

# ***EXPLORATORY ANALYSIS OF AGRICULTURE IN MAHARASHTRA***

**A project report submitted in partial fulfillment of the requirements  
for the degree of Masters in Science (Computer Science)**

By  
***Yogesh Dhanraj Pokale(34354)***  
***Siddharth Manoj Chandurkar(34347)***

**Department of Computer Science  
M.Sc.(Computer Science) Part-I**

**Guide**

**Dr. Sushmita Bhambure**

**Sir Parashurambhau College (Autonomous),  
Pune.**

**2022-23**

Shikshana Prasarak Mandali's

Accredited with A+ Grade by NAAC

**SIR PARASHURAMBHAU COLLEGE**  
**(AUTONOMOUS)**

TILAK ROAD, PUNE – 411 030.

Department of Computer Science

# *Certificate*

*This is to certify that \_\_\_\_\_ has  
presented a mini project titled \_\_\_\_\_  
in partial fulfillment of the requirements of M.Sc.  
(Computer Science) Part I (semester I) course.*

*Date:    /    /*

*Teacher In-Charge*

*Head,*

*Department of Computer Science*

Shikshana Prasarak Mandali's

Accredited with A+ Grade by NAAC

**SIR PARASHURAMBHAU COLLEGE**  
**(AUTONOMOUS)**

TILAK ROAD, PUNE – 411 030.

Department of Computer Science

## ***Certificate***

*This is to certify that \_\_\_\_\_ has  
presented a mini project titled \_\_\_\_\_  
in partial fulfillment of the requirements of M.Sc.  
(Computer Science) Part I (semester I) course.*

*Date:    /    /*

*Teacher In-Charge*

*Head,*

*Department of Computer Science*

## **Acknowledgement**

The success and final outcome of this project required a lot of guidance and assistance from many people and we are extremely privileged to have got this all along the completion of our project. All that we have done is only due to such supervision and assistance and we would not forget to thank them.

We owe our deep gratitude to our project guide Mrs.SUSHMITA BHAMBURE, who took keen interest on our project work and guided us all along, till the completion of our project work by providing all the necessary information for developing a good system.

We are thankful to and fortunate enough to get constant encouragement, support and guidance from all Teaching staffs of Computer Science Department, S.P. College. which helped us in successfully completing our project work. Also, We would like to extend our sincere esteems to all staff in laboratory for their timely support.

## **Index:**

1. Description of the data
2. Data observation techniques
3. Techniques used for preprocessing data
4. Python Code
5. Data Visualization and Interpretation
6. Bibliography

## *DESCRIPTION OF DATA*

Agriculture is one of the dominant sectors of economy in Maharashtra. This sector contributes 13.2% of the total GDP of the state and demands 51% of total labour force by occupation.

About 82% of the rural population in Maharashtra depends on agriculture for livelihood. Both food crops and cash crops are grown in the state.

Principal crops grown in the State are rice, bajra, wheat, tur, mung, urad, gram and other pulses. The State is a major producer of oilseeds. Groundnut, sunflower, soyabean are the major oil seed crops. The important cash crops are cotton, sugarcane, turmeric and vegetables. State is pioneer in onion production in the country. It is today emerging as an important horticultural State in the country. Different types of soil, diverse agro climatic conditions, adequate technical manpower, well developed communication facilities, increasing trend in drip irrigation, green house, use of cool chain facilities and vibrant farmer organisation offer wide opportunities to grow different horticultural crops in the State.

The following dataset contains yearly production of major crops in various districts of Maharashtra over the last 50 years from 1966 to 2016.

Columns:

- Year : respective year between 1966 to 2016 from which the data is based.
- Dist name: name of the subsequent district
- Production: total production of given crop with a unit taken per 1000 metric tonnes.
- Yield: Total yield of a given crop with a unit taken in kgs per hectare
- Area: Total area where the given crop was cultivated with a unit taken per 1000 hectares.

The following dataset is taken from :

<http://data.icrisat.org/dld/src/crops.html>

Using the given information we can understand and study the trends in various crops produced in various districts. We can also get to know which crop yield is more economically favourable in the subsequent districts.

## *DATA OBSERVATION TECHNIQUES*

In this project we are using Python language, Including following packages for visualization.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import time as tm
```

To read the data:

To read available dataset (.csv) , we first collect it in data variable and then convert it into dataframe using pandas

```
df = pd.read_csv("Crop.csv")
print(df)
```

Total numbers of rows and columns of data:

```
count_row = df.shape[0]
count_col = df.shape[1]
print("No of rows in data:-", count_row)
print("No of columns in data:-", count_col)
```



# *TECHNIQUES USED FOR PREPROCESSING DATA*

## **Cleaning:**

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

Steps involved in data cleaning:



- Removal of unwanted observations :

This includes deleting duplicate/ redundant or irrelevant values from your dataset. Duplicate observations most frequently arise during data collection and Irrelevant observations are those that don't actually fit the specific problem that you're trying to solve.

- Fixing Structural errors :

The errors that arise during measurement, transfer of data, or other similar situations are called structural errors. Structural errors include typos in the name of features, the same attribute with a different name, mislabeled classes, i.e. separate classes that should really be the same, or inconsistent capitalization.

- Managing Unwanted outliers :

Outliers can cause problems with certain types of models. Generally, we should not remove outliers until we have a legitimate reason to remove them. Sometimes, removing them improves performance, sometimes not.

- Handling missing data :

Missing data is a deceptively tricky issue in machine learning. We cannot just ignore or remove the missing observation. They must be handled carefully as they can be an indication of something important.

1. Substitution:

Impute the value from a new individual who was not selected to be in the sample.

2. Regression imputation:

The predicted value obtained by regressing the missing variable on other variables.

3. Imputation Using (Mean/Median) Values:

This works by calculating the mean/median of the non-missing values in a column and then replacing the missing values within each column separately and

independently from the others. It can only be used with numeric data.

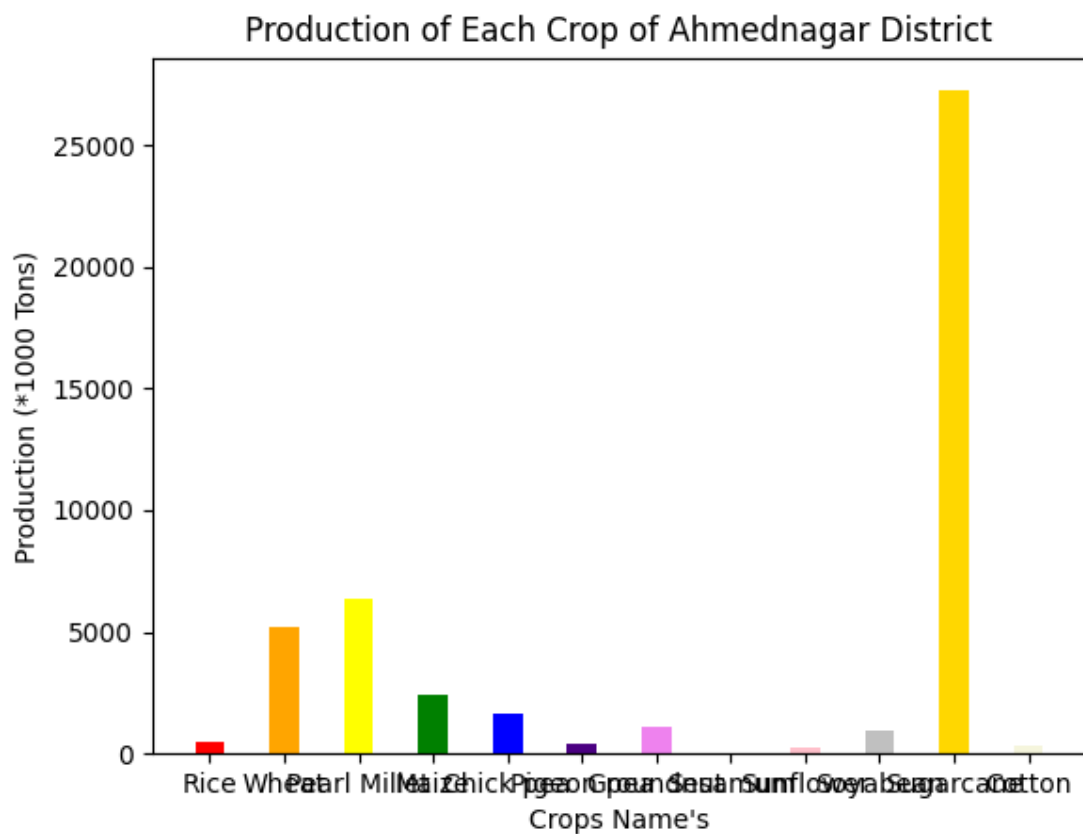
#### 4. Imputation Using (Most Frequent) or (Zero/Constant) Values:

Most Frequent is another statistical strategy to impute missing values and YES!! It works with categorical features (strings or numerical representations) by replacing missing data with the most frequent values within each column.

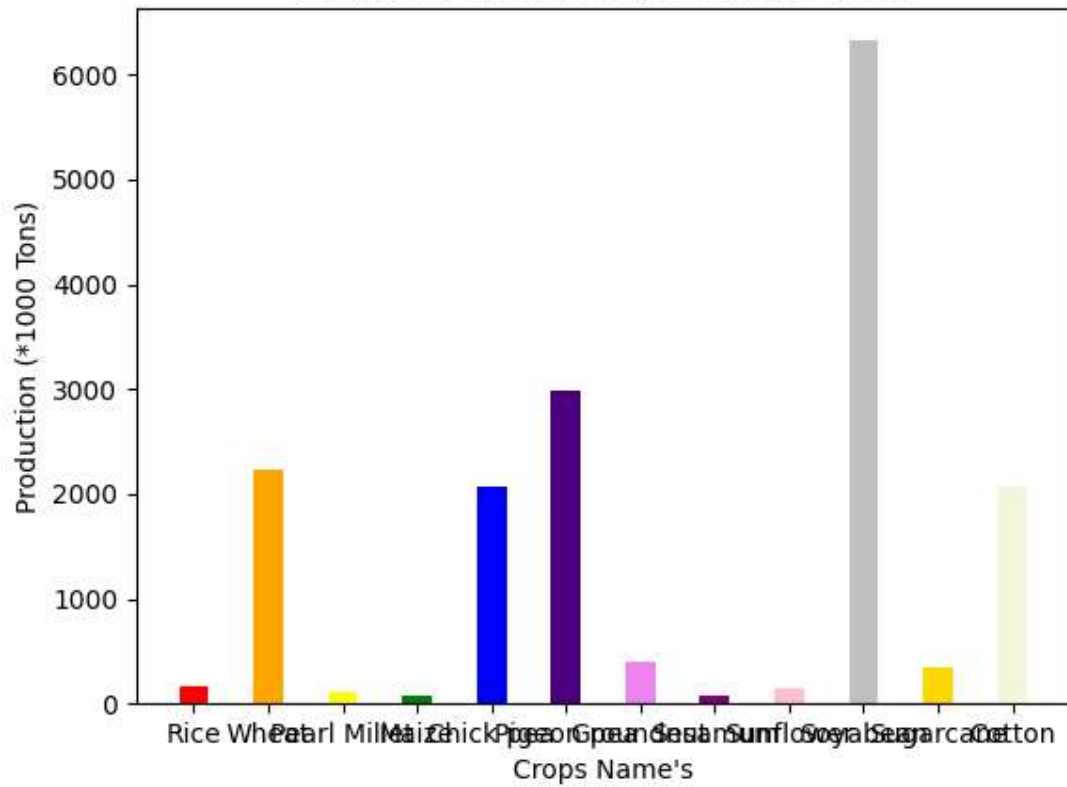
**PythonCode:-** <https://github.com/yogeshpokale9696/DEAP-Project.git>

## Data Visualization and Interpretation:-

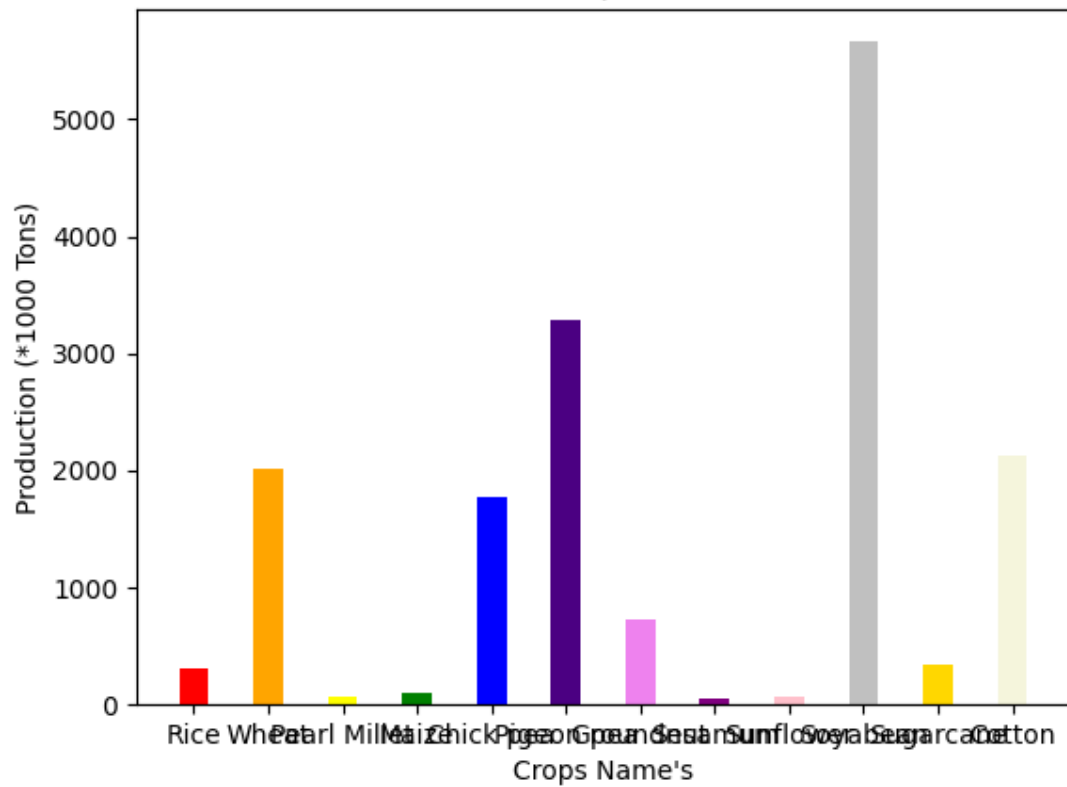
1. Bar Plot is one of the useful technique for data visualization, So we used Bar Plot for Showing Production of each Crop divided in Districts.

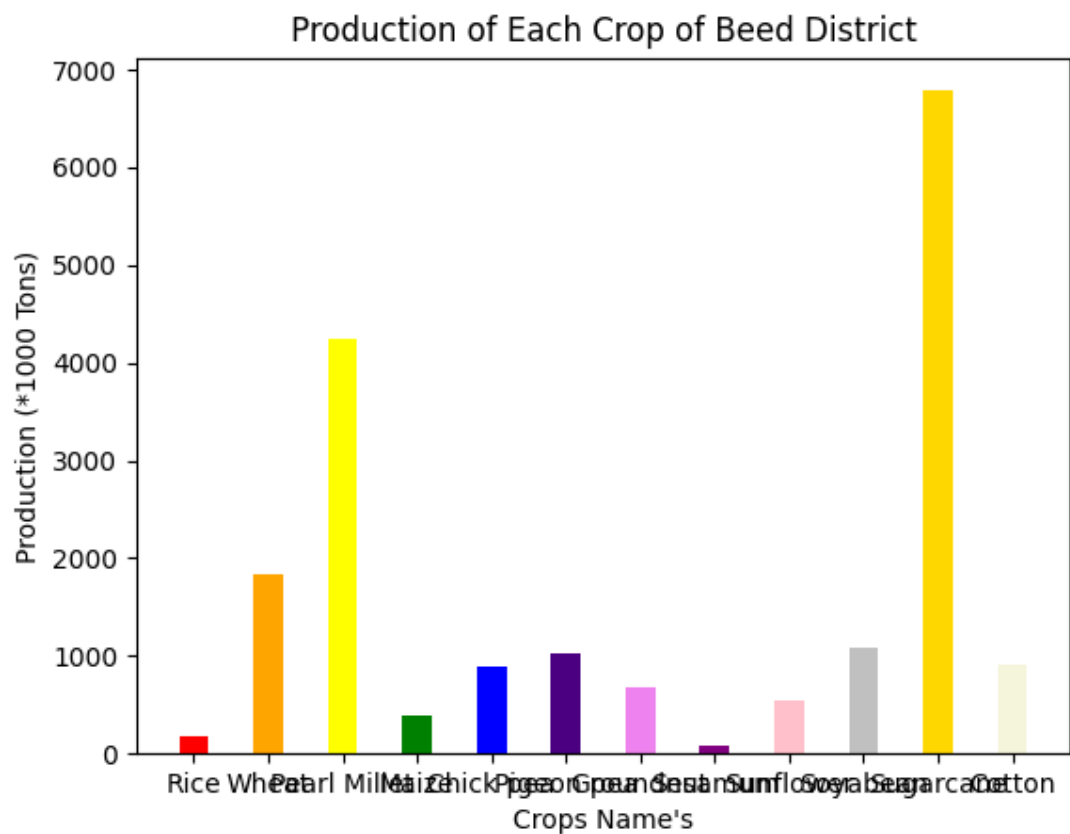
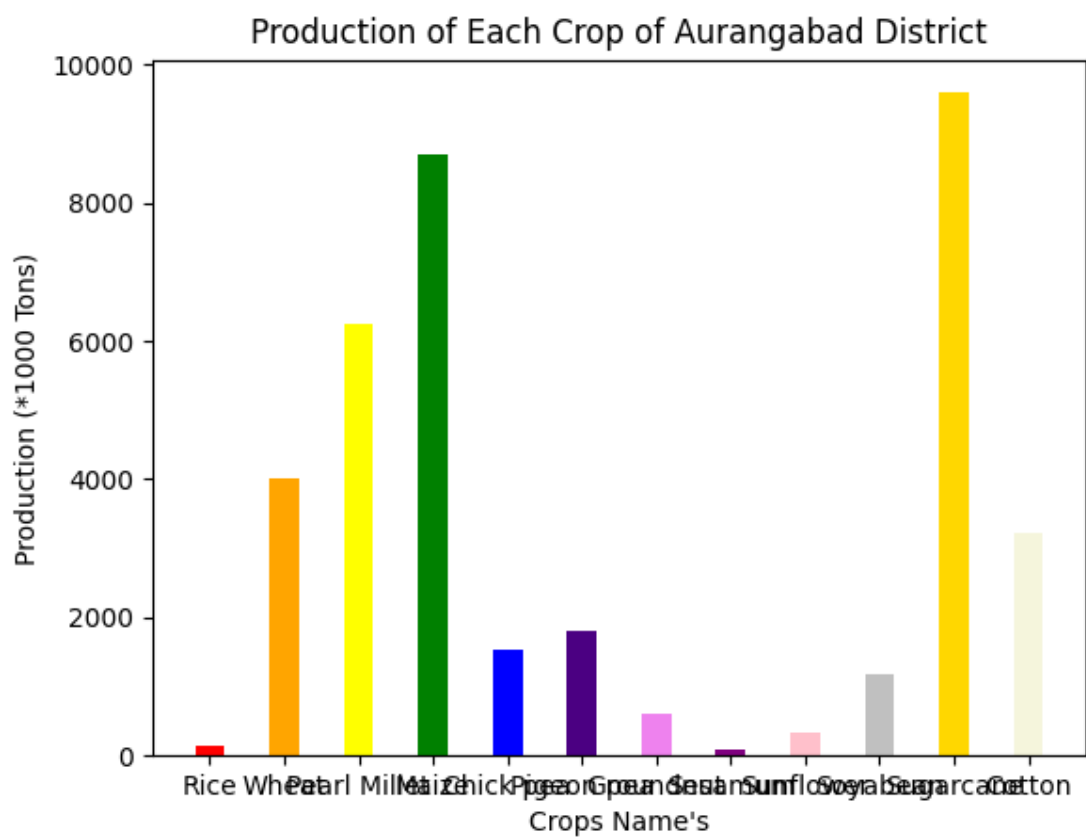


Production of Each Crop of Akola District

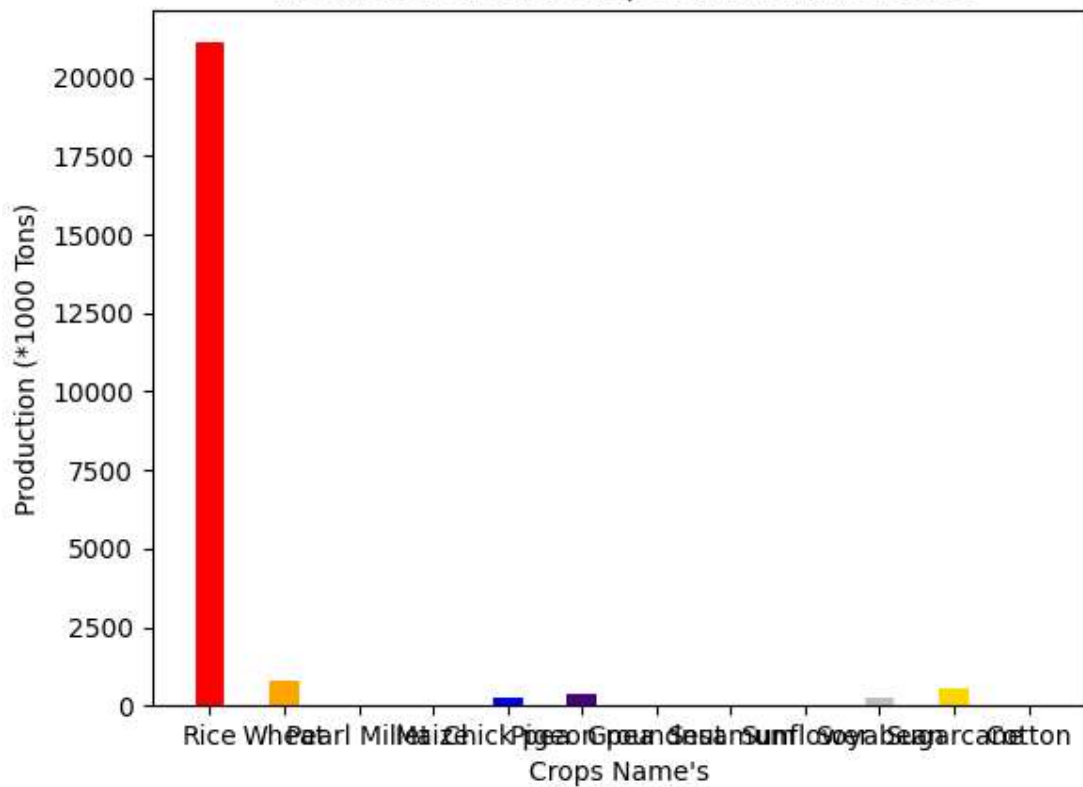


Production of Each Crop of Amarawati District

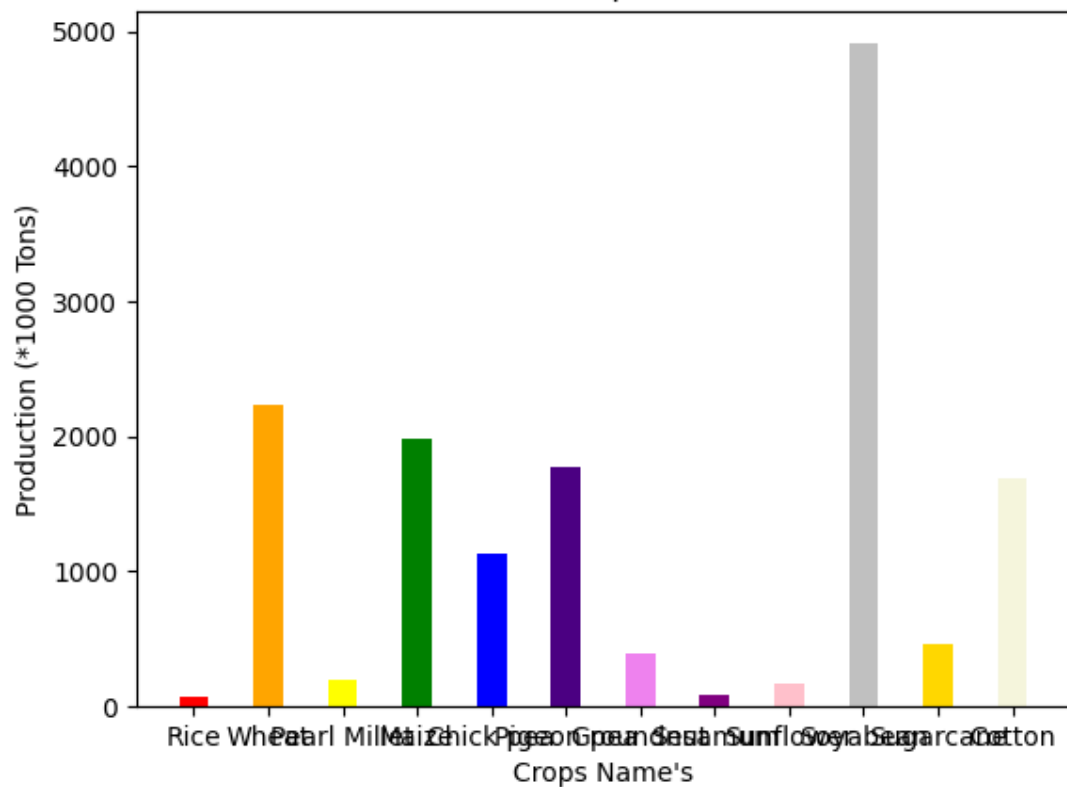




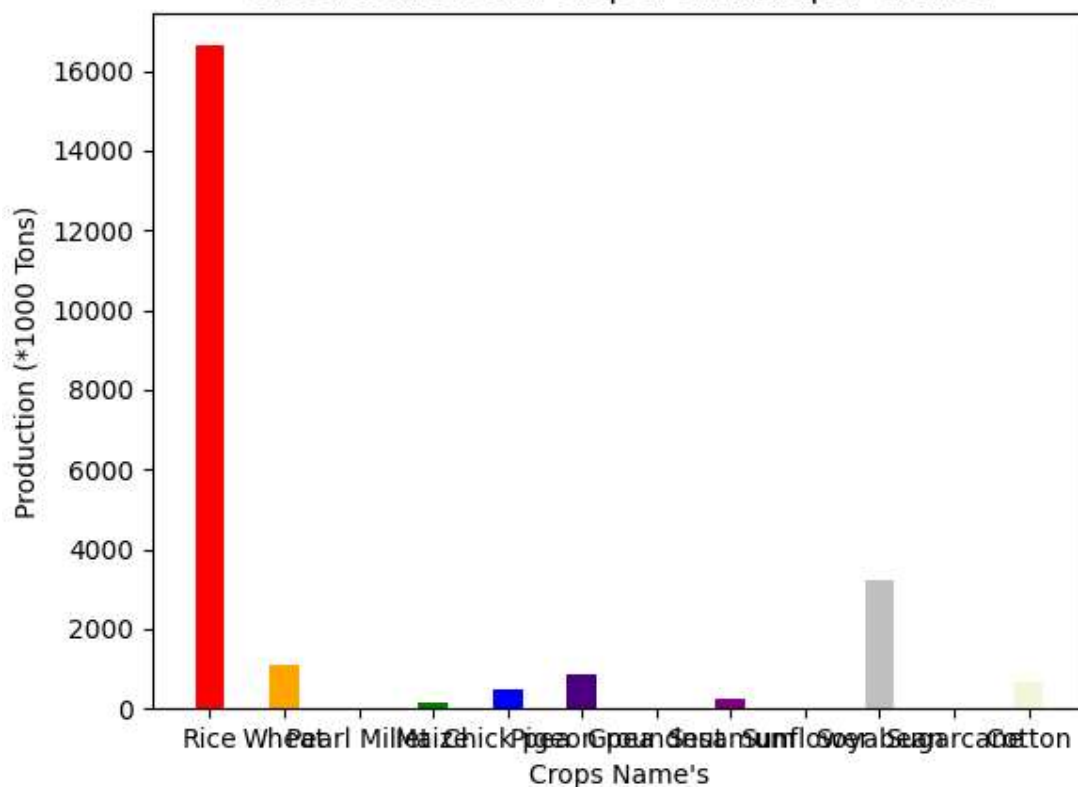
Production of Each Crop of Bhandara District



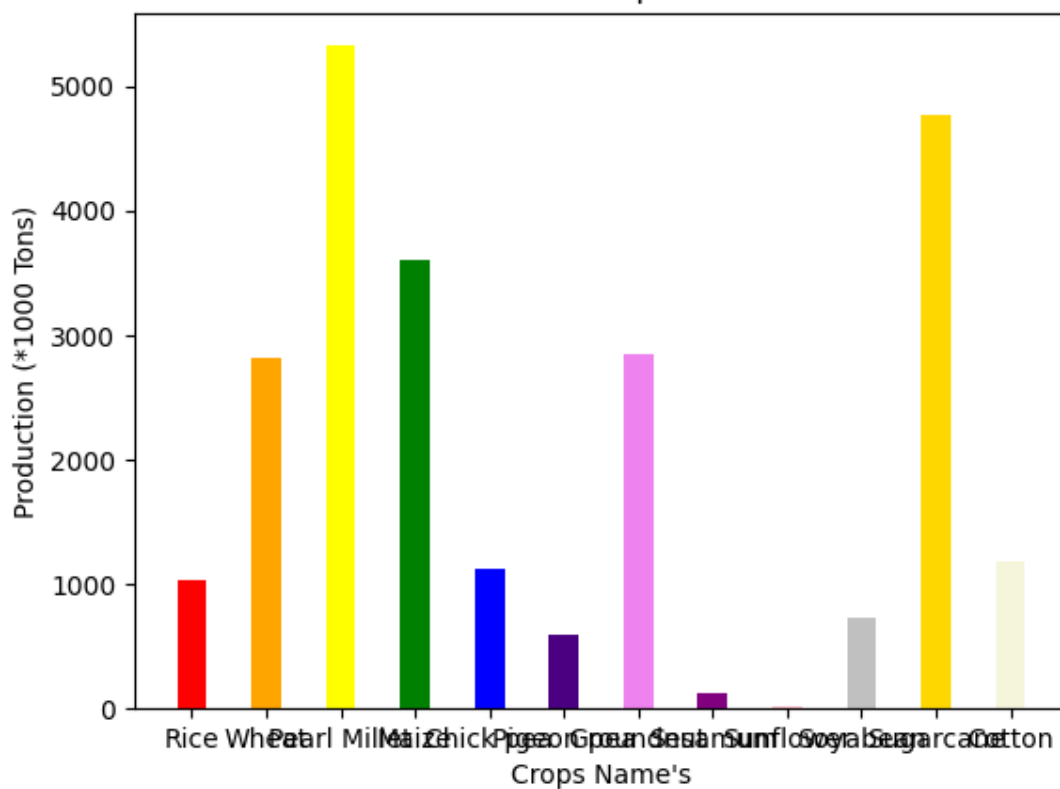
Production of Each Crop of Buldhana District



Production of Each Crop of Chandrapur District

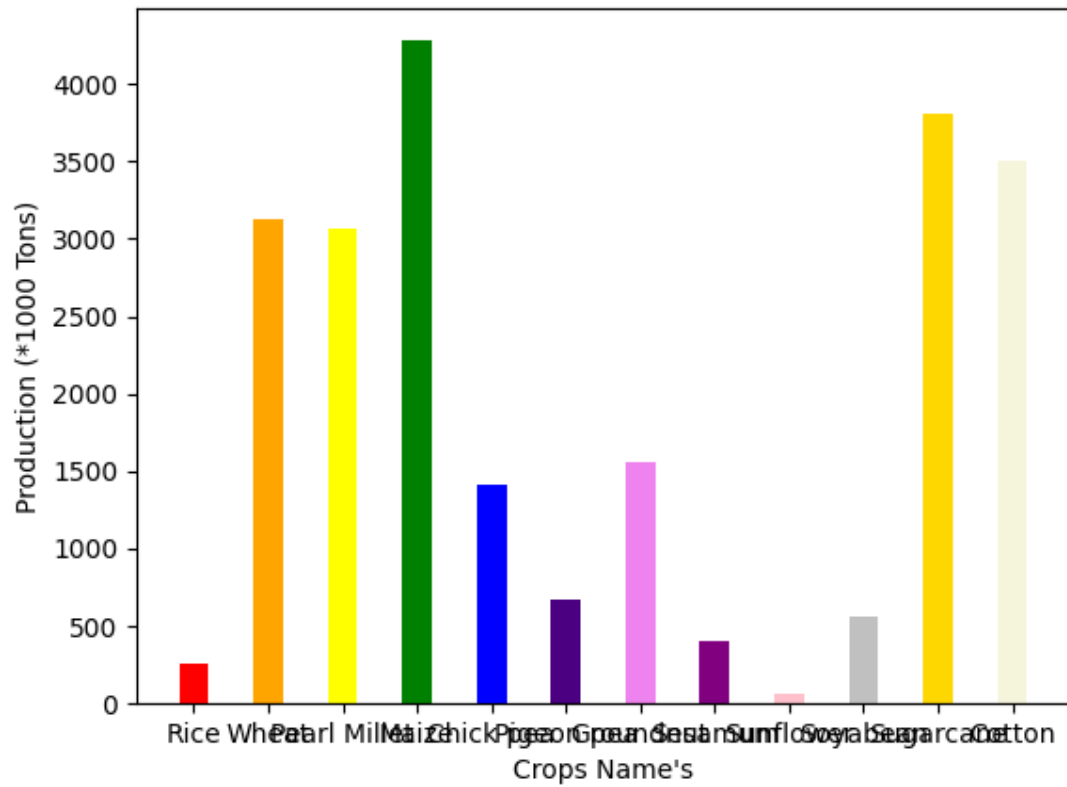


Production of Each Crop of Dhule District

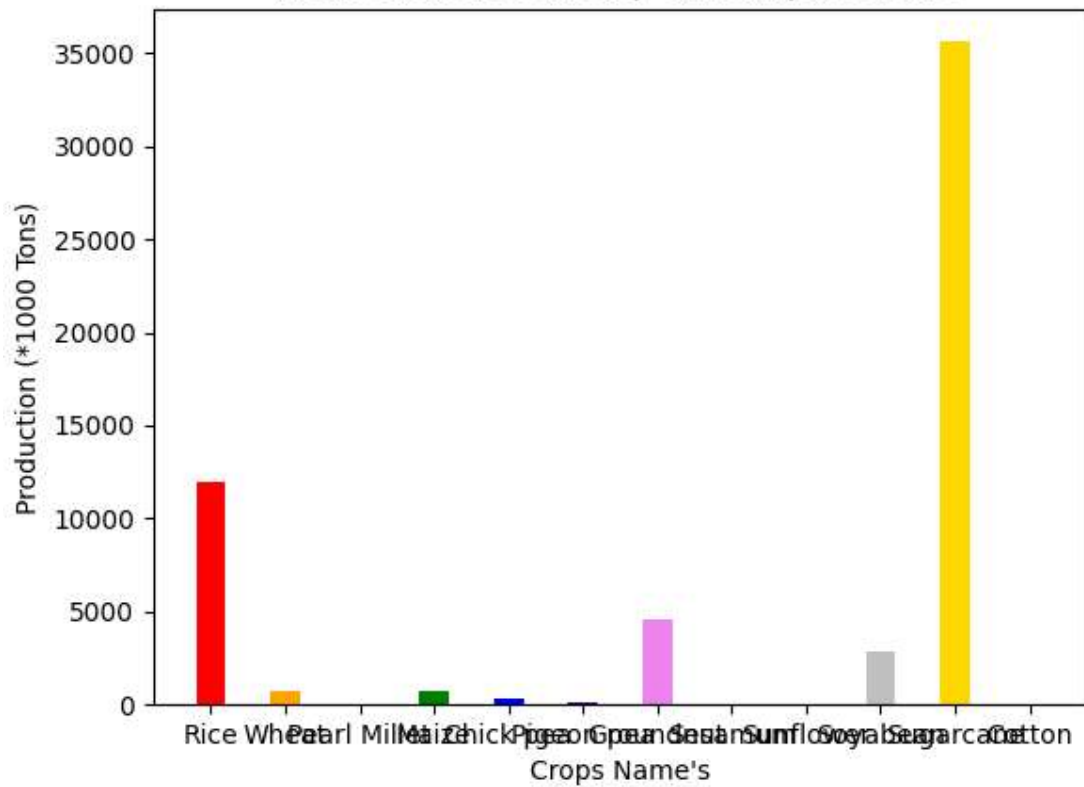




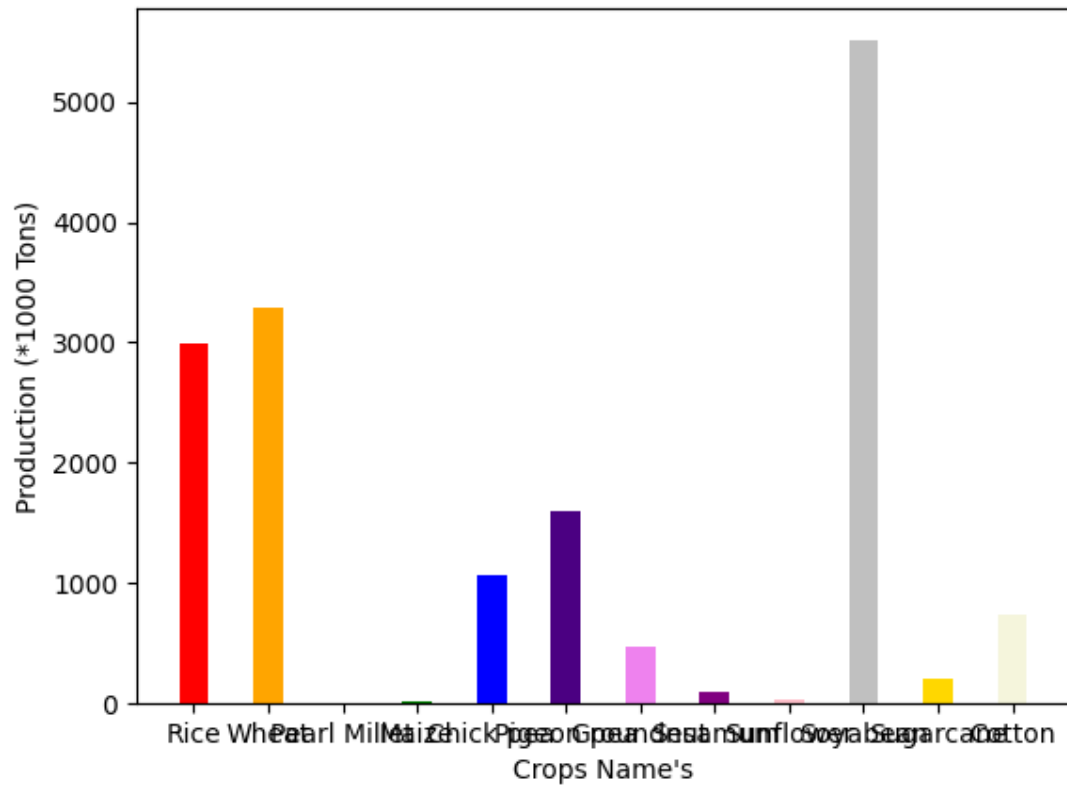
Production of Each Crop of Jalgaon District



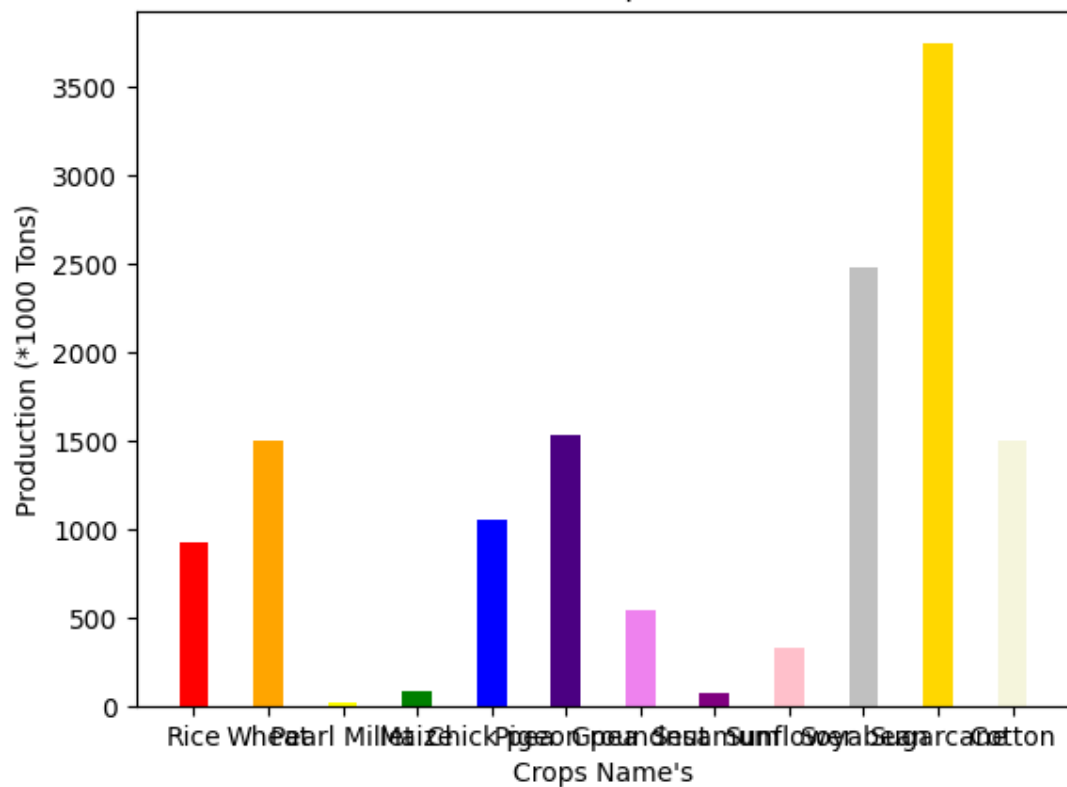
Production of Each Crop of Kolhapur District



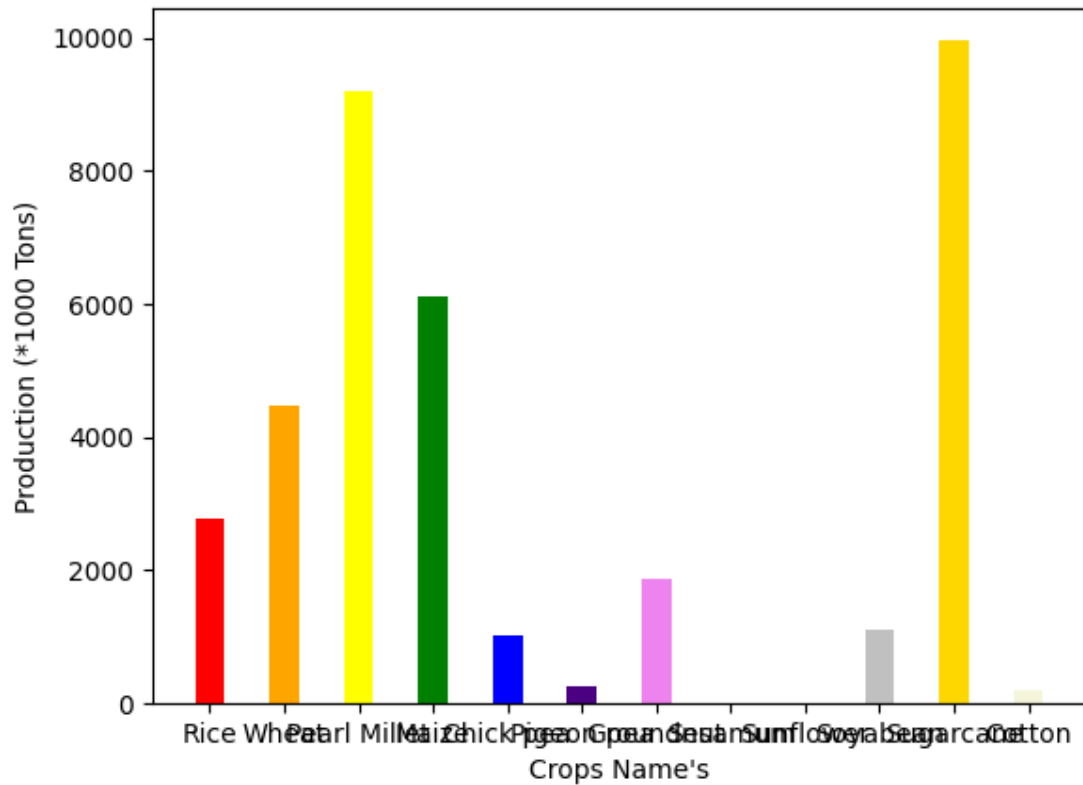
Production of Each Crop of Nagpur District



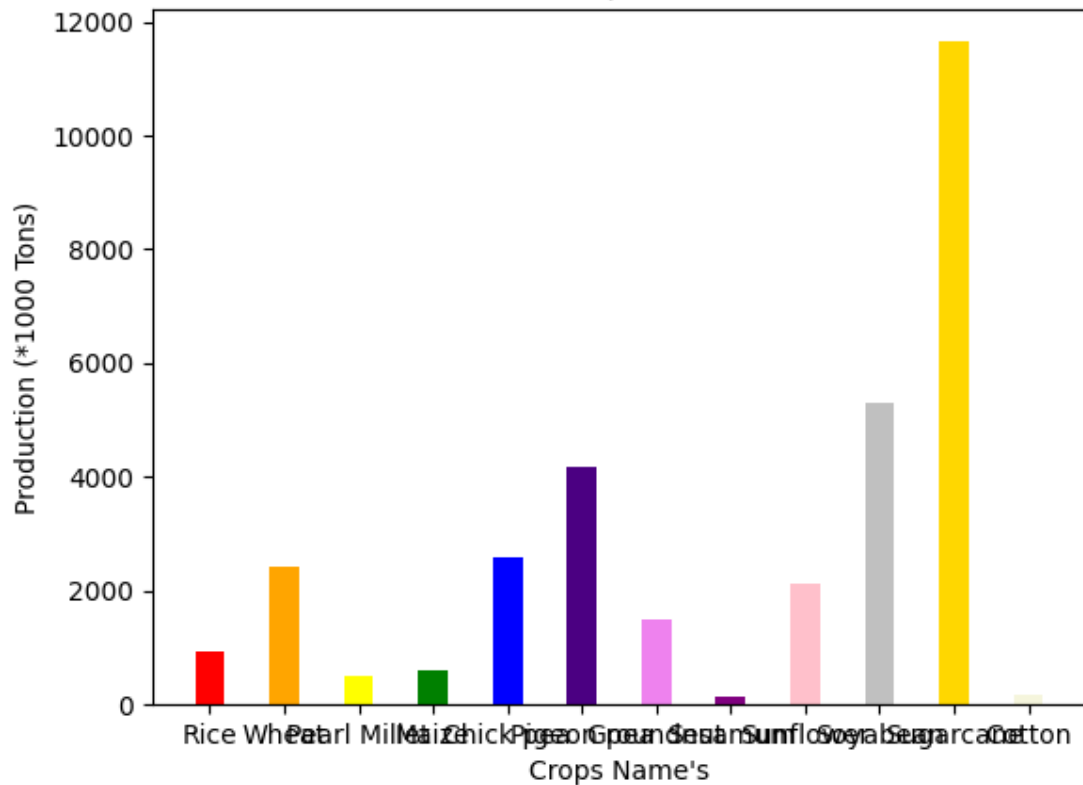
Production of Each Crop of Nanded District



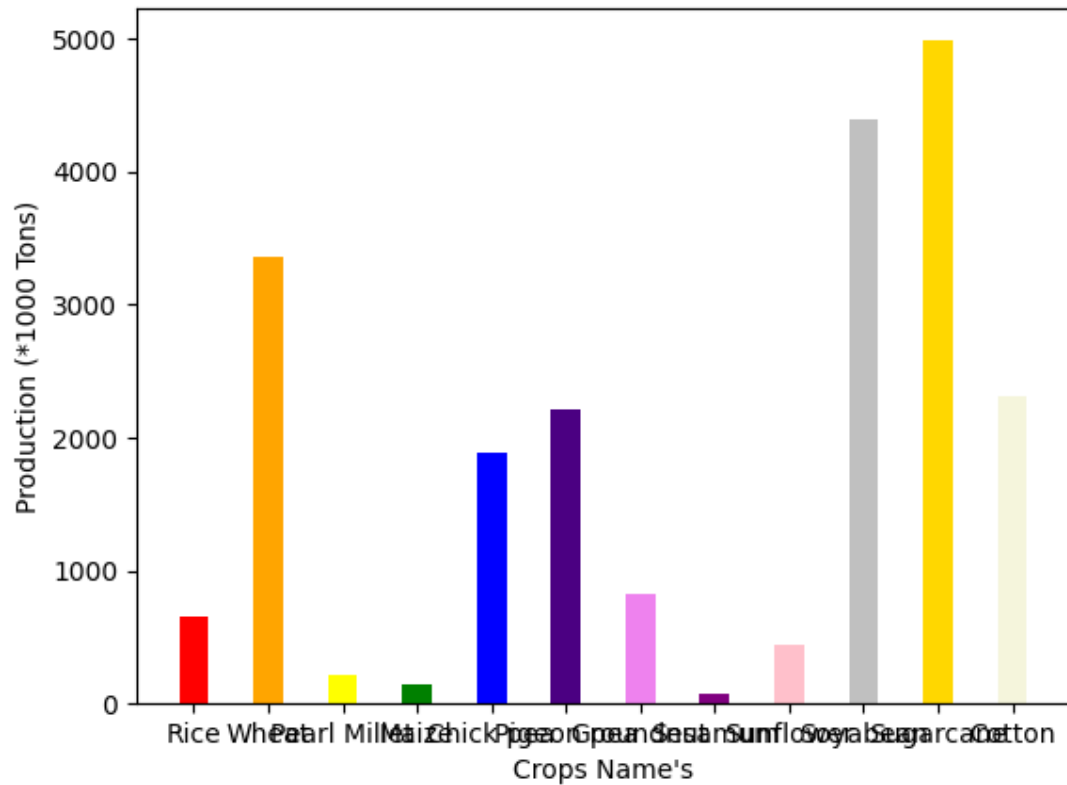
Production of Each Crop of Nasik District



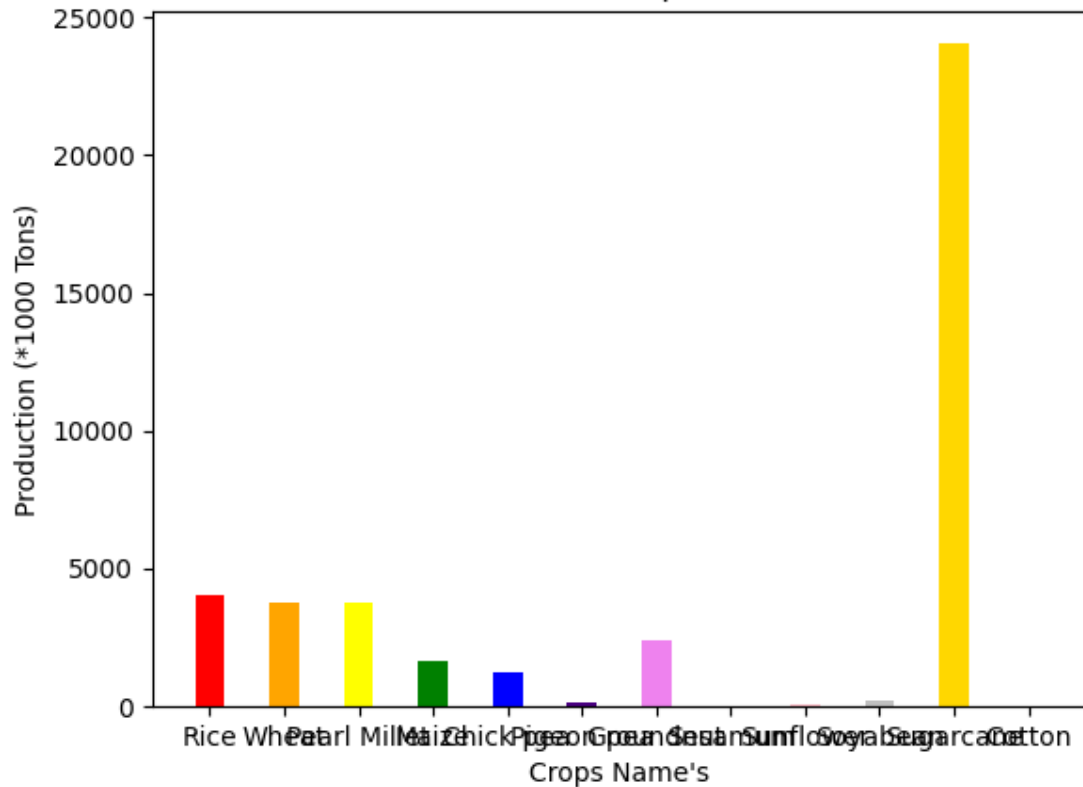
Production of Each Crop of Osmanabad District



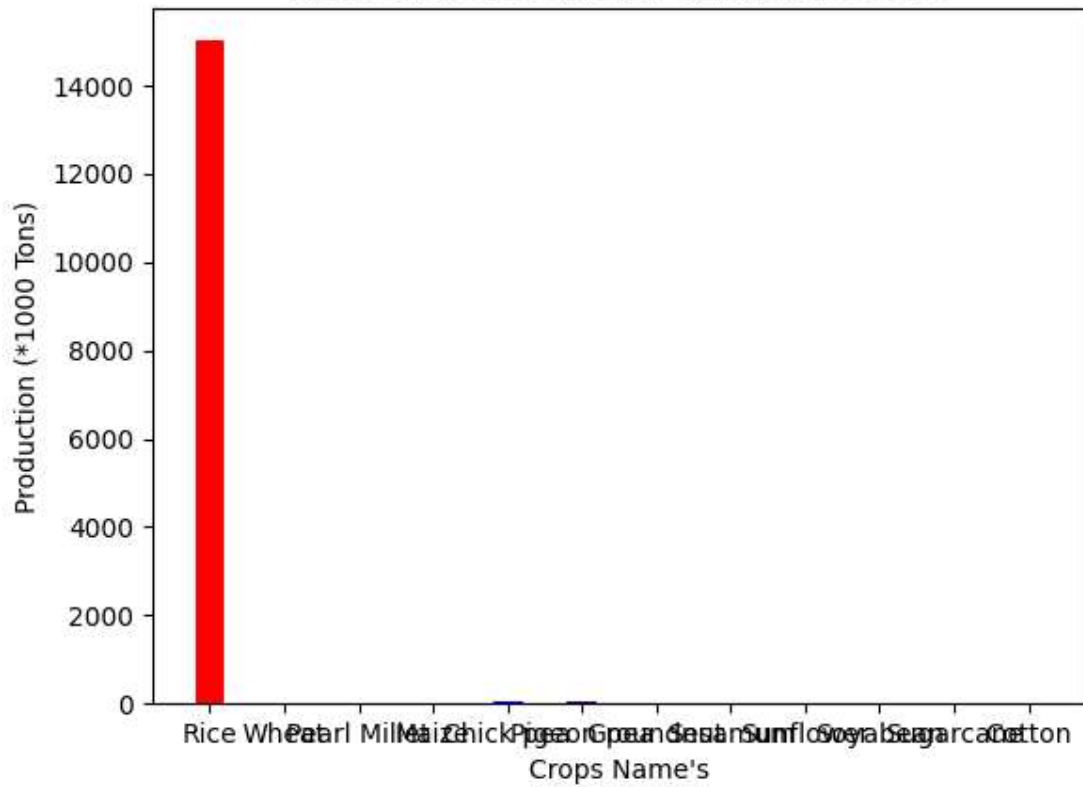
Production of Each Crop of Parbhani District



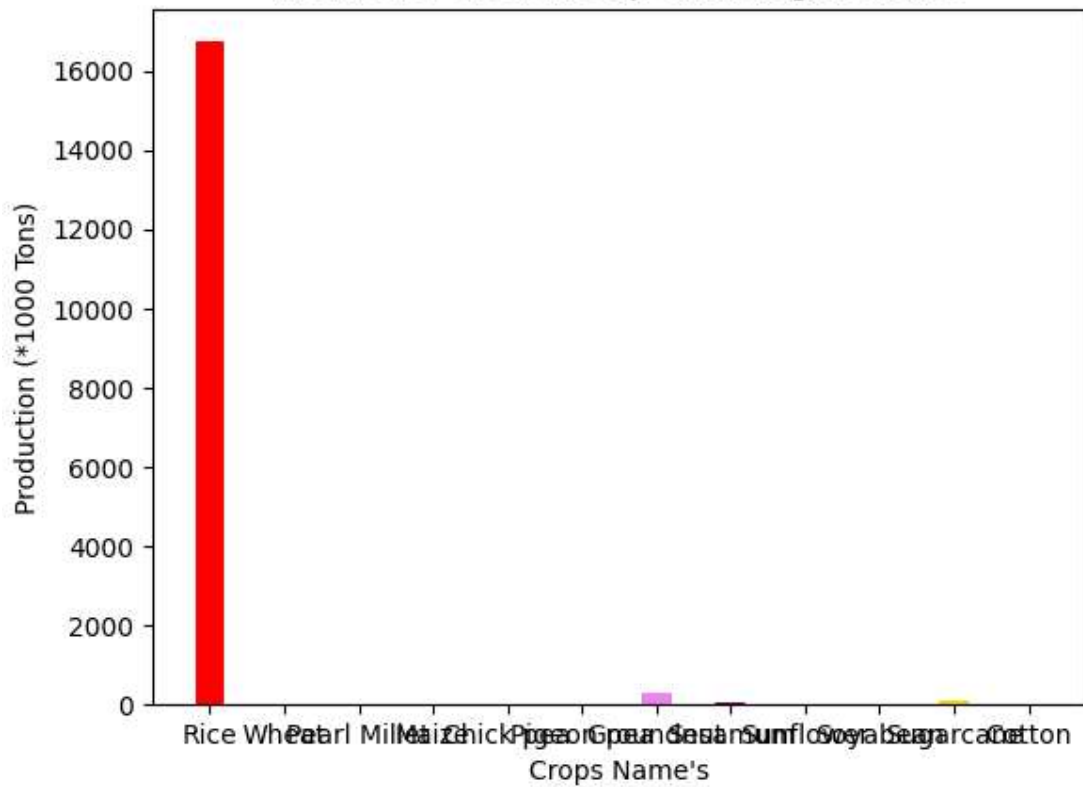
Production of Each Crop of Pune District



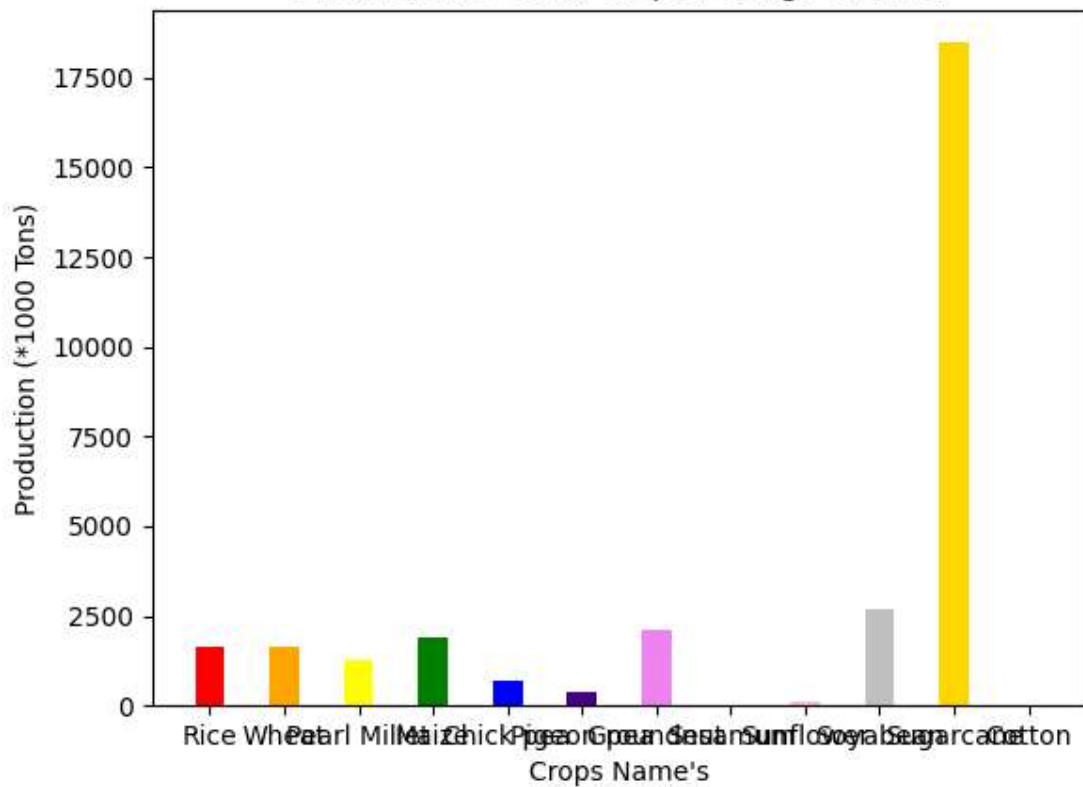
Production of Each Crop of Raigad District



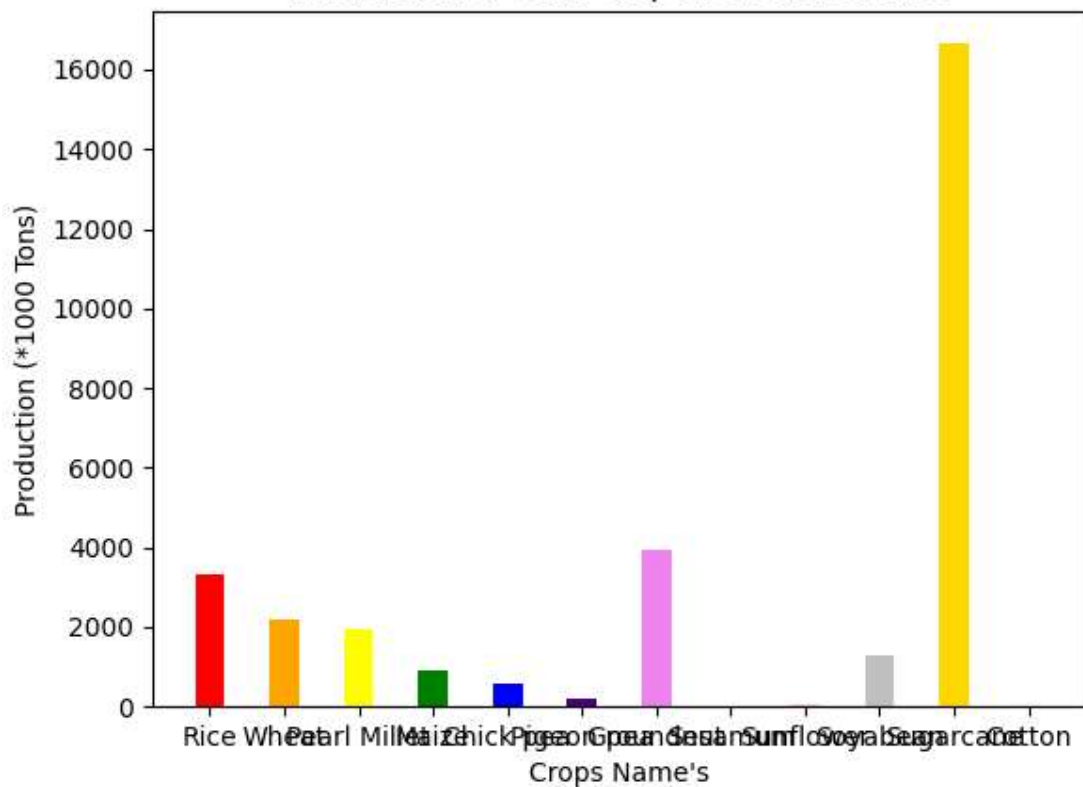
Production of Each Crop of Ratnagiri District



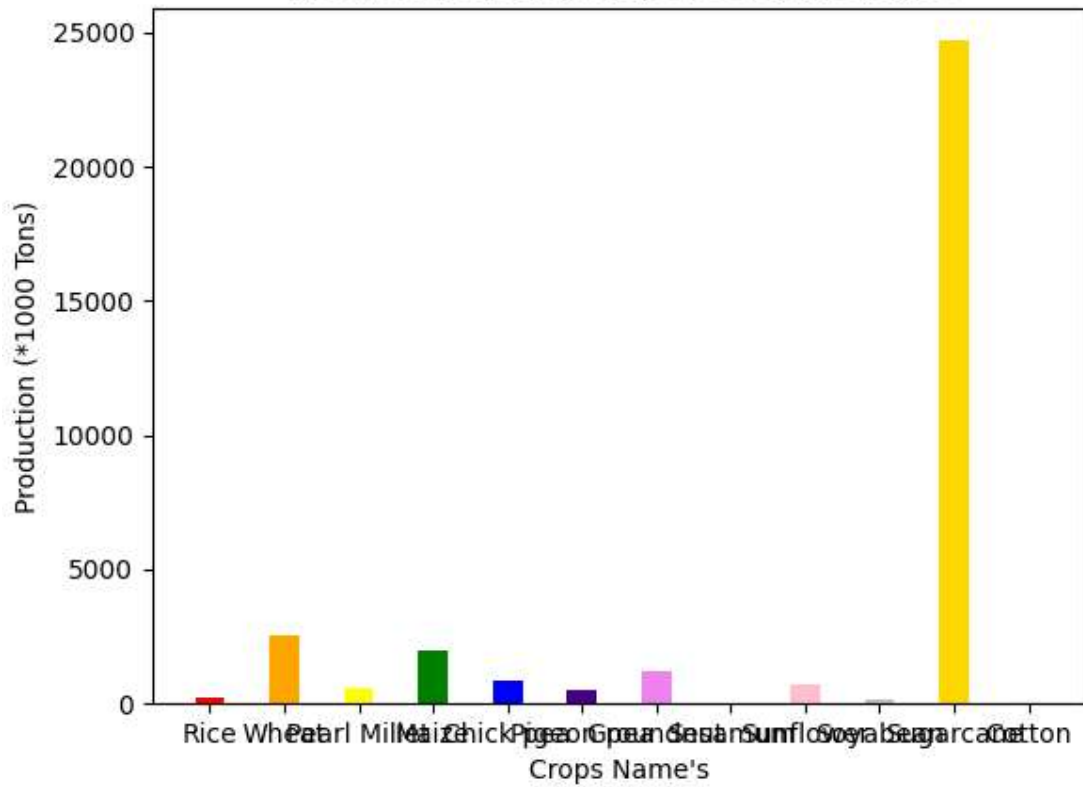
Production of Each Crop of Sangli District



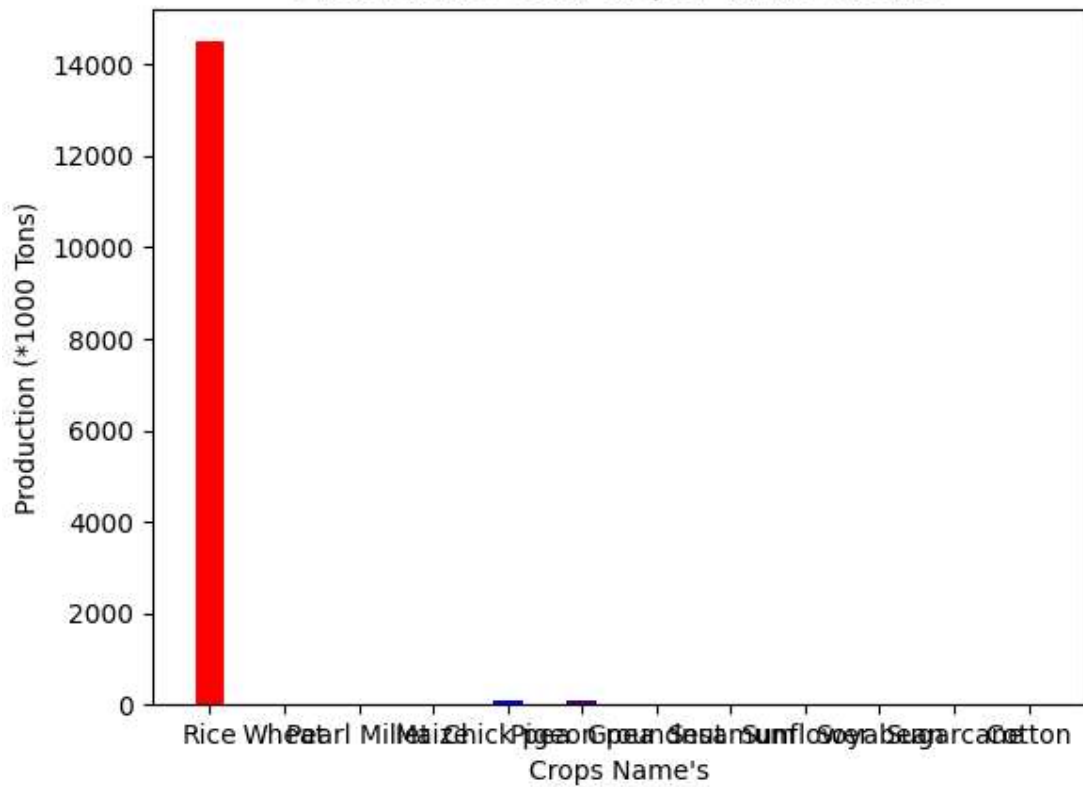
Production of Each Crop of Satara District



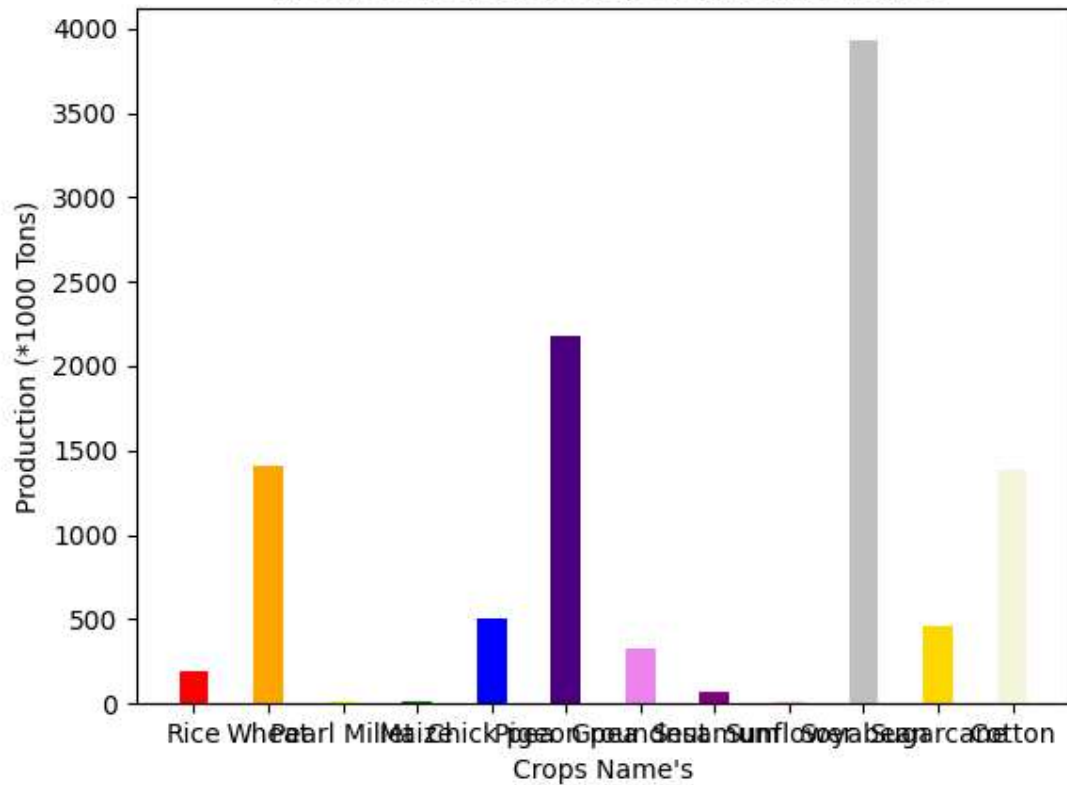
Production of Each Crop of Solapur District



Production of Each Crop of Thane District



Production of Each Crop of Wardha District



Production of Each Crop of Yeotmal District

