

TEAM ID :PNT2022TMID12429

Code:

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

raw_data = pd.read_csv('D:\crop_production.csv')

raw_data

raw_data.isnull().sum()

raw_data.shape

#Dropping Null values in production column
data = raw_data.dropna()
data.shape

test_data = raw_data[~raw_data["Production"].notna()].drop("Production",axis=1)
test_data

sum_maxProduction = data["Production"].sum()
sum_maxProduction
```

```
data["percent_of_production"]=data["Production"].map(lambda x:(x/sum_maxProduction)*100)
```

```
data
```

```
data[:5]
```

```
pip install plotly
```

```
import plotly.express as px ## Visualization
```

```
import plotly.graph_objects as go ## Visualization
```

```
import matplotlib.pyplot as plt ## Visualization
```

```
import plotly as py ## Visuaization
```

```
from plotly import tools ## Visualization
```

```
import os
```

```
plt.figure(figsize=(25,10))
```

```
sns.barplot(data["State_Name"],data["Production"])
```

```
plt.xticks(rotation=90)
```

```
sns.lineplot(data["Crop_Year"],data["Production"])
```

```
sns.barplot(data["Season"],data["Production"])
```

```
data.groupby("Season",axis=0).agg( {"Production":np.sum} )
```

```
data["Crop"].value_counts()
```

```
top_crop_production =  
data.groupby("Crop")["Production"].sum().reset_index().sort_values(by='Production',ascending=  
False)
```

```
top_crop_production
```

```
temp = data.groupby(by='State_Name')['Production'].sum().reset_index().sort_values(by='Production')
px.bar(temp, 'State_Name', 'Production')
```

#PRODUCTIVITY OF DIFFERENT STATES

```
temp = data.groupby('State_Name')['Area', 'Production'].sum().reset_index()
temp['Production_Per_Unit_Area'] = temp['Production']/temp['Area']
temp = temp.sort_values(by='Production_Per_Unit_Area')
px.bar(temp, 'State_Name', 'Production_Per_Unit_Area', color='Production_Per_Unit_Area', )
```

```
croptyear_production = data.groupby(by='Crop_Year')['Production'].sum().reset_index()
px.line(croptyear_production, 'Crop_Year', 'Production')
```

```
croptyear_area = data.groupby(by='Crop_Year')['Area'].mean().reset_index()
px.scatter(croptyear_area, 'Crop_Year', 'Area', color='Area', size='Area')
```

```
statename_areaproductio = data.groupby('State_Name')['Area',
'Production'].sum().reset_index()
```

```
statename_areaproductio['Production_Per_Unit_Area'] =
statename_areaproductio['Production']/statename_areaproductio['Area']
```

```
statename_areaproductio =
statename_areaproductio.sort_values(by='Production_Per_Unit_Area')
```

```
px.bar(statename_areaproductio, 'State_Name', 'Production_Per_Unit_Area',
color='Production_Per_Unit_Area')
```

```
crop_production = data.groupby(by='Crop')['Production'].sum().reset_index().sort_values(by='Production')
px.bar(crop_production.tail(50), 'Crop', 'Production')
```

```
crop_production[crop_production['Production']==0]
```

```
import plotly.subplots as ps
```

```
#COCONUT
```

```
coconut = data[data['Crop']=='Coconut ']
```

```
fig = py.subplots.make_subplots(rows=1,cols=2,subplot_titles=('Coconut production in different states', 'Coconut crop area in states'))
```

```
temp = coconut.groupby(by='State_Name')['Production'].sum().reset_index().sort_values(by='Production')
n')
```

```
trace0 = go.Bar(x=temp['State_Name'], y=temp['Production'])
```

```
temp = coconut.groupby(by='State_Name')['Area'].mean().reset_index().sort_values(by='Area')
```

```
trace1 = go.Bar(x=temp['State_Name'], y=temp['Area'])
```

```
fig.append_trace(trace0, 1,1)
```

```
fig.append_trace(trace1, 1,2)
```

```
fig.show()
```

```
fig = ps.make_subplots(rows=1,cols=2,
                        subplot_titles=('Highest crop producing districts', 'Least overall crop producing districts'))
```

```
temp = data.groupby(by='District_Name')['Production'].sum().reset_index().sort_values(by='Production')
)
```

