

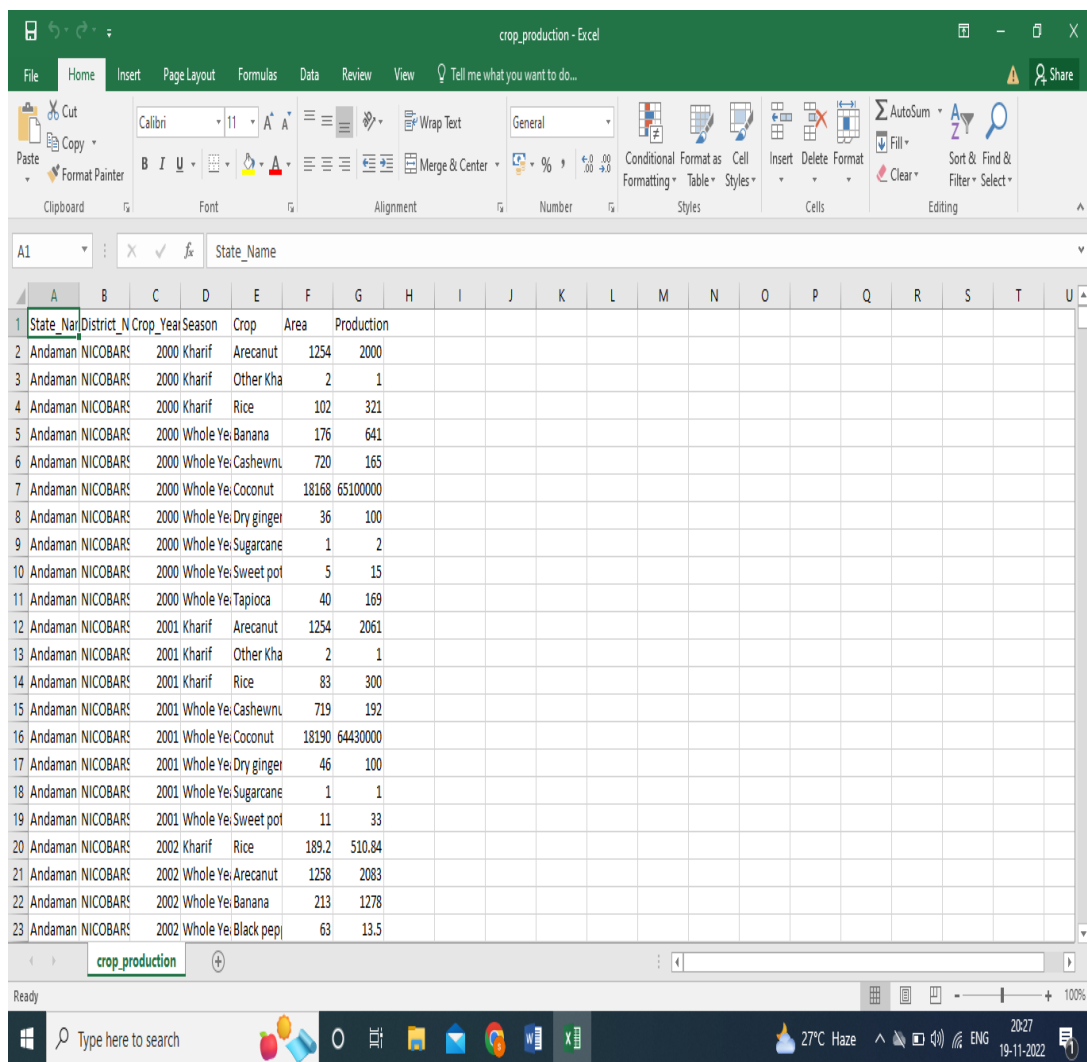
PROJECT DEVELOPMENT PHASE

SPRINT-2

Date	03 November 2022
Team ID	PNT2022TMID51723
Project Name	Estimate the crop yield using data analytics

WORKING WITH DATASETS:

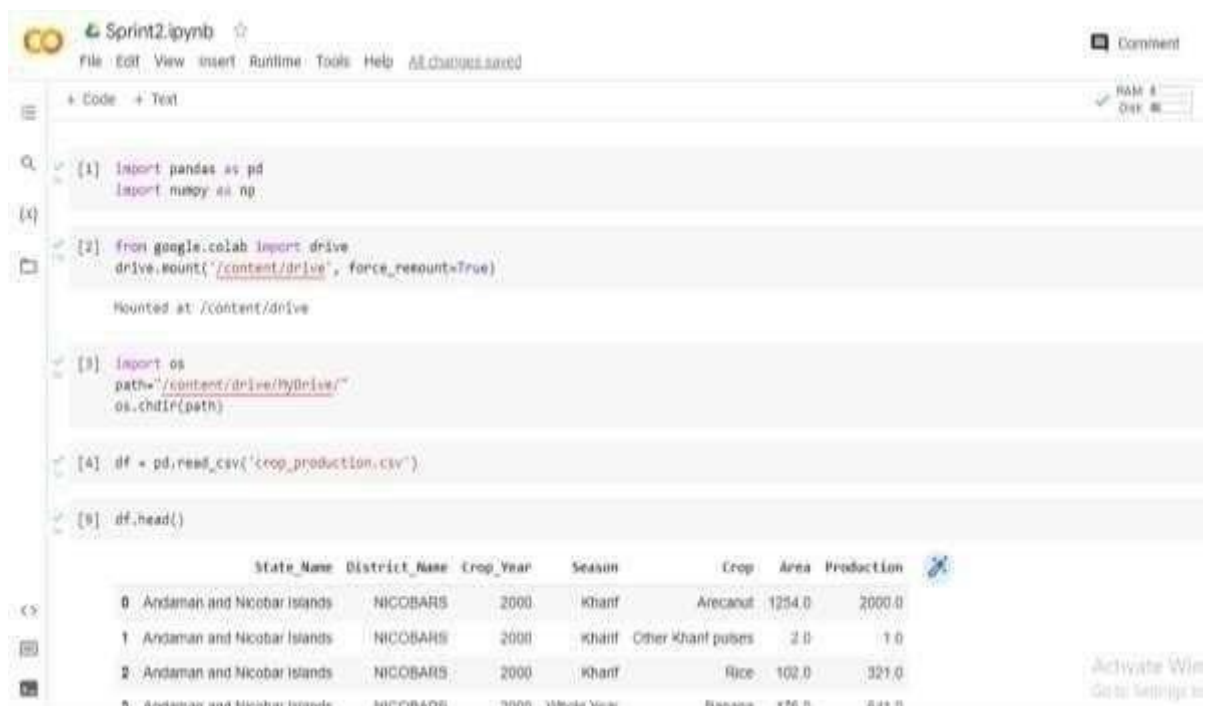
➤ Dataset used is as follows,



State_Name	District_Name	Crop_YearSeason	Crop	Area	Production
Andaman NICOBARS	2000 Kharif	Areca nut	1254	2000	
Andaman NICOBARS	2000 Kharif	Other Kharif	2	1	
Andaman NICOBARS	2000 Kharif	Rice	102	321	
Andaman NICOBARS	2000 Whole Year	Banana	176	641	
Andaman NICOBARS	2000 Whole Year	Cashew nut	720	165	
Andaman NICOBARS	2000 Whole Year	Coconut	18168	65100000	
Andaman NICOBARS	2000 Whole Year	Dry ginger	36	100	
Andaman NICOBARS	2000 Whole Year	Sugarcane	1	2	
Andaman NICOBARS	2000 Whole Year	Sweet potato	5	15	
Andaman NICOBARS	2000 Whole Year	Tapioca	40	169	
Andaman NICOBARS	2001 Kharif	Areca nut	1254	2061	
Andaman NICOBARS	2001 Kharif	Other Kharif	2	1	
Andaman NICOBARS	2001 Kharif	Rice	83	300	
Andaman NICOBARS	2001 Whole Year	Cashew nut	719	192	
Andaman NICOBARS	2001 Whole Year	Coconut	18190	64430000	
Andaman NICOBARS	2001 Whole Year	Dry ginger	46	100	
Andaman NICOBARS	2001 Whole Year	Sugarcane	1	1	
Andaman NICOBARS	2001 Whole Year	Sweet potato	11	33	
Andaman NICOBARS	2002 Kharif	Rice	189.2	510.84	
Andaman NICOBARS	2002 Whole Year	Areca nut	1258	2083	
Andaman NICOBARS	2002 Whole Year	Banana	213	1278	
Andaman NICOBARS	2002 Whole Year	Black pepper	63	13.5	

This dataset consists of columns like State Name, District Name, Crop, Season, Crop year, Area and Production.

