In [2]:
```
import pandas as pd
```

In [3]:
```
df = pd.read_csv(r"C:\Users\JAGADISH\Downloads\Crop Production data.csv")
df
```

Out[3]:

	Unnamed: 0	State_Name	District_Name	Crop_Year	Season	Crop	Are
0	0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254
1	1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2
2	2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102
3	3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176
4	4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720
...	...	...	...	...	...	...	
246086	246086	West Bengal	PURULIA	2014	Summer	Rice	306
246087	246087	West Bengal	PURULIA	2014	Summer	Sesamum	627
246088	246088	West Bengal	PURULIA	2014	Whole Year	Sugarcane	324
246089	246089	West Bengal	PURULIA	2014	Winter	Rice	279151
246090	246090	West Bengal	PURULIA	2014	Winter	Sesamum	175

246091 rows × 8 columns

In [4]:
```
# Fill missing values in the 'Production' column with the median value
df['Production'].fillna(df['Production'].median(), inplace=True)
```

In [5]:
```
# Remove duplicate rows
df_cleaned = df.drop_duplicates()
```

In [6]:
```
# Verify the cleaned df
print(df_cleaned.info())
print(df_cleaned.head())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 246091 entries, 0 to 246090
Data columns (total 8 columns):
 #   Column         Non-Null Count   Dtype  
---  ------         --------------   -----  
 0   Unnamed: 0     246091 non-null  int64  
 1   State_Name     246091 non-null  object 
 2   District_Name  246091 non-null  object 
 3   Crop_Year      246091 non-null  int64  
 4   Season         246091 non-null  object 
 5   Crop           246091 non-null  object 
 6   Area           246091 non-null  float64
 7   Production     246091 non-null  float64
dtypes: float64(2), int64(2), object(4)
memory usage: 15.0+ MB
None
```

	Unnamed: 0	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254.0	2000.0
1	1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2.0	1.0
2	2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102.0	321.0
3	3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176.0	641.0
4	4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720.0	165.0

In [7]:
```
df_cleaned.to_csv('crop production data.csv')
df_cleaned
```

Out[7]:

	Unnamed: 0	State_Name	District_Name	Crop_Year	Season	Crop	Are
0	0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254
1	1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2
2	2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102
3	3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176
4	4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720
...	...	...	...	...	...	...	
246086	246086	West Bengal	PURULIA	2014	Summer	Rice	306
246087	246087	West Bengal	PURULIA	2014	Summer	Sesamum	627
246088	246088	West Bengal	PURULIA	2014	Whole Year	Sugarcane	324
246089	246089	West Bengal	PURULIA	2014	Winter	Rice	279151
246090	246090	West Bengal	PURULIA	2014	Winter	Sesamum	175

246091 rows × 8 columns

In [8]:
```
print(df_cleaned.columns)
```

```
Index(['Unnamed: 0', 'State_Name', 'District_Name', 'Crop_Year', 'Season',
       'Crop', 'Area', 'Production'],
      dtype='object')
```

In [ ]: