```
Attempting uninstall: scikit-learn
    Found existing installation: scikit-learn 1.2.2
    Uninstalling scikit-learn-1.2.2:
        Successfully uninstalled scikit-learn-1.2.2
Successfully installed scikit-learn-1.4.dev0
```

[ ]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn import metrics
```

[ ]:
```python
import io
from google.colab import files
url = 'https://raw.githubusercontent.com/santhoshpkumar/StudentAdmissionsKeras/master/student_data.c
df = pd.read_csv(url, header = 0, names = ["admit", 'gre', 'gpa', 'rank'])
```

[ ]:
```python
df.describe()
```

[ ]:

	admit	gre	gpa	rank
count	400.000000	398.000000	398.00000	399.000000
mean	0.317500	588.040201	3.39093	2.486216
std	0.466087	115.628513	0.38063	0.945333
min	0.000000	220.000000	2.26000	1.000000
25%	0.000000	520.000000	3.13000	2.000000
50%	0.000000	580.000000	3.39500	2.000000
75%	1.000000	660.000000	3.67000	3.000000
max	1.000000	800.000000	4.00000	4.000000

[ ]:
```python
# Generate a cross-tabulation (frequency table by default) of the factors; here we use prestige
pd.crosstab(df['admit'], df['rank'], rownames=['Admission'])
```

[ ]:

rank	1.0	2.0	3.0	4.0
Admission				
0	28	97	93	55
1	33	53	28	12

[ ]:
```python
sns.set_color_codes('muted')
df.hist(color='g')
plt.show()
```
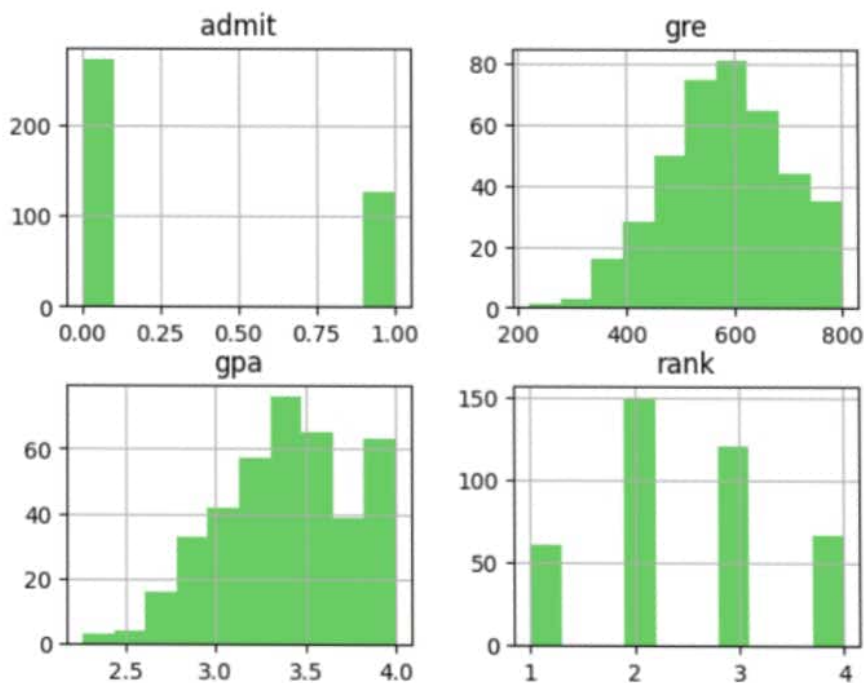
```
dummy_ranks = pd.get_dummies(df['rank'], prefix="rank")
dummy_ranks.head()
```

	rank_1.0	rank_2.0	rank_3.0	rank_4.0
0	0	0	1	0
1	0	0	1	0
2	1	0	0	0
3	0	0	0	1
4	0	0	0	1

```
columns1 = ['admit', 'gre', 'gpa']
data1 = df[columns1]
columns2 = ['rank_1.0','rank_2.0','rank_3.0']
data2 = dummy_ranks[columns2]
data = pd.merge(data1, data2, how="outer", left_index=True, right_index=True)
data
```

	admit	gre	gpa	rank_1.0	rank_2.0	rank_3.0
0	0	380.0	3.61	0	0	1
1	1	660.0	3.67	0	0	1
2	1	800.0	4.00	1	0	0
3	1	640.0	3.19	0	0	0
4	0	520.0	2.93	0	0	0
...	...	...	...	...	...	...
395	0	620.0	4.00	0	1	0
396	0	560.0	3.04	0	0	1
397	0	460.0	2.63	0	1	0
398	0	700.0	3.65	0	1	0
399	0	600.0	3.89	0	0	1

400 rows × 6 columns