```
temp1 = temp.tail()
```

```
trace1 = go.Bar(x= temp1['District_Name'], y=temp1['Production'])
```

```
temp1=temp.head()
```

```
trace2 = go.Bar(x= temp1['District_Name'], y=temp1['Production'])
```

```
fig.append_trace(trace1,1,1)
```

```
fig.append_trace(trace2,1,2)
```

```
fig.show()
```

```
del temp,temp1
```

```
coconut_production = coconut.groupby(by='Crop_Year')['Production'].sum().reset_index()
```

```
px.line(coconut_production, 'Crop_Year', 'Production', title='Coconut production over the years')
```

```
kerala = data[data['State_Name']=='Tamil Nadu']
```

```
crop_production = kerala.groupby(by='Crop')['Production'].mean().reset_index().sort_values(by='Production').tail(50)
```

```
px.bar(crop_production, 'Crop', 'Production', title = 'Avg. Crop Production')
```

```
kerala = kerala[~(kerala['Crop']=='Coconut ')]
```

```
crop_production = kerala.groupby(by='Crop')['Production'].sum().reset_index().sort_values(by='Production')
```

```
px.bar(crop_production, 'Crop', 'Production', title='AVG. Crop Production excluding coconut')
```

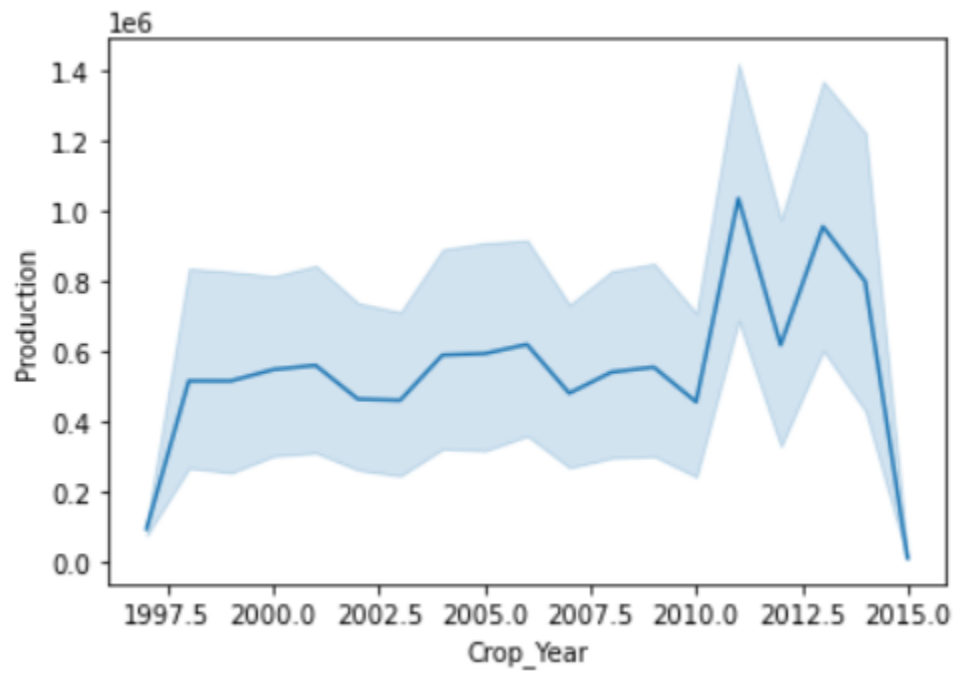
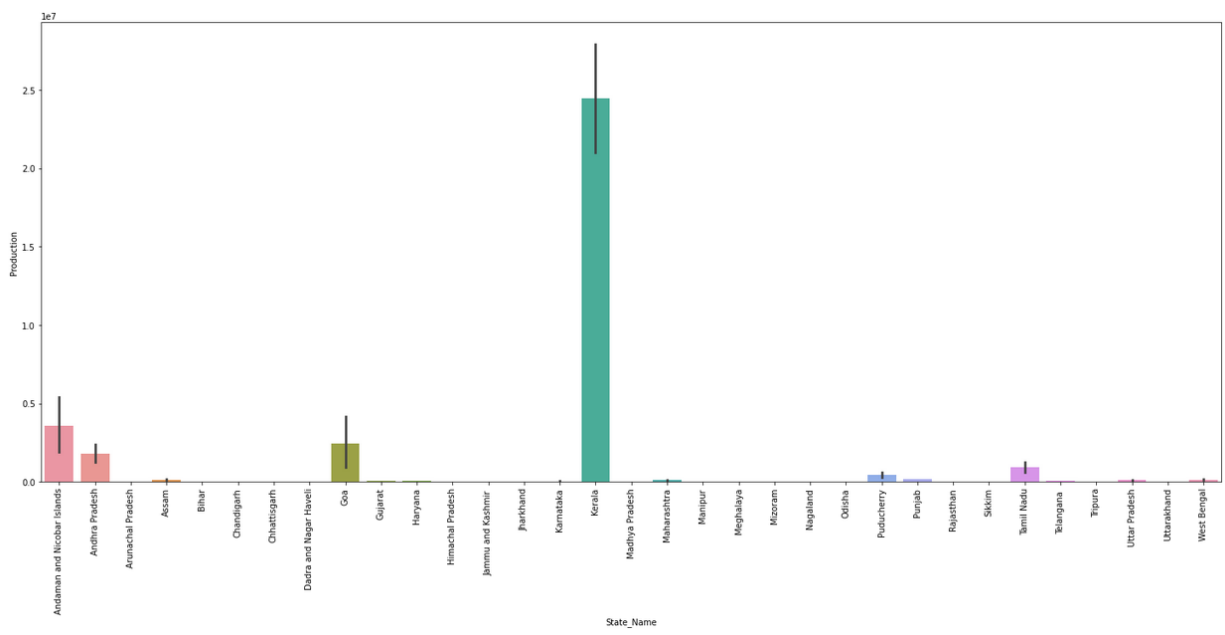
```
data1 = data[~((data['State_Name']=='Kerala') | (data['Crop']=='Coconut '))]
```

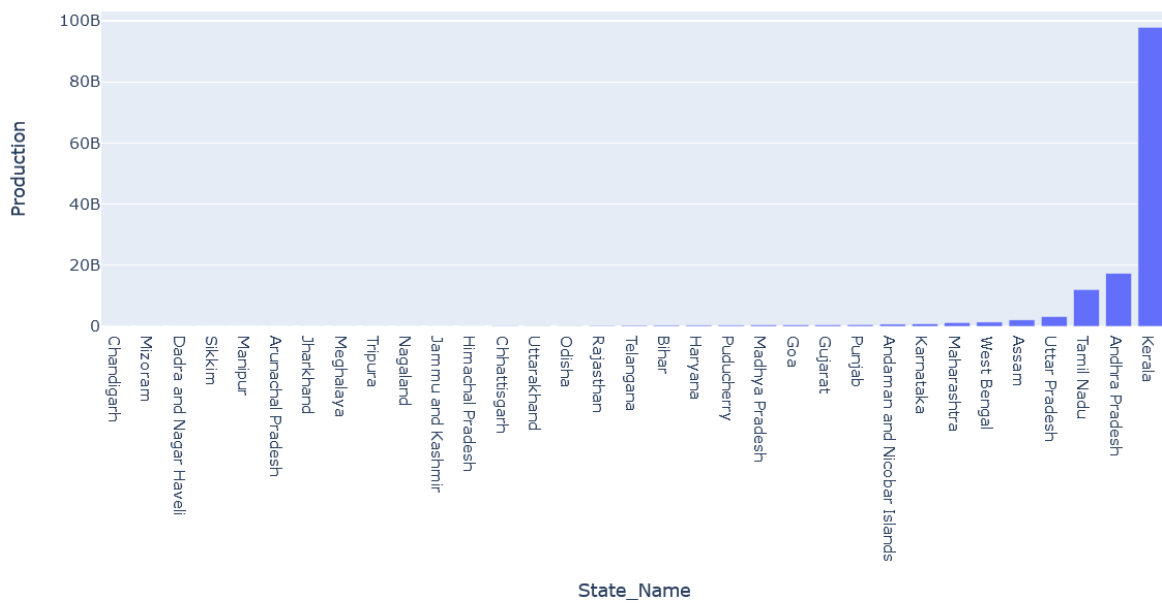
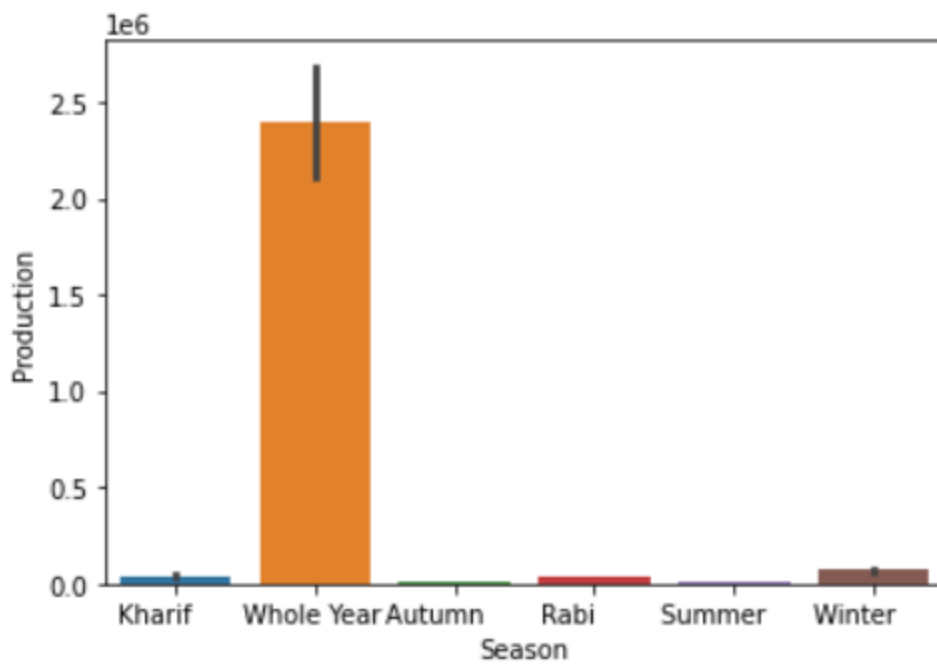
```
data1
```

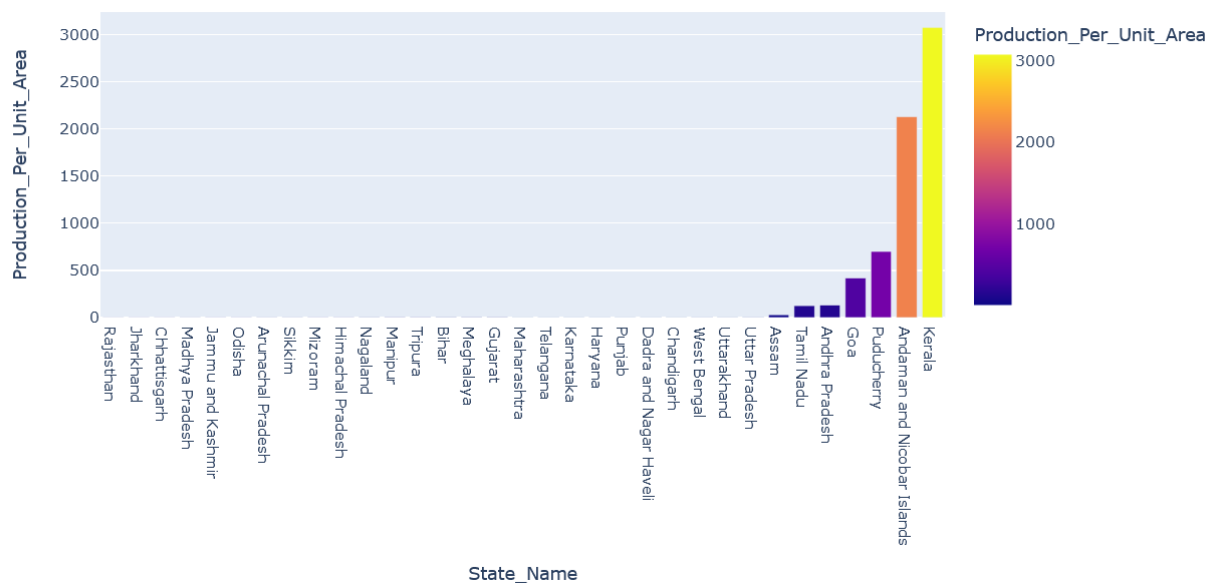
