Analysis of Indian Census-2011

Chethan M
Computer science
PES UNIVERSITY
Bangalore,India
chethanmiit@gmail.com

Anirudh R
Computer science
PES UNIVERSITY
Bangalore,India
anirudhr@pesu.pes.edu

Sree Kaustub
Computer science
PES UNIVERSITY
Bangalore,India
sreekaustub2000@gmail.com

G Sai Charan Reddy
Computer science
PES UNIVERSITY
Bangalore,India
charangreddy@gmail.com

***Abstract*—This project is to assess the literacy rate of the Indian population by finding out the various factors that affect education and how it affects the development of the country based on the Indian population census-2011 data released by the Indian government.**

I. Introduction

The Census Data provides time series data on the economic and social changes in the country, considered as the best source of information on demography, census emphasizes only on population numbers, not on individual/household consumption.

Education Plays a major role in the development of a country. It has been observed that female labor force participation and maternal education is maximized by education. Infant mortality is negatively related to both education and labor force. Excessive female feticide in some parts is also inversely related to female education and labor work force. Unemployment on the other hand is strongly correlated to literacy rate. The percentage of unemployment is used to assess the development of a country, more the unemployment, the country is considered is considered to be less developed. Education is important not only for the overall economic growth of the country but also making individuals self-sufficient. It also helps in minimizing the social taboos and making the society a better place to live in.

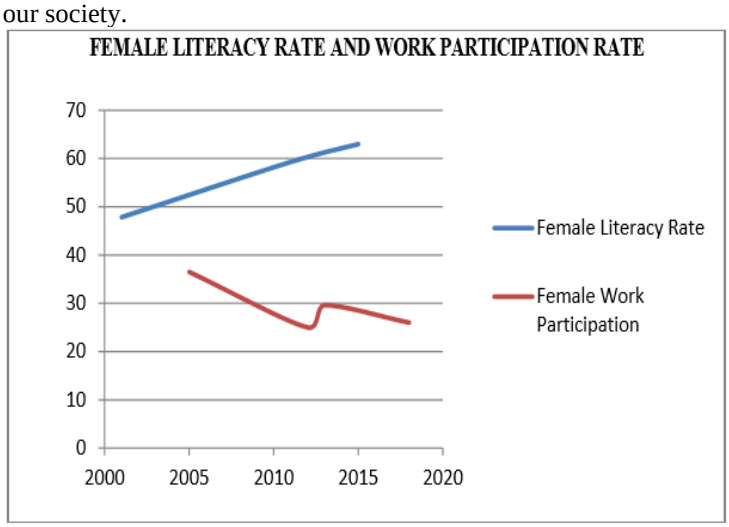
II. Inferences of Literature Survey

A. Analysing women work participation and literacy

The main aim of this study is to depict the picture of women empowerment and work participation in India, shows the prevailing status of women in our country. The increasing graph of literacy rate shows us a society that is gradually eliminating the social taboos, but the decreasing graph of work participation in last few years suggest that the insecurities are still not completely gone.

The author has plotted various line graphs to verify the increasing trend of women literacy. All the graphs show an increasing trend in the literacy rate. While it's not the same with the work participation, almost all the states and union territories have decreasing trend except for states like Haryana and Tamil Nadu, Pondicherry in the list of Union territories.

As it has been mentioned, an increasing literacy rate should result in women empowerment and work participation. But an exception to this indicates the insecurities still prevail in our society.



Thus it can be said that literacy work participation and education rate in India doesn't have a positive correlation.

This research paper was extremely helpful for a better understanding of our datasets which will help us to build appropriate regression models and for an efficient prediction.

B. Trends and Structure Of Census In India

Introduction:

India has a rich history of data and data collection. Taking into account, the literatures of RIG VEDA to the modern analytics of population, the most crucial part is played by census.

Census in ancient times were mostly collected for military and political purposes. Census in the modern era are mostly collected to know about the demographics, mortality, education, government records etc.

Census – Latin word 'Censere' meaning 'to asses or to rate'.

Methods of census:

1. De-Facto method

De-facto – Latin word – 'Concerning fact'

A date will be fixed for the census of the whole country (census night).

There is something called as reference point in which census are taken. So populations means all persons available in the country at that point.

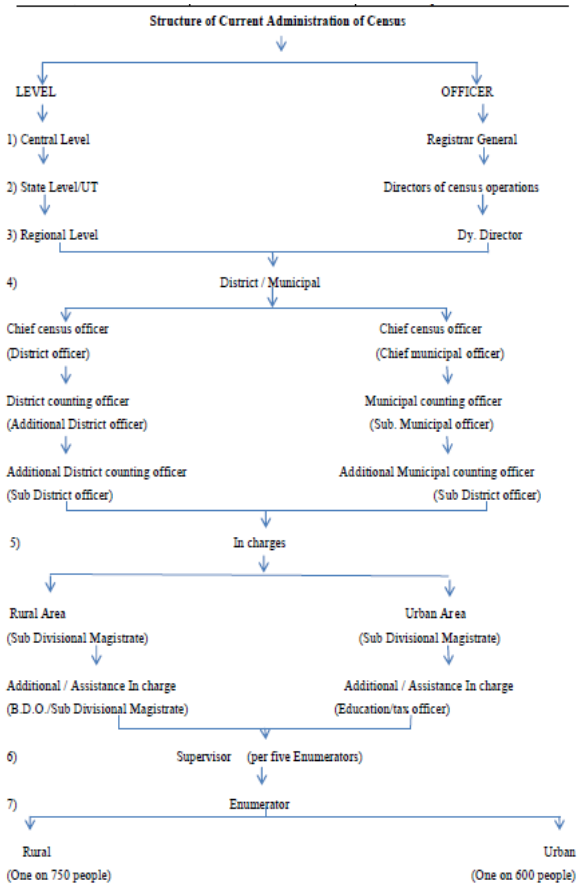
People are cautioned not to leave their homes during the census night.

2. De-jure method

De-jure in latin mreans 'Concerning facts'.

The population that lives in a country at a particular reference time-point will be considered as the population of the country.

The census work is carried on for more than 2 weeks. This period in which the census is carried out is called as the enumeration period.



Some Important result of Census of India (2001-2011)

Year / Unit	2001 census	2011 census	Decade change
Total population	1028737436	1210569573	17.7% (growth)
Urban population (in %)	285354954	377106125	31.8% (growth)
Rural population (in %)	741660293	833463448	12.3% (growth)
District	593	640	+47
Sub-District	5463	5924	+461
Village	938588	640930	+297658
Sex ratio (per1000)	933	943	+ 10
Urban sex ratio	900	929	+ 29
Rural sex ratio	946	949	+ 3
Literacy ratio (in %)	64.84	74.04	+ 9.2
Urban literacy ratio	79.9	85.0	+ 5.1
Rural literacy ratio (in %)	58.7	68.9	+ 10.2
Density (person/km)	325	382	+ 57
Life Expectancy (in year)	65.3	64.5	- 0.8

Drawbacks of Indian census:

Firstly Enumerators efficiency is the key tool. It has a large take on purity of the census since they form the primary basis for any kind of census.

Second and last one is that the enumerators are lazy. They actually don't do the field work rather they fill values themselves without even conducting the census nor asking the respondents.

RELEVANCY

In this dataset, the author has built a clear background picture about census from its history to current scenario and also how census have been taken in India. This helped us in knowing the authenticity of the dataset (India Census 2011).

The author further explained the types of census taking methods that defined the how efficient the dataset is with respect to the methods.

C. Birth Rate Trends In India:1985-2007

Introduction:

Reduction within the birth rate to curb the growth in population is the major concern in India. This concern was a major reason for adopting various birth control policies and led to the launch of first birth limitation programme in 1950.We cannot exactly say till what extent this programme has succeeded in reducing the birth rate.

The year 1985 imprints the start of a continuous however critical move in India's legitimate methodology towards conception prevention. This move is portrayed by expanding ascendency of wellbeing concerns over segment goals in arranging also, execution of legitimate birth decrease endeavours.

Methodology:

The strategy utilized in this examination joins fertility at the degree of populace fertility to singular fertility.

Population fertility is estimated by the birth rate which in blend with the passing rate decides the pace of regular increment.

The connection between birth rate and individual fertility isn't immediate on the grounds that population fertility is additionally impacted by the age structure of the populace

and examples of marriage while fertility of wedded ladies is autonomous of these variables.

Fertility Transition In India:

The most recent data accessible through the SRS proposes that the birth rate in India was 23 live births for each 1000 populace around the year 2006 while the individual fertility was around 4.4 live births per wedded woman. This implies that wedded couples, in India, actually have, on a normal, in excess of 4 live births during their whole conceptive period.

Levels of population fertility and individual f

India/States	Crude Birth Rate	Total Fertili Rate
Andhra Pradesh	18.9	2.0
Assam	24.6	2.8
Bihar	29.9	4.1
Gujarat	23.4	2.7
Haryana	23.9	2.7
Karnataka	20.2	2.1
Kerala	14.9	1.7
Madhya Pradesh	29.0	3.5
Maharashtra	18.5	2.1
Orissa	21.9	2.5
Punjab	17.8	2.1
Rajasthan	28.3	3.5
Tamil Nadu	16.2	1.7
Uttar Pradesh	30.0	4.1
India	23.5	2.8

Source: Sample Registration System.

Two observations can be made from the Table:

To begin with, the pace of abatement in population fertility (estimated as far as b) seems to possess eased back down over energy for India in general just as for its provincial and metropolitan zones as the decrease in population fertility (b) has been generally quick during the period 1986-91.

Second, and additionally uncovering, perception is that the lessening in the total martial fertility rate or, proportionately, unweighted normal average marital fertility rate seems to have turned around during the period 1996-2001.

From the above table we can see during the period 1986 the pace of abatement in population fertility (b) was quicker than the pace of reduction in individual fertility. Nonetheless, during the period 196 the diminishing in individual fertility was more quick than the lessening in population fertility.

Decomposition of Change in Population Fertility:

The table below disintegrates the adjustment in population fertility (b) into changes in its four segment determinants - individual fertility (a), age circulation of wedded ladies in the conceptive age gathering (d), extent of wedded ladies to all ladies in the conceptive age gathering (m), and extent of regenerative time of ladies to the total population (p). Somewhere in the range of 1986 and 2006, the population fertility in the nation diminished by around 9 births for each 1000 population. It represented around 6.7 of the abatement of 9 births for every 1000 population saw in population fertility among 1986 and 2006. Then again, the adjustment in the age circulation of wedded ladies in conceptive age gathering (d) represented 62 The Journal of Family Welfare about 2.5 of these births; the adjustment in the extent of wedded ladies to all ladies in the conceptive age gathering (m) represented about 1.6 percent of these births yet the adjustment in the extent of conceptive time of ladies in the population added to an expansion, not decline, of about 1.6 births per 1000 population.

State-wise trends in rate of decrease in the Crude Birth Rate and Total Marital Fertility Rate in Indian states: 1986-2006

Country	Fertility indicator	1986-91	1991-96	1996-2001	2001-06
Andhra Pradesh	*b*	-3.695	-1.702	-1.941	-2.229
	g	-3.050	-3.481	0.618	-1.274
Assam	*b*	-2.364	-1.037	-1.401	-1.626
	g	-1.867	-1.145	0.336	-2.683
Bihar	*b*	-3.189	-0.076	-.0468	-0.881
	g	-2.514	-1.163	1.222	-1.887
Gujarat	*b*	-2.150	-1.611	-0.825	-1.207
	g	-2.022	-1.944	0.664	-0.667
Haryana	*b*	-1.426	-2.547	-1.492	-2.072
	g	-1.353	-2.094	0.041	-0.693
Karnataka	*b*	-1.403	-3.339	-0.833	-1.649
	g	-1.608	-3.160	0.613	-1.747
Kerala	*b*	-4.270	-0.343	-0.279	-3.644
	g	-4.482	-0.195	-1.096	3.343
Madhya Pradesh	*b*	-0.777	-2.203	-0.181	-1.151
	g	-0.694	-1.684	0.916	-1.831
Maharashtra	*b*	-2.360	-2.365	-2.737	-2.030
	g	-2.519	-0.941	-1.879	-1.409
Orissa	*b*	-2.123	-1.375	-2.695	-1.338
	g	-1.891	-0.891	-1.356	-0.898
Punjab	*b*	-0.738	-3.014	-2.195	-3.409
	g	-1.276	-1.933	1.939	-4.039
Rajasthan	*b*	-0.781	-0.939	-1.107	-1.784
	g	-0.739	-1.719	-0.212	-1.507
Tamil Nadu	*b*	-3.047	-1.606	-0.131	-3.362
	g	-2.514	-2.509	1.646	-.0795
Uttar Pradesh	*b*	-0.775	-1.184	-1.288	-1.264
	g	-0.091	-1.469	0.776	-1.695
West Bengal	*b*	-2.238	-3.209	-2.120	-1.951
	g	-2.115	-3.773	0.009	-2.928

Source: Author's calculations.

Stalling of Individual Fertility Decline:

The foregoing investigation demonstrates that the diminishing in individual fertility (estimated as far as total martial fertility rate or, equally, unweighted average marital fertility rate) in India and in various states practically slowed down after the presentation of the objective free or network needs appraisal approach. The purpose behind this might be investigated by examining the adjustment in the age explicit conjugal fertility rates during the period 1996-2001. Evaluations accessible through the SRS recommends that during the period 1996-2001, fertility of wedded ladies between the age group 15-24 years in India expanded, while that of wedded ladies of age 25 years or more diminished.

It is additionally clear from underneath indicated Table that the total expansion in the fertility of wedded ladies matured 15-24 years was hardly higher than the total abatement in the fertility of wedded ladies aged 25-49 years in so individual fertility in India expanded barely during the period 1996-

2001. A comparative circumstance won in 9 of the 15 significant states remembered for this investigation.

State-wise decomposition of the change in Birth Rate: 1986-2006

State	Period	b	a	D	m	p
Andhra Pradesh	1986-1991	-5.00	-4.29	-0.77	-0.83	0.89
	1991-1996	-2.43	-4.18	0.43	0.96	0.35
	1996-2001	-2.17	0.41	-2.68	-0.94	1.05
	2001-2006	-2.10	-1.32	-0.75	-0.56	0.54
	1986-2006	-11.70	-9.38	-3.77	-1.37	2.83
Assam	1986-1991	-3.93	-3.1	-1	0.03	0.14
	1991-1996	-2.10	-1.6	-0.78	1.49	-1.21
	1996-2001	-1.53	0.31	-2.97	-0.79	1.92
	2001-2006	-2.20	-2.69	0.6	0.53	-0.63
	1986-2006	-9.76	-7.09	-4.15	1.26	0.21
Bihar	1986-1991	-5.00	-3.73	-0.39	-1.7	0.82
	1991-1996	-0.00	-1.38	0.99	0.42	-0.03
	1996-2001	-0.63	1.36	-1.51	-1.34	0.85
	2001-2006	-1.43	-2.3	2	0.71	-1.84
	1986-2006	-7.07	-6.04	1.09	-1.91	-0.21
Gujarat	1986-1991	-3.60	-3.22	-0.91	-0.44	0.97
	1991-1996	-2.40	-2.57	-0.29	0.21	0.25
	1996-2001	-1.03	0.62	-1.71	-0.64	0.7
	2001-2006	-1.57	-0.59	-1.07	0.64	-0.55
	1986-2006	-8.60	-5.76	-3.99	-0.22	1.37
Haryana	1986-1991	-2.83	-2.75	-0.75	-0.46	1.12
	1991-1996	-3.33	-2.94	-0.28	0.2	-0.31
	1996-2001	-2.23	-0.2	-2.6	-1.28	1.85
	2001-2006	-2.90	-1.32	-1.38	-0.51	0.3
	1986-2006	-11.30	-7.21	-5.01	-2.05	2.97
Karnataka	1986-1991	-2.10	-2.21	-0.7	0.64	0.16
	1991-1996	-3.80	-3.28	-1.36	0.19	0.66
	1996-2001	-1.17	0	-1.61	-0.53	0.97
	2001-2006	-1.90	-1.53	-0.37	0	-0.01
	1986-2006	-8.97	-7.02	-4.04	0.31	1.78
Kerala	1986-1991	-3.97	-3.85	-1	0.57	0.31
	1991-1996	-0.57	-0.48	-1.55	0.97	0.5
	1996-2001	-0.60	-0.96	-0.31	0.66	0.01
	2001-2006	-2.50	2.13	-4.21	0.13	-0.55
	1986-2006	-7.63	-3.16	-7.08	2.34	0.27
Madhya Pradesh	1986-1991	-1.73	-1.49	-0.4	-0.96	1.11
	1991-1996	-3.47	-2.95	0.14	-0.21	-0.45
	1996-2001	-1.53	1.01	-2.04	-1.55	1.05
	2001-2006	-1.93	-2.74	1.14	-0.4	0.07
	1986-2006	-8.67	-6.18	-1.15	-3.11	1.79
Maharashtra	1986-1991	-3.00	-3.04	0.03	-0.38	0.39
	1991-1996	-2.67	-0.98	-1.32	-0.11	-0.25
	1996-2001	-3.00	-2.04	-1.08	-0.74	0.86
	2001-2006	-2.13	-1.61	-0.48	-0.2	0.16
	1986-2006	-10.80	-7.67	-2.85	-1.44	1.16
Orissa	1986-1991	-2.53	-2.33	-0.71	-1.21	1.72
	1991-1996	-1.77	-1.35	-0.14	0.34	-0.62
	1996-2001	-3.43	-1.47	-1.85	-1.2	1.09
	2001-2006	-1.77	-1.22	-0.46	0.05	-0.14
	1986-2006	-9.50	-6.37	-3.16	-2.01	2.04
Punjab	1986-1991	-1.17	-1.1	-0.55	0.25	0.23
	1991-1996	-3.57	-2.32	-1.64	-0.21	0.61
	1996-2001	-2.70	1.44	-4.33	-0.8	1
	2001-2006	-3.37	-3.18	0.11	0.08	-0.38
	1986-2006	-10.80	-5.16	-6.41	-0.68	1.46
Rajasthan	1986-1991	-2.57	-2.35	-0.54	-1.45	1.78
	1991-1996	-1.90	-3.03	0.96	0.34	-0.17
	1996-2001	-1.57	-0.43	-1.09	-1.17	1.13
	2001-2006	-2.77	-1.67	0.38	-0.8	-0.67
	1986-2006	-8.80	-7.49	-0.3	-3.08	2.06
Tamil Nadu	1986-1991	-3.13	-2.36	-0.78	-0.35	0.36
	1991-1996	-1.43	-2.19	-0.21	0.4	0.57
	1996-2001	-0.63	1.15	-1.76	-0.29	0.27
	2001-2006	-2.80	-0.44	-2.03	-0.44	0.11
	1986-2006	-8.00	-3.84	-4.78	-0.69	1.31
Uttar Pradesh	1986-1991	-1.80	-0.58	-0.45	-1.59	0.82
	1991-1996	-1.77	-2.29	1.28	-0.07	-0.68
	1996-2001	-1.93	1.11	-1.82	-2.38	1.16
	2001-2006	-2.17	-2.47	0.73	-0.41	-0.01
	1986-2006	-7.67	-4.24	-0.26	-4.45	1.29
West Bengal	1986-1991	-3.27	-3.29	-0.71	-0.19	0.92
	1991-1996	-3.73	-4.45	-0.47	0.78	0.4
	1996-2001	-2.33	-0.36	-2.61	-0.54	1.18
	2001-2006	-2.23	-2.6	0.22	0.46	-0.31
	1986-2006	-11.57	-10.69	-3.58	0.52	2.19

Conclusions:

Population fertility (estimated regarding birth rate) is diminishing in our nation and in its constituent states, in any case, individual fertility stays high and the decrease in the individual fertility slowed down in the nation and expanded in numerous states in the five years after the presentation of the target free or community needs assessment approach.

