 ISSN: 2454-4248
193 – 199

DATA Analytics of Agriculture Production, Wages and Income in Rural Areas of India using Big Data and Python Matplot Lib

K. Raghavendra Swamy
Technical Trainer
BDPS PVT LTD.
VIJAYAWADA.
swamy1712@gmail.com

G. Venkataprasad
Assistant professor
Department of CSE
sphoorthy Engineering College
Hyderabad.
prasad.venkata8@gmail.com

ABSTRACT: Agriculture Sector is the major contribution in GDP growth rate of India and Most of the Rural India it will become major resource of Income generator it contains different sectors like paddy, poultry, fisheries, Milk, and other crops. In this paper we studied general, commercial, dairy and other related Agricultural out comes and their Incomes and wages.

In this paper we are performing different Data Analytics by taking parameters Daily wages, Income and production of Rural India. In this we are using Big Data Hive and Python Matplotlib to produce Graphical Analytical Reports. and finding results of different crops and daily wages of rural workers. The results we are finding year of production, crop wise production, crop wise and sector wise wages and Income of different crops.

In this paper we collected Data and sample Analytical Reports from Agriculture Statistics Ministry of Agriculture, Co operation & Farmers Welfare and Data.gov.in . we are revealing different Analytical Reports regarding wages, Income and Production.

Keywords *: Hadoop, Hive, python, MatplotLib*

*****

I. Introduction:

The Hadoop Frame work[1][2] supports different Big Data tools it is a large eco system contains Hive, pig ,Mahout, spark, flume, hbase, mongo ,impala and Extension Apache Spark . In Hadoop we have Map reduce program which will be used to process massive parallel execution Mapping and Reducing Data. The Hive working like as a SQL , it will minimize writing map Reduce programs . it is top on Hadoop environment it will generate execution plan and executed under Hadoop Map Reduce program. Python is a object oriented programming and supports open source modules we can develop any kind of applications with support of open source connectivity and modules.

II. METHODOLOGY:

In this paper we are mapping the same related attribute values as a group and performing aggregate functions in Hive and using Matplotlib modules to plot graphical reports in python.

Apache Hive :

Hive is a data warehouse software project built on top of Hadoop for providing data summarization, query, and analysis[3]. Hive gives an sql-like interface to query data stored in various databases and file systems that integrate with Hadoop. Traditional SQL queries must be implemented in the Map Reduce Java API to execute SQL applications and queries over distributed data. Hive provides the necessary SQL abstraction to integrate SQL-like queries (into the underlying Java without the need to implement queries in the low-level Java API. Since most data warehousing applications work with SQL-based querying languages, Hive aids portability of SQL-based applications to Hadoop[3].

Season Wise Production Analytics:

hive> select season,sum(production) from agri9 group by season;

2017-11-16 07:26:18,723 Stage-1 map = 100%, reduce = 0%
2017-11-16 07:26:28,111 Stage-1 map = 100%, reduce = 100%
Ended Job = job_201711160649_0001
OK

 ISSN: 2454-4248
193 – 199

Autumn 60413505
Kharif 4019624546
Rabi 2050580770
Summer 169091570
Whole Year 134397509428
Winter 416605045
Time taken: 18.72 seconds

Result: In the season of Rabi we have maximum production and in season of summer we have less production.

Year wise production Analytics:

hive>select year,sum(production) as totalpro from agri9 group by year order by totalpro desc;

2011 14303715404	2004 8168236155	2009 7656703299	2010 6302449249
2013 12902042178	2005 8041215684	2001 7464342486	1998 5825203507
2006 8679179768	2003 7914327046	2000 7447965265	1997 851232906
2014 8663337286	2008 7713066074	2007 6876791657	2015 6029574
2012 8169692424	2002 7694255067	1999 6434039835	

By observing output results in the year of 2011,2013 we have highest production Rate and in the year of 1997,2015 we have lowest production .

State wise production Analytics:

hive> select state,sum(production) as totalpro from agri9 group by state order by totalpro asc;

Chandigarh 63955	Odisha 142024428	Andaman and Nicobar Islands 718026736
Mizoram 1661278	Rajasthan 281320269	Karnataka 839336401
Dadra and Nagar Haveli 1847871	Telangana 335147930	Maharashtra 1263640603
Jharkhand 2257129	Bihar 366454569	West Bengal 1396437372
Sikkim 2435735	Haryana 381273890	Assam 2111751759
Manipur 5230917	Puducherry 384724502	Uttar Pradesh 3234492659
Arunachal Pradesh 6687736	Madhya Pradesh 448840736	Tamil Nadu 12076441353
Meghalaya 12112496	Goa 505755733	Andhra Pradesh 17324590296
Jammu and Kashmir 12141562	Gujarat 524291337	Kerala 97872227313
Tripura 12522917	Punjab 586385001	
Nagaland 12765950		
Himachal Pradesh 17805168		
Chhattisgarh 100951908		
Uttarakhand 132177355		

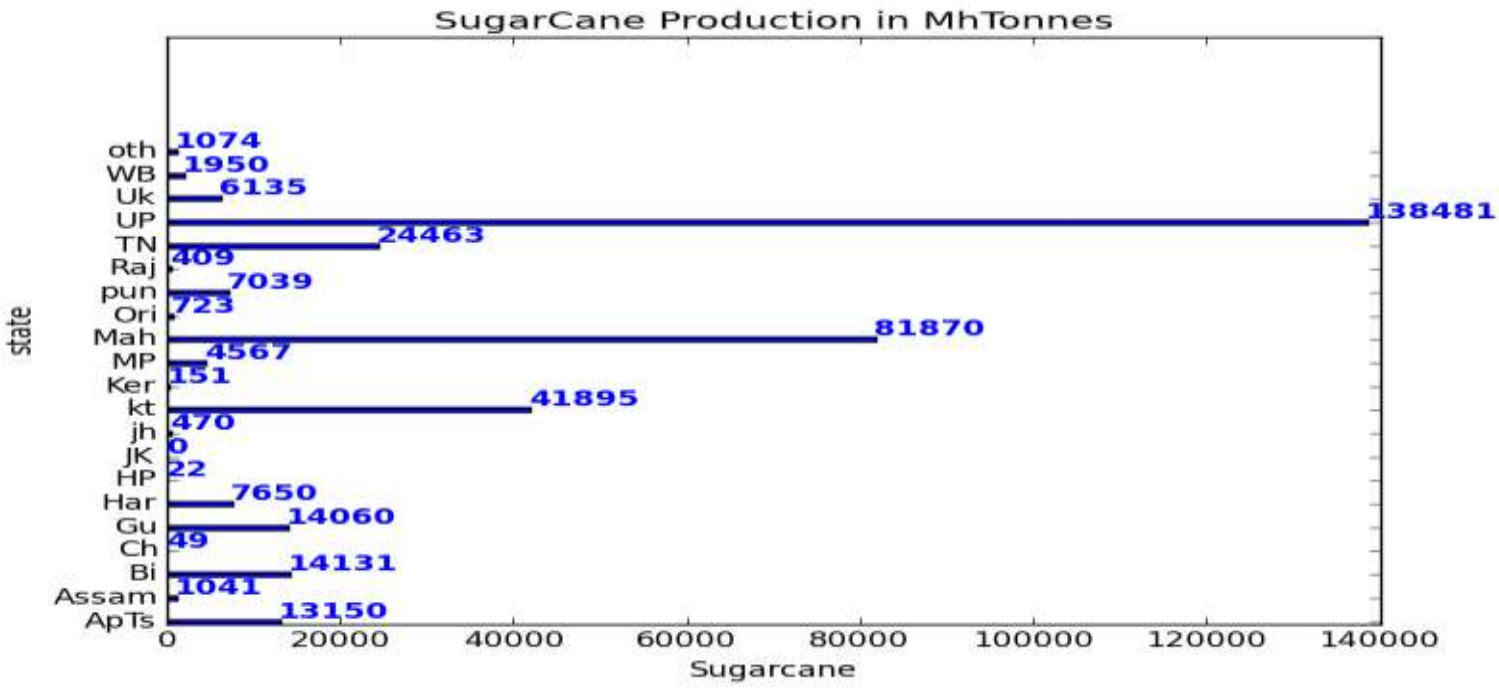
III. Result:

