```python
import pandas as pd

url = 'https://stats.idre.ucla.edu/stat/data/binary.csv'
df = pd.read_csv(url)

print(df.info())
print(df.head())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 400 entries, 0 to 399
Data columns (total 4 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   admit   400 non-null    int64
 1   gre     400 non-null    int64
 2   gpa     400 non-null    float64
 3   rank    400 non-null    int64
dtypes: float64(1), int64(3)
memory usage: 12.6 KB
None
   admit  gre   gpa  rank
0      0  380  3.61     3
1      1  660  3.67     3
2      1  800  4.00     1
3      1  640  3.19     4
4      0  520  2.93     4
```

```python
print(df.isna().sum())
```

```
admit    0
gre      0
gpa      0
rank     0
dtype: int64
```

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

```
            admit         gre         gpa        rank
count  400.000000  400.000000  400.000000  400.00000
mean     0.317500  587.700000    3.389900    2.48500
std      0.466087  115.516536    0.380567    0.94446
min      0.000000  220.000000    2.260000    1.00000
25%      0.000000  520.000000    3.130000    2.00000
50%      0.000000  580.000000    3.395000    2.00000
75%      1.000000  660.000000    3.670000    3.00000
max      1.000000  800.000000    4.000000    4.00000
```

```python
print(pd.crosstab(df['admit'], df['rank']))
```

```
rank    1   2   3   4
admit
```

```
0      28  97  93  55
1      33  54  28  12
```

```
import pandas as pd
import statsmodels.api as sm

url = 'https://stats.idre.ucla.edu/stat/data/binary.csv'
df = pd.read_csv(url)

df['rank'] = df['rank'].astype('category')

X = df[['gre', 'gpa', 'rank']]
X = sm.add_constant(X)

y = df['admit']

logit_model = sm.Logit(y, X).fit()

print(logit_model.summary())
```

```
Optimization terminated successfully.
         Current function value: 0.574302
         Iterations 6
                           Logit Regression Results

==============================================================================
========
Dep. Variable:                    admit   No. Observations:
400
Model:                            Logit   Df Residuals:
396
Method:                             MLE   Df Model:
3
Date:                  Thu, 13 Feb 2025   Pseudo R-squ.:
0.08107
Time:                          00:56:42   Log-Likelihood:
-229.72
converged:                         True   LL-Null:
-249.99
Covariance Type:             nonrobust   LLR p-value:
8.207e-09
==============================================================================
========
                 coef    std err          z      P>|z|      [0.025
0.975]
------------------------------------------------------------------------------
--------
const         -3.4495      1.133     -3.045      0.002      -5.670
-1.229
gre            0.0023      0.001      2.101      0.036       0.000
```

```
0.004
gpa             0.7770      0.327       2.373       0.018        0.135
1.419
rank           -0.5600      0.127      -4.405       0.000       -0.809
-0.311
==============================================================================
========

new_data = pd.DataFrame({'gre': [790], 'gpa': [3.8], 'rank': [1]})

new_data['rank'] = new_data['rank'].astype('category')

new_data = sm.add_constant(new_data, has_constant='add')

probability = logit_model.predict(new_data)
print(probability)

0    0.680332
dtype: float64
```