**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', )

cropyear\_production = data.groupby(by='Crop\_Year')['Production'].sum().reset\_index()

px.line(cropyear\_production, 'Crop\_Year', 'Production')

cropyear\_area = data.groupby(by='Crop\_Year')['Area'].mean().reset\_index()

px.scatter(cropyear\_area, 'Crop\_Year', 'Area', color='Area', size='Area')

statename\_areaproduction = data.groupby('State\_Name')['Area', 'Production'].sum().reset\_index()

statename\_areaproduction['Production\_Per\_Unit\_Area'] = statename\_areaproduction['Production']/statename\_areaproduction['Area']

statename\_areaproduction = statename\_areaproduction.sort\_values(by='Production\_Per\_Unit\_Area')

px.bar(statename\_areaproduction, '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')

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')

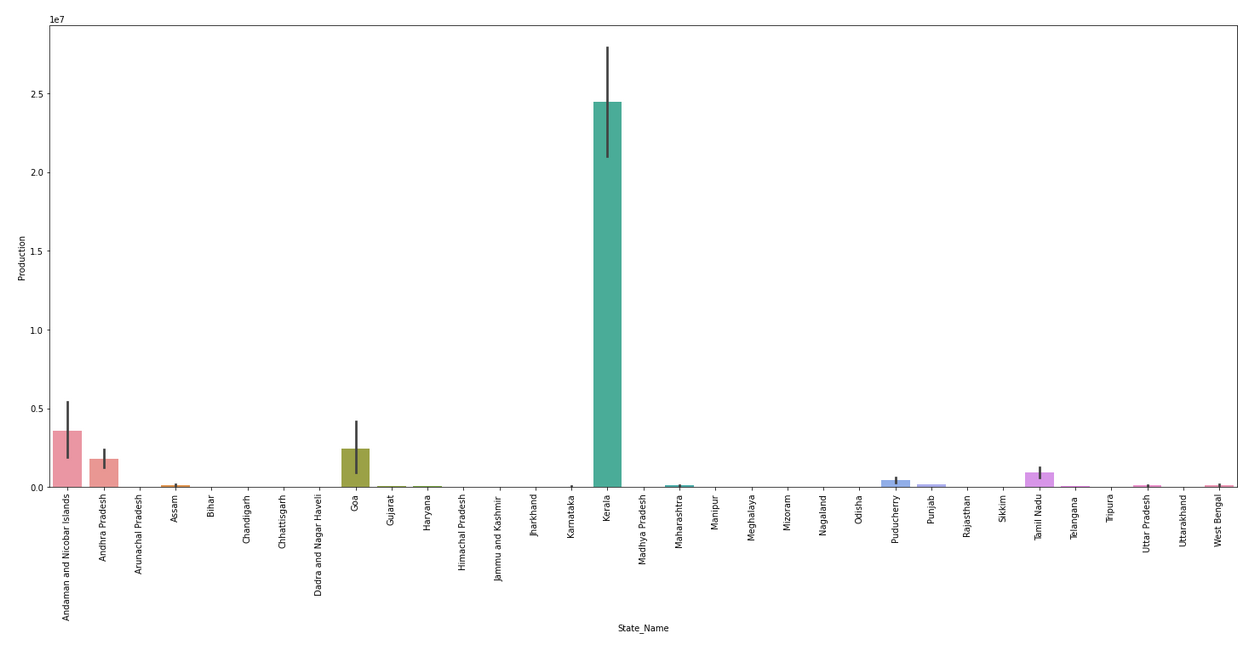
data1 = data[~((data['State\_Name']=='Kerala') | (data['Crop']=='Coconut '))]