- Importing and cleaning the dataset in Google Colab platform,



The screenshot shows a Google Colab notebook titled 'Sprint2.ipynb'. The code cells are as follows:

```
[1] import pandas as pd
import numpy as np

[2] from google.colab import drive
drive.mount('/content/drive', force_remount=True)

Mounted at /content/drive

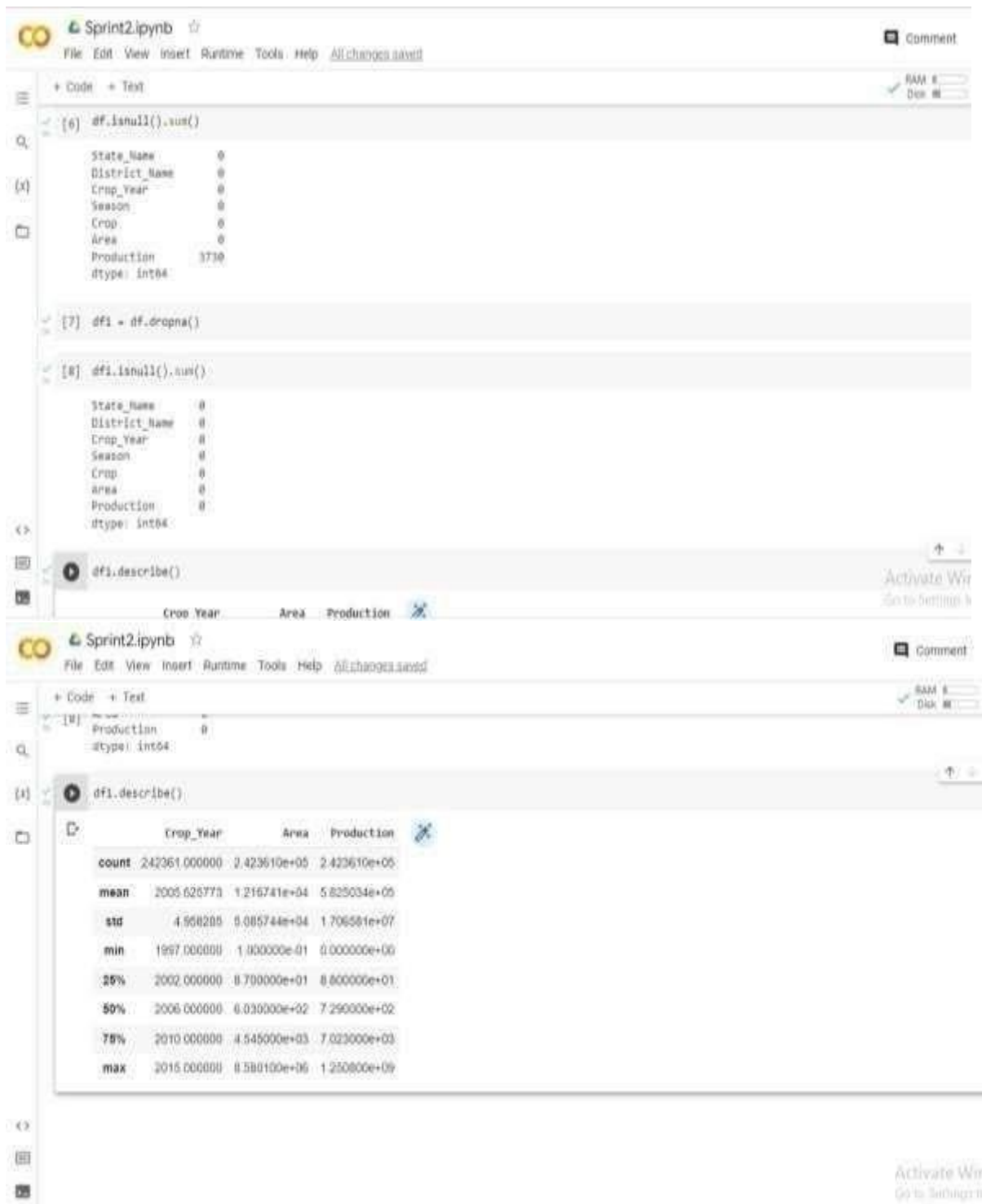
[3] import os
path='/content/drive/MyDrive/'
os.chdir(path)

[4] df = pd.read_csv('crop_production.csv')

[5] df.head()
```

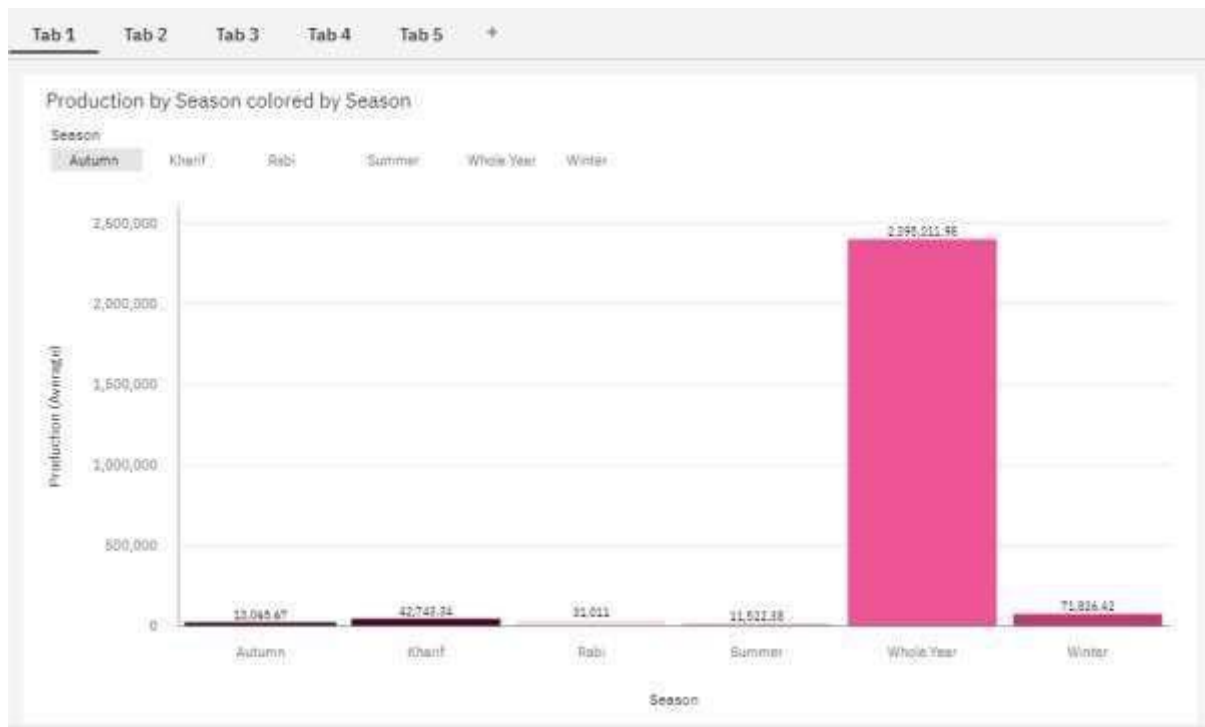
The output of the code shows the first five rows of the 'crop_production.csv' dataset:

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Areca nut	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	Winter Year	Rabosee	176.0	641.0



DATAVISUALIZATIONCHARTS:

