Prodigy Infotech Internship Task 1

Name : Pavan yadav

Task : Create a bar plot or histogram to visualize the distribution of a categorical or continous variable

In [21]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [27]:
```
df = pd.read_csv('metadata.csv')
df
```

Out[27]:

	Country Code	Region	IncomeGroup	SpecialNotes	TableName	Unnamed: 5
0	ABW	Latin America & Caribbean	High income	NaN	Aruba	NaN
1	AFE	NaN	NaN	26 countries, stretching from the Red Sea in t...	Africa Eastern and Southern	NaN
2	AFG	South Asia	Low income	The reporting period for national accounts dat...	Afghanistan	NaN
3	AFW	NaN	NaN	22 countries, stretching from the westernmost ...	Africa Western and Central	NaN
4	AGO	Sub-Saharan Africa	Lower middle income	The World Bank systematically assesses the app...	Angola	NaN
...	...	...	...	...	...	...
260	XKX	Europe & Central Asia	Upper middle income	NaN	Kosovo	NaN
261	YEM	Middle East & North Africa	Low income	The World Bank systematically assesses the app...	Yemen, Rep.	NaN
262	ZAF	Sub-Saharan Africa	Upper middle income	Fiscal year end: March 31; reporting period fo...	South Africa	NaN
263	ZMB	Sub-Saharan Africa	Lower middle income	National accounts data were rebased to reflect...	Zambia	NaN
264	ZWE	Sub-Saharan Africa	Lower middle income	National Accounts data are reported in Zimbabw...	Zimbabwe	NaN

265 rows × 6 columns

Out[23]:

	Country	Unnamed:

In [23]:
```
df.head()
```

	Code	Region	IncomeGroup	SpecialNotes	TableName	5
0	ABW	Latin America & Caribbean	High income	NaN	Aruba	NaN
1	AFE	NaN	NaN	26 countries, stretching from the Red Sea in t...	Africa Eastern and Southern	NaN
2	AFG	South Asia	Low income	The reporting period for national accounts dat...	Afghanistan	NaN
3	AFW	NaN	NaN	22 countries, stretching from the westernmost ...	Africa Western and Central	NaN
4	AGO	Sub-Saharan Africa	Lower middle income	The World Bank systematically assesses the app...	Angola	NaN

In [24]:
```
df.tail()
```

Out[24]:

	Country Code	Region	IncomeGroup	SpecialNotes	TableName	Unnamed: 5
260	XKX	Europe & Central Asia	Upper middle income	NaN	Kosovo	NaN
261	YEM	Middle East & North Africa	Low income	The World Bank systematically assesses the app...	Yemen, Rep.	NaN
262	ZAF	Sub-Saharan Africa	Upper middle income	Fiscal year end: March 31; reporting period fo...	South Africa	NaN
263	ZMB	Sub-Saharan Africa	Lower middle income	National accounts data were rebased to reflect...	Zambia	NaN
264	ZWE	Sub-Saharan Africa	Lower middle income	National Accounts data are reported in Zimbabw...	Zimbabwe	NaN

In [25]:
```
##Count the apperance of each region
region_counts = df['Region'].value_counts()
region_counts
```

Out[25]:
```
Europe & Central Asia         58
Sub-Saharan Africa            48
Latin America & Caribbean     42
East Asia & Pacific           37
Middle East & North Africa    21
South Asia                     8
North America                  3
Name: Region, dtype: int64
```

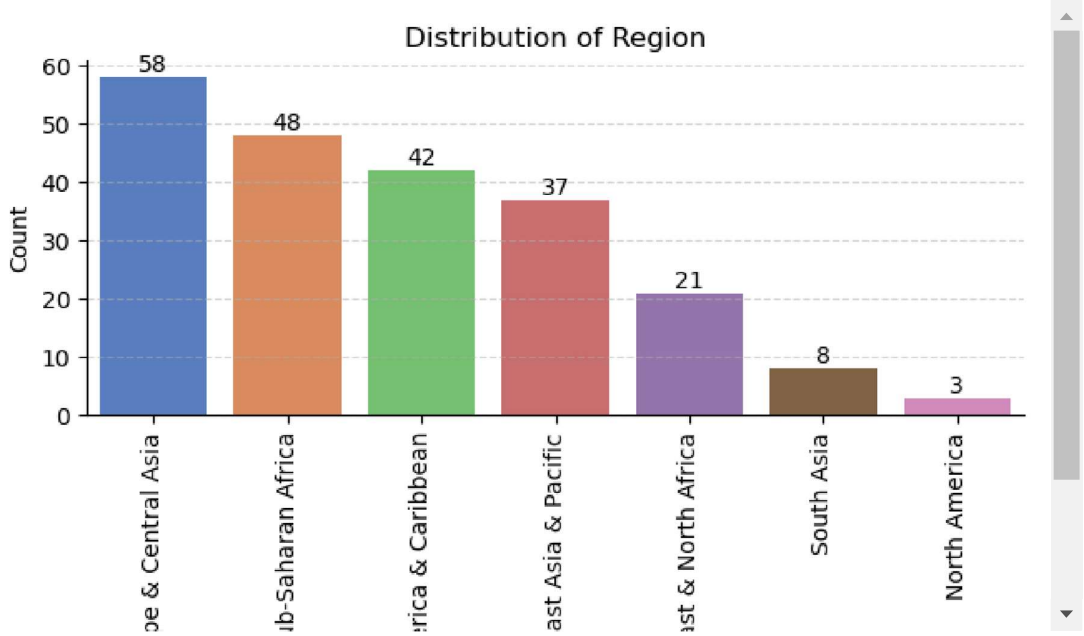
In [26]:
```
sns.barplot(x = region_counts.index, y = region_counts.values , palette = 'm
plt.xlabel('Region')
plt.ylabel('Count')
plt.title('Distribution of Region')
plt.xticks(rotation=90)

#add
for x,y in enumerate(region_counts.values):
  plt.text(x,  y, str(y),ha = 'center' , va = 'bottom')

#remove the top and right spines
plt.gca().spines['top'].set_visible(False)
plt.gca().spines['right'].set_visible(False)

#adding grid to the plot
plt.grid(axis = 'y' , linestyle = '--' , alpha = 0.5)

#display the chart
plt.tight_layout()
plt.show()
```



Thank You