```python
# Step 1: Import Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, confusion_matrix,
classification_report

# Step 2: Load Dataset
# Upload your CSV file from local or Google Drive
# Example: df = pd.read_csv("/content/student_dropout.csv")
df = pd.read_csv("student_dropout_dataset.csv")  # replace file name
if different
print("✅ Dataset Loaded Successfully\n")
print(df.head())

# Step 3: Check for Null Values
print("\nMissing Values in Dataset:\n", df.isnull().sum())

# Step 4: Encode Categorical Columns
label = LabelEncoder()
for column in df.select_dtypes(include=['object']).columns:
    df[column] = label.fit_transform(df[column])

print("\n✅ Encoded categorical columns successfully\n")
print(df.head())

# Step 5: Define Features (X) and Target (y)
X = df.drop('dropout_status', axis=1)
y = df['dropout_status']

# Step 6: Split the Dataset
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Step 7: Feature Scaling
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# Step 8: Train the Model (Random Forest)
model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Step 9: Make Predictions
y_pred = model.predict(X_test)
```

```
# Step □: Evaluate the Model
print("□ Model Accuracy:", accuracy_score(y_test, y_pred))
print("\n□ Confusion Matrix:\n", confusion_matrix(y_test, y_pred))
print("\n□ Classification Report:\n", classification_report(y_test,
y_pred))

# Step 11: Feature Importance Visualization
importances = model.feature_importances_
features = X.columns
plt.figure(figsize=(8,5))
plt.barh(features, importances)
plt.title("Feature Importance in Dropout Prediction")
plt.xlabel("Importance")
plt.ylabel("Features")
plt.show()

print("\n□ Project Completed Successfully - Online Course Dropout
Prediction")
# ------------------------------------
# □ Visual Dashboard – Dropout Insights
# ------------------------------------
import matplotlib.pyplot as plt
import pandas as pd

# 1 Dropout Distribution
plt.figure(figsize=(5,4))
df['dropout_status'].value_counts().plot(kind='bar',
color=['skyblue','salmon'])
plt.title('Dropout vs Completion Count')
plt.xlabel('Dropout Status (0 = Completed, 1 = Dropped Out)')
plt.ylabel('Number of Students')
plt.show()

# 2 Dropout by Gender
if 'gender' in df.columns:
    plt.figure(figsize=(6,4))
    pd.crosstab(df['gender'], df['dropout_status']).plot(kind='bar')
    plt.title('Dropout Status by Gender')
    plt.xlabel('Gender')
    plt.ylabel('Count')
    plt.show()

# 3 Dropout vs Motivation Level
if 'motivation_level' in df.columns:
    plt.figure(figsize=(6,4))
    pd.crosstab(df['motivation_level'],
df['dropout_status']).plot(kind='bar')
    plt.title('Motivation Level vs Dropout')
    plt.xlabel('Motivation Level')
    plt.ylabel('Number of Students')
```

```
    plt.show()

# 4 Average Attendance Rate by Dropout Status
if 'attendance_rate' in df.columns:
    plt.figure(figsize=(6,4))
    df.groupby('dropout_status')
['attendance_rate'].mean().plot(kind='bar')
    plt.title('Average Attendance Rate by Dropout Status')
    plt.xlabel('Dropout Status')
    plt.ylabel('Average Attendance Rate (%)')
    plt.show()

# 5 Average Login Frequency by Dropout Status
if 'login_frequency' in df.columns:
    plt.figure(figsize=(6,4))
    df.groupby('dropout_status')
['login_frequency'].mean().plot(kind='bar')
    plt.title('Average Login Frequency by Dropout Status')
    plt.xlabel('Dropout Status')
    plt.ylabel('Average Logins')
    plt.show()

# 6 Correlation Heatmap
plt.figure(figsize=(10,8))
corr = df.corr(numeric_only=True)
plt.imshow(corr, cmap='coolwarm', interpolation='none')
plt.colorbar()
plt.title('Correlation Heatmap — Dropout Factors')
plt.xticks(range(len(corr.columns)), corr.columns, rotation=90)
plt.yticks(range(len(corr.columns)), corr.columns)
plt.show()

print("\n Dashboard Visualizations Generated Successfully!")
