ps1

April 6, 2024

[ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib
import tensorflow as tf
```

Clubbed the Table 1 and 2 into single df....Subhranil

[ ]:
```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

[ ]:
```
from openpyxl import load_workbook


wb = load_workbook("/content/drive/MyDrive/BrainDead/
 ↪State_wise_rice_production_in_India.xlsx")


ws = wb["Table 2"]


ws.delete_rows(2)
ws.delete_rows(3)


wb.save("rice_production_modified.xlsx")
```

[ ]:
```
df_table1 = pd.read_excel("rice_production_modified.xlsx", sheet_name="Table␣
 ↪1", header=1)
df_table1.head()
```

[ ]:

	State/Union Territory	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10	\
0	Andhra Pradesh	9601	11704	11872	13324	14241	10538	
1	Arunachal Pradesh	135	146.2	146.2	158.1	163.9	215.8	
2	Assam	3470.7	3552.5	2916	3319	4008.5	4335.9	

```
3                Bihar  2472.2  3495.5  4989.3  4418.1  5590.3  3599.3
4          Chhattisgarh  4383.3  5011.6  5041.4  5426.6  4391.8  4110.4

   2010-11  2011-12  2012-13
0   7882.4   7746.2   6862.4
1    234.0    255.0    263.0
2   4736.6   4516.3   5128.5
3   3102.1   7162.6   7529.3
4   6159.0   6028.4   6608.8
```

[ ]:
```
df_table2 = pd.read_excel("rice_production_modified.xlsx", sheet_name="Table 2", header=1).drop(labels = 'State/Union Territory', axis=1)
df_table2.head()
```

[ ]:
```
   2013-14  2014-15  2015-16  2016-17  2017-18  2018-19  2019-20  2020-21  \
0   6969.7   7233.9   7488.7   7452.4   8166.2   8234.7   8658.9   7882.9
1    276.2    285.0    204.0    220.0    233.3    240.0    244.7    247.1
2   4927.1   5222.7   5125.1   4727.4   5283.7   5220.6   4984.6   5214.8
3   5505.8   6356.7   6802.2   8239.3   8093.1   6155.5   6298.0   6747.0
4   6716.4   6322.1   5789.4   8048.4   4930.8   6526.9   6774.8   7161.2

   2021-22 2022-23*
0   7763.6   8542.3
1    252.4        -
2   4382.1   4979.8
3   7717.0   6725.2
4   8021.7   8238.3
```

[ ]:
```
print(df_table1.shape)
print(df_table2.shape)
```

```
(32, 10)
(32, 10)
```

[ ]:
```
df = pd.concat([df_table1, df_table2], axis=1)
df.set_index('State/Union Territory', inplace=True)
df.head()
```

[ ]:
```
                     2004-05 2005-06 2006-07 2007-08 2008-09 2009-10  \
State/Union Territory
Andhra Pradesh          9601   11704   11872   13324   14241   10538
Arunachal Pradesh        135   146.2   146.2   158.1   163.9   215.8
Assam                 3470.7  3552.5    2916    3319  4008.5  4335.9
Bihar                 2472.2  3495.5  4989.3  4418.1  5590.3  3599.3
Chhattisgarh          4383.3  5011.6  5041.4  5426.6  4391.8  4110.4

                     2010-11  2011-12  2012-13  2013-14  2014-15  2015-16  \
```

```
State/Union Territory
Andhra Pradesh          7882.4   7746.2   6862.4   6969.7   7233.9   7488.7
Arunachal Pradesh        234.0    255.0    263.0    276.2    285.0    204.0
Assam                   4736.6   4516.3   5128.5   4927.1   5222.7   5125.1
Bihar                   3102.1   7162.6   7529.3   5505.8   6356.7   6802.2
Chhattisgarh            6159.0   6028.4   6608.8   6716.4   6322.1   5789.4

                       2016-17  2017-18  2018-19  2019-20  2020-21  2021-22  \
State/Union Territory
Andhra Pradesh          7452.4   8166.2   8234.7   8658.9   7882.9   7763.6
Arunachal Pradesh        220.0    233.3    240.0    244.7    247.1    252.4
Assam                   4727.4   5283.7   5220.6   4984.6   5214.8   4382.1
Bihar                   8239.3   8093.1   6155.5   6298.0   6747.0   7717.0
Chhattisgarh            8048.4   4930.8   6526.9   6774.8   7161.2   8021.7

                      2022-23*
State/Union Territory
Andhra Pradesh          8542.3
Arunachal Pradesh            -
Assam                   4979.8
Bihar                   6725.2
Chhattisgarh            8238.3
```

[ ]:
```
df_T = df.transpose()
df_T.head()
```

[ ]:
```
State/Union Territory Andhra Pradesh Arunachal Pradesh   Assam   Bihar  \
2004-05                         9601              135  3470.7  2472.2
2005-06                        11704            146.2  3552.5  3495.5
2006-07                        11872            146.2    2916  4989.3
2007-08                        13324            158.1    3319  4418.1
2008-09                        14241            163.9  4008.5  5590.3

State/Union Territory Chhattisgarh NCT of Delhi    Goa Gujarat Haryana  \
2004-05                    4383.3         14.3  145.2  1238.2    3023
2005-06                    5011.6           24  147.3    1298    3210
2006-07                    5041.4         31.1  130.3    1390    3371
2007-08                    5426.6         31.4  121.6    1474    3613
2008-09                    4391.8         31.4  123.3    1303    3298

State/Union Territory Himachal Pradesh  … Punjab Rajasthan Sikkim  \
2004-05                           122  …  10437     150.4   21.6
2005-06                         112.1  …  10193       153   21.5
2006-07                         123.5  …  10138     169.8   21.5
2007-08                         121.5  …  10489     259.6   22.9
2008-09                         118.3  …  11000     241.1   21.7
```

```
State/Union Territory Tamil Nadu Telangana Tripura Uttar Pradesh Uttarakhand  \
2004-05                   5062.2         .   545.1        9555.6         572
2005-06                     5220         .   552.9       11133.7         590
2006-07                   6610.6         .   620.5         11124         556
2007-08                   5040.2         .   624.6         11780         593
2008-09                   5182.7         .   627.1         13097         582

State/Union Territory West Bengal ALL INDIA
2004-05                  14884.8   83131.7
2005-06                  14510.8   91793.4
2006-07                  14745.9   93355.3
2007-08                  14719.5   96692.9
2008-09                  15037.3   99182.5

[5 rows x 32 columns]
```

[ ]:
```
df_T
```

[ ]:
```
State/Union Territory Andhra Pradesh Arunachal Pradesh   Assam   Bihar  \
2004-05                       9601              135  3470.7  2472.2
2005-06                      11704            146.2  3552.5  3495.5
2006-07                      11872            146.2    2916  4989.3
2007-08                      13324            158.1    3319  4418.1
2008-09                      14241            163.9  4008.5  5590.3
2009-10                      10538            215.8  4335.9  3599.3
2010-11                     7882.4            234.0  4736.6  3102.1
2011-12                     7746.2            255.0  4516.3  7162.6
2012-13                     6862.4            263.0  5128.5  7529.3
2013-14                     6969.7            276.2  4927.1  5505.8
2014-15                     7233.9            285.0  5222.7  6356.7
2015-16                     7488.7            204.0  5125.1  6802.2
2016-17                     7452.4            220.0  4727.4  8239.3
2017-18                     8166.2            233.3  5283.7  8093.1
2018-19                     8234.7            240.0  5220.6  6155.5
2019-20                     8658.9            244.7  4984.6  6298.0
2020-21                     7882.9            247.1  5214.8  6747.0
2021-22                     7763.6            252.4  4382.1  7717.0
2022-23*                    8542.3                -  4979.8  6725.2

State/Union Territory Chhattisgarh NCT of Delhi    Goa Gujarat  Haryana  \
2004-05                   4383.3         14.3  145.2  1238.2     3023
2005-06                   5011.6           24  147.3    1298     3210
2006-07                   5041.4         31.1  130.3    1390     3371
2007-08                   5426.6         31.4  121.6    1474     3613
2008-09                   4391.8         31.4  123.3    1303     3298
2009-10                   4110.4         19.3  100.6    1292     3625
2010-11                   6159.0         19.6  115.0  1496.6   3472.0
```

2011-12	6028.4	19.8	121.8	1790.0	3759.0
2012-13	6608.8	19.7	122.8	1541.0	3976.0
2013-14	6716.4	29.6	126.5	1636.0	3998.0
2014-15	6322.1	25.9	120.5	1830.9	4006.0
2015-16	5789.4	17.3	115.1	1702.0	4145.0
2016-17	8048.4	17.3	113.2	1930.0	4453.0
2017-18	4930.8	16.8	103.0	1890.9	4523.38
2018-19	6526.9	16.8	98.8	1912.1	4516.1
2019-20	6774.8	16.8	90.4	1983.1	4824.3
2020-21	7161.2	19.8	87.3	2145.7	4424.9
2021-22	8021.7	19.0	90.4	2101.1	4618.0
2022-23*	8238.3	-	-	2395.2	5406.9

State/Union Territory	Himachal Pradesh	…	Punjab	Rajasthan	Sikkim	\
2004-05	122	…	10437	150.4	21.6	
2005-06	112.1	…	10193	153	21.5	
2006-07	123.5	…	10138	169.8	21.5	
2007-08	121.5	…	10489	259.6	22.9	
2008-09	118.3	…	11000	241.1	21.7	
2009-10	105.9	…	11236	228.3	24.3	
2010-11	128.9	…	10837.0	265.5	21.0	
2011-12	131.6	…	10542.0	253.4	20.9	
2012-13	125.3	…	11374.0	222.5	21.3	
2013-14	120.8	…	11267.0	312.6	20.3	
2014-15	125.2	…	11107.0	366.7	20.1	
2015-16	129.9	…	11823.0	369.8	13.1	
2016-17	146.6	…	11586.2	452.7	19.7	
2017-18	114.79	…	13381.79	450.87	17.63	
2018-19	114.9	…	12821.6	453.2	17.2	
2019-20	143.8	…	11779.3	480.5	16.1	
2020-21	140.5	…	12783.7	634.0	16.2	
2021-22	167.5	…	12885.5	478.6	16.0	
2022-23*	119.2	…	13146.7	577.4	-	

State/Union Territory	Tamil Nadu	Telangana	Tripura	Uttar Pradesh	Uttarakhand	\
2004-05	5062.2	.	545.1	9555.6	572	
2005-06	5220	.	552.9	11133.7	590	
2006-07	6610.6	.	620.5	11124	556	
2007-08	5040.2	.	624.6	11780	593	
2008-09	5182.7	.	627.1	13097	582	
2009-10	5665.2	.	640	10807.1	608	
2010-11	5792.4	6535.6	702.5	11992.0	550.4	
2011-12	7458.7	5148.8	718.3	14022.0	594.0	
2012-13	4049.9	4647.6	713.2	14416.0	579.8	
2013-14	5349.8	5755.0	711.8	14636.0	578.6	
2014-15	5727.8	4440.8	747.0	12167.9	603.7	
2015-16	7517.1	3047.0	794.8	12501.0	639.1	

```
2016-17                    2369.4    5173.4   814.6       13754.0       630.0
2017-18                    6638.9    6262.2   812.1       13274.0       646.7
2018-19                    6130.9    6670.0   793.2       15545.3       617.6
2019-20                    7171.1    7427.8   810.2       15517.8       658.4
2020-21                    6881.2   10217.1   803.1       15520.0       714.9
2021-22                    7906.6   12409.6   811.0       15271.5       716.1
2022-23*                   7850.6   16013.9       -       15171.3       641.7

State/Union Territory West Bengal ALL INDIA
2004-05                   14884.8   83131.7
2005-06                   14510.8   91793.4
2006-07                   14745.9   93355.3
2007-08                   14719.5   96692.9
2008-09                   15037.3   99182.5
2009-10                   14340.7   89092.9
2010-11                   13045.9   95979.8
2011-12                   14605.8  105310.9
2012-13                   15023.7  105231.6
2013-14                   15370.7  106645.5
2014-15                   14677.2  104798.5
2015-16                   15953.9  104408.2
2016-17                   15302.5  109698.4
2017-18                   14967.0  112757.6
2018-19                   16242.2  116477.8
2019-20                   15881.4  118870.3
2020-21                   16524.4  124368.3
2021-22                   16728.7  129471.4
2022-23*                  15636.9    135542

[19 rows x 32 columns]
```

[ ]:
```
df = df.apply(pd.to_numeric, errors='coerce')
df_T = df_T.apply(pd.to_numeric, errors='coerce')
```

[ ]:
```
df_T.dtypes
```

[ ]:
```
State/Union Territory
Andhra Pradesh       float64
Arunachal Pradesh    float64
Assam                float64
Bihar                float64
Chhattisgarh         float64
NCT of Delhi         float64
Goa                  float64
Gujarat              float64
Haryana              float64
Himachal Pradesh     float64
```

```
Jammu & Kashmir      float64
Jharkhand            float64
Karnataka            float64
Kerala               float64
Madhya Pradesh       float64
Maharashtra          float64
Manipur              float64
Meghalaya            float64
Mizoram              float64
Nagaland             float64
Odisha               float64
Puducherry           float64
Punjab               float64
Rajasthan            float64
Sikkim               float64
Tamil Nadu           float64
Telangana            float64
Tripura              float64
Uttar Pradesh        float64
Uttarakhand          float64
West Bengal          float64
ALL INDIA            float64
dtype: object
```

[ ]:
```
df_T
```

[ ]:
```
State/Union Territory  Andhra Pradesh  Arunachal Pradesh   Assam   Bihar  \
2004-05                         9601.0              135.0  3470.7  2472.2
2005-06                        11704.0              146.2  3552.5  3495.5
2006-07                        11872.0              146.2  2916.0  4989.3
2007-08                        13324.0              158.1  3319.0  4418.1
2008-09                        14241.0              163.9  4008.5  5590.3
2009-10                        10538.0              215.8  4335.9  3599.3
2010-11                         7882.4              234.0  4736.6  3102.1
2011-12                         7746.2              255.0  4516.3  7162.6
2012-13                         6862.4              263.0  5128.5  7529.3
2013-14                         6969.7              276.2  4927.1  5505.8
2014-15                         7233.9              285.0  5222.7  6356.7
2015-16                         7488.7              204.0  5125.1  6802.2
2016-17                         7452.4              220.0  4727.4  8239.3
2017-18                         8166.2              233.3  5283.7  8093.1
2018-19                         8234.7              240.0  5220.6  6155.5
2019-20                         8658.9              244.7  4984.6  6298.0
2020-21                         7882.9              247.1  5214.8  6747.0
2021-22                         7763.6              252.4  4382.1  7717.0
2022-23*                        8542.3                NaN  4979.8  6725.2
```

State/Union Territory	Chhattisgarh	NCT of Delhi	Goa	Gujarat	Haryana	\
2004-05	4383.3	14.3	145.2	1238.2	3023.00	
2005-06	5011.6	24.0	147.3	1298.0	3210.00	
2006-07	5041.4	31.1	130.3	1390.0	3371.00	
2007-08	5426.6	31.4	121.6	1474.0	3613.00	
2008-09	4391.8	31.4	123.3	1303.0	3298.00	
2009-10	4110.4	19.3	100.6	1292.0	3625.00	
2010-11	6159.0	19.6	115.0	1496.6	3472.00	
2011-12	6028.4	19.8	121.8	1790.0	3759.00	
2012-13	6608.8	19.7	122.8	1541.0	3976.00	
2013-14	6716.4	29.6	126.5	1636.0	3998.00	
2014-15	6322.1	25.9	120.5	1830.9	4006.00	
2015-16	5789.4	17.3	115.1	1702.0	4145.00	
2016-17	8048.4	17.3	113.2	1930.0	4453.00	
2017-18	4930.8	16.8	103.0	1890.9	4523.38	
2018-19	6526.9	16.8	98.8	1912.1	4516.10	
2019-20	6774.8	16.8	90.4	1983.1	4824.30	
2020-21	7161.2	19.8	87.3	2145.7	4424.90	
2021-22	8021.7	19.0	90.4	2101.1	4618.00	
2022-23*	8238.3	NaN	NaN	2395.2	5406.90	

State/Union Territory	Himachal Pradesh	…	Punjab	Rajasthan	Sikkim	\
2004-05	122.00	…	10437.00	150.40	21.60	
2005-06	112.10	…	10193.00	153.00	21.50	
2006-07	123.50	…	10138.00	169.80	21.50	
2007-08	121.50	…	10489.00	259.60	22.90	
2008-09	118.30	…	11000.00	241.10	21.70	
2009-10	105.90	…	11236.00	228.30	24.30	
2010-11	128.90	…	10837.00	265.50	21.00	
2011-12	131.60	…	10542.00	253.40	20.90	
2012-13	125.30	…	11374.00	222.50	21.30	
2013-14	120.80	…	11267.00	312.60	20.30	
2014-15	125.20	…	11107.00	366.70	20.10	
2015-16	129.90	…	11823.00	369.80	13.10	
2016-17	146.60	…	11586.20	452.70	19.70	
2017-18	114.79	…	13381.79	450.87	17.63	
2018-19	114.90	…	12821.60	453.20	17.20	
2019-20	143.80	…	11779.30	480.50	16.10	
2020-21	140.50	…	12783.70	634.00	16.20	
2021-22	167.50	…	12885.50	478.60	16.00	
2022-23*	119.20	…	13146.70	577.40	NaN	

State/Union Territory	Tamil Nadu	Telangana	Tripura	Uttar Pradesh	\
2004-05	5062.2	NaN	545.1	9555.6	
2005-06	5220.0	NaN	552.9	11133.7	
2006-07	6610.6	NaN	620.5	11124.0	
2007-08	5040.2	NaN	624.6	11780.0	

```
2008-09                    5182.7        NaN    627.1        13097.0
2009-10                    5665.2        NaN    640.0        10807.1
2010-11                    5792.4     6535.6    702.5        11992.0
2011-12                    7458.7     5148.8    718.3        14022.0
2012-13                    4049.9     4647.6    713.2        14416.0
2013-14                    5349.8     5755.0    711.8        14636.0
2014-15                    5727.8     4440.8    747.0        12167.9
2015-16                    7517.1     3047.0    794.8        12501.0
2016-17                    2369.4     5173.4    814.6        13754.0
2017-18                    6638.9     6262.2    812.1        13274.0
2018-19                    6130.9     6670.0    793.2        15545.3
2019-20                    7171.1     7427.8    810.2        15517.8
2020-21                    6881.2    10217.1    803.1        15520.0
2021-22                    7906.6    12409.6    811.0        15271.5
2022-23*                   7850.6    16013.9      NaN        15171.3

State/Union Territory  Uttarakhand  West Bengal  ALL INDIA
2004-05                      572.0      14884.8    83131.7
2005-06                      590.0      14510.8    91793.4
2006-07                      556.0      14745.9    93355.3
2007-08                      593.0      14719.5    96692.9
2008-09                      582.0      15037.3    99182.5
2009-10                      608.0      14340.7    89092.9
2010-11                      550.4      13045.9    95979.8
2011-12                      594.0      14605.8   105310.9
2012-13                      579.8      15023.7   105231.6
2013-14                      578.6      15370.7   106645.5
2014-15                      603.7      14677.2   104798.5
2015-16                      639.1      15953.9   104408.2
2016-17                      630.0      15302.5   109698.4
2017-18                      646.7      14967.0   112757.6
2018-19                      617.6      16242.2   116477.8
2019-20                      658.4      15881.4   118870.3
2020-21                      714.9      16524.4   124368.3
2021-22                      716.1      16728.7   129471.4
2022-23*                     641.7      15636.9   135542.0

[19 rows x 32 columns]
```

[ ]:
```
df_T.isnull().sum().sum()
```

[ ]:
```
17
```

[ ]:
```
df_T.isnull().sum()
```

[ ]:
```
State/Union Territory
Andhra Pradesh       0
```

```
Arunachal Pradesh    1
Assam                0
Bihar                0
Chhattisgarh         0
NCT of Delhi         1
Goa                  1
Gujarat              0
Haryana              0
Himachal Pradesh     0
Jammu & Kashmir      1
Jharkhand            0
Karnataka            0
Kerala               0
Madhya Pradesh       0
Maharashtra          0
Manipur              1
Meghalaya            1
Mizoram              1
Nagaland             1
Odisha               0
Puducherry           1
Punjab               0
Rajasthan            0
Sikkim               1
Tamil Nadu           0
Telangana            6
Tripura              1
Uttar Pradesh        0
Uttarakhand          0
West Bengal          0
ALL INDIA            0
dtype: int64
```

[ ]:
```
df_T.iloc[18].isnull().sum()
```

[ ]:
```
11
```

[ ]:
```
df_T.describe()
```

[ ]:
```
State/Union Territory  Andhra Pradesh  Arunachal Pradesh        Assam  \
count                       19.000000          18.000000    19.000000
mean                      9061.278947         217.772222  4529.047368
std                       2212.017631          47.838943   739.191158
min                       6862.400000         135.000000  2916.000000
25%                       7617.450000         173.925000  4172.200000
50%                       8166.200000         233.650000  4736.600000
75%                      10069.500000         251.075000  5126.800000
```

```
max                      14241.000000         285.000000  5283.700000

State/Union Territory        Bihar  Chhattisgarh  NCT of Delhi         Goa  \
count                    19.000000     19.000000     18.000000   18.000000
mean                   5842.026316   6089.015789     21.661111  115.172222
std                    1737.536829   1259.750102      5.719126   17.431506
min                    2472.200000   4110.400000     14.300000   87.300000
25%                    4703.700000   5026.500000     17.300000  101.200000
50%                    6298.000000   6159.000000     19.650000  117.800000
75%                    6982.400000   6745.600000     25.425000  123.175000
max                    8239.300000   8238.300000     31.400000  147.300000

State/Union Territory      Gujarat      Haryana  Himachal Pradesh  …  \
count                    19.000000    19.000000         19.000000  …
mean                   1702.621053  4013.820000        126.962632  …
std                     333.273049   627.124925         14.401712  …
min                    1238.200000  3023.000000        105.900000  …
25%                    1432.000000  3542.500000        118.750000  …
50%                    1702.000000  3998.000000        123.500000  …
75%                    1921.050000  4484.550000        130.750000  …
max                    2395.200000  5406.900000        167.500000  …

State/Union Territory        Punjab   Rajasthan     Sikkim   Tamil Nadu  \
count                    19.000000   19.000000  18.000000    19.000000
mean                   11517.252105  343.156316  19.612778  5980.278947
std                     1037.304623  144.756229   2.916061  1397.244387
min                    10138.000000  150.400000  13.100000  2369.400000
25%                    10689.500000  234.700000  17.307500  5201.350000
50%                    11267.000000  312.600000  20.600000  5792.400000
75%                    12303.350000  452.950000  21.500000  7026.150000
max                    13381.790000  634.000000  24.300000  7906.600000

State/Union Territory     Telangana     Tripura  Uttar Pradesh  Uttarakhand  \
count                     13.000000   18.000000      19.000000    19.000000
mean                    7211.446154  713.444444   13225.589474   614.315789
std                     3621.205336   92.236845    1849.559244    46.838425
min                     3047.000000  545.100000    9555.600000   550.400000
25%                     5148.800000  630.325000   11886.000000   580.900000
50%                     6262.200000  715.750000   13274.000000   603.700000
75%                     7427.800000  801.025000   14903.650000   640.400000
max                    16013.900000  814.600000   15545.300000   716.100000

State/Union Territory   West Bengal      ALL INDIA
count                     19.000000      19.000000
mean                   15168.384211  106463.631579
std                      867.613563   13946.740460
min                    13045.900000   83131.700000
```

```
25%                   14698.350000   96336.350000
50%                   15023.700000  105231.600000
75%                   15759.150000  114617.700000
max                   16728.700000  135542.000000

[8 rows x 32 columns]
```

[ ]:
```
for state in df_T.columns:
  if df_T[state].isnull().sum() != 0:
    df_T[state] = df_T[state].fillna(df_T[state].mean())
```

[ ]:
```
df_T.isnull().sum().sum()
```

[ ]:
```
0
```

[ ]:
```
df_T
```

[ ]:
```
State/Union Territory  Andhra Pradesh  Arunachal Pradesh   Assam   Bihar  \
2004-05                         9601.0         135.000000  3470.7  2472.2
2005-06                        11704.0         146.200000  3552.5  3495.5
2006-07                        11872.0         146.200000  2916.0  4989.3
2007-08                        13324.0         158.100000  3319.0  4418.1
2008-09                        14241.0         163.900000  4008.5  5590.3
2009-10                        10538.0         215.800000  4335.9  3599.3
2010-11                         7882.4         234.000000  4736.6  3102.1
2011-12                         7746.2         255.000000  4516.3  7162.6
2012-13                         6862.4         263.000000  5128.5  7529.3
2013-14                         6969.7         276.200000  4927.1  5505.8
2014-15                         7233.9         285.000000  5222.7  6356.7
2015-16                         7488.7         204.000000  5125.1  6802.2
2016-17                         7452.4         220.000000  4727.4  8239.3
2017-18                         8166.2         233.300000  5283.7  8093.1
2018-19                         8234.7         240.000000  5220.6  6155.5
2019-20                         8658.9         244.700000  4984.6  6298.0
2020-21                         7882.9         247.100000  5214.8  6747.0
2021-22                         7763.6         252.400000  4382.1  7717.0
2022-23*                        8542.3         217.772222  4979.8  6725.2

State/Union Territory  Chhattisgarh  NCT of Delhi         Goa  Gujarat  \
2004-05                      4383.3     14.300000  145.200000   1238.2
2005-06                      5011.6     24.000000  147.300000   1298.0
2006-07                      5041.4     31.100000  130.300000   1390.0
2007-08                      5426.6     31.400000  121.600000   1474.0
2008-09                      4391.8     31.400000  123.300000   1303.0
2009-10                      4110.4     19.300000  100.600000   1292.0
2010-11                      6159.0     19.600000  115.000000   1496.6
2011-12                      6028.4     19.800000  121.800000   1790.0
```

2012-13	6608.8	19.700000	122.800000	1541.0
2013-14	6716.4	29.600000	126.500000	1636.0
2014-15	6322.1	25.900000	120.500000	1830.9
2015-16	5789.4	17.300000	115.100000	1702.0
2016-17	8048.4	17.300000	113.200000	1930.0
2017-18	4930.8	16.800000	103.000000	1890.9
2018-19	6526.9	16.800000	98.800000	1912.1
2019-20	6774.8	16.800000	90.400000	1983.1
2020-21	7161.2	19.800000	87.300000	2145.7
2021-22	8021.7	19.000000	90.400000	2101.1
2022-23*	8238.3	21.661111	115.172222	2395.2

State/Union Territory	Haryana	Himachal Pradesh	…	Punjab	Rajasthan	\
2004-05	3023.00	122.00	…	10437.00	150.40	
2005-06	3210.00	112.10	…	10193.00	153.00	
2006-07	3371.00	123.50	…	10138.00	169.80	
2007-08	3613.00	121.50	…	10489.00	259.60	
2008-09	3298.00	118.30	…	11000.00	241.10	
2009-10	3625.00	105.90	…	11236.00	228.30	
2010-11	3472.00	128.90	…	10837.00	265.50	
2011-12	3759.00	131.60	…	10542.00	253.40	
2012-13	3976.00	125.30	…	11374.00	222.50	
2013-14	3998.00	120.80	…	11267.00	312.60	
2014-15	4006.00	125.20	…	11107.00	366.70	
2015-16	4145.00	129.90	…	11823.00	369.80	
2016-17	4453.00	146.60	…	11586.20	452.70	
2017-18	4523.38	114.79	…	13381.79	450.87	
2018-19	4516.10	114.90	…	12821.60	453.20	
2019-20	4824.30	143.80	…	11779.30	480.50	
2020-21	4424.90	140.50	…	12783.70	634.00	
2021-22	4618.00	167.50	…	12885.50	478.60	
2022-23*	5406.90	119.20	…	13146.70	577.40	

State/Union Territory	Sikkim	Tamil Nadu	Telangana	Tripura	\
2004-05	21.600000	5062.2	7211.446154	545.100000	
2005-06	21.500000	5220.0	7211.446154	552.900000	
2006-07	21.500000	6610.6	7211.446154	620.500000	
2007-08	22.900000	5040.2	7211.446154	624.600000	
2008-09	21.700000	5182.7	7211.446154	627.100000	
2009-10	24.300000	5665.2	7211.446154	640.000000	
2010-11	21.000000	5792.4	6535.600000	702.500000	
2011-12	20.900000	7458.7	5148.800000	718.300000	
2012-13	21.300000	4049.9	4647.600000	713.200000	
2013-14	20.300000	5349.8	5755.000000	711.800000	
2014-15	20.100000	5727.8	4440.800000	747.000000	
2015-16	13.100000	7517.1	3047.000000	794.800000	
2016-17	19.700000	2369.4	5173.400000	814.600000	

```
2017-18                17.630000      6638.9   6262.200000  812.100000
2018-19                17.200000      6130.9   6670.000000  793.200000
2019-20                16.100000      7171.1   7427.800000  810.200000
2020-21                16.200000      6881.2  10217.100000  803.100000
2021-22                16.000000      7906.6  12409.600000  811.000000
2022-23*               19.612778      7850.6  16013.900000  713.444444

State/Union Territory  Uttar Pradesh  Uttarakhand  West Bengal  ALL INDIA
2004-05                       9555.6        572.0      14884.8    83131.7
2005-06                      11133.7        590.0      14510.8    91793.4
2006-07                      11124.0        556.0      14745.9    93355.3
2007-08                      11780.0        593.0      14719.5    96692.9
2008-09                      13097.0        582.0      15037.3    99182.5
2009-10                      10807.1        608.0      14340.7    89092.9
2010-11                      11992.0        550.4      13045.9    95979.8
2011-12                      14022.0        594.0      14605.8   105310.9
2012-13                      14416.0        579.8      15023.7   105231.6
2013-14                      14636.0        578.6      15370.7   106645.5
2014-15                      12167.9        603.7      14677.2   104798.5
2015-16                      12501.0        639.1      15953.9   104408.2
2016-17                      13754.0        630.0      15302.5   109698.4
2017-18                      13274.0        646.7      14967.0   112757.6
2018-19                      15545.3        617.6      16242.2   116477.8
2019-20                      15517.8        658.4      15881.4   118870.3
2020-21                      15520.0        714.9      16524.4   124368.3
2021-22                      15271.5        716.1      16728.7   129471.4
2022-23*                     15171.3        641.7      15636.9   135542.0

[19 rows x 32 columns]
```

except telangana we have only one missing values for other states

[ ]:
```
df_ip = df_T.interpolate(method="linear")
df_ip.head(33)
```

[ ]:
```
State/Union Territory  Andhra Pradesh  Arunachal Pradesh   Assam   Bihar  \
2004-05                        9601.0         135.000000  3470.7  2472.2
2005-06                       11704.0         146.200000  3552.5  3495.5
2006-07                       11872.0         146.200000  2916.0  4989.3
2007-08                       13324.0         158.100000  3319.0  4418.1
2008-09                       14241.0         163.900000  4008.5  5590.3
2009-10                       10538.0         215.800000  4335.9  3599.3
2010-11                        7882.4         234.000000  4736.6  3102.1
2011-12                        7746.2         255.000000  4516.3  7162.6
2012-13                        6862.4         263.000000  5128.5  7529.3
2013-14                        6969.7         276.200000  4927.1  5505.8
2014-15                        7233.9         285.000000  5222.7  6356.7
```

2015-16	7488.7	204.000000	5125.1	6802.2
2016-17	7452.4	220.000000	4727.4	8239.3
2017-18	8166.2	233.300000	5283.7	8093.1
2018-19	8234.7	240.000000	5220.6	6155.5
2019-20	8658.9	244.700000	4984.6	6298.0
2020-21	7882.9	247.100000	5214.8	6747.0
2021-22	7763.6	252.400000	4382.1	7717.0
2022-23*	8542.3	217.772222	4979.8	6725.2

State/Union Territory	Chhattisgarh	NCT of Delhi	Goa	Gujarat	\
2004-05	4383.3	14.300000	145.200000	1238.2	
2005-06	5011.6	24.000000	147.300000	1298.0	
2006-07	5041.4	31.100000	130.300000	1390.0	
2007-08	5426.6	31.400000	121.600000	1474.0	
2008-09	4391.8	31.400000	123.300000	1303.0	
2009-10	4110.4	19.300000	100.600000	1292.0	
2010-11	6159.0	19.600000	115.000000	1496.6	
2011-12	6028.4	19.800000	121.800000	1790.0	
2012-13	6608.8	19.700000	122.800000	1541.0	
2013-14	6716.4	29.600000	126.500000	1636.0	
2014-15	6322.1	25.900000	120.500000	1830.9	
2015-16	5789.4	17.300000	115.100000	1702.0	
2016-17	8048.4	17.300000	113.200000	1930.0	
2017-18	4930.8	16.800000	103.000000	1890.9	
2018-19	6526.9	16.800000	98.800000	1912.1	
2019-20	6774.8	16.800000	90.400000	1983.1	
2020-21	7161.2	19.800000	87.300000	2145.7	
2021-22	8021.7	19.000000	90.400000	2101.1	
2022-23*	8238.3	21.661111	115.172222	2395.2	

State/Union Territory	Haryana	Himachal Pradesh	…	Punjab	Rajasthan	\
2004-05	3023.00	122.00	…	10437.00	150.40	
2005-06	3210.00	112.10	…	10193.00	153.00	
2006-07	3371.00	123.50	…	10138.00	169.80	
2007-08	3613.00	121.50	…	10489.00	259.60	
2008-09	3298.00	118.30	…	11000.00	241.10	
2009-10	3625.00	105.90	…	11236.00	228.30	
2010-11	3472.00	128.90	…	10837.00	265.50	
2011-12	3759.00	131.60	…	10542.00	253.40	
2012-13	3976.00	125.30	…	11374.00	222.50	
2013-14	3998.00	120.80	…	11267.00	312.60	
2014-15	4006.00	125.20	…	11107.00	366.70	
2015-16	4145.00	129.90	…	11823.00	369.80	
2016-17	4453.00	146.60	…	11586.20	452.70	
2017-18	4523.38	114.79	…	13381.79	450.87	
2018-19	4516.10	114.90	…	12821.60	453.20	
2019-20	4824.30	143.80	…	11779.30	480.50	

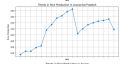
```
2020-21                4424.90            140.50  …  12783.70     634.00
2021-22                4618.00            167.50  …  12885.50     478.60
2022-23*               5406.90            119.20  …  13146.70     577.40

State/Union Territory     Sikkim  Tamil Nadu     Telangana     Tripura  \
2004-05                21.600000      5062.2   7211.446154  545.100000
2005-06                21.500000      5220.0   7211.446154  552.900000
2006-07                21.500000      6610.6   7211.446154  620.500000
2007-08                22.900000      5040.2   7211.446154  624.600000
2008-09                21.700000      5182.7   7211.446154  627.100000
2009-10                24.300000      5665.2   7211.446154  640.000000
2010-11                21.000000      5792.4   6535.600000  702.500000
2011-12                20.900000      7458.7   5148.800000  718.300000
2012-13                21.300000      4049.9   4647.600000  713.200000
2013-14                20.300000      5349.8   5755.000000  711.800000
2014-15                20.100000      5727.8   4440.800000  747.000000
2015-16                13.100000      7517.1   3047.000000  794.800000
2016-17                19.700000      2369.4   5173.400000  814.600000
2017-18                17.630000      6638.9   6262.200000  812.100000
2018-19                17.200000      6130.9   6670.000000  793.200000
2019-20                16.100000      7171.1   7427.800000  810.200000
2020-21                16.200000      6881.2  10217.100000  803.100000
2021-22                16.000000      7906.6  12409.600000  811.000000
2022-23*               19.612778      7850.6  16013.900000  713.444444

State/Union Territory  Uttar Pradesh  Uttarakhand  West Bengal  ALL INDIA
2004-05                       9555.6        572.0      14884.8    83131.7
2005-06                      11133.7        590.0      14510.8    91793.4
2006-07                      11124.0        556.0      14745.9    93355.3
2007-08                      11780.0        593.0      14719.5    96692.9
2008-09                      13097.0        582.0      15037.3    99182.5
2009-10                      10807.1        608.0      14340.7    89092.9
2010-11                      11992.0        550.4      13045.9    95979.8
2011-12                      14022.0        594.0      14605.8   105310.9
2012-13                      14416.0        579.8      15023.7   105231.6
2013-14                      14636.0        578.6      15370.7   106645.5
2014-15                      12167.9        603.7      14677.2   104798.5
2015-16                      12501.0        639.1      15953.9   104408.2
2016-17                      13754.0        630.0      15302.5   109698.4
2017-18                      13274.0        646.7      14967.0   112757.6
2018-19                      15545.3        617.6      16242.2   116477.8
2019-20                      15517.8        658.4      15881.4   118870.3
2020-21                      15520.0        714.9      16524.4   124368.3
2021-22                      15271.5        716.1      16728.7   129471.4
2022-23*                     15171.3        641.7      15636.9   135542.0

[19 rows x 32 columns]
```

[ ]:
```
y = df_T['ALL INDIA']
df_T_A = df_T.drop(columns=['ALL INDIA'])
```

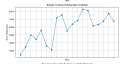
1 Trends Over Time

[ ]:
```
states = df_T.columns


fig, axs = plt.subplots(nrows=len(states), ncols=1, figsize=(8, 4*len(states)),␣
 ↪sharex=True)

for i, state in enumerate(states):
    axs[i].plot(df_T.index, df_T[state], marker='o', linestyle='-')
    axs[i].set_title(f'Trends in Rice Production in {state}')
    axs[i].set_ylabel('Rice Production')
    axs[i].set_xlabel('Year')
    axs[i].grid(True)
    axs[i].tick_params(axis='x', rotation=45)

fig.text(0.5, 0.04, 'Year', ha='center', va='center')


plt.tight_layout()
plt.show()
```

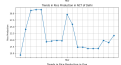

Trends in Rice Production in Andhra Pradesh
Year
Rice Production

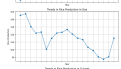

Trends in Rice Production in Arunachal Pradesh
Year
Rice Production

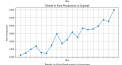

Trends in Rice Production in Assam
Year
Rice Production


Trends in Rice Production in Bihar
Year
Rice Production

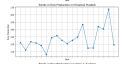

Trends in Rice Production in Chhattisgarh
Year
Rice Production

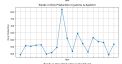

Trends in Rice Production in NCT of Delhi
Year
Rice Production

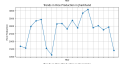

Trends in Rice Production in Goa
Year
Rice Production

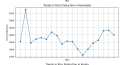

Trends in Rice Production in Gujarat
Year
Rice Production

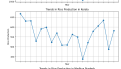

Trends in Rice Production in Haryana
Year
Rice Production

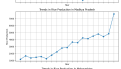

Trends in Rice Production in Himachal Pradesh
Year
Rice Production

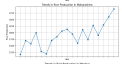

Trends in Rice Production in Jammu & Kashmir
Year
Rice Production

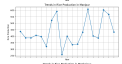

Trends in Rice Production in Jharkhand
Year
Rice Production


Trends in Rice Production in Karnataka
Year
Rice Production


Trends in Rice Production in Kerala
Year
Rice Production


Trends in Rice Production in Madhya Pradesh
Year
Rice Production

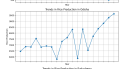

Trends in Rice Production in Maharashtra
Year
Rice Production

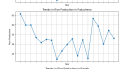

Trends in Rice Production in Manipur
Year
Rice Production

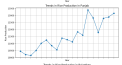

Trends in Rice Production in Meghalaya
Year
Rice Production


Trends in Rice Production in Mizoram
Year
Rice Production


Trends in Rice Production in Nagaland
Year
Rice Production

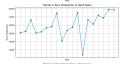

Trends in Rice Production in Odisha
Year
Rice Production

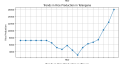

Trends in Rice Production in Puducherry
Year
Rice Production

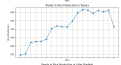

Trends in Rice Production in Punjab
Year
Rice Production

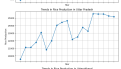

Trends in Rice Production in Rajasthan
Year
Rice Production

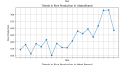

Trends in Rice Production in Sikkim
Year
Rice Production


Trends in Rice Production in Tamil Nadu
Year
Rice Production


Trends in Rice Production in Telangana
Year
Rice Production


Trends in Rice Production in Tripura
Year
Rice Production

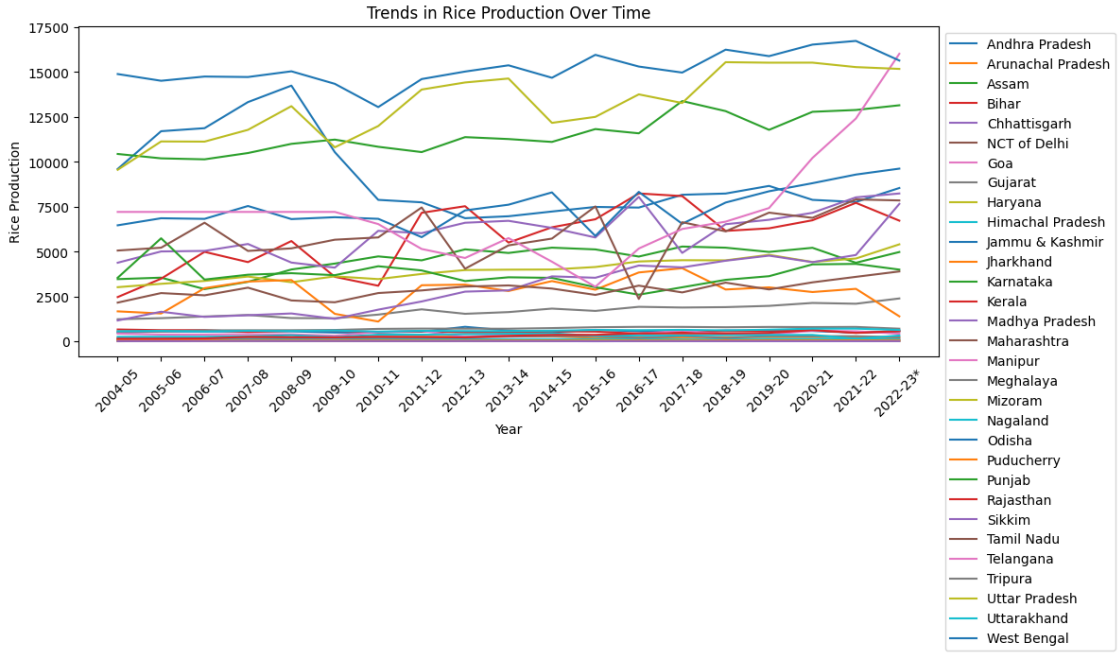

Trends in Rice Production in Uttar Pradesh
Year
Rice Production


Trends in Rice Production in Uttarakhand
Year
Rice Production


Trends in Rice Production in West Bengal
Year
Rice Production

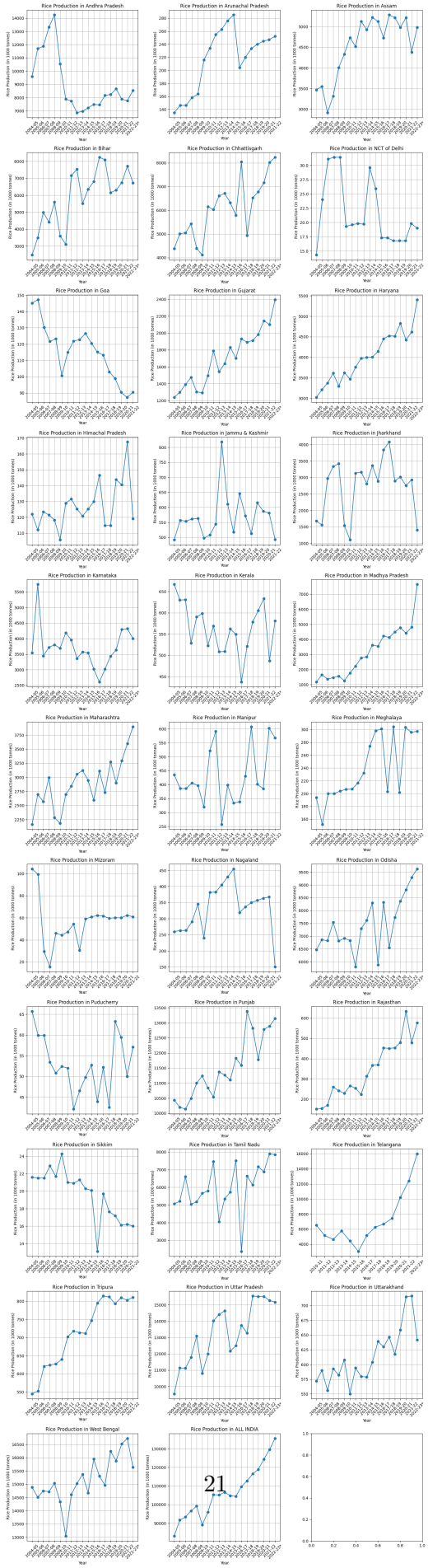

Trends in Rice Production in Goa, Bihar
Year
Rice Production

[ ]:
```
fig, ax = plt.subplots(figsize=(12, 6))


for state in df_T_A.columns:
    ax.plot(df_T_A.index, df_T_A[state], label=state)

ax.set_xlabel('Year')
ax.set_ylabel('Rice Production')
ax.set_title('Trends in Rice Production Over Time')


ax.legend(loc='upper left', bbox_to_anchor=(1, 1))


plt.xticks(rotation=45)


plt.tight_layout()
plt.show()
```



[ ]:
```
num_states = df.shape[0]
num_cols = 3
num_rows = (num_states + num_cols - 1) // num_cols
```

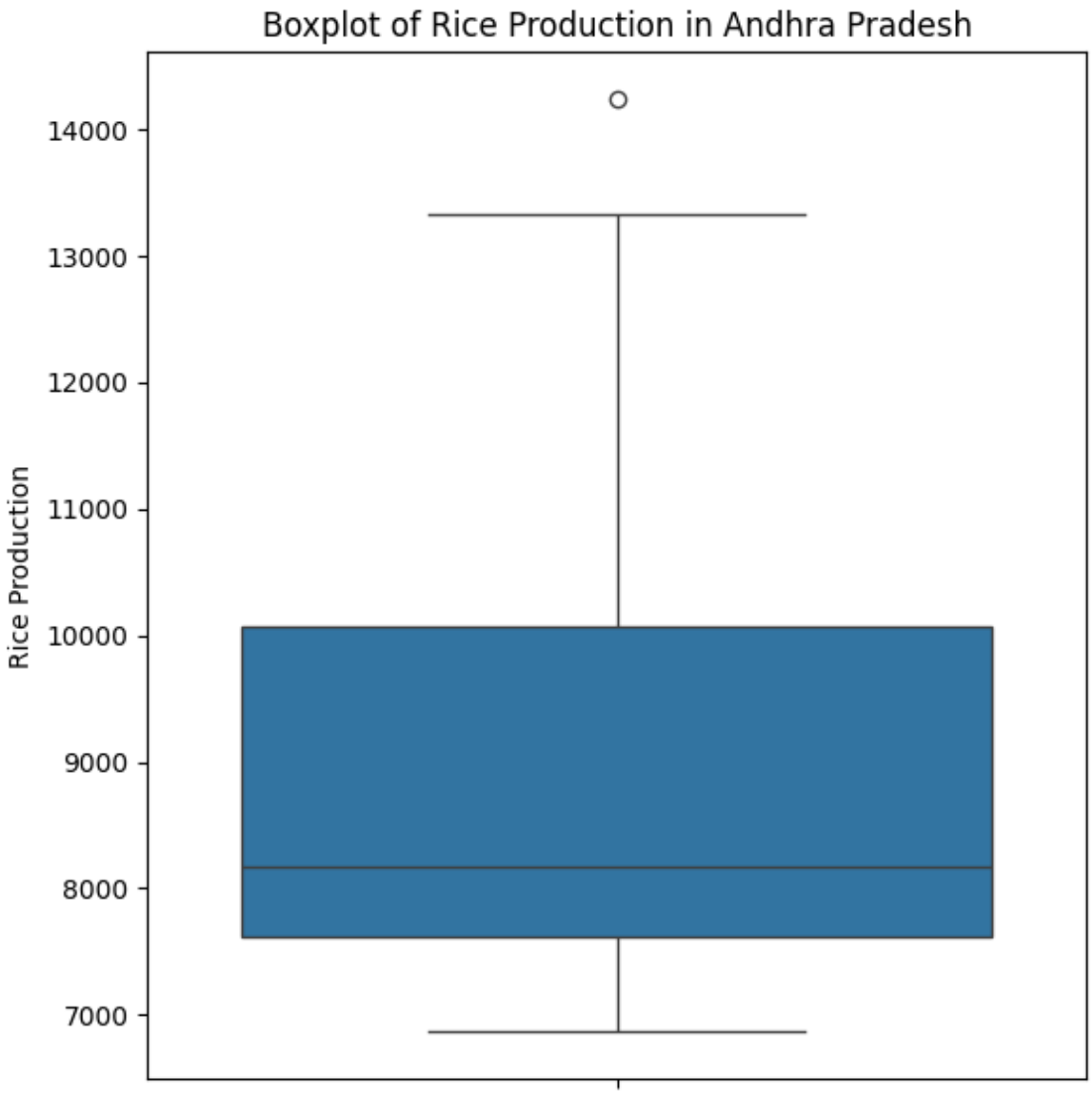
```
fig, axes = plt.subplots(num_rows, num_cols, figsize=(15, 5*num_rows))


if num_rows == 1:
    axes = [axes]

for i, state in enumerate(df.index):
    row_idx = i // num_cols
    col_idx = i % num_cols
    axes[row_idx][col_idx].plot(df.columns, df.loc[state].values, marker='o',␣
 ↪linestyle='-')
    axes[row_idx][col_idx].set_title(f'Rice Production in {state}')
    axes[row_idx][col_idx].set_xlabel('Year')
    axes[row_idx][col_idx].set_ylabel('Rice Production (in 1000 tonnes)')
    axes[row_idx][col_idx].tick_params(axis='x', rotation=45)
    axes[row_idx][col_idx].grid(True)


plt.tight_layout()
plt.show()
```

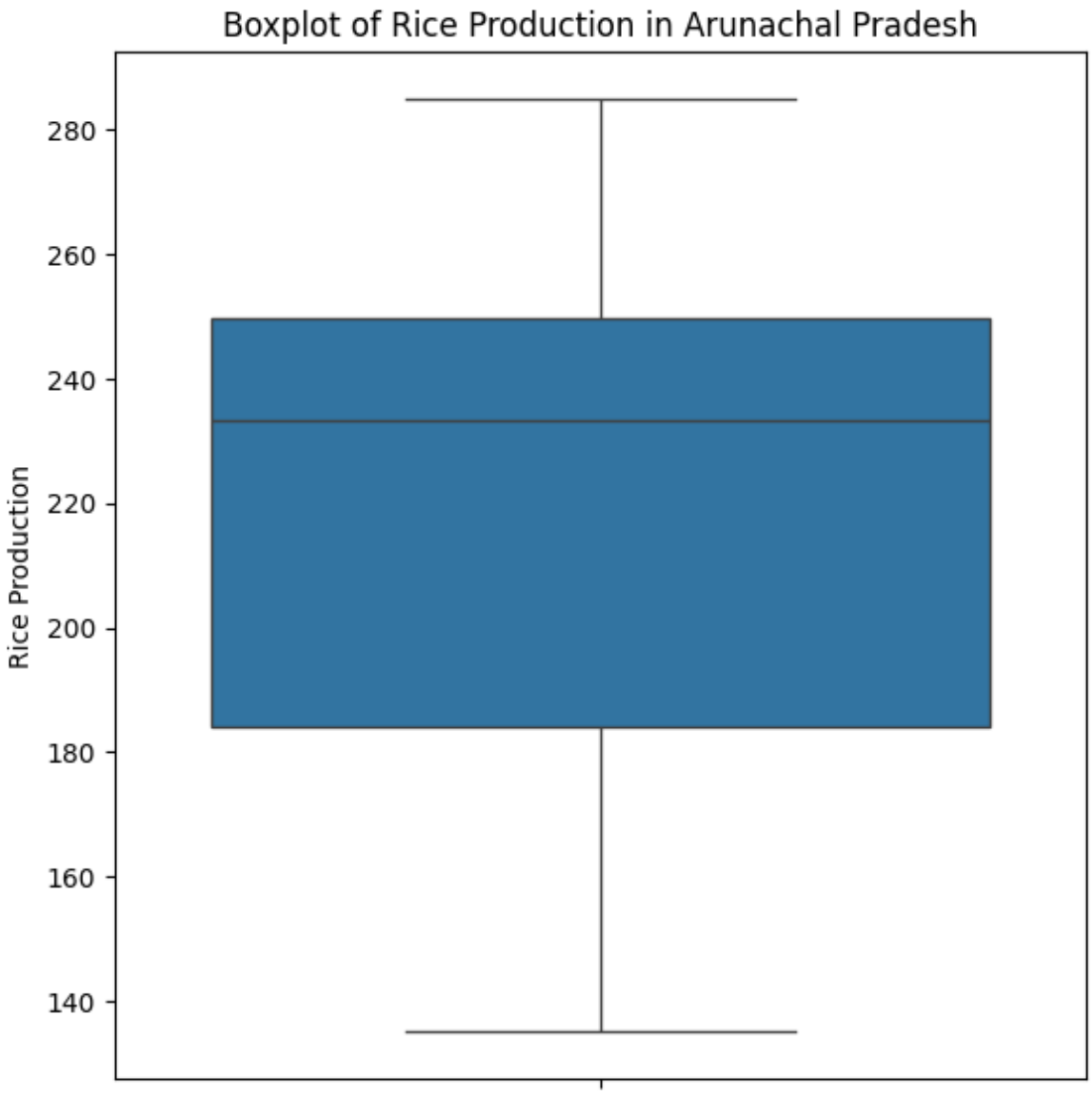

Rice Production in Andhra Pradesh
Rice Production in Arunachal Pradesh
Rice Production in Assam
Rice Production in Bihar
Rice Production in Chhattisgarh
Rice Production in NCT of Delhi
Rice Production in Goa
Rice Production in Gujarat
Rice Production in Haryana
Rice Production in Himachal Pradesh
Rice Production in Jammu & Kashmir
Rice Production in Jharkhand
Rice Production in Karnataka
Rice Production in Kerala
Rice Production in Madhya Pradesh
Rice Production in Maharashtra
Rice Production in Manipur
Rice Production in Meghalaya
Rice Production in Mizoram
Rice Production in Nagaland
Rice Production in Odisha
Rice Production in Puducherry
Rice Production in Punjab
Rice Production in Rajasthan
Rice Production in Sikkim
Rice Production in Tamil Nadu
Rice Production in Telangana
Rice Production in Tripura
Rice Production in Uttar Pradesh
Rice Production in Uttarakhand
Rice Production in West Bengal
Rice Production in ALL INDIA
Year

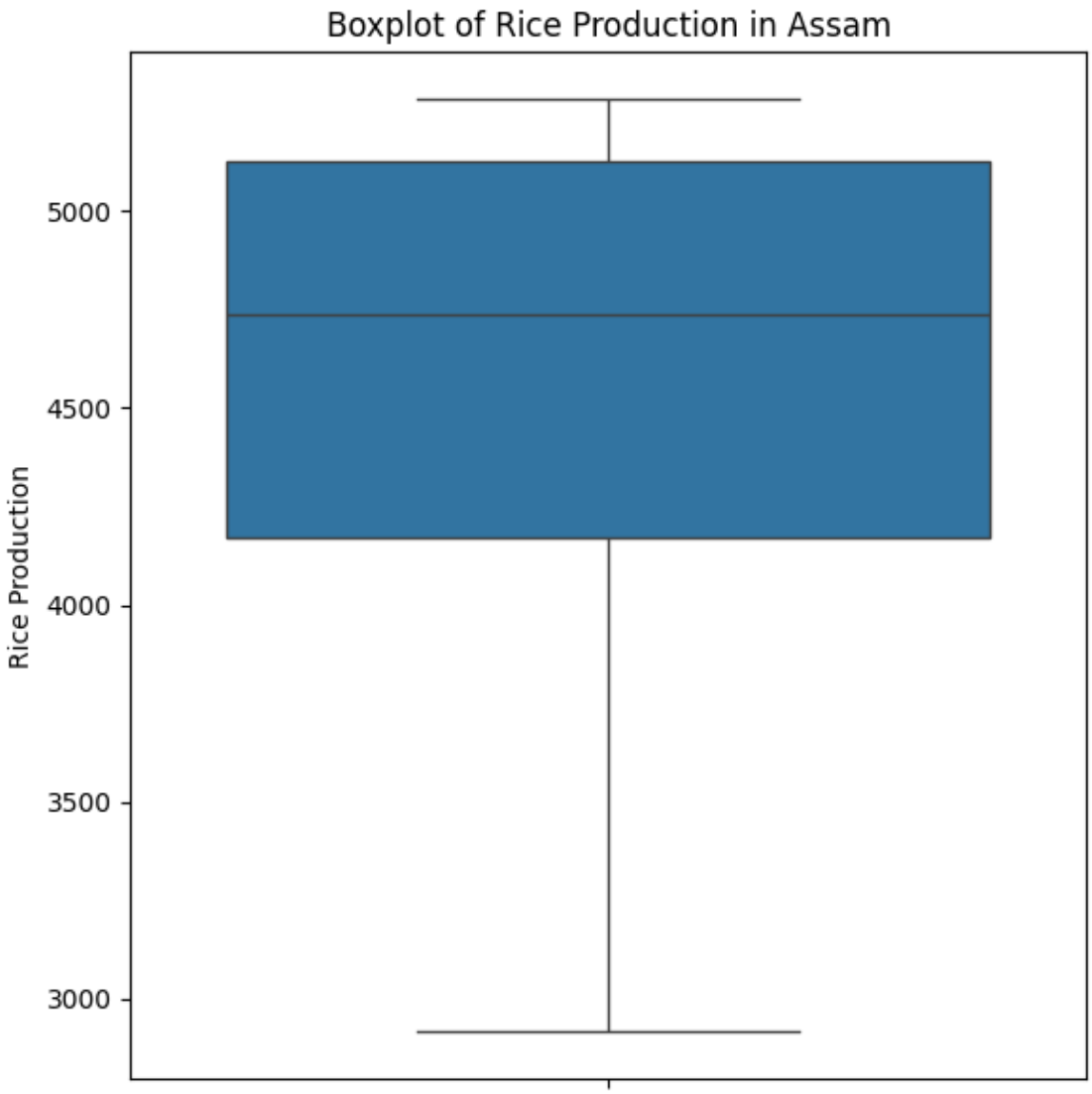
Rice prod in Andhra pradesh, Delhi NCT, Goa, J and K, Jharkhand, nagaland, Sikkim has fallen over the years and needs to increase

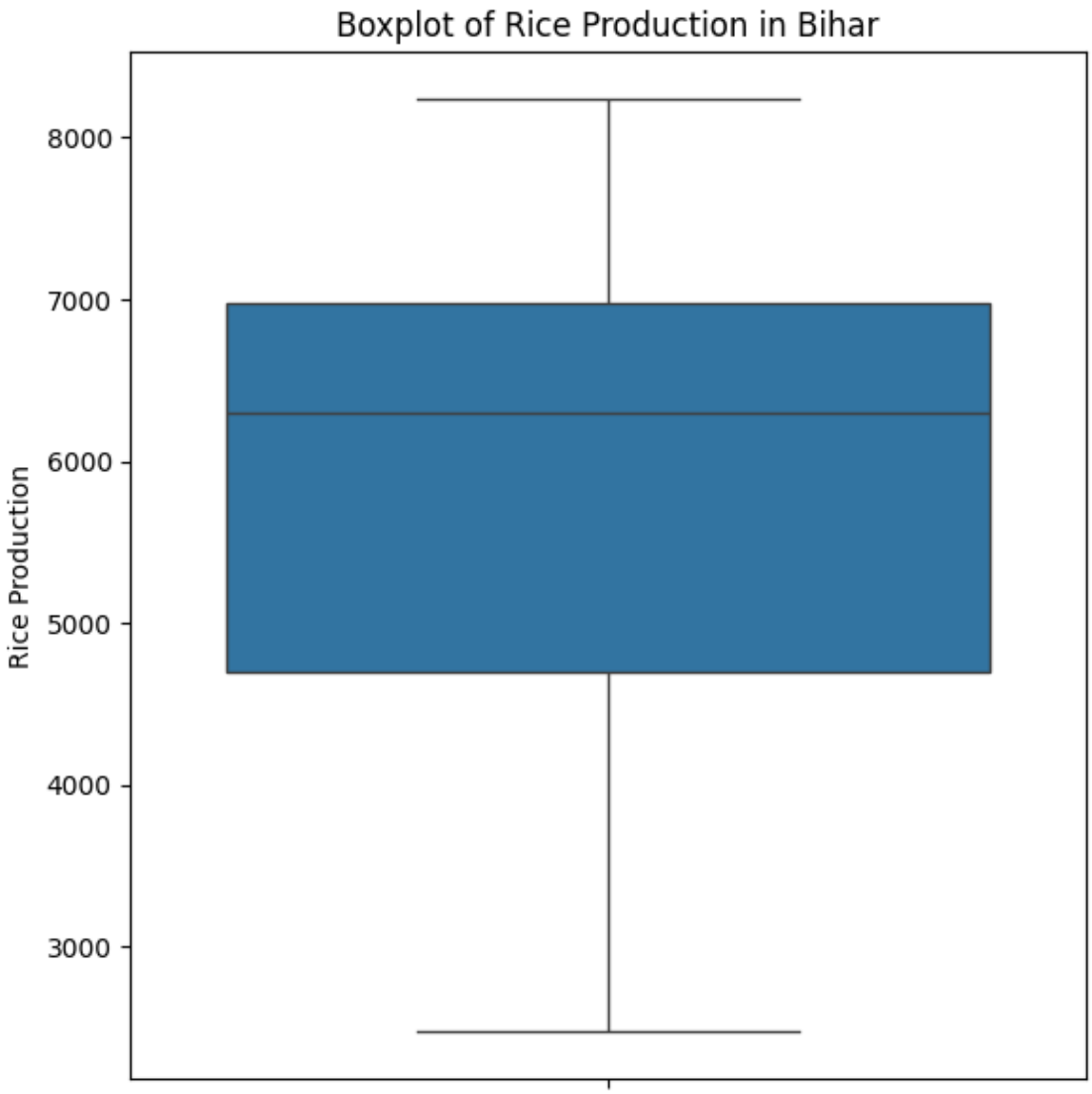
2 Outlier Detection

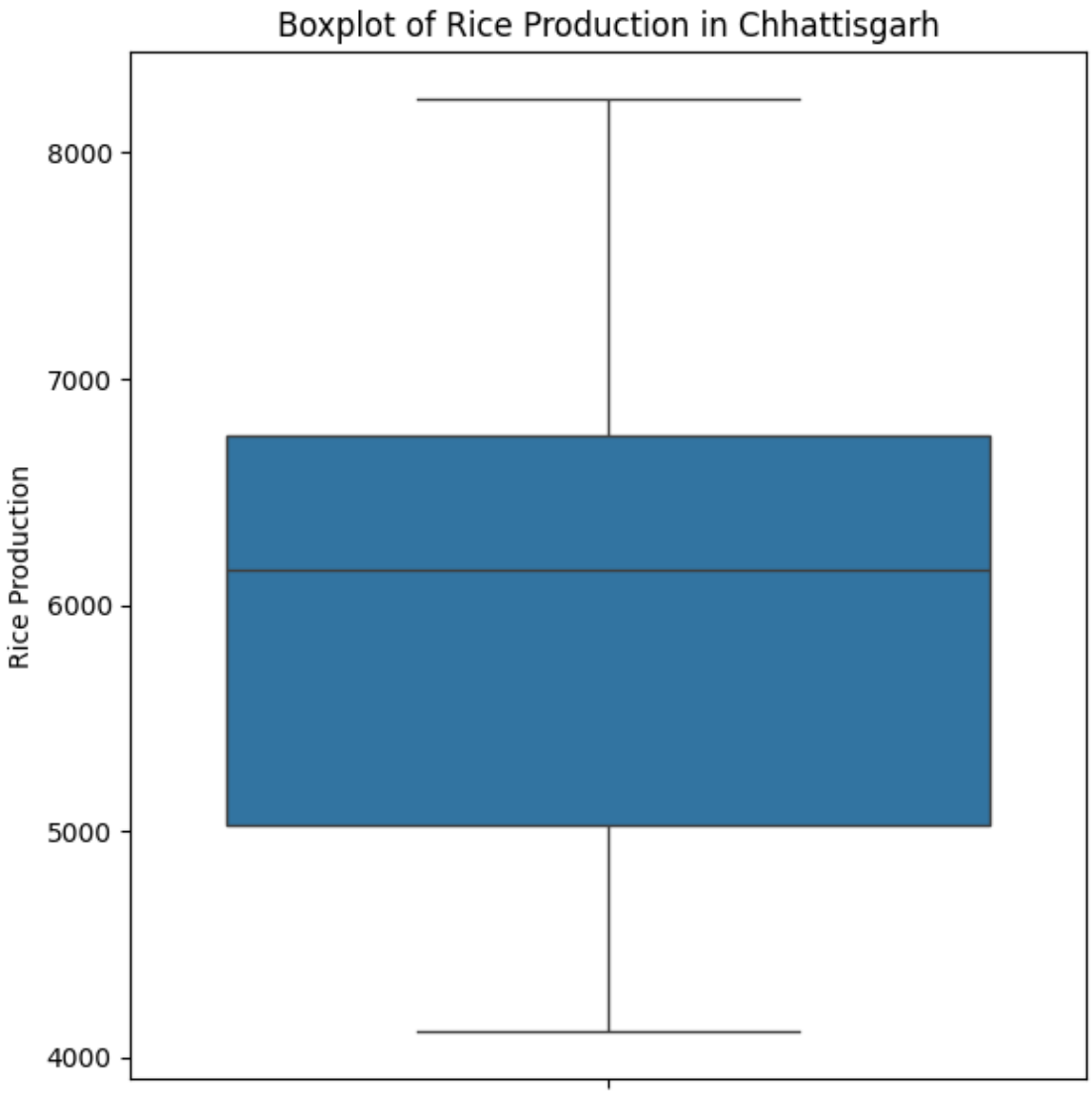
[ ]:
```
states = df_T.columns


for state in states:

    fig, ax = plt.subplots(figsize=(6, 6))


    sns.boxplot(y=df_T[state], ax=ax)


    ax.set_ylabel('Rice Production')
    ax.set_title(f'Boxplot of Rice Production in {state}')


    plt.tight_layout()
    plt.show()
```

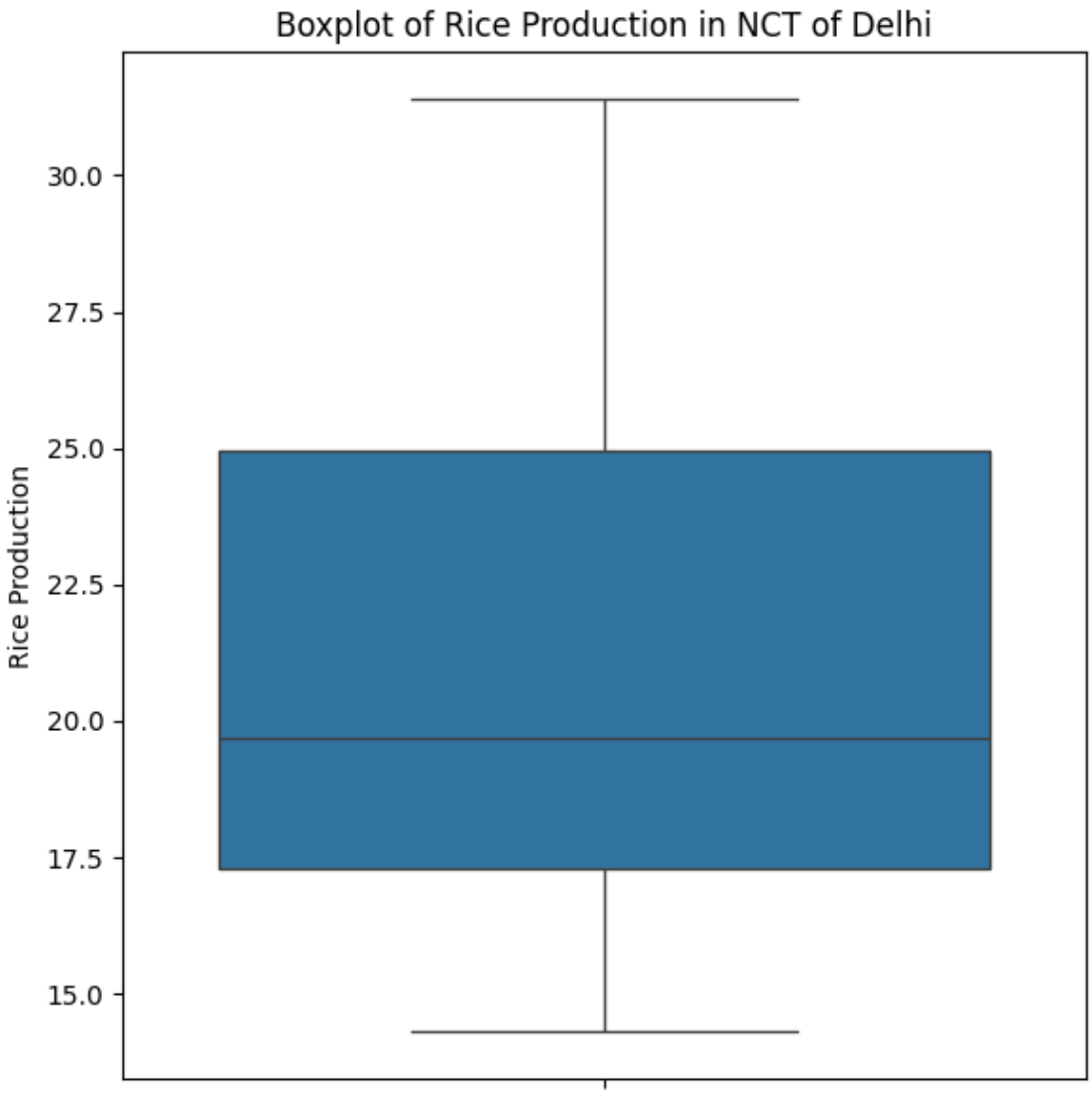

Boxplot of Rice Production in Andhra Pradesh
Rice Production
14000
13000
12000
11000
10000
9000
8000
7000

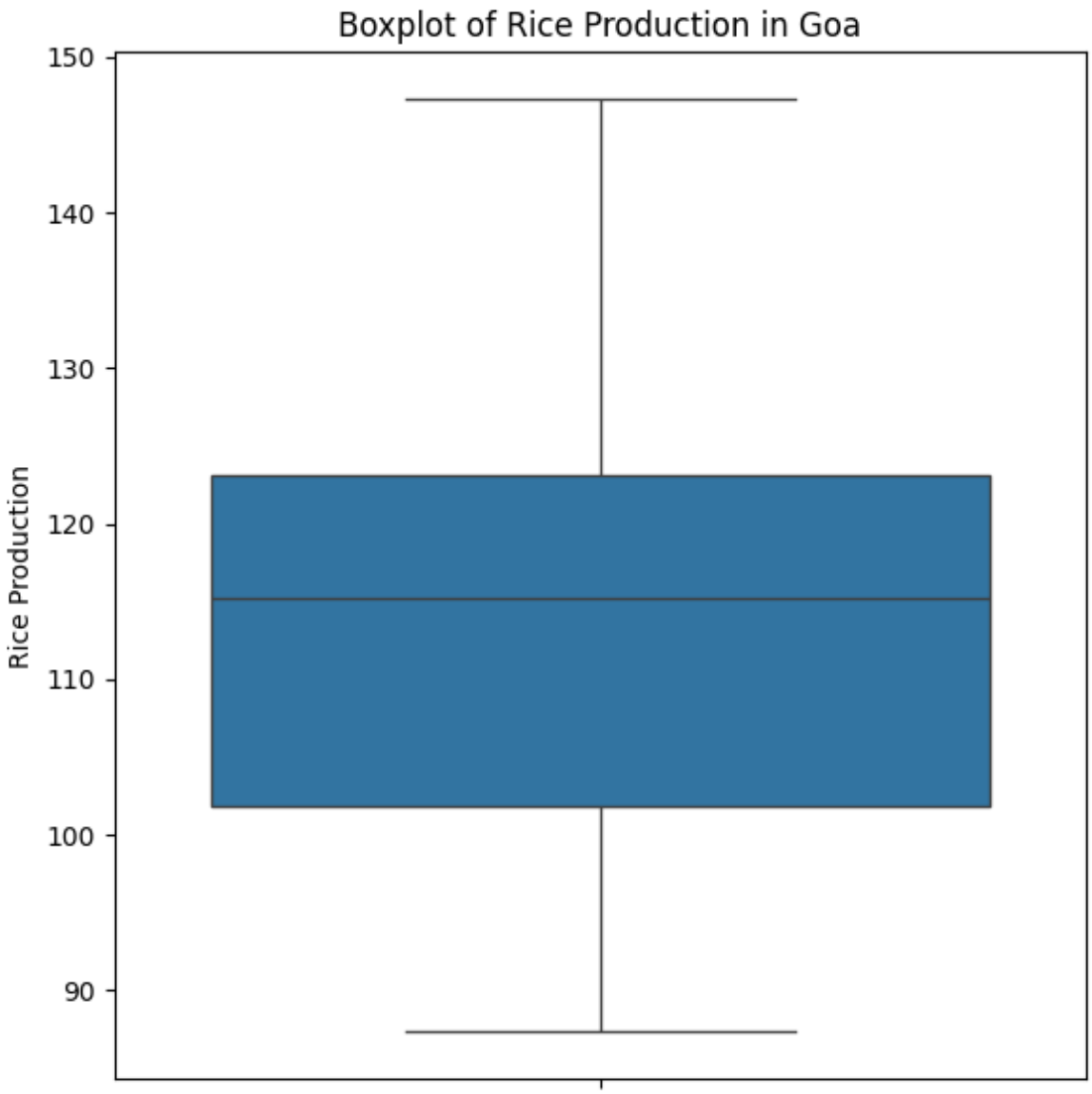

Boxplot of Rice Production in Arunachal Pradesh
Rice Production
280
260
240
220
200
180
160
140

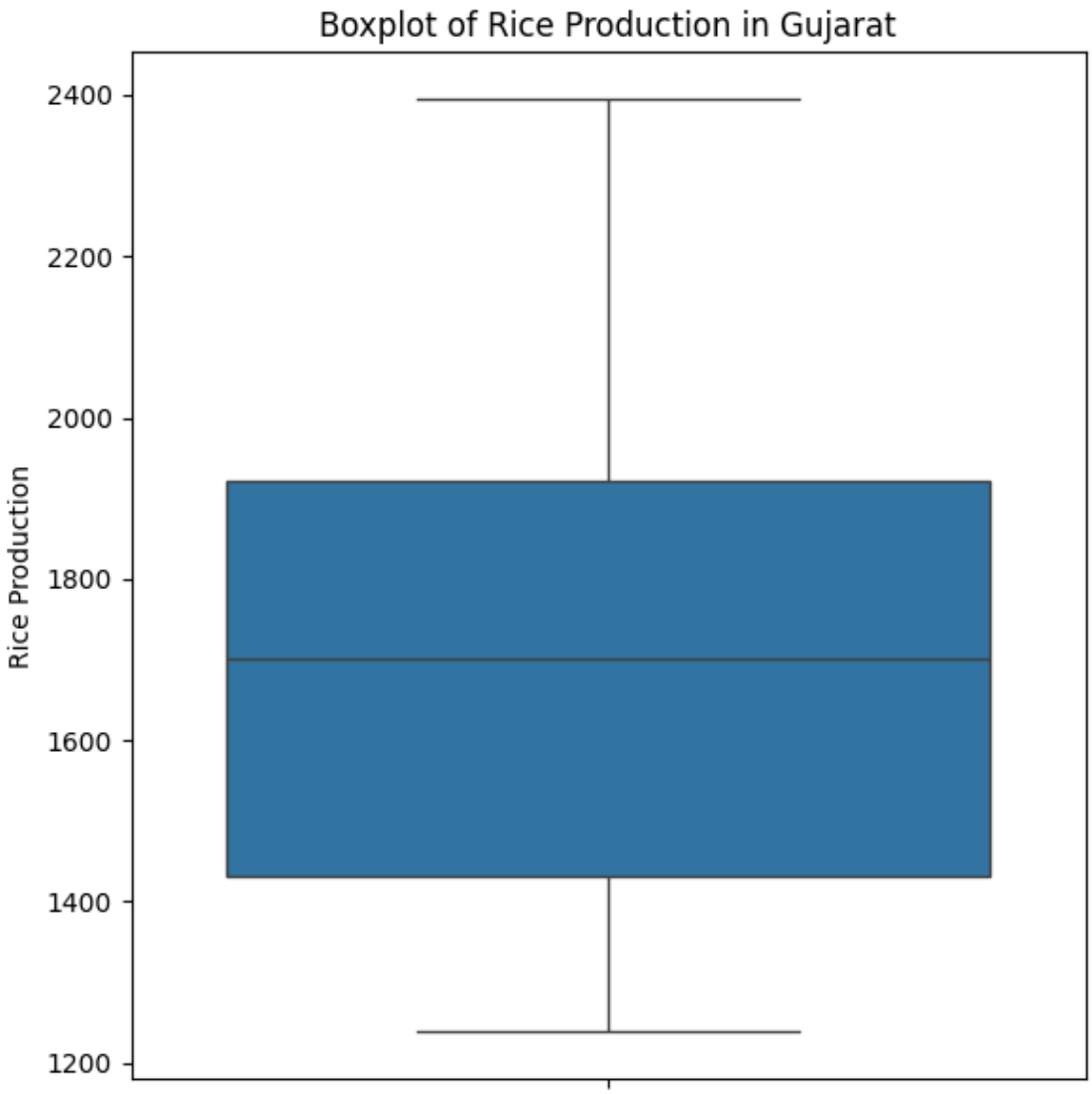

Boxplot of Rice Production in Assam
Rice Production
5000
4500
4000
3500
3000

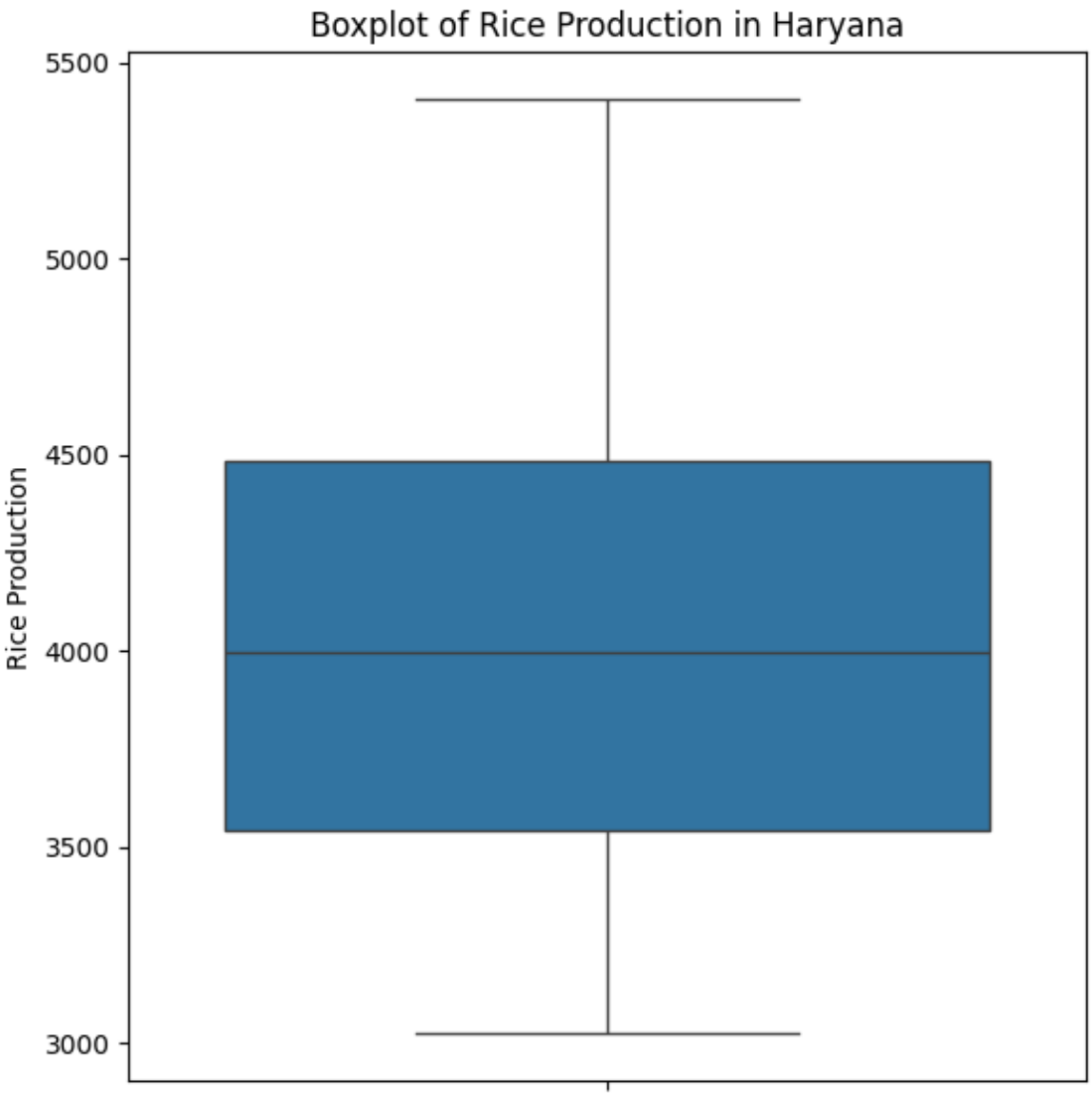

Boxplot of Rice Production in Bihar
Rice Production
8000
7000
6000
5000
4000
3000

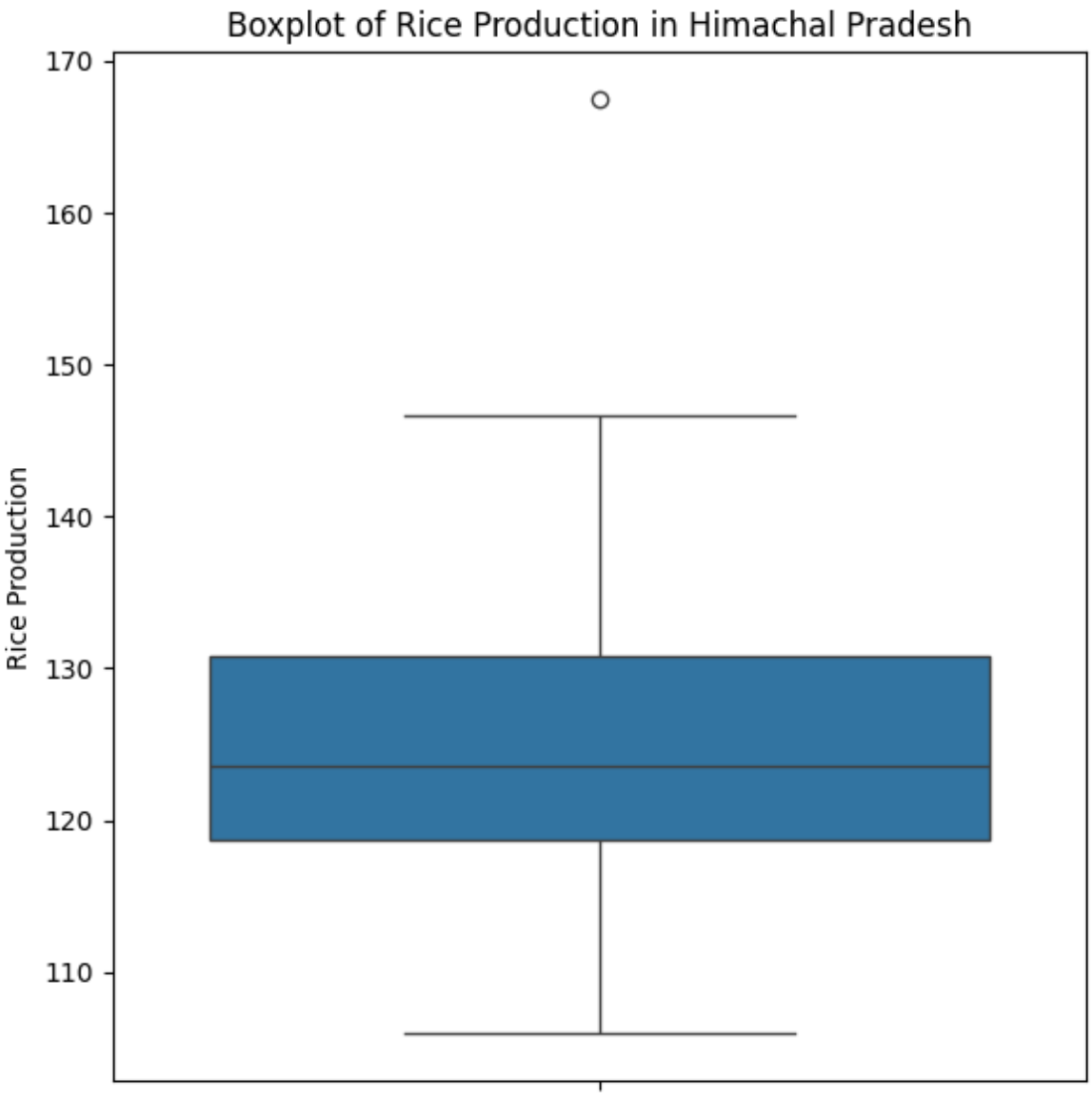

Boxplot of Rice Production in Chhattisgarh
Rice Production
8000
7000
6000
5000
4000

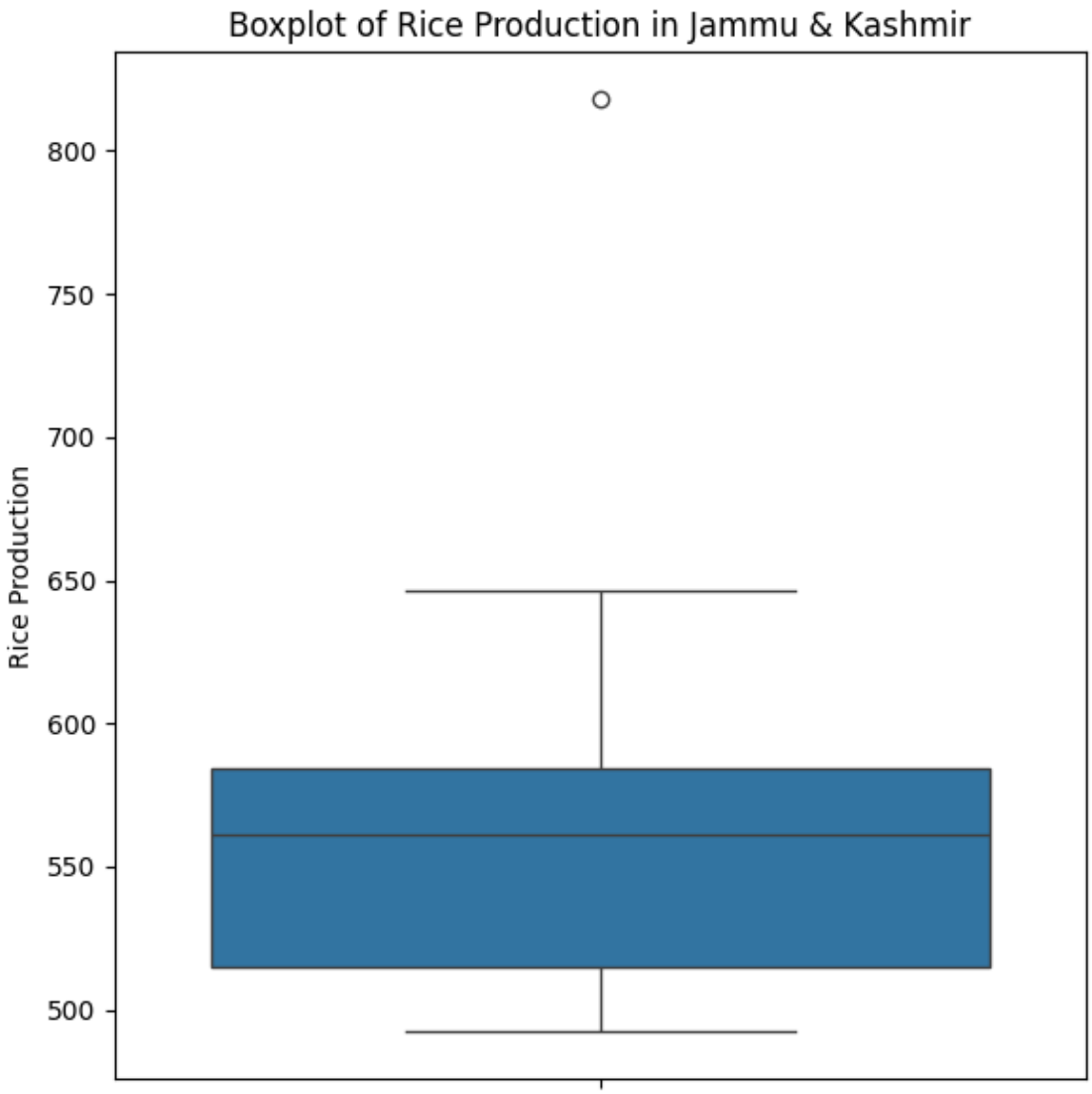

Boxplot of Rice Production in NCT of Delhi
Rice Production
30.0
27.5
25.0
22.5
20.0
17.5
15.0

