Importing Data

```
import pandas as pd
#importing panda library with an alias pd

#Reading csv files into a variable df
df = pd.read_csv("crop_production.csv")
df
```

```
                        State_Name District_Name  Crop_Year
Season  \
0       Andaman and Nicobar Islands      NICOBARS       2000  Kharif

1       Andaman and Nicobar Islands      NICOBARS       2000  Kharif

2       Andaman and Nicobar Islands      NICOBARS       2000  Kharif

3       Andaman and Nicobar Islands      NICOBARS       2000  Whole
Year
4       Andaman and Nicobar Islands      NICOBARS       2000  Whole
Year
...                             ...           ...        ...
...
246086                  West Bengal       PURULIA       2014  Summer

246087                  West Bengal       PURULIA       2014  Summer

246088                  West Bengal       PURULIA       2014  Whole
Year
246089                  West Bengal       PURULIA       2014  Winter

246090                  West Bengal       PURULIA       2014  Winter


                     Crop      Area  Production
0                Arecanut    1254.0      2000.0
1       Other Kharif pulses       2.0         1.0
2                    Rice     102.0       321.0
3                  Banana     176.0       641.0
4               Cashewnut     720.0       165.0
...                   ...       ...         ...
246086               Rice     306.0       801.0
246087             Sesamum     627.0       463.0
246088          Sugarcane     324.0     16250.0
246089               Rice  279151.0    597899.0
246090             Sesamum     175.0        88.0

[246091 rows x 7 columns]
```

```
df.head(10)
```

```
                  State_Name District_Name  Crop_Year
Season  \
0  Andaman and Nicobar Islands      NICOBARS       2000  Kharif

1  Andaman and Nicobar Islands      NICOBARS       2000  Kharif

2  Andaman and Nicobar Islands      NICOBARS       2000  Kharif

3  Andaman and Nicobar Islands      NICOBARS       2000  Whole Year

4  Andaman and Nicobar Islands      NICOBARS       2000  Whole Year

5  Andaman and Nicobar Islands      NICOBARS       2000  Whole Year

6  Andaman and Nicobar Islands      NICOBARS       2000  Whole Year

7  Andaman and Nicobar Islands      NICOBARS       2000  Whole Year

8  Andaman and Nicobar Islands      NICOBARS       2000  Whole Year

9  Andaman and Nicobar Islands      NICOBARS       2000  Whole Year


                 Crop     Area  Production
0             Arecanut   1254.0      2000.0
1  Other Kharif pulses      2.0         1.0
2                 Rice    102.0       321.0
3               Banana    176.0       641.0
4            Cashewnut    720.0       165.0
5             Coconut   18168.0  65100000.0
6           Dry ginger     36.0       100.0
7            Sugarcane      1.0         2.0
8         Sweet potato      5.0        15.0
9              Tapioca     40.0       169.0

df.tail(10)

         State_Name District_Name  Crop_Year       Season
Crop  \
246081  West Bengal       PURULIA       2014  Rabi         Rapeseed
&Mustard
246082  West Bengal       PURULIA       2014  Rabi
Safflower
246083  West Bengal       PURULIA       2014  Rabi
Urad
246084  West Bengal       PURULIA       2014  Rabi
Wheat
246085  West Bengal       PURULIA       2014  Summer
Maize
```

```
246086  West Bengal       PURULIA       2014  Summer
Rice
246087  West Bengal       PURULIA       2014  Summer
Sesamum
246088  West Bengal       PURULIA       2014  Whole Year
Sugarcane
246089  West Bengal       PURULIA       2014  Winter
Rice
246090  West Bengal       PURULIA       2014  Winter
Sesamum

            Area  Production
246081    1885.0      1508.0
246082      54.0        37.0
246083     220.0       113.0
246084    1622.0      3663.0
246085     325.0      2039.0
246086     306.0       801.0
246087     627.0       463.0
246088     324.0     16250.0
246089  279151.0    597899.0
246090     175.0        88.0
```

```
df.sample(10)
```

```
           State_Name     District_Name  Crop_Year       Season  \
109430  Madhya Pradesh             GUNA       2002  Whole Year
3729    Andhra Pradesh           KADAPA       2009  Kharif
37765            Bihar       MUZAFFARPUR       2014  Autumn
141932       Meghalaya  WEST KHASI HILLS       2014  Kharif
120990  Madhya Pradesh          SHAHDOL       2002  Whole Year
209910   Uttar Pradesh       FARRUKHABAD       2010  Summer
57521          Gujarat         AHMADABAD       2012  Kharif
215654   Uttar Pradesh           JAUNPUR       2013  Rabi
166342       Rajasthan            BARMER       2008  Rabi
112621  Madhya Pradesh             KATNI       1999  Whole Year

                     Crop     Area  Production
109430         Pome Fruit     62.0         0.0
3729                 Rice  52423.0    153757.0
37765                Rice  31319.0     69515.0
141932             Banana    521.0      3032.0
120990   Other Vegetables    262.0         0.0
209910  Moong(Green Gram)    768.0       789.0
57521             Sesamum   1600.0       800.0
215654              Onion    366.0      7444.0
166342            Linseed     13.0        12.0
112621             Garlic     51.0       158.0
```

