

```
In [40]: #importing relevant libraries and importing data
```

```
In [51]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

```
In [2]: df=pd.read_csv("C:/Users/csasi/OneDrive/Desktop/myprojects/datasets/Crop_Production_data.csv", header=0)
```

```
In [42]: # observing, cleaning and analyzing data
```

```
In [5]: df.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 [6]: pd.isnull(df)
```

```
Out[6]:
```

	State Name	District Name	Crop_Year	Season	Crop	Area	Production
0	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False
...
246086	False	False	False	False	False	False	False
246087	False	False	False	False	False	False	False
246088	False	False	False	False	False	False	False
246089	False	False	False	False	False	False	False
246090	False	False	False	False	False	False	False

246091 rows x 7 columns

```
In [7]: pd.isnull(df).sum()
```

```
Out[7]: State Name      0
District_Name    0
Crop_Year        0
Season           0
Crop             0
Area             0
Production      3730
dtype: int64
```

```
In [43]: df.Production=df.Production.fillna(df.Production.mean()) #replacing null values in'Production' field with its mean value
```

```
In [9]: df.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  246091 non-null float64
dtypes: float64(2), int64(1), object(4)
memory usage: 13.1+ MB
```

```
In [11]: df.describe()
```

```
Out[11]:
```

	Crop_Year	Area	Production
count	246091.000000	2.460910e+05	2.460910e+05
mean	2005.643018	1.200282e+04	5.825034e+05
std	4.952164	5.052340e+04	1.693599e+07
min	1997.000000	4.000000e+02	0.000000e+00
25%	2002.000000	8.000000e+01	9.100000e+01
50%	2006.000000	5.620000e+02	7.880000e+02
75%	2010.000000	4.392000e+03	8.000000e+03
max	2015.000000	8.580100e+06	1.250800e+09

```
In [15]: df.columns
```

```
Out[15]: Index(['State Name', 'District Name', 'Crop_Year', 'Season', 'Crop', 'Area',
'Production'],
dtype='object')
```

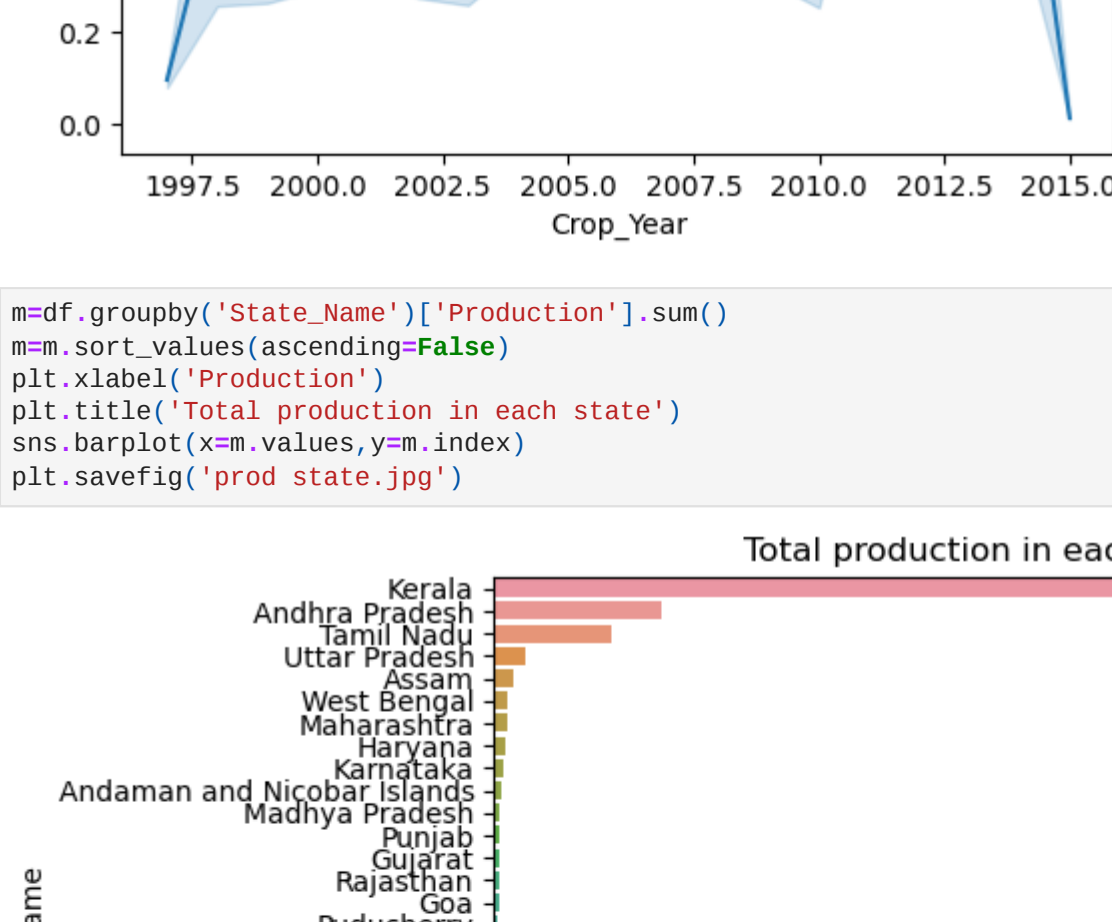
```
In [4]: df.shape
```

```
Out[4]: (246091, 7)
```

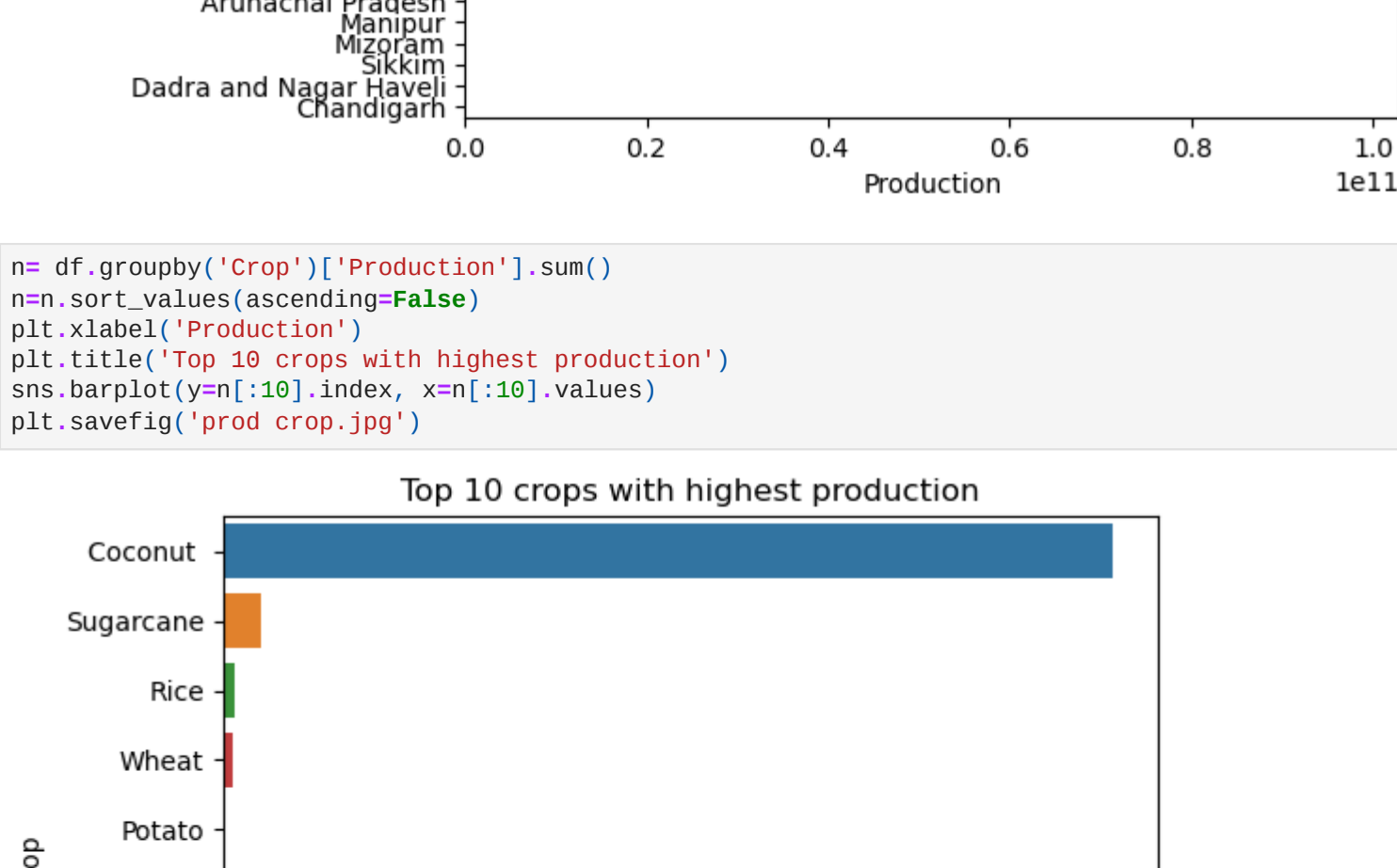
```
In [44]: # conducting eda
```

Production metrics

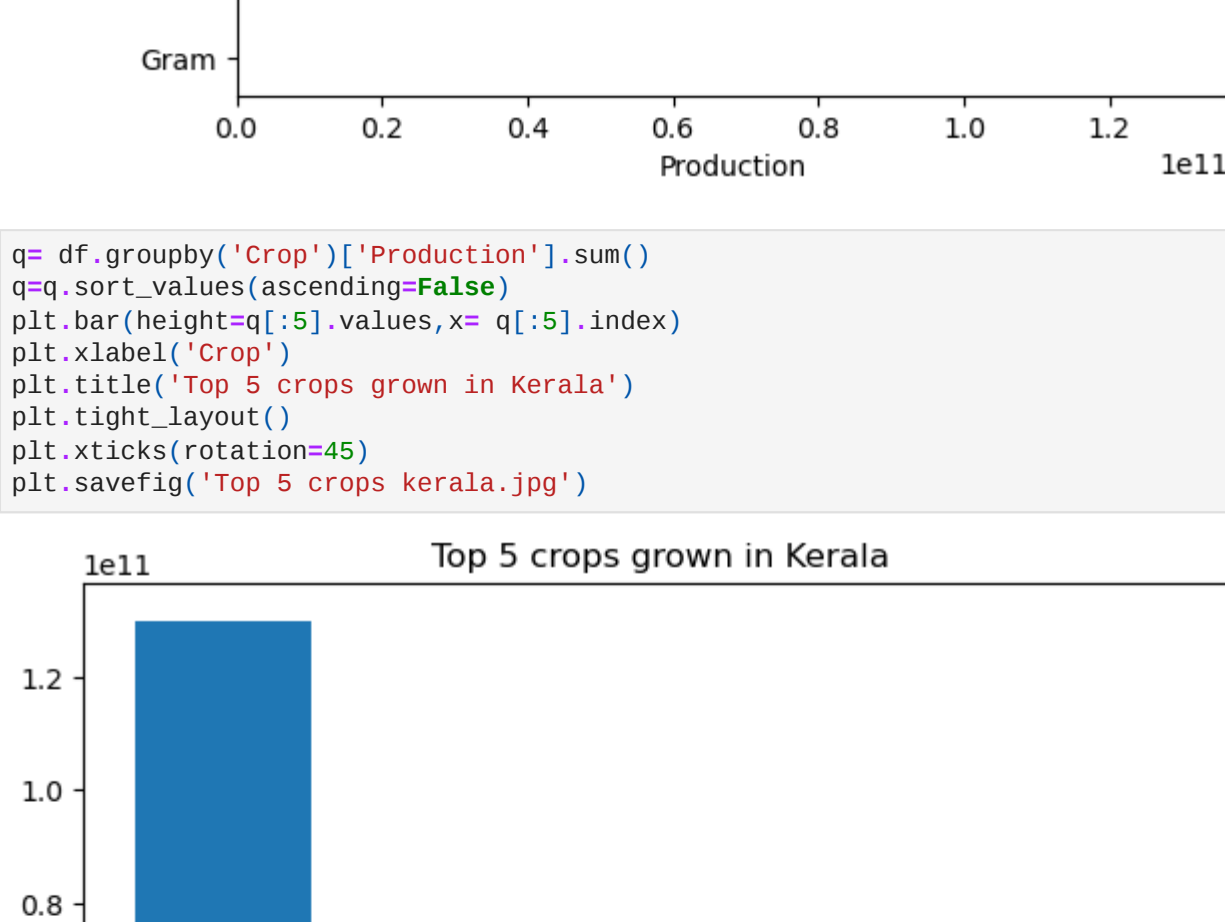
```
In [31]: sns.lineplot(x=df['Crop_Year'], y=df['Production'])
plt.xlabel('Crop_Year')
plt.title('Total production in each year')
plt.savefig('prod in yr.jpg')
```



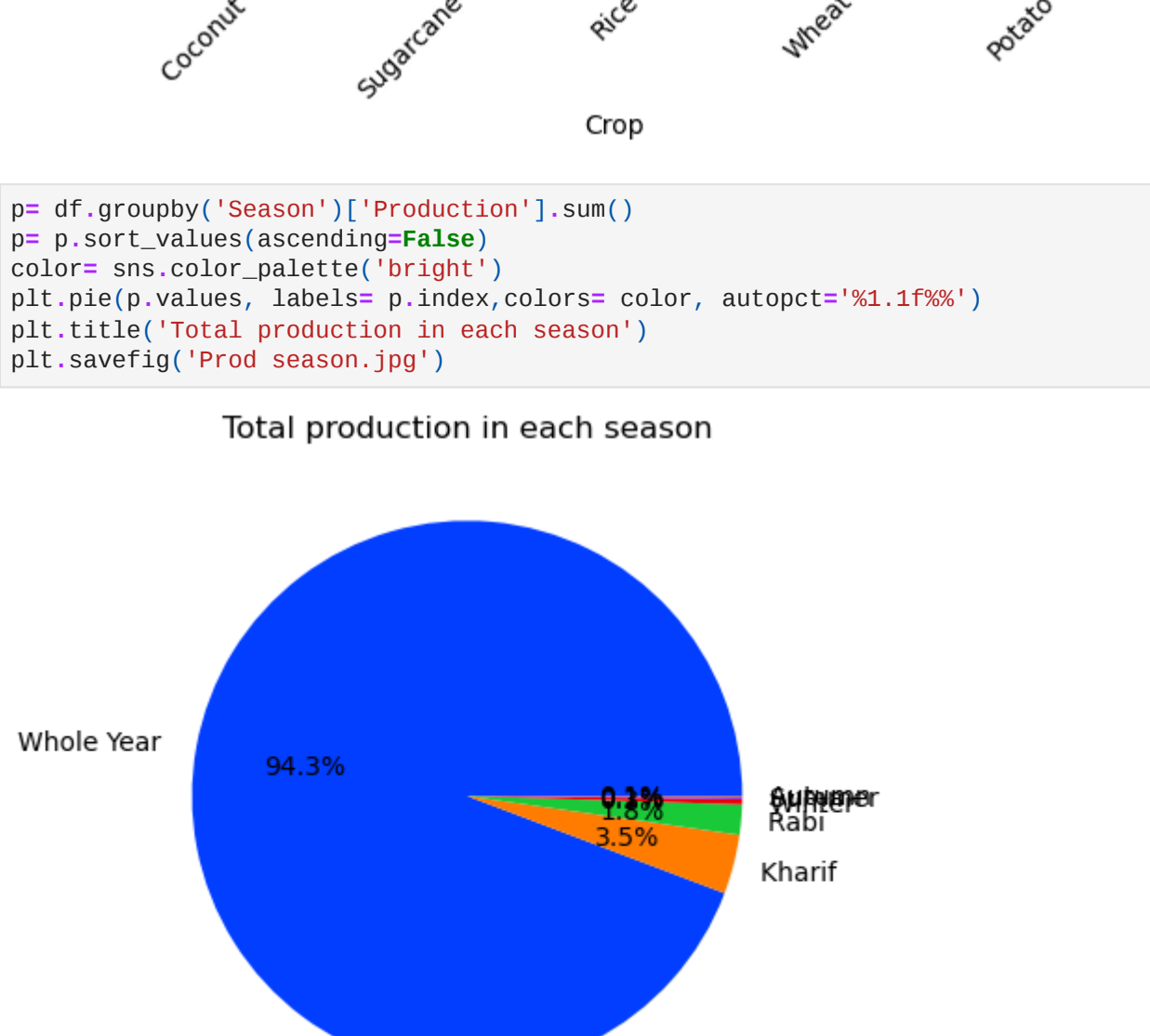
```
In [32]: m=df.groupby('State Name')['Production'].sum()
m=m.sort_values(ascending=False)
plt.xlabel('Production')
plt.title('Total production in each state')
sns.barplot(x=m.values, y=m.index)
plt.savefig('prod state.jpg')
```



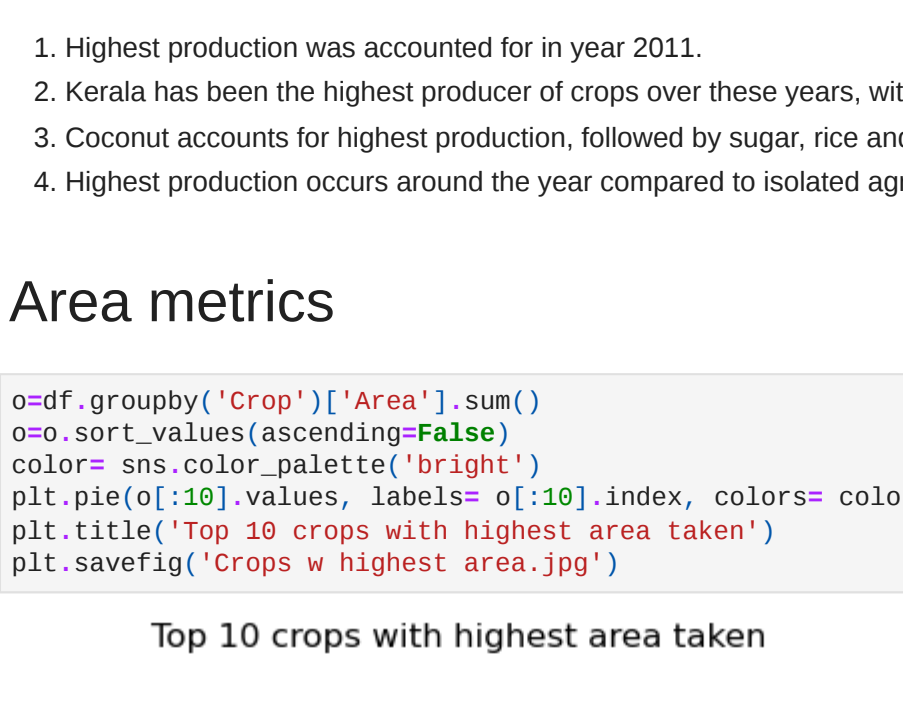
```
In [39]: m=df.groupby('Crop')['Production'].sum()
m=m.sort_values(ascending=False)
plt.xlabel('Production')
plt.title('Top 10 crops with highest production')
sns.barplot(y=m[:10].index, x=m[:10].values)
plt.savefig('prod crop.jpg')
```



```
In [66]: q=df.groupby('Crop')['Production'].sum()
q=q.sort_values(ascending=False)
plt.bar(height=q[:5].values, x=q[:5].index)
plt.xlabel('Crop')
plt.title('Top 5 crops grown in Kerala')
plt.tight_layout()
plt.xticks(rotation=45)
plt.savefig('Top 5 crops kerala.jpg')
```



```
In [48]: p=df.groupby('Season')['Production'].sum()
p=p.sort_values(ascending=False)
colors=sns.color_palette('bright')
plt.pie(p.values, labels=q.index, colors= color, autopct='%1.1f%%')
plt.title('Total production in each season')
plt.savefig('Prod season.jpg')
```



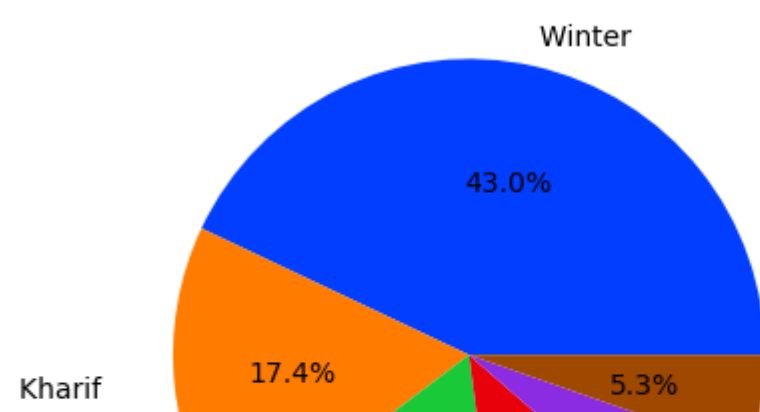
Observations

- Highest production was accounted for in year 2011.
- Kerala has been the highest producer of crops over these years, with states like Andhra Pradesh, Tamil Nadu and Uttar Pradesh being significant contributors. (although at a greater deviation from Kerala)
- Coconut accounts for highest production, followed by sugar, rice and wheat. Coconut is also the staple production in Kerala along with sugar, rice, wheat and potato.
- Highest production occurs around the year compared to isolated agri seasons.

Area metrics

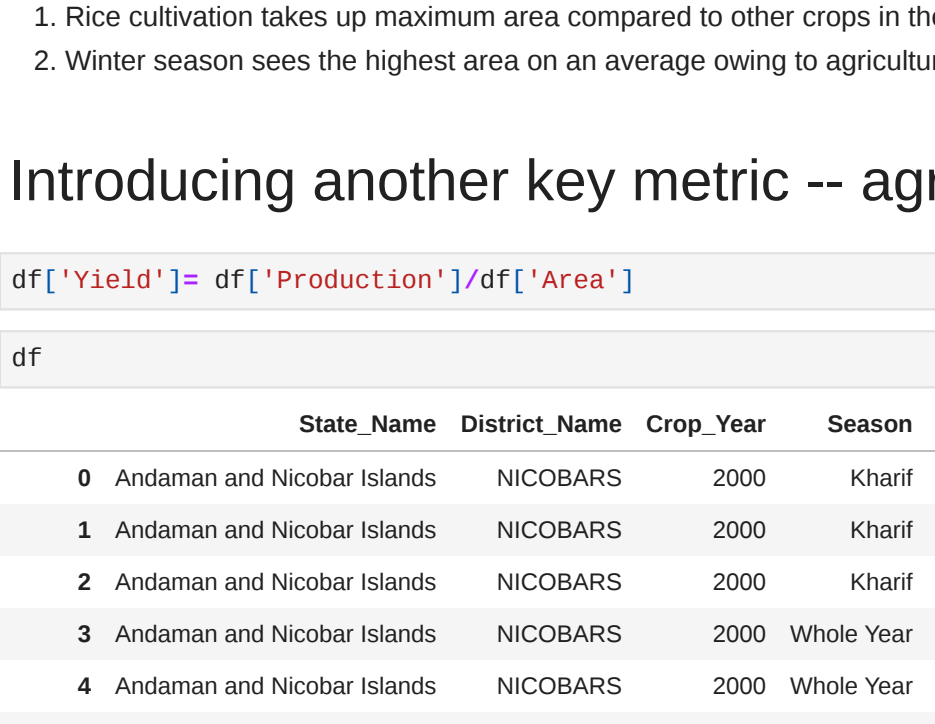
```
In [34]: o=df.groupby('Crop')['Area'].sum()
o=o.sort_values(ascending=False)
colors=sns.color_palette('bright')
plt.pie(o[:10].values, labels=q.index, colors= color, autopct='%1.1f%%')
plt.title('Top 10 crops with highest area taken')
plt.savefig('Crops w highest area.jpg')
```

Top 10 crops with highest area taken



```
In [35]: q=df.groupby('Season')['Area'].mean()
q=q.sort_values(ascending=False)
plt.pie(q.values, labels=q.index, colors= sns.color_palette('bright'), autopct='%1.1f%%')
plt.title('Average area sown in each season')
plt.savefig('Avg area season.jpg')
```

Average area sown in each season



Observations

- Rice cultivation takes up maximum area compared to other crops in the entire country.
- Winter season sees the highest area on an average owing to agriculture.

Introducing another key metric -- agricultural yield to the dataframe

```
In [48]: df['Yield']=df['Production']/df['Area'] # calculated agr. yield and added that as a new column.
```

```
In [47]: df
```

```
Out[47]:
```

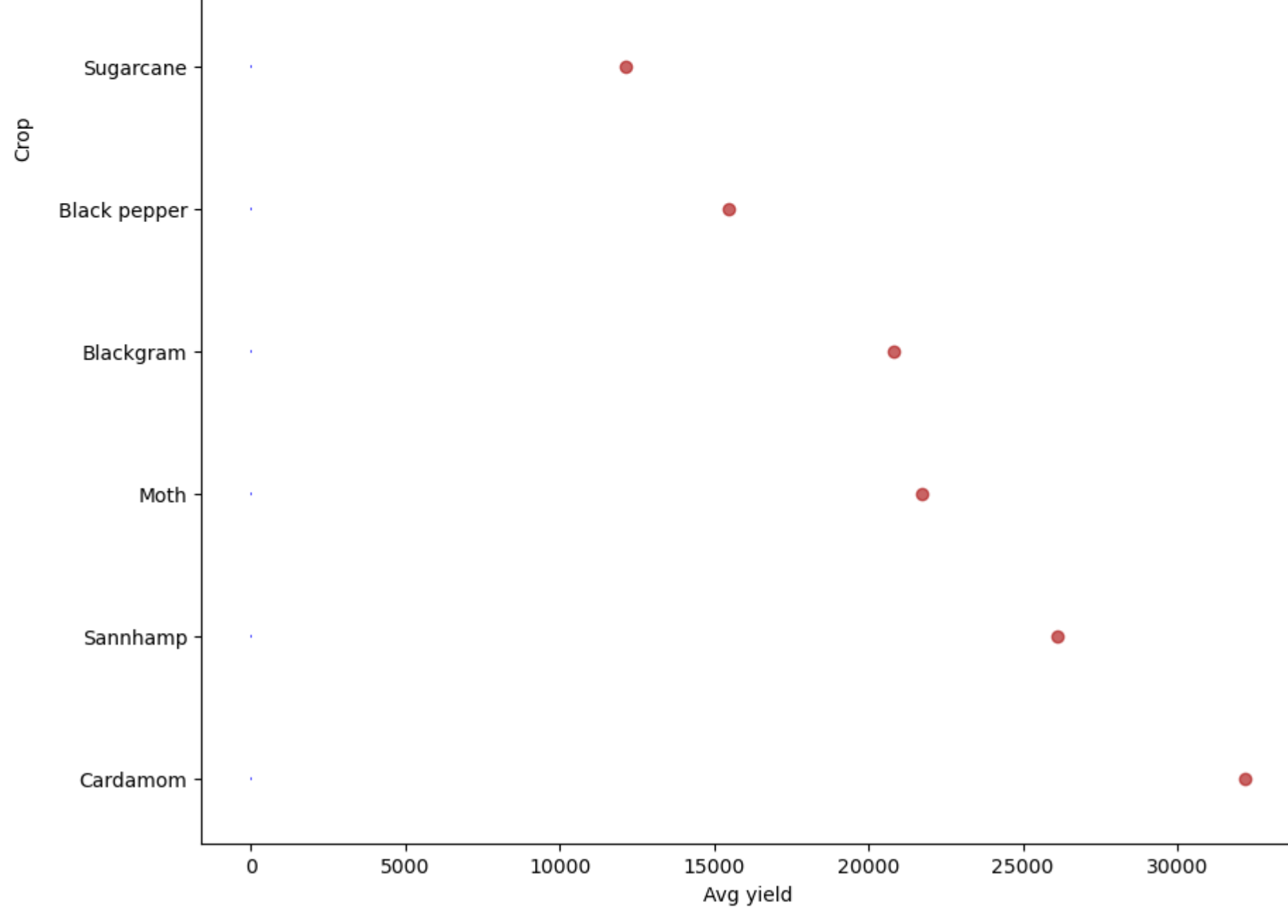
	State Name	District Name	Crop_Year	Season	Crop	Area	Production	Yield
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Areacanut	1254.0	2000.0	1.594896
1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2.0	1.0	0.500000
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102.0	321.0	3.147059
3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176.0	641.0	3.642045
4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720.0	165.0	0.229167
...
246086	West Bengal	PURULIA	2014	Summer	Rice	306.0	801.0	2.617647
246087	West Bengal	PURULIA	2014	Summer	Sesamum	627.0	463.0	0.738437
246088	West Bengal	PURULIA	2014	Whole Year	Sugarcane	324.0	16250.0	50.154321
246089	West Bengal	PURULIA	2014	Winter	Rice	279151.0	597899.0	2.141848
246090	West Bengal	PURULIA	2014	Winter	Sesamum	175.0	88.0	0.502857

246091 rows x 8 columns

```
In [65]: r=df.groupby('Crop')['Yield'].mean()
r=r.sort_values(ascending=False)
fig, ax=plt.subplots(figsize=(10,13))
ax.lines(y=r[:10].index, linestyle='dashed',xmin=0, xmax=10, color= 'blue', alpha= 0.7)
ax.scatter(x=r[:10].values, y=r[:10].index, alpha=0.7, color='firebrick')
```

```
ax.set_xlabel('Avg yield')
ax.set_ylabel('Crop')
ax.set_title('Average yield of crops')
```

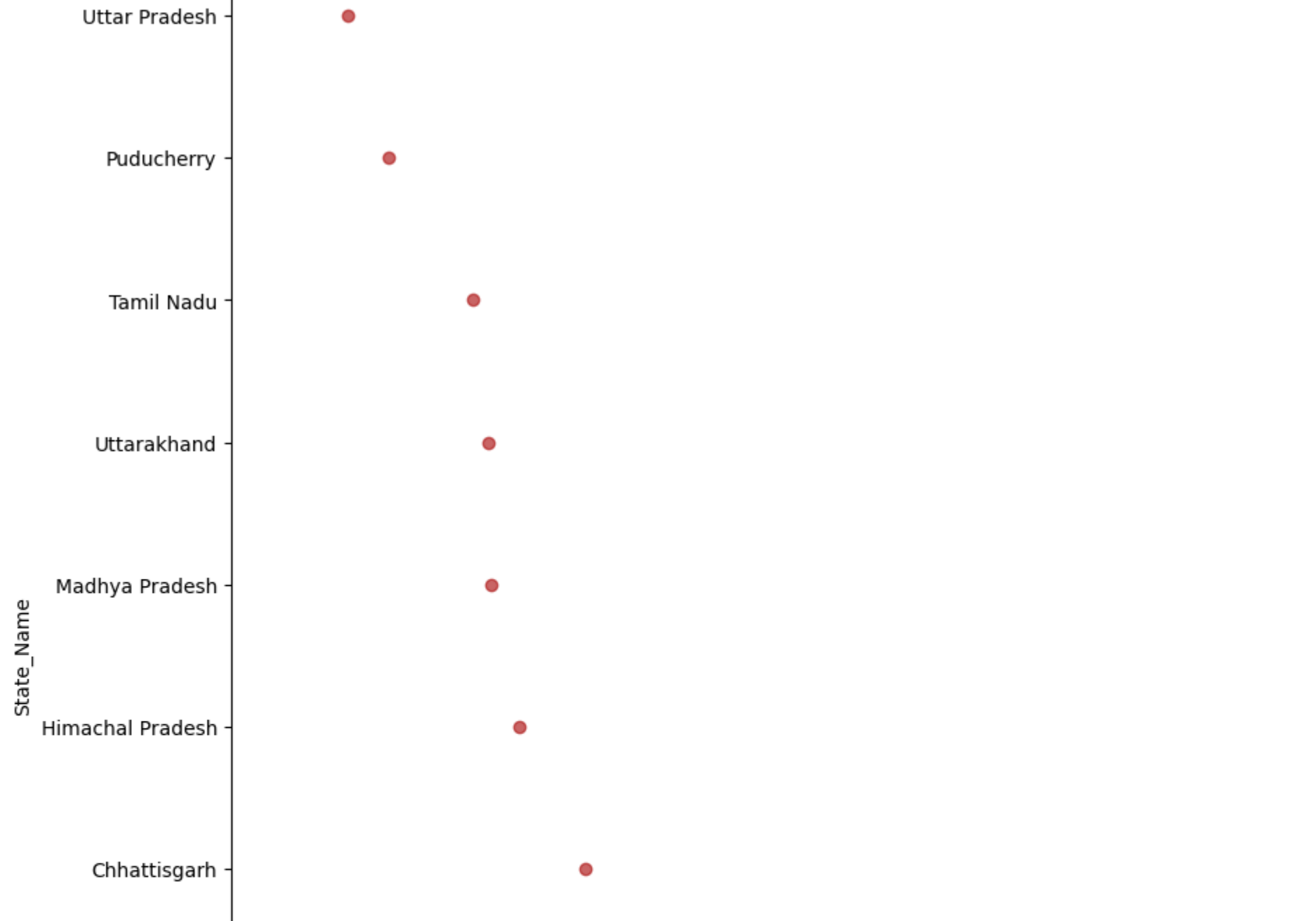
```
Out[65]: Text(0.5, 1.0, 'Average yield of crops')
```



```
In [62]: r=df.groupby('State Name')['Yield'].mean()
r=r.sort_values(ascending=False)
fig, ax=plt.subplots(figsize=(10,13))
ax.lines(y=r[:10].index, linestyle='dashed',xmin=0, xmax=10, color= 'blue', alpha= 0.7)
ax.scatter(x=r[:10].values, y=r[:10].index, alpha=0.7, color='firebrick')
```

```
ax.set_xlabel('Avg yield')
ax.set_ylabel('State Name')
ax.set_title('Average yield of crops in each state')
```

```
Out[62]: Text(0.5, 1.0, 'Average yield of crops in each state')
```



Observations

- Cardamom has the highest crop yield and so, is the most sustainable crop relative to others.
- Kerala has the highest yield and productivity compared to other states and hence, contributes to greater turnover in production with regards to any other state.

Conclusion

- States have greater variation among one another as far as production units are concerned. Southern states have the highest production.
- States like Kerala, Haryana seem to use sustainable sowing and cultivation practices owing to the fact that they have impressive agri.yields.
- Crops like cardamom, sannhap, moth seem to be ecologically sustainable and rewarding crops using lesser area and yielding greater returns.