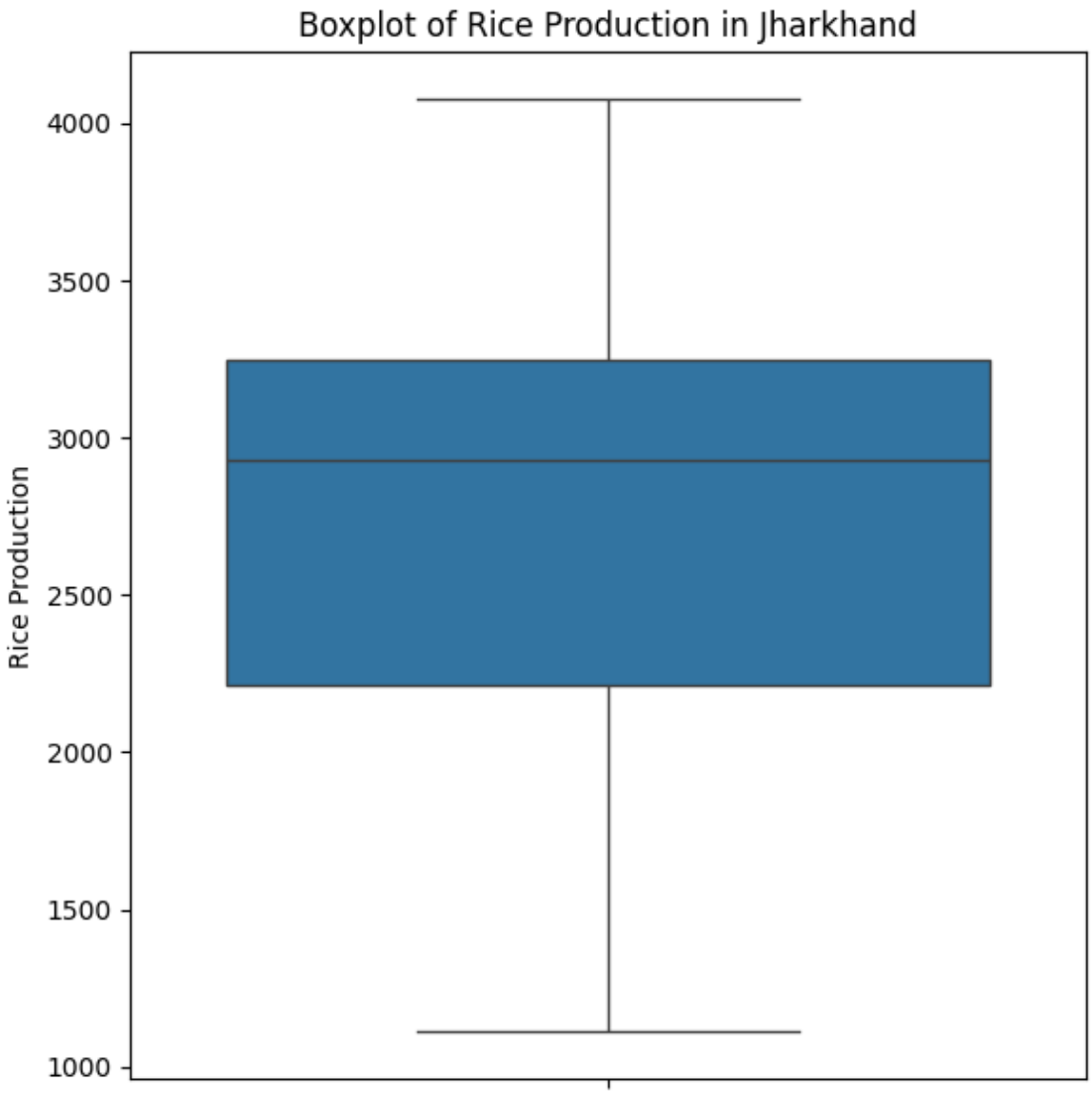

Boxplot of Rice Production in Goa
Rice Production
150
140
130
120
110
100
90

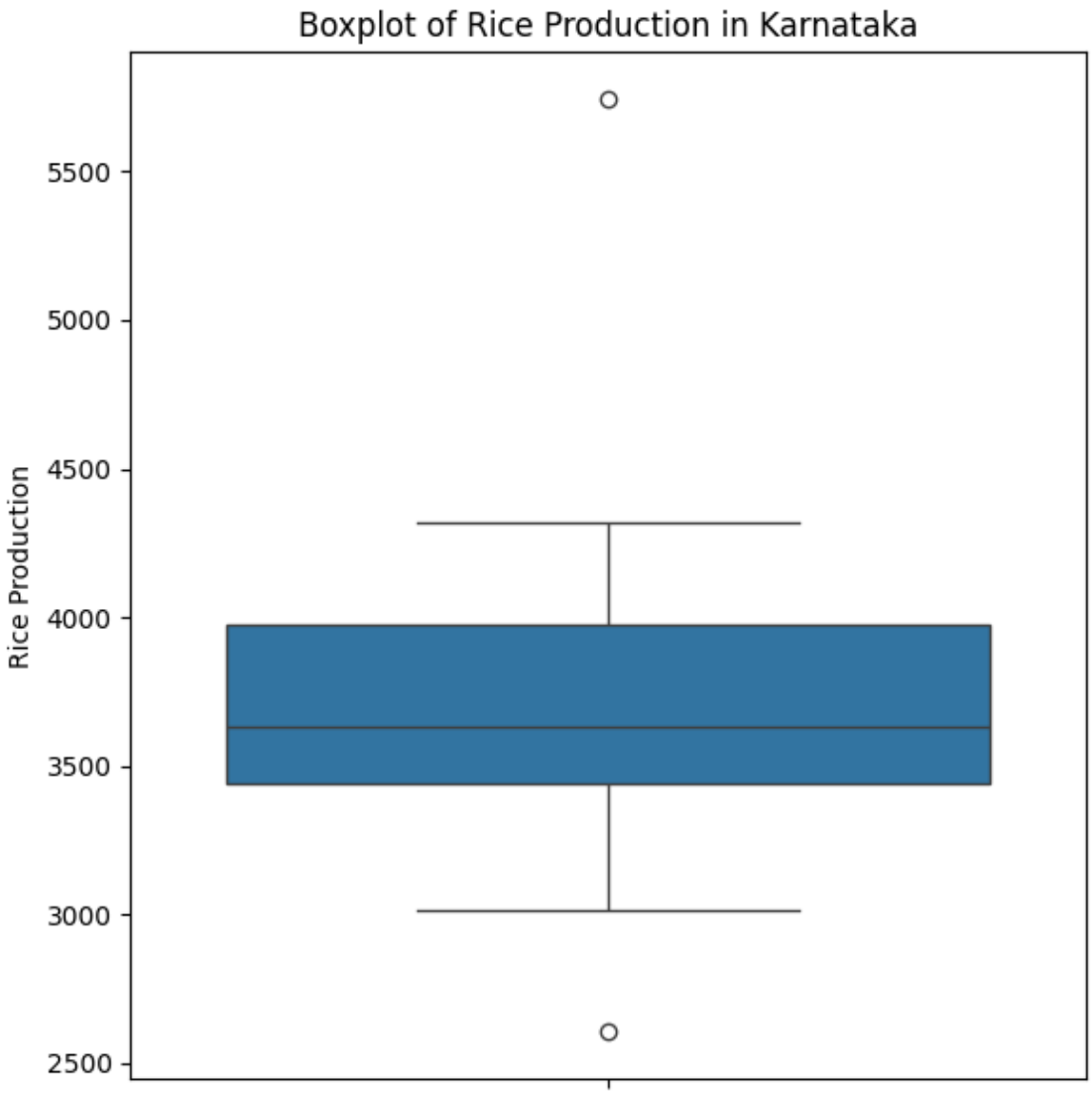

Boxplot of Rice Production in Gujarat
Rice Production
2400
2200
2000
1800
1600
1400
1200

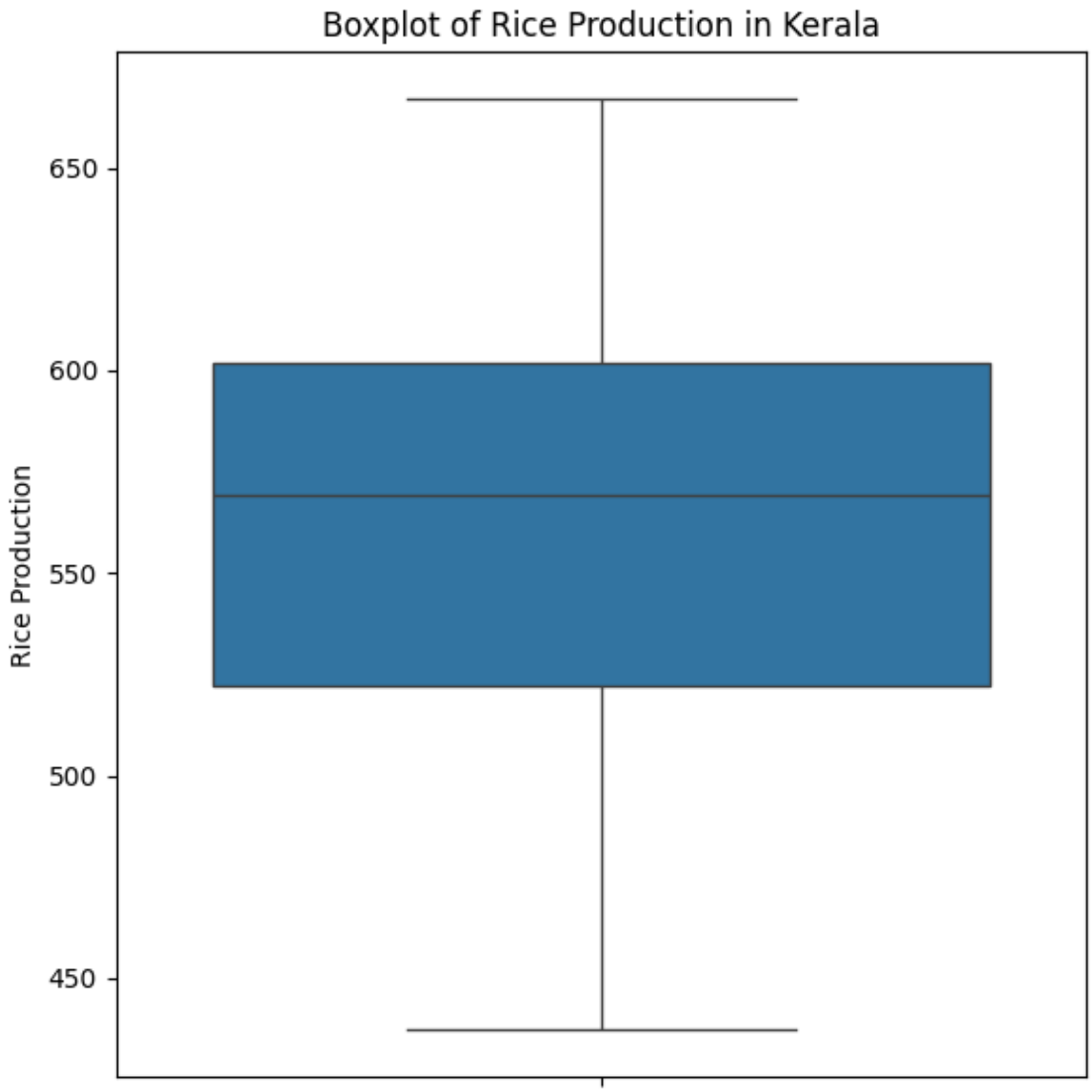

Boxplot of Rice Production in Haryana
Rice Production
5500
5000
4500
4000
3500
3000

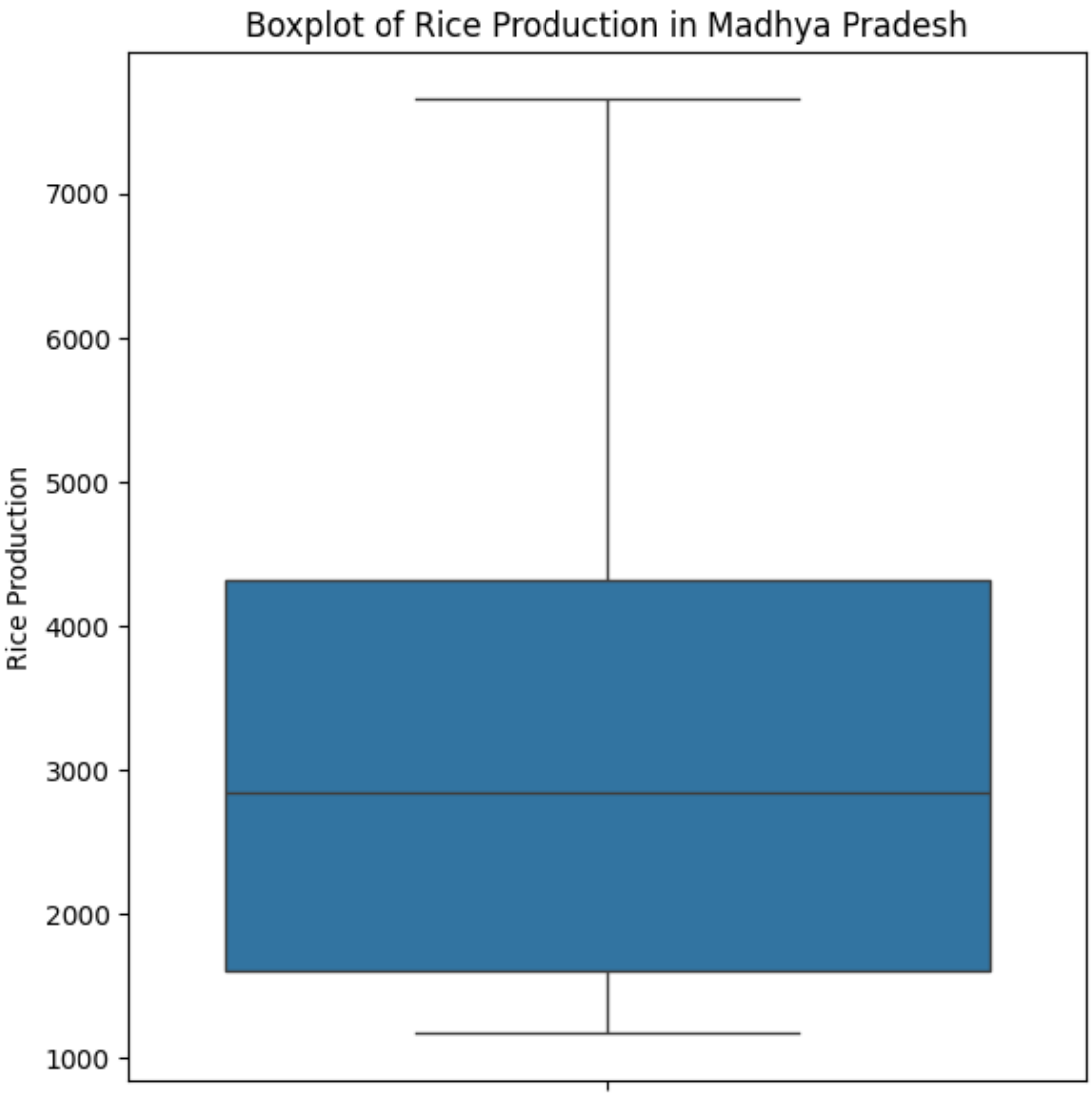

Boxplot of Rice Production in Himachal Pradesh
Rice Production
170
160
150
140
130
120
110

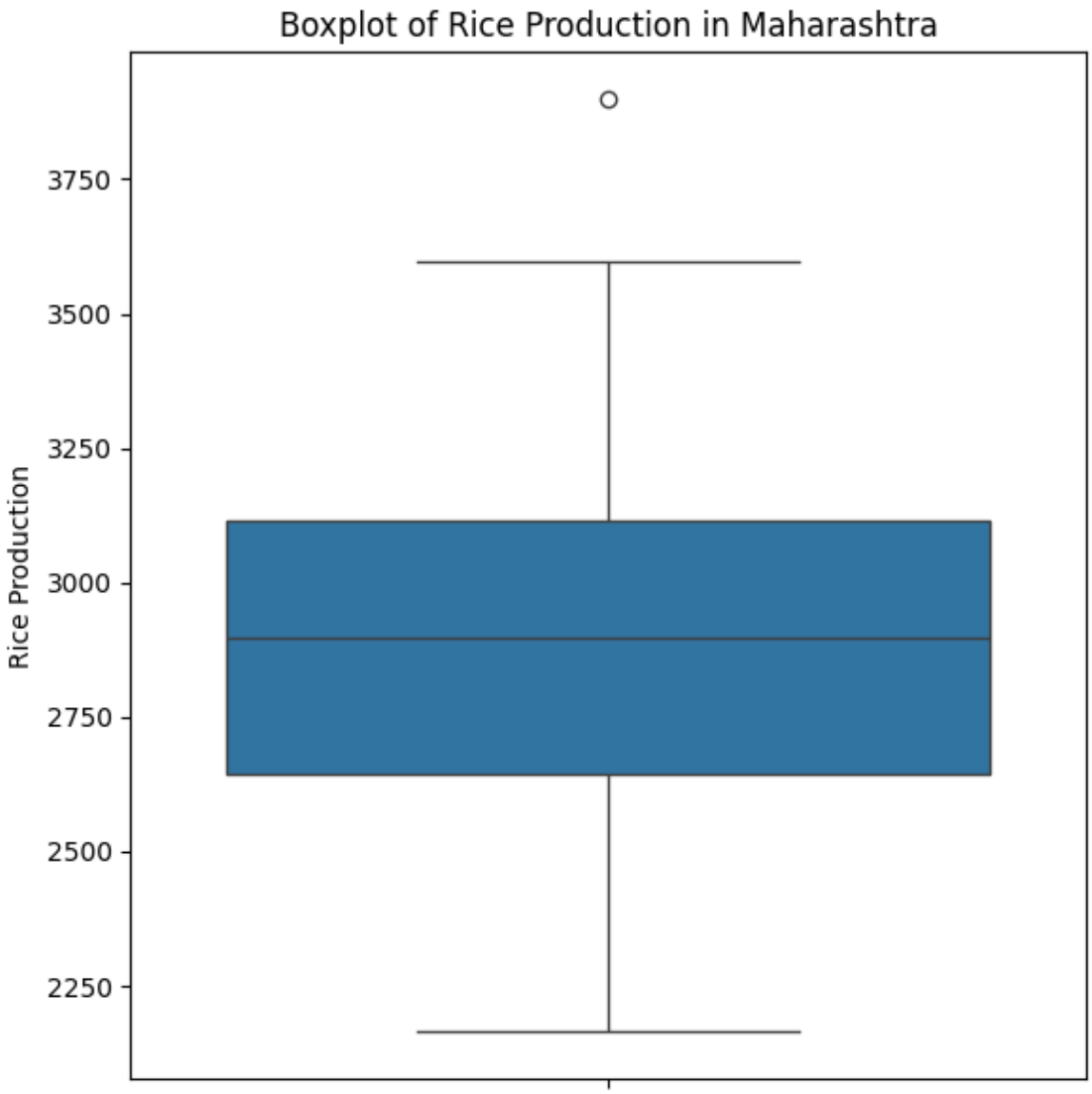

Boxplot of Rice Production in Jammu & Kashmir
Rice Production
800
750
700
650
600
550
500

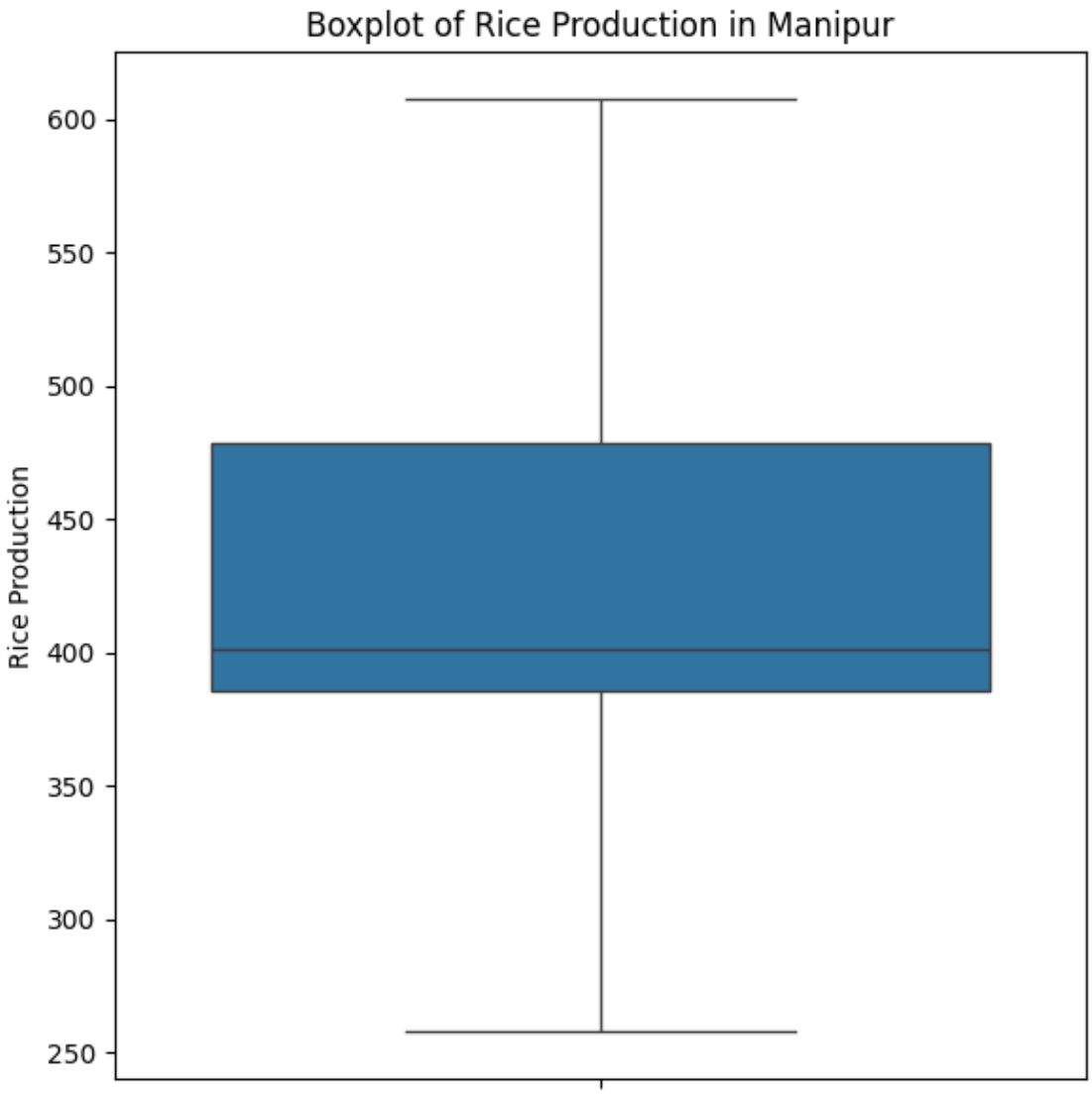

Boxplot of Rice Production in Jharkhand
Rice Production
4000
3500
3000
2500
2000
1500
1000

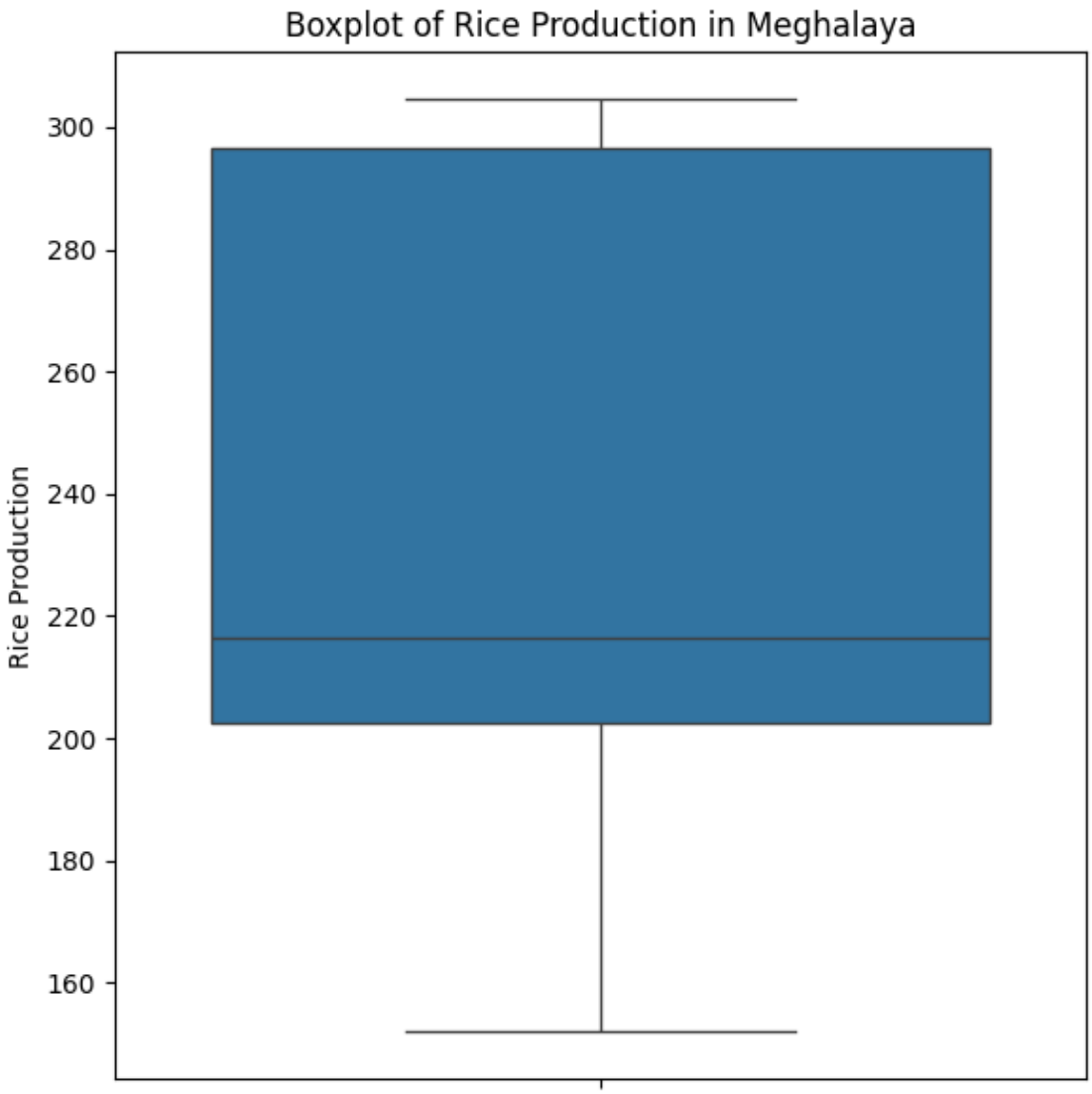

Boxplot of Rice Production in Karnataka
Rice Production
5500
5000
4500
4000
3500
3000
2500

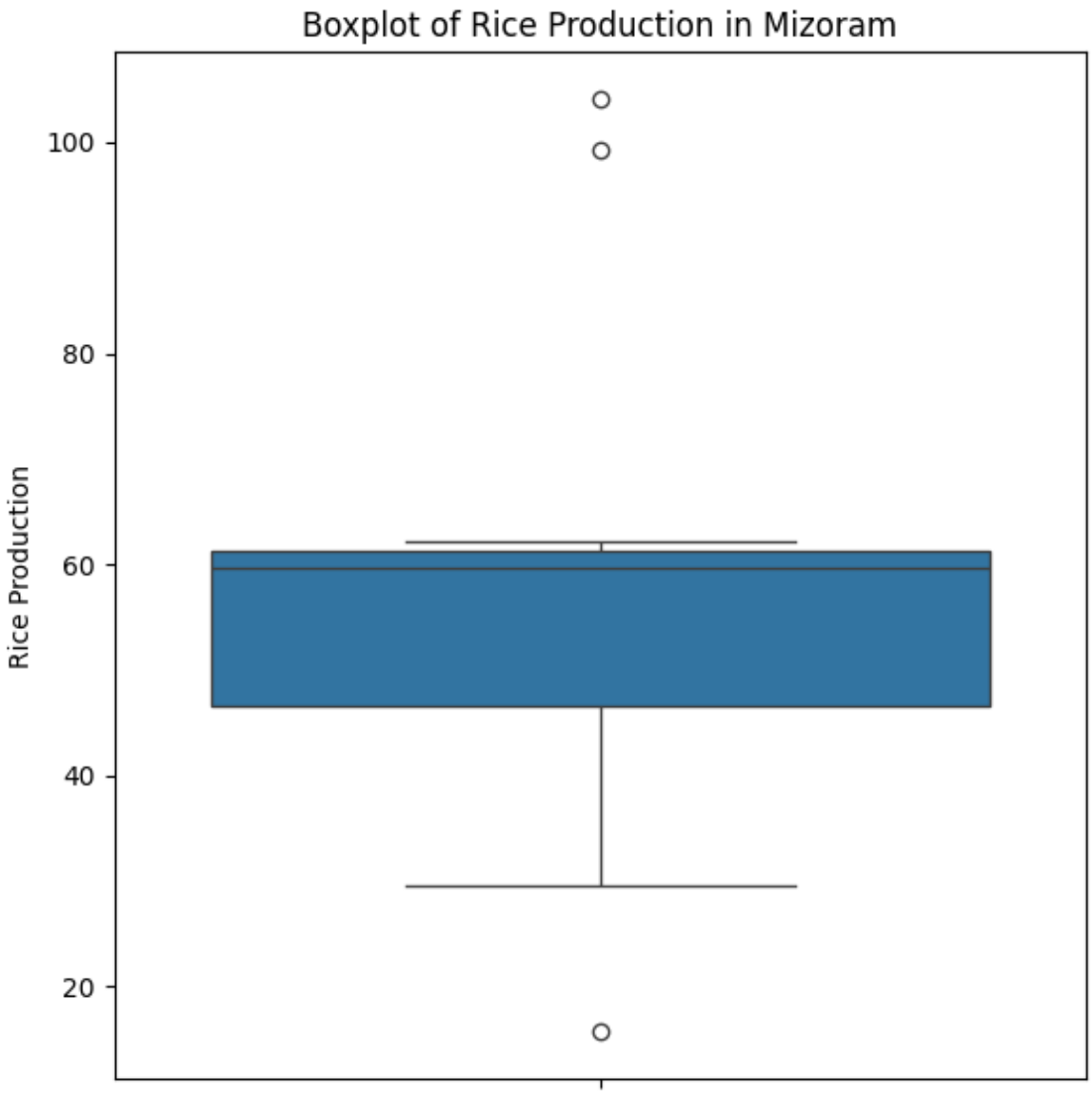

Boxplot of Rice Production in Kerala
Rice Production
650
600
550
500
450

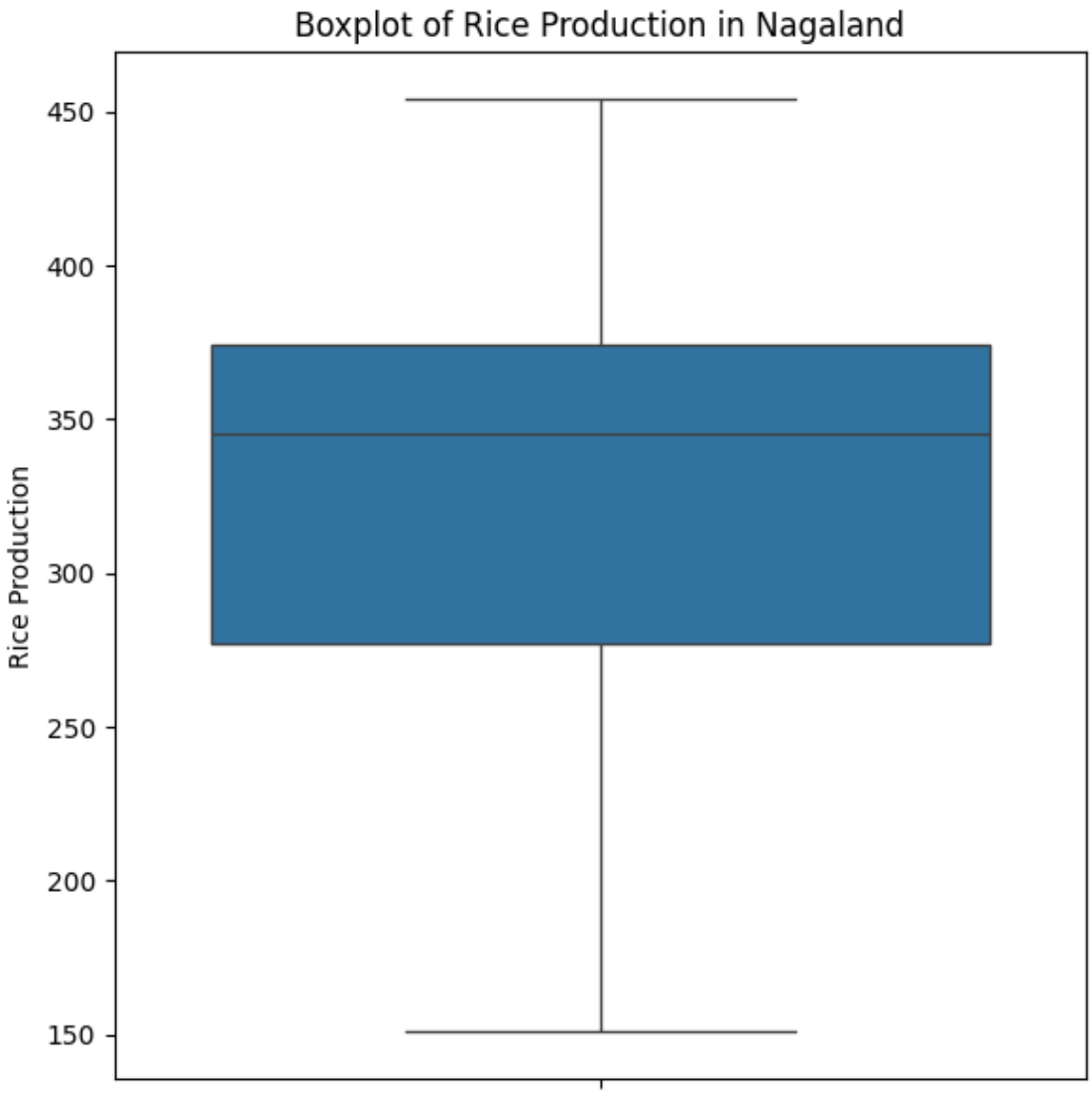

Boxplot of Rice Production in Madhya Pradesh
Rice Production
7000
6000
5000
4000
3000
2000
1000

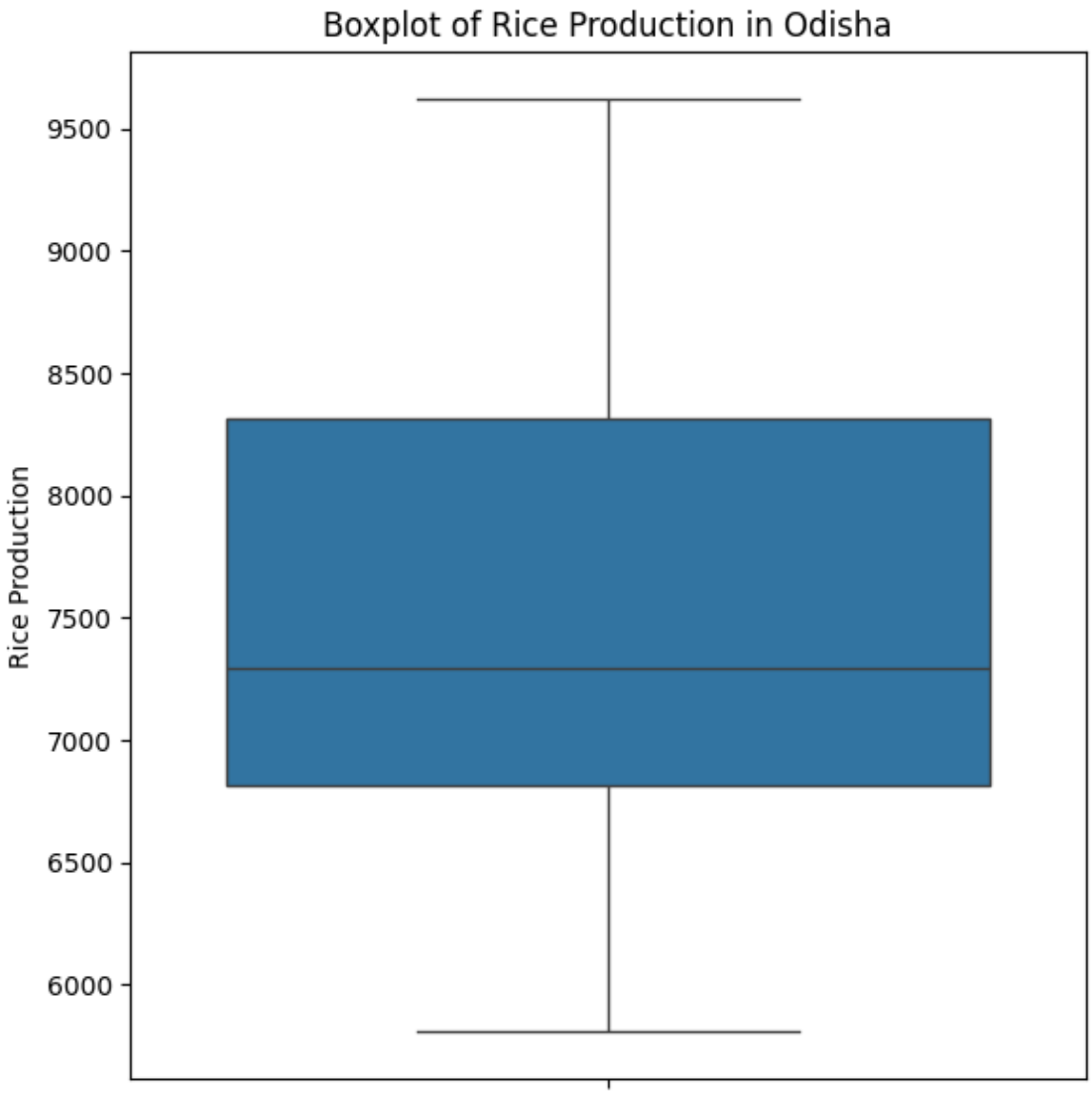

Boxplot of Rice Production in Maharashtra
Rice Production
3750
3500
3250
3000
2750
2500
2250

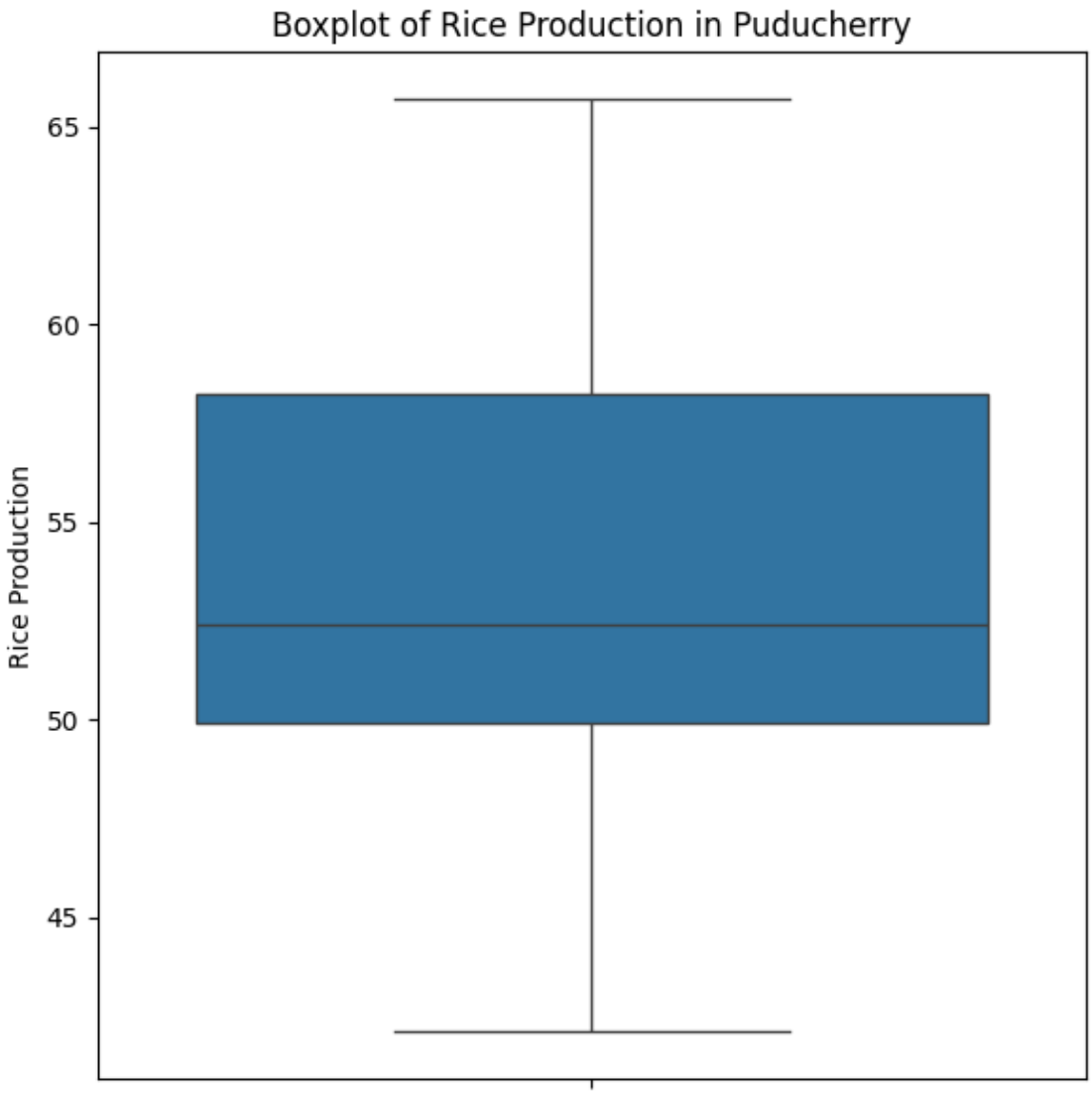

Boxplot of Rice Production in Manipur
Rice Production
600
550
500
450
400
350
300
250

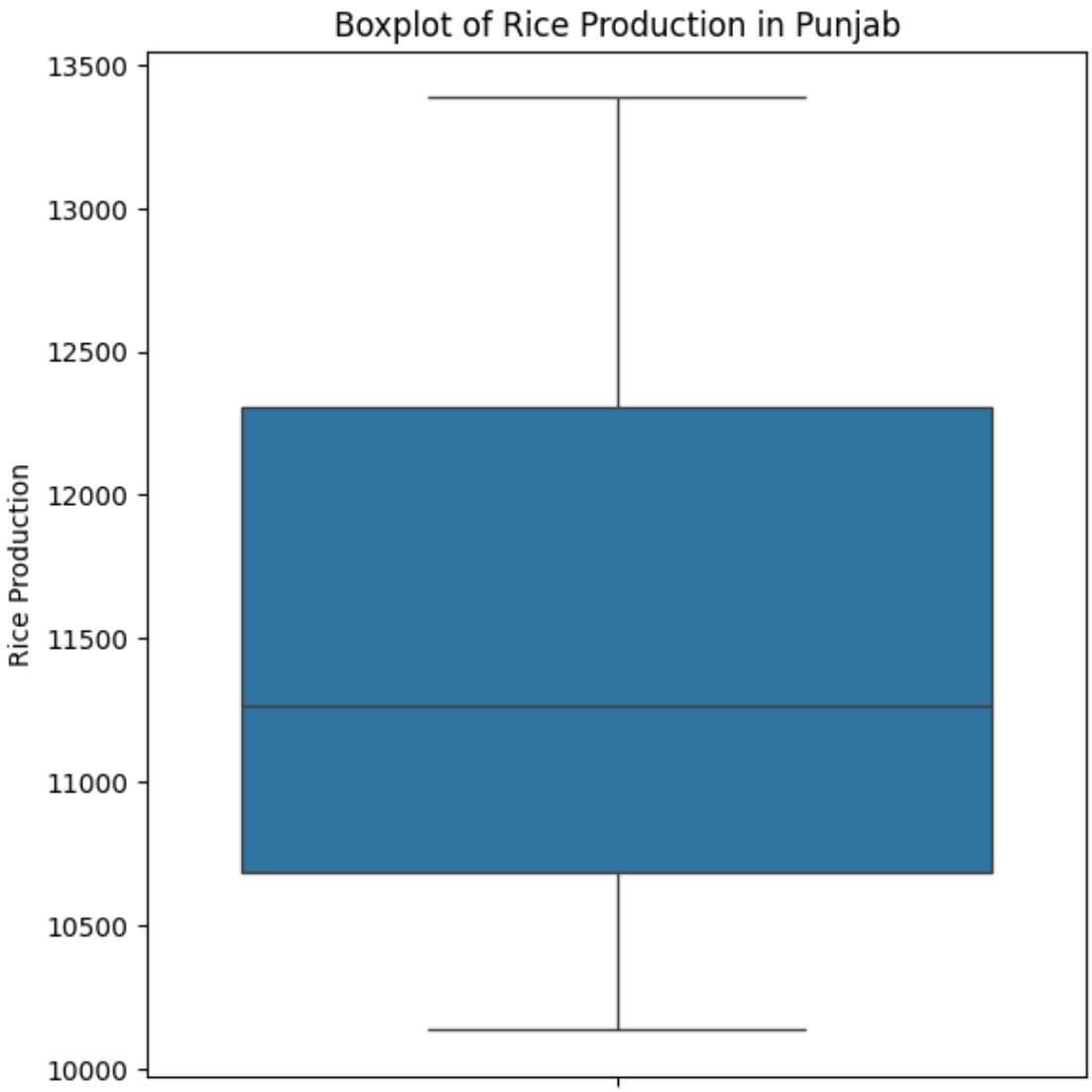

Boxplot of Rice Production in Meghalaya
Rice Production
300
280
260
240
220
200
180
160

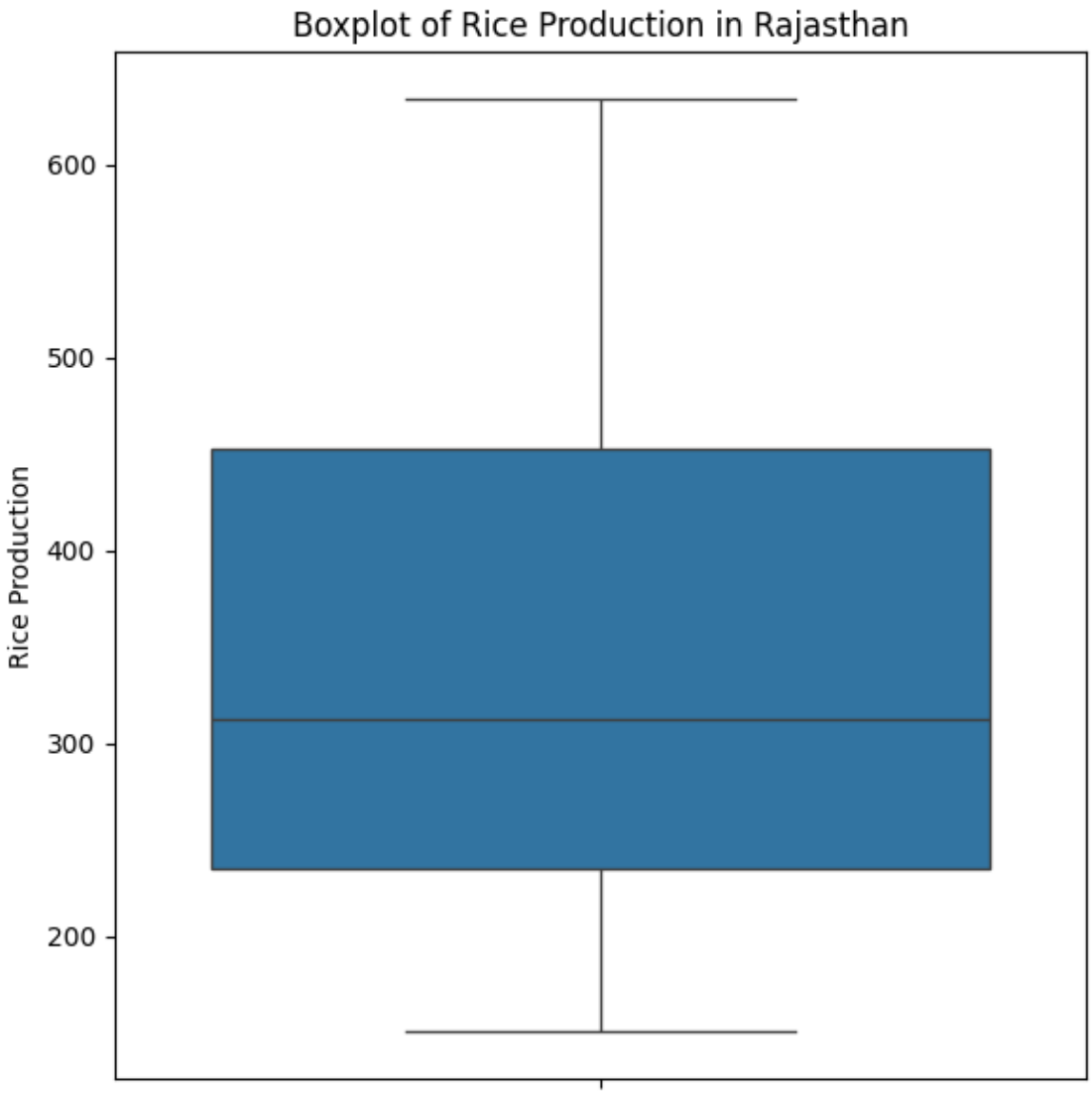

Boxplot of Rice Production in Mizoram
Rice Production
100
80
60
40
20

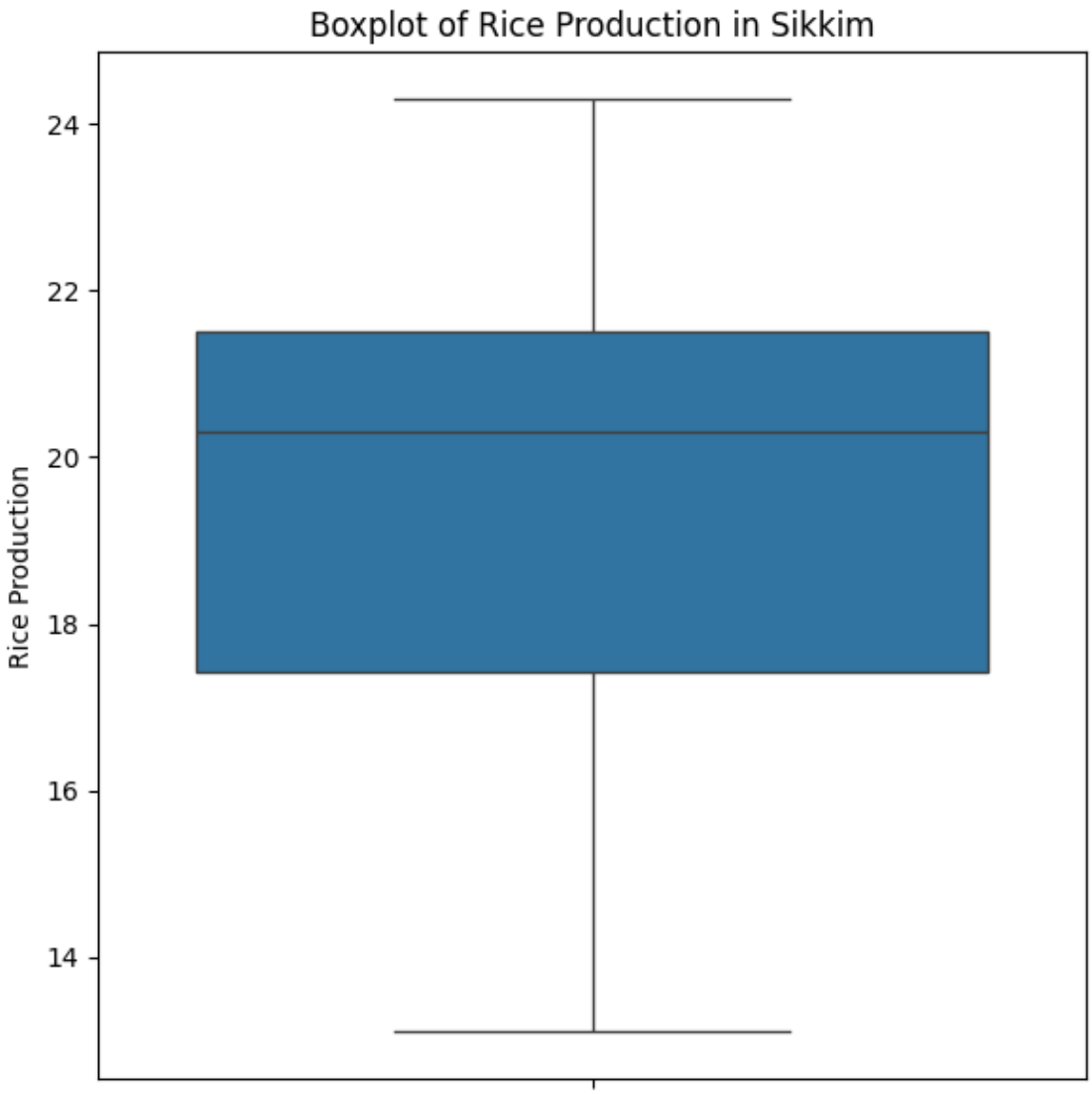

Boxplot of Rice Production in Nagaland
Rice Production
450
400
350
300
250
200
150

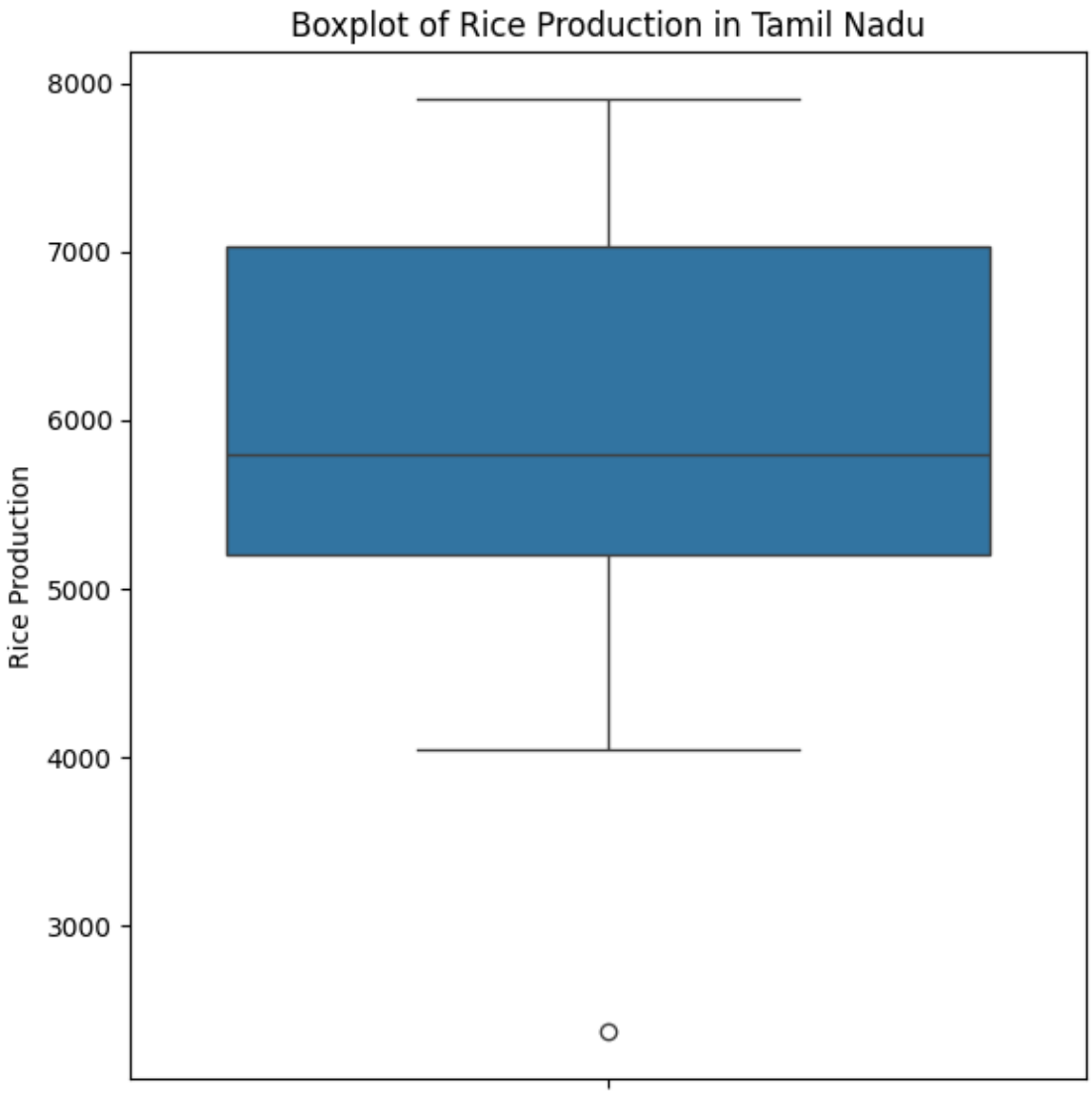

Boxplot of Rice Production in Odisha
Rice Production
9500
9000
8500
8000
7500
7000
6500
6000

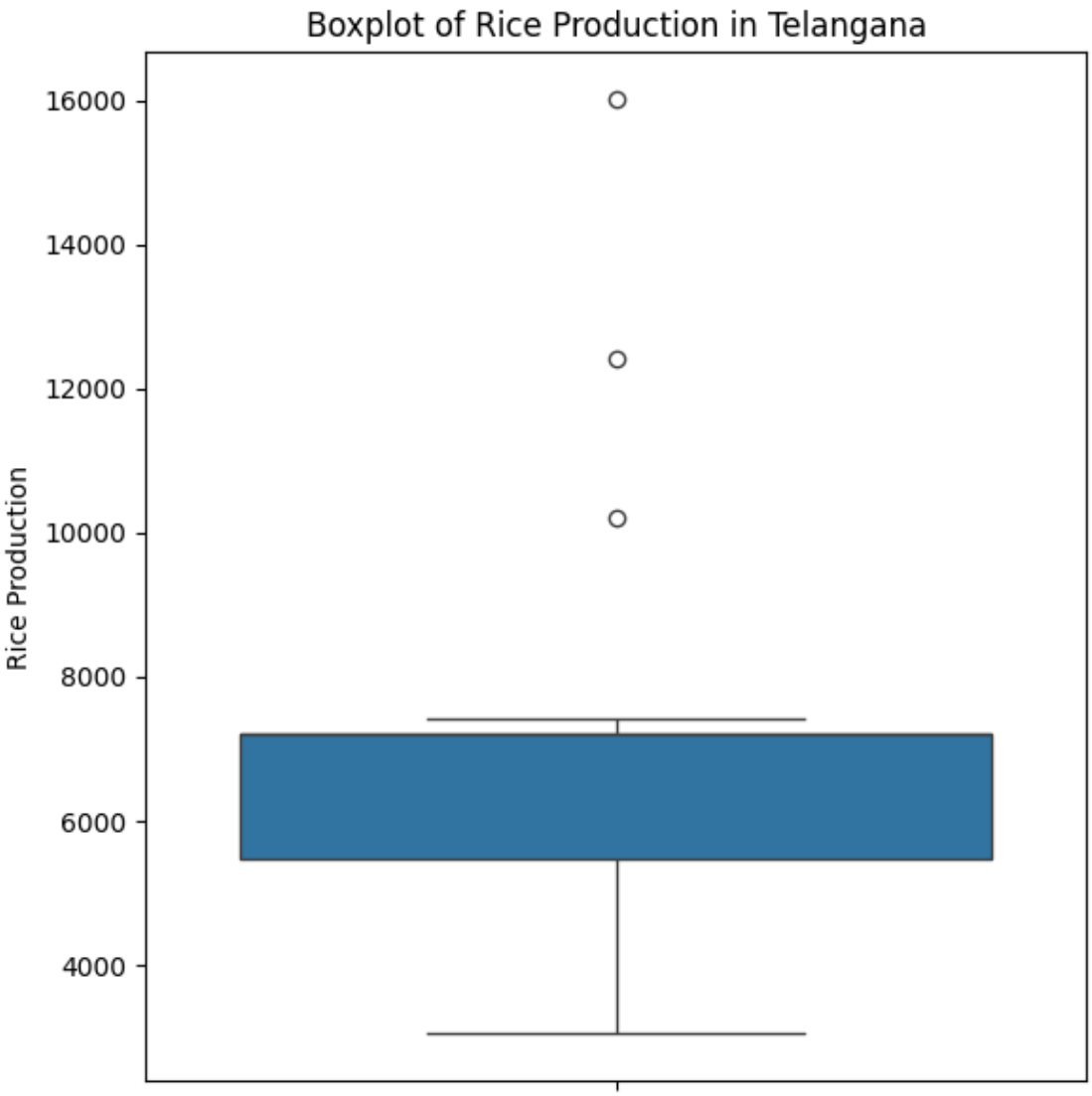

Boxplot of Rice Production in Puducherry
Rice Production
65
60
55
50
45

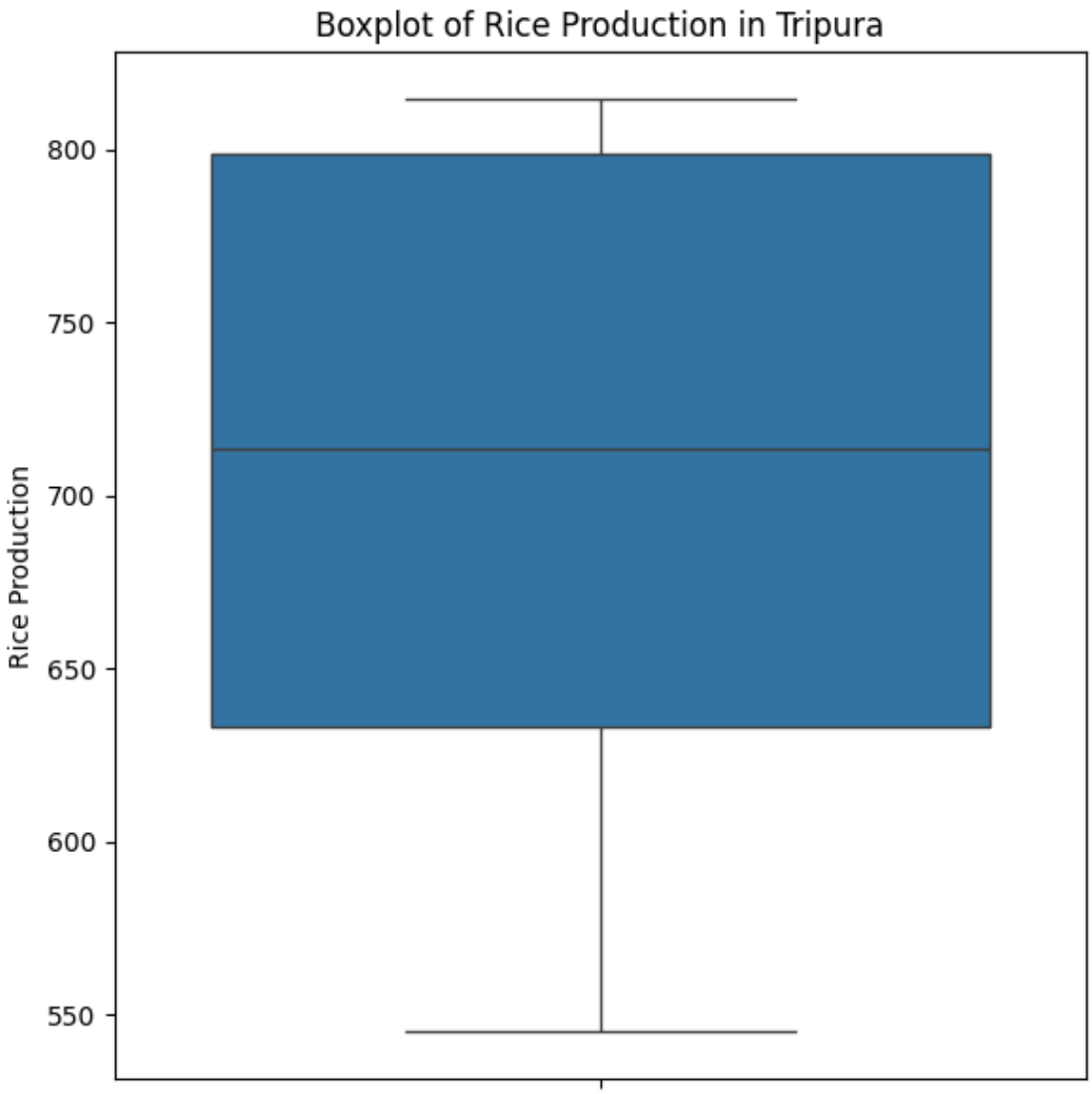

Boxplot of Rice Production in Punjab
Rice Production
13500
13000
12500
12000
11500
11000
10500
10000

