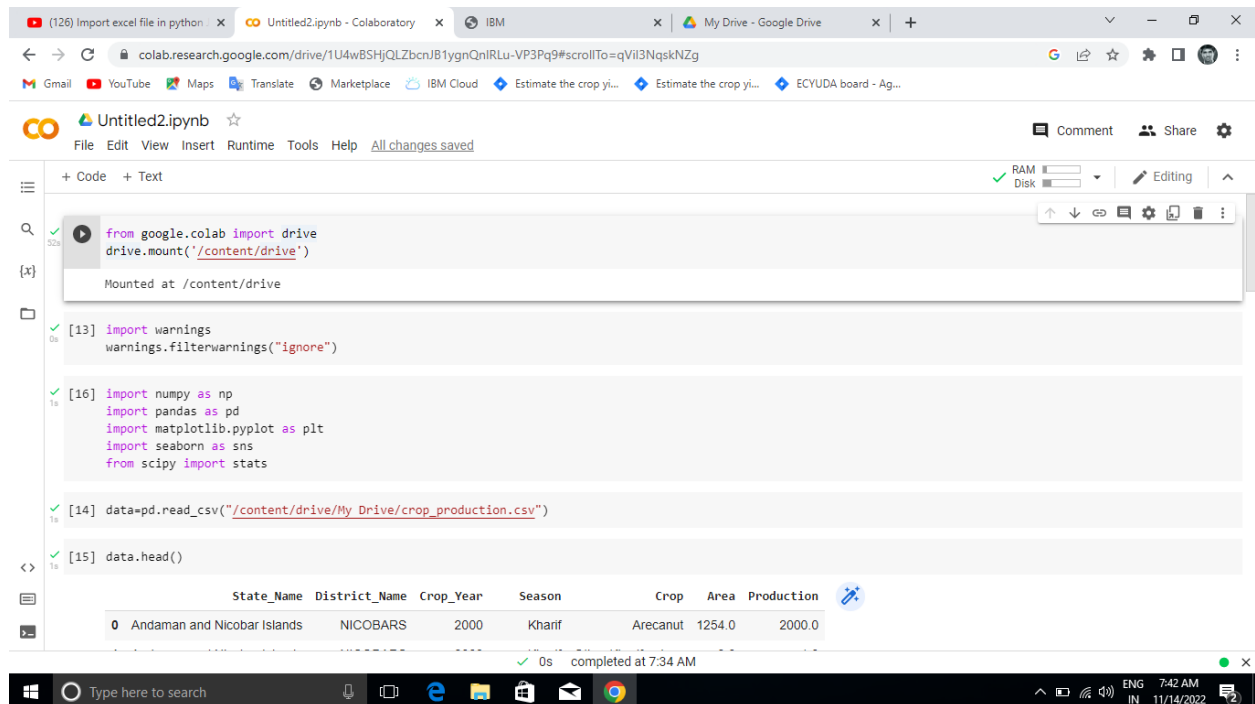


PROJECT DEVELOPMENT PHASE

SPRINT – 2

Date	31 October 2022
Team Id	PNT2022TMID07696
Project Name	Estimate the Crop Yield Using Data Analytics

DATA EXPLORATION INPUT & OUTPUT:



The screenshot displays a Google Colab notebook interface. The top navigation bar includes tabs for 'Untitled2.ipynb - Colaboratory', 'IBM', and 'My Drive - Google Drive'. The address bar shows the URL: `colab.research.google.com/drive/1U4wBSHjQLZbcnJ81ygnQnIRLu-VP3Pq9#scrollTo=qVil3NqskNZg`. The notebook's menu bar includes 'File', 'Edit', 'View', 'Insert', 'Runtime', 'Tools', and 'Help'. The left sidebar shows a file explorer with 'Untitled2.ipynb' selected. The main workspace contains the following code cells:

```
from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive

[13] import warnings
warnings.filterwarnings("ignore")

[16] import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats

[14] data=pd.read_csv("/content/drive/My Drive/crop_production.csv")

[15] data.head()
```

The output of the final cell is a table showing the first row of data:

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254.0	2000.0

The bottom status bar indicates the notebook is 'completed at 7:34 AM'.

Colaboratory interface showing the first five rows of a dataset. The code cell [15] contains `data.head()`. The output is a table with 7 columns: State_Name, District_Name, Crop_Year, Season, Crop, Area, and Production. The data is for Andaman and Nicobar Islands in 2000.

	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

The code cell [18] contains `data.info()`. The output shows the data is a pandas DataFrame with 246091 entries and 7 columns. The dtypes are: State_Name (object), District_Name (object), Crop_Year (int64), Season (object), Crop (object), Area (float64), and Production (float64).

Colaboratory interface showing the last five rows of a dataset. The code cell [19] contains `data.columns`. The output is a list of column names: ['State_Name', 'District_Name', 'Crop_Year', 'Season', 'Crop', 'Area', 'Production']. The code cell [21] contains `data.tail()`. The output is a table with 7 columns: State_Name, District_Name, Crop_Year, Season, Crop, Area, and Production. The data is for West Bengal in 2014.

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
246086	West Bengal	PURULIA	2014	Summer	Rice	306.0	801.0
246087	West Bengal	PURULIA	2014	Summer	Sesamum	627.0	463.0
246088	West Bengal	PURULIA	2014	Whole Year	Sugarcane	324.0	16250.0
246089	West Bengal	PURULIA	2014	Winter	Rice	279151.0	597899.0
246090	West Bengal	PURULIA	2014	Winter	Sesamum	175.0	88.0

Colaboratory interface showing the execution of a Python script. The script includes a data description and a null check.

```
[22] data.describe()
```

	Crop_Year	Area	Production
count	246091.000000	2.460910e+05	2.423610e+05
mean	2005.643018	1.200282e+04	5.825034e+05
std	4.952164	5.052340e+04	1.706581e+07
min	1997.000000	4.000000e-02	0.000000e+00
25%	2002.000000	8.000000e+01	8.800000e+01
50%	2006.000000	5.820000e+02	7.290000e+02
75%	2010.000000	4.392000e+03	7.023000e+03
max	2015.000000	8.580100e+06	1.250800e+09

```
[23] data.isnull().sum()
```

State_Name	0
District_Name	0
Crop_Year	0
Season	0
Crop	0

completed at 7:34 AM

Colaboratory interface showing the execution of a Python script. The script includes a data description and a null check.

```
[22] data.describe()
```

	Crop_Year	Area	Production
count	246091.000000	2.460910e+05	2.423610e+05
mean	2005.643018	1.200282e+04	5.825034e+05
std	4.952164	5.052340e+04	1.706581e+07
min	1997.000000	4.000000e-02	0.000000e+00
25%	2002.000000	8.000000e+01	8.800000e+01
50%	2006.000000	5.820000e+02	7.290000e+02
75%	2010.000000	4.392000e+03	7.023000e+03
max	2015.000000	8.580100e+06	1.250800e+09

```
[23] data.isnull().sum()
```

State_Name	0
District_Name	0
Crop_Year	0
Season	0
Crop	0
Area	0
Production	3730
dtype:	int64

completed at 7:34 AM

Untitled2.ipynb - Colaboratory

colab.research.google.com/drive/1U4wBSHjQLZbcnJB1ygnQnRLu-VP3Pq9#scrollTo=-MrMNq_wNkuv

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

```
[ ] data.corr()
```

	Crop_Year	Area	Production
Crop_Year	1.000000	-0.026022	0.006989
Area	-0.026022	1.000000	0.040587
Production	0.006989	0.040587	1.000000

```
[ ] data.cov()
```

	Crop_Year	Area	Production
Crop_Year	24.523927	-6.510591e+03	5.914148e+05
Area	-6510.590664	2.552614e+09	3.522683e+10
Production	591414.831146	3.522683e+10	2.912420e+14

```
[ ] data.dtypes
```

State_Name	object
District_Name	object
Crop_Year	int64

0s completed at 9:11 AM

Untitled2.ipynb - Colaboratory

colab.research.google.com/drive/1U4wBSHjQLZbcnJB1ygnQnRLu-VP3Pq9#scrollTo=-MrMNq_wNkuv

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

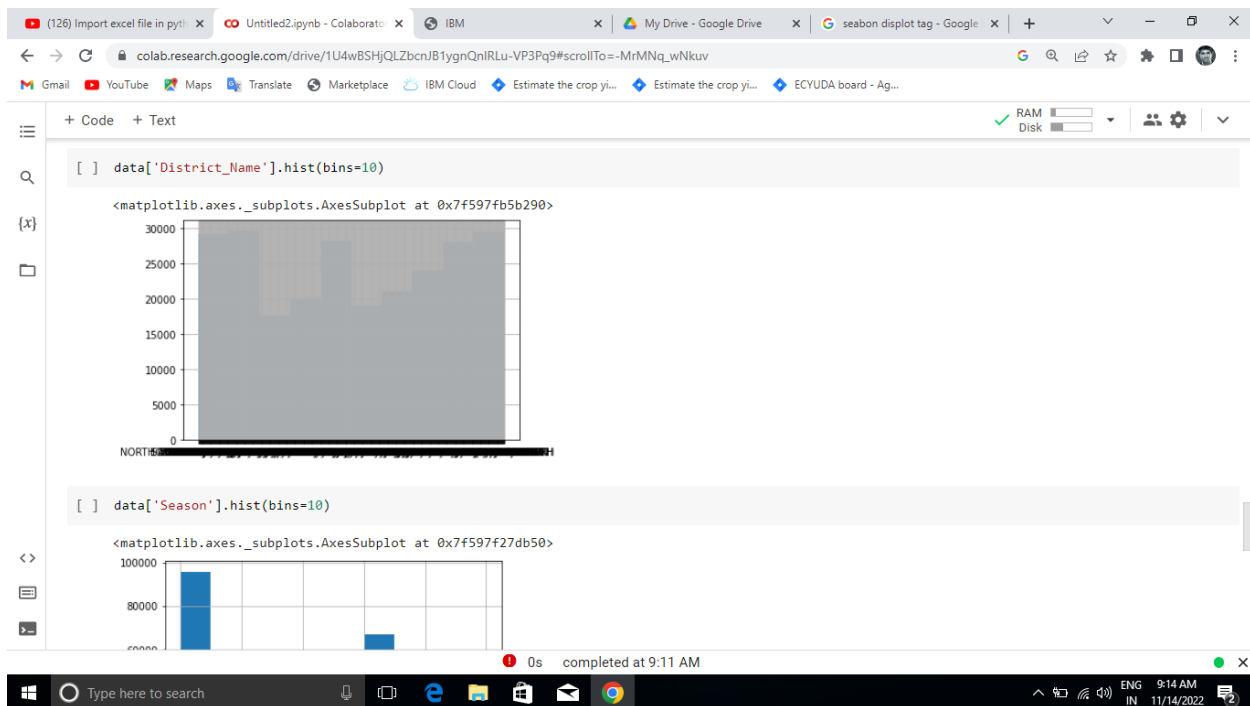
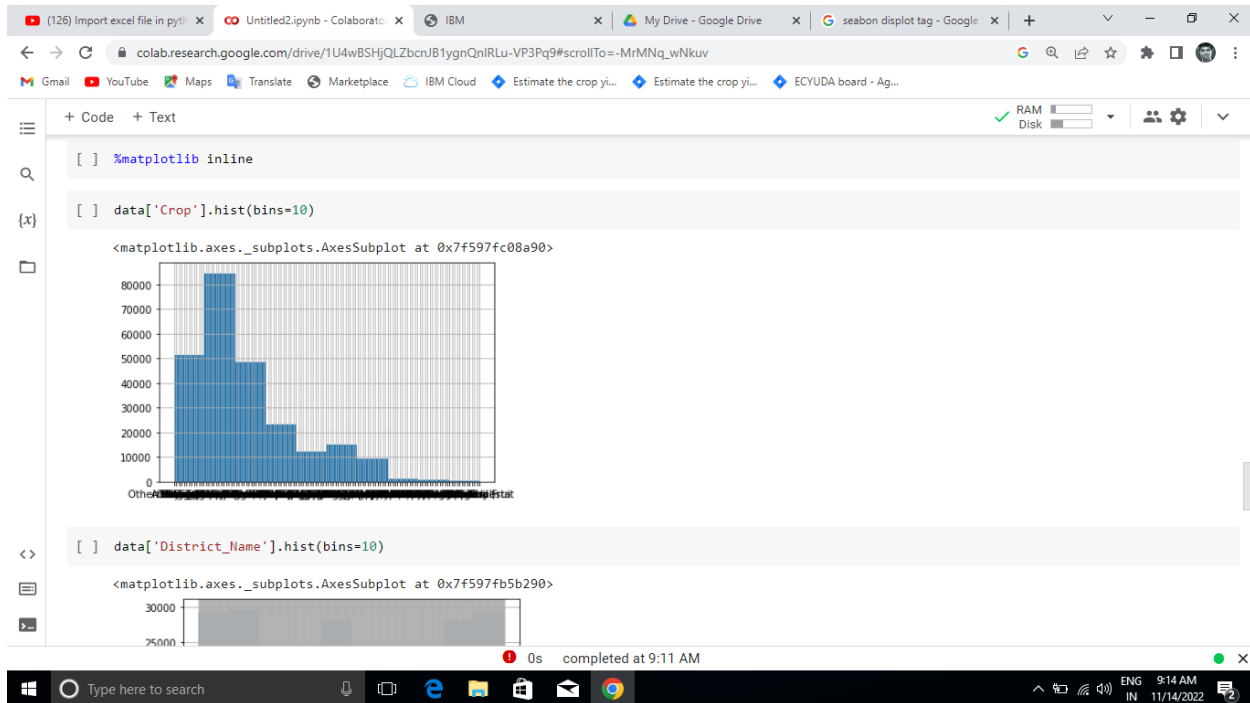
```
[ ] correlation Matrix
```

```
[ ] %matplotlib inline
```

```
[ ] data['Crop'].hist(bins=10)
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f597f08a90b>

0s completed at 9:11 AM



Browser tabs: (126) Import excel file in pytho... x, Untitled2.ipynb - Colaborato... x, IBM x, My Drive - Google Drive x, seaborne displot tag - Google x, +, -, x

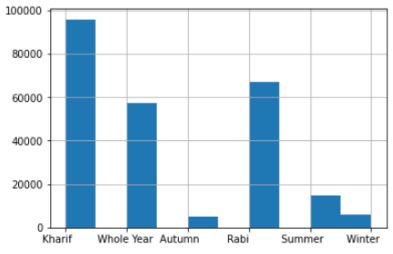
Address bar: colab.research.google.com/drive/1U4wBSHjQLZbcnJB1ygnQnIRLu-VP3Pq9#scrollTo=-MrMNq_wNkuv

Navigation bar: Gmail, YouTube, Maps, Translate, Marketplace, IBM Cloud, Estimate the crop yi..., Estimate the crop yi..., ECVUDA board - Ag...

Code editor: + Code + Text, RAM Disk, Settings, Close

```
[ ] data['Season'].hist(bins=10)
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f597f27db50>



Season	Production (approx.)
Kharif	95000
Whole Year	58000
Autumn	5000
Rabi	68000
Summer	15000
Winter	8000

```
[ ] data.columns
```

Index(['State_Name', 'District_Name', 'Crop_Year', 'Season', 'Crop', 'Area', 'Production'], dtype='object')

[]

0s completed at 9:11 AM

Windows taskbar: Type here to search, File Explorer, Edge, Mail, Chrome, System tray: ENG IN, 9:14 AM, 11/14/2022