INDIAN CROPS DATA REPORT

DATA CLEANING AND PREPROCESSING

Hiren T Modi -12110077
Sarabjit Rangi-12110011
Paras Jain -12110032
Vaibhav Sethi -12110105

PROJECT LINK : https://isbhydmoh-my.sharepoint.com/:f:/g/personal/hirent_modi_ampba2022s_isb_edu/EgrUpzCuYFhEgi11Ovfat9YBBsASFnJL7ItQ2kJvM-34eA?e=dR0Eg3

Executive Summary

Problem statement

- To provide basic understanding of various aspects of crops in India.
- To Collect, Study and Analysis of Indian Crops on varied attributes like acreage, production, consumption, crop-names, seasons, yield, use of fertiliser, rainfall, credit, water level for bore - well /dug well, moisture, pesticides', land holding pattern of Indian farmer etc .
- To emphasise the importance of data collection technique like scrapping and crawling.
- Visualize collected data to bring fruitful insights.

Proposed Solution

- We scrapped Govt Websites and retrieved data as per attributes of our data frames.
- We did Exploratory Data Analysis on our data frame permuting and combining various attributes.
- Since single seed data source in un-available, we first have to collect data from varied sources, post have to merge in unified way to create a meaningful data frame.
- We have to use various data collection techniques like web scrapping, crawling etc. for segregating and aggregating data as per attribute-oriented data context.
- Data frame cells with non – availability of data, none values, absurd values, negative values etc. etc. need data pre -processing which mostly done manually.

Challenges Faced

- Data seed not available in a single URL.
- Not all crops' data is recorded in India.
- Need to calibrate Scraper and Crawler for each website.
- Most of data available in pdf for this we have to use scraper that can extract data from pdf.
- Numerous data sources required for data collection.
- Different names for different crops which hampers data collection process.
- Need to merge large number of different files to consolidate data in one file.
- Google search and amazon(prices) not allowing web crawlers and banned our IP address.
- Most data sources include only production information and not include other attributes.
- Some states do not record Agriculture data or record data for some most common crops grown in their states.
- Non standardised structure for attributes of varied crops like wheat data is different than apple data which is again different than herbs data.

Hiren T Modi -12110077
Sarabjit Rangi-12110011
Paras Jain -12110032
Vaibhav Sethi -12110105

Table of contents

Hiren T Modi -12110077
Sarabjit Rangi-12110011
Paras Jain -12110032
Vaibhav Sethi -12110105

1 - The Chosen Domain and Seed Sources.

Motivation for the chosen domain: -

The main motivation behind picking the topic of Indian Crops was to be able to work on our collected data in the future and put it to good use.

Either in terms of -

- Further analytics on the dataset so as to understand growth patterns or,
- For predictive analytics and forecast with statistical evidence, on various parameters.

Two-thirds of India's population is engaged in agricultural activities. Given that this vast country such diverse dynamics, we were bound to find various food and non-food crops which are cultivated. Largely, India's crop system is divided into three main cropping seasons which are rabi, kharif and Zaid.

Agriculture, with its allied sectors, are main sources of income of livelihoods in India. India is the largest cultivator, (25% of world manufacturing), customer (27% of global wide consumption) and importer (14%) of pulses globally. India's yearly milk production was at 185MT (2017-18), making it the most important manufacturer of milk, jute and pulses. India is the penultimate manufacturer of rice, wheat, sugarcane, cotton and groundnuts, It is also the second-biggest fruit and vegetable manufacturer, accounting for 10.9% and 8.6% of the sector fruit and vegetable manufacturing, respectively. (STATS by FAO.org)

The importance and the impact our effort could have, solidified our problem statement.

Analysis conducted and the Seed source:-

We followed the following strategy

1) To keep things simple;
2) To plan the entire process;
3) To ensure reliable, credible and valid data;
4) To be ethical with data collection.

- We decided to split the work into batches which involved the following steps

1) To collect relevant data while keeping the strategy in mind.
2) To ideate on how to bring the data together.
3) To clean the data for ease of merging and potential usage.
4) To present the data efficiently.

Data Cleaning and preprocessing – Group 24 Assignment

Major Challenges while trying to kick start things : -

- Our biggest challenge, was to identify workable source of data.
- The obvious websites were the government portals, but these portals and all major government surveys track only 56 major crops and that would mean that our dataset would lose its expanse and depth.
- With extensive search, we realised that the government has split the crop surveys and information into area, yield and production information. This information seemed reliable and that became our first step of sourcing data.
- The main idea was to be able to work on the seed data, identify the crops about which we could get information (valuable information) and then crawl, download and scrape the web for expanse and depth of the dataset.
- **The work was divided based on crop categories to avoid overlap of research. We followed a MECE approach (Mutually exclusive and cumulatively exhaustive).**
- **The research for the seed was split into a) Grains and general crops, Fruits and vegetables, Herbs and other plants, Oil seeds.**

Working on the seed source.

While extensive research was done on various government surveys, we found the following sources which we believed were good starting points for the seed data.

- http://mospi.nic.in/agriculture-statistics - Ministry of Statistics & Programme Implementation.
- https://krishi.icar.gov.in/ - KRISHI Agricultural Knowledge Resources and Information System Hub for Innovations, is an initiative of Indian Council of Agricultural Research (ICAR)
- https://data.gov.in/sector/agriculture - Open government data platform for India.

The seed source provided by default seemed to be a little sparse for Indian crops in specific and hence we had to develop a new seed for a general overview and to streamline future efforts. While https://world-crops.com/ seemed like a good source, we decided to look for more structured sources for our initial seed so as to establish a key column very early on.

Having spent majority of our time trying to set the initial seed right, we found that Directorate of Economics and Statistics, Department of Agriculture, Cooperation and Farmers Welfare, Ministry of Agriculture and Farmers Welfare publishes a handbook every year called AGRICULTURAL STATISTICS AT A GLANCE.

Hiren T Modi -12110077
Paras Jain -12110032
Sarabjit Rangi-12110011
Vaibhav Sethi -12110105

Data Cleaning and preprocessing – Group 24 Assignment

This source turned out to be a game changer because : -

1) We could now search for yearly patterns and parameters.
2) Make a seed based on common parameters to be used as keys in the future.
3) Get an idea about the kind of attributes we can scrape, download or crawl for.

SEED SOURCE : - www.agricoop.nic.in & http:// eands.dacnet.nic.in

The first seed that we started off with was the following page : -

Table 4.1(a): Target and Achievement of Production of Major Crops during Tenth Five Year Plan

(Million Tonnes)

Crop	2002-03		2003-04		2004-05		2005-06		2006-07		Xth Plan	
	Target	Achievement	Target	Achievement	Target	Achievement	Target	Achievement	Target	Achievement	Target	Achievement
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Rice	93.00	71.82	93.00	88.53	93.50	83.13	87.80	91.79	92.80	93.35	460.10	428.62
Wheat	78.00	65.76	78.00	72.15	79.50	68.64	75.53	69.35	75.53	75.81	386.56	351.71
Coarse Cereals	33.00	26.07	34.00	37.60	36.80	33.46	36.52	34.06	36.52	33.92	176.84	165.11
Pulses	16.00	11.13	15.00	14.91	15.30	13.13	15.15	13.39	15.15	14.20	76.60	66.76
Foodgrains	220.00	174.77	220.00	213.19	225.10	198.36	215.00	208.60	220.00	217.28	1100.10	1012.20
Oilseeds	27.00	14.84	24.70	25.19	26.20	24.35	26.58	27.98	29.40	24.29	133.88	116.65
Sugarcane	320.00	287.38	320.00	233.86	270.00	237.09	237.50	281.17	270.00	355.52	1417.50	1395.02
Cotton #	15.00	8.62	15.00	13.73	15.00	16.43	16.50	18.50	18.50	22.63	80.00	79.91
Jute & Mesta@	12.00	11.28	12.00	11.17	11.80	10.27	11.28	10.84	11.28	11.27	58.36	54.83

