

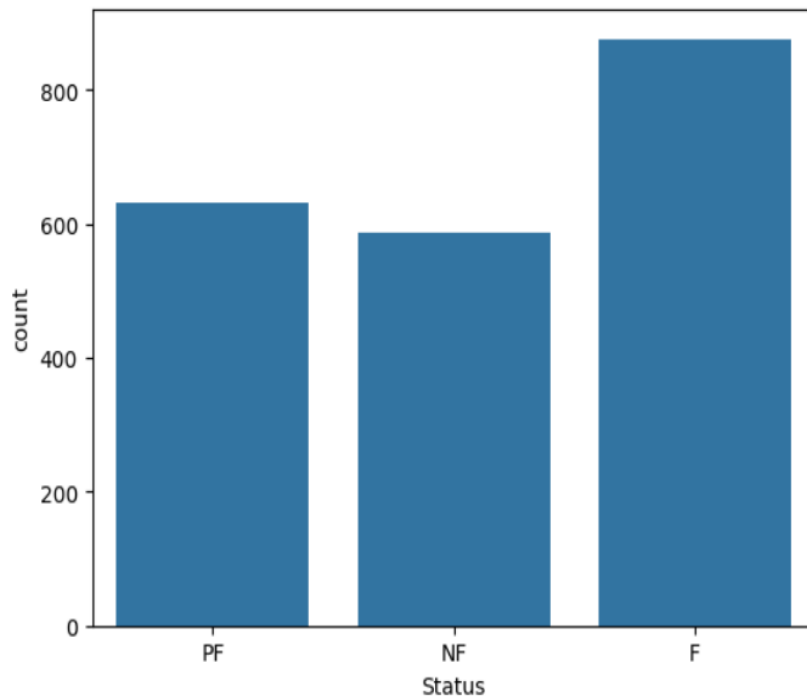
Section	Description												
Data Overview													
		Edition	PR rating	CL rating	A1	A2	A3	A	B1	B2	B3	...	F3
	count	2095.000000	2095.000000	2095.000000	2095.000000	2095.000000	2095.000000	2095.000000	2095.000000	2095.000000	2095.000000	...	2095.000000
	mean	2017.503103	3.575656	3.458234	2.412888	2.523628	2.391885	7.328401	2.599523	2.466826	2.253461	...	2.000477
	std	2.873464	2.211561	1.932492	1.615921	1.527616	1.419146	4.425660	1.401493	1.572119	1.435066	...	1.390547
	min	2013.000000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
	25%	2015.000000	1.000000	2.000000	1.000000	1.000000	1.000000	3.000000	1.000000	1.000000	1.000000	...	1.000000
	50%	2018.000000	3.000000	3.000000	3.000000	3.000000	3.000000	9.000000	3.000000	3.000000	2.000000	...	2.000000
	75%	2020.000000	6.000000	5.000000	4.000000	4.000000	4.000000	12.000000	4.000000	4.000000	4.000000	...	3.000000
	max	2022.000000	7.000000	7.000000	4.000000	4.000000	4.000000	12.000000	4.000000	4.000000	4.000000	...	4.000000
8 rows × 38 columns													

F3	F4	F	G1	G2	G3	G4	G	CL	Total
2095.000000	2095.000000	2095.000000	2095.000000	2095.000000	2095.000000	2095.000000	2095.000000	2095.000000	2095.000000
2.000477	1.850597	7.850119	2.633890	2.364200	2.291169	2.002387	9.291647	34.949403	57.798091
1.390547	1.101843	4.941150	1.254256	1.102466	1.046755	1.057332	4.153617	17.126707	30.288533
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	-1.000000
1.000000	1.000000	4.000000	2.000000	2.000000	2.000000	1.000000	6.000000	20.500000	30.000000
2.000000	2.000000	7.000000	3.000000	2.000000	2.000000	2.000000	9.000000	36.000000	62.000000
3.000000	3.000000	12.000000	4.000000	3.000000	3.000000	3.000000	13.000000	51.000000	87.000000
4.000000	4.000000	16.000000	4.000000	4.000000	4.000000	4.000000	16.000000	60.000000	100.000000

Univariate Analysis

```
sns.countplot(data=data,x='Status')
```

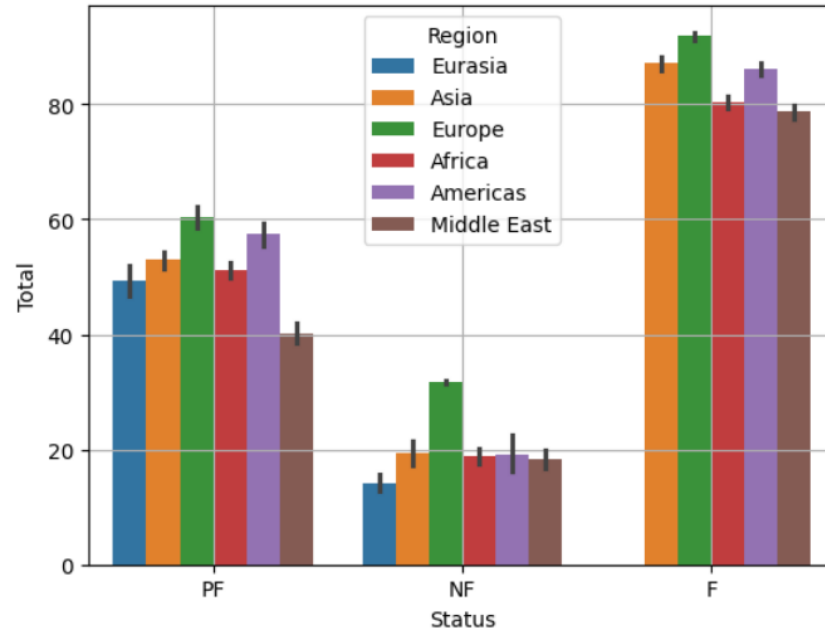
<Axes: xlabel='Status', ylabel='count'>



Bivariate Analysis

```
import seaborn as sns
import matplotlib.pyplot as plt

sns.barplot(data=data, x='Status', y='Total', hue='Region')
plt.grid(True)
plt.show()
```

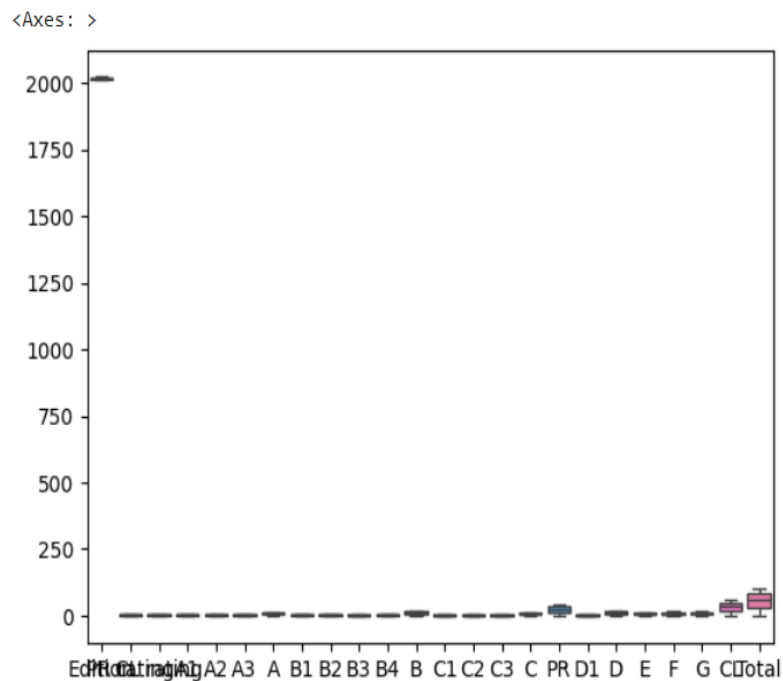


Multivariate Analysis

```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd # Import pandas for data manipulation

# Assuming 'Country/Territory' is the column with non-numerical values
numerical_data = data.select_dtypes(include=['number']) # Select only numerical columns

sns.heatmap(numerical_data.corr(), annot=True)
plt.show()
```



Data Preprocessing Code Screenshots

Loading Data

```
data=pd.read_csv('/content/Freedom in the World 2013-2022 Dataset (Ver 2.18.23).csv')
```

```
data.head()
```

	Country/Territory	Region	C/T	Edition	Status	PR rating	CL rating	A1	A2	A3	...	F3	F4	F	G1	G2	G3	G4	G	CL	Total
0	Abkhazia	Eurasia	t	2022	PF	5	5	2	2	1	...	1	1	4	1	1	2	1	5	23	40
1	Afghanistan	Asia	c	2022	NF	7	6	0	0	0	...	0	0	0	0	1	0	1	2	9	10
2	Albania	Europe	c	2022	PF	3	3	3	3	3	...	2	3	9	3	2	2	2	9	39	67
3	Algeria	Africa	c	2022	NF	6	5	1	1	1	...	2	2	6	2	2	2	1	7	22	32
4	Andorra	Europe	c	2022	F	1	1	4	4	4	...	4	3	15	4	4	3	4	15	55	93

5 rows × 42 columns

```
data.tail()
```

	Country/Territory	Region	C/T	Edition	Status	PR rating	CL rating	A1	A2	A3	...	F3	F4	F	G1	G2	G3	G4	G	CL	Total
2090	West Bank	Middle East	t	2013	NF	6	5	0	0	2	...	0	1	5	1	1	2	1	5	24	30
2091	Western Sahara	Africa	t	2013	NF	7	7	0	0	0	...	0	0	0	1	1	2	0	4	7	5
2092	Yemen	Middle East	c	2013	NF	6	6	1	0	2	...	0	1	2	2	1	1	1	5	16	25
2093	Zambia	Africa	c	2013	PF	3	4	3	3	3	...	2	2	8	3	2	1	2	8	34	62
2094	Zimbabwe	Africa	c	2013	NF	6	6	1	1	1	...	0	0	1	1	1	1	1	4	14	25

5 rows × 42 columns

Handling
Missing
values

```
[ ] test_data.isna().sum()
```

```
no_of_adults      0
no_of_children    0
no_of_weekend_nights  0
no_of_week_nights  0
type_of_meal_plan  0
required_car_parking_space  0
room_type_reserved  0
lead_time         0
arrival_year      0
arrival_month     0
arrival_date      0
market_segment_type  0
repeated_guest    0
no_of_previous_cancellations  0
no_of_previous_bookings_not_canceled  0
avg_price_per_room  0
no_of_special_requests  0
dtype: int64
```

```
[ ] train_data.isna().sum()
```

```
no_of_adults      0
no_of_children    0
no_of_weekend_nights  0
no_of_week_nights  0
type_of_meal_plan  0
required_car_parking_space  0
room_type_reserved  0
lead_time         0
arrival_year      0
arrival_month     0
arrival_date      0
market_segment_type  0
repeated_guest    0
no_of_previous_cancellations  0
no_of_previous_bookings_not_canceled  0
avg_price_per_room  0
no_of_special_requests  0
booking_status    0
dtype: int64
```

Save
Processed
Data

```
import joblib  
joblib.dump(knn, 'model.pkl')  
  
['model.pkl']
```