

The Battle of Neighborhoods – Week2

Introduction:

Background:

At 157.35 million hectares, **India** holds the second largest **agricultural** land globally. **Agriculture** is the primary source of livelihood for 58 per cent of the country's population. **India** is the **largest** producer of spices, pulses, milk, tea, cashew and jute; and the **second largest** producer of wheat, rice, fruits and vegetables, sugarcane, cotton and oilseeds. Further, **India** is **second** in global **production** of fruits and vegetables, and is the **largest** producer of mango and banana.

Indian agribusiness; this should be evaluated on a nonstop premise considering quality and shortcomings of the nation. Inquires about in this unit would target setting up India's farming part in the developing universal exchange request. The difficulties and open doors for nation frequently fluctuate crosswise over products; in this manner area and were explicit investigations surveying potential, potential dangers and likely reaction of the nation would be sought after in the unit. Specifically, the investigations would investigate the potential for progressively compelling utilization of the current provisos (like 'Unique Products' or 'Uncommon Safeguards') and the requirement for new arrangements.

Problem:

Cultivating is a difficult activity for farmers to choose best reasonable harvest dependent on the season and area for greater efficiency and income. Utilizing the Data science approach and machine learning procedures like clustering, this project aims to give the answers for organizations to encourage the necessary seeds and fertilizers to the farmers depending on the season and location. This thus impact to the Indian economy and its gross GDP.

Target Audience:

To prescribe the best appropriate harvest at offered area to the administration, private firms and free ranchers, which builds the efficiency and enables people to gross income that thusly infers to the nation's principle wellspring of economy. This would interest any former who wants to start the best crop which generates more benefits in any selected Indian State.

Data acquisition and cleaning

Data Source :

The data refers to district wise, crop wise, season wise and year wise data on crop covered area and production. The data is being used to study and analyze crop production, production contribution to district/State/country, Agro-climatic zone wise performance, and high yield production order for crops, crop growing pattern and diversification. One state will be analyzed in this project: **Andhra Pradesh**.

We will be using the below datasets for analyzing **Andhra Pradesh Crop data**

Data 1:

Crop dataset has a total of 29 States and 69 different crops. In order to segment the crops and explore them, we will essentially need a dataset that contains the 13 districts of Andhra Pradesh and the crop that exist in each district.

This dataset exists for free on the web.

Link to the dataset is: <https://data.gov.in/resources/district-wise-season-wise-crop-production-statistics-1997>

Data 2:

Andhra Pradesh state geographical coordinates data will be utilized as input for the Foursquare API, which will be leveraged to provision crop information for each district of State Andhra Pradesh. We will use the Foursquare API to explore districts in **Andhra Pradesh** State.

Approach:

- Collect the Crop dataset from <https://data.gov.in/resources/district-wise-season-wise-crop-production-statistics-1997>
- Using Foursquare API we will find all geographical coordinates of each State.
- Filter out crops that are in Andhra Pradesh State.
- Find coordinates and crops for each district in the Andhra Pradesh State using Foursquare API.
- Using production for each crop, we will sort that data.

- Visualize the Ranking of crop and neighborhoods using folium library(python)

Questions that can be asked using the above mentioned datasets:

- What is best location in Andhra Pradesh for given Crop?
- Which areas have potential Corps?
- What are the seeds and fertilizers supplies needed for Farmers?
- Which crop gives the better production for Farmers?

Exploratory Data Analysis:

Finding the longitude & Latitude of Andhra Pradesh using the [Nominatium](#)

Code:

```
geolocator = Nominatim(user_agent="my-application1")
location = geolocator.geocode(state_name)
latitude = location.latitude
longitude = location.longitude
print("State : {}, Latitude: {}, Longitude: {}".format(state_name, latitude, longitude))
```

Andhra Pradesh, Latitude: 15.9240905, Longitude: 80.1863809



Andhra Pradesh has below list of 13 districts and each of it having different weather conditions for farming.

['ANANTAPUR', 'CHITTOOR', 'EAST GODAVARI', 'GUNTUR', 'KADAPA', 'KRISHNA', 'KURNOOL', 'PRAKASAM', 'SPSR NELLORE', 'SRIKAKULAM', 'VISAKHAPATNAM', 'VIZIANAGARAM', 'WEST GODAVARI']

Coordinates of the each District in Andhra Pradesh:

```
# Get the langitude and latitude for each district
district_names=[]
district_lats=[]
district lngs= []
#geolocator = Nominatim(user_agent="my-application1")
for district in districts:
    #print(" " + district)
    location = geolocator.geocode(district)
    latitude = location.latitude
    longitude = location.longitude
    #print('The georapical coordinate of {} are {}, {}.'.format(district, latitude, longitude))
    district_lats.append(latitude)
```

```
district lngs.append(longitude)
district names.append(district)

district_lats
district lngs
district_names

for dist, lat, lon in zip(district_names, district_lats, district lngs):
    print ('{} - {}, {}'.format(dist, lat, lon))

ANANTAPUR - 14.55, 77.416667

CHITTOOR - 13.30374115, 78.9877010160408

EAST GODAVARI - 17.233496, 81.7225986

GUNTUR - 16.2915189, 80.4541588

KADAPA - 14.4671491, 78.8228877

KRISHNA - 16.6691525, 80.7190024

KURNOOL - 15.8309251, 78.0425373

PRAKASAM - 15.5, 79.5

SPSR NELLORE - 14.4493717, 79.9873763

SRIKAKULAM - 18.32002205, 83.9160771993717

VISAKHAPATNAM - 17.7231276, 83.3012842

VIZIANAGARAM - 18.1120819, 83.4052196224888

WEST GODAVARI - 17.0, 81.166667
```

Current Data set for the Andhra Pradesh contains a total of **9628** records for different districts.

Source dataset shape - (9628, 70)

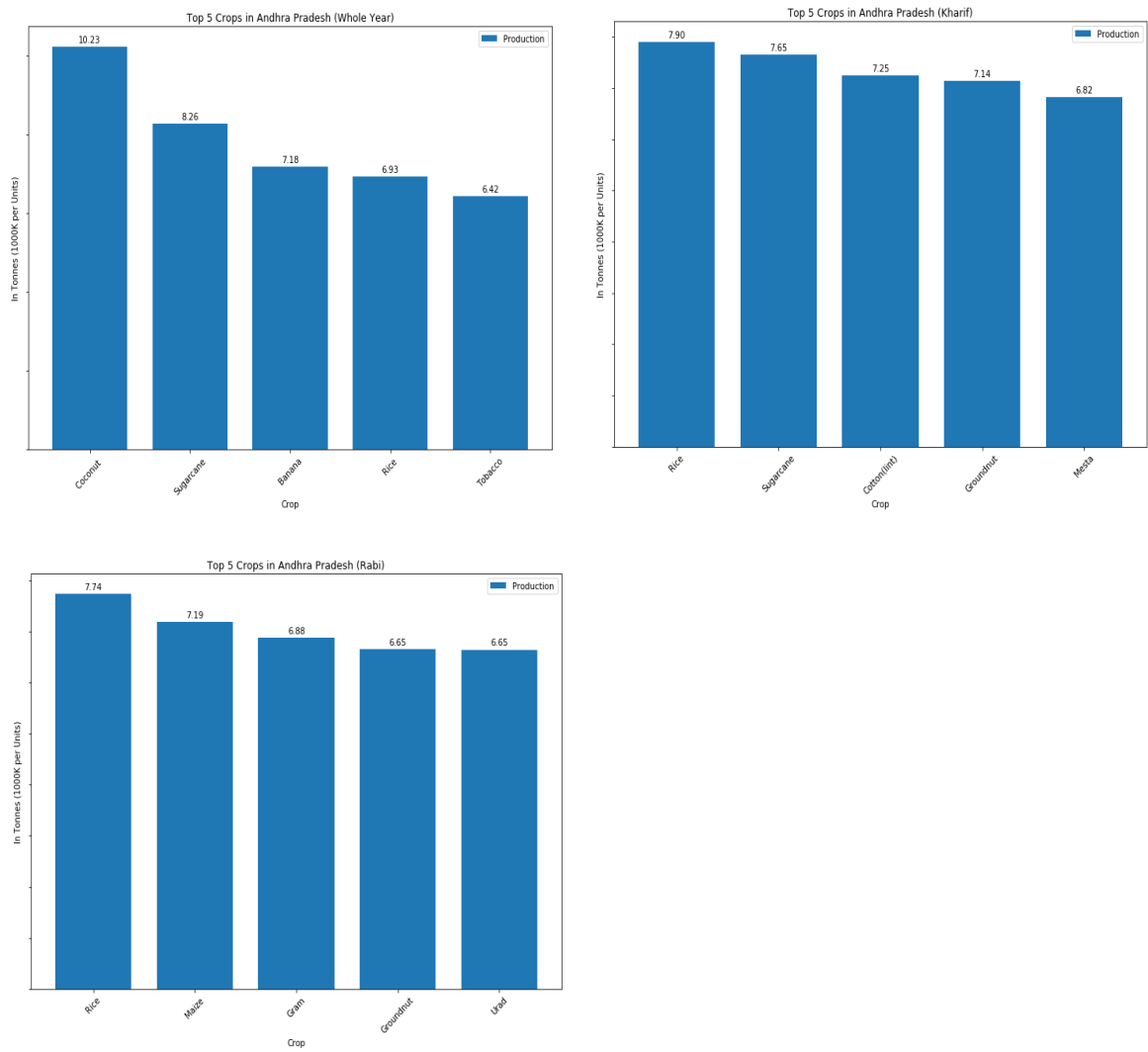
	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	Andhra Pradesh	ANANTAPUR	1997	Kharif	Arhar/Tur	21400	2600.0
1	Andhra Pradesh	ANANTAPUR	1997	Kharif	Bajra	1400	500.0
2	Andhra Pradesh	ANANTAPUR	1997	Kharif	Castor seed	1000	100.0
3	Andhra Pradesh	ANANTAPUR	1997	Kharif	Cotton(lint)	7300	9400.0
4	Andhra Pradesh	ANANTAPUR	1997	Kharif	Dry chillies	3700	7100.0

Based on the soli and weather conditions, each district is suitable for different crops. In our dataset we have below listed agriculture products:

['Arecanut', 'Arhar/Tur', 'Bajra', 'Banana', 'Beans & Mutter(Vegetable)', 'Bhindi', 'Bottle Gourd', 'Brinjal', 'Cabbage', 'Cashewnut', 'Castor seed', 'Citrus Fruit', 'Coconut ', 'Coriander', 'Cotton(lint)', 'Cowpea(Lobia)', 'Cucumber', 'Dry chillies', 'Dry ginger', 'Garlic', 'Ginger', 'Gram', 'Grapes', 'Groundnut', 'Horse-gram', 'Jowar', 'Korra', 'Lemon', 'Linseed', 'Maize', 'Mango', 'Masoor', 'Mesta', 'Moong(Green Gram)', 'Niger seed', 'Onion', 'Orange', 'Other Rabi pulses', 'Other Fresh Fruits', 'Other Kharif pulses', 'Other Vegetables', 'Papaya', 'Peas (vegetable)', 'Pome Fruit', 'Pome Granet', 'Potato', 'Ragi', 'Rapeseed &Mustard', 'Rice', 'Safflower', 'Samai', 'Sannhamp', 'Sapota', 'Sesamum', 'Small millets', 'Soyabean', 'Sugarcane', 'Sunflower', 'Sweet potato', 'Tapioca', 'Tobacco', 'Tomato', 'Turmeric', 'Urad', 'Varagu', 'Wheat', 'other fibres', 'other misc. pulses', 'other oilseeds']

Top 5 Crop Productions in the State (Whole Year):

Coconut production top most followed by Sugarcane, Banana, Rice and Tobacco in the sequence. Production units are measured in the tones, in the graph 1 unit represents the1000K tons production.



It was noticed that the whole year production is different from the Season based crop production.

If you look into the Kharif season Rice, Sugarcane, Cotton, Groundnut and Mesta are the top 5 Crops in the production. In Rabi season Rice, Maize, Gram, Coconut and Urad are the top 5 Crops in the Production

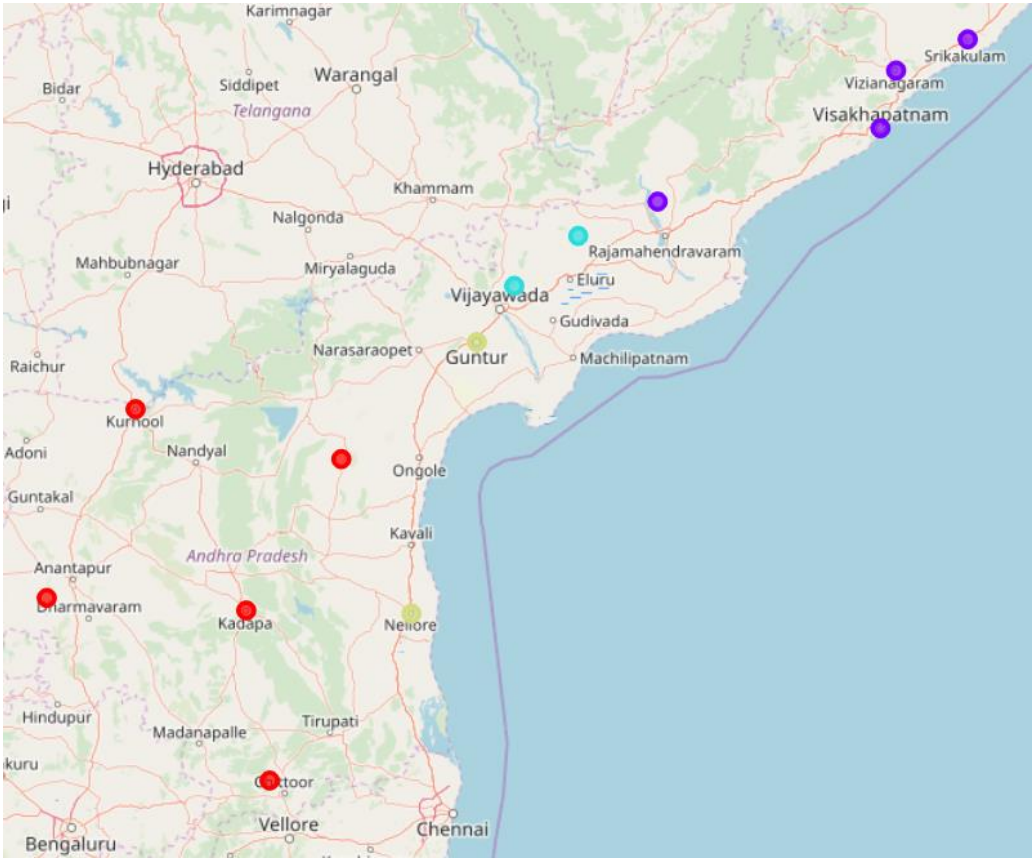
Top 10 crops by each district:

	District_Name	1st Most Common Crop	2nd Most Common Crop	3rd Most Common Crop	4th Most Common Crop	5th Most Common Crop	6th Most Common Crop	7th Most Common Crop	8th Most Common Crop	9th Most Common Crop	10th Most Common Crop
0	ANANTAPUR	Rice	Maize	Groundnut	Jowar	Ragi	Horse-gram	Bajra	Dry chillies	Moong(Green Gram)	Onion
1	CHITTOOR	Rice	Horse-gram	Groundnut	Ragi	Maize	Bajra	Moong(Green Gram)	Sesamum	Urad	Dry chillies
2	EAST GODAVARI	Rice	Sesamum	Maize	Dry chillies	Urad	Groundnut	Moong(Green Gram)	Ragi	Jowar	Onion
3	GUNTUR	Rice	Moong(Green Gram)	Urad	Maize	Sesamum	Dry chillies	Groundnut	Jowar	Arhar/Tur	Onion
4	KADAPA	Rice	Jowar	Groundnut	Horse-gram	Bajra	Ragi	Dry chillies	Sesamum	Urad	Maize

Predictive Modeling:

Top 10 crops are mapped to 4 different clusters and marked to the AP map which is generated from the Foliium.

- In first cluster 5 districts and Rice is the 1st common crop.
- In the second cluster 4 districts and Rice is the 1st Common crop.
- In third cluster 2 districts and Rice is the 1st Common crop.
- In fourth cluster 2 districts and Rice is the 1st Common crop.



Methodology:

For this report I used a few different maps that could help a farmer to select the suitable crops, Fertilizer firm to supply required amount of supply and administration to make sure the availability of the seeds to get the maximum production and profits in the neighborhood/district based on its crop production and the resource availability. In order to do that I've used the 1997-2014 crop production dataset from open government data platform India, combined with Folium maps to visually display the neighborhoods/districts and Foursquare data to display the cluster in the Andhra Pradesh region.

Conclusions:

This report may be helpful for someone (administration, Fertilizer firms, Farmers) to focus on the suitable crop, by comparing the current crop production in different districts in Andhra Pradesh, however it may not cover all variables such as weather condition, rain fall, Soil based crop, so it shall not be used as a single decision making tool. Based on the dataset below conclusions made

- Top 5 Crop production in **Kharif** and **Rabi** are not same
- Most Common Crop is **Rice** in the neighborhoods/districts in Andhra Pradesh.
- Coconut is the highest crop production in the state which is mostly in the costal districts/areas.

Limitations:

- The accuracy of data depends purely depends on the data provided by data.gov.in site team
- Data is available from 1997-2012 and weather forecast is also impact the crop production which way results the wrong predictions
- Soli is the most important elements in the Crops production and it is missing in the current dataset.