Million Bales of 170 kg. each.
@ Million Bales of 180 kg. each.
Source: Directorate of Economics & Statistics, DAC&FW

74 Agricultural Statistics at a Glance 2017

Table4.1(b): Target and Achievement of Production of Major Crops during Eleventh and Twelfth Five Year Plans

(Million Tonnes)

Crop	2007-08		2008-09		2009-10		2010-11		2011-12		XIth Plan		2012-13		2013-14		2014-15		2015-16		2016-17*	
	Target	Achieve-ment	Target	Achieve-ment	Target	Achieve-ment	Target	Achieve-ment	Target	Achieve-ment	Target	Achieve-ment	Target	Achieve-ment	Target	Achieve-ment	Target	Achieve-ment	Target	Achieve-ment	Target	Achieve-ment
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)
Rice	93.00	96.69	97.00	99.18	100.50	89.09	102.00	95.98	102.00	105.30	494.50	486.24	104.00	105.24	105.00	106.65	106.00	105.48	106.10	104.41	108.50	110.15
Wheat	75.50	78.57	78.50	80.68	79.00	80.80	82.00	86.87	84.00	94.88	399.00	421.80	88.00	93.51	92.50	95.85	94.00	86.53	94.75	92.29	96.50	98.38
Coarse Cereals	37.50	40.76	42.00	40.03	43.10	33.55	44.00	43.68	42.00	42.01	208.60	200.03	44.00	40.04	42.50	43.29	41.50	42.86	43.20	38.52	44.35	44.19
Pulses	15.50	14.76	15.50	14.57	16.50	14.66	16.50	18.24	17.00	17.09	81.00	79.32	18.24	18.34	19.00	19.25	19.50	17.15	20.05	16.35	20.75	22.95
Foodgrains	221.50	230.78	233.00	234.47	239.10	218.11	244.50	244.78	245.00	259.29	1183.10	1187.43	254.24	257.13	259.00	265.04	261.00	252.02	264.10	251.57	270.10	275.68
Oilseeds	30.00	29.76	31.75	27.72	31.60	24.88	33.20	32.48	33.60	30.01	160.15	144.85	33.50	30.94	31.00	32.75	33.00	27.51	33.00	25.25	35.00	32.10
Sugarcane	310.00	348.19	340.00	285.03	340.00	292.30	315.00	342.38	350.00	361.04	1655.00	1628.94	352.00	341.20	340.00	352.14	345.00	362.33	355.00	348.45	355.00	306.72
Cotton #	22.00	25.88	26.00	22.28	26.00	24.02	26.00	33.00	34.00	35.20	134.00	140.38	35.00	34.22	35.00	35.90	35.00	34.80	35.15	30.01	36.00	33.09
Jute & Mesta@	11.00	11.21	11.00	10.37	11.20	11.82	11.50	10.62	12.30	11.40	57.00	55.42	12.00	10.93	12.00	11.69	11.20	11.13	11.70	10.52	11.70	10.60

Source: Directorate of Economics & Statistics, DAC&FW
* 4th Advance Estimates.

Area, Production an

Hiren T Modi -12110077
Paras Jain -12110032
Sarabjit Rangi-12110011
Vaibhav Sethi -12110105

2 - The structured and unstructured sources from open domain/ internal sources:

Sources and the reasoning behind the picks : -

a) We did so by downloading all the agricultural statistics at a glance summary that were released by the government year after year.
b) Every member of the group was given the task of trying to determine an attribute that could become a column of our data set.
c) As we started collecting these PDFs and started extracting the respective Excel files using the Python code (given below), we realized that most of these data sources had a very specific pattern that it repeated.
d) Data was either group based on your or based on states or based on the various districts across the country.
e) The government pays major focus to the principal crops that act as a backbone of import-export and the agricultural and allied services sector of the country.
f) Most of the data available was regarding these principal crops and a few of the data points were regarding various horticulture crop's vegetation and plant growth.
g) The focus was then to pay most attention to these principal crops and the horticulture crops for which we could gather a lot of reliable data and consistent data.
h) We began extracting records from the 1990s up till the current year. We hit a major roadblock because most of the statistical units of the country have failed at collaborating with each other to form a combined data space after 2017. Hence, we began restricting our data downloading crawling and collection process up till 2017.
i) We decided to collect other sparse data sets up till the 2020 and started storing the same in a separate knowledge base to then be able to create a sparse data set that might have semi structured or unstructured data for the latest years.
j) The idea behind scraping different websites for our data did not make a lot of contribution because the type of attributes we were looking for were not necessarily available in any open domain.
k) The team ended up collecting close to 20 sets of data having common attributes such as the year the state name the district name.
l) The further process was to eliminate data which didn't necessarily fit into our common attributes. Do note that special attention was paid to having some common attributes so that it should be easier to then merge these different data sets and to form a combined bigger data set with credible information.
m) Like previously mentioned, the additional data sets which were sparse and contained information about crops and horticulture branch plantations of the latest year were collected separately and even here we tried to maintain the common attributes so that it would be easier to merge later on if need be.

Data Cleaning and preprocessing – Group 24 Assignment

The list of all the structured and unstructured sources from where we had either scraped, downloaded the PDFs or the CSVs have now been given below : -

1. https://www.fao.org/3/s2022e/s2022e02.htm
2. https://agricoop.nic.in/en/statistics/state-level
3. https://pmfby.gov.in/
4. https://aps.dac.gov.in/LUS/Public/Reports.aspx
5. https://indiadataportal.com/?language=&location=Haryana
6. https://farmer.gov.in/BooKLetView/FarmerFriendlyMaterial.aspx?DocType=Booklets%20and%20Flyers&SCode=33
7. https://eands.dacnet.nic.in/
8. http://mospi.nic.in/agriculture-statistics
9. https://world-crops.com/
10. https://www.britannica.com/topic/list-of-herbs-and-spices-2024392
11. https://en.wikipedia.org/wiki/List_of_Indian_spices
12. http://www.indianspices.com/spices-development/spice-catalogue.html
13. https://www.nmpb.nic.in/
14. https://www.ibef.org/exports/spice-industry-indias.aspx
15. http://www.spices.res.in/annual-report
16. https://indianculture.gov.in/food-and-culture/spices-and-herbs
17. https://www.spicesinc.com/t-list-of-spices.aspx
18. https://www.spicesofindia.co.uk/acatalog/All-Herbs-and-Spices.html
19. http://vanneman.umd.edu/districts/codebook/notecrop.html
20. http://vdsa.icrisat.ac.in/Include/document/all-apportioned-web-document.pdf
21. https://www.google.co.in/intl/en/about/products?tab=wh
22. https://agricoop.gov.in/sites/default/files/agristatglance2018.pdf
23. https://aps.dac.gov.in/
24. https://data.gov.in/catalog/district-wise-season-wise-crop-production-statistics
25. https://www.tn.gov.in/crop/stat.htm
26. https://agricoop.nic.in/sites/default/files/pocketbook_0.pdf
27. https://eands.dacnet.nic.in/FHP(District).htm
28. http://mospi.nic.in/agriculture-statistics
29. http://data.icrisat.org/dld/src/crops.html
30. http://indiastat.com