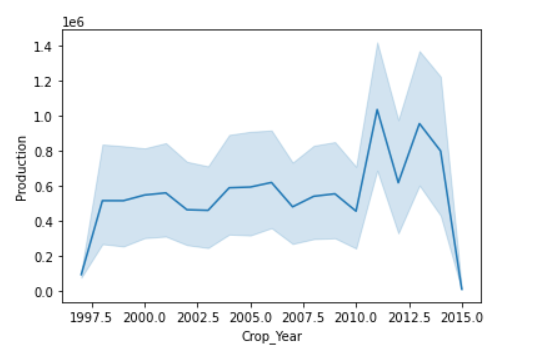
data1

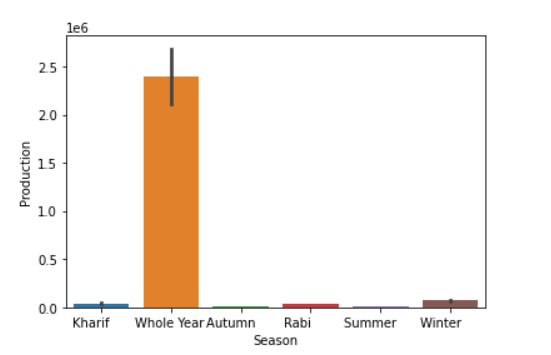
temp=data1.groupby('Crop')['Production'].sum().reset\_index().sort\_values(by='Production').tail(50)

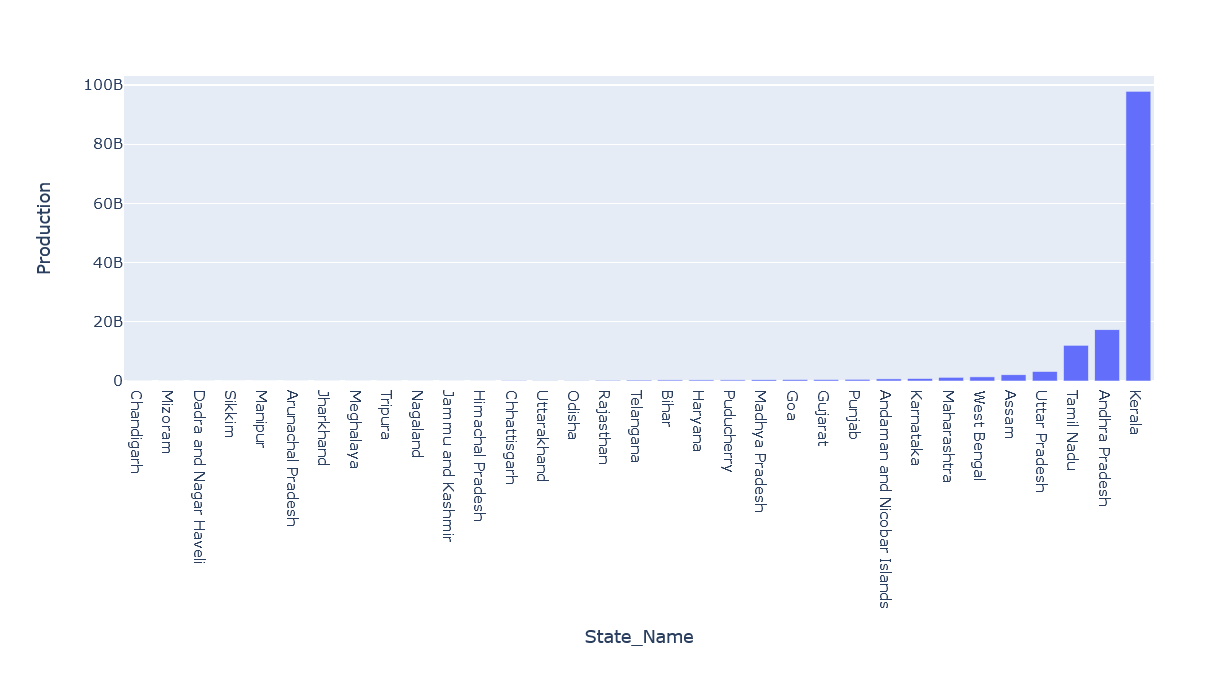
px.bar(temp, 'Crop', 'Production', title='Overall production of crops')

