

Exploratory Data Analysis

May 29, 2024

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

```
[2]: df=pd.read_csv(r'C:\Users\ART\Downloads\All the Intermediate Colleges in _
sPakistan.csv')
```

```
[3]: df.shape
```

[3]: (1630, 6)

```
[4]: df.head(10)
```

```
[4]:
```

	Name \
0	Campus.pk Lahore
1	Allama Iqbal Open University Aiou Islamabad
2	Government College University Gcu Lahore
3	Govt College University GCU Faisalabad
4	Lahore College For Women University (Lcwu) Lahore
5	University Of Agriculture Faisalabad
6	Pir Mehr Ali Shah Arid Agriculture University ...
7	International Islamic University Iiu Islamabad
8	Kinnaird College For Women Lahore
9	The Women University Multan

	Location \
0	Office 206, 2nd Floor Siddique Trade Center Ma...
1	Sector H-8, Islamabad-44000
2	GC University, Katchery Road Lahore 54000
3	Kotwali Rd, Faisalabad 38000
4	Near Wapda Flats, Jail Road, Lahore
5	University of Agriculture, Agriculture Univers...
6	PMAS-Arid Agriculture University Rawalpindi, S...
7	International Islamic University Sector H-10, ...
8	93- Jail Road, Lahore
9	Kutchery Campus, L.M.Q. Road, Multan.

	Study Program	Sector Affiliation \
0	FSc Pre Engineering Fsc Pre Medical Private	ads
1	FSc Pre Engineering ICs Computer Science FSc P...	Public NaN
2	FSc Pre Engineering ICS Computer Science FSc P...	Public HEC
3	FSc Pre-Engineering ICS FSc Pre-Medical Icom F...	Public HEC
4	FA Arts FSc Pre Engineering ICS Physics ICS St...	Public NaN
5	FSc Pre Agriculture	Public NaN
6	FSc Pre Agriculture	Public NaN
7	FSc Pre-Engineering ICS FSc Pre-Medical Icom F...	Public NaN
8	FSc Pre Engineering ICS ICS Statistics ICS Phy...	Public NaN
9	FSc Pre Engineering ICS FSc Pre Medical Genera...	Public NaN

Rating

0	0*
1	5*
2	4*
3	3*
4	5*
5	5*
6	5*
7	5*
8	5*
9	0*

[5]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1630 entries, 0 to 1629 Data
columns (total 6 columns):
```

#	Column	Non-Null Count	Dtype
0	Name	1630 non-null	object
1	Location	1626 non-null	object
2	Study Program	1630 non-null	object
3	Sector	1101 non-null	object
4	Affiliation	155 non-null	object
5	Rating	1630 non-null	object

[6]:

```
dtypes: object(6)
memory usage: 76.5+ KB
```

```
pd.isnull(df)
```

0	False	False	False	False	False	False
1	False	False	False	False	True	False
2	False	False	False	False	False	False

3	False	False	False	False	False	False	False
4	False	False	False	False	False	True	False
...
1625	False	False	False	False	False	True	False
1626	False	False	False	False	False	True	False
1627	False	False	False	False	False	True	False
1628	False	False	False	False	False	True	False
1629	False	False	False	False	False	True	False

[1630 rows x 6 columns]

```
[7] : pd.isnull(df).sum()
```

```
[7]: Name          0
     Location      4
     Study Program  0
     Sector        529
     Affiliation    1475
     Rating         0
     dtype: int64
```

```
[8] : df.dropna(inplace=True)
```

```
[9] : pd.isnull(df).sum()
```

```
[9] : Name          0
     Location      0
     Study Program  0
     Sector        0
     Affiliation    0
     Rating         0
     dtype: int64
```

```
[10] : df['Rating'] = df['Rating'].str.replace('*', '').astype(float)
```

```
[11] : print(df)
```

	Name \
0	Campus.pk Lahore
2	Government College University Gcu Lahore
3	Govt College University GCU Faisalabad
12	Punjab Group of Colleges PGC Lahore
19	Govt College of Science Wahdat Road Lahore
...	...
1455	Little Oxford College of Commerce
1476	Stars Group of Colleges Shahdara 1523
	Govt Inter College Boys Bahtar Fateh Jang Attock

1560 Govt Inter College Narrah Kanjoor Chhab Jand A... 1588
Govt Institute Of Commerce Pindigheb Attock

	Location \		
0	Office 206, 2nd Floor Siddique Trade Center Ma...		
2	GC University, Katchery Road Lahore 54000		
3	Kotwali Rd, Faisalabad 38000		
12	1 Noon Avenue, New Muslim Town at Canal Bank L...		
19	Govt. College of Science, Wahdat Road Lahore,		
...	...		
1455	154/1, Bank Road, Rawalpindi Cantt,		
1476	35-A GT Road Shahdara Lahore. 1523		
Govt Inter College Boys Bahtar Fateh Jang Attock 1560	PO: CHHAB		
TEHSIL:JAND ATTOCK.			
1588	Govt Institute Of Commerce Pindigheb Attock		
	Study Program	Sector \	
0	FSc Pre Engineering Fsc Pre Medical Private		
2	FSc Pre Engineering ICS Computer Science FSc P...	Public	
3	FSc Pre-Engineering ICS FSc Pre-Medical Icom F...	Public	
12	FSc Pre Engineering ICS Computer Science FSc P...	Private	
19	FSC pre Engineering ICS FSC pre medical ICOM F...	Public	
...	
1455	FA Icom ICS Computer Science Fsc Pre Engineeri...	Private	1476 ICS
Computer Science FSc Pre Medical FSc Pre E...	Private	1523	FA Fsc Pre-
Engineering FSc Pre-Medical ICS FA ...		Public	
1560	FA FSc Pre Engineering FSc Pre Medical	Public	
1588		DCom	Public
	Affiliation	Rating	
0		ads	0.0
2		HEC	4.0
3		HEC	3.0
12	University of the Punjab Lahore \n BISE Lahore		2.0
19	Board of Intermediate Secondary Education Laho...		3.0
...	
1455	FBISE IslamabadUniversity of Gujrat		0.0
1476	BISE Lahore Board & Punjab University Lahore		0.0
1523	University of punjab		0.0
1560	University of punjab		0.0
1588	Punjab University		0.0

[12] :

[75 rows x 6 columns]

[12] : df['Rating'].dtypes

```
[13] : df.columns
```

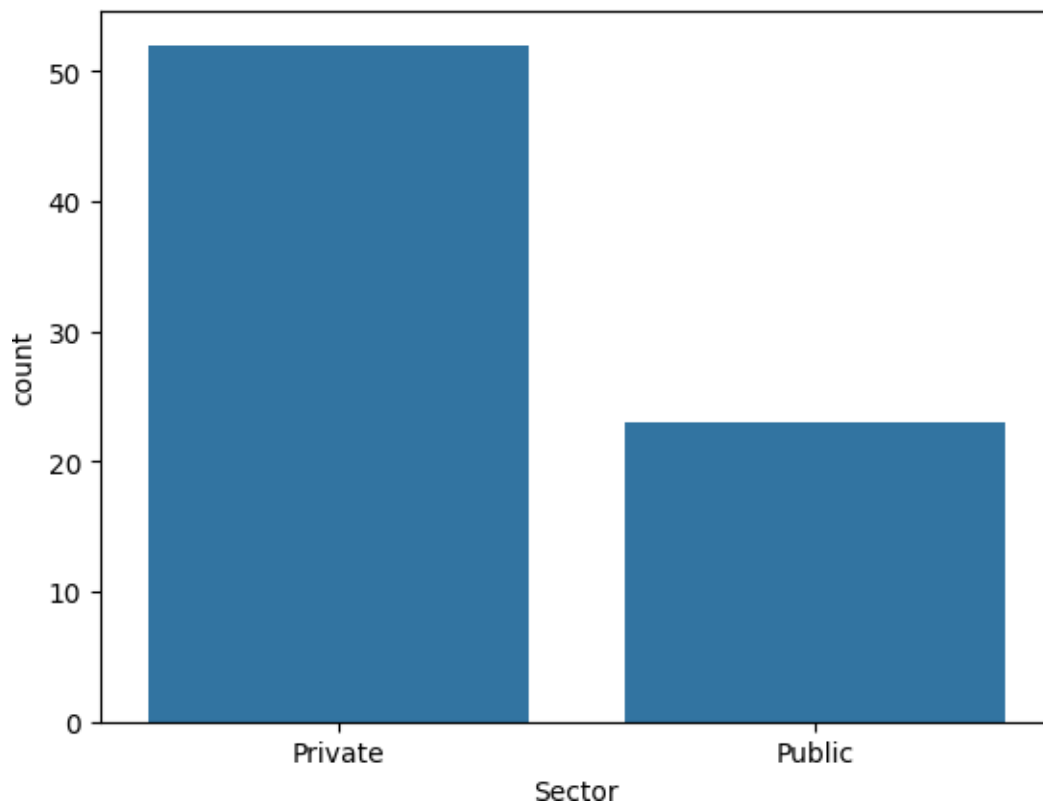
```
[13] : Index(['Name', 'Location', 'Study Program', 'Sector', 'Affiliation', 'Rating'], dtype='object')
```

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

