

# *Relation of Literacy Rate of a State and Wealth of People*

Sachin Kohli  
CS20123

## Table of Contents

1. Abstract
2. Introduction
3. Methodology
4. Data Cleaning
5. Imported Libraries
6. Dataset Used
7. Data Summary
8. Calculating Different Literacy Rates
9. Plotting Literacy Rate of Different States in Order
10. Calculating Data for Wealth Metric
11. Dimensionality Reduction using PCA
12. Result
  - Correlation
  - Wealth vs Literacy Rate
  - Wealth vs Male Literacy Rate
  - Wealth vs Female Literacy Rate
13. Conclusion
14. Dataset Reference

## Abstract

In this document we will try to find a relationship between Literacy Rate of different States with the wealth of it's people using the Census of 2011. Literacy Rate is a very straight-forward metric to calculate but same cannot be said for the metric of Wealth.

In this document to calculate the Wealth of people we use attributes such as commodities owned by a particular household. However such metric doesn't give us the exact wealth of the household but gives us a rough estimate about the wealth of the household which is good enough to find relation between the two.

## Introduction

- Census is the process of collection and compilation of data of population of a country.
- It is a reflection of truth and facts as they exist in a country about its people.
- It is conducted once in every ten years, the last census was conducted in the year 2011.
- Its comprises of population count of different sections of society with their geographic location, details of different commodity owned by households, etc. But for the project we are only concerned with the two above mentioned usecases.

- Here we are going to see whether literacy rate relates to Wealth or not by graphically plotting their respective metrics with each other.

## Methodology

- **Data Source Used**
  - The Census Data of 2011 is Used
- **For Calculation of Literacy Rate**
  - First we calculate the Total population then,
  - calculate Literate population the finally,
  - Literacy Rate = Literate Population / Total Population
- **For Calculation of Wealth Metric**
  - Calculate the households with Electricity, Television, Computer, Two-Wheeler and Four-Wheeler,
  - Divide the above by Total households
  - Now we get five columns of values, we reduce this to two column values by PCA
  - Atlast we plot graph between the most significant principle component, which is Wealth with Literacy Rate

## Data Cleaning

- Data was manually cleaned through MS Excel.
- Certain names were change according to preference.

## Imported Libraries

```
In [288... import pandas as pd
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
import numpy as np
```

## Dataset Used

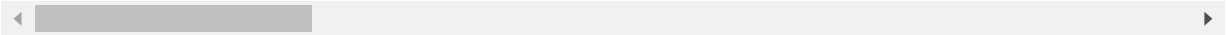
```
In [289... df = pd.read_csv('census_2011.csv')
df.head()
```

```
Out[289... 
```

	District code	State name	District name	Population	Male	Female	Literate	Male_Literate	Female_Liter
0	1	JAMMU AND KASHMIR	Kupwara	870354	474190	396164	439654	282823	1568
1	2	JAMMU AND KASHMIR	Badgam	753745	398041	355704	335649	207741	1279
2	3	JAMMU AND KASHMIR	Leh(Ladakh)	133487	78971	54516	93770	62834	309

	District code	State name	District name	Population	Male	Female	Literate	Male_Literate	Female_Literate
3	4	JAMMU AND KASHMIR	Kargil	140802	77785	63017	86236	56301	29935
4	5	JAMMU AND KASHMIR	Punch	476835	251899	224936	261724	163333	98391

5 rows × 118 columns



## Data Summary

In [290...

df.describe()

Out[290...

	District code	Population	Male	Female	Literate	Male_Literate	Female_Literate
count	640.000000	6.400000e+02	6.400000e+02	6.400000e+02	6.400000e+02	6.400000e+02	6.4000
mean	320.500000	1.891961e+06	9.738598e+05	9.181011e+05	1.193186e+06	6.793182e+05	5.1386
std	184.896367	1.544380e+06	8.007785e+05	7.449864e+05	1.068583e+06	5.924144e+05	4.8018
min	1.000000	8.004000e+03	4.414000e+03	3.590000e+03	4.436000e+03	2.614000e+03	1.8220
25%	160.750000	8.178610e+05	4.171682e+05	4.017458e+05	4.825982e+05	2.764365e+05	2.0089
50%	320.500000	1.557367e+06	7.986815e+05	7.589200e+05	9.573465e+05	5.483525e+05	4.0385
75%	480.250000	2.583551e+06	1.338604e+06	1.264277e+06	1.602260e+06	9.188582e+05	6.6415
max	640.000000	1.106015e+07	5.865078e+06	5.195070e+06	8.227161e+06	4.591396e+06	3.6357

8 rows × 116 columns



## Calculating Different Literacy Rates

In [291...

```
states = pd.Series(df['State name']).unique().tolist()
pop = [0 for i in states]
f_pop = [0 for i in states]
m_pop = [0 for i in states]
lit = [0 for i in states]
f_lit = [0 for i in states]
m_lit = [0 for i in states]
f_rate = [0 for i in states]
m_rate = [0 for i in states]
rate = [0 for i in states]

for index,row in df.iterrows():
    sn = row['State name']
    i = states.index(sn)
    pop[i] += row['Population']
    f_pop[i] += row['Female']
    m_pop[i] += row['Male']
    lit[i] += row['Literate']
    f_lit[i] += row['Female_Literate']
    m_lit[i] += row['Male_Literate']
```

```

for i in range(35):
    f_rate[i] = f_lit[i]/f_pop[i]
    m_rate[i] = m_lit[i]/m_pop[i]
    rate[i] = lit[i]/pop[i]

dic = {'State':states, 'Literacy Rate':rate, 'Female Literacy Rate':f_rate, 'Male Li
new_df = pd.DataFrame(dic)

new_df

```

Out[291]...

	State	Literacy Rate	Female Literacy Rate	Male Literacy Rate
0	JAMMU AND KASHMIR	0.563517	0.474959	0.642206
1	HIMACHAL PRADESH	0.734163	0.676125	0.790549
2	PUNJAB	0.674293	0.631194	0.712871
3	CHANDIGARH	0.763123	0.716304	0.801405
4	UTTARAKHAND	0.682208	0.609727	0.752020
5	HARYANA	0.654755	0.573929	0.725770
6	NCT OF DELHI	0.758745	0.710574	0.800556
7	RAJASTHAN	0.558368	0.442061	0.666322
8	UTTAR PRADESH	0.572525	0.484231	0.653088
9	BIHAR	0.504369	0.419430	0.582334
10	SIKKIM	0.728740	0.673664	0.777754
11	ARUNACHAL PRADESH	0.553581	0.486906	0.616138
12	NAGALAND	0.678510	0.648399	0.706541
13	MANIPUR	0.668282	0.612908	0.722833
14	MIZORAM	0.773032	0.755990	0.789660
15	TRIPURA	0.763431	0.724304	0.800997
16	MEGHALAYA	0.601642	0.590571	0.612588
17	ASSAM	0.614569	0.563950	0.663049
18	WEST BENGAL	0.674199	0.623371	0.722485
19	JHARKHAND	0.555596	0.463671	0.642783
20	ORISSA	0.637120	0.561259	0.711370
21	CHHATTISGARH	0.602067	0.516982	0.686353
22	MADHYA PRADESH	0.590019	0.504843	0.669311
23	GUJARAT	0.679907	0.608616	0.745441
24	DAMAN AND DIU	0.774546	0.686022	0.829289
25	DADRA AND NAGAR HAVELI	0.649474	0.538243	0.735554
26	MAHARASHTRA	0.725738	0.670531	0.777047
27	ANDHRA PRADESH	0.597733	0.529337	0.665641
28	KARNATAKA	0.665310	0.602047	0.726862
29	GOA	0.799075	0.764054	0.833162

	State	Literacy Rate	Female Literacy Rate	Male Literacy Rate
30	LAKSHADWEEP	0.815116	0.782456	0.846028
31	KERALA	0.842237	0.830382	0.855091
32	TAMIL NADU	0.718498	0.660862	0.775929
33	PONDICHERRY	0.767103	0.723797	0.812031
34	ANDAMAN AND NICOBAR ISLANDS	0.773241	0.730989	0.810254

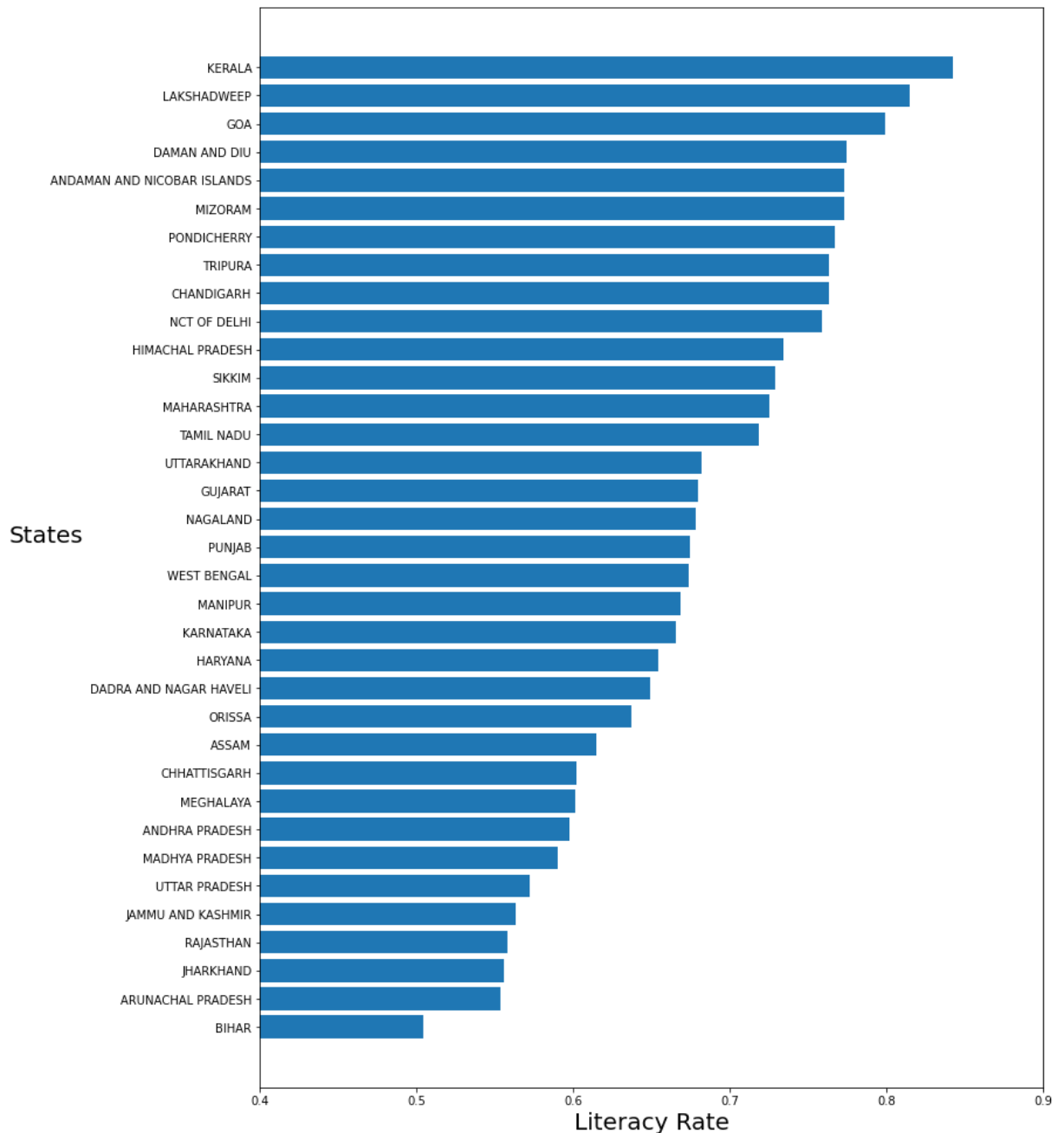
## Plotting Literacy Rate of Different States in Order

In [292...

```

df3 = new_df.sort_values(by=['Literacy Rate'])
rate = df3['Literacy Rate'].tolist()
states = df3['State'].tolist()
fig = plt.figure(figsize = (12, 17))
plt.barh(states,rate)
plt.ylabel("States", size = 20, rotation='horizontal')
plt.xlabel("Literacy Rate" , size = 20)
plt.xlim([0.4, 0.9])
#plt.title("", size = 20)
plt.show()

```



## Calculating Data for Wealth Metric

```
In [293... states = pd.Series(df['State name']).unique().tolist()
hous = [0 for i in states]
elec = [0 for i in states]
tv = [0 for i in states]
comp = [0 for i in states]
mobile = [0 for i in states]
two = [0 for i in states]
four = [0 for i in states]

for index, row in df.iterrows():
    sn = row['State name']
    i = states.index(sn)
    hous[i] += row['Households']
    elec[i] += row['Households_with_Electric_Lighting']
    tv[i] += row['Households_with_Television']
    comp[i] += row['Households_with_Computer']
    mobile[i] += row['Households_with_Telephone_Mobile_Phone']
    two[i] += row['Households_with_Scooter_Motorcycle_Moped']
    four[i] += row['Households_with_Car_Jeep_Van']
```

```

for i in range(35):
    elec[i] /= hous[i]
    tv[i] /= hous[i]
    comp[i] /= hous[i]
    mobile[i] /= hous[i]
    two[i] /= hous[i]
    four[i] /= hous[i]

new_df['Electricity'] = elec
new_df['T.V.'] = tv
new_df['Computer'] = comp
new_df['Mobile'] = mobile
new_df['2-Wheeler'] = two
new_df['4-Wheeler'] = four

new_df

```

Out[293...

	State	Literacy Rate	Female Literacy Rate	Male Literacy Rate	Electricity	T.V.	Computer	Mobile	2-Wheeler
0	JAMMU AND KASHMIR	0.563517	0.474959	0.642206	0.476063	0.016088	0.046720	0.388554	0.072234
1	HIMACHAL PRADESH	0.734163	0.676125	0.790549	0.487030	0.014041	0.042042	0.414167	0.078089
2	PUNJAB	0.674293	0.631194	0.712871	0.665923	0.037224	0.087939	0.566138	0.327087
3	CHANDIGARH	0.763123	0.716304	0.801405	0.774898	0.148418	0.261802	0.703104	0.367677
4	UTTARAKHAND	0.682208	0.609727	0.752020	0.513735	0.018630	0.064538	0.440265	0.135087
5	HARYANA	0.654755	0.573929	0.725770	0.602468	0.034999	0.087887	0.528126	0.221890
6	NCT OF DELHI	0.758745	0.710574	0.800556	0.718873	0.127878	0.211255	0.658287	0.281917
7	RAJASTHAN	0.558368	0.442061	0.666322	0.466496	0.012534	0.048139	0.491546	0.167489
8	UTTAR PRADESH	0.572525	0.484231	0.653088	0.268310	0.013499	0.058984	0.487834	0.142913
9	BIHAR	0.504369	0.419430	0.582334	0.132329	0.007069	0.056997	0.448897	0.065542
10	SIKKIM	0.728740	0.673664	0.777754	0.701554	0.025346	0.087764	0.553129	0.021476
11	ARUNACHAL PRADESH	0.553581	0.486906	0.616138	0.457019	0.013640	0.056514	0.336121	0.097593
12	NAGALAND	0.678510	0.648399	0.706541	0.610760	0.012977	0.066861	0.397112	0.047292
13	MANIPUR	0.668282	0.612908	0.722833	0.568288	0.017847	0.074934	0.478233	0.164689
14	MIZORAM	0.773032	0.755990	0.789660	0.690919	0.020803	0.124893	0.596650	0.113480
15	TRIPURA	0.763431	0.724304	0.800997	0.540638	0.008072	0.058073	0.379725	0.065110
16	MEGHALAYA	0.601642	0.590571	0.612588	0.454571	0.011198	0.056651	0.320748	0.039999
17	ASSAM	0.614569	0.563950	0.663049	0.259948	0.011039	0.065401	0.335837	0.071234
18	WEST BENGAL	0.674199	0.623371	0.722485	0.431473	0.017505	0.065845	0.389805	0.067484
19	JHARKHAND	0.555596	0.463671	0.642783	0.362988	0.011682	0.054697	0.380769	0.127292
20	ORISSA	0.637120	0.561259	0.711370	0.325719	0.010611	0.038487	0.301500	0.109817
21	CHHATTISGARH	0.602067	0.516982	0.686353	0.622373	0.009974	0.038380	0.253872	0.129124

	State	Literacy Rate	Female Literacy Rate	Male Literacy Rate	Electricity	T.V.	Computer	Mobile	2-Wheeler
22	MADHYA PRADESH	0.590019	0.504843	0.669311	0.542990	0.011486	0.047766	0.372278	0.152194
23	GUJARAT	0.679907	0.608616	0.745441	0.628464	0.021777	0.061488	0.479330	0.237297
24	DAMAN AND DIU	0.774546	0.686022	0.829289	0.665236	0.018726	0.062249	0.580147	0.211133
25	DADRA AND NAGAR HAVELI	0.649474	0.538243	0.735554	0.639420	0.018505	0.055312	0.422612	0.171415
26	MAHARASHTRA	0.725738	0.670531	0.777047	0.595898	0.041089	0.094550	0.490285	0.176998
27	ANDHRA PRADESH	0.597733	0.529337	0.665641	0.757071	0.021461	0.068902	0.518161	0.152923
28	KARNATAKA	0.665310	0.602047	0.726862	0.663671	0.035472	0.094019	0.524105	0.187477
29	GOA	0.799075	0.764054	0.833162	0.542440	0.071220	0.174416	0.498940	0.318795
30	LAKSHADWEEP	0.815116	0.782456	0.846028	0.502260	0.015394	0.070662	0.471848	0.193438
31	KERALA	0.842237	0.830382	0.855091	0.649385	0.043111	0.108278	0.616829	0.165602
32	TAMIL NADU	0.718498	0.660862	0.775929	0.745253	0.033335	0.084459	0.597876	0.258017
33	PONDICHERRY	0.767103	0.723797	0.812031	0.758262	0.046614	0.105562	0.627911	0.361733
34	ANDAMAN AND NICOBAR ISLANDS	0.773241	0.730989	0.810254	0.567644	0.022827	0.057837	0.551262	0.162880

Correlation

In [294...

new\_df.corr()

Out[294...

	Literacy Rate	Female Literacy Rate	Male Literacy Rate	Electricity	T.V.	Computer	Mobile	2-Wheeler	4-Wheeler
Literacy Rate	1.000000	0.981689	0.974847	0.578507	0.448626	0.511363	0.592412	0.418484	0.417322
Female Literacy Rate	0.981689	1.000000	0.916683	0.545021	0.435113	0.520344	0.553752	0.359320	0.423355
Male Literacy Rate	0.974847	0.916683	1.000000	0.589097	0.441286	0.474855	0.602418	0.471178	0.380522
Electricity	0.578507	0.545021	0.589097	1.000000	0.490812	0.478130	0.624161	0.538415	0.460455
T.V.	0.448626	0.435113	0.441286	0.490812	1.000000	0.955996	0.669908	0.690434	0.916622
Computer	0.511363	0.520344	0.474855	0.478130	0.955996	1.000000	0.698258	0.638185	0.918795
Mobile	0.592412	0.553752	0.602418	0.624161	0.669908	0.698258	1.000000	0.674648	0.604385
2-Wheeler	0.418484	0.359320	0.471178	0.538415	0.690434	0.638185	0.674648	1.000000	0.602711



	Literacy Rate	Female Literacy Rate	Male Literacy Rate	Electricity	T.V.	Computer	Mobile	2-Wheeler	4-Wheeler
4-Wheeler	0.417323	0.423359	0.380522	0.460452	0.916622	0.918798	0.604385	0.602710	1.000000

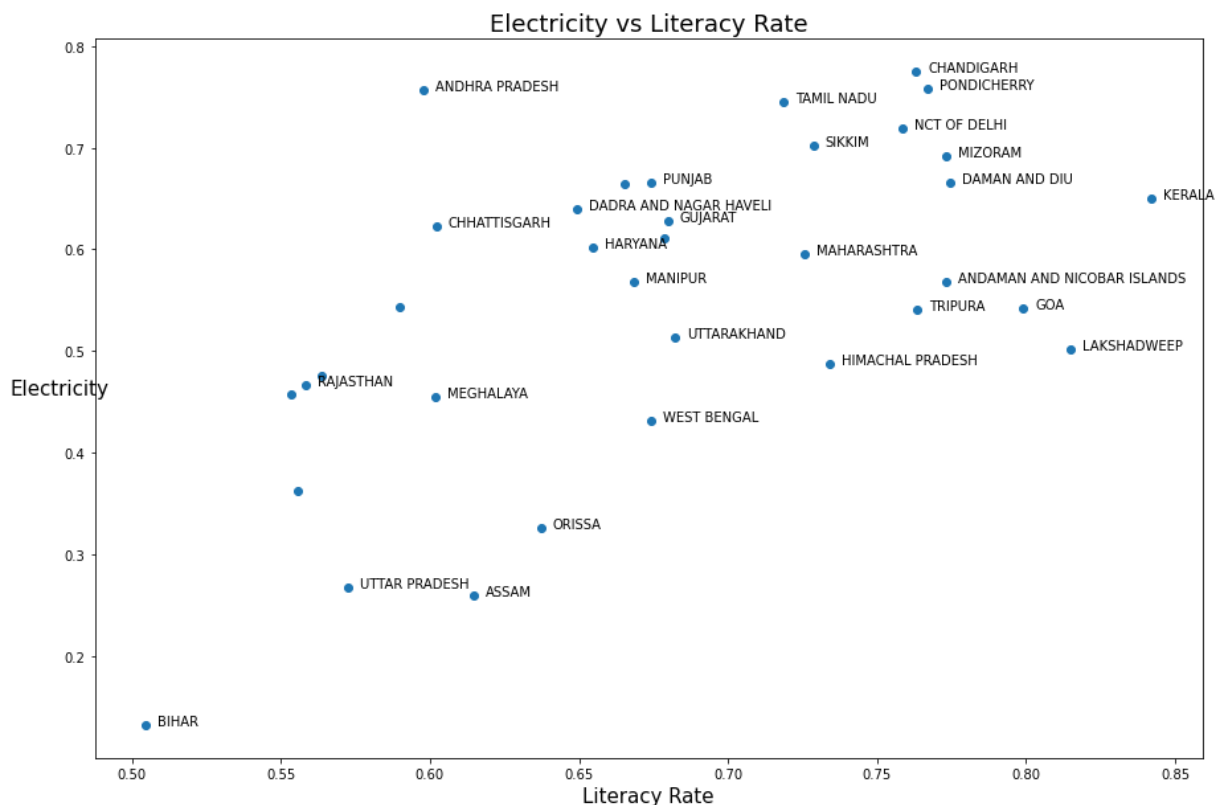
## Electricity per Household vs Literacy Rate

```
In [295... x = new_df['Literacy Rate'].tolist()
y = new_df['Electricity'].tolist()

fig = plt.figure(figsize = (15, 10))
plt.scatter(x,y)
plt.xlabel("Literacy Rate", size = 15)
plt.ylabel("Electricity" , size = 15,rotation='horizontal')
plt.title("Electricity vs Literacy Rate", size = 18)

for i, txt in enumerate(states):
    if txt in ['HIMACHAL PRADESH', 'PUNJAB', 'CHANDIGARH', 'UTTARAKHAND', 'HARYANA',
               'RAJASTHAN', 'UTTAR PRADESH', 'BIHAR', 'SIKKIM', 'MANIPUR', 'MIZORAM',
               'TRIPURA', 'ASSAM', 'WEST BENGAL', 'ORISSA', 'CHHATTISGARH',
               'GUJARAT', 'DAMAN AND DIU', 'DADRA AND NAGAR HAVELI', 'MAHARASHTRA',
               'GOA', 'LAKSHADWEEP', 'KERALA', 'MEGHALAYA', 'TAMIL NADU', 'PONDICHERY']:
        plt.annotate(txt, (x[i]+.004, y[i]))

plt.show()
```



## Mobile Commodity vs Literacy Rate

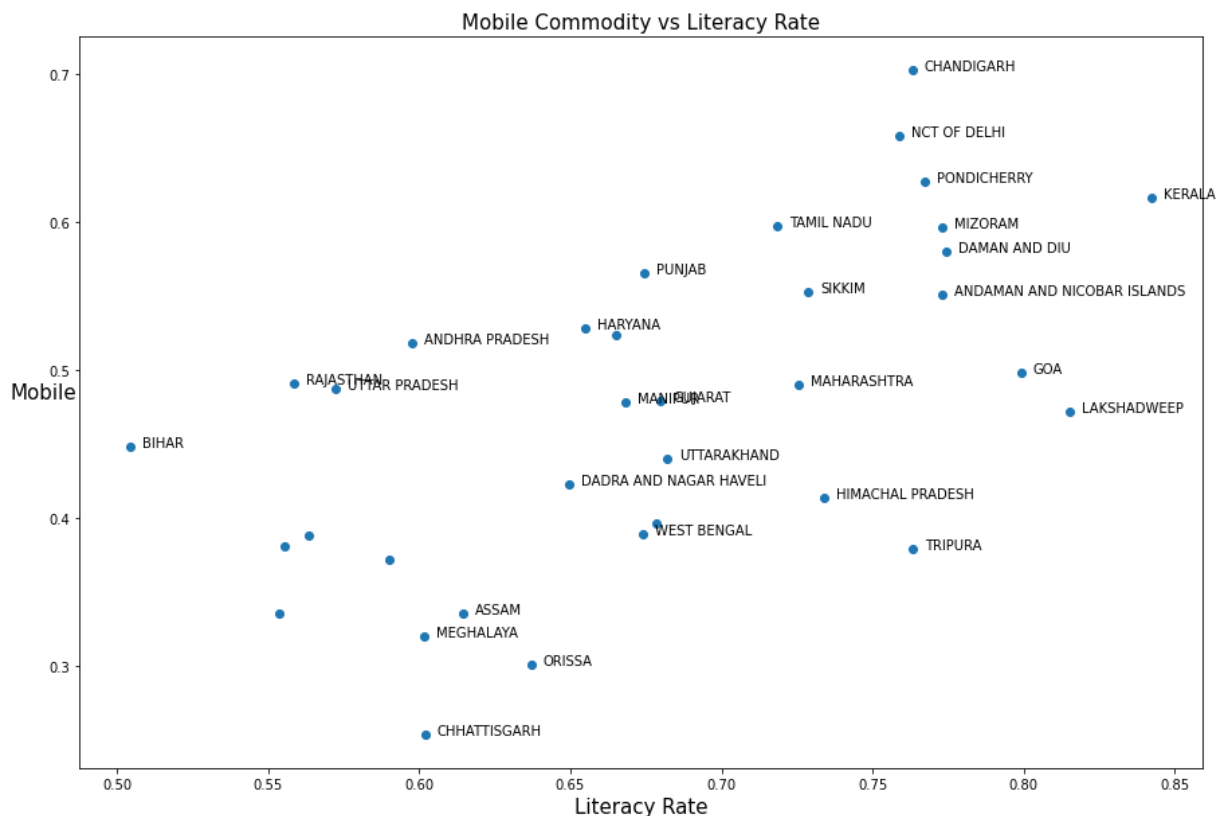
```
In [296... x = new_df['Literacy Rate'].tolist()
y = new_df['Mobile'].tolist()

fig = plt.figure(figsize = (15, 10))
plt.scatter(x,y)
```

```
plt.xlabel("Literacy Rate", size = 15)
plt.ylabel("Mobile" , size = 15,rotation='horizontal')
plt.title("Mobile Commodity vs Literacy Rate", size = 15)

for i, txt in enumerate(states):
    if txt in ['HIMACHAL PRADESH', 'PUNJAB', 'CHANDIGARH', 'UTTARAKHAND', 'HARYANA',
               'RAJASTHAN', 'UTTAR PRADESH', 'BIHAR', 'SIKKIM', 'MANIPUR', 'MIZORAM',
               'TRIPURA', 'ASSAM', 'WEST BENGAL', 'ORISSA', 'CHHATTISGARH',
               'GUJARAT', 'DAMAN AND DIU', 'DADRA AND NAGAR HAVELI', 'MAHARASHTRA',
               'GOA', 'LAKSHADWEEP', 'KERALA', 'MEGHALAYA', 'TAMIL NADU', 'PONDICHERRY']:
        plt.annotate(txt, (x[i]+.004, y[i]))

plt.show()
```



## Dimensionality Reduction using PCA

- ### Standardize the Data

```
In [297... features = ['Electricity','T.V.','Computer','Mobile','2-Wheeler','4-Wheeler']
# Separating out the features
x = new_df.loc[:, features].values
# Separating out the target
y = new_df.loc[:,['Literacy Rate','Female Literacy Rate','Male Literacy Rate']].valu
# Standardizing the features
x = StandardScaler().fit_transform(x)
```

- ### Principle Component Analysis (PCA)

```
In [298... pca = PCA(n_components=2)
principalComponents = pca.fit_transform(x)
principalDf = pd.DataFrame(data = principalComponents, columns = ['principal compone

final_df = pd.concat([new_df[['Literacy Rate','Female Literacy Rate','Male Literacy
final_df
```

Out[298...

	<b>Literacy Rate</b>	<b>Female Literacy Rate</b>	<b>Male Literacy Rate</b>	<b>principal component 1</b>	<b>principal component 2</b>
<b>0</b>	0.563517	0.474959	0.642206	-1.473560	0.417109
<b>1</b>	0.734163	0.676125	0.790549	-1.404316	0.218053
<b>2</b>	0.674293	0.631194	0.712871	1.921930	-0.709739
<b>3</b>	0.763123	0.716304	0.801405	7.363155	1.665735
<b>4</b>	0.682208	0.609727	0.752020	-0.777305	0.019139
<b>5</b>	0.654755	0.573929	0.725770	0.937953	-0.243887
<b>6</b>	0.758745	0.710574	0.800556	5.355753	1.226566
<b>7</b>	0.558368	0.442061	0.666322	-0.840522	-0.222528
<b>8</b>	0.572525	0.484231	0.653088	-1.318396	0.808652
<b>9</b>	0.504369	0.419430	0.582334	-2.351515	1.547685
<b>10</b>	0.728740	0.673664	0.777754	0.159886	-0.429329
<b>11</b>	0.553581	0.486906	0.616138	-1.400297	0.773743
<b>12</b>	0.678510	0.648399	0.706541	-0.937171	0.097241
<b>13</b>	0.668282	0.612908	0.722833	-0.162031	-0.238194
<b>14</b>	0.773032	0.755990	0.789660	0.947477	-0.573082
<b>15</b>	0.763431	0.724304	0.800997	-1.673354	-0.072410
<b>16</b>	0.601642	0.590571	0.612588	-1.908544	0.813521
<b>17</b>	0.614569	0.563950	0.663049	-2.199959	1.515229
<b>18</b>	0.674199	0.623371	0.722485	-1.640458	0.548920
<b>19</b>	0.555596	0.463671	0.642783	-1.710171	0.643795
<b>20</b>	0.637120	0.561259	0.711370	-2.422797	0.904483
<b>21</b>	0.602067	0.516982	0.686353	-1.837769	-0.327760
<b>22</b>	0.590019	0.504843	0.669311	-1.321570	-0.279831
<b>23</b>	0.679907	0.608616	0.745441	0.133274	-0.832197
<b>24</b>	0.774546	0.686022	0.829289	0.407124	-1.302118
<b>25</b>	0.649474	0.538243	0.735554	-0.492645	-0.653156
<b>26</b>	0.725738	0.670531	0.777047	0.442966	-0.137945
<b>27</b>	0.597733	0.529337	0.665641	0.037475	-1.476491
<b>28</b>	0.665310	0.602047	0.726862	0.717629	-0.603401
<b>29</b>	0.799075	0.764054	0.833162	3.209606	1.489571
<b>30</b>	0.815116	0.782456	0.846028	-0.693652	-0.408111
<b>31</b>	0.842237	0.830382	0.855091	1.447032	-0.364120
<b>32</b>	0.718498	0.660862	0.775929	1.218721	-1.596377
<b>33</b>	0.767103	0.723797	0.812031	2.302001	-1.649568
<b>34</b>	0.773241	0.730989	0.810254	-0.035948	-0.569200

## Result

- ### Correlation

In [299... final\_df.corr()

Out[299...

	Literacy Rate	Female Literacy Rate	Male Literacy Rate	principal component 1	principal component 2
Literacy Rate	1.000000	0.981689	0.974847	5.720551e-01	-2.649433e-01
Female Literacy Rate	0.981689	1.000000	0.916683	5.488191e-01	-2.073006e-01
Male Literacy Rate	0.974847	0.916683	1.000000	5.683620e-01	-3.226811e-01
principal component 1	0.572055	0.548819	0.568362	1.000000e+00	-1.915884e-16
principal component 2	-0.264943	-0.207301	-0.322681	-1.915884e-16	1.000000e+00

- ### Wealth vs Literacy Rate

In [300...

```
x = final_df['Literacy Rate'].tolist()
y = final_df['principal component 1'].tolist()

# Linear Regression
Sum = 0
for i in x:
    Sum += i;
x_mean = Sum/len(x);

Sum = 0
for i in y:
    Sum += i;
y_mean = Sum/len(y);

nume = 0
deno = 0
for i in range(len(x)):
    nume += (x[i]-x_mean)*(y[i]-y_mean)
    deno += pow((x[i] - x_mean),2)

# y = mx + c
m = nume/deno
c = y_mean - m*x_mean

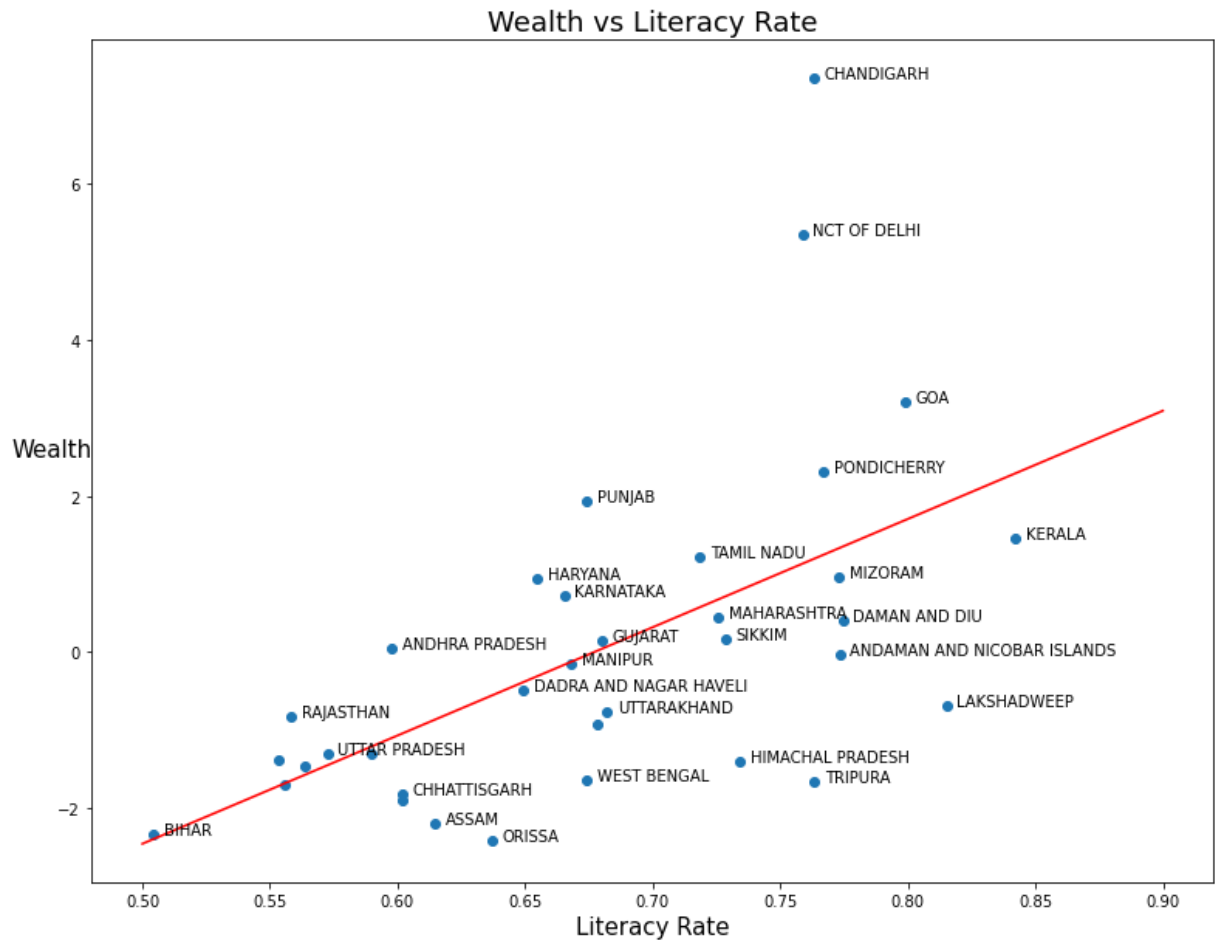
xx = np.array([0.5,0.9])
plt.figure(figsize = (13,10))
plt.plot(xx, xx*m + c, color = 'red')

plt.scatter(x,y)
plt.xlabel("Literacy Rate", size = 15)
plt.ylabel("Wealth" , size = 15,rotation='horizontal')
plt.title("Wealth vs Literacy Rate", size = 18)

for i, txt in enumerate(states):
    if txt in ['HIMACHAL PRADESH', 'PUNJAB', 'CHANDIGARH', 'UTTARAKHAND', 'HARYANA',
               'RAJASTHAN', 'UTTAR PRADESH', 'BIHAR', 'SIKKIM', 'MANIPUR', 'MIZORAM',
               'TRIPURA', 'ASSAM', 'WEST BENGAL', 'ORISSA', 'CHHATTISGARH',
               'GUJARAT', 'DAMAN AND DIU', 'DADRA AND NAGAR HAVELI', 'MAHARASHTRA',
```

```
'GOA', 'LAKSHADWEEP', 'KERALA', 'TAMIL NADU', 'PONDICHERRY', 'ANDAMAN  
plt.annotate(txt, (x[i]+.004, y[i]))
```

```
plt.show()
```



- ### Wealth vs Male Literacy Rate

```
In [301... x = final_df['Male Literacy Rate'].tolist()
y = final_df['principal component 1'].tolist()

# Linear Regression
Sum = 0
for i in x:
    Sum += i;
x_mean = Sum/len(x);

Sum = 0
for i in y:
    Sum += i;
y_mean = Sum/len(y);

nume = 0
deno = 0
for i in range(len(x)):
    nume += (x[i]-x_mean)*(y[i]-y_mean)
    deno += pow((x[i] - x_mean),2)

# y = mx + c
m = nume/deno
c = y_mean - m*x_mean

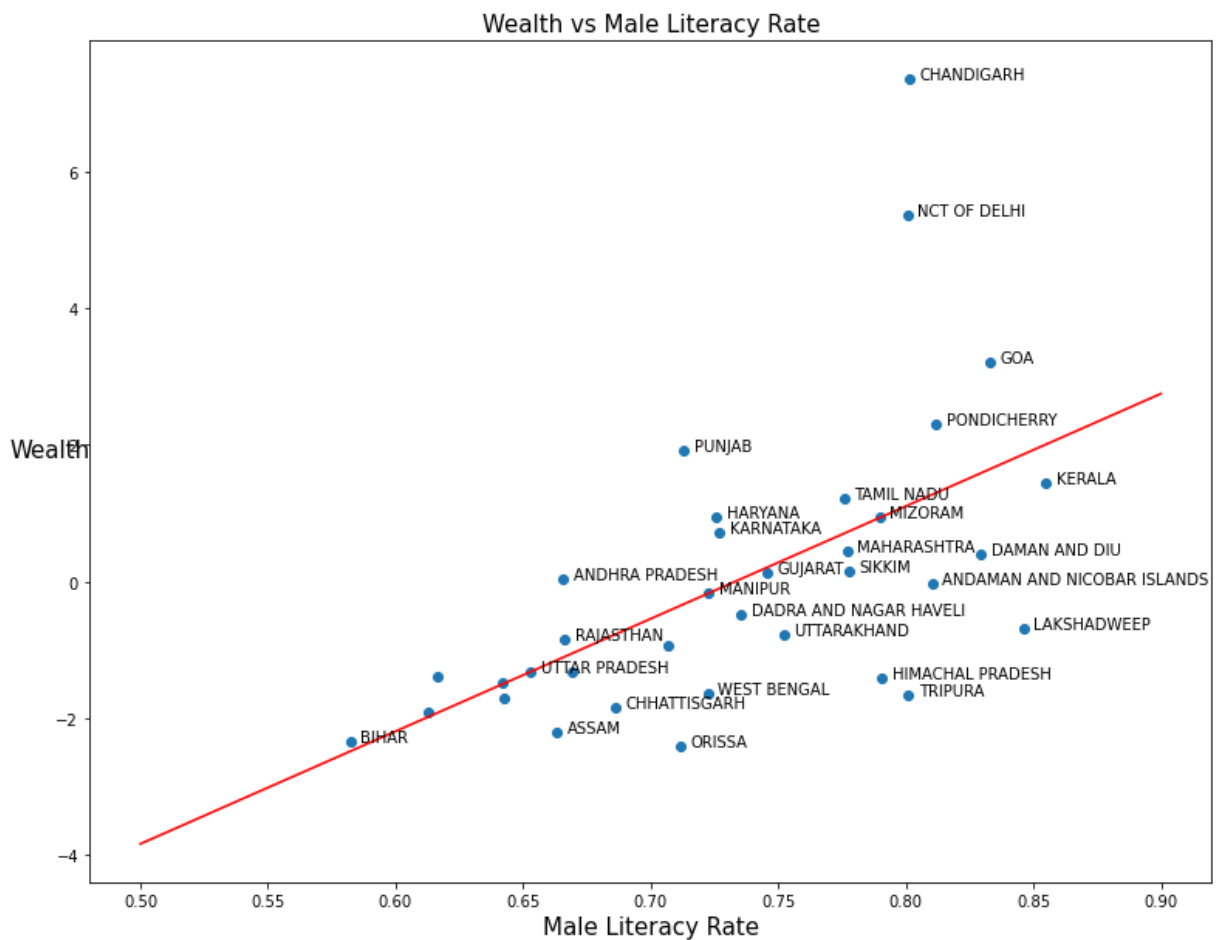
xx = np.array([0.5,0.9])
plt.figure(figsize = (13,10))
```

```
plt.plot(xx, xx*m + c, color = 'red')

plt.scatter(x,y)
plt.xlabel("Male Literacy Rate", size = 15)
plt.ylabel("Wealth" , size = 15,rotation='horizontal')
plt.title("Wealth vs Male Literacy Rate", size = 15)

for i, txt in enumerate(states):
    if txt in ['HIMACHAL PRADESH', 'PUNJAB', 'CHANDIGARH', 'UTTARAKHAND', 'HARYANA',
               'RAJASTHAN', 'UTTAR PRADESH', 'BIHAR', 'SIKKIM', 'MANIPUR', 'MIZORAM',
               'TRIPURA', 'ASSAM', 'WEST BENGAL', 'ORISSA', 'CHHATTISGARH',
               'GUJARAT', 'DAMAN AND DIU', 'DADRA AND NAGAR HAVELI', 'MAHARASHTRA',
               'GOA', 'LAKSHADWEEP', 'KERALA', 'TAMIL NADU', 'PONDICHERRY', 'ANDAMAN
               plt.annotate(txt, (x[i]+.004, y[i]))

plt.show()
```



- ### Wealth vs Female Literacy Rate

In [302...

```
x = final_df['Female Literacy Rate'].tolist()
y = final_df['principal component 1'].tolist()

# Linear Regression
Sum = 0
for i in x:
    Sum += i;
x_mean = Sum/len(x);

Sum = 0
for i in y:
    Sum += i;
y_mean = Sum/len(y);
```

```

nume = 0
deno = 0
for i in range(len(x)):
    nume += (x[i]-x_mean)*(y[i]-y_mean)
    deno += pow((x[i] - x_mean),2)

# y = mx + c
m = nume/deno
c = y_mean - m*x_mean

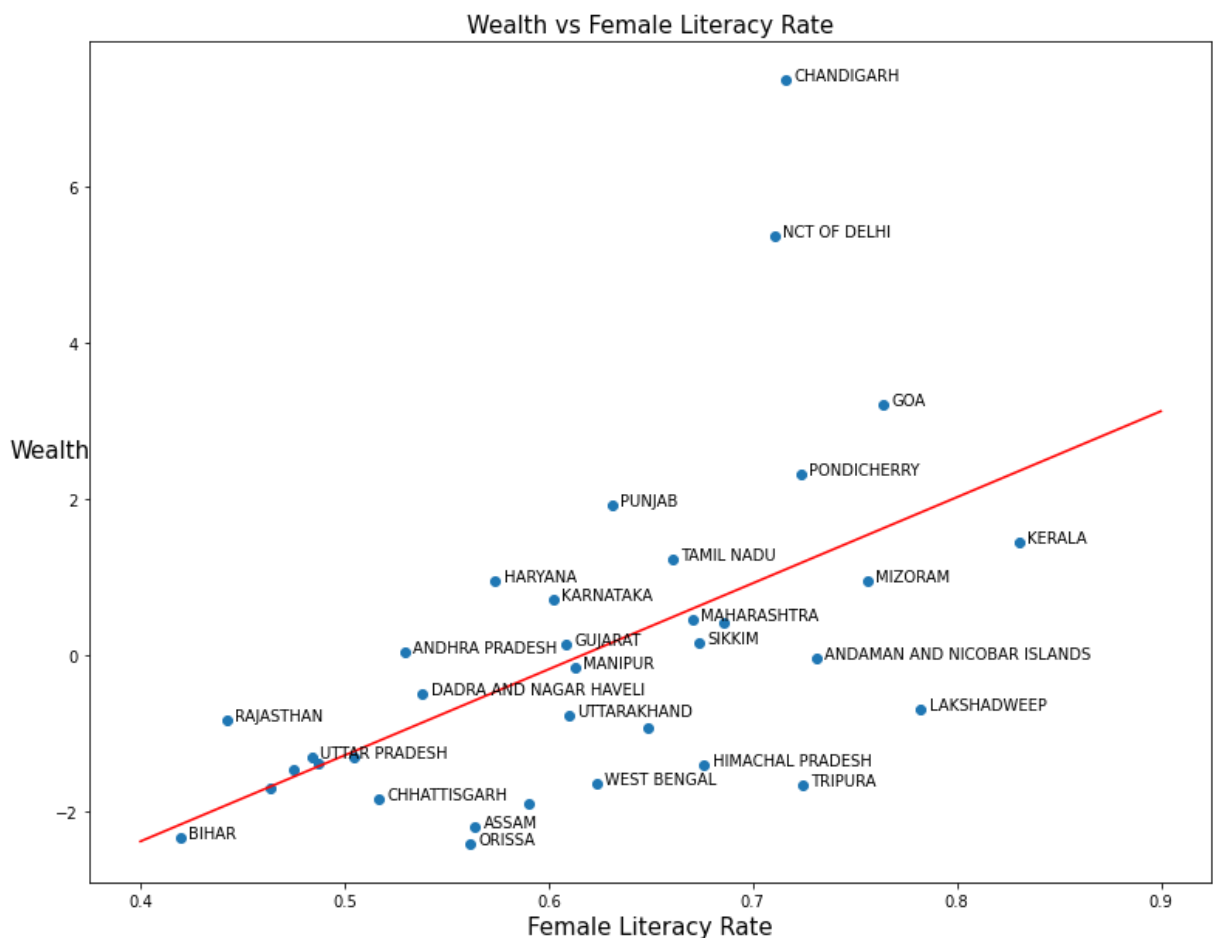
xx = np.array([0.4,0.9])
plt.figure(figsize = (13,10))
plt.plot(xx, xx*m + c, color = 'red')

plt.scatter(x,y)
plt.xlabel("Female Literacy Rate", size = 15)
plt.ylabel("Wealth" , size = 15,rotation='horizontal')
plt.title("Wealth vs Female Literacy Rate", size = 15)

for i, txt in enumerate(states):
    if txt in ['HIMACHAL PRADESH', 'PUNJAB', 'CHANDIGARH', 'UTTARAKHAND', 'HARYANA',
               'RAJASTHAN', 'UTTAR PRADESH', 'BIHAR', 'SIKKIM', 'MANIPUR', 'MIZORAM',
               'TRIPURA', 'ASSAM', 'WEST BENGAL', 'ORISSA', 'CHHATTISGARH',
               'GUJARAT', 'DADRA AND NAGAR HAVELI', 'MAHARASHTRA', 'ANDHRA PRADESH',
               'GOA', 'LAKSHADWEEP', 'KERALA', 'TAMIL NADU', 'PONDICHERRY', 'ANDAMAN
               AND NICOBAR ISLANDS']:
        plt.annotate(txt, (x[i]+.004, y[i]))

plt.show()

```



## Conclusion

- The Correlation between Literacy Rate and Wealth comes out to be 0.57

- The Scatter Graphs Plotted gives us impression that the data follow a linear order with few inconsistencies.
- Further Linear Regression gives tells us the relation between Literacy Rate is somewhat linear with few inconsistencies.
- The Male Literacy Rate is the most related compared to Female Literacy Rate or Total Literacy Rate to the Wealth.

## Dataset Reference

- <https://www.kaggle.com/danofer/india-census?select=india-districts-census-2011.csv>