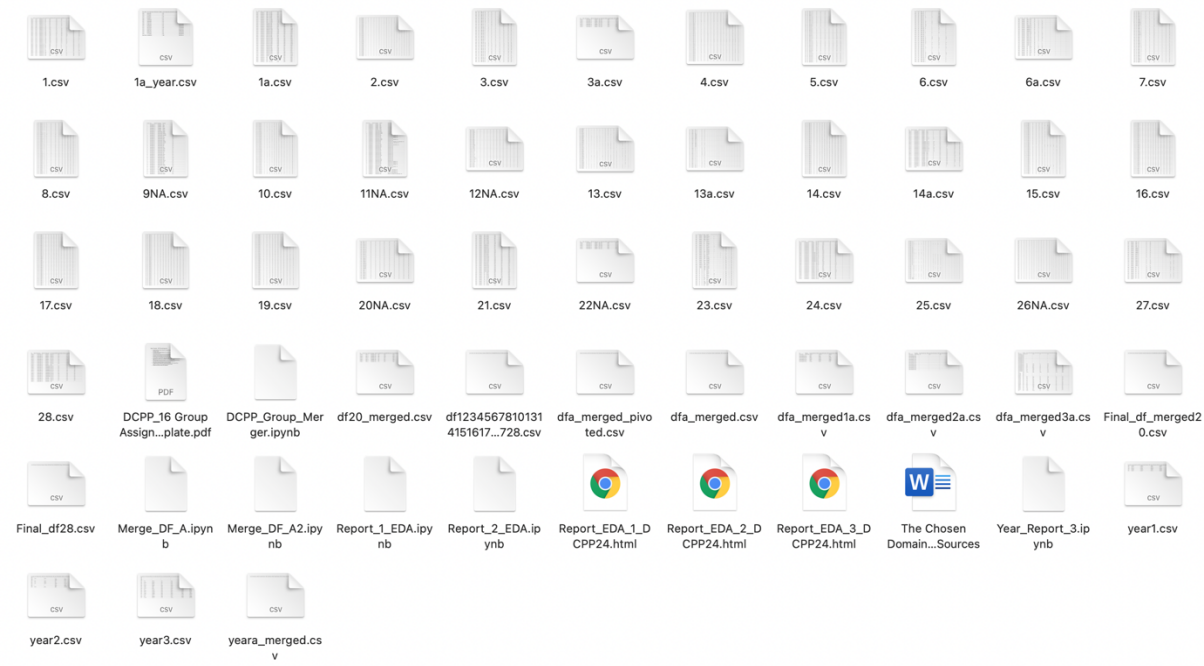
The combined knowledge base of all our downloads and scraped data is present in the source folder of the repository.

ISB's LRC also granted access to Indiastat.com website and this turned out to be the best source of structured excel sheets.

Hiren T Modi -12110077
Paras Jain -12110032
Sarabjit Rangi-12110011
Vaibhav Sethi -12110105

Data Cleaning and preprocessing – Group 24 Assignment

A snapshot of our combined knowledge base has been attached below : -



Reasoning behind the choices : -

1. Like mentioned above the idea was always to have a common knowledge base with data sets that could be merged at a later stage.

2. It always had to be kept in mind that at least 60% of the entire data set was filled individually so that we do not end up with a lot of null values and if we are trying to compute any aggregations during a pivot table these values do not turn out to be unreliable.

3. While trying to download crawl and scrape the data we always kept in mind that the work was divided amongst the team attribute wise so that there was no overlap in the kind of data we brought back to the table.

4. Regular team meets were conducted so as to be able to understand the progress and if we were able to not just attain the minimal requirement of criteria but also have data on with some credible EDA could be carried out later on.

5. The reason why most attention was being paid to the PDFs is because most of our data was available in statistical hand books which are in the PDF format.

Hiren T Modi -12110077
Sarabjit Rangi-12110011
Paras Jain -12110032
Vaibhav Sethi -12110105

3 - Download/ crawl/ collect data from all the sources:

- The Python code for the various ways that we scraped the data and how we converted the PDFs into the respective CSVs/XLSXs have all been attached in the code repository.

Attached below is a sample screenshot of the type of excel sheet we would get if we were to scrape the PDFs that word downloaded from the various seed sources mentioned above : -

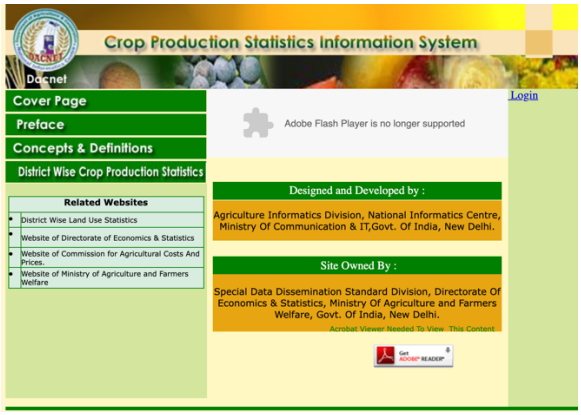
OUTPUT FILE

	Crop	Unnamed: 0	2002-03	Unnamed: 1	2003-04	Unnamed: 2	2004-05	Unnamed: 3	2005-06	Unnamed: 4	2006-07	Xth Plan
0		Target	Achievement	Target	Achievement	Target	Achievement	Target	Achievement	Target	Achievement	Target Achievement
1	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12) (13)
2	Rice	93.00	71.82	93.00	88.53	93.50	83.13	87.80	91.79	92.80	93.35	460.10 428.62
3	Wheat	78.00	65.76	78.00	72.15	79.50	68.64	75.53	69.35	75.53	75.81	386.56 351.71
4	Coarse Cereals	33.00	26.07	34.00	37.60	36.80	33.46	36.52	34.06	36.52	33.92	176.84 165.11
5	Pulses	16.00	11.13	15.00	14.91	15.30	13.13	15.15	13.39	15.15	14.20	76.60 66.76
6	Foodgrains	220.00	174.77	220.00	213.19	225.10	198.36	215.00	208.60	220.00	217.28	1100.10 1012.20
7	Oilseeds	27.00	14.84	24.70	25.19	26.20	24.35	26.58	27.98	29.40	24.29	133.88 116.65
8	Sugarcane	320.00	287.38	320.00	233.86	270.00	237.09	237.50	281.17	270.00	355.52	1417.50 1395.02
9	Cotton #	15.00	8.62	15.00	13.73	15.00	16.43	16.50	18.50	18.50	22.63	80.00 79.91
10	Jute & Mesta@	12.00	11.28	12.00	11.17	11.80	10.27	11.28	10.84	11.28	11.27	58.36 54.83