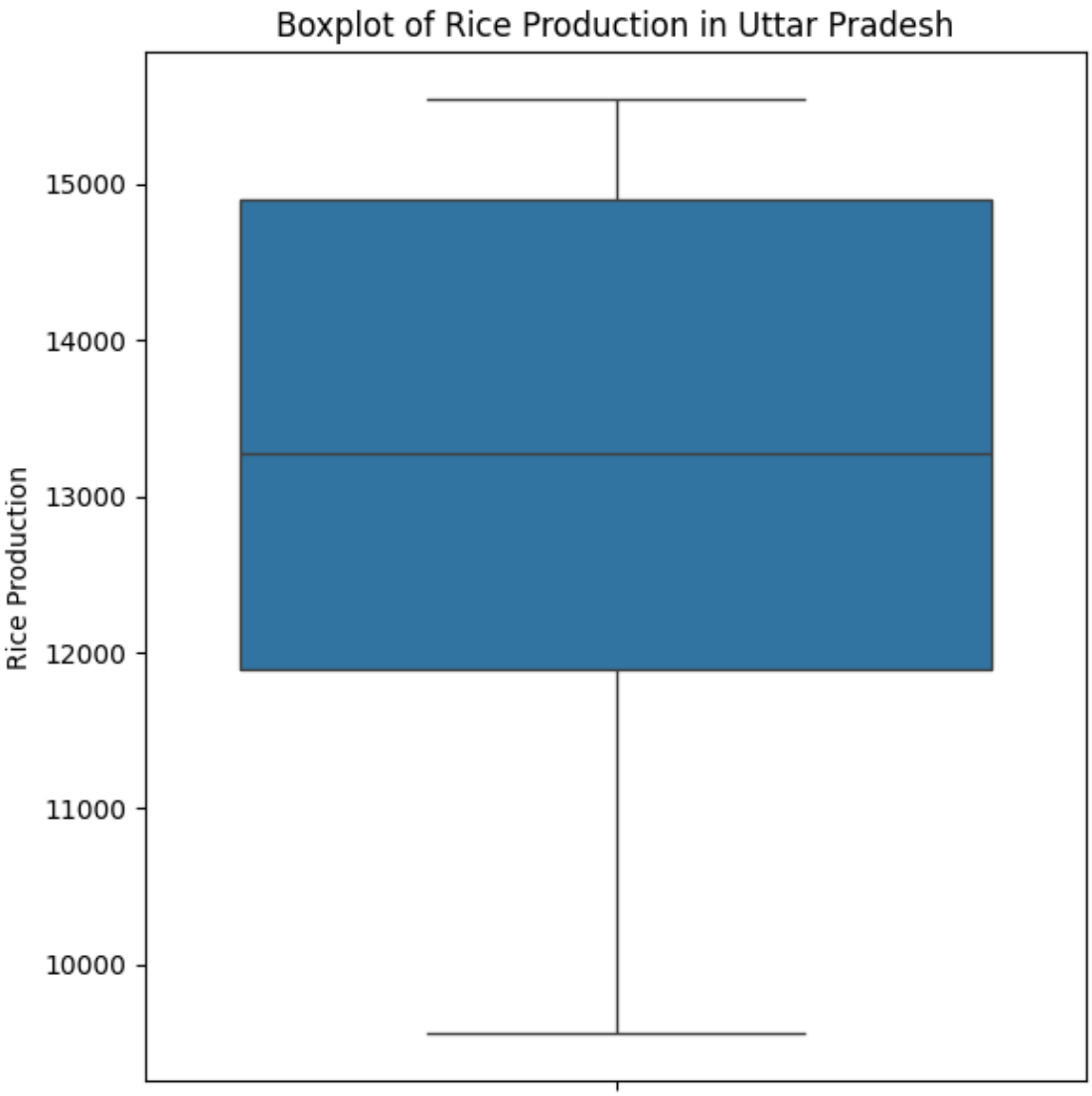

Boxplot of Rice Production in Rajasthan
Rice Production
600
500
400
300
200

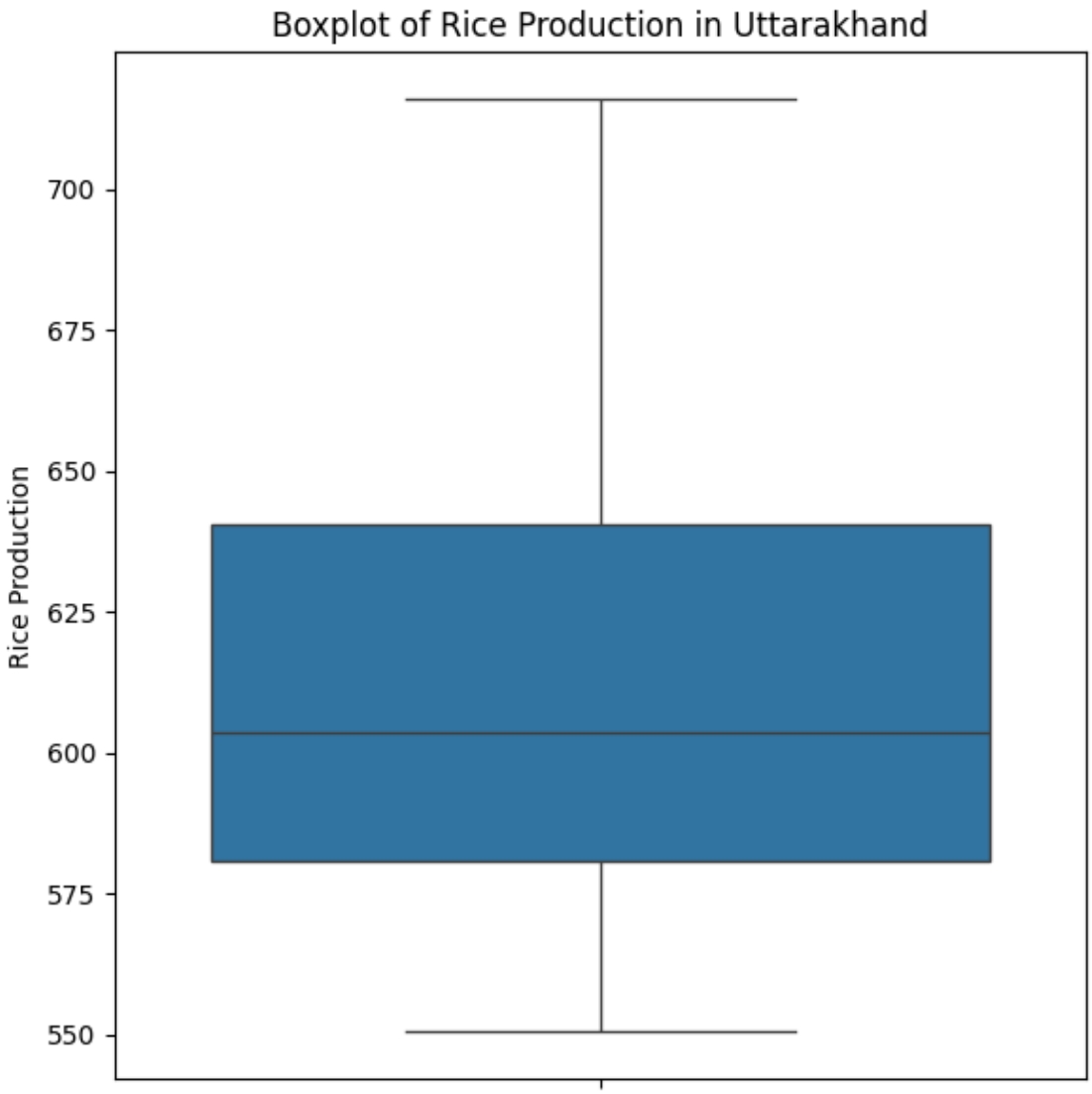

Boxplot of Rice Production in Sikkim
Rice Production
24
22
20
18
16
14

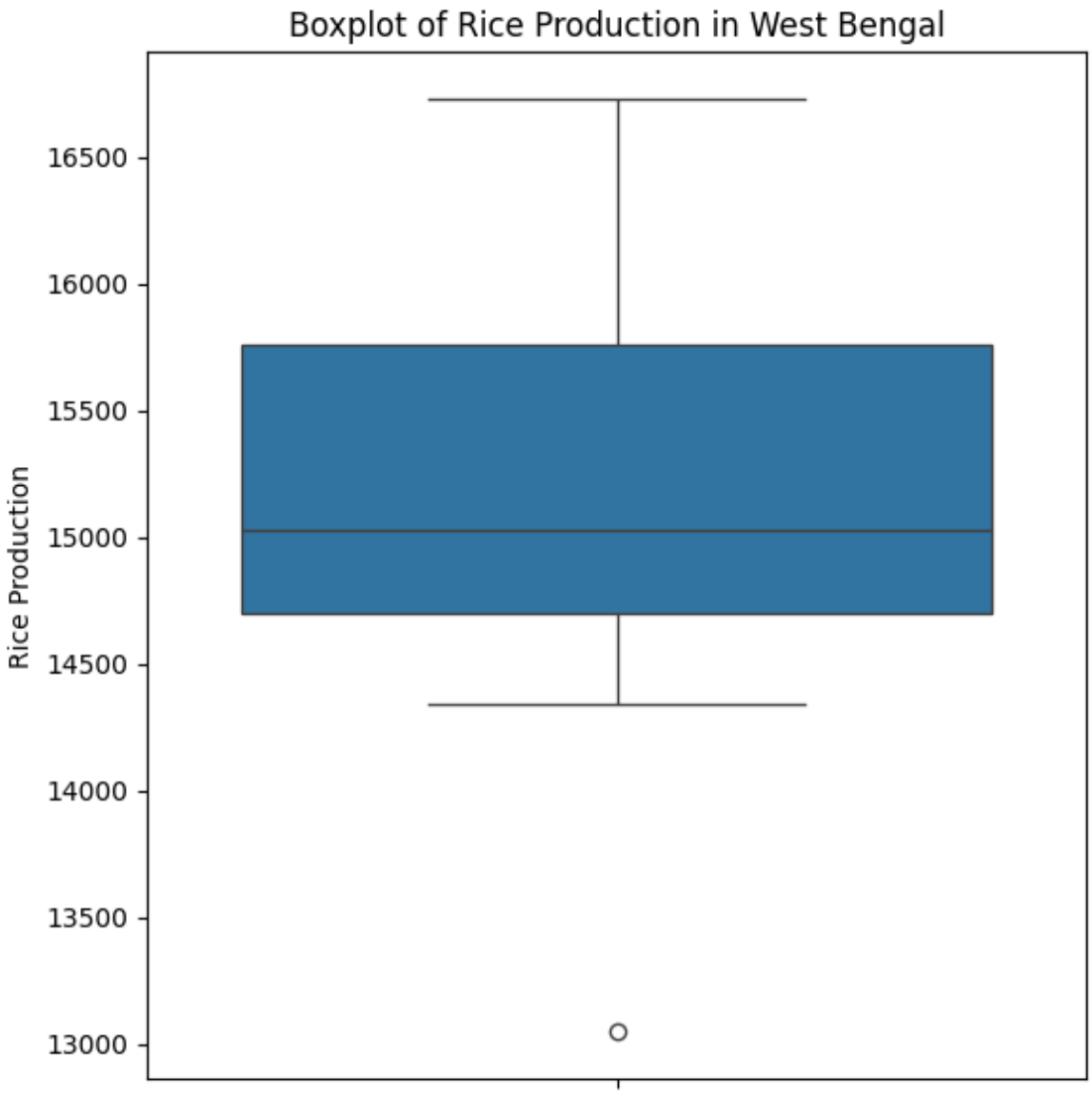

Boxplot of Rice Production in Tamil Nadu
Rice Production
8000
7000
6000
5000
4000
3000

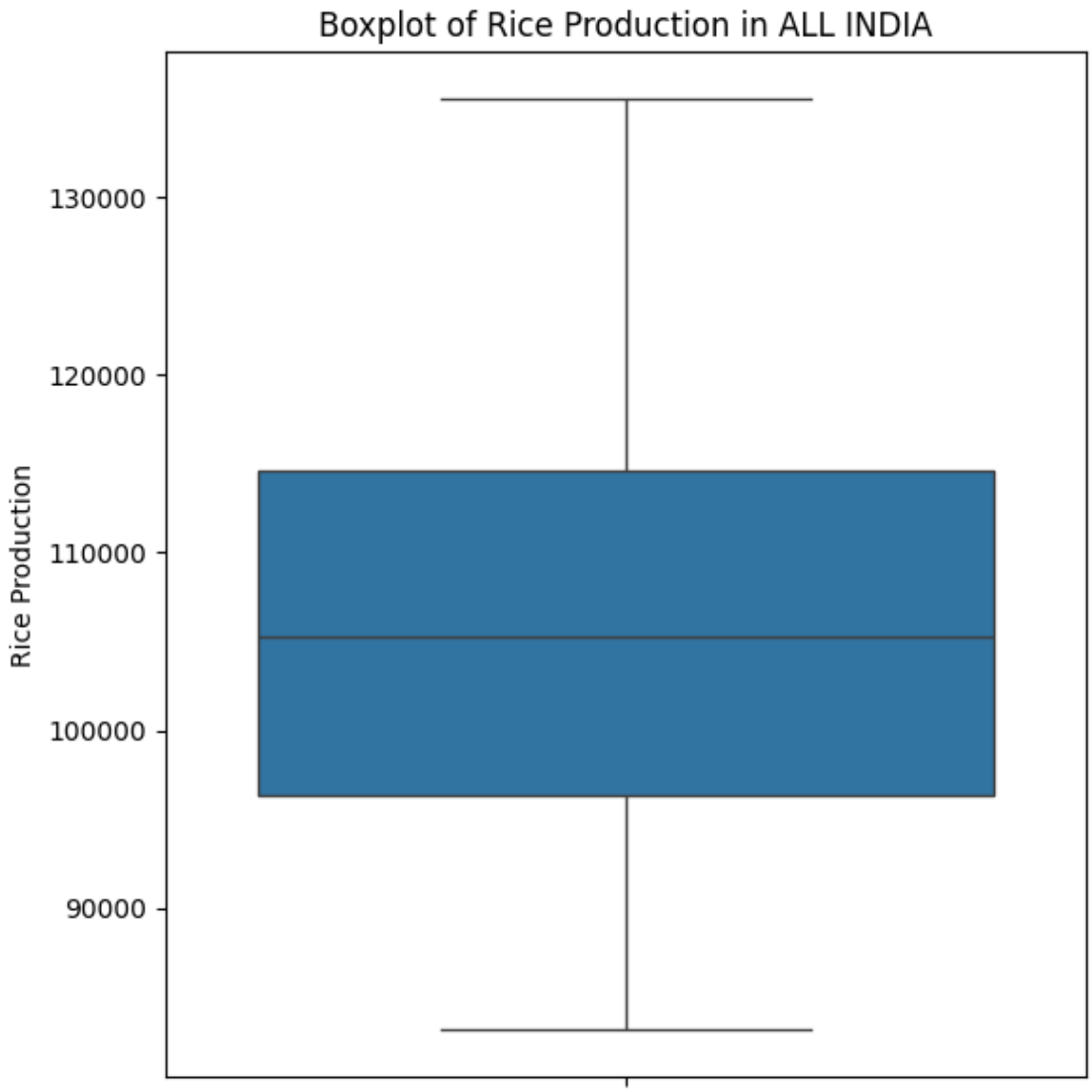

Boxplot of Rice Production in Telangana
Rice Production
16000
14000
12000
10000
8000
6000
4000


Boxplot of Rice Production in Tripura
Rice Production
800
750
700
650
600
550


Boxplot of Rice Production in Uttar Pradesh
Rice Production
15000
14000
13000
12000
11000
10000

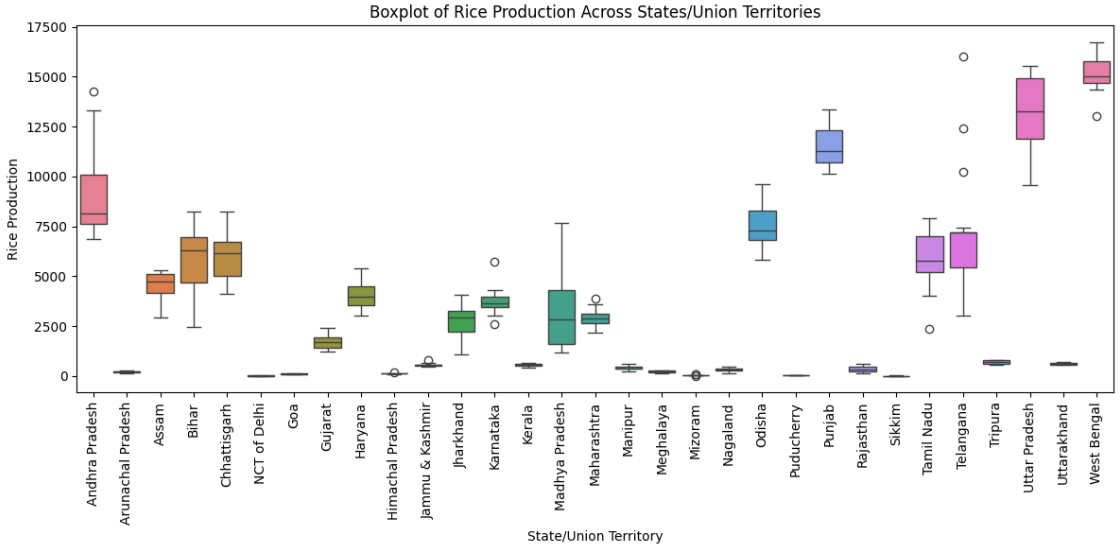

Boxplot of Rice Production in Uttarakhand
Rice Production
700
675
650
625
600
575
550


Boxplot of Rice Production in West Bengal
Rice Production
16500
16000
15500
15000
14500
14000
13500
13000



[ ]:
```
fig, ax = plt.subplots(figsize=(12, 6))

#
sns.boxplot(data=df_T_A, ax=ax)


ax.set_xlabel('State/Union Territory')
ax.set_ylabel('Rice Production')
ax.set_title('Boxplot of Rice Production Across States/Union Territories')


plt.xticks(rotation=90)
```

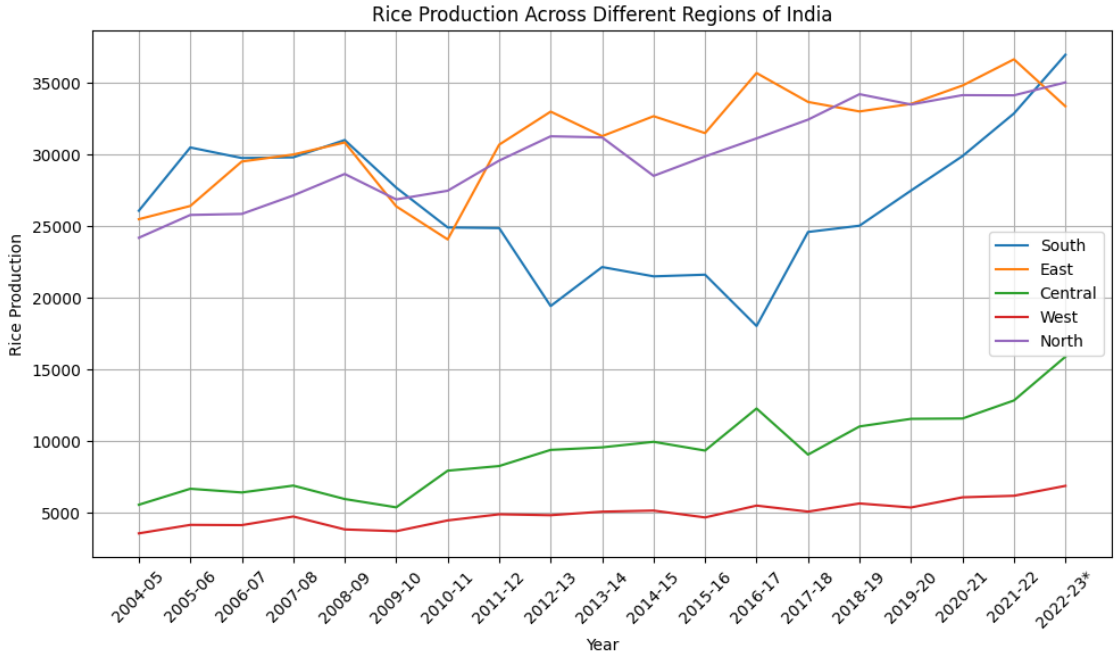
```
plt.tight_layout()
plt.show()
```



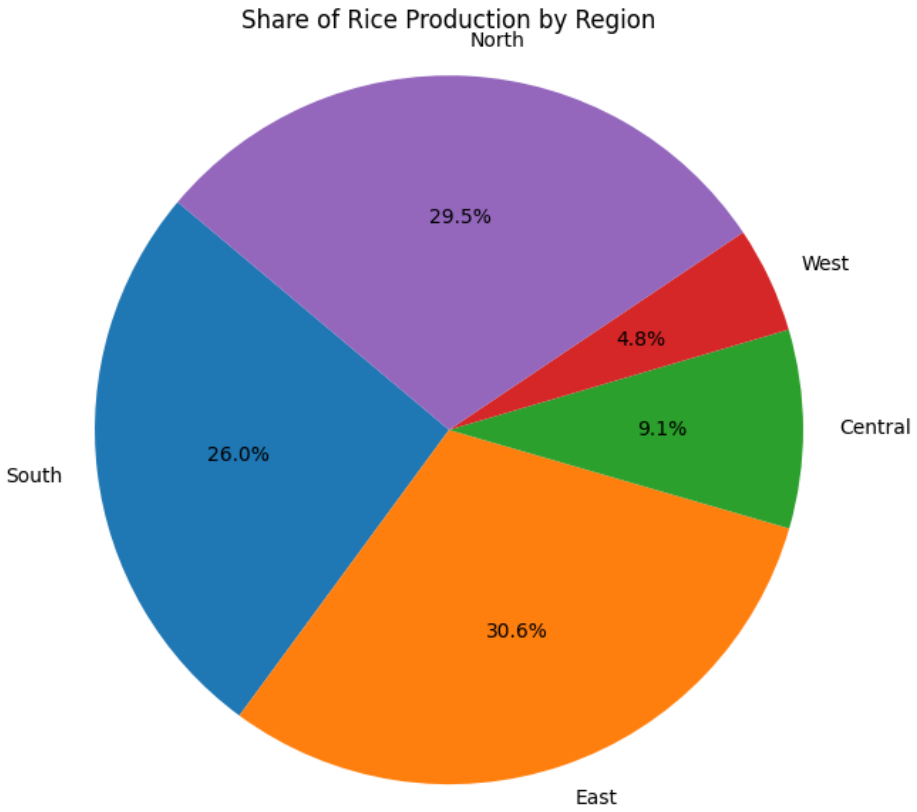
3 Regional Comparisons

[ ]:
```
north_states = ['Delhi', 'Haryana', 'Himachal Pradesh', 'Jammu & Kashmir',␣
 ↪'Punjab', 'Uttarakhand', 'Uttar Pradesh']
south_states = ['Andhra Pradesh', 'Karnataka', 'Kerala', 'Tamil Nadu',␣
 ↪'Telangana']
east_states = ['Bihar', 'Jharkhand', 'Odisha', 'West Bengal']
west_states = ['Gujarat', 'Maharashtra', 'Rajasthan']
central_states = ['Chhattisgarh', 'Madhya Pradesh']


state_to_region = {}
for state in df_T.columns:
    if state in north_states:
        state_to_region[state] = 'North'
    elif state in south_states:
        state_to_region[state] = 'South'
    elif state in east_states:
        state_to_region[state] = 'East'
    elif state in west_states:
        state_to_region[state] = 'West'
    elif state in central_states:
        state_to_region[state] = 'Central'
```

```
region_production = {}
for state, region in state_to_region.items():
    if region not in region_production:
        region_production[region] = df_T[state]
    else:
        region_production[region] += df_T[state]


plt.figure(figsize=(10, 6))
for region, production in region_production.items():
    plt.plot(df_T.index, production, label=region)


plt.xlabel('Year')
plt.ylabel('Rice Production')
plt.title('Rice Production Across Different Regions of India')
plt.legend()
plt.xticks(rotation=45)
plt.grid(True)

plt.tight_layout()
plt.show()
```

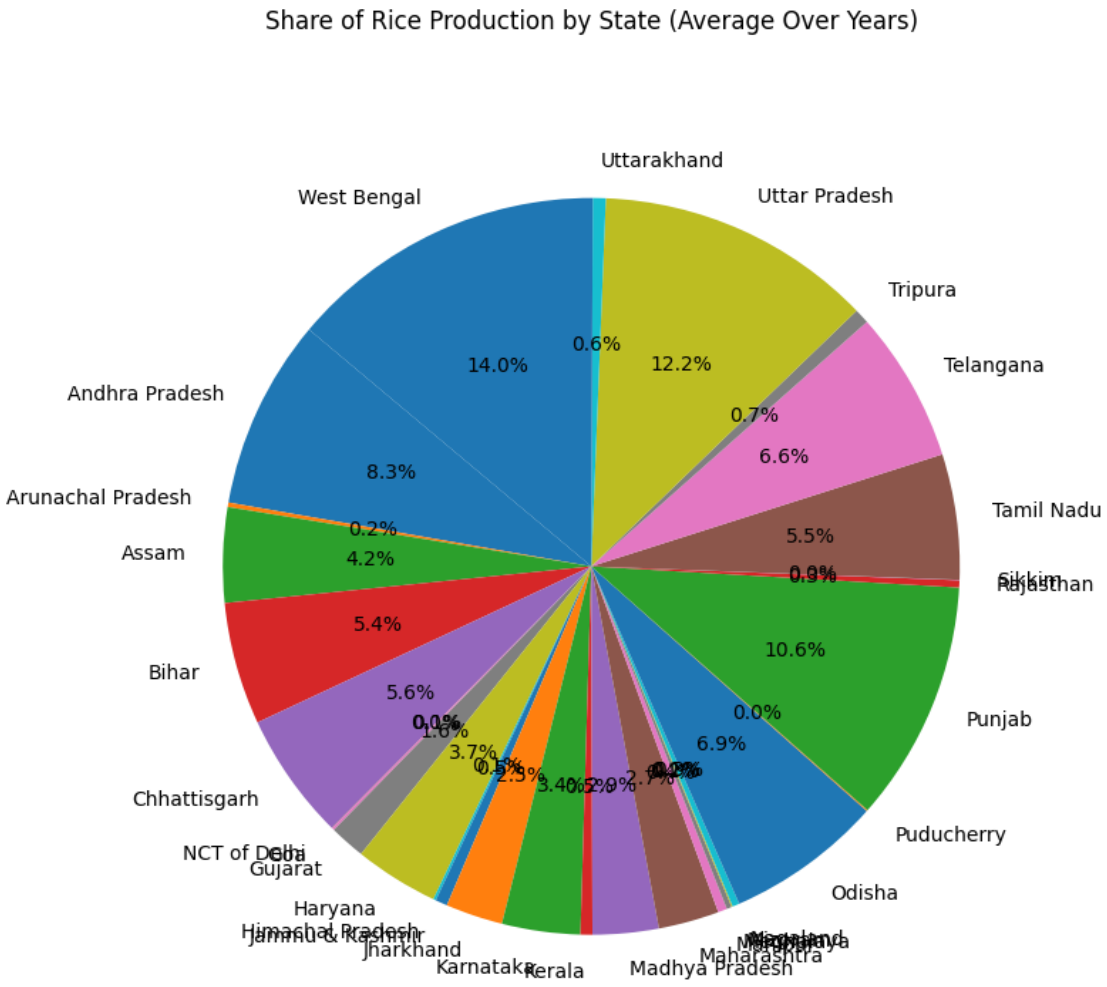


[ ]:
```
total_production = {region: production.sum() for region, production in␣
↪region_production.items()}


plt.figure(figsize=(8, 6))
plt.pie(total_production.values(), labels=total_production.keys(), autopct='%1.
↪1f%%', startangle=140)
plt.title('Share of Rice Production by Region')
plt.axis('equal')

plt.tight_layout()
plt.show()
```



[ ]:
```
average_production = df_T_A.mean()


plt.figure(figsize=(8, 8))
plt.pie(average_production, labels=average_production.index, autopct='%1.1f%%',␣
↪startangle=140)
plt.title('Share of Rice Production by State (Average Over Years)')
```

```
plt.axis('equal')
plt.tight_layout()
plt.show()
```



[ ]:
```
average_production = df_T_A.mean()


highest_state = average_production.idxmax()
lowest_state = average_production.idxmin()


highest_production = average_production[highest_state]
lowest_production = average_production[lowest_state]
```

```
print(f"Highest Producing State: {highest_state}, Production:␣
 ↪{highest_production:.2f} (Average Over Years)")
print(f"Lowest Producing State: {lowest_state}, Production: {lowest_production:.
 ↪2f} (Average Over Years)")
```

```
Highest Producing State: West Bengal, Production: 15168.38 (Average Over Years)
Lowest Producing State: Sikkim, Production: 19.61 (Average Over Years)
```

4 Seasonality

[ ]:
```
rainfall_data = {
    'Year': ['2004-05', '2005-06', '2006-07', '2007-08', '2008-09', '2009-10',␣
 ↪'2010-11', '2011-12', '2012-13', '2013-14', '2014-15', '2015-16', '2016-17',␣
 ↪'2017-18', '2018-19', '2019-20', '2021-22', '2022-23'],
    'Annual Rainfall': [790, 875.2, 926.8, 970.9, 902.8, 714.2, 910.7, 915.4,␣
 ↪841.7, 945.8, 784.2, 765.4, 863.7, 843.7, 802.4, 969.4, 863.8, 868.5]
}
rainfall_df = pd.DataFrame(rainfall_data)

rainfall_df.set_index('Year', inplace=True)

df_T_with_rainfall = pd.merge(df_T, rainfall_df, left_index=True,␣
 ↪right_index=True, how='left')


print(df_T_with_rainfall)
```

```
         Andhra Pradesh  Arunachal Pradesh   Assam     Bihar  Chhattisgarh  \
2004-05     26088.746154         135.000000  3470.7  25500.00        5552.3
2005-06     30509.346154         146.200000  3552.5  26423.30        6667.9
2006-07     29771.046154         146.200000  2916.0  29527.70        6409.8
2007-08     29821.146154         158.100000  3319.0  30014.70        6888.5
2008-09     31027.446154         163.900000  4008.5  30860.50        5951.5
2009-10     27703.946154         215.800000  4335.9  26395.90        5371.0
2010-11     24921.100000         234.000000  4736.6  24085.70        7931.1
2011-12     24877.700000         255.000000  4516.3  30706.00        8255.7
2012-13     19432.200000         263.000000  5128.5  33013.40        9383.8
2013-14     22156.300000         276.200000  4927.1  31300.50        9561.2
2014-15     21505.600000         285.000000  5222.7  32694.00        9947.4
2015-16     21623.100000         204.000000  5125.1  31513.70        9336.1
2016-17     18037.100000         220.000000  4727.4  35709.50       12275.2
2017-18     24605.700000         233.300000  5283.7  33689.44        9054.7
2018-19     25044.900000         240.000000  5220.6  33025.30       11021.6
2019-20     27497.900000         244.700000  4984.6  33552.60       11553.0
```

2020-21	29906.800000	247.100000	5214.8	34834.60	11575.0
2021-22	32885.200000	252.400000	4382.1	36667.00	12836.6
2022-23*	36989.500000	217.772222	4979.8	33383.10	15895.5

