

1. Introduction

December 14, 2020

0.0.1 Importing libraries:

```
[1]: import pandas as pd
```

0.0.2 Indian crops dataset:

```
[2]: df = pd.read_csv('Indian crops dataset.csv')
df.head()
```

```
[2]:
```

	Dist Code	Year	State Code	State Name	Dist Name	RICE AREA (1000 ha)	\
0	1	1966	14	Chhattisgarh	Durg	548.0	
1	1	1967	14	Chhattisgarh	Durg	547.0	
2	1	1968	14	Chhattisgarh	Durg	556.3	
3	1	1969	14	Chhattisgarh	Durg	563.4	
4	1	1970	14	Chhattisgarh	Durg	571.6	

	RICE PRODUCTION (1000 tons)	RICE YIELD (Kg per ha)	\
0	185.0	338	
1	409.0	748	
2	468.0	841	
3	400.8	711	
4	473.6	829	

	RICE SHARE IN TOTAL (Percent)	WHEAT AREA (1000 ha)	...	\
0	50.63	44.0	...	
1	42.16	50.0	...	
2	41.59	53.7	...	
3	42.04	49.4	...	
4	41.33	44.2	...	

	COTTON YIELD (Kg per ha)	COTTON SHARE IN TOTAL (Percent)	\
0	0	0.0	
1	0	0.0	
2	0	0.0	
3	0	0.0	
4	0	0.0	

	FRUITS AREA (1000 ha)	FRUITS SHARE IN TOTAL (Percent)	\
--	-----------------------	---------------------------------	---

0	5.95	0.55
1	5.77	0.44
2	5.41	0.40
3	5.52	0.41
4	5.45	0.39

	VEGETABLES AREA (1000 ha)	VEGETABLES SHARE IN TOTAL (Percent) \
0	6.64	0.61
1	7.24	0.56
2	7.40	0.55
3	7.16	0.53
4	7.19	0.52

	FRUITS AND VEGETABLES AREA (1000 ha)	POTATOES AREA (1000 ha) \
0	12.59	0.01
1	13.02	0.01
2	12.81	0.10
3	12.69	0.01
4	12.64	0.02

	ONION AREA (1000 ha)	FODDER AREA (1000 ha)
0	0.60	0.47
1	0.56	1.23
2	0.58	1.02
3	0.56	0.84
4	0.52	0.42

[5 rows x 109 columns]

0.0.3 Checking for empty cells:

```
[4]: df.isnull().sum()
```

```
[4]: Dist Code      0
     Year          0
     State Code     0
     State Name     0
     Dist Name      0
     ..
     VEGETABLES SHARE IN TOTAL (Percent)  0
     FRUITS AND VEGETABLES AREA (1000 ha)  0
     POTATOES AREA (1000 ha)              0
     ONION AREA (1000 ha)                  0
     FODDER AREA (1000 ha)                 0
     Length: 109, dtype: int64
```

0.0.4 Data types:

```
[5]: df.dtypes
```

```
[5]: Dist Code          int64
     Year              int64
     State Code        int64
     State Name        object
     Dist Name         object
     ...
     VEGETABLES SHARE IN TOTAL (Percent) float64
     FRUITS AND VEGETABLES AREA (1000 ha) float64
     POTATOES AREA (1000 ha)            float64
     ONION AREA (1000 ha)               float64
     FODDER AREA (1000 ha)              float64
     Length: 109, dtype: object
```

0.0.5 Details of the data:

```
[6]: df.describe(include='all')
```

```
[6]:
```

	Dist Code	Year	State Code	State Name	Dist Name \
count	15508.000000	15508.000000	15508.000000	15508	15508
unique	NaN	NaN	NaN	20	311
top	NaN	NaN	NaN	Uttar Pradesh	Srikakulam
freq	NaN	NaN	NaN	2300	50
mean	270.201315	1990.481945	9.573253	NaN	NaN
std	278.562132	14.433817	4.990594	NaN	NaN
min	1.000000	1966.000000	1.000000	NaN	NaN
25%	78.000000	1978.000000	6.000000	NaN	NaN
50%	156.000000	1990.000000	10.000000	NaN	NaN
75%	242.000000	2003.000000	12.000000	NaN	NaN
max	917.000000	2015.000000	20.000000	NaN	NaN