INPUT FILE

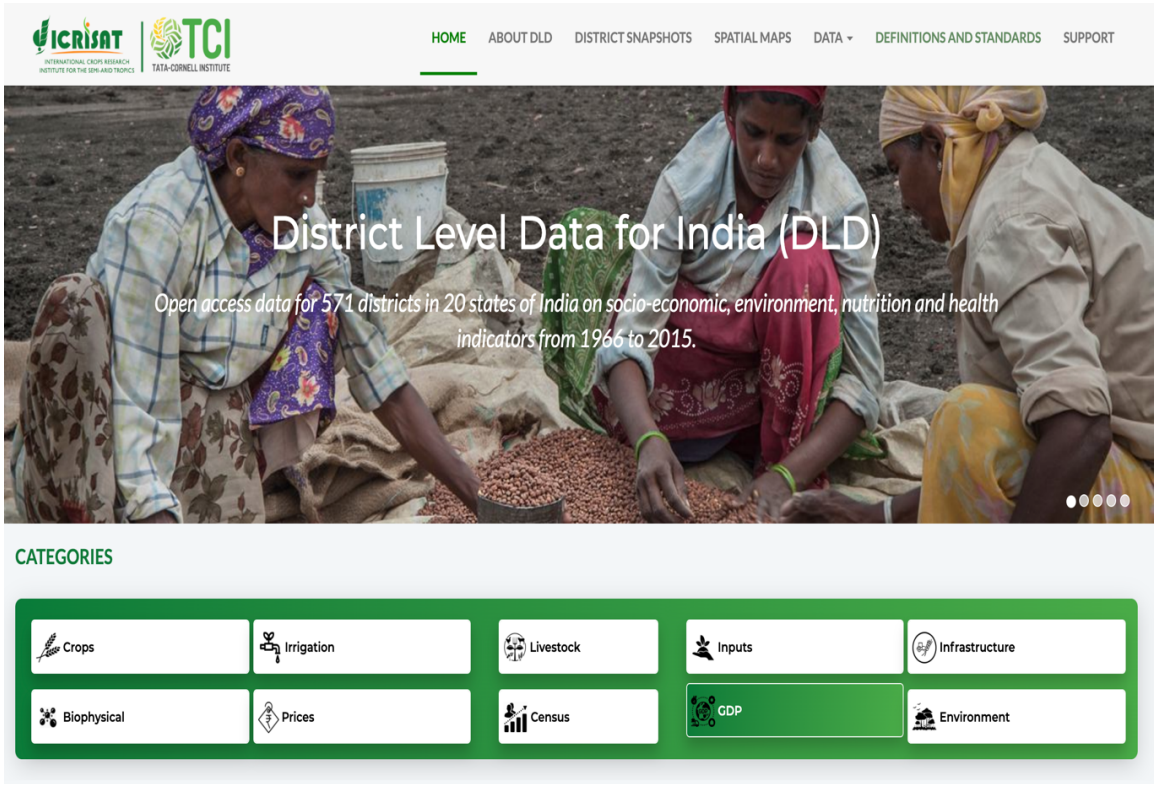
Crop	2002-03		2003-04		2004-05		2005-06		2006-07		Xth Plan	
	Target	Achievement	Target	Achievement	Target	Achievement	Target	Achievement	Target	Achievement	Target	Achievement
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Rice	93.00	71.82	93.00	88.53	93.50	83.13	87.80	91.79	92.80	93.35	460.10	428.62
Wheat	78.00	65.76	78.00	72.15	79.50	68.64	75.53	69.35	75.53	75.81	386.56	351.71
Coarse Cereals	33.00	26.07	34.00	37.60	36.80	33.46	36.52	34.06	36.52	33.92	176.84	165.11
Pulses	16.00	11.13	15.00	14.91	15.30	13.13	15.15	13.39	15.15	14.20	76.60	66.76
Foodgrains	220.00	174.77	220.00	213.19	225.10	198.36	215.00	208.60	220.00	217.28	1100.10	1012.20
Oilseeds	27.00	14.84	24.70	25.19	26.20	24.35	26.58	27.98	29.40	24.29	133.88	116.65
Sugarcane	320.00	287.38	320.00	233.86	270.00	237.09	237.50	281.17	270.00	355.52	1417.50	1395.02
Cotton #	15.00	8.62	15.00	13.73	15.00	16.43	16.50	18.50	18.50	22.63	80.00	79.91
Jute & Mesta@	12.00	11.28	12.00	11.17	11.80	10.27	11.28	10.84	11.28	11.27	58.36	54.83

- Screenshots of the two major sources of structured data have been given below



Hiren T Modi -12110077 Paras Jain -12110032
Sarabjit Rangi-12110011 Vaibhav Sethi -12110105

Data Cleaning and preprocessing – Group 24 Assignment



Challenges faced during download/ crawl/ collect data from all the sources:

1. Data seed not available in a single URL.
2. Not all crops' data is recorded in India.
3. Need to calibrate Scraper and Crawler for each website.
4. Most of data available in pdf for this we must use scraper that can extract data from pdf.
5. Numerous data sources required for data collection.
6. Different names for different crops which hampers data collection process.
7. Need to merge large number of different files to consolidate data in one file.
8. Google search and amazon(prices) not allowing web crawlers and banned our IP address.
9. Most data sources include only production information and not include other attributes.
10. Some states do not record Agriculture data or record data for some most common crops grown in their states.
11. Non standardised structure for attributes of varied crops like wheat data is different than apple data which is again different than herbs data.

BIGGEST CHALLENGE WAS THE LACK OF A COMMON ATTRIBUTE IN THE DATA TO ENABLE AN EASY WAY TO MERGE THE DATA. THIS MEANT THAT OUT CLEANING PROCESS WOULD BE EXTREMELY DIFFICULT.

Hiren T Modi -12110077
Sarabjit Rangi-12110011
Paras Jain -12110032
Vaibhav Sethi -12110105

4 - Convert data from original sources (Webpages, pdf files, CSV files, ...) to structured data fields:

- Most of our information was either directly converted from PDF file to excel or CSV files or was directly downloaded as CSV files from the Internet.
- This meant that the data was structured in a semi tabular or tabular format already and we did not have to necessarily work on highly unstructured data.
- However the semi structured to structured data that we were working with had a lot of cleaning that was necessary because none of the column attributes were the attributes that we could match in theme with our master seed data set and none of the Rose could match the rows so as to be able to merge with the seed data set.
- The two most important Python libraries that we found helpful were the tabula library and the PDFtables library.
- The next part which was important was to be able to clean the data so as to merge it into a single large master file and to further subdivide the master file into smaller chunks so as to be able to derive insights from it.

5 - Data cleaning/pre-processing as needed

Data cleaning and pre-processing turned out to be the biggest challenge for our group.

We then decided to split the entire Indian crops data set into two major files first set would contain all the data that we were able to confidently scrape and download and merge the second set would contain data about horticulture crops, fruits, vegetables alongside data of the principal crops however this set had a lot of sparse values and merge of these data sets wasn't the most easy task.

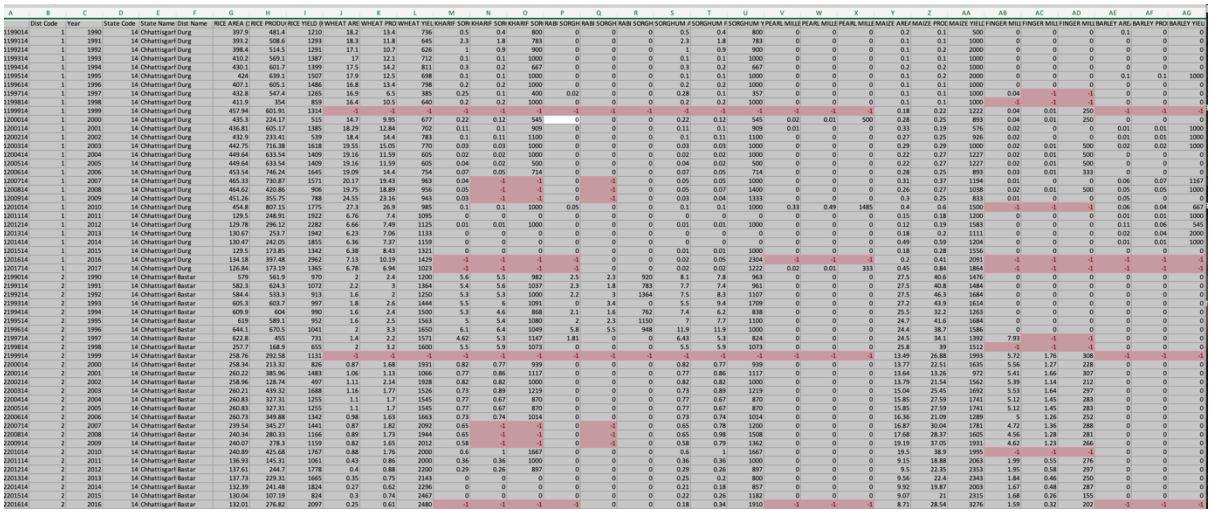
Following word the most obvious points that we had to take care of while trying to clean and pre process our data sets.

The data sets could be merged only after they were completely cleaned and were credible sources of information individually.

Cleaning techniques used and pre-processing required.