	NCT of Delhi	Goa	Gujarat	Haryana	Himachal Pradesh	\
2004-05	14.300000	145.200000	3552.60	24201.800000	122.00	
2005-06	24.000000	147.300000	4146.00	25795.600000	112.10	
2006-07	31.100000	130.300000	4128.80	25866.500000	123.50	
2007-08	31.400000	121.600000	4729.60	27157.800000	121.50	
2008-09	31.400000	123.300000	3828.10	28658.400000	118.30	
2009-10	19.300000	100.600000	3703.30	26879.400000	105.90	
2010-11	19.600000	115.000000	4458.10	27488.000000	128.90	
2011-12	19.800000	121.800000	4884.40	29593.300000	131.60	
2012-13	19.700000	122.800000	4820.50	31289.200000	125.30	
2013-14	29.600000	126.500000	5068.60	31211.300000	120.80	
2014-15	25.900000	120.500000	5143.60	28527.000000	125.20	
2015-16	17.300000	115.100000	4664.80	29884.400000	129.90	
2016-17	17.300000	113.200000	5492.20	31142.000000	146.60	
2017-18	16.800000	103.000000	5072.57	32453.800000	114.79	
2018-19	16.800000	98.800000	5641.00	34231.300000	114.90	
2019-20	16.800000	90.400000	5361.20	33510.600000	143.80	
2020-21	19.800000	87.300000	6071.40	34165.500000	140.50	
2021-22	19.000000	90.400000	6177.80	34151.500000	167.50	
2022-23*	21.661111	115.172222	6871.70	35054.263333	119.20	

	…	Rajasthan	Sikkim	Tamil Nadu	Telangana	Tripura	\
2004-05	…	150.40	21.600000	5062.2	7211.446154	545.100000	
2005-06	…	153.00	21.500000	5220.0	7211.446154	552.900000	
2006-07	…	169.80	21.500000	6610.6	7211.446154	620.500000	
2007-08	…	259.60	22.900000	5040.2	7211.446154	624.600000	
2008-09	…	241.10	21.700000	5182.7	7211.446154	627.100000	
2009-10	…	228.30	24.300000	5665.2	7211.446154	640.000000	
2010-11	…	265.50	21.000000	5792.4	6535.600000	702.500000	
2011-12	…	253.40	20.900000	7458.7	5148.800000	718.300000	
2012-13	…	222.50	21.300000	4049.9	4647.600000	713.200000	
2013-14	…	312.60	20.300000	5349.8	5755.000000	711.800000	
2014-15	…	366.70	20.100000	5727.8	4440.800000	747.000000	
2015-16	…	369.80	13.100000	7517.1	3047.000000	794.800000	
2016-17	…	452.70	19.700000	2369.4	5173.400000	814.600000	
2017-18	…	450.87	17.630000	6638.9	6262.200000	812.100000	
2018-19	…	453.20	17.200000	6130.9	6670.000000	793.200000	
2019-20	…	480.50	16.100000	7171.1	7427.800000	810.200000	
2020-21	…	634.00	16.200000	6881.2	10217.100000	803.100000	
2021-22	…	478.60	16.000000	7906.6	12409.600000	811.000000	
2022-23*	…	577.40	19.612778	7850.6	16013.900000	713.444444	

	Uttar Pradesh	Uttarakhand	West Bengal	ALL INDIA	Annual Rainfall
2004-05	9555.6	572.0	14884.8	83131.7	790.0

```
2005-06         11133.7        590.0       14510.8     91793.4            875.2
2006-07         11124.0        556.0       14745.9     93355.3            926.8
2007-08         11780.0        593.0       14719.5     96692.9            970.9
2008-09         13097.0        582.0       15037.3     99182.5            902.8
2009-10         10807.1        608.0       14340.7     89092.9            714.2
2010-11         11992.0        550.4       13045.9     95979.8            910.7
2011-12         14022.0        594.0       14605.8    105310.9            915.4
2012-13         14416.0        579.8       15023.7    105231.6            841.7
2013-14         14636.0        578.6       15370.7    106645.5            945.8
2014-15         12167.9        603.7       14677.2    104798.5            784.2
2015-16         12501.0        639.1       15953.9    104408.2            765.4
2016-17         13754.0        630.0       15302.5    109698.4            863.7
2017-18         13274.0        646.7       14967.0    112757.6            843.7
2018-19         15545.3        617.6       16242.2    116477.8            802.4
2019-20         15517.8        658.4       15881.4    118870.3            969.4
2020-21         15520.0        714.9       16524.4    124368.3              NaN
2021-22         15271.5        716.1       16728.7    129471.4            863.8
2022-23*        15171.3        641.7       15636.9    135542.0              NaN

[19 rows x 33 columns]
```

[ ]:
```
df_T_with_rainfall['Annual Rainfall'] = df_T_with_rainfall['Annual Rainfall'].
↪fillna(df_T_with_rainfall['Annual Rainfall'].mean())
```

[ ]:
```
df_T_with_rainfall
```

[ ]:
```
           Andhra Pradesh  Arunachal Pradesh   Assam     Bihar  Chhattisgarh  \
2004-05     26088.746154         135.000000  3470.7  25500.00        5552.3
2005-06     30509.346154         146.200000  3552.5  26423.30        6667.9
2006-07     29771.046154         146.200000  2916.0  29527.70        6409.8
2007-08     29821.146154         158.100000  3319.0  30014.70        6888.5
2008-09     31027.446154         163.900000  4008.5  30860.50        5951.5
2009-10     27703.946154         215.800000  4335.9  26395.90        5371.0
2010-11     24921.100000         234.000000  4736.6  24085.70        7931.1
2011-12     24877.700000         255.000000  4516.3  30706.00        8255.7
2012-13     19432.200000         263.000000  5128.5  33013.40        9383.8
2013-14     22156.300000         276.200000  4927.1  31300.50        9561.2
2014-15     21505.600000         285.000000  5222.7  32694.00        9947.4
2015-16     21623.100000         204.000000  5125.1  31513.70        9336.1
2016-17     18037.100000         220.000000  4727.4  35709.50       12275.2
2017-18     24605.700000         233.300000  5283.7  33689.44        9054.7
2018-19     25044.900000         240.000000  5220.6  33025.30       11021.6
2019-20     27497.900000         244.700000  4984.6  33552.60       11553.0
2020-21     29906.800000         247.100000  5214.8  34834.60       11575.0
2021-22     32885.200000         252.400000  4382.1  36667.00       12836.6
2022-23*    36989.500000         217.772222  4979.8  33383.10       15895.5
```

	NCT of Delhi	Goa	Gujarat	Haryana	Himachal Pradesh	\
2004-05	14.300000	145.200000	3552.60	24201.800000	122.00	
2005-06	24.000000	147.300000	4146.00	25795.600000	112.10	
2006-07	31.100000	130.300000	4128.80	25866.500000	123.50	
2007-08	31.400000	121.600000	4729.60	27157.800000	121.50	
2008-09	31.400000	123.300000	3828.10	28658.400000	118.30	
2009-10	19.300000	100.600000	3703.30	26879.400000	105.90	
2010-11	19.600000	115.000000	4458.10	27488.000000	128.90	
2011-12	19.800000	121.800000	4884.40	29593.300000	131.60	
2012-13	19.700000	122.800000	4820.50	31289.200000	125.30	
2013-14	29.600000	126.500000	5068.60	31211.300000	120.80	
2014-15	25.900000	120.500000	5143.60	28527.000000	125.20	
2015-16	17.300000	115.100000	4664.80	29884.400000	129.90	
2016-17	17.300000	113.200000	5492.20	31142.000000	146.60	
2017-18	16.800000	103.000000	5072.57	32453.800000	114.79	
2018-19	16.800000	98.800000	5641.00	34231.300000	114.90	
2019-20	16.800000	90.400000	5361.20	33510.600000	143.80	
2020-21	19.800000	87.300000	6071.40	34165.500000	140.50	
2021-22	19.000000	90.400000	6177.80	34151.500000	167.50	
2022-23*	21.661111	115.172222	6871.70	35054.263333	119.20	

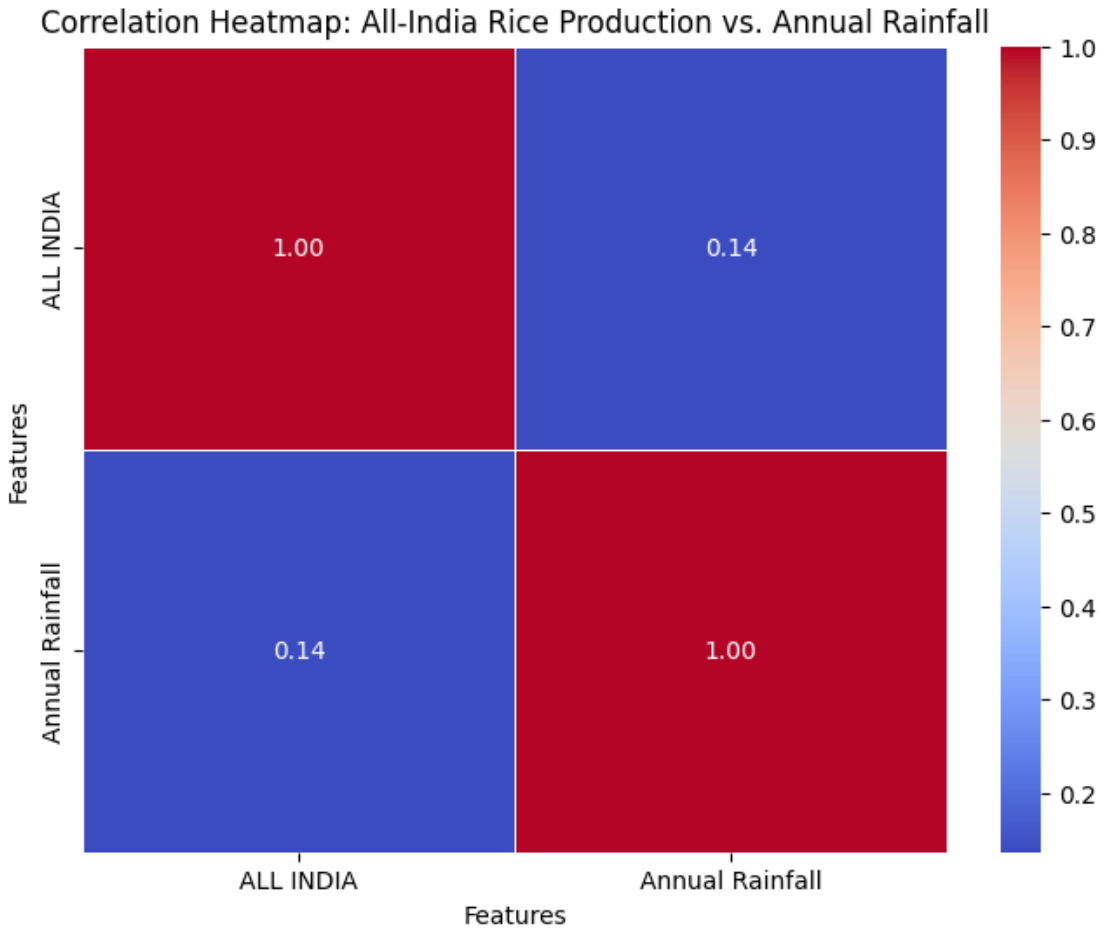
	…	Rajasthan	Sikkim	Tamil Nadu	Telangana	Tripura	\
2004-05	…	150.40	21.600000	5062.2	7211.446154	545.100000	
2005-06	…	153.00	21.500000	5220.0	7211.446154	552.900000	
2006-07	…	169.80	21.500000	6610.6	7211.446154	620.500000	
2007-08	…	259.60	22.900000	5040.2	7211.446154	624.600000	
2008-09	…	241.10	21.700000	5182.7	7211.446154	627.100000	
2009-10	…	228.30	24.300000	5665.2	7211.446154	640.000000	
2010-11	…	265.50	21.000000	5792.4	6535.600000	702.500000	
2011-12	…	253.40	20.900000	7458.7	5148.800000	718.300000	
2012-13	…	222.50	21.300000	4049.9	4647.600000	713.200000	
2013-14	…	312.60	20.300000	5349.8	5755.000000	711.800000	
2014-15	…	366.70	20.100000	5727.8	4440.800000	747.000000	
2015-16	…	369.80	13.100000	7517.1	3047.000000	794.800000	
2016-17	…	452.70	19.700000	2369.4	5173.400000	814.600000	
2017-18	…	450.87	17.630000	6638.9	6262.200000	812.100000	
2018-19	…	453.20	17.200000	6130.9	6670.000000	793.200000	
2019-20	…	480.50	16.100000	7171.1	7427.800000	810.200000	
2020-21	…	634.00	16.200000	6881.2	10217.100000	803.100000	
2021-22	…	478.60	16.000000	7906.6	12409.600000	811.000000	
2022-23*	…	577.40	19.612778	7850.6	16013.900000	713.444444	

	Uttar Pradesh	Uttarakhand	West Bengal	ALL INDIA	Annual Rainfall
2004-05	9555.6	572.0	14884.8	83131.7	790.000000
2005-06	11133.7	590.0	14510.8	91793.4	875.200000
2006-07	11124.0	556.0	14745.9	93355.3	926.800000
2007-08	11780.0	593.0	14719.5	96692.9	970.900000

2008-09	13097.0	582.0	15037.3	99182.5	902.800000
2009-10	10807.1	608.0	14340.7	89092.9	714.200000
2010-11	11992.0	550.4	13045.9	95979.8	910.700000
2011-12	14022.0	594.0	14605.8	105310.9	915.400000
2012-13	14416.0	579.8	15023.7	105231.6	841.700000
2013-14	14636.0	578.6	15370.7	106645.5	945.800000
2014-15	12167.9	603.7	14677.2	104798.5	784.200000
2015-16	12501.0	639.1	15953.9	104408.2	765.400000
2016-17	13754.0	630.0	15302.5	109698.4	863.700000
2017-18	13274.0	646.7	14967.0	112757.6	843.700000
2018-19	15545.3	617.6	16242.2	116477.8	802.400000
2019-20	15517.8	658.4	15881.4	118870.3	969.400000
2020-21	15520.0	714.9	16524.4	124368.3	863.888235
2021-22	15271.5	716.1	16728.7	129471.4	863.800000
2022-23*	15171.3	641.7	15636.9	135542.0	863.888235

[19 rows x 33 columns]

[ ]:
```
correlation_matrix = df_T_with_rainfall[['ALL INDIA', 'Annual Rainfall']].corr()

plt.figure(figsize=(8, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f",␣
 ↪linewidths=0.5)
plt.title('Correlation Heatmap: All-India Rice Production vs. Annual Rainfall')
plt.xlabel('Features')
plt.ylabel('Features')
plt.show()
```



[ ]:
```
# Imports
import numpy as np
import pandas as pd

import sklearn
import seaborn as sns
import matplotlib.pyplot as plt
```

[ ]:
```
# Load data

rainfall_data = pd.read_csv('/content/Monthly_Rainfall_From_1901_to_2017.csv')

yield_data = pd.read_excel('/content/State_wise_rice_production_in_India.xlsx')
```

[ ]:
```
rainfall_data
```

[ ]:

```
      Unnamed: 0                States/UTs  YEAR   JAN    FEB   MAR    APR  \
0              0  Andaman & Nicobar Islands  1901  49.2   87.1  29.2    2.3
1              1  Andaman & Nicobar Islands  1902   0.0  159.8  12.2    0.0
2              2  Andaman & Nicobar Islands  1903  12.7  144.0   0.0    1.0
3              3  Andaman & Nicobar Islands  1904   9.4   14.7   0.0  202.4
4              4  Andaman & Nicobar Islands  1905   1.3    0.0   3.3   26.9
…            …                          …   …    …     …    …     …
4182        4182                Lakshadweep  2012  19.2    0.1   1.6   76.8
4183        4183                Lakshadweep  2013  26.2   34.4  37.5    5.3
4184        4184                Lakshadweep  2014  53.2   16.1   4.4   14.9
4185        4185                Lakshadweep  2015   2.2    0.5   3.7   87.1
4186        4186                Lakshadweep  2016  59.6   12.1   3.2    2.6

        MAY    JUN    JUL    AUG    SEP    OCT    NOV    DEC  ANNUAL
0     528.8  517.5  365.1  481.1  332.6  388.5  558.2   33.6  3373.2
1     446.1  537.1  228.9  753.7  666.2  197.2  359.0  160.5  3520.7
2     235.1  479.9  728.4  326.7  339.0  181.2  284.4  225.0  2957.4
3     304.5  495.1  502.0  160.1  820.4  222.2  308.7   40.1  3079.6
4     279.5  628.7  368.7  330.5  297.0  260.7   25.4  344.7  2566.7
…     …    …    …    …    …    …    …    …      …
4182   21.2  327.0  231.5  381.2  179.8  145.9   12.4    8.8  1405.5
4183   88.3  426.2  296.4  154.4  180.0   72.8   78.1   26.7  1426.3
4184   57.4  244.1  116.1  466.1  132.2  169.2   59.0   62.3  1395.0
4185  133.1  296.6  257.5  146.4  160.4  165.4  231.0  159.0  1642.9
4186   77.4  321.1  262.6   86.2   75.6   58.6   32.0   74.7  1065.7

[4187 rows x 16 columns]
```

[ ]:

```
rainfall_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4187 entries, 0 to 4186
Data columns (total 16 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Unnamed: 0  4187 non-null   int64
 1   States/UTs  4187 non-null   object
 2   YEAR        4187 non-null   int64
 3   JAN         4183 non-null   float64
 4   FEB         4184 non-null   float64
 5   MAR         4181 non-null   float64
 6   APR         4183 non-null   float64
 7   MAY         4184 non-null   float64
 8   JUN         4182 non-null   float64
 9   JUL         4180 non-null   float64
 10  AUG         4183 non-null   float64
 11  SEP         4181 non-null   float64
```

```
 12  OCT         4180 non-null   float64
 13  NOV         4176 non-null   float64
 14  DEC         4177 non-null   float64
 15  ANNUAL      4161 non-null   float64
dtypes: float64(13), int64(2), object(1)
memory usage: 523.5+ KB
```

[ ]:
```
rainfall_data.drop(['Unnamed: 0'],axis = 1,inplace=True)
```

[ ]:
```
rainfall_data.head()
```

[ ]:
```
                States/UTs  YEAR   JAN    FEB   MAR    APR    MAY    JUN  \
0  Andaman & Nicobar Islands  1901  49.2   87.1  29.2    2.3  528.8  517.5
1  Andaman & Nicobar Islands  1902   0.0  159.8  12.2    0.0  446.1  537.1
2  Andaman & Nicobar Islands  1903  12.7  144.0   0.0    1.0  235.1  479.9
3  Andaman & Nicobar Islands  1904   9.4   14.7   0.0  202.4  304.5  495.1
4  Andaman & Nicobar Islands  1905   1.3    0.0   3.3   26.9  279.5  628.7

     JUL    AUG    SEP    OCT    NOV    DEC  ANNUAL
0  365.1  481.1  332.6  388.5  558.2   33.6  3373.2
1  228.9  753.7  666.2  197.2  359.0  160.5  3520.7
2  728.4  326.7  339.0  181.2  284.4  225.0  2957.4
3  502.0  160.1  820.4  222.2  308.7   40.1  3079.6
4  368.7  330.5  297.0  260.7   25.4  344.7  2566.7
```

[ ]:
```
rainfall_data.describe()
```

[ ]:
```
              YEAR          JAN          FEB          MAR          APR  \
count  4187.000000  4183.000000  4184.000000  4181.000000  4183.000000
mean   1959.206831    18.937772    21.604374    27.395503    43.081712
std      33.709294    33.797148    35.734572    46.920068    68.156144
min    1901.000000     0.000000     0.000000     0.000000     0.000000
25%    1930.000000     0.600000     0.500000     1.000000     3.000000
50%    1959.000000     5.900000     6.500000     7.900000    15.500000
75%    1988.000000    22.000000    26.600000    31.300000    49.700000
max    2017.000000   583.700000   403.500000   605.600000   595.100000

               MAY          JUN          JUL          AUG          SEP  \
count  4184.000000  4182.000000  4180.000000  4183.000000  4181.000000
mean     85.646511   230.047704   347.063780   289.762156   197.311863
std     122.706290   234.276638   268.867991   188.444169   135.563551
min       0.000000     0.400000     0.000000     0.000000     0.100000
25%       8.600000    70.800000   175.825000   155.800000   100.400000
50%      36.850000   138.750000   285.050000   258.500000   173.600000
75%      97.725000   304.700000   418.525000   377.650000   266.200000
max    1168.600000  1609.900000  2362.800000  1664.600000  1222.000000
```

	OCT	NOV	DEC	ANNUAL
count	4180.000000	4176.000000	4177.000000	4161.000000
mean	95.314713	39.490685	18.934858	1409.370031
std	99.204613	68.365883	43.003479	902.693445
min	0.000000	0.000000	0.000000	62.300000
25%	14.600000	0.600000	0.100000	803.000000
50%	65.100000	9.400000	3.000000	1120.200000
75%	148.300000	45.300000	17.500000	1642.900000
max	948.300000	648.900000	617.500000	6331.100000

[ ]:
```
len(rainfall_data['YEAR'].unique())
```

[ ]: 117

[ ]:
```
rainfall_data.head(12)
```

[ ]:

	States/UTs	YEAR	JAN	FEB	MAR	APR	MAY	JUN	\
0	Andaman & Nicobar Islands	1901	49.2	87.1	29.2	2.3	528.8	517.5	
1	Andaman & Nicobar Islands	1902	0.0	159.8	12.2	0.0	446.1	537.1	
2	Andaman & Nicobar Islands	1903	12.7	144.0	0.0	1.0	235.1	479.9	
3	Andaman & Nicobar Islands	1904	9.4	14.7	0.0	202.4	304.5	495.1	
4	Andaman & Nicobar Islands	1905	1.3	0.0	3.3	26.9	279.5	628.7	
5	Andaman & Nicobar Islands	1906	36.6	0.0	0.0	0.0	556.1	733.3	
6	Andaman & Nicobar Islands	1907	110.7	0.0	113.3	21.6	616.3	305.2	
7	Andaman & Nicobar Islands	1908	20.9	85.1	0.0	29.0	562.0	693.6	
8	Andaman & Nicobar Islands	1910	26.6	22.7	206.3	89.3	224.5	472.7	
9	Andaman & Nicobar Islands	1911	0.0	8.4	0.0	122.5	327.3	649.0	
10	Andaman & Nicobar Islands	1912	583.7	0.8	0.0	21.9	140.7	549.8	
11	Andaman & Nicobar Islands	1913	84.8	0.5	1.3	2.5	190.7	530.0	

	JUL	AUG	SEP	OCT	NOV	DEC	ANNUAL
0	365.1	481.1	332.6	388.5	558.2	33.6	3373.2
1	228.9	753.7	666.2	197.2	359.0	160.5	3520.7
2	728.4	326.7	339.0	181.2	284.4	225.0	2957.4
3	502.0	160.1	820.4	222.2	308.7	40.1	3079.6
4	368.7	330.5	297.0	260.7	25.4	344.7	2566.7
5	247.7	320.5	164.3	267.8	128.9	79.2	2534.4
6	443.9	377.6	200.4	264.4	648.9	245.6	3347.9
7	481.4	699.9	428.8	170.7	208.1	196.9	3576.4
8	264.3	337.4	626.6	208.2	267.3	153.5	2899.4
9	253.0	187.1	464.5	333.8	94.5	247.1	2687.2
10	468.9	370.3	386.2	318.7	117.2	2.3	2960.5
11	280.8	205.8	580.1	288.8	133.0	67.5	2365.8

[ ]:
```
rainfall_df = rainfall_data.groupby(['YEAR'],as_index = False,axis = 0)
```

[ ]:
```
rainfall_df.head()
```

[ ]:
```
                    States/UTs  YEAR   JAN    FEB   MAR    APR    MAY    JUN  \
0    Andaman & Nicobar Islands  1901  49.2   87.1  29.2    2.3  528.8  517.5
1    Andaman & Nicobar Islands  1902   0.0  159.8  12.2    0.0  446.1  537.1
2    Andaman & Nicobar Islands  1903  12.7  144.0   0.0    1.0  235.1  479.9
3    Andaman & Nicobar Islands  1904   9.4   14.7   0.0  202.4  304.5  495.1
4    Andaman & Nicobar Islands  1905   1.3    0.0   3.3   26.9  279.5  628.7
..                         …   …   …    …   …    …    …    …
609       Gangetic West Bengal  1948  13.2   44.9  48.4   52.6  135.9  200.4
615       Gangetic West Bengal  1954  12.3    5.9   0.3   11.6   88.6  211.1
616       Gangetic West Bengal  1955   2.9    2.3  11.4   24.3   48.0  174.7
617       Gangetic West Bengal  1956   4.5   33.9  50.7   25.6  141.6  343.3
687                     Orissa  1909   5.7   11.9   4.8  148.4   53.1  288.2

       JUL    AUG    SEP    OCT    NOV    DEC  ANNUAL
0    365.1  481.1  332.6  388.5  558.2   33.6  3373.2
1    228.9  753.7  666.2  197.2  359.0  160.5  3520.7
2    728.4  326.7  339.0  181.2  284.4  225.0  2957.4
3    502.0  160.1  820.4  222.2  308.7   40.1  3079.6
4    368.7  330.5  297.0  260.7   25.4  344.7  2566.7
..     …    …    …    …    …    …    …
609  268.4  323.6  214.3  116.7  113.6    0.0  1532.1
615  201.7  215.1  247.1   56.8    1.5   28.0  1080.0
616  338.8  251.9  188.1  151.9   73.9    0.1  1268.2
617  253.9  295.4  391.1  157.6   12.7    3.9  1714.2
687  452.7  234.3  234.0   42.3    1.8   58.1  1535.5

[585 rows x 15 columns]
```

[ ]:
```
yield_data.head()
```

[ ]:
```
  State/Union Territory 2004-05 2005-06 2006-07 2007-08 2008-09 2009-10  \
0        Andhra Pradesh    9601   11704   11872   13324   14241   10538
1     Arunachal Pradesh     135   146.2   146.2   158.1   163.9   215.8
2                 Assam  3470.7  3552.5    2916    3319  4008.5  4335.9
3                 Bihar  2472.2  3495.5  4989.3  4418.1  5590.3  3599.3
4          Chhattisgarh  4383.3  5011.6  5041.4  5426.6  4391.8  4110.4

   2010-11  2011-12  2012-13
0   7882.4   7746.2   6862.4
1    234.0    255.0    263.0
2   4736.6   4516.3   5128.5
3   3102.1   7162.6   7529.3
4   6159.0   6028.4   6608.8
```

[ ]:
```
rainfall_data.
 ↪drop(['JAN'        ,'FEB',        'MAR',        'APR',        'MAY',        'JUN'        ,'
 ↪= 1,inplace=True)
```

```
[ ]: rainfall_data.head()
```

```
[ ]:                   States/UTs  YEAR  ANNUAL
    0  Andaman & Nicobar Islands  1901  3373.2
    1  Andaman & Nicobar Islands  1902  3520.7
    2  Andaman & Nicobar Islands  1903  2957.4
    3  Andaman & Nicobar Islands  1904  3079.6
    4  Andaman & Nicobar Islands  1905  2566.7
```

```
[ ]: rainfall_data = rainfall_data[rainfall_data['YEAR']>=2004]
```

```
[ ]: rainfall_data
```

```
[ ]:                      States/UTs  YEAR  ANNUAL
    98    Andaman & Nicobar Islands  2004  2460.1
    99    Andaman & Nicobar Islands  2005  2954.7
    100   Andaman & Nicobar Islands  2006  2404.7
    101   Andaman & Nicobar Islands  2007  2748.0
    102   Andaman & Nicobar Islands  2008  3374.8
    …                           …     …       …
    4182                Lakshadweep  2012  1405.5
    4183                Lakshadweep  2013  1426.3
    4184                Lakshadweep  2014  1395.0
    4185                Lakshadweep  2015  1642.9
    4186                Lakshadweep  2016  1065.7

    [503 rows x 3 columns]
```

```
[ ]: rainfall_data.columns
```

```
[ ]: Index(['States/UTs', 'YEAR', 'ANNUAL'], dtype='object')
```

```
[ ]: yield_data.head()
```

```
[ ]:   State/Union Territory 2004-05 2005-06 2006-07 2007-08 2008-09 2009-10  \
    0        Andhra Pradesh    9601   11704   11872   13324   14241   10538
    1     Arunachal Pradesh     135   146.2   146.2   158.1   163.9   215.8
    2                 Assam  3470.7  3552.5    2916    3319  4008.5  4335.9
    3                 Bihar  2472.2  3495.5  4989.3  4418.1  5590.3  3599.3
    4          Chhattisgarh  4383.3  5011.6  5041.4  5426.6  4391.8  4110.4

       2010-11  2011-12  2012-13
    0   7882.4   7746.2   6862.4
    1    234.0    255.0    263.0
    2   4736.6   4516.3   5128.5
    3   3102.1   7162.6   7529.3
    4   6159.0   6028.4   6608.8
```

[ ]:
```
yield_data.columns
```

[ ]:
```
Index(['State/Union Territory', '2004-05', '2005-06', '2006-07', '2007-08',
       '2008-09', '2009-10', '2010-11', '2011-12', '2012-13'],
      dtype='object')
```

[ ]:
```
df = yield_data
```

[ ]:
```
melted_df = pd.melt(df, id_vars=["State/Union Territory"], var_name="YEAR",␣
↪value_name="PRODUCTION")

# Print the melted DataFrame
print(melted_df.to_string(index=False))
```

```
State/Union Territory    YEAR PRODUCTION
       Andhra Pradesh 2004-05       9601
    Arunachal Pradesh 2004-05        135
                Assam 2004-05     3470.7
                Bihar 2004-05     2472.2
         Chhattisgarh 2004-05     4383.3
         NCT of Delhi 2004-05       14.3
                  Goa 2004-05      145.2
              Gujarat 2004-05     1238.2
              Haryana 2004-05       3023
     Himachal Pradesh 2004-05        122
      Jammu & Kashmir 2004-05      492.2
            Jharkhand 2004-05       1677
            Karnataka 2004-05       3547
               Kerala 2004-05      667.1
       Madhya Pradesh 2004-05       1169
          Maharashtra 2004-05       2164
              Manipur 2004-05      435.9
            Meghalaya 2004-05      193.7
              Mizoram 2004-05      104.1
             Nagaland 2004-05      259.8
               Odisha 2004-05       6466
           Puducherry 2004-05       65.7
               Punjab 2004-05      10437
            Rajasthan 2004-05      150.4
               Sikkim 2004-05       21.6
           Tamil Nadu 2004-05     5062.2
            Telangana 2004-05          .
              Tripura 2004-05      545.1
        Uttar Pradesh 2004-05     9555.6
          Uttarakhand 2004-05        572
          West Bengal 2004-05    14884.8
            ALL INDIA 2004-05    83131.7
```

Andhra Pradesh	2005-06	11704
Arunachal Pradesh	2005-06	146.2
Assam	2005-06	3552.5
Bihar	2005-06	3495.5
Chhattisgarh	2005-06	5011.6
NCT of Delhi	2005-06	24
Goa	2005-06	147.3
Gujarat	2005-06	1298
Haryana	2005-06	3210
Himachal Pradesh	2005-06	112.1
Jammu & Kashmir	2005-06	556.8
Jharkhand	2005-06	1558
Karnataka	2005-06	5744
Kerala	2005-06	629.9
Madhya Pradesh	2005-06	1656.3
Maharashtra	2005-06	2695
Manipur	2005-06	386.1
Meghalaya	2005-06	151.9
Mizoram	2005-06	99.2
Nagaland	2005-06	263.1
Odisha	2005-06	6859
Puducherry	2005-06	59.9
Punjab	2005-06	10193
Rajasthan	2005-06	153
Sikkim	2005-06	21.5
Tamil Nadu	2005-06	5220
Telangana	2005-06	.
Tripura	2005-06	552.9
Uttar Pradesh	2005-06	11133.7
Uttarakhand	2005-06	590
West Bengal	2005-06	14510.8
ALL INDIA	2005-06	91793.4
Andhra Pradesh	2006-07	11872
Arunachal Pradesh	2006-07	146.2
Assam	2006-07	2916
Bihar	2006-07	4989.3
Chhattisgarh	2006-07	5041.4
NCT of Delhi	2006-07	31.1
Goa	2006-07	130.3
Gujarat	2006-07	1390
Haryana	2006-07	3371
Himachal Pradesh	2006-07	123.5
Jammu & Kashmir	2006-07	554
Jharkhand	2006-07	2967.8
Karnataka	2006-07	3446
Kerala	2006-07	631
Madhya Pradesh	2006-07	1368.4
Maharashtra	2006-07	2569

