Crop Production Analysis in India

Neenu E S

PROJECT DETAILS

Project Title	Crop Production Analysis in India
Technology	Data Science
Domain	Agriculture
Project Difficulty level	Advanced
Programming Language Used	Python
Tools Used	Jupyter Notebook, MS-Excel

Objective

- Predict crop production.

- Find important insights highlighting key indicators.

- Find metrics that influence crop production.

Problem Statement

- The Agriculture business domain, as a vital part of the overall supply chain, is expected to highly evolve in the upcoming years via the developments, which are taking place on the side of the Future Internet.

- This paper presents a novel Business-to-Business collaboration platform from the agri-food sector perspective, which aims to facilitate the collaboration of numerous stakeholders belonging to associated business domains, in an effective and flexible manner.

- This dataset provides a huge amount of information on crop production in India ranging from several years.

- Based on the Information the ultimate goal would be to predict crop production and find important insights highlighting key indicators and metrics that influence crop production.

DATA TRANSFORMATION AND EDA

```
ead in the input file as a Pandas DataFrame

= pd.read_csv('Crop Production data.csv')

imensions of the dataset is 246091 rows and 7 columns
.shape #(246091, 7)

091, 7)

.info()

ss 'pandas.core.frame.DataFrame'>
eIndex: 246091 entries, 0 to 246090
 columns (total 7 columns):
 Column        Non-Null Count    Dtype
 ------        --------------    -----
 State_Name    246091 non-null   object
 District_Name 246091 non-null   object
 Crop_Year     246091 non-null   int64
 Season        246091 non-null   object
 Crop          246091 non-null   object
 Area          246091 non-null   float64
 Production    242361 non-null   float64
es: float64(2), int64(1), object(4)
ry usage: 13.1+ MB

, we have

four categorical variables - 'State_Name','District_Name', 'Season' and 'Crop type'
three continuous variables in float - 'Area' and 'Production'
nteger - 'Crop_Year' format.
```

```
# drop the duplicated rows
data = data.drop_duplicates()
data.shape

(246091, 7)
```

They is no duplicated rows in the data.

Checking for Missing Data

```
data.isnull().sum()

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

'Production' variable shows huge number of missing values - 3730 and no missing values for other variables. We can drop missing value samples as we have sufficiently big dataset.

```
data.shape

(246091, 7)
```

```
## Here we remove missing value from the production
data.dropna(subset=["Production"],axis=0,inplace=True)
```

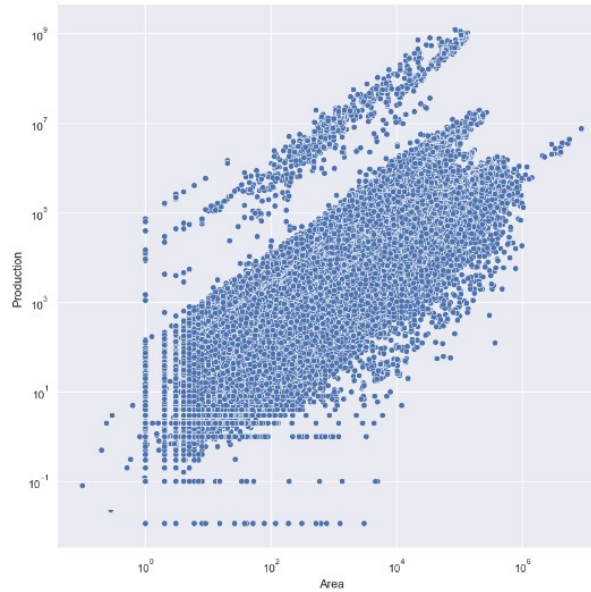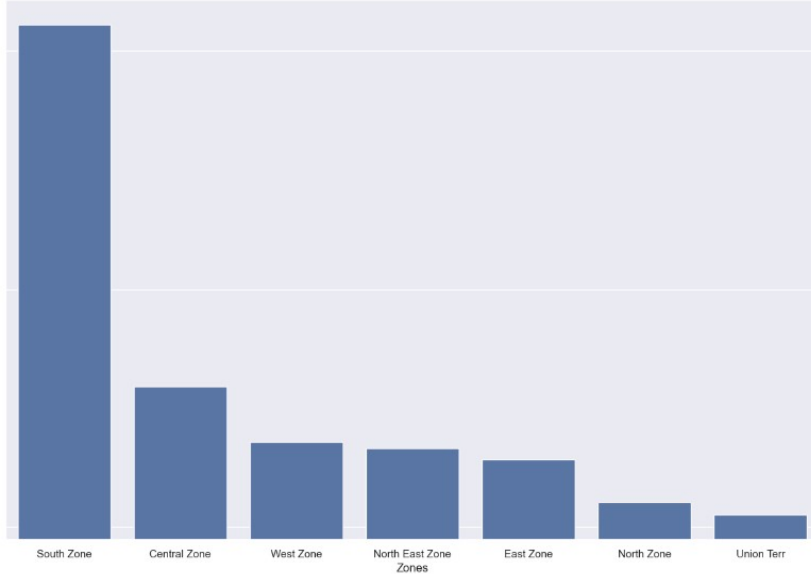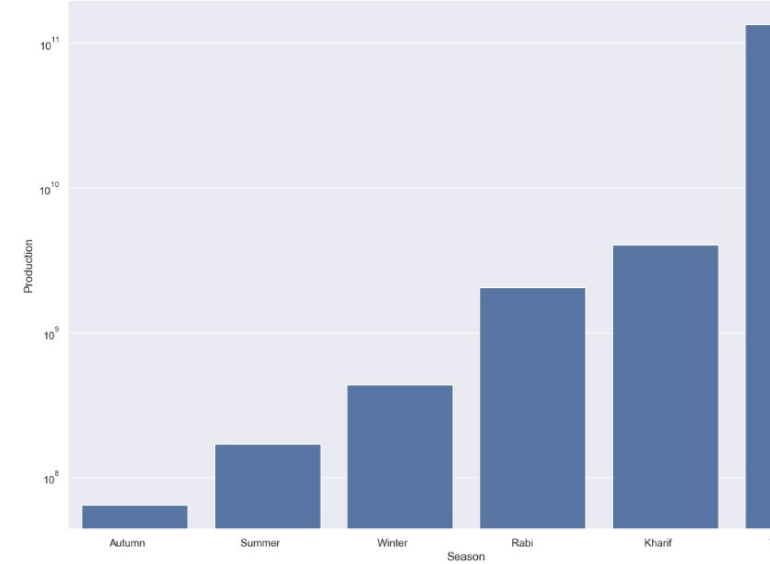```
data.shape

(242361, 7)
```

```
data.isnull().sum()

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

Now, there is now missing values.
