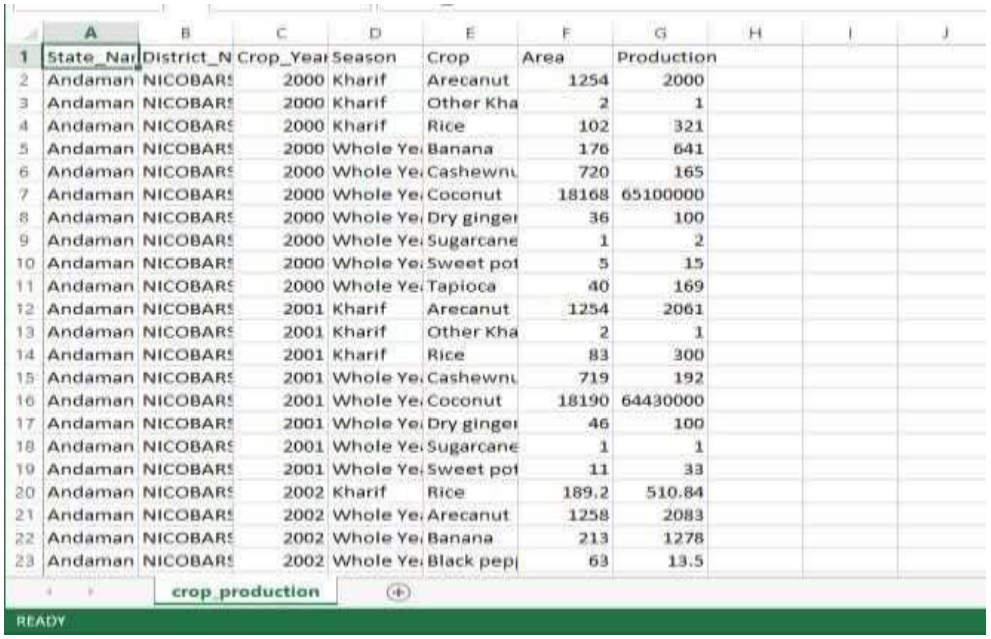
PROJECT DEVELOPMENT PHASE

SPRINT-2

Date	03 November 2022
Team ID	PNT2022TMID09610
Project Name	ESTIMATION OF CROP YIELD USING DATA ANALYTICS

WORKING WITH DATASETS:

- Dataset used is as follows,



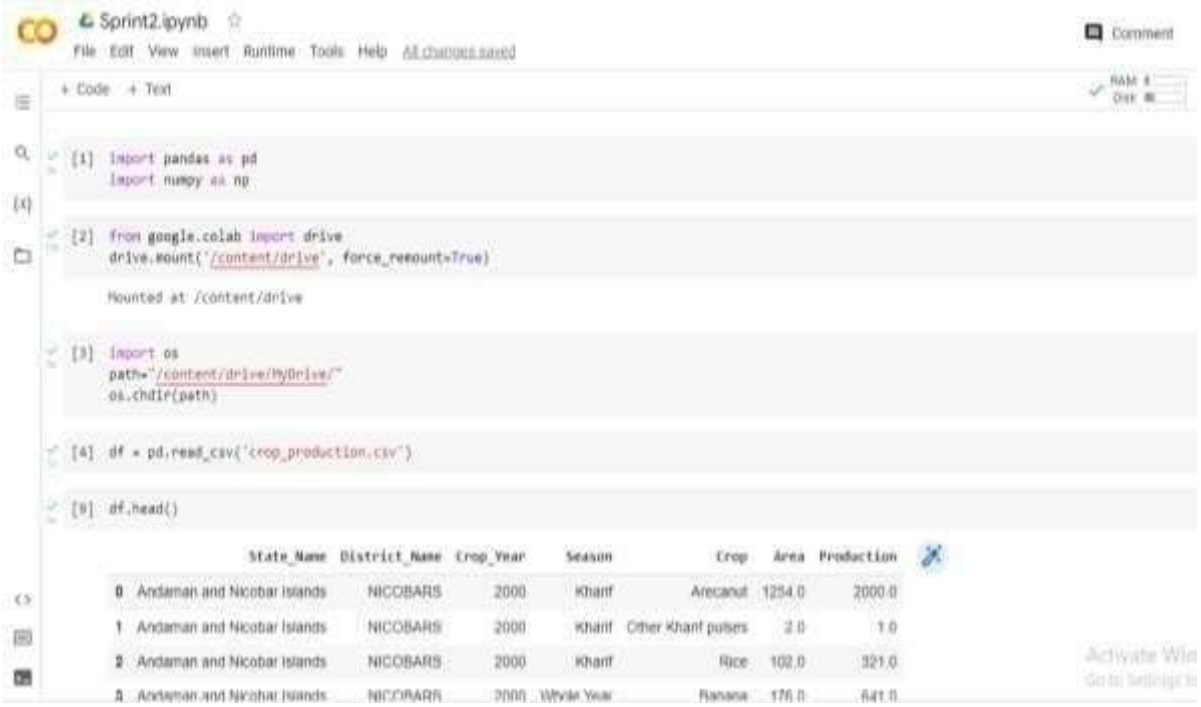
	A	B	C	D	E	F	G	H	I	J
1	State_Nar	District_N	Crop_Year	Season	Crop	Area	Production			
2	Andaman	NICOBARS	2000	Kharif	Arecanut	1254	2000			
3	Andaman	NICOBARS	2000	Kharif	Other Kha	2	1			
4	Andaman	NICOBARS	2000	Kharif	Rice	102	321			
5	Andaman	NICOBARS	2000	Whole Ye	Banana	176	641			
6	Andaman	NICOBARS	2000	Whole Ye	Cashewnu	720	165			
7	Andaman	NICOBARS	2000	Whole Ye	Coconut	18168	65100000			
8	Andaman	NICOBARS	2000	Whole Ye	Dry ginger	36	100			
9	Andaman	NICOBARS	2000	Whole Ye	Sugarcane	1	2			
10	Andaman	NICOBARS	2000	Whole Ye	Sweet pot	5	15			
11	Andaman	NICOBARS	2000	Whole Ye	Tapioca	40	169			
12	Andaman	NICOBARS	2001	Kharif	Arecanut	1254	2061			
13	Andaman	NICOBARS	2001	Kharif	Other Kha	2	1			
14	Andaman	NICOBARS	2001	Kharif	Rice	83	300			
15	Andaman	NICOBARS	2001	Whole Ye	Cashewnu	719	192			
16	Andaman	NICOBARS	2001	Whole Ye	Coconut	18190	64430000			
17	Andaman	NICOBARS	2001	Whole Ye	Dry ginger	46	100			
18	Andaman	NICOBARS	2001	Whole Ye	Sugarcane	1	1			
19	Andaman	NICOBARS	2001	Whole Ye	Sweet pot	11	33			
20	Andaman	NICOBARS	2002	Kharif	Rice	189.2	510.84			
21	Andaman	NICOBARS	2002	Whole Ye	Arecanut	1258	2083			
22	Andaman	NICOBARS	2002	Whole Ye	Banana	213	1278			
23	Andaman	NICOBARS	2002	Whole Ye	Black pep	63	13.5			

crop_production

READY

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,



Sprint2.ipynb

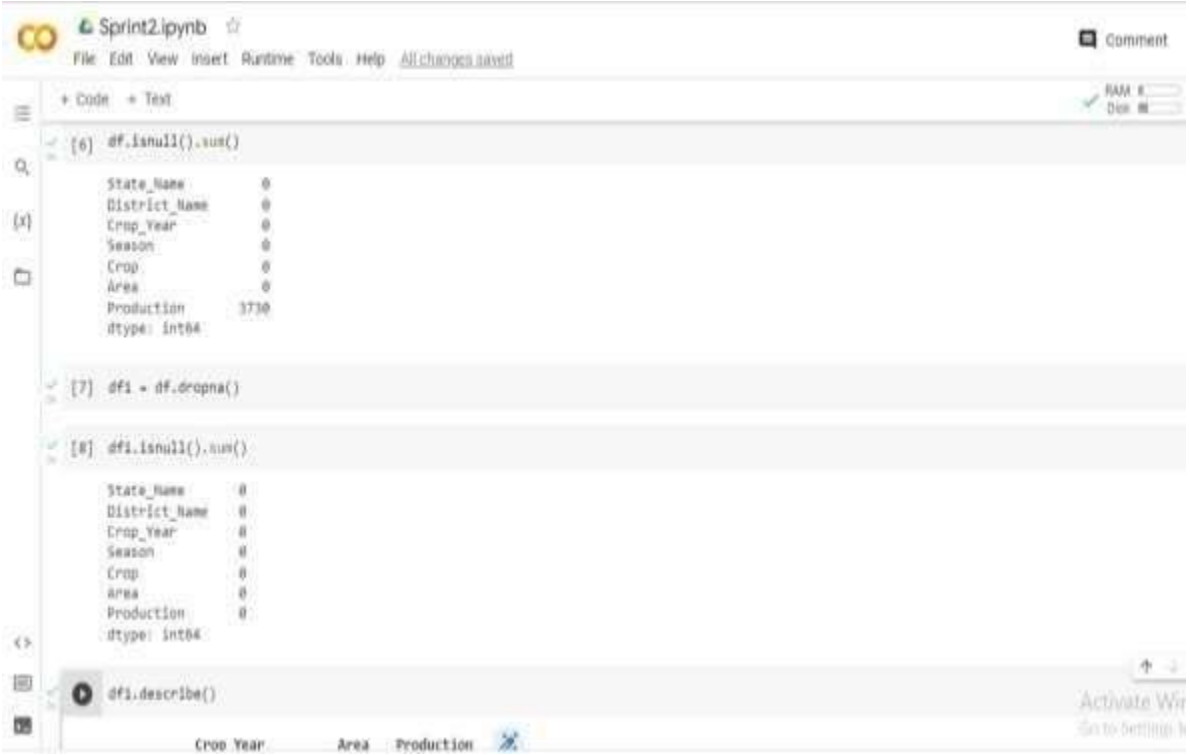
File Edit View Insert Runtime Tools Help All changes saved

Comment

+ Code + Text

```
[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()
```

	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



Sprint2.ipynb

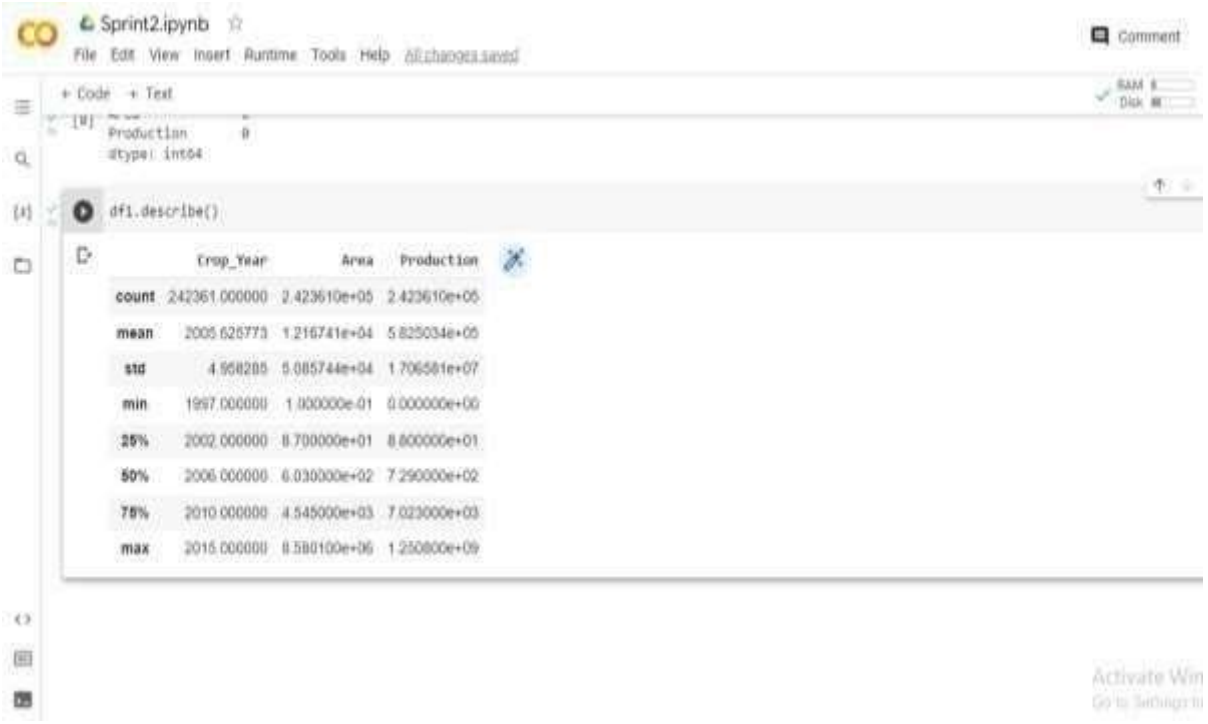
File Edit View Insert Runtime Tools Help All changes saved

Comment

+ Code + Text

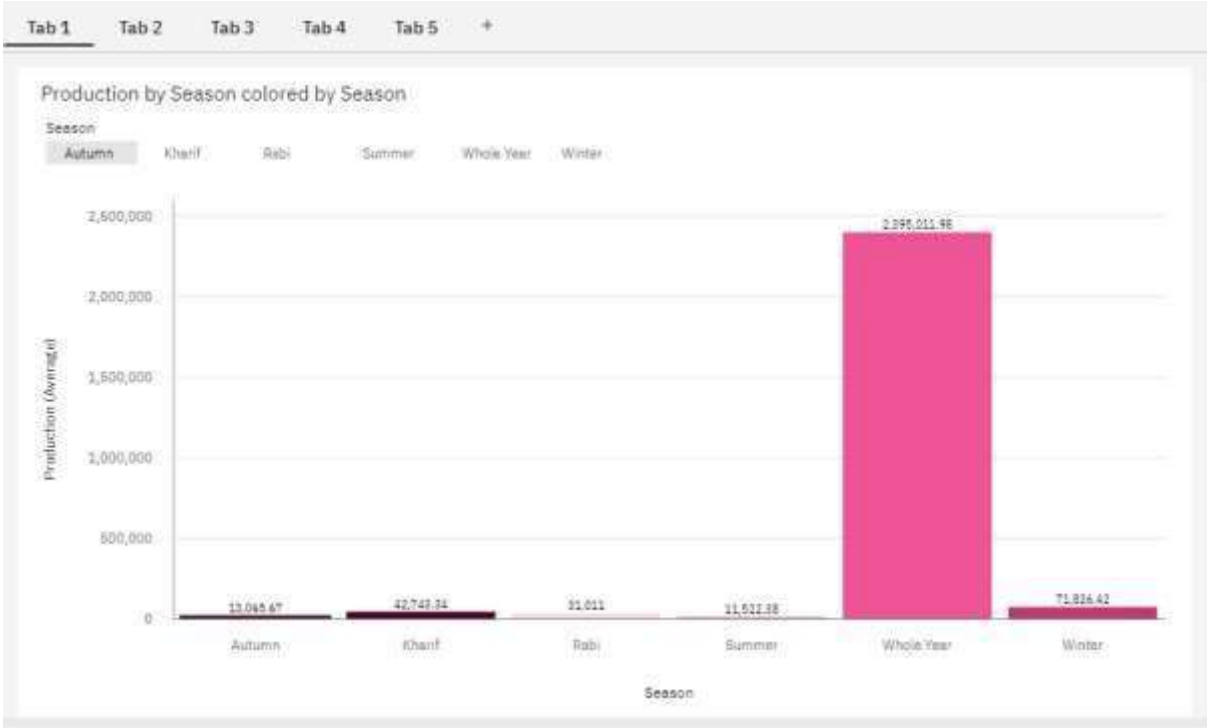
```
[6] df.isnull().sum()

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

[7] df1 = df.dropna()

[8] df1.isnull().sum()

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

    df1.describe()
```

	Crop_Year	Area	Production

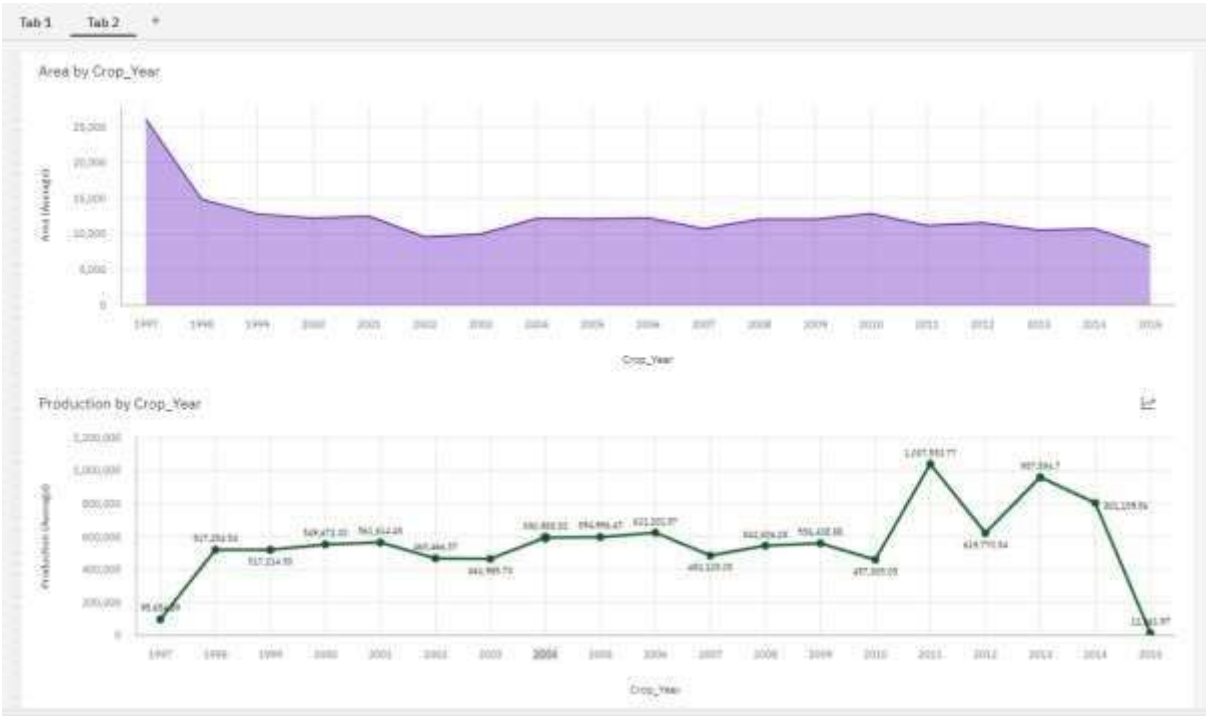


DATA VISUALIZATION CHARTS:

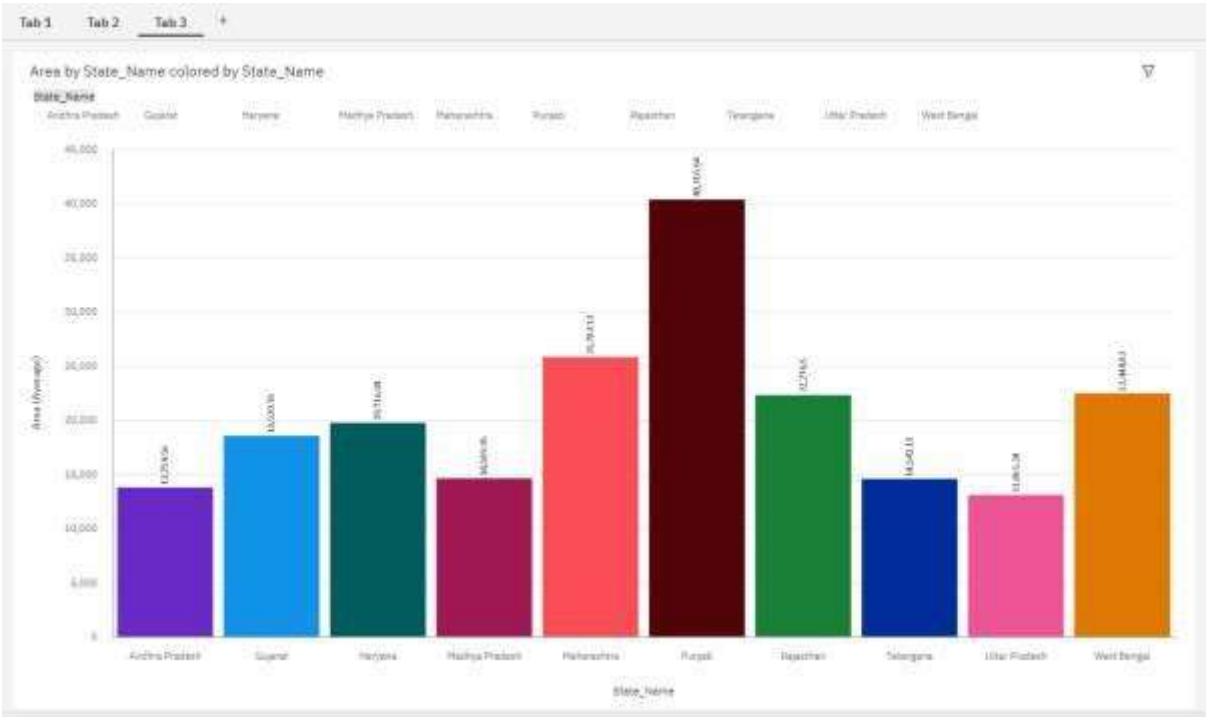
- Visualization on Seasons with average Production



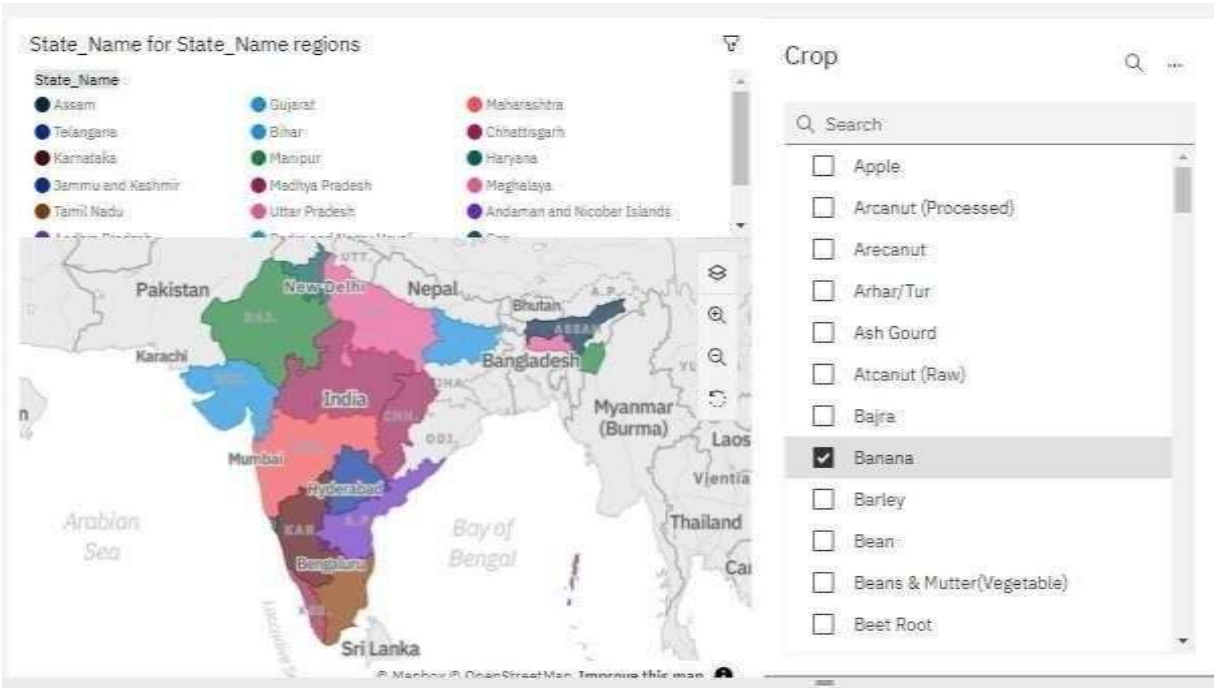
- Visualization with years usage of Area and Production



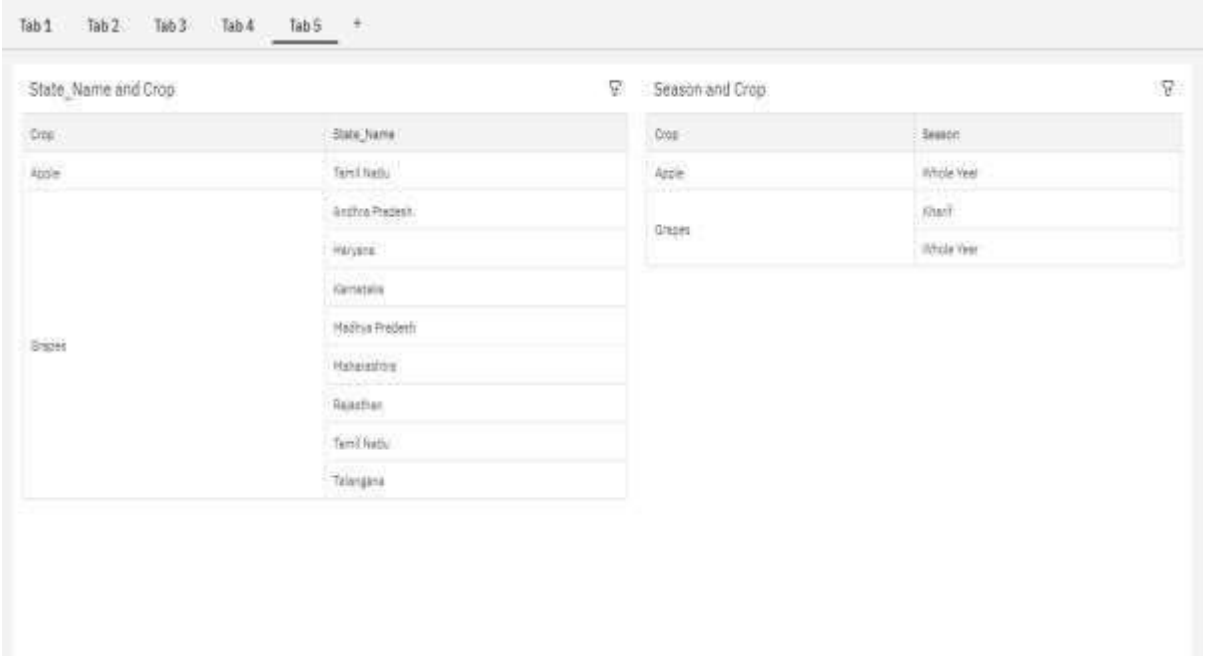
- Visualization on top 10 States with most Area



- Visualization on State with Crop Production



- Visualization on States with the Crop Production along with Season



The above visualizations are created using the COGNOS platform.