# **Honey Production Data Set Advance Python**

[ ]

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

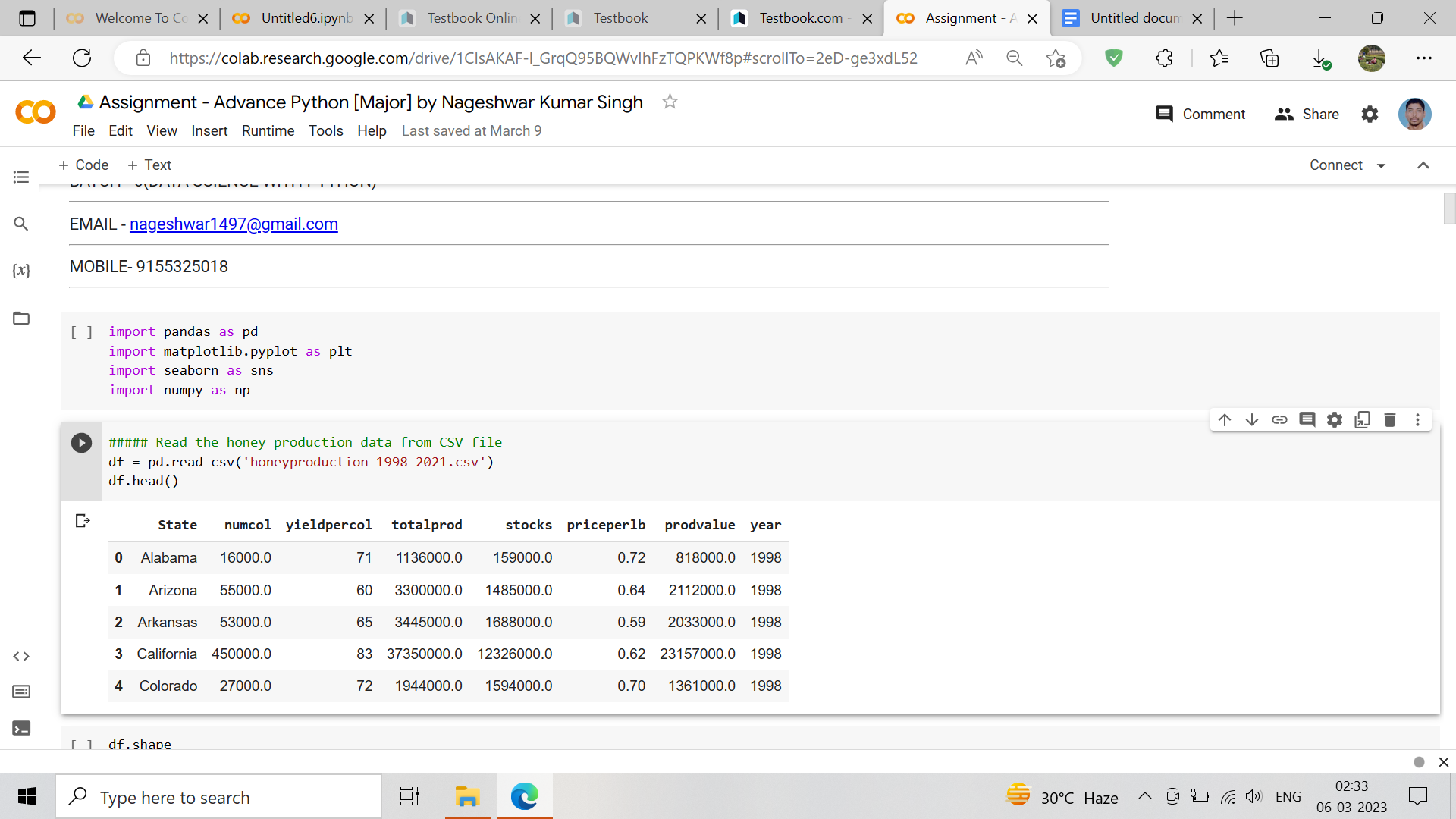
import numpy as np

[ ]

##### Read the honey production data from CSV file

df = pd.read\_csv('honeyproduction 1998-2021.csv')

df.head()



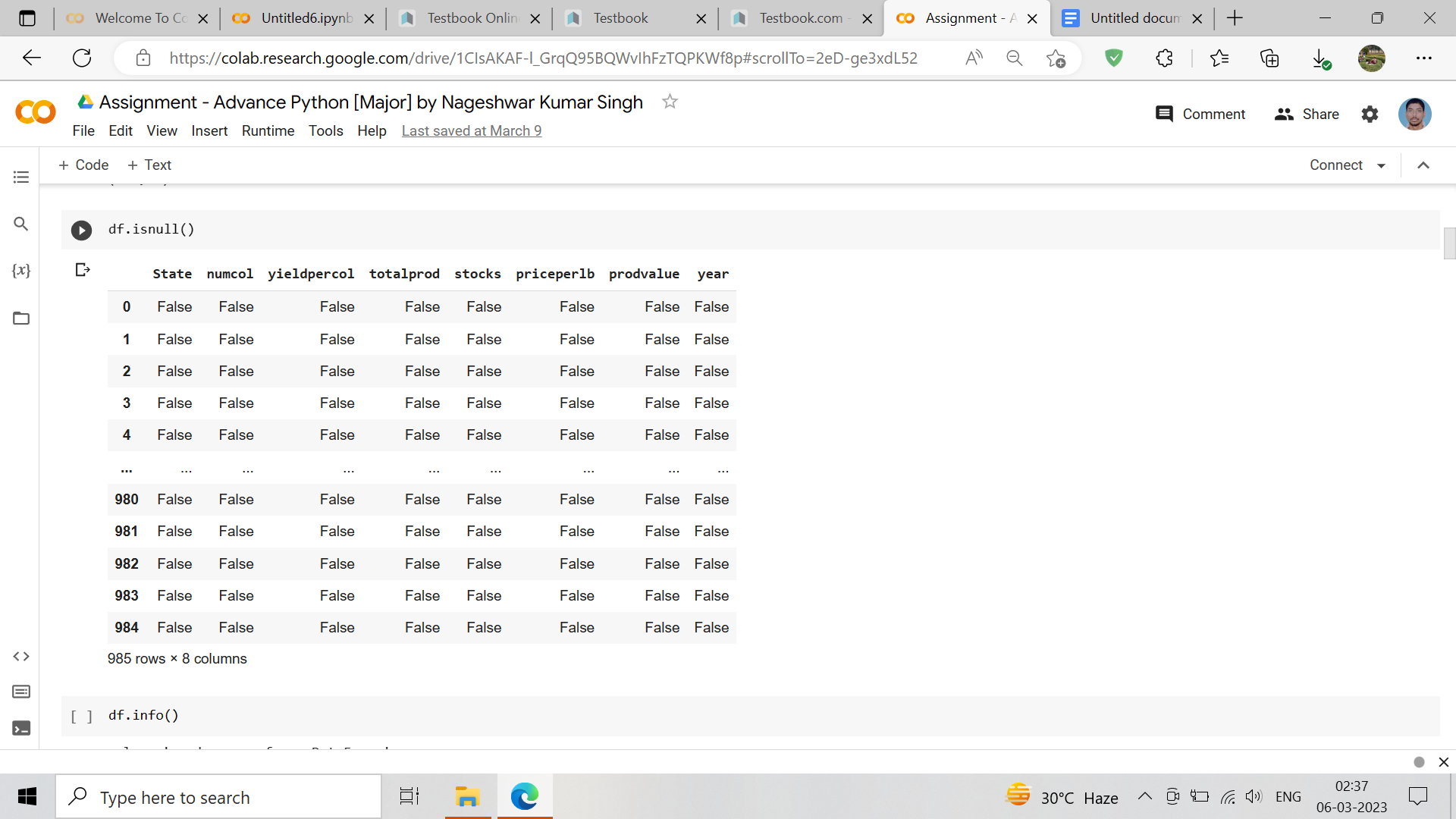
[ ]

df.shape

(985, 8)

[ ]

df.isnull()



[ ]

df.info()

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 985 entries, 0 to 984

Data columns (total 8 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 State 985 non-null object

1 numcol 985 non-null float64

2 yieldpercol 985 non-null int64

3 totalprod 985 non-null float64

4 stocks 985 non-null float64

5 priceperlb 985 non-null float64

6 prodvalue 985 non-null float64

7 year 985 non-null datetime64[ns]

dtypes: datetime64[ns](1), float64(5), int64(1), object(1)

memory usage: 61.7+ KB

[ ]

df.dtypes

State object

numcol float64

yieldpercol int64

totalprod float64

stocks float64

priceperlb float64

prodvalue float64

year int64

dtype: object

[ ]

# Convert the 'year' column to datetime format

df['year'] = pd.to\_datetime(df['year'], format='%Y')

[ ]

df.dtypes

State object

numcol float64

yieldpercol int64

totalprod float64

stocks float64

priceperlb float64

prodvalue float64

year datetime64[ns]

dtype: object

# **Q1. How has honey production yield changed from 1998 to 2021?**

[ ]

#### Create a line plot of honey production over time

sns.set\_style('whitegrid')

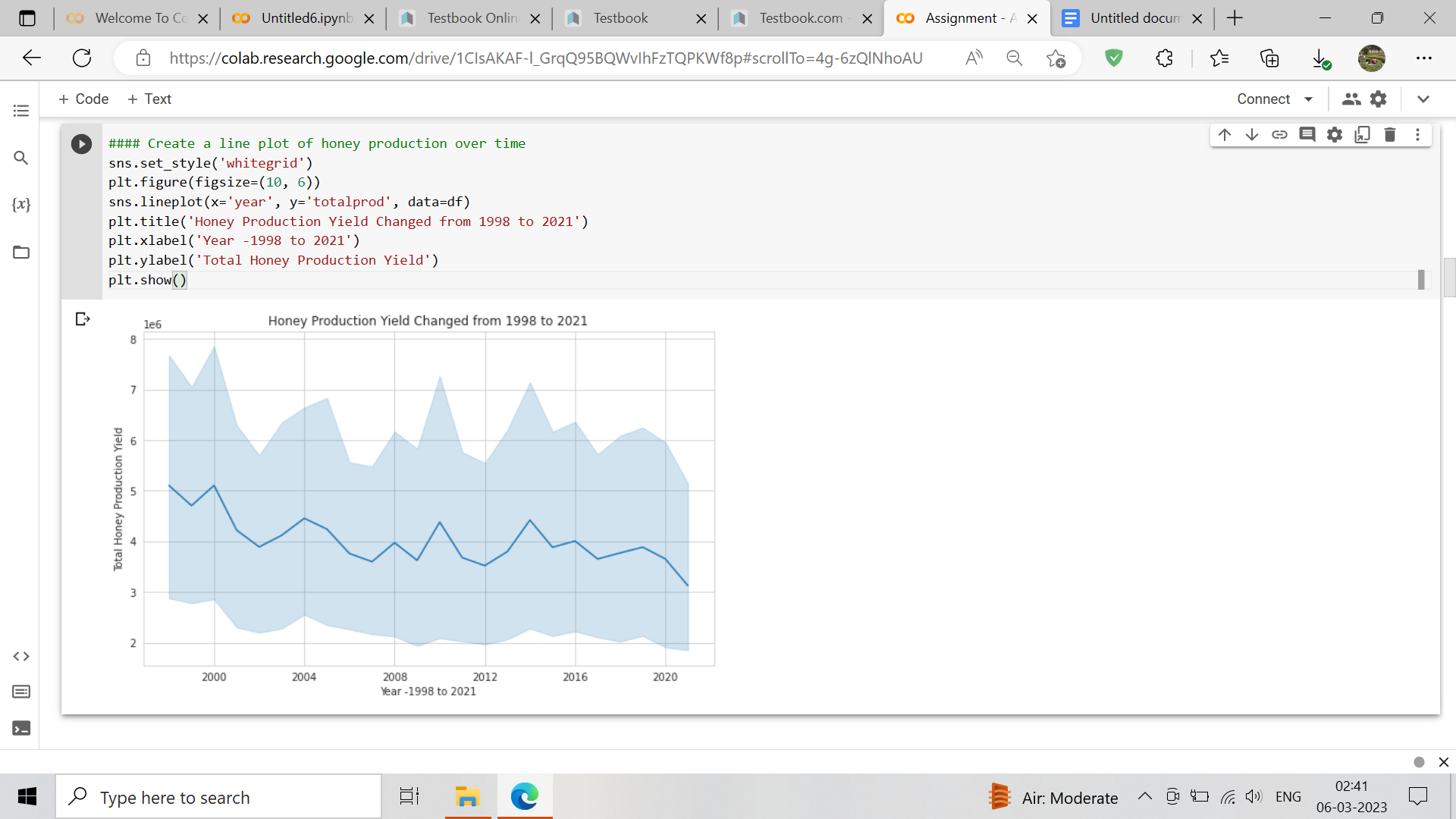
plt.figure(figsize=(10, 6))

sns.lineplot(x='year', y='totalprod', data=df)

plt.title('Honey Production Yield Changed from 1998 to 2021')

plt.xlabel('Year -1998 to 2021')

plt.ylabel('Total Honey Production Yield')

plt.show()

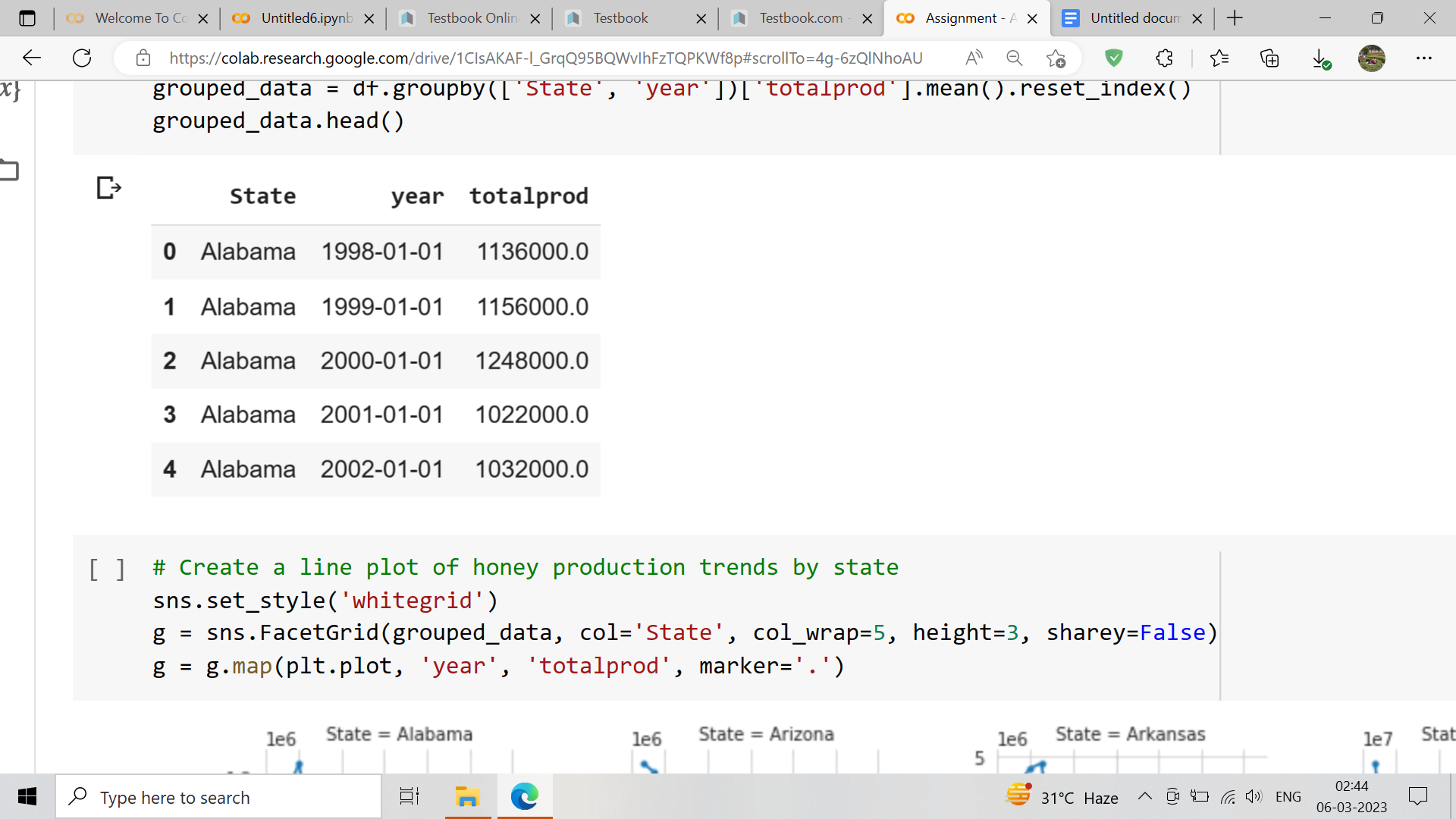
# **Q2.Overtime, what are the major production trends across the states?**

