

```
In [1]: #libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

#filterwarnings and autosave
import warnings
warnings.filterwarnings("ignore")
%autosave 1

Autosaving every 1 seconds
```

```
In [2]: df = pd.read_csv("admission_data.csv")
df
```

```
Out[2]:
```

	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
0	337	118	4	4.5	4.5	9.65	1	0.92
1	324	107	4	4.0	4.5	8.87	1	0.76
2	316	104	3	3.5	3.5	8.00	1	0.72
3	322	110	3	3.5	2.5	8.67	1	0.80
4	314	103	2	2.0	3.0	8.21	0	0.65
...
495	332	108	5	4.5	4.0	9.02	1	0.87
496	337	117	5	5.0	5.0	9.87	1	0.96
497	330	120	5	4.5	5.0	9.56	1	0.93
498	312	103	4	4.0	5.0	8.43	0	0.73
499	327	113	4	4.5	4.5	9.04	0	0.84

```
500 rows x 8 columns
```

```
In [3]: df.info()
```

```
Out[3]:
```

```
In [4]: df.isna().sum()
```

```
Out[4]:
```

	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
#	0	0	0	0	0	0	0	0
Count	500	500	500	500	500	500	500	500
Non-Null Count	500	500	500	500	500	500	500	500
Dtype	int64	int64	int64	float64	float64	float64	float64	float64

```
In [5]: #finding target variable
#check for the unique values so that we can divide the data into categorical and numerical
df.uniquie().sort_values()
```

```
Out[5]:
```

	Research	University Rating	SOP	LOR	CGPA	Chance of Admit
Count	2	5	9	9	61	184
unique	0	5	9	9	61	184
dtype	int64	int64	int64	int64	float64	float64

```
In [6]: #so from the above we can set Research as target variable
```

```
In [7]:
```

```
for col in df.columns:
    if col != "Research":
        plt.figure(figsize=(10,5))
        sns.histplot(data=df, x=col,multiple="dodge",hue="Research",kde=True,bins=5)
        plt.title(f"{col}\n",color="red")
        plt.show()
        print()
```

```
GRE Score
```

```
TOEFL Score
```

```
University Rating
```

```
SOP
```

```
LOR
```

```
CGPA
```

```
Chance of Admit
```