crop-production-analysis

[1]:
```
# Open the image
from IPython.display import Image
Image(filename="C:/Users/Rajesh/Downloads/crop production.png")
```

[1]:

Introduction

India is a major player in agriculture on the global stage, offers a fascinating landscape for understanding crop prediction, seasonal impacts, regional influences on yield, and the evolving trends in agricultural forecasting.The main objectives of this project include data exploration, visualization, and addressing various analytical questions. So, join me on this exploration of Indian agriculture. Let's go!

What is Crop Production?

Crop production is the process of growing crops for domestic and commercial purposes. Some of the crops produced on a large scale include rice, wheat, maize, jute, etc.

Crop production is a common agricultural practice followed by worldwide farmers to grow and produce crops to use as food and fibre. This practice includes all the feed sources that are required to maintain and produce crops. Listed below are few practices used during crop production.

- Preparation of Soil.
- Sowing of Seeds.
- Irrigation.
- Application of manure, pesticides, and fertilizers to the crops.
- Protecting and Harvesting Crops.
- Storage and Preserving the produced Crops.

About The Data:

This dataset, holds a wealth of valuable information sourced from the Indian government's Area Production Statistics (APS) database. Maintained by the Ministry of Agriculture and Farmers Welfare, the APS database offers comprehensive and detailed data on crop production, yield, and cultivated areas across various states and districts in India.

Columns in dataset

State_Name

District_Name

Crop_Year

Season

Crop

Area

Production

Importing Libraries and Datasets The libraries used are :

Pandas: This library helps to load the data frame in a 2D array format and has multiple functions to perform analysis tasks in one go.

Seaborn/Matplotlib: For data visualization.

plotly.express: Plotly Express provides functions to visualize a variety of types of data.

Numpy: Numpy arrays are very fast and can perform large computations in a very short time.

[2]:
```
# import library
import pandas as pd
```

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import warnings
warnings.filterwarnings("ignore")
```

[3]:
```
# Read Attrition data .csv file and print first 5 records
df = pd.read_csv("Crop Production data.csv")
df.head()
```

[3]:
```
                    State_Name District_Name  Crop_Year       Season  \
0  Andaman and Nicobar Islands      NICOBARS       2000  Kharif
1  Andaman and Nicobar Islands      NICOBARS       2000  Kharif
2  Andaman and Nicobar Islands      NICOBARS       2000  Kharif
3  Andaman and Nicobar Islands      NICOBARS       2000  Whole Year
4  Andaman and Nicobar Islands      NICOBARS       2000  Whole Year

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

1.3 Check total number of columns

[4]:
```
print(f"Total colums-",len(df.columns))
print(f"Total entries-",df.size)
```

```
Total colums- 7
Total entries- 1722637
```

Total colums- 7 and total entries are 1722637

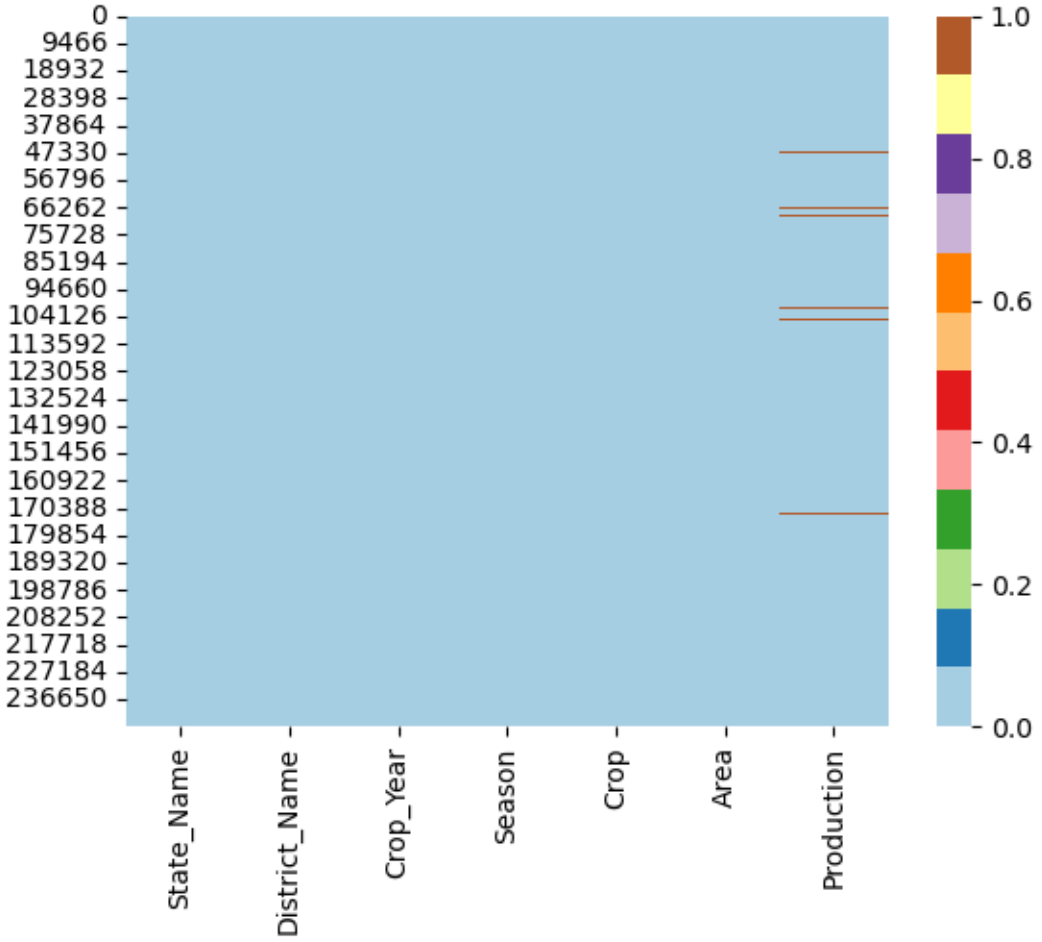
[5]:
```
# df. shape method provides information about the number of rows and columns in␣
 ↪a DataFrame quickly and easily.
df.shape
```

[5]:
```
(246091, 7)
```

EDA (Exploratory Data Analysis)

Checking for missing values

[6]:
```
sns.heatmap(df.isnull(),cmap="Paired")
plt.show()
```



Find null values

[7]:
```
# Now, let's have a look at whether this dataset has any null values or not
print(df.isnull().sum())
print()
print()
print(df.isna().sum())
```

```
State_Name          0
District_Name       0
```

```
Crop_Year           0
Season              0
Crop                0
Area                0
Production       3730
dtype: int64


State_Name          0
District_Name       0
Crop_Year           0
Season              0
Crop                0
Area                0
Production       3730
dtype: int64
```

[8]:
```
print(df.isnull().sum().sum())
print()
print(df.isna().sum().sum())
```

```
3730

3730
```

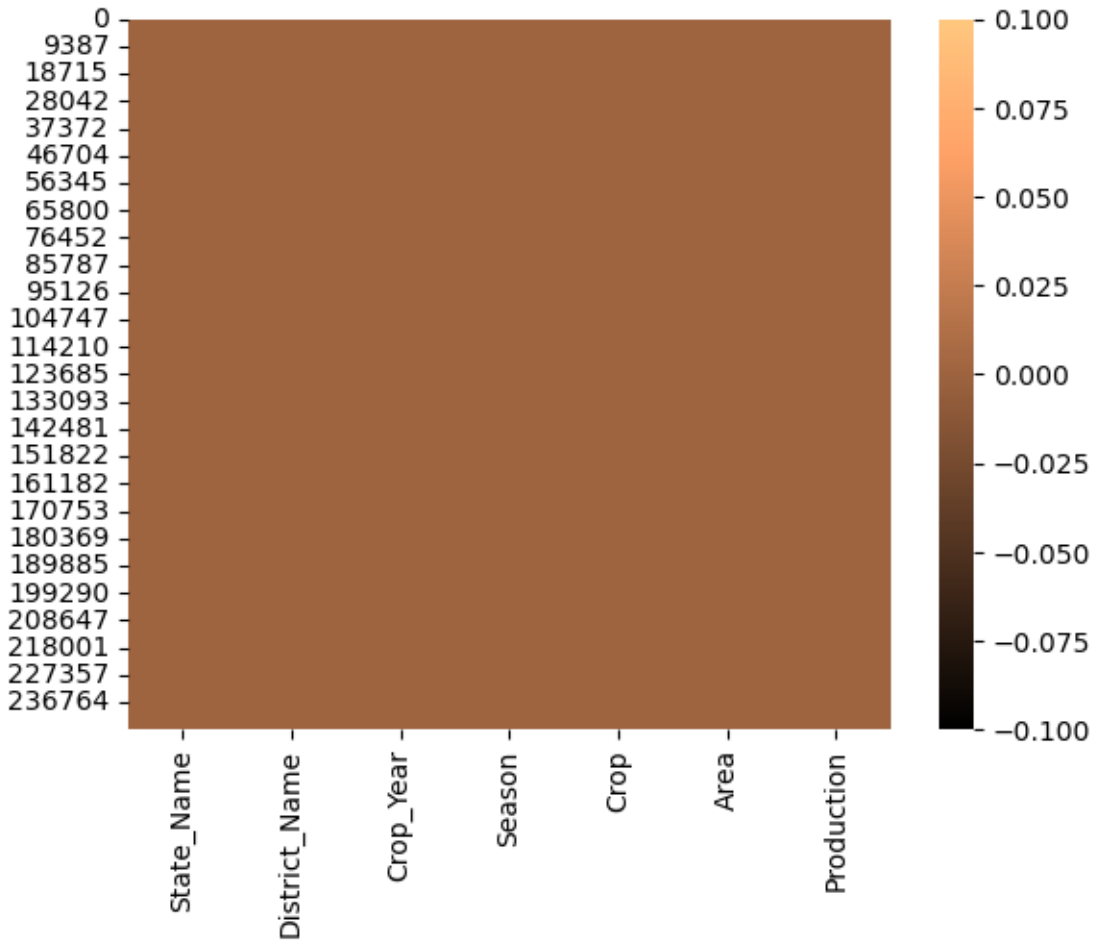
Null values are very few, we can drop them without affecting data set

[9]:
```
#Drop null values
df=df.dropna(axis=0)
```

[10]:
```
# df. shape method provides information about the number of rows and columns in␣
 ↪a DataFrame quickly and easily.
df.shape
```

[10]:
```
(242361, 7)
```

[11]:
```
sns.heatmap(df.isnull(),cmap="copper")
plt.show()
```



To print the information of the data we can use data.info() command.

[12]:
```
# view the data types and missing values in each column
print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 242361 entries, 0 to 246090
Data columns (total 7 columns):
 #   Column         Non-Null Count   Dtype
---  ------         --------------   -----
 0   State_Name     242361 non-null  object
 1   District_Name  242361 non-null  object
 2   Crop_Year      242361 non-null  int64
 3   Season         242361 non-null  object
 4   Crop           242361 non-null  object
```

```
 5   Area           242361 non-null  float64
 6   Production     242361 non-null  float64
dtypes: float64(2), int64(1), object(4)
memory usage: 14.8+ MB
None
```

Let's see the mean, count , minimum and maximum values of the data

[13]:
```
#The describe() method returns description of the data in the DataFrame.
# view the summary statistics for each column
df.describe().style.background_gradient(cmap='binary')
```

[13]: <pandas.io.formats.style.Styler at 0x252441f0340>

[14]:
```
# Now, let's have a look at whether this dataset has any null values or not
print(df.isnull().sum())
print()
print()
print(df.isna().sum())
```

```
State_Name       0
District_Name    0
Crop_Year        0
Season           0
Crop             0
Area             0
Production       0
dtype: int64


State_Name       0
District_Name    0
Crop_Year        0
Season           0
Crop             0
Area             0
Production       0
dtype: int64
```

Data Visualization

In this section, we will try to understand and compare all columns.

Let's count the columns with different datatypes like Category, Integer, Float.

Crop Production Trends Over Years

[15]:
```
df.dtypes
```

[15]:
```
State_Name        object
District_Name     object
Crop_Year          int64
Season            object
Crop              object
Area             float64
Production       float64
dtype: object
```

[16]:
```
print(f"Number of categorical columns:", len(df.select_dtypes(include='object').
 ↪columns))
print(f"Number of integer columns:", len(df.select_dtypes(include='int').
 ↪columns))
print(f"Number of float columns:", len(df.select_dtypes(include='float').
 ↪columns))
```

```
Number of categorical columns: 4
Number of integer columns: 1
Number of float columns: 2
```

[17]:
```
# Exploring Crop type
df.Crop.value_counts()
```

[17]:
```
Rice                15082
Maize               13787
Moong(Green Gram)   10106
Urad                 9710
Sesamum              8821
                    …
Litchi                  6
Coffee                  6
Apple                   4
Peach                   4
Other Dry Fruit         1
Name: Crop, Length: 124, dtype: int64
```

[18]:
```
df.Season.value_counts()
```

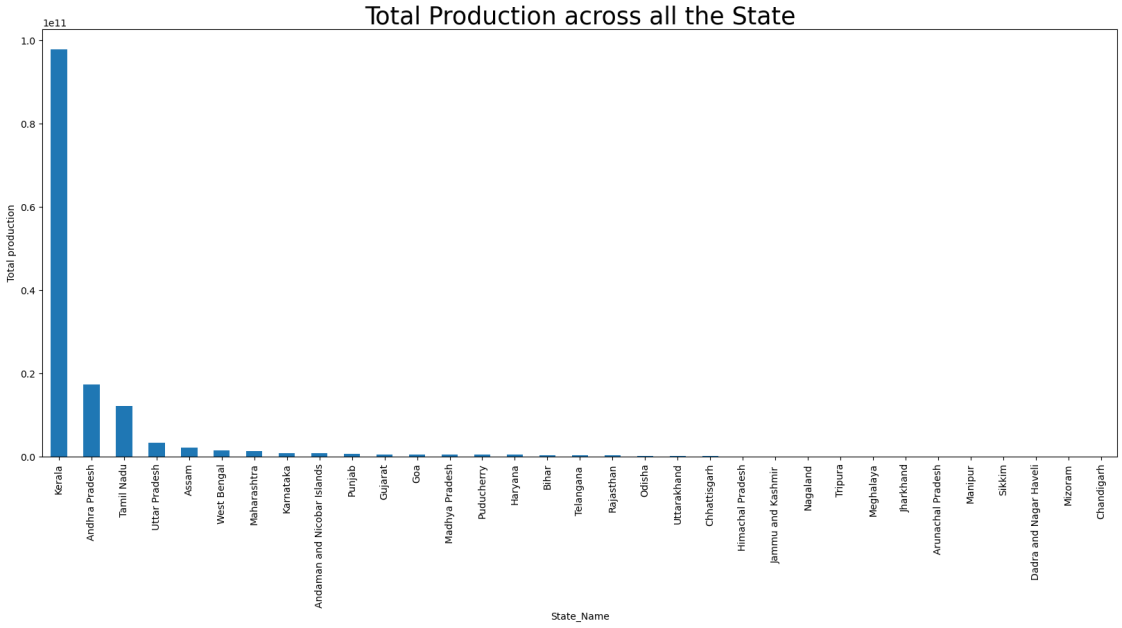
[18]:
```
Kharif         94283
Rabi           66160
Whole Year     56127
Summer         14811
Winter          6050
Autumn          4930
Name: Season, dtype: int64
```

Now we start The Visualization Part

we find out wich factors affecting Crop

Overall Crop Production By state

[19]:
```
plt.figure(figsize=(20,8))
df.groupby('State_Name')["Production"].sum().sort_values(ascending=False).
  ↪plot(kind='bar')
plt.title('Total Production across all the State', size=25)
plt.ylabel('Total production');
```



Kerala is highest crop producing state whereas chandigarh is least and Top 3 crop producing states are from south India

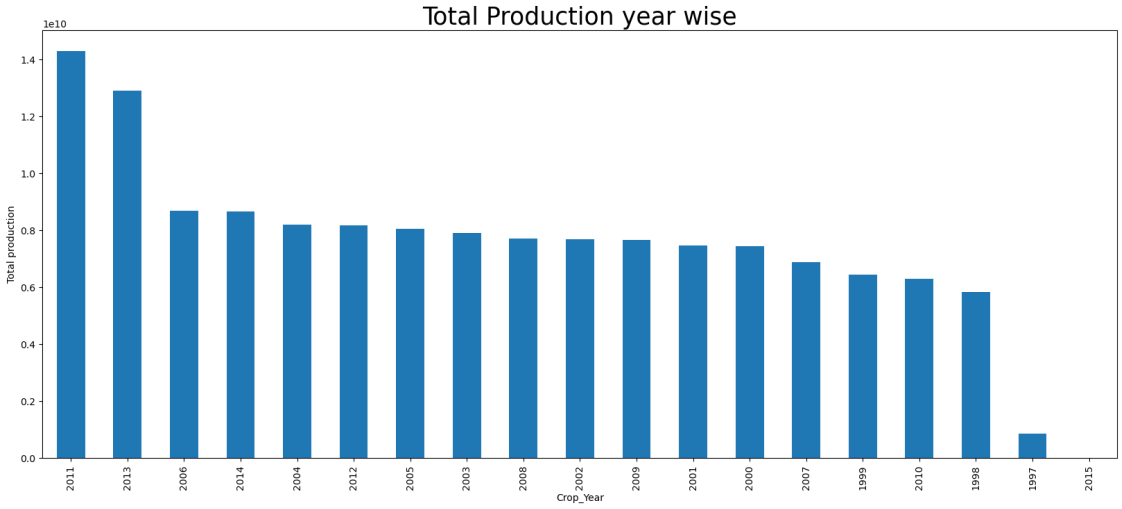
Production per unit area according to state

[20]:
```
temp1 = df.groupby('State_Name')['Area', 'Production'].sum().reset_index()
temp1['Production_Per_Unit_Area'] = temp1['Production']/temp1['Area']
temp1 = temp1.sort_values(by='Production_Per_Unit_Area')
px.bar(temp1, 'State_Name', 'Production_Per_Unit_Area',␣
  ↪color='Production_Per_Unit_Area')
```

Kerala is the most productive state in terms of production by area. Even small states like Andaman and nikobar islands, Goa, Panduchery and many other states which are low in overall production, have high productivity.

Crop Production in India yearly

[21]:
```
plt.figure(figsize=(20,8))
df.groupby('Crop_Year')["Production"].sum().sort_values(ascending=False).
 ↪plot(kind='bar')
plt.title('Total Production year wise', size=25)
plt.ylabel('Total production');
```



Hence 2013 is the most productive year where as 1997 is less productive

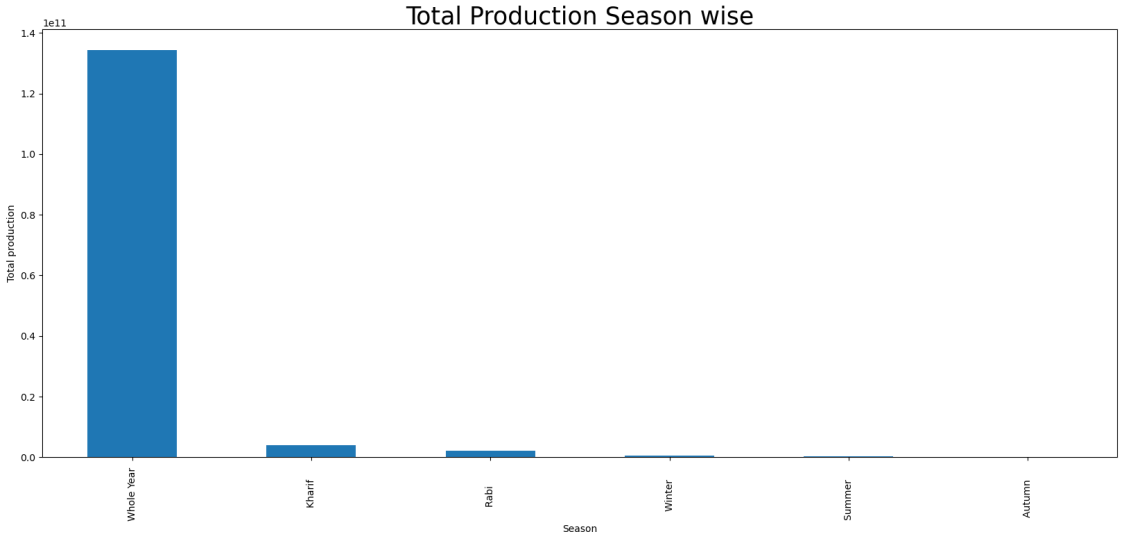
Production per unit area according to state

[22]:
```
temp_crop = df.groupby('Crop_Year')['Area', 'Production'].sum().reset_index()
temp_crop['Production_Per_Unit_Area'] = temp_crop['Production']/
 ↪temp_crop['Area']
temp_crop = temp_crop.sort_values(by='Production_Per_Unit_Area')
px.bar(temp_crop, 'Crop_Year', 'Production_Per_Unit_Area',␣
 ↪color='Production_Per_Unit_Area')
```

yearly production per unit are is getting increased, may be due to occupancy of larger agricultural land across the country

Total Production Season wise

[23]:
```
plt.figure(figsize=(20,8))
df.groupby('Season')["Production"].sum().sort_values(ascending=False).
 ↪plot(kind='bar')
plt.title('Total Production Season wise', size=25)
plt.ylabel('Total production');
```



Kharif is the season where india is growing major crops

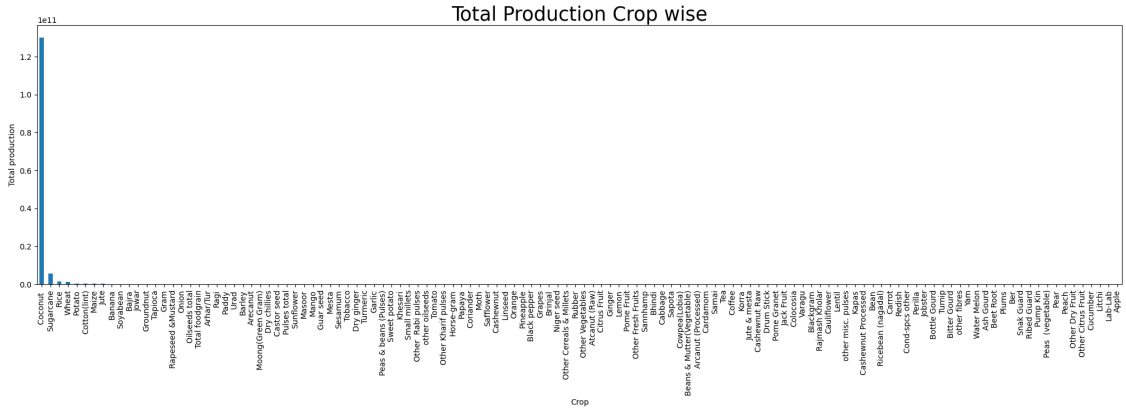
Production per unit area according to Season

[24]:
```
temp_Season = df.groupby('Season')['Area', 'Production'].sum().reset_index()
temp_Season['Production_Per_Unit_Area'] = temp_Season['Production']/
 ↪temp_Season['Area']
temp_Season = temp_Season.sort_values(by='Production_Per_Unit_Area')
px.bar(temp_Season, 'Season', 'Production_Per_Unit_Area',␣
 ↪color='Production_Per_Unit_Area')
```

Production per area is constant through out but little lesser in Autumn.