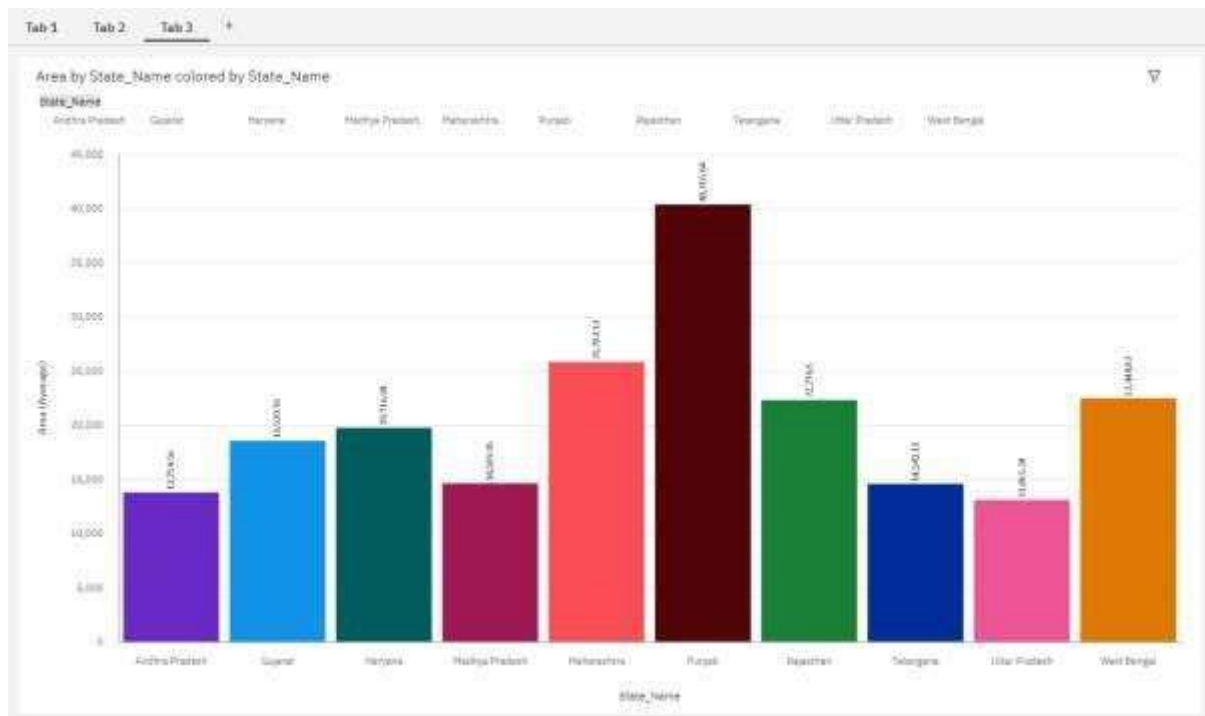
- Visualization on Seasons with average Production



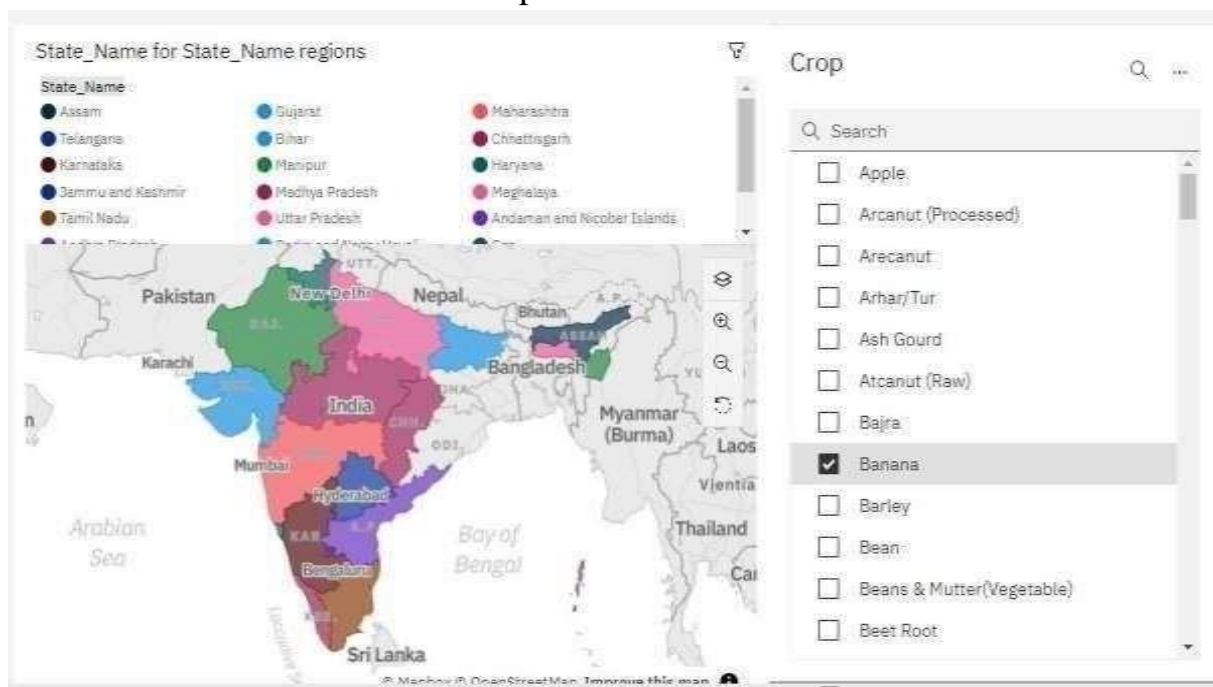
- Visualization with years usage of Area and Production



- Visualization on top 10 States with most Area



- Visualization on State with Crop Production



- Visualization on States with the Crop Production along with Season