[ ]

# Group the data by state and year and calculate the mean yield

grouped\_data = df.groupby(['State', 'year'])['totalprod'].mean().reset\_index()

grouped\_data.head()



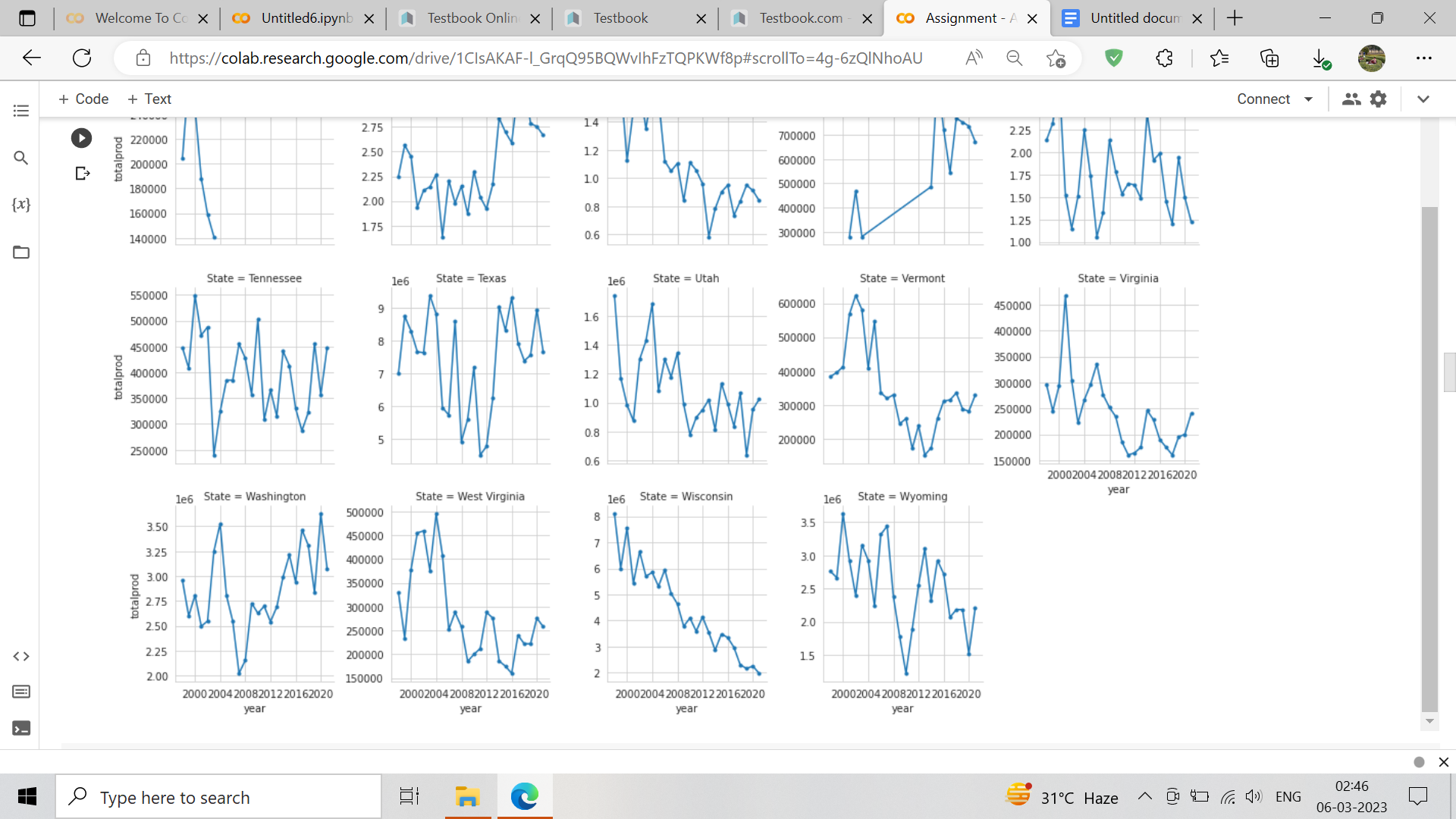
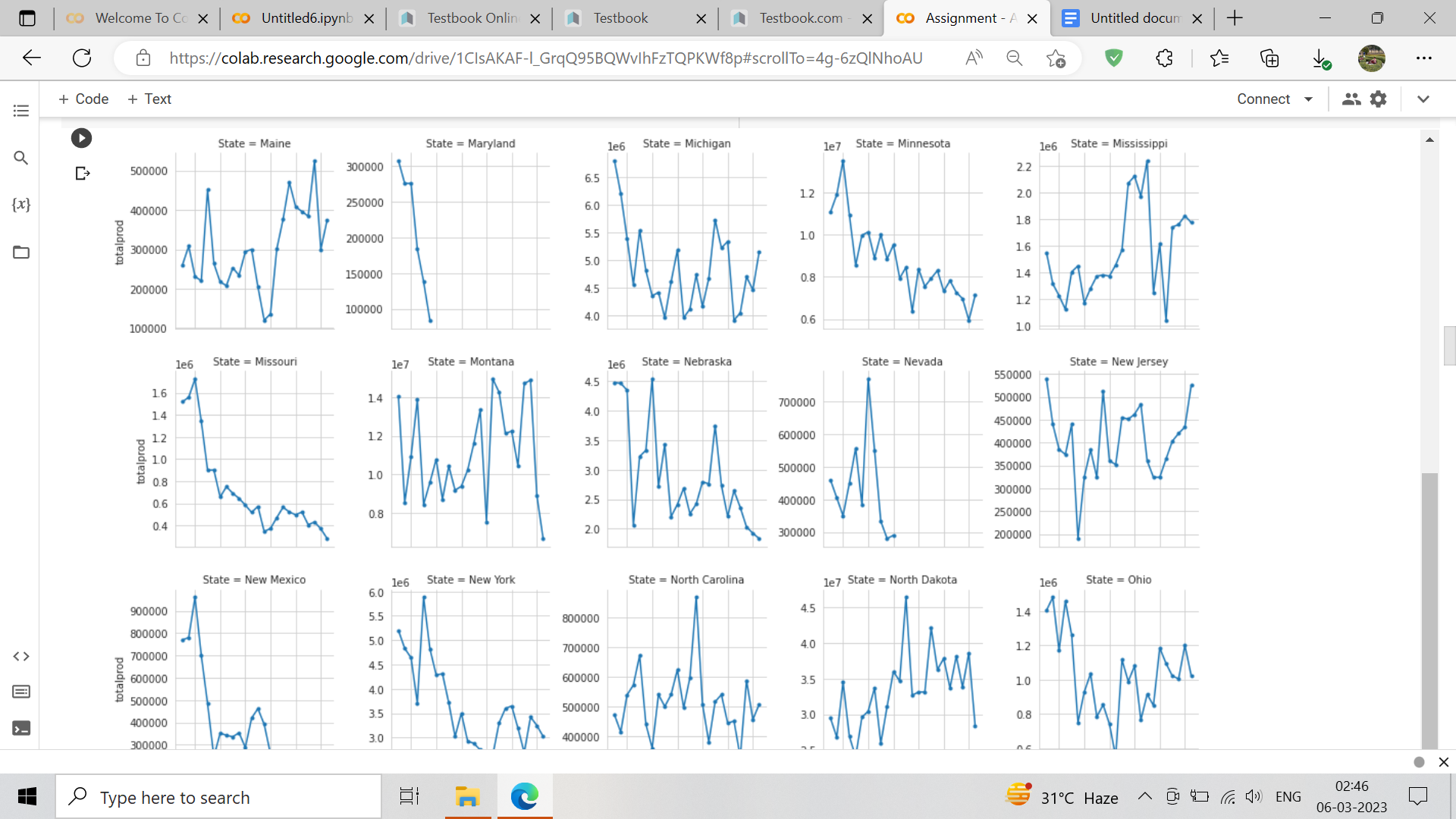
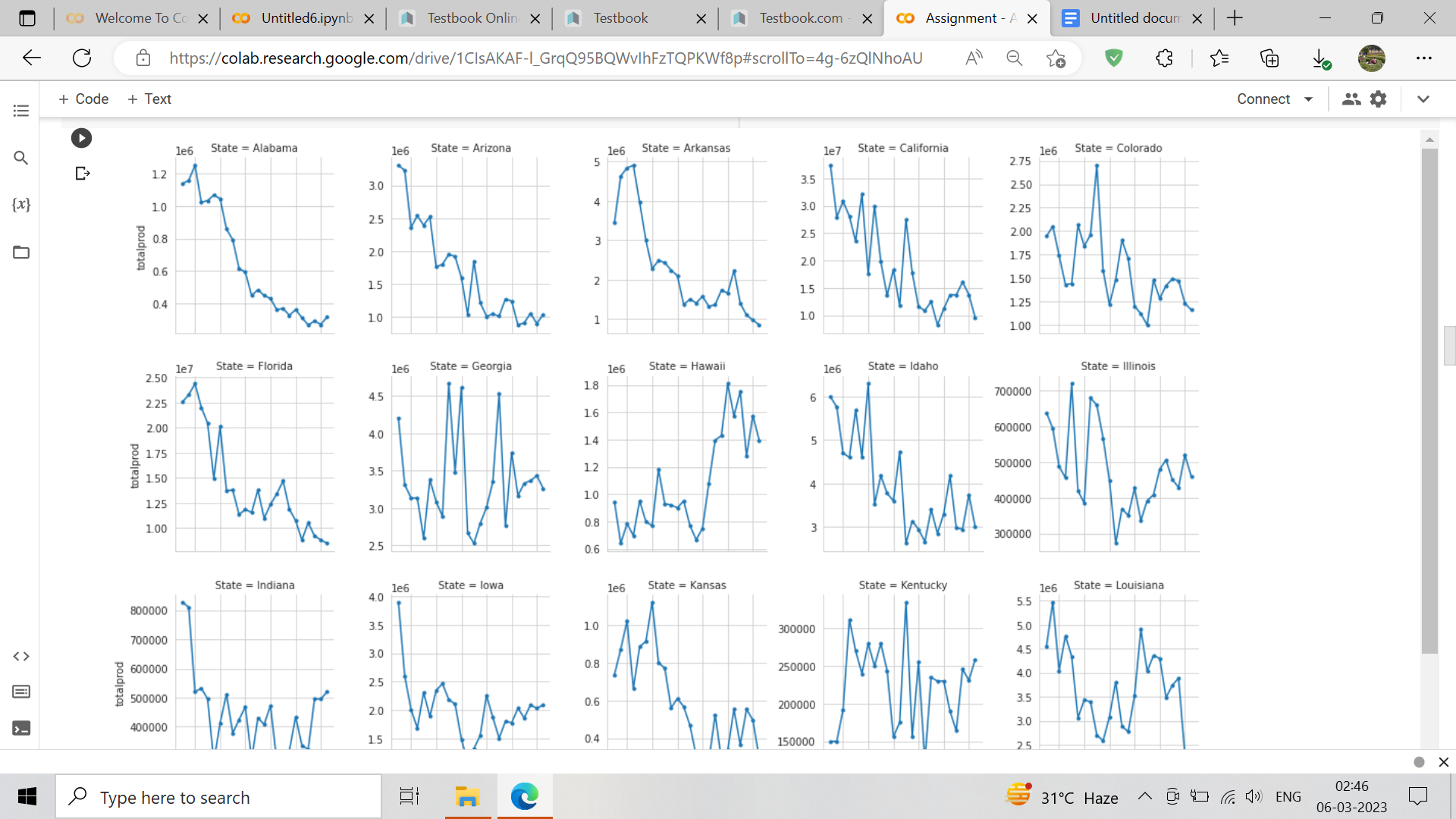
[ ]

# Create a line plot of honey production trends by state

sns.set\_style('whitegrid')

g = sns.FacetGrid(grouped\_data, col='State', col\_wrap=5, height=3, sharey=False)

g = g.map(plt.plot, 'year', 'totalprod', marker='.')



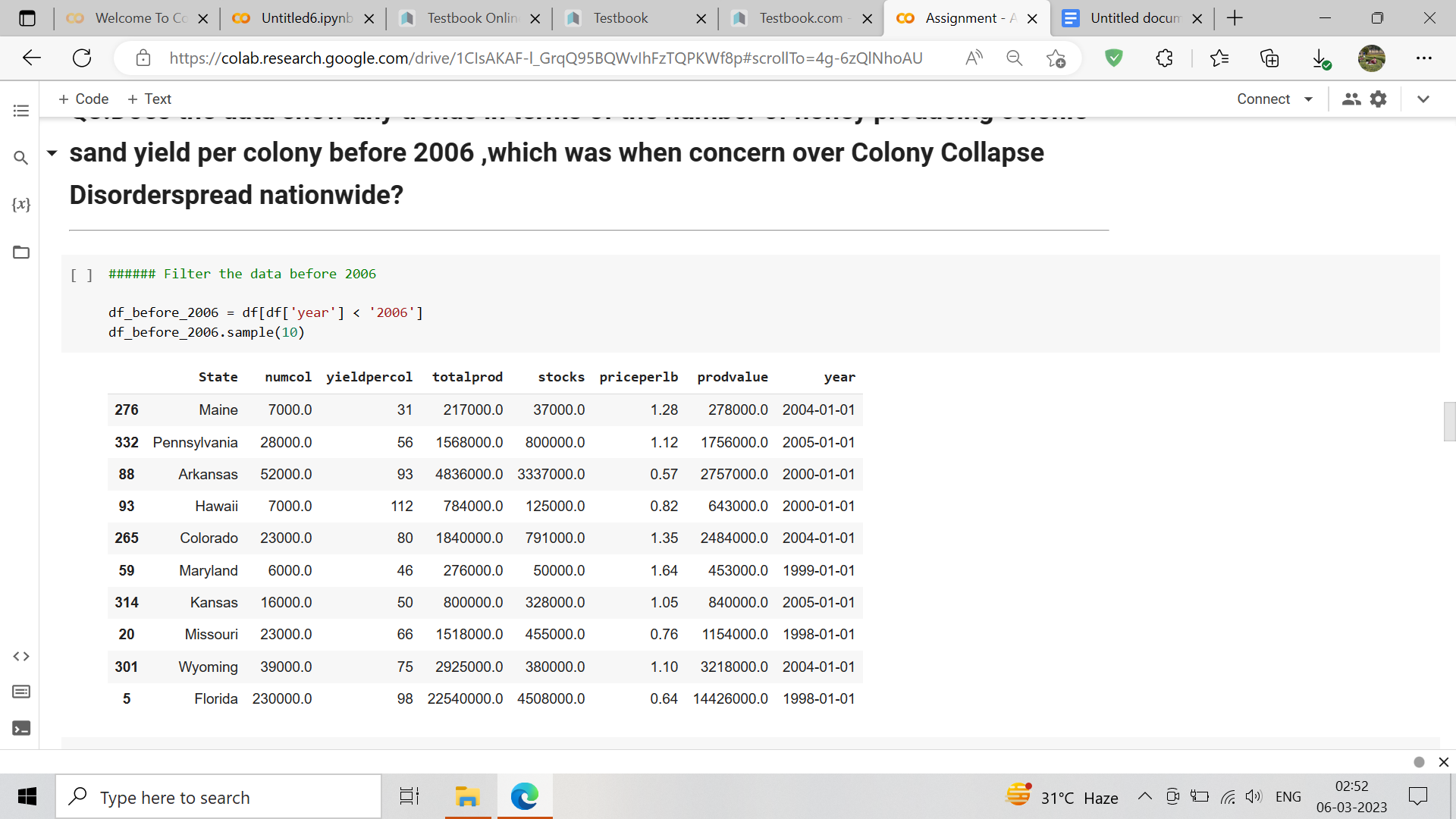
# **Q3.Does the data show any trends in terms of the number of honey producing colonie sand yield per colony before 2006 ,which was when concern over Colony Collapse Disorderspread nationwide?**

[ ]

###### Filter the data before 2006

df\_before\_2006 = df[df['year'] < '2006']

df\_before\_2006.sample(10)



[ ]

#### Calculate summary statistics

state\_colonies = df\_before\_2006.groupby('State')['numcol'].sum()

state\_yield = df\_before\_2006.groupby('State')['yieldpercol'].mean()

[ ]

# Visualize trends in colonies and yield

plt.figure(figsize=(12,6))

sns.barplot(x=state\_colonies.index, y=state\_colonies.values)

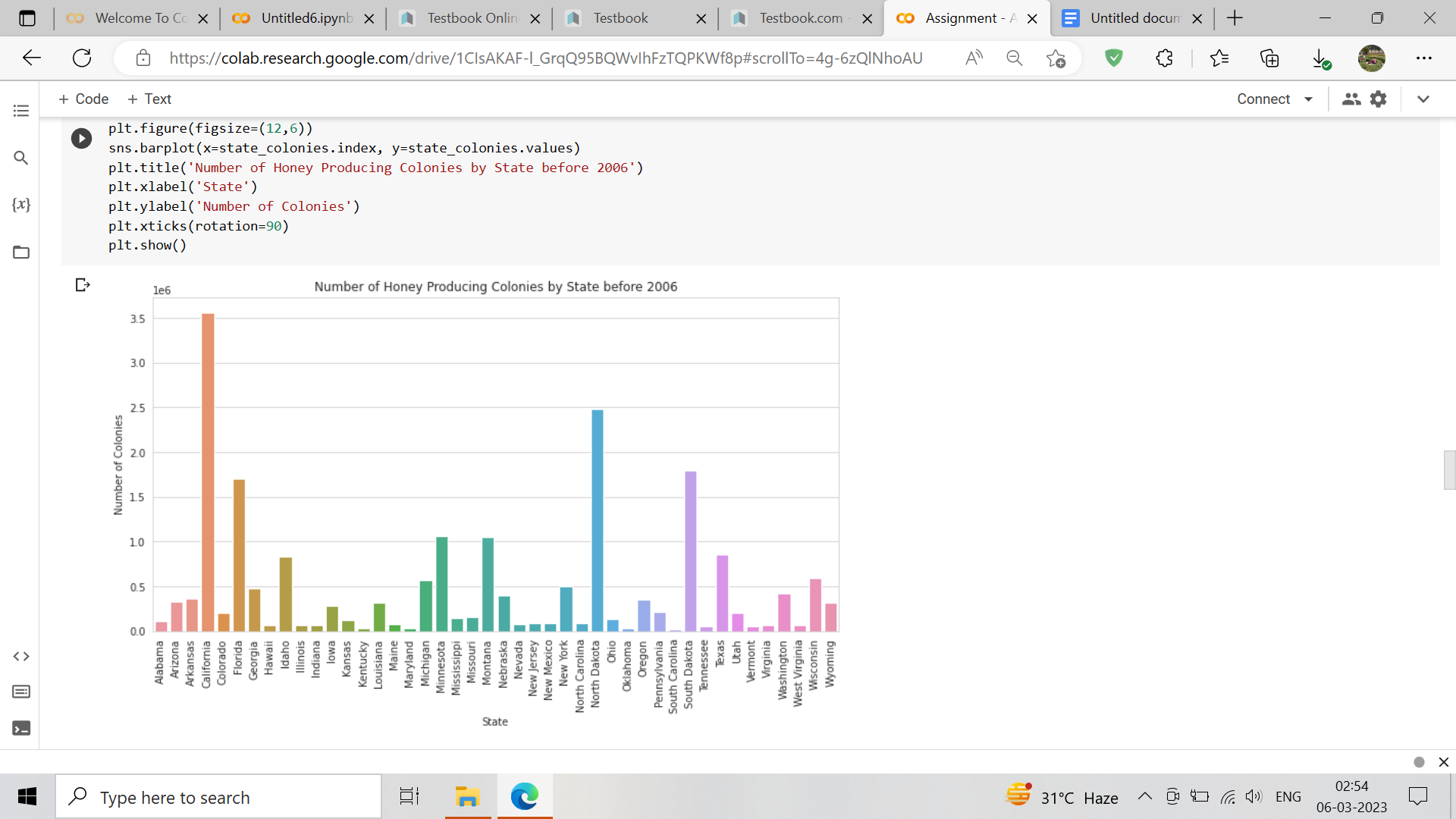
plt.title('Number of Honey Producing Colonies by State before 2006')

plt.xlabel('State')

plt.ylabel('Number of Colonies')

plt.xticks(rotation=90)

plt.show()



[ ]

# Yield per Colony by State before 2006

plt.figure(figsize=(12,6))

sns.barplot(x=state\_yield.index, y=state\_yield.values)

plt.title('Yield per Colony by State before 2006')

plt.xlabel('State')

plt.ylabel('Yield per Colony')

plt.xticks(rotation=90)

plt.show()



# **Q4. Are there any patterns that can be observed between total honey production and value of productione very year?**

[ ]

#### Visualize production trends and value of production

plt.figure(figsize=(12,6))

sns.lineplot(x='year', y='totalprod', data=df)

sns.lineplot(x='year', y='prodvalue', data=df)

plt.title('Total US Honey Production and Value of Production over Time')

plt.xlabel('Year')

plt.ylabel('Total Production/Value')

plt.legend(['Total Production', 'Value of Production'])

plt.show()

# 

# **Q5.How has the value of production, which in some sense could be tied to demand, changed every year?**

[ ]

#·Visualize·value·of·production·trends

plt.figure(figsize=(12,6))

sns.lineplot(x='year',·y='prodvalue',·data=df)

plt.title('Value·of·US·Honey·Production·over·Time')

plt.xlabel('Year')

plt.ylabel('Value·of·Production')

plt.show()



# **Q6.Constructs the related plots using Seaborn and Matplot apply customization and derive insights from the visualization**

[ ]

# Plot the distribution of honey production by state in 2016

plt.figure(figsize=(10,6))

sns.boxplot(x='State', y='totalprod', data=df[df['year'].dt.year == 2016])

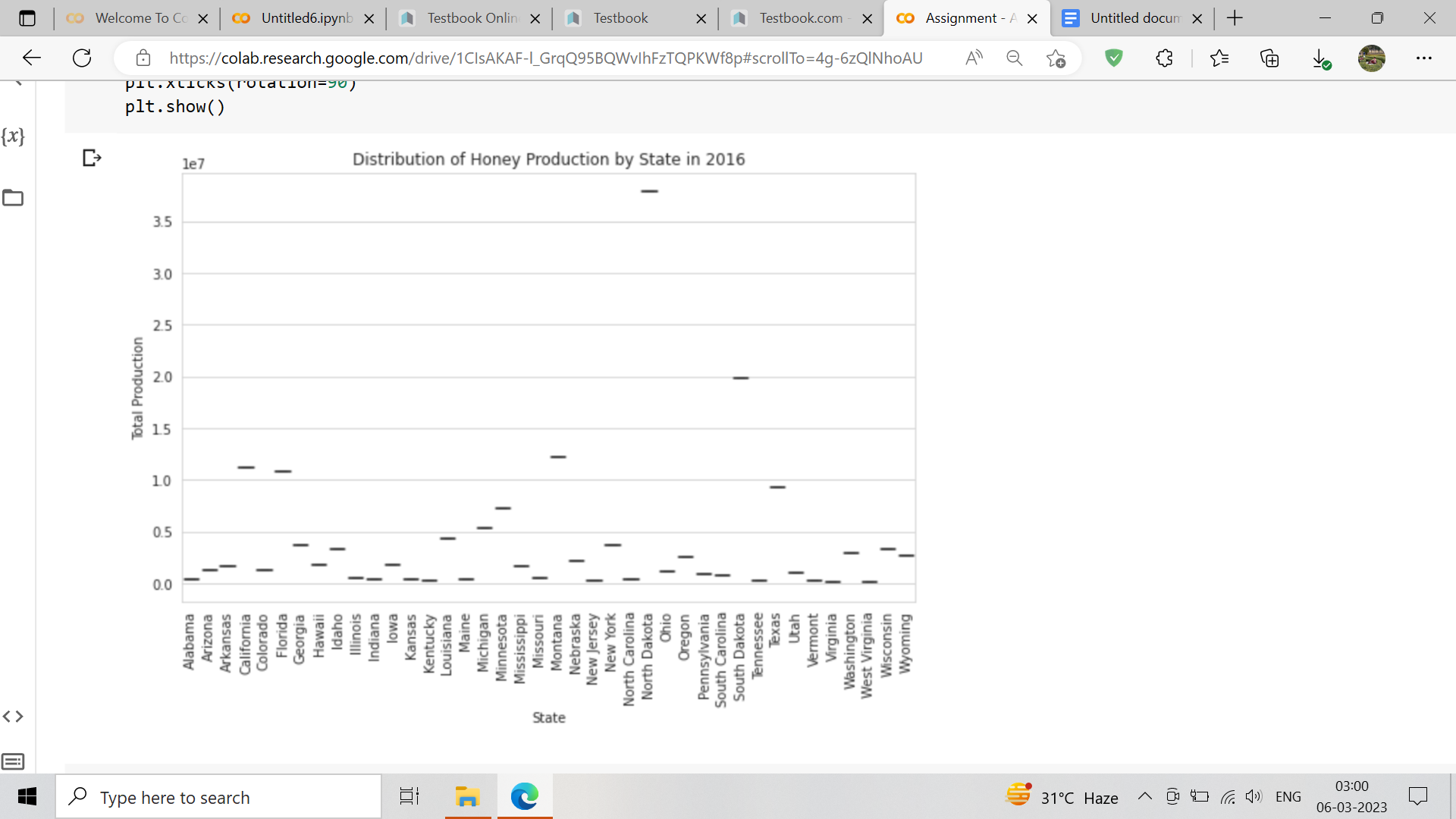
plt.title('Distribution of Honey Production by State in 2016')

plt.xlabel('State')

plt.ylabel('Total Production')

plt.xticks(rotation=90)

plt.show()



[ ]

# Plot the correlation between the price per pound and the value of production in 2008

plt.figure(figsize=(10,6))

sns.scatterplot(x='priceperlb', y='prodvalue', data=df[df['year'].dt.year == 2008])

plt.title('Correlation between Price per Pound and Value of Production in 2008')

plt.xlabel('Price per Pound')

plt.ylabel('Value of Production')

plt.show()