	RICE AREA (1000 ha)	RICE PRODUCTION (1000 tons) \
count	15508.000000	15508.000000
unique	NaN	NaN
top	NaN	NaN
freq	NaN	NaN
mean	128.424233	219.389249
std	159.949357	318.034684
min	-1.000000	-1.000000
25%	10.587500	9.400000
50%	66.605000	94.200000
75%	190.602500	308.082500
max	1154.230000	3215.010000

	RICE YIELD (Kg per ha)	RICE SHARE IN TOTAL (Percent) \
count	15508.000000	15508.000000
unique	NaN	NaN
top	NaN	NaN
freq	NaN	NaN
mean	1458.956796	23.739390
std	936.796983	23.531236
min	-1.000000	-1.000000
25%	791.000000	2.507500
50%	1304.500000	16.980000
75%	2062.000000	38.610000
max	6547.000000	98.120000

	WHEAT AREA (1000 ha) ...	COTTON YIELD (Kg per ha) \
count	15508.000000 ...	15508.000000
unique	NaN ...	NaN
top	NaN ...	NaN
freq	NaN ...	NaN
mean	76.257169 ...	120.563580
std	99.288499 ...	193.747941
min	-1.000000 ...	-1.000000
25%	1.860000 ...	0.000000
50%	36.525000 ...	0.000000
75%	121.647500 ...	200.000000
max	879.490000 ...	5000.000000

	COTTON SHARE IN TOTAL (Percent)	FRUITS AREA (1000 ha) \
count	15508.000000	15508.000000
unique	NaN	NaN
top	NaN	NaN
freq	NaN	NaN
mean	3.490417	7.733721
std	8.511718	13.453608
min	-1.000000	-1.000000
25%	0.000000	0.340000
50%	0.010000	2.300000
75%	1.740000	8.800000
max	74.780000	159.540000

	FRUITS SHARE IN TOTAL (Percent)	VEGETABLES AREA (1000 ha) \
count	15508.000000	15508.000000
unique	NaN	NaN
top	NaN	NaN
freq	NaN	NaN
mean	1.766446	10.951978
std	3.584192	17.618990
min	-1.000000	-1.000000

25%	0.080000	1.620000
50%	0.410000	4.770000
75%	1.740000	12.372500
max	51.660000	200.060000

	VEGETABLES SHARE IN TOTAL (Percent) \
count	15508.000000
unique	NaN
top	NaN
freq	NaN
mean	2.245756
std	3.749053
min	-1.000000
25%	0.330000
50%	1.000000
75%	2.690000
max	71.200000

	FRUITS AND VEGETABLES AREA (1000 ha)	POTATOES AREA (1000 ha) \
count	15508.000000	15508.000000
unique	NaN	NaN
top	NaN	NaN
freq	NaN	NaN
mean	18.520090	3.092763
std	25.405096	7.744254
min	-1.000000	-1.000000
25%	2.600000	0.000000
50%	9.000000	0.390000
75%	23.352500	3.110000
max	240.990000	111.610000

	ONION AREA (1000 ha)	FODDER AREA (1000 ha)
count	15508.000000	15508.000000
unique	NaN	NaN
top	NaN	NaN
freq	NaN	NaN
mean	1.165945	21.679561
std	4.212277	59.190416
min	-1.000000	-1.000000
25%	0.060000	0.000000
50%	0.280000	1.255000
75%	0.870000	17.330000
max	131.350000	1162.660000

[11 rows x 109 columns]