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

from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import OneHotEncoder
from sklearn.preprocessing import OrdinalEncoder
from sklearn.preprocessing import StandardScaler

train_df=pd.read_csv('train.csv')
test_df=pd.read_csv('test.csv')

train_df=train_df.drop("id",axis='columns')
test_df=test_df.drop("id",axis='columns')

train_df.head()
```


```
{"summary":"{\n  \"name\": \"train_df\",\n  \"rows\": 15533,\n
\"fields\": [\n    {\n      \"column\": \"Gender\",\n
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n
\"Female\",\n          \"Male\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Age\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 5.663166982529945,\n
\"min\": 14.0,\n        \"max\": 61.0,\n        \"num_unique_values\":
1602,\n        \"samples\": [\n          21.501721,\n
19.054008\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Height\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 0.08767002604931125,\n        \"min\": 1.45,\n
\"max\": 1.975663,\n        \"num_unique_values\": 1723,\n
\"samples\": [\n          1.75091,\n          1.773594\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Weight\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\":
26.369144218105923,\n        \"min\": 39.0,\n        \"max\":
165.057269,\n        \"num_unique_values\": 1836,\n
\"samples\": [\n          120.699119,\n          131.572387\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"family_history_with_overweight\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          \"no\",\n          \"yes\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"FAVC\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          \"no\",\n          \"yes\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
```

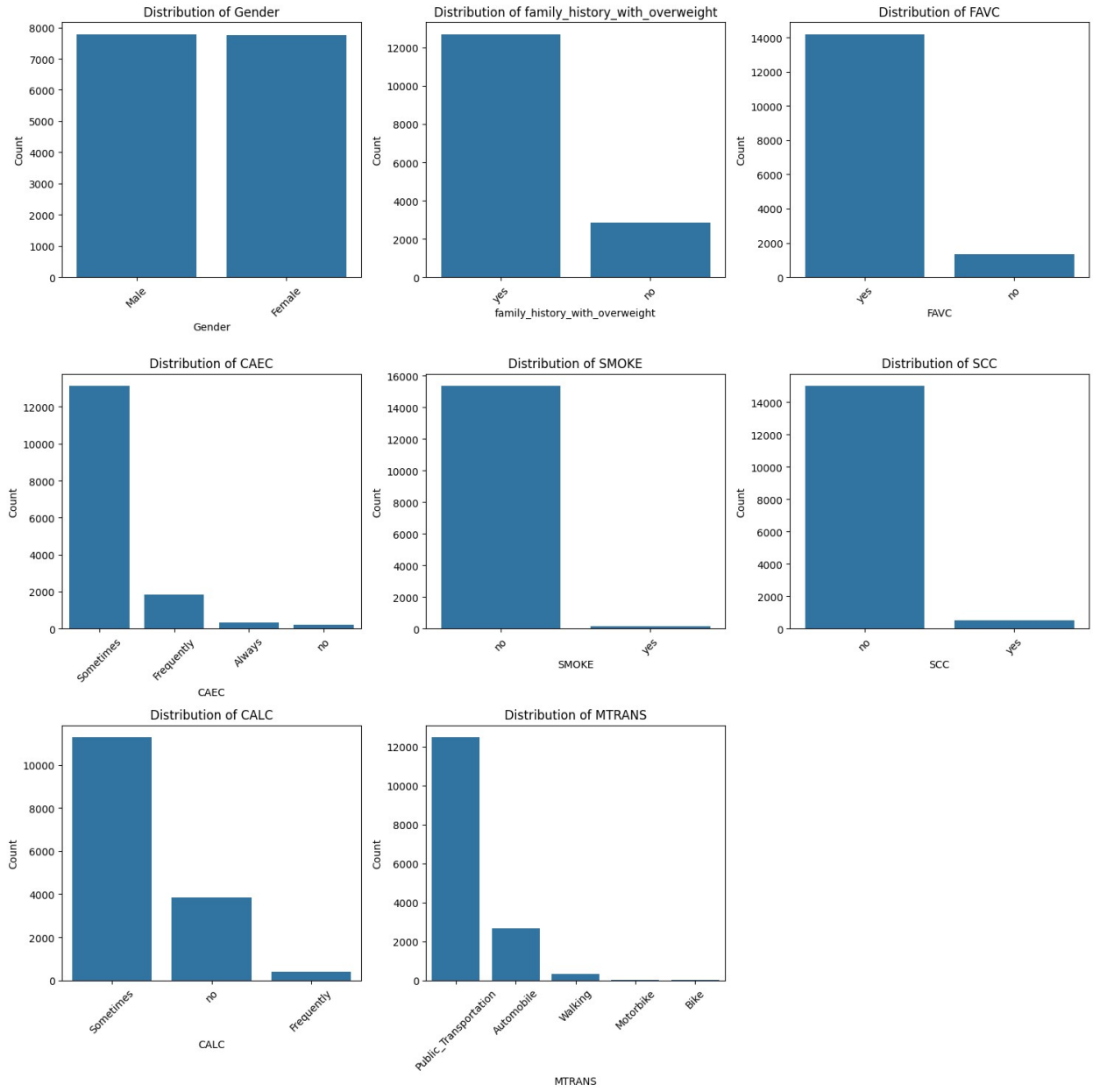
```
n    },\n    {\n      \"column\": \"FCVC\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 0.5308946652933533,\n
\"min\": 1.0,\n        \"max\": 3.0,\n        \"num_unique_values\":
872,\n        \"samples\": [\n          2.038774,\n          1.973499\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"NCP\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 0.7064628491410607,\n        \"min\": 1.0,\n        \"max\":
4.0,\n        \"num_unique_values\": 645,\n        \"samples\": [\n
1.146794,\n          1.044628\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"CAEC\",\n      \"properties\": {\n        \"dtype\":
\"category\",\n        \"num_unique_values\": 4,\n        \"samples\":
[\n          \"Frequently\",\n          \"Always\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"SMOKE\",\n      \"properties\": {\
n        \"dtype\": \"category\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          \"yes\",\n          \"no\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"CH2O\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 0.6077326835588633,\n
\"min\": 1.0,\n        \"max\": 3.0,\n        \"num_unique_values\":
1408,\n        \"samples\": [\n          2.435489,\n
2.299156\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"SCC\",\n      \"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n          \"yes\",\
n          \"no\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"FAF\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 0.8368409659505609,\n        \"min\": 0.0,\n        \"max\":
3.0,\n        \"num_unique_values\": 1274,\n        \"samples\": [\n
0.06275,\n          0.294763\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"TUE\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 0.6022227179241944,\n        \"min\":
0.0,\n        \"max\": 2.0,\n        \"num_unique_values\": 1207,\n
\"samples\": [\n          0.58284,\n          0.808599\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"CALC\",\n      \"properties\": {\n
\"dtype\": \"category\",\n        \"num_unique_values\": 3,\n
\"samples\": [\n          \"Sometimes\",\n          \"no\"\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"MTRANS\",\n      \"properties\": {\n        \"dtype\":
\"category\",\n        \"num_unique_values\": 5,\n        \"samples\":
[\n          \"Automobile\",\n          \"Bike\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"WeightCategory\",\n
```

```
\"properties\": {\n        \"dtype\": \"category\",\n
\"num_unique_values\": 7,\n        \"samples\": [\n
\"Overweight_Level_II\",\n          \"Normal_Weight\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe","variable_name":"train_df"}


from scipy.stats import zscore
from sklearn.preprocessing import MinMaxScaler

# Check for null values
print("Missing values in train_df:")
print(train_df.isnull().sum())

print("\nMissing values in test_df:")
print(test_df.isnull().sum())

print("\nPrimitive processing steps checked.")
```

```
Missing values in train_df:
Gender                            0
Age                               0
Height                            0
Weight                            0
family_history_with_overweight    0
FAVC                              0
FCVC                              0
NCP                               0
CAEC                              0
SMOKE                             0
CH2O                              0
SCC                               0
FAF                               0
TUE                               0
CALC                              0
MTRANS                            0
WeightCategory                    0
dtype: int64

Missing values in test_df:
Gender                            0
Age                               0
Height                            0
Weight                            0
family_history_with_overweight    0
FAVC                              0
FCVC                              0
NCP                               0
CAEC                              0
SMOKE                             0
```

```
CH2O                                0
SCC                                 0
FAF                                 0
TUE                                 0
CALC                                0
MTRANS                              0
dtype: int64

Primitive processing steps checked.
```

```
# Examine categorical feature distributions

categorical_features = ['Gender', 'family_history_with_overweight',
'FAVC', 'CAEC', 'SMOKE', 'SCC', 'CALC', 'MTRANS']
numerical_features = ['Age',
'Height','Weight','FCVC','NCP','CH2O','FAF','TUE']


n_features = len(categorical_features)
n_cols = 3  # You can adjust the number of columns as needed
n_rows = (n_features + n_cols - 1) // n_cols

fig, axes = plt.subplots(n_rows, n_cols, figsize=(15, n_rows * 5))
axes = axes.flatten() # Flatten the 2D array of axes for easy
iteration

for i, feature in enumerate(categorical_features):
    sns.countplot(data=train_df, x=feature,
order=train_df[feature].value_counts().index, ax=axes[i])
    axes[i].set_title(f'Distribution of {feature}')
    axes[i].set_xlabel(feature)
    axes[i].set_ylabel('Count')
    axes[i].tick_params(axis='x', rotation=45)

# just hiding
for j in range(i + 1, len(axes)):
    fig.delaxes(axes[j])

plt.tight_layout()
plt.show()
```

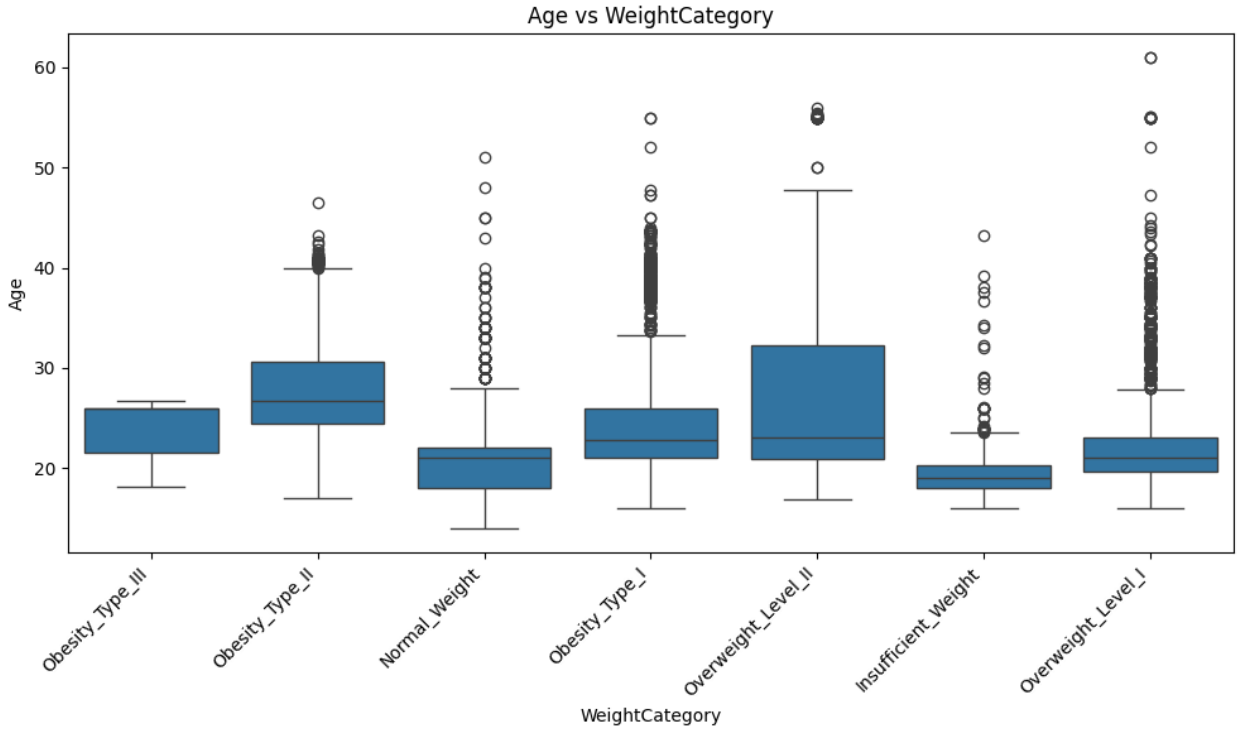


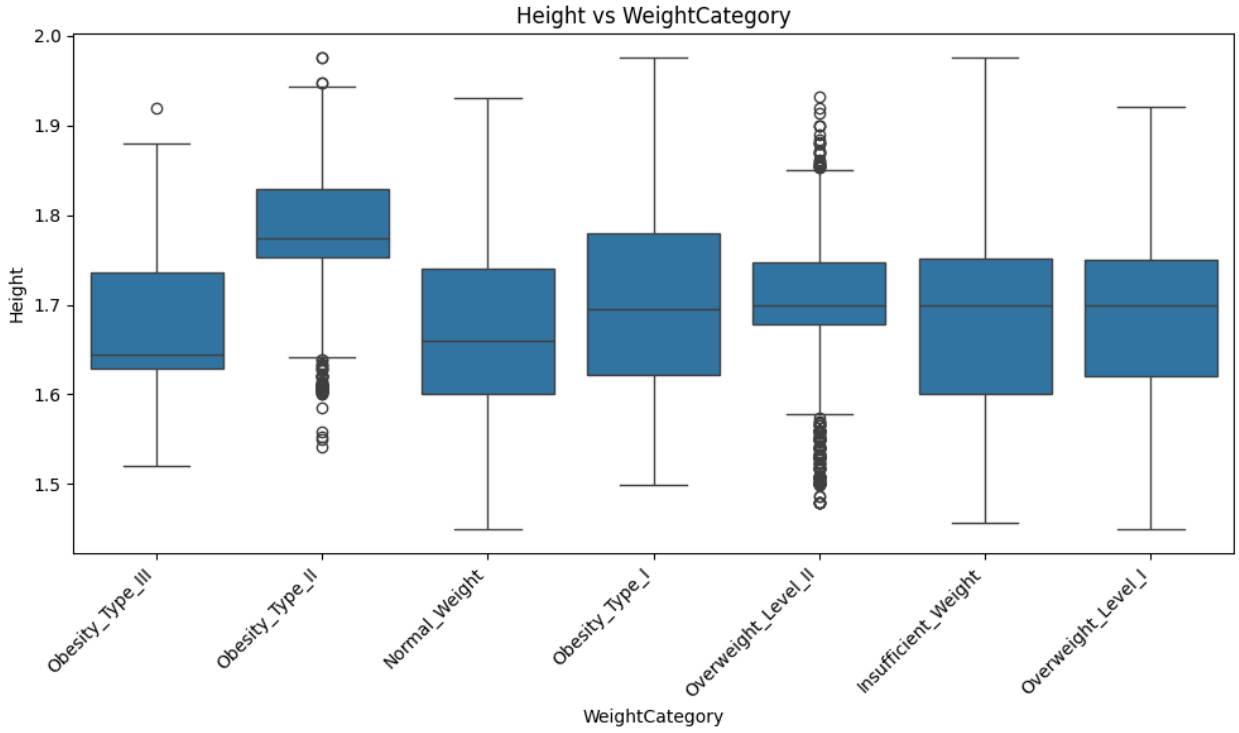