```

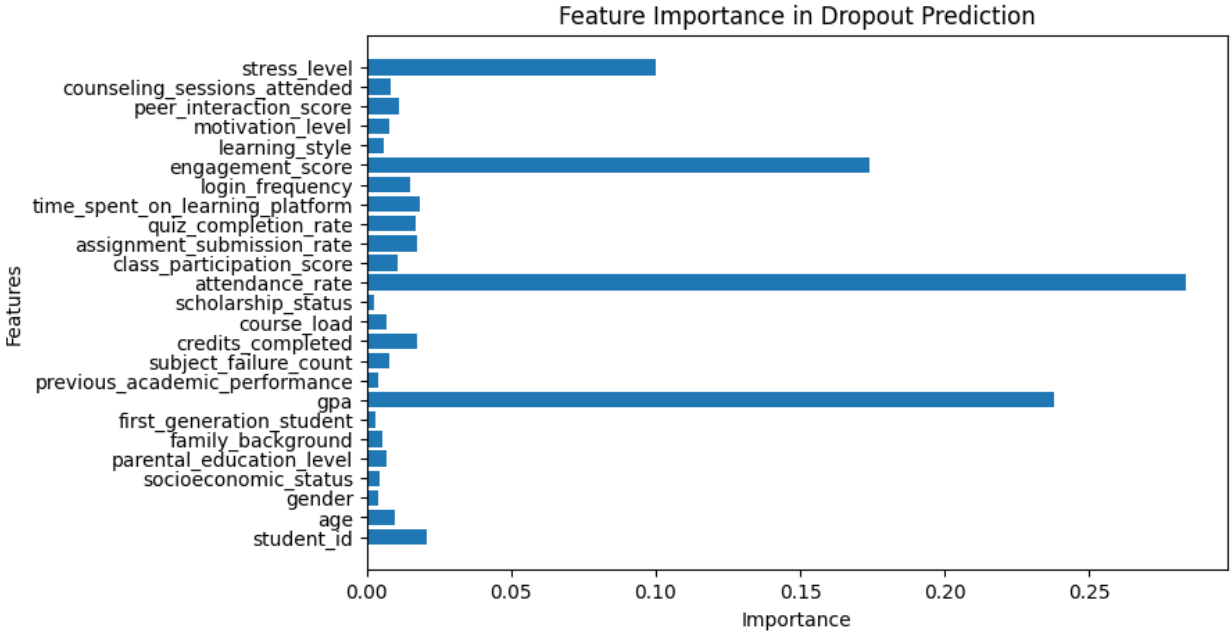
```
 Dataset Loaded Successfully

  student_id  age  gender socioeconomic_status
parental_education_level  \
0      S0001   21    Male                 High
Primary
1      S0002   25  Female                  Low
Secondary
2      S0003   19    Male                  Low
Primary
3      S0004   21  Female               Medium
NaN
4      S0005   20    Male                  Low
Graduate

  family_background first_generation_student   gpa  \
```

```
0           Nuclear                       Yes  2.39
1             Joint                        No  1.78
2          Guardian                        No  3.88
3             Joint                       Yes  2.44
4             Joint                       Yes  3.08