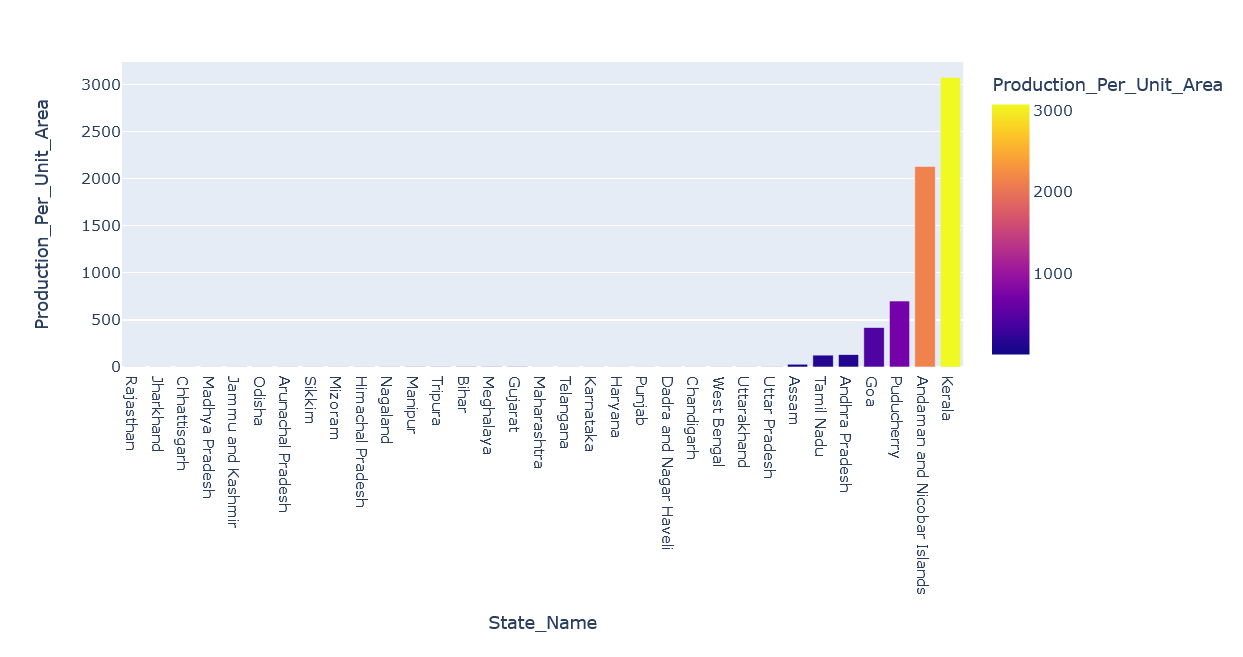
**Output:**

****

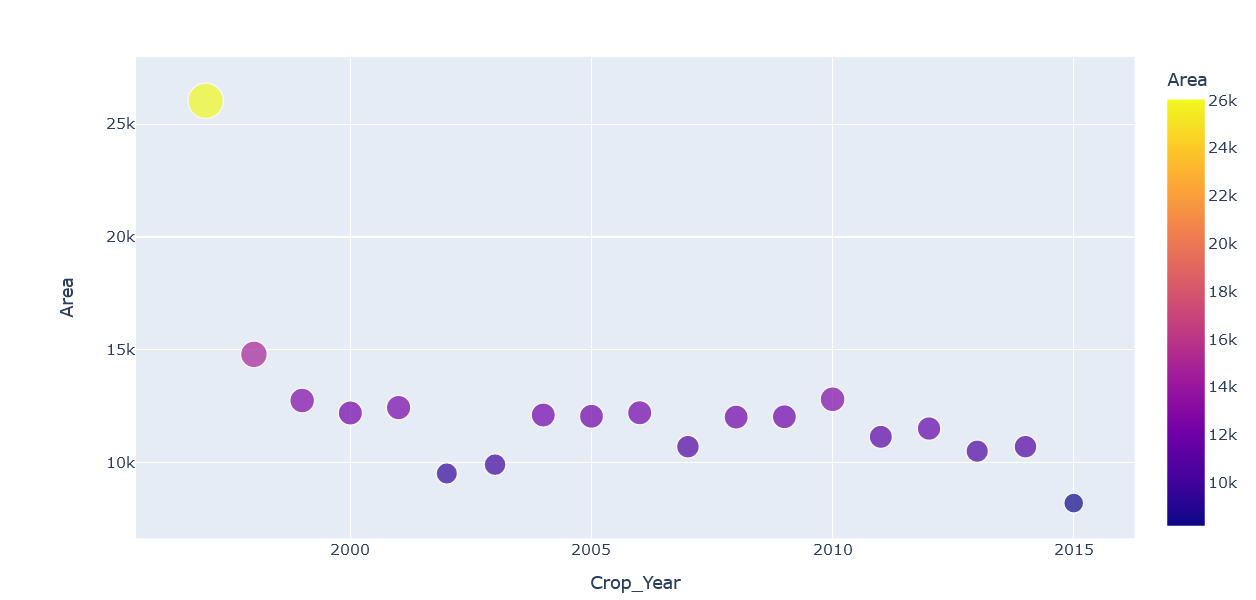
****

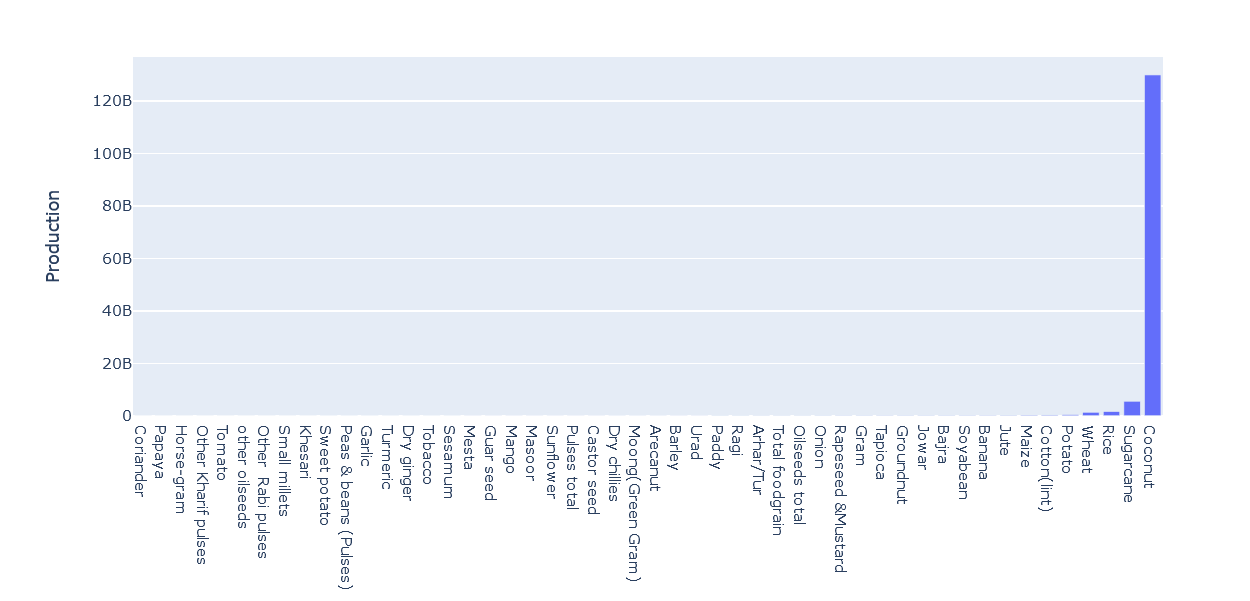
****

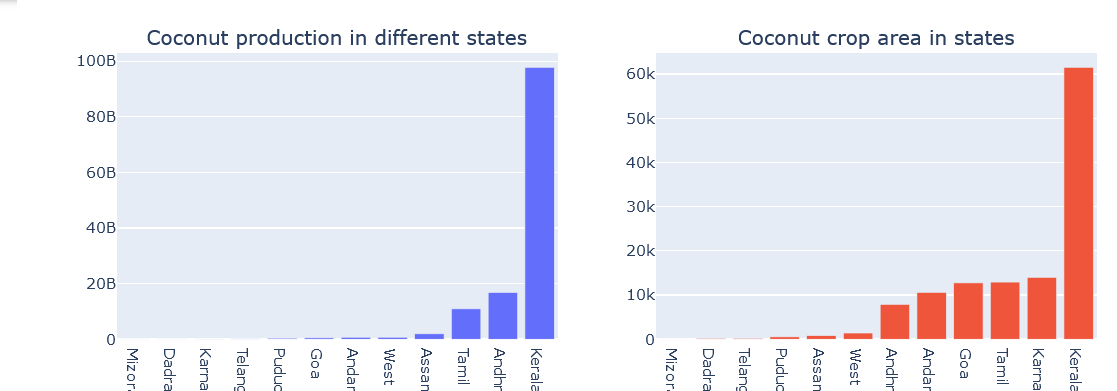
****

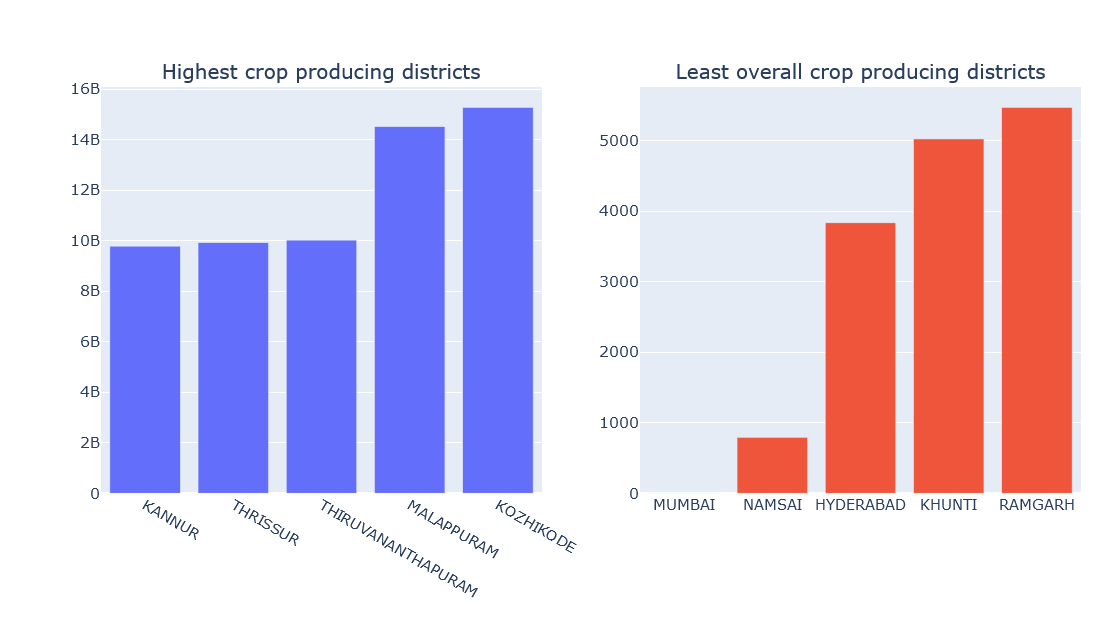
****

****

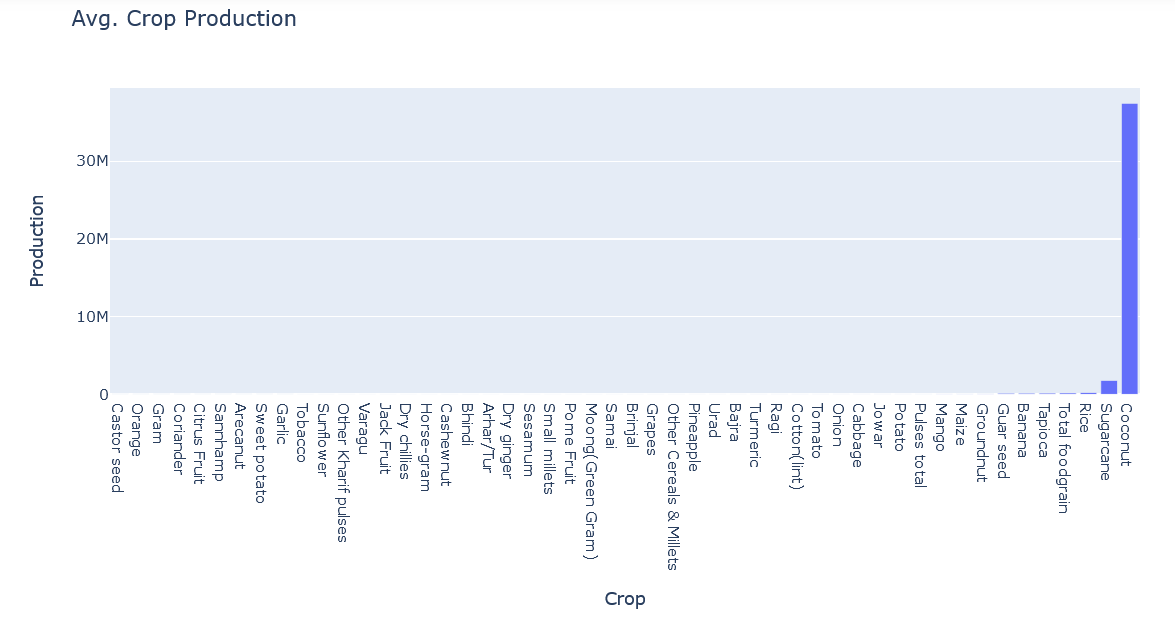
****

****

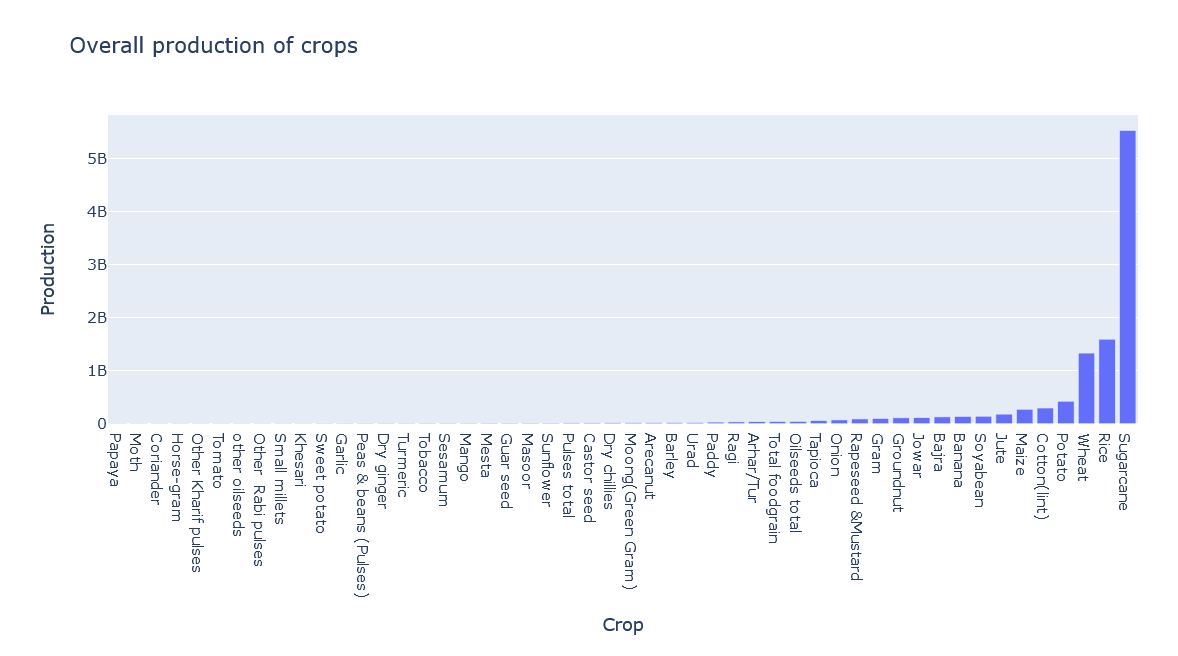
****

****

****

****

****

****