```
x=train_df.drop("WeightCategory",axis='columns')
y=train_df['WeightCategory']
y = np.array(y)
y.shape
```

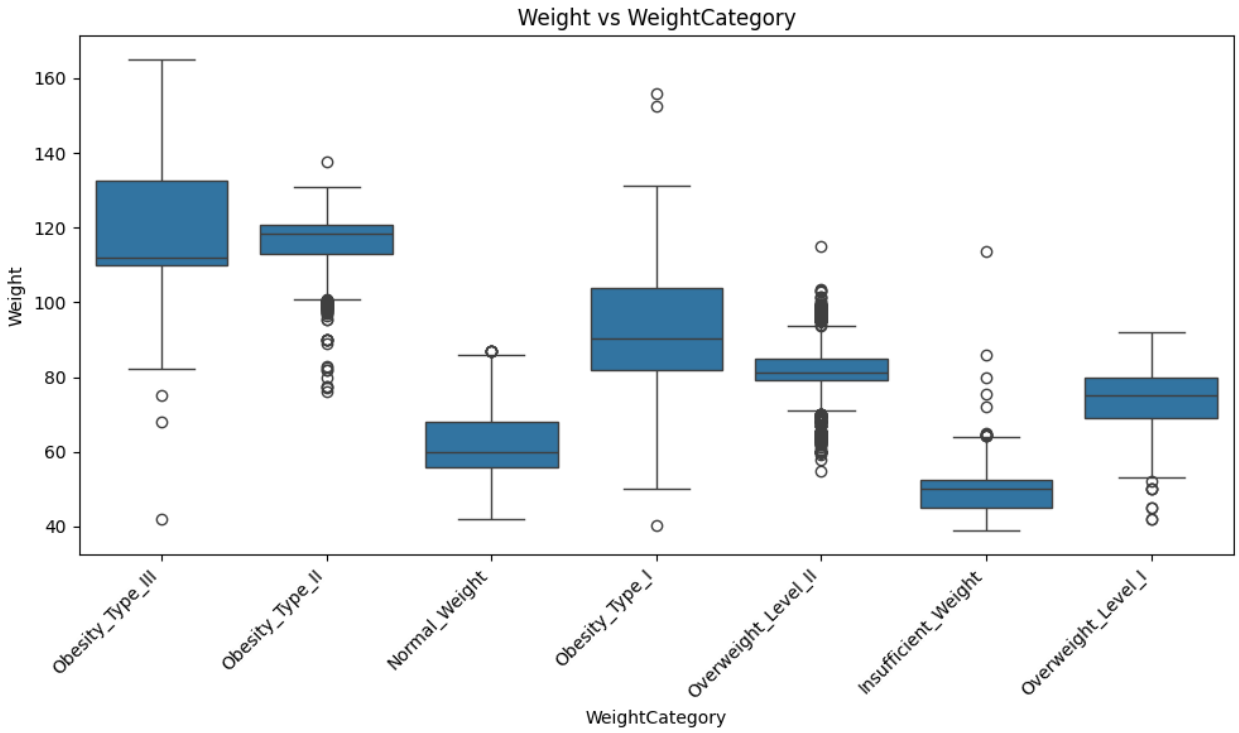
```
(15533,)
```

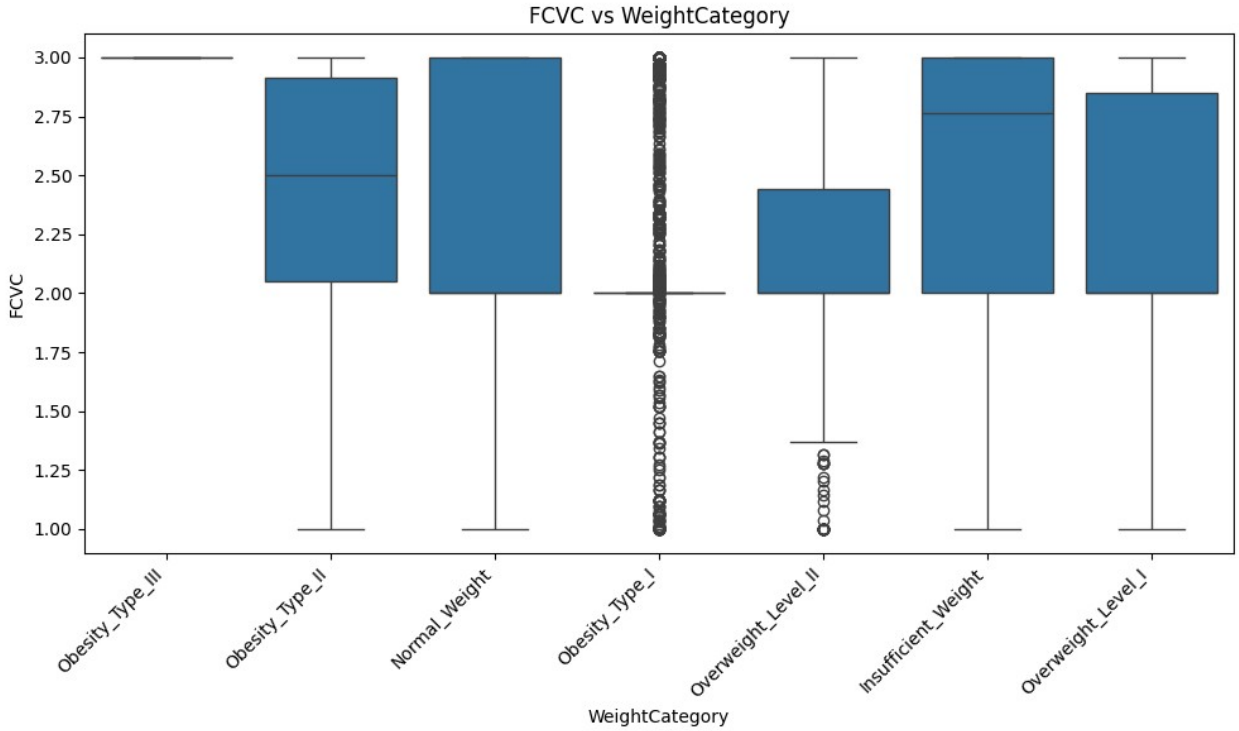
```
target = 'WeightCategory'