Data Cleaning

Checking wheather there is any null values

```
#use ".isnull().sum()" to check for null values

df.isnull().sum()
```

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

As we can see we have null values present in production, we can either drop the entire rows that contains null values or handle null values by using mean, median or mode methods. Here we intend to delete the rows to keep up with the data integrity.

```
df = df.dropna()
df.isnull().sum()
```

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

Now we have handled the Null values present in dataset

Checting for wrong data form

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 242361 entries, 0 to 246090
Data columns (total 7 columns):
 #   Column         Non-Null Count   Dtype
---  ------         --------------   -----
 0   State_Name     242361 non-null  object
 1   District_Name  242361 non-null  object
 2   Crop_Year      242361 non-null  int64
 3   Season         242361 non-null  object
 4   Crop           242361 non-null  object
 5   Area           242361 non-null  float64
 6   Production     242361 non-null float64
dtypes: float64(2), int64(1), object(4)
memory usage: 14.8+ MB
```

```
df['State_Name'].unique()
# Checking for any wrong data present under state_name columns
```

```
array(['Andaman and Nicobar Islands', 'Andhra Pradesh',
       'Arunachal Pradesh', 'Assam', 'Bihar', 'Chandigarh',
       'Chhattisgarh', 'Dadra and Nagar Haveli', 'Goa', 'Gujarat',
       'Haryana', 'Himachal Pradesh', 'Jammu and Kashmir ',
'Jharkhand',
       'Karnataka', 'Kerala', 'Madhya Pradesh', 'Maharashtra',
'Manipur',
       'Meghalaya', 'Mizoram', 'Nagaland', 'Odisha', 'Puducherry',
       'Punjab', 'Rajasthan', 'Sikkim', 'Tamil Nadu', 'Telangana ',
       'Tripura', 'Uttar Pradesh', 'Uttarakhand', 'West Bengal'],
      dtype=object)
```

```
df['District_Name'].unique()
# Checking for any wrong data present under District_name columns
```