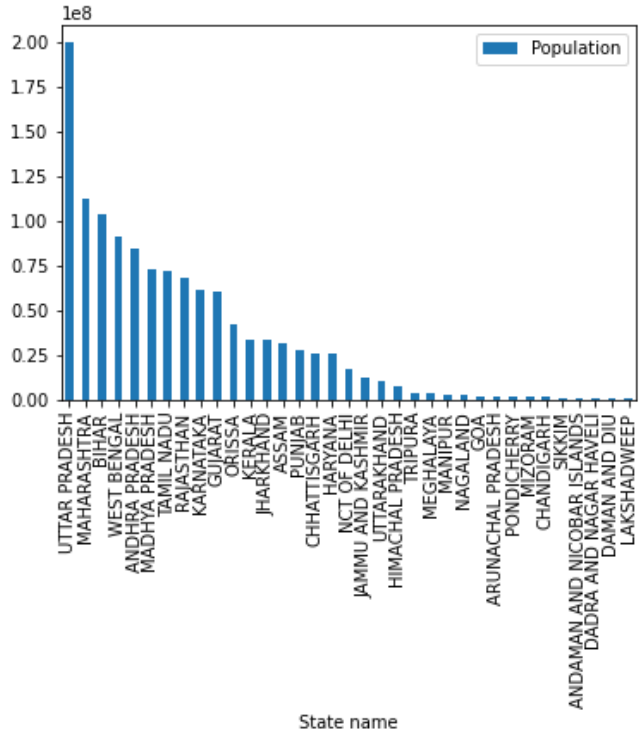
Managing fertility of wedded ladies under 25 years old seems basic to diminishing martial fertility in India. The official family management program needs to situate itself to meeting the fertility guideline needs of youthful couples.

How is this paper related to your work?

Since our dataset is related to population stats, the author in this paper examines about how the birth rate influences the population growth

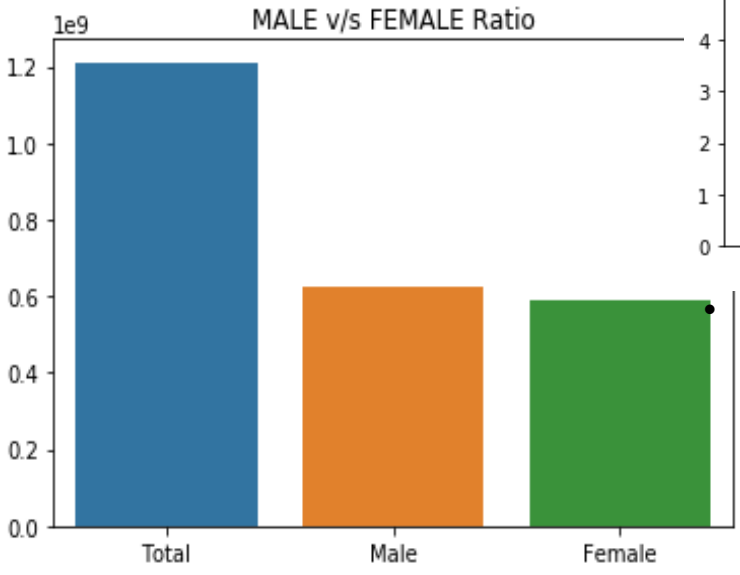
III. Exploratory Data Analysis

A. State Wise Population



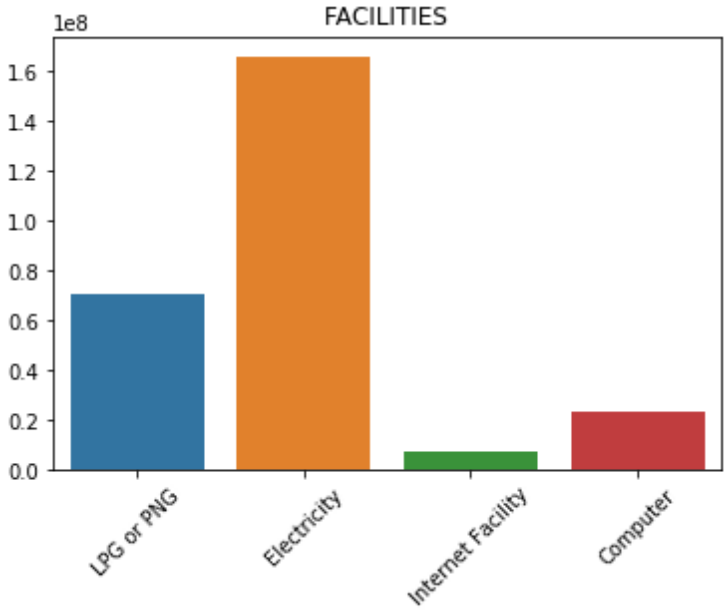
- Uttar Pradesh records the highest population. Though in terms of area ,it is the fourth biggest state. The reason for its huge population is the fertile indo-gangetic plains which is suitable for agriculture resulting in most of the people in the 50's-70's to settle here. The establishment of industries has further increased the count. Another reason is the lack of awareness of family planning, even now the fertility rate stands at 3 per family.

B. Male vs Female Ratio



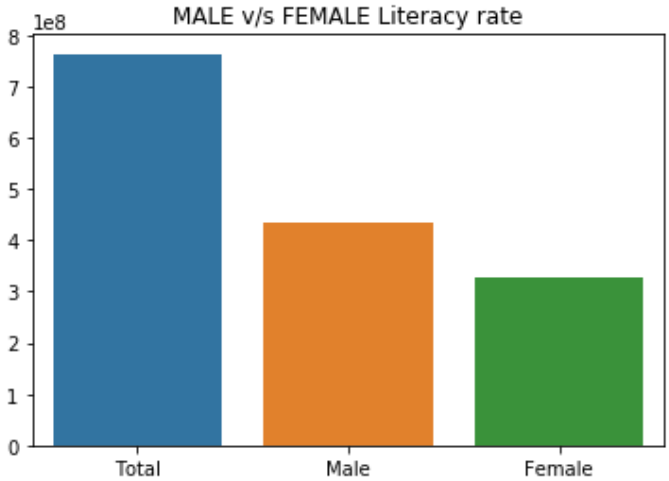
- Kerala has the highest female sex ratio and Haryana has the lowest, in contrary Kerala has the lowest male sex ratio and Haryana has the highest.

C. Facilities



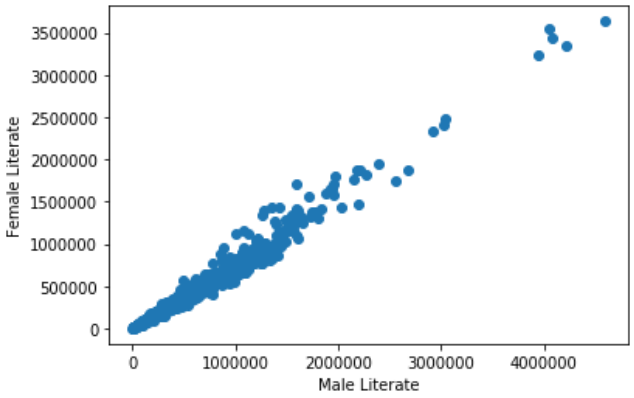
- Most of the population doesn't have access to the internet and computer which has an effect on the education and the population using LPG for cooking is also relatively less depicts that there are villages which are still backward.

D. Male vs Female Literacy Rate



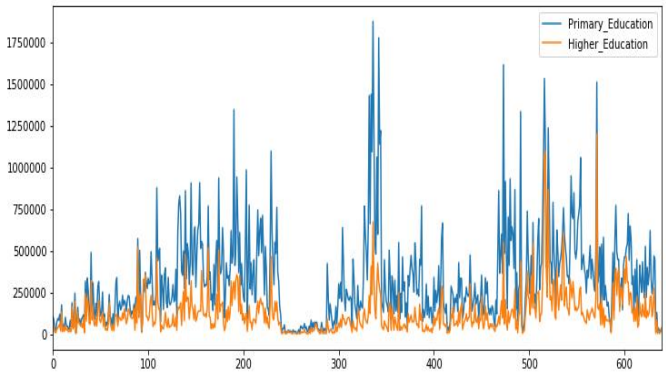
- This graph gives an insight on the male and female literacy rate. From the graph we can infer that overall male literacy ratio is greater than the female literacy rate.

E. Scatter plot between male and female literacy rate



- This graph gives an insight on the correlation of male and female literacy rate. Inference is that they are strongly correlated.

F. Comparing population completed primary education vs higher education



- This graph compares the population that has completed primary education and higher education. It is clear from the graph that very small percentage of the population continues to pursue higher education.

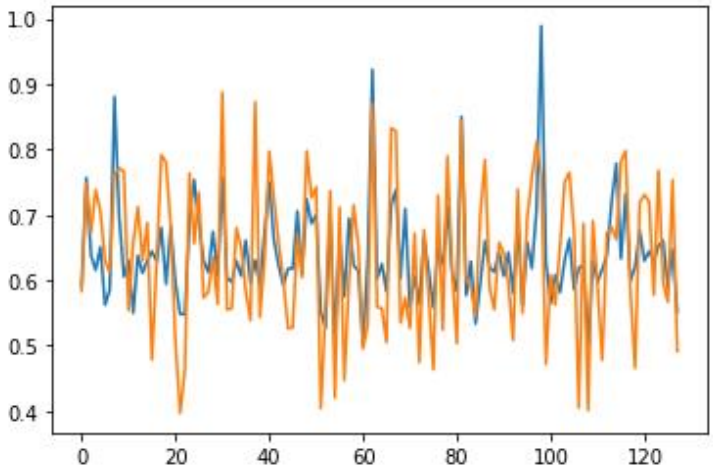
IV. PROPOSED SOLUTION USING MODEL BUILDING

A. Parameters considered for the model

- **Households with Telephone/mobile**-This variable suggests us how developed a place is. Telephone is a basic requirement, we try to correlate development and literacy.
- **Households with internet**-Internet has a vital role to play in the field of education in this digital world. People get to learn from various sources through internet, it has helped in decreasing the gap between student and teacher and also made it closer to people in remote areas.
- **Households with Computer**-Internet and computer co-exist in education, households with computers are more likely to study is our assumption.
- **Age group(0-29)**-97% of people in this age group are the ones who opt for education, age of the people may have an effect on the literacy rate.
- **Households with Car/jeep**-This parameter implies that the places with more car/jeep have motor-able roads, which means there is no problem of transportation and also that people with car can afford education is another assumption.
- **Target Variable :** Literacy rate

B. Model Building using Multiple Linear Regression

Multiple linear regression is a linear approach to modeling the relationship between a scalar response and one or more explanatory variables.



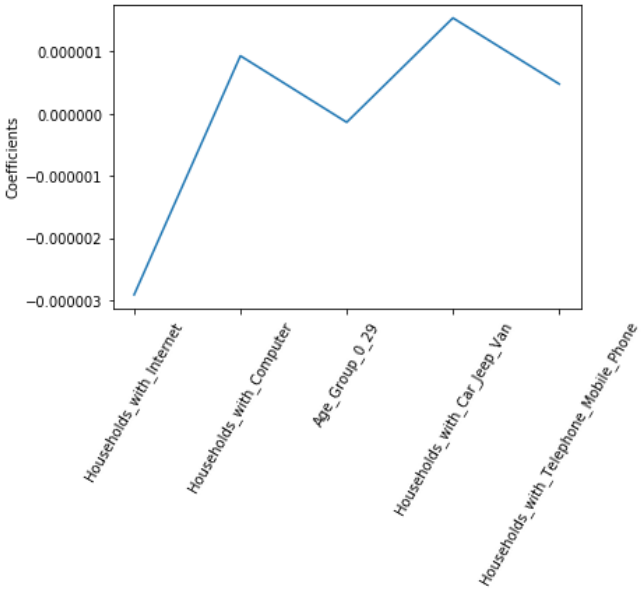
MODEL EVALUATION

Mean Squared Error=0.0074
Root Mean Squared Error=0.086
Mean Absolute Error=0.07177.

The ideal MSE,RMSE and MAE is 0. The model obtained above has been the best with the error values mentioned above.

C. Model Building using Lasso Regression

The above MLR model is carried further to find the priority (importance) of each independent variable used in the model using the widely used concept of Lasso regression.



The peak suggests that Households_with_Car_jeep_Van has the highest priority among the variables. This implies that people who are mostly rich have the highest access to education and followed by people having computers.
One more important factor is the age group i.e 0-29 years. Districts having high youth population tend to have high literacy rates. This is because they are more exposed to schools and colleges making the literacy rate of those districts marginally higher.

V. Conclusion And inferences

1. Literacy rate is largely affected by the sex ratio. Districts having high male population tend to have high literacy rates.

2. Districts having good infrastructure and very less people in BPL (Below Poverty Line) have high litercy rates. This means that they have marginally high access to education when compared to other districts.

3. States with less population density and more work population have high literacy rates.

4. The most important thing affecting the literacy rates in India is education non-continuity i.e most people in highly-dense Indian states like Uttar Pradesh stop their education after primary classes and doesn't take it further hence affecting the Literacy rates of those states.

5. Women given less access to graduate education in remote districts of India. This increses the "non-literate' tag among these women and hence there is downfall of Literacy rate.