1) In all our data sources -1 was present as a value which indicated that information wasn't relevant for that geographic location. The -1 value was handled by removing the entire value and keeping that specific cell as 0. The reason why the cell was left 0 is because it would help in future aggregation of data in case we were to try and determine the average of the column or the sum of that entire column.
2) The second error was the "." . The "." error straight away indicated the absence of a value and the inability of that data source to provide that information. The simplest solution to this problem was to eliminate that entire cell and state quite clearly as a blank cell so that it does not hinder future calculations.
3) The third set of errors was the completely blank cell which would also include the combination of all the "." errors that were remove in the earlier correction. Python was able to assign end a value to all these data cells virtual totally blank however aggregation calculations had to be done before we allowed Python to assign any values to these cells.
4) Attached below are the screenshots of these common error types that we notice.

Data Cleaning and preprocessing – Group 24 Assignment



ite Code	State name	District code	District nam	Year	Potato Area	Potato Prodi	Sweet Potat	Sweet Potat	Onion Area (	Onion Produ	Green onion	Green onion	Tomato Are	Tomato Proc	Ladys finger	Ladys finger	Brir
1	Andhra Pradi	44	Srikakulam	2007	0	0	0.06	1.24	1.6	27.25	.	.	0.79	15.03	0.59	8.82	
1	Andhra Pradi	44	Srikakulam	2008	0	0	0.11	2.12	1.39	23.56	.	.	0.74	14.7	0.69	10.28	
1	Andhra Pradi	44	Srikakulam	2009	0	0	0.11	2.2	1.6	27.2	.	.	0.79	15.03	0.59	8.82	
1	Andhra Pradi	44	Srikakulam	2010	0	0	0.04	0.81	2.62	44.57	.	.	6.86	137.26	2.82	42.31	
1	Andhra Pradi	44	Srikakulam	2011	0	0	0.04	0.81	2.66	45.24	.	.	6.97	139.32	2.86	42.94	
1	Andhra Pradi	44	Srikakulam	2012	0	0	0.02	0.46	4.6	82.8	.	.	5.89	117.82	2.59	38.82	
1	Andhra Pradi	44	Srikakulam	2013	0	0	0.03	0.51	5.11	91.91	.	.	7.26	145.17	2.99	44.81	
1	Andhra Pradi	44	Srikakulam	2014	0.01	0.3	0.03	0.56	0.57	10.17	.	.	0.73	16.84	0.31	3.13	
1	Andhra Pradi	44	Srikakulam	2015	0	0	0.03	0.62	0.64	11.53	.	.	0.28	7	0.38	5.7	
1	Andhra Pradi	44	Srikakulam	2016	.	.	0.03	0.65	0.64	14.1	.	.	0.31	28.08	0.39	6.7	
1	Andhra Pradi	45	Visakhapatar	2007	0.25	4.94	0.24	4.8	0.42	7.14	.	.	2.2	41.88	1.74	26.03	
1	Andhra Pradi	45	Visakhapatar	2008	0.15	2.9	0.24	4.8	0.51	8.59	.	.	2.31	46.26	1.74	26.03	
1	Andhra Pradi	45	Visakhapatar	2009	0.28	5.6	0.24	4.7	0.29	4.9	.	.	1.62	30.78	1.85	27.75	
1	Andhra Pradi	45	Visakhapatar	2010	0.16	3.16	0.08	1.58	0.73	12.48	.	.	3.2	64.08	3.11	46.59	
1	Andhra Pradi	45	Visakhapatar	2011	0.16	3.21	0.08	1.61	0.75	12.67	.	.	3.25	65.04	3.15	47.29	
1	Andhra Pradi	45	Visakhapatar	2012	0.3	6	0.05	0.9	1.29	23.18	.	.	2.75	55	2.85	42.75	
1	Andhra Pradi	45	Visakhapatar	2013	0.33	6.66	0.05	1	1.43	25.73	.	.	3.39	67.77	3.29	49.35	
1	Andhra Pradi	45	Visakhapatar	2014	0.38	7.6	0.08	1.51	0.71	11.28	.	.	2.89	57.77	2.29	34.34	
1	Andhra Pradi	45	Visakhapatar	2015	0.3	1.8	0.08	1.51	0.71	12.68	.	.	2.89	57.77	2.29	34.34	
1	Andhra Pradi	45	Visakhapatar	2016	0.53	7.95	0.03	0.53	0.11	2.49	.	.	1.72	154.98	0.59	10.1	
1	Andhra Pradi	46	East Godavai	2007	0	0	0.05	1.06	0.59	10.1	.	.	1.08	20.46	2.05	30.81	
1	Andhra Pradi	46	East Godavai	2008	0	0	0.05	1.06	0.5	8.45	.	.	1.45	29	2.05	30.81	
1	Andhra Pradi	46	East Godavai	2009	0	0	0.05	1.06	0.5	8.52	.	.	0.72	13.59	1.48	22.14	
1	Andhra Pradi	46	East Godavai	2010	0	0	0.04	0.81	1.03	17.44	.	.	2.93	58.65	2.7	40.48	
1	Andhra Pradi	46	East Godavai	2011	0	0	0.04	0.82	1.04	17.7	.	.	2.98	59.53	2.74	41.09	
1	Andhra Pradi	46	East Godavai	2012	0	0	0.02	0.46	1.8	32.4	.	.	2.52	50.34	2.48	37.14	
1	Andhra Pradi	46	East Godavai	2013	0	0	0.03	0.51	2	35.96	.	.	3.1	62.02	2.86	42.87	
1	Andhra Pradi	46	East Godavai	2014	.	.	0.09	1.72	0.73	11.36	.	.	1	20	1.5	19.55	

5) After the removal of these problems the most important part turned out to be trying to find an attribute upon which all these different files could be merged. This process turned out to be an extremely difficult task since we could have chosen to go with either crops as names of rows attributes as names of rows or even the different years, states, districts as the names of the row.
6) This step was extremely important because if we were not able to join all the data sources present in the combined knowledge base we will never be able to make a comprehensive data set.
7) We decided to give each state a specific code and each district within a state a specific code. The reason behind this is that we wanted to create a common attribute column the common attribute column would be a concatenated string in other words a key of the district code year code and the state code respectively.
8) The combination of these three entities was always unique regardless of the data source that we were looking at and this could become an important factor while trying to merge all the data sources. We then calibrated our process of cleaning in a way that each and every file would contain the following three columns as their first columns.

Hiren T Modi -12110077
Sarabjit Rangi-12110011
Paras Jain -12110032
Vaibhav Sethi -12110105

	A	B	C	D	E	F
	Key	State Code	State name	District code	District nam	Year
2	4420071	1	Andhra Prade	44	Srikakulam	2007
3	4420081	1	Andhra Prade	44	Srikakulam	2008
4	4420091	1	Andhra Prade	44	Srikakulam	2009
5	4420101	1	Andhra Prade	44	Srikakulam	2010
6	4420111	1	Andhra Prade	44	Srikakulam	2011
7	4420121	1	Andhra Prade	44	Srikakulam	2012
8	4420131	1	Andhra Prade	44	Srikakulam	2013
9	4420141	1	Andhra Prade	44	Srikakulam	2014
10	4420151	1	Andhra Prade	44	Srikakulam	2015
11	4420161	1	Andhra Prade	44	Srikakulam	2016
12	4520071	1	Andhra Prade	45	Visakhapatar	2007
13	4520081	1	Andhra Prade	45	Visakhapatar	2008
14	4520091	1	Andhra Prade	45	Visakhapatar	2009
15	4520101	1	Andhra Prade	45	Visakhapatar	2010
16	4520111	1	Andhra Prade	45	Visakhapatar	2011
17	4520121	1	Andhra Prade	45	Visakhapatar	2012
18	4520131	1	Andhra Prade	45	Visakhapatar	2013
19	4520141	1	Andhra Prade	45	Visakhapatar	2014
20	4520151	1	Andhra Prade	45	Visakhapatar	2015
21	4520161	1	Andhra Prade	45	Visakhapatar	2016
22	4620071	1	Andhra Prade	46	East Godavar	2007
23	4620081	1	Andhra Prade	46	East Godavar	2008
24	4620091	1	Andhra Prade	46	East Godavar	2009
25	4620101	1	Andhra Prade	46	East Godavar	2010
26	4620111	1	Andhra Prade	46	East Godavar	2011
27	4620121	1	Andhra Prade	46	East Godavar	2012
28	4620131	1	Andhra Prade	46	East Godavar	2013
29	4620141	1	Andhra Prade	46	East Godavar	2014

28 +

9) All the days data sets slowly started taking shape when we were able to assign a specific keys column and luckily even the sparse data sets for horticulture crops had similar district codes state codes and your codes respectively.
10) The key to all the districts States and the years has also been attached below for reference.

Hiren T Modi -12110077
Sarabjit Rangi-12110011
Paras Jain -12110032
Vaibhav Sethi -12110105

State Code	State Name	District Code	District Name
1	Andhra Prad	44	Srikakulam
1	Andhra Prad	45	Visakhapatnam
1	Andhra Prad	46	East Godavari
1	Andhra Prad	47	West Godavari
1	Andhra Prad	48	Krishna
1	Andhra Prad	49	Guntur
1	Andhra Prad	50	S.P.S.Nellore
1	Andhra Prad	51	Kurnool
1	Andhra Prad	52	Anantapur
1	Andhra Prad	53	Kadapa YSR
1	Andhra Prad	54	Chittoor
1	Andhra Prad	503	Vizianagaram
1	Andhra Prad	504	Prakasam
2	Bihar	902	Muzaffarpur
2	Bihar	903	Darbhanga
2	Bihar	904	Saharsa
2	Bihar	905	Purnea
2	Bihar	906	Saran
2	Bihar	907	Patna
2	Bihar	908	Mungair
2	Bihar	909	Bhagalpur
2	Bihar	912	Gaya
2	Bihar	918	Nalanda
2	Bihar	919	Nawada
2	Bihar	920	Aurangabad

11) After assigning these specific code and the keys to the various data sets we merged using the python's left join on our key column.
12) The Python code for the left join merge and also the entire CSV for the codes we had given is available in the repository.
13) The process hence gave us six different data sets that we could use the information about each of these data sets have been given below

- SET 1 - This set contains data of all the principal crops from the year 1990 to the year 2017 across all the States and its districts. Name : G24_PC_D

- SET 2 - This set contains data of all principal crops as well as horticulture crops, vegetables, fruits and other crops and spices across all the States and districts. The data in this set is sparse and hence could not be merged with SET 1/ Name : G24_OC_D

- SET 3 - This data set contains pivoted information of all the states from the year 1990 to 2017 for the principal crops Name : G24_PC_S

- SET 4 - This data set contains pivoted information of all principal crops as well as horticulture crops, vegetables, fruits and other crops and spices across all the states year wise. Name: G24_PC_S

- SET 5 - MASTER DATA SET - Name: G24_Master_D

- SET 6 - MASTER DATA SET PIVOTED BY STATE AND YEAR - Name : G24_Master_S

The master data set is a combination of all 28 smaller seed data sets that we worked on and also is accumulative knowledge base for the entire data that we were able to collect the master set has also been pivoted into states and the information across each of the years is for the that specific state year wise

14) The two data sets under consideration have also been pivoted based on their respective States and years because we may not always require data that is subjected to each and every district of the states across the years of consideration.

15) A total of four data sets have hence been uploaded in the final Jason format

Hiren T Modi -12110077
Sarabjit Rangi-12110011
Paras Jain -12110032
Vaibhav Sethi -12110105

Data Cleaning and preprocessing – Group 24 Assignment

All the data sets have been stored in the repository in the dataset folder as the JSON format required.

- The main data set that we are using for our analysis and consideration is set number 1 which contains data of all the districts and states from the 1990 to 2017 for the principal crops for which the government always provides credible and elaborate data across all their statistical forums.
- Python code for all the merge activity and has been stored in the code folder of the repository.

Hiren T Modi -12110077
Sarabjit Rangi-12110011
Paras Jain -12110032
Vaibhav Sethi -12110105

6 - Observations/ Insights and Analysis on the data collected

Insight 1:

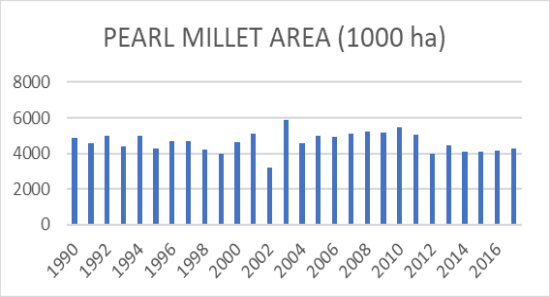
Referring to State Andhra Pradesh, we observed that Rice Production increased by approx. 35 % from 1991 to 2017, however production area got reduced by 15 %.

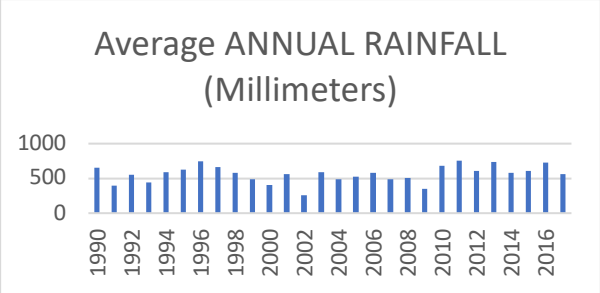
We hypothesised, the rationale behind could be increased consumption of fertilisers, improvised techniques of agriculture & irrigation

Improved consumption of fertiliser can be verified from data frame.

Insight 2:

Referring to State Rajasthan, we observed that millet production is increased to 828% in 2003 as compared to 2002.



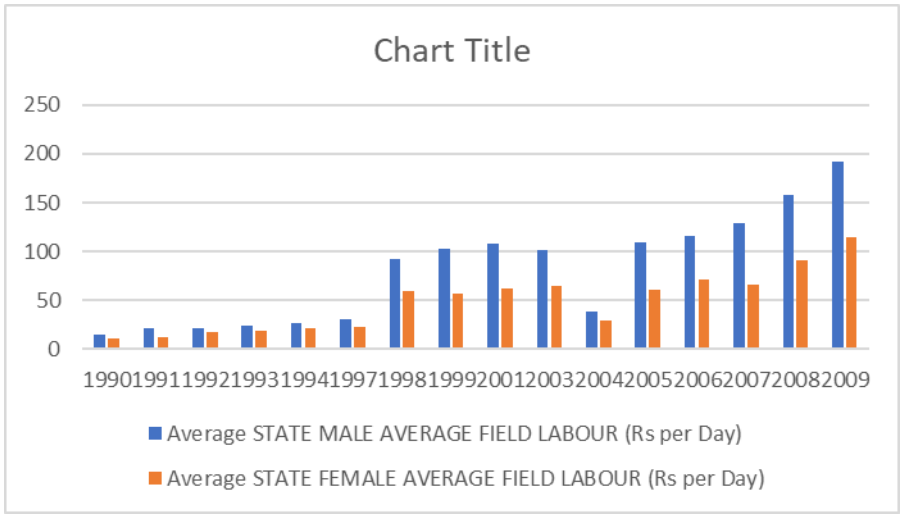


We hypothesise, the possible reason for this could be low rainfall in 2002, so people started sowing pearl millet as it requires less water.

Insight 3:

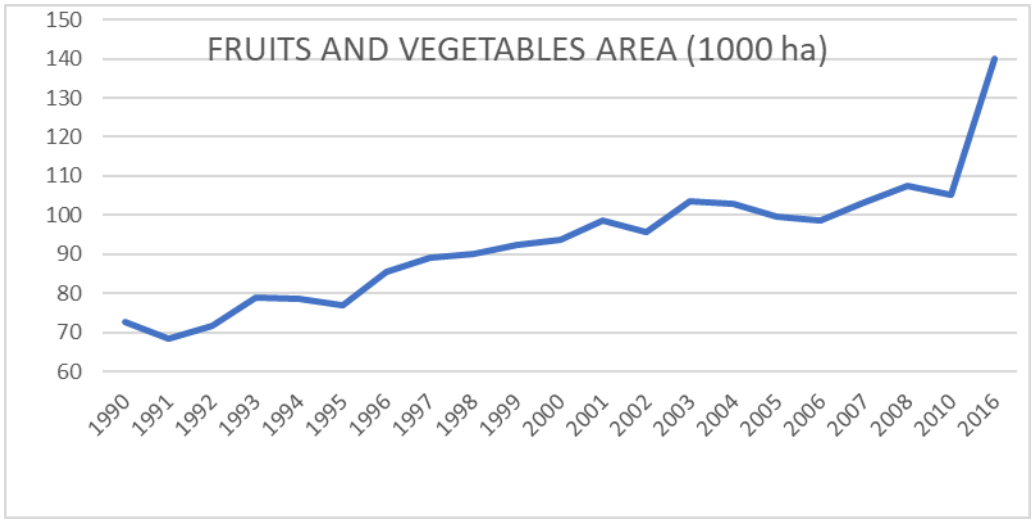
Referring to state Haryana, we observed that, difference in average wages between male and female is increasing year in year basis.

Hiren T Modi -12110077
Sarabjit Rangi-12110011
Paras Jain -12110032
Vaibhav Sethi -12110105



Insight 4:

Referring to State, Himachal Pradesh we observed that FRUITS AND VEGETABLES AREA (1000 ha) is linearly increasing on YoY basis.



Lots of more analysis can be done from the permutation and combination of varied attributes. The insights would be price less.

Hiren T Modi -12110077
Sarabjit Rangi-12110011
Paras Jain -12110032
Vaibhav Sethi -12110105

7 - Strategy to enhance the data with crowd sourcing methods:

Crowd Sourcing - Definition:

A Definition in Context Crowdsourcing is when information is sourced from a group of people in response to an open call, a request for specific information, or for an exchange, organized by a central organizer/ organizing body. In the context of this paper, the crowd generally refers to smallholder farmers in rural India, although it can also apply to other players in the value chain, such as buyers, warehouses, transporters, etc.

Crowdsourcing was first coined in 2006.1 Worldwide, it commonly refers to outsourcing a function done locally to a large, disconnected group of people. It broadly collectivizes a role otherwise done by an individual or small group by using ICT. A well-known example is the online platform Kickstarter. It facilitates the sourcing of capital for creative projects from the general public. Since 2009, it has fuelled $600 million in funding for more than 100,000 projects.

Crowd Sourcing - Process:

When considering if crowdsourcing is applicable for potential new initiatives it is therefore necessary to ask yourself the following questions:

1. What is the fundamental role to be done or replaced by a group?

2. Do target users have access to and literacy in using mobile phones?

3. What is the incentive for people, aka 'the crowd,' to participate?

crowdsourcing applications can be used within different agricultural development initiatives, particularly those that are working with smallholder farmers. They include increasing farmer access to information, promoting market access and farmer collectivization, tracking pest outbreaks, and sharing weather information, a few examples & not a comprehensive list.

Indian Context:

In India, Kisan call centres (toll free number **1800-180-1551**) are already active & we can make use of them to source information from farmers like their **farm size, crops grown, fertilizer used, any prevalent crop disease in the area**, in exchange of vital information like weather data, best crop to grow scientifically on the soil available on the farmer's land. This way we can very easily source the information for our project in future.

Detailed Process To be Followed:

Tracking Pest and Disease Outbreaks Crop loss from pest and diseases are severe threats to smallholder farmers. Yet, tracking these outbreaks before they have already spread and tracking their movements in a timely manner once there is a known problem can be challenging. **Farmers can greatly benefit from knowing the status of pests and diseases in their region**. When they are informed about outbreaks as they unfold, government agencies, NGOs, and farmers can act quickly on the ground to head them off before they spread. Crowdsourcing the reporting of outbreaks therefore, holds significant potential. To date, crowdsourced pest and disease services have been value added offerings to existing relationships and/or as part of a greater integrated service model. For instance, the Centre for Agricultural BioScience International (CABI) launched **Plant wise**, an initiative offering farmers crop and pest diagnostic tools and

other resources. Plant wise collects information on soil health and other risk factors**. In exchange for participating by providing data, farmers receive technical assistance via SMS and voice messages on how to remove or avoid pests, thereby reducing crop loss**. Plant wise has crowdsourced the part of the role of its agricultural technicians, a role previously completed by a few individuals, now done by thousands of farmers in 24 countries, including DR Congo, Ghana, Kenya, Rwanda, Sierra Leone, and Uganda. Instead of sending a technician into the field to diagnose plant disease, farmers are trained by agricultural experts to identify and transmit this data to Plant Wise's central database, called the Knowledge Bank.

Verifying Local Weather Information Access to accurate weather information is highly valuable to farmers. While predictive models are improving, hyper localized weather information in much of rural Africa is still often spotty at best. National governments and international organizations are, however, beginning to crowdsource the collection of localized weather information to verify forecast accuracy and to improve their modelling. In Zambia, for example, the Meteorological Department's RANET project uses crowdsourcing to collect local rainfall and weather information. Weather data is collected by remote sensors distributed to 3,000 community.

Plant wise Knowledge Bank (2013) 5 Grameen Foundation & Palantir: Partners for Food Security (2012) members in rural areas.6 Locals, including farmers, are trained to read and remit data to a central database via mobile phones. This weather information is later disseminated back into the community, broadcast by community radio stations. While a valuable service for communities, weather information is also desired by planners and policy makers. Since benefits to farmers can often seem to be indirect, **RANET incentivizes participation by providing free mobile phones and subsidizing air time**. Oxfam in Eastern Ethiopia administers another example. This SMS-based effort monitors water points and provides early drought warnings. Participating farmers enter data into a phone with specially designed survey software. Air time is subsidized, facilitating easy transmission of water availability in the region.

In exchange for providing data, farmers receive local water information, reducing their travel in drought conditions.

Another example is iCow, a mobile application from the Kenyan app development Green Dream Tech Ltd., that helps 42,000 dairy farmers daily to track livestock.

iCow aggregates and maps its crowdsourced data, helping dairy farmers and others find one another by accessing a publicly-available map online or sending an SMS to iCow with a location request . It also increases bargaining power by facilitating collective buying of inputs from suppliers in bulk and enabling larger sales volumes in aggregate.

Incentivizing the Farmers:

Nothing works like Cash or near to cash benefits, hence we propose to incentivize the Indian farmers as below:

1. In exchange of vital information, we can compensate farmers by providing them with talk time on their mobiles.
2. Another factor that can motivate the farmers is to provide them with technical assistance like SMS & voice message on how to remove or control pests, thereby reducing crop loss.
3. In exchange for providing data, farmers may receive local water information, reducing their travel in drought conditions.

Challenges:

1. Ensure Quick Sign-up & Usability The online user interface should be simple and the sign-up process fast. In the context of smallholder farmers, mobile phone registration through SMS or voice is highly favoured over e-mail or websites. Participant self-registration by sending an SMS can also reduce the practitioner's logistical burden of registration.
2. Build Local Partnerships It is valuable to coordinate crowdsourcing initiatives with local entities and governments. This is particularly important for anything that could be deemed a public good, such as weather, pest and disease outbreaks, and market prices. Sharing this information with the government, can empower them to make decisions and form policy on data and proof what is really going on.
3. Establish Local Physical Presence A physical presence encourages adoption and ongoing participation in crowdsourcing efforts. Programs that offer technical assistance as an incentive, however, do face higher costs and require on-the-ground infrastructure. Local training is a cost, but also an opportunity, as a means to reach remote people from whom data is often difficult to collect. This presence augments the crowdsourcing effort in a number of ways.
4. Human Resources One of the greatest challenges for crowdsourcing initiatives targeting smallholder farmers is ensuring skilled expertise
5. Data Quality Crowdsourcing in development can provide access to a breadth and depth of the population rarely available for the same cost with other mechanisms.
6. Encourage Farmer Participation Programs interested in farmer participation grapple with a common challenge: older farmers are often more reticent to participate than younger farmers in ICT-based efforts. Lack of older farmer participation can be problematic, as they have specific knowledge and experience that could be overlooked. Failing to record their input can adversely affect the quality of information, and thereby also adoption rates. To address this issue, iCow introduced a free customer service phone line that responds to questions in real time, a service they found encourages participation by more farmers than SMS services alone.
7. Monitoring and Evaluation It is difficult for crowdsourcing programs to measure and evaluate impact because the cost for mobile airtime leaves farmers little incentive to provide feedback. To gather feedback via mobile, programs will generally need to reimburse this cost. Beyond the cost of airtime, other barriers exist to using mobile phones alone for robust M&E efforts, though applications do exist to facilitate with this process.

Hiren T Modi -12110077 **Paras Jain -12110032**
Sarabjit Rangi-12110011 **Vaibhav Sethi -12110105**

Data Cleaning and preprocessing – Group 24 Assignment

Some Examples of Crowd Sourcing From Real World:

Example 1:

Crowdsourcing with ClimMob:

How it works?

- Statistical approaches:
  Each farmer receives a package of three different varieties. The farmer has to note which of the three is best and which worst on a list of characteristics that they develop together with the researchers. The varieties are drawn from a pool of several varieties, so while one farmer receives A, B and C, another receives A, B and D and so on.
- Even though no farmer compared A and D directly, statistical methods can reveal whether A or D is better. Additional variables, such as whether a farmer
- has access to irrigation, or the altitude of the plot, can also be examined to see whether they affect the performance of the varieties.
- An additional benefit is that the varieties are grown in the farmers' fields rather than a trial plot, allowing a greater number of farmers to take part, and to capture other data such as performance at different altitudes or in varying climatic conditions.

- Looking ahead, ClimMob can be used to gather big data on farmers' varietal preferences, and to share that information with relevant actors to create a 2-sided business platform allowing for small quantities of a diverse selection of planting materials to be marketed to targeted consumers.

Example 2:

Crowdsourcing Knowledge: An Extension Approach for Remunerative and Sustainable Home Garden Farming Systems in Kerala

Kerala has pioneered development models through farmer participatory approaches. Involving the grassroots in scientific studies can help improve crop variety adoption and refined technology recommendations. Participatory tools like Participatory Rural Appraisal (PRA), Participatory Technology Development (PTD), On Farm Testing (OFT), Front Line Demonstrations (FLD) and Participatory Breeding Programmes (PBP) are all time-tested examples of its successful use in researches. Crowdsourced citizen science approach called tricot – 'triadic comparisons of technologies' is popular today, where farmers are made to adopt three crop varieties or technologies randomly assigned to them from a broader set of varieties/technologies for final choice aimed at continuous adoption. The results of this study conducted during 2019-2020 revealed that more than 80% of the farmers fully adopted the technology prescribed in the checklist as a result of crowdsourcing knowledge. The results on attitude of farmers towards crowdsourcing revealed that majority of the farmers possessed favourable attitude towards crowdsourcing approach.

Hiren T Modi -12110077 Paras Jain -12110032
Sarabjit Rangi-12110011 Vaibhav Sethi -12110105

Example 3:

Citizen Scientist Approach in India:

- Farmers in three countries, including in India, have turned into citizen scientists, helping generate data on crop varieties that adapt best to potential climatic changes.
- In a study spanning India, Nicaragua and Ethiopia, researchers have demonstrated how farmers' involvement in scientific studies can improve and accelerate crop variety recommendations.
- Scientists applied a nifty crowdsourced citizen science approach called tricot – triadic comparisons of technologies – in which each farmer plants seeds of three crop varieties randomly assigned to them from a broader set of varieties. The farmer then ranks the varieties according to different characteristics such as early vigour, yield, and grain quality. In other formats for on-farm experimentation, farmers only get snapshots of the crop. Another aspect is that farmers contribute with their land and effort, reducing the costs of the trials. Seeds are generally provided for free, but other formats of experimentation often require renting land.
- Trials were carried out between 2012 and 2016 during three cropping seasons in Ethiopia, five cropping seasons in Nicaragua, and four cropping seasons in India (Uttar Pradesh and Bihar).
- The rank-based feedback format allowed even those with low literacy skills to contribute their evaluation data through various channels, including mobile telephones. Scientists then linked the farmer-generated data with agroclimatic and soil data.
- We combined the data from farmers with data on the seasonal climate of each plot," Van Etten explained. "**Farmers told us when they planted, and we had GPS coordinates of their farms. This allowed us to link the relevant weather data that occurred during the trial on each plot. We then used statistical analysis to see how the seasonal climate influenced how well the varieties did on each farm."**

References:

https://agrilinks.org/sites/default/files/resource/files/Crowdsourcing_Applications_for_Agricultural_Development_in_Africa.pdf

https://www.bioversityinternational.org/innovations/seeds-for-needs/crowdsourcing/

https://scroll.in/article/921392/crowdsourcing-knowledge-indian-farmers-become-citizen-scientists-to-adapt-to-climate-change

8 - References and Sources used for this Assignment

- https://agrilinks.org/sites/default/files/resource/files/Crowdsourcing_Applications_for_Agricultural_Development_in_Africa.pdf
- https://www.bioversityinternational.org/innovations/seeds-for-needs/crowdsourcing/
- https://scroll.in/article/921392/crowdsourcing-knowledge-indian-farmers-become-citizen-scientists-to-adapt-to-climate-change
- https://www.fao.org/3/s2022e/s2022e02.htm
- https://agricoop.nic.in/en/statistics/state-level
- https://pmfby.gov.in/
- https://aps.dac.gov.in/LUS/Public/Reports.aspx
- https://indiadataportal.com/?language=&location=Haryana
- https://farmer.gov.in/BooKLetView/FarmerFriendlyMaterial.aspx?DocType=Booklets%20and%20Flyers&SCode=33
- https://eands.dacnet.nic.in/
- http://mospi.nic.in/agriculture-statistics
- https://world-crops.com/
- https://www.britannica.com/topic/list-of-herbs-and-spices-2024392
- https://en.wikipedia.org/wiki/List_of_Indian_spices
- http://www.indianspices.com/spices-development/spice-catalogue.html
- https://www.nmpb.nic.in/
- https://www.ibef.org/exports/spice-industry-indias.aspx
- http://www.spices.res.in/annual-report
- https://indianculture.gov.in/food-and-culture/spices-and-herbs
- https://www.spicesinc.com/t-list-of-spices.aspx
- https://www.spicesofindia.co.uk/acatalog/All-Herbs-and-Spices.html
- http://vanneman.umd.edu/districts/codebook/notecrop.html
- http://vdsa.icrisat.ac.in/Include/document/all-apportioned-web-document.pdf
- https://www.google.co.in/intl/en/about/products?tab=wh
- https://agricoop.gov.in/sites/default/files/agristatglance2018.pdf
- https://aps.dac.gov.in/
- https://data.gov.in/catalog/district-wise-season-wise-crop-production-statistics
- https://www.tn.gov.in/crop/stat.htm
- https://agricoop.nic.in/sites/default/files/pocketbook_0.pdf
- https://eands.dacnet.nic.in/FHP(District).htm
- http://mospi.nic.in/agriculture-statistics
- http://data.icrisat.org/dld/src/crops.html
- http://indiastat.com
- https://www.geeksforgeeks.org/how-to-convert-pdf-file-to-excel-file-using-python/
- https://mango-is.com/tools/csv-to-json/#resultsPane
- https://www.customguide.com/excel/pivot-table-multiple-columns

…. And many more