```
array(['NICOBARS', 'NORTH AND MIDDLE ANDAMAN', 'SOUTH ANDAMANS',
       'ANANTAPUR', 'CHITTOOR', 'EAST GODAVARI', 'GUNTUR', 'KADAPA',
       'KRISHNA', 'KURNOOL', 'PRAKASAM', 'SPSR NELLORE', 'SRIKAKULAM',
       'VISAKHAPATANAM', 'VIZIANAGARAM', 'WEST GODAVARI', 'ANJAW',
       'CHANGLANG', 'DIBANG VALLEY', 'EAST KAMENG', 'EAST SIANG',
       'KURUNG KUMEY', 'LOHIT', 'LONGDING', 'LOWER DIBANG VALLEY',
       'LOWER SUBANSIRI', 'NAMSAI', 'PAPUM PARE', 'TAWANG', 'TIRAP',
       'UPPER SIANG', 'UPPER SUBANSIRI', 'WEST KAMENG', 'WEST SIANG',
       'BAKSA', 'BARPETA', 'BONGAIGAON', 'CACHAR', 'CHIRANG',
'DARRANG',
       'DHEMAJI', 'DHUBRI', 'DIBRUGARH', 'DIMA HASAO', 'GOALPARA',
       'GOLAGHAT', 'HAILAKANDI', 'JORHAT', 'KAMRUP', 'KAMRUP METRO',
       'KARBI ANGLONG', 'KARIMGANJ', 'KOKRAJHAR', 'LAKHIMPUR',
'MARIGAON',
       'NAGAON', 'NALBARI', 'SIVASAGAR', 'SONITPUR', 'TINSUKIA',
       'UDALGURI', 'ARARIA', 'ARWAL', 'AURANGABAD', 'BANKA',
'BEGUSARAI',
       'BHAGALPUR', 'BHOJPUR', 'BUXAR', 'DARBHANGA', 'GAYA',
'GOPALGANJ',
       'JAMUI', 'JEHANABAD', 'KAIMUR (BHABUA)', 'KATIHAR', 'KHAGARIA',
       'KISHANGANJ', 'LAKHISARAI', 'MADHEPURA', 'MADHUBANI', 'MUNGER',
       'MUZAFFARPUR', 'NALANDA', 'NAWADA', 'PASHCHIM CHAMPARAN',
'PATNA',
       'PURBI CHAMPARAN', 'PURNIA', 'ROHTAS', 'SAHARSA', 'SAMASTIPUR',
       'SARAN', 'SHEIKHPURA', 'SHEOHAR', 'SITAMARHI', 'SIWAN',
'SUPAUL',
       'VAISHALI', 'CHANDIGARH', 'BALOD', 'BALODA BAZAR', 'BALRAMPUR',
       'BASTAR', 'BEMETARA', 'BIJAPUR', 'BILASPUR', 'DANTEWADA',
       'DHAMTARI', 'DURG', 'GARIYABAND', 'JANJGIR-CHAMPA', 'JASHPUR',
       'KABIRDHAM', 'KANKER', 'KONDAGAON', 'KORBA', 'KOREA',
'MAHASAMUND',
       'MUNGELI', 'NARAYANPUR', 'RAIGARH', 'RAIPUR', 'RAJNANDGAON',
       'SUKMA', 'SURAJPUR', 'SURGUJA', 'DADRA AND NAGAR HAVELI',
```

```
       'NORTH GOA', 'SOUTH GOA', 'AHMADABAD', 'AMRELI', 'ANAND',
       'BANAS KANTHA', 'BHARUCH', 'BHAVNAGAR', 'DANG', 'DOHAD',
       'GANDHINAGAR', 'JAMNAGAR', 'JUNAGADH', 'KACHCHH', 'KHEDA',
       'MAHESANA', 'NARMADA', 'NAVSARI', 'PANCH MAHALS', 'PATAN',
       'PORBANDAR', 'RAJKOT', 'SABAR KANTHA', 'SURAT',
'SURENDRANAGAR',
       'TAPI', 'VADODARA', 'VALSAD', 'AMBALA', 'BHIWANI', 'FARIDABAD',
       'FATEHABAD', 'GURGAON', 'HISAR', 'JHAJJAR', 'JIND', 'KAITHAL',
       'KARNAL', 'KURUKSHETRA', 'MAHENDRAGARH', 'MEWAT', 'PALWAL',
       'PANCHKULA', 'PANIPAT', 'REWARI', 'ROHTAK', 'SIRSA', 'SONIPAT',
       'YAMUNANAGAR', 'CHAMBA', 'HAMIRPUR', 'KANGRA', 'KINNAUR',
'KULLU',
       'LAHUL AND SPITI', 'MANDI', 'SHIMLA', 'SIRMAUR', 'SOLAN',
'UNA',
       'ANANTNAG', 'BADGAM', 'BANDIPORA', 'BARAMULLA', 'DODA',
       'GANDERBAL', 'JAMMU', 'KARGIL', 'KATHUA', 'KISHTWAR', 'KULGAM',
       'KUPWARA', 'LEH LADAKH', 'POONCH', 'PULWAMA', 'RAJAURI',
'RAMBAN',
       'REASI', 'SAMBA', 'SHOPIAN', 'SRINAGAR', 'UDHAMPUR', 'BOKARO',
       'CHATRA', 'DEOGHAR', 'DHANBAD', 'DUMKA', 'EAST SINGHBUM',
'GARHWA',
       'GIRIDIH', 'GODDA', 'GUMLA', 'HAZARIBAGH', 'JAMTARA', 'KHUNTI',
       'KODERMA', 'LATEHAR', 'LOHARDAGA', 'PAKUR', 'PALAMU',
'RAMGARH',
       'RANCHI', 'SAHEBGANJ', 'SARAIKELA KHARSAWAN', 'SIMDEGA',
       'WEST SINGHBHUM', 'BAGALKOT', 'BANGALORE RURAL', 'BELGAUM',
       'BELLARY', 'BENGALURU URBAN', 'BIDAR', 'CHAMARAJANAGAR',
       'CHIKBALLAPUR', 'CHIKMAGALUR', 'CHITRADURGA', 'DAKSHIN KANNAD',
       'DAVANGERE', 'DHARWAD', 'GADAG', 'GULBARGA', 'HASSAN',
'HAVERI',
       'KODAGU', 'KOLAR', 'KOPPAL', 'MANDYA', 'MYSORE', 'RAICHUR',
       'RAMANAGARA', 'SHIMOGA', 'TUMKUR', 'UDUPI', 'UTTAR KANNAD',
       'YADGIR', 'ALAPPUZHA', 'ERNAKULAM', 'IDUKKI', 'KANNUR',
       'KASARAGOD', 'KOLLAM', 'KOTTAYAM', 'KOZHIKODE', 'MALAPPURAM',
       'PALAKKAD', 'PATHANAMTHITTA', 'THIRUVANANTHAPURAM', 'THRISSUR',
       'WAYANAD', 'AGAR MALWA', 'ALIRAJPUR', 'ANUPPUR', 'ASHOKNAGAR',
       'BALAGHAT', 'BARWANI', 'BETUL', 'BHIND', 'BHOPAL', 'BURHANPUR',
       'CHHATARPUR', 'CHHINDWARA', 'DAMOH', 'DATIA', 'DEWAS', 'DHAR',
       'DINDORI', 'GUNA', 'GWALIOR', 'HARDA', 'HOSHANGABAD', 'INDORE',
       'JABALPUR', 'JHABUA', 'KATNI', 'KHANDWA', 'KHARGONE', 'MANDLA',
       'MANDSAUR', 'MORENA', 'NARSINGHPUR', 'NEEMUCH', 'PANNA',
'RAISEN',
       'RAJGARH', 'RATLAM', 'REWA', 'SAGAR', 'SATNA', 'SEHORE',
'SEONI',
       'SHAHDOL', 'SHAJAPUR', 'SHEOPUR', 'SHIVPURI', 'SIDHI',
'SINGRAULI',
       'TIKAMGARH', 'UJJAIN', 'UMARIA', 'VIDISHA', 'AHMEDNAGAR',
'AKOLA',
       'AMRAVATI', 'BEED', 'BHANDARA', 'BULDHANA', 'CHANDRAPUR',
'DHULE',
```

```
       'GADCHIROLI', 'GONDIA', 'HINGOLI', 'JALGAON', 'JALNA',
'KOLHAPUR',
       'LATUR', 'MUMBAI', 'NAGPUR', 'NANDED', 'NANDURBAR', 'NASHIK',
       'OSMANABAD', 'PALGHAR', 'PARBHANI', 'PUNE', 'RAIGAD',
'RATNAGIRI',
       'SANGLI', 'SATARA', 'SINDHUDURG', 'SOLAPUR', 'THANE', 'WARDHA',
       'WASHIM', 'YAVATMAL', 'BISHNUPUR', 'CHANDEL', 'CHURACHANDPUR',
       'IMPHAL EAST', 'IMPHAL WEST', 'SENAPATI', 'TAMENGLONG',
'THOUBAL',
       'UKHRUL', 'EAST GARO HILLS', 'EAST JAINTIA HILLS',
       'EAST KHASI HILLS', 'NORTH GARO HILLS', 'RI BHOI',
       'SOUTH GARO HILLS', 'SOUTH WEST GARO HILLS',
       'SOUTH WEST KHASI HILLS', 'WEST GARO HILLS', 'WEST JAINTIA
HILLS',
       'WEST KHASI HILLS', 'AIZAWL', 'CHAMPHAI', 'KOLASIB',
'LAWNGTLAI',
       'LUNGLEI', 'MAMIT', 'SAIHA', 'SERCHHIP', 'DIMAPUR', 'KIPHIRE',
       'KOHIMA', 'LONGLENG', 'MOKOKCHUNG', 'MON', 'PEREN', 'PHEK',
       'TUENSANG', 'WOKHA', 'ZUNHEBOTO', 'ANUGUL', 'BALANGIR',
       'BALESHWAR', 'BARGARH', 'BHADRAK', 'BOUDH', 'CUTTACK',
'DEOGARH',
       'DHENKANAL', 'GAJAPATI', 'GANJAM', 'JAGATSINGHAPUR', 'JAJAPUR',
       'JHARSUGUDA', 'KALAHANDI', 'KANDHAMAL', 'KENDRAPARA',
'KENDUJHAR',
       'KHORDHA', 'KORAPUT', 'MALKANGIRI', 'MAYURBHANJ',
'NABARANGPUR',
       'NAYAGARH', 'NUAPADA', 'PURI', 'RAYAGADA', 'SAMBALPUR',
'SONEPUR',
       'SUNDARGARH', 'KARAIKAL', 'MAHE', 'PONDICHERRY', 'YANAM',
       'AMRITSAR', 'BARNALA', 'BATHINDA', 'FARIDKOT', 'FATEHGARH
SAHIB',
       'FAZILKA', 'FIROZEPUR', 'GURDASPUR', 'HOSHIARPUR', 'JALANDHAR',
       'KAPURTHALA', 'LUDHIANA', 'MANSA', 'MOGA', 'MUKTSAR',
'NAWANSHAHR',
       'PATHANKOT', 'PATIALA', 'RUPNAGAR', 'S.A.S NAGAR', 'SANGRUR',
       'TARN TARAN', 'AJMER', 'ALWAR', 'BANSWARA', 'BARAN', 'BARMER',
       'BHARATPUR', 'BHILWARA', 'BIKANER', 'BUNDI', 'CHITTORGARH',
       'CHURU', 'DAUSA', 'DHOLPUR', 'DUNGARPUR', 'GANGANAGAR',
       'HANUMANGARH', 'JAIPUR', 'JAISALMER', 'JALORE', 'JHALAWAR',
       'JHUNJHUNU', 'JODHPUR', 'KARAULI', 'KOTA', 'NAGAUR', 'PALI',
       'PRATAPGARH', 'RAJSAMAND', 'SAWAI MADHOPUR', 'SIKAR', 'SIROHI',
       'TONK', 'UDAIPUR', 'EAST DISTRICT', 'NORTH DISTRICT',
       'SOUTH DISTRICT', 'WEST DISTRICT', 'ARIYALUR', 'COIMBATORE',
       'CUDDALORE', 'DHARMAPURI', 'DINDIGUL', 'ERODE', 'KANCHIPURAM',
       'KANNIYAKUMARI', 'KARUR', 'KRISHNAGIRI', 'MADURAI',
'NAGAPATTINAM',
       'NAMAKKAL', 'PERAMBALUR', 'PUDUKKOTTAI', 'RAMANATHAPURAM',
'SALEM',
       'SIVAGANGA', 'THANJAVUR', 'THE NILGIRIS', 'THENI',
'THIRUVALLUR',
```

```
       'THIRUVARUR', 'TIRUCHIRAPPALLI', 'TIRUNELVELI', 'TIRUPPUR',
       'TIRUVANNAMALAI', 'TUTICORIN', 'VELLORE', 'VILLUPURAM',
       'VIRUDHUNAGAR', 'ADILABAD', 'HYDERABAD', 'KARIMNAGAR',
'KHAMMAM',
       'MAHBUBNAGAR', 'MEDAK', 'NALGONDA', 'NIZAMABAD', 'RANGAREDDI',
       'WARANGAL', 'DHALAI', 'GOMATI', 'KHOWAI', 'NORTH TRIPURA',
       'SEPAHIJALA', 'SOUTH TRIPURA', 'UNAKOTI', 'WEST TRIPURA',
'AGRA',
       'ALIGARH', 'ALLAHABAD', 'AMBEDKAR NAGAR', 'AMETHI', 'AMROHA',
       'AURAIYA', 'AZAMGARH', 'BAGHPAT', 'BAHRAICH', 'BALLIA',
'BANDA',
       'BARABANKI', 'BAREILLY', 'BASTI', 'BIJNOR', 'BUDAUN',
       'BULANDSHAHR', 'CHANDAULI', 'CHITRAKOOT', 'DEORIA', 'ETAH',
       'ETAWAH', 'FAIZABAD', 'FARRUKHABAD', 'FATEHPUR', 'FIROZABAD',
       'GAUTAM BUDDHA NAGAR', 'GHAZIABAD', 'GHAZIPUR', 'GONDA',
       'GORAKHPUR', 'HAPUR', 'HARDOI', 'HATHRAS', 'JALAUN', 'JAUNPUR',
       'JHANSI', 'KANNAUJ', 'KANPUR DEHAT', 'KANPUR NAGAR', 'KASGANJ',
       'KAUSHAMBI', 'KHERI', 'KUSHI NAGAR', 'LALITPUR', 'LUCKNOW',
       'MAHARAJGANJ', 'MAHOBA', 'MAINPURI', 'MATHURA', 'MAU',
'MEERUT',
       'MIRZAPUR', 'MORADABAD', 'MUZAFFARNAGAR', 'PILIBHIT', 'RAE
BARELI',
       'RAMPUR', 'SAHARANPUR', 'SAMBHAL', 'SANT KABEER NAGAR',
       'SANT RAVIDAS NAGAR', 'SHAHJAHANPUR', 'SHAMLI', 'SHRAVASTI',
       'SIDDHARTH NAGAR', 'SITAPUR', 'SONBHADRA', 'SULTANPUR',
'UNNAO',
       'VARANASI', 'ALMORA', 'BAGESHWAR', 'CHAMOLI', 'CHAMPAWAT',
       'DEHRADUN', 'HARIDWAR', 'NAINITAL', 'PAURI GARHWAL',
'PITHORAGARH',
       'RUDRA PRAYAG', 'TEHRI GARHWAL', 'UDAM SINGH NAGAR', 'UTTAR
KASHI',
       '24 PARAGANAS NORTH', '24 PARAGANAS SOUTH', 'BANKURA',
'BARDHAMAN',
       'BIRBHUM', 'COOCHBEHAR', 'DARJEELING', 'DINAJPUR DAKSHIN',
       'DINAJPUR UTTAR', 'HOOGHLY', 'HOWRAH', 'JALPAIGURI', 'MALDAH',
       'MEDINIPUR EAST', 'MEDINIPUR WEST', 'MURSHIDABAD', 'NADIA',
       'PURULIA'], dtype=object)

df['Season'].unique()
# Checking for any wrong data present under Season columns

array(['Kharif     ', 'Whole Year ', 'Autumn     ', 'Rabi       ',
       'Summer     ', 'Winter     '], dtype=object)

df['Crop'].unique()
# Checking for any wrong data present under Crop columns

array(['Arecanut', 'Other Kharif pulses', 'Rice', 'Banana',
'Cashewnut',
       'Coconut ', 'Dry ginger', 'Sugarcane', 'Sweet potato',
'Tapioca',
```

```
       'Black pepper', 'Dry chillies', 'other oilseeds', 'Turmeric',
       'Maize', 'Moong(Green Gram)', 'Urad', 'Arhar/Tur', 'Groundnut',
       'Sunflower', 'Bajra', 'Castor seed', 'Cotton(lint)', 'Horse-
gram',
       'Jowar', 'Korra', 'Ragi', 'Tobacco', 'Gram', 'Wheat', 'Masoor',
       'Sesamum', 'Linseed', 'Safflower', 'Onion', 'other misc.
pulses',
       'Samai', 'Small millets', 'Coriander', 'Potato',
       'Other Rabi pulses', 'Soyabean', 'Beans & Mutter(Vegetable)',
       'Bhindi', 'Brinjal', 'Citrus Fruit', 'Cucumber', 'Grapes',
'Mango',
       'Orange', 'other fibres', 'Other Fresh Fruits', 'Other
Vegetables',
       'Papaya', 'Pome Fruit', 'Tomato', 'Mesta', 'Cowpea(Lobia)',
       'Lemon', 'Pome Granet', 'Sapota', 'Cabbage', 'Rapeseed
&Mustard',
       'Peas  (vegetable)', 'Niger seed', 'Bottle Gourd', 'Varagu',
       'Garlic', 'Ginger', 'Oilseeds total', 'Pulses total', 'Jute',
       'Peas & beans (Pulses)', 'Blackgram', 'Paddy', 'Pineapple',
       'Barley', 'Sannhamp', 'Khesari', 'Guar seed', 'Moth',
       'Other Cereals & Millets', 'Cond-spcs other', 'Turnip',
'Carrot',
       'Redish', 'Arcanut (Processed)', 'Atcanut (Raw)',
       'Cashewnut Processed', 'Cashewnut Raw', 'Cardamom', 'Rubber',
       'Bitter Gourd', 'Drum Stick', 'Jack Fruit', 'Snak Guard',
'Tea',
       'Coffee', 'Cauliflower', 'Other Citrus Fruit', 'Water Melon',
       'Total foodgrain', 'Kapas', 'Colocosia', 'Lentil', 'Bean',
       'Jobster', 'Perilla', 'Rajmash Kholar', 'Ricebean (nagadal)',
       'Ash Gourd', 'Beet Root', 'Lab-Lab', 'Ribed Guard', 'Yam',
       'Pump Kin', 'Apple', 'Peach', 'Pear', 'Plums', 'Litchi', 'Ber',
       'Other Dry Fruit', 'Jute & mesta'], dtype=object)
```

```
df['Crop_Year'].unique()
```

```
array([2000, 2001, 2002, 2003, 2004, 2005, 2006, 2010, 1997, 1998,
1999,
       2007, 2008, 2009, 2011, 2012, 2013, 2014, 2015], dtype=int64)
```

```
df['Area'].unique()
```

```
array([1.25400e+03, 2.00000e+00, 1.02000e+02, ..., 3.02274e+05,
       1.14930e+04, 2.79151e+05])
```