  previous_academic_performance  subject_failure_count  ...  \
0                        Medium                      1  ...
1                          High                      5  ...
2                        Medium                      1  ...
3                           Low                      5  ...
4                           Low                      1  ...

   quiz_completion_rate  time_spent_on_learning_platform
login_frequency  \
0                    27                             19.0
12
1                    35                             17.2
29
2                    27                             15.7
2
3                    94                             19.9
24
4                    30                              5.3
29

   engagement_score  learning_style  motivation_level
peer_interaction_score  \
0                 1          Visual                 2
4
1                 6     Kinesthetic                 2
6
2                10           Mixed                 4
0
3                 5          Visual                 5
1
4                 1        Auditory                 1
9

   counseling_sessions_attended  stress_level  dropout_status
0                             2        Medium               0
1                             1          High               1
2                             1        Medium               0
3                             2           Low               0
4                             0        Medium               1

[5 rows x 26 columns]

Missing Values in Dataset:
 student_id                          0
```

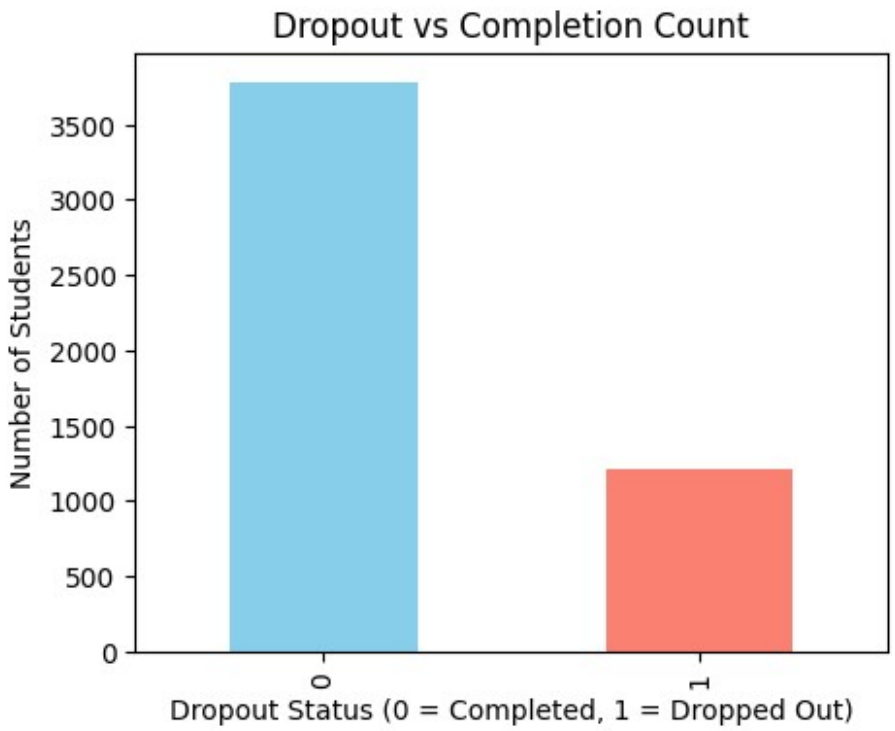
```
age                                   0
gender                                0
socioeconomic_status                  0
parental_education_level            519
family_background                     0
first_generation_student              0
gpa                                   0
previous_academic_performance         0
subject_failure_count                 0
credits_completed                     0
course_load                           0
scholarship_status                    0
attendance_rate                       0
class_participation_score             0
assignment_submission_rate            0
quiz_completion_rate                  0
time_spent_on_learning_platform       0
login_frequency                       0
engagement_score                      0
learning_style                        0
motivation_level                      0
peer_interaction_score                0
counseling_sessions_attended          0
stress_level                          0
dropout_status                        0
dtype: int64

 Encoded categorical columns successfully

   student_id  age  gender  socioeconomic_status
parental_education_level  \
0           0   21       1                     0
2
1           1   25       0                     1
3
2           2   19       1                     1
2
3           3   21       0                     2
4
4           4   20       1                     1
0

   family_background  first_generation_student   gpa  \
0                  2                         1  2.39
1                  1                         0  1.78
2                  0                         0  3.88
3                  1                         1  2.44
4                  1                         1  3.08

   previous_academic_performance  subject_failure_count  ...  \
```

```
0                                2                       1  ...
1                                0                       5  ...
2                                2                       1  ...
3                                1                       5  ...
4                                1                       1  ...

   quiz_completion_rate  time_spent_on_learning_platform
login_frequency  \
0                    27                             19.0
12
1                    35                             17.2
29
2                    27                             15.7
2
3                    94                             19.9
24
4                    30                              5.3
29

   engagement_score  learning_style  motivation_level
peer_interaction_score  \
0                 1               3                 2
4
1                 6               1                 2
6
2                10               2                 4
0
3                 5               3                 5
1
4                 1               0                 1
9

   counseling_sessions_attended  stress_level  dropout_status
0                             2             2               0
1                             1             0               1
2                             1             2               0
3                             2             1               0
4                             0             2               1

[5 rows x 26 columns]
 Model Accuracy: 1.0

 Confusion Matrix:
 [[764   0]
 [  0 236]]

 Classification Report:
               precision    recall  f1-score   support

           0       1.00      1.00      1.00       764
```

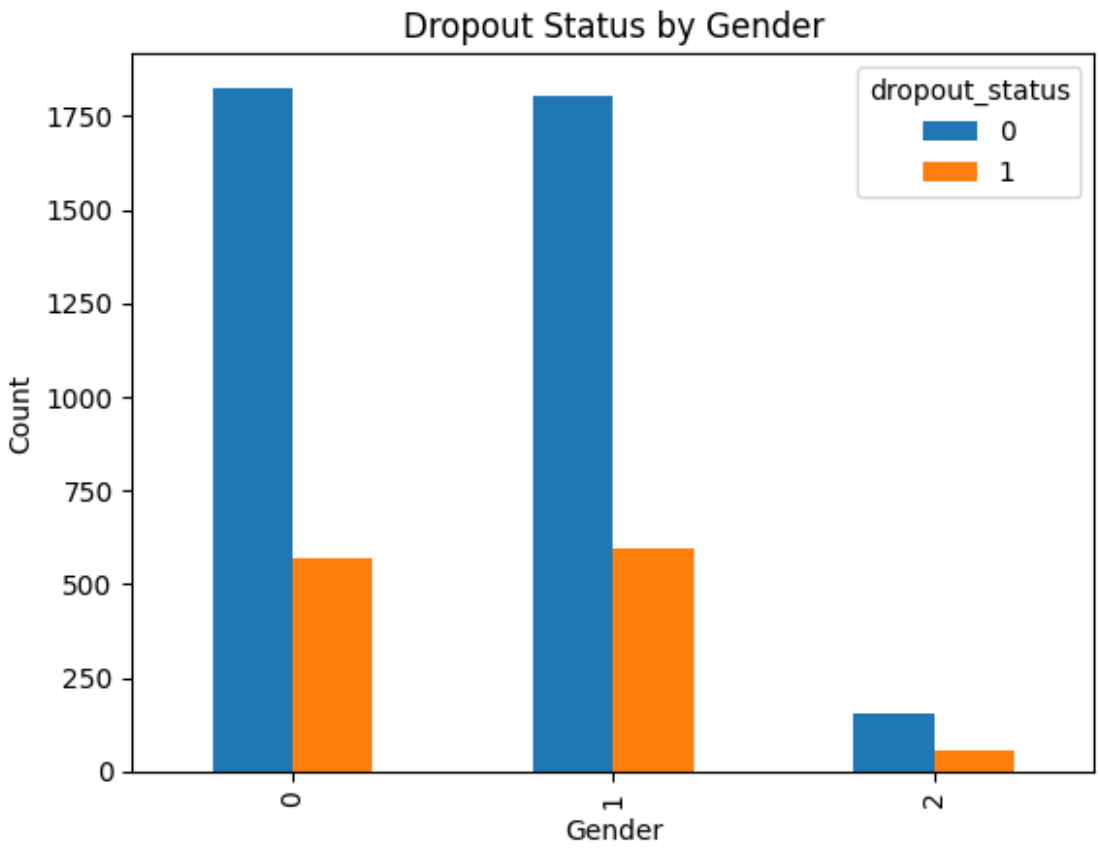
```
           1       1.00      1.00      1.00       236

    accuracy                           1.00      1000
   macro avg       1.00      1.00      1.00      1000
weighted avg       1.00      1.00      1.00      1000
```



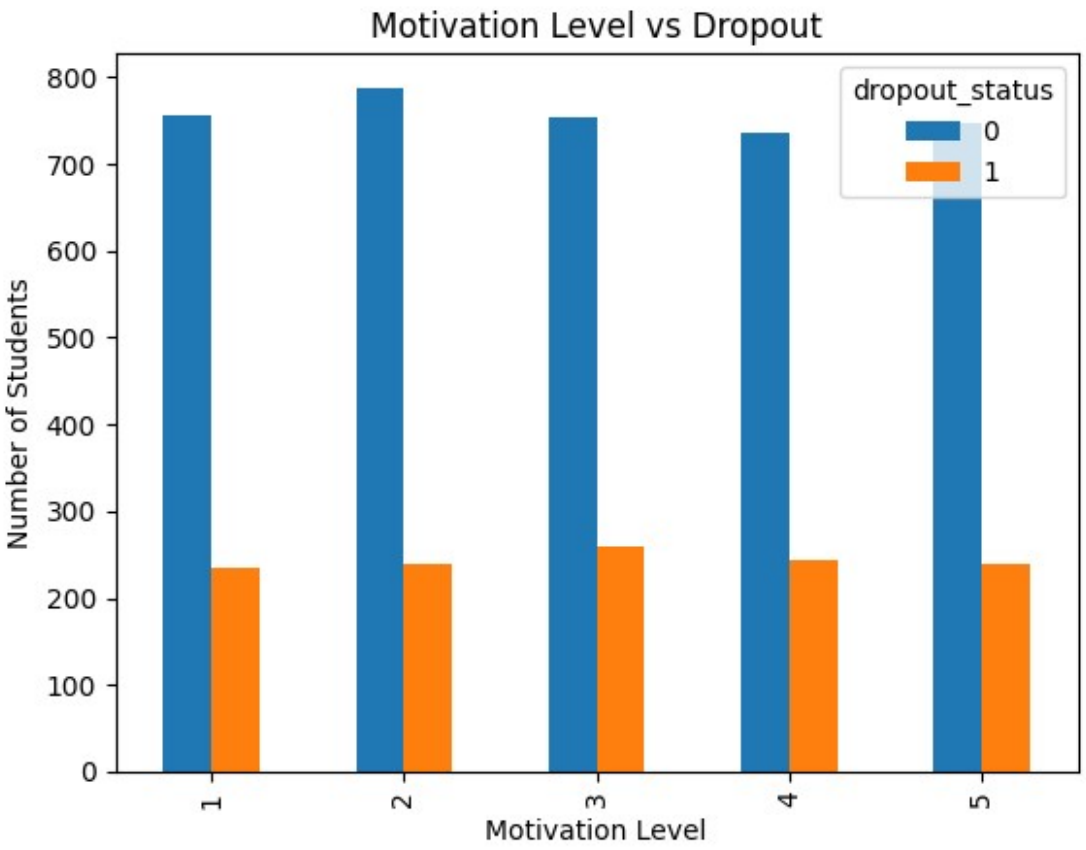

```
 Project Completed Successfully - Online Course Dropout Prediction
```

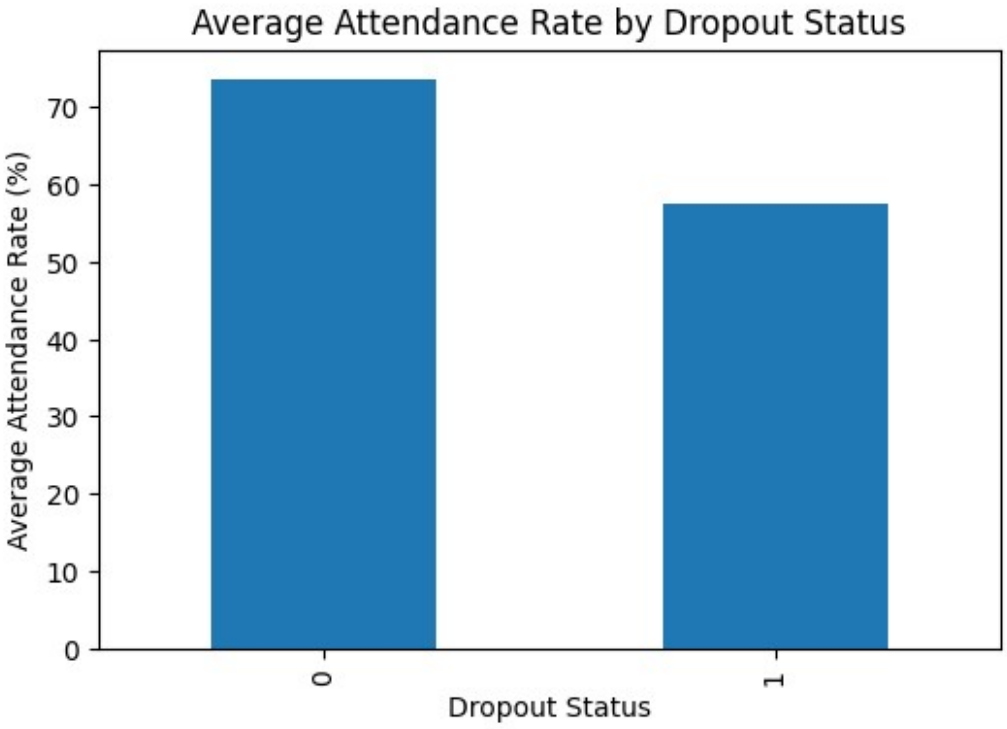


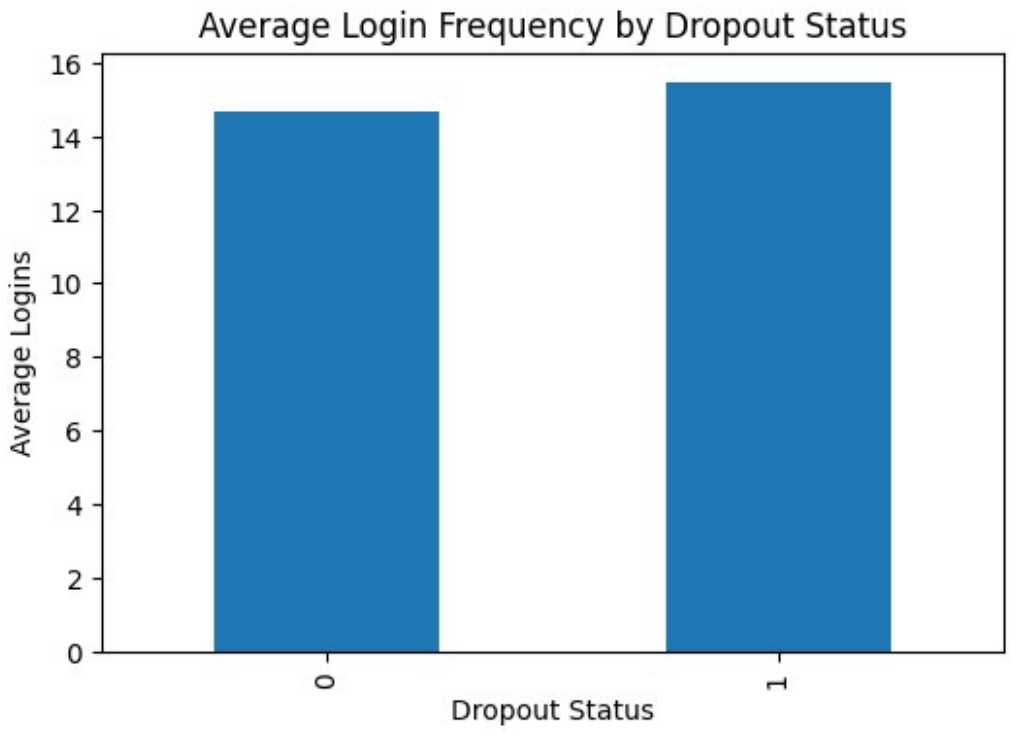

```
<Figure size 600x400 with 0 Axes>
```

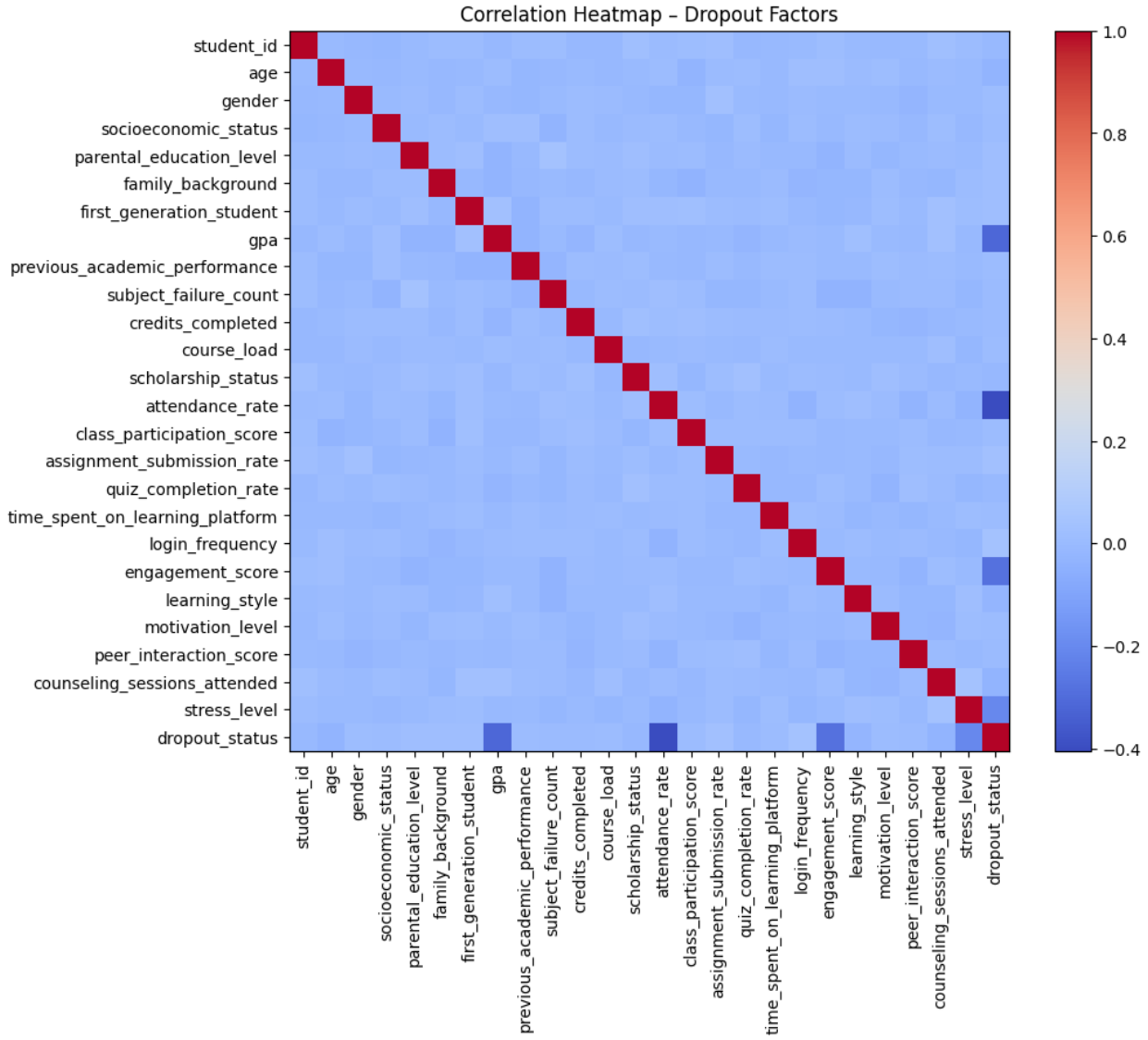



```
<Figure size 600x400 with 0 Axes>
```


Motivation Level vs Dropout
dropout_status
0
1
Number of Students
800
700
600
500
400
300
200
100
0
1
2
3
4
5
Motivation Level


Average Attendance Rate by Dropout Status
Average Attendance Rate (%)
70
60
50
40
30
20
10
0
0
1
Dropout Status


Average Login Frequency by Dropout Status
Average Logins
16
14
12
10
8
6
4
2
0
0
1
Dropout Status




```
 Dashboard Visualizations Generated Successfully!
```