SDG12: RESPONSIBLE CONSUMPTION AND PRODUCTION

TITLE: SUSTAINABLE LIVING

ES1101: Computational Data Analysis

Faculty Guide:

Santosh Kumar Verma

Dr. Richa Sharma

SUBMITTED BY:

Ritika kuwera (2021BTech097)

Manasvi Jain(2021BTech070)

Rajika Patel (2021BTech094)

Prapti Bhatt(2021BTech088)

Institute of Engineering and Technology (IET)

JK LAKSHMIPAT UNIVERSITY, JAIPUR

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The main objective of this report is to analyse the production and consumption of various entities like electricity, food grains, medical equipment. On analysing the above-mentioned resource consuming entities, we aim to analyse the quantity of waste treated out of the total waste generated. To achieve the objectives, we have used the inferential and the descriptive statistics. Mean, Standard Deviation, Statistical tests like the- single mean tests, difference of means, linear regression and corelation are some of the methods taken in use. The hypothesis tests have been used to compare the means whether it is intra-sample or not. One of the underlying objectives of this report is to spread awareness regarding the appropriate or over or bad usage and production of resources. From the discussions, the very objective of goal 12- Responsible Consumption and production gets highlighted.

Keywords: resource consuming entities, consumption, production, statistics, goal 12- Responsible Consumption and production.

1. INTRODUCTION

Fig 1: Sustainable Development Goals

Source: http://www.wordsintodeeds.org/youth-voices-to-promote-sustainability/

The sustainable development goals as embraced by the United Nations in 2015 is a practice of a universal call against the bad practices and the situations like poverty, hunger, AIDS, unequal distribution of resources, polluted environment, crimes like discrimination towards women etc.

It tries to ensure that by 2030, people gain happiness and peace, and until then these 17 goals are followed.

There are 17 Sustainable Development Goals, and their main motive are: -

- **No Poverty:** Poverty is when the people live without basic things, they need for living like food, water, shelter, or money to buy clothes, education and health care facilities. Governments have invested in economic growth; charities have raised money and awareness through campaigns and individual have come up with innovative new solutions.

- **No Hunger:** Hunger is something that affects people everywhere. Through SDG the hunger going halved over the past 25 years with this, Agricultural productivity has improved; economic growth has enabled countries to tackle problems like undernourishment, food insecurity.
- **Good Health and Well-Being:** In this goal, SDG ensured that everybody being fit, and proper way of sanitation and hygiene should be maintained.
- **Quality Education:** To make sure that every child in any country has been provided the fair minded and in-depth type of quality education.
- **Gender Equality:** It ensure that every individual has right to equality in every sense of caste discrimination and woman empowerment.
- **Clean Water and Sanitation:** It ensures that water must be clean if it is for drinking, or any other uses and it may maintain the sanitation. So, that people will not affect by any harmful diseases.
- **Affordable and Clean Energy:** Adopting cost- efficient standards. So, that technology can also reduce the wastage of electricity and people can easily afford the electricity consumption.
- **Decent Work and Economic Growth:** This goal has focused on economic productivity by increasing the infrastructure, employment and decent employs.
- **Industry, Innovation and Infrastructure:** To ensure increasing the industrialization, to increase the resources for betterment of industries and technologies and for making the strong foundations.
- **Reduce Inequalities:** To spread awareness among people and creating the laws to prohibit the racism, caste, color and more status.
- **Sustainable cities and Communities**: Encourage increasing the urbanization for the development of nation and improves the living standards and ensure the safe living in society.

- **Responsible Consumption and Production:** To ensure the proper consumption and production of goods and services.

- **Climate Change:** Move forward to aware and educate people on climate change and the disasters occurs from the climate change.

- **Life below Water:** To conserve the marine resources and maintain the sustainability and cleanliness under water.

- **Life on Land:** To secure the forests and conserve the terrestrial ecosystem.

- **Justice, Peace and Strong institution:** Ensure that everyone has the fundamental rights, and all have equal approach for the justice.

- **Partnership for the goals:** To spread the knowledge, increase in technologies, goes for more urbanization and for the increasing the economic growth.

SDG 12: RESPONSIBLE CONSUMPTION AND PRODUCTION

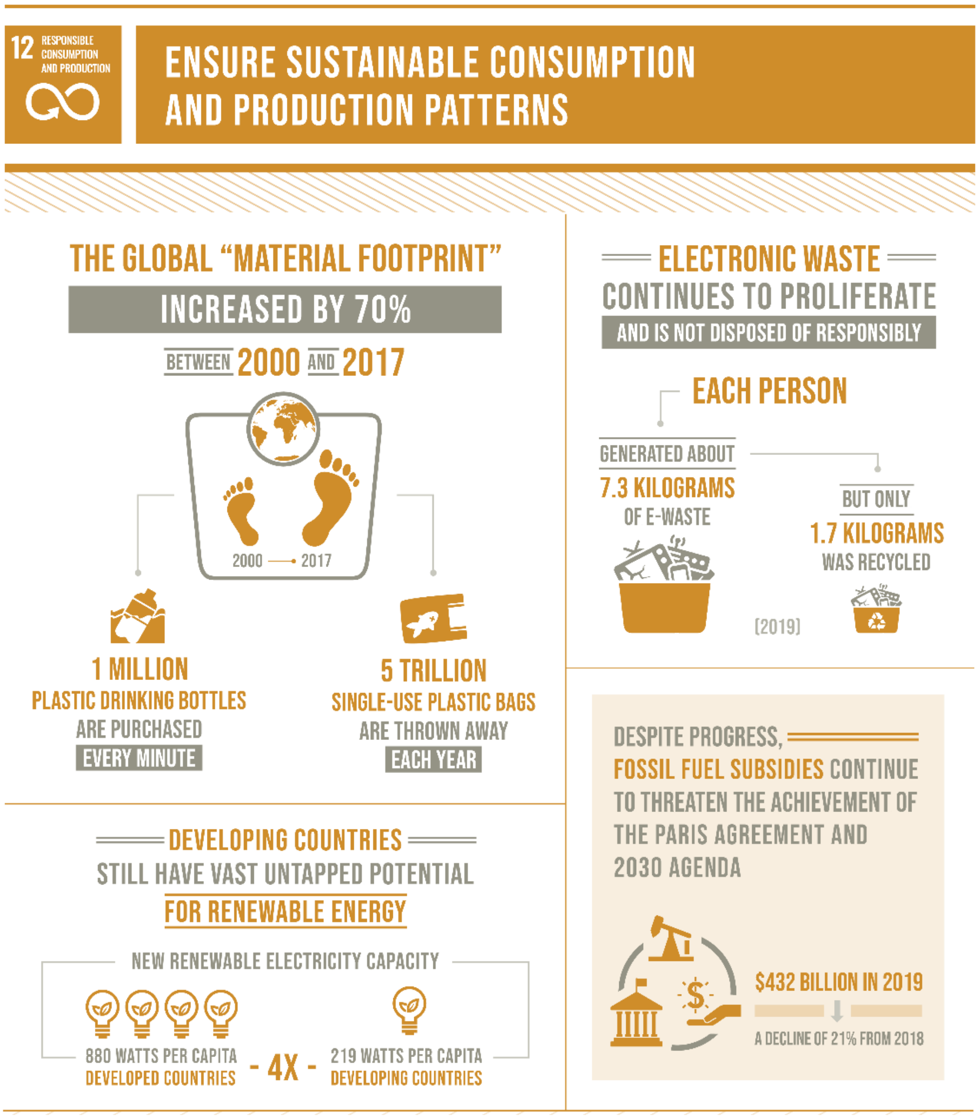


Fig 2: Goal 12: Responsible Consumption and Production

Source: https://sdgs.un.org/goals/goal12

SDGs are divided into Social, Economic, Governance, and Environment goals. Goal 12 comes under the environment one.

This goal aims to achieve some of the practices till 2030 like- sustainable management and efficient use of natural resources, halving the global food waste, changing the way we consume and other resources like medical facilities and the raw food items like grains, cereals etc.

2. SURVEY OF LITERATURE

Santosh Kumar Sahu (2008) in the paper ‘Trends and Patterns of Energy Consumption in India ‘conducted various studies on the pattern on energy production and consumption and the factors affecting it. He mentioned that around the world, in all the countries, energy is something which serves as a measure to calculate a country’s overall growth. This growth and development can be the economic or human. As compared to some of the highly matured economies, India’s energy intensity is higher. The targets and the ideas regarding the measures for the betterment of the economy and society are supported well by the government of India too. Many studies have shown how the factors like population, GDP, insufficiency etc. are connected with the energy intensity of a country. There can be scarcity or dearth of different forms of energy too which may lead to the increase in prices and ultimately tremor for the economies leading to the declining of the growth rates. He mentioned that, along with the betterment of the country, and its growth in terms of economy, there is another aspect of power generation which gets neglected the most, which is harm to the environment. During construction, generation and disposal from these power plants often cause water usage, emissions, local pollution, and wildlife displacement. This becomes a barrier to sustainable development Along with the requirements, the economies make the factors like literacy rate, infant mortality rate, sex ratio and other factors change. With India’s increasing demands for the energy, there has been a positive impact for the same factors which get influenced. For the past decades, there have been variations in the consumption rates. The economy gets advanced as the industrialization paces. Some of the production and the consumption entities of energy include coal, electricity, renewable energy sources, natural gas, petroleum, small hydro power, wind power, solar power& biomass generated power.

Michael A. William in the paper ‘Energy Efficiency retrofitting strategies for Healthcare’ discusses the parametric and economic analysis of healthcare facilities. In India when the first wave of Covid -19 pandemic was getting down late last year when India had more than 1.5 million isolation beds and 15,375 dedicated treatment facilities for 1.38 billion of population. Which count over 1 bed per 1,000 people. Out of which only 18% were oxygen supported beds, each of these dedicated facilities had about 5.2(ICU) beds and total of 80,583 ICU beds for covid patients, half of these were ventilator beds. Beds in some prestigious public hospitals continue to be inaccessible to the masses as they were reserved for politicians, members of the judiciary, bureaucrats, other well-connected people and their families. There is an increased use and demand of single-use-plastics amid of this corona pandemic due to which there is a rapid increase in waste production such as PPE. In addition, it accompanied by the disposal of other single-used-plastic(sup). States says the demand of plastic is increased by 40 % in packing and 17% in other application which also include medical uses. The medical waste during the COVID-19 increased dramatically out of those the king Abdullah university hospital in Jordan produced tenfold higher medical waste, Which is approximately 650 kg per day, when considering an occupation of 95 covid-19 patients. Due to this irresponsible degradation of medical waste is overloading the capacity of various countries to treat it adequately.

Med Ram Verma K in the paper ‘Diversification of Food Production and Consumption Patterns in India’ has discussed that, to supplement farm revenue, India’s agricultural diversification is progressively increasing toward high-value crops and animal operations. Some of the characteristics that influence the type and rate of agriculture growth from dietary staples to high value crops include technological advancements in crop production, improved rural infrastructure, and variety in food demand. Due to vast variety in agricultural and social economic factors, the nature of agriculture development varies by area. In rainfed areas, diversity in support of horticulture and animal commodities is more evident. With rising earnings, people’s eating habits shift away from a primary cereal-based meal and toward non-cereal products. The livestock sub-sector has expanded throughout diverse areas as a consequence of rising demand for animal products such as milk, meat, eggs, and so on. In the paper ‘Relationship between Food Production and Consumption Diversity in India’ the authors P.Venkatesh, and Premlata Singh give the Empirical Evidences from cross section analysis. The study looked at (i) the diversity of food production and consumption across regions, (ii) patterns in food intake and nutrient intake, as well as changes in related socioeconomic characteristics, and (iii) the relationship between consumption and production diversity, as well as other major determinants. It reflects the nutritional consumption, which has decreased in terms of calories and protein while increasing in terms of fat intake throughout this time. Simultaneously, increased urbanization has resulted in large increases with per income and food commodity availability. The study found that local production variety has a substantial influence on consumption patterns.

Snigdha Chakrabarti and Prasen ji Sarkhel in the paper Economies of Solid Waste Management have mentioned that Solid waste generation access approximation the quantity of waste erects by accommodation over a certain amount of time. Waste Generation involves all the stuff throw away or whether scrap or tempt in a landfill. Today, waste is obtained as a report of society's venture to solve the domain of problems like water and air pollution. Some of these high amounts of waste stand up with new issues such as sediments and remaining from purifying of flue gases. Major problems of Waste generation are "Diseases". Many diseases attack to children. Economy Growth is also affects from the solid waste generation in a good manner. Globally, from waste they obtain power and many other useful things which improves the GDP of a country which likely echo national differences in energy supply systems. Let's take an example of fly-ash, it's a remaining product but it also used in Hydro-electric power stations. There are three points through which we can reduce the waste generation. 1) Reduce, 2) Reuse, 3) Recycle.Solid waste management in developing countries, As increases in urban sprawl. In present waste generation, daily per capita generation is very less in developing countries than in developed countries. Variation is also occurred between high- and low-income countries in terms of waste formation and physical aspect of wastes.

3. OBJECTIVES

To examine the utilisation of various resources like power, medical equipment, and food grains through which the total treatment of waste is analysed against the total waste generated.

Objective 1: To Analyse the State-Wise Energy Consumption and Production in different states of India.

1.1 To check if the population's demand for mean energy units from 2016 to 2021**.**

1.2 To study the effect of energy units demand on the production of energy units in India.

1.3 To predict the consumption of energy units in next five years based on current trend in energy consumption.

1.4 To evaluate the significance of variance in the production of power units from the year 2016-17 to 2020-21 for 36 states and union territories of India.

Objective 2: To Analyse the Consumption and Production of Medical Facilities in different states of India.

2.1 To designate different states based on the no. of beds provided in the year 2019.

2.2 To check if the number of mean medical beds provided in the hospitals has significantly increased from 2019 to 2020.

2.3 To study the dependence of population on number of beds.

Objective 3: To Analyse the Production and Consumption of Food Grains in different states of India.

3.1 To check if the production of rice and pulses has significantly increased from 2018-19 to 2019-20.

3.2 To study the effect on production of rice and pluses in 2019-20 with respect to the population of same year.

3.3 To designate different states with respect to sum of production of food grains (rice and pulses) in year 2018-2019 and 2019-2020.

3.4 To analyse the sum of consumption of rice and pulse in 2019-2020**.**

Objective 4: To Analyse the Generation, Collection and hence the treatment of solid waste produced in different states.

4.1 To designate different states in India based on total waste generated from year 2016-17 to 2017-18.

4.2 To check if there is a significant difference between the means of waste collected and treated in the years 2016-17 to 2017-18.

4.3 To study the relation between the waste generated and collected in India in the years 2016-2020.

4.4 To predict the generation of solid waste in the next five years on the basis of current trend.

4. DATA COLLECTION

<u>Table 1.1</u>: Demand of power units (net crore units) in the years from 2016-17 to 2020-21 shown in particular state/UT

state/UT	2016-17	2017-18	2018-19	2019-20	2020-21
Andaman and Nicobar Islands	24	33	35	35	35
Andhra Pradesh	5430	5838	6374	6545	6208
Arunachal Pradesh	73	80	86	75	72
Assam	902	910	953	980	1019
Bihar	2571	2702	3005	3163	3417
Chandigarh	165	161	156	173	152
Chhattisgarh	2375	2592	2613	3011	3047
Dadra and Nagar Haveli	602	617	632	653	550
Daman and Diu	240	253	256	257	222
Delhi	3083	3183	3230	3309	2956
Goa	432	412	428	435	408
Gujarat	10370	10999	11659	11394	11162
Haryana	4889	5078	5367	5451	5316
Himachal Pradesh	883	940	985	1042	1019
Jammu and Kashmir & Ladakh	1740	1881	1959	2003	1977
Jharkhand	796	791	868	894	995
Karnataka	6690	6787	7176	7280	6885
Kerala	2430	2500	2503	2632	2512
Lakshadweep	5	5	5	5	6
Madhya Pradesh	6576	6993	7567	7617	8344
Maharashtra	13929	14976	15829	15517	15068
Manipur	76	87	92	92	97
Meghalaya	171	156	196	211	203
Mizoram	51	50	67	65	73
Nagaland	76	80	90	81	83
Odisha	2676	2880	3181	2969	2985
Puducherry	255	267	276	285	264
Punjab	5310	5481	5529	5678	5845
Rajasthan	6784	7119	7983	8128	8531
Sikkim	47	49	51	55	55
Tamil Nadu	10451	10601	10938	10882	10119
Telangana	5303	6032	6670	6831	6700
Tripura	164	260	186	154	148
Uttar Pradesh	10757	12005	11710	12255	12437
Uttarakhand	1307	1346	1385	1447	1383
West Bengal	4795	5076	5224	5295	5164
India	112428	119220	125264	126899	125457

Table 1.2: **Production of power units (net crore units) in the years from 2016-17 to 2020-21 shown in particular state/UT**

state/UT	2016-17	2017-18	2018-19	2019-20	2020-21
Andaman and Nicobar Islands	52	52	57	57	75
Andhra Pradesh	22298	23674	24184	24854	25484
Arunachal Pradesh	274	301	337	379	765
Assam	1600	1571	1710	1757	1794
Bihar	3608	4341	4566	5792	6321
Chandigarh	177	193	198	207	215
Chhattisgarh	12060	13527	14044	12835	13076
Dadra and Nagar Haveli	242	255	276	493	509
Daman and Diu	182	190	206	224	256
Delhi	7839	7180	7237	7582	7590
Goa	541	549	567	581	596
Gujarat	30716	31043	32291	35211	37893
Haryana	11056	11260	11268	12290	12481
Himachal Pradesh	3893	4047	4068	4007	4155
Jammu and Kashmir & Ladakh	3274	3389	3394	3417	3478
Jharkhand	2237	1771	1774	2514	2548
Karnataka	21317	27157	28075	29825	30091
Kerala	4999	5083	5083	5696	5837
Lakshadweep	1	1	1	1	1
Madhya Pradesh	19618	21728	23334	24821	25489
Maharashtra	41410	43255	44144	43496	44166
Manipur	233	242	259	268	262
Meghalaya	518	565	582	609	616
Mizoram	137	196	206	212	207
Nagaland	155	159	170	175	179
Odisha	8538	7377	7654	8567	8594
Puducherry	367	369	370	375	380
Punjab	14162	13432	13432	14205	14389
Rajasthan	19776	21684	22589	25360	26045
Sikkim	758	962	962	674	678
Tamil Nadu	29112	30327	31059	32840	33695
Telangana	11501	15139	15826	16024	17218
Tripura	727	733	752	764	724
Uttar Pradesh	23662	24909	25130	26162	27896
Uttarakhand	3313	3399	3403	3550	3731
West Bengal	10383	10518	10568	11027	11037
India	310736	330578	339776	356851	368471

<u>Table 2.1</u>: Table showing no. of beds and hospitals in the year 2019 and 2020 and the population in 2020.

States	Government Hospitals2019	Beds in Government Hospitals2019	Government Hospitals2020	Beds in Government Hospitals2020	Population 2020
Andhra Pradesh	6,234	86,721	6,240	86,900	53903393
Bihar	2,132	29,339	2,150	30,333	124799926
Chandigarh	49	3,758	49	3,758	961587
Chhattisgarh	247	9,610	256	9,830	29436231
Goa	*43*	*3,086*	49	3,296	1586250
Gujarat	2,245	29,402	2,249	29,556	63872399
Haryana	678	12,590	678	12,590	28204692
Himachal Pradesh	822	14,782	824	14,832	7451955
Jharkhand	4,463	14,891	4,463	14,891	38593948
Karnataka	2,842	70,474	2,844	70,564	67562686
Kerala	1,284	38,097	1,292	38,230	35699443
Madhya Pradesh	465	31,106	480	31,348	85358965
Maharashtra	514	33,028	520	33,260	12314223
Meghalaya	154	4,467	154	4,467	3366710
NCT of Delhi	107	27,154	107	27,154	18710922
Odisha	1,806	18,519	1,810	18,611	46356334
Puducherry	44	4,768	47	4,829	1413542
Punjab	816	21,241	822	21,512	30141373
Rajasthan	2,849	46,778	2,849	46,778	81032689
Tamil Nadu	2,507	99,435	2,507	99,435	77841267
Telangana	677	5,094	677	5,094	38510982
West Bengal	155	4,343	155	4,343	4169794
Uttar Pradesh	4,683	66,700	4,683	66,700	237882725
Uttarakhand	618	8,106	622	85,043	11250858
Andaman & Nicobar	30	1,202	30	1,202	398774
Arunachal Pradesh	218	2,404	218	2,404	1570458
Lakshadweep	12	250	12	250	65998
Mizoram	99	2,022	99	2,022	1239244
Sikkim	33	2,260	35	2,260	690251
Nagaland	36	1,880	38	2,139	22496334
Tripura	156	4,429	156	4,429	4169794

<u>Table 3.1</u>: Production of Rice and Pulses (in thousand tons) in the years from 2018-19 to 2019-20 shown for state/UT

States/UTs	Rice		Pulses		Population
	2018-2019	**2019-2020**	**2018-2019**	**2019-2020**	**2019-2020**
Andhra Pradesh	8234.7	8658.9	739.6	1166.7	5,39,03,393
Arunachal Pradesh	240.0	244.7	12.0	14.2	15,70,458
Assam	5220.6	4984.6	113.5	106.1	3,56,07,039
Bihar	6155.5	6298.0	453.5	334.4	12,47,99,926
Chhattisgarh	6526.9	6774.8	537.5	241.3	2,94,36,231
Delhi	16.8	16.8	0.0	0.0	1,87,10,922
Goa	98.8	90.4	6.3	3.9	15,86,250
Gujarat	1912.1	1983.1	681.3	1057.3	6,38,72,399
Haryana	4516.1	4824.3	82.1	64.4	2,82,04,692
Himachal Pradesh	114.9	143.8	54.0	55.2	74,51,955
Jammu and Kashmir	615.8	587	10.6	44.2	1,36,06,320
Jharkhand	2893.9	3012.8	735.2	814.9	3,85,93,948
Karnataka	3431.0	3634.5	1773.9	2155.9	6,75,62,686
Kerala	578.3	605.6	2.3	2.2	3,56,99,443
Madhya Pradesh	4494.7	4778.2	6045.4	4108.4	8,53,58,965
Maharashtra	3275.7	2897.6	2682.5	3736.0	12,31,44,223
Manipur	401.6	385.5	29.5	25.2	30,91,545
Meghalaya	202.0	303.4	13.4	12.0	33,66,710
Mizoram	60.0	60	5.9	5.5	12,39,244
Nagaland	356.7	363.3	46.4	46.8	22,49,695
Odisha	7733.7	8360.4	412.1	432.5	4,63,56,334
Puducherry	63.3	59.4	0.7	0.5	14,13,542
Punjab	12821.6	11779.3	27.7	29.2	3,01,41,373
Rajasthan	453.2	480.5	3759.4	4497.1	8,10,32,689
Sikkim	17.2	16.1	4.8	5.0	6,90,251
Tamil Nadu	6130.9	7171.1	551.2	605.4	7,78,41,267
Telangana	6670.0	7427.8	440.1	549.2	3,85,10,982
Tripura	793.2	810.2	18.9	18.7	41,69,794
Uttar Pradesh	15545.3	15517.9	2408.0	2447.3	23,78,82,725
Uttarakhand	617.6	658.4	55.3	57.8	1,12,50,858
West Bengal	16242.2	15881.4	368.4	384.9	9,96,09,303

Table 3.2: **Total production of rice and pulses (in thousand tons) of years 2018-19 and 2019-20 shown state/UT**

	States/UTs	**Total production of rice and pulses (2018-19)**	**Total production of rice and pulses (2019-20)**
1	Andhra Pradesh	8974.3	9825.6
2	Arunachal Pradesh	252.0	258.9
3	Assam	5334.1	5090.7
4	Bihar	6609.0	6632.4
5	Chhattisgarh	7064.4	7016.1
6	Delhi	16.8	16.8
7	Goa	105.1	94.3
8	Gujarat	2593.4	3040.4
9	Haryana	4598.2	4888.7
10	Himachal Pradesh	168.9	199.0
11	Jammu and Kashmir	626.4	631.2
12	Jharkhand	3629.1	3827.7
13	Karnataka	5204.9	5790.4
14	Kerala	580.6	607.8
15	Madhya Pradesh	10540.1	8886.6
16	Maharashtra	5958.2	6633.6
17	Manipur	431.1	410.7
18	Meghalaya	215.4	315.4
19	Mizoram	65.9	65.5
20	Nagaland	403.1	410.1
21	Odisha	8145.8	8792.9
22	Puducherry	64.0	59.9
23	Punjab	12849.3	11808.5
24	Rajasthan	4212.6	4977.6
25	Sikkim	22.0	21.1
26	Tamil Nadu	6682.1	7776.5
27	Telangana	7110.1	7977.0
28	Tripura	812.1	828.9
29	Uttar Pradesh	17953.3	17965.2
30	Uttarakhand	672.9	716.2
31	West Bengal	16610.6	16266.3

Table 4.1: Waste generated, collected and treated by different states in (2016-2017)

Sl. No.	State (2016-17)	Solid waste Generated	Collected	Treated
1.	Madhya Pradesh	6678	Nil	Nil
2.	Maharashtra	21867.27	21867.27	6993.2
3.	Manipur	-	-	-
4.	Meghalaya	187	156	36
5.	Mizoram	-	-	-
6.	Nagaland	-	-	-
7.	Orissa	2574.7	2283.9	30
8.	Punjab	4456.2	4435	3.72
9.	Rajasthan	-	-	-
10.	Sikkim	-	-	-
11.	Tamil Nadu	230	210	-
12.	Telangana	6628	6625	3175
13.	Tripura	414	368.2	250.4
14.	Uttarakhand	917	917	Nil
15.	Uttar Pradesh	15192	11394	1857
16.	West Bengal	-	-	-
17.	Andaman and Nicobar	6440	6331	500
18.	Chandigarh	370	360	Nil
19.	Jammu and Kashmir	1634.5	1388.7	3.45
20.	Karnataka	8842	7716	3584
21.	Lakshadweep	-	-	-
22.	Delhi	9620	-	-
23.	Goa	-	-	-
24.	Gujarat	-	10480	2565
25.	Puducherry	-	513	10

Table 4.2: **Waste generated, collected and treated by different states in (2017-2018)**

Sl. No.	State (2017-18)	Solid waste Generated	Collected	Treated
1.	Madhya Pradesh	6773	5480	1141
2.	Maharashtra	23449.66	23079	7543
3.	Manipur	-	-	-
4.	Meghalaya	187	156	36
5.	Mizoram	159.88	159.88	0
6.	Nagaland	337	255	28
7.	Odisha	18.55	14.28	30
8.	Punjab	4544.35	4520.3 5	39.17 5
9.	Rajasthan	-	-	-
10.	Sikkim	76.04	62.10	11.05
11.	Tamil Nadu	14658.5	14416. 629	4776. 218
12.	Telangana	-	-	-
13.	Tripura	428.09	379.2	134.4
14.	Uttarakhand	1180	1180	0
15.	Uttar Pradesh	15500	12000	3115
16.	West Bengal	140005	12600	830
17.	Andaman and Nicobar	115	115	26.45
18.	Chandigarh	450	450 7	61.15
19.	Jammu and Kashmir	-	-	-
20.	Karnataka	11186	9706	3475
21.	Lakshadweep	-	-	-
22.	Delhi	-	-	-
23.	Goa	226. 8	218.8	0
24.	Gujarat	-	10527	757
25.	Puducherry	398.5	398.5	10.0

Table 4.3: Waste generated, collected and treated by different states in (2018-2019)

Sl. No.	State (2018-19)	Solid waste Generated	Collected	Treated
1.	Madhya Pradesh	7212	6537	2272.5
2.	Maharashtra	-	-	-
3.	Manipur	-	-	-
4.	Meghalaya	210	175	36
5.	Mizoram	-	-	-
6.	Nagaland	539.44(only nineteen ULBs)	471.58 (only nineteen ULBs)	Nil
7.	Odisha			
8.	Punjab	-	-	-
9.	Rajasthan	-	-	-
10.	Sikkim	73.34 Nil	63.00	7.50
11.	Tamil Nadu	15174.612	14209.127	5307.146
12.	Telangana	7804	7023	4795
13.	Tripura	433.2	372.50	148.40
14.	Uttarakhand	1099	1099	Nil
15.	Uttar Pradesh	-	-	-
16.	West Bengal	-	-	-
17.	Andaman and Nicobar	-	-	-
18.	Chandigarh	500	463	14287
19.	Jammu and Kashmir	Jammu666.68 Kashmir930.60	Jammu617.83 Kashmir833.30	Nil
20.	Karnataka	-	-	-
21.	Lakshadweep	-	-	-
22.	Delhi	-	-	-
23.	Goa	-	-	-
24.	Gujarat	Nil	11,119	1,127
25.	Puducherry	-	-	-

Table 4.4: Waste generated, collected and treated by different states in (2019-2020)

Sl. No.	State (2019-20)	Solid waste Generated	Collected	Treated
1.	Madhya Pradesh	8000	7500	6100
2.	Maharashtra	23844.551	23675.7	12623.33
3.	Manipur	218.6	126.63	80.00
4.	Meghalaya	170.63	170.63	8.72
5.	Mizoram	251.42	213.07	29.22
6.	Nagaland	339.5	216.9	135.8
7.	Orissa	2564.43	2255.32	91.63
8.	Punjab	4634.48	4574.93	917.56
9.	Rajasthan	6625.56	6475.39	780.18
10.	Sikkim	75.1	67.1	13.05
11.	Tamil Nadu	13968	12850	7196
12.	Telangana	8497	8360	5747
13.	Tripura	445.72	389.46	150.10
14.	Uttarakhand	1527.458	1437.4	524
15.	Uttar Pradesh	17377.3	17329.4	4615
16.	West Bengal	14613.3	13064.63	916
17.	Andaman and Nicobar	120	117	65.1
18.	Chandigarh	470	458.52	150
19.	Jammu and Kashmir	Jammu702.2 Kashmir828.33 TOTAL1530.53	Jammu648.4 Kashmir804.46 TOTAL1452.86	-
20.	Karnataka	11958	10011	4515
21.	Lakshadweep	35	18	18
22.	Delhi	10817	10614	5714
23.	Goa	236.41	235.90	154.71
24.	Gujarat	-	11,119	1,127
25.	Puducherry	599.25	505	24

5. METHODOLOGY

STATISTICS: It is the science or method of collecting the data and analysing it, for the purpose of inferencing something. It deals with the practice of collection, organization, analysis, interpretation, and the presentation of data as well.

It can be divided into two parts:

STATISTICS

DESCRIPTIVE	INFERENTIAL
Method of organizing, Summarizing, and Presenting the data in Informative way Sample is random subset of population.	Method to determine something about the sample on the basis of Population.
Mean, Median, Mode is calculated for central tendency. Range, variance , and standard deviation is calculated for variability measure.	Estimation and hypothesis testing are there.

Descriptive Statistics Methods

Mean ($\bar{X}$)

It is the sum of all the observations divided by the total number of observations.

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n} Xi \qquad \ldots (1)$$

X_i = i^{th} observation

N= Total number of observations.

Median

It is defined as the mid value in a sample of values.

If n is odd:
$(n+1)^{th}$ observation

If n is even:
Average of $(n/2)^{th}$ and $(n/2 + 1)^{th}$ observation

Here, n= Total number of observations.

Mode

It is defined as the value which occurs maximum no. of times.

Standard deviation(σ)

It is the value which tells that how much the data is scattered from the average value. Lesser SD infers that the data is more precise.

$$\boldsymbol{\sigma = \sqrt{\frac{\Sigma(x_i - \mu)^2}{N}}} \qquad \ldots (2)$$

N= total number of observations
x_i= i^{th} observation
μ= mean of the sample

Variance(σ^2)

It is the square of the standard deviation, and this term gives a statistical inference about the spread between numbers in the data set.
It is the average of the squared differences from the Mean.

$$\sigma^2 = \sum_{i=1}^{N} \frac{(x-\mu)^2}{N} \qquad \ldots (3)$$

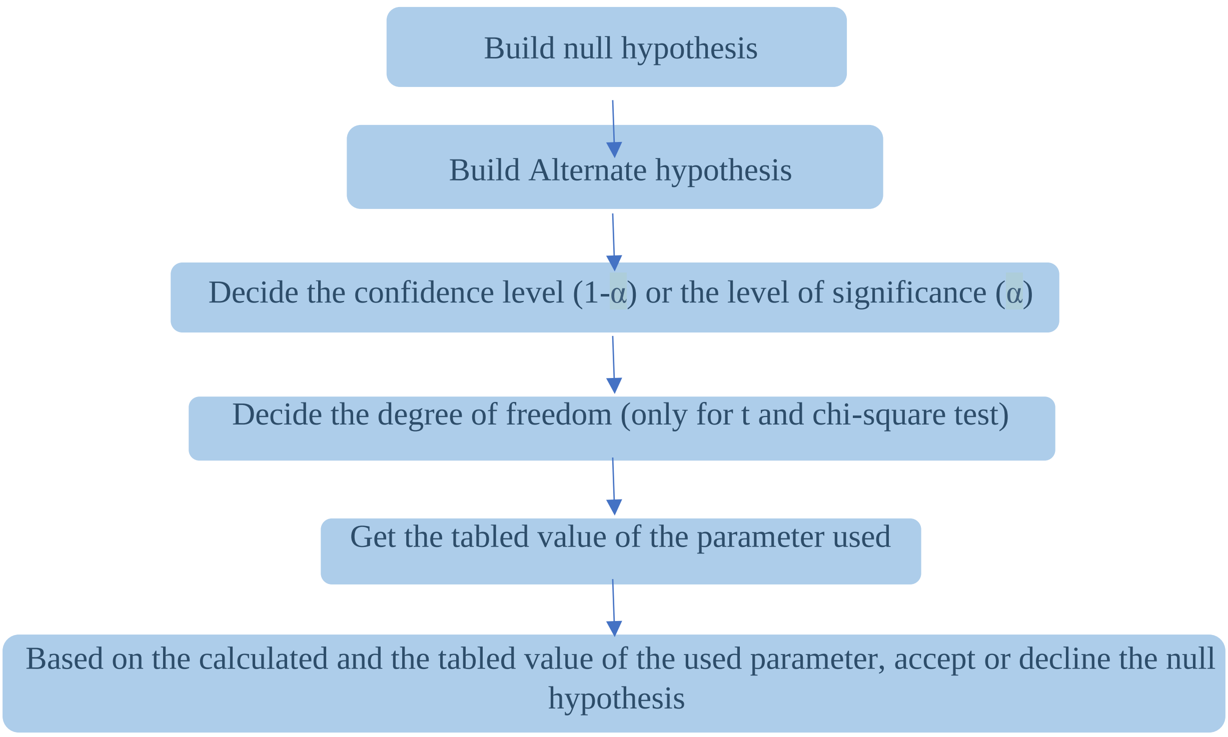
N= total number of observations
x_i= i^{th} observation
μ= mean of the sample

Inferential Statistics Methods

Hypothesis Testing

Hypothesis is an assumption that we make about the population parameters. Hypothesis testing is a way for testing a claim about a particular parameter in a population, using the sample data. It consists of mainly two attributes which are- Null and Alternate hypothesis. We validate our assumptions on the basis of these two assumptions.

STEPS TO HYPOTHESIS TESTING



Null Hypothesis (H_0) –It is a statistical hypothesis that depicts that there is no difference between a parameter and a specific value. (Always has the = sign)

Alternative Hypothesis (H_1) – It is a statistical hypothesis that depicts that there is a difference between a parameter and a specific value.

Level of significance (α) – It is the maximum probability of committing an error where the null hypothesis is rejected when it is true.

Critical or Rejection Region –It depicts the range of values for the test value that include the significant difference and infer that the null hypothesis should be rejected.

Non-critical or Non-rejection Region –It depicts the range of values for the test value that include the significant difference and infer that the null hypothesis should not be rejected.

Critical Value –This value separates the critical region from the non-critical region.

One Tailed tests: 1. Right Tailed Test:

$H_0 : \mu \leq k$

$H_1 : \mu > k$

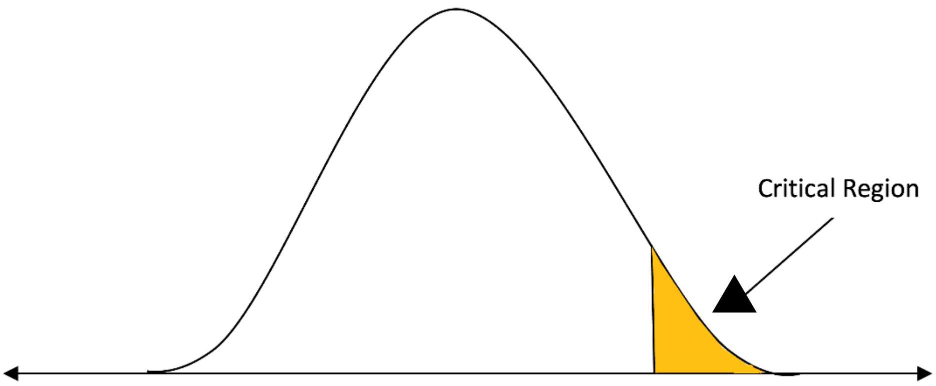


Fig 3: Bell curve for right tailed test

2. Left Tailed Test:

$H_0 : \mu \leq k$

$H_1 : \mu > k$

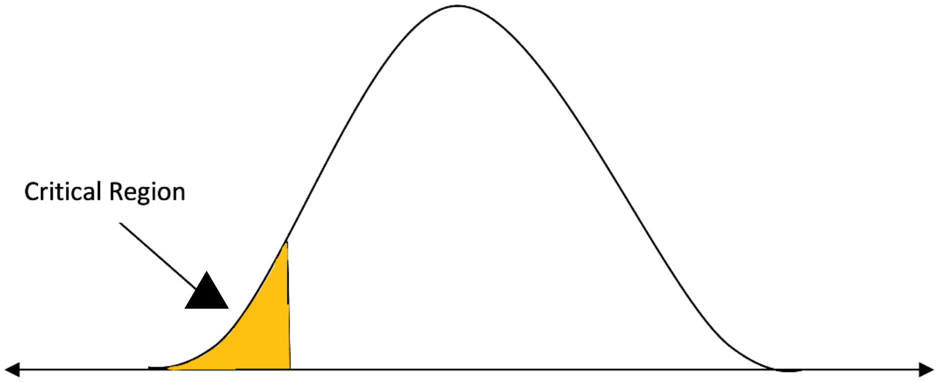


Fig 4: Bell Curve for left tailed test

Two Tailed Test:

$H_0 : \mu = k$

$H_1 : \mu \neq k$

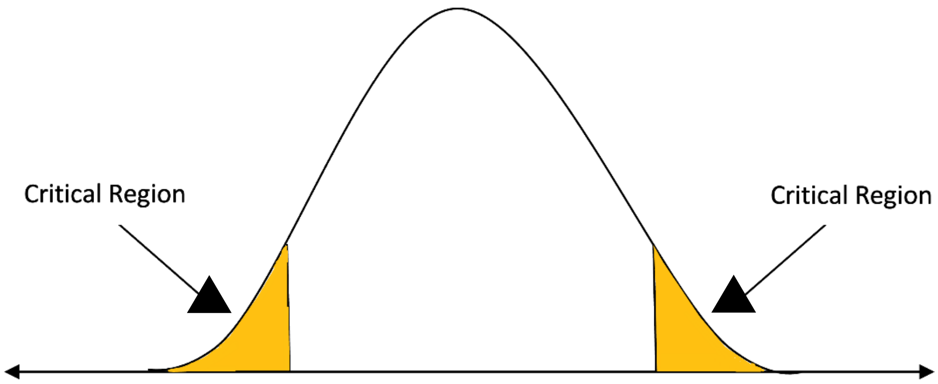


Fig 5: Bell curve for Two tailed test

Hypothesis Testing

Parametric	**Non-Parametric**
Population information like the mean, median and mode is known.	Population information is not known.

Normal probability Distribution	Arbitrary probability Distribution
Comprises of: • Z-Test • T-Test • ANOVA Test	Comprises of: • Contingency Test • Chi-Square Test

T-Test: It is a parametric statistical test which is used to compare and infer whether the two population's means differ from one another or not, when the standard deviation is not known.

It is used when sample size is not large (n<30).

For difference of means **(two populations**):

$$\boldsymbol{t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}} \qquad \ldots (4)$$

Where ,

$$\boldsymbol{S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \qquad \ldots (5)$$

Here, S_1^2 and S_2^2 are the sample variance of sample 1 and sample 2.

N_1 and n_2 are the population size of two samples.

For **one population**:

$$t = \boldsymbol{\frac{\bar{x} - \mu}{s/\sqrt{n}}} \qquad \ldots (6)$$

Where,

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1} \qquad \dots (7)$$

Here, n is the population size.

X̅=mean of the sample

μ= population mean

s=Standard deviation of the sample

Z-Test: It is a parametric statistical test which is used to compare and infer whether the two population's means differ from one another or not, when the standard deviation is known.

It is used when sample size is large (n>30).

For **one population:**

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \qquad \dots (8)$$

Here, n is the population size.

X̅=mean of the sample

μ= population mean

s=Standard deviation of the population

For difference of means **(two populations)**:

$$z = \frac{(\bar{x}_1 - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad \dots (9)$$

Here, n_1 and n_2 is the sample sizes.

$\overline{X}_1$ and $\overline{x}_2$ are the means of the two samples

μ_1 and μ_2 are the population means

σ_1 and σ_2 Standard deviation of the population

ANOVA Test (Analysis of variance test): this is a statistics test which is concerned with comparing the means of three or more populations. It infers about the difference in the means of those populations and to what extent they differ.

For this kind of testing, we use the F distribution tables.

There are various steps to 2 get the calculated value of F.

Calculate total t of observation in the sample

↓

calculate correction factor

↓

calculate total sum of squares of deviation SST

↓

calculate sum of squared difference between the sample mean SSB

↓

calculate sum of squares of difference within the sample SSW

↓

make ANOVA table based on the two degrees of freedom obtained

↓

find out the F value from the table

accept or reject the null hypothesis by analysing the calculated and the tabled value of F.

Source of variation	**Sum of squares**	**Difference of squares**	**Mean squares**	**Test of states/F value**
Between Sample	SSB	k-1	MSB=SSB/k-1	F=MSB/MSW
Within Sample	SSW	n-k	MSW=SSW/n-k	
Total	SST	n-1		

Table 1: The way to annotate an ANOVA table is shown

- Total T (sum of the observation) of the sample.

$$T=x_1+x_2+\dots\dots\dots+x_{k.} \qquad \dots (10)$$

- correlation factor CF.

$$CF=T^2 / n \qquad \dots (11)$$

- total sum of squares of deviation SST.

$$SST = (\Sigma x_1^2+ \Sigma x_2^2+ \Sigma x_3^2+\dots\dots \Sigma x_k^2+) - CF \qquad \dots (12)$$

- sum of squared difference between the sample means SSB.

$$SSB = ((\Sigma x_i)^2 / n_i) - CF \qquad \dots (13)$$

- sum of squares of difference within the sample SSW.

$$SSW = SST - SSB \qquad \dots (14)$$

Regression: This method is used to predict the relation b/w two variables by the help of a linear equation of the observed data. The y value is predicted on the basis of x value whose line equation is as follows:

$$\boldsymbol{y = \bar{y} + \beta(x - \bar{x})} \qquad \dots (14)$$

Where,

$$\boldsymbol{\beta = \frac{n\Sigma xy - \sum x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}} \qquad \dots (15)$$

Corelation: This term signifies the measure to which the two variables are linearly related to each other.

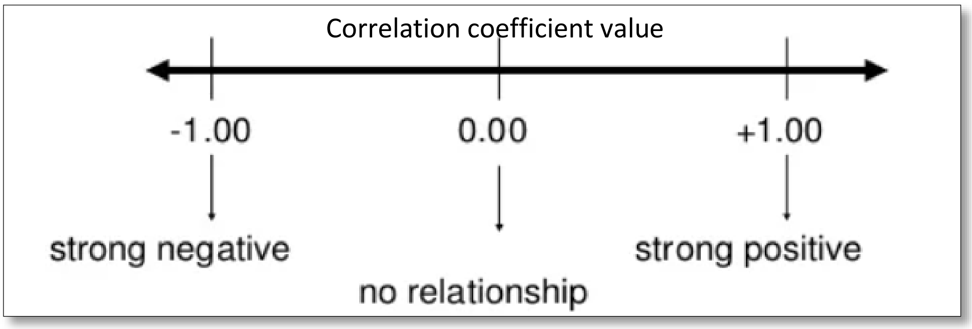


Fig 6: Number line depicting the kind of correlation based on the correlation coefficient value.

- **Pearson Correlation Coefficient (r):** This factor evaluates a linear relationship between the two variables. It ranges from -1 to 1.

If x and y are two variables, then

$$\boldsymbol{r = \frac{\frac{1}{n}\Sigma xy - \overline{xy}}{S_x S_y}} \qquad \dots (16)$$

Where,

$$\boldsymbol{s_x = \sqrt{\frac{1}{n}\sum x^2 - (\bar{x})^2}} \qquad \ldots (17)$$

And

$$\boldsymbol{s_y = \sqrt{\frac{1}{n}\Sigma y^2 - (\bar{y})^2}} \qquad \ldots (18)$$

- **Spearman Correlation Coefficient (r_s):** This Factor evaluates a monotonic relationship between the two variables. $-1 \le r_s \le 1$.

$$\boldsymbol{r_S = 1 - \frac{6\Sigma\, d^2}{n(n^2-1)}} \qquad \ldots (19)$$

Where d=x-y, is the difference in ranking.

LINEAR ALGEBRA: It is a study which deals with linear combinations. It encompasses the study of mappings, lines and spaces, vector spaces, etc. It includes the study of sets of linear equations and their transformations.

Power Method: This method is used to find the dominant eigen value of a matrix.

In this method the given square matrix of order n*n is multiplied with a column matrix of the order n*1 consisting of any values. From the resultant matrix, the highest value is taken as common, and the column matrix so obtained is multiplied again with the square matrix. This process is repeated until the same consequent column matrices are obtained. At this stage, the highest value so obtained in the last obtained column matrix is the dominant eigen value of the initial square matrix, and the column matrix obtained after dividing each value by the dominant eigen value is known as the dominant eigen vector.

STEPS TO FIND THE RANK USING POWER METHOD

Form a square matrix of the data

↓

Follow the power method (stated above) and obtain the dominant eigen vector of the same.

↓

In the dominant eigen vector, give ranking based on the descending order of values.

6.RESULTS AND DISCUSSIONS

Analysis 1

1.1

Claim: Population's demand for energy units has increased from 2016 to 2021. (Five-year time span)

For the same procedure, table 1.1 has been used and the columns used are first(state/UT), second (2016-17), and sixth (2020-21).

Step 1: Formulation of Hypothesis:

If $\mu 1$ and $\mu 2$ are the means of demand of energy units in 2021 and 2016 respectively, then the null and the alternate hypothesis are as follows:

Null Hypothesis:

$\mu 1 \geq \mu 2$
The mean of demand of energy units in 2021 is greater than or equal to (increased) the mean energy units demand in 2016.

Alternative Hypothesis:

$\mu 1 < \mu 2$
The mean of demand of energy units in 2021 lesser than the mean energy units demand in 2016.

Test Statistics:

Now as our sample size i.e., n >30 and the population standard deviation is known, so we will use two population z-test in order to find the difference of mean. The formulas used are (1), (2) and (8).

Here, $\mu 1 \geq \mu 2$, so the right tail test shall be used.

Terminology:

Level of significance: α

Sample size: n1, n2

$\bar{X}_1$=sample mean of energy units demand in 2021

$\bar{X}_2$=sample mean of energy units demand in 2016

Z=z score

$\sigma_{1=}$ standard deviation of energy units demand in 2021

$\sigma_{2=}$ standard deviation of energy units demand in 2016

Step 2: Finding the value of z using python code by deciding the value of α:

```
Means and Standard Deviation using Python code
Mean for 2020-21 data is 3484.9166666666665
mean for 2016-17 data is 3123.0
standard deviation for 2016-17 data is 4099.951207897828
standard deviation for column 2 data is 3721.2092120707216
the value of z is 0.3921
```

So,

Level of significance: $\alpha=0.05$

Sample size: n1=36, n2=36 (for both the columns)

$\bar{X}_1$=3484.9 net crore units (column2 – table 2.1 used)

$\bar{X}_2$=3123 net crore units (column6 – table 2.1 used)

$\sigma_{1=}$ 4099.951(column2 – table 2.1 used)

$\sigma_{2=}$ 3721.209(column6 – table 2.1 used)

From the figures, formulas and data as mentioned above, the calculated value of z obtained by python implementation is **z=0.3921**

Step 3: Get the tabled value of z_α based on α value decided:

For α=0.05, the tabled value of **Z_α=1.645.**

Step 4: Reject or accept the null hypothesis based on calculated and tabled value of z.

Result:

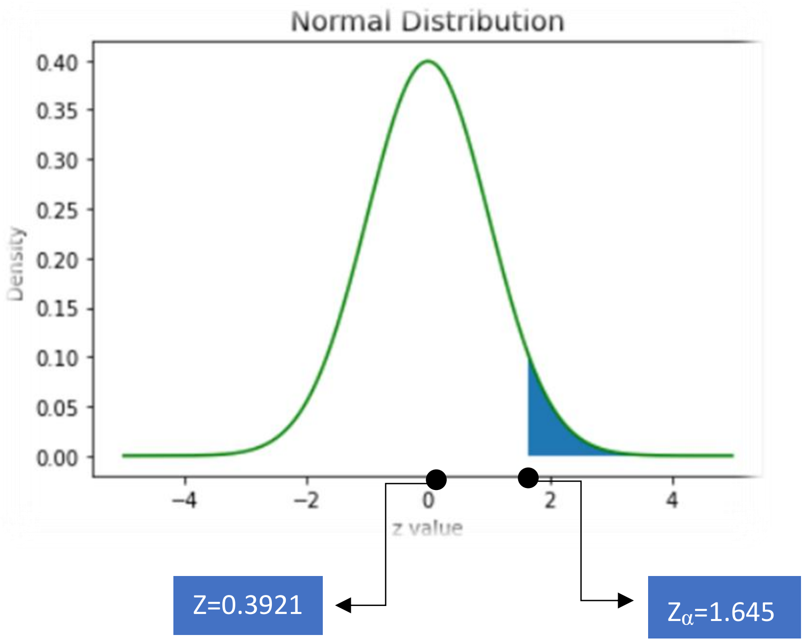


Fig 7: Bell curve depicting the critical value, calculated z value and the critical region.

From the previous calculations, the calculated and the tabled values are 0.3921 and 1.645 respectively. Since calculated value of z = 0.3921<1.645 (tabled value), so we accept the null hypothesis and reject the alternate hypothesis that the mean of demand of energy units in 2021 is greater than or equal to (increased) the mean energy units demand in 2016.

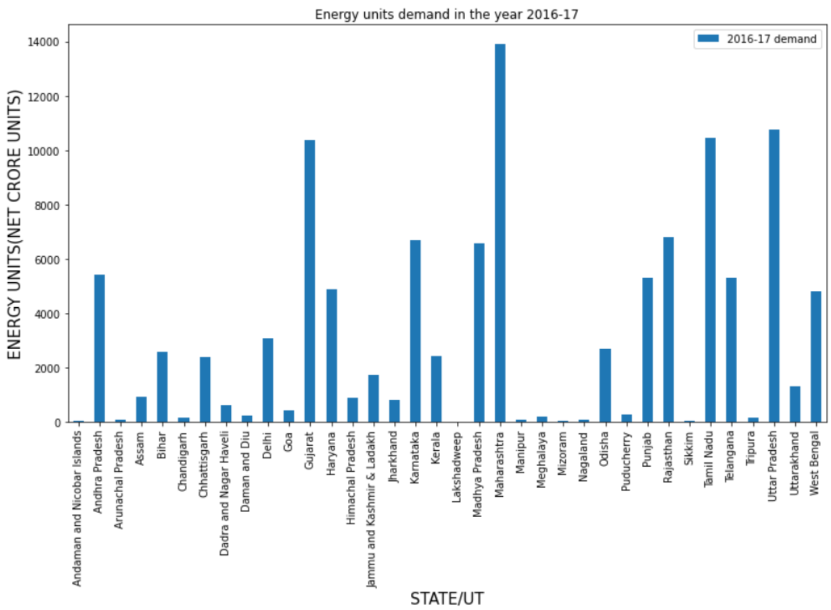


Fig 8: Graph showing Energy units demand in the year 2016-17

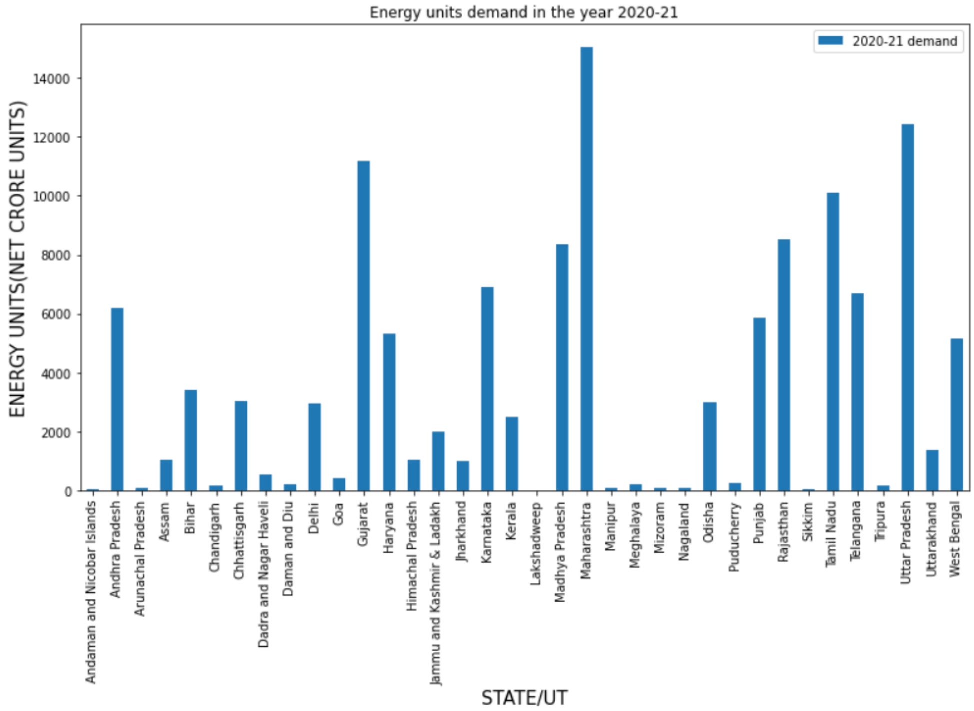


Fig 9: Graph showing Energy units demand in the year 2020-21

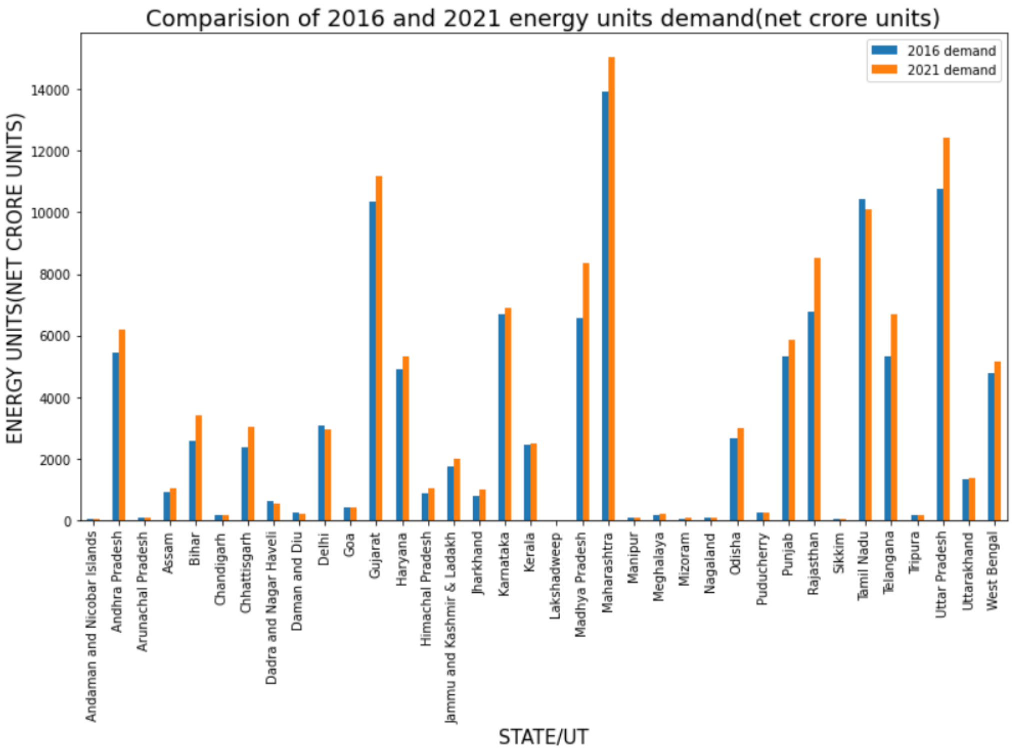


Fig 10: Graph showing Comparison of 2016 and 2021 energy units demand (net crore units)

- From the analysis 1.1 of this report, and the graphs shown above, we can conclude that the demand for power units has increased significantly from the year 2016 to 2021.

- It can infer that the use resources have significantly increased in these years.

1.2:

The data that we have used is quantitative and parametric.

Since only the data and not the ranks are being used to calculate the correlation, so we use the Pearson's rank correlation method.

For the same procedure table 1.1 and 1.2 have been used. From both tables the column used is 6^{th}. The formula used is (12).

Terminology:

r = Pearson coefficient

∑ = summation of symbol

n = number of observations

X=correlation data 1 values

Y= correlation data 2 values

Values obtained using python code:

```
Values obtained using python code
2020-21 demand(X)     125457
2020-21 production(Y) 368471
X^2                   1025543187
Y^2                   9396514195
XY                    3045338804
```

So, n = 36

Σx = 125457

Σy = 368471

Σx2 = 1025543187

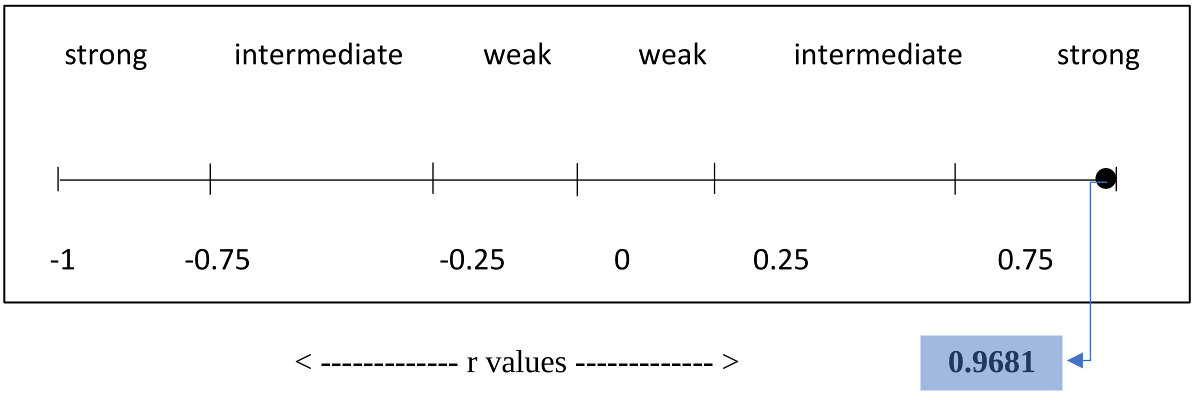
Σy2 = 9396514195

Σxy = 3045338804

calculation:

```
Values obtained using python code
0.9681
```

So, r = 0.9681



Since the value of r is very close to one so we can infer about the strong correlation between demand and production of power units.

Result:

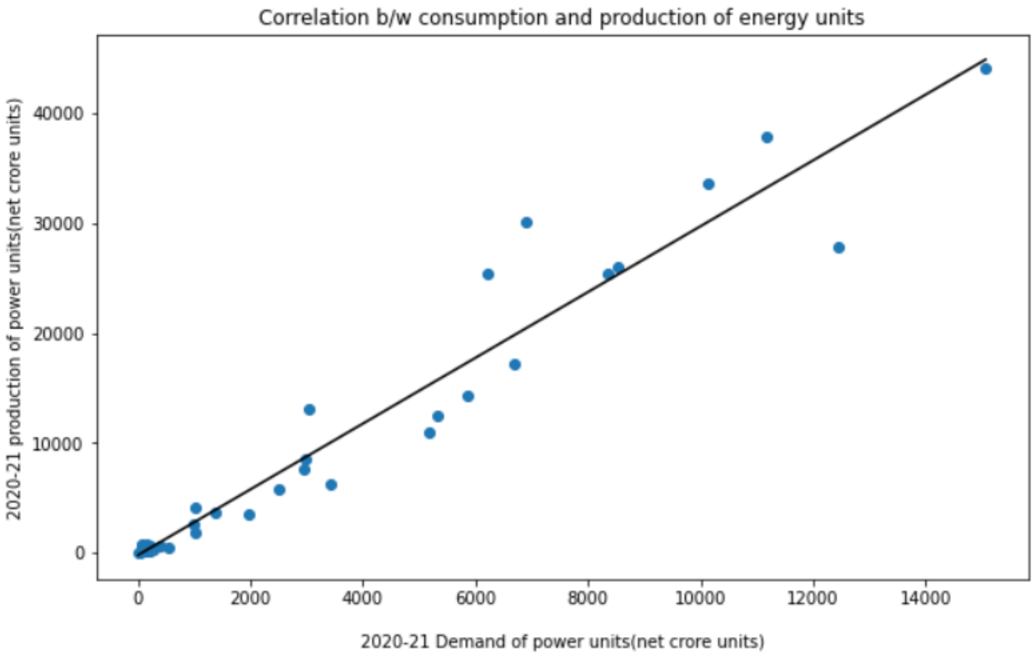


Fig 11: Graph showing Correlation b/w consumption and production of energy units

- The above scatter graph shows the strong positive bivariate linear correlation between the 2 attributes.
- As the value of correlation coefficient is quite high so we can conclude that there is a strong correlation between the demand of power units and the production of it.
- As the demand increases, the production also increases, and there is more use of resources.

1.3:

Linear regression is applied to predict the mean of total demand of power in the next 5 years 2021-22 to 2026-17 based on the current statistics.

For the same procedure, all the columns of the table 1.1 have been used. The formulas used are (14) and (15).

For the prediction for the next five years, we ought to apply the regression line expression which includes ‘b’ value. This ‘b’ value is calculated using the terms which are calculated below using the python code whose output has been shown below:

```
                Y     X      XY     X^2
2016-17    11242    1   112428     1
2017-18    119220   2   238440     4
2018-19    125264   3   375792     9
2019-20    126899   4   507596    16
2020-21    125457   5   627285    25
Sigma      60926   15  1861541    55
Mean    121853.6 3.0 372308.2 11.0

the value of b = 3373.7
```

Therefore, the regression equation is obtained as follows:

$$y=\bar{Y}+b(x-\bar{X})$$

$$Y=121853.6 + 3373.7(x - 3)$$

$$Y=121853.6 + 3373.7X - 10121.1$$

$$\boxed{Y=111732.5 + 3373.7X}$$

From this equation, we can predict the values of demand of power units in the next five years.

The python output for the predicted values of demand of power units are:

```
         Predicted demand units
2021-22                131974.7
2022-23                135348.4
```

2023-24	138722.1
2024-25	142095.8
2025-26	145469.5

Result:

Current statistics:

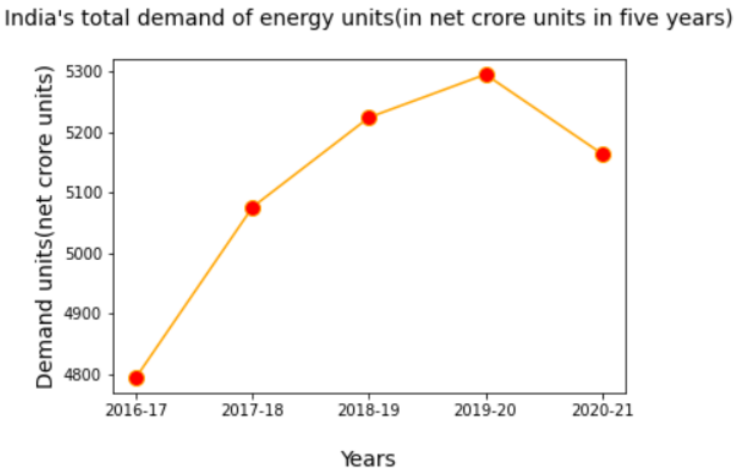


Fig 12: Graph showing the county's demand for power in the years from 2016-17 to 2020-21.

Current and predicted statistics:

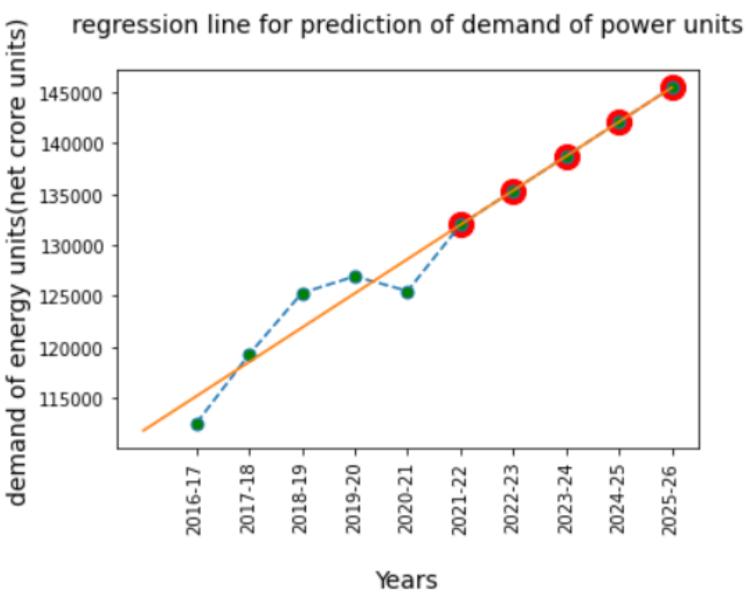


Fig 13: Graph showing county's predicted values for demand for power units in the next five years from 2021-22 to 2025-26.

- The above graph shows that in the last year there is a little decline in the way demand for power consumption used to be. It also shows that the demand for power had been continuously increasing over the years.
- The above graph shows the future predicted values of the demand of power units which may take place. The markers in red and green circles, show the next predicted values. The line in yellow depicts the regression line for the regression equation derived above.

1.4:

Since we need the analysis for 36 states and union territories over a 5-year period and we are using more than 2 variables here, so we use the ANOVA test.

For the same procedure all the columns of the table 1.2 has been used. The formula used are (10), (11), (12), (13), (14).

Step 1: Formulation of Hypothesis:

Let $\mu1$, $\mu2$, $\mu3$, $\mu4$…… $\mu36$ are the mean value of the region 1,2,3,4…...36 respectively, then the null and the alternate hypothesis are as follows:

Null Hypothesis:

- $\mu1 = \mu2 = \mu3 = \mu4 = \ldots \mu35 = \mu36 =$ k.
- There is no significant difference of all population(regions) mean.

Alternative Hypothesis:

- $\mu1 \neq \mu2 = \mu3 = \mu4 = \ldots \mu35 = \mu36 =$ k.
- Not all μ_j are equal to k, where j belongs to [1,36].

Step 2: calculate total T of the observation in the sample.

From python calculation,

```
T value is equal to 1706412
```

So, the value of T = 1706412.

Step 3: calculate correlation factor CF.

From python calculation,

```
The value of CF 16176899520.8
```

So, the value of CF is 16176899520.80.

Step 4: Calculate total sum of squares of deviation SST.

From python calculation,

```
Value of SST is 24477767781.2
```

So, the value of SST is 24477767781.2

Step 5: calculate sum of squared difference between the sample means SSB.

From python calculation,

```
Value of SSB is 24272781345.999996
```

So, the value of SSB is 24272781345.999996

Step 6: calculate sum of squares of difference within the sample SSW.

From python calculation**,**

```
value of SSW is 204986435.20000458
```

So, the value of SSW is 204986435.20000458

Step 7: Make ANOVA table.

After putting the values obtained in the previous steps (2 to 6), we can obtain the ANOVA table as shown below.

Source of variation	**Sum of squares**	**Difference of squares**	**Mean squares**	**Test of states/F value**
Between Sample	24272781345.999996	36-1=35	693508038.45	487.18
Within Sample	204986435.20000458	180-36=144	1423516.91	

Total	24477767781.2	180-1=179		

Table 1.3: Table showing the ANOVA table after putting the actuals values of different variables.

Step 8: Accept or reject the null hypothesis by analysing the calculated and the tabled value of F.

From previous calculations, the calculated value of F is 487.18 and the tabled F value is 1.505.

Since, the calculated value of F is much greater than the tabled value, so we reject the null hypothesis and accept the alternate hypothesis.

So, there is a significant difference of means of all populations(regions) in the five-year time span.

Result:

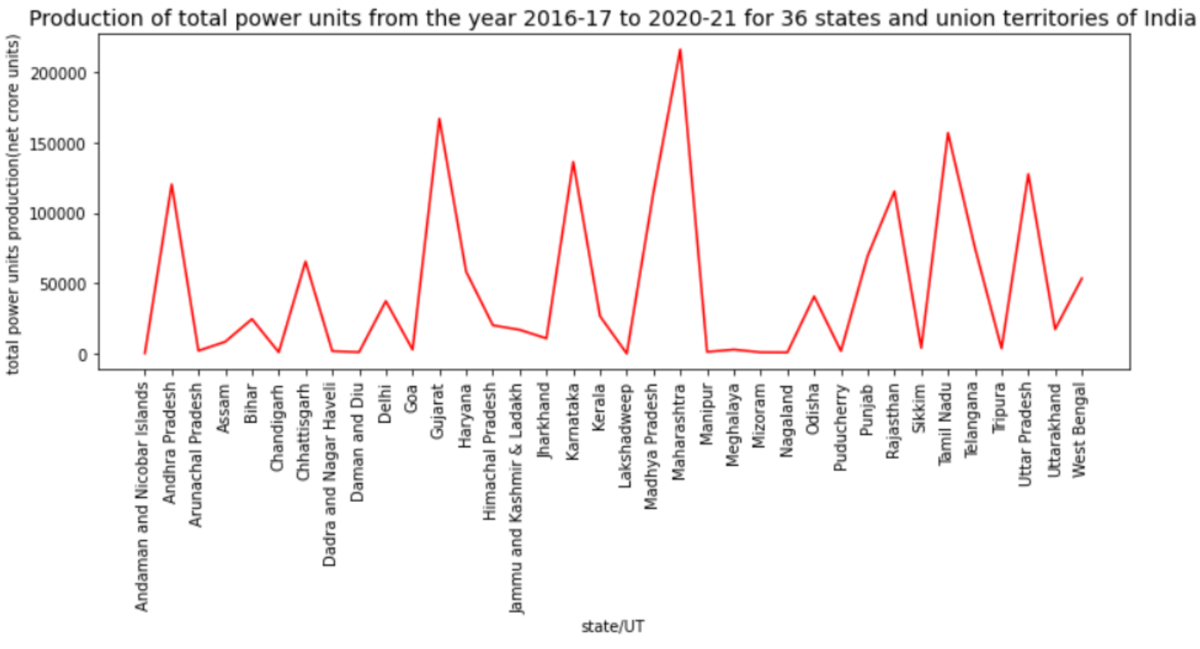


Fig 14: Graph showing Production of total power units from the year 2016-17 to 2020-21 for 36 states and union territories of India

From analysis 1.4, and the graph shown above, we can conclude that there is a significant difference between the total production of all the states/Ut in India

- It is not produced in the same number in every state.
- So, there is more risk of environmental issues in the states where the production is more.

Analysis 2

2.1:

To analyse the same objective, we have used the power method to determine the rank of various states on the basis of number of beds provided.

(Alphabet in table 2.2 represents states in the same order of table 2.1)

In the table 2.2, we mentioned the matrix (31X31) of 31 different states.

- In order to determine the rank, we first formed the matrix of (31X31) containing elements 0,1,2.
- Based on our data collection of years (2019), we are comparing state by state number of beds in hospitals.
- If both the states are similar, we consider the element as 0.
- If the states which are present in row has greater number of beds present the state in column, we consider the element as 1.
- If the states which are present in column has lesser number of beds present, then we consider the element as 2.

Table 2.2 : Matrix of no. of beds in different states.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	a	b	c	d	e
A	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2
B	1	0	2	2	2	1	2	2	2	1	1	1	1	1	2	2	2	2	2	1	1	1	1	2	2	2	2	2	2	2	2
C	1	1	0	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	1
D	1	1	2	0	2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	2	2	1	2	2	2	2	2	2	2	2
E	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	1
F	1	2	2	2	2	0	2	2	2	1	1	1	1	1	2	2	2	2	2	1	1	2	2	1	2	2	2	2	2	2	2
G	1	1	2	2	2	1	0	2	1	1	1	1	1	1	2	1	1	2	1	1	1	2	1	2	2	2	2	2	2	2	2
H	1	1	2	2	2	1	2	0	1	1	1	1	1	1	1	1	2	1	1	1	2	2	1	2	2	2	2	2	2	2	2
I	1	1	2	2	2	1	2	2	0	1	1	1	1	2	1	1	2	1	1	1	2	2	1	2	2	2	2	2	2	2	2
J	1	2	2	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2
K	1	2	2	2	2	2	2	2	2	1	0	2	2	2	2	2	2	2	1	1	1	2	1	2	2	2	2	2	2	2	2
L	1	2	2	2	2	2	2	2	2	1	1	0	1	2	2	2	2	2	1	1	2	2	1	2	2	2	2	2	2	2	2
M	1	2	2	2	2	2	2	2	2	1	1	2	0	2	2	2	2	2	1	1	2	2	1	2	2	2	2	2	2	2	2
N	1	1	2	1	2	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	2	1	2	2	2	2	2	2	2	2
O	1	1	2	2	2	1	2	2	2	1	1	1	1	2	0	2	2	2	1	1	2	2	1	2	2	2	2	2	2	2	2
P	1	1	2	2	2	1	2	2	2	1	1	1	1	2	1	0	2	1	1	1	2	2	1	2	2	2	2	2	2	2	2
Q	1	1	2	1	2	1	1	1	1	1	1	1	1	2	1	1	0	1	1	1	1	2	1	1	2	2	2	2	2	2	2
R	1	1	2	2	2	1	2	2	2	1	1	1	1	2	1	2	2	0	1	1	2	2	1	2	2	2	2	2	2	2	2
S	1	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	0	1	2	2	1	2	2	2	2	2	2	2	2
T	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2
U	1	1	2	1	2	1	1	1	1	1	1	1	1	2	1	1	2	1	1	1	0	2	1	1	2	2	2	2	2	2	2

V	1	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	2	2	2	2	2	2	1
W	1	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	1	2	2	0	2	2	2	2	2	2	2	2
X	1	1	2	1	2	1	1	1	1	1	1	1	1	2	1	1	2	1	1	1	2	2	1	0	2	2	2	2	2	2	2
Y	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	2	1	1	1	1
Z	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	0	2	2	2	2	1
a	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
b	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	0	1	2	1
c	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	0	2	1
d	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	1	0	1
e	1	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	2	2	2	2	2	2	0

Here, we use python code to find the dominant eigen value and eigen vectors.

The the last two steps are as follows.

Iteration:

```
Iteration 7: [[0.9778] [0.795] [0.5979] [0.685] [0.5823] [0.8326] [0.7226]
[0.7169] [0.748] [0.9562] [0.8791] [0.8551] [0.8745] [0.655] [0.8 ] [0.765 ]
[0.6543] [0.7823] [0.9142] [1.    ] [0.6692] [0.6114] [0.935] [0.6843]
[0.5229] [0.5717] [0.5113] [0.5469] [0.5469] [0.5348] [0.6253]] [43.1068]
```

```
Iteration 8: [[0.9778] [0.795] [0.5979] [0.685] [0.5823] [0.8326] [0.7226]
[0.7169] [0.748] [0.9562] [0.8791] [0.8551] [0.8745] [0.655] [0.8   ] [0.765 ]
[0.6543] [0.7823] [0.9142][1.    ] [0.6692] [0.6114] [0.935] [0.6843] [0.5229]
[0.5717] [0.5113] [0.5469] [0.5469] [0.5348] [0.6253]] [43.1068]
```

So, now we can find our rank of 31 different states

Rank 1 – Tamil Nadu

Rank 2 – Andhra Pradesh

Rank 3 – Karnataka

Rank 4 – utter Pradesh

Rank 5 – Rajasthan

Rank 6 - Kerala

Rank 7 –Maharashtra

Rank 8 – Madhya Pradesh

Rank 9 –Gujarat

Rank 10 – Bihar

Rank 11- NCT of Delhi

Rank 12- Punjab

Rank 13 - Odisha

Rank 14- Jharkhand

Rank 15- Himachal Pradesh

Rank 16-Haryana

Rank 17- Chhattisgarh

Rank 18- Uttarakhand

Rank 19- Telangana

Rank 20- Puducherry

Rank 21- Meghalaya

Rank 22- Tripura

Rank 23- West Bengal

Rank 24- Chandigarh

Rank 25- Goa

Rank 26- Arunachal Pradesh

Rank 27- Sikkim

Rank 28- Mizoram

Rank 29-Andaman and Nicobar

Rank 30-Lakshadeweep

Rank 31- Nagaland

Since each column is compared with every row**.** So, we have ranked the state in the order of state have highest beds at the top in comparison with the Eigen vector.

2.2

Claim: Number of mean beds in respective hospitals has increased significantly from 2019 to 2020.

For the same procedure, table 3.1 has been used and the columns used are first(states), third (**Beds in Government Hospitals2019)** and fifth (**Beds in Government Hospitals2020)**

Step 1: Formulation of Hypothesis:

If $\mu 1$ and $\mu 2$ are the means of number of beds in hospitals in 2020 and 2019 respectively, then the null and the alternate hypothesis are as follows:

Null Hypothesis:

- $\mu 1 \geq \mu 2$
- The mean number of beds in hospitals in 2020 is greater than or equal to the number of beds in hospitals in the year 2019.

Alternative Hypothesis:

- $\mu 1 < \mu 2$
- The mean number of beds in hospitals in 2020 is lesser than the beds in hospitals in the year 2019.

Terminology:

Level of significance: α

Sample size: n1, n2

$\bar{X}_1$=sample mean of beds in hospitals in 2020

$\bar{X}_2$=sample mean of beds in hospitals in 2019

Z=z score

$\sigma_{1=}$ standard deviation of beds in hospitals in 2020

$\sigma_{2=}$ standard deviation of beds in hospitals in 2019

Test Statistics:

Now as our sample size i.e., n >30 and the population standard deviation is known, so we will use z- test in order to find the difference of mean.

Here, $\mu 1 \geq \mu 2$, so the left tail test shall be used.

The formula used in this analysis are same as that used in analysis 2.1 of this report.

Step 2: Finding the value of z by deciding the value of α:

Means and standard deviation using python code

```
mean for 2019 data is 22514.064516129034
```

```
mean for 2020 data is 25099.935483870966

standard deviation for 2019 data is 26440.036718249194

standard deviation for 2020 data is 28576.64893689191

The value of Z is 0.37
```

Level of significance: $\alpha=0.05$

Sample size: n1=31, n2= 31(for both the columns)

$\bar{X}_1$= 22514.064516129034 net crore units (column5 – table 3.1 used)

$\bar{X}_2$= 25099.935483870966 net crore units (column3– table 3.1 used)

σ_1= 26440.036718249194 (column5 – table 3.1 used)

σ_2= 28576.64893689191 (column3 – table 3.1 used)

From the figures, formulas and data as mentioned above, the calculated of z obtained by python implementation is **z= 0.37**

Step 3: Get the tabled value of z_α based on α value decided:

For $\alpha=0.05$, the tabled value of $\mathbf{Z_\alpha= -1.645}$

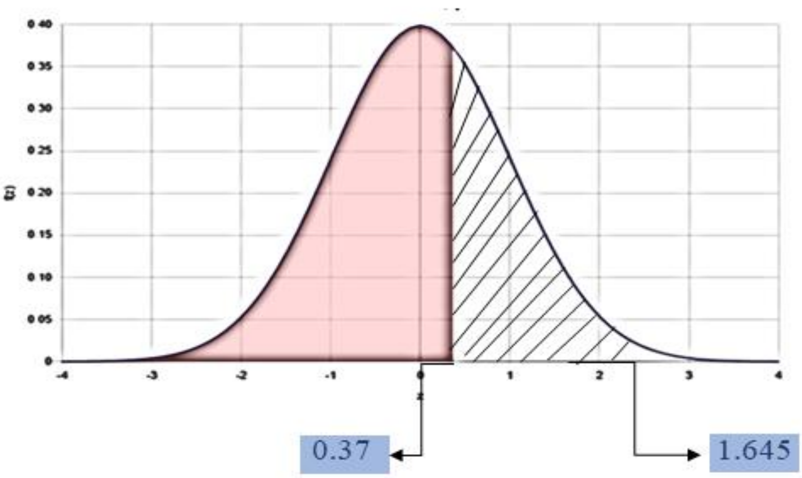


Fig 15: bell curve depicting the critical value , calculated z value and the critical region.

Step 4: Reject or accept the null hypothesis based on calculated and tabled value of z.

From the above calculations, the calculated and the tabled values are 0.37 and 1.645 respectively. Since calculated value of z = 0.37<1.645 (tabled value), so it is concluded that the null hypothesis is not rejected, and the number of beds has significantly increased.

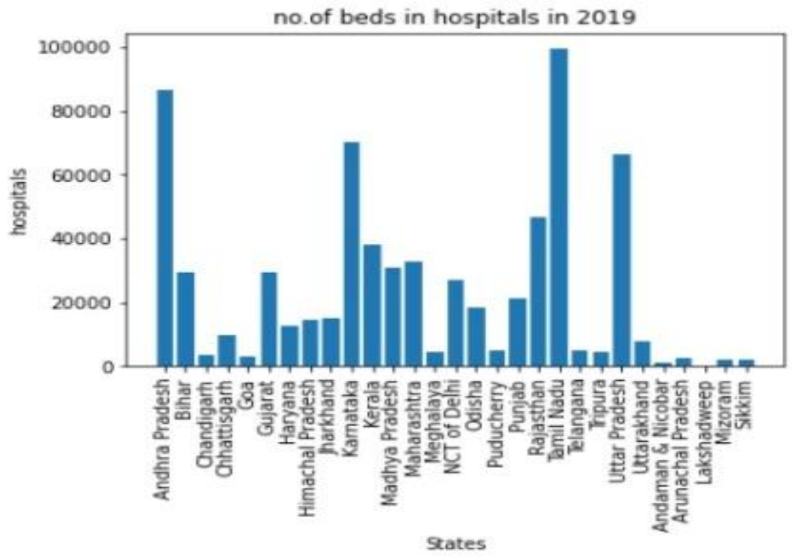


Fig 16: Graph showing no. Of beds in hospitals in year 2019.

- From the above graph we can conclude that number of beds in Tamil Nadu is highest followed by Andhra Pradesh.

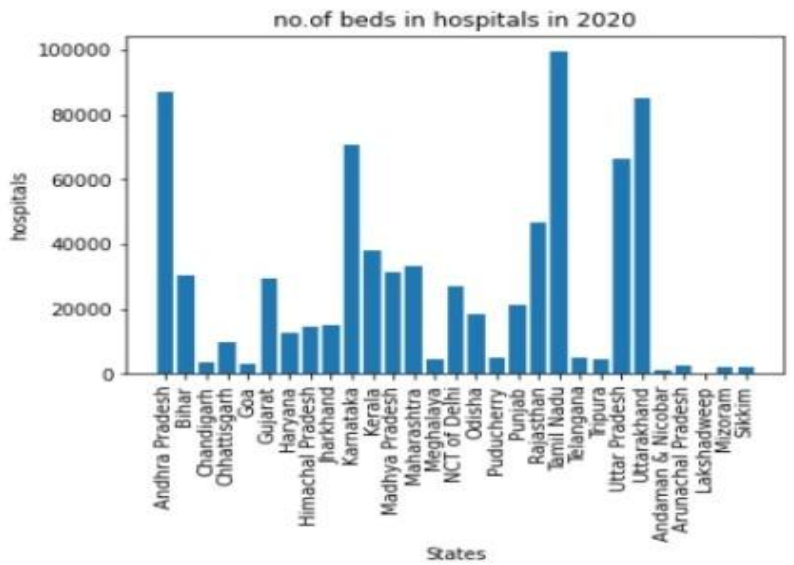


Fig 17: Graph showing no. Of beds in hospitals in year 2020.

- From the above graph we can conclude that beds in Tamil Nadu is highest in the year 2020 followed by Andhra Pradesh.
-

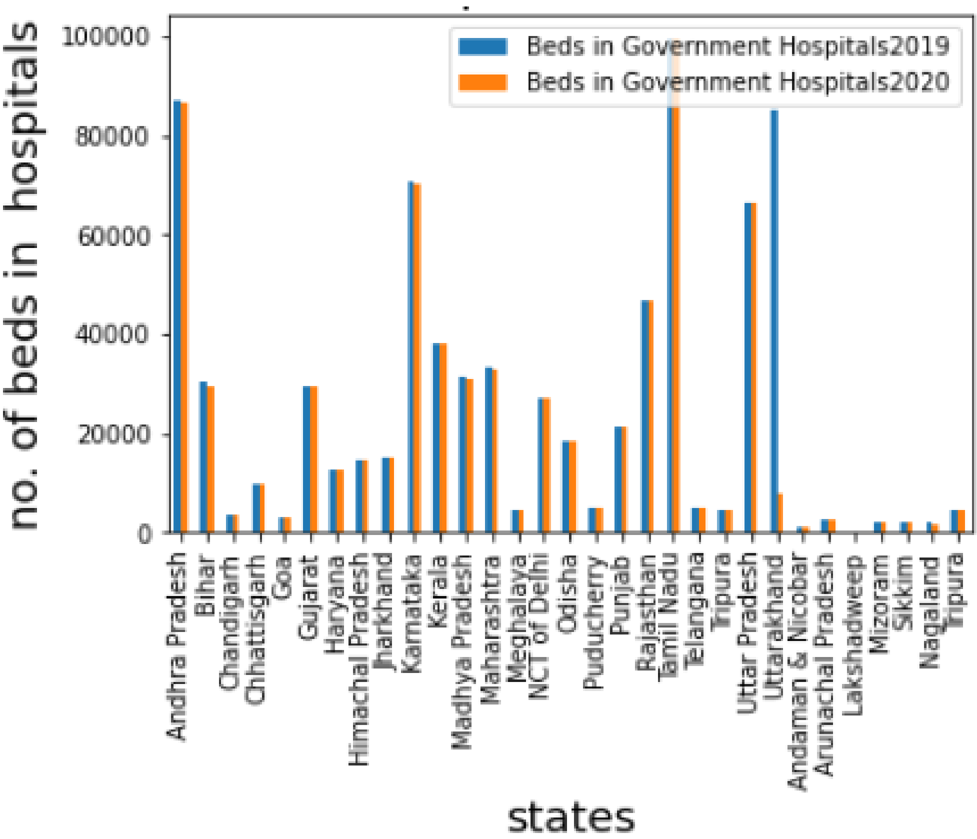


Fig 18: Graph showing Comparison of no. Of beds in hospitals in year 2019 and 2020.

From the analysis 2.2 of this report, and the graphs shown above, we can conclude that the number of beds in hospitals has significantly increased from year 2019 to 2020. Therefore, the consumption of medical resources (beds) has also increased.

2.3:

For the same procedure table 3.1 have been used. From the tables the column used is 6th.

The methodology and the formula used in this analysis are same as that used in analysis 1.2 of this report.

Terminology:

r = Pearson coefficient

∑ = summation of symbol

n = number of observations

X=data column 5 values

Y=data column 6 values

Values used:

$n = 31$

$\Sigma x = 778060$

$\Sigma y = 1131053747$

$\Sigma x2 = 197005101$

$\Sigma y2 = 4.1558E+16$

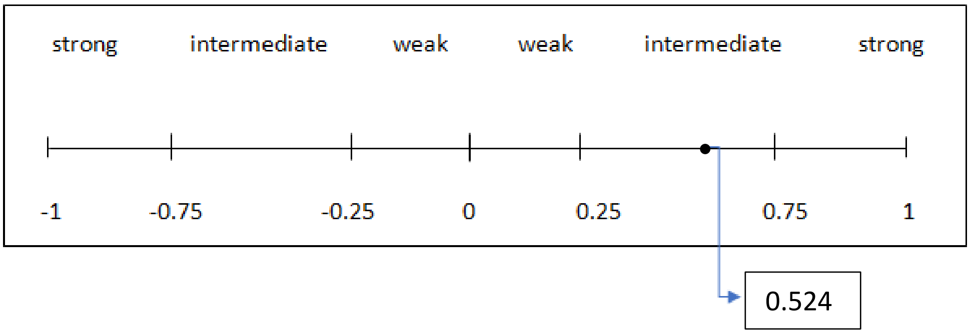
$\Sigma xy = 2.22996E+13$

calculation:

Value obtained using python code

So, $r = 0.5504$

So, there is an intermediate correlation between the number of beds provided and population.



< ------------- r values ------------- >

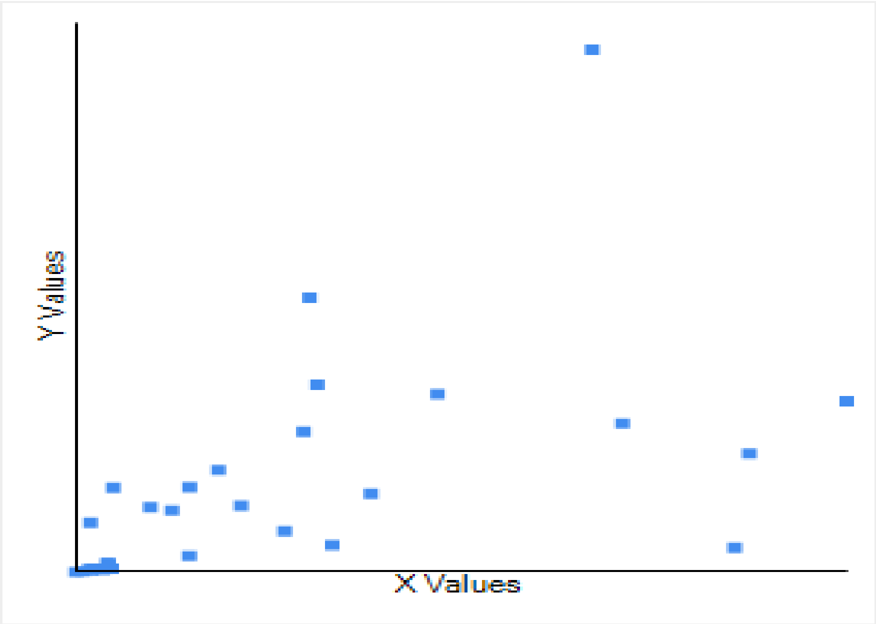


Fig 19: Graph showing Correlation b/w number of beds in hospitals and population.

- The above scatter graph shows the strong positive bivariate linear correlation between the 2 attributes.
  - As the value of correlation coefficient is quite high so we can conclude that there is a intermediate correlation between the number of beds in hospital and population.
  - As the population increases, the production of beds also increases, and there is more use of resources.

Analysis 3

To analyse production and consumption of food grains in India

3.1

The data consist of two tables which we have collected to analysis the production and consumption of food. It includes one table 3.1, which comprises of production of rice of year 2018-19 and 2019-20 of 31 states. It also contains the population in 2019-20. Similarly, in the table 3.2 the total production of rice and pulses of both the years of 31 states is shown.

For the above test, column 4 and 3 from table 3.2 have been used which are Total production of rice and pulses (2018-19), and Total production of rice and pulses (2019-20).

Step 1: Formulation of Hypothesis:

If $\mu1$ and $\mu2$ are the means of total production of rice and pulses in 2020 and 2019 respectively, then the null and the alternate hypothesis are as follows:

Null Hypothesis:

- $\mu1 \geq \mu2$
- The mean of total production of rice and pulses in 2020 is greater than or equal to (increased) the mean of total production of rice and pulses in 2019.

Alternative Hypothesis:

- $\mu1 < \mu2$
- The mean of total production of rice and pulses in 2020 lesser than the mean of total production of rice and pulses in 2019.

Terminology:

Level of significance: α

Sample size: n1, n2

$\bar{X}_1$=sample mean of total production of rice and pulses in 2020

$\bar{X}_2$=sample mean of total production of rice and pulses in 2019

Z=z score

$\sigma_{1=}$ standard deviation of total production of rice and pulses in 2020

$\sigma_{2=}$ standard deviation of total production of rice and pulses in 2019

Test Statistics:

The formulas and the methodology used for this objective are same as those used in analysis 1.1 of this report.

Step 2: Finding the value of z by deciding the value of α:

Mean and standard deviation through python code

```
Mean of Total production of rice and pulses (2019-20):
4575.225806451612
Mean of Total production of rice and pulses (2018-19):
4467.9290322580655
Standard deviation of Total production of rice and pulses (2019-20):
4921.26792083556
Standard deviation of Total production of rice and pulses(2018-19):
4978.1387168427755
```

Level of significance: $\alpha=0.05$

Sample size: n1=31, n2=31 (for both the columns)

$\bar{X}_1$= 4575.225806451612 in thousand tons (column 4 – table 3.2 used)

$\bar{X}_2$= 4467.9290322580655 in thousand tons (column 3 – table 3.2 used)

$\sigma_{1}=$ 4921.26792083556 (column 4 – table 3.2 used)

$\sigma_{2}=$ 4978.1387168427755 (column 3 – table 3.2 used)

From the formulas and data as mentioned above, the calculated of z obtained by python implementation is **z=0.085**

Step 3: Get the tabled value of z_α based on α value decided:

For $\alpha=0.05$, the tabled value of **Z_α= 1.645.**

Step 4: Reject or accept the null hypothesis based on calculated and tabled value of z

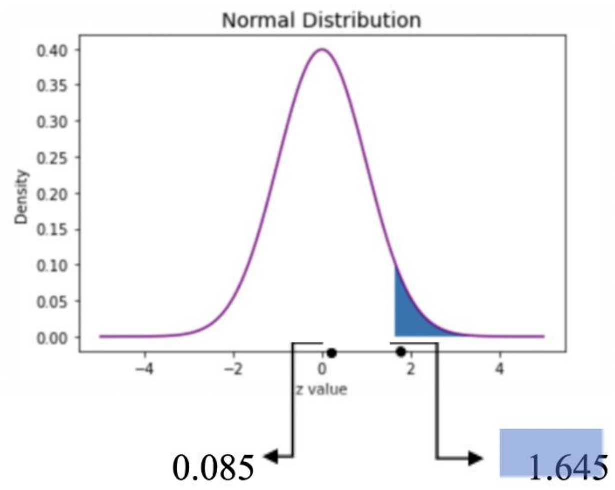


Fig 20: Bell curve depicting the critical value, calculated z value and the critical region.

Since calculated value of z = 0.085<1.645 (tabled value), so we accept the null hypothesis and reject the alternate hypothesis.

So, the mean of total production of rice and pulses in 2020 is greater than or equal to (increased) the mean of total production of rice and pulses in 2019.

Result:

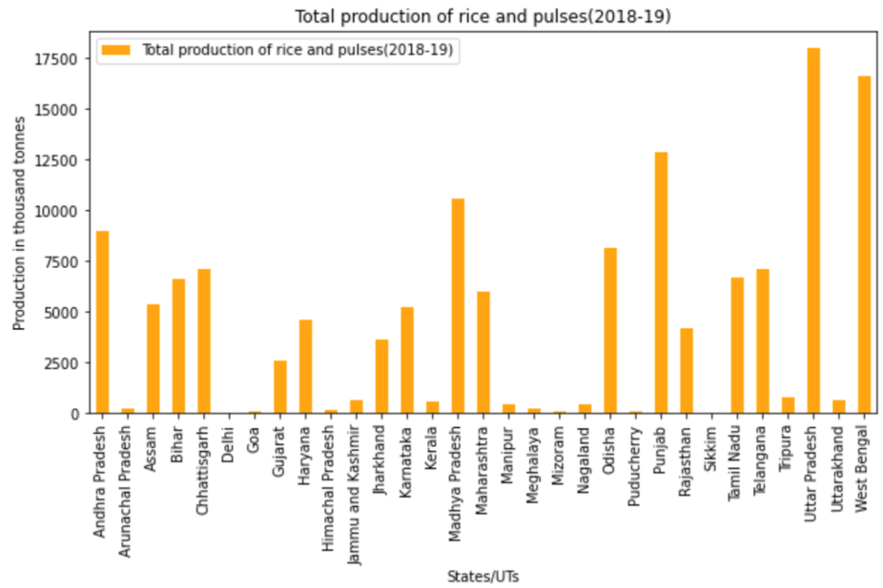


Fig 21: Graph showing Total production of rice and pulses(2018-19)

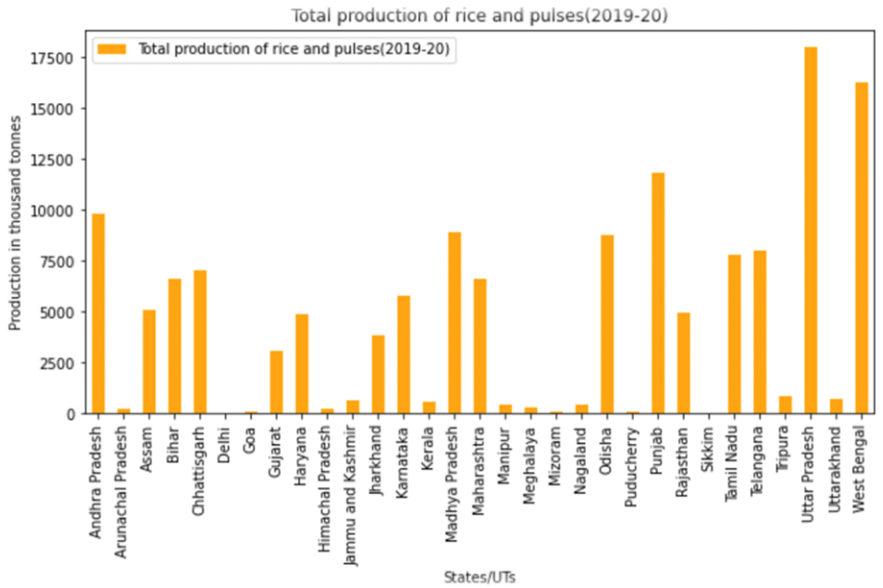


Fig 22: Graph showing Total production of rice and pulses(2019-20)

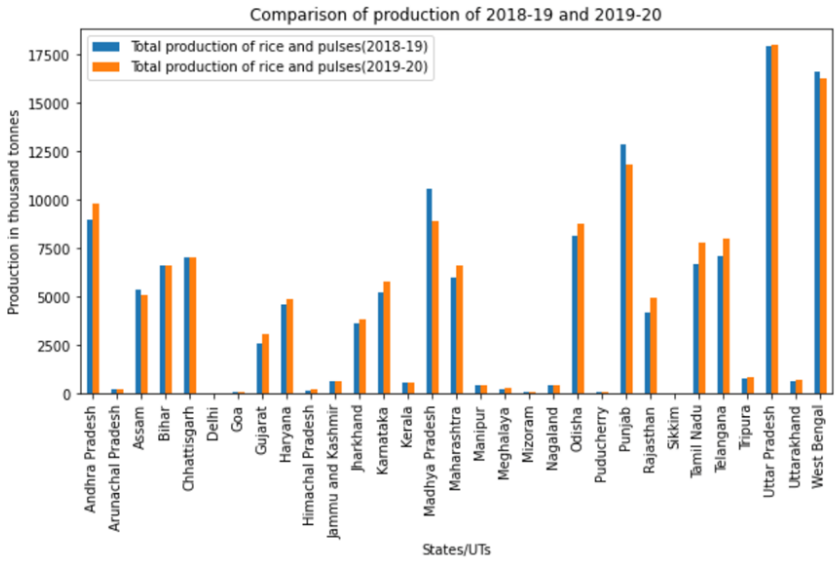


Fig 23: Graph showing comparison of production of 2018-19 and 2019-20

- With the help of the above graphs and analysis 3.1 we can say that the production of rice and pulses has significantly increased from 2018-19 to 2019-20.
- Therefore, it can infer that food waste has significantly increased from 2018-19 to 2019-20 years.

<u>3.2</u>

To study the effect on production of rice and pluses in 2019-20 with respect to the population of same year.

For the same procedure table 3.1 and 3.2 have been used. From table 3.1 the column used is 6^{th} and from table 3.2 the column used is 4^{th}.

The formulas, methodology and terminology used for this objective have been mentioned in analysis 1.2 of this report.

Values:

Values obtained through python code:

```
Population (2019-20) (X) 1.367955e+09
Total production of rice and pulses(2019-20)(Y) 1.418320e+05
X^2 1.397154e+17
Y^2 1.375480e+09
XY 1.221250e+13
```

So, N = 31

$\Sigma X = 1.367955e+09$

$\Sigma Y = 1.418320e+05$

ΣX^2 =1.397154e+17

ΣY^2 = 1.375480e+09

ΣXY = 1.221250e+13

Calculation:

Value of r obtained through python code:

```
Pearson’s correlation: 0.784
```

So, r = 0.784

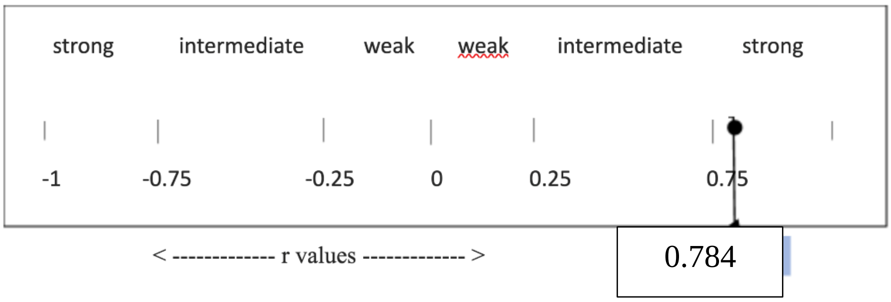


Fig 24: Correlation table showing the calculated r value.

Since the value of r is close to the 1 so we can infer about the strong correlation between population and the production of food grain.

Result:

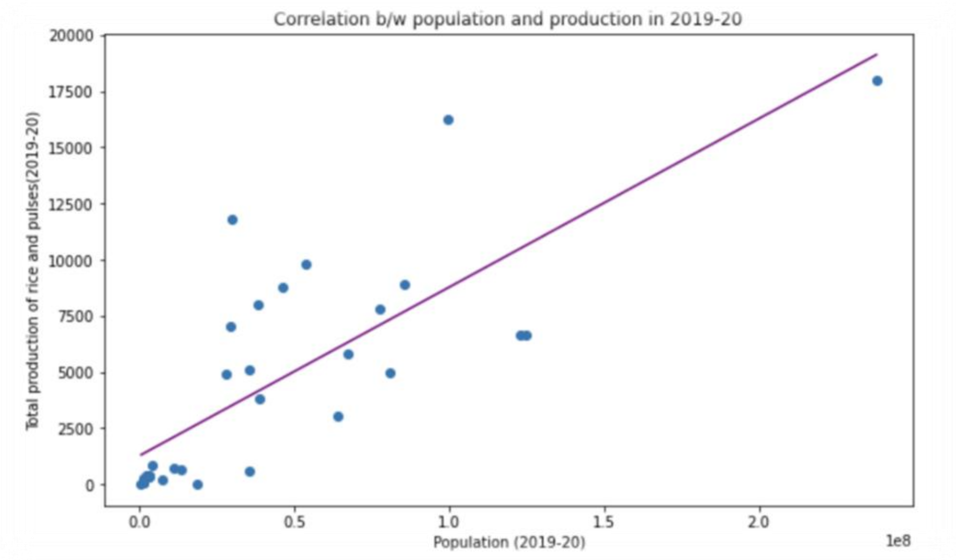


Fig 25: Graph showing correlation between population and production of rice and pulses in 2019-20.

- The above graph shows the strong positive correlation and from 3.2 analysis the correlation coefficient is high so we can conclude that there is the strong correlation between population and production of rice and pulses in the year 2019-20.
- As the population increases, the demand of food grains (rice and pulses) will also increase.

3.3

To designate different states with respect to sum of production of food grains (rice and pulses) in year 2018-2019 and 2019-2020(combined).

Table 3.3: Table shows States/UTs and denoted letters.

States/UTs	Denoted letters
Andhra Pradesh	A
Arunachal Pradesh	B
Assam	C
Bihar	D
Chhattisgarh	E
Delhi	F
Goa	G
Gujarat	H
Haryana	I
Himachal Pradesh	J
Jammu and Kashmir	K
Jharkhand	L
Karnataka	M
Kerala	N
Madhya Pradesh	O
Maharashtra	P
Manipur	Q
Meghalaya	R
Mizoram	S
Nagaland	T
Odisha	U
Puducherry	V
Punjab	W
Rajasthan	X
Sikkim	Y
Tamil Nadu	Z
Telangana	AA
Tripura	AB

Uttar Pradesh	AC
Uttarakhand	AD
West Bengal	AE

<u>Table 3.4</u>: Table(matrix) of production of rice and pulses.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE
A	0	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	1	2	2	2	2	2	1	2	1
B	1	0	1	1	1	2	2	1	1	2	1	1	1	1	1	1	1	1	2	1	1	2	1	1	2	1	1	1	1	1	1
C	1	2	0	1	1	2	2	2	2	2	2	2	1	2	1	1	2	2	2	2	1	2	1	2	2	1	1	2	1	2	1
D	1	2	2	0	1	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	1	2	1	2	2	1	1	2	1	2	1
E	1	2	2	2	0	2	2	1	2	2	2	2	2	2	1	2	2	2	2	2	1	2	1	2	2	1	1	2	1	2	1
F	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
G	1	1	1	1	1	2	0	2	1	1	1	1	1	1	1	1	1	1	2	1	1	2	1	1	2	1	1	1	1	1	1
H	1	2	1	1	1	2	2	0	1	2	2	1	1	2	1	1	2	2	2	2	1	2	1	1	2	1	1	2	1	2	1
I	1	2	1	1	1	2	2	2	0	2	2	2	1	2	1	1	2	2	2	2	1	2	1	2	2	1	1	2	1	2	1
J	1	1	1	1	1	2	2	1	1	0	1	1	1	1	1	1	1	1	2	1	1	2	1	1	2	1	1	1	1	1	1
K	1	2	1	1	1	2	2	1	1	2	0	1	1	2	1	1	2	2	2	2	1	2	1	1	2	1	1	1	1	1	1
L	1	2	1	1	1	2	2	2	1	2	2	0	1	2	1	1	2	2	2	2	1	2	1	1	2	1	1	2	1	2	1
M	1	2	2	1	1	2	2	2	2	2	2	2	0	2	1	1	2	2	2	2	1	2	1	2	2	1	1	2	1	2	1
N	1	2	1	1	1	2	2	1	1	2	1	1	1	0	1	1	2	2	2	2	1	2	1	1	2	1	1	1	1	1	1
O	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	2	2	2	2	2	2	2	1	2	2	2	2	2	1	2	1
P	1	2	2	1	1	2	2	2	2	2	2	2	2	2	1	0	2	2	2	2	1	2	1	2	2	1	1	2	1	2	1
Q	1	2	1	1	1	2	2	1	1	2	1	1	1	1	1	1	0	2	2	2	1	2	1	1	2	1	1	1	1	1	1
R	1	2	1	1	1	2	2	1	1	2	1	1	1	1	1	1	1	0	2	1	1	2	1	1	2	1	1	1	1	1	1
S	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	2	1	1	2	1	1	1	1	1	1
T	1	2	1	1	1	2	2	1	1	2	1	1	1	1	1	1	1	2	2	0	1	2	1	1	2	1	1	1	1	1	1
U	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	0	2	1	2	2	2	2	2	1	2	1
V	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	2	1	1	1	1	1	1
W	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	2	2	2	2	2	1	2	1
X	1	2	1	1	1	2	2	2	1	2	2	2	1	2	1	1	2	2	2	2	1	2	1	0	2	1	1	2	1	2	1
Y	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
Z	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	1	2	1	2	2	0	1	2	1	2	1
AA	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	1	2	1	2	2	2	0	2	1	2	1
AB	1	2	1	1	1	2	2	1	1	2	2	1	1	2	1	1	2	2	2	2	1	2	1	1	2	1	1	0	1	2	1
AC	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	2	2
AD	1	2	1	1	1	2	2	1	1	2	2	1	1	2	1	1	2	2	2	2	1	2	1	1	2	1	1	1	1	0	1
AE	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	0

The methodology used for the formation of this table is same as that used in analysis 2.1 of this report.

Eigen value and vector through python code

```
Iteration 1:
[0.9333] [0.6 ][0.8 ] [0.85 ] [0.85 ] [0.5 ][0.5833] [0.7333]
[0.7833][0.5833] [0.6833] [0.75 ] [0.8167] [0.6667] [0.95 ] [0.8333]
[0.65 ][0.6167] [0.55 ] [0.6333] [0.9167] [0.5333] [0.9667] [0.7667]
[0.5167][0.8833] [0.9 ][0.7167] [1. ] [0.7 ] [0.9833] [60]
```

…

…

…

```
Iteration 7 :

[[0.9145] [0.5851] [0.765 ] [0.8181] [0.8208][0.5113] [0.5749]
[0.6996] [0.7481] [0.5721] [0.6542][0.7154] [0.7823] [0.6398]
[0.9351] [0.8 ]
[0.6256] [0.5983] [0.5468] [0.6118] [0.8943] [0.5347] [0.9563]
[0.7316][0.5229] [0.8552] [0.8745] [0.6842] [1. ][0.669 ] [0.9779]
[43.2386]


Iteration 8 :
[0.9145] [0.5851] [0.765 ] [0.8181] [0.8208] [0.5113] [0.5749]
[0.6996][0.7481] [0.5721] [0.6542] [0.7154] [0.7823] [0.6398]
[0.9351] [0.8 ]
[0.6256] [0.5983] [0.5468] [0.6118] [0.8943] [0.5347] [0.9563]
[0.7316][0.5229] [0.8552] [0.8745] [0.6841] [1. 0 ] [0.669 ]
[0.9779] [43.2388]
```

So, according to the above eigen values following states are ranked.

Rank1= Uttar Pradesh

Rank 2= West Bengal

Rank 3= Punjab

Rank 4= Madhya Pradesh

Rank 5= Andhra Pradesh

Rank 6= Odisha

Rank 7= Telangana

Rank 8= Tamil Nadu

Rank 9= Chhattisgarh

Rank 10=Bihar

Rank 11= Maharashtra

Rank 12= Karnataka

Rank 13= Assam

Rank 14= Haryana

Rank 15= Rajasthan

Rank 16= Jharkhand

Rank 17= Gujarat

Rank 18= Tripura

Rank 19= Uttarakhand

Rank 20= Jammu and Kashmir

Rank 21= Kerala

Rank 22= Manipur

Rank 23= Nagaland

Rank 24= Meghalaya

Rank 25= Arunachal Pradesh

Rank 26= Goa

Rank 27= Himachal Pradesh

Rank 28= Mizoram

Rank 29= Puducherry

Rank 30= Sikkim

Rank 31=Delhi

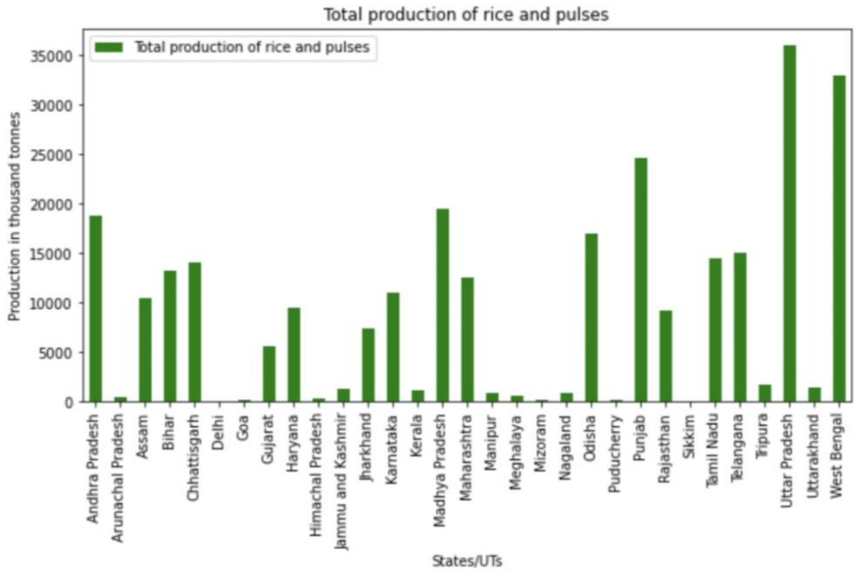


Fig 26: Graph showing total production of rice and pulses.

- From analysis 3.3 of this report and from the above graph, it is noticeable that Uttar Pradesh is the highest total producer of rice and pulses, and West Bengal is second.
- Whereas Sikkim and Delhi have the lowest total production of rice and pulses.

3.4

To analyse the consumption of rice and pulse 2019-2020 in different states of India and draw inference about the same**.**

To find the total production of food grains in the year 2019-20, we have used the python code to form the following dataset table.

Table 3.5: Population and Total production of rice and pulses (in thousand tones) in the year 2019-20 is shown for different Indian states/UT.

	States/UTs	**Population (2019-20)**	**Total production of rice and pulses (2019-20)**
1	Andhra Pradesh	53903393	9825.6
2	Arunachal Pradesh	1570458	258.9
3	Assam	35607039	5090.7
4	Bihar	124799926	6632.4
5	Chhattisgarh	29436231	7016.1
6	Delhi	18710922	16.8
7	Goa	1586250	94.3

8	Gujarat	63872399	3040.4
9	Haryana	28204692	4888.7
10	Himachal Pradesh	7451955	199.0
11	Jammu and Kashmir	13606320	631.2
12	Jharkhand	38593948	3827.7
13	Karnataka	67562686	5790.4
14	Kerala	35699443	607.8
15	Madhya Pradesh	85358965	8886.6
16	Maharashtra	123144223	6633.6
17	Manipur	3091545	410.7
18	Meghalaya	3366710	315.4
19	Mizoram	1239244	65.5
20	Nagaland	2249695	410.1
21	Odisha	46356334	8792.9
22	Puducherry	1413542	59.9
23	Punjab	30141373	11808.5
24	Rajasthan	81032689	4977.6
25	Sikkim	690251	21.1
26	Tamil Nadu	77841267	7776.5
27	Telangana	38510982	7977.0
28	Tripura	4169794	828.9
29	Uttar Pradesh	237882725	17965.2
30	Uttarakhand	11250858	716.2
31	West Bengal	99609303	16266.3

For the same procedure table 3.5 shall be used. From table 3.5 the column used is 2 and 3.

Facts and figures:

According to the research a person in India consumes an average of 75.284535 Kgs food grain (rice and pulses) every year.

For a particular state,

Total Consumption of food grains in thousand tons = (Population*average consumption by a person)/1000000

Using the above-mentioned formula, the derived table which consists of information regarding production and consumption of food grains is shown below.

Table 3.6: Total consumption and production of food gains (in thousand tons) in the year 2019-20 for different Indian states/UT.

	States/UTs	Total consumption of rice and pulses (2019-20)	Total production of rice and pulses (2019-20)
1	Andhra Pradesh	4058.091877	9825.6
2	Arunachal Pradesh	118.231200	258.9
3	Assam	2680.659374	5090.7
4	Bihar	9395.504397	6632.4
5	Chhattisgarh	2216.092963	7016.1
6	Delhi	1408.643062	16.8
7	Goa	119.420094	94.3
8	Gujarat	4808.603858	3040.4
9	Haryana	2123.377122	4888.7
10	Himachal Pradesh	561.016967	199.0
11	Jammu and Kashmir	1024.345474	631.2
12	Jharkhand	2905.527429	3827.7
13	Karnataka	5086.425399	5790.4
14	Kerala	2687.615966	607.8
15	Madhya Pradesh	6426.209988	8886.6
16	Maharashtra	9270.855566	6633.6
17	Manipur	232.745528	410.7
18	Meghalaya	253.461197	315.4
19	Mizoram	93.295908	65.5
20	Nagaland	169.367242	410.1
21	Odisha	3489.915049	8792.9
22	Puducherry	106.417852	59.9
23	Punjab	2269.179251	11808.5
24	Rajasthan	6100.508311	4977.6
25	Sikkim	51.965226	21.1
26	Tamil Nadu	5860.243590	7776.5
27	Telangana	2899.281372	7977.0
28	Tripura	313.921002	828.9
29	Uttar Pradesh	17908.890336	17965.2
30	Uttarakhand	847.015613	716.2
31	West Bengal	7499.040058	16266.3

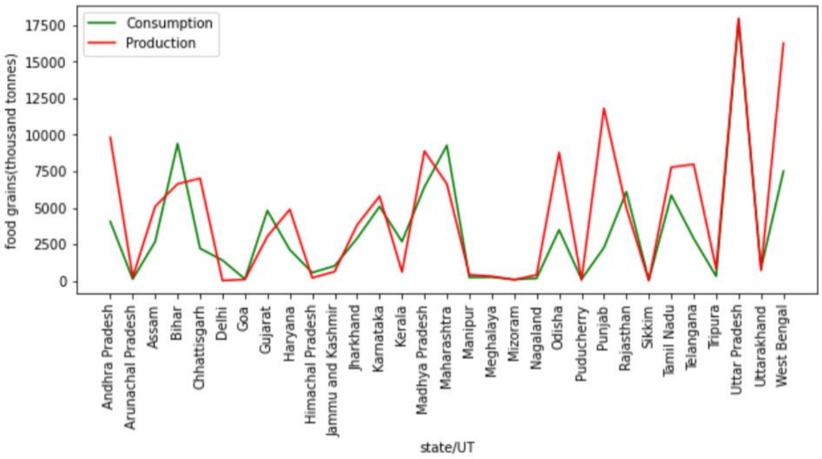


Fig 27: Graph showing the comparison of consumption and production of rice and pulses in 2019-20.

- From analysis 3.4 of this report and from the above graph, we can conclude that Uttar Pradesh was the largest producer as well as consumer of rice and pulse in 2019-20.
- It is also noticeable that the consumption of food grain (rice and pulses) is much more than the production in some states.

Analysis 4

To Analyse the Generation, Collection and hence the treatment of solid waste produced in different states.

Data Collection:

The data contains four tables of the waste generation, waste collection and waste treated in tonnes per day by the different states and union territories of India in year 2016-17 to 2019-20 for 25 states and union territories of India.

4.1

To designate different states in India based on total waste generated from year 2016-17 to 2019-20.

We can determine the rank of 25 different states of India on the basis of waste generated in 2016-2020.

In the below table, we mentioned the matrix (25X25) of 25 different states.

- In order to determine the rank, we first formed the matrix of (25X25) containing elements 0,1,2.
- Based on our data collection of years (2016-17 to 2019-20), we are comparing state by state data values of all four years (2016-17 to 2019-20).

- If both the states are similar, we consider the element as 0.
- If the states which are present in row is smaller than states which are present in column, we consider the element as 1.
- If the states which are present in column is greater than we will consider the element
-

Table 4.5: Matrix formed by the waste generated data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
Madhya Pradesh(A)	0	2	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	2	1	1	1	1	1
Maharashtra (B)	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Manipur(C)	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	1	2
Meghalaya(D)	2	2	1	0	1	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	1	2	2	1	2
Mizoram(E)	2	2	1	2	0	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	1	2	2	1	2
Nagaland(F)	2	2	1	1	1	0	2	2	2	1	2	2	2	2	2	2	2	2	2	2	1	2	2	1	1
Odisha(G)	2	2	1	1	1	2	0	2	2	1	2	2	2	2	2	2	2	2	2	2	1	2	2	1	1
Punjab(H)	2	2	1	1	1	1	1	0	1	2	2	2	1	1	2	2	1	1	1	2	1	2	1	1	1
Rajasthan(I)	2	2	1	1	1	2	2	2	0	1	2	2	2	2	2	2	2	2	2	2	1	2	1	1	1
Sikkim(J)	2	2	1	2	2	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	1	2	2	1	2

Tamil Nadu(K)	1	2	1	1	1	1	1	1	1	1	0	1	1	1	2	1	1	1	1	2	1	1	1	1	1
Telangana(L)	2	2	1	1	1	1	1	1	1	1	1	0	1	1	2	1	1	1	1	2	1	1	1	1	1
Tripura(M)	2	2	1	1	1	1	2	2	2	1	2	2	0	1	2	1	1	1	1	2	1	1	1	1	1
Uttarakhand (N)	2	2	1	1	1	1	2	2	2	1	2	2	1	0	2	2	2	1	2	2	1	2	1	1	1
Uttar Pradesh(O)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
West Bengal(P)	2	2	1	1	1	1	1	1	1	1	2	2	1	1	2	0	1	1	1	2	1	2	1	1	1
Andaman and Nicobar(Q)	2	2	1	1	1	1	1	2	1	1	2	2	1	1	2	2	0	1	1	2	1	2	1	1	1
Chandigarh (R)	2	2	1	1	1	1	2	2	2	1	2	2	1	2	2	2	2	0	2	2	1	2	1	1	1
Jammu and Kashmir(S)	2	2	1	1	1	1	2	2	2	1	2	2	1	1	2	2	2	1	0	2	1	2	1	1	1
Karnataka(T)	1	2	1	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	1	0	1	1	1	1	1
Lakshadweep(U)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	2	2	1	2
Delhi(V)	2	2	1	1	1	1	1	1	1	1	2	2	1	1	2	1	1	1	1	2	1	0	2	1	1
Goa(W)	2	2	1	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	1	2	0	1	2
Gujarat(X)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	0	2
Puducherry (Y)	2	2	1	1	1	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	1	2	1	1	0

Calculation:

Here, we use python code to find the dominant eigen value and eigen vectors.

The first and the last two steps are as follows.

```
Iteration 1:

[[0.5625] [0.5] [0.9583] [0.8958] [0.9167] [0.8542] [0.8542]
[0.6875] [0.8333] [0.9375] [0.5625] [0.5833] [0.6875] [0.7708] [0.5
] [0.6458] [0.6875] [0.7917] [0.75  ] [0.9792] [0.6458] [0.9167] [1.
] [0.8542]]      [48]
```

…

…

…

```
Iteration 4:

[[0.5591][0.5139] [0.9468] [0.8722] [0.8964] [0.8264] [0.8264]
[0.665 ] [0.8022] [0.9212] [0.5734] [0.6683] [0.7413] [0.5139]
```

```
[0.6222] [0.6579] [0.7619] [0.7213] [0.5579] [0.973 ] [0.6296]
[0.8964] [1.    ] [0.8245]]      [35.0506]
```

```
Iteration 5:
```

```
[[0.5591] [0.5139] [0.9468] [0.8722] [0.8963] [0.8264] [0.8264]
[0.665 ] [0.8022] [0.9212] [0.5733] [0.6683]
[0.6683][0.7413][0.5139][0.6222][0.6579] [0.7619] [0.7213] [0.5578]
[0.973 ] [0.6296] [0.8963] [1.    ] [0.8244]]     [35.0582][0.5139]
[0.9468] [0.8722]
```

So, now we can find our rank of 25 different states.

Rank 1 – Maharashtra

Rank 2 – Uttar Pradesh

Rank 3 – Karnataka

Rank 4 - Tamil Nadu

Rank 5 – Madhya Pradesh

Rank 6 - Telangana

Rank 7 –West Bengal

Rank 8 - Delhi

Rank 9 – Andaman and Nicobar

Rank 10 – Punjab

Rank 11- Tripura

Rank 12- Jammu and Kashmir

Rank 13 - Uttarakhand

Rank 14- Chandigarh

Rank 15- Puducherry

Rank 16- Rajasthan

Rank 17- Orissa

Rank 18- Nagaland

Rank 19- Meghalaya

Rank 20- Goa

Rank 21- Mizoram

Rank 22- Sikkim

Rank 23- Manipur

Rank 24- Lakshadweep

Rank 25- Gujarat

Since each column is compared with every row**.** So, we have ranked the state in the ascending order of values in the Eigen vector.

Result:

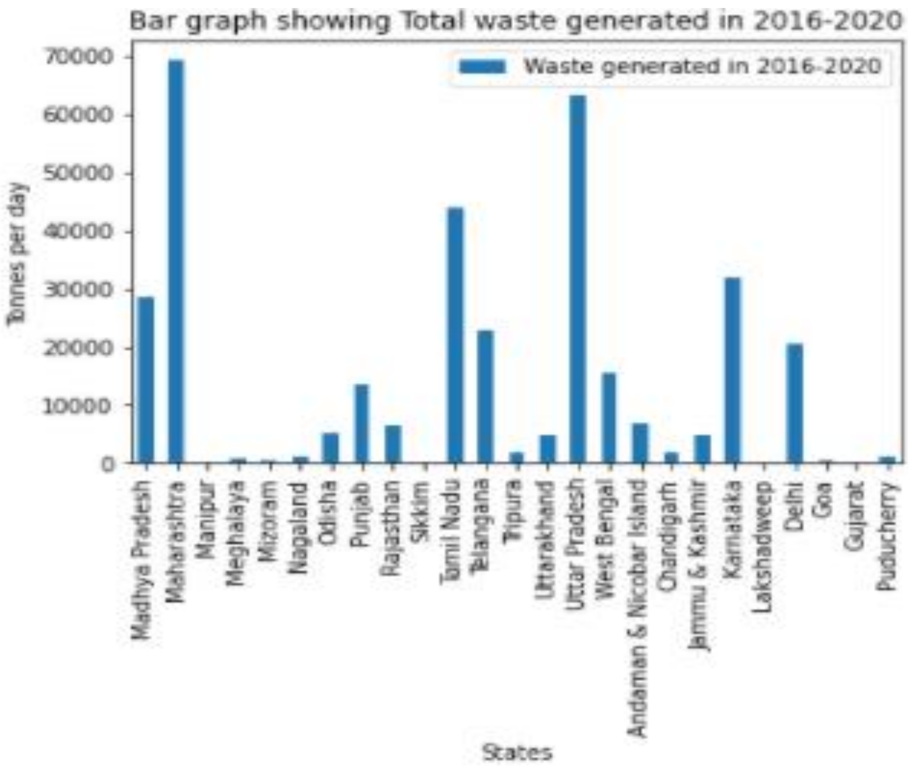


Fig 28*: Graph shows the total waste generated the year 2016-2020*

- From the power method performed in the analysis 4.1 and the graph shown above, we concluded that Maharashtra ranked highest in solid waste generated in year 2016-2020 and Gujarat ranked lowest in solid waste generation in 2016-2020. Basically, Maharashtra generated highest solid waste in year 2016-2020.

4.2

Claim: The mean of total solid waste treated is significantly less than the total solid waste collected in the years 2016-17 to 2019-20.

For the same procedure, column 4 of each table 4.1, 4.2, 4.3 and 4.4 have been used.

Step 1: Formation of Hypothesis

- Let μ_1 and μ_2 be the waste treated and collected in years from 2016 to 2020.

- Let S_1 and S_2 be the standard deviation of mean waste treated and collected from year 2016 to 2020.

- μ_1 = Mean of the Total Waste Treated.
- μ_2 = Mean of the Total Waste Collected.

- Null hypothesis (H_o) = $\mu_2 - \mu_1 = 0$
  There is a no significant difference between the means of the total waste treated and total waste collected.

- Alternate Hypothesis (H_a) = $\mu_2 - \mu_1 \neq 0$
  There is significant difference between the means of the total waste treated and total waste collected.

Step 2: Test statistics

Now as our sample size i.e., n<30 and there is a statistically significant difference between waste collected and waste treated (independent samples). Therefore, t-test can be applied. The formulas which have used are (4), (5).

Here, $\mu_2 - \mu_1 = 0$. So, right tailed test shall be used.

Step 3: Calculation

To find the value of Degree of Freedom:

Here,
Value of $t\alpha$
$\alpha = 0.05$
$df = n_1 + n_2 - 2$
$df = 25+25-2 = 48$

Value of $t\alpha$ from the table:
$t_\alpha = 1.677$

<u>Calculating mean and standard deviation by python code:</u>

```
Mean of the waste collected for the year 2016-2020 - 3548.00775
Standard Deviation of the waste collected for the year 2016-2020 - 1278.81
```

```
Mean of the waste treated for the year 2016-2020 - 1340.702499999999
```

```
Standard Deviation of the waste treated for the year 2016-2020 - 352.63
```

- For finding t, we have to find s_p

Value of s_p: `938.0`

Now, find the value of t:
Value of t is 8.32
So,
Level of significance is: $\alpha = 0.05$
Sample size: n1=25, n2=25 (for both the column)
$\bar{X}1$ = 3548.00775 in tonnes per day (column 2 – table 4.1,4.2,4.3,4.4 used)
$\bar{X}_2$= 1340.702499999999 in tonnes per day (column 3 – table 4.1,4.2,4.3,4.4 used)
$\sigma_{1=}$ 1278.81 (column 2 – table 4.1,4.2,4.3,4.4 used)
$\sigma_{2=}$ 352.63 (column 3 – table 4.1,4.2,4.3,4.4 used)

From the figures, formulas and data as mentioned above, the calculated value of t obtained by python implementation is **t is 8.32**

Step 4: Rejecting or accepting Null hypothesis

From the Right Tailed Test:
H_0: $\mu \leq k$
$H_{1:}$ $\mu > k$
Tabled value of t:
1.677

Result:

Reject Null Hypothesis

Since t > tα (1.677<8.32) the null hypothesis is rejected through the rule of right – tailed test, and the alternative hypothesis is accepted.

There is a significant difference between the means of the total waste treated and total waste collected.

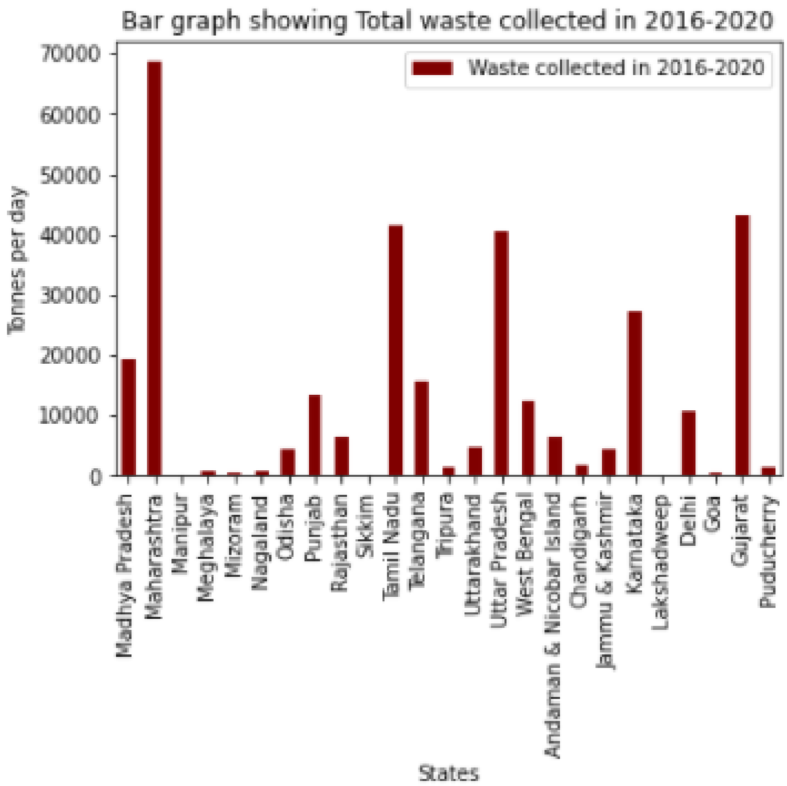


Fig 29: Graph shows the total waste collected the year 2016-2020.

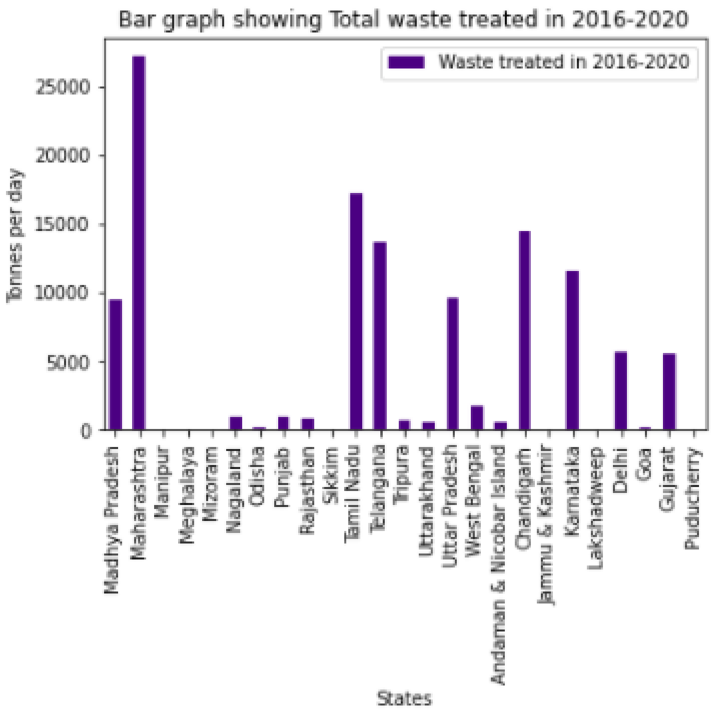


Fig 30: Graph shows the total waste treated the year 2016-2020

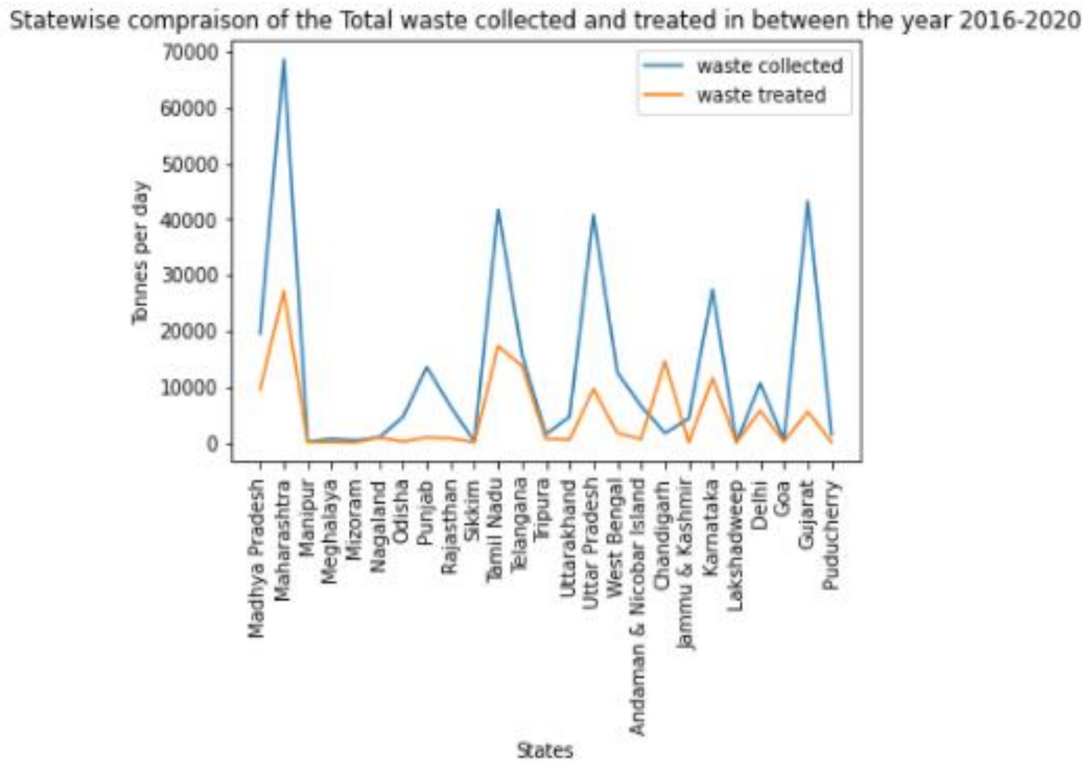


Fig 31: Graph shows the total waste collected and treated in between the year 2016-2020.

- Graph shows that for most of the states total waste Collected is very much higher than the total waste treated in years (2016-2020). Basically, the waste gets collected but gets treated in lesser amounts.
- This can pose various threats to the environment.

<u>4.3</u>

From the tables 4.1,4.2,4.3,4.4, first and the second columns have been used.

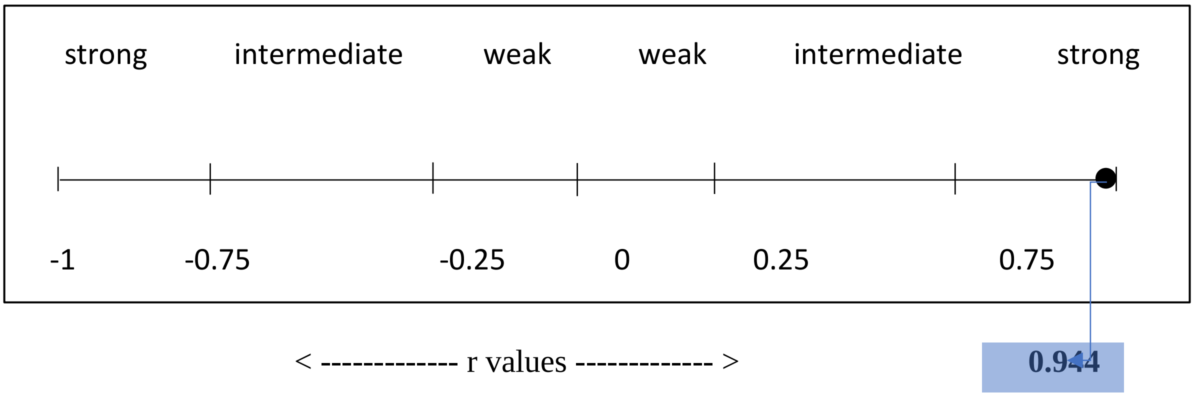
The data that we have used is quantitative and parametric.

Here, we used Pearson correlation because it only needs data instead of ranks in correlation.

Calculation:

Here, we use python code for obtaining the value of r.

```
r = 0.94400842
```



Result:

Since the value of r is close to the 1 so we can infer about the strong correlation between the waste generated and collected.

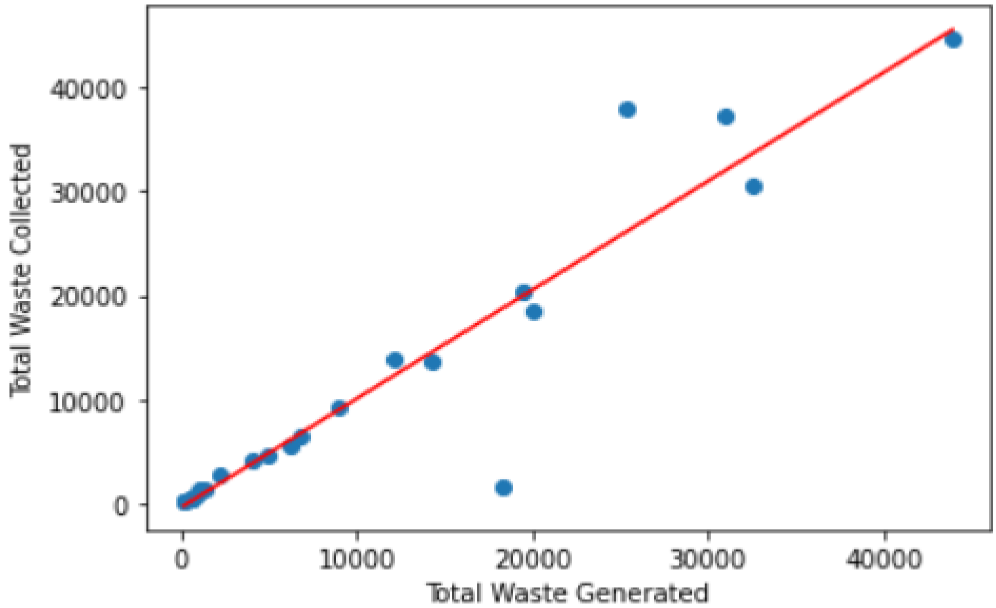


Fig 32: Graph shows the total waste generated and collected in between the year 2016-2020

- From the figure25. we can conclude that as the total waste generation increases then the waste collection also increases and there is a direct positive correlation between them.

4.4

To predict the generation of solid waste in the next five years on the basis of current trend.

Linear regression is applied to forecast the mean of total waste generated in next five years (2016-2020) based on the current trends.

For the same procedure, 1st the columns of the table 4.1 ,4.2, 4.3 & 4.4 have been used.

For the prediction for the next five years, we ought to apply the regression line expression which includes ‘b’ value. This ‘b’ value is calculated using the terms which are calculated below using the python code whose output has been shown below

	Year	Y	X	YX	X^2
0	2015.0	64183.67000	1.0	64183.670	1
1	2016.0	89693.37000	2.0	179386.740	4
2	2017.0	90209.67000	3.0	270629.010	9
3	2018.0	98919.23900	4.0	395676.956	16
Sigma	8066.0	343005.94900	10.0	909876.376	30
Mean	2016.5	85751.48725	2.5	227469.094	7.5

```
the value of b = 22386.4478
```

$y=\bar{Y}+b(x-\bar{X})$

$Y= 85751.48725+ 22386.45\ (x-2.5)$

$Y=85751.48725+22386.4478\ x - 55966.125$

$\boxed{Y=29785.36225 + 22386.4478x}$

From this equation, we can predict the values of waste generation in year (20-2020-21 to 2025-2026).

The python output for the predicted values of waste generated are:

```
        prediction values of waste generation
  2020-2021          168717.60675000004,
  2021-2022          191504.05455000003,
  2022-2023          214290.50235000005,
  2024-2025          237076.95015000005,
  2025-2026          259863.39795000007
```

Current and predicted statistics: -

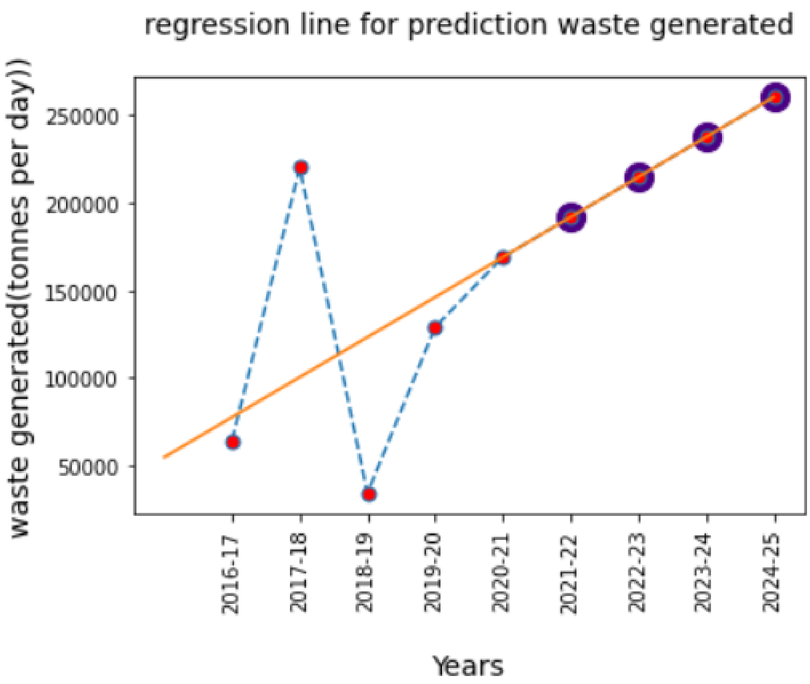


Fig 33: Graph shows the total waste generated and collected in between the year 2016-2020.

1. As per the above data, we can predict the values of waste generation that how year by year it shall increase.
2. Although, it got boosted in the year (2017-2019) but we further notice that in future, waste will increase gradually.

7. CONCLUSION

- After conducting various tests and studying the data we highlight the very objective and the title of this report which is Sustainable Living. From the first 3 objective we understand how the consumption of resources is increasing and is also estimated to increase in upcoming years. it also states how did over production of some resources like food grains in India we also see how the population of India affect the usage of this resources which is increasing with the increase in population. In some states, the use of these resources is very much as mentioned in the conclusions.

- But, in the 4th objective we see how in the past few years the waste generation has been increasing. Although there is a strong correlation between waste collected and generated, but then too the waste treated is much less than the waste generated.

- If we continue to follow the same trend, then the waste generation is also expected to increase.

- So, with pure evidence, we can say that in order to have a sustainable living, and ensuring the resources availability for the future generations, we need to have a proper check of the production of resources. We shall try to limit our consumption and wisely produce these resources-consuming entities.

8.REFERENCES

[1] Sahu, S. (2008). Trends and Patterns of Energy consumption in India

Rani, R., & Kumar, N. (2017). Does nuclear energy consumption improve environment? Empirical evidence from India. *Nature Environment and Pollution Technology*, *16*(4), 1075-1079.

[2]Reserve Bank of India - Handbook of Statistics on Indian States (rbi.org.in)

[3]William, M. A., Elharidi, A. M., Hanafy, A. A., Attia, A., & Elhelw, M. (2020). Energy-efficient retrofitting strategies for healthcare facilities in hot-humid climate: Parametric and economical analysis. Alexandria Engineering Journal, 59(6), 4549-4562.

[4]https://m.rbi.org.in/Scripts/PublicationsView.aspx?id=20793

[5]https://www.mohfw.gov.in/

[6]Verma, M. R., Datta, K. K., Mandal, S., & Tripathi, A. K. (2007). Diversification of food production and consumption patterns in India. Journal of Agricultural & Food Information, 8(3), 87-100.

[7]Venkatesh, P., Sangeetha, V., & Singh, P. (2016). Relationship between food production and consumption diversity in India–Empirical evidences from cross section analysis. Agricultural economics research review, 29(347-2016-17235), 139-148.

[8]https://www.indiaonlinepages.com/population/state-wise-population-of-india.html

[9]https://uidai.gov.in/images/state-wise-aadhaar-saturation.pdf

[10]https://m.rbi.org.in/Scripts/AnnualPublications.aspx?head=Handbook%20of%20Statistics%20on%20Indian%20States

[11]Chakrabarti, S., & Sarkhel, P. (2003). Economics of solid waste management: A survey of existing literature. *Economic Research Unit Statistical Institute. Kolkata* ..

[12]https://cpcb.nic.in/status-of-implementation-of-solid-waste-rules/

APPENDIX

1

1.1

Z-test code (calculations):

```
import pandas as pd

dem_data=pd.read_csv('demand.csv')

dem_data.index=range(1,38)

import statistics

import math

print ('Means and Standard Deviation using Python code\n')

print ('mean for 2020-21 data is’, dem_data['2020-21']. mean ())

print ('mean for 2016-17 data is’, dem_data['2016-17']. mean ())

print ('standard deviation for 2016-17 data is’,
statistics.stdev(dem_data['2020-21']))

print ('standard deviation for column 2 data is’,
statistics.stdev(dem_data['2016-17']))

print ('the value of z is’, dem_data['2020-21'].mean()-
dem_data['2016-
17'].mean()/math.sqrt((statistics.stdev(dem_data['2020-
21'])**2/36)+(statistics.stdev(dem_data['2016-17'])**2/36)))

print ('the value of z is',0.3921)
```

(graphs):

```
import pandas as pd

import matplotlib.pyplot as plt

import matplotlib
```

#1 in conclusions

```
df=pd.read_csv('demand.csv')

df1=pd.DataFrame({'2016-17 demand':list(df['2016-
17'])},index=list(df['state/UT']))

df1.plot.bar()
```

```
plt.xlabel("STATE/UT", size=15)

plt.ylabel("ENERGY UNITS(NET CRORE UNITS)", size=15)

plt.title('Energy units demand in the year 2016-17')
```

#2 in conclusions

```
df=pd.read_csv('demand.csv')

df1=pd.DataFrame({'2020-21 demand':list(df['2020-
21'])},index=list(df['state/UT']))

df1.plot.bar ()

plt.xlabel("STATE/UT", size=15)

plt.ylabel("ENERGY UNITS(NET CRORE UNITS)", size=15)

plt.title('Energy units demand in the year 2020-21')
```

#3 in conclusions

```
df=pd.read_csv('demand.csv')

a=df['2016-17']

b=df['2020-21']

c=df['state/UT']

df1=pd.DataFrame({'2016 demand':list(a),'2021
demand':list(b)},index=list(c))

df1.plot.bar ()

plt.xlabel("STATE/UT", size=15)

plt.ylabel("ENERGY UNITS(NET CRORE UNITS)", size=15)

plt.title("Comparision of 2016 and 2021 energy units demand(net
crore units)", size=18)
```

#4 bell curve

```
# Z Alpha value

import scipy.stats as st

print(st.norm.ppf(.95))
```

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

#x-axis ranges from -5 and 5 with .001 steps
x = np.arange(-5, 5, 0.001)
#Define multiple normal distributions
plt.plot(x, norm.pdf(x, 0, 1), color='green')

#add axes labels and a title
plt.ylabel('Density')
plt.xlabel('z value')
plt.title('Normal Distribution', fontsize=14)
plt.fill_between(x,norm.pdf(x,0,1), where=(1.645 <x) & (x < 4))
```

1.2

#Correlation values table

```
import pandas as pd
dem=pd.read_csv('demand.csv')
prod=pd.read_csv('production.csv')
x=pd.DataFrame()
x['2020-21 demand(X)']=dem['2020-21']
x['2020-21 production(Y)']=prod['2020-21']
x['X^2']=dem['2020-21']**2
x['Y^2']=prod['2020-21']**2
x['XY']=dem['2020-21']*prod['2020-21']
x.index=dem['state/UT']
s=pd.DataFrame(x.sum())
s.columns=['Sigma']
c=pd.concat([x,s.T])
```

Correlation curve

```
import numpy as np

df1=pd.read_csv('demand.csv')

df2=pd.read_csv('production.csv')

dem=df1['2020-21']

pro=df2['2020-21']

plt.title('Correlation b/w consumption and production of energy units')

plt.scatter(dem,pro)


plt.plot(np.unique(dem),

         np. poly1d(np.polyfit(dem, pro, 1))

         (np.unique(dem)), color='black')


# Labelling axes

plt.xlabel('2020-21 Demand of power units(net crore units)')

plt.ylabel('2020-21 production of power units(net crore units)')
```

1.3

Formation of regression dataset

```
import pandas as pd

t=pd.DataFrame(dem_data.iloc[36,1:])

t.columns=['Y']

t['X']=range(1,6)

t['XY']=t['Y']*t['X']

t['X^2']=t['X']**2


s=pd.DataFrame(t.sum())
```

```
s.columns=['Sigma']

m=pd.DataFrame(t.mean())
m.columns=['Mean']
t=pd.concat([t,s.T,m.T])
```

b value calculation

```
numerator=t['XY']. loc['Sigma']-
(t['X'].loc['Sigma']*t['Y'].loc['Sigma'])/5
print(numerator)
denominator=t['X^2'].loc['Sigma']-t['X'].loc['Sigma']**2/5
beta=numerator/denominator
print('the value of b =',beta)
```

regression curve:

```
import matplotlib.pyplot as plt
def regression(xi):
    yi=t['Y'].loc['Mean']+beta*(xi-t['X'].loc['Mean'])
    return yi
x=[]
y=[]
for xi in range(6,11):
    x.append(xi)
    y.append(regression(xi))
t1=pd.DataFrame(dem_data.iloc[36,1:])
t1.columns=['Y']
t1.loc['2021-22']=y[0]
t1.loc['2022-23']=y[1]
t1.loc['2023-24']=y[2]
t1.loc['2024-25']=y[3]
```

```
t1.loc['2025-26']=y[4]
plt.plot(t1.index,t1['Y'],marker='o',linestyle='--',mfc='g',ms=7)
xl=[]
yl=[]
for a in range(-1,10):
    xl.append(a)
    yl.append(regression(a+1))
plt.plot(xl,yl)
plt.xlabel('\nYears',fontsize=14)
plt.ylabel("demand of energy units(net crore units)",fontsize=14)
plt.title('regression line for prediction of demand of power
units\n',fontsize=14)
plt.xticks(rotation=90)
plt.scatter(xl[6:],yl[6:],color='red',s=200)
```

values prediction:

```
pred_sh=pd.DataFrame({'Predicted demand units':y})
pred_sh. index= ['2021-22','2022-23','2023-24','2024-25','2025-26']
```

1.4

Dataset formation

```
import pandas as pd
prod=pd.read_csv('production.csv')
prod ['sigma all years']=prod['2016-17']+prod['2017-18']+prod['2018-
19']+prod['2019-20']+prod['2020-21']
print ('T value is equal to',prod['sigma all years'].sum())
```

t, SST ,SSW ,SSB values

```
T=prod ['sigma all years'].sum()
print ('The value of CF',(T**2)/180)
```

```
CF=(T**2)/180
print (CF)
prod ['sig all year sq']=prod['sigma all years']**2
prod ['sig x sq / n']=prod['sig all year sq']/5
SSB=prod ['sig x sq / n'].sum()-CF
print ('Value of SSB is',SSB)
prod['A']=prod['2016-17']**2
prod['B']=prod['2017-18']**2
prod['C']=prod['2018-19']**2
prod['D']=prod['2019-20']**2
prod['E']=prod['2020-21']**2
SA=prod['A'].sum()
SB=prod['B'].sum()
SC=prod['C'].sum()
SD=prod['D'].sum()
SE=prod['E'].sum()
SST=SA+SB+SC+SD+SE-CF
print('Value of SST is',SA+SB+SC+SD+SE-CF)

print('value of SSW is',SST-SSB)
#sum of sigma all year sq column values
print(prod['sig all year sq'].sum())
```

Graph

```
import pandas as pd
import matplotlib.pyplot as plt
prod=pd.read_csv('production.csv')
prod['sigma all years']=prod['2016-17']+prod['2017-18']+prod['2018-
19']+prod['2019-20']+prod['2020-21']
```

```
plt.figure(figsize=(12,4))
plt.plot(prod['state/UT'],prod['sigma all years'],color='red')
plt.xticks(rotation=90)
plt.xlabel('state/UT')
plt.ylabel('total power units production(net crore units)')
```

2.1

Power method

```
import pandas as pd
matrix =pd.read_csv("rajikamatrix.csv")
mat=pd.DataFrame(matrix.iloc[:,1:])

Mat

import numpy as np


x = []
eigen = [0]


x.append(np.array([[1],[1],[1],[1],[1],[1],[1],[1],[1],[1],[1],[1],[
1],[1],[1],[1],[1],[1],[1],
[1],[1],[1],[1],[1],[1],[1],[1],[1],[1],[1],[1]]))


print('----------------------------------------------')


i=0
```

```
while len(eigen) == 1 or np.round(eigen[i],4)!=np.round(eigen[i-
1],4):

    x.append(np.matmul(mat,x[i]))

    eigen.append(max(x[i+1]))

    x[i+1] = np.round(x[i+1]/eigen[i+1],4)

    print('Iteration',i+1,':')

    print(x[i+1],'\t',eigen[i+1])

    print()

    i+=1
```

2.2

Z-test code (calculations):

```
import pandas as pd

dem_data=pd.read_csv('cda data 2020,2019.csv')

dem_data.index=range(1,38)

import statistics

import math

print ('the value of z is’, dem_data['2019'].mean()-
dem_data['2020'].mean()/math.sqrt((statistics.stdev(dem_data['2019']
)**2/31)+(statistics.stdev(dem_data['2020'])**2/31)))

print ('the value of z is',0.554)
```

mean

```
import pandas as pd

import statistics

d=pd.read_csv("cda data 2020,2019.csv")

print("mean for 2019 data is",statistics.mean(d['Beds in Government
Hospitals2019']))

import pandas as pd
```

```
import statistics
d=pd.read_csv("cda data 2020,2019.csv")
print("mean for 2020 data is",statistics.mean(d['Beds in Government Hospitals2020']))
```

Standard deviation

```
import pandas as pd
import statistics
d=pd.read_csv("cda data 2020,2019.csv")
print("standard deviation for 2019 data is",statistics.stdev(d['Beds in Government Hospitals2019']))
import pandas as pd
import statistics
d=pd.read_csv("cda data 2020,2019.csv")
print("standard deviation for 2020 data is",statistics.stdev(d['Beds in Government Hospitals2020']))
```

Graphs

```
import matplotlib
import pandas as pd
d=pd.read_csv("2019 cda data.csv")
import matplotlib.pyplot as plt
plt.bar(d["States"],d["Government Hospitals2019"])
plt.xlabel("States")
plt.ylabel("hospitals")
plt.title("no.of hospitals in 2019")
plt.xticks(rotation =90)
plt.show()

import matplotlib
import pandas as pd
d=pd.read_csv("2019 cda data.csv")
```

```
import matplotlib.pyplot as plt
plt.bar(d["States"],d["Beds in Government Hospitals2019"])
plt.xlabel("States")
plt.ylabel("hospitals")
plt.title("no.of beds in hospitals in 2019")
plt.xticks(rotation =90)
plt.show()

import matplotlib
import pandas as pd
d=pd.read_csv("cda data 2020,2019.csv")
import matplotlib.pyplot as plt
plt.bar(d["States"],d["Government Hospitals2020"])
plt.xlabel("States")
plt.ylabel("hospitals")
plt.title("no.of hospitals in 2020")
plt.xticks(rotation =90)
plt.show()

import matplotlib
import pandas as pd
d=pd.read_csv("cda data 2020,2019.csv")
import matplotlib.pyplot as plt
plt.bar(d["States"],d["Beds in Government Hospitals2020"])
plt.xlabel("States")
plt.ylabel("hospitals")
plt.title("no.of beds in hospitals in 2020")
plt.xticks(rotation =90)
plt.show()
```

bell curve

```
import scipy.stats as st
print(st.norm.ppf(.55))

import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

#x-axis ranges from -5 and 5 with .001 steps
x = np.arange(-5, 5, 0.001)
#Define multiple normal distributions
plt.plot(x, norm.pdf(x, 0, 1), color='green')

#add axes labels and a title
plt.ylabel('Density')
plt.xlabel('z value')
plt.title('Normal Distribution', fontsize=14)
plt.fill_between(x,norm.pdf(x,0,1), where=(1.645 <x) & (x < 4))
```

2.3

```
# Correlation
import numpy as np
import matplotlib.pyplot as plt
df=pd.read_csv("cda data 2020,2019.csv")
x = df['Beds in Government Hospitals2020']
y = df['population in 2020']

print(np.corrcoef(x,y))
```

```
plt.scatter(x,y)

plt.xlabel(' Beds in Government Hospitals2020')

plt.ylabel('population in 2020')

plt.plot(np.unique(x),
np.poly1d(np.polyfit(x,y,1))(np.unique(x)),color = 'r')

plt.show()
```

3.1

Z-test code (calculations):

#To create table of total production of 2018-19 and 2019-20

```
import pandas as pd

d=pd.read_csv('raw table.csv')

a=d['Production of Rice(2018-19)']+d['Production of Pulses(2018-
19)']

b=d['Production of Rice(2019-20)']+d['Production of Pulses(2019-
20)']

dnew=pd.DataFrame()

dnew['States/UTs']=d['States/UTs']

dnew['Total production of rice and pulses(2018-19)']=a

dnew['Total production of rice and pulses(2019-20)']=b

dnew.index=range(1,32)

Dnew

import statistics

print("Mean of Total production of rice and pulses(2019-
20):",dnew["Total production of rice and pulses(2019-20)"].mean())

print("Mean of Total production of rice and pulses(2018-
19):",dnew["Total production of rice and pulses(2018-19)"].mean())

print("Standard deviation of Total production of rice and
pulses(2019-20):",statistics.stdev(dnew["Total production of rice
and pulses(2019-20)"]))

print("Standard deviation of Total production of rice and
pulses(2018-19):",statistics.stdev(dnew["Total production of rice
and pulses(2018-19)"]))
```

#Graph

#Conclusion 1

```
import matplotlib.pyplot as plt

df=pd.read_csv('raw table2.csv')

df1=pd.DataFrame({'Total production of rice and pulses(2018-
19)':list(df['Total production of rice and pulses(2018-
19)'])},index=list(df['States/UTs']))

df1.plot.bar(figsize=(10,5),color="orange")

plt.xlabel('States/UTs')

plt.ylabel('Production in thousand tonnes')

plt.title('Total production of rice and pulses(2018-19)')
```

#Conclusion 2

```
import matplotlib.pyplot as plt

df=pd.read_csv('raw table2.csv')

df1=pd.DataFrame({'Total production of rice and pulses(2019-
20)':list(df['Total production of rice and pulses(2019-
20)'])},index=list(df['States/UTs']))

df1.plot.bar(figsize=(10,5),color="orange")

plt.xlabel('States/UTs')

plt.ylabel('Production in thousand tonnes')

plt.title('Total production of rice and pulses(2019-20)')
```

#Conclusion 3

```
df=pd.read_csv('raw table2.csv')

a=df['Total production of rice and pulses(2018-19)']

b=df['Total production of rice and pulses(2019-20)']

c=df['States/UTs']

df1=pd.DataFrame({'Total production of rice and pulses(2018-
19)':list(a),'Total production of rice and pulses(2019-
20)':list(b)},index=list(c))

df1.plot.bar(figsize=(10,5))
```

```
plt.xlabel('States/UTs')
plt.ylabel('Production in thousand tonnes')
plt.title('Comparison of production of 2018-19 and 2019-20')
```

#Bell curve

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm
#x-axis ranges from -5 and 5 with .001 steps
x = np.arange(-5, 5, 0.001)
#define multiple normal distributions
plt.plot(x, norm.pdf(x, 0, 1), color='purple')
#add legend to plot
#add axes labels and a title
plt.ylabel('Density')
plt.xlabel('z value')
plt.title('Normal Distribution', fontsize=14)
plt.fill_between(x,norm.pdf(x,0,1), where=(1.645 <x) & (x < 4))
```

3.2

#Correlation values table

```
import pandas as pd
popu=pd.read_csv('raw table.csv')
prod=pd.read_csv('raw table2.csv')
x=pd.DataFrame()
x['Population(2019-20)(X)']=popu['Population (2019-20)']
x['Total production of rice and pulses(2019-20)(Y)']=prod['Total production of rice and pulses(2019-20)']
x['X^2']=popu['Population (2019-20)']**2
x['Y^2']=prod['Total production of rice and pulses(2019-20)']**2
x['XY']=popu['Population (2019-20)']*prod['Total production of rice and pulses(2019-20)']
x.index=popu['States/UTs']
```

```
s=pd.DataFrame(x.sum())
s.columns=['Sigma']
c=pd.concat([x,s.T])
c.loc['Sigma']
```

#Correlation curve

```
import numpy as np
df1=pd.read_csv('raw table.csv')
df2=pd.read_csv('raw table2.csv')
b=df1['Population (2019-20)']
a=df2['Total production of rice and pulses(2019-20)']
plt.figure(figsize=(10,6))
plt.title('Correlation b/w population and production in 2019-20')
plt.scatter(b,a)
plt.plot(np.unique(b),np.poly1d(np.polyfit(b,a, 1))
(np.unique(b)), color='purple')
# Labelling axes
plt.xlabel('Population (2019-20)')
plt.ylabel('Total production of rice and pulses(2019-20)')
```

3.3

#import the csv file

```
import pandas as pd
matrix = pd.read_csv("raw table4.csv")
mat=pd.DataFrame(matrix.iloc[:,1:])
mat = matrix.iloc[:,1:].to_numpy()
Mat
```

#Find the eigen value and eigen vector

```
import numpy as np
x = []
eigen = [0]
x.append(np.array([[1],[1],[1],[1],[1],[1],[1],[1],[1],[1],[1],[1],[
1],[1],[1],[1],[1],[1],[1],
[1],[1],[1],[1],[1],[1],[1],[1],[1],[1],[1],[1]]))
```

```
print('-----------------------------------------------')
i=0
while len(eigen) == 1 or np.round(eigen[i],4)!=np.round(eigen[i-1],4):
x.append(np.matmul(mat,x[i]))
eigen.append(max(x[i+1]))
x[i+1] = np.round(x[i+1]/eigen[i+1],4)
print('Iteration',i+1,':')
print(x[i+1],'\t',eigen[i+1])
print()
i+=1
```

#Graph

```
import pandas as pd
d=pd.read_csv('raw table.csv')
a=d['Production of Rice(2018-19)']+d['Production of Pulses(2018-19)']+d['Production of Rice(2019-20)']+d['Production of Pulses(2019-20)']
dnew=pd.DataFrame()
dnew['States/UTs']=d['States/UTs']
dnew['Total production of rice and pulses']=a
dnew.index=range(1,32)
Dnew
import matplotlib.pyplot as plt
dnew=pd.DataFrame({'Total production of rice and pulses':list(dnew['Total production of rice and pulses'])},index=list(dnew['States/UTs']))
dnew.plot.bar(figsize=(10,5),color="green")
plt.xlabel('States/UTs')
plt.ylabel('Production in thousand tonnes')
plt.title('Total production of rice and pulses')
```

3.4

#To make the table of population and production of 2019-20

```
import pandas as pd
```

```
d=pd.read_csv('raw table.csv')
dnew=pd.DataFrame()
dnew['States/UTs']=d['States/UTs']
dnew['Population(2019-20)']=d['Population (2019-20)']
dnew['Total production of rice and pulses(2019-20)']=b
dnew.index=range(1,32)
dnew
#To make the table of consumption and production of 2019-20
import pandas as pd
d=pd.read_csv('raw table3.csv')
a=(75.284535 *d['Population(2019-20)'])/1000000
b=d["Total production of rice and pulses(2019-20)"]
dnew=pd.DataFrame()
dnew['States/UTs']=d['States/UTs']
dnew['Total consumption of rice and pulses(2019-20)']=a
dnew['Total production of rice and pulses(2019-20)']=b
dnew.index=range(1,32)
dnew
```

#Comparison graph of consumption and production of 2019-20

```
import matplotlib.pyplot as plt
plt.figure(figsize=(10,4))
plt.plot(dnew['States/UTs'],dnew['Total consumption of rice and
pulses(2019-20)'],color='green')
plt.plot(dnew['States/UTs'],dnew['Total production of rice and
pulses(2019-20)'],color='red')
plt.legend(['Consumption','Production'])
plt.xticks(rotation=90)
plt.xlabel('state/UT')
plt.ylabel('food grains(thousand tonnes)')
plt
```

4.1.1 To import our csv file of rank:-

```
import pandas as pd
matrix = pd.read_csv("rank.csv")
mat=pd.DataFrame(matrix.iloc[:,1:])
#mat = matrix.iloc[:,1:].to_numpy()
mat
```

4.1.2 To find the eigen values and eigen vectors to determine the rank.

```
import numpy as np
x = []
eigen = [0]
x.append(np.array([[1],[1],[1],[1],[1],[1],[1],[1],[1],[1],[1],[1],[
1],[1],[1],[1],[1],[1],[1],
[1],[1],[1],[1],[1],[1]]))
print('-----------------------------------------------')
i=0
while len(eigen) == 1 or np.round(eigen[i],4)!=np.round(eigen[i-
1],4):
    x.append(np.matmul(mat,x[i]))
    eigen.append(max(x[i+1]))
    x[i+1] = np.round(x[i+1]/eigen[i+1],4)
    print('Iteration',i+1,':')
    print(x[i+1],'\t',eigen[i+1])
    print()
    i+=1
```

4.2

```
import statistics
from statistics import mean

list1 = [4124.259, 3230.565, 1675.745, 5161.462]
```

```
mean1 = mean(list1)

std1 = round(statistics.pstdev(list1),2)


print("Mean of the waste collected for the year 2016-2020 - ",mean1)

print("Standard Deviation of the waste collected for the year 2016-
2020 - ",std1)
```

mean and standard deviation of total waste treated in year (2016-2020)

```
import statistics

from statistics import mean


list2 = [1727.665, 880.537, 1119.221, 1635.387]


mean2 = mean(list2)

std2 = round(statistics.pstdev(list2),2)


print("Mean of the waste treated for the year 2016-2020 - ",mean2)

print("Standard Deviation of the waste treated for the year 2016-
2020 - ",std2)
```

To find the value of Sp

```
import math

Sp = round(math.sqrt((pow(std1,2) + pow(std2,2))/2),2)

Sp


To find the value of t

t = round((mean1-mean2)/(Sp*math.sqrt(0.04 + 0.04)),2)

t
```

To reject the Null hypothesis.

```
a = 1.677
print(ta)

if (t>ta):
    print("Reject Null Hypothesis")
else:
    print("Null Hypothesis Accepted")
```

4.2.3 Graph for the T- test.

```
import pandas as pd
from matplotlib import pyplot as plt

df = pd.DataFrame("data2.csv")

plt.plot(df.index,df["Waste collected"],label='waste collected')

plt.plot(df.index,df["Waste treated"],label='waste treated')
plt.legend()

plt.xlabel("States")
plt.ylabel("Tonnes per day",)
plt.xticks(rotation=90)
plt.title("Statewise compraison of the Total waste collected and tre
ated in between the year 2016-2020")
```

4.3

4.3.1 To find the correlation between waste generated and treated in year(2016-2020)

```
# Correlation
import numpy as np
```

```
import matplotlib.pyplot as plt

x = df['Total Waste Generated']
y = df['Total Waste Collected']

print(np.corrcoef(x,y))
```

4.3.2 To draw the graph of correlation

```
plt.scatter(x,y)
plt.xlabel('Physical Working Capital')
plt.ylabel('Productive Capital')
plt.plot(np.unique(x),
np.poly1d(np.polyfit(x,y,1))(np.unique(x)),color = 'r')
plt.show()
```

4.3

4.3.1 To find the sigma and mean of the data.

```
import pandas as pd
rdata=pd.read_csv('Regression.csv')
rdata['X']=range(1,5)
rdata['XY']=rdata['Y']*rdata['X']
rdata['X^2']=rdata['X']**2
s=pd.DataFrame(rdata.sum())
s.columns=['Sigma']

m=pd.DataFrame(rdata.mean())
```

```
m.columns=['Mean']
t=pd.concat([rdata,s.T,m.T])
t
```

4.3.2 To find the value of b

```
numerator=t['XY'].loc['Sigma']-
(t['X'].loc['Sigma']*t['Y'].loc['Sigma'])/5
print(numerator)
denominator=t['X^2'].loc['Sigma']-t['X'].loc['Sigma']**2/5
beta=numerator/denominator
print('the value of b =',beta)
```

4.3.3 plot the graph

```
import matplotlib.pyplot as plt
def regression(xi):
    yi=t['Y'].loc['Mean']+beta*(xi-t['X'].loc['Mean'])
    return yi
x=[]
y=[]
for xi in range(5,10):
    x.append(xi)
    y.append(regression(xi))

t1=pd.DataFrame(w_gen.iloc[3,:])
t1.columns=['Y']
t1.loc['2020-21']=y[0]
t1.loc['2021-22']=y[1]
t1.loc['2022-23']=y[2]
t1.loc['2023-24']=y[3]
```

```
t1.loc['2024-25']=y[4]
plt.plot(t1.index,t1['Y'],marker='o',linestyle='--',mfc='r',ms=7)

xl=[]
yl=[]
for a in range(-1,9):
    xl.append(a)
    yl.append(regression(a+1))
plt.plot(xl,yl)

plt.xlabel('\nYears',fontsize=14)
plt.ylabel("demand of energy units(net crore units)",fontsize=14)
plt.title('regression line for prediction of demand of power
units\n',fontsize=14)
plt.xticks(rotation=90)
plt.scatter(xl[6:],yl[6:],color='indigo',s=200)
```

 # Graph for waste generation in years(2016-2020)

```
 import pandas as pd
from matplotlib import pyplot as plt

df = pd.DataFrame("data1.csv")
ax = df.plot.bar(x='States', y='Waste generated in 2016-
2020', rot=90)
plt.ylabel("Tonnes per day")
plt.title("Bar graph showing Total waste generated in 2016-2020")
```

 # Graph for waste collected in years(2016-2020)

```
import pandas as pd
```

```
from matplotlib import pyplot as plt


df = pd.DataFrame(data3.csv)

ax = df.plot.bar(x='States', y='Waste collected in 2016-
2020', rot=90, color= 'maroon')

plt.ylabel("Tonnes per day")

plt.title("Bar graph showing Total waste collected in 2016-2020")
```

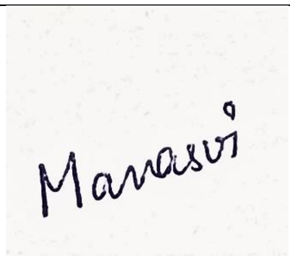
 # Graph for waste treated in years(2016-2020)

```
import pandas as pd

from matplotlib import pyplot as plt


df = pd.DataFrame("data4.csv")

ax = df.plot.bar(x='States', y='Waste treated in 2016-
2020', rot=90, color= 'indigo')

plt.ylabel("Tonnes per day")

plt.title("Bar graph showing Total waste treated in 2016-2020")
```

Computational Data Analysis –Project
Individual Contribution Report

Team No. 15 Section: B Project Title: Sustainable Living

SDG Goal No. Planned:12

Name	Objective(s)	Summary on Contribution (Brief method, result summary and conclusion)	Signature
Manasvi Jain 2021BTECH070	**To Analyse the State-Wise Energy Consumption and Production in different states of India.**	Objective 1, abstract, methodology, Compilation of report, formatting.	
Ritika Kuwera 2021BTECH097	**To Analyse the Generation, Collection and hence the treatment of solid waste produced in different states.**	Objective 4, Introduction.	
Rajika Patel 2021BTECH094	**To Analyse the Consumption and Production of Medical Facilities in different states of India**	Objective 2, formatting.	
Prapti Bhatt 2021BTECH088	**To Analyse the Production and Consumption of Food Grains in different states of India.**	Objective 3, Figures of methodology, Indexing.