

**Presented by:-**  
Ved Kadiya(0803785)  
Deep Patel(0805115)  
Nihar Patel(0806045)

# Market Segmentation for Healthcare Provider

Segmentation using K-means clustering



---

# Table of Contents

---

## Presentation Agenda

---

- 01 What is market Segmentation and targeting strategies
  - 02 Data Importing and cleaning
  - 03 Exploring Insights
  - 04 Model Creation
- 

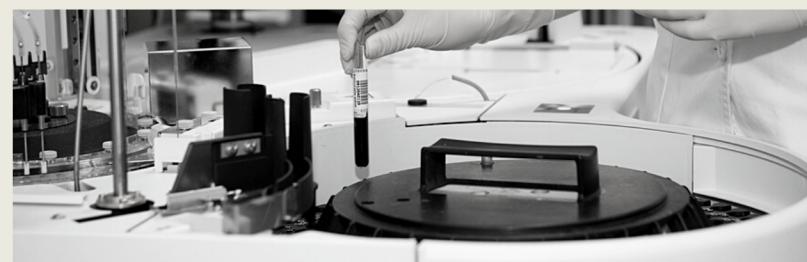


# What is market Segmentation and targeting strategies?

---

# Segmentation, Targeting and Positioning

---



## Segmentation

It groups customers with similar needs together and then determines the characteristics of those customers.

---

## Targeting

Companies will determine this base on the attractiveness of the segment.

---

## Positioning

creating a value proposition for the company that will appeal to the selected customer segment.



# Data Importing and Cleaning

# Data Importing

- Here, we used online healthcare provider data .
- For importing data we used pandas library.

In [2]: `data = pd.read_csv('C:/Users/vedka/Downloads/india-districts-census-2011.csv')  
data.head()`

out[2]:

	District code	State name	District name	Population	Male	Female	Literate	Male_Literate	Female_Literate	SC	...	Power_Parity_Rs_90000_150000
0	1	JAMMU AND KASHMIR	Kupwara	870354	474190	396164	439654	282823	156831	1048	...	94
1	2	JAMMU AND KASHMIR	Badgam	753745	398041	355704	335649	207741	127908	368	...	126
2	3	JAMMU AND KASHMIR	Leh(Ladakh)	133487	78971	54516	93770	62834	30936	488	...	46
3	4	JAMMU AND KASHMIR	Kargil	140802	77785	63017	86236	56301	29935	18	...	27
4	5	JAMMU AND KASHMIR	Punch	476835	251899	224936	261724	163333	98391	556	...	78

5 rows × 118 columns



# Data Cleaning

- Checking Nulls
- Dumping unwanted rows

- For this dataset we used just 2 cleaning techniques.
- We checked for null values, but there are no null values for this dataset.
- and , then after we dumped unwanted rows.

## Checking for null values

```
In [5]: data.isnull().sum()
```

```
Out[5]: District code      0  
State name        0  
District name      0  
Population         0  
Male              0  
..  
Power_Parity_Rs_330000_425000 0  
Power_Parity_Rs_425000_545000 0  
Power_Parity_Rs_330000_545000 0  
Power_Parity_Above_Rs_545000 0  
Total_Power_Parity       0  
Length: 118, dtype: int64
```

- There are no null values so carrying forward with our analysis

```
In [6]: data.drop(['SC','Male_SC','Female_SC','ST','Male_ST','Female_ST','Male_Workers','Female_Workers','Hindus','Muslims','Christians','Households_with_Electric_Lighting'], axis=1, inplace=True)
```

Here we deleted columns that were not suitable for the purpose of the analysis for our business problem

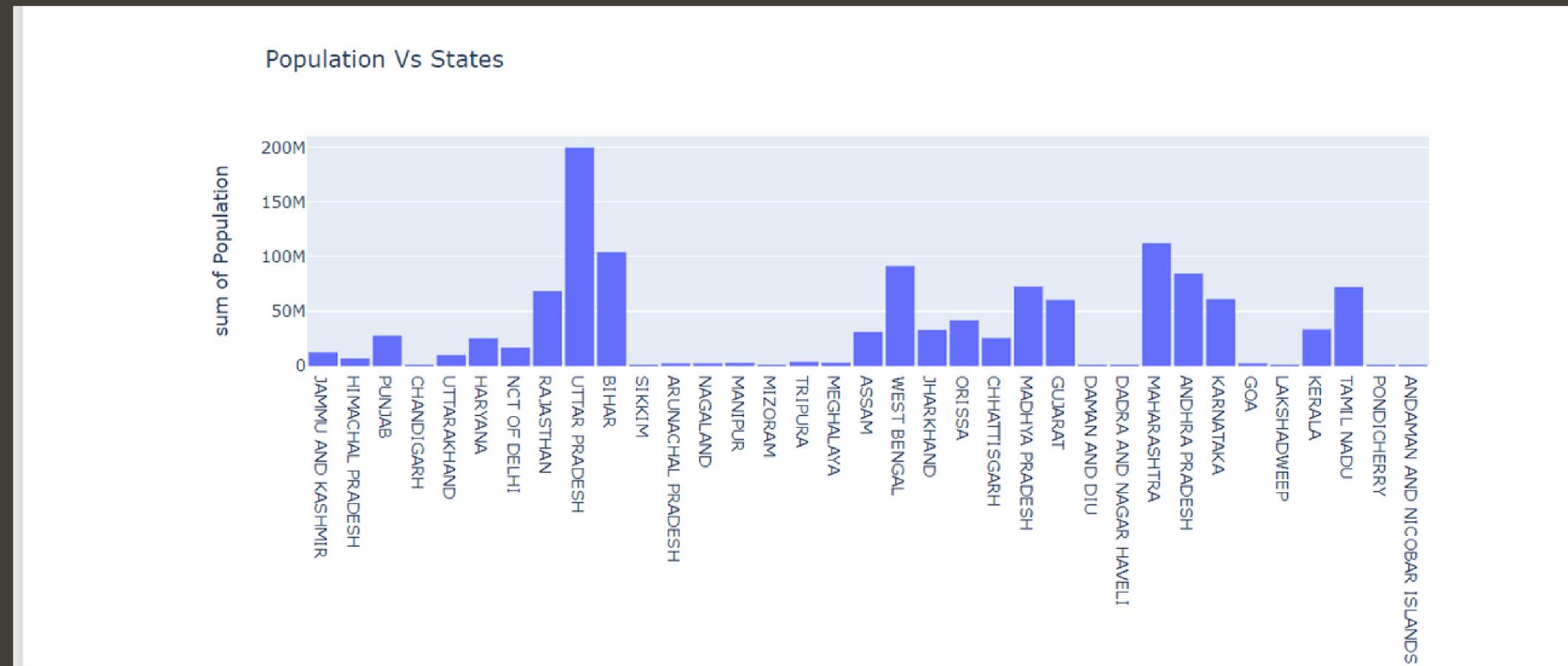
```
In [7]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 640 entries, 0 to 639  
Data columns (total 48 columns):  
 #   Column           Non-Null Count Dtype  
 ---  --  
 0   District code    640 non-null   int64  
 1   State name      640 non-null   object  
 2   District name    640 non-null   object  
 3   Population       640 non-null   int64  
 4   Male             640 non-null   int64  
 5   Female           640 non-null   int64  
 6   Literate          640 non-null   int64  
 7   Male_Literate    640 non-null   int64  
 8   Female_Literate  640 non-null   int64  
 9   Workers           640 non-null   int64  
 10  Main_Workers     640 non-null   int64  
 11  Marginal_Workers 640 non-null   int64  
 12  Non_Workers      640 non-null   int64  
 13  Cultivator_Workers 640 non-null   int64  
 14  Agricultural_Workers 640 non-null   int64  
 15  Household_Workers 640 non-null   int64  
 16  Other_Workers     640 non-null   int64  
 17  Households_with_Internet 640 non-null   int64  
 18  Households_with_Computer 640 non-null   int64
```



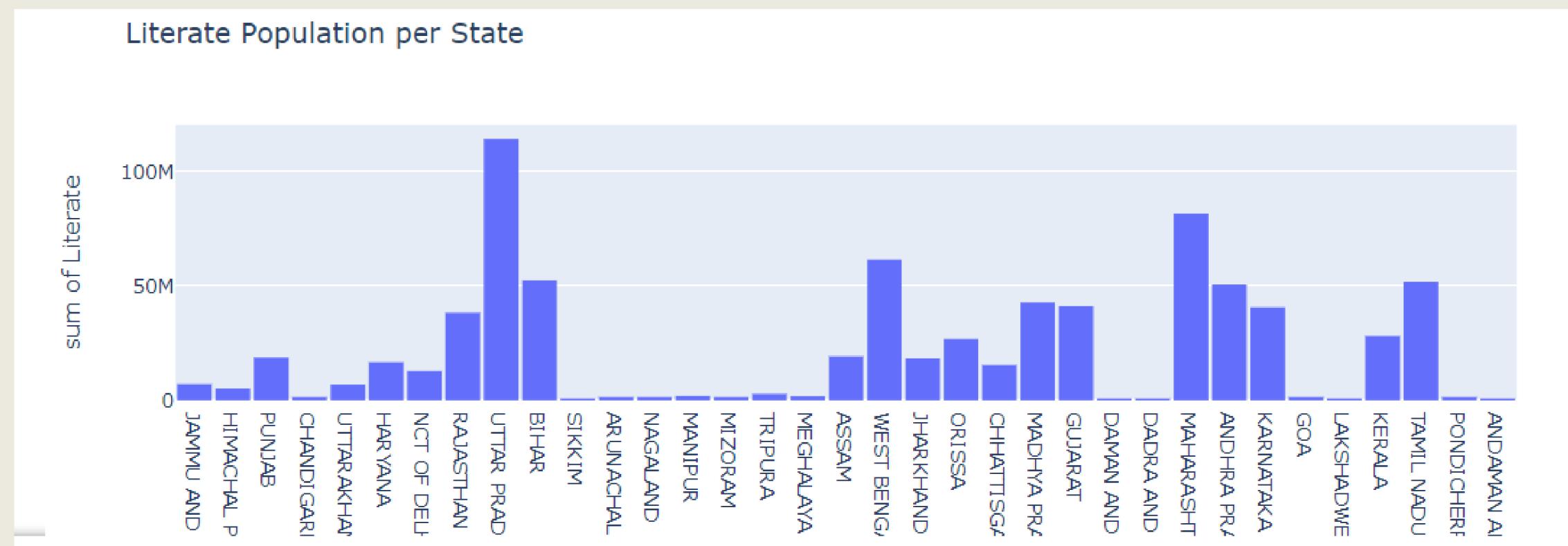
# Exploring Insights

# Population Count



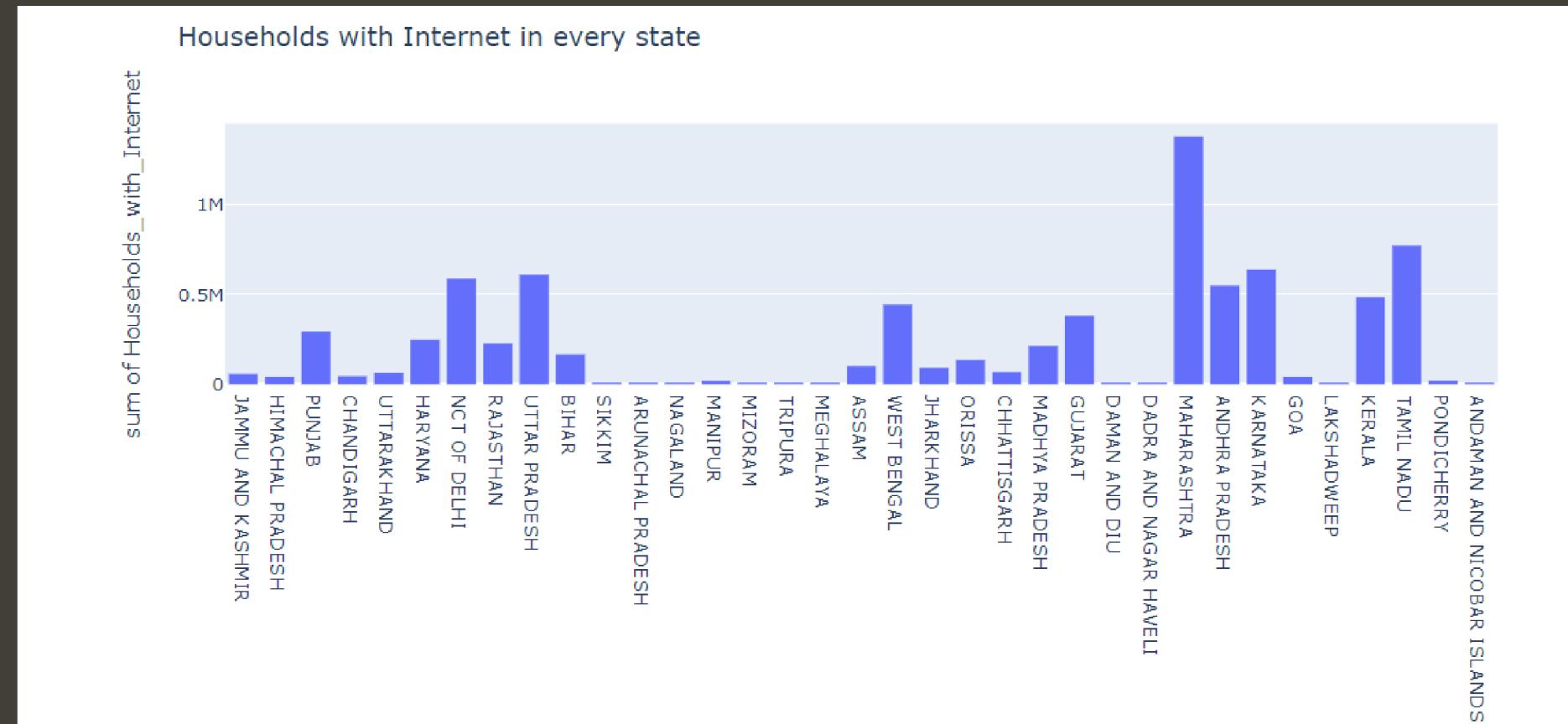
- So there are many states which can be selected for segmentation to start their services in solely based on population count.
- These are states with population greater than 50 Millions and this does not visualize whole scenario it is just a speculation based on Total population count of the above given states.

# Literacy Rate



- Number of literate people residing in every state as literacy rate is directly correlated by regular medical check ups.

