UNIFIED MENTOR

PROJECT-4

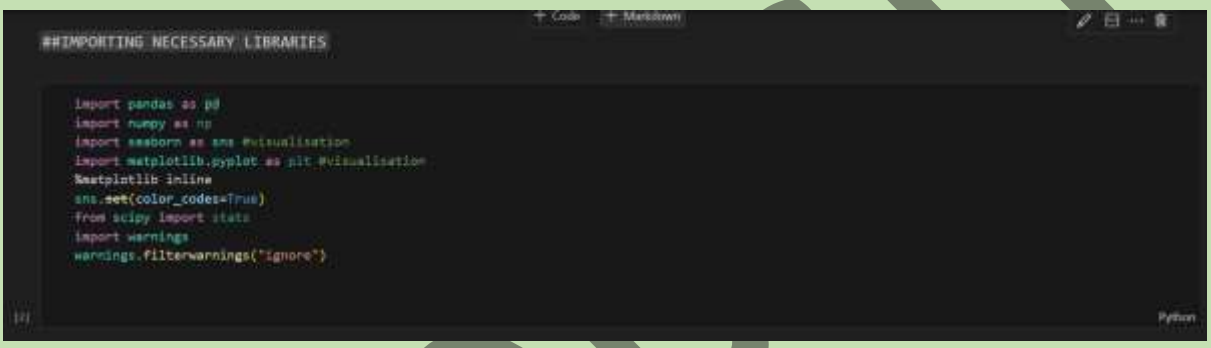
REPORT

ON

CROP PRODUCTION ANALYSIS IN INDIA

STARTING FROM EXAMINING THE FEATURES OF DATA USING JUPYTER NOTEBOOK

-IMPORTING LIBRARIES TO CARRU OUT THE FUNCTIONS ASSOCIATED



-LOAD THE DATA INTO KERNEL

UNIFIED MENTOR

```
## load the csv file
df = pd.read_csv("C:/Users/shamb/Downloads/Crop Production data.csv")
df
```

✓ 0.2s

	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
...	...	...	...	...	...	...	...
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

246091 rows × 7 columns

BRIEF ACCOUNTANCE

```
df.head()
```

✓ 0.0s

	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

RUN THROUGH THE DATA AND ITS DESCRIPTION

```
[4]  df.info()
     ✓ 0.0s
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 246091 entries, 0 to 246090
Data columns (total 7 columns):
 #   Column         Non-Null Count   Dtype  
---  ------         --------------   -----  
 0   State_Name     246091 non-null  object 
 1   District_Name  246091 non-null  object 
 2   Crop_Year      246091 non-null  int64  
 3   Season         246091 non-null  object 
 4   Crop           246091 non-null  object 
 5   Area           246091 non-null  float64
 6   Production     242361 non-null  float64
dtypes: float64(2), int64(1), object(4)
memory usage: 13.1+ MB
```

GOING THROUGH EACH FIELD

Now we observe the each features present in the dataset.

`State_Name:` Name of the state where the crop is grown.
`District_Name:` Name of the district within the state where the crop is cultivated.
`Crop_Year:` Year in which the crop was harvested.
`Season:` The growing season during which the crop was produced (e.g., Rabi, Kharif).
`CropArea:` The total area of land used for growing the crop.
`Production:` The total amount of crop produced in the given area.

CHECKING IF DUPLICATED FIEDLS PRESENT

Dropping the duplicate rows

```
data = df.drop_duplicates()
data.to_csv('C:/Users/shamb/Downloads/Crop Production data.csv', index=False)

# Save the cleaned dataset to a new file
data
```

[7] ✓ 0.7s

	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
...	...	...	...	...	...	...	...
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

	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
...	...	...	...	...	...	...	...
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

246091 rows × 7 columns

No duplicates present as rows and columns are equal to values before we have applied the function.

CHECKING IF ANY VALUE IS MISSING OR NOT

```
# Drop rows with any missing values
data1 = data.dropna()
data1.to_csv('C:/Users/shamb/Downloads/Crop Production data.csv', index=False)
data1
```

[8] ✓ 0.7s

	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
...	...	...	...	...	...	...	...
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

242361 rows × 7 columns

Correct data to be updated after removing the missing values.

BRIEF STATISTICS TERMS TO BE CALCULATED.

```
numerical_stats = data1.describe()
numerical_stats
```

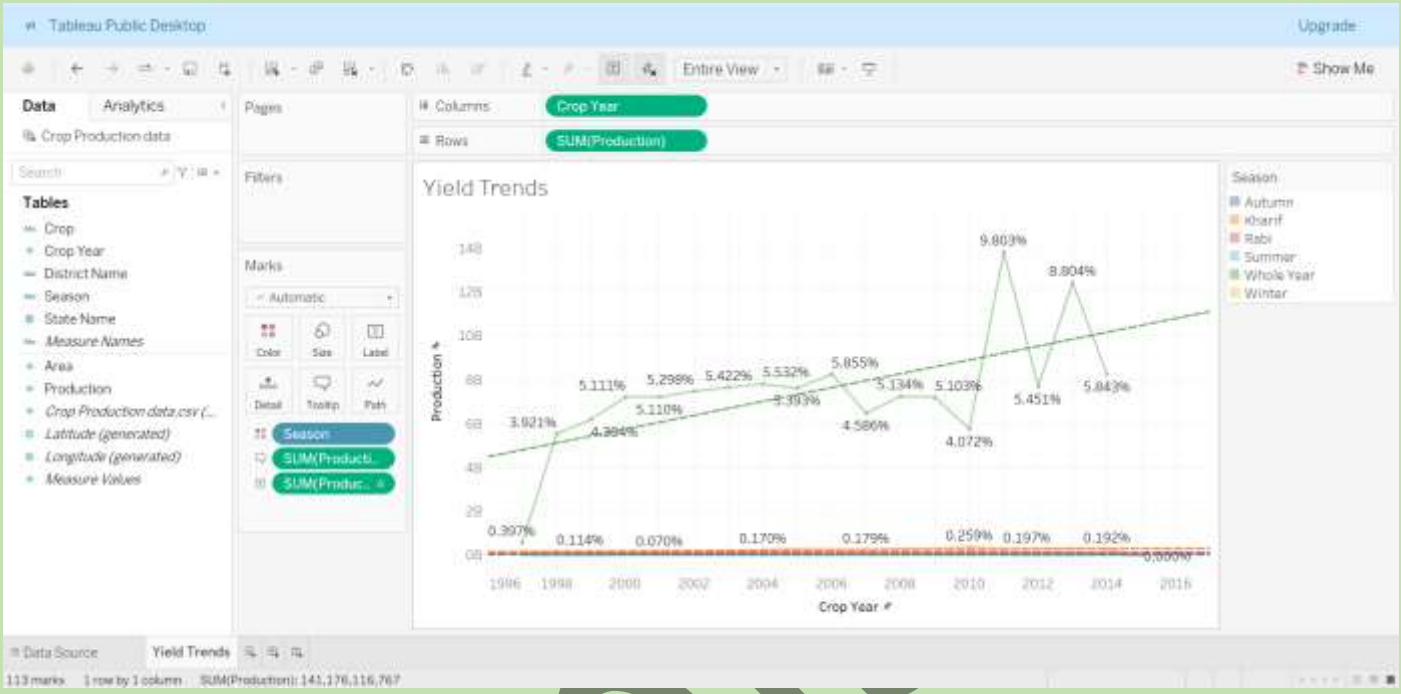
[9] ✓ 0.0s

	Crop_Year	Area	Production
count	242361.000000	2.423610e+05	2.423610e+05
mean	2005.625773	1.216741e+04	5.825034e+05
std	4.958285	5.085744e+04	1.706581e+07
min	1997.000000	1.000000e-01	0.000000e+00
25%	2002.000000	8.700000e+01	8.800000e+01
50%	2006.000000	6.030000e+02	7.290000e+02
75%	2010.000000	4.545000e+03	7.023000e+03
max	2015.000000	8.580100e+06	1.250800e+09

BUILDING UP THE VISUALIZATION THROUGH TABLEAU

1)YIELD TRENDS

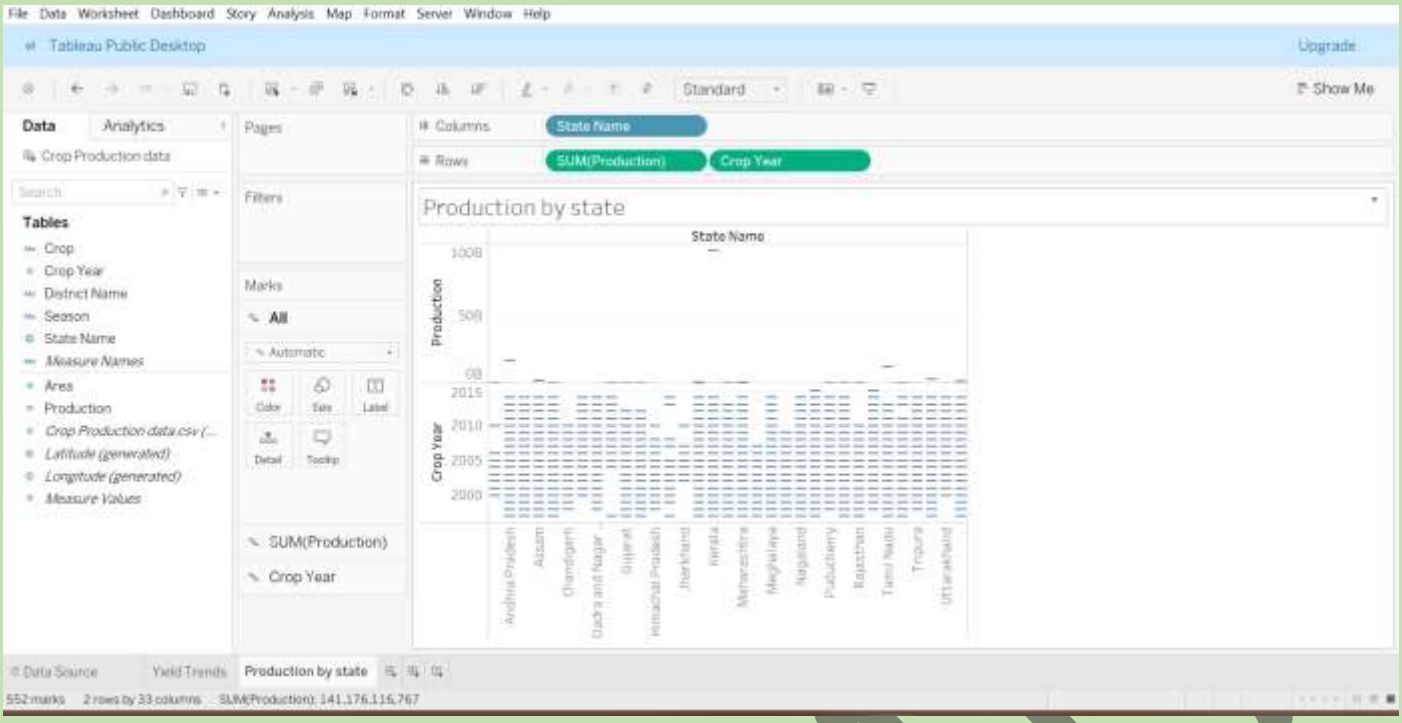
- **Seasonal Patterns**: Identify trends in crop yields across different seasons and years. This can help predict future production and plan for seasonal fluctuations.
- **Geographic Variations**: Compare yields across different regions to determine which areas are most productive and why.



2)PRODUCTION BY STATE

• **Strategic Resource Allocation:**

- **Focus on High-Profit Areas:** Allocate resources and investment towards sales channels, item types, and regions that show high profitability to maximize returns.
- **Cost Management in Low-Profit Areas:** Implement cost control measures in areas with lower profitability to improve margins and overall financial health.

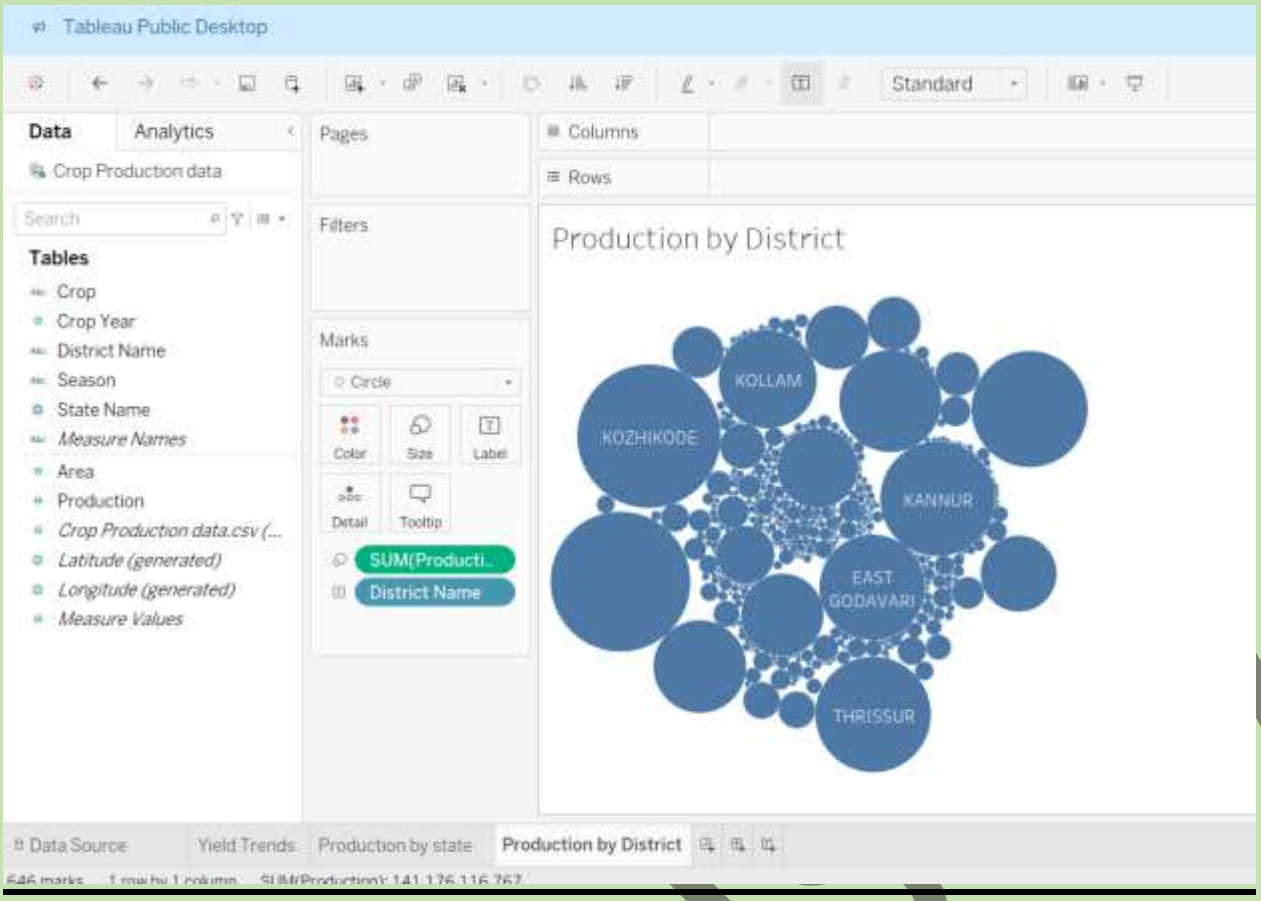


Production by District

Production by District

Insights:

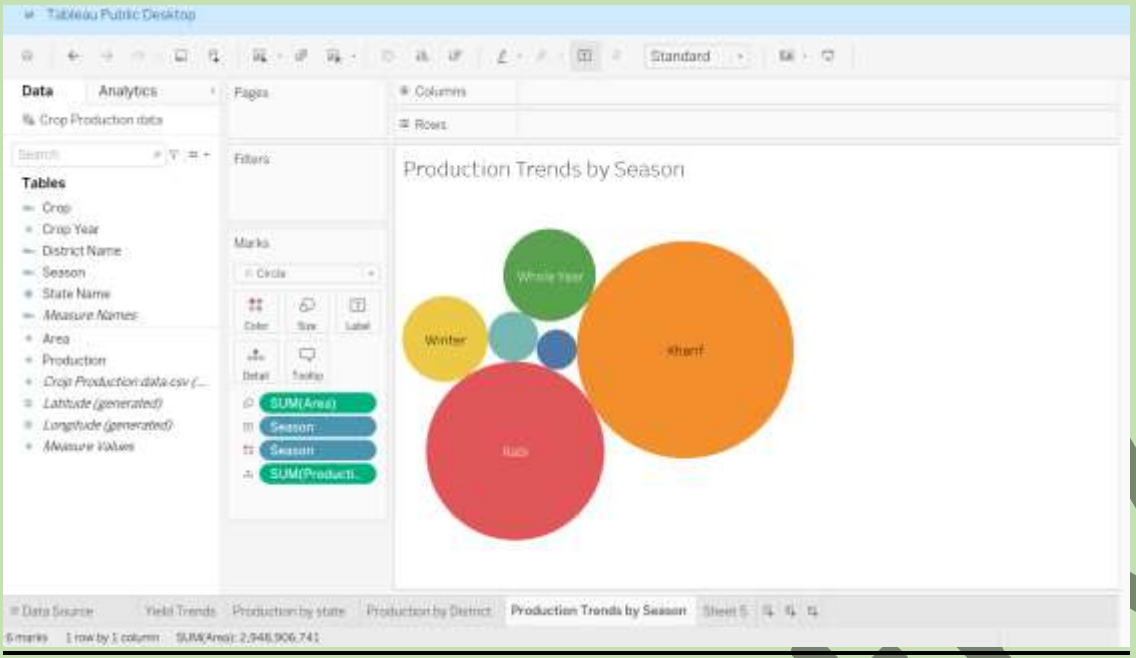
1. **District-Level Production Comparison**: Evaluate how different districts within a state contribute to the overall production. This can identify high-performing districts and areas that may need more support or resources.
2. **District Production Variability**: Analyze variations in production levels across districts to understand local differences in crop yield. This insight helps in pinpointing regions with unusual production patterns or inefficiencies.
3. **Impact of District Characteristics**: Explore how the characteristics of different districts (e.g., land area, soil quality) might influence production levels. This can provide insights into the factors driving production success or challenges.



PRODUCTION BY SEASON

- **Seasonal Yield Variations**: Examine how production levels vary by season within the same year. This can highlight which seasons are more productive and help optimize planting and harvesting schedules.

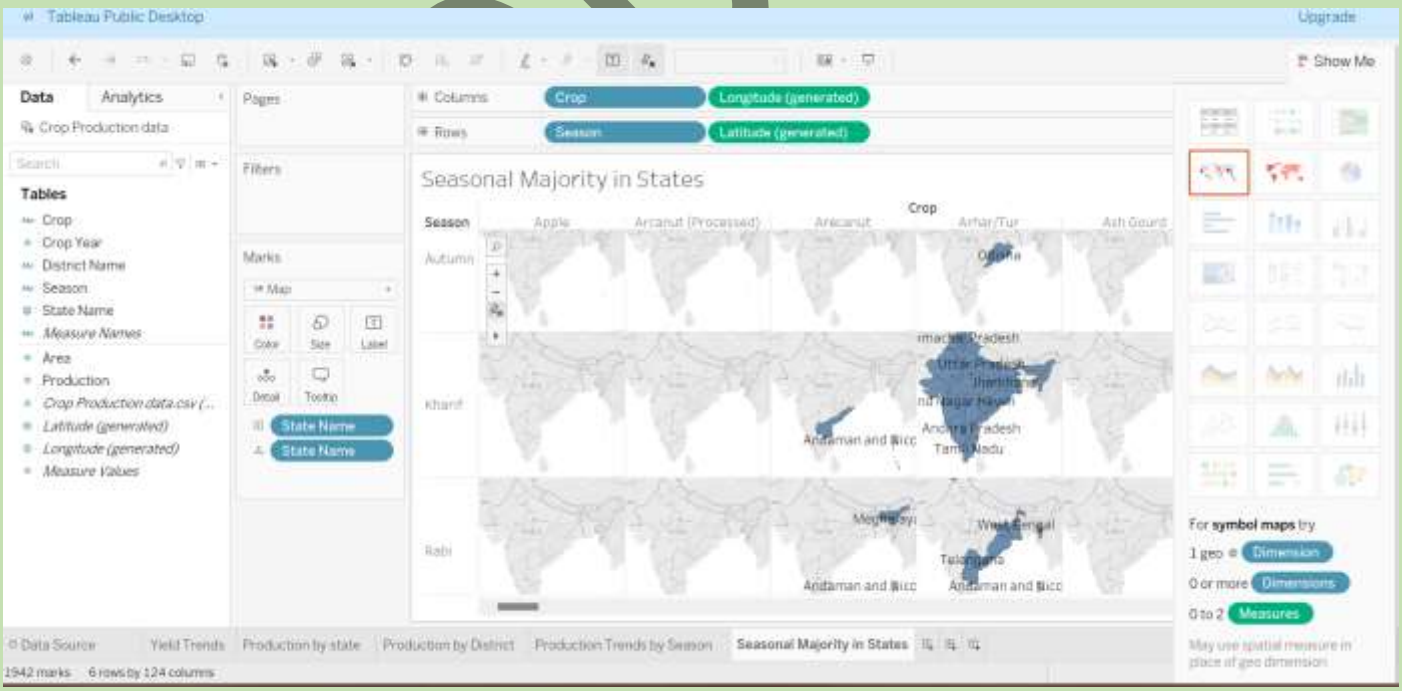
- **Seasonal Effectiveness Over Years**: Analyze if certain seasons consistently result in higher production over multiple years. This insight is valuable for understanding the impact of seasonal growing conditions on overall yield.



Seasonal Majority in States

- **Seasonal Production Differences**: Compare production levels across different seasons within a single year to identify which seasons are most productive. This can inform decisions on crop scheduling and resource allocation.

- **Effectiveness of Each Season**: Analyze how different seasons contribute to production to understand their relative effectiveness. This helps in planning for future crop cycles and improving seasonal strategies.



Area visualization

1. **egional Land Use Patterns**

Insight: **Understand regional variations in land use for crop production**.

- **Details**: Observing land area distribution by state can reveal regional patterns or clusters of high and low land use. This insight can be useful for identifying agricultural zones or regions with significant land use changes that may be influenced by local agricultural policies, climate conditions, or economic factors.

2. **Trends in Land Use Over Time**

Insight: **Analyze how land area used for cropping has changed over time within each state**.

- **Details**: By visualizing changes in crop area over multiple years, you can identify trends such as expanding or contracting agricultural land. This can indicate shifts in agricultural practices, such as increased investment in certain states or changes in land availability.

Land utilisation efficiency

- **efficiency by State**

Insight:**Identify which states have the highest and lowest land utilization efficiency**.

- **Details**: By comparing production levels relative to land area across states, you can identify which states are using their agricultural land most efficiently. States with high production per unit of land are more efficient, while those with lower production relative to land area may have room for improvement or face challenges.

- **Trends in Efficiency Over Time**

Insight: **Analyze how land utilization efficiency has changed over the years**.

- **Details**: Tracking changes in efficiency over time can reveal trends such as improvements in land use practices, changes in agricultural productivity, or impacts of policies and investments. This helps in understanding whether land use efficiency is improving or declining over time.

Dashboard

UNIFIED MENTOR