8/17/24, 5:20 PM crop production

```
import pandas as pd
In [2]:
        df = pd.read_csv(r"C:\Users\JAGADISH\Downloads\Crop Production data.csv")
In [3]:
Out[3]:
                 Unnamed:
                            Season
                                                                                Crop
                                                                                          Ar€
                               Andaman
              0
                             and Nicobar
                                                            2000
                                                                    Kharif
                                             NICOBARS
                                                                             Arecanut
                                                                                        1254
                                 Islands
                               Andaman
                                                                                Other
              1
                             and Nicobar
                                                                                            2
                                             NICOBARS
                                                            2000
                                                                    Kharif
                                                                               Kharif
                                 Islands
                                                                               pulses
                               Andaman
              2
                             and Nicobar
                                             NICOBARS
                                                            2000
                                                                    Kharif
                                                                                 Rice
                                                                                         102
                                 Islands
                               Andaman
                                                                    Whole
              3
                             and Nicobar
                                             NICOBARS
                                                            2000
                                                                              Banana
                                                                                         176
                                                                      Year
                                 Islands
                               Andaman
                                                                    Whole
              4
                             and Nicobar
                                             NICOBARS
                                                            2000
                                                                           Cashewnut
                                                                                         720
                                                                      Year
                                 Islands
         246086
                    246086
                             West Bengal
                                              PURULIA
                                                            2014
                                                                  Summer
                                                                                 Rice
                                                                                         306
         246087
                    246087
                                                                  Summer
                                                                            Sesamum
                             West Bengal
                                              PURULIA
                                                            2014
                                                                                         627
                                                                    Whole
                             West Bengal
         246088
                    246088
                                              PURULIA
                                                            2014
                                                                            Sugarcane
                                                                                         324
                                                                      Year
         246089
                                                            2014
                                                                    Winter
                                                                                      279151
                    246089
                             West Bengal
                                              PURULIA
                                                                                 Rice
                            West Bengal
         246090
                    246090
                                              PURULIA
                                                            2014
                                                                    Winter
                                                                                         175
                                                                            Sesamum
        246091 rows × 8 columns
        # Fill missing values in the 'Production' column with the median value
In [4]:
        df['Production'].fillna(df['Production'].median(), inplace=True)
        # Remove duplicate rows
In [5]:
        df_cleaned = df.drop_duplicates()
In [6]: # Verify the cleaned df
         print(df_cleaned.info())
        print(df_cleaned.head())
```

8/17/24, 5:20 PM crop production

```
<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
                  246091 non-null int64
3 Crop_Year
4
   Season
                  246091 non-null object
5
    Crop
                  246091 non-null object
                  246091 non-null float64
6
    Area
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 \
          O Andaman and Nicobar Islands
                                             NICOBARS
           1 Andaman and Nicobar Islands
                                             NICOBARS
2
           2 Andaman and Nicobar Islands
                                             NICOBARS
           3 Andaman and Nicobar Islands
                                             NICOBARS
           4 Andaman and Nicobar Islands
                                             NICOBARS
```

```
Season
                           Crop
                                Area Production
0 Kharif
                       Arecanut 1254.0
                                          2000.0
1 Kharif
              Other Kharif pulses 2.0
                                             1.0
2 Kharif
                           Rice 102.0
                                            321.0
3 Whole Year
                         Banana 176.0
                                            641.0
4 Whole Year
                      Cashewnut 720.0
                                            165.0
```

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

2000

2000

2000

2000

2000

8/17/24, 5:20 PM crop production

Out[7]: **Unnamed:** State_Name District_Name Crop_Year Season Crop Are Andaman 0 and Nicobar 2000 0 **NICOBARS** Kharif Arecanut 1254 Islands Andaman Other 2 1 and Nicobar 2000 Kharif 1 **NICOBARS** Kharif Islands pulses Andaman 2 and Nicobar 2000 **NICOBARS** Kharif Rice 102 Islands Andaman Whole 3 and Nicobar **NICOBARS** 2000 Banana 176 Year Islands Andaman Whole 4 and Nicobar **NICOBARS** 2000 Cashewnut 720 Year Islands 246086 246086 West Bengal **PURULIA** 2014 Summer Rice 306 246087 246087 **PURULIA** 2014 Sesamum 627 West Bengal Summer Whole 246088 2014 246088 West Bengal **PURULIA** Sugarcane 324 Year 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 []: