

Contents

1. Chapter 1	
Introduction	3
1.1 Introduction.....	3
1.2 Literature Review.....	3
1.3 Objectives.....	4
1.4 Scope of the study.....	4
2. Chapter 2	
Methodology	5
2.1 Materials and Methods.....	5
2.2 About the data.....	5
2.3 Statistical Techniques for data analysis.....	5
2.3.1 Bar chart.....	6
2.3.2 Line chart.....	6
2.3.3 Pie chart.....	6
2.3.4 T test.....	6
2.3.5 One-way Analysis of Variance (ANOVA)	7
3. Chapter 3	
Results and Discussion	8
3.1 Analysis of factor.....	8
3.1.1 Best season for crop production in Karnataka state	8
3.1.2 The most common growing crop in Karnataka for agriculture.....	9
3.1.3 Analysis of highest and lowest yield in an year.....	10
3.1.4 Analysis of district wise crop production.....	11
3.1.5 Analysis of annual crop production from 1997-98 to 2019-20.....	11
3.1.6 Highest crop production ever received in a year in Districts.....	13
3.1.7 Analysis of crop production in coastal region.....	14
3.1.7.1 Annual crop production in Dakshina Kannada.....	14
3.1.7.2 Annual crop production in Udupi.....	15
3.1.7.3 Annual crop production in Uttara Kannada.....	15
3.1.7.4 Most common growing crop in coastal region.....	16
3.1.7.5 Analysis of distribution of total production of coconut crop in coastal region.....	17
3.1.7.6 Testing significance difference between production of coconut in districts of coastal region.....	18
3.1.7.7 Analysis of distribution of total production of arecanut crop in coastal region.....	18
3.1.7.8 Testing significance difference between production of arecanut in districts of coastal region.....	19
3.1.8 Analysis of crop production on North Karnataka and South Karnataka.....	19
3.1.8.1 Analysis of distribution of total production of maize in North Karnataka and South Karnataka.....	20

3.1.8.2 Testing significance difference between mean production of maize in North Karnataka and South Karnataka.....	20
4. Chapter 4	
Conclusion	21
4.1 Conclusion.....	21
5. Chapter 5	22
5.1 Summary.....	22
6. Chapter 6	23
6.1 Bibliography.....	23
7. Chapter 7	
Appendix	24
7.1 Python code for analysis.....	24

1 Chapter 1

Introduction:

1.1 Introduction

The history of Indian agriculture dates back to 10,000 years. Indian agriculture began during 9000 BC as a result of early cultivation of plants and domestication of crops and animals. The agricultural sector employed 60 per cent of the total workforce in India, and despite a steady decline of its share in the GDP, it still remains the largest economic sector. Agriculture plays a significant role in the overall socioeconomic development of India

Karnataka is India's eighth largest state in geographical area covering 1.92 lakh sq km and accounting for 6.3 per cent of the geographical area of the country. The state is delineated into 30 districts and 176 taluks spread over 27,481 villages. In Karnataka, agriculture is the major occupation for a majority of the rural population. As per the population Census 2011, agriculture supports 13.74 million workers, of which 23.61 per cent are cultivators and 25.67 per cent agricultural workers. A total of 123,100 km² of land is cultivated in Karnataka constituting 64.6% of the total geographical area of the state. Agriculture plays an important role in the overall growth of Karnataka's economy despite a fall in its share in the state domestic product. In Karnataka, horticulture crops occupy about 15.21 lakh hectares with an annual production of about 96.60 lakh tonnes.

About Agriculture Census

Agriculture census is conducted by the "Ministry of Agriculture and Farmers Welfare". It is , conducted every 5 years , which is undertaken now after delay due to Covid-19 Pandemic. Entire Census operation is conducted in three phases and operational holding is taken as a statistical unit at micro level for data collection. Based on the agriculture census data collected in three phases, the Department brings out three detailed reports analysing trends on various parameters at All India and States/UTs level. District/Tehsil level reports are prepared by the respective States/UTs. Agricultural Census is the main source of information on a variety of agricultural parameters at a relatively minute level, such as the number and area of operational holdings, their size, class-wise distribution, land use, tenancy and cropping pattern, etc.

1.2 Literature Review

Statistical Analysis Software for Agricultural Research Data Analysis Jayade K. G.*
Deshmukh P. D. Khot P. G. Asst. Professor in Computer Science, Asst. Professor in Statistics
Professor Dr. PDKV, College of Agriculture, Dr. PDKV, College of Agriculture Dept. of
Statistics, RTM Nagpur , Maharashtra, India Nagpur, Maharashtra, India University Nagpur,
Maharashtra, - Agriculture is the backbone of India and agriculture research is required for
sustainable and modern agriculture. India is the agricultural country and the rural India is
depending on the agriculture. In India, many organizations are working for research and
development in agriculture. For sustainable agriculture, statistical analysis on the research
data is most important. With the use of computer in agriculture sector, statistical analysis
becomes easier. This software's are very easy, gives fast result and are reliable. Statistical

analysis is used for analysis of agricultural research data. Many software's have been used since the use of computer in agriculture. This research paper is the study of statistical software used by the agricultural scientists in agriculture research for analysis of data.

1.3 Objectives

- To study the best season for crop production in Karnataka
- To study most and least choice crop in Karnataka for agriculture
- To determine Annual production of crops from 2010-11 to 2019-20
- To study about district wise total production of crops.
- To determine the highest Crop production ever received in a District
- Analysis of crop production in coastal region of Karnataka.
- Analysis of crop production in North Karnataka and South Karnataka

1.4 Scope of the study

The results of this project can be used for researchers and also government agencies to predict how agriculture in Karnataka helps in Economic growth and also predict how agriculture gives employment . It may help government to release some agriculture schemes and also may useful in time of budget.

The results of this project can be used for study how agriculture activity changes as year passes and also for study the region of Karnataka help more for agriculture activity.

2 Chapter 2

Methodology

2.1 Materials and Methods

A secondary data has been collected from the “Kaggle” website. The data contained features like state, district ,crop, year, season, area units, production, production units, yield. The data consists of 9 columns and 27493 records.

2.2 About the data

- **State** - Karnataka state
- **Districts** – Districts of Karnataka State
- **Crops** – Crops for which production is recorded.
- **Year** - From 1997-98 to 2019-20
- **Season** - Whole year , Kharif , Summer , Rabi
- **Area** - Crop production area in Hectare
- **Production** – Total production count.
- **Production Units** – Tonnes (measurement 100KG = 1 Ton)
- **Yield** - Yield are production per unit of harvested area for crop production.

2.3 Statistical Techniques used for Data Analysis

The programming language ‘Python’ has been used to carry out the analysis of the data. The statistical methods considered in order to carry out the analysis are given as follows:

2.3.1 Bar chart

A bar chart is a way of summarizing a set of categorical data (continuous data can be made categorical by auto-binning). The bar chart displays data using a number of bars, each representing a particular category. The height of each bar is proportional to a specific aggregation (for example the sum of the values in the category it represents). The categories could be something like an age group or a geographical location. It is also possible to color or split each bar into another categorical column in the data, which enables you to see the contribution from different categories to each bar or group of bars in the bar chart.

2.3.2 Line chart

A line chart also known as curve chart is a type of chart which displays information as a series of data points called 'markers' connected by straight line segments. It is a basic type of chart commonly in many field. It is similar to a scatter plot expect that the measurement points are ordered(typically by their x-axis value) and joined with straight line segments

2.3.3 Pie chart

A pie chart is a statistical graph, which is divided into slices to illustrate numerical propotion. Pie chart represents the data in a circular graph. The slice of pie show the relative size of the data and it is a type of pictorial.

2.3.3 T-test

A t-test is a type of inferential statistic used to determine if there is a significant difference between the means of two groups, which may be related in certain features.

The t-test is one of many tests used for the purpose of hypothesis testing in statistics. It can be used to check relationship between categorical predictor variable and numerical target variable.

The formula for the two-sample t-test (the Student's t-test) is shown below.

T-test formula:

$$t = \frac{\bar{X}_1 - \bar{X}_2 - d}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

where \bar{X}_1 is the mean of sample 1, \bar{X}_2 is the mean of sample 2, d is the hypothesized difference between population means, s_1 is the standard deviation of sample 1, s_2 is the standard deviation of sample 2, and n_1 is the size of the sample 1, and n_2 is the size of sample 2. A larger t-value shows that the difference between group means is greater than the pooled standard error, indicating a more significant difference between the groups.

2.3.4 One-way Analysis of Variance (ANOVA)

A key statistical test in research fields including biology, economics and psychology, Analysis of Variance (ANOVA) is very useful for analyzing datasets. It allows comparisons to be made between three or more groups of data. The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups.

One-way ANOVA is a statistical method to test the null hypothesis (H_0) that three or more population means are equal vs. the alternative hypothesis (H_1) that at least one mean is different.

ANOVA determines whether the groups created by the levels of the independent variable are statistically different by calculating whether the means of the treatment levels are different from the overall mean of the dependent variable. If any of the group means is significantly different from the overall mean, then the null hypothesis is rejected. ANOVA uses the F-test for statistical significance. This allows for comparison of multiple means at once, because the error is calculated for the whole set of comparisons rather than for each individual two-way comparison (which would happen with a t-test).

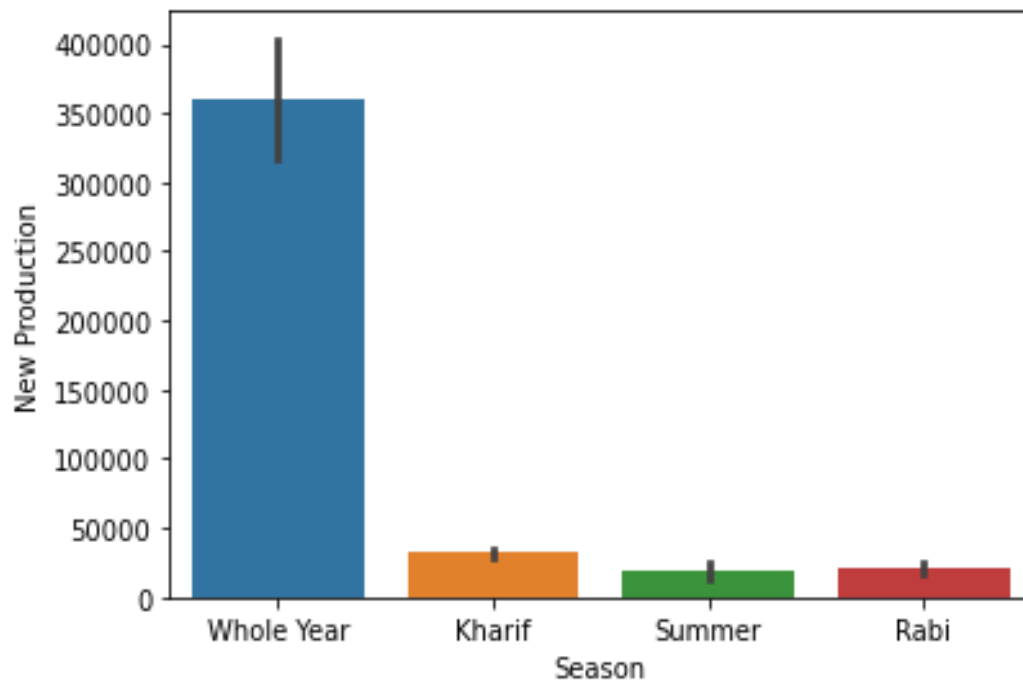
The F-test compares the variance in each group mean from the overall group variance. If the variance within groups is smaller than the variance between groups, the F-test will find a higher F-value, and therefore a higher likelihood that the difference observed is real and not due to chance.

3 Chapter 3

Result and Discussion

3.1 Analysis of factor

3.1.1 Best season for crop production in Karnataka state



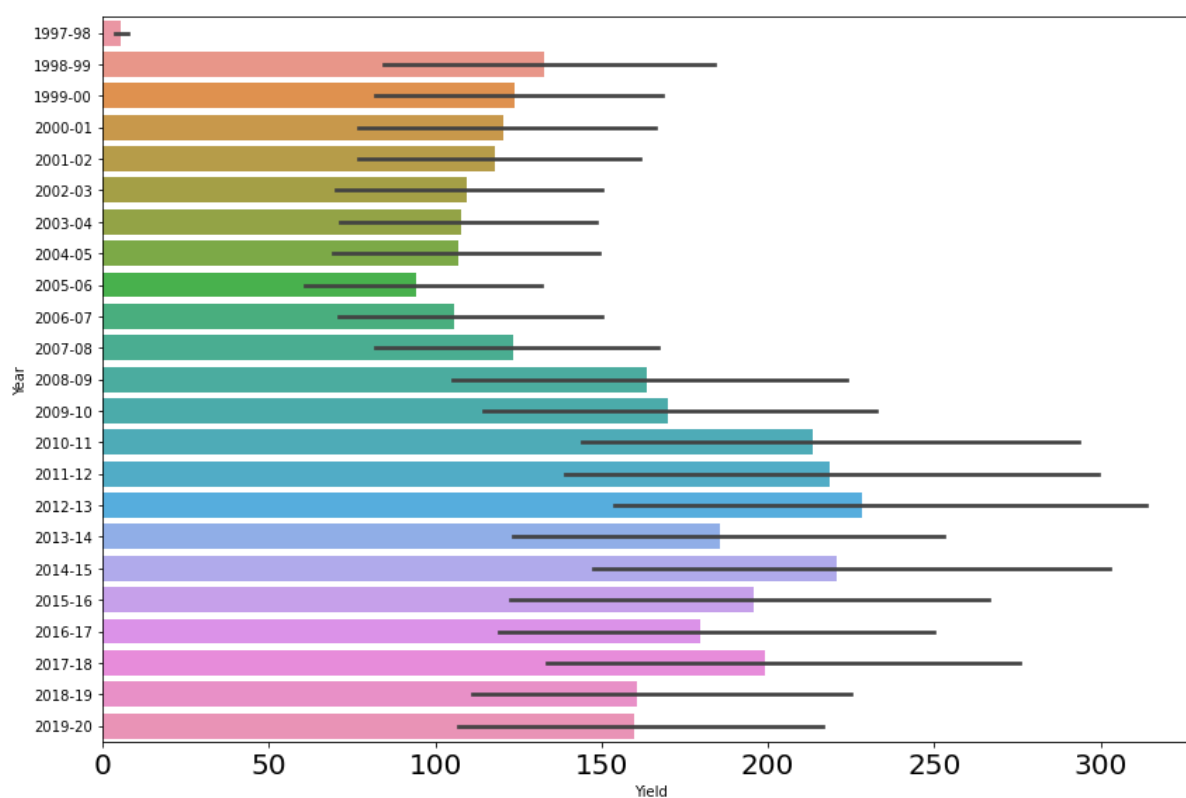
From the above graph we understood that whole year season have maximum crop production in Karnataka.

3.1.2 The most common growing crop in Karnataka for agriculture.

Maize	1736
Rice	1538
Sunflower	1420
Jowar	1242
Dry chillies	1225
Onion	1191
Groundnut	1170
Horse-gram	1167
Ragi	1112
Moong(Green Gram)	897
Urad	773
Potato	673
Cowpea(Lobia)	672
Cotton(lint)	636
Coconut	627
Gram	590
Arhar/Tur	586
Sesamum	574
Other Kharif pulses	541
Sugarcane	533
Niger seed	526
Bajra	505
Other Rabi pulses	485
Castor seed	483
Turmeric	472
Coriander	461
Ginger	444
Arecanut	443
Small millets	421
Rapeseed & Mustard	380
Wheat	360
Garlic	350
Safflower	340
Banana	333
Sweet potato	318
Black pepper	309
Linseed	301
Tobacco	287
Soyabean	270
Cashewnut	268
Peas & beans (Pulses)	211
Tapioca	194
Cardamom	184
Sannhamp	127
Mesta	118

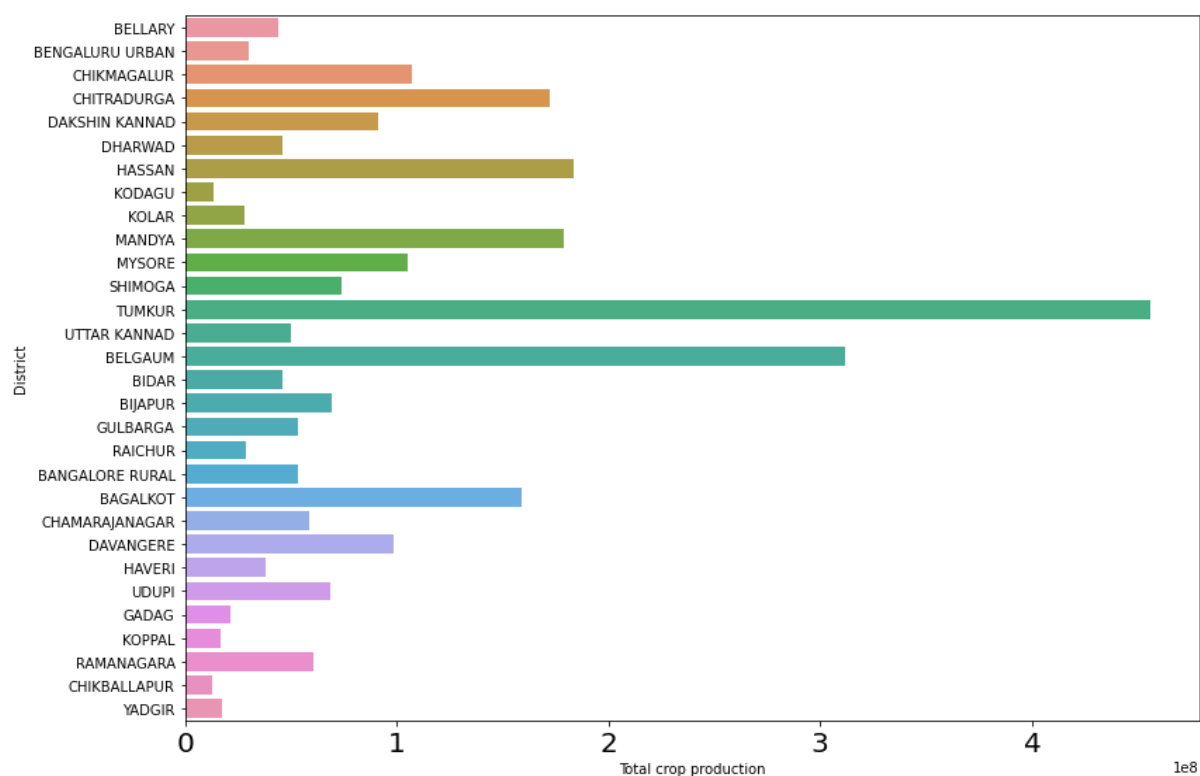
The above table shows that ‘Maize’ is the most growing and ‘Mesta’ is the least growing crop in Karnataka from 1997-2020

3.1.3 Analysis of highest and lowest yield in an year



The above graph shows that in the year ‘2019-20’ has highest yield and ‘1998-99’ has lowest yield in Karnataka.

3.1.4 Analysis of District wise crop production

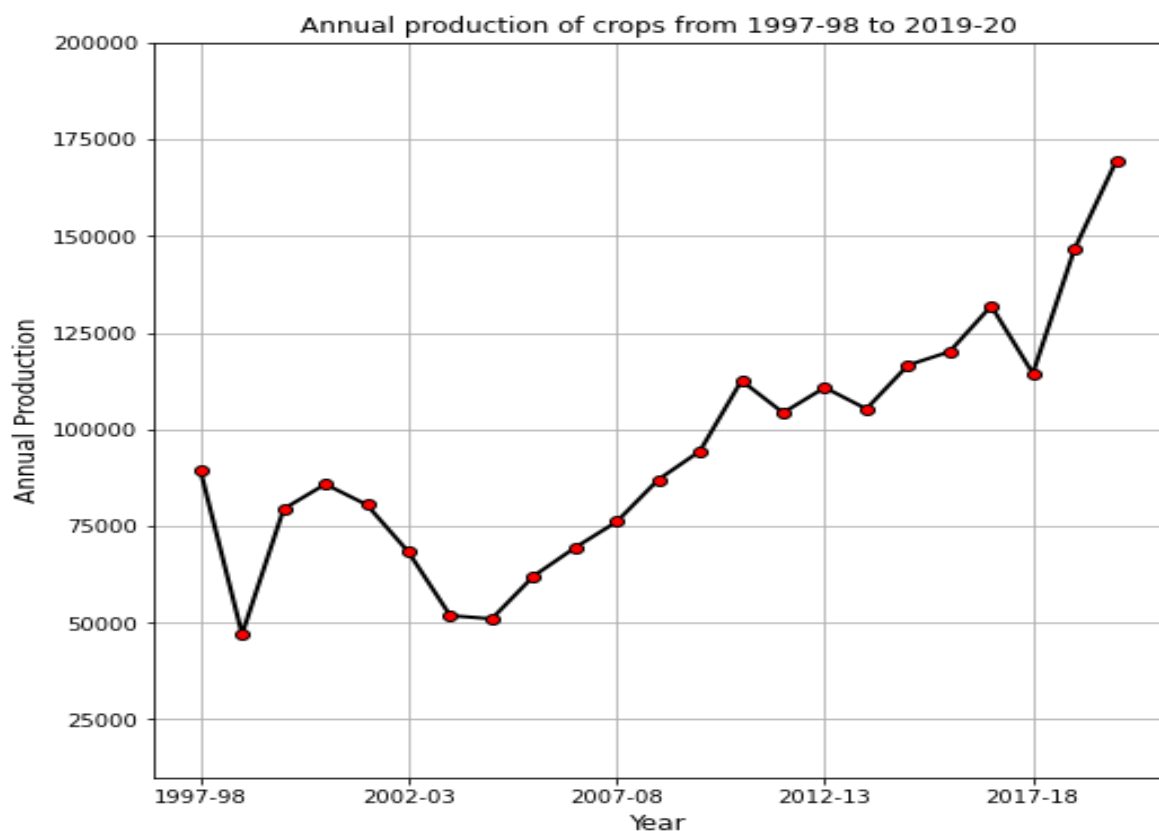


From the graph we can observe that Tumakuru district has the highest crop production and Chikballapur has lowest crop production in Karnataka. Tumakuru district has 4.561525e+08 tonnes of total crop production and Chikballapur district has 1.265161e+07 tonnes of total crop production from 1997-98 to 2019-20.

3.1.5 Analysis of annual crop production from 1997-98 to 2019-20

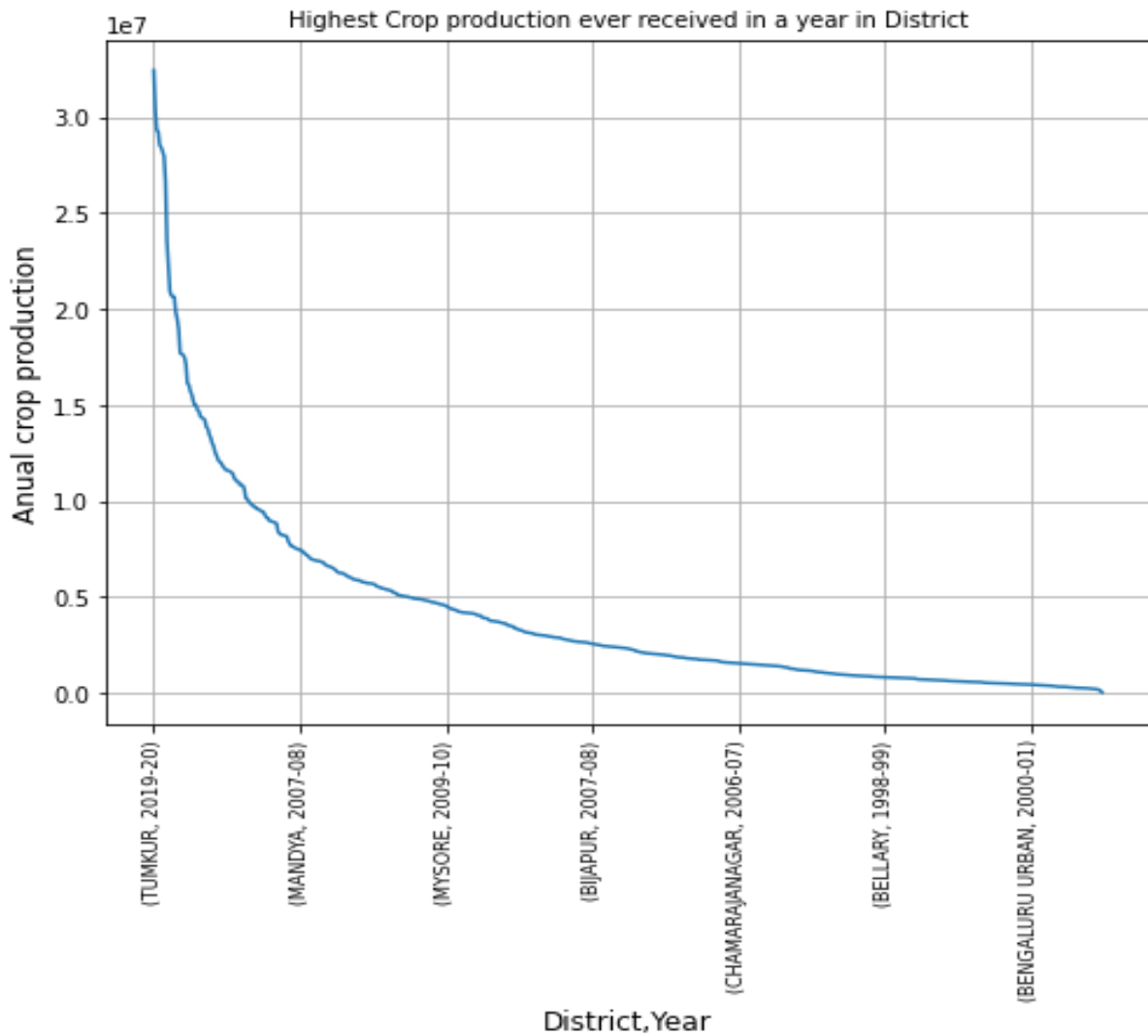
Year	Total production (Tonnes)
1998-99	47045.477428
2004-05	51009.836764
2003-04	51845.236267
2005-06	62031.937067
2002-03	68250.732845
2006-07	69417.916913
2007-08	76104.534211
1999-00	79432.289469
2001-02	80483.104981
2000-01	85835.849276
2008-09	87012.385565
1997-98	89330.874704
2009-10	94379.888290

2011-12	104298.302944
2013-14	105291.618471
2012-13	110827.720734
2010-11	112616.523280
2017-18	114475.347003
2014-15	116609.803567
2015-16	120097.536202
2016-17	131933.063375
2018-19	146595.916606
2019-20	169330.944245



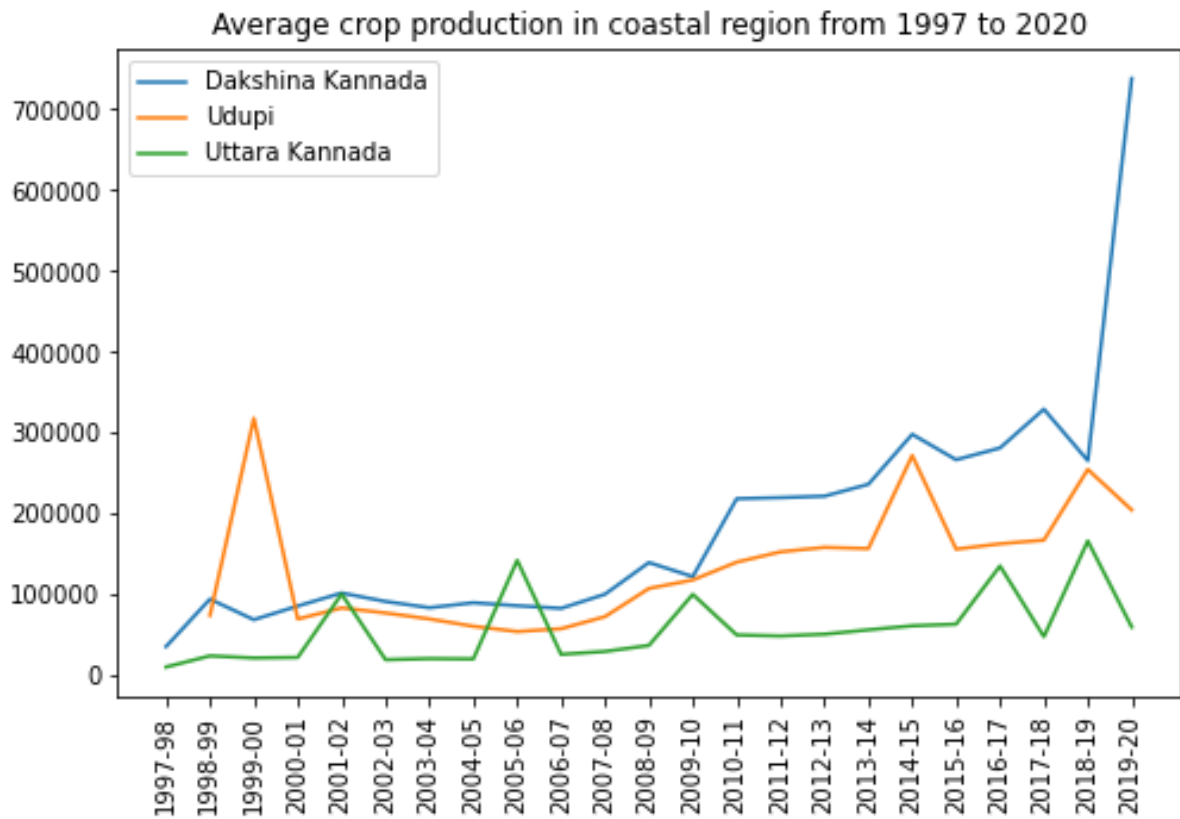
The graph above depicts Highest crop production recorded in year 2019-2020 and lowest crop production recorded in the year 1998-99

3.1.6 Highest Crop production ever received in a year in District

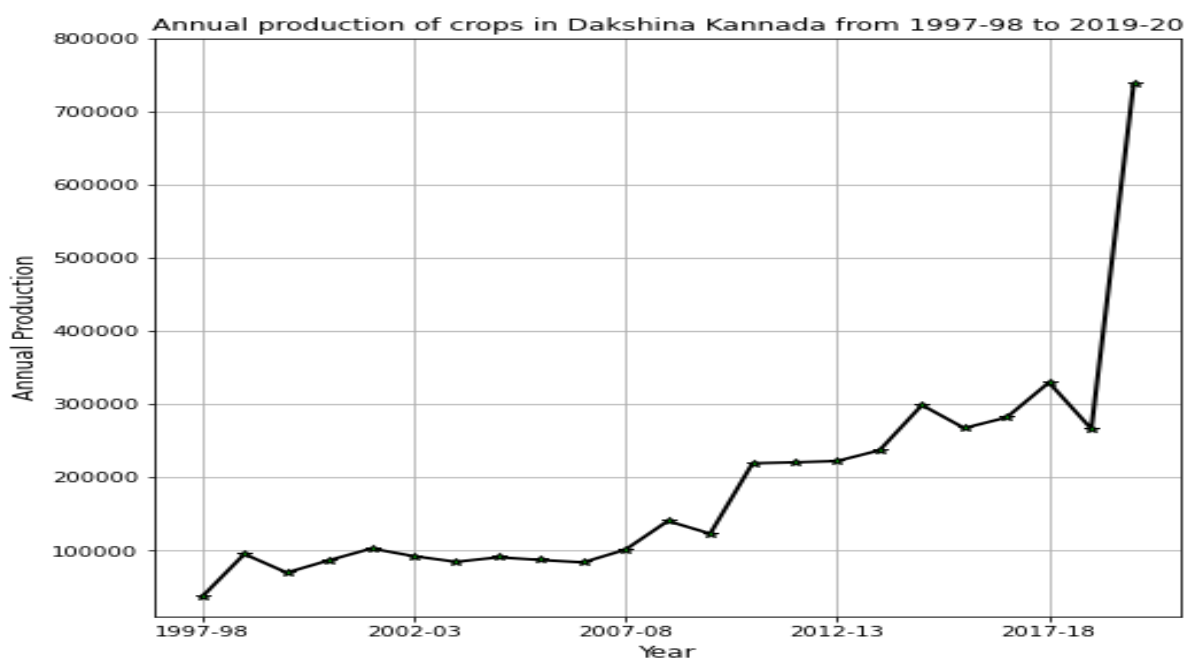


From the graph we analysed Tumakur is the district in the year 2019-20 produced highest ever crop in Karnataka. Total 3.244502e+07 tonnes of crops produced in 2019-20 in Tumakur district.

3.1.7 Analysis of Crop production in coastal region.

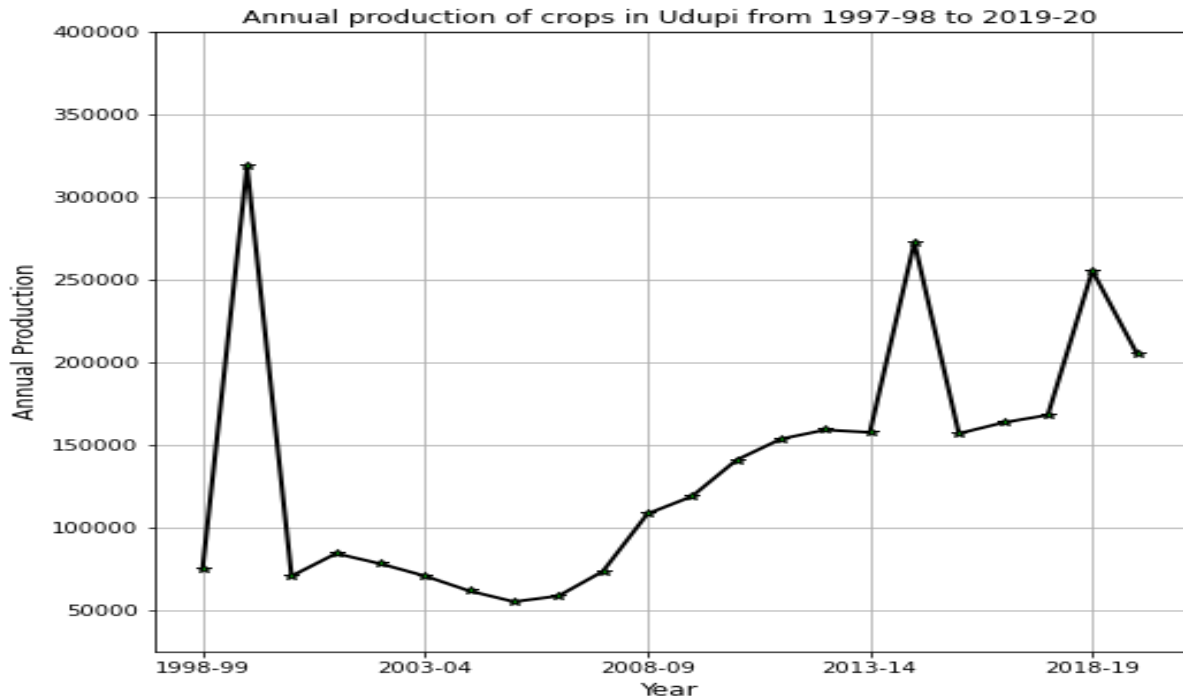


3.1.7.1 Annual production of crops in Dakshina Kannada



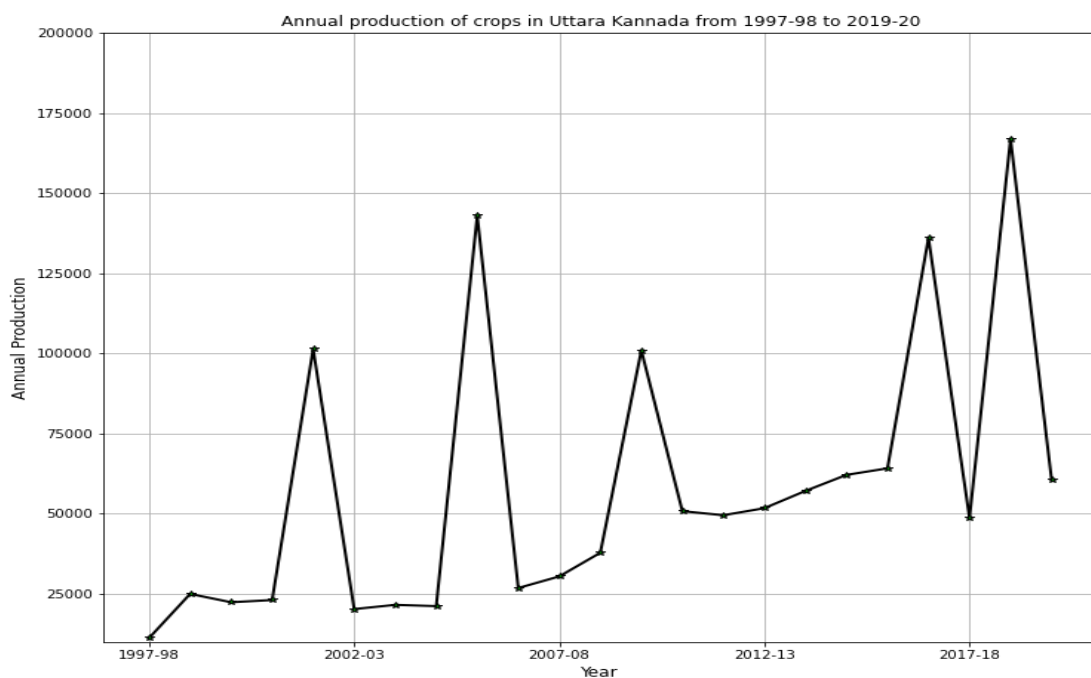
The graph above depicts Highest crop production recorded in year 2019-20 and lowest crop production recorded in the year 1997-98 in Dakshina Kannada district.

3.1.7.2 Annual production of crops in Udupi



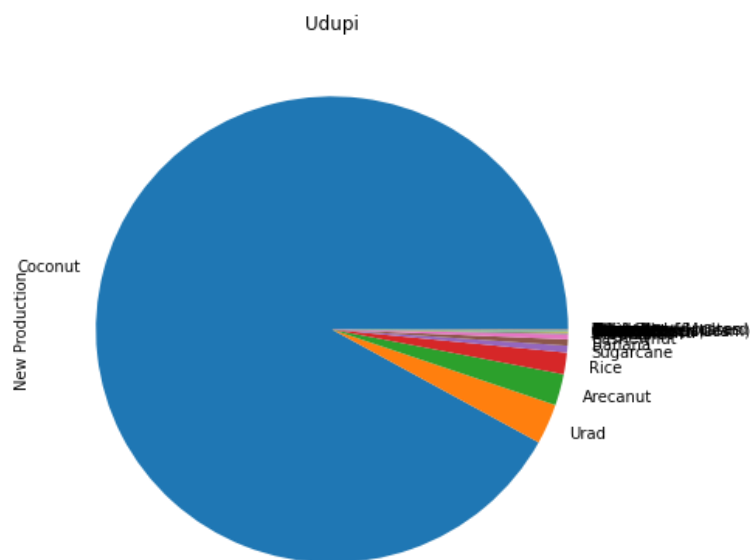
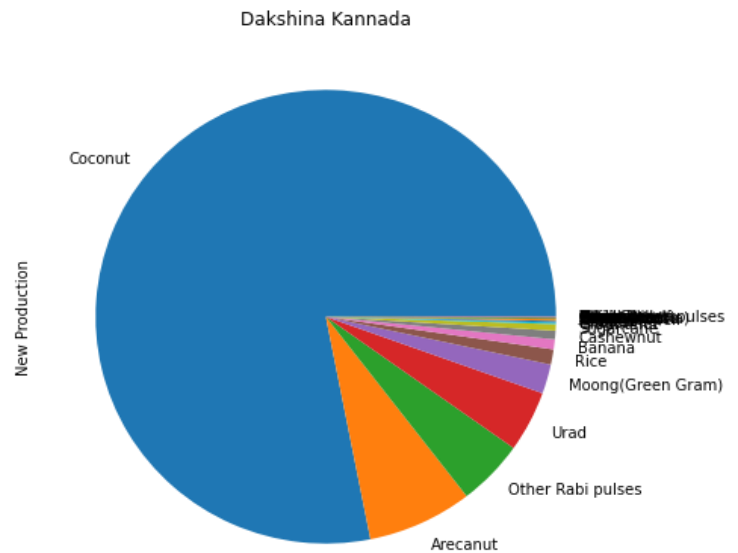
The graph above depicts Highest crop production recorded in year 1999-00 and lowest crop production recorded in the year 2005-06 in Udupi district.

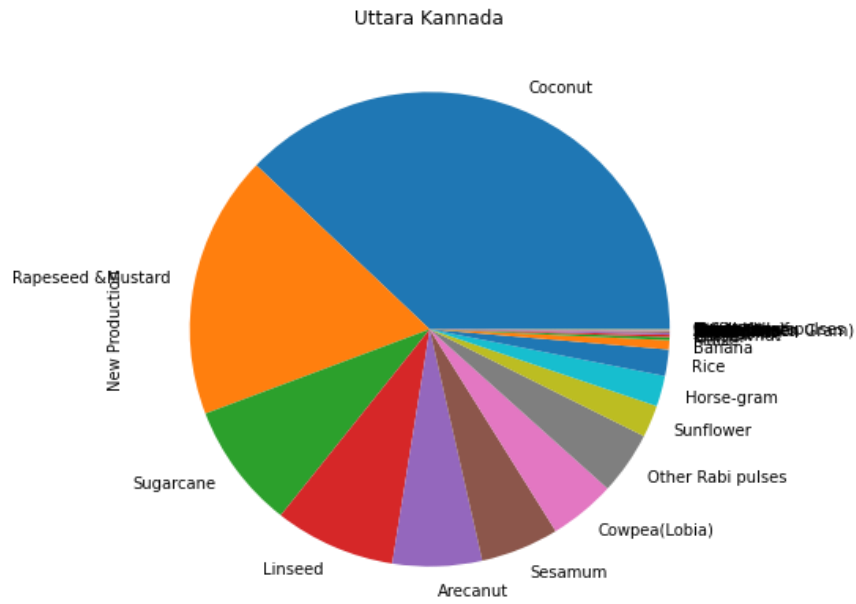
3.1.7.3 Annual production of crops in Uttara Kannada



The graph above depicts Highest crop production recorded in year 2018-19 and lowest crop production recorded in the year 1997-98 in Uttara Kannada district

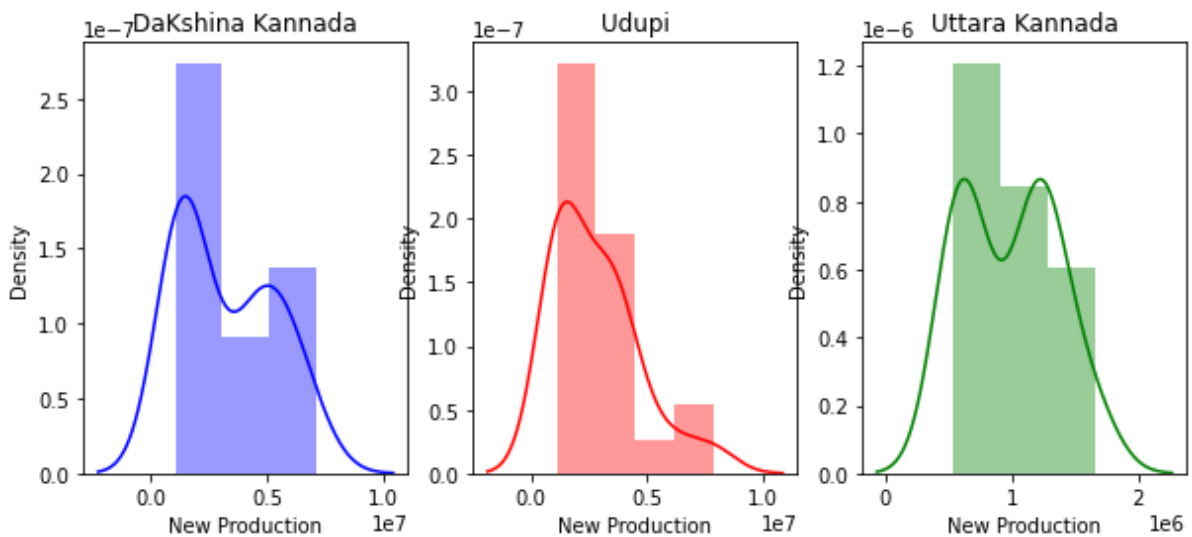
3.1.7.4 The most commom growing crop in costal region of Karnataka.





The graphs above depicts that Coconut is the most common growing crop in Coastal region.

3.1.7.5 Analysing distribution of Total production of ‘Coconut’ crop in Coastal region of Karnataka



From the graph, we analysed that Total production of ‘coconut’ crops in coastal region of Karnataka is almost normally distributed.

3.1.7.6 Testing the significance difference between mean production of ‘Coconut’ in districts of Coastal region of Karnataka

Hypothesis:

H_0 : There is no significance difference between mean production of "Arecanut" in district of Coastal region of Karnataka

H_1 : There is significance difference between mean production of ‘Coconut’ in district of Coastal region of Karnataka

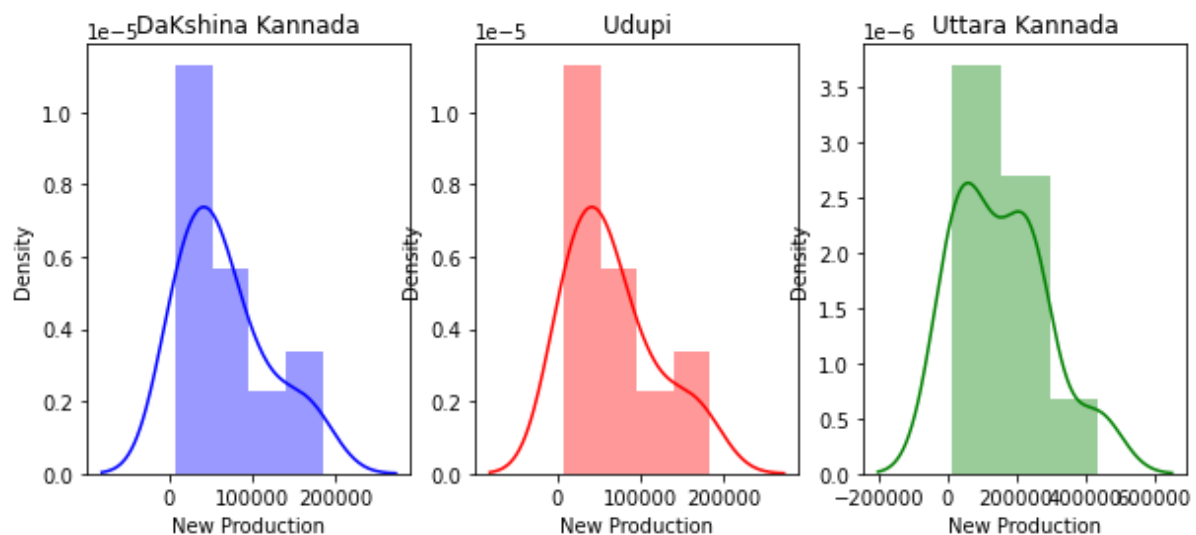
F value is : 12.016677147895756

p value is : 3.794259883400737e-05

We reject the null hypothesis if p value<0.05. Hence we reject null hypothesis

There is significance difference between mean production of "Coconut" in districts of Coastal region of Karnataka

3.1.7.7 Analysing distribution of Total production of ‘Areca nut’ crop in Coastal region of Karnataka



From the graph, we analysed that Total production of ‘Arecanut’ crops in coastal region of Karnataka is almost normally distributed.

3.1.7.8 Testing the significance difference between mean production of 'Arecanut' in district of Coastal region of Karnataka

Hypothesis:

H_0 : There is no significance difference between mean production of 'Arecanut' in district of Coastal region of Karnataka

H_1 : There is significance difference between mean production of 'Arecanut' in district of Coastal region of Karnataka

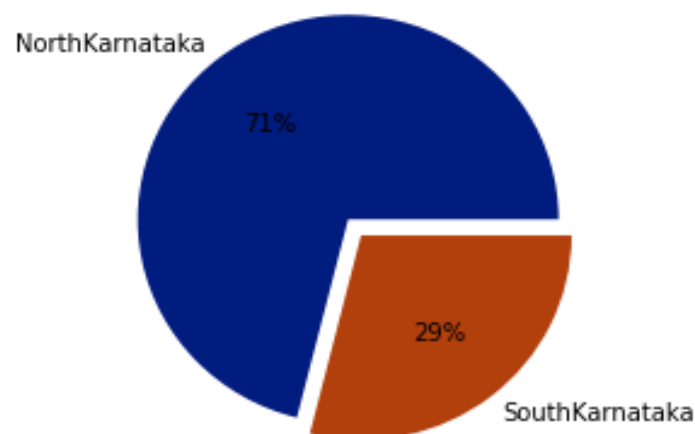
F value is : 7.206816178603528

p value is : 0.0016018511703318472

We reject the null hypothesis if p value < 0.05. Hence we reject null hypothesis

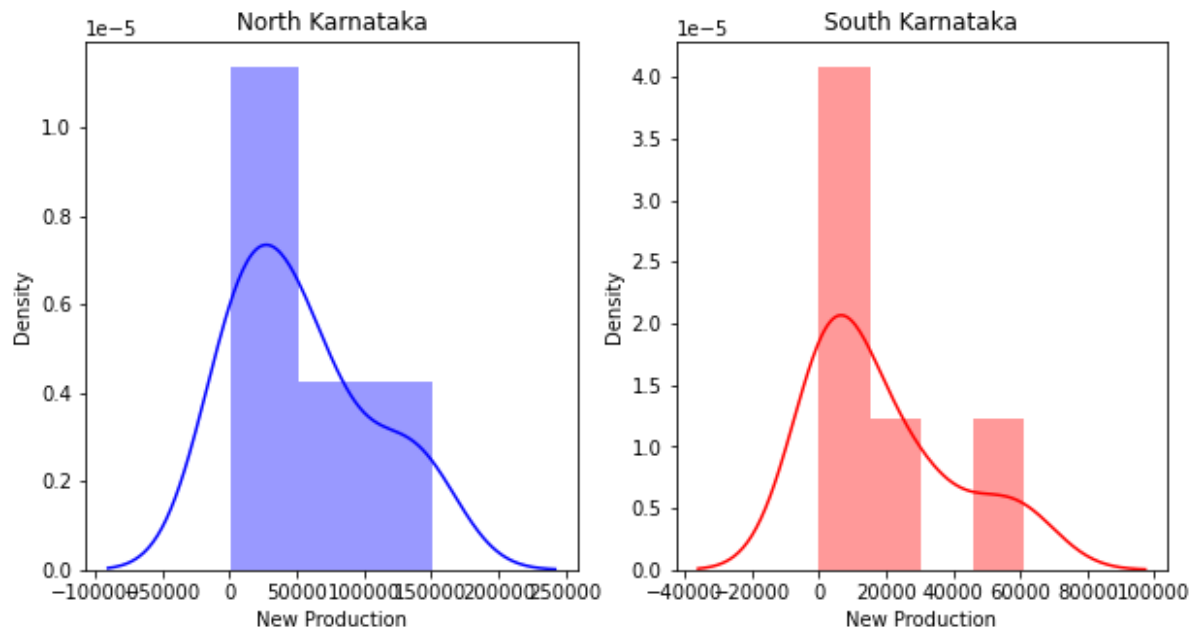
There is significance difference between mean production of "Arecanut" in district of Coastal region of Karnataka

3.1.8 Analysis of crop production in North Karnataka and South Karnataka



From the graph we analysed that North Karnataka is the most agriculture depending region of Karnataka.

3.1.8.1 Analyzing distribution of Total mean production of 'Maize' in North Karnataka and South Karnataka



The graph above depicts Total production of 'Maize' crops in South Karnataka and North Karnataka is almost normally distributed.

3.1.8.2 Testing Significance difference between mean production of "Maize" in North Karnataka and South Karnataka

Hypothesis:

H_0 : There is no significance difference between mean production of "Maize" in North Karnataka and South Karnataka

H_1 : There is significance difference between mean production of "Maize" in North Karnataka and South Karnataka

statistic=2.528

p value =0.017

We reject null hypothesis if p value < 0.05. Hence we reject null hypothesis.

There is significance difference between mean production of "Maize" in North Karnataka and South Karnataka.

4 Chapter 4

Conclusion

4.1 Conclusion

Following are the overall conclusion:

- The whole year season have maximum crop production in Karnataka.
- The ‘Maize’ is the most growing and ‘Mesta’ is the least growing crop in Karnataka from 1997-2020.
- Highest crop production recorded in year 2019-2020 and lowest crop production recorded in the year 1998-99
- Tumakur is the district in the year 2019-20 produced highest ever crop in Karnataka. Total 3.244502e+07 tonnes of crops produced in 2019-20 in Tumakur district.
- ‘Coconut’ is the most growing crop in coastal region of Karnataka.
 - i. In Dakshina Kannada, highest crop production recorded in year 2019-2020 and lowest crop production recorded in the year 1997-98.
 - ii. In Udupi, highest crop production recorded in year 1999-00 and lowest crop production recorded in the year 2005-06.
 - iii. In Uttara Kannada, highest crop production recorded in year 2018-19 and lowest crop production recorded in the year 1997-98.
- There is difference in mean production of "Coconut" and “Arecanut” crop in districts of Coastal region of Karnataka.
- North Karnataka is the most agriculture depending area of Karnataka
- There is difference in mean production of "Maize" in North Karnataka and South Karnataka.

5 Chapter 5

5.1 Summary

A project entitled “Analysis of Karnataka crop production from 1997-2020” has been done. A secondary data has been collected from the website of “Kaggle”. The data contained features like State, district, crop, year, season, area units, production, production units, yield . The data consists of 9 columns and 27493 records. The analysis and interpretation of the data is done by using some of the statistical methods like two sample t-test, one way anova.

From the results we came to know whole year season is the best season for crop production and Maize is the most growing crop in Karnataka. Tumakuru is the most agriculture depending district in Karnataka. We got to know that 2019-2020 is the year produced highest crop and 1998-99 is the year produced lowest crop in Karnataka. Tumakur is the district in the year 2019-20 produced highest ever crop in Karnataka. Total 3.244502e+07 tonnes of crops produced in 2019-20 in Tumakur district.

Dakshina Kannada ,Udupi and Uttara Kannada are the coastal region of Karnataka, where Coconut was the highest growing crop. In Dakshina Kannada District 2019-2020 was the year recorded highest crop production and 1997-98 was lowest crop production similarly in Udupi 1999-00 was the year recorded highest and 2005-06 was the year lowest crop production. Also in Uttara Kannada 2018-19 was the year has highest crop production and 1997-98 was the lowest crop production. We got to know that there is difference in production of Coconut and also Arecanut crop in coastal region.

In Karnataka, North Karnataka is most agriculture depending region and also there is a difference in production of Maize in North Karnataka and South Karnataka.

5 Chapter 5

Bibliography

- <https://www.kaggle.com/datasets/pyatakov/india-agriculture-crop-production>
- <https://www.analyticsvidhya.com/blog/2021/07/t-test-performing-hypothesis-testing-with-python/>
- <https://www.statology.org/one-way-anova-python/>

6 Chapter 6

Appendix

Python code for analysis

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Read csv file

```
df=pd.read_csv("AgricultureKar.csv")
df.head()
```

Best season for crop production

```
sns.barplot(x='Season',y='New Production',data=df)
```

The most common growing crop in Karnataka for agriculture.

```
df["Crop"].value_counts()
```

Analysis of highest and lowest yield in an year

```
plt.figure(figsize=(12,8))
ax=sns.barplot(y=df["Year"],x=df["Yield"])
plt.setp(ax.get_xticklabels(),fontsize=20)
plt.tight_layout()
plt.show()
```


Analysis of District wise crop production

```
plt.figure(figsize=(12,8))

ax=sns.barplot(y=total_production_df["District"],x=total_production_df["Total crop
production"])

plt.setp(ax.get_xticklabels(),fontsize=20)

plt.tight_layout()

plt.show()
```

Annual crop production from 1997-98 to 2019-20

```
df.groupby("Year").mean()["New
Production"].plot(ylim=(10000,200000),color='k',marker='o',

                    markerfacecolor='red',linestyle='-',linewidth=2,figsize=(8,8))

plt.xlabel('Year',fontsize=12)

plt.ylabel("Annual Production",fontsize=12)

plt.title('Annual production of crops from 1997-98 to 2019-20',fontsize=12)

plt.grid()
```

Highest Crop production ever received in a year in District

```
plt.figure(figsize=(8,6))

df.groupby(["District","Year"])[["New
Production"].sum().sort_values(ascending=False).plot()

plt.grid()

plt.xticks(rotation='vertical',size=8)

plt.xlabel("District,Year",fontsize=12)

plt.ylabel("Anual crop production",fontsize=12)

plt.title("Highest Crop production ever received in a year in District",fontsize=10)
```

Analysis of Crop production in coastal region.

```
plt.figure(figsize=(8,5))

plt.plot(dk_1["Year"],dk_1["Average Production DK"],label="Dakshina Kannada")

plt.plot(udp_1["Year"],udp_1["Average Production Udupi"],label="Udupi")

plt.plot(uk_1["Year"],uk_1["Average Production UK"],label="Uttara Kannada")

plt.xticks(rotation='vertical',size=10)

plt.title("Average crop production in coastal region from 2010 to 2020")

plt.legend(loc='upper left')

plt.show
```

Annual production of crops in Dakshina Kannada

```
dk.groupby("Year").mean()["New
Production"].plot(ylim=(10000,800000),color='k',marker='*',

                    markerfacecolor='green',linestyle='-',
                    linewidth=2,figsize=(8,8))

plt.xlabel('Year',fontsize=12)

plt.ylabel("Annual Production",fontsize=12)

plt.title('Annual production of crops in Dakshina Kannada from 1997-98 to 2019-
20',fontsize=12)

plt.grid()
```

Annual production of crops in Udupi

```
udp.groupby("Year").mean()["New
Production"].plot(ylim=(25000,400000),color='k',marker='*',

                    markerfacecolor='green',linestyle='-',
                    linewidth=2,figsize=(8,8))

plt.xlabel('Year',fontsize=12)

plt.ylabel("Annual Production",fontsize=12)
```

```
plt.title('Annual production of crops in Udupi from 1997-98 to 2019-20',fontsize=12)
```

```
plt.grid()
```

Annual production of crops in Uttara Kannada

```
uk.groupby("Year").mean()["New  
Production"].plot(ylim=(10000,200000),color='k',marker='*',  
  
markerfacecolor='green',linestyle='-  
,linewidth=2,figsize=(12,10))
```

```
plt.xlabel('Year',fontsize=12)
```

```
plt.ylabel("Annual Production",fontsize=12)
```

```
plt.title('Annual production of crops in Uttara Kannada from 1997-98 to 2019-  
20',fontsize=12)
```

```
plt.grid()
```

The most common growing crop in coastal region for agriculture

```
dk.groupby("Crop")["New  
Production"].mean().sort_values(ascending=False).plot(kind='pie',stacked=True,figsize=(15,  
7))
```

```
plt.title('Dakshina Kannada',fontsize=12)
```

```
plt.show()
```

```
udp.groupby("Crop")["New  
Production"].mean().sort_values(ascending=False).plot(kind='pie',stacked=True,figsize=(15,  
7))
```

```
plt.title('Udupi',fontsize=12)
```

```
plt.show()
```

```
uk.groupby("Crop")["New  
Production"].mean().sort_values(ascending=False).plot(kind='pie',stacked=True,figsize=(15,  
7))
```

```
plt.title('Uttara Kannada',fontsize=12)
```

```
plt.show()
```

Testing the significance difference between mean production of 'Coconut' in districts of Coastal region of Karnataka

```
import scipy.stats as stats

data_1=dk_coconut["New Production"]

data_2=udp_coconut["New Production"]

data_3=uk_coconut["New Production"]

f_value,p_value=stats.f_oneway(data_1,data_2,data_3)

print("F value is : ",f_value)

print("p value is : ",p_value)
```

Analysis of crop production in North Karnataka and South Karnataka

```
part=["NorthKarnataka","SouthKarnataka"]

area=[183335667,74571732]

explode=[0.1,0]

palette_color=seaborn.color_palette('dark')

plt.pie(area,labels=part,colors=palette_color,

        explode=explode,autopct='%.0f%%')

plt.show()
```

Testing Significance difference between mean production of "Maize" in North Karnataka and South Karnataka

```
from scipy.stats import ttest_ind

b1=north_1["New Production"]

b2=south_1["New Production"]

stat,p=ttest_ind(b1,b2)

print('statistic=%.3f , p_value =%.3f % (stat ,p))
```