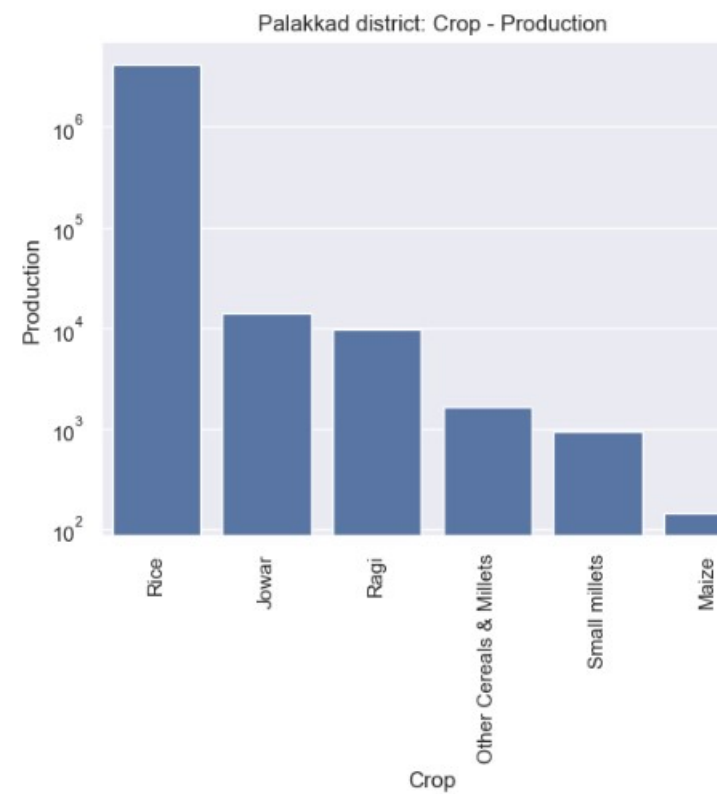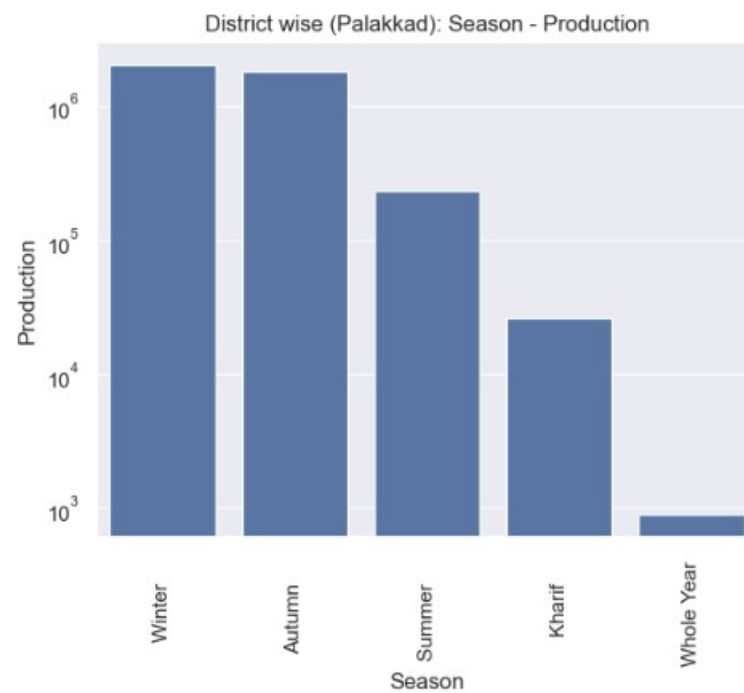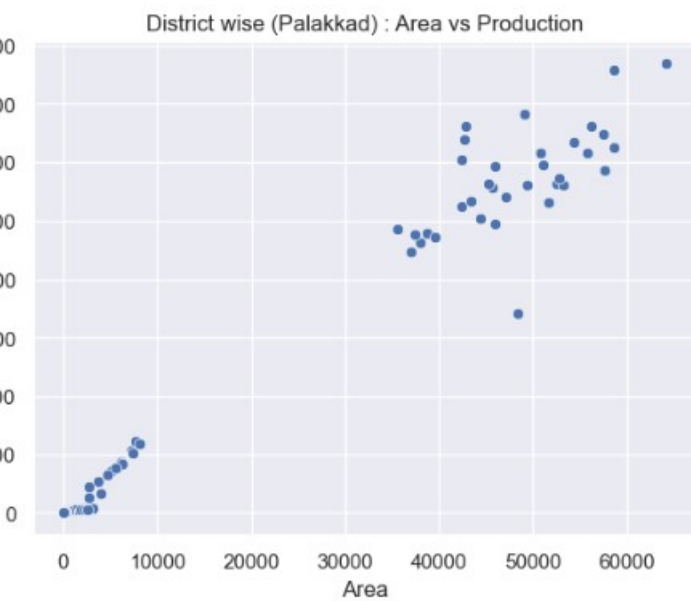
Crop Production Analysis in India

Visual Analytics And Key Insights

• This illustrates how the variables of regions, seasons, and zones impact crop productivity.

• Based on the data, we can conclude that these three factors are critical to crop productivity in India.

Crop Production Analysis in India
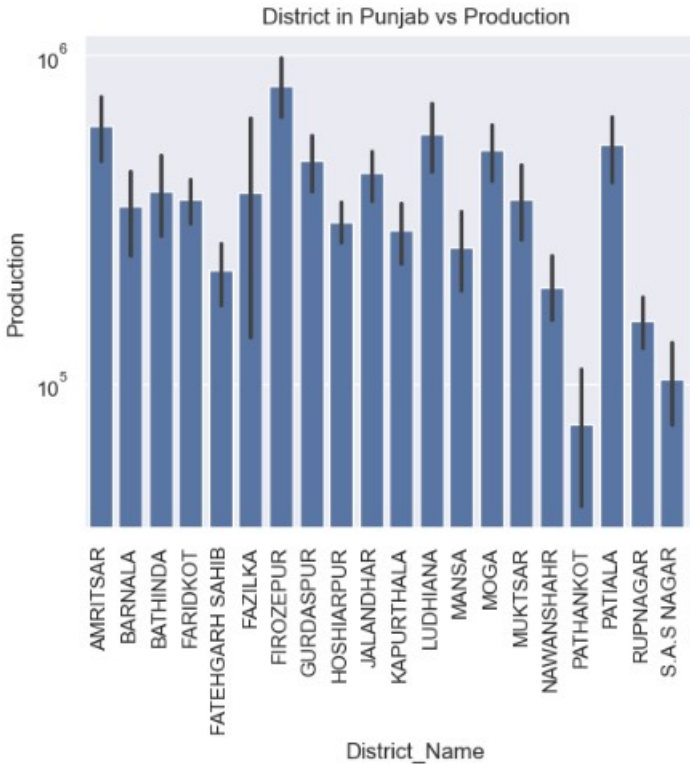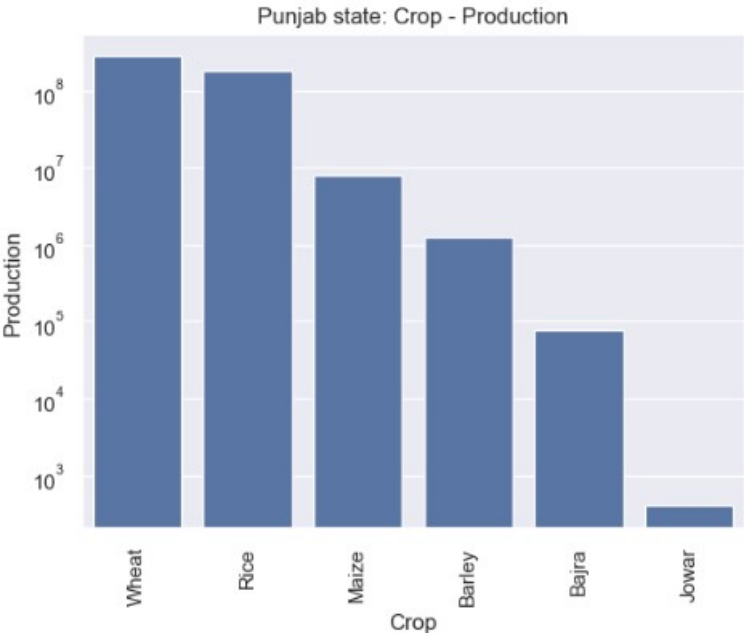
Zones (south zone)
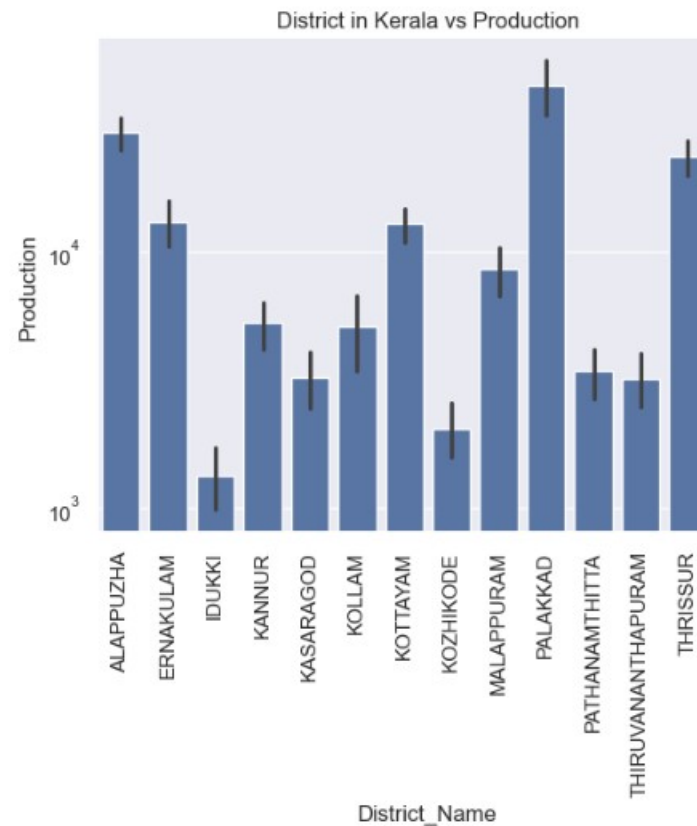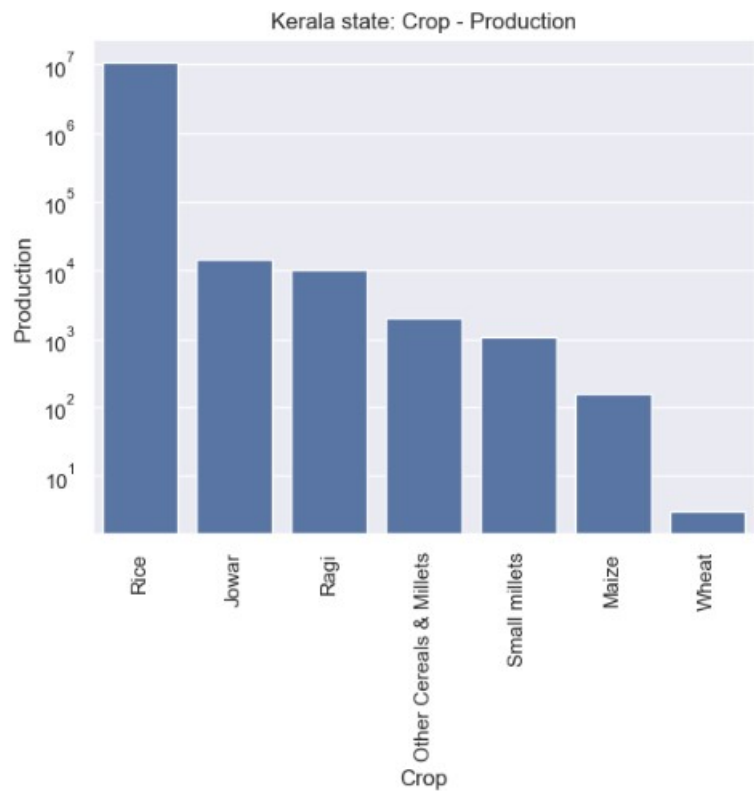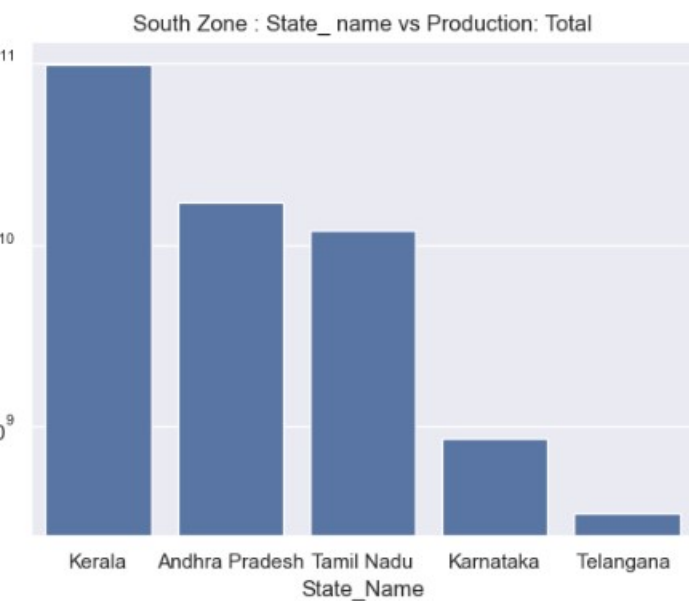
South Zone : State_name vs Production: Total

Kerala state: Crop - Production

District in Kerala vs Production

Crop Production Analysis in India

District wise (Palakkad) : Area vs Production

District wise (Palakkad): Season - Production
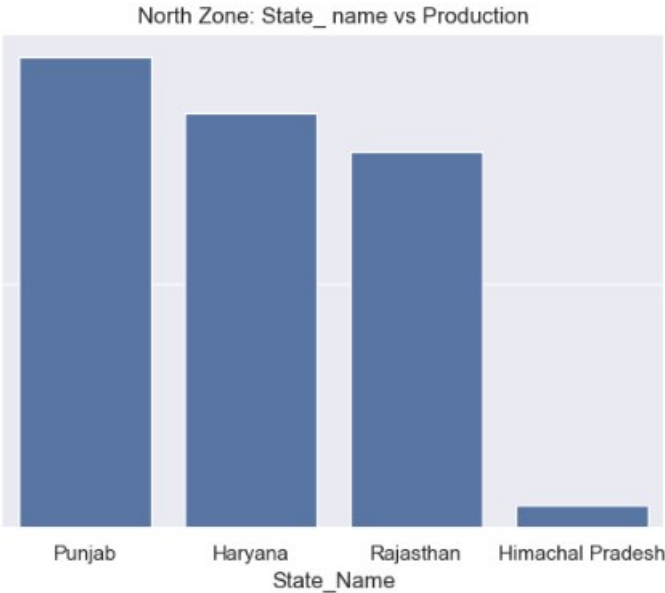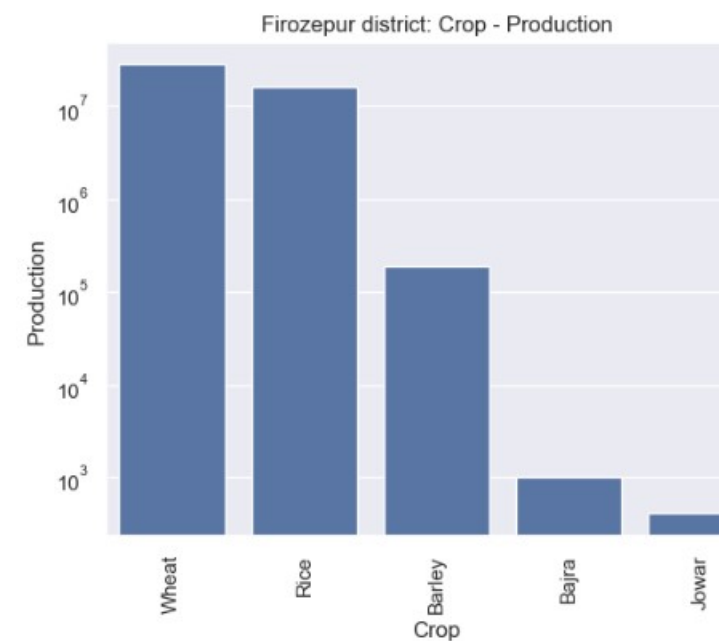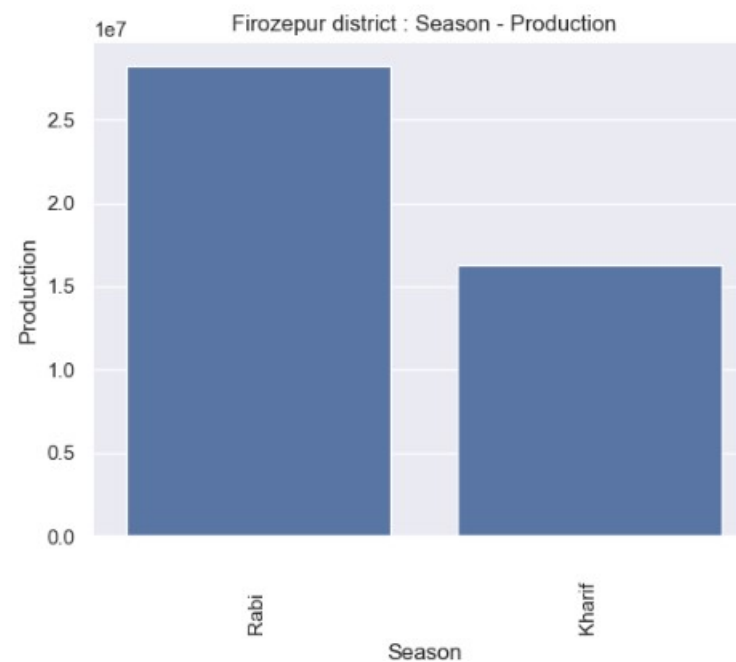
Palakkad district: Crop - Production

•Area has the impact in the production of crop then other variable such zone and season.

Crop Production Analysis in India

Zones (north zone)

Crop Production Analysis in India

District wise (Firozepur) : Area vs Production — Firozepur district : Season - Production — Firozepur district: Crop - Production
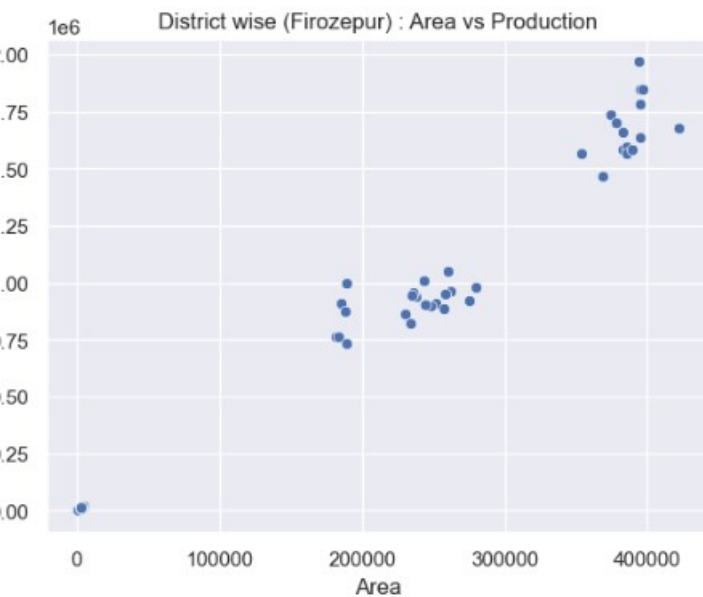
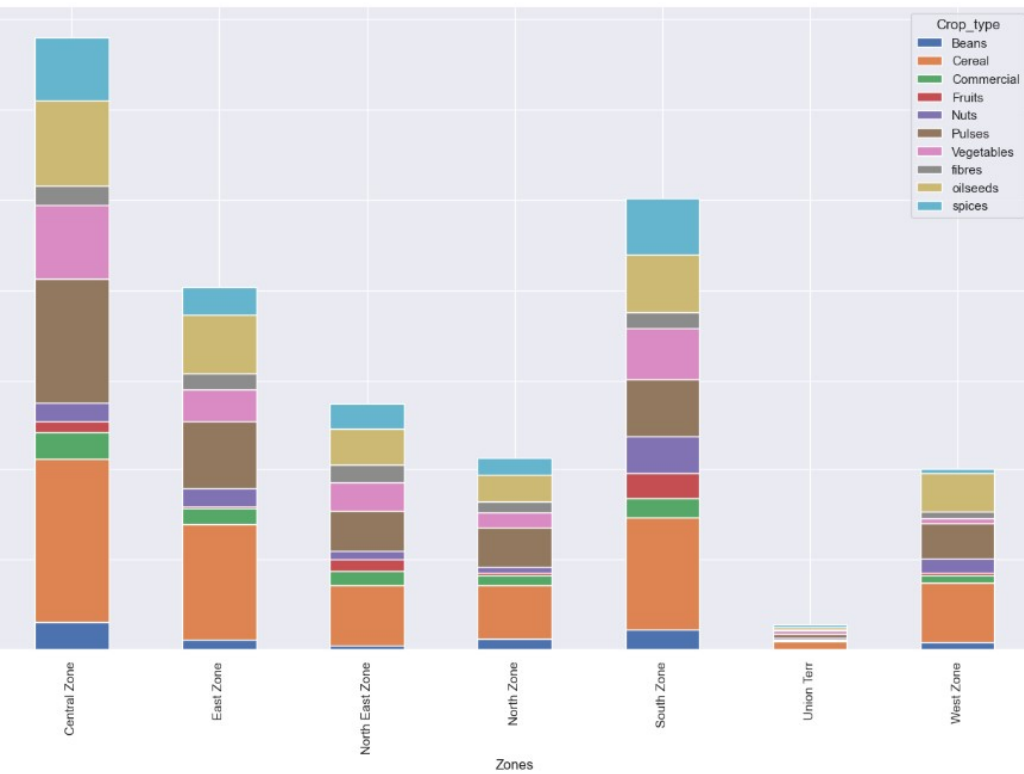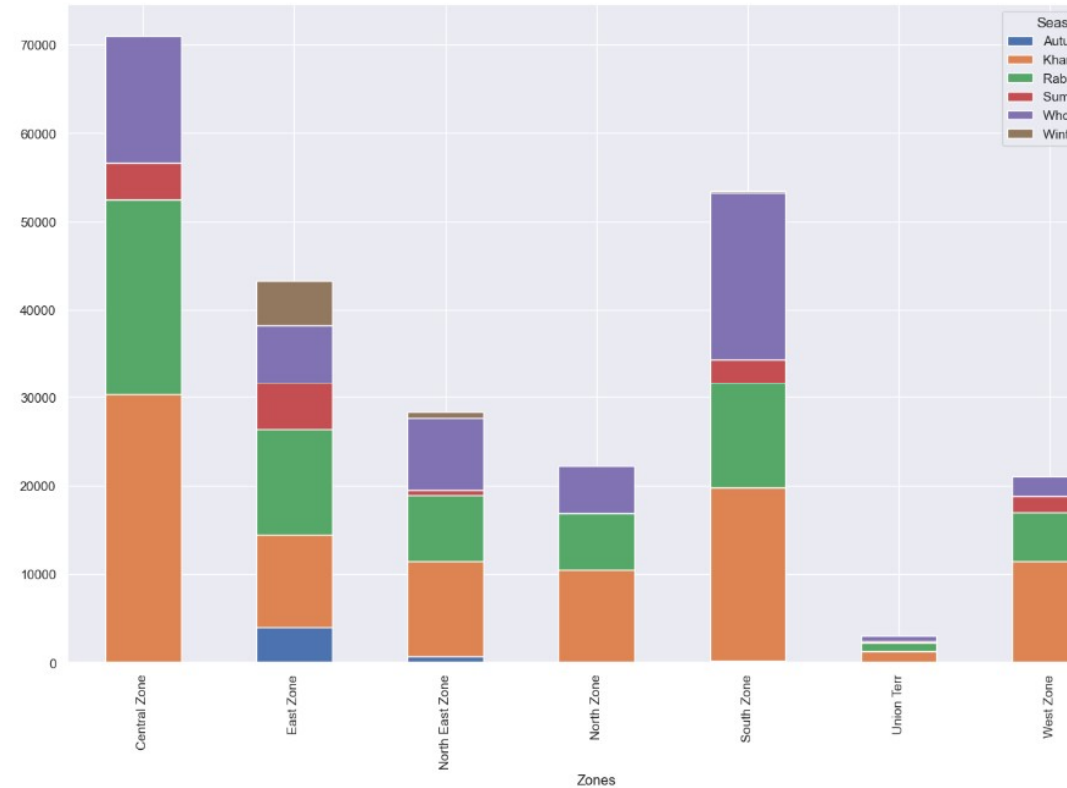•Area has the impact in the production of crop then other variable such zone and season. As similar that got from the south zone data.
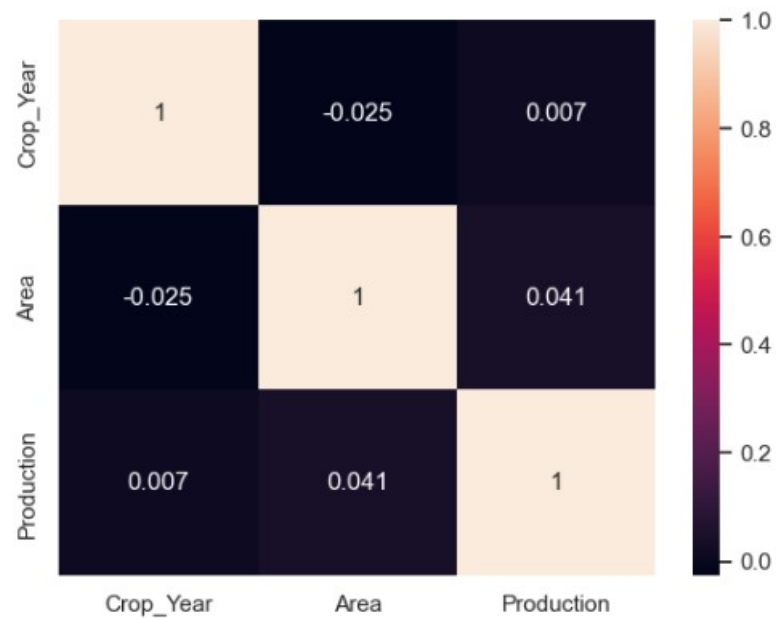
Zones vs. crop type

Zones vs. Season

• From the all data analysis we can find the production of a crop in different state within different zone.

Crop Production Analysis in India

Relation between the variables

Conclusions:

• Crop production from 1997 to 2015 can be examined.

•We used basic visualisations to gain an idea of the distribution of data based on the provided data.

•We only need to subset data sets according to state name, district name and crops type to do basic analyses.

•Area, Season and zone (state, district) all have a significant impact on agricultural yield.

•Analyses how different crops are produced in India's zones as well as how crop output is impacted by zone-specific seasons.

Crop Production Analysis in India

Thank you