Total Production Crop wise

[25]:
```
plt.figure(figsize=(25,6))
df.groupby('Crop')["Production"].sum().sort_values(ascending=False).
 ↪plot(kind='bar')
plt.title('Total Production Crop wise', size=25)
plt.ylabel('Total production');
```



Coconut is the highest producing crop in India

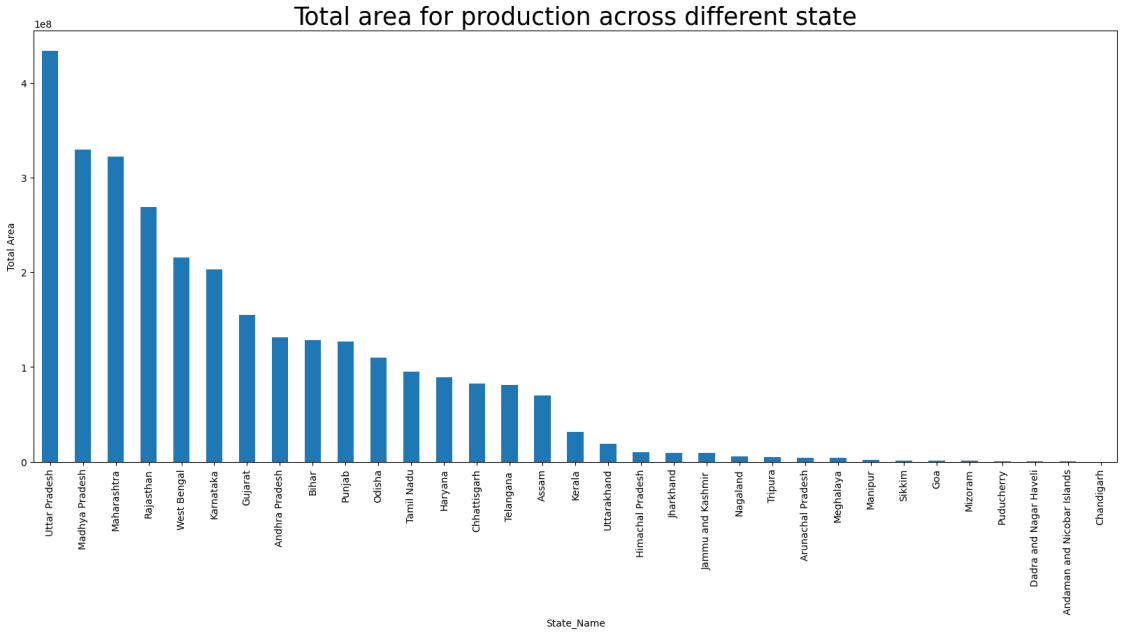
Production per unit area according to Crop

[26]:
```
temp_Crop = df.groupby('Crop')['Area', 'Production'].sum().reset_index()
temp_Crop['Production_Per_Unit_Area'] = temp_Crop['Production']/
 ↪temp_Crop['Area']
temp_Crop = temp_Crop.sort_values(by='Production_Per_Unit_Area')
px.bar(temp_Crop.tail(5), 'Crop', 'Production_Per_Unit_Area',␣
 ↪color='Production_Per_Unit_Area')
```

As production of coconut is higher, same the production rate is also high for coconut

Total area for production across different state

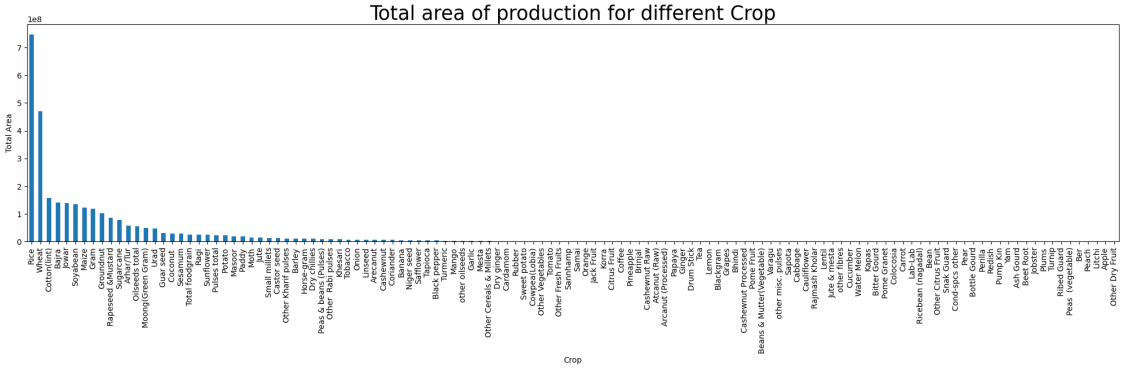
[27]:
```
plt.figure(figsize=(20,8))
df.groupby('State_Name')["Area"].sum().sort_values(ascending=False).
 ↪plot(kind='bar')
plt.title('Total area for production across different state', size=25)
plt.ylabel('Total Area');
```



Uttarpradesh has higher area for crop production

Total area of production for different Crop

[28]:
```
plt.figure(figsize=(25,5))
df.groupby('Crop')["Area"].sum().sort_values(ascending=False).plot(kind='bar')
plt.title('Total area of production for different Crop', size=25)
plt.ylabel('Total Area');
```



Hence rice is grown in abbudent area across India

[29]:
```
temp2 = df.groupby(by='Season')['Area'].mean().reset_index()
px.line(temp2, 'Season', 'Area')
```

Season wise area is getting increased for crop production

[30]:
```
temp3 = df.groupby(by='Crop_Year')['Area'].mean().reset_index()
px.line(temp3, 'Crop_Year', 'Area')
```

Average Crop Area has decresed over the years due to urbanisation, indusatrilisation and resedential areas across the states or may be due to lack of knowledege or intrest towards agriculture. Lowest Average Crop area are in the Year 2002 and 2003 after that slightly increased due to providing proper KT towards the agricultural products and providing benefits to farmer. (We have very comparitively very less data of year 2015 so, we'll not consider that)

Top 5 Crop producting States

[31]:
```
temp_top5_production = df.groupby(by='State_Name')['Production'].sum().
↪reset_index().sort_values(by = 'Production')
px.bar(temp_top5_production.tail(5), 'State_Name', 'Production')
```

5 least Crop producting State in India

[32]:
```
temp_top5_production = df.groupby(by='State_Name')['Production'].sum().
↪reset_index().sort_values(by = 'Production')
px.bar(temp_top5_production.head(5), 'State_Name', 'Production')
```

Top 5 Crop produced

[33]:
```
temp_top5_production_crop = df.groupby(by='Crop')['Production'].sum().
↪reset_index().sort_values(by = 'Production')
px.bar(temp_top5_production_crop.tail(5), 'Crop', 'Production')
```

[34]:
```
temp_top5_production_crop[temp_top5_production_crop['Production']==0]
```

[34]:

	Crop	Production
0	Apple	0.0
79	Pear	0.0
72	Other Dry Fruit	0.0
71	Other Citrus Fruit	0.0
80	Peas (vegetable)	0.0
84	Plums	0.0

```
89              Pump Kin         0.0
58                Litchi         0.0
54               Lab-Lab         0.0
94          Ribed Guard          0.0
35              Cucumber         0.0
104          Snak Guard          0.0
78                Peach          0.0
11            Beet Root          0.0
121        other fibres          0.0
12                  Ber          0.0
120                 Yam          0.0
4             Ash Gourd          0.0
118         Water Melon          0.0
```

0 producing crop in India(may be due to lack of data)

Top 5 Crop producting District

[35]:
```
temp_top5_production_district = df.groupby(by='District_Name')['Production'].
↪sum().reset_index().sort_values(by = 'Production')
px.bar(temp_top5_production_district.tail(5), 'District_Name', 'Production')
```

Least 5 Crop producting States

[36]:
```
px.bar(temp_top5_production_district.head(5), 'District_Name', 'Production')
```

Top 3 highest crop produced in India Analysis

[37]:
```
temp_top5_production_crop = df.groupby(by='Crop')['Production'].sum().
↪reset_index().sort_values(by = 'Production')
px.bar(temp_top5_production_crop.tail(3), 'Crop', 'Production')
```

1. Coconut

[38]:
```
Coconut_df = df[df["Crop"]=="Coconut "]
Coconut_df
```

[38]:
```
                        State_Name District_Name  Crop_Year       Season  \
5       Andaman and Nicobar Islands      NICOBARS       2000  Whole Year
14      Andaman and Nicobar Islands      NICOBARS       2001  Whole Year
23      Andaman and Nicobar Islands      NICOBARS       2002  Whole Year
32      Andaman and Nicobar Islands      NICOBARS       2003  Whole Year
41      Andaman and Nicobar Islands      NICOBARS       2004  Whole Year
```