# Internet users



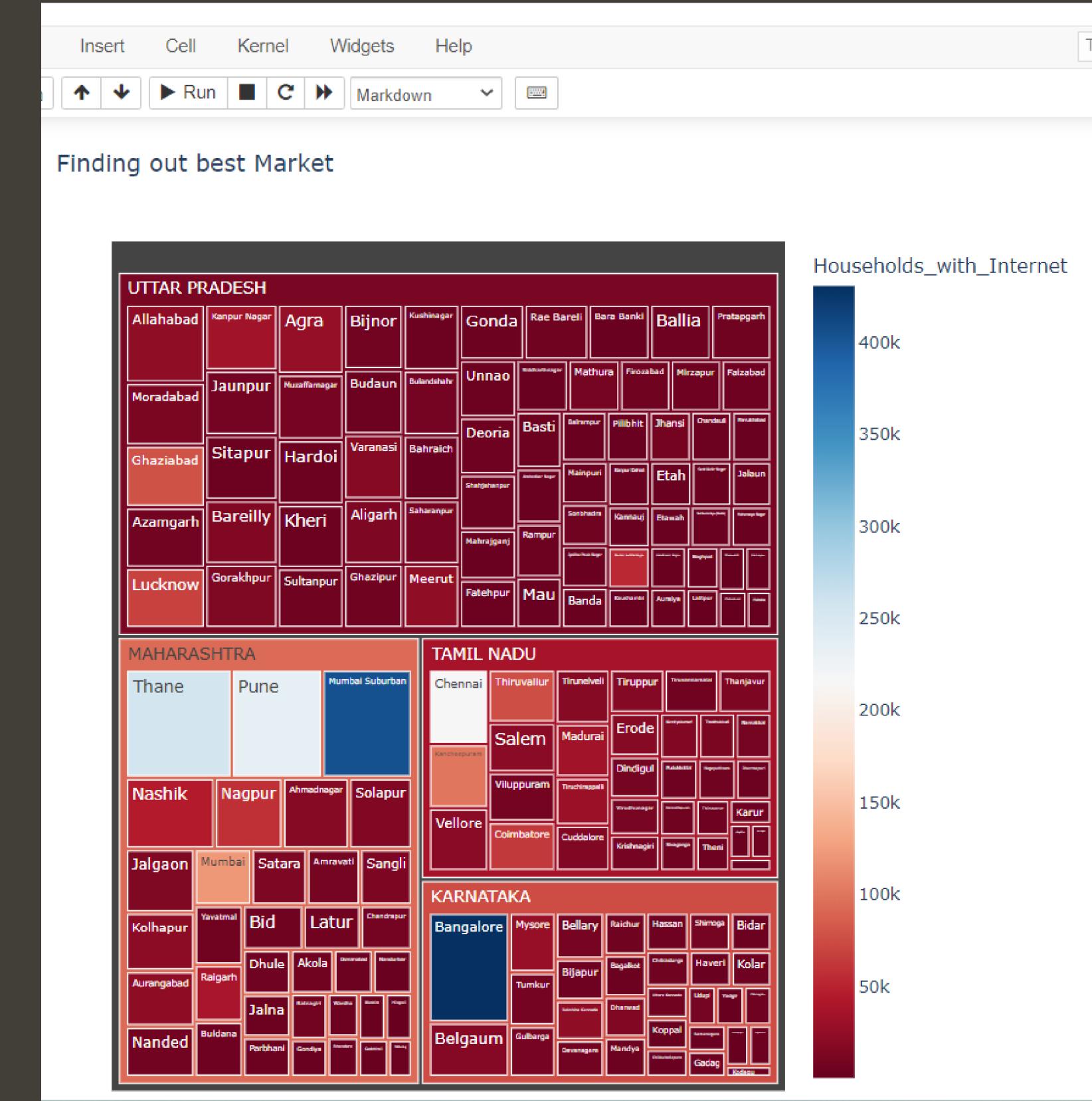
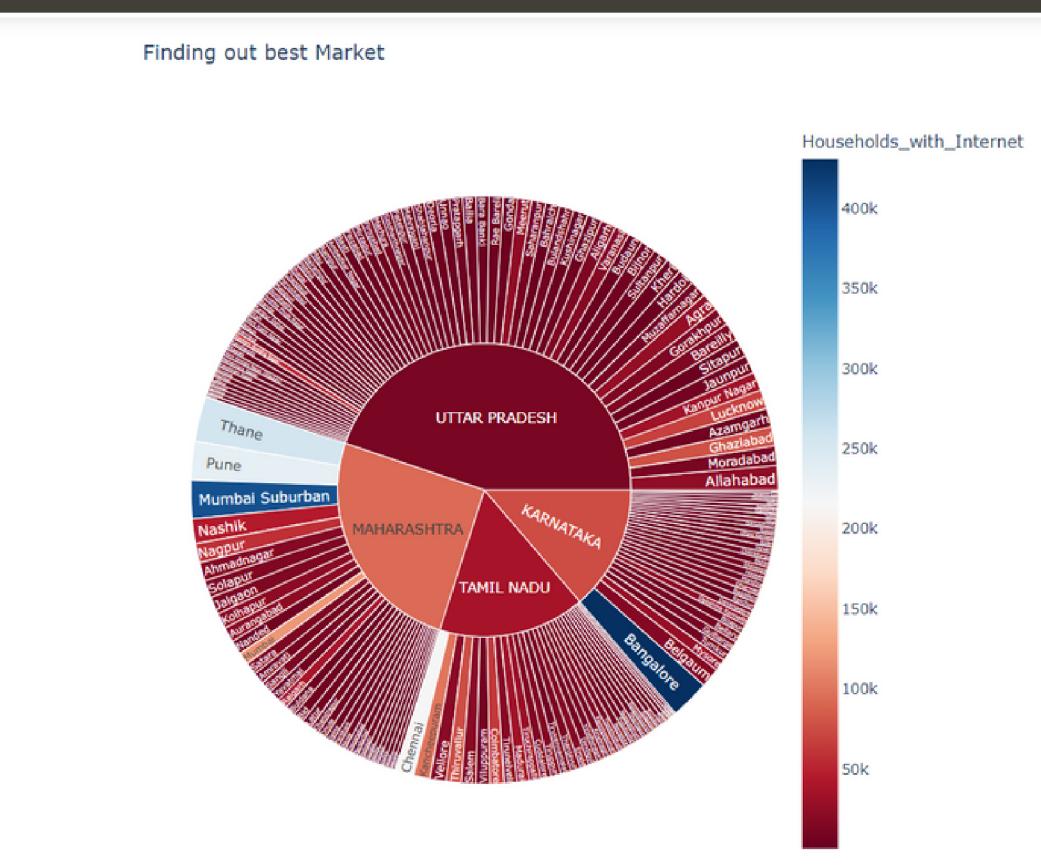
- So there are many states which can be selected for segmentation to start their services in solely based on population count.
- These are states with population greater than 50 Millions and this does not visualize whole scenario it is just a speculation based on Total population count of the above given states.

# Gathered the Explored dataset

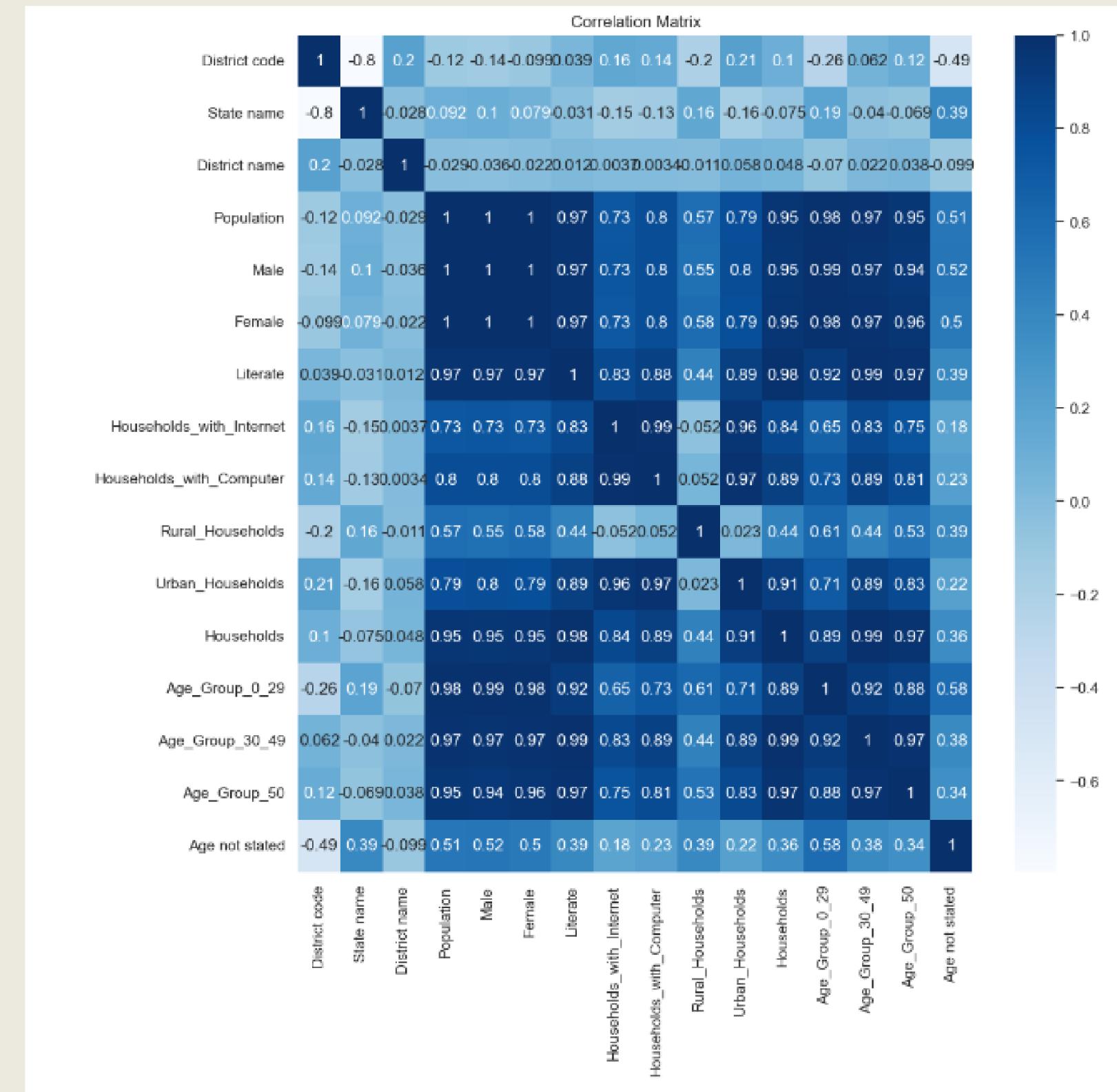
Selected_States = pd.concat([Uttar_Pradesh, Maharashtra, Tamil_Nadu, Karnataka], axis=0)												
Selected_States												
Out[18]:												
District code	State name	District name	Population	Male	Female	Literate	Male_Literate	Female_Literate	Workers	...	Power_Parity_Rs_90000_150000	...
131	132	UTTAR PRADESH	Saharanpur	3466382	1834106	1632276	2077108	1220114	856994	1037344	...	974
132	133	UTTAR PRADESH	Muzaffarnagar	4143512	2193434	1950078	2417339	1448528	968811	1291644	...	1114
133	134	UTTAR PRADESH	Bijnor	3682713	1921215	1761498	2135393	1241471	893922	1088036	...	930
134	135	UTTAR PRADESH	Moradabad	4772006	2503186	2268820	2263848	1357435	906413	1417811	...	1250
135	136	UTTAR PRADESH	Rampur	2335819	1223889	1111930	1043666	630408	413258	737261	...	474
...	...	...	...	...	...	...	...	...	...	...	...	...
579	580	KARNATAKA	Yadgir	1174271	590329	583942	510003	306751	203252	547696	...	563
580	581	KARNATAKA	Kolar	1536401	776396	760005	1016219	564110	452109	717872	...	799
581	582	KARNATAKA	Chikkaballapura	1255104	636437	618667	783222	442158	341064	639778	...	121
582	583	KARNATAKA	Bangalore Rural	990923	509172	481751	688749	385311	303438	459891	...	368
583	584	KARNATAKA	Ramanagara	1082636	548008	534628	674758	378461	296297	531459	...	520

# Finding best market

- Here, we plotted households with internet.
  - Segmented this tree map based on district and correlate with netizens.