Manipur	2006-07	386.1
Meghalaya	2006-07	200.2
Mizoram	2006-07	29.5
Nagaland	2006-07	263.5
Odisha	2006-07	6824.7
Puducherry	2006-07	59.9
Punjab	2006-07	10138
Rajasthan	2006-07	169.8
Sikkim	2006-07	21.5
Tamil Nadu	2006-07	6610.6
Telangana	2006-07	.
Tripura	2006-07	620.5
Uttar Pradesh	2006-07	11124
Uttarakhand	2006-07	556
West Bengal	2006-07	14745.9
ALL INDIA	2006-07	93355.3
Andhra Pradesh	2007-08	13324
Arunachal Pradesh	2007-08	158.1
Assam	2007-08	3319
Bihar	2007-08	4418.1
Chhattisgarh	2007-08	5426.6
NCT of Delhi	2007-08	31.4
Goa	2007-08	121.6
Gujarat	2007-08	1474
Haryana	2007-08	3613
Himachal Pradesh	2007-08	121.5
Jammu & Kashmir	2007-08	561.3
Jharkhand	2007-08	3336.4
Karnataka	2007-08	3717
Kerala	2007-08	528.5
Madhya Pradesh	2007-08	1461.9
Maharashtra	2007-08	2996
Manipur	2007-08	406.2
Meghalaya	2007-08	200
Mizoram	2007-08	15.7
Nagaland	2007-08	290.6
Odisha	2007-08	7540.7
Puducherry	2007-08	53.4
Punjab	2007-08	10489
Rajasthan	2007-08	259.6
Sikkim	2007-08	22.9
Tamil Nadu	2007-08	5040.2
Telangana	2007-08	.
Tripura	2007-08	624.6
Uttar Pradesh	2007-08	11780
Uttarakhand	2007-08	593
West Bengal	2007-08	14719.5
ALL INDIA	2007-08	96692.9

Andhra Pradesh	2008-09	14241
Arunachal Pradesh	2008-09	163.9
Assam	2008-09	4008.5
Bihar	2008-09	5590.3
Chhattisgarh	2008-09	4391.8
NCT of Delhi	2008-09	31.4
Goa	2008-09	123.3
Gujarat	2008-09	1303
Haryana	2008-09	3298
Himachal Pradesh	2008-09	118.3
Jammu & Kashmir	2008-09	563.1
Jharkhand	2008-09	3420.2
Karnataka	2008-09	3802
Kerala	2008-09	590.3
Madhya Pradesh	2008-09	1559.7
Maharashtra	2008-09	2284
Manipur	2008-09	397
Meghalaya	2008-09	203.9
Mizoram	2008-09	46
Nagaland	2008-09	345.1
Odisha	2008-09	6812.7
Puducherry	2008-09	50.8
Punjab	2008-09	11000
Rajasthan	2008-09	241.1
Sikkim	2008-09	21.7
Tamil Nadu	2008-09	5182.7
Telangana	2008-09	.
Tripura	2008-09	627.1
Uttar Pradesh	2008-09	13097
Uttarakhand	2008-09	582
West Bengal	2008-09	15037.3
ALL INDIA	2008-09	99182.5
Andhra Pradesh	2009-10	10538
Arunachal Pradesh	2009-10	215.8
Assam	2009-10	4335.9
Bihar	2009-10	3599.3
Chhattisgarh	2009-10	4110.4
NCT of Delhi	2009-10	19.3
Goa	2009-10	100.6
Gujarat	2009-10	1292
Haryana	2009-10	3625
Himachal Pradesh	2009-10	105.9
Jammu & Kashmir	2009-10	497.4
Jharkhand	2009-10	1538.4
Karnataka	2009-10	3691
Kerala	2009-10	598.3
Madhya Pradesh	2009-10	1260.6
Maharashtra	2009-10	2183

Manipur	2009-10	319.9
Meghalaya	2009-10	206.7
Mizoram	2009-10	44.3
Nagaland	2009-10	240.3
Odisha	2009-10	6917.5
Puducherry	2009-10	52.4
Punjab	2009-10	11236
Rajasthan	2009-10	228.3
Sikkim	2009-10	24.3
Tamil Nadu	2009-10	5665.2
Telangana	2009-10	.
Tripura	2009-10	640
Uttar Pradesh	2009-10	10807.1
Uttarakhand	2009-10	608
West Bengal	2009-10	14340.7
ALL INDIA	2009-10	89092.9
Andhra Pradesh	2010-11	7882.4
Arunachal Pradesh	2010-11	234.0
Assam	2010-11	4736.6
Bihar	2010-11	3102.1
Chhattisgarh	2010-11	6159.0
NCT of Delhi	2010-11	19.6
Goa	2010-11	115.0
Gujarat	2010-11	1496.6
Haryana	2010-11	3472.0
Himachal Pradesh	2010-11	128.9
Jammu & Kashmir	2010-11	507.7
Jharkhand	2010-11	1110.0
Karnataka	2010-11	4188.0
Kerala	2010-11	522.7
Madhya Pradesh	2010-11	1772.1
Maharashtra	2010-11	2696.0
Manipur	2010-11	521.7
Meghalaya	2010-11	207.0
Mizoram	2010-11	47.2
Nagaland	2010-11	381.4
Odisha	2010-11	6827.7
Puducherry	2010-11	52.0
Punjab	2010-11	10837.0
Rajasthan	2010-11	265.5
Sikkim	2010-11	21.0
Tamil Nadu	2010-11	5792.4
Telangana	2010-11	6535.6
Tripura	2010-11	702.5
Uttar Pradesh	2010-11	11992.0
Uttarakhand	2010-11	550.4
West Bengal	2010-11	13045.9
ALL INDIA	2010-11	95979.8

Andhra Pradesh	2011-12	7746.2
Arunachal Pradesh	2011-12	255.0
Assam	2011-12	4516.3
Bihar	2011-12	7162.6
Chhattisgarh	2011-12	6028.4
NCT of Delhi	2011-12	19.8
Goa	2011-12	121.8
Gujarat	2011-12	1790.0
Haryana	2011-12	3759.0
Himachal Pradesh	2011-12	131.6
Jammu & Kashmir	2011-12	544.7
Jharkhand	2011-12	3130.6
Karnataka	2011-12	3955.0
Kerala	2011-12	569.0
Madhya Pradesh	2011-12	2227.3
Maharashtra	2011-12	2841.0
Manipur	2011-12	591.0
Meghalaya	2011-12	216.5
Mizoram	2011-12	54.3
Nagaland	2011-12	382.4
Odisha	2011-12	5807.0
Puducherry	2011-12	42.1
Punjab	2011-12	10542.0
Rajasthan	2011-12	253.4
Sikkim	2011-12	20.9
Tamil Nadu	2011-12	7458.7
Telangana	2011-12	5148.8
Tripura	2011-12	718.3
Uttar Pradesh	2011-12	14022.0
Uttarakhand	2011-12	594.0
West Bengal	2011-12	14605.8
ALL INDIA	2011-12	105310.9
Andhra Pradesh	2012-13	6862.4
Arunachal Pradesh	2012-13	263.0
Assam	2012-13	5128.5
Bihar	2012-13	7529.3
Chhattisgarh	2012-13	6608.8
NCT of Delhi	2012-13	19.7
Goa	2012-13	122.8
Gujarat	2012-13	1541.0
Haryana	2012-13	3976.0
Himachal Pradesh	2012-13	125.3
Jammu & Kashmir	2012-13	818.1
Jharkhand	2012-13	3164.9
Karnataka	2012-13	3364.0
Kerala	2012-13	508.3
Madhya Pradesh	2012-13	2775.0
Maharashtra	2012-13	3057.0

```
              Manipur 2012-13      257.6
            Meghalaya 2012-13      232.0
              Mizoram 2012-13       30.5
             Nagaland 2012-13      405.2
               Odisha 2012-13     7295.5
           Puducherry 2012-13       46.5
               Punjab 2012-13    11374.0
            Rajasthan 2012-13      222.5
               Sikkim 2012-13       21.3
           Tamil Nadu 2012-13     4049.9
            Telangana 2012-13     4647.6
              Tripura 2012-13      713.2
        Uttar Pradesh 2012-13    14416.0
          Uttarakhand 2012-13      579.8
          West Bengal 2012-13    15023.7
            ALL INDIA 2012-13   105231.6
```

[ ]:
```
melted_df
```

[ ]:
```
     State/Union Territory     YEAR PRODUCTION
0           Andhra Pradesh  2004-05       9601
1        Arunachal Pradesh  2004-05        135
2                    Assam  2004-05     3470.7
3                    Bihar  2004-05     2472.2
4             Chhattisgarh  2004-05     4383.3
..                     …      …         …
283               Tripura  2012-13      713.2
284         Uttar Pradesh  2012-13    14416.0
285           Uttarakhand  2012-13      579.8
286           West Bengal  2012-13    15023.7
287             ALL INDIA  2012-13   105231.6

[288 rows x 3 columns]
```

[ ]:
```
melted_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 288 entries, 0 to 287
Data columns (total 3 columns):
 #   Column                 Non-Null Count  Dtype 
---  ------                 --------------  ----- 
 0   State/Union Territory  288 non-null    object
 1   YEAR                   288 non-null    object
 2   PRODUCTION             288 non-null    object
dtypes: object(3)
memory usage: 6.9+ KB
```

```
[ ]: melted_df = melted_df.rename(columns={"State/Union Territory": "State/UTs"})
```

```
[ ]: melted_df.head()
```

```
[ ]:            State/UTs     YEAR PRODUCTION
     0     Andhra Pradesh  2004-05       9601
     1  Arunachal Pradesh  2004-05        135
     2              Assam  2004-05     3470.7
     3              Bihar  2004-05     2472.2
     4       Chhattisgarh  2004-05     4383.3
```

```
[ ]: melted_df['YEAR'] = melted_df['YEAR'].str.split('-').str[0].astype(int)
```

```
[ ]: melted_df.head()
     melted_df.info()
```

```
     <class 'pandas.core.frame.DataFrame'>
     RangeIndex: 288 entries, 0 to 287
     Data columns (total 3 columns):
      #   Column      Non-Null Count  Dtype
     ---  ------      --------------  -----
      0   State/UTs   288 non-null    object
      1   YEAR        288 non-null    int64
      2   PRODUCTION  288 non-null    object
     dtypes: int64(1), object(2)
     memory usage: 6.9+ KB
```

```
[ ]: melted_df.head()
```

```
[ ]:            State/UTs  YEAR PRODUCTION
     0     Andhra Pradesh  2004       9601
     1  Arunachal Pradesh  2004        135
     2              Assam  2004     3470.7
     3              Bihar  2004     2472.2
     4       Chhattisgarh  2004     4383.3
```

```
[ ]: merged_df = pd.merge(melted_df, rainfall_data, how='inner', left_on=['State/
      ↪UTs', 'YEAR'], right_on=['States/UTs', 'YEAR'])
```

```
[ ]: merged_df.head()
```

```
[ ]:            State/UTs  YEAR PRODUCTION         States/UTs  ANNUAL
     0  Arunachal Pradesh  2004        135  Arunachal Pradesh  2545.7
     1              Bihar  2004     2472.2              Bihar  1147.8
     2       Chhattisgarh  2004     4383.3       Chhattisgarh  1144.5
     3   Himachal Pradesh  2004        122   Himachal Pradesh   878.5
     4    Jammu & Kashmir  2004      492.2    Jammu & Kashmir   944.9
```

[ ]:
```
merged_df.drop(columns=['States/UTs'], inplace=True)
```

[ ]:
```
merged_df.head()
```

[ ]:
```
            State/UTs  YEAR PRODUCTION  ANNUAL
0  Arunachal Pradesh  2004        135  2545.7
1              Bihar  2004     2472.2  1147.8
2       Chhattisgarh  2004     4383.3  1144.5
3   Himachal Pradesh  2004        122   878.5
4    Jammu & Kashmir  2004      492.2   944.9
```

[ ]:
```
merged_df = merged_df.rename(columns={"ANNUAL": "ANNUAL RAINFALL (mm)"})
```

[ ]:
```
merged_df.head()
```

[ ]:
```
            State/UTs  YEAR PRODUCTION  ANNUAL RAINFALL (mm)
0  Arunachal Pradesh  2004        135                2545.7
1              Bihar  2004     2472.2                1147.8
2       Chhattisgarh  2004     4383.3                1144.5
3   Himachal Pradesh  2004        122                 878.5
4    Jammu & Kashmir  2004      492.2                 944.9
```

[ ]:
```
df = merged_df
```

[ ]:
```
df.head()
```

[ ]:
```
            State/UTs  YEAR PRODUCTION  ANNUAL RAINFALL (mm)
0  Arunachal Pradesh  2004        135                2545.7
1              Bihar  2004     2472.2                1147.8
2       Chhattisgarh  2004     4383.3                1144.5
3   Himachal Pradesh  2004        122                 878.5
4    Jammu & Kashmir  2004      492.2                 944.9
```

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

[ ]:
```
State/UTs               0
YEAR                    0
PRODUCTION              0
ANNUAL RAINFALL (mm)    1
dtype: int64
```

[ ]:
```
replacement_value = 0  # You can replace NaN with any desired value

# Replace NaN values with the specified value
df['ANNUAL RAINFALL (mm)'].fillna(replacement_value, inplace=True)
```

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

[ ]:
```
State/UTs               0
YEAR                    0
PRODUCTION              0
ANNUAL RAINFALL (mm)    0
dtype: int64
```

[ ]:
```
grouped_df = df.groupby(['State/UTs', 'YEAR']).sum()

# Reset index to make the grouped columns into regular columns
grouped_df.reset_index(inplace=True)

# Print the grouped DataFrame
print(grouped_df)
```

```
            State/UTs  YEAR PRODUCTION  ANNUAL RAINFALL (mm)
0   Arunachal Pradesh  2004        135                2545.7
1   Arunachal Pradesh  2005      146.2                2335.5
2   Arunachal Pradesh  2006      146.2                2259.6
3   Arunachal Pradesh  2007      158.1                3020.7
4   Arunachal Pradesh  2008      163.9                2244.4
..                ...   ...        ...                   ...
94        Uttarakhand  2008        582                1363.7
95        Uttarakhand  2009        608                 847.8
96        Uttarakhand  2010      550.4                1424.0
97        Uttarakhand  2011      594.0                1564.7
98        Uttarakhand  2012      579.8                1223.9

[99 rows x 4 columns]
```

[ ]:
```
grouped_df.head()
```

[ ]:
```
            State/UTs  YEAR PRODUCTION  ANNUAL RAINFALL (mm)
0  Arunachal Pradesh  2004        135                2545.7
1  Arunachal Pradesh  2005      146.2                2335.5
2  Arunachal Pradesh  2006      146.2                2259.6
3  Arunachal Pradesh  2007      158.1                3020.7
4  Arunachal Pradesh  2008      163.9                2244.4
```

[ ]:
```
grouped_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99 entries, 0 to 98
Data columns (total 4 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   State/UTs             99 non-null     object
 1   YEAR                  99 non-null     int64
 2   PRODUCTION            99 non-null     object
```

```
 3   ANNUAL RAINFALL (mm)  99 non-null     float64
dtypes: float64(1), int64(1), object(2)
memory usage: 3.2+ KB
```

[ ]:
```python
import pandas as pd

# Assuming 'PRODUCTION' is the name of the column you want to convert
# Replace 'PRODUCTION' with the actual name of your column

# Convert float values to integers by rounding to the nearest integer
grouped_df['PRODUCTION'] = grouped_df['PRODUCTION'].round()

# Convert the column to integer type
grouped_df['PRODUCTION'] = grouped_df['PRODUCTION'].astype('Int64')
```

[ ]:
```python
# Import the necessary libraries
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import OneHotEncoder

# Method 1: Label Encoding
label_encoder = LabelEncoder()
grouped_df['State/UTs_LabelEncoded'] = label_encoder.
 ↪fit_transform(grouped_df['State/UTs'])

# # Method 2: One-Hot Encoding
# one_hot_encoder = OneHotEncoder(sparse=False)
# encoded_states = one_hot_encoder.fit_transform(grouped_df[['State/UTs']])
# encoded_states_df = pd.DataFrame(encoded_states, columns=one_hot_encoder.
 ↪get_feature_names(['State/UTs']))
# grouped_df = pd.concat([grouped_df, encoded_states_df], axis=1)

# Print the DataFrame with encoded columns
print(grouped_df)
```

```
            State/UTs  YEAR  PRODUCTION  ANNUAL RAINFALL (mm)  \
0   Arunachal Pradesh  2004         135                2545.7
1   Arunachal Pradesh  2005         146                2335.5
2   Arunachal Pradesh  2006         146                2259.6
3   Arunachal Pradesh  2007         158                3020.7
4   Arunachal Pradesh  2008         164                2244.4
..                ...   ...         ...                   ...
94        Uttarakhand  2008         582                1363.7
95        Uttarakhand  2009         608                 847.8
96        Uttarakhand  2010         550                1424.0
97        Uttarakhand  2011         594                1564.7
98        Uttarakhand  2012         580                1223.9

    State/UTs_LabelEncoded
```

```
0                        0
1                        0
2                        0
3                        0
4                        0
..                     …
94                      10
95                      10
96                      10
97                      10
98                      10

[99 rows x 5 columns]
```

[ ]:
```
# Drop rows with missing target values
grouped_df.dropna(subset=['PRODUCTION'], inplace=True)
```

[ ]:
```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

# Define features and target variable
# Assume 'PRODUCTION' is the target variable and the rest are features
X = grouped_df.drop(['PRODUCTION','State/UTs'], axis=1)  # Features
y = grouped_df['PRODUCTION']  # Target variable

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
 ↪random_state=42)

# Train a linear regression model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions on the testing data
y_pred = model.predict(X_test)

# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
print("Mean Squared Error:", mse)
```

```
Mean Squared Error: 6936444.969812487
```

[ ]:
```
from sklearn.metrics import mean_squared_error, mean_absolute_error
import numpy as np

# Calculate Mean Squared Error (MSE)
```

```
mse = mean_squared_error(y_test, y_pred)
print("Mean Squared Error (MSE):", mse)

# Calculate Root Mean Squared Error (RMSE)
rmse = np.sqrt(mse)
print("Root Mean Squared Error (RMSE):", rmse)

# Calculate Mean Absolute Error (MAE)
mae = mean_absolute_error(y_test, y_pred)
print("Mean Absolute Error (MAE):", mae)
```

```
Mean Squared Error (MSE): 6936444.969812487
Root Mean Squared Error (RMSE): 2633.713152530565
Mean Absolute Error (MAE): 2100.4235675540003
```

[ ]:
```
from sklearn.metrics import r2_score

# Calculate R-squared (R^2) score
r2 = r2_score(y_test, y_pred)
print("R-squared (R^2) score:", r2)
```

```
R-squared (R^2) score: 0.15557464894927664
```

[ ]:
```
from sklearn.linear_model import LinearRegression

# Assuming 'PRODUCTION' is the name of the production column and 'ANNUAL␣
 ↪RAINFALL (mm)' is the name of the rainfall column
# Assume 'State/UTs' is the name of the state column

# Initialize lists to store predictions for each state
predictions = []

# Iterate over each state
for state, state_data in grouped_df.groupby('State/UTs'):
    # Sort the data by year
    state_data = state_data.sort_values(by='YEAR')

    # Prepare X (features) and y (target) data for the state
    X_state = state_data[['YEAR', 'ANNUAL RAINFALL (mm)']].iloc[:-1]  # Exclude␣
 ↪the last year for prediction
    y_state = state_data['PRODUCTION'].iloc[1:]  # Production for the next year

    # Train a linear regression model
    model = LinearRegression()
    model.fit(X_state, y_state)

    # Prepare data for prediction (data from the last year)
```

```
    X_pred = state_data[['YEAR', 'ANNUAL RAINFALL (mm)']].iloc[[-1]]

    # Make prediction for the next year's production
    next_year_production = model.predict(X_pred)

    # Append prediction to the list
    predictions.append({'State/UTs': state, 'Next Year Production Prediction':␣
 ↪next_year_production[0]})

# Convert predictions list to DataFrame
predictions_df = pd.DataFrame(predictions)

# Print the predictions
print(predictions_df)
```

	State/UTs	Next Year Production Prediction
0	Arunachal Pradesh	273.070994
1	Bihar	6763.864481
2	Chhattisgarh	6378.915161
3	Himachal Pradesh	124.805271
4	Jammu & Kashmir	683.010642
5	Jharkhand	2681.837691
6	Kerala	543.996919
7	Punjab	11180.255093
8	Tamil Nadu	4138.392176
9	Telangana	4892.609354
10	Uttarakhand	578.355259

[ ]:
```
len(grouped_df['State/UTs'].unique())
```

[ ]: 11

[ ]:
```
from sklearn.linear_model import LinearRegression

# Initialize lists to store predictions for each state
predictions = []

# Iterate over each state
for state, state_data in grouped_df.groupby('State/UTs'):
    # Sort the data by year
    state_data = state_data.sort_values(by='YEAR')

    # Check if the state has enough historical data points
    if len(state_data) >= 6:  # Assuming you want to predict for the next five␣
 ↪years

        # Prepare X (features) and y (target) data for the state
```

```
        X_state = state_data[['YEAR', 'ANNUAL RAINFALL (mm)']].iloc[:-5]  # Exclude the last five years for prediction
        y_state = state_data['PRODUCTION'].iloc[5:]  # Production for the next five years

        # Train a linear regression model
        model = LinearRegression()
        model.fit(X_state, y_state)

        # Prepare data for prediction (data from the last year)
        X_pred = state_data[['YEAR', 'ANNUAL RAINFALL (mm)']].iloc[[-1]]

        # Make predictions for the next five years' production
        next_five_years_production = []
        for i in range(1, 6):
            year = X_pred.iloc[0]['YEAR'] + i
            X_pred['YEAR'] = year
            prediction = model.predict(X_pred)
            rounded_prediction = round(prediction[0], 2)  # Round to 2 decimal places (adjust as needed)
            next_five_years_production.append(rounded_prediction)
        # Append predictions to the list
        predictions.append({'State/UTs': state, 'Next Five Years Production Prediction': next_five_years_production})

# Convert predictions list to DataFrame
predictions_df = pd.DataFrame(predictions)

# Print the predictions
print(predictions_df)
```

```
          State/UTs                Next Five Years Production Prediction
0  Arunachal Pradesh            [364.78, 400.28, 453.53, 524.52, 613.26]
1              Bihar  [15902.06, 18756.92, 23039.21, 28748.94, 35886…
2       Chhattisgarh  [9377.26, 10066.81, 11101.15, 12480.26, 14204.14]
3   Himachal Pradesh             [142.69, 149.99, 160.94, 175.55, 193.8]
4    Jammu & Kashmir        [1399.18, 1601.51, 1905.02, 2309.7, 2815.54]
5           Jharkhand       [2903.32, 3259.97, 3794.96, 4508.28, 5399.93]
6              Kerala     [-578.44, -779.34, -1080.7, -1482.5, -1984.76]
7              Punjab  [14376.58, 15220.78, 16487.09, 18175.51, 20286…
8         Tamil Nadu        [3924.31, 3093.2, 1846.54, 184.32, -1893.45]
9        Uttarakhand             [581.57, 580.3, 578.39, 575.84, 572.66]
```

[ ]:
```
predictions_df.head(11)
```

[ ]:
```
          State/UTs                Next Five Years Production Prediction
0  Arunachal Pradesh            [364.78, 400.28, 453.53, 524.52, 613.26]
```

```
1              Bihar  [15902.06, 18756.92, 23039.21, 28748.94, 35886…
2       Chhattisgarh  [9377.26, 10066.81, 11101.15, 12480.26, 14204.14]
3   Himachal Pradesh             [142.69, 149.99, 160.94, 175.55, 193.8]
4    Jammu & Kashmir       [1399.18, 1601.51, 1905.02, 2309.7, 2815.54]
5           Jharkhand      [2903.32, 3259.97, 3794.96, 4508.28, 5399.93]
6              Kerala     [-578.44, -779.34, -1080.7, -1482.5, -1984.76]
7              Punjab  [14376.58, 15220.78, 16487.09, 18175.51, 20286…
8          Tamil Nadu       [3924.31, 3093.2, 1846.54, 184.32, -1893.45]
9         Uttarakhand            [581.57, 580.3, 578.39, 575.84, 572.66]
```

[ ]:
```
grouped_df.isnull().sum()
```

[ ]:
```
State/UTs                0
YEAR                     0
PRODUCTION               0
ANNUAL RAINFALL (mm)     0
State/UTs_LabelEncoded   0
dtype: int64
```

[ ]:
```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score

# Initialize lists to store accuracy metrics for each state
mse_list = []
rmse_list = []
mae_list = []
r2_list = []

# Iterate over each state
for state, state_data in grouped_df.groupby('State/UTs'):
    # Sort the data by year
    state_data = state_data.sort_values(by='YEAR')

    # Check if the state has enough historical data points
    if len(state_data) >= 6:  # Assuming you want to predict for the next five␣
 ↪years

        # Prepare X (features) and y (target) data for the state
        X_state = state_data[['YEAR', 'ANNUAL RAINFALL (mm)']].iloc[:-5]  #␣
 ↪Exclude the last five years for prediction
        y_state = state_data['PRODUCTION'].iloc[5:]  # Production for the next␣
 ↪five years

        # Split the data into training and testing sets (80% train, 20% test)
        X_train, X_test, y_train, y_test = train_test_split(X_state, y_state,␣
 ↪test_size=0.2, random_state=42)
```

```
        # Train a linear regression model
        model = LinearRegression()
        model.fit(X_train, y_train)

        # Make predictions on the testing data
        y_pred = model.predict(X_test)

        # Calculate evaluation metrics
        mse = mean_squared_error(y_test, y_pred)
        rmse = np.sqrt(mse)
        mae = mean_absolute_error(y_test, y_pred)
        r2 = r2_score(y_test, y_pred)

        # Append accuracy metrics to lists
        mse_list.append(mse)
        rmse_list.append(rmse)
        mae_list.append(mae)
        r2_list.append(r2)

        print(mse)
        print(rmse)
        print(mae)
        print(r2)
        print("Score",model.score(X_test, y_test))

# Calculate average accuracy metrics across all states
avg_mse = np.mean(mse_list)
avg_rmse = np.mean(rmse_list)
avg_mae = np.mean(mae_list)
avg_r2 = np.mean(r2_list)

# Print the average accuracy metrics
print("Average Mean Squared Error (MSE):", avg_mse)
print("Average Root Mean Squared Error (RMSE):", avg_rmse)
print("Average Mean Absolute Error (MAE):", avg_mae)
print("Average R-squared (R^2) score:", avg_r2)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_regression.py:918:
UndefinedMetricWarning: R^2 score is not well-defined with less than two
samples.
  warnings.warn(msg, UndefinedMetricWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_regression.py:918:
UndefinedMetricWarning: R^2 score is not well-defined with less than two
samples.
  warnings.warn(msg, UndefinedMetricWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_regression.py:918:
```

```
UndefinedMetricWarning: R^2 score is not well-defined with less than two
samples.
  warnings.warn(msg, UndefinedMetricWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_regression.py:918:
UndefinedMetricWarning: R^2 score is not well-defined with less than two
samples.
  warnings.warn(msg, UndefinedMetricWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_regression.py:918:
UndefinedMetricWarning: R^2 score is not well-defined with less than two
samples.
  warnings.warn(msg, UndefinedMetricWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_regression.py:918:
UndefinedMetricWarning: R^2 score is not well-defined with less than two
samples.
  warnings.warn(msg, UndefinedMetricWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_regression.py:918:
UndefinedMetricWarning: R^2 score is not well-defined with less than two
samples.
  warnings.warn(msg, UndefinedMetricWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_regression.py:918:
UndefinedMetricWarning: R^2 score is not well-defined with less than two
samples.
  warnings.warn(msg, UndefinedMetricWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_regression.py:918:
UndefinedMetricWarning: R^2 score is not well-defined with less than two
samples.
  warnings.warn(msg, UndefinedMetricWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_regression.py:918:
UndefinedMetricWarning: R^2 score is not well-defined with less than two
samples.
  warnings.warn(msg, UndefinedMetricWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_regression.py:918:
UndefinedMetricWarning: R^2 score is not well-defined with less than two
samples.
  warnings.warn(msg, UndefinedMetricWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_regression.py:918:
UndefinedMetricWarning: R^2 score is not well-defined with less than two
samples.
  warnings.warn(msg, UndefinedMetricWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_regression.py:918:
UndefinedMetricWarning: R^2 score is not well-defined with less than two
samples.
  warnings.warn(msg, UndefinedMetricWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_regression.py:918:
UndefinedMetricWarning: R^2 score is not well-defined with less than two
samples.
  warnings.warn(msg, UndefinedMetricWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_regression.py:918:
```

```
UndefinedMetricWarning: R^2 score is not well-defined with less than two
samples.
  warnings.warn(msg, UndefinedMetricWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_regression.py:918:
UndefinedMetricWarning: R^2 score is not well-defined with less than two
samples.
  warnings.warn(msg, UndefinedMetricWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_regression.py:918:
UndefinedMetricWarning: R^2 score is not well-defined with less than two
samples.
  warnings.warn(msg, UndefinedMetricWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_regression.py:918:
UndefinedMetricWarning: R^2 score is not well-defined with less than two
samples.
  warnings.warn(msg, UndefinedMetricWarning)
5.541737262456535
2.354089476306399
2.354089476306399
nan
Score nan
7492234.576973229
2737.1946545639075
2737.1946545639075
nan
Score nan
842237.8507985848
917.7351746547502
917.7351746547502
nan
Score nan
1.7083823435637717
1.3070510103143533
1.3070510103143533
nan
Score nan
17756.824890855412
133.25473684209283
133.25473684209283
nan
Score nan
31469.195151489355
177.3955894364044
177.3955894364044
nan
Score nan
1012.8928311892082
31.82597730139969
```

```
31.82597730139969
nan
Score nan
64987.9420658834
254.9273270284757
254.9273270284757
nan
Score nan
59513083.102298334
7714.472315220162
7714.472315220162
nan
Score nan
2463.970132257002
49.638393731636825
49.638393731636825
nan
Score nan
Average Mean Squared Error (MSE): 6796525.360526143
Average Root Mean Squared Error (RMSE): 1202.010530926545
Average Mean Absolute Error (MAE): 1202.010530926545
Average R-squared (R^2) score: nan

/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_regression.py:918:
UndefinedMetricWarning: R^2 score is not well-defined with less than two
samples.
  warnings.warn(msg, UndefinedMetricWarning)
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_regression.py:918:
UndefinedMetricWarning: R^2 score is not well-defined with less than two
samples.
  warnings.warn(msg, UndefinedMetricWarning)
```

[ ]: