

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
data = pd.read_csv('Crop Production data.csv')
data
```

		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
	3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176.0	641.0
	4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720.0	165.0
	...	...	...	...	...	...	...	...
	153172	Odisha	KALAHANDI	2000	Whole Year	Sugarcane	1141.0	50345.5
	153173	Odisha	KALAHANDI	2000	Whole Year	Sweet potato	1010.0	8520.0
	153174	Odisha	KALAHANDI	2000	Whole Year	Tobacco	430.0	110.0
	153175	Odisha	KALAHANDI	2000	Whole Year	Turmeric	110.0	250.0
	153176	Odisha	KALAHANDI	2000	Winter	NaN	NaN	NaN

153177 rows × 7 columns

```
data.isnull().sum()
```

State\_Name0  
District\_Name0  
Crop\_Year0  
Season0  
Crop1  
Area1  
Production2678  
dtype: int64

```
data=data.dropna()
data
```

		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
	3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176.0	641.0
	4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720.0	165.0
	...	...	...	...	...	...	...	...
	153171	Odisha	KALAHANDI	2000	Whole Year	Sannhamp	490.0	2230.0
	153172	Odisha	KALAHANDI	2000	Whole Year	Sugarcane	1141.0	50345.5
	153173	Odisha	KALAHANDI	2000	Whole Year	Sweet potato	1010.0	8520.0
	153174	Odisha	KALAHANDI	2000	Whole Year	Tobacco	430.0	110.0
	153175	Odisha	KALAHANDI	2000	Whole Year	Turmeric	110.0	250.0

150499 rows × 7 columns

```
data.isnull().sum()
```

State\_Name0  
District\_Name0  
Crop\_Year0  
Season0  
Crop0  
Area0  
Production0  
dtype: int64

```
import pandas as pd
df=pd.DataFrame(data)
df
```

	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
3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176.0	641.0
4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720.0	165.0
...	...	...	...	...	...	...	...
153171	Odisha	KALAHANDI	2000	Whole Year	Sannhamp	490.0	2230.0
153172	Odisha	KALAHANDI	2000	Whole Year	Sugarcane	1141.0	50345.5
153173	Odisha	KALAHANDI	2000	Whole Year	Sweet potato	1010.0	8520.0
153174	Odisha	KALAHANDI	2000	Whole Year	Tobacco	430.0	110.0
153175	Odisha	KALAHANDI	2000	Whole Year	Turmeric	110.0	250.0

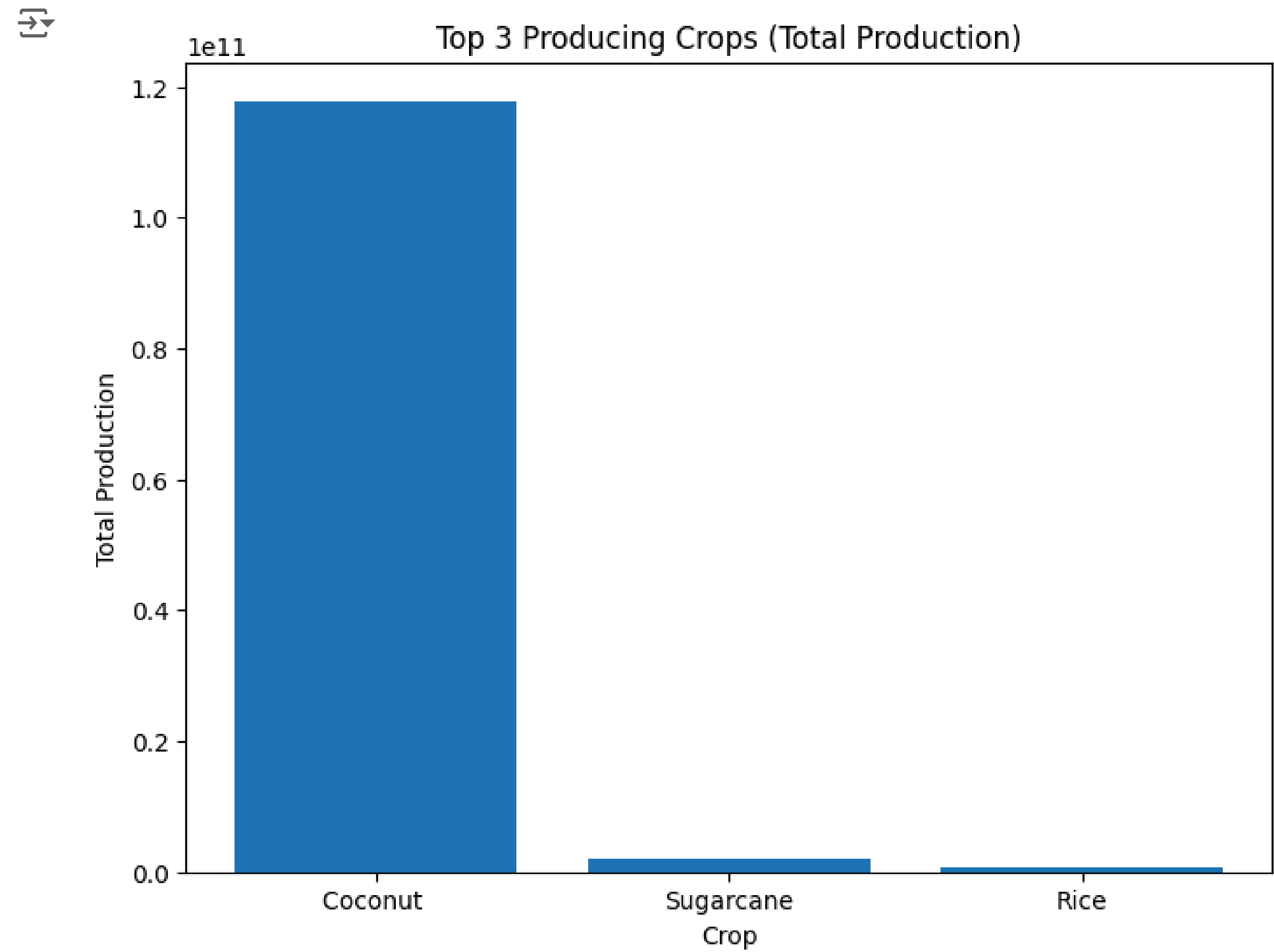
150499 rows × 7 columns

```
import pandas as pd
import matplotlib.pyplot as plt

# Calculate Yield (Production / Area)
df['Yield'] = df['Production'] / df['Area']

# Top 3 Producing Crops (by Total Production)
top_crops = df.groupby('Crop')['Production'].sum().sort_values(ascending=False).head(3)

plt.figure(figsize=(8, 6)) # Adjust figure size for better readability
plt.bar(top_crops.index, top_crops.values)
plt.xlabel('Crop')
plt.ylabel('Total Production')
plt.title('Top 3 Producing Crops (Total Production)')
plt.show()
```

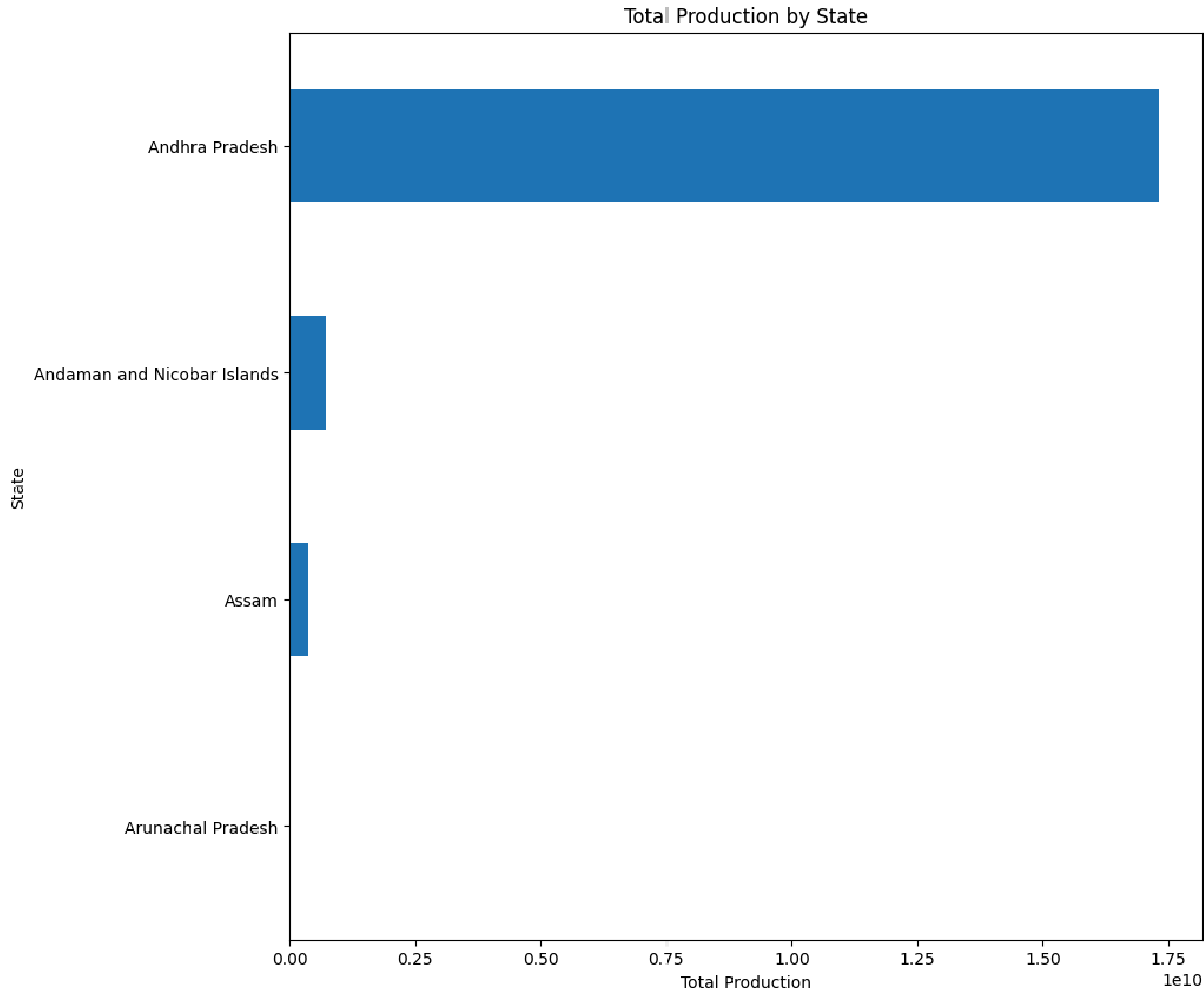


Start coding or [generate](#) with AI.

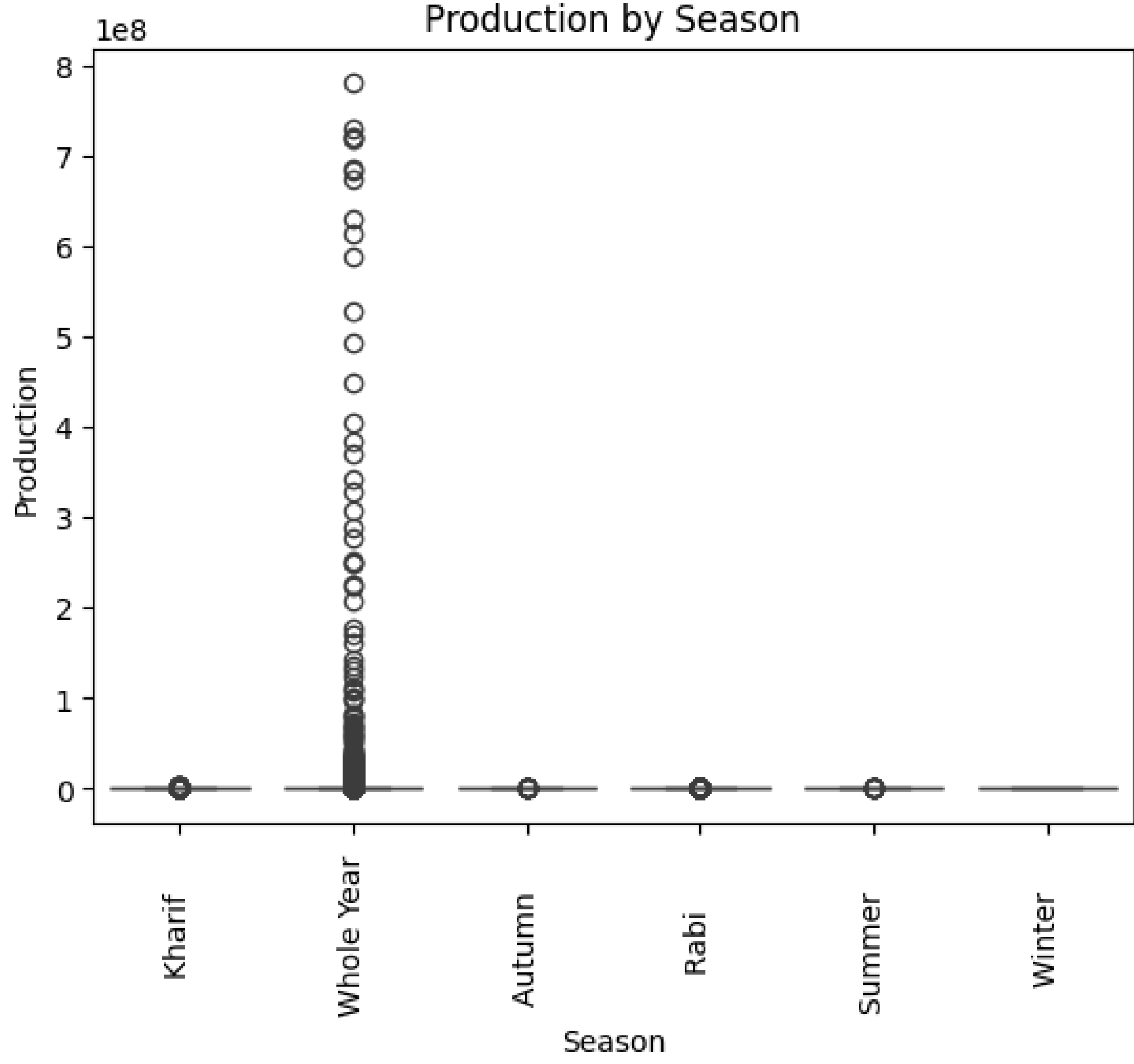
```
# Crop production over the years
df.groupby('Crop_Year')['Production'].sum().plot(kind='line')
plt.title('Crop Production Over the Years')
plt.xlabel('Year')
plt.ylabel('Total Production')
plt.show()
```



```
# Production by state
df.groupby('State_Name')['Production'].sum().sort_values().plot(kind='barh', figsize=(10, 10))
plt.title('Total Production by State')
plt.xlabel('Total Production')
plt.ylabel('State')
plt.show()
```



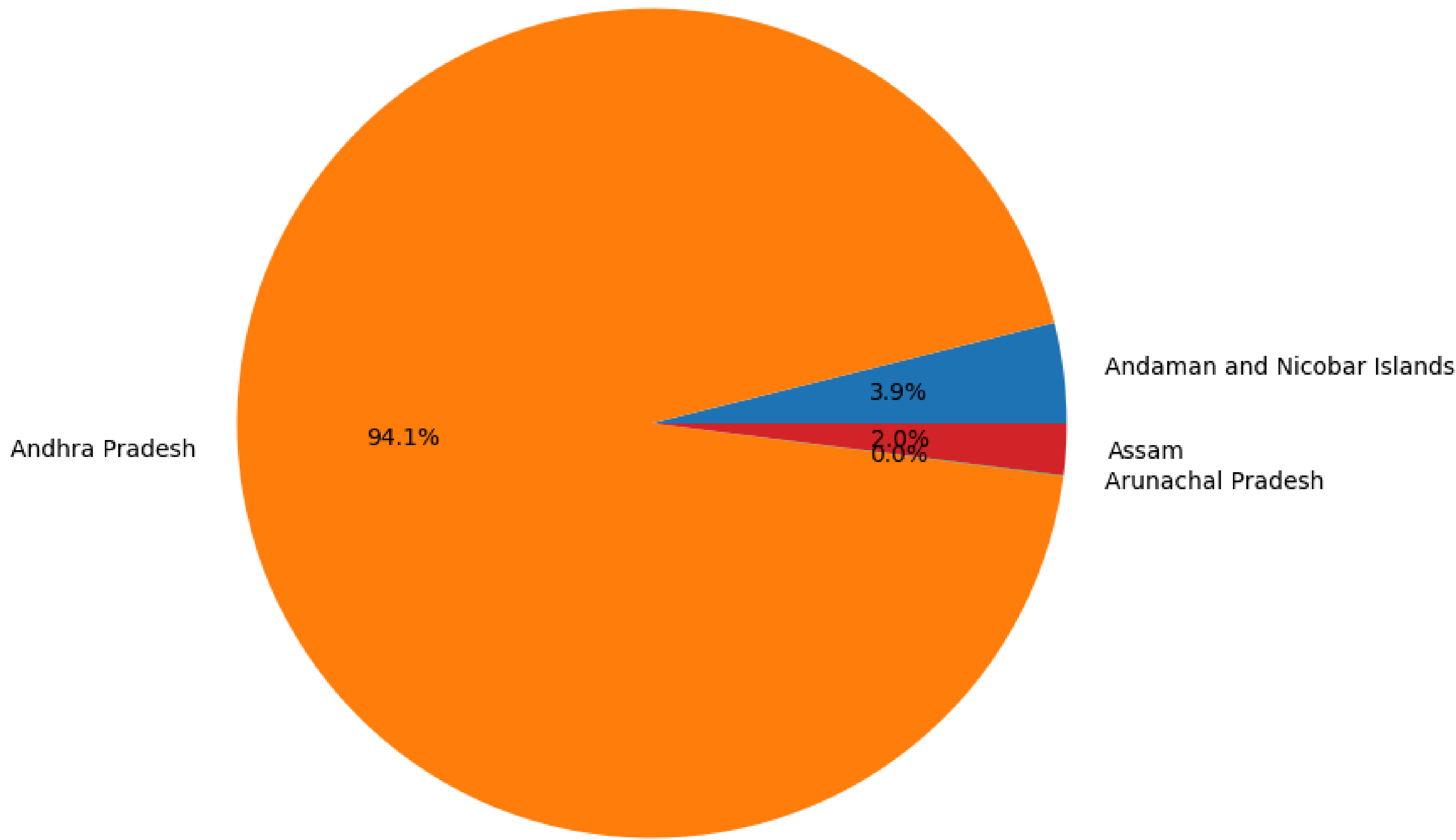
```
# Box plot: Production by Season
sns.boxplot(x='Season', y='Production', data=df)
plt.title('Production by Season')
plt.xlabel('Season')
plt.ylabel('Production')
plt.xticks(rotation=90)
plt.show()
```



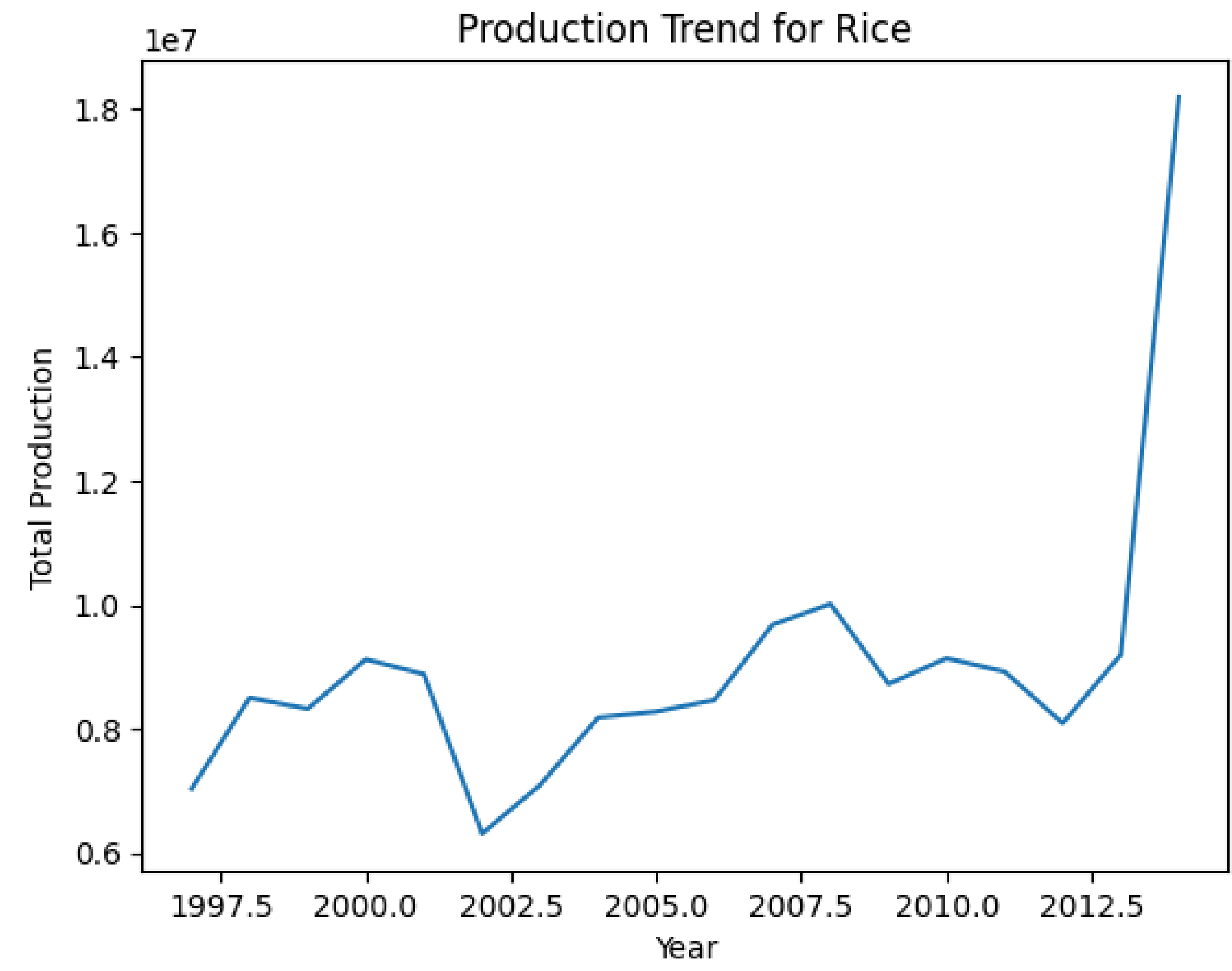
```
# Pie chart: Production share by state
state_prod = df.groupby('State_Name')['Production'].sum()
state_prod.plot(kind='pie', autopct='%1.1f%%', figsize=(8, 8))
plt.title('Production Share by State')
plt.ylabel('')
plt.show()
```



Production Share by State



```
# Time series analysis for a specific crop (e.g., 'Rice')
crop_name = 'Rice'
df_crop = df[df['Crop'] == crop_name]
df_crop.groupby('Crop_Year')['Production'].sum().plot(kind='line')
plt.title(f'Production Trend for {crop_name}')
plt.xlabel('Year')
plt.ylabel('Total Production')
plt.show()
```

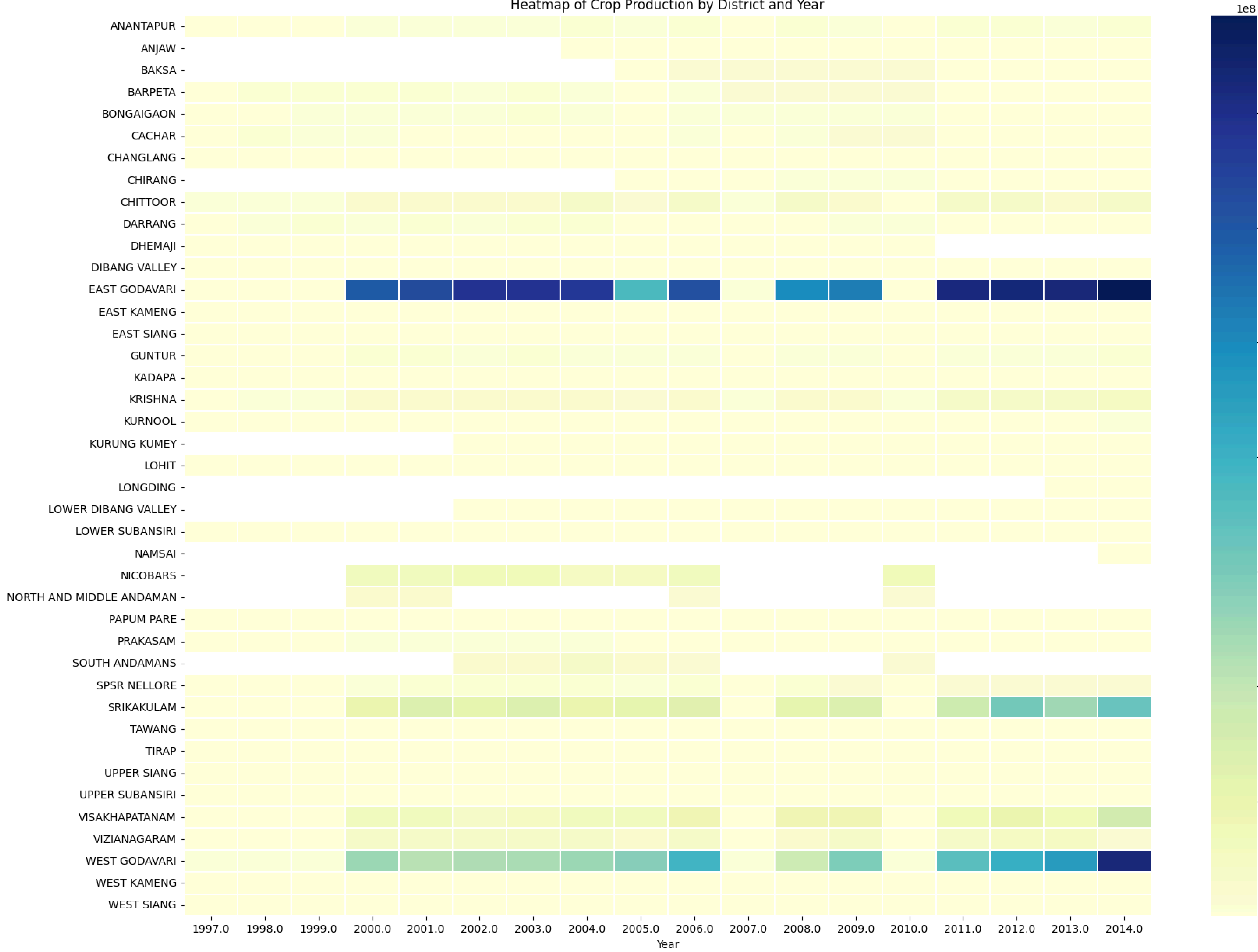


```
# Heatmap: Crop production by district and year
pivot_table = df.pivot_table(values='Production', index='District_Name', columns='Crop_Year', aggfunc='sum')
plt.figure(figsize=(20, 15))
sns.heatmap(pivot_table, cmap='YlGnBu', linecolor='white', linewidths=0.1)
plt.title('Heatmap of Crop Production by District and Year')
plt.xlabel('Year')
plt.ylabel('District')
plt.show()
n
```



District

Heatmap of Crop Production by District and Year



1e8

7

6

5

4

3

2

1



```
import pandas as pd
import matplotlib.pyplot as plt

# Group by Crop_Year and Crop, then sum the Production
year_crop_prod = df.groupby(['Crop_Year', 'Crop'])['Production'].sum().reset_index()

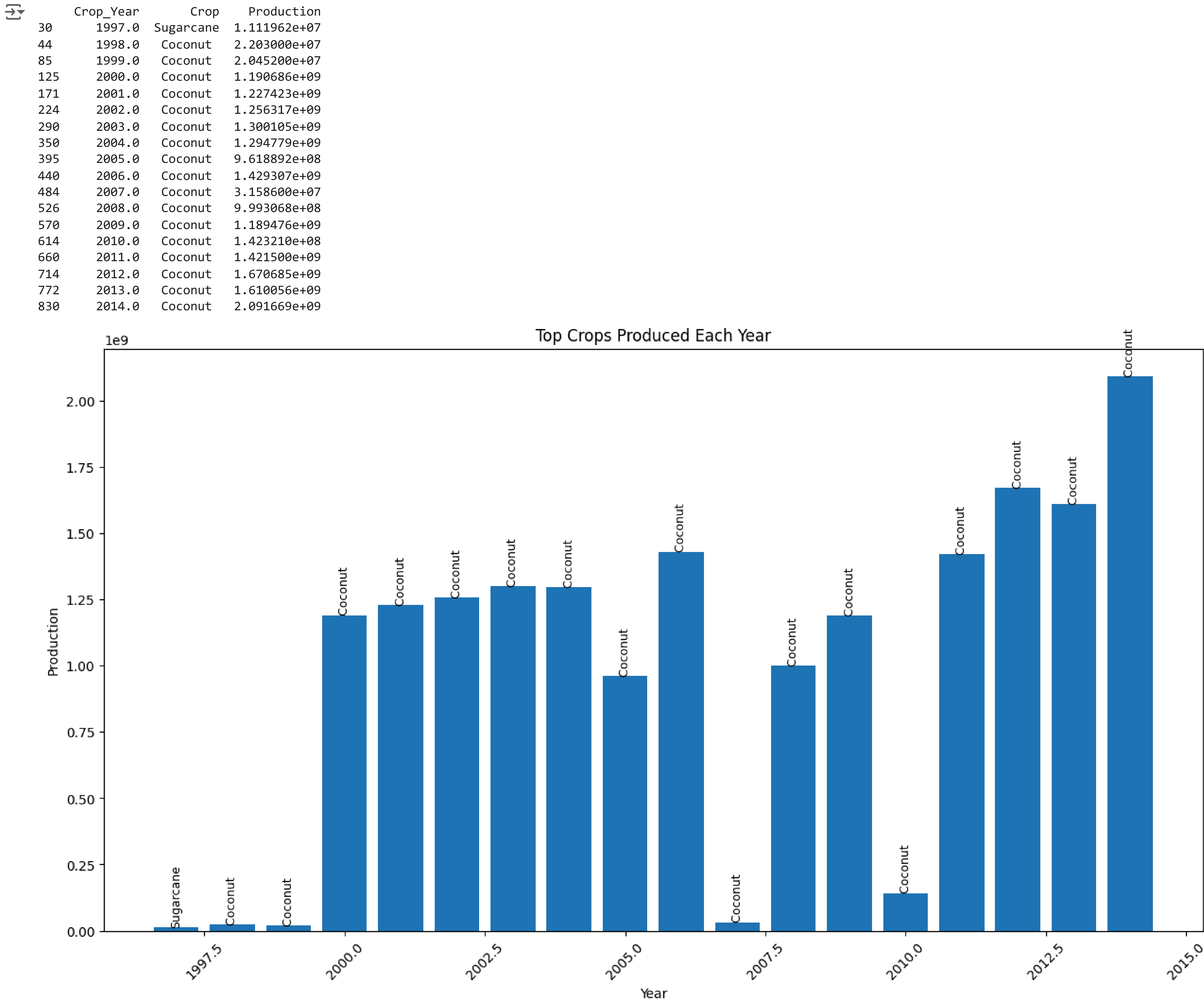
# For each year, find the crop with the highest production
top_crops_each_year = year_crop_prod.loc[year_crop_prod.groupby('Crop_Year')['Production'].idxmax()]

# Print the top crops for each year
print(top_crops_each_year)

# Plotting
plt.figure(figsize=(15, 8))
bars = plt.bar(top_crops_each_year['Crop_Year'], top_crops_each_year['Production'])

# Annotate the bars with the crop names
for bar, crop in zip(bars, top_crops_each_year['Crop']):
    yval = bar.get_height()
    plt.text(bar.get_x() + bar.get_width()/2, yval, crop, ha='center', va='bottom', fontsize=9, rotation=90)

plt.xlabel('Year')
plt.ylabel('Production')
plt.title('Top Crops Produced Each Year')
plt.xticks(rotation=45)
plt.show()
```



```
import pandas as pd
import matplotlib.pyplot as plt

# Filter for Andhra Pradesh
ap_data = df[df['State_Name'] == 'Andhra Pradesh']

# Group by District_Name and sum the Production
district_prod = ap_data.groupby('District_Name')['Production'].sum().reset_index()

# Identify the district with the highest production
highest_prod_district = district_prod.loc[district_prod['Production'].idxmax()]

# Identify the district with the lowest production
lowest_prod_district = district_prod.loc[district_prod['Production'].idxmin()]


# Print the districts
print("District with the highest production in Andhra Pradesh:")
print(highest_prod_district)
print("\nDistrict with the lowest production in Andhra Pradesh:")
print(lowest_prod_district)

# Visualize the results
plt.figure(figsize=(10, 5))

# Bar plot for highest production district
plt.subplot(1, 2, 1)
plt.bar(highest_prod_district['District_Name'], highest_prod_district['Production'], color='green')
plt.title('Highest Production District')
plt.xlabel('District')
plt.ylabel('Production')

# Bar plot for lowest production district
plt.subplot(1, 2, 2)
plt.bar(lowest_prod_district['District_Name'], lowest_prod_district['Production'], color='red')
plt.title('Lowest Production District')
plt.xlabel('District')
plt.ylabel('Production')

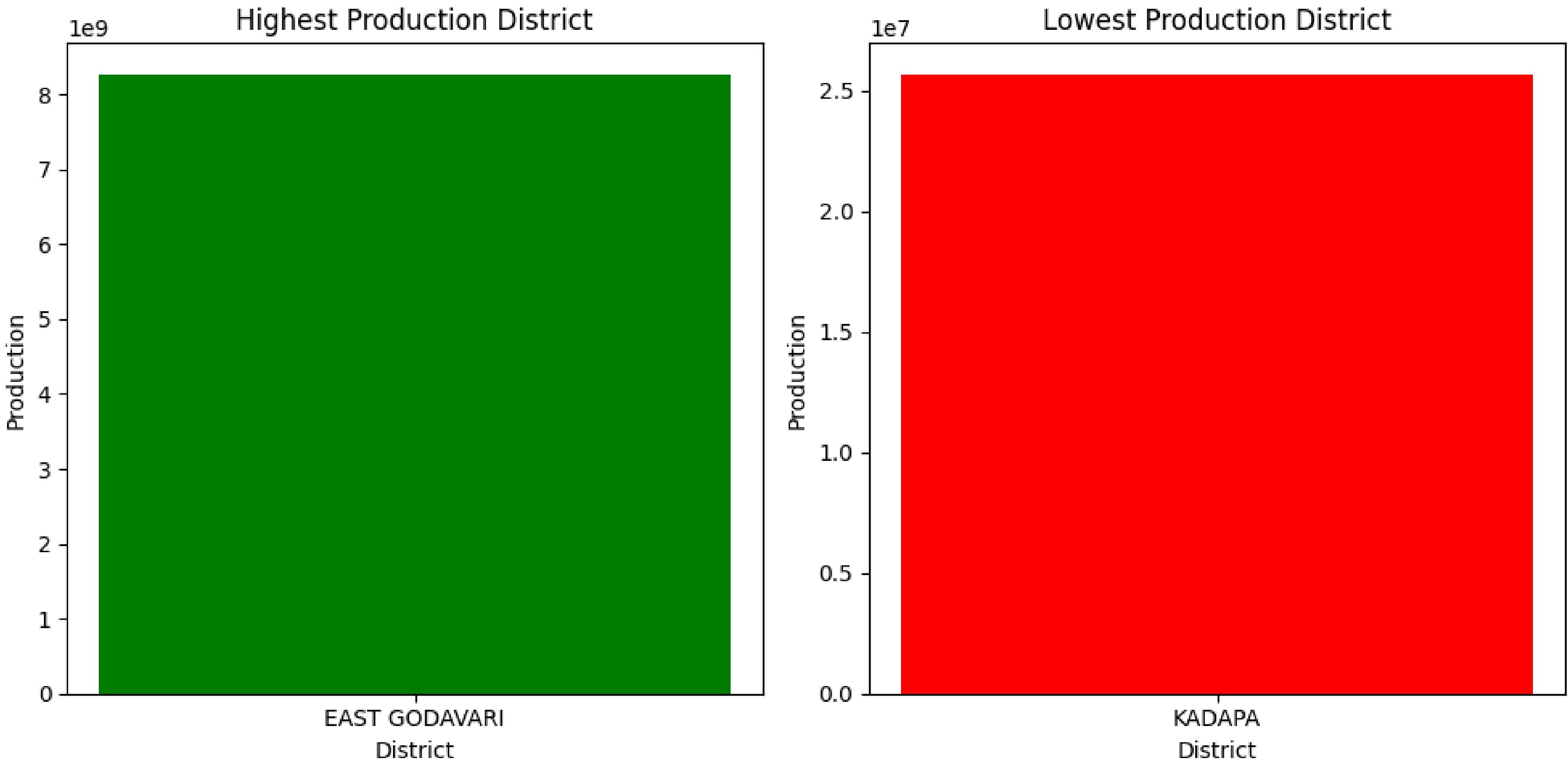
plt.tight_layout()
plt.show()
```

 District with the highest production in Andhra Pradesh:

District_Name	EAST GODAVARI
Production	8271057200.0
Name: 2, dtype: object	

District with the lowest production in Andhra Pradesh:

District_Name	KADAPA
Production	25662234.0
Name: 4, dtype: object	



```
import pandas as pd
import matplotlib.pyplot as plt

# Group by State_Name and District_Name and sum the Area
state_district_area = df.groupby(['State_Name', 'District_Name'])['Area'].sum().reset_index()

# Identify the state and district with the highest area
```