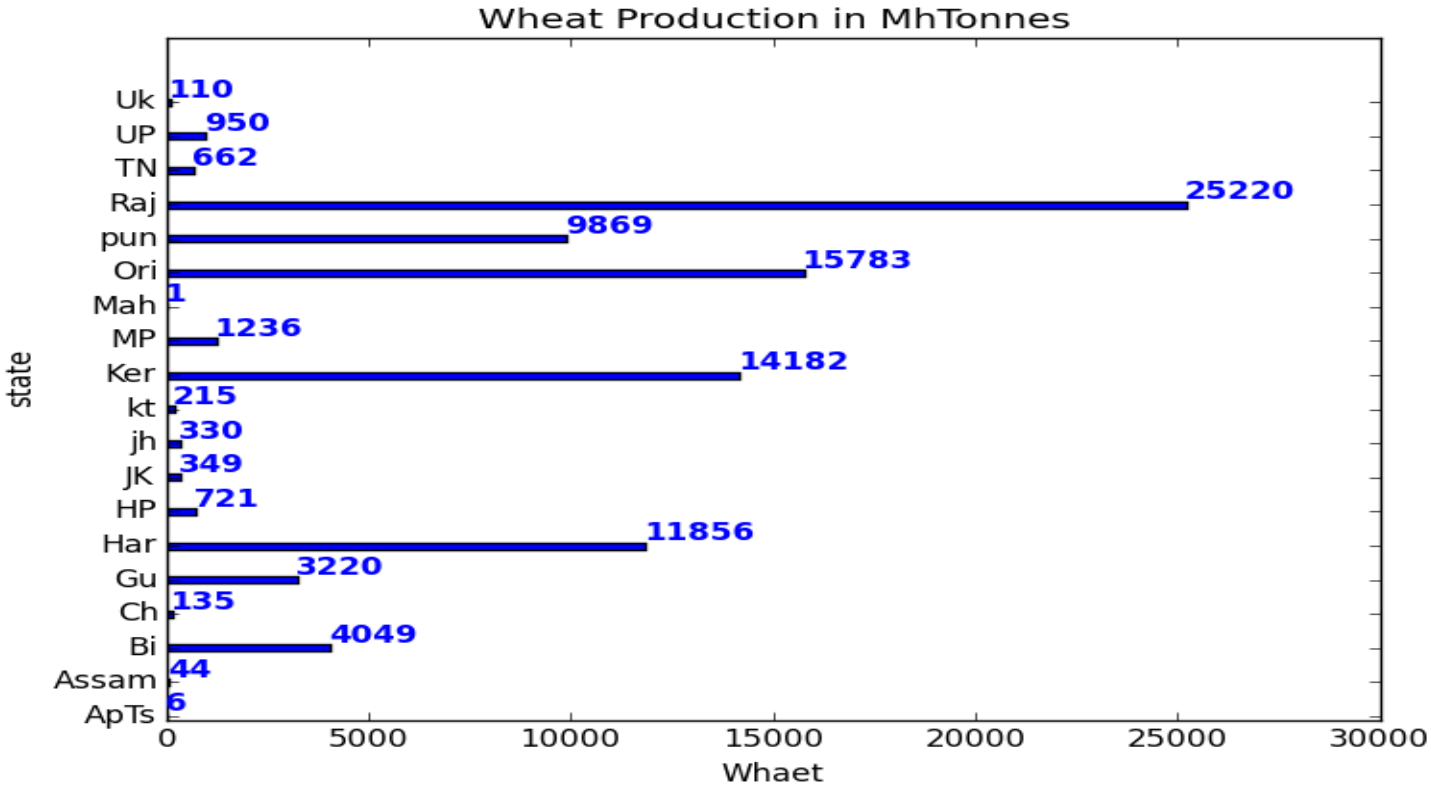
By Analysis of output Data Chandigarh, Mizoram, Dadra and Nagar Haveli, jharkhand having low production Rate and West Bengal, Assam, TamilNadu, AP, Kerala and UP having highest production Rate.

 ISSN: 2454-4248
193 – 199

Matplotlib[7][6] is a plotting library for the python programming language and its numerical mathematics extension Numpy. It provides an object oriented for embedding plots into applications using general-purpose.

SELECTED Item Wise Production Analytics using Python Matplotlib:





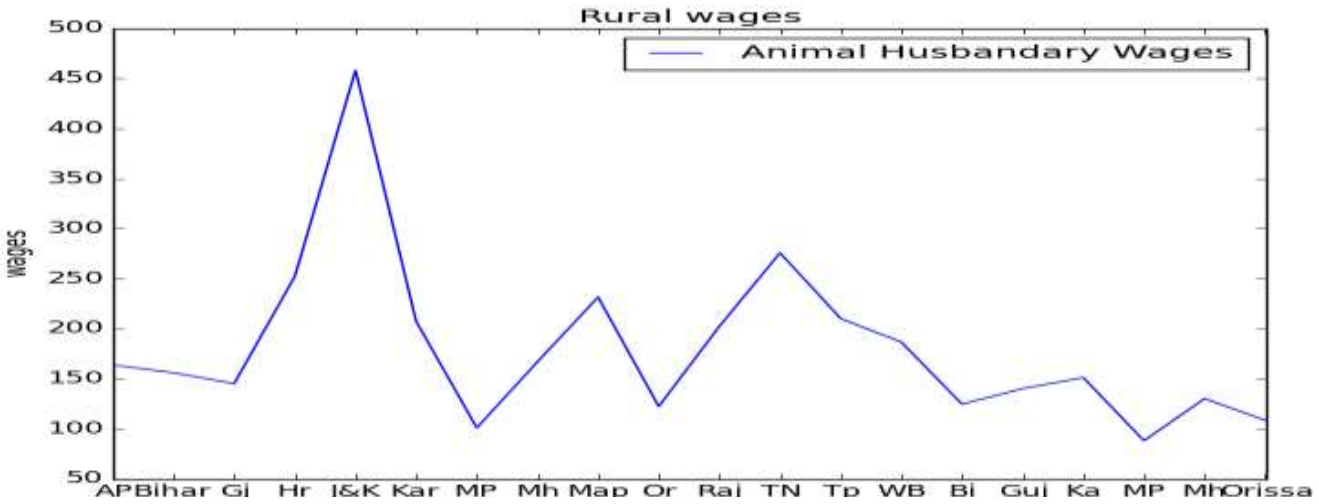
 ISSN: 2454-4248
193 – 199

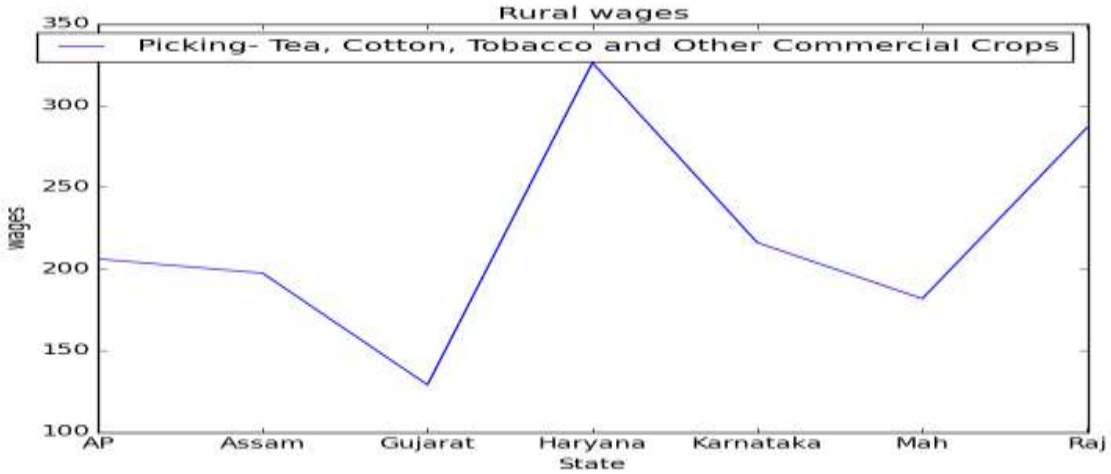
Sample Item wise Maximum and minimum production states Analysis by observing above plotting graphs.