for feature in numerical_features:
    plt.figure(figsize=(10, 6))
    sns.boxplot(data=train_df, x=target, y=feature,
order=train_df[target].value_counts().index)
```

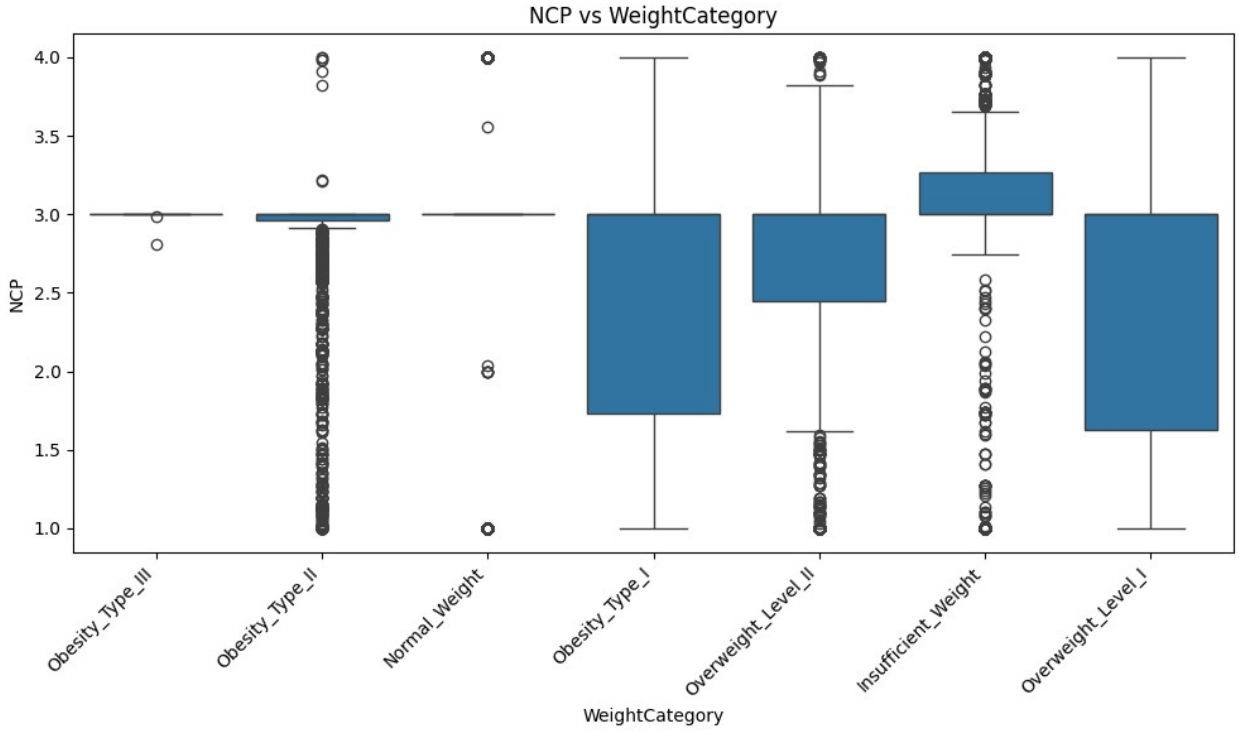
```
    plt.title(f'{feature} vs {target}')
    plt.xlabel(target)
    plt.ylabel(feature)
    plt.xticks(rotation=45, ha='right')
    plt.tight_layout()
    plt.show()
```

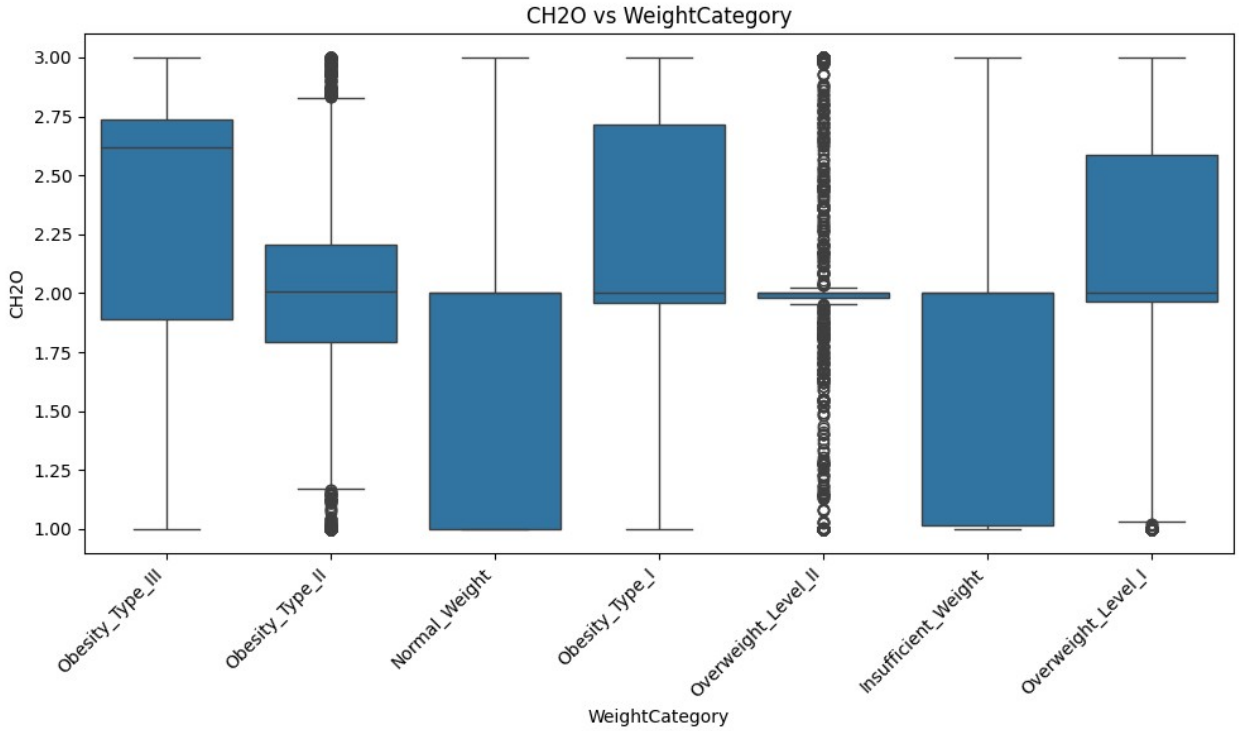


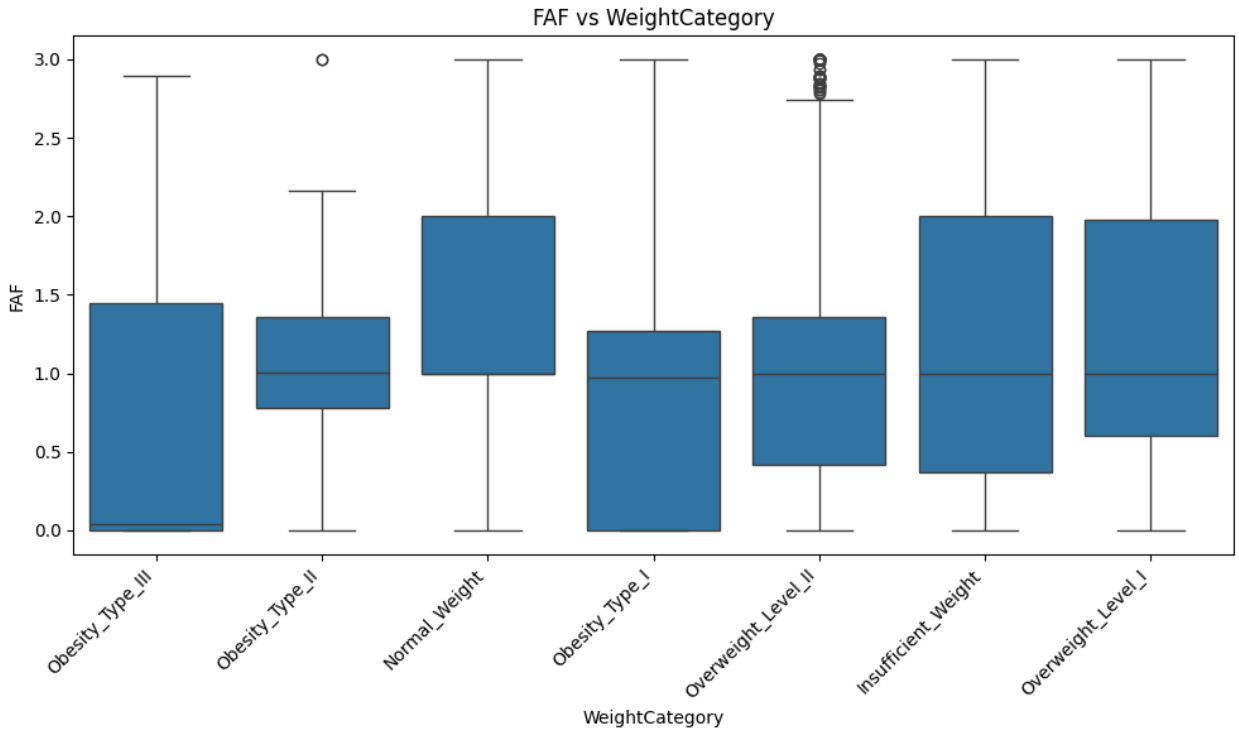

Height vs WeightCategory
Height
2.0
1.9
1.8
1.7
1.6
1.5
Obesity_Type_III
Obesity_Type_II
Normal_Weight
Obesity_Type_I
Overweight_Level_II
Insufficient_Weight
Overweight_Level_I
WeightCategory

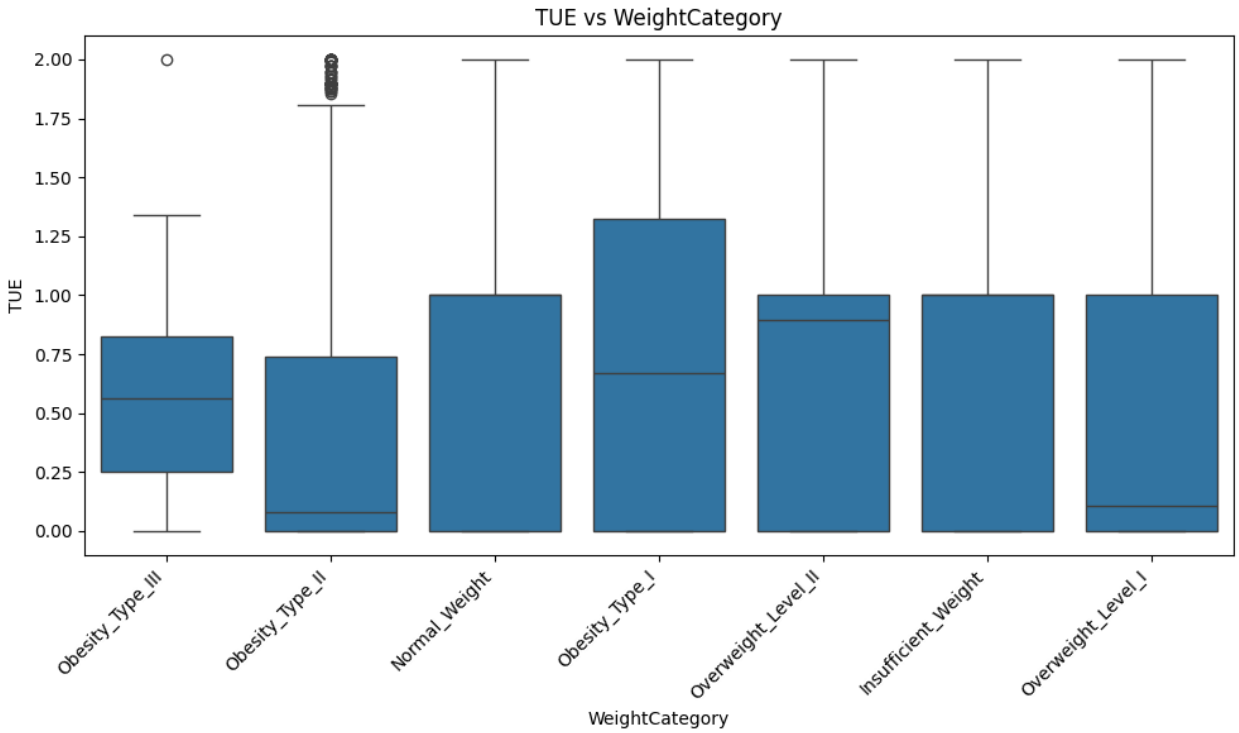

Weight vs WeightCategory
Weight
160
140
120
100
80
60
40
Obesity_Type_III
Obesity_Type_II
Normal_Weight
Obesity_Type_I
Overweight_Level_II
Insufficient_Weight
Overweight_Level_I
WeightCategory


FCVC vs WeightCategory
FCVC
3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
Obesity_Type_III
Obesity_Type_II
Normal_Weight
Obesity_Type_I
Overweight_Level_II
Insufficient_Weight
Overweight_Level_I
WeightCategory

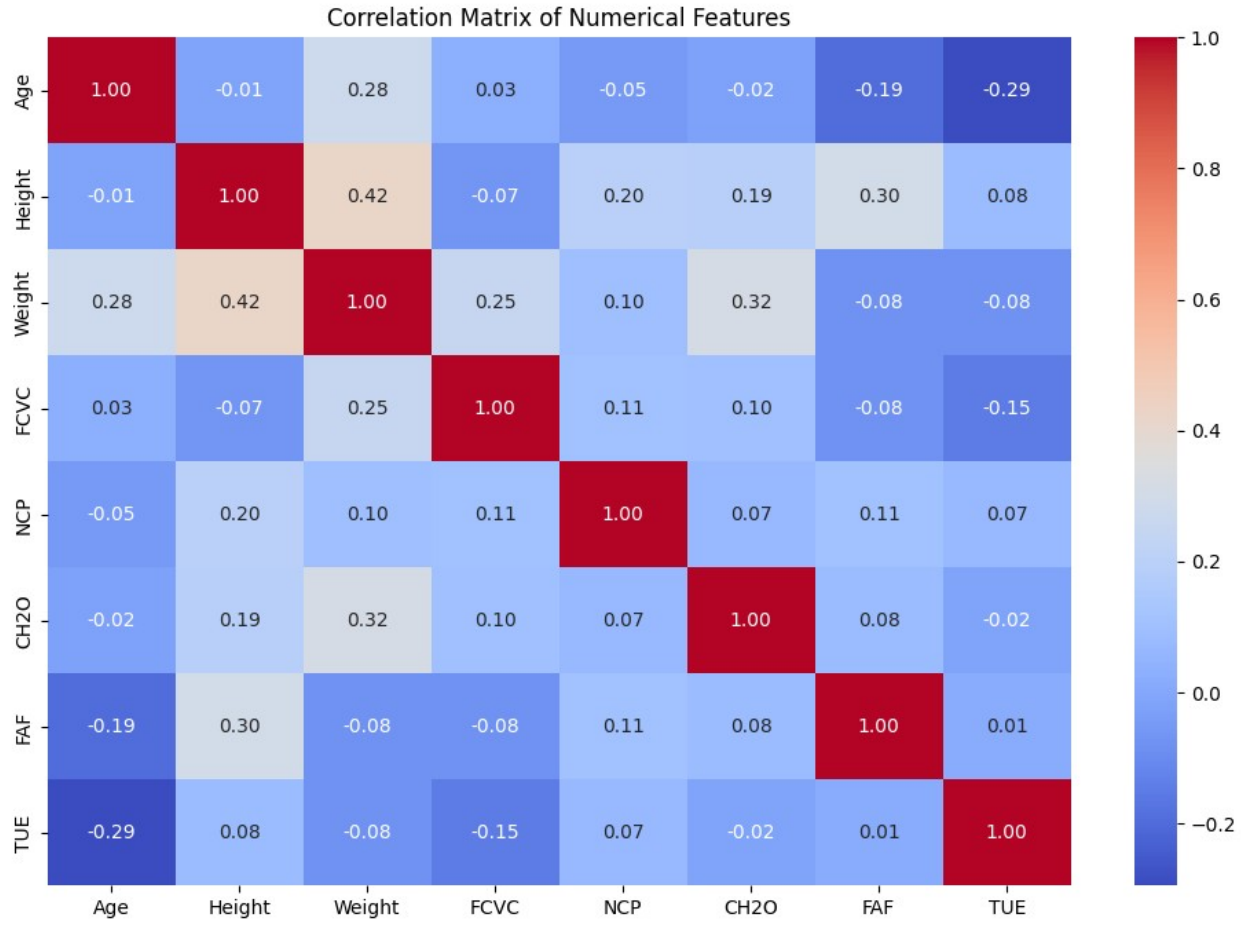

NCP vs WeightCategory
NCP
4.0
3.5
3.0
2.5
2.0
1.5
1.0
Obesity_Type_III
Obesity_Type_II
Normal_Weight
Obesity_Type_I
Overweight_Level_II
Insufficient_Weight
Overweight_Level_I
WeightCategory


CH2O vs WeightCategory
CH2O
3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
Obesity_Type_III
Obesity_Type_II
Normal_Weight
Obesity_Type_I
Overweight_Level_II
Insufficient_Weight
Overweight_Level_I
WeightCategory


FAF vs WeightCategory
FAF
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Obesity_Type_III
Obesity_Type_II
Normal_Weight
Obesity_Type_I
Overweight_Level_II
Insufficient_Weight
Overweight_Level_I
WeightCategory

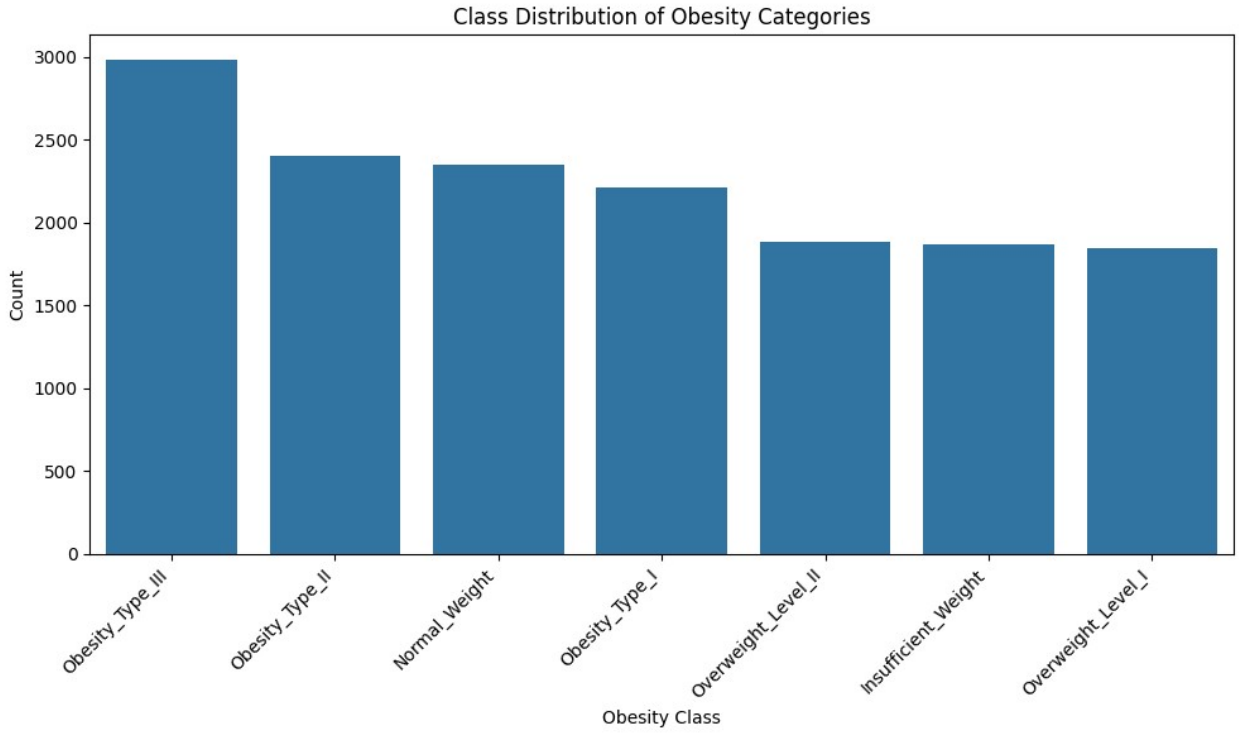



```
plt.figure(figsize=(12, 8))
sns.heatmap(train_df[numerical_features].corr(), annot=True,
cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix of Numerical Features')
plt.show()
```

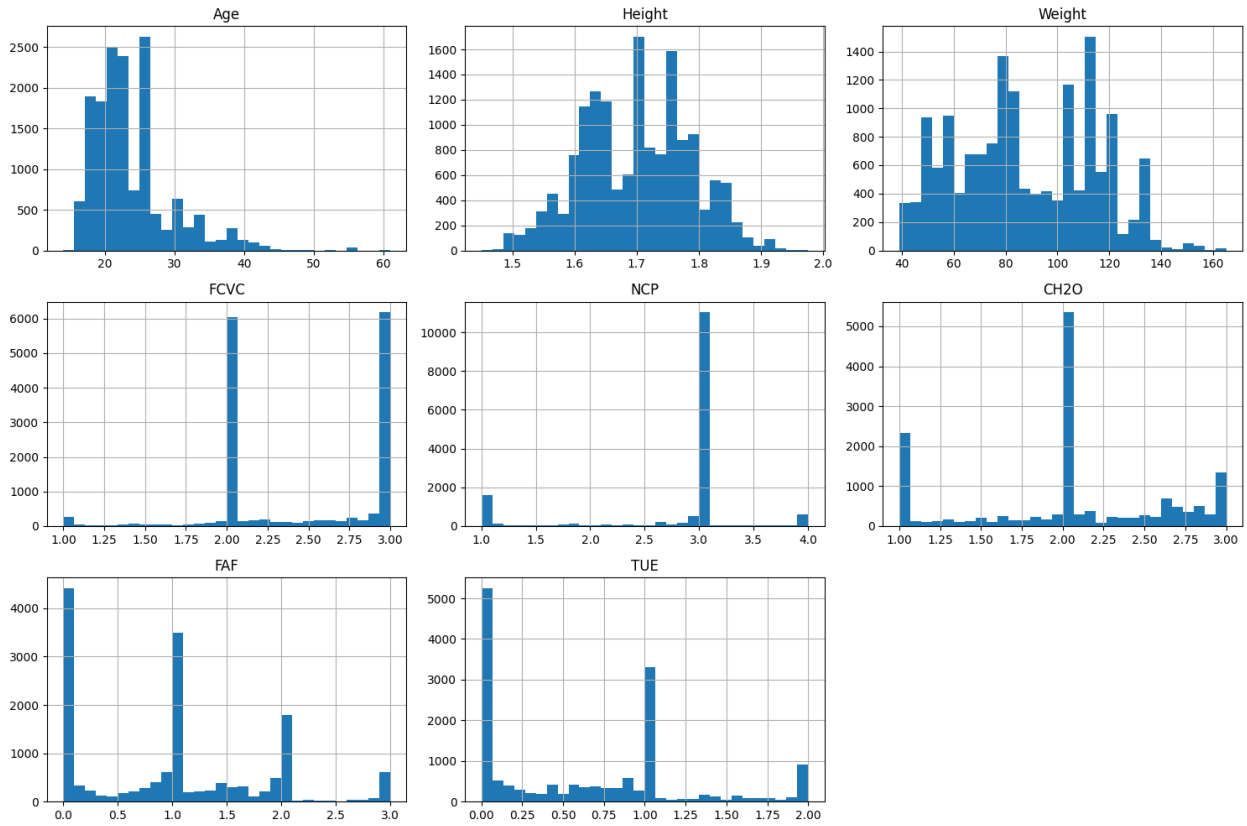



```
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd # Import pandas to ensure train_df is accessible

# Assuming 'y' is the target variable (WeightCategory)
# If y is a pandas Series, we can directly use it, otherwise convert
it
if not isinstance(y, pd.Series):
  y_series = pd.Series(y, name="WeightCategory")
else:
  y_series = y

plt.figure(figsize=(10, 6)) # Adjusted figure size for better
readability
sns.countplot(data=None, x=y_series,
order=y_series.value_counts().index) # Use data=None and pass the
series directly, order by count
plt.title("Class Distribution of Obesity Categories")
plt.xlabel("Obesity Class")
plt.ylabel("Count")
plt.xticks(rotation=45, ha='right') # Rotate labels for better
```

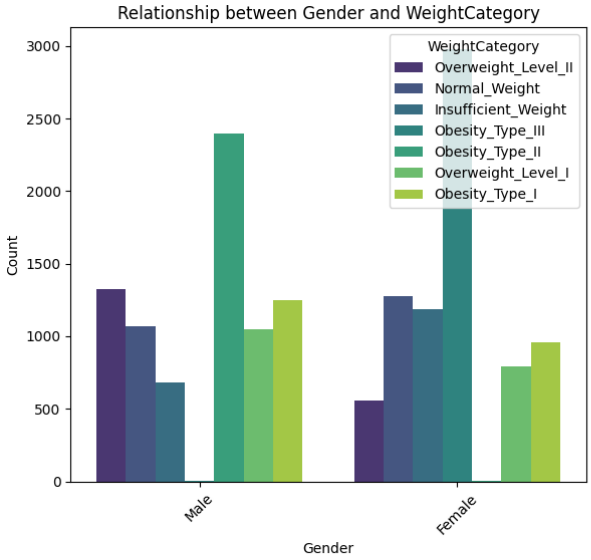
```
readability if many classes
plt.tight_layout()
plt.show()
```

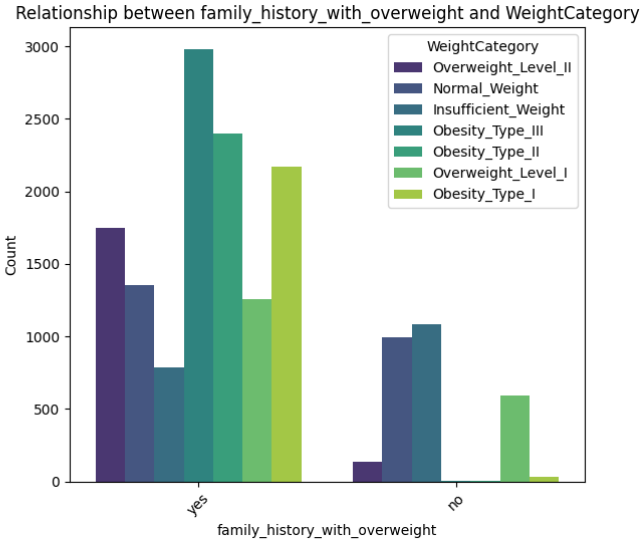


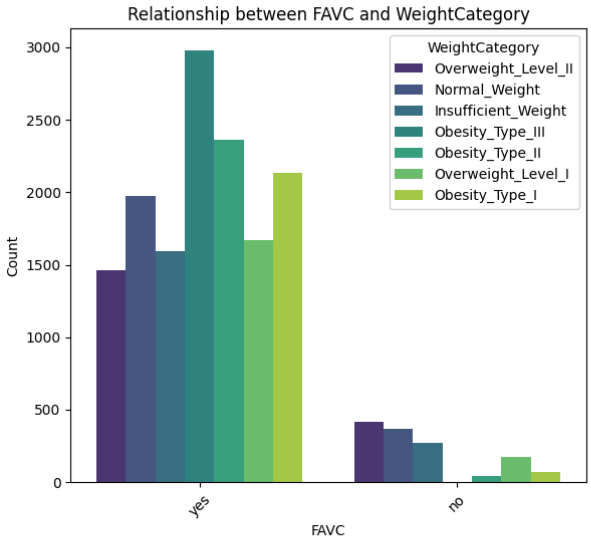

```
train_df[numerical_features].hist(bins=30, figsize=(15, 10))
plt.tight_layout()
plt.show()
```

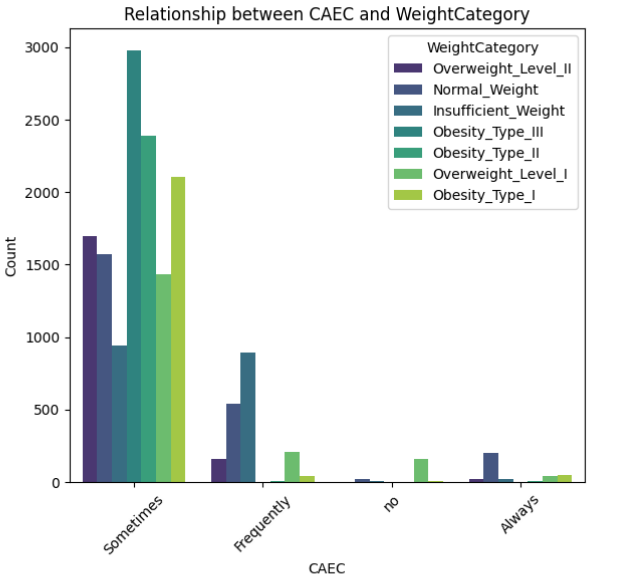


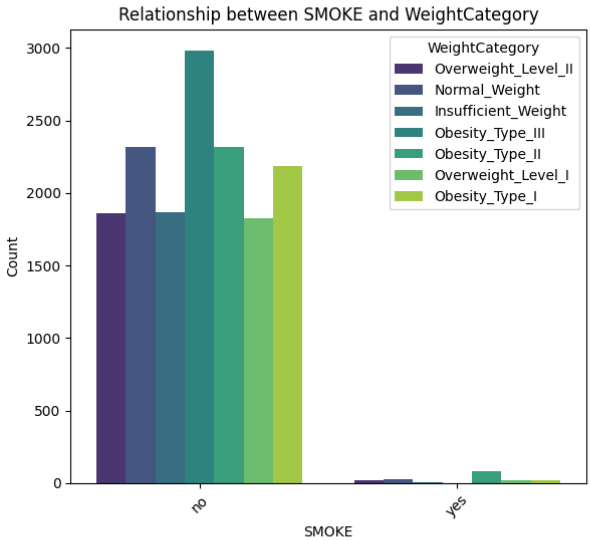

```
categorical_features = ['Gender', 'family_history_with_overweight',
'FAVC', 'CAEC', 'SMOKE', 'SCC', 'CALC', 'MTRANS']
target = 'WeightCategory'

n_features = len(categorical_features)
n_cols = 2
n_rows = (n_features + n_cols - 1) // n_cols

fig, axes = plt.subplots(n_rows, n_cols, figsize=(12, n_rows * 6)) #
Adjust figsize as needed
axes = axes.flatten()
for i, feature in enumerate(categorical_features):
    sns.countplot(data=train_df, x=feature, hue=target,
palette='viridis', ax=axes[i])
    axes[i].set_title(f'Relationship between {feature} and {target}')
    axes[i].set_xlabel(feature)
    axes[i].set_ylabel('Count')
    axes[i].tick_params(axis='x', rotation=45)
    axes[i].legend(title=target)

for j in range(i + 1, len(axes)):
    fig.delaxes(axes[j])

plt.tight_layout()
plt.show()
```

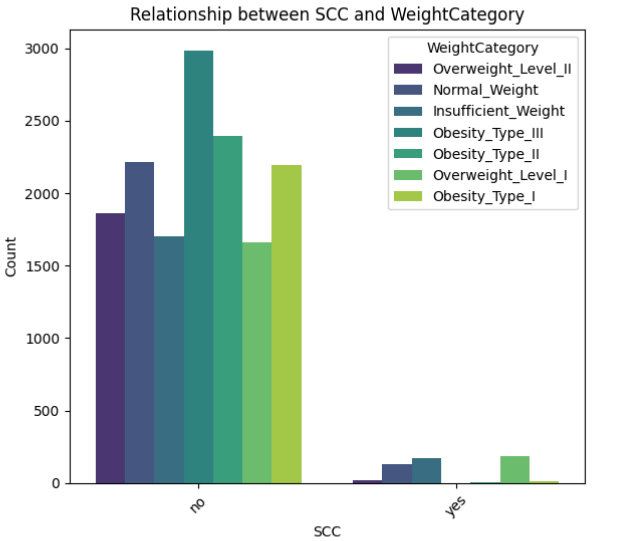

Relationship between Gender and WeightCategory
WeightCategory
Overweight_Level_II
Normal_Weight
Insufficient_Weight
Obesity_Type_III
Obesity_Type_II
Overweight_Level_I
Obesity_Type_I
Count
Male
Female
Gender


Relationship between family_history_with_overweight and WeightCategory
WeightCategory
Overweight_Level_II
Normal_Weight
Insufficient_Weight
Obesity_Type_III
Obesity_Type_II
Overweight_Level_I
Obesity_Type_I
Count
yes
no
family_history_with_overweight


Relationship between FAVC and WeightCategory
WeightCategory
Overweight_Level_II
Normal_Weight
Insufficient_Weight
Obesity_Type_III
Obesity_Type_II
Overweight_Level_I
Obesity_Type_I
Count
yes
no
FAVC


Relationship between CAEC and WeightCategory
WeightCategory
Overweight_Level_II
Normal_Weight
Insufficient_Weight
Obesity_Type_III
Obesity_Type_II
Overweight_Level_I
Obesity_Type_I
Count
Sometimes
Frequently
no
Always
CAEC


Relationship between SMOKE and WeightCategory
WeightCategory
Overweight_Level_II
Normal_Weight
Insufficient_Weight
Obesity_Type_III
Obesity_Type_II
Overweight_Level_I
Obesity_Type_I
Count
no
yes
SMOKE


Relationship between SCC and WeightCategory
WeightCategory
Overweight_Level_II
Normal_Weight
Insufficient_Weight
Obesity_Type_III
Obesity_Type_II
Overweight_Level_I
Obesity_Type_I
Count
no
yes
SCC

```
transformer = ColumnTransformer(transformers=[
    ('t1',OneHotEncoder(drop='first'),['CAEC']),
    ('t2',OneHotEncoder(drop='first'),['CALC']),
    ('t3', StandardScaler(), numerical_features),

    ('t4',OneHotEncoder(drop='first'),
['Gender','family_history_with_overweight','FAVC','SMOKE','SCC','MTRAN
S'])
],remainder='passthrough')

from sklearn.preprocessing import OrdinalEncoder

encoder = OrdinalEncoder()
y_encoded = encoder.fit_transform(np.array(y).reshape(-1, 1))
y_encoded.shape
```

```
(15533, 1)
```

```
from xgboost import XGBClassifier

def objective(trial):
    params = {
        'n_estimators': trial.suggest_int('n_estimators', 460, 480),
        'learning_rate': trial.suggest_float('learning_rate', 0.05,
0.065),
        'max_depth': trial.suggest_int('max_depth', 3, 6),
        'min_child_weight': trial.suggest_int('min_child_weight', 2,
5),
        'gamma': trial.suggest_float('gamma', 0.2, 0.4),
        'subsample': trial.suggest_float('subsample', 0.6, 1.0),
        'colsample_bytree': trial.suggest_float('colsample_bytree',
0.5, 0.8),
        'reg_alpha': trial.suggest_float('reg_alpha', 0.4, 0.6),
        'reg_lambda': trial.suggest_float('reg_lambda', 0, 0.5),
    }

    model = XGBClassifier(**params, objective='multi:softmax')

    score = cross_val_score(model, x_transformed, y_encoded, cv=3,
scoring='accuracy').mean()

    return score

x_transformed = transformer.fit_transform(x)

import optuna
from sklearn.model_selection import cross_val_score
from xgboost import XGBClassifier
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import accuracy_score
```

```
study = optuna.create_study(direction='maximize')

# Encode y to int
label_encoder = LabelEncoder()
y_encoded_labels = label_encoder.fit_transform(y)

# 100 trials optimization - 50 and 25 didnt yeild the same performance
study.optimize(lambda trial: objective(trial), n_trials=100,
show_progress_bar=True)


print("Best hyperparameters:", study.best_params)
print("Best accuracy:", study.best_value)

final_model = XGBClassifier(**study.best_params, random_state=42,
objective='multi:softmax', num_class=len(label_encoder.classes_))
final_model.fit(x_transformed, y_encoded_labels)

# pred
y_pred = final_model.predict(x_transformed)

# accuracy
accuracy = accuracy_score(y_encoded_labels, y_pred)
print(f"Accuracy on the training data: {accuracy}")


[I  2025-10-28 16:24:49,938] A new study created in memory with name:
no-name-c9fed308-6f2c-4c46-891a-660a680d5c08

{"model_id":"1b66c1224acd44d496b57e15440f319d","version_major":2,"vers
ion_minor":0}

[I  2025-10-26 14:47:05,642] A new study created in memory with name:
no-name-ccac80b1-f4cf-4a11-801c-63c9c6b4b46e
[I  2025-10-26 14:47:19,185] Trial 0 finished with value:
0.9080024951156758 and parameters: {'n_estimators': 473,
'learning_rate': 0.05381262950414241, 'max_depth': 5,
'min_child_weight': 5, 'gamma': 0.2782339027287223, 'subsample':
0.71774986558411 59, 'colsample_bytree': 0.5870083902786944,
'reg_alpha': 0.4604845337127979, 'reg_lambda': 0.32735488630179993}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:47:34,513] Trial 1 finished with value:
0.9036889278545539 and parameters: {'n_estimators': 464,
'learning_rate': 0.05906161873739976, 'max_depth': 6,
'min_child_weight': 4, 'gamma': 0.3430873781146919, 'subsample':
0.6152497315216234, 'colsample_bytree': 0.7963797002576285,
'reg_alpha': 0.401346914594285, 'reg_lambda': 0.05932862122029109}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:47:47,533] Trial 2 finished with value:
0.9071011965030679 and parameters: {'n_estimators': 462,
```

```
'learning_rate': 0.061256753641571834, 'max_depth': 6,
'min_child_weight': 2, 'gamma': 0.30429871311536094, 'subsample':
0.8066210662784381, 'colsample_bytree': 0.5374706452876434,
'reg_alpha': 0.5750999148389956, 'reg_lambda': 0.36544222748525607}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:47:58,388] Trial 3 finished with value:
0.9056206230930157 and parameters: {'n_estimators': 476,
'learning_rate': 0.05677290581288154, 'max_depth': 3,
'min_child_weight': 4, 'gamma': 0.20386263151315134, 'subsample':
0.9621717248868199, 'colsample_bytree': 0.6027354640522151,
'reg_alpha': 0.5312343937697294, 'reg_lambda': 0.02942472619553138}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:48:12,380] Trial 4 finished with value:
0.9043972633608672 and parameters: {'n_estimators': 467,
'learning_rate': 0.05761756968450838, 'max_depth': 6,
'min_child_weight': 3, 'gamma': 0.3622306289798835, 'subsample':
0.6990378384251419, 'colsample_bytree': 0.7605927562975652,
'reg_alpha': 0.5349194264200875, 'reg_lambda': 0.015222789102392997}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:48:26,391] Trial 5 finished with value:
0.9048478629280021 and parameters: {'n_estimators': 478,
'learning_rate': 0.06453052401068039, 'max_depth': 5,
'min_child_weight': 2, 'gamma': 0.24744842137745487, 'subsample':
0.7505215895317898, 'colsample_bytree': 0.7233830038005558,
'reg_alpha': 0.4790605075460840 4, 'reg_lambda': 0.008756371539279717}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:48:36,775] Trial 6 finished with value:
0.9056205236146777 and parameters: {'n_estimators': 475,
'learning_rate': 0.06054522251337999, 'max_depth': 3,
'min_child_weight': 5, 'gamma': 0.22473513199873946, 'subsample':
0.9794337450907834, 'colsample_bytree': 0.6948609046682118,
'reg_alpha': 0.4425084873908079, 'reg_lambda': 0.1116515138315573}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:48:47,774] Trial 7 finished with value:
0.9064573478294163 and parameters: {'n_estimators': 472,
'learning_rate': 0.06152467701013528, 'max_depth': 5,
'min_child_weight': 3, 'gamma': 0.27215592661565724, 'subsample':
0.8756561386648939, 'colsample_bytree': 0.6465465597132676,
'reg_alpha': 0.5198829220317744, 'reg_lambda': 0.25618602935570084}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:49:00,406] Trial 8 finished with value:
0.9069080593096815 and parameters: {'n_estimators': 474,
'learning_rate': 0.05759806123727607, 'max_depth': 5,
'min_child_weight': 5, 'gamma': 0.28756136594305726, 'subsample':
0.8523209436061027, 'colsample_bytree': 0.6355087530793623,
'reg_alpha': 0.4562875017408737, 'reg_lambda': 0.13235708218222758}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:49:14,489] Trial 9 finished with value:
0.907101159198691 and parameters: {'n_estimators': 470,
```

```
'learning_rate': 0.05548119112252349, 'max_depth': 6,
'min_child_weight': 3, 'gamma': 0.31920479517598416, 'subsample':
0.7477643464201206, 'colsample_bytree': 0.5482191847937812,
'reg_alpha': 0.5680202992752462, 'reg_lambda': 0.12104695683440686}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:49:26,812] Trial 10 finished with value:
0.9072299090377537 and parameters: {'n_estimators': 480,
'learning_rate': 0.051286813675359554, 'max_depth': 4,
'min_child_weight': 5, 'gamma': 0.3929431362069141, 'subsample':
0.6014904194203177, 'colsample_bytree': 0.5089536217330015,
'reg_alpha': 0.40188685223141146, 'reg_lambda': 0.498368351743533}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:49:39,103] Trial 11 finished with value:
0.9076161585549419 and parameters: {'n_estimators': 480,
'learning_rate': 0.05090894803808087, 'max_depth': 4,
'min_child_weight': 5, 'gamma': 0.3966728654659811, 'subsample':
0.6389397782861664, 'colsample_bytree': 0.5019171595983015,
'reg_alpha': 0.4006433577883363, 'reg_lambda': 0.4927127289959342}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:49:51,814] Trial 12 finished with value:
0.9072943461312463 and parameters: {'n_estimators': 480,
'learning_rate': 0.05104439802261947, 'max_depth': 4,
'min_child_weight': 4, 'gamma': 0.2661078201593408, 'subsample':
0.6713240148120199, 'colsample_bytree': 0.5808226378363642,
'reg_alpha': 0.43164768709043194, 'reg_lambda': 0.4608350645108664}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:50:14,157] Trial 13 finished with value:
0.906199810846914 and parameters: {'n_estimators': 469,
'learning_rate': 0.053572037758079935, 'max_depth': 4,
'min_child_weight': 5, 'gamma': 0.32867074082033326, 'subsample':
0.6738167474124458, 'colsample_bytree': 0.5035993588233398,
'reg_alpha': 0.48377709083794435, 'reg_lambda': 0.3665410788922982}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:50:26,502] Trial 14 finished with value:
0.9062642355056143 and parameters: {'n_estimators': 473,
'learning_rate': 0.0534979017329977, 'max_depth': 4,
'min_child_weight': 5, 'gamma': 0.3928760431915049, 'subsample':
0.7303853156215465, 'colsample_bytree': 0.5810942414091742,
'reg_alpha': 0.42718146214783986, 'reg_lambda': 0.3645091783155846}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:50:39,819] Trial 15 finished with value:
0.9069079847009279 and parameters: {'n_estimators': 477,
'learning_rate': 0.05343969320451253, 'max_depth': 5,
'min_child_weight': 4, 'gamma': 0.36008800332558943, 'subsample':
0.6481999590168339, 'colsample_bytree': 0.5576159748234909,
'reg_alpha': 0.46426838113002716, 'reg_lambda': 0.2507598220782447}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:50:52,140] Trial 16 finished with value:
0.907616170989734 and parameters: {'n_estimators': 466,
```

```
'learning_rate': 0.050160157068483265, 'max_depth': 4,
'min_child_weight': 5, 'gamma': 0.25430174846305464, 'subsample':
0.7932492143621047, 'colsample_bytree': 0.6147069813732862,
'reg_alpha': 0.5065501630793224, 'reg_lambda': 0.4254211192236438}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:51:00,907] Trial 17 finished with value:
0.905556173564731 and parameters: {'n_estimators': 466,
'learning_rate': 0.050144698460788145, 'max_depth': 3,
'min_child_weight': 4, 'gamma': 0.24306316005724452, 'subsample':
0.8185133611942305, 'colsample_bytree': 0.6772974195475189,
'reg_alpha': 0.5010139443591028, 'reg_lambda': 0.314627737949061}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:51:14,004] Trial 18 finished with value:
0.9080024826808835 and parameters: {'n_estimators': 460,
'learning_rate': 0.05277535733077376, 'max_depth': 5,
'min_child_weight': 5, 'gamma': 0.28106086510474654, 'subsample':
0.7758481738581695, 'colsample_bytree': 0.6122939687787609,
'reg_alpha': 0.5534226941452232, 'reg_lambda': 0.43375752534152434}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:51:26,403] Trial 19 finished with value:
0.9074230586659323 and parameters: {'n_estimators': 460,
'learning_rate': 0.05546817269717135, 'max_depth': 5,
'min_child_weight': 4, 'gamma': 0.2855984980444312, 'subsample':
0.9194207320871004, 'colsample_bytree': 0.6079526820474466,
'reg_alpha': 0.5515696404529827, 'reg_lambda': 0.30625358978476475}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:51:39,869] Trial 20 finished with value:
0.9065217600533244 and parameters: {'n_estimators': 471,
'learning_rate': 0.05261884946058405, 'max_depth': 5,
'min_child_weight': 5, 'gamma': 0.3043519546790307, 'subsample':
0.7745001031063636, 'colsample_bytree': 0.6794334585946151,
'reg_alpha': 0.5455558864550013, 'reg_lambda': 0.19425740996351826}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:51:52,846] Trial 21 finished with value:
0.9069080095705124 and parameters: {'n_estimators': 460,
'learning_rate': 0.055071466648562106, 'max_depth': 5,
'min_child_weight': 5, 'gamma': 0.2526738524107699, 'subsample':
0.8419429766352303, 'colsample_bytree': 0.6206716875765201,
'reg_alpha': 0.5970637850373184, 'reg_lambda': 0.4317361236533005}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:52:05,556] Trial 22 finished with value:
0.9067149221162953 and parameters: {'n_estimators': 464,
'learning_rate': 0.05205773682195099, 'max_depth': 4,
'min_child_weight': 5, 'gamma': 0.22642345891123417, 'subsample':
0.7078958031320008, 'colsample_bytree': 0.581135107470662,
'reg_alpha': 0.507537414620069, 'reg_lambda': 0.4026733448028652}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:52:18,894] Trial 23 finished with value:
0.9065218097924933 and parameters: {'n_estimators': 466,
```

```
'learning_rate': 0.054340377421608486, 'max_depth': 5,
'min_child_weight': 5, 'gamma': 0.2734738983808681, 'subsample':
0.7859200162676956, 'colsample_bytree': 0.6638959734390055,
'reg_alpha': 0.4831704371093454, 'reg_lambda': 0.4257445095949918}.
Best is trial 0 with value: 0.9080024951156758.
[I  2025-10-26 14:52:31,060] Trial 24 finished with value:
0.9085175442110955 and parameters: {'n_estimators': 469,
'learning_rate': 0.05013541007985127, 'max_depth': 4,
'min_child_weight': 4, 'gamma': 0.28858182186421116, 'subsample':
0.8912893860750811, 'colsample_bytree': 0.629795674376255,
'reg_alpha': 0.5134153687270296, 'reg_lambda': 0.3925722471346489}.
Best is trial 24 with value: 0.9085175442110955.
[I  2025-10-26 14:52:43,768] Trial 25 finished with value:
0.9072298592985847 and parameters: {'n_estimators': 469,
'learning_rate': 0.05231534144129121, 'max_depth': 5,
'min_child_weight': 4, 'gamma': 0.2909865595540487, 'subsample':
0.9055102393162788, 'colsample_bytree': 0.7114199993389471,
'reg_alpha': 0.5628726247605382, 'reg_lambda': 0.3144492660005853}.
Best is trial 24 with value: 0.9085175442110955.
[I  2025-10-26 14:52:57,159] Trial 26 finished with value:
0.9065217351837399 and parameters: {'n_estimators': 463,
'learning_rate': 0.052048722892413746, 'max_depth': 6,
'min_child_weight': 3, 'gamma': 0.31842404820431125, 'subsample':
0.9237829149659362, 'colsample_bytree': 0.636210597914112,
'reg_alpha': 0.5868952559929538, 'reg_lambda': 0.3908373168798027}.
Best is trial 24 with value: 0.9085175442110955.
[I  2025-10-26 14:53:10,228] Trial 27 finished with value:
0.9089681562130227 and parameters: {'n_estimators': 473,
'learning_rate': 0.05654568747242938, 'max_depth': 4,
'min_child_weight': 4, 'gamma': 0.27855847794120936, 'subsample':
0.8811061714182321, 'colsample_bytree': 0.5936850968585458,
'reg_alpha': 0.47089089488880537, 'reg_lambda': 0.2945379203470715}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 14:53:21,088] Trial 28 finished with value:
0.9067149967250487 and parameters: {'n_estimators': 473,
'learning_rate': 0.05656502073794932, 'max_depth': 3,
'min_child_weight': 4, 'gamma': 0.30275645797123296, 'subsample':
0.8846898656439277, 'colsample_bytree': 0.5720598203472708,
'reg_alpha': 0.4596122760313717, 'reg_lambda': 0.20287587091486675}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 14:53:33,499] Trial 29 finished with value:
0.9057492734537403 and parameters: {'n_estimators': 468,
'learning_rate': 0.059238697224179716, 'max_depth': 4,
'min_child_weight': 4, 'gamma': 0.3510793475680346, 'subsample':
0.9429642036629325, 'colsample_bytree': 0.7850952573352685,
'reg_alpha': 0.4713603686910982, 'reg_lambda': 0.27758864231641556}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 14:53:45,594] Trial 30 finished with value:
0.9081312076503617 and parameters: {'n_estimators': 471,
```

'learning_rate': 0.05893871185286817, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.22994982702264954, 'subsample':
0.8385936771964192, 'colsample_bytree': 0.533075066643256,
'reg_alpha': 0.4921682245850608, 'reg_lambda': 0.3393416711252024}.
Best is trial 27 with value: 0.9089681562130227.
[I 2025-10-26 14:53:57,695] Trial 31 finished with value:
0.9080024826808835 and parameters: {'n_estimators': 471,
'learning_rate': 0.05909395342494775, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.23225365933243503, 'subsample':
0.8394137803409543, 'colsample_bytree': 0.530075870641647,
'reg_alpha': 0.49018652196761525, 'reg_lambda': 0.33953225771304746}.
Best is trial 27 with value: 0.9089681562130227.
[I 2025-10-26 14:54:10,951] Trial 32 finished with value:
0.9075518831137486 and parameters: {'n_estimators': 472,
'learning_rate': 0.05819012283174161, 'max_depth': 4,
'min_child_weight': 2, 'gamma': 0.2056660302244556, 'subsample':
0.8780496218899865, 'colsample_bytree': 0.5349310225122411,
'reg_alpha': 0.5181074724176962, 'reg_lambda': 0.3374280666729625}.
Best is trial 27 with value: 0.9089681562130227.
[I 2025-10-26 14:54:19,826] Trial 33 finished with value:
0.9069081587880197 and parameters: {'n_estimators': 475,
'learning_rate': 0.05616907531148024, 'max_depth': 3,
'min_child_weight': 3, 'gamma': 0.2591398744948589, 'subsample':
0.819228750331526, 'colsample_bytree': 0.5909341104538051,
'reg_alpha': 0.49510948090267964, 'reg_lambda': 0.28544715061461096}.
Best is trial 27 with value: 0.9089681562130227.
[I 2025-10-26 14:54:31,638] Trial 34 finished with value:
0.9073586713116087 and parameters: {'n_estimators': 470,
'learning_rate': 0.05972969328451243, 'max_depth': 4,
'min_child_weight': 4, 'gamma': 0.31447774115233035, 'subsample':
0.8600772464276778, 'colsample_bytree': 0.5602770814994273,
'reg_alpha': 0.44477086721730213, 'reg_lambda': 0.19523859298759016}.
Best is trial 27 with value: 0.9089681562130227.
[I 2025-10-26 14:54:42,343] Trial 35 finished with value:
0.9065863339295319 and parameters: {'n_estimators': 474,
'learning_rate': 0.061491243479173445, 'max_depth': 3,
'min_child_weight': 3, 'gamma': 0.33386198176730175, 'subsample':
0.8953374724791056, 'colsample_bytree': 0.5917147693240741,
'reg_alpha': 0.5201115033389065, 'reg_lambda': 0.22535133234585286}.
Best is trial 27 with value: 0.9089681562130227.
[I 2025-10-26 14:54:53,964] Trial 36 finished with value:
0.90703689619229 and parameters: {'n_estimators': 468,
'learning_rate': 0.06318855171707524, 'max_depth': 4,
'min_child_weight': 2, 'gamma': 0.21390131844457688, 'subsample':
0.9492367916832115, 'colsample_bytree': 0.5257756039591628,
'reg_alpha': 0.47533385321393584, 'reg_lambda': 0.3514153261287924}.
Best is trial 27 with value: 0.9089681562130227.
[I 2025-10-26 14:55:04,916] Trial 37 finished with value:
0.906071185355774 and parameters: {'n_estimators': 476,

'learning_rate': 0.058424484993585685, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.2397393687301232, 'subsample':
0.999913477318824, 'colsample_bytree': 0.6253153671897431,
'reg_alpha': 0.5358105496798875, 'reg_lambda': 0.38569979432839374}.
Best is trial 27 with value: 0.9089681562130227.
[I 2025-10-26 14:55:17,331] Trial 38 finished with value:
0.9068437838684883 and parameters: {'n_estimators': 472,
'learning_rate': 0.06032379264681409, 'max_depth': 6,
'min_child_weight': 4, 'gamma': 0.2955833903537824, 'subsample':
0.9172557951993019, 'colsample_bytree': 0.5613665362080215,
'reg_alpha': 0.44928310517165276, 'reg_lambda': 0.28769567261593754}.
Best is trial 27 with value: 0.9089681562130227.
[I 2025-10-26 14:55:26,150] Trial 39 finished with value:
0.9065218097924933 and parameters: {'n_estimators': 474,
'learning_rate': 0.05718257864340665, 'max_depth': 3,
'min_child_weight': 3, 'gamma': 0.26585304405942245, 'subsample':
0.8116645665700939, 'colsample_bytree': 0.6603522602485068,
'reg_alpha': 0.4305973327270227, 'reg_lambda': 0.32626828263932384}.
Best is trial 27 with value: 0.9089681562130227.
[I 2025-10-26 14:55:38,475] Trial 40 finished with value:
0.9078736831026517 and parameters: {'n_estimators': 478,
'learning_rate': 0.054461468743805475, 'max_depth': 4,
'min_child_weight': 2, 'gamma': 0.27567116784369644, 'subsample':
0.8687790995778162, 'colsample_bytree': 0.5963026151050698,
'reg_alpha': 0.4697720869841581, 'reg_lambda': 0.26616307906262343}.
Best is trial 27 with value: 0.9089681562130227.
[I 2025-10-26 14:55:51,730] Trial 41 finished with value:
0.9077449581331737 and parameters: {'n_estimators': 462,
'learning_rate': 0.05467658289606458 5, 'max_depth': 5,
'min_child_weight': 4, 'gamma': 0.2855351209707797, 'subsample':
0.7364804101399105, 'colsample_bytree': 0.6346595708846641,
'reg_alpha': 0.526428566930388, 'reg_lambda': 0.4602125582137038}.
Best is trial 27 with value: 0.9089681562130227.
[I 2025-10-26 14:56:05,034] Trial 42 finished with value:
0.9074230835355168 and parameters: {'n_estimators': 471,
'learning_rate': 0.05289256663336095, 'max_depth': 5,
'min_child_weight': 5, 'gamma': 0.28195990119553305, 'subsample':
0.7639386520378645, 'colsample_bytree': 0.6050903950389664,
'reg_alpha': 0.5470270118497647, 'reg_lambda': 0.4551712907591976}.
Best is trial 27 with value: 0.9089681562130227.
[I 2025-10-26 14:56:18,475] Trial 43 finished with value:
0.9065860852336867 and parameters: {'n_estimators': 473,
'learning_rate': 0.05609233666236062, 'max_depth': 5,
'min_child_weight': 4, 'gamma': 0.3090205524118714, 'subsample':
0.7132357650113083, 'colsample_bytree': 0.656694320384187,
'reg_alpha': 0.4919894784749264, 'reg_lambda': 0.3977580024377436}.
Best is trial 27 with value: 0.9089681562130227.
[I 2025-10-26 14:56:31,710] Trial 44 finished with value:
0.9074230835355168 and parameters: {'n_estimators': 475,

'learning_rate': 0.05142260467726477, 'max_depth': 5,
'min_child_weight': 5, 'gamma': 0.2641124806337731, 'subsample':
0.8322761065802726, 'colsample_bytree': 0.5483592950273882,
'reg_alpha': 0.4547805789309431, 'reg_lambda': 0.375687910377971}.
Best is trial 27 with value: 0.9089681562130227.
[I 2025-10-26 14:56:45,996] Trial 45 finished with value:
0.9058135240253491 and parameters: {'n_estimators': 470,
'learning_rate': 0.053740043557339544, 'max_depth': 6,
'min_child_weight': 4, 'gamma': 0.2971763953638843, 'subsample':
0.6884192765700052, 'colsample_bytree': 0.6442880868947506,
'reg_alpha': 0.5101435513719836, 'reg_lambda': 0.23422093331057706}.
Best is trial 27 with value: 0.9089681562130227.
[I 2025-10-26 14:56:58,200] Trial 46 finished with value:
0.9072299339073382 and parameters: {'n_estimators': 476,
'learning_rate': 0.05771525522784893, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.27622038145820754, 'subsample':
0.7972319113968565, 'colsample_bytree': 0.5710837489456442,
'reg_alpha': 0.4819385177997296, 'reg_lambda': 0.35496050075185104}.
Best is trial 27 with value: 0.9089681562130227.
[I 2025-10-26 14:57:10,616] Trial 47 finished with value:
0.9074231830138549 and parameters: {'n_estimators': 465,
'learning_rate': 0.06279515185100343, 'max_depth': 5,
'min_child_weight': 5, 'gamma': 0.32861745135883325, 'subsample':
0.7674977319257367, 'colsample_bytree': 0.5181567451366881,
'reg_alpha': 0.4201482315586666, 'reg_lambda': 0.05530867231959957}.
Best is trial 27 with value: 0.9089681562130227.
[I 2025-10-26 14:57:22,821] Trial 48 finished with value:
0.907423095970309 and parameters: {'n_estimators': 478,
'learning_rate': 0.05698860927149707, 'max_depth': 4,
'min_child_weight': 5, 'gamma': 0.24804720891270216, 'subsample':
0.8617089397338614, 'colsample_bytree': 0.5471966212674557,
'reg_alpha': 0.5603572010077237, 'reg_lambda': 0.4116193994912466}.
Best is trial 27 with value: 0.9089681562130227.
[I 2025-10-26 14:57:36,460] Trial 49 finished with value:
0.9071655341182224 and parameters: {'n_estimators': 468,
'learning_rate': 0.050713139136791575, 'max_depth': 5,
'min_child_weight': 4, 'gamma': 0.2163009657239446, 'subsample':
0.7252336657964157, 'colsample_bytree': 0.6239266311270306,
'reg_alpha': 0.5731928226133617, 'reg_lambda': 0.2972618360520062}.
Best is trial 27 with value: 0.9089681562130227.
[I 2025-10-26 14:57:48,890] Trial 50 finished with value:
0.9072942715224928 and parameters: {'n_estimators': 469,
'learning_rate': 0.051783502122241386, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.2821792758794816, 'subsample':
0.7527022392427434, 'colsample_bytree': 0.6081000004887369,
'reg_alpha': 0.4644800628651835, 'reg_lambda': 0.4412243854399002}.
Best is trial 27 with value: 0.9089681562130227.
[I 2025-10-26 14:58:00,854] Trial 51 finished with value:
0.9080024205069223 and parameters: {'n_estimators': 471,

```
'learning_rate': 0.05833848835578824, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.2359500227275941, 'subsample':
0.833802065300648, 'colsample_bytree': 0.5364928069183397,
'reg_alpha': 0.4906232602158351, 'reg_lambda': 0.3412919687296036}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 14:58:12,819] Trial 52 finished with value:
0.9079381077613521 and parameters: {'n_estimators': 473,
'learning_rate': 0.06062311512422782, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.23190242682206455, 'subsample':
0.8506156350489616, 'colsample_bytree': 0.5194282001900414,
'reg_alpha': 0.5029519459302393, 'reg_lambda': 0.32712839884167155}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 14:58:24,901] Trial 53 finished with value:
0.9067148475075416 and parameters: {'n_estimators': 471,
'learning_rate': 0.05911403346515178, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.22228324287333917, 'subsample':
0.8914804952040176, 'colsample_bytree': 0.569453067953705,
'reg_alpha': 0.48480276140720713, 'reg_lambda': 0.4854601758499848}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 14:58:37,019] Trial 54 finished with value:
0.9069080220053047 and parameters: {'n_estimators': 472,
'learning_rate': 0.05572650310303125, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.25863192718696476, 'subsample':
0.8242998968288268, 'colsample_bytree': 0.5860914249521798,
'reg_alpha': 0.5154952728872778, 'reg_lambda': 0.3705101420197111}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 14:58:50,103] Trial 55 finished with value:
0.9045902762063309 and parameters: {'n_estimators': 461,
'learning_rate': 0.06001286543627005, 'max_depth': 5,
'min_child_weight': 4, 'gamma': 0.24574725188995986, 'subsample':
0.8427409736919494, 'colsample_bytree': 0.7383692750064972,
'reg_alpha': 0.5283658848832283, 'reg_lambda': 0.30737585099273007}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 14:59:02,381] Trial 56 finished with value:
0.9078093206179126 and parameters: {'n_estimators': 470,
'learning_rate': 0.053030706424053185, 'max_depth': 4,
'min_child_weight': 5, 'gamma': 0.2674344640608036, 'subsample':
0.8016607203514021, 'colsample_bytree': 0.6158215796562874,
'reg_alpha': 0.4955011380823517, 'reg_lambda': 0.41337435226885016}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 14:59:15,914] Trial 57 finished with value:
0.9070368091487442 and parameters: {'n_estimators': 474,
'learning_rate': 0.05041181112465467, 'max_depth': 5,
'min_child_weight': 2, 'gamma': 0.29251941046939905, 'subsample':
0.7821494451399057, 'colsample_bytree': 0.5492157588482436,
'reg_alpha': 0.4381636070330574, 'reg_lambda': 0.16067901554468816}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 14:59:27,871] Trial 58 finished with value:
0.9072943212616619 and parameters: {'n_estimators': 467,
```

```
'learning_rate': 0.05401568240203184, 'max_depth': 4,
'min_child_weight': 4, 'gamma': 0.23052710472664398, 'subsample':
0.9023242247608824, 'colsample_bytree': 0.5082164342269219,
'reg_alpha': 0.5391337081850803, 'reg_lambda': 0.2641195816618723}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 14:59:41,109] Trial 59 finished with value:
0.906586060364102 and parameters: {'n_estimators': 472,
'learning_rate': 0.0550400396998648, 'max_depth': 5,
'min_child_weight': 5, 'gamma': 0.3790419668349649, 'subsample':
0.6571249759166224, 'colsample_bytree': 0.5993499816424933,
'reg_alpha': 0.46520319470956445, 'reg_lambda': 0.3593006335349509}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 14:59:49,925] Trial 60 finished with value:
0.9058136856776485 and parameters: {'n_estimators': 473,
'learning_rate': 0.05871900948204201, 'max_depth': 3,
'min_child_weight': 3, 'gamma': 0.25466276411685995, 'subsample':
0.9355452465463002, 'colsample_bytree': 0.6847533522121361,
'reg_alpha': 0.4877606074694653, 'reg_lambda': 0.4796842354528999}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:00:02,046] Trial 61 finished with value:
0.9078094200962507 and parameters: {'n_estimators': 472,
'learning_rate': 0.058364402984996365, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.2034335707879117, 'subsample':
0.8311929963682836, 'colsample_bytree': 0.5340575398242885,
'reg_alpha': 0.47290915756862967, 'reg_lambda': 0.33855697800410334}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:00:13,978] Trial 62 finished with value:
0.9081312449547384 and parameters: {'n_estimators': 470,
'learning_rate': 0.06086956306176626, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.2311392536121867, 'subsample':
0.870087796109512, 'colsample_bytree': 0.5389992754731373,
'reg_alpha': 0.47751661480861585, 'reg_lambda': 0.34931229366747935}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:00:25,872] Trial 63 finished with value:
0.9080668949047918 and parameters: {'n_estimators': 469,
'learning_rate': 0.06098147183461081, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.21915812550745833, 'subsample':
0.8761876987034176, 'colsample_bytree': 0.5241148833757541,
'reg_alpha': 0.4774148294966291, 'reg_lambda': 0.38347742581302596}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:00:37,835] Trial 64 finished with value:
0.9076805832136423 and parameters: {'n_estimators': 469,
'learning_rate': 0.06216047429203136, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.21199968372588301, 'subsample':
0.8741213956130306, 'colsample_bytree': 0.5005747056422744,
'reg_alpha': 0.47837328299964677, 'reg_lambda': 0.38065259149298714}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:00:49,560] Trial 65 finished with value:
0.9068436346509811 and parameters: {'n_estimators': 470,
```

```
'learning_rate': 0.061093461545572506, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.21750838749691298, 'subsample':
0.9086780904711053, 'colsample_bytree': 0.5166602507261815,
'reg_alpha': 0.4558937573430007, 'reg_lambda': 0.3212948335137018}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:01:01,807] Trial 66 finished with value:
0.9065218346620779 and parameters: {'n_estimators': 467,
'learning_rate': 0.06397862015105117, 'max_depth': 4,
'min_child_weight': 5, 'gamma': 0.20035373548250762, 'subsample':
0.9616528039442593, 'colsample_bytree': 0.5761213164649814,
'reg_alpha': 0.45898302485956094, 'reg_lambda': 0.4365377576969052}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:01:13,762] Trial 67 finished with value:
0.9078093827918741 and parameters: {'n_estimators': 475,
'learning_rate': 0.061991134464043626, 'max_depth': 4,
'min_child_weight': 4, 'gamma': 0.22545498814074588, 'subsample':
0.8827268115178533, 'colsample_bytree': 0.5436923069045135,
'reg_alpha': 0.49840110440394086, 'reg_lambda': 0.40455490578696174}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:01:24,937] Trial 68 finished with value:
0.9067149221162952 and parameters: {'n_estimators': 465,
'learning_rate': 0.061228338385600316, 'max_depth': 5,
'min_child_weight': 2, 'gamma': 0.3081304135297199, 'subsample':
0.85774712687897, 'colsample_bytree': 0.5592299699327908, 'reg_alpha':
0.479499034336154, 'reg_lambda': 0.3871857380146938}. Best is trial 27
with value: 0.9089681562130227.
[I  2025-10-26 15:01:35,767] Trial 69 finished with value:
0.9071011716334834 and parameters: {'n_estimators': 469,
'learning_rate': 0.060525419591757786, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.2707478650944307, 'subsample':
0.8691931629087278, 'colsample_bytree': 0.6317282820034641,
'reg_alpha': 0.4498618638537874, 'reg_lambda': 0.2991278553582622}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:01:46,629] Trial 70 finished with value:
0.9051055615628036 and parameters: {'n_estimators': 468,
'learning_rate': 0.05130770972753499, 'max_depth': 3,
'min_child_weight': 4, 'gamma': 0.23988050734833116, 'subsample':
0.9268947696185695, 'colsample_bytree': 0.6095563613291812,
'reg_alpha': 0.5867182359931615, 'reg_lambda': 0.3639660712863501}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:01:58,814] Trial 71 finished with value:
0.9071655838573914 and parameters: {'n_estimators': 471,
'learning_rate': 0.0593250767363866, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.2090190630564194, 'subsample':
0.8490820389304323, 'colsample_bytree': 0.5294328890445239,
'reg_alpha': 0.46976746243417217, 'reg_lambda': 0.3403642384347983}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:02:10,750] Trial 72 finished with value:
0.9079381201961444 and parameters: {'n_estimators': 471,
'learning_rate': 0.059718128618570405, 'max_depth': 4,
```

```
'min_child_weight': 3, 'gamma': 0.22122224090882509, 'subsample':
0.8908816492677514, 'colsample_bytree': 0.5218077026588849,
'reg_alpha': 0.5047810408478168, 'reg_lambda': 0.349845414323115}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:02:23,463] Trial 73 finished with value:
0.9060709615295132 and parameters: {'n_estimators': 470,
'learning_rate': 0.05963166105739637, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.23031956955433663, 'subsample':
0.6304694024299006, 'colsample_bytree': 0.6460130179510288,
'reg_alpha': 0.5115689896281197, 'reg_lambda': 0.41908719537543315}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:02:35,643] Trial 74 finished with value:
0.9074874335854637 and parameters: {'n_estimators': 474,
'learning_rate': 0.05776675919777195, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.2810981558083179, 'subsample':
0.8144615044770446, 'colsample_bytree': 0.5642643784559365,
'reg_alpha': 0.48664934229356727, 'reg_lambda': 0.3920689865041434}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:02:47,479] Trial 75 finished with value:
0.9080024453765067 and parameters: {'n_estimators': 472,
'learning_rate': 0.060937729827454615, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.2899534051983242, 'subsample':
0.8404098007467715, 'colsample_bytree': 0.5134992935181226,
'reg_alpha': 0.4767915526946482, 'reg_lambda': 0.2782073606004608}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:02:59,160] Trial 76 finished with value:
0.9081312449547384 and parameters: {'n_estimators': 473,
'learning_rate': 0.056515538384622976, 'max_depth': 4,
'min_child_weight': 5, 'gamma': 0.23744974818840792, 'subsample':
0.9124191312288359, 'colsample_bytree': 0.540485361234337,
'reg_alpha': 0.4667159696635432, 'reg_lambda': 0.3208994546821477}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:03:08,947] Trial 77 finished with value:
0.9063286601643149 and parameters: {'n_estimators': 473,
'learning_rate': 0.05721138463020763, 'max_depth': 4,
'min_child_weight': 5, 'gamma': 0.2503889443043257, 'subsample':
0.9098051830877745, 'colsample_bytree': 0.5557061049081367,
'reg_alpha': 0.4661801313132824, 'reg_lambda': 0.31733445097043894}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:03:22,267] Trial 78 finished with value:
0.9074874833246328 and parameters: {'n_estimators': 474,
'learning_rate': 0.05665017214035018, 'max_depth': 6,
'min_child_weight': 5, 'gamma': 0.25866823136301526, 'subsample':
0.8804256888631752, 'colsample_bytree': 0.5408324079479204,
'reg_alpha': 0.4395526378138762, 'reg_lambda': 0.44630890728152584}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:03:35,291] Trial 79 finished with value:
0.9078093454874973 and parameters: {'n_estimators': 479,
'learning_rate': 0.05632434860956904, 'max_depth': 5,
```

```
'min_child_weight': 5, 'gamma': 0.24187379398737002, 'subsample':
0.8697939582359405, 'colsample_bytree': 0.5830141355151387,
'reg_alpha': 0.45226875534865973, 'reg_lambda': 0.29294086957553794}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:03:46,905] Trial 80 finished with value:
0.9077450451767194 and parameters: {'n_estimators': 463,
'learning_rate': 0.05575047574985764, 'max_depth': 4,
'min_child_weight': 5, 'gamma': 0.30213886025839043, 'subsample':
0.8993560258944477, 'colsample_bytree': 0.5925088493281754,
'reg_alpha': 0.4695037842116105, 'reg_lambda': 0.24434985752787874}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:03:58,694] Trial 81 finished with value:
0.9072299463421305 and parameters: {'n_estimators': 472,
'learning_rate': 0.05795267503087275, 'max_depth': 4,
'min_child_weight': 5, 'gamma': 0.23623561877903657, 'subsample':
0.9148574766651032, 'colsample_bytree': 0.5285636328830046,
'reg_alpha': 0.4612534438416163, 'reg_lambda': 0.3299903326950642}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:04:10,670] Trial 82 finished with value:
0.9076161958593186 and parameters: {'n_estimators': 473,
'learning_rate': 0.058843827337618296, 'max_depth': 4,
'min_child_weight': 5, 'gamma': 0.2215743962116025, 'subsample':
0.8594631790172119, 'colsample_bytree': 0.5528085694234834,
'reg_alpha': 0.491866970465197, 'reg_lambda': 0.37240730738251365}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:04:22,503] Trial 83 finished with value:
0.908131269824323 and parameters: {'n_estimators': 471,
'learning_rate': 0.06179174973950056, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.2354483175331867, 'subsample':
0.8916453572153938, 'colsample_bytree': 0.510576928866652,
'reg_alpha': 0.49841134540379944, 'reg_lambda': 0.3077394834244541}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:04:34,105] Trial 84 finished with value:
0.9076162580332799 and parameters: {'n_estimators': 476,
'learning_rate': 0.06161978270254131, 'max_depth': 4,
'min_child_weight': 4, 'gamma': 0.27770859195851216, 'subsample':
0.8882674052379, 'colsample_bytree': 0.5105358914157336, 'reg_alpha':
0.4982352125891988, 'reg_lambda': 0.312589039786891}. Best is trial 27
with value: 0.9089681562130227.
[I  2025-10-26 15:04:46,061] Trial 85 finished with value:
0.907423095970309 and parameters: {'n_estimators': 469,
'learning_rate': 0.052507726230418605, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.22687591599564952, 'subsample':
0.9295040388711716, 'colsample_bytree': 0.5401830078538065,
'reg_alpha': 0.5238058214985762, 'reg_lambda': 0.3451732344108055}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:04:56,023] Trial 86 finished with value:
0.9074874584550482 and parameters: {'n_estimators': 470,
'learning_rate': 0.06314885662804279, 'max_depth': 5,
```

```
'min_child_weight': 5, 'gamma': 0.27124491516202753, 'subsample':
0.9580011894510142, 'colsample_bytree': 0.5254489290041796,
'reg_alpha': 0.44561611016240266, 'reg_lambda': 0.2727502107267696}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:05:06,758] Trial 87 finished with value:
0.9073588080943237 and parameters: {'n_estimators': 467,
'learning_rate': 0.0623952018042361, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.2618519109270419, 'subsample':
0.9397063104236831, 'colsample_bytree': 0.5068254247350841,
'reg_alpha': 0.4744982388775791, 'reg_lambda': 0.30231750094388615}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:05:19,275] Trial 88 finished with value:
0.9078093081831203 and parameters: {'n_estimators': 475,
'learning_rate': 0.054950887743818455, 'max_depth': 4,
'min_child_weight': 4, 'gamma': 0.2358738765212859, 'subsample':
0.6849992932022525, 'colsample_bytree': 0.6185586908667415,
'reg_alpha': 0.48033630066779315, 'reg_lambda': 0.361045390063125}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:05:30,727] Trial 89 finished with value:
0.9071011965030679 and parameters: {'n_estimators': 471,
'learning_rate': 0.060151565143621556, 'max_depth': 5,
'min_child_weight': 4, 'gamma': 0.2957503660547396, 'subsample':
0.9738926335844249, 'colsample_bytree': 0.6685933522353787,
'reg_alpha': 0.48511398922484783, 'reg_lambda': 0.47247623577040676}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:05:41,575] Trial 90 finished with value:
0.9056849358385858 and parameters: {'n_estimators': 468,
'learning_rate': 0.053199868359386365, 'max_depth': 3,
'min_child_weight': 5, 'gamma': 0.2463868648267856, 'subsample':
0.8965562878899832, 'colsample_bytree': 0.5642595257176909,
'reg_alpha': 0.5538889547097657, 'reg_lambda': 0.32815114045603483}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:05:53,550] Trial 91 finished with value:
0.9085818693914579 and parameters: {'n_estimators': 471,
'learning_rate': 0.06093762285519829, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.21875619619473838, 'subsample':
0.8648353117135467, 'colsample_bytree': 0.5351955371320672,
'reg_alpha': 0.49291119527482313, 'reg_lambda': 0.28462242147548705}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:06:05,686] Trial 92 finished with value:
0.9069080593096815 and parameters: {'n_estimators': 472,
'learning_rate': 0.061805997923857874, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.21910841724927987, 'subsample':
0.8777392288688356, 'colsample_bytree': 0.6400471596447667,
'reg_alpha': 0.5028295442101585, 'reg_lambda': 0.26217310512905634}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:06:17,916] Trial 93 finished with value:
0.908259945054632 and parameters: {'n_estimators': 473,
'learning_rate': 0.06066718097976369, 'max_depth': 4,
```

```
'min_child_weight': 3, 'gamma': 0.21179942957054923, 'subsample':
0.8637648648868944, 'colsample_bytree': 0.534767250250937,
'reg_alpha': 0.5131162063772005, 'reg_lambda': 0.2859740004004305}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:06:29,649] Trial 94 finished with value:
0.9074874211506714 and parameters: {'n_estimators': 471,
'learning_rate': 0.060748025936919206, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.20672239736486378, 'subsample':
0.8633857817767561, 'colsample_bytree': 0.5362762989215323,
'reg_alpha': 0.514202249748965, 'reg_lambda': 0.2462732978173794}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:06:39,895] Trial 95 finished with value:
0.9089037688586993 and parameters: {'n_estimators': 473,
'learning_rate': 0.0626240383525263, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.21346170985183197, 'subsample':
0.8545793107600768, 'colsample_bytree': 0.5227820799494255,
'reg_alpha': 0.5096117129212134, 'reg_lambda': 0.28305828189406573}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:06:51,903] Trial 96 finished with value:
0.9075518458093716 and parameters: {'n_estimators': 473,
'learning_rate': 0.06409167391078528, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.21299501953246813, 'subsample':
0.8839870678839793, 'colsample_bytree': 0.5235784612013887,
'reg_alpha': 0.5079453200967148, 'reg_lambda': 0.2846431258261355}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:07:03,857] Trial 97 finished with value:
0.907809295748328 and parameters: {'n_estimators': 470,
'learning_rate': 0.06268120253351009, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.22776897811545285, 'subsample':
0.8510735952896923, 'colsample_bytree': 0.5152052311186343,
'reg_alpha': 0.4950884538620786, 'reg_lambda': 0.25637754221317904}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:07:15,811] Trial 98 finished with value:
0.9081955701351007 and parameters: {'n_estimators': 469,
'learning_rate': 0.0632786947824558, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.21007337811519916, 'subsample':
0.8222377508741362, 'colsample_bytree': 0.5312243813003265,
'reg_alpha': 0.5339541656294005, 'reg_lambda': 0.28545689777073435}.
Best is trial 27 with value: 0.9089681562130227.
[I  2025-10-26 15:07:27,947] Trial 99 finished with value:
0.9076805459092654 and parameters: {'n_estimators': 470,
'learning_rate': 0.06481522866534237, 'max_depth': 4,
'min_child_weight': 3, 'gamma': 0.21129599874531252, 'subsample':
0.8288335069479867, 'colsample_bytree': 0.5320438540755655,
'reg_alpha': 0.5193626947735854, 'reg_lambda': 0.27468557277828676}.
Best is trial 27 with value: 0.9089681562130227.
Best hyperparameters: {'n_estimators': 473, 'learning_rate':
0.05654568747242938, 'max_depth': 4, 'min_child_weight': 4, 'gamma':
0.27855847794120936, 'subsample': 0.8811061714182321,
'colsample_bytree': 0.5936850968585458, 'reg_alpha':
```

```
0.47089089488880537, 'reg_lambda': 0.2945379203470715}Best accuracy:
0.9089681562130227
Accuracy on the training data: 0.940706882121934


# fit tarsnform predict
x_transformed = transformer.fit_transform(x)
test_transformed = transformer.transform(test_df)
test_predictions_encoded = final_model.predict(test_transformed)

# decode
test_predictions =
encoder.inverse_transform(test_predictions_encoded.reshape(-1, 1))

# submission
submission_df = pd.DataFrame({'id': test_df.index + 15533,
'Weight_Category': test_predictions.flatten()})
submission_df.to_csv('xgb_optuna.csv', index=False)

print("xgb_optuna.csv created successfully with predictions from the
optimized model!")

xgb_optuna.csv created successfully with predictions from the
optimized model!

!pip install optuna

Requirement already satisfied: optuna in
/usr/local/lib/python3.12/dist-packages (4.5.0)
Requirement already satisfied: alembic>=1.5.0 in
/usr/local/lib/python3.12/dist-packages (from optuna) (1.17.0)
Requirement already satisfied: colorlog in
/usr/local/lib/python3.12/dist-packages (from optuna) (6.10.1)
Requirement already satisfied: numpy in
/usr/local/lib/python3.12/dist-packages (from optuna) (2.0.2)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.12/dist-packages (from optuna) (25.0)
Requirement already satisfied: sqlalchemy>=1.4.2 in
/usr/local/lib/python3.12/dist-packages (from optuna) (2.0.44)
Requirement already satisfied: tqdm in /usr/local/lib/python3.12/dist-
packages (from optuna) (4.67.1)
Requirement already satisfied: PyYAML in
/usr/local/lib/python3.12/dist-packages (from optuna) (6.0.3)
Requirement already satisfied: Mako in /usr/local/lib/python3.12/dist-
packages (from alembic>=1.5.0->optuna) (1.3.10)
Requirement already satisfied: typing-extensions>=4.12 in
/usr/local/lib/python3.12/dist-packages (from alembic>=1.5.0->optuna)
(4.15.0)
Requirement already satisfied: greenlet>=1 in
/usr/local/lib/python3.12/dist-packages (from sqlalchemy>=1.4.2-
>optuna) (3.2.4)
```

```
Requirement already satisfied: MarkupSafe>=0.9.2 in
/usr/local/lib/python3.12/dist-packages (from Mako->alembic>=1.5.0-
>optuna) (3.0.3)
```