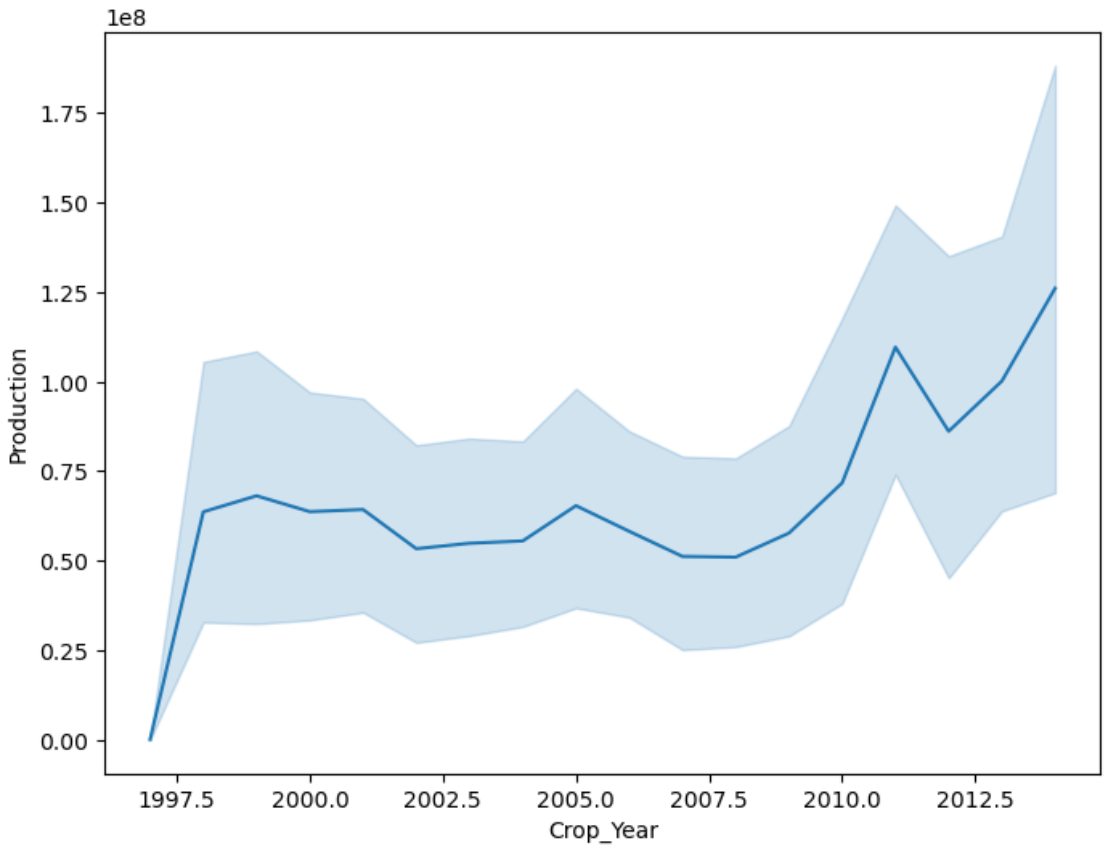
```
…                              …          …          …          …
245719                  West Bengal    PURULIA       2004  Whole Year
245756                  West Bengal    PURULIA       2005  Whole Year
245792                  West Bengal    PURULIA       2006  Whole Year
245837                  West Bengal    PURULIA       2007  Whole Year
245879                  West Bengal    PURULIA       2008  Whole Year

            Crop      Area  Production
5       Coconut   18168.00  65100000.0
14      Coconut   18190.00  64430000.0
23      Coconut   18240.00  67490000.0
32      Coconut   18284.74  68580000.0
41      Coconut   18394.70  52380000.0
…           …         …          …
245719  Coconut      66.00      296.1
245756  Coconut      74.00      311.0
245792  Coconut      73.00   365000.0
245837  Coconut      58.00   898000.0
245879  Coconut      58.00      598.0

[1958 rows x 7 columns]
```

[39]:
```
Coconut_production = Coconut_df.groupby(by='State_Name')['Production'].sum().
↪reset_index().sort_values(by = 'Production')
px.bar(Coconut_production, 'State_Name', 'Production')
```

[40]:
```
Coconut_production = Coconut_df.groupby(by='District_Name')['Production'].sum().
↪reset_index().sort_values(by = 'Production')
px.bar(Coconut_production, 'District_Name', 'Production')
```

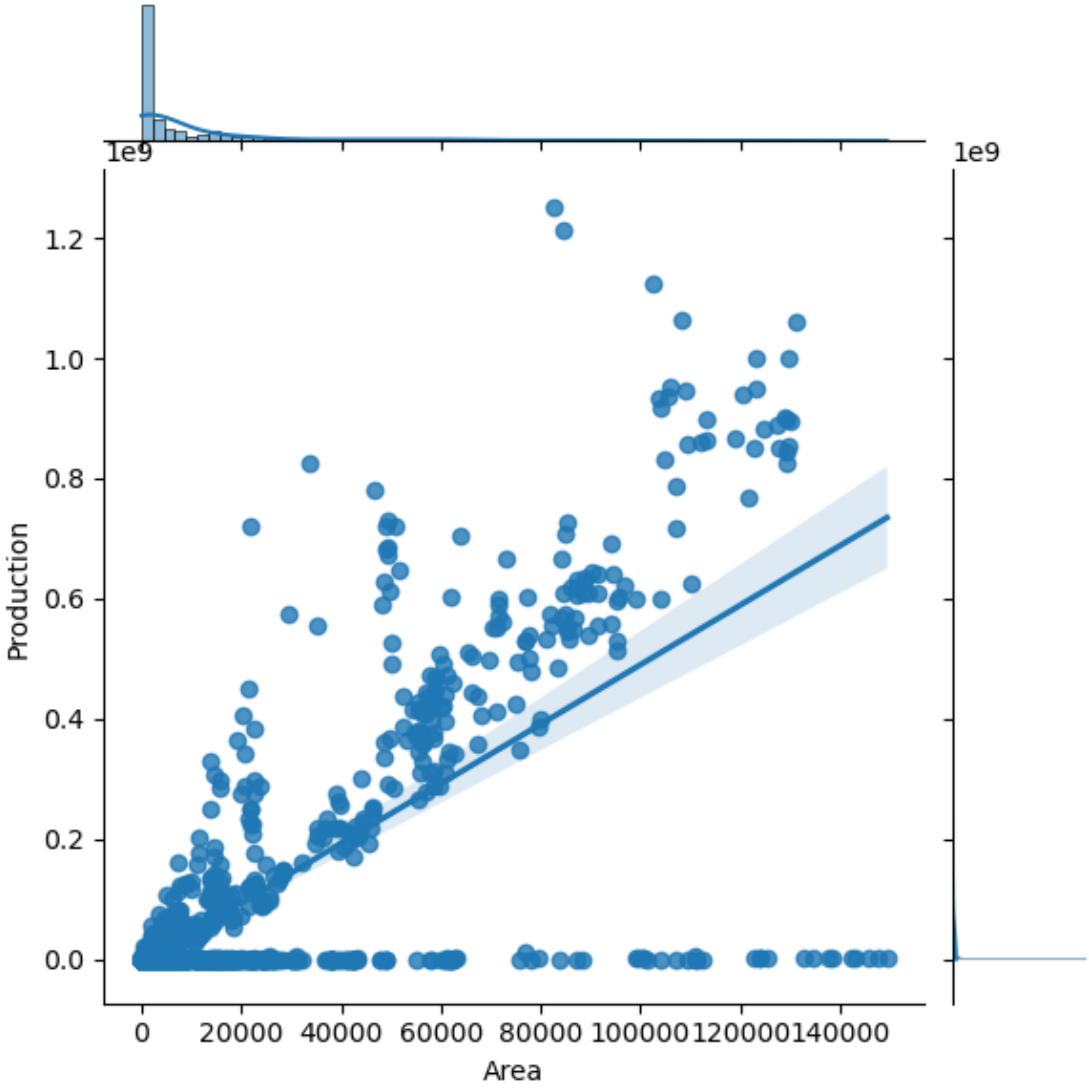
[41]:
```
plt.figure(figsize=(8,6),dpi=100)
sns.lineplot(data=Coconut_df,x='Crop_Year',y='Production')
```

[41]:
```
<Axes: xlabel='Crop_Year', ylabel='Production'>
```



[42]:
```
Coconut_production = Coconut_df.groupby(by='Season')['Production'].sum().
  ↪reset_index().sort_values(by = 'Production')
px.bar(Coconut_production, 'Season', 'Production')
```

[43]:
```
sns.jointplot(x = "Area",y = "Production",data=Coconut_df,kind="reg")
plt.show()
```



Coconut is majorly produced in South Indian states where there is high rainfall. Over the year the production of coconut is getting increased. It is mainly grown during Kharfi season across the countries. From Data Visualization: coconut production is mostly depends on Season, Area, State(place).

2. Sugarcane

[44]:
```
Sugarcane_df = df[df["Crop"]=="Sugarcane"]
Sugarcane_df
```

[44]:

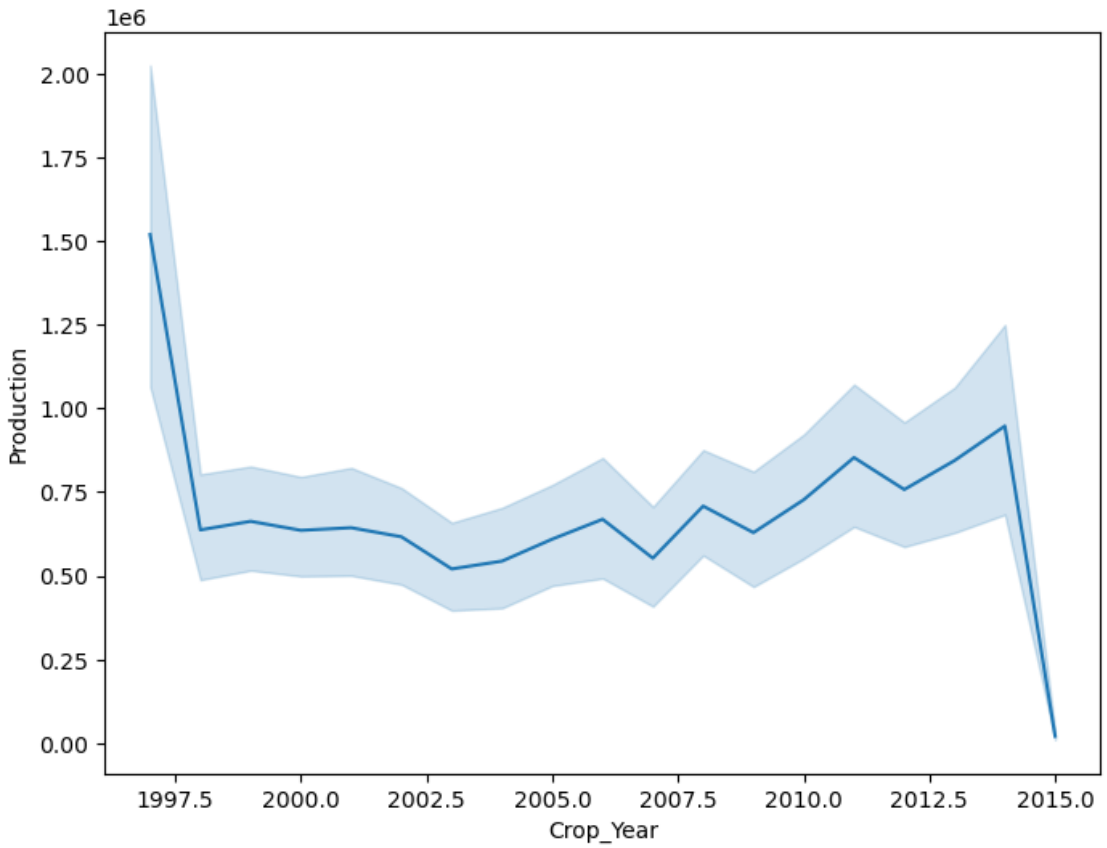
	State_Name	District_Name	Crop_Year	Season	\
7	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	

```
16      Andaman and Nicobar Islands      NICOBARS       2001  Whole Year
26      Andaman and Nicobar Islands      NICOBARS       2002  Whole Year
60      Andaman and Nicobar Islands      NICOBARS       2006  Whole Year
65      Andaman and Nicobar Islands      NICOBARS       2010  Autumn
…                               …          …       …        …
245953                 West Bengal       PURULIA        2010  Whole Year
245984                 West Bengal       PURULIA        2011  Whole Year
246016                 West Bengal       PURULIA        2012  Whole Year
246051                 West Bengal       PURULIA        2013  Whole Year
246088                 West Bengal       PURULIA        2014  Whole Year

             Crop   Area  Production
7       Sugarcane    1.0        2.00
16      Sugarcane    1.0        1.00
26      Sugarcane    5.0       40.00
60      Sugarcane    0.2        0.50
65      Sugarcane   13.4       41.75
…           …    …          …
245953  Sugarcane  303.0    11541.00
245984  Sugarcane  303.0    20264.00
246016  Sugarcane  314.0    11199.00
246051  Sugarcane  161.0     7948.00
246088  Sugarcane  324.0    16250.00

[7827 rows x 7 columns]
```