Item	Maximum	Minimum
RICE	WB,UP,PUNJAB,AP&TS	Hp,gujarath,RAJASTHAN,UK
SUGARCANE	UP,MAHARASTAR,KARNATAKA	MP,CH,WB,MP
WHEAT	RAJASHAN,PUNJAB,ORISSA,KERALA	AP&TS,MAHARS,TN

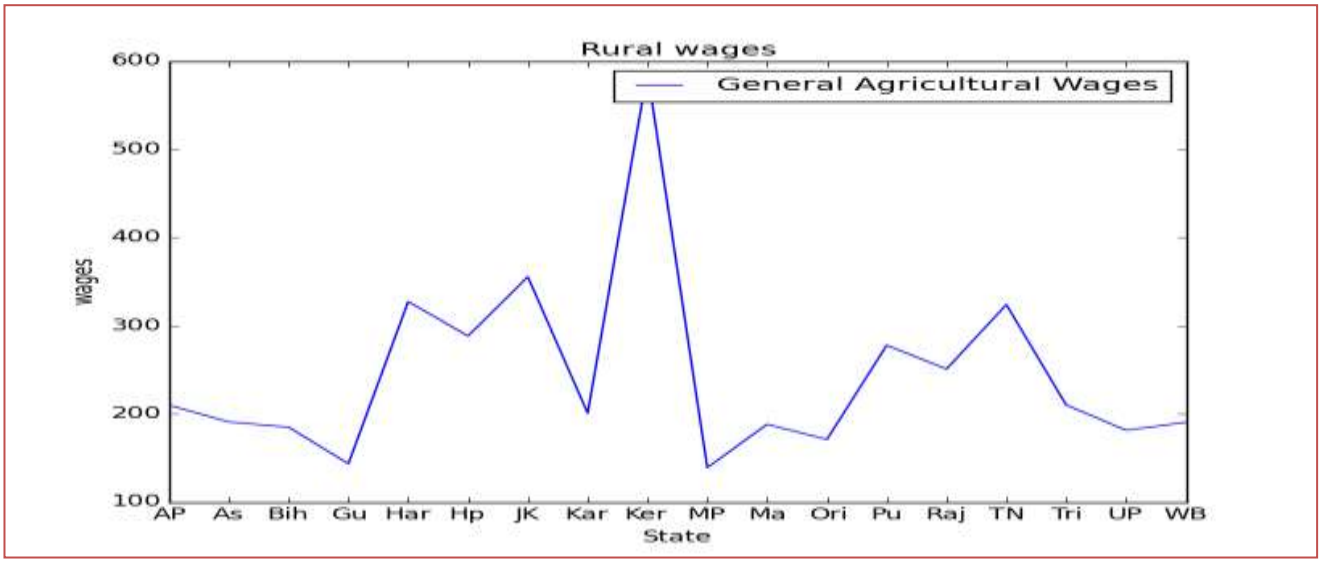
Agricultural Analytics by considering Daily Wages in different sectors:

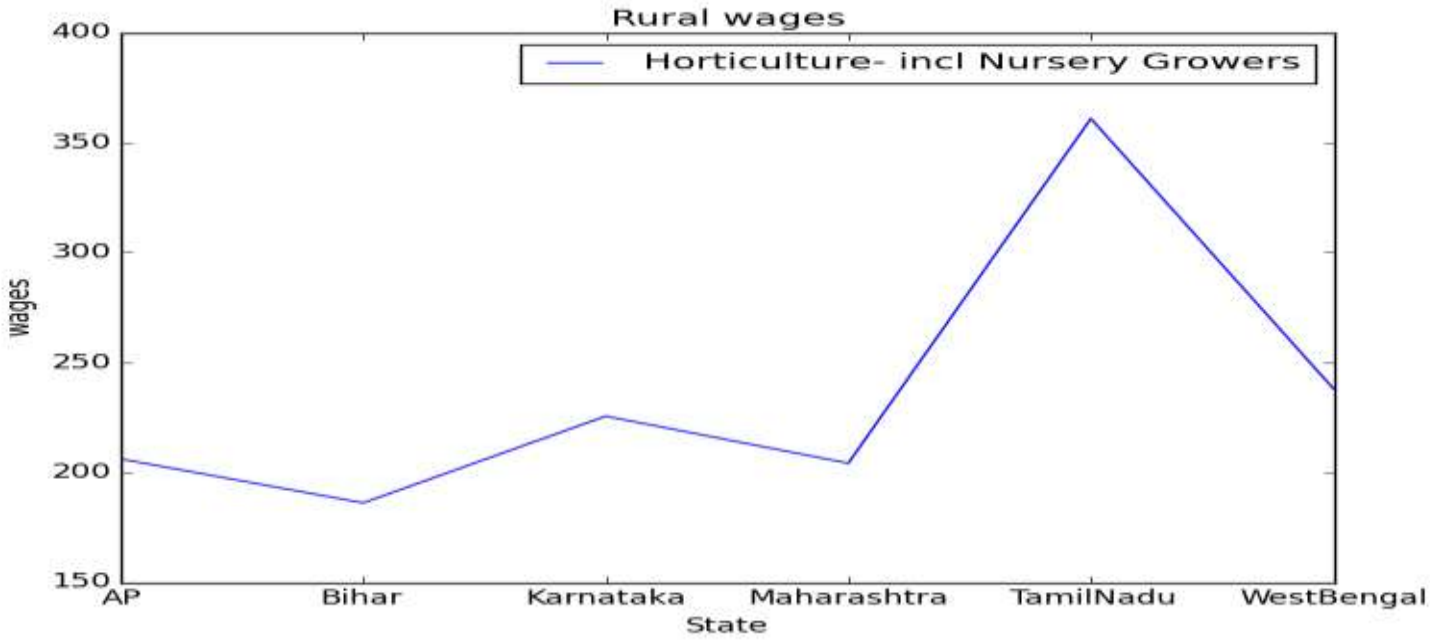
The daily wage rate data received from the different villages are first normalized for eight hours [4]working day and then the simple arithmetic average of these normalized daily wage rates is worked out. State-wise averages are restricted only to those occupations where the number of quotations is five or more in order to avoid inconsistency in wages paid to different categories of workers on account of difference in number of quotations. The average wage rates at all-India level are derived by dividing the sum total of wages of the measured states by the number of quotations. At all-India level also, the number of quotations for working out occupation-wise averages are restricted to five or more. State-wise as well as All-India average [4,5]daily wage rates in respect of agricultural and non agricultural occupations in Rural India have been furnished in data. Agricultural occupations include Animal Husbandry incl Poultry, Dairy and Herdsman, General Agricultural incl Watering and Irrigation, Harvesting, Winnowing, Threshing, Herdsman, Horticulture incl Nursery Growers and Fishermen Inland, Loggers and Wood Cutters, Packaging,Picking, Plant Protection of Pesticides and Treating Seeds, Ploughing, Tilling, Sowing including Planting, Transplanting and Weeding, Threshing, Transplanting, Weeding, Well Digging, Cane Crushing and Winnowing etc. Non agricultural occupations include Blacksmith, Carpenter, Bamboo and Cane Basket Weavers, Handicraft, Cobbler, construction includes Roads, Dams, Industrial and Project Construction,Well Diggers, Electrician, Plumbers, Porters, Loaders, Sweeper, Sweeping, Cleaning, Tractor Driver, Unskilled Labors[4][5]





 ISSN: 2454-4248
193 – 199





Selected wages Analytics from above graphs as shown below:

sector	HIGH WAGES	LOW WAGES
General	Kerala,Haryana,TN	Wb,Gujarath,Orissa,MP
Animal Husbandary	Haryana,j&k,kerala,TN	MP,ORISSA,Maharashtra
HortiCulture	TN	AP,Bihar
Commercial	Haryana	Gujarath

 ISSN: 2454-4248
193 – 199

Data Analytics on Income of Farmers:

we have collected Analytical data from Department of Stastics and Economics

we find that growth rates of total income in the decade have been highest in Haryana (8.3%), Rajasthan (8.1%) and Odisha (7.6%) while it is lowest in the states of Assam(-0.3%), Bihar(-0.8%) and West Bengal (-1.3%). Haryana's growth has largely come from incomes from cultivation (8.8%) while that of Rajasthan and Odisha has come through growth in incomes from livestock (45.1% and 36.1% respectively). In the low growth states, Assam has suffered deceleration in nonfarm business incomes (-7.8%) and wage incomes (-4%). Bihar's low income comes from deceleration in all sectors except wage income. West Bengal has seen major deceleration in incomes from cultivation (-5%) and nonfarm business (-2.2%) during this period.

STATES	INCOME FROM FARMING	INCOME FROM LIVESTOCK	FROM NONFARM BUSINESS	FROM WAGES/ SALARY	TOTAL ANNUAL INCOME
ANDHRA PRADESH	5.89%	14.35%	-0.36%	3.78%	5.45%
ASSAM	0.70%	9.47%	-7.77%	-3.99%	-0.34%
BIHAR	-0.78%	-3.64%	-6.29%	1.95%	-0.75%
CHHATTISGARH	6.34%	---	-52.74%	1.74%	3.98%
GUJARAT	1.40%	7.10%	2.28%	2.81%	3.12%
HARYANA	8.77%	---	-5.87%	2.29%	8.32%
JAMMU & KASHMIR	-5.51%	1.04%	1.04%	4.86%	0.66%
JHARKHAND	-2.53%	20.13%	-6.17%	-0.95%	0.87%
KARNATAKA	5.76%	9.46%	5.28%	1.51%	4.48%
KERALA	3.64%	7.23%	5.05%	1.75%	3.20%
MADHYA PRADESH	6.10%	---	-5.28%	0.83%	6.91%
MAHARASHTRA	3.32%	8.96%	3.97%	2.09%	3.46%
ODISHA	6.48%	36.08%	5.89%	3.13%	7.57%
PUNJAB	5.64%	12.27%	-2.70%	4.00%	5.13%
RAJASTHAN	13.84%	45.11%	4.65%	2.17%	8.10%
TAMIL NADU	2.84%	15.53%	9.64%	1.82%	4.47%
UTTAR PRADESH	4.44%	16.32%	-0.55%	-0.63%	3.31%
WEST BENGAL	-5.01%	3.82%	-2.16%	0.90%	-1.25%

IV. CONCLUSION:

we conclude in this analytical Report identified hidden patterns by using Matplotlib and Apache Hive Tool to understand agriculture production and wages in different general and commercial crops. we can extend this research report by incorporating Machine learning Algorithms and produce more realistic analytical Reports it is very helpful for farmer welfare and strengthen and guide them regarding different seasonal crops and Incomes.

*Ethics approval and consent to participate NOT APPLICABLE
*Consent for publication NOT APPLICABLE
*Availability of data and material NOT APPLICABLE
*Competing interests NOT APPLICABLE
*Funding NOT APPLICABLE
*Authors' contributions NOT APPLICABLE
*Authors' information (optional) NOT APPLICABLE
*Acknowledgements NOT APPLICABLE

 ISSN: 2454-4248
193 – 199

REFERENCES:

[1] Oreilly Hadoop The Definitive Guide, Third *Edition* by Tom White
[2] Hadoop for Dummies Paperback – 2014 by Dirk Deroos,, Paul C. Zikopoulos, Roman B. Melnyk
[3] Oreilly Programming Hive,Third Edition
[4] Resources from Department of Agriculture & Farmers Well Fare, Director of Economics and Statistics http://eands.dacnet.nic.in/
[5] https://data.gov.in/
[6] https://matplotlib.org/users/pyplot_tutorial.html
[7] https://pythonprogramming.net/matplotlib-python-3-basics-tutorial/

*