```
[14]:
```

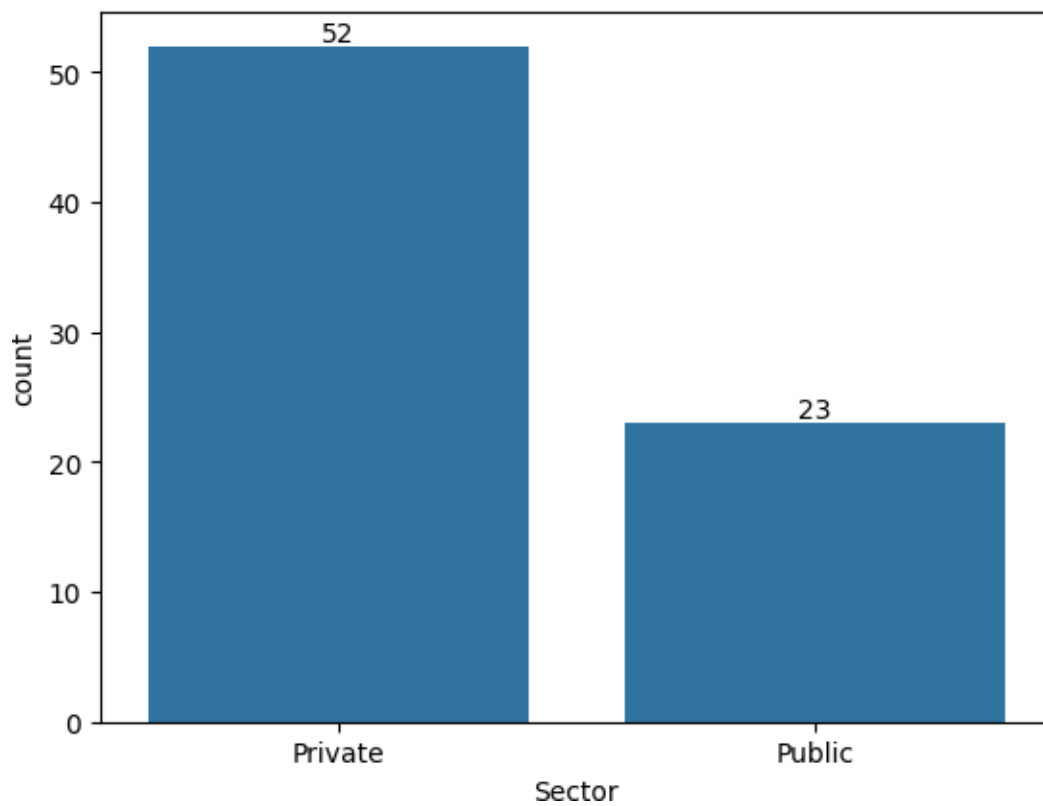
	Rating
count	75.000000
mean	1.516667
std	2.022397
min	0.000000
25%	0.000000
50%	0.000000
75%	3.625000
max	5.000000

