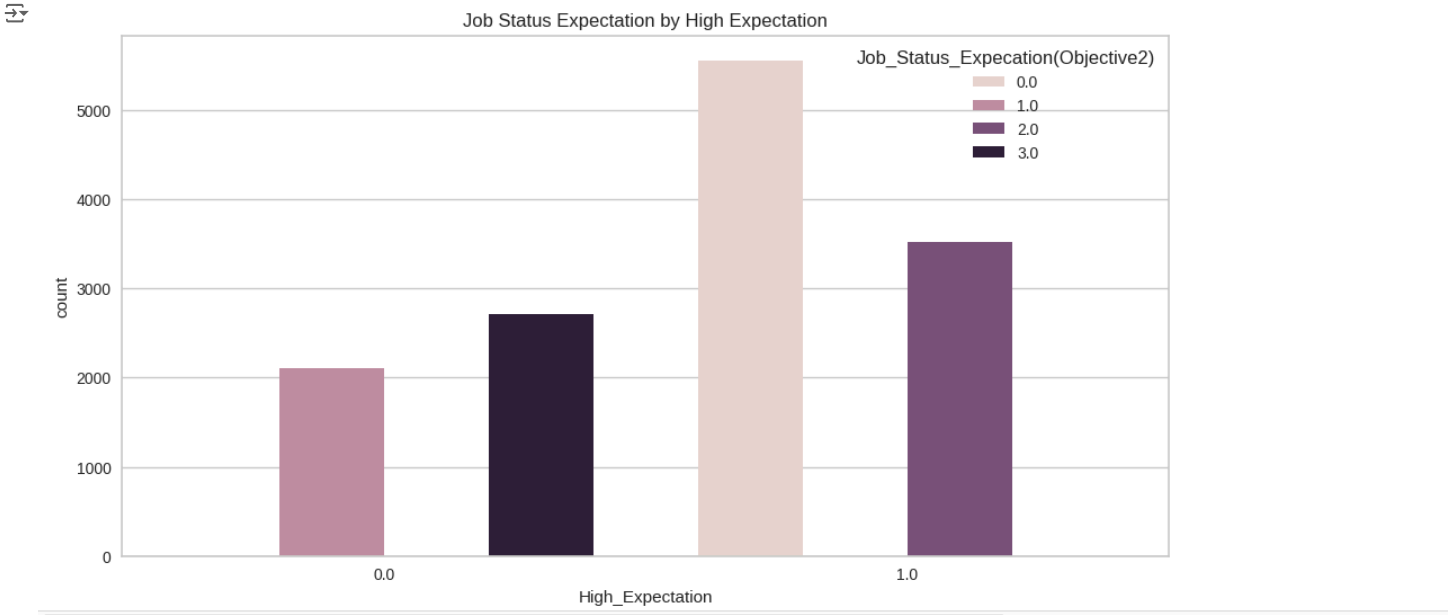

```
plt.figure(figsize = (10, 8))
sns.heatmap(df.corr(), annot = True, fmt = ".2f", cmap = "coolwarm")
plt.title("Correlation Matrix")
plt.show()
```




```
sns.pairplot(df, hue = "Job_Status_Expecation(Objective2)", vars = numerical_features)
plt.show()
```

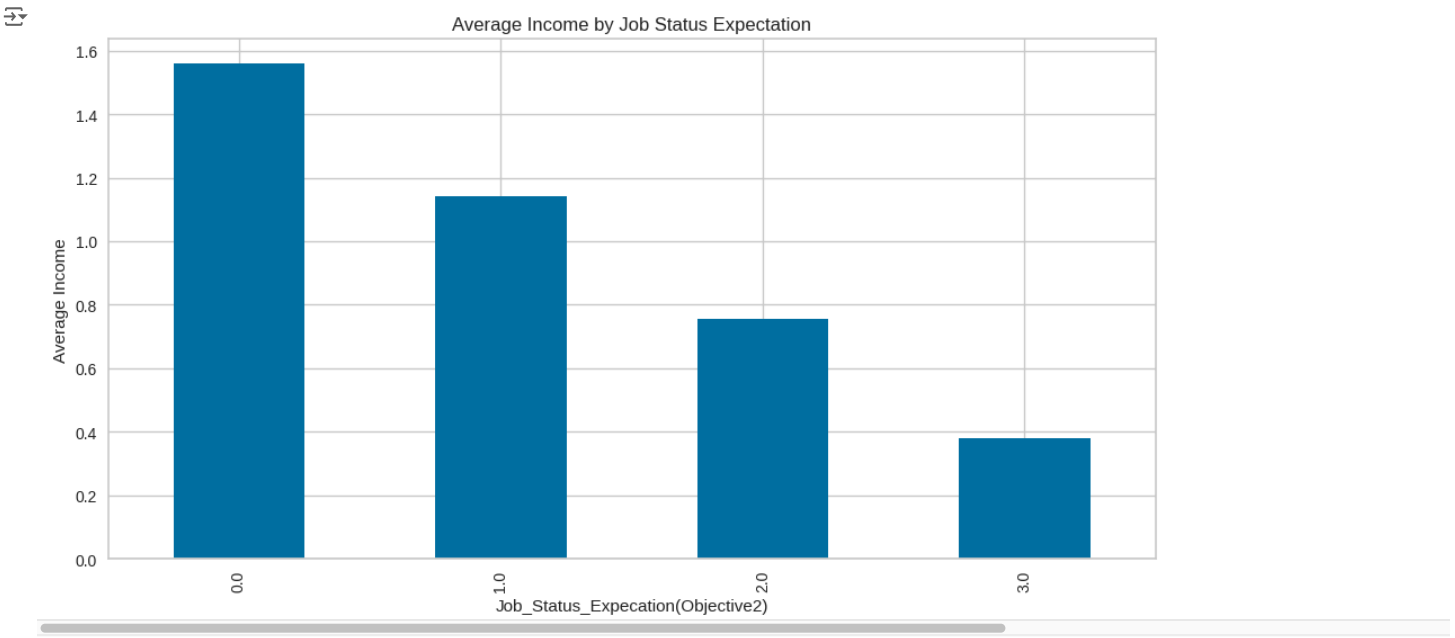



```
plt.figure(figsize = (12, 6))
sns.countplot(x = "High_Expectation", hue = "Job_Status_Expecation(Objective2)", data = df)
plt.title("Job Status Expectation by High Expectation")
plt.show()
```




```
plt.figure(figsize = (12, 6))
df.groupby("Job_Status_Expecation(Objective2)")["Income"].mean().plot(kind = "bar")
```

```
plt.title("Average Income by Job Status Expectation")
plt.ylabel("Average Income")
plt.show()
```