```
temp=data1.groupby('Crop')['Production'].sum().reset_index().sort_values(by='Production').tail(50)
```

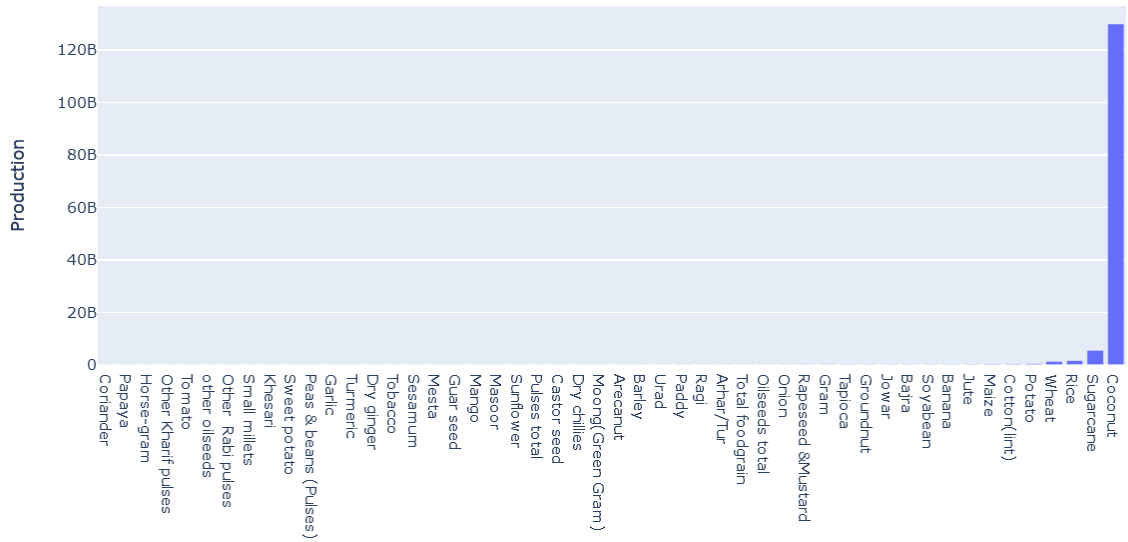
```
px.bar(temp, 'Crop', 'Production', title='Overall production of crops')
```

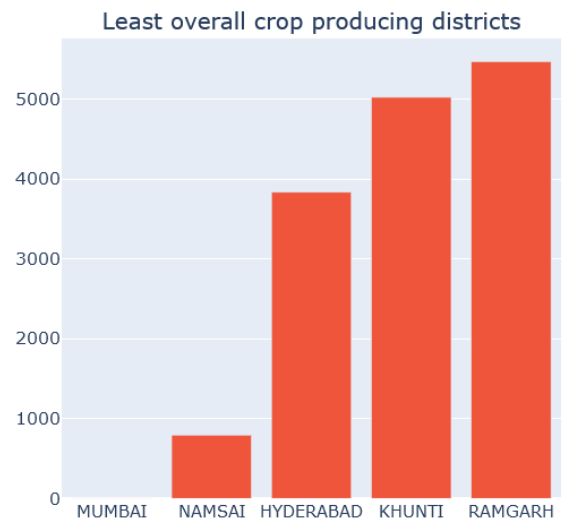
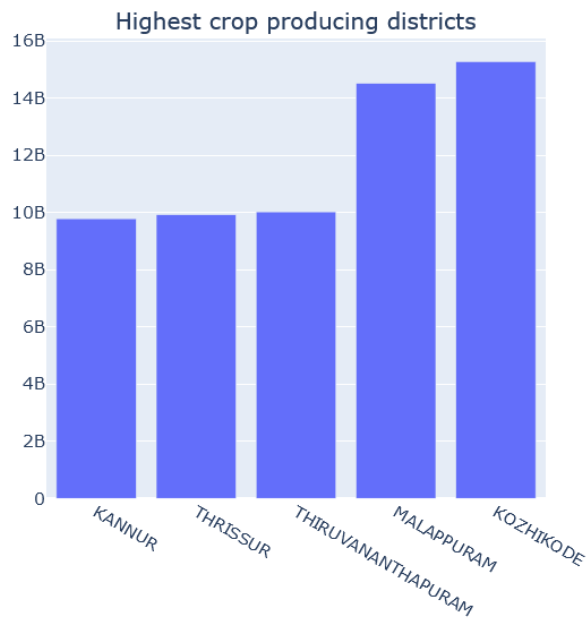
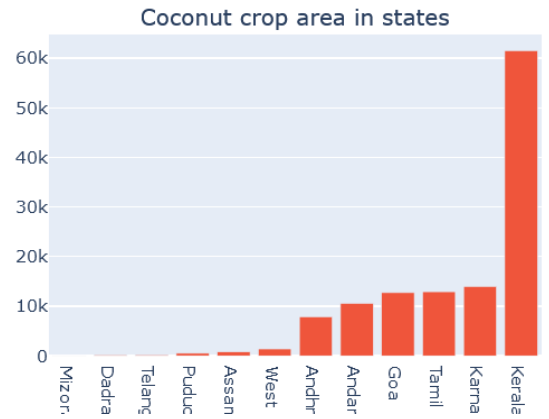
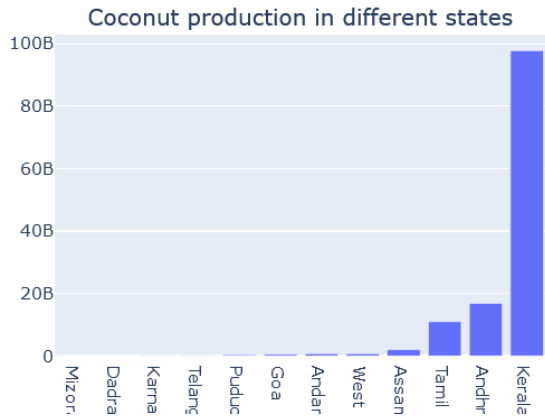
Output:



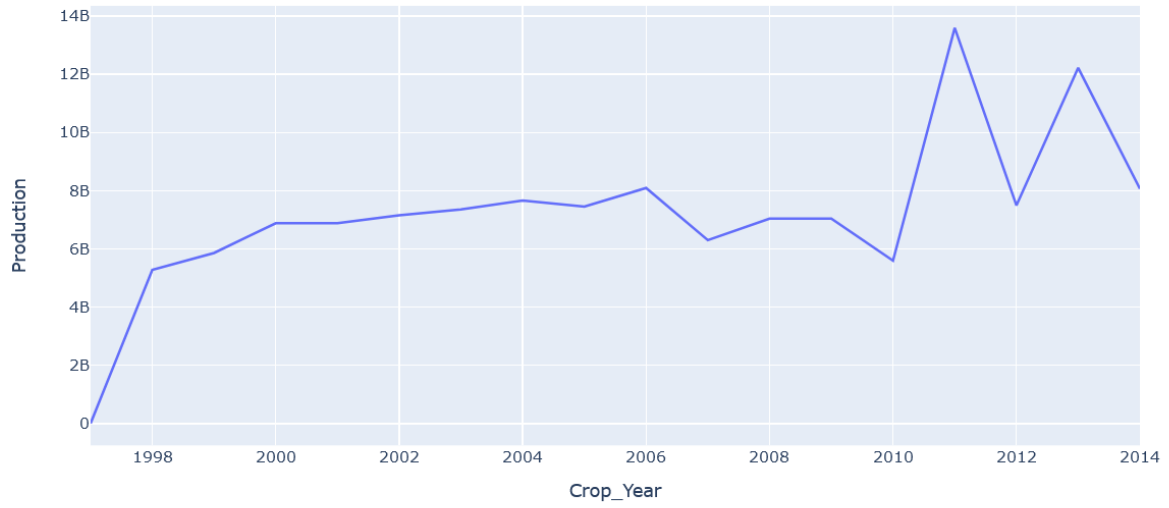




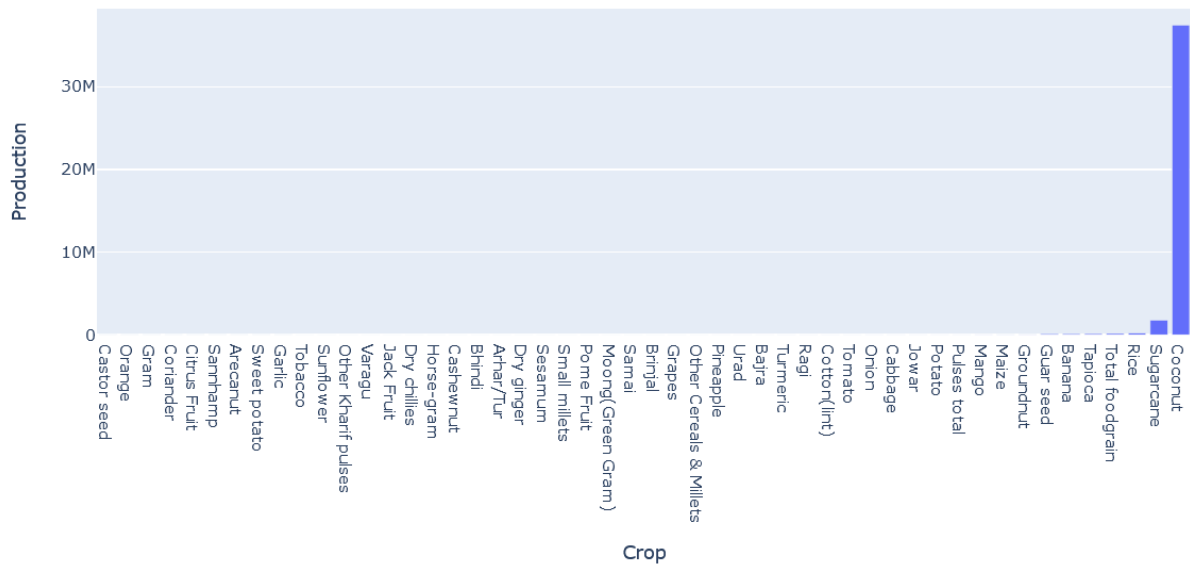




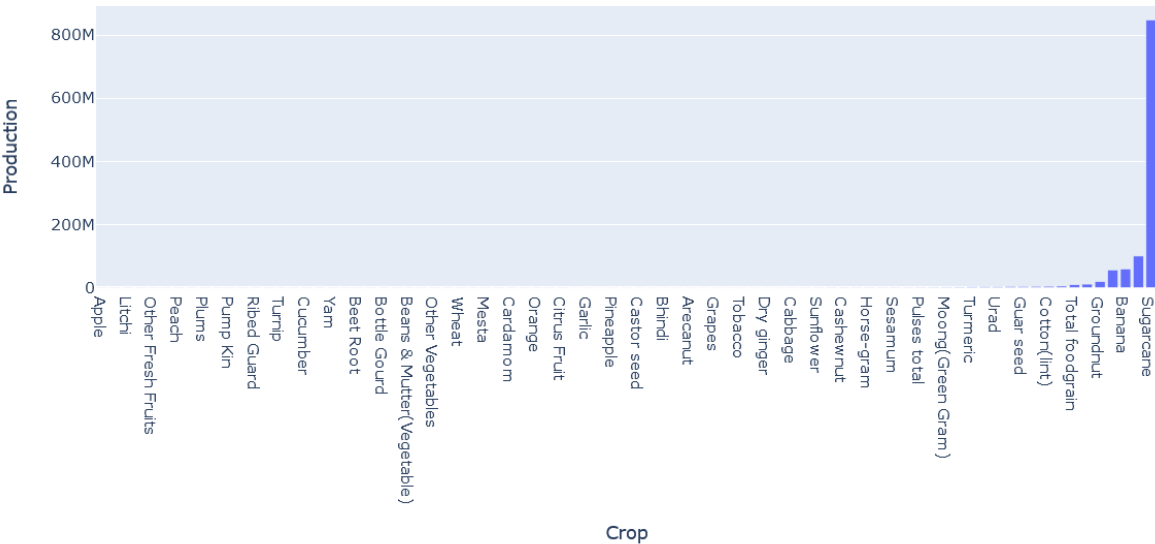
Coconut production over the years



Avg. Crop Production



AVG. Crop Production excluding coconut



Overall production of crops