[45]:
```
Sugarcane_production = Sugarcane_df.groupby(by='State_Name')['Production'].
  ↪sum().reset_index().sort_values(by = 'Production')
px.bar(Sugarcane_production, 'State_Name', 'Production')
```

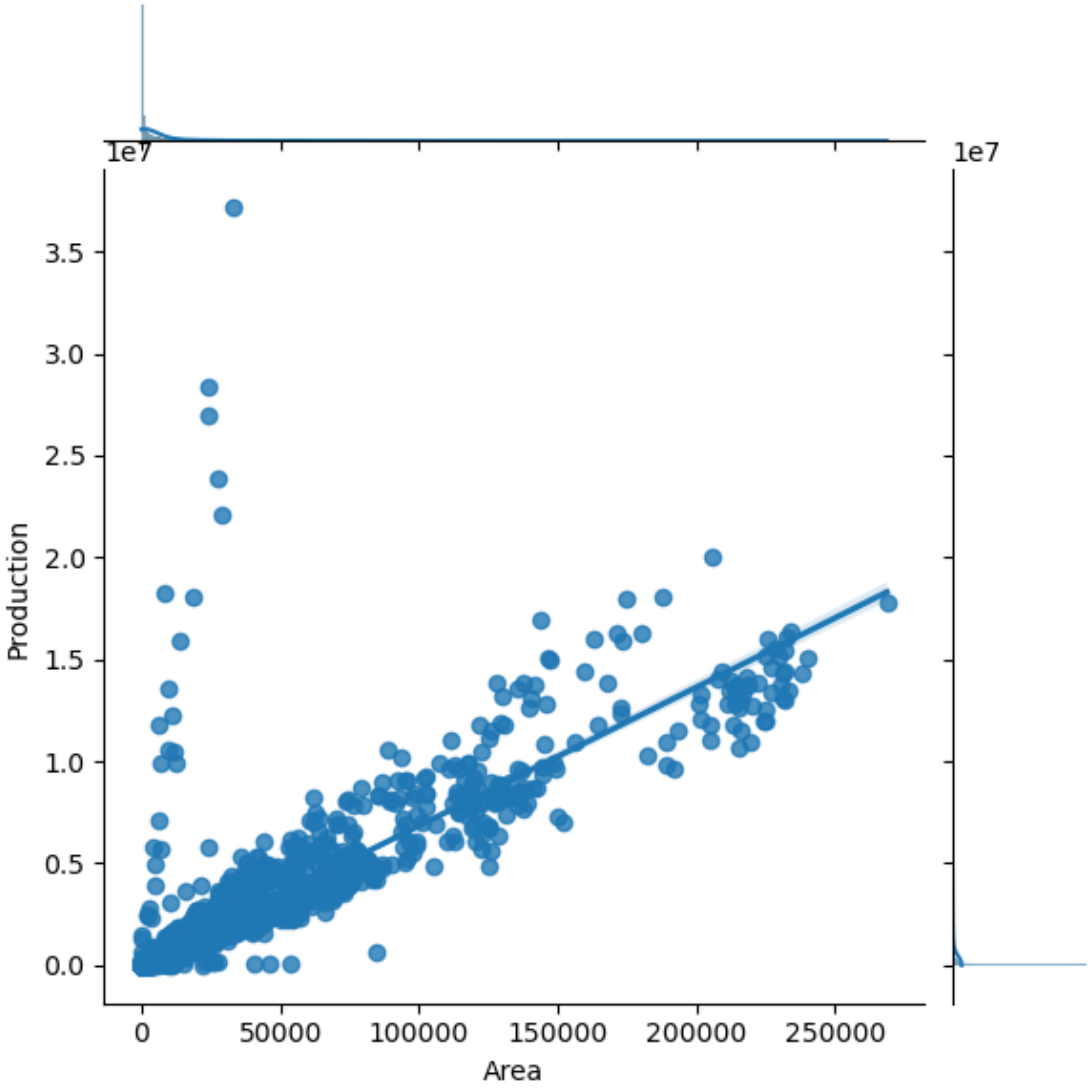
[46]:
```
plt.figure(figsize=(8,6),dpi=100)
sns.lineplot(data=Sugarcane_df,x='Crop_Year',y='Production')
```

[46]:
```
<Axes: xlabel='Crop_Year', ylabel='Production'>
```



[47]:
```
Sugarcane_production = Sugarcane_df.groupby(by='District_Name')['Production'].
  ↪sum().reset_index().sort_values(by = 'Production')
px.bar(Sugarcane_production, 'District_Name', 'Production')
```

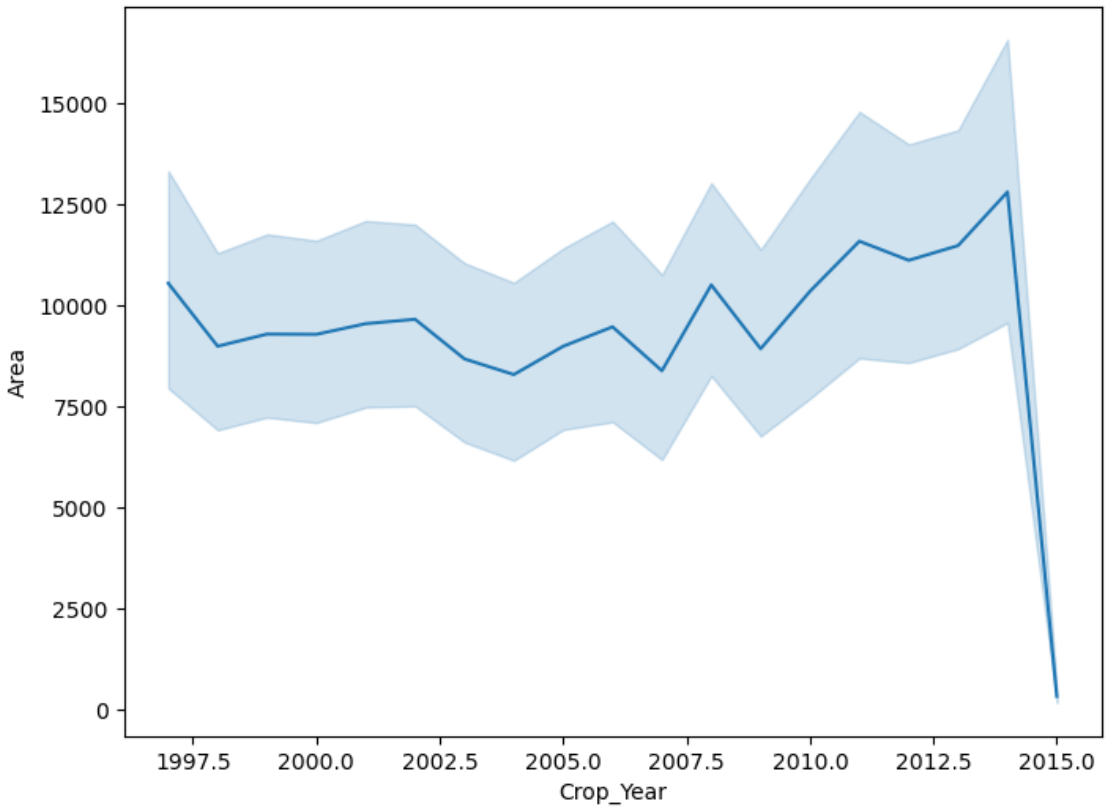
[48]:
```
sns.jointplot(x = "Area",y = "Production",data=Sugarcane_df,kind="reg")
plt.show()
```



[49]:
```
Sugarcane_production = Sugarcane_df.groupby(by='Season')['Production'].sum().
↪reset_index().sort_values(by = 'Production')
px.bar(Sugarcane_production, 'Season', 'Production')
```

[50]:
```
plt.figure(figsize=(8,6),dpi=100)
sns.lineplot(data=Sugarcane_df,x='Crop_Year',y='Area')
```

[50]:
```
<Axes: xlabel='Crop_Year', ylabel='Area'>
```



Sugarcane is largly produced in Uttarpradesh and other than that it is majorly produced in South Indian districts. After 1998 the production of Sugarcane is getting increased. It is mainly grown during Kharif season across the countries. Also as the area increases and the production is getting increased. From Data Visualization: Sugarcane production is mostly depends on Season, Area, State(place).

3.Rice

[51]:
```
Rice_df = df[df["Crop"]=="Rice"]
Rice_df
```

[51]:

	State_Name	District_Name	Crop_Year	Season	\
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	
12	Andaman and Nicobar Islands	NICOBARS	2001	Kharif	
18	Andaman and Nicobar Islands	NICOBARS	2002	Kharif	
27	Andaman and Nicobar Islands	NICOBARS	2003	Kharif	
36	Andaman and Nicobar Islands	NICOBARS	2004	Kharif	
…	…	…	…	…	
246049	West Bengal	PURULIA	2013	Summer	
246052	West Bengal	PURULIA	2013	Winter	

```
246058                 West Bengal       PURULIA       2014  Autumn
246086                 West Bengal       PURULIA       2014  Summer
246089                 West Bengal       PURULIA       2014  Winter

        Crop       Area  Production
2       Rice     102.00      321.00
12      Rice      83.00      300.00
18      Rice     189.20      510.84
27      Rice      52.00       90.17
36      Rice      52.94       72.57
…       …         …           …
246049  Rice     516.00     1274.00
246052  Rice  302274.00   730136.00
246058  Rice     264.00      721.00
246086  Rice     306.00      801.00
246089  Rice  279151.00   597899.00

[15082 rows x 7 columns]
```

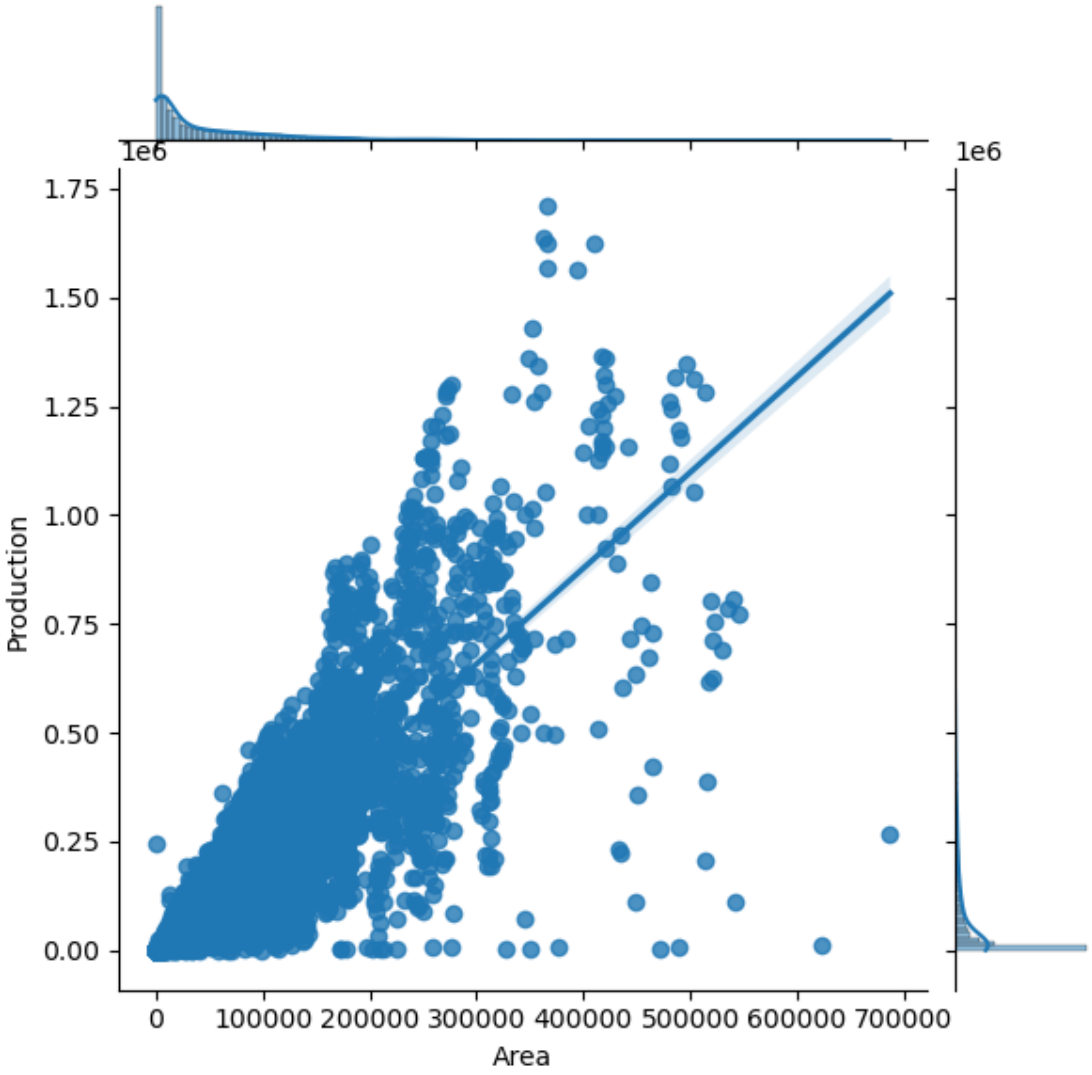
[52]:
```
Rice_production = Rice_df.groupby(by='State_Name')['Production'].sum().
 ↪reset_index().sort_values(by = 'Production')
px.bar(Rice_production, 'State_Name', 'Production')
```

[53]:
```
Rice_production = Rice_df.groupby(by='District_Name')['Production'].sum().
 ↪reset_index().sort_values(by = 'Production')
px.bar(Rice_production, 'District_Name', 'Production')
```

[54]:
```
plt.figure(figsize=(8,6),dpi=100)
sns.lineplot(data=Rice_df,x='Crop_Year',y='Production')
```

[54]:
```
<Axes: xlabel='Crop_Year', ylabel='Production'>
```

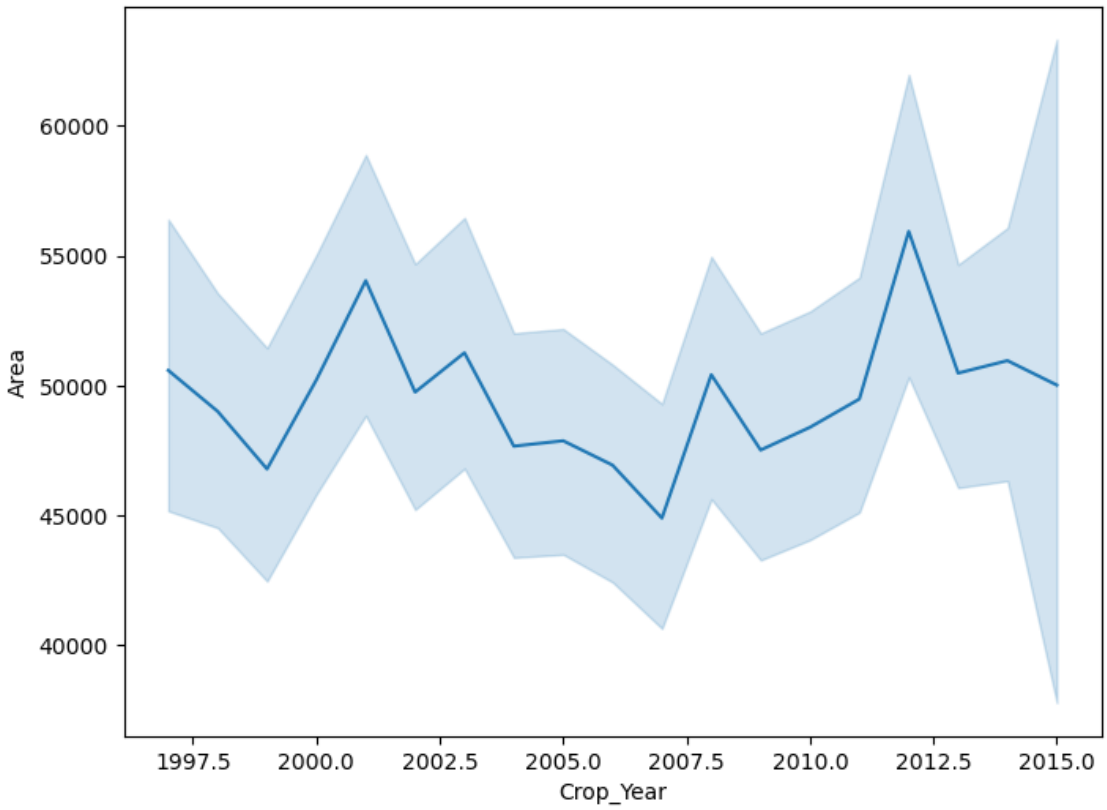
[55]:
```
sns.jointplot(x = "Area",y = "Production",data=Rice_df,kind="reg")
plt.show()
```



[56]:
```
Rice_production = Rice_df.groupby(by='Season')['Production'].sum().
  ↪reset_index().sort_values(by = 'Production')
px.bar(Rice_production, 'Season', 'Production')
```

[57]:
```
plt.figure(figsize=(8,6),dpi=100)
sns.lineplot(data=Rice_df,x='Crop_Year',y='Area')
```

[57]:
```
<Axes: xlabel='Crop_Year', ylabel='Area'>
```



Rice is largly produced in norther part of India. The production of Sugarcane is getting increased. It is mainly grown during Kharif season across the countries. Also as the area for production is getting increased and the production is also getting increased. From Data Visualization: Rice production is also mostly depends on Season, Area, State(place).

Seasonal crops Analysis

[58]:
```
df["Season"].unique()
```

[58]:
```
array(['Kharif     ', 'Whole Year ', 'Autumn     ', 'Rabi       ',
       'Summer     ', 'Winter     '], dtype=object)
```

Kharif

Kharif Crops -: These crops are sown in the early monsoon season, which generally varies by crop and region of cultivation. In India, Kharif crops are sown at the beginning of the rainy season, between the month of June and July. These crops are harvested at the end of monsoon season, between the month of September and October. Paddy is the main Kharif crop.

July – October

Harvest – September to October

A.k.a Monsoon Crops

Such crops require a lot of water

Example: rice, sorghum, maize, tea, rubber, coffee, guar, Sesame, cereals such as Arhar Dhal, pearl millet, soybeans, cotton, oilseeds, etc.

[59]:
```
Kharif_df = df[df["Season"]=="Kharif     "]
Kharif_df
```

[59]:

	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
10	Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Arecanut	1254.0	2061.0
11	Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Other Kharif pulses	2.0	1.0
…	…	…	…	…	…	…	…
246066	West Bengal	PURULIA	2014	Kharif	Other Kharif pulses	79.0	39.0
246067	West Bengal	PURULIA	2014	Kharif	Sannhamp	171.0	727.0
246068	West Bengal	PURULIA	2014	Kharif	Soyabean	18.0	7.0
246069	West Bengal	PURULIA	2014	Kharif	Sunflower	46.0	42.0
246070	West Bengal	PURULIA	2014	Kharif	Urad	11493.0	3287.0

[94283 rows x 7 columns]

[60]:
```
Kharif_production = Kharif_df.groupby(by='State_Name')['Production'].sum().
↪reset_index().sort_values(by = 'Production')
px.bar(Kharif_production, 'State_Name', 'Production')
```

[61]:
```
Kharif_production = Kharif_df.groupby(by='State_Name')['Production'].mean().
↪reset_index().sort_values(by = 'Production')
px.bar(Kharif_production, 'State_Name', 'Production')
```

[62]:
```
Kharif_production = Kharif_df.groupby(by='Crop')['Production'].sum().
↪reset_index().sort_values(by = 'Production')
px.bar(Kharif_production, 'Crop', 'Production')
```

[63]:
```
Kharif_production = Kharif_df.groupby(by='Crop')['Production'].mean().
↪reset_index().sort_values(by = 'Production')
px.bar(Kharif_production, 'Crop', 'Production')
```

Hence during Kharif season majorly crops are grown in uttarpradesh but on for all crops average crop grown is in punjab. During kharif season major crop grown is Sugarcane

Autumn

The four seasons are spring, summer, autumn, and winter. The word 'Autumn', is derived from a Latin word 'autumnus', and first appeared in English in the late 14th century. This article will share some insights into the autumn season.

For many farming cultures, autumn ushers in a time of celebration.

Autumn is the time of year that acts as a transition between summer and winter.

The autumn months are the time of the harvest season. Autumn is generally regarded as the end of the growing season.

In the autumn season, the daylight grows shorter, and animals prepare for the long, cold months ahead.

The temperature starts becoming cooler during autumn.

Leaves on the trees will turn yellow, orange, red and brown during autumn.

In the United States, autumn is also known as fall.

[64]:
```
Autumn_df = df[df["Season"]=="Autumn     "]
Autumn_df
```

[64]:

	State_Name	District_Name	Crop_Year	\
64	Andaman and Nicobar Islands	NICOBARS	2010	
65	Andaman and Nicobar Islands	NICOBARS	2010	
111	Andaman and Nicobar Islands	NORTH AND MIDDLE ANDAMAN	2010	
112	Andaman and Nicobar Islands	NORTH AND MIDDLE ANDAMAN	2010	
185	Andaman and Nicobar Islands	SOUTH ANDAMANS	2010	
…	…	…	…	
246056	West Bengal	PURULIA	2014	
246057	West Bengal	PURULIA	2014	
246058	West Bengal	PURULIA	2014	
246059	West Bengal	PURULIA	2014	
246060	West Bengal	PURULIA	2014	

```
             Season           Crop     Area  Production
64       Autumn                Rice     3.50       10.00
65       Autumn           Sugarcane    13.40       41.75
111      Autumn                Rice  6791.00    20118.00
112      Autumn           Sugarcane    73.33      889.20
185      Autumn                Rice  1595.50     3788.00
…             …              …      …          …
246056   Autumn               Maize  6317.00    13337.00
246057   Autumn                Ragi   112.00       44.00
246058   Autumn                Rice   264.00      721.00
246059   Autumn             Sesamum   170.00       87.00
246060   Autumn       Small millets   154.00       41.00

[4930 rows x 7 columns]
```

[65]:
```
Autumn_production = Autumn_df.groupby(by='State_Name')['Production'].sum().
↪reset_index().sort_values(by = 'Production')
px.bar(Autumn_production, 'State_Name', 'Production')
```

[66]:
```
Autumn_production = Autumn_df.groupby(by='State_Name')['Production'].mean().
↪reset_index().sort_values(by = 'Production')
px.bar(Autumn_production, 'State_Name', 'Production')
```

[67]:
```
Autumn_production = Autumn_df.groupby(by='Crop')['Production'].sum().
↪reset_index().sort_values(by = 'Production')
px.bar(Autumn_production, 'Crop', 'Production')
```

[68]:
```
Autumn_production = Autumn_df.groupby(by='Crop')['Production'].mean().
↪reset_index().sort_values(by = 'Production')
px.bar(Autumn_production, 'Crop', 'Production')
```

Hence during Autumn season majorly crops are grown in Northen part of India. During Autumn season major crop grown are Rice, Paddy and Maize

Rabi

Rabi Crops -: These crops are sown during winter and after the monsoon, which is between the month of October and November. In India, Rabi crops are harvested during the spring between the month of March and April. Wheat is the main Rabi crop.

Sowing between October and November

Harvest – February to April

A.k.a Winter Season Crops

Need cold weather for growth

Need less water

Example: wheat, oats, barley, pulses, cereals, oilseeds, linseed, etc.

[69]:
```
Rabi_df = df[df["Season"]=="Rabi       "]
Rabi_df
```

[69]:
```
                        State_Name District_Name  Crop_Year       Season  \
66      Andaman and Nicobar Islands      NICOBARS       2010  Rabi
67      Andaman and Nicobar Islands      NICOBARS       2010  Rabi
68      Andaman and Nicobar Islands      NICOBARS       2010  Rabi
69      Andaman and Nicobar Islands      NICOBARS       2010  Rabi
70      Andaman and Nicobar Islands      NICOBARS       2010  Rabi
…                               …           …         …          …
246080                  West Bengal       PURULIA       2014  Rabi
246081                  West Bengal       PURULIA       2014  Rabi
246082                  West Bengal       PURULIA       2014  Rabi
246083                  West Bengal       PURULIA       2014  Rabi
246084                  West Bengal       PURULIA       2014  Rabi

                     Crop    Area  Production
66               Arecanut   944.0     1610.00
67           Black pepper    23.0        8.50
68              Cashewnut  1000.5      260.50
69           Dry chillies    12.0       25.00
70             Dry ginger     7.0        9.64
…                     …      …          …
246080              Potato   477.0     9995.00
246081  Rapeseed &Mustard  1885.0     1508.00
246082          Safflower    54.0       37.00
246083               Urad   220.0      113.00
246084              Wheat  1622.0     3663.00

[66160 rows x 7 columns]
```

[70]:
```
Rabi_production = Rabi_df.groupby(by='State_Name')['Production'].sum().
 ↪reset_index().sort_values(by = 'Production')
px.bar(Rabi_production, 'State_Name', 'Production')
```

[71]:
```
Rabi_production = Rabi_df.groupby(by='State_Name')['Production'].mean().
 ↪reset_index().sort_values(by = 'Production')
px.bar(Rabi_production, 'State_Name', 'Production')
```

[72]:
```
Rabi_production = Rabi_df.groupby(by='Crop')['Production'].sum().reset_index().
 ↪sort_values(by = 'Production')
px.bar(Rabi_production, 'Crop', 'Production')
```

[73]:
```
Rabi_production = Rabi_df.groupby(by='Crop')['Production'].mean().reset_index().
 ↪sort_values(by = 'Production')
```

```
px.bar(Rabi_production, 'Crop', 'Production')
```

Hence during Rabi season majorly crops are grown in Northen part of India. During Rabi season major crop grown are Wheat & Potato

Winter

[74]:
```
Winter_df = df[df["Season"]=="Winter     "]
Winter_df
```

[74]:

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
12406	Assam	BAKSA	2005	Winter	Rice	66229.0	93287.0
12438	Assam	BAKSA	2006	Winter	Rice	52195.0	67915.0
12472	Assam	BAKSA	2007	Winter	Rice	58558.0	73246.0
12506	Assam	BAKSA	2008	Winter	Rice	69495.0	124614.0
12540	Assam	BAKSA	2009	Winter	Rice	68453.0	109965.0
…	…	…	…	…	…	…	…
246018	West Bengal	PURULIA	2012	Winter	Sesamum	224.0	88.0
246052	West Bengal	PURULIA	2013	Winter	Rice	302274.0	730136.0
246053	West Bengal	PURULIA	2013	Winter	Sesamum	208.0	101.0
246089	West Bengal	PURULIA	2014	Winter	Rice	279151.0	597899.0
246090	West Bengal	PURULIA	2014	Winter	Sesamum	175.0	88.0

[6050 rows x 7 columns]

[75]:
```
Winter_production = Winter_df.groupby(by='State_Name')['Production'].sum().
  ↪reset_index().sort_values(by = 'Production')
px.bar(Winter_production, 'State_Name', 'Production')
```

[76]:
```
Winter_production = Winter_df.groupby(by='State_Name')['Production'].mean().
  ↪reset_index().sort_values(by = 'Production')
px.bar(Winter_production, 'State_Name', 'Production')
```

[77]:
```
Winter_production = Winter_df.groupby(by='Crop')['Production'].sum().
↪reset_index().sort_values(by = 'Production')
px.bar(Winter_production, 'Crop', 'Production')
```

[78]:
```
Winter_production = Winter_df.groupby(by='Crop')['Production'].mean().
↪reset_index().sort_values(by = 'Production')
px.bar(Winter_production, 'Crop', 'Production')
```

Hence during Winter season majorly crops are grown in Northen part of India. During Rabi season major crop grown are Rice, Paddy & Sugarcane

Summer

[79]:
```
Summer_df = df[df["Season"]=="Summer     "]
Summer_df
```

[79]:

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
12396	Assam	BAKSA	2005	Summer	Rice	12650.0	17027.0
12426	Assam	BAKSA	2006	Summer	Rice	12737.0	25691.0
12458	Assam	BAKSA	2007	Summer	Rice	7124.0	14341.0
12492	Assam	BAKSA	2008	Summer	Rice	11712.0	20949.0
12526	Assam	BAKSA	2009	Summer	Rice	10524.0	16463.0
…	…	…	…	…	…	…	…
246049	West Bengal	PURULIA	2013	Summer	Rice	516.0	1274.0
246050	West Bengal	PURULIA	2013	Summer	Sesamum	697.0	356.0
246085	West Bengal	PURULIA	2014	Summer	Maize	325.0	2039.0
246086	West Bengal	PURULIA	2014	Summer	Rice	306.0	801.0
246087	West Bengal	PURULIA	2014	Summer	Sesamum	627.0	463.0

[14811 rows x 7 columns]

[80]:
```
Summer_production = Summer_df.groupby(by='State_Name')['Production'].sum().
↪reset_index().sort_values(by = 'Production')
```

```
px.bar(Summer_production, 'State_Name', 'Production')
```

[81]:
```
Summer_production = Summer_df.groupby(by='State_Name')['Production'].mean().
↪reset_index().sort_values(by = 'Production')
px.bar(Summer_production, 'State_Name', 'Production')
```

[82]:
```
Summer_production = Summer_df.groupby(by='Crop')['Production'].sum().
↪reset_index().sort_values(by = 'Production')
px.bar(Summer_production, 'Crop', 'Production')
```

[83]:
```
Summer_production = Summer_df.groupby(by='Crop')['Production'].mean().
↪reset_index().sort_values(by = 'Production')
px.bar(Summer_production, 'Crop', 'Production')
```

Hence during Summer season on an average crops are grown in Northen part of India. During Summer season major crop grown are Rice, Paddy and Bajra

Top Crop Producing State anlysis

[84]:
```
Kerala_df = df[df["State_Name"]=="Kerala"]
Kerala_df
```

[84]:

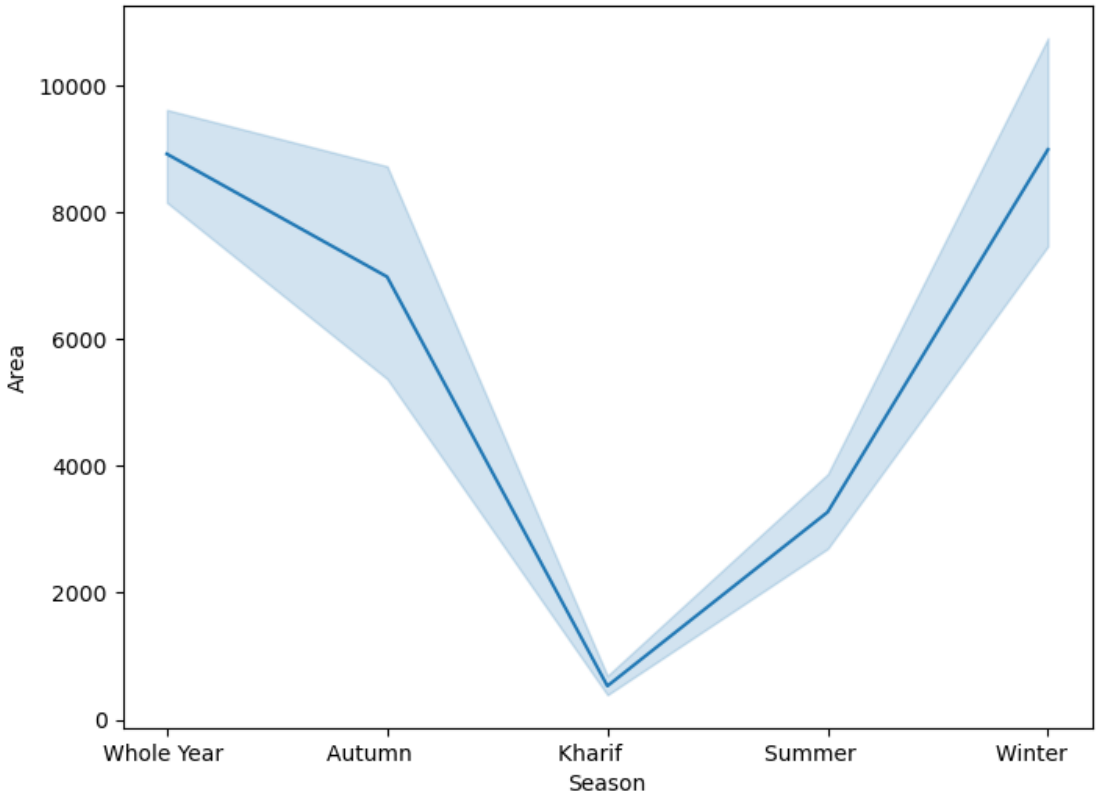
	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
97987	Kerala	ALAPPUZHA	1997	Whole Year	Arecanut	2253.0	1518.00
97988	Kerala	ALAPPUZHA	1997	Whole Year	Black pepper	2235.0	248.00
97989	Kerala	ALAPPUZHA	1997	Whole Year	Cashewnut	7205.0	2358.00
97990	Kerala	ALAPPUZHA	1997	Whole Year	Coconut	77893.0	399000.00
97991	Kerala	ALAPPUZHA	1997	Whole Year	Tapioca	3911.0	74110.00
…	…	…	…	…	…	…	…
102242	Kerala	WAYANAD	2014	Whole Year	Sesamum	1.0	0.45
102244	Kerala	WAYANAD	2014	Whole Year	Sweet potato	8.0	127.00
102245	Kerala	WAYANAD	2014	Whole Year	Tapioca	2327.0	99788.00
102246	Kerala	WAYANAD	2014	Whole Year	Turmeric	161.0	481.00
102247	Kerala	WAYANAD	2014	Winter	Rice		

```
102247    8651.0    23324.00

[4003 rows x 7 columns]
```

[85]:
```
Kerala_production = Kerala_df.groupby(by='District_Name')['Production'].sum().
  ↪reset_index().sort_values(by = 'Production')
px.bar(Kerala_production, 'District_Name', 'Production')
```

[86]:
```
plt.figure(figsize=(8,6),dpi=100)
sns.lineplot(data=Kerala_df,x='Season',y='Area')
```

[86]:
```
<Axes: xlabel='Season', ylabel='Area'>
```



[87]:
```
Kerala_production = Kerala_df.groupby(by='Crop')['Production'].sum().
  ↪reset_index().sort_values(by = 'Production')
px.bar(Kerala_production, 'Crop', 'Production')
```

[88]:
```
Kerala_production = Kerala_df.groupby(by='Crop_Year')['Production'].mean().
  ↪reset_index().sort_values(by = 'Production')
px.bar(Kerala_production, 'Crop_Year', 'Production')
```

Hence Kerala's major crop is coconut Tirunavantampuram is the district in Kerala where crop is produced more Area for production is getting decreased in Kerala but still the production is at peak in India Kerala is mainintaing Average crop production from 1998 to 2015 with 1 or 2 up downs

Maharashtra Analysis

[89]:
```
df["State_Name"].unique()
```

[89]:
```
array(['Andaman and Nicobar Islands', 'Andhra Pradesh',
       'Arunachal Pradesh', 'Assam', 'Bihar', 'Chandigarh',
       'Chhattisgarh', 'Dadra and Nagar Haveli', 'Goa', 'Gujarat',
       'Haryana', 'Himachal Pradesh', 'Jammu and Kashmir ', 'Jharkhand',
       'Karnataka', 'Kerala', 'Madhya Pradesh', 'Maharashtra', 'Manipur',
       'Meghalaya', 'Mizoram', 'Nagaland', 'Odisha', 'Puducherry',
       'Punjab', 'Rajasthan', 'Sikkim', 'Tamil Nadu', 'Telangana ',
       'Tripura', 'Uttar Pradesh', 'Uttarakhand', 'West Bengal'],
      dtype=object)
```

[90]:
```
Maharashtra_df = df[df["State_Name"]=="Maharashtra"]
Maharashtra_df
```

[90]:

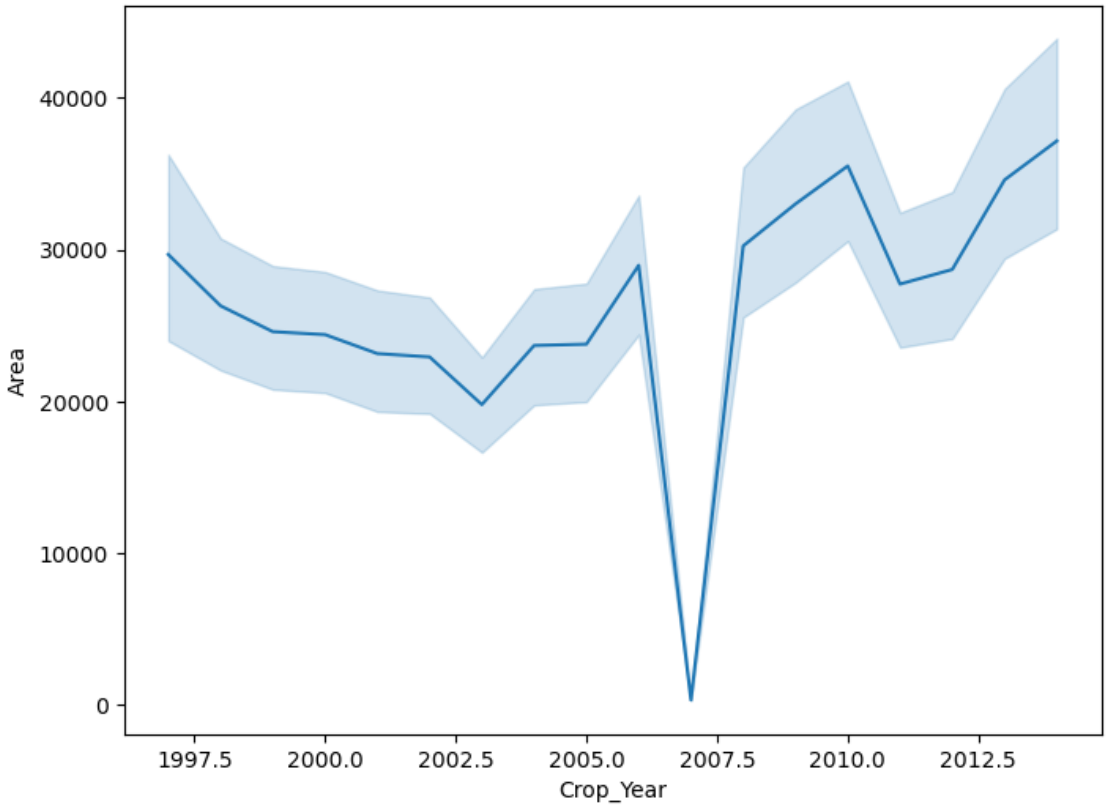
	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
125191	Maharashtra	AHMEDNAGAR	1997	Autumn	Maize	1.0	1113.0
125192	Maharashtra	AHMEDNAGAR	1997	Kharif	Arhar/Tur	17600.0	6300.0
125193	Maharashtra	AHMEDNAGAR	1997	Kharif	Bajra	274100.0	152800.0
125194	Maharashtra	AHMEDNAGAR	1997	Kharif	Gram	40800.0	18600.0
125195	Maharashtra	AHMEDNAGAR	1997	Kharif	Jowar	900.0	1100.0
…	…	…	…	…	…	…	…
137814	Maharashtra	YAVATMAL	2014	Rabi	Jowar	4000.0	3300.0
137815	Maharashtra	YAVATMAL	2014	Rabi	Maize	1300.0	200.0
137816	Maharashtra	YAVATMAL	2014	Rabi	Wheat	29100.0	26800.0
137817	Maharashtra	YAVATMAL	2014	Summer	Groundnut	9400.0	11500.0
137818	Maharashtra	YAVATMAL	2014	Whole Year	Sugarcane	8100.0	553700.0

```
[12496 rows x 7 columns]
```

[91]:
```
Maharashtra_production = Maharashtra_df.
↪groupby(by='District_Name')['Production'].sum().reset_index().sort_values(by␣
↪= 'Production')
px.bar(Maharashtra_production, 'District_Name', 'Production')
```

[92]:
```
plt.figure(figsize=(8,6),dpi=100)
sns.lineplot(data=Maharashtra_df,x='Crop_Year',y='Area')
```

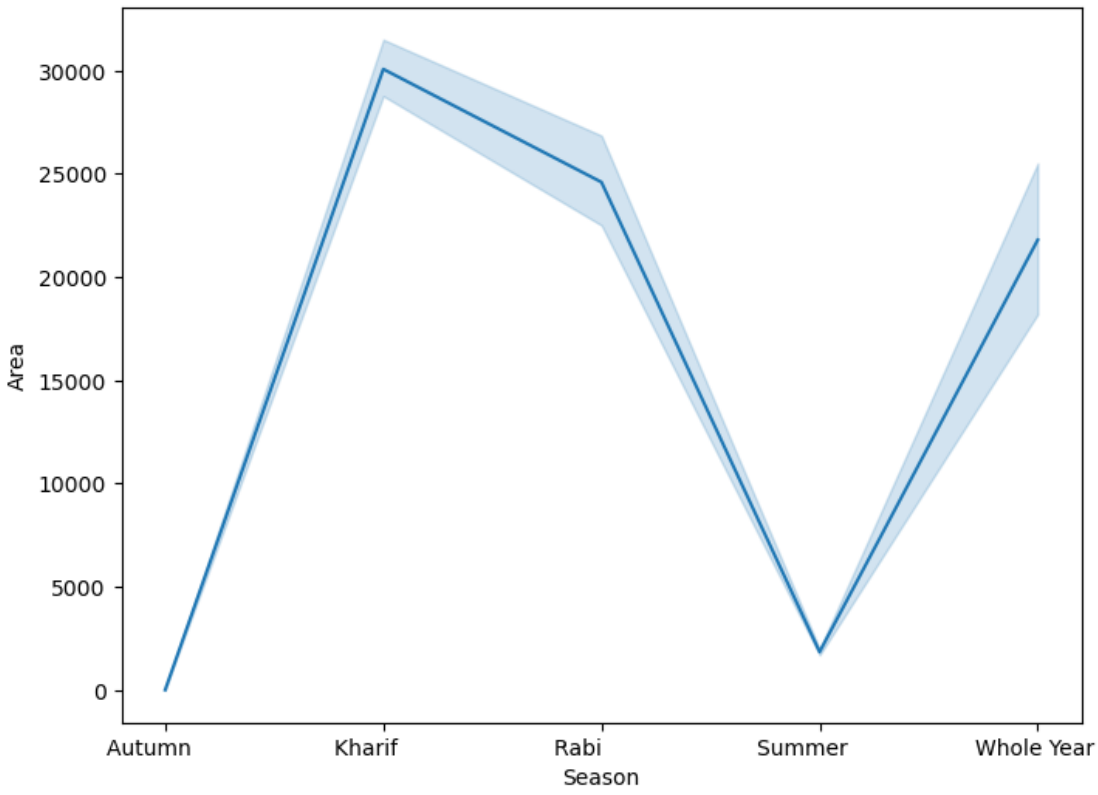
[92]:
```
<Axes: xlabel='Crop_Year', ylabel='Area'>
```



[93]:
```
Maharashtra_production = Maharashtra_df.groupby(by='Crop_Year')['Production'].
↪mean().reset_index().sort_values(by = 'Production')
px.bar(Maharashtra_production, 'Crop_Year', 'Production')
```

[94]:
```
plt.figure(figsize=(8,6),dpi=100)
sns.lineplot(data=Maharashtra_df,x='Season',y='Area')
```

[94]:
```
<Axes: xlabel='Season', ylabel='Area'>
```



[95]:
```
Maharashtra_production = Maharashtra_df.groupby(by='Crop')['Production'].sum().
↪reset_index().sort_values(by = 'Production')
px.bar(Maharashtra_production, 'Crop', 'Production')
```

Hence Maharashtra's major crop is Sugarcane. Khollpur is the district in Maharashtra where crop is produced more. Area for production is slightly getting increased. Maharastra avrearge production is getting constant after 2008.

What factors control crop production?

The factors that control crop production include: 1. Temperature 2. Precipitation 3. Solar radiation 4. Wind velocity 5. Soil moisture

Visualization Done On:

1: State Wise, District, Year wise, Season Wise and type of crop wise Production

2: Checked for Rate of production vs all entities like State wise production vs production rate, crop wise production vs production rate. 3: Top Crops Production In India are Sugarcane, Rice and Sugarcane and same all 3 crop analysis are done accordingly and found all the crop depends on Season, Area and Place. 4: Season Wise crop production Analysis are done for different seasons

Like Kharif, Rabi, Summer, Winter and Rabi and found crop production depends on Season and area where it is grown.

5: Top Crop producing state crop production analysis

6: Maharashtra's crop production analysis

Conclusion

This project provides valuable insights into crop production patterns and trends. By exploring and visualizing the data, we have gained a better understanding of the most commonly grown crops, the distribution of crop production across seasons and states, and the highest production in different years. These insights can be utilized for decision-making, resource allocation, and future planning in the agricultural sector. Additionally, the project highlights the importance of data exploration, visualization, and analytical techniques in understanding and utilizing agricultural data effectively.

End project

By Rajesh Hinduja

THANK YOU!!