```
[15] : ax=sns.countplot(x='Sector',data=df)
```

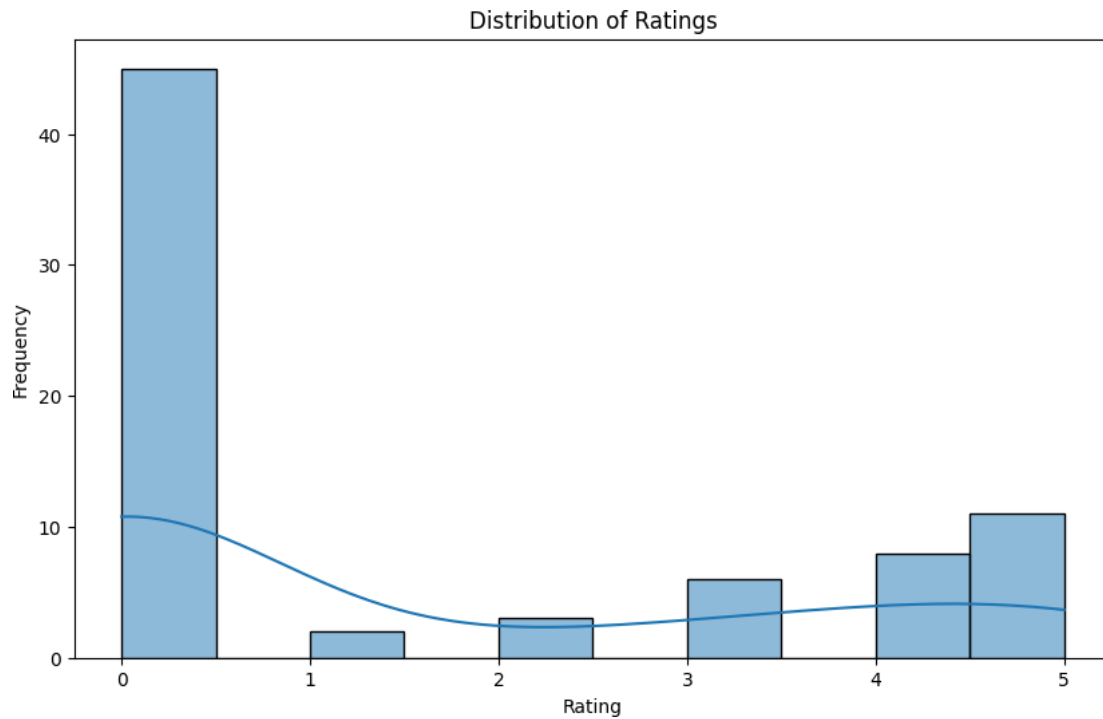


```
[16] : ax=sns.countplot(x='Sector',data=df)
```

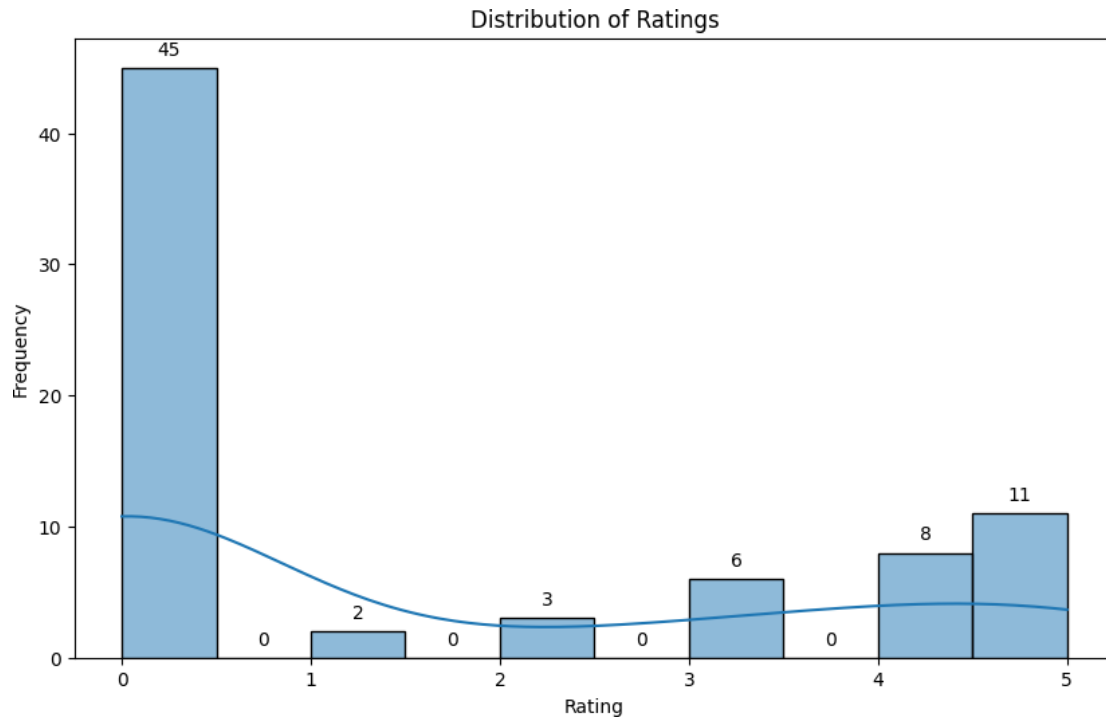
```
for bars in ax.containers: ax.bar_label(bars)
```



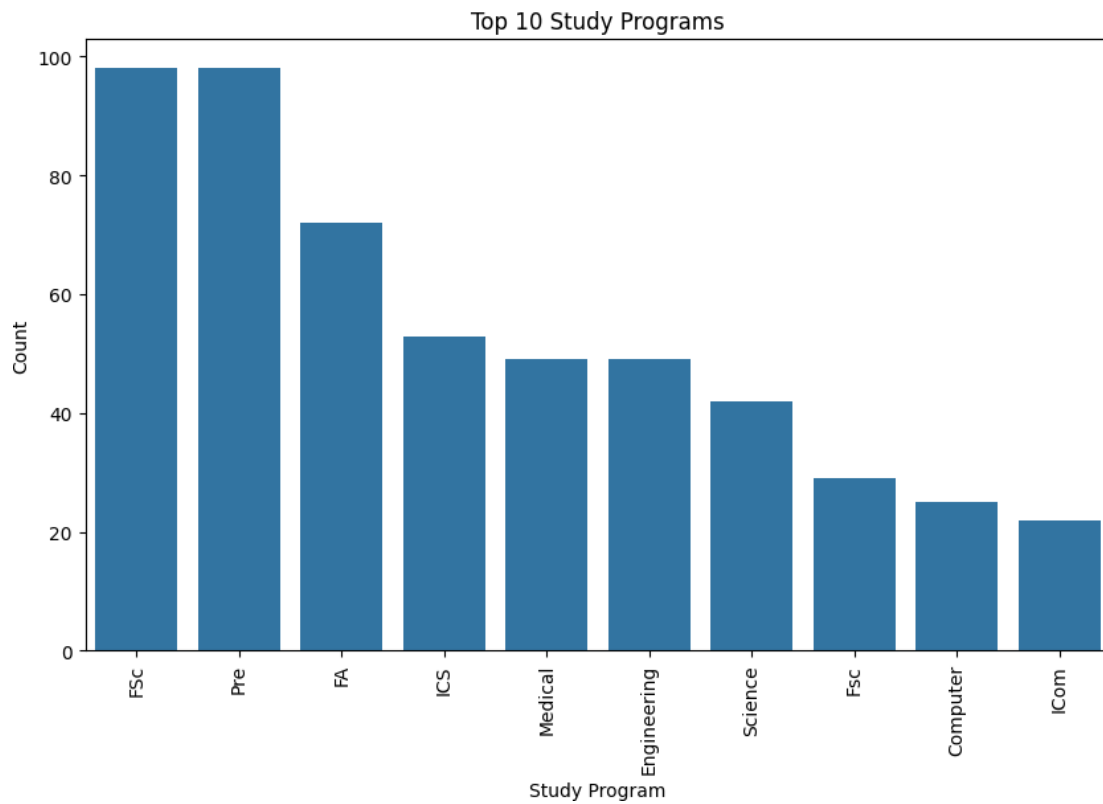
```
[17] : plt.figure(figsize=(10, 6)) sns.histplot(df['Rating'].astype(float), bins=10, kde=True)
plt.title('Distribution of Ratings')
plt.xlabel('Rating')
plt.ylabel('Frequency')
plt.show()
```



```
[19]: plt.figure(figsize=(10,6))
ax = sns.histplot(df['Rating'], bins=10, kde=True)
for p in ax.patches:
    height = p.get_height()
    ax.text(p.get_x() + p.get_width() / 2., height + 1, '{:1.0f}'.
            format(height), ha="center")
plt.title('Distribution of Ratings')
plt.xlabel('Rating')
plt.ylabel('Frequency')
plt.show()
```



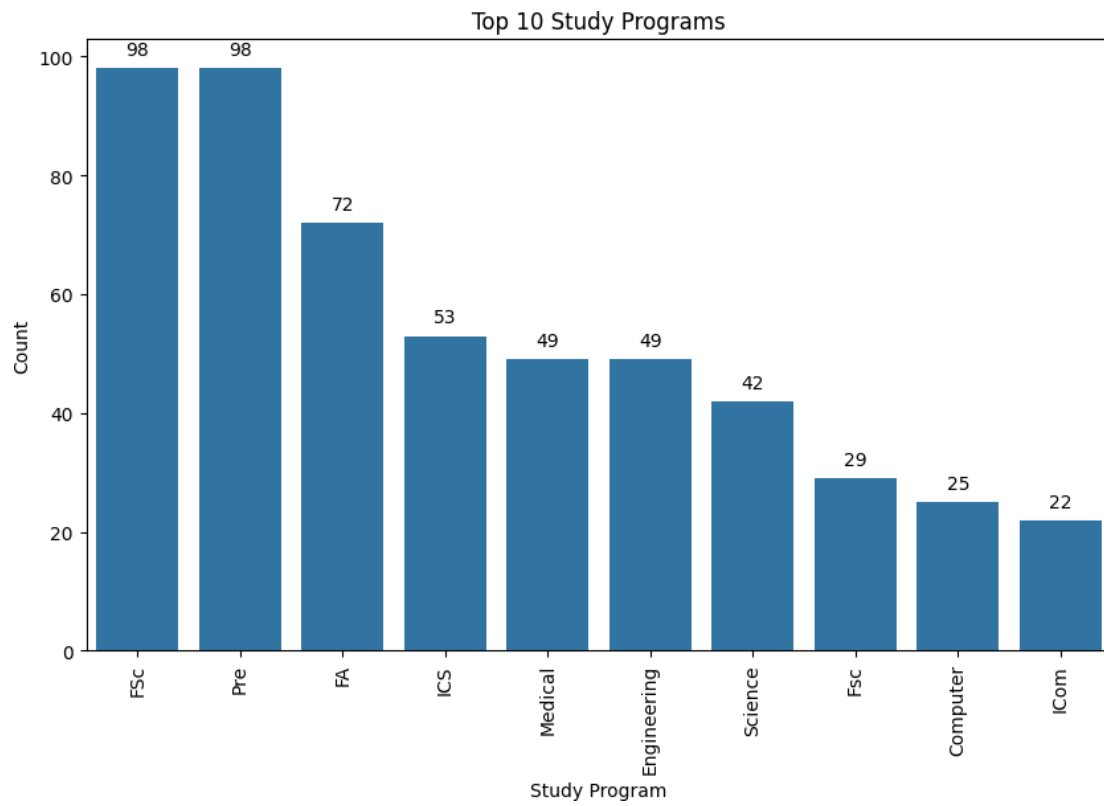
```
[23]: df['Study Programs'] = df['Study Program'].str.split() df_exploded =
df.explode('Study Programs')
top_study_programs = df_exploded['Study Programs'].value_counts().head(10)
plt.figure(figsize=(10, 6))
sns.barplot(x=top_study_programs.index, y=top_study_programs.values) plt.xticks(rotation=90)
plt.title('Top 10 Study Programs')
plt.xlabel('Study Program') plt.ylabel('Count')
plt.show()
```

```
[24]: df['Study Programs'] = df['Study Program'].str.split() df_exploded =
df.explode('Study Programs')
top_study_programs = df_exploded['Study Programs'].value_counts().head(10)

plt.figure(figsize=(10,6))
ax=sns.barplot(x=top_study_programs.index,y=top_study_programs.values) plt.xticks(rotation=90)
plt.title('Top 10 Study Programs')
plt.xlabel('Study Program') plt.ylabel('Count')
for p in ax.patches: ax.annotate(format(p.get_height(), '.0f'),
                                (p.get_x() + p.get_width() / 2., p.get_height()), ha = 'center', va =
                                'center',
                                xytext = (0, 10),
                                textcoords = 'offset points')

plt.show()
```



[: