Data Analyst Project using python on crop production in india

Problem Statement:

The Agriculture business domain, as a vital part of the overall supply chain, is expected to highly evolve in the upcoming years via the developments, which are taking place on the side of the Future Internet. This paper presents a novel Business-to-Business collaboration platform from the agri-food sector perspective, which aims to facilitate the collaboration of numerous stakeholders belonging to associated business domains, in an effective and flexible manner. This dataset provides a huge amount of information on crop production in India ranging from several years. Based on the Information the ultimate goal would be to predict crop production and find important insights highlighting key indicators and metrics that influence crop production. Make views and dashboards first and also make a story out of it.

Now let's start this task by importing the necessary Python libraries and the dataset:

```
In [44]: import pandas as pd
         import numpy as np
         import seaborn as sns
         import matplotlib.pyplot as plt
         %matplotlib inline
         #we have to show charts in jupyter notebook only
```

Loading the dataset:

```
In [45]: dataset=pd.read_csv(r"C:\Users\saini\OneDrive\Desktop\Crop Production data.csv")
```

Head View of the data:

```
In [46]: dataset.head(3)
```

Out[46]:

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254.0	2000.0
1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2.0	1.0
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102.0	321.0

Tail View of the data:

Out[47]:

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
246088	West Bengal	PURULIA	2014	Whole Year	Sugarcane	324.0	16250.0
246089	West Bengal	PURULIA	2014	Winter	Rice	279151.0	597899.0
246090	West Bengal	PURULIA	2014	Winter	Sesamum	175.0	88.0

Columns Details:

```
In [48]: dataset.columns
Out[48]: Index(['State_Name', 'District_Name', 'Crop_Year', 'Season', 'Crop', 'Area',
                'Production'],
               dtype='object')
```

Information about the data:

```
In [49]: dataset.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 246091 entries, 0 to 246090
Data columns (total 7 columns):
 #   Column         Non-Null Count   Dtype
---  ------         --------------   -----
 0   State_Name     246091 non-null  object
 1   District_Name  246091 non-null  object
 2   Crop_Year      246091 non-null  int64
 3   Season         246091 non-null  object
 4   Crop           246091 non-null  object
 5   Area           246091 non-null  float64
 6   Production     242361 non-null  float64
dtypes: float64(2), int64(1), object(4)
memory usage: 13.1+ MB
```

```
In [50]: dataset.describe()
```

Out[50]:

	Crop_Year	Area	Production
count	246091.000000	2.460910e+05	2.423610e+05
mean	2005.643018	1.200282e+04	5.825034e+05
std	4.952164	5.052340e+04	1.706581e+07
min	1997.000000	4.000000e-02	0.000000e+00
25%	2002.000000	8.000000e+01	8.800000e+01
50%	2006.000000	5.820000e+02	7.290000e+02
75%	2010.000000	4.392000e+03	7.023000e+03
max	2015.000000	8.580100e+06	1.250800e+09

Checking whether any columns contain null values:

```
In [51]: dataset.isnull().sum()

Out[51]: State_Name          0
         District_Name       0
         Crop_Year           0
         Season              0
         Crop                0
         Area                0
         Production       3730
         dtype: int64
```

Removing the null values:

```
In [52]: dataset.dropna(inplace=True)
```

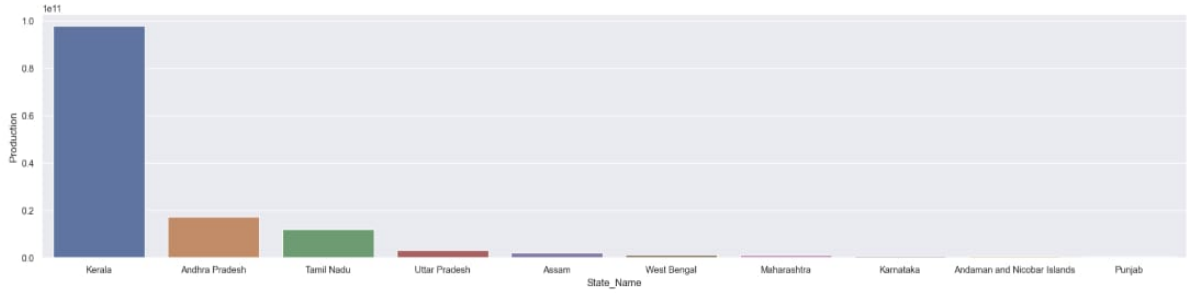
<u>Visualization:</u>

EXPLORATORY DATA ANALYSIS

```
In [54]: taset= dataset.groupby(['State_Name'], as_index=False)['Production'].sum().sort_values(by='Production', ascending=False).head(10)

         s.set(rc={'figure.figsize':(25,5)})
         s.barplot(data = dataset, x = 'State_Name',y= 'Production')
```

Out[54]: <Axes: xlabel='State_Name', ylabel='Production'>



Over the years, there has been a noticeable increase/decrease in the production of certain crops. For example, the production of rice and wheat has shown an upward trend

NAME: YASH GAUTAM

EMAIL: yashr7143@gmail.com