```
df[df['Production'] == 0]
```

```
           State_Name       District_Name  Crop_Year       Season  \
397     Andhra Pradesh           ANANTAPUR       2002  Kharif
424     Andhra Pradesh           ANANTAPUR       2002  Whole Year
428     Andhra Pradesh           ANANTAPUR       2002  Whole Year
430     Andhra Pradesh           ANANTAPUR       2002  Whole Year
```

```
481     Andhra Pradesh            ANANTAPUR       2003  Whole Year
...                ...                   ...        ...          ...
221362   Uttar Pradesh             MAINPURI       2003  Rabi
222672   Uttar Pradesh               MEERUT       2002  Kharif
224493   Uttar Pradesh             PILIBHIT       2002  Kharif
227377   Uttar Pradesh  SANT RAVIDAS NAGAR       2003  Rabi
231266   Uttar Pradesh              VARANASI       2003  Kharif

                     Crop    Area  Production
397              Soyabean    51.0         0.0
424              Cucumber    18.0         0.0
428          other fibres   132.0         0.0
430      Other Vegetables  1767.0         0.0
481              Cucumber    51.0         0.0
...                   ...     ...         ...
221362            Linseed     1.0         0.0
222672  Moong(Green Gram)     1.0         0.0
224493  Moong(Green Gram)     1.0         0.0
227377            Linseed     1.0         0.0
231266               Moth     1.0         0.0

[3523 rows x 7 columns]
```

```
df = df[df['Production']!=0]
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 238838 entries, 0 to 246090
Data columns (total 7 columns):
 #   Column         Non-Null Count   Dtype
---  -------        --------------   -----
 0   State_Name     238838 non-null  object
 1   District_Name  238838 non-null  object
 2   Crop_Year      238838 non-null  int64
 3   Season         238838 non-null  object
 4   Crop           238838 non-null  object
 5   Area           238838 non-null  float64
 6   Production     238838 non-null float64
dtypes: float64(2), int64(1), object(4)
memory usage: 14.6+ MB
```

Check for duplicates

```
df.duplicated()
df.drop_duplicates()
```

```
                         State_Name District_Name  Crop_Year
Season  \
0       Andaman and Nicobar Islands      NICOBARS       2000  Kharif

1       Andaman and Nicobar Islands      NICOBARS       2000  Kharif
```

```
2       Andaman and Nicobar Islands      NICOBARS       2000  Kharif

3       Andaman and Nicobar Islands      NICOBARS       2000  Whole
Year
4       Andaman and Nicobar Islands      NICOBARS       2000  Whole
Year
...                             ...           ...        ...
...
246086                  West Bengal       PURULIA       2014  Summer

246087                  West Bengal       PURULIA       2014  Summer

246088                  West Bengal       PURULIA       2014  Whole
Year
246089                  West Bengal       PURULIA       2014  Winter

246090                  West Bengal       PURULIA       2014  Winter


                     Crop      Area  Production
0                Arecanut    1254.0      2000.0
1       Other Kharif pulses       2.0         1.0
2                    Rice     102.0       321.0
3                  Banana     176.0       641.0
4               Cashewnut     720.0       165.0
...                   ...       ...         ...
246086               Rice     306.0       801.0
246087             Sesamum     627.0       463.0
246088           Sugarcane     324.0     16250.0
246089               Rice  279151.0    597899.0
246090             Sesamum     175.0        88.0

[238838 rows x 7 columns]
```

Understanding the dataset

```
df.corr()
```

```
            Crop_Year      Area Production
Crop_Year    1.000000 -0.027791    0.006668
Area        -0.027791  1.000000    0.040487
Production   0.006668  0.040487    1.000000
```

```
df.cov()
```

```
                Crop_Year       Area       Production
Crop_Year       24.755320  -7.080846e+03   5.703444e+05
Area     -7080.846215     2.622432e+09     3.564302e+10
Production  570344.353860  3.564302e+10  2.955329e+14
```

Exporting the dataset

```
df.to_csv("Crop_yield.csv")
```

Uploading Dataset into IBM Cognos