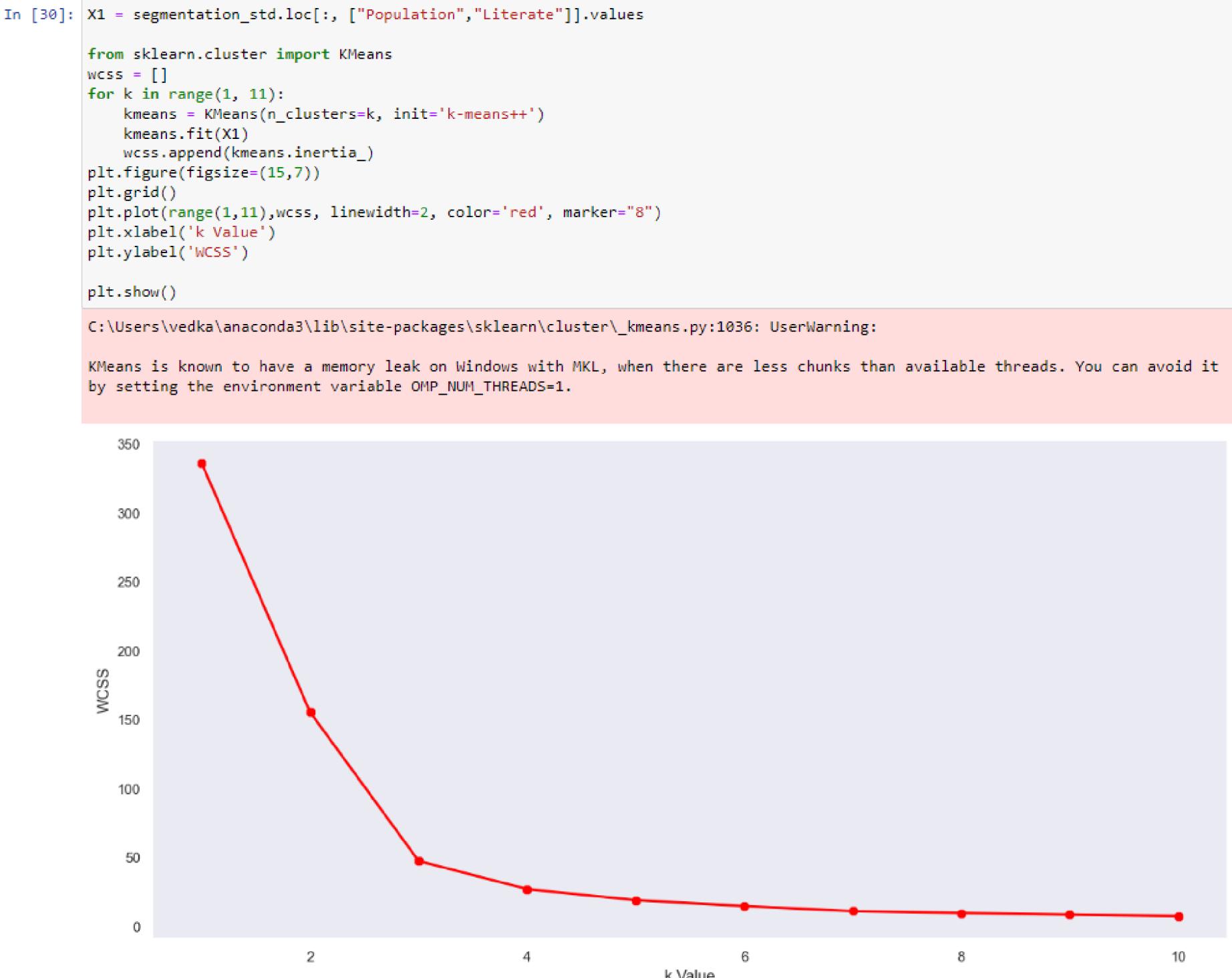


# Correlation Matrix



# Model Creation

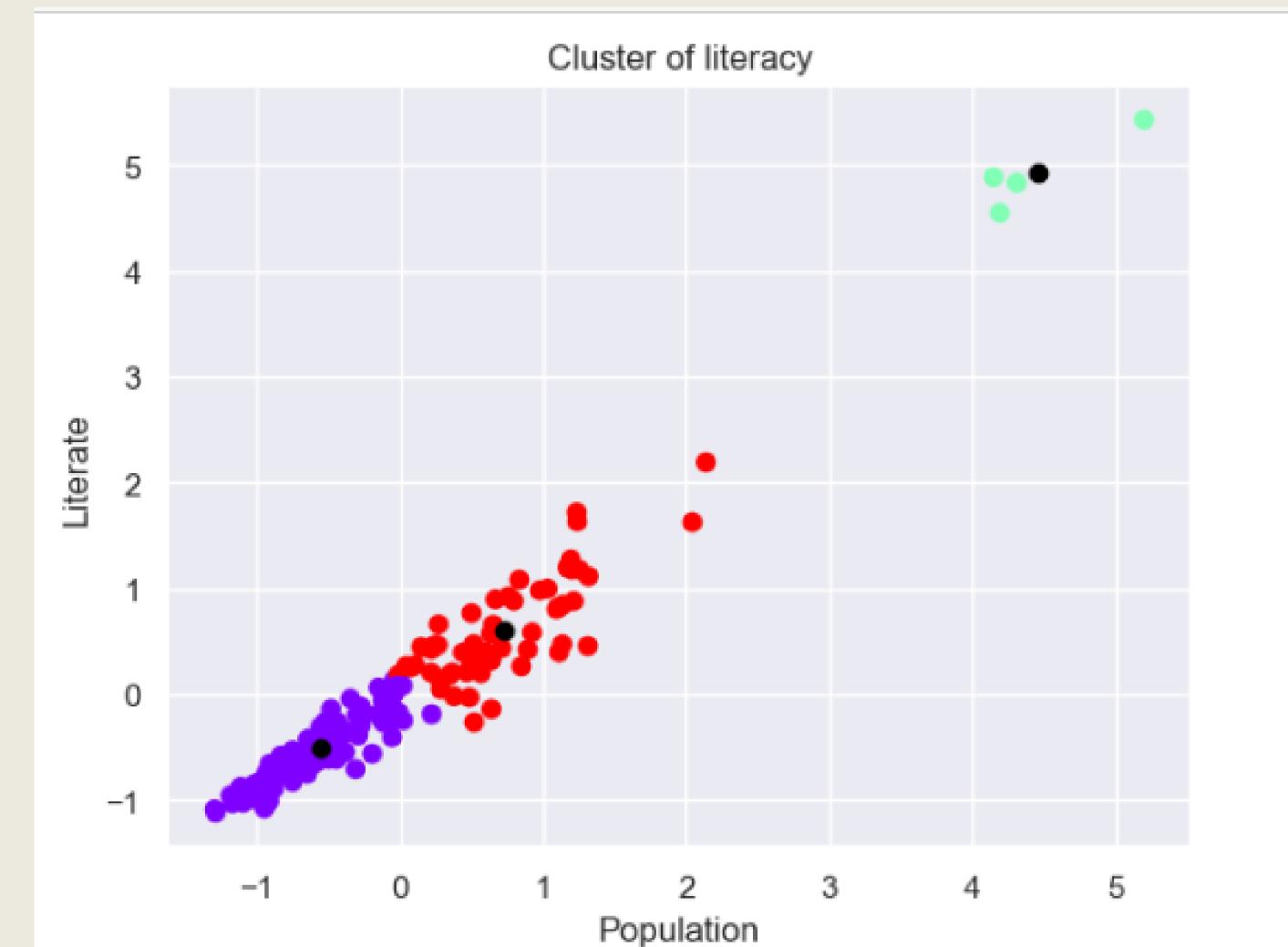
# Finding K value



# Clusters

## Encoding and clusters

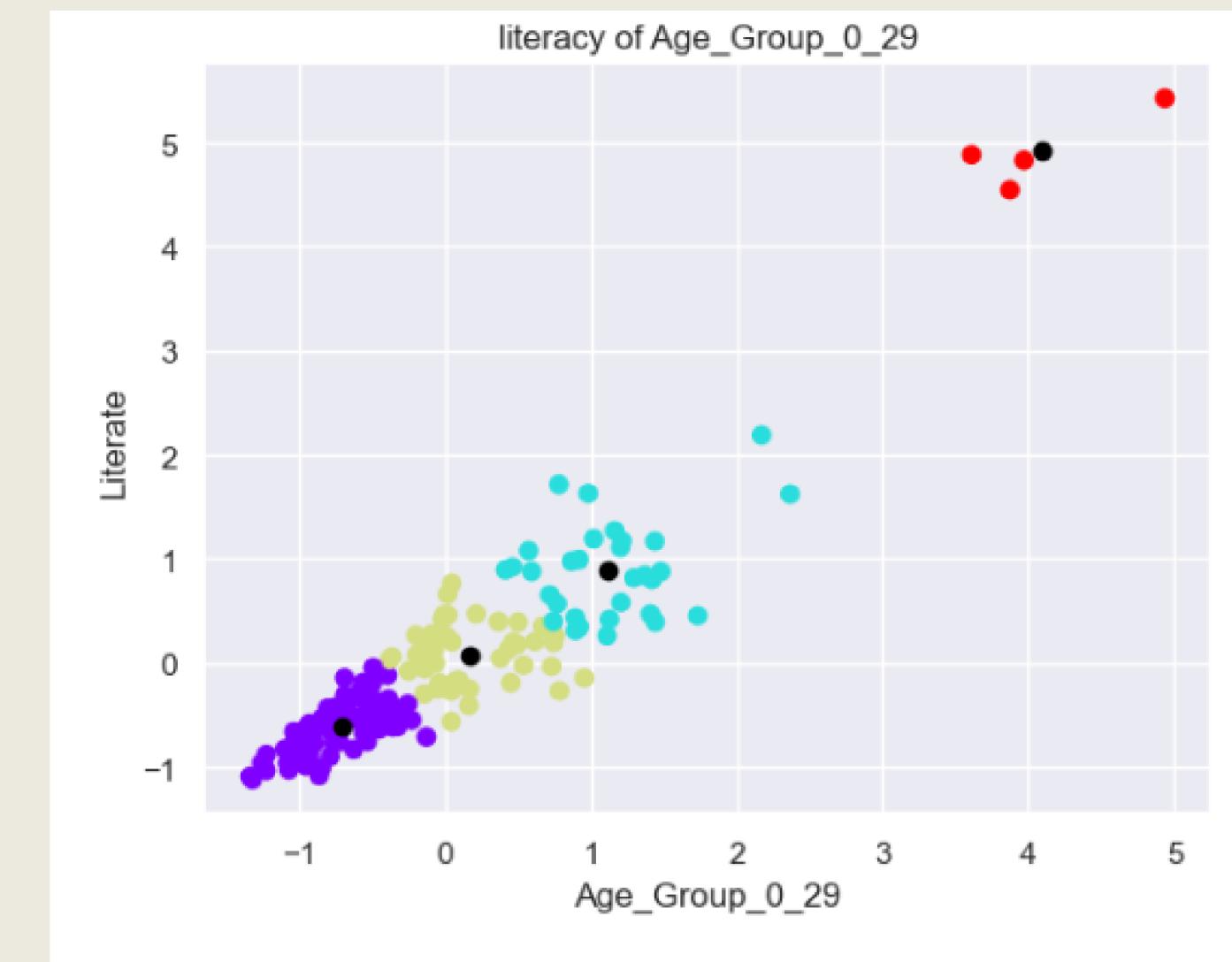
- Why encoding is needed?
  - a. To compare two different parameters we have to encode it into one dimensions.
- Clusters based on literacy
  - b. From this clusters we can identify that literacy rate and population .
  - c. Here, we encode this 2 parameters. so, that's why this data is linearly aligned.



# Clusters

## Based on age group

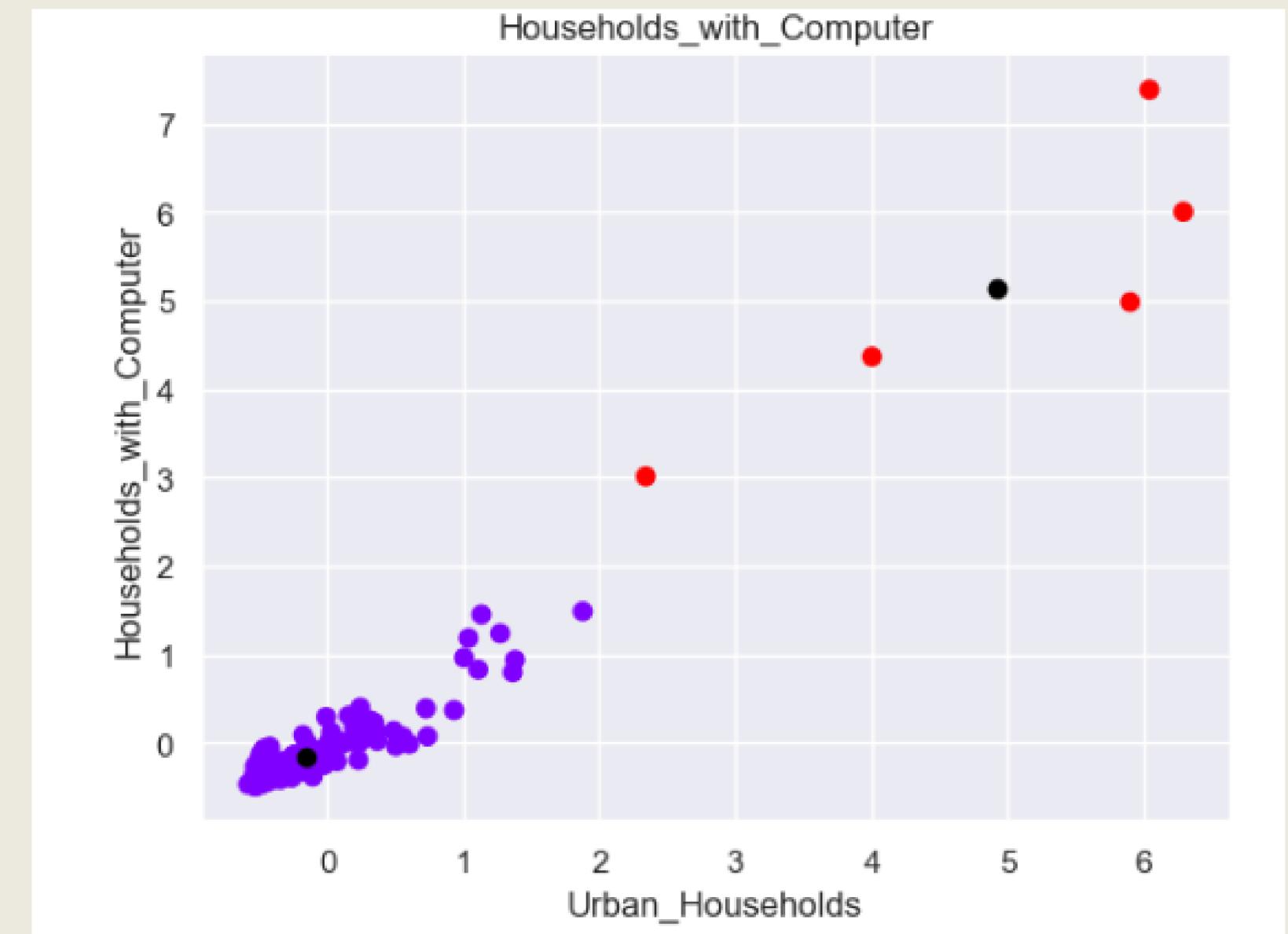
- Here, we created 3 clusters based on literacy and age groups .
- We can target this customers based on their age and literacy.



# Clusters

Based on household in urban area

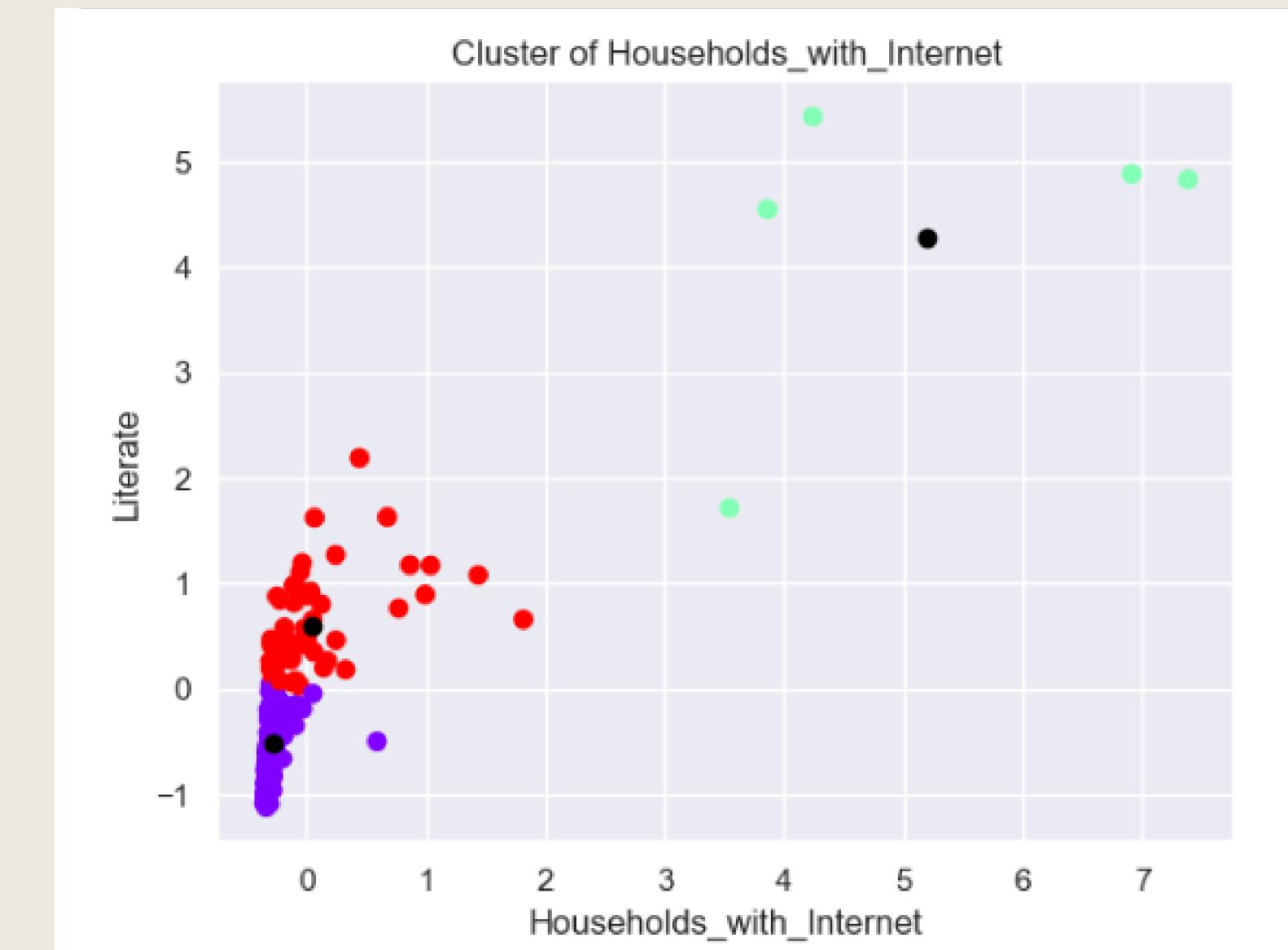
- We can say that very rare number of people in India has computer.
- So, we can't create segment based on this parameters.



# Clusters

## Based on Internet

- Here we segmented data in two clusters that we can target.



Thank you