Multiclassification

Objective 2 via Random Forest Classifier

```
# Define target variable and features for Objective 2
y_obj2 = df["Job_Status_Expecation(Objective2)"]
X_obj2 = df.drop(columns=["Job_Status_Income(Objective1)", "Job_Status_Expecation(Objective2)",
                          "Field_Working_Education", "Field_Working_Others",
                          "Field_Working_Self-employed", "Field_Working_Software development and IT",
                          "Field_Working_unemployed", "Employment_Status"])

# Check for correlated features
print(X_obj2.corr())

# Split the data into training and testing sets
X_train_obj2, X_test_obj2, y_train_obj2, y_test_obj2 = train_test_split(X_obj2, y_obj2, random_state=0, stratify=y_obj2)

# Define parameter grid for Grid Search
param_grid_obj2 = {
    "n_estimators": [50, 100, 200],
    "max_depth": [None, 5, 10],
    "min_samples_split": [2, 5, 10]
}

# Perform grid search
grid_search_obj2 = GridSearchCV(RandomForestClassifier(random_state=0), param_grid_obj2, cv=5)
grid_search_obj2.fit(X_train_obj2, y_train_obj2)

# Output the best parameters and score
print("Best parameters for Objective 2 found: ", grid_search_obj2.best_params_)
print("Best cross-validation score for Objective 2: {:.3f}".format(grid_search_obj2.best_score_))

# Train the final model with best parameters
clf_obj2 = RandomForestClassifier(**grid_search_obj2.best_params_).fit(X_train_obj2, y_train_obj2)

# Evaluate the model on training and test sets
print("Accuracy of RF classifier on training set for Objective 2: {:.3f}".format(clf_obj2.score(X_train_obj2, y_train_obj2)))
print("Accuracy of RF classifier on test set for Objective 2: {:.3f}".format(clf_obj2.score(X_test_obj2, y_test_obj2)))

# Predictions and evaluation
y_pred_obj2 = clf_obj2.predict(X_test_obj2)
print("Confusion Matrix:")
print(confusion_matrix(y_test_obj2, y_pred_obj2))
print("Classification Report:")
print(classification_report(y_test_obj2, y_pred_obj2))

# Cross-validation scores
scores_obj2 = cross_val_score(clf_obj2, X_obj2, y_obj2, cv=5)
print("Cross-validation scores for Objective 2:", scores_obj2)
print("Mean cross-validation score for Objective 2:", scores_obj2.mean())

# Calculate predicted probabilities
y_prob_obj2 = clf_obj2.predict_proba(X_test_obj2)
```

```
# Compute ROC curve and AUC for each class
n_classes_obj2 = len(np.unique(y_obj2))  # Number of classes
fpr_obj2 = dict()
tpr_obj2 = dict()
roc_auc_obj2 = dict()

for i in range(n_classes_obj2):
    # Compute ROC curve
    fpr_obj2[i], tpr_obj2[i], _ = roc_curve((y_test_obj2 == np.unique(y_obj2)[i]).astype(int), y_prob_obj2[:, i])
    roc_auc_obj2[i] = auc(fpr_obj2[i], tpr_obj2[i])

# Plot ROC curves
plt.figure(figsize=(10, 8))

for i in range(n_classes_obj2):
    plt.plot(fpr_obj2[i], tpr_obj2[i], lw=2, label="ROC curve for class {0} (area = {1:0.2f})".format(np.unique(y_obj2)[i], roc_auc_obj2[i]))

plt.plot([0, 1], [0, 1], color="red", lw=2, linestyle="--")  # Diagonal line
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel("False Positive Rate")
plt.ylabel("True Positive Rate")
plt.title("Receiver Operating Characteristic for Objective 2")
plt.legend(loc="lower right")
plt.grid()
plt.show()

# Feature Importance
importances_obj2 = clf_obj2.feature_importances_
indices_obj2 = np.argsort(importances_obj2)[::-1]

# Print feature ranking
print("Feature ranking for Objective 2:")
for f in range(X_obj2.shape[1]):
    print(f"{f + 1}. feature {X_obj2.columns[indices_obj2[f]]} (importance: {importances_obj2[indices_obj2[f]]:.4f})")

# Plot feature importances
plt.figure(figsize=(12, 6))
plt.title("Feature Importances for Objective 2")
plt.bar(range(X_obj2.shape[1]), importances_obj2[indices_obj2], align="center")
plt.xticks(range(X_obj2.shape[1]), X_obj2.columns[indices_obj2], rotation=90)
plt.xlim([-1, X_obj2.shape[1]])
plt.show()
```

	Intuition_Encoded	Age
Intuition_Encoded	1.000000	0.090742
Age	0.090742	1.000000
Income	0.131062	0.471341
High_Expectation	0.174528	0.175655
Industry_Experience	0.050012	0.066886
Highest_Degree_Ordinal	0.073184	-0.025134
Hours_Learning_Weekly	0.087854	-0.002358
Months_Programming	-0.007396	0.076105
Money_Spent	0.005001	0.020583
Count_Learning_Methods	0.047164	0.044477
Count_Online_Resources	0.086844	0.034200
In-person Events	0.005315	0.029103
Listen_Podcasts	0.038554	0.067487
Youtube_Channels	0.058031	-0.051340
Months_Finding_New_Job	-0.065653	-0.185327
Laid_Off_Potential	0.064379	0.144076
Replacable_Job_Potential	-0.057078	-0.203498
Study_Field_Computer-related	0.067606	-0.067479
Study_Field_Not applicable	-0.135339	-0.265250
Study_Field_Other Science & Engineering	0.033110	0.128653
Study_Field_Others	0.039473	0.185737

	Income	High_Expectation
Intuition_Encoded	0.131062	0.174528
Age	0.471341	0.175655
Income	1.000000	0.320677
High_Expectation	0.320677	1.000000
Industry_Experience	0.113756	0.026562
Highest_Degree_Ordinal	-0.021749	0.054430
Hours_Learning_Weekly	-0.051388	0.042847
Months_Programming	0.035111	0.000465
Money_Spent	0.008016	0.009563
Count_Learning_Methods	0.005351	0.048319
Count_Online_Resources	0.003865	0.073476
In-person Events	0.023046	0.018133
Listen_Podcasts	0.108968	0.059565
Youtube_Channels	-0.079271	0.042425
Months_Finding_New_Job	-0.307871	-0.150979
Laid_Off_Potential	0.195592	0.047880
Replacable_Job_Potential	-0.241540	-0.067869
Study_Field_Computer-related	-0.114055	-0.005276
Study_Field_Not applicable	-0.203981	-0.090295
Study_Field_Other Science & Engineering	0.121281	0.027416
Study_Field_Others	0.174031	0.064232

	Industry_Experience
Intuition_Encoded	0.050012
Age	0.066886
Income	0.113756
High_Expectation	0.026562
Industry_Experience	1.000000
Highest_Degree_Ordinal	-0.049779
Hours_Learning_Weekly	0.009426
Months_Programming	0.138711
Money_Spent	0.012964
Count_Learning_Methods	0.218269
Count_Online_Resources	0.241969
In-person Events	0.207979
Listen_Podcasts	0.138218
Youtube_Channels	0.142192
Months_Finding_New_Job	-0.074526
Laid_Off_Potential	0.086616
Replacable_Job_Potential	-0.059131
Study_Field_Computer-related	0.175408
Study_Field_Not applicable	-0.109062
Study_Field_Other Science & Engineering	0.000826
Study_Field_Others	-0.046254

	Highest_Degree_Ordinal
Intuition_Encoded	0.073184
Age	-0.025134
Income	-0.021749
High_Expectation	0.054430
Industry_Experience	-0.049779
Highest_Degree_Ordinal	1.000000
Hours_Learning_Weekly	0.035547
Months_Programming	-0.017025
Money_Spent	-0.002959
Count_Learning_Methods	0.003320
Count_Online_Resources	0.014774
In-person Events	-0.015117
Listen_Podcasts	-0.012901
Youtube_Channels	0.046848
Months_Finding_New_Job	-0.044586
Laid_Off_Potential	0.017083
Replacable_Job_Potential	-0.013267
Study_Field_Computer-related	0.112653
Study_Field_Not applicable	-0.034542
Study_Field_Other Science & Engineering	-0.051385
Study_Field_Others	-0.012601

	Hours_Learning_Weekly
Intuition_Encoded	0.087854
Age	-0.002358
Income	-0.051388
High_Expectation	0.042847
Industry_Experience	0.009426
Highest_Degree_Ordinal	0.035547
Hours_Learning_Weekly	1.000000

```
Months_Programming                                0.011929
Money_Spent                                       0.008927
Count_Learning_Methods                            0.169200
Count_Online_Resources                            0.216975
In-person Events                                  0.045737
Listen_Podcasts                                   0.058270
Youtube_Channels                                  0.199963
Months_Finding_New_Job                           -0.002772
Laid_Off_Potential                               -0.021352
Replacable_Job_Potential                          0.040330
Study_Field_Computer-related                      0.070794
Study_Field_Not applicable                       -0.027270
Study_Field_Other Science & Engineering          -0.008946
Study_Field_Others                               -0.025606

                                  Months_Programming  Money_Spent  \
Intuition_Encoded                          -0.007396     0.005001
Age                                         0.076105     0.020583
Income                                      0.035111     0.008016
High_Expectation                            0.000465     0.009563
Industry_Experience                         0.138711     0.012964
Highest_Degree_Ordinal                     -0.017025    -0.002959
Hours_Learning_Weekly                       0.011929     0.008927
Months_Programming                          1.000000     0.036536
Money_Spent                                 0.036536     1.000000
Count_Learning_Methods                      0.100665     0.042062
Count_Online_Resources                      0.079477     0.016566
In-person Events                            0.068390     0.035577
Listen_Podcasts                             0.040186     0.003813
Youtube_Channels                            0.031311     0.023405
Months_Finding_New_Job                     -0.011318    -0.007255
Laid_Off_Potential                          0.032761     0.027032
Replacable_Job_Potential                   -0.021825    -0.021519
Study_Field_Computer-related                0.060797    -0.000275
Study_Field_Not applicable                 -0.036656    -0.009969
Study_Field_Other Science & Engineering    -0.005456     0.011678
Study_Field_Others                         -0.011467    -0.001706

                                  Count_Learning_Methods  ...  \
Intuition_Encoded                               0.047164  ...
Age                                             0.044477  ...
Income                                          0.005351  ...
High_Expectation                                0.048319  ...
Industry_Experience                             0.218269  ...
Highest_Degree_Ordinal                          0.003320  ...
Hours_Learning_Weekly                           0.169200  ...
Months_Programming                              0.100665  ...
Money_Spent                                     0.042062  ...
Count_Learning_Methods                          1.000000  ...
Count_Online_Resources                          0.433880  ...
In-person Events                                0.347024  ...
Listen_Podcasts                                 0.175651  ...
Youtube_Channels                                0.236176  ...
Months_Finding_New_Job                         -0.008991  ...
Laid_Off_Potential                              0.038853  ...
Replacable_Job_Potential                        0.007198  ...
Study_Field_Computer-related                    0.189215  ...
Study_Field_Not applicable                     -0.120510  ...
Study_Field_Other Science & Engineering        -0.014986  ...
Study_Field_Others                             -0.031482  ...

                                  In-person Events  Listen_Podcasts  \
Intuition_Encoded                         0.005315         0.038554
Age                                       0.029103         0.067487
Income                                    0.023046         0.108968
High_Expectation                          0.018133         0.059565
Industry_Experience                       0.207979         0.138218
Highest_Degree_Ordinal                   -0.015117        -0.012901
Hours_Learning_Weekly                     0.045737         0.058270
Months_Programming                        0.068390         0.040186
Money_Spent                               0.035577         0.003813
Count_Learning_Methods                    0.347024         0.175651
Count_Online_Resources                    0.181424         0.214580
In-person Events                          1.000000         0.132269
Listen_Podcasts                           0.132269         1.000000
Youtube_Channels                          0.084551         0.186873
Months_Finding_New_Job                   -0.011455        -0.024116
Laid_Off_Potential                        0.027291         0.053699
Replacable_Job_Potential                 -0.000718        -0.030184
Study_Field_Computer-related              0.115755         0.015514
Study_Field_Not applicable               -0.067439        -0.047360
Study_Field_Other Science & Engineering  -0.037425        -0.006020
Study_Field_Others                        0.002579         0.038158

                                  Youtube_Channels  \
Intuition_Encoded                         0.058031
Age                                      -0.051340
Income                                   -0.079271
High_Expectation                          0.042425
Industry_Experience                       0.142192
Highest_Degree_Ordinal                    0.046848
Hours_Learning_Weekly                     0.199963
Months_Programming                        0.031311
Money_Spent                               0.023405
Count_Learning_Methods                    0.236176
Count_Online_Resources                    0.462027
In-person Events                          0.084551
Listen_Podcasts                           0.186873
Youtube_Channels                          1.000000
Months_Finding_New_Job                    0.020082
Laid_Off_Potential                        0.018605
```

Replacable_Job_Potential	0.012908
Study_Field_Computer-related	0.138035
Study_Field_Not applicable	-0.039804
Study_Field_Other Science & Engineering	-0.036192
Study_Field_Others	-0.044273

	Months_Finding_New_Job
Intuition_Encoded	-0.065653
Age	-0.185327
Income	-0.307871
High_Expectation	-0.150979
Industry_Experience	-0.074526
Highest_Degree_Ordinal	-0.044586
Hours_Learning_Weekly	-0.002772
Months_Programming	-0.011318
Money_Spent	-0.007255
Count_Learning_Methods	-0.008991
Count_Online_Resources	-0.010833
In-person Events	-0.011455
Listen_Podcasts	-0.024116
Youtube_Channels	0.020082
Months_Finding_New_Job	1.000000
Laid_Off_Potential	-0.160414
Replacable_Job_Potential	0.184638
Study_Field_Computer-related	0.005433
Study_Field_Not applicable	0.124407
Study_Field_Other Science & Engineering	-0.062243
Study_Field_Others	-0.062784

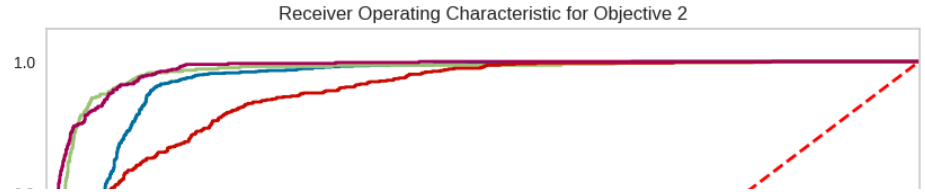
	Laid_Off_Potential
Intuition_Encoded	0.064379
Age	0.144076
Income	0.195592
High_Expectation	0.047880
Industry_Experience	0.086616
Highest_Degree_Ordinal	0.017083
Hours_Learning_Weekly	-0.021352
Months_Programming	0.032761
Money_Spent	0.027032
Count_Learning_Methods	0.038853
Count_Online_Resources	0.033114
In-person Events	0.027291
Listen_Podcasts	0.053699
Youtube_Channels	0.018605
Months_Finding_New_Job	-0.160414
Laid_Off_Potential	1.000000
Replacable_Job_Potential	-0.425111
Study_Field_Computer-related	0.009062
Study_Field_Not applicable	-0.084437
Study_Field_Other Science & Engineering	0.033385
Study_Field_Others	0.040434

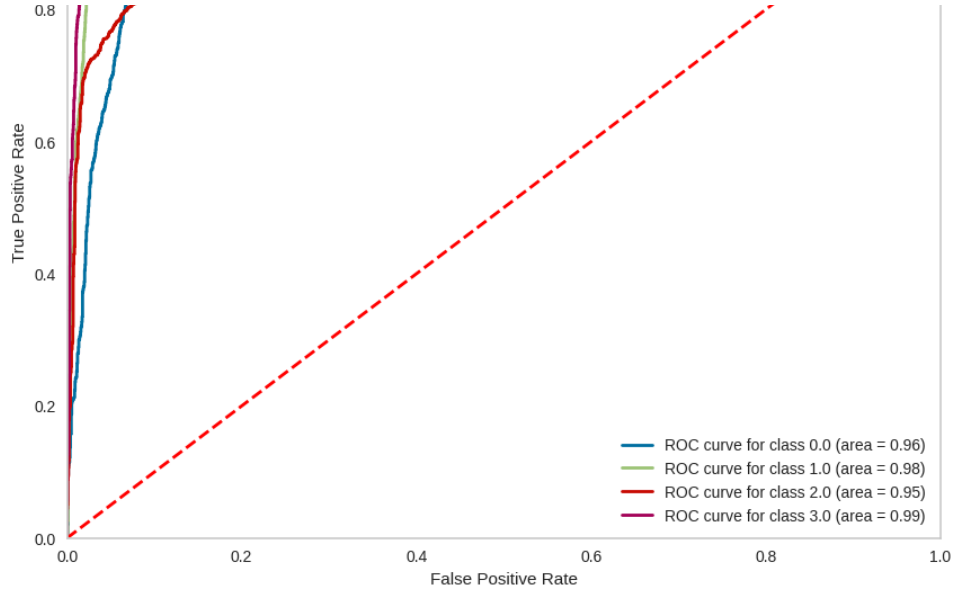
	Replacable_Job_Potential
Intuition_Encoded	-0.057078
Age	-0.203498
Income	-0.241540
High_Expectation	-0.067869
Industry_Experience	-0.059131
Highest_Degree_Ordinal	-0.013267
Hours_Learning_Weekly	0.040330
Months_Programming	-0.021825
Money_Spent	-0.021519
Count_Learning_Methods	0.007198
Count_Online_Resources	0.011354
In-person Events	-0.000718
Listen_Podcasts	-0.030184
Youtube_Channels	0.012908
Months_Finding_New_Job	0.184638
Laid_Off_Potential	-0.425111
Replacable_Job_Potential	1.000000
Study_Field_Computer-related	0.018849
Study_Field_Not applicable	0.096419
Study_Field_Other Science & Engineering	-0.043255
Study_Field_Others	-0.066114

	Study_Field_Computer-related
Intuition_Encoded	0.067606
Age	-0.067479
Income	-0.114055
High_Expectation	-0.005276
Industry_Experience	0.175408
Highest_Degree_Ordinal	0.112653
Hours_Learning_Weekly	0.070794
Months_Programming	0.060797
Money_Spent	-0.000275
Count_Learning_Methods	0.189215
Count_Online_Resources	0.173375
In-person Events	0.115755
Listen_Podcasts	0.015514
Youtube_Channels	0.138035
Months_Finding_New_Job	0.005433
Laid_Off_Potential	0.009062
Replacable_Job_Potential	0.018849
Study_Field_Computer-related	1.000000
Study_Field_Not applicable	-0.276736
Study_Field_Other Science & Engineering	-0.292653
Study_Field_Others	-0.301858

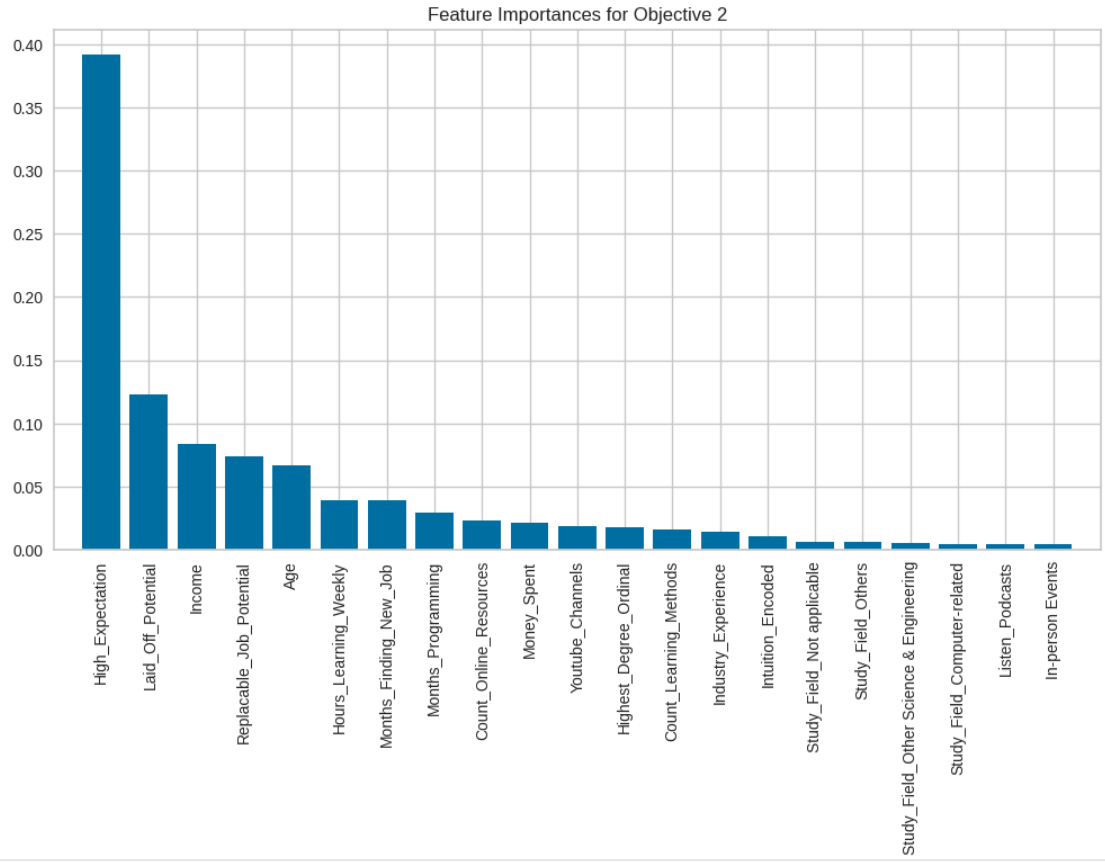
	Study_Field_Not applicable
Intuition_Encoded	-0.135339

```
Age                                                -0.265250
Income                                             -0.203981
High_Expectation                                   -0.090295
Industry_Experience                                -0.109062
Highest_Degree_Ordinal                             -0.034542
Hours_Learning_Weekly                              -0.027270
Months_Programming                                 -0.036656
Money_Spent                                        -0.009969
Count_Learning_Methods                             -0.120510
Count_Online_Resources                             -0.128761
In-person Events                                   -0.067439
Listen_Podcasts                                    -0.047360
Youtube_Channels                                   -0.039804
Months_Finding_New_Job                              0.124407
Laid_Off_Potential                                 -0.084437
Replacable_Job_Potential                            0.096419
Study_Field_Computer-related                       -0.276736
Study_Field_Not applicable                          1.000000
Study_Field_Other Science & Engineering            -0.358687
Study_Field_Others                                 -0.369969

                                   Study_Field_Other Science & Engineering  \
Intuition_Encoded                                                0.033110
Age                                                              0.128653
Income                                                           0.121281
High_Expectation                                                 0.027416
Industry_Experience                                              0.000826
Highest_Degree_Ordinal                                          -0.051385
Hours_Learning_Weekly                                           -0.008946
Months_Programming                                              -0.005456
Money_Spent                                                      0.011678
Count_Learning_Methods                                          -0.014986
Count_Online_Resources                                           0.000177
In-person Events                                                -0.037425
Listen_Podcasts                                                 -0.006020
Youtube_Channels                                                -0.036192
Months_Finding_New_Job                                          -0.062243
Laid_Off_Potential                                               0.033385
Replacable_Job_Potential                                        -0.043255
Study_Field_Computer-related                                    -0.292653
Study_Field_Not applicable                                      -0.358687
Study_Field_Other Science & Engineering                          1.000000
Study_Field_Others                                              -0.391248

                                   Study_Field_Others
Intuition_Encoded                            0.039473
Age                                          0.185737
Income                                       0.174031
High_Expectation                             0.064232
Industry_Experience                         -0.046254
Highest_Degree_Ordinal                      -0.012601
Hours_Learning_Weekly                       -0.025606
Months_Programming                          -0.011467
Money_Spent                                 -0.001706
Count_Learning_Methods                      -0.031482
Count_Online_Resources                      -0.024946
In-person Events                             0.002579
Listen_Podcasts                              0.038158
Youtube_Channels                            -0.044273
Months_Finding_New_Job                      -0.062784
Laid_Off_Potential                           0.040434
Replacable_Job_Potential                    -0.066114
Study_Field_Computer-related                -0.301858
Study_Field_Not applicable                  -0.369969
Study_Field_Other Science & Engineering     -0.391248
Study_Field_Others                           1.000000

[21 rows x 21 columns]
Best parameters for Objective 2 found:  {'max_depth': None, 'min_samples_split': 5, 'n_estimators': 200}
Best cross-validation score for Objective 2: 0.857
Accuracy of RF classifier on training set for Objective 2: 0.973
Accuracy of RF classifier on test set for Objective 2: 0.861
Confusion Matrix:
[[1296    0   93    0]
 [   0  461    0   66]
 [ 232    0  649    0]
 [   0   91    0  587]]
Classification Report:
              precision    recall  f1-score   support

         0.0       0.85      0.93      0.89      1389
         1.0       0.84      0.87      0.85       527
         2.0       0.87      0.74      0.80       881
         3.0       0.90      0.87      0.88       678

    accuracy                           0.86      3475
   macro avg       0.86      0.85      0.86      3475
weighted avg       0.86      0.86      0.86      3475

Cross-validation scores for Objective 2: [0.76906475 0.80395683 0.86906475 0.85858222 0.73839511]
Mean cross-validation score for Objective 2: 0.8078127321820127
```






```
Feature ranking for Objective 2:
1. feature High_Expectation (importance: 0.3919)
2. feature Laid_Off_Potential (importance: 0.1233)
3. feature Income (importance: 0.0842)
4. feature Replacable_Job_Potential (importance: 0.0743)
5. feature Age (importance: 0.0665)
6. feature Hours_Learning_Weekly (importance: 0.0388)
7. feature Months_Finding_New_Job (importance: 0.0388)
8. feature Months_Programming (importance: 0.0295)
9. feature Count_Online_Resources (importance: 0.0232)
10. feature Money_Spent (importance: 0.0212)
11. feature Youtube_Channels (importance: 0.0187)
12. feature Highest_Degree_Ordinal (importance: 0.0178)
13. feature Count_Learning_Methods (importance: 0.0164)
14. feature Industry_Experience (importance: 0.0147)
15. feature Intuition_Encoded (importance: 0.0108)
16. feature Study_Field_Not applicable (importance: 0.0059)
17. feature Study_Field_Others (importance: 0.0058)
18. feature Study_Field_Other Science & Engineering (importance: 0.0052)
19. feature Study_Field_Computer-related (importance: 0.0048)
20. feature Listen_Podcasts (importance: 0.0042)
21. feature In-person Events (importance: 0.0041)
```



Outcome Observations

Model Performance

1. Best Parameters:

- The model selected max_depth = 10, which restricts the depth of the trees. This helps prevent overfitting and allows for better generalization.
- min_samples_split = 10 means that a node must have at least 10 samples to be considered for splitting, which is a common strategy to avoid overfitting.
- n_estimators = 100 indicates that the model uses 100 trees, a reasonable choice that balances performance and computation time.

2. Cross-Validation Score:

- The best cross-validation score of 0.857 indicates that the model performs well across different subsets of the training data, but there is room for improvement.

3. Accuracy:

- The training accuracy is 0.928, suggesting that the model fits the training data reasonably well.
- The test accuracy is 0.863, showing that the model maintains decent performance on unseen data, but it is slightly lower than the training accuracy, which could suggest some overfitting.

Confusion Matrix

- True Positives (TP): High TPs for class 0 and class 3, indicating strong performance in these categories.
- False Negatives (FN): Class 2 has a significant number of FNs (230), suggesting that the model struggles to correctly identify this class.
- False Positives (FP): Class 1 has a notable number of misclassifications (124), indicating that instances of this class are often confused with class 3.

Classification Report

1. Precision, Recall, F1 Score:

- Class 0 shows good precision (0.85) and high recall (0.93), indicating it is well-classified.
- Class 1 has moderate performance with precision (0.79) and recall (0.80), indicating that while it is reasonably classified, there is room for improvement.
- Class 2 has high precision (0.91) and recall (0.83), suggesting it is generally well identified but with some room for improvement in recall.
- Class 3 has good performance with both precision (0.89) and recall (0.89).

2. Macro and Weighted Averages:

- The macro average F1-score of 0.86 shows balanced performance across classes.
- The weighted average F1-score of 0.87 indicates good performance considering the class distribution.

Feature Importance

1. Top Features:

- The most important feature is High_Expectation (importance: 0.4996), suggesting it plays a significant role in determining the job status.
- Other important features include Laid_Off_Potential and Income, indicating these factors are also relevant predictors.

2. Lower Importance Features:

- Features such as Listen_Podcasts, In-person Events, and others have very low importance, suggesting they contribute less to the model's predictions.

Summary

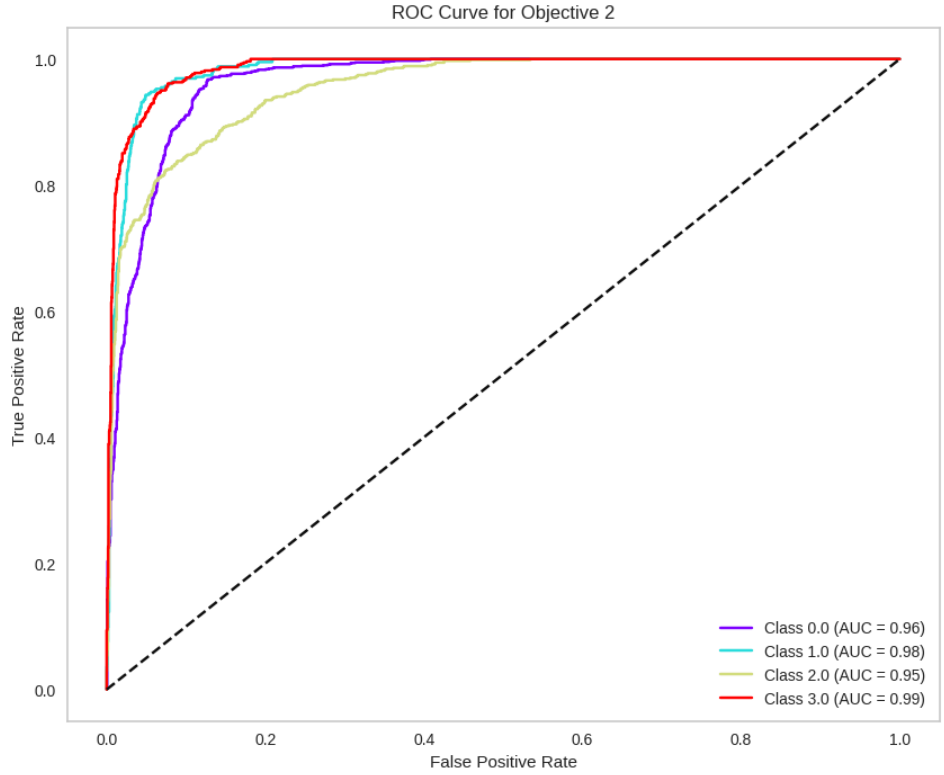
Overall, the Random Forest model for Objective 2 shows solid performance with an accuracy of 0.869 on the test set and a reasonable cross-validation score. The confusion matrix reveals that while some classes are well-identified, others, particularly class 1, could benefit from further refinement. The feature importance analysis highlights key factors influencing predictions, indicating that focusing on "High_Expectation" and related features could lead to improved model performance.

Objective 2 via Gradient Boosting

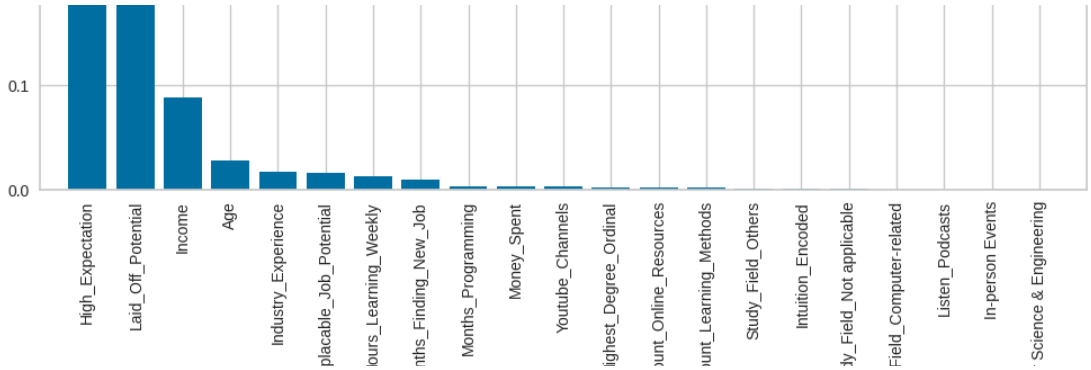
```
from sklearn.impute import SimpleImputer
from sklearn.pipeline import make_pipeline
from sklearn.ensemble import GradientBoostingClassifier

# Define target variable and features for Objective 2
y_obj2 = df["Job_Status_Expecation(Objective2)"]
X_obj2 = df.drop(columns=["Job_Status_Income(Objective1)", "Job_Status_Expecation(Objective2)",
                          "Field_Working_Education", "Field_Working_Others",
                          "Field_Working_Self-employed", "Field_Working_Software development and IT",
                          "Field_Working_unemployed", "Employment_Status"])

# Split the data into training and testing sets
X_train_obj2, X_test_obj2, y_train_obj2, y_test_obj2 = train_test_split(X_obj2, y_obj2, test_size = 0.2, random_state = 0, stratify = y_obj2)

# Create an imputer for categorical data
imputer = SimpleImputer(strategy="most_frequent")

# Define a smaller parameter grid for Grid Search
param_grid = {
```

```
param_grid = {
    "gradientboostingclassifier__n_estimators": [50, 100],
    "gradientboostingclassifier__max_depth": [3, 5],
    "gradientboostingclassifier__learning_rate": [0.1]  # Keeping it simple
}

# Create a pipeline with the imputer and Gradient Boosting Classifier
pipeline_obj2 = make_pipeline(imputer, GradientBoostingClassifier(random_state=0))

# Perform grid search
grid_search_obj2 = GridSearchCV(pipeline_obj2, param_grid, cv=3)  # Reduced to 3 folds
grid_search_obj2.fit(X_train_obj2, y_train_obj2)

# Output the best parameters and score
print("Best parameters found for Objective 2: ", grid_search_obj2.best_params_)
print("Best cross-validation score for Objective 2: {:.3f}".format(grid_search_obj2.best_score_))

# Fit the pipeline with the best parameters
best_pipeline_obj2 = grid_search_obj2.best_estimator_

# Make predictions on the test set
y_pred_obj2 = best_pipeline_obj2.predict(X_test_obj2)

# Generate a classification report
print("Classification Report for Objective 2:")
print(classification_report(y_test_obj2, y_pred_obj2))

# Confusion matrix
conf_matrix_obj2 = confusion_matrix(y_test_obj2, y_pred_obj2)
print("Confusion Matrix for Objective 2:")
print(conf_matrix_obj2)

# Calculate probabilities for ROC Curve
y_pred_proba_obj2 = best_pipeline_obj2.predict_proba(X_test_obj2)

# Plot ROC Curve for each class
plt.figure(figsize=(10, 8))
n_classes_obj2 = len(np.unique(y_obj2))  # Number of classes
colors = plt.cm.rainbow(np.linspace(0, 1, n_classes_obj2))  # Generate colors

for i in range(n_classes_obj2):
    fpr, tpr, _ = roc_curve(y_test_obj2 == np.unique(y_obj2)[i], y_pred_proba_obj2[:, i])  # True positive rate and false positive rate
    roc_auc = auc(fpr, tpr)  # Area under the curve
    plt.plot(fpr, tpr, color=colors[i], label=f"Class {np.unique(y_obj2)[i]} (AUC = {roc_auc:.2f})")

plt.plot([0, 1], [0, 1], "k--")  # Diagonal line
plt.xlabel("False Positive Rate")
plt.ylabel("True Positive Rate")
plt.title("ROC Curve for Objective 2")
plt.legend(loc = "best")
plt.grid()
plt.show()

# Visualize feature importances
feature_importances_obj2 = best_pipeline_obj2.named_steps["gradientboostingclassifier"].feature_importances_
indices = np.argsort(feature_importances_obj2)[::-1]

# Print feature ranking
print("Feature ranking for Objective 2:")
for f in range(X_obj2.shape[1]):
    print(f"{f + 1}. feature {X_obj2.columns[indices[f]]} (importance: {feature_importances_obj2[indices[f]]:.4f})")

# Plot feature importances
plt.figure(figsize=(12, 6))
plt.title("Feature Importances from Gradient Boosting Classifier (Objective 2)")
plt.bar(range(X_obj2.shape[1]), feature_importances_obj2[indices], align="center")
plt.xticks(range(X_obj2.shape[1]), X_obj2.columns[indices], rotation=90)
plt.xlim([-1, X_obj2.shape[1]])
plt.show()
```

```
Best parameters found for Objective 2:  {'gradientboostingclassifier__learning_rate': 0.1, 'gradientboostingclassifier__max_depth': 5, 'gradientboost
Best cross-validation score for Objective 2: 0.858
Classification Report for Objective 2:
              precision    recall  f1-score   support

         0.0       0.85      0.93      0.89      1111
         1.0       0.82      0.88      0.85       422
         2.0       0.87      0.74      0.80       705
         3.0       0.90      0.85      0.88       542

    accuracy                           0.86      2780
   macro avg       0.86      0.85      0.85      2780
weighted avg       0.86      0.86      0.86      2780

Confusion Matrix for Objective 2:
[[1033    0   78    0]
 [   0  372    0   50]
 [ 180    0  525    0]
 [   0   81    0  461]]
```




```
Feature ranking for Objective 2:
1. feature High_Expectation (importance: 0.4735)
2. feature Laid_Off_Potential (importance: 0.3306)
3. feature Income (importance: 0.0881)
4. feature Age (importance: 0.0281)
5. feature Industry_Experience (importance: 0.0176)
6. feature Replacable_Job_Potential (importance: 0.0165)
7. feature Hours_Learning_Weekly (importance: 0.0130)
8. feature Months_Finding_New_Job (importance: 0.0097)
9. feature Months_Programming (importance: 0.0035)
10. feature Money_Spent (importance: 0.0033)
11. feature Youtube_Channels (importance: 0.0030)
12. feature Highest_Degree_Ordinal (importance: 0.0027)
13. feature Count_Online_Resources (importance: 0.0025)
14. feature Count_Learning_Methods (importance: 0.0024)
15. feature Study_Field_Others (importance: 0.0016)
16. feature Intuition_Encoded (importance: 0.0014)
17. feature Study_Field_Not applicable (importance: 0.0007)
18. feature Study_Field_Computer-related (importance: 0.0006)
19. feature Listen_Podcasts (importance: 0.0004)
20. feature In-person Events (importance: 0.0004)
21. feature Study_Field_Other Science & Engineering (importance: 0.0003)
```


0.1
0.0
High_Expectation
Laid_Off_Potential
Income
Age
Industry_Experience
placable_Job_Potential
lours_Learning_Weekly
nths_Finding_New_Job
Months_Programming
Money_Spent
Youtube_Channels
ighest_Degree_Ordinal
unt_Online_Resources
unt_Learning_Methods
Study_Field_Others
Intuition_Encoded
ly_Field_Not applicable
Field_Computer-related
Listen_Podcasts
In-person Events
Science & Engineering