

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

```
In [2]: df = pd.read_csv('ICRISAT_District_Level_Data.csv')
#df.info()
```

```
In [3]: crop = []
dist_code = []
year = []
state_code = []
state_name = []
dist_name = []
area = []
production = []
yields = []
cols_to_rows = ['RICE', 'WHEAT', 'KHARIF SORGHUM', 'RABI SORGHUM', 'SORGHUM', 'PEARL MILLET', 'MAIZE', 'FINGER MILLET',
                'BARLEY', 'CHICKPEA', 'PIGEONPEA', 'MINOR PULSES', 'GROUNDNUT', 'SESAMUM', 'RAPESEED AND MUSTARD',
                'SAFFLOWER', 'CASTOR', 'LINSEED', 'SUNFLOWER', 'SOYABEAN', 'OILSEEDS', 'SUGARCANE', 'COTTON']

extra_cols = ['FRUITS', 'VEGETABLES', 'FRUITS AND VEGETABLES', 'POTATOES', 'ONION', 'FODDER']
```

```
In [4]: for i in range(len(cols_to_rows)):
    for j in range(len(df)):
        crop.append(df[cols_to_rows[i]][j])
        dist_code.append(df['Dist Code'][j])
        year.append(df['Year'][j])
        state_code.append(df['State Code'][j])
        state_name.append(df['State Name'][j])
        dist_name.append(df['Dist Name'][j])
        area.append(df[cols_to_rows[i]+' AREA (1000 ha)'][j])
        production.append(df[cols_to_rows[i]+' PRODUCTION (1000 tons)'][j])
        yields.append(df[cols_to_rows[i]+' YIELD (Kg per ha)'][j])

    for i in range(len(extra_cols)):
        for j in range(len(df)):
            crop.append(extra_cols[i])
            dist_code.append(df['Dist Code'][j])
            year.append(df['Year'][j])
            state_code.append(df['State Code'][j])
            state_name.append(df['State Name'][j])
            dist_name.append(df['Dist Name'][j])
            area.append(df[extra_cols[i]+' AREA (1000 ha)'][j])
            production.append('')
            yields.append('')
```

```
In [5]: df2 = pd.DataFrame({'Plant':crop, 'Dist Code':dist_code, 'Year':year, 'State Code':state_code, 'State Name':state_name,
                          'Dist Name':dist_name, 'Area (1000 ha)':area, 'Production (1000 tons)':production, 'Yield (Kg per ha)':yields})
```

```
In [6]: crop_production = ['RICE', 'WHEAT']
coarse_grains = ['KHARIF SORGHUM', 'RABI SORGHUM', 'SORGHUM', 'PEARL MILLET', 'MAIZE', 'FINGER MILLET']
pulses = ['CHICKPEA', 'PIGEONPEA', 'MINOR PULSES']
oilseeds = ['GROUNDNUT', 'SESAMUM', 'RAPESEED AND MUSTARD', 'SAFFLOWER', 'CASTOR', 'LINSEED', 'SUNFLOWER', 'SOYABEAN', 'OILSEEDS']
other_crops = ['SUGARCANE', 'COTTON', 'FRUITS', 'VEGETABLES', 'FRUITS AND VEGETABLES', 'POTATOES', 'ONION', 'FODDER']
```

```
In [8]: category = []
for i in range(len(df2)):
    if df2['Plant'][i] in crop_production:
        category.append('crop production')
    elif df2['Plant'][i] in coarse_grains:
        category.append('coarse grains')
    elif df2['Plant'][i] in pulses:
        category.append('pulses')
    elif df2['Plant'][i] in oilseeds:
        category.append('oilseeds')
    else:
        category.append('other crops')
```

```
In [11]: df2['category'] = category
```

```
In [12]: df2.tail()
```

```
Out[12]:
```

	Plant	Dist Code	Year	State Code	State Name	Dist Name	Area (1000 ha)	Production (1000 tons)	Yield (Kg per ha)	category
468229	FODDER	917	2013	15	Jharkhand	Singhbhum	0.0			other crops
468230	FODDER	917	2014	15	Jharkhand	Singhbhum	0.0			other crops
468231	FODDER	917	2015	15	Jharkhand	Singhbhum	0.0			other crops
468232	FODDER	917	2016	15	Jharkhand	Singhbhum	0.0			other crops
468233	FODDER	917	2017	15	Jharkhand	Singhbhum	0.0			other crops

```
In [13]: df2.to_csv('ICRISAT_data.csv', index = False)
```

```
In [ ]:
```