Market Segment Analysis of Ed-Tech Market in India

Contributor: *Rishi Kumar, Vinay Wadhwa, Owais, TNV Sai Siddhartha*

Problem Statement

The global edtech market is a diverse and rapidly growing industry with a large runway across the business lifecycle, from early-stage startups to middle-market companies to publicly traded companies. It continues to garner a lot of attention not only from long-time industry veterans but from generalist investors as well. Across the three main education categories (PreK-12, post-secondary, and corporate training), technology is infused throughout the life of a learner. Experienced investors know that each category operates as its own sub-segment within the edtech market. Each category is unique and has different end users, buyers, and funding mechanisms.

Due to differences in methodology and categorizations, various publications have released investment figures that can allow us to triangulate and get a sense for the size, scope, and growth of the edtech market:

Data Sources

Open budgets India: https://openbudgetsindia.org/dataset?groups=education

Data world: https://data.world/inderz/india-district-level-school-report-card#

Statista: https://www.statista.com/statistics/1235210/india-edtech-market-size-by-segment/

Kaggle: https://www.kaggle.com/rajanand/education-in-india

data.gov: https://data.gov.in/search/site?query=education

Ibef: https://www.ibef.org/blogs/india-to-become-the-edtech-capital-of-the-world

Geeksforgeeks: https://www.geeksforgeeks.org/

India Data Labs @ Observer Research Foundation, National Sample Survey Organization, Periodic Labor Force Survey (PLFS), July, 2017- June, 2018

Data is information that can be analyzed to make business strategies. With the developing technology, various types of data are continuously being collected. With this vast plethora of data being available, new strategies, and also questions are being raised about privacy and new ethical regulations.

Data Pre-processing

Preprocessing in Data Mining:

Data preprocessing is a data mining technique which is used to transform the raw data in a useful and efficient format.

```
data = pd.read_csv(r"F:\Feynn labs\New folder\data\2015_16_Statewise_Elementary.csv")
data.head()
```

	AC_YEAR	STATCD	STATNAME	DISTRICTS	BLOCKS	VILLAGES	CLUSTERS	TOTPOPULAT	P_URB_POP	POPULATION_0_6	...	USCR35	NOTCH_ASS	TCHINV	TOTCLS1G	TOTC
0	2015-16	1	JAMMU & KASHMIR	22	201	7263	1628	12549	20.05	16.01	...	1014	20491	1946	41171	
1	2015-16	2	HIMACHAL PRADESH	12	124	10120	2243	6857	8.69	11.14	...	652	36054	3053	38307	
2	2015-16	3	PUNJAB	22	146	13197	1780	27704	29.82	10.62	...	2754	88618	14200	56222	
3	2015-16	4	CHANDIGARH	1	20	84	20	1055	97.25	11.18	...	117	1994	218	135	
4	2015-16	5	UTTARAKHAND	13	95	11989	995	10117	21.54	13.14	...	1147	12608	1024	48215	

5 rows × 816 columns

Python libraries such as **NumPy**, **Pandas**, **Scikit-learn** and **SciPy** are used for the workflow and the results obtained are ensured to be reproducible

```
Data. Columns
```

Out [5]:

```
Index (['AC_YEAR', 'STATCD', 'STATNAME', 'DISTRICTS', 'BLOCKS', 'VILLAGES',
       'CLUSTERS', 'TOTPOPULAT', 'P_URB_POP', 'POPULATION_0_6',
       ...
       'USCR35', 'NOTCH_ASS', 'TCHINV', 'TOTCLS1G', 'TOTCLS2G', 'TOTCLS3G',
       'TOTCLS4G', 'TOTCLS5G', 'TOTCLS6G', 'TOTCLS7G'],
      dtype='object', length=816)
```

.

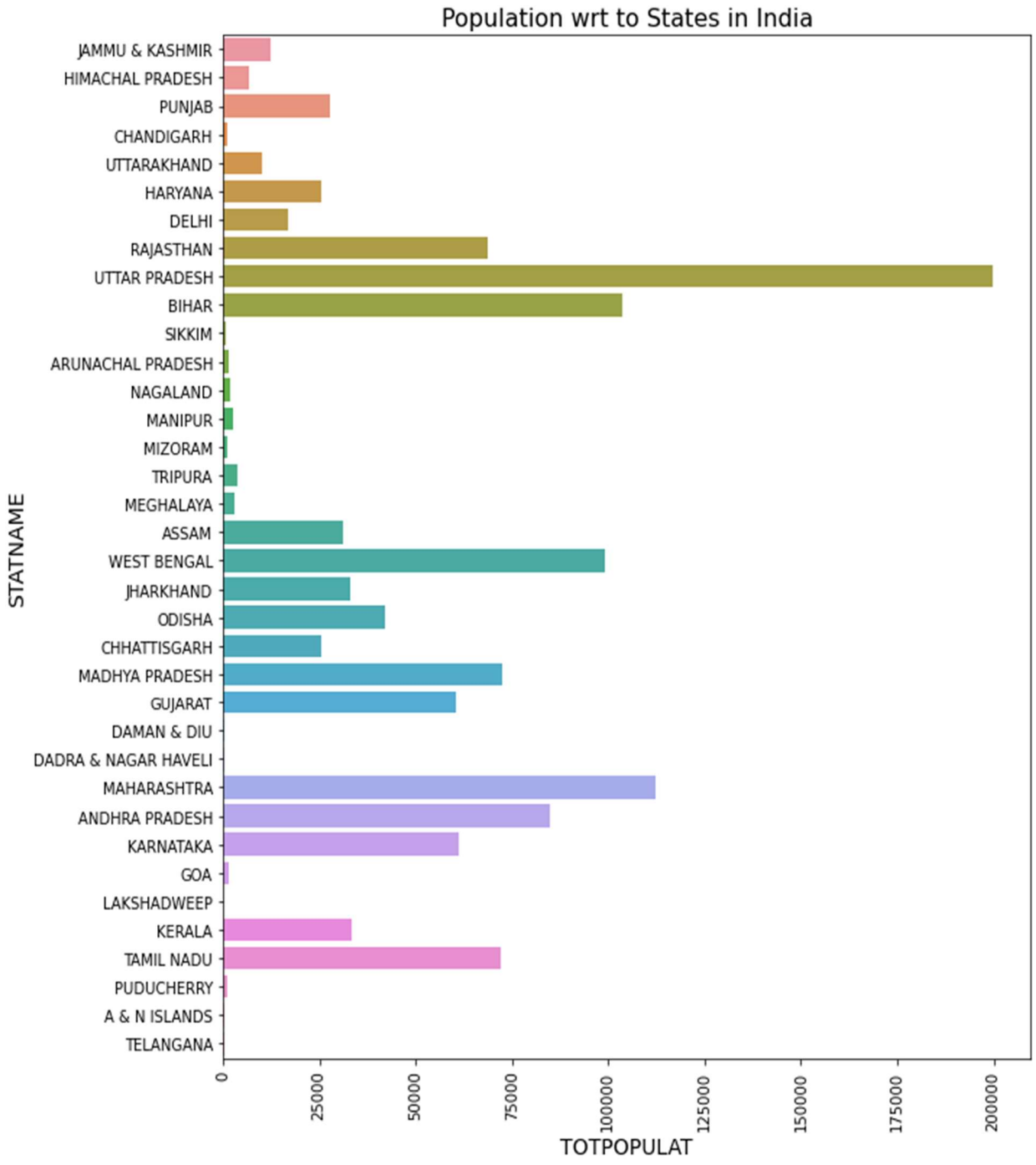
For check the null values:

```
data.isnull ().sum ()
```

Out [9]:

```
AC_YEAR       0
STATCD        0
STATNAME      0
DISTRICTS     0
BLOCKS        0
             ..
TOTCLS3G      0
TOTCLS4G      0
TOTCLS5G      0
TOTCLS6G      0
TOTCLS7G      0
Length: 816, dtype: int64
```

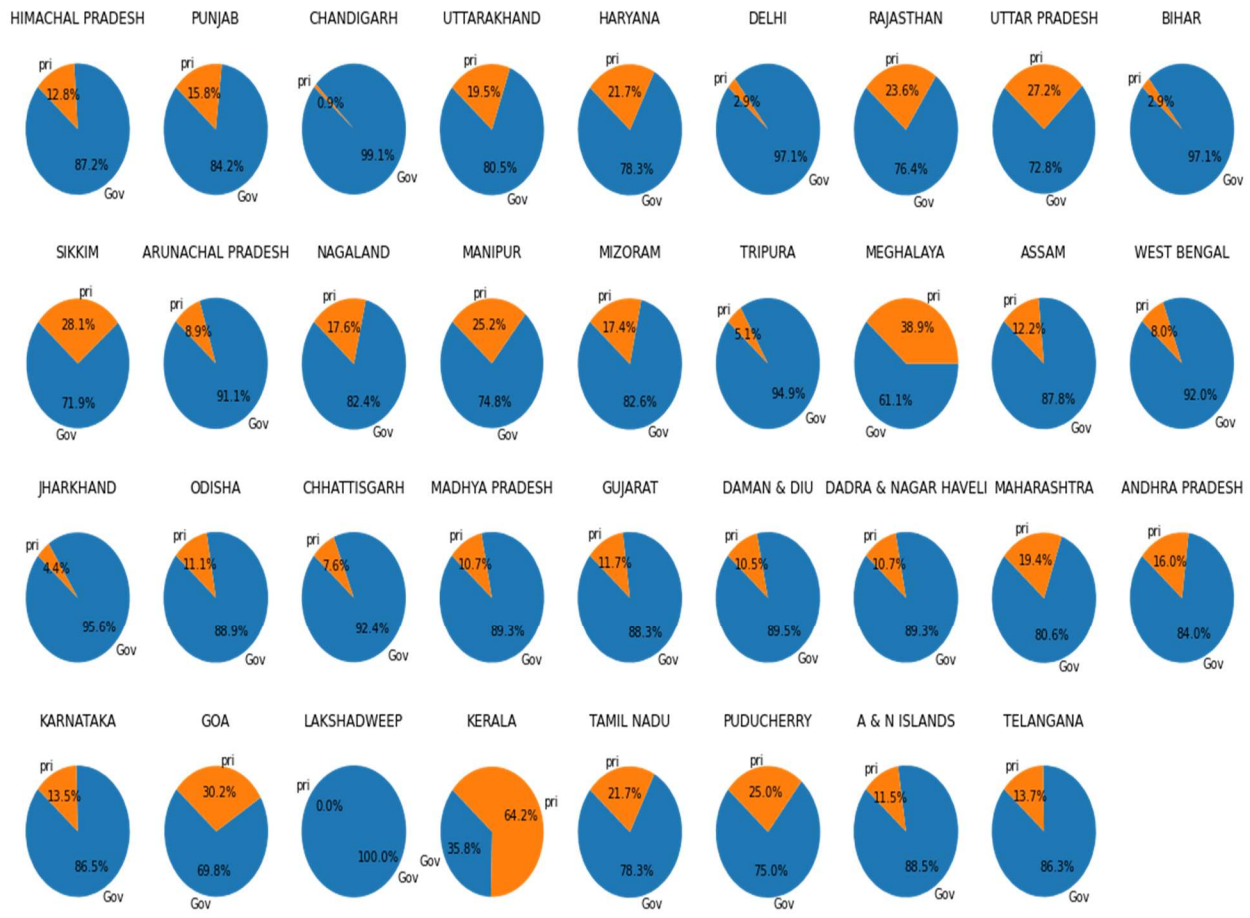
Seaborn is a high level interface for drawing statistical graphics with **Matplotlib.** It aims to make visualization a central part of exploring and understanding complex datasets.



Private vs public education:

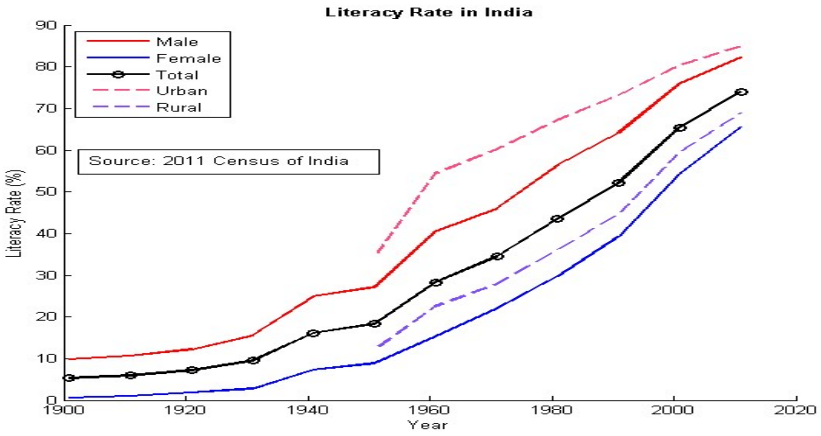
The debate around public vs private education never gets old. There is a set of experts who think that private education should be encouraged and it will lead to more opportunities and healthy competition. The opponents of this school of thought believe education must be a government subject and provided free for all.

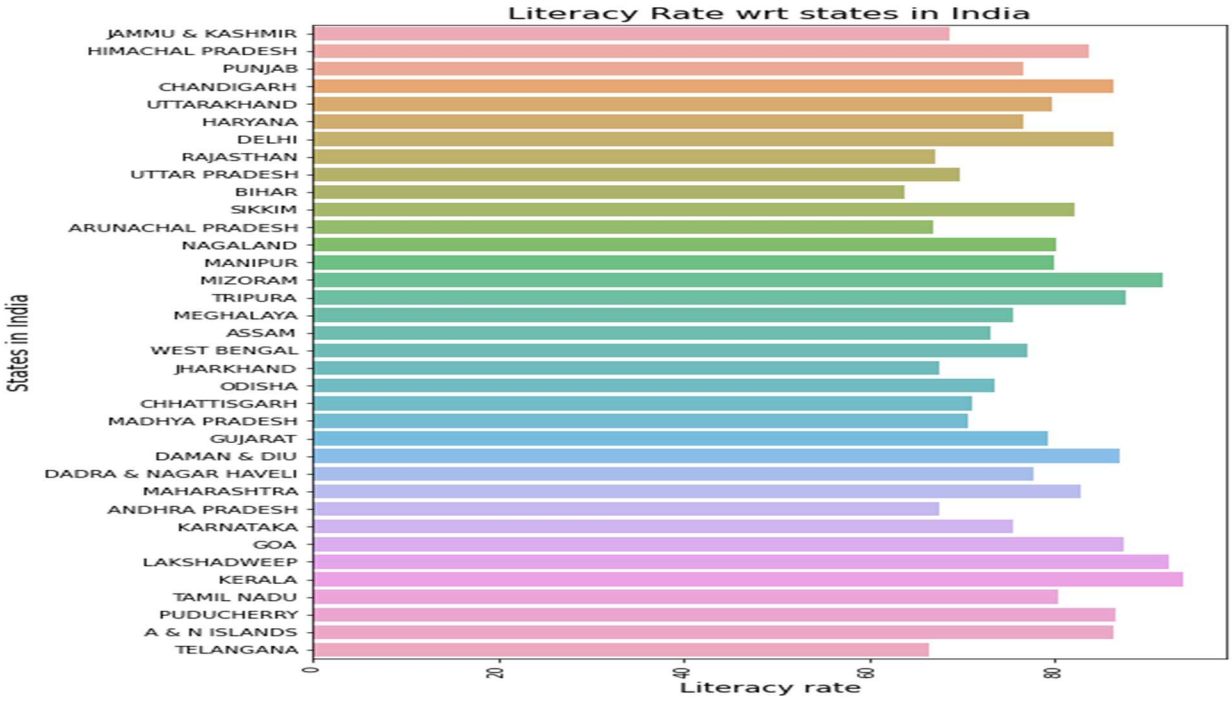
```
plt.figure(figsize=(20,12))
for i in range(1,len(data1)):
    plt.subplot (4, 9, i)
    plt.title (data1 ['STATNAME'][i])
    top = ['Gov','pri']
    uttar = data1.loc[data1['STATNAME'] == data1['STATNAME'][i],:]
    value
=[float(uttar['SCHTOTG']/uttar['SCHTOT'])*100,float(uttar['SCHTOTPR']/uttar['
SCHTOT'])*100]
    plt.pie(value, labels=top, autopct='%1.1f%%',startangle=140)
    plt.axis('equal')
plt.show ()
```



Literacy rate of India

India's literacy rate is at **75%**. Kerala has achieved a literacy rate of 93%. Bihar is the least literate state in India, with a literacy of 63.82%.





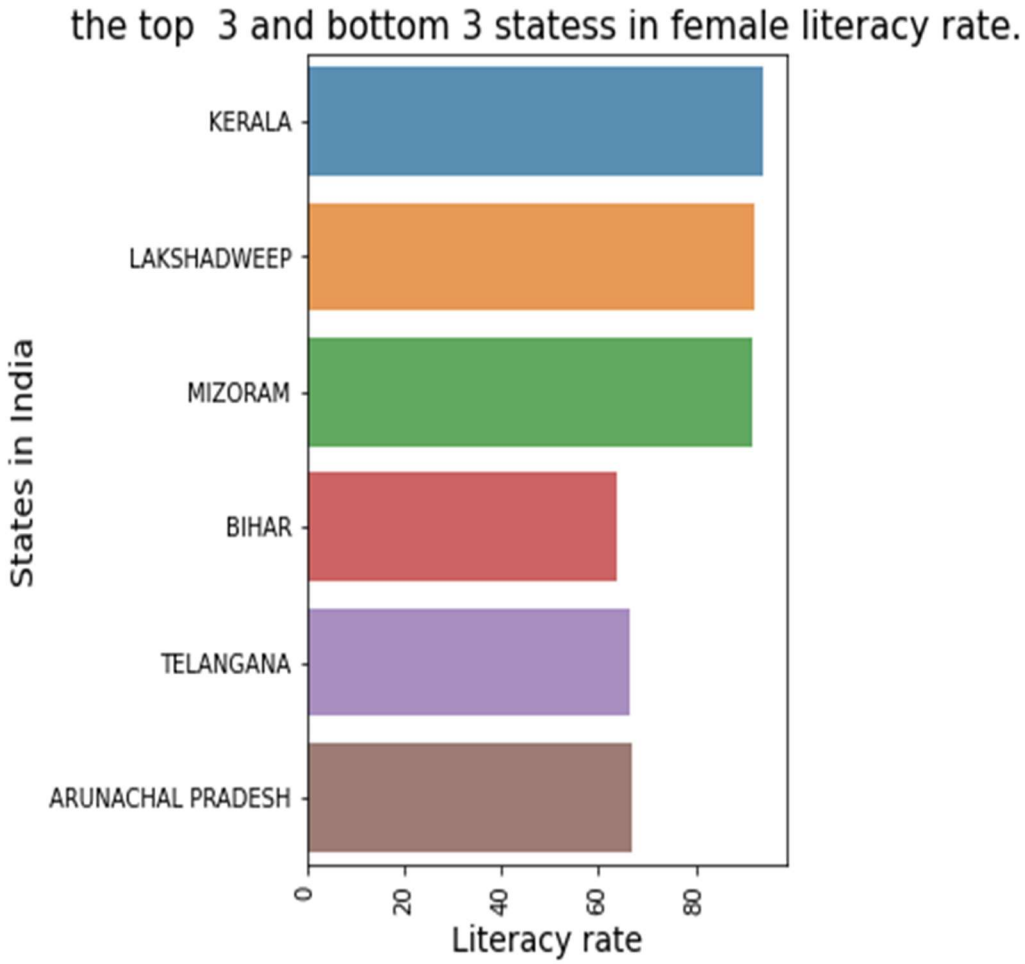
Let us find the difference between the top 3 and bottom 3 states in female literacy rate.

```
top_3=data1.sort_values (by='OVERALL_LI', ascending=False).head (3)
bottom_3=data1.sort_values (by='OVERALL_LI', ascending=True).head (3)
```

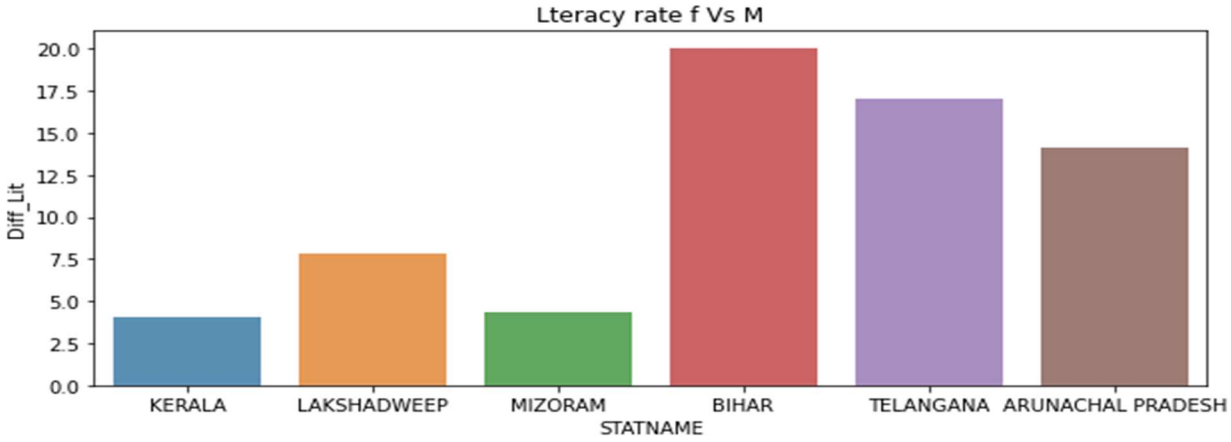
LET US CONCATENATE TOP AND BOTTOM STATES

```
top_bottom = pd.concat ([top_3, bottom_3], axis=0, sort=False)
top_bottom
```

Today, the female literacy levels according to the Literacy Rate 2011 census are 65.46% where the male literacy rate is over 80%.



According to National Statistical Office (NSO) data **India's average literacy rate is 77.70% & male literacy at the India level in 2021 stands at 84.70% & female literacy stands at 70.30%**



1. With low literacy states have high differences.

2. Bottom states have high male literacy rates but female literacy rates on these states is low,

3. Which makes over all literacy rate down.

Overall Gender Gap in Literacy

About three decades ago, the adult male literacy rate in India was almost twice that for adult females. While this gap has narrowed substantially over the years, adult male literacy rate still surpasses the adult female literacy rate by 17 percentage points. (Table 1.)

Figure 1: Trends in male and female literacy

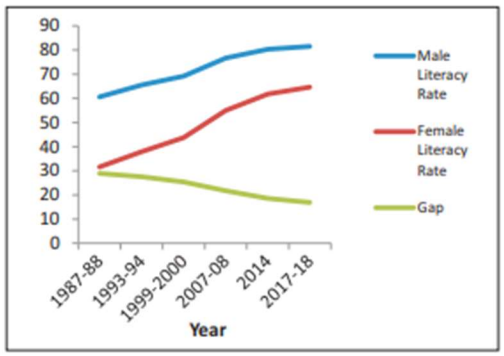


Table 1: Trends in male and female literacy

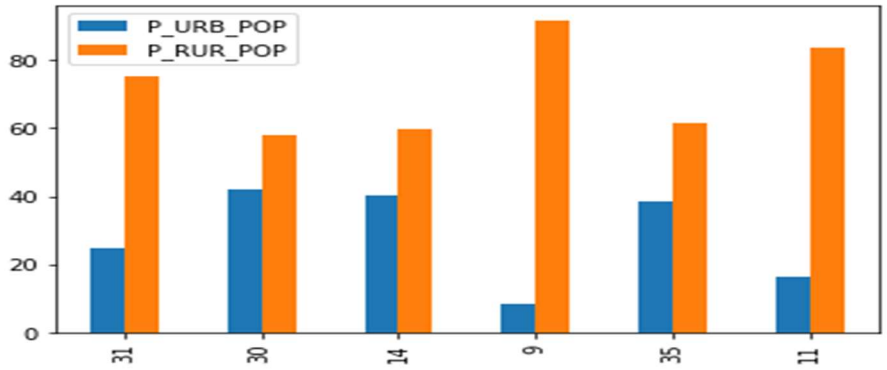
	1987-88	1993-94	1999-00	2007-08	2014	2017-18
Male Literacy Rate	60.6	65.5	69.2	76.6	80.3	81.5
Female Literacy Rate	31.6	37.9	43.8	54.9	61.8	64.6
Gap	28.9	27.6	25.4	21.7	18.5	16.9

Source: National Sample Survey @ Observer Research Foundation's India Data Labs

Literary Rate Urban Vs Rural 20

The literacy rate among persons (aged 7 years and above) in India was about 77.7%. In rural areas, literacy rate was 73.5% compared to 87.7% in urban areas.

```
#3.Let us check urban and rural population in terms of literacy rate.
#we don't have data of rural population. Therefore, rural=100-urban
```



States	Rural Literacy Rate			Urban Literacy Rate		
	Male	Female	Average	Male	Female	Average
Andhra Pradesh	67.5	53.4	60.4	86.3	73.1	79.6
Assam	89.4	79.9	84.9	96.1	91.4	93.8
Bihar	78.6	58.7	69.5	89.3	75.9	83.1
Chhattisgarh	84	65.6	75	91.8	82.3	87.2
Delhi				94.1	83.4	89.4
Gujarat	85.7	68	77	95.2	86.3	91.1
Haryana	85.8	66.4	77	92.5	81.2	87.3
Himachal Pradesh	92.3	79.2	85.6	97.8	93	95.5
Jammu & Kashmir	84.9	66	75.8	88.5	75.7	82.6
Jharkhand	80.6	61.4	71.4	92.6	78.6	86.1
Karnataka	78.2	63.1	71	92.5	83.7	88.3
Kerala	96.7	94.1	95.4	98.2	96.4	97.3
Madhya Pradesh	77.9	61	69.8	91.4	79.5	85.8
Maharashtra	87	71.4	79.4	95.3	87.6	91.7
Odisha	82	67.3	74.9	94.4	85.9	90.2
Punjab	85.5	74	80	93.8	86.7	90.5
Rajasthan	77.6	52.6	65.5	91.1	74.6	83.5
Tamil Nadu	84.2	70.8	77.5	92.3	85.9	89
Telangana	70.6	53.7	62.1	91.7	79	85.5
Uttarakhand	93.1	79	86.1	97.4	85.9	92
Uttar Pradesh	80.5	60.4	70.8	86.8	74.9	81.2
West Bengal	82	72.6	77.4	91.4	84.7	88.1
All-India	**81.5**	**65**	**73.5**	**92.2**	**82.8**	**87.7**

Completed educational level of population for different age-groups

Percentage distribution of rural persons (ages 15 years & above by highest completed levels of education)

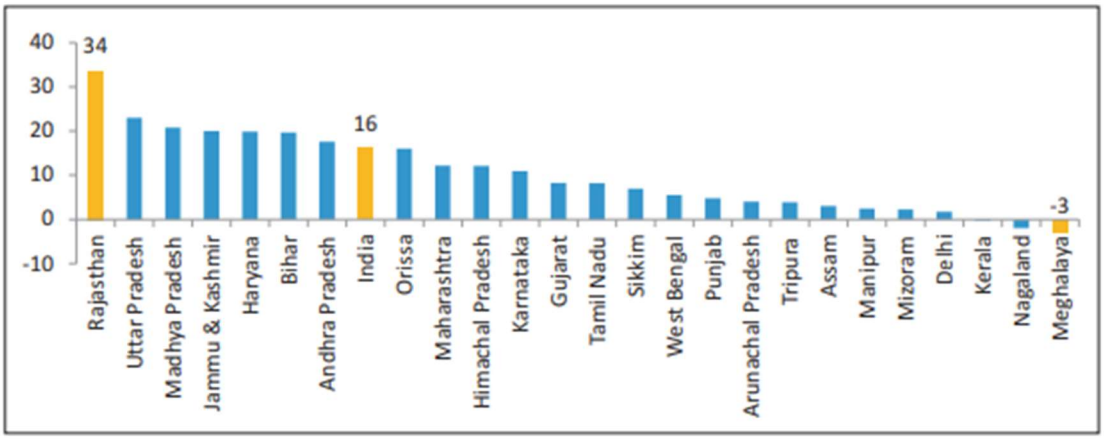
- 31.5% were not literate,
- 20.9% were literates up to primary
- 17.2% were of level upper primary/middle,
- 24.9% were of levels secondary and higher secondary
- 5.7% were graduates & above.

Percentage distribution of urban persons (age 15 years & above by highest completed level of education)

- 13.9% were not literate,
- 14.7% were literates up to primary,
- 14.0% were of level upper primary/middle,
- 35.8% were of level secondary and higher secondary
- 21.7% were graduate& above

Source: survey by National Statistical Office (NSO)

Figure 8: Reduction in the Literacy Gender Gap over 30 Years in the Age Group 6-14 Years (1987-88 to 2017-18)

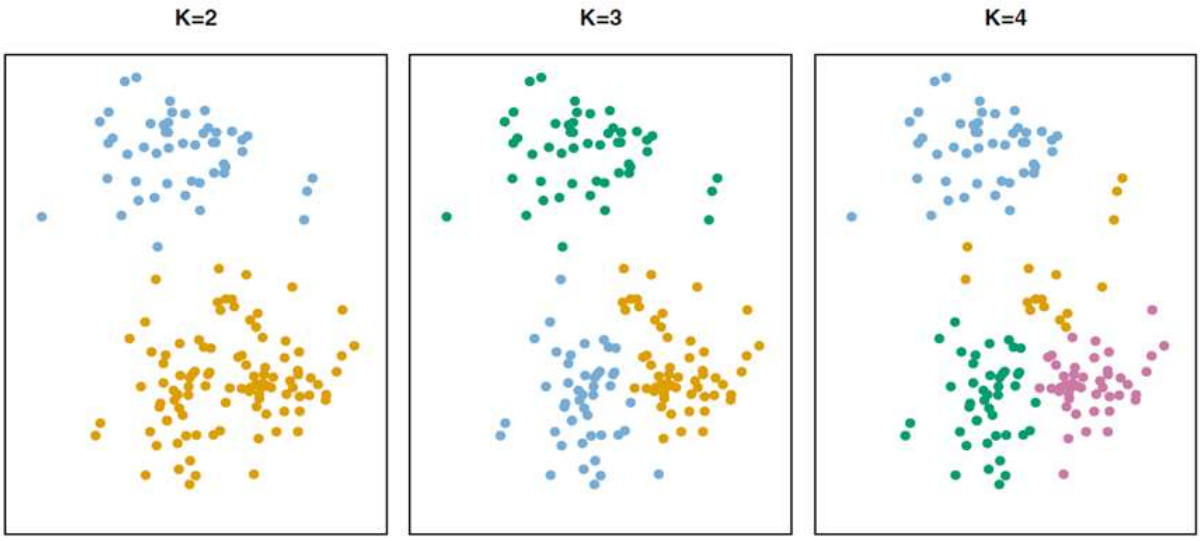


Source: National Sample Survey @ Observer Research Foundation's India Data Labs

Segment Extraction

What is **K-Means algorithm** and how it works

K-means clustering is a simple and elegant approach for partitioning a data set into K distinct, no overlapping clusters. To perform K-means clustering, we must first specify the desired number of clusters K; then, the K-means algorithm will assign each observation to exactly one of the K clusters. The below figure shows the results obtained from performing K-means clustering on a simulated example consisting of 150 observations in two dimensions, using three different values of K.

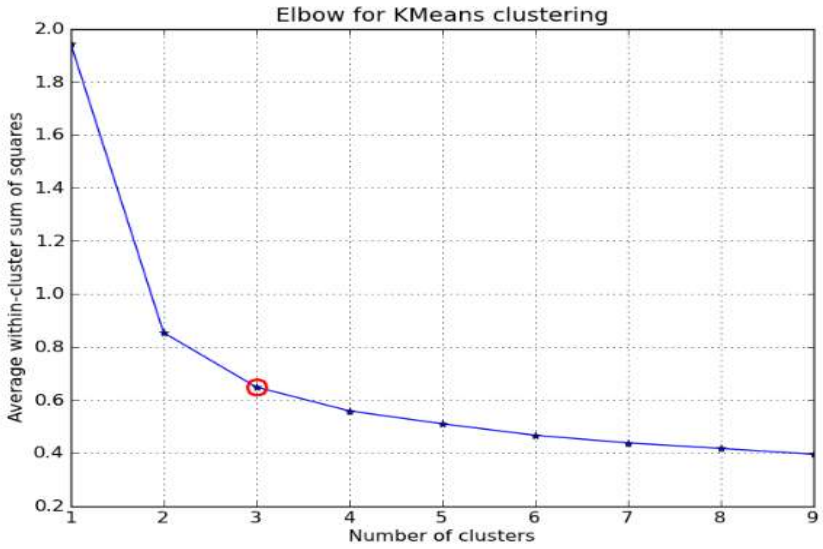


What is K-Means Clustering?
K-Means clustering is an unsupervised learning algorithm. There is no labelled data for this clustering, unlike in supervised learning. K-Means performs the division of objects into clusters that share similarities and are dissimilar to the objects belonging to another cluster.

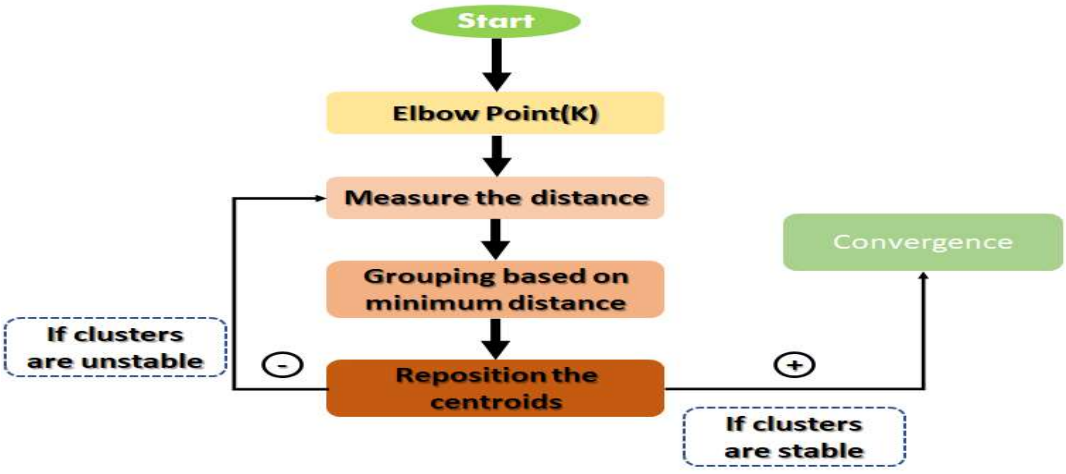
$$\underset{C_1,\ldots,C_k}{minimize}\left\{\sum_{k=1}^{K}\frac{1}{|C_k|}\sum_{i,i'\in C_k}\sum_{j=1}^{p}\left(x_{ij}-x_{i'j}\right)^2\right\}$$

Elbow Method

The elbow method involves finding a metric to evaluate how good a clustering outcome is for various values of K and finding the **elbow point**. Initially, the quality of clustering improves rapidly when changing the value of K but eventually stabilizes. The elbow point is the point where the relative improvement is not very high anymore. This is shown pictorially in the two graphs below for the metric average within-cluster sum of squares.



The flowchart below shows how k-means clustering works:



Profiling and describing potential segments

He we using python libbers for uploading data set and make a model we use machine learning technics (K-means clustering)

In [16]:
```
import plotly.express as px
```

In [45]:
```
import pandas as pd
import numpy as np
import matplotlib as mpl
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import scipy.stats as stats
from sklearn import ensemble, tree, linear_model, preprocessing
import missingno
import missingno as msno
```

In [46]:
```
District_wise = pd.read_csv(r'F:\Feynn labs\New folder\data\2015_16_Districtwise.csv')
State_wise_elementry = pd.read_csv(r'F:\Feynn labs\New folder\data\2015_16_Statewise_Elementary.csv')
State_wise_secondary = pd.read_csv(r'F:\Feynn labs\New folder\data\2015_16_Statewise_Secondary.csv')
```

In [47]:
```
District_wise_met = pd.read_csv(r'F:\Feynn labs\New folder\data\2015_16_Districtwise_Metadata.csv')
State_wise_elementry_met = pd.read_csv(r'F:\Feynn labs\New folder\data\2015_16_Statewise_Elementary_Metadata.csv')
State_wise_secondary_met = pd.read_csv(r'F:\Feynn labs\New folder\data\2015_16_Statewise_Secondary_Metadata.csv')
```

In District data set we select the columns and plot the scatter plot by using scatter Metrix libbers.

]:
```
District_wise_fac.drop(['Schools_with_Ramp_(where_needed): Total','Schools_where_Ramp_is_Required: Total'], axis=1, inplace=True)
```

]:
```
District_wise_fac.head()
```

]:

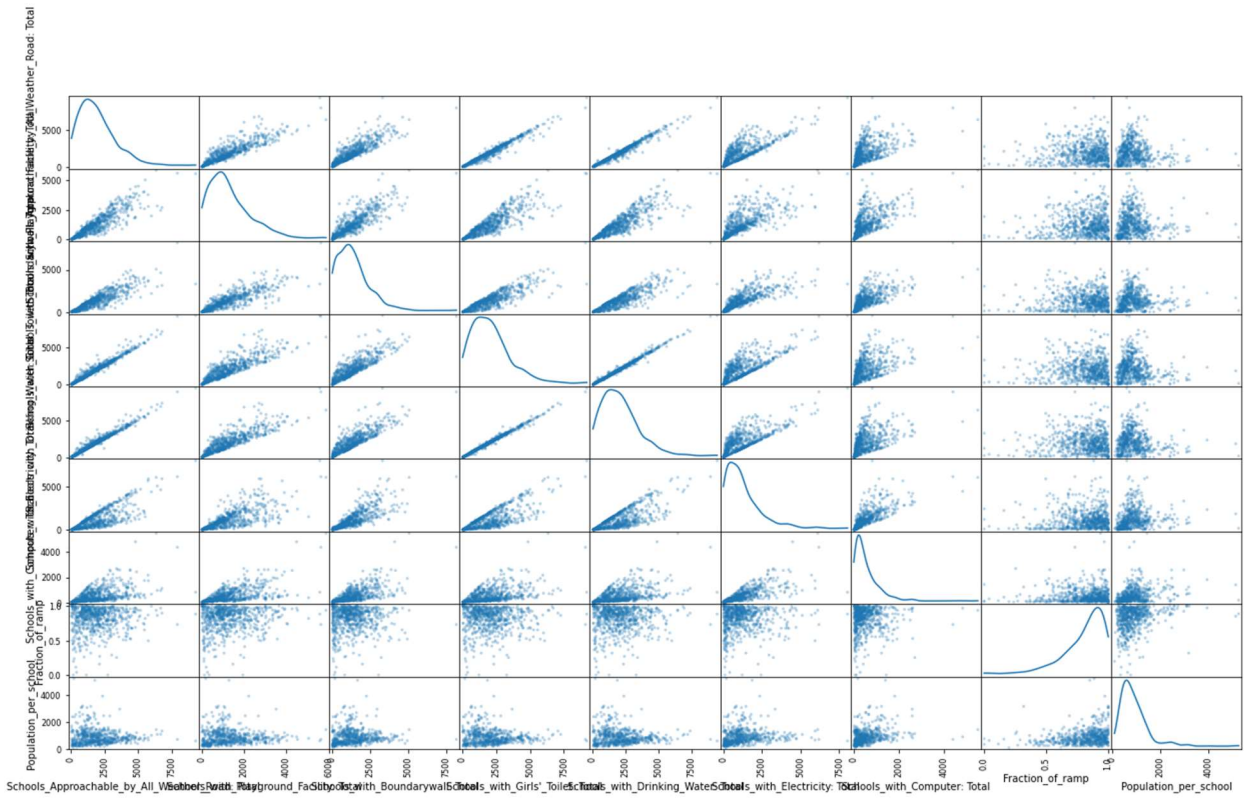
	Schools_Approachable_by_All_Weather_Road: Total	Schools_with_Playground_Facility: Total	Schools_with_Boundarywall: Total	Schools_with_Girls'_Toilet: Total	Schools_with_Drinking_Water: Total	Schools_with_El
0	1611	424	560	1897	1678	
1	2071	761	595	2296	2109	
2	967	607	836	843	970	
3	1438	490	520	1459	1468	
4	965	367	435	889	989	

]:
```
District_wise_fac['Population_per_school'] = District_wise_total['Basic_data_from_Census_2011: Total_Population(in_1000\'s)']/District_wise_total['Scho
```

]:
```
District_wise_fac.drop('Schools_with_Boys\'_Toilet: Total', axis=1, inplace = True)
```

]:
```
pd.plotting.scatter_matrix(District_wise_fac, alpha = 0.3, figsize = (21,12), diagonal = 'kde')
```

]:
```
array([[<AxesSubplot:xlabel='Schools_Approachable_by_All_Weather_Road: Total', ylabel='Schools_Approachable_by_All_Weather_Road: Total'>,
        <AxesSubplot:xlabel='Schools_with_Playground_Facility: Total', ylabel='Schools_Approachable_by_All_Weather_Road: Total'>,
        <AxesSubplot:xlabel='Schools_with_Boundarywall: Total', ylabel='Schools_Approachable_by_All_Weather_Road: Total'>,
```



Draw a scatterplot where one variable is categorical.

A strip plot can be drawn on its own, but it is also a good complement to a box or violin plot in cases where you want to show all observations along with some representation of the underlying distribution.

In [59]:
```
Ramp = District_wise_fac['Fraction_of_ramp']
```

In [60]:
```
District_wise_fac.drop('Fraction_of_ramp', axis=1, inplace=True)
```

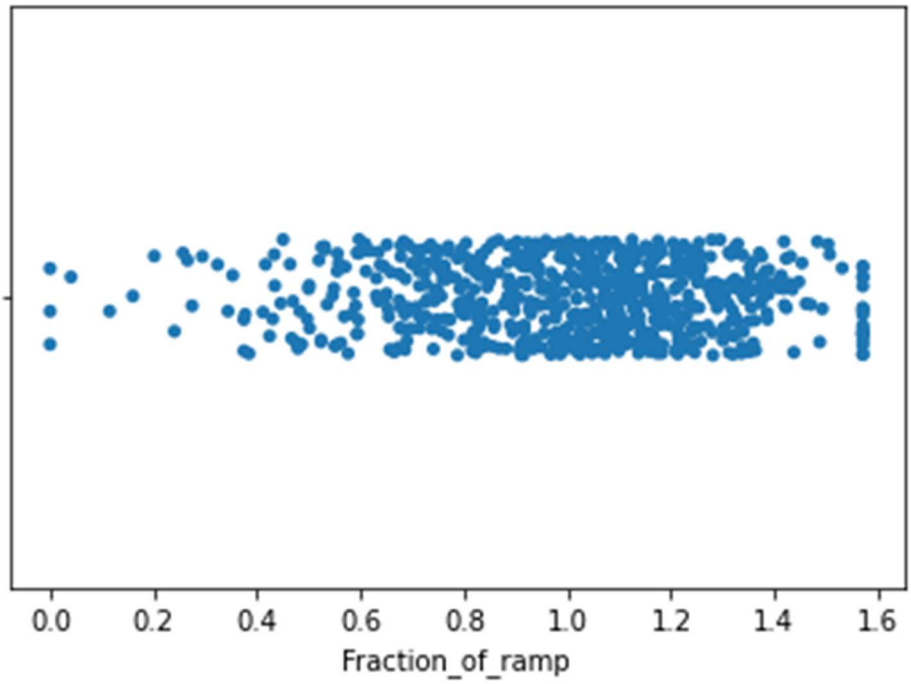
In [ ]:
```
for name in District_wise_fac.columns:
    District_wise_fac[name]=np.log(District_wise_fac[name])
```

In [61]:
```
Ramp = np.arcsin(Ramp)
```

In [62]:
```
sns.stripplot(Ramp)
```

```
C:\Users\rishi\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12,
the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
<AxesSubplot:xlabel='Fraction_of_ramp'>
```

Fraction of ramp District wise.



Description:

A strip plot is a graphical data analysis technique for summarizing a univariate data set. The strip plot consists of:

Horizontal axis = the value of the response variable;
Vertically axis = all values are set to 1.

That is, a strip plot is simply a plot of the sorted response values along one axis. The strip plot is an alternative to a histogram or a density plot. It is typically used for small data sets (histograms and density plots are typically preferred for larger data sets).

There are a few variations supported for the strip plot.

One problem with strip plots is how to display multiple points with the same value. Data plot provides two options to address this.

1. With the jitter option, a small amount of random noise is added to the vertical coordinate.
2. With the stack option, repeated values add a fixed increment to the vertical coordinate. So if there are 3 points with the same value, the y coordinates might be 1, 1.1, and 1.2. This gives the strip plot a histogram-like appearance.

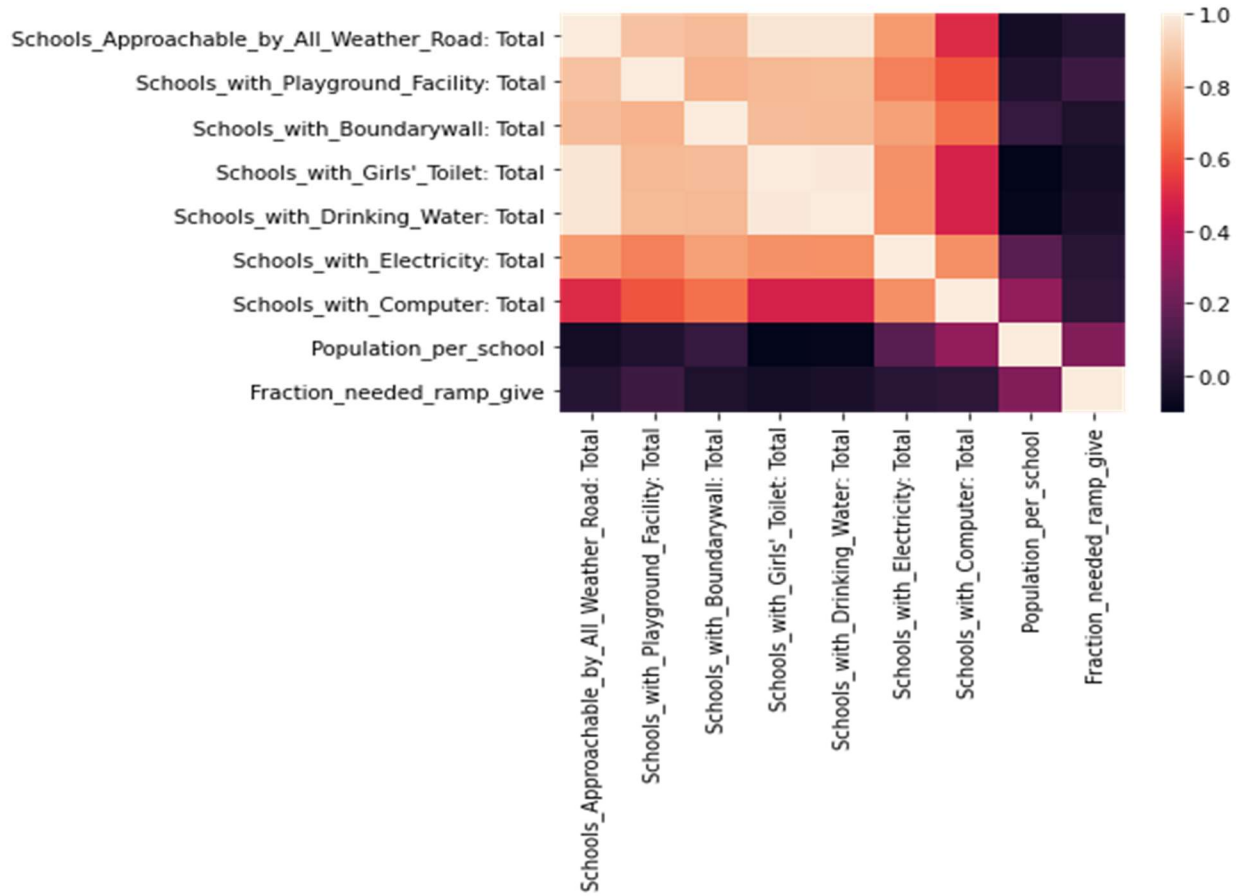
```
n [72]: District_wise_scaled.describe()
```

ut[72]:

	Schools_Approachable_by_All_Weather_Road: Total	Schools_with_Playground_Facility: Total	Schools_with_Boundarywall: Total	Schools_with_Girls'_Toilet: Total	Schools_with_Drinking_Water: Total	Schools_wi
count	6.800000e+02	6.800000e+02	6.800000e+02	6.800000e+02	6.800000e+02	
mean	-5.469481e-17	8.979745e-17	1.358391e-16	-1.005731e-16	-1.611456e-16	
std	1.000736e+00	1.000736e+00	1.000736e+00	1.000736e+00	1.000736e+00	
min	-1.417108e+00	-1.328631e+00	-1.325540e+00	-1.455482e+00	-1.455887e+00	-
25%	-7.357954e-01	-7.372966e-01	-7.337602e-01	-7.483869e-01	-7.388136e-01	
50%	-1.757891e-01	-2.383261e-01	-1.686990e-01	-1.503518e-01	-1.325819e-01	
75%	5.266204e-01	5.395417e-01	5.029914e-01	5.050831e-01	5.385459e-01	
max	5.563663e+00	4.717840e+00	6.722870e+00	5.119675e+00	5.278274e+00	

```
n [73]: District_wise_scaled = District_wise_scaled[District_wise_scaled['Population_per_school'].notna()]

n [74]: sns.heatmap(District_wise_scaled.corr())
```

Heatmap is defined as a graphical representation of data using colors to visualize the value of the matrix. In this, to represent more common values or higher activities brighter colors basically reddish colors are used and to represent less common or activity values, darker colors are preferred. Heatmap is also defined by the name of the shading matrix. Heatmaps in Seaborn can be plotted by using the seaborn.heatmap () function.



PCA is a linear dimensionality reduction technique. It transforms a set of correlated variables (p) into a smaller k (k<p) number of uncorrelated variables called ***principal components*** while retaining as much of the variation in the original dataset as possible.

In [80]:
```
#Fitting the PCA algorithm with our Data
from sklearn.decomposition import pca
#pca = PCA().fit(District_wise_scaled)
```

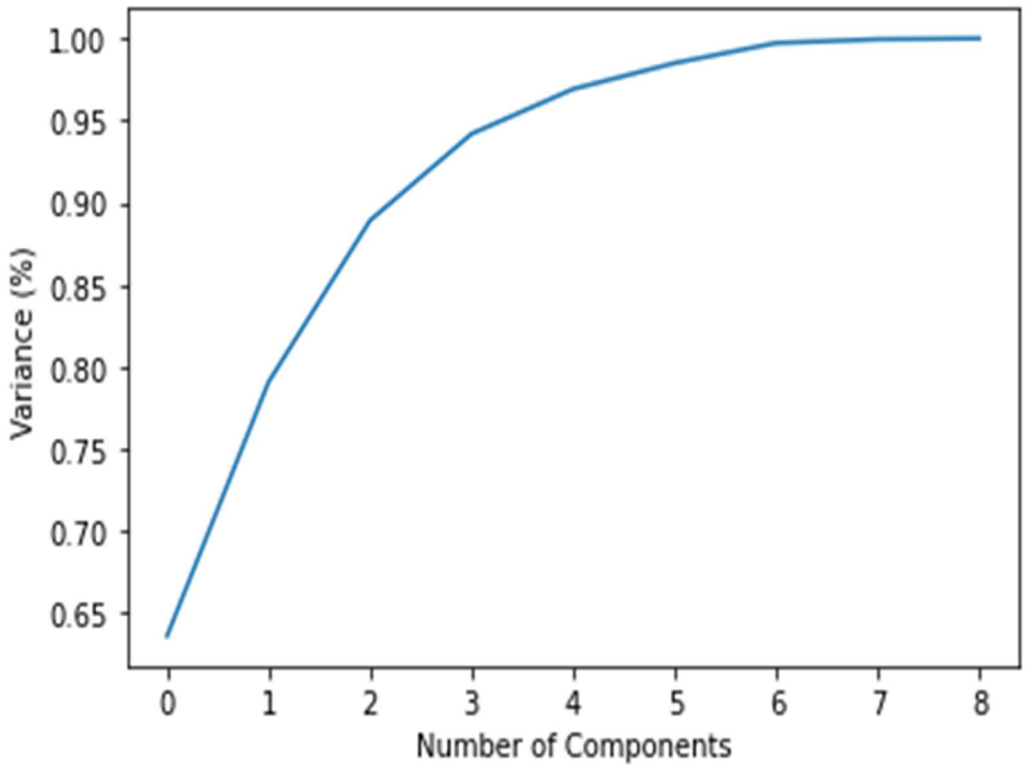
```
C:\Users\rishi\anaconda3\lib\site-packages\sklearn\utils\deprecation.py:143: FutureWarning: The sklearn.decomposition.pca module is  deprecated in versi
on 0.22 and will be removed in version 0.24. The corresponding classes / functions should instead be imported from sklearn.decomposition. Anything that
cannot be imported from sklearn.decomposition is now part of the private API.
  warnings.warn(message, FutureWarning)
```

In [82]:
```
# Importing standardscalar module
from sklearn.preprocessing import StandardScaler

scalar = StandardScaler()
```

In [83]:
```
# fitting
scalar.fit(District_wise_scaled)
scaled_data = scalar.transform(District_wise_scaled)
```

In [84]:
```
pca = PCA().fit(District_wise_scaled)
```

In [85]:
```
#Plotting the Cumulative Summation of the Explained Variance
plt.figure()
plt.plot(np.cumsum(pca.explained_variance_ratio_))
plt.xlabel('Number of Components')
plt.ylabel('Variance (%)') #for each component
#plt.title('Pulsar Dataset Explained Variance')
plt.show()
```



Since the K nearest neighbors algorithm makes predictions about a data point by using the observations that are closest to it, the scale of the features within a data set matters a lot.

Because of this, machine learning practitioners typically standardize the data set, which means adjusting every x value so that they are roughly on the same scale.

Fortunately, scikit-learn includes some excellent functionality to do this with very little headache.

To start, we will need to import the StandardScaler class from scikit-learn. Add the following command to your Python script to do this:

[90]:
```
from sklearn.cluster import KMeans
from sklearn import metrics
from scipy.spatial.distance import cdist
```

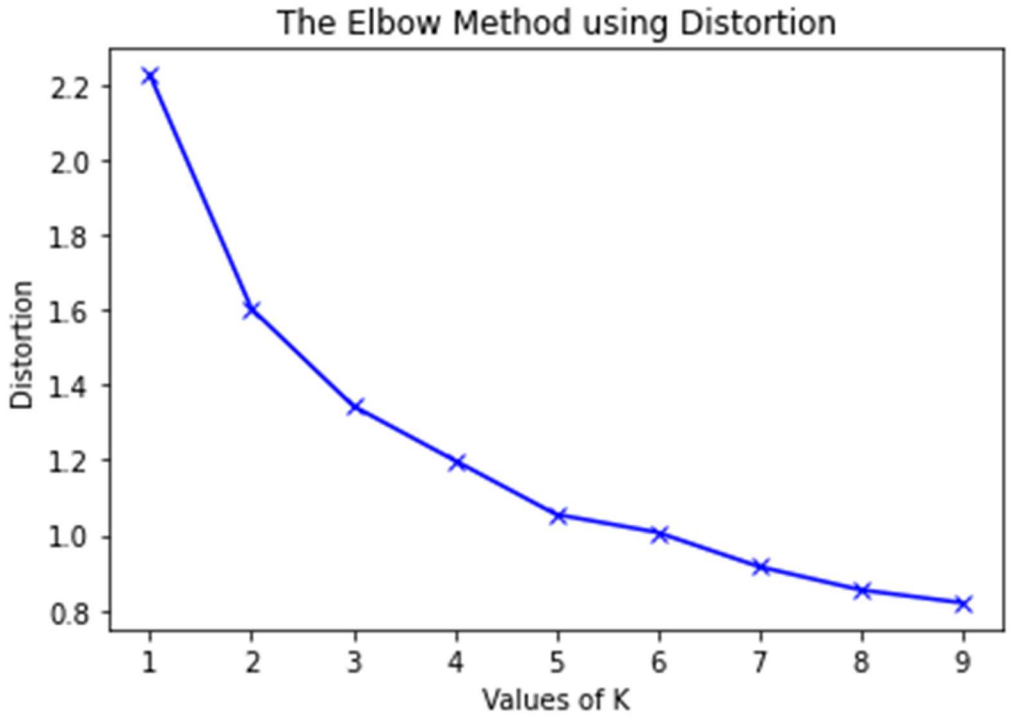
[91]:
```
distortions = []
inertias = []
mapping1 = {}
mapping2 = {}
```

[92]:
```
K = range(1,10)

for k in K:
    #Building and fitting the model
    kmeanModel = KMeans(n_clusters=k).fit(for_train)
    kmeanModel.fit(for_train)

    distortions.append(sum(np.min(cdist(for_train, kmeanModel.cluster_centers_,
                      'euclidean'),axis=1)) / for_train.shape[0])
    inertias.append(kmeanModel.inertia_)

    mapping1[k] = sum(np.min(cdist(for_train, kmeanModel.cluster_centers_,
                'euclidean'),axis=1)) / for_train.shape[0]
    mapping2[k] = kmeanModel.inertia_
```

Mapping and plot the Elbow Method Using Distortion:

```
In [94]: plt.plot(K, distortions, 'bx-')
         plt.xlabel('Values of K')
         plt.ylabel('Distortion')
         plt.title('The Elbow Method using Distortion')
         plt.show()
```

The picture below shows the elbow curve and the elbow point for the % of variance explained metrics.

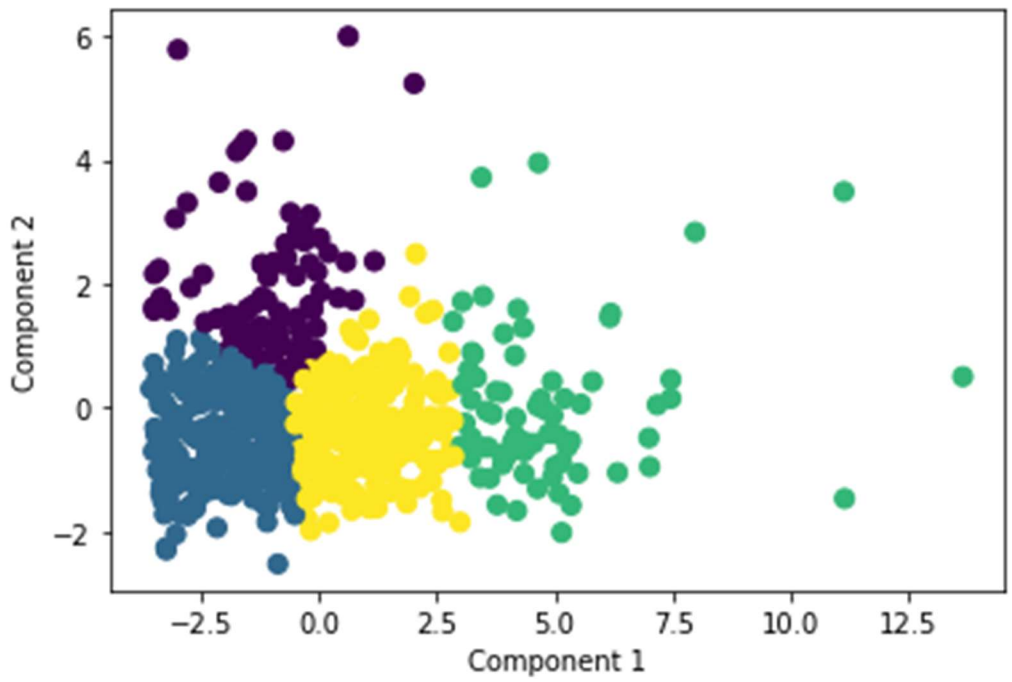


We now iterate the values of k from 2 to 5. We assume that no practical data exists for which all the data points can be optimally clustered into 1 cluster.

We construct the following tables for each value of k:-

```
[95]: kmeanModel = KMeans(n_clusters=4).fit(for_train)
      kmeanModel.fit(for_train)

t[95]: KMeans(n_clusters=4)

[96]: y_kmeans = kmeanModel.predict(for_train)

[97]: plt.scatter(for_train['1'], for_train['2'], c=y_kmeans, s=50)
      plt.xlabel('Component 1')
      plt.ylabel('Component 2')

      Text(0, 0.5, 'Component 2')
```

We see that the highest value of s (i) exists for k = 4. Therefore we conclude that the optimal number of clusters for the given data is 4.



Here we show two states according to Total population Bihar and Tamil Nadu.

Fig 1 – Bihar

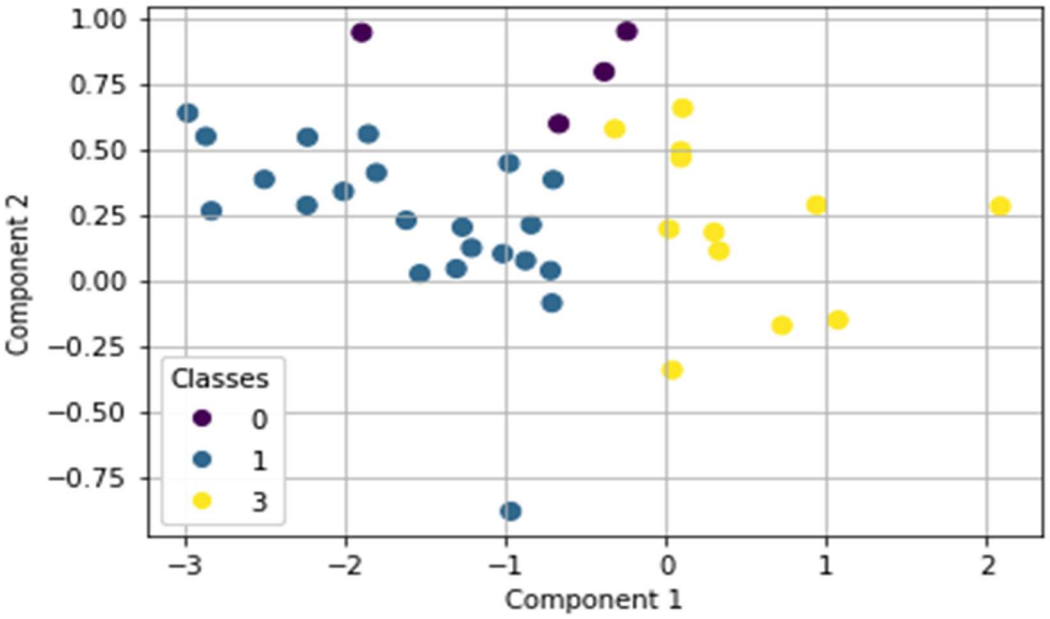
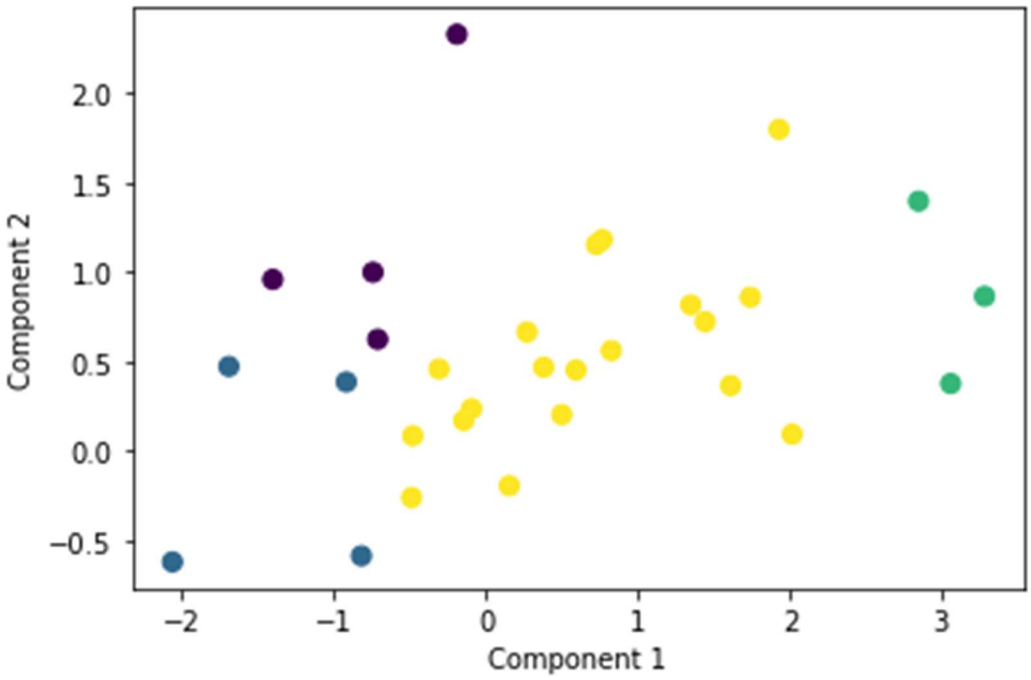
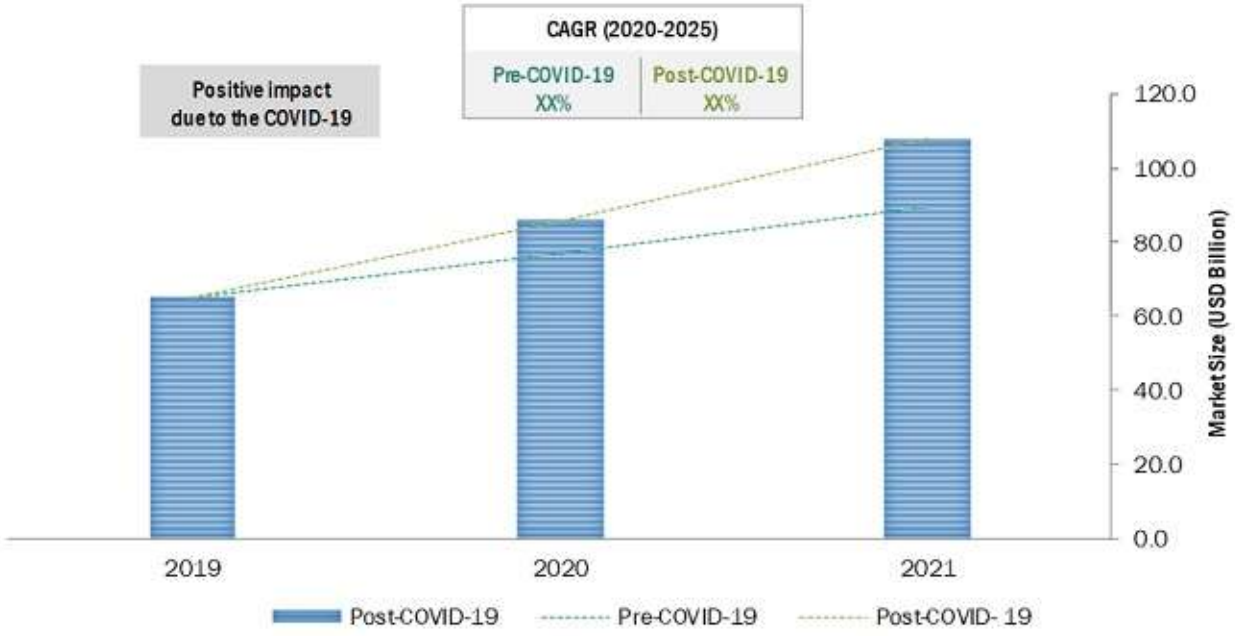


Fig 2 - Tamil Nadu



Selection of Target Segment

In EdTech, target marketing is the technique of promoting your Teaching abilities to a specific audience using tactics like audience research, segmentation, and more. You can better reach your ideal client or student with EdTech target marketing.

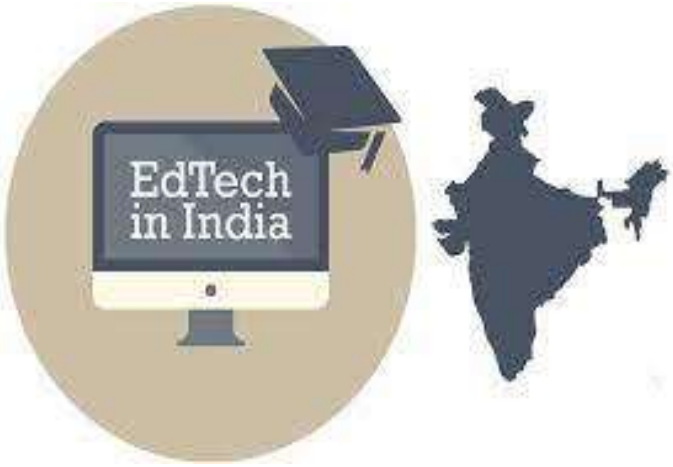


e: estimated; p: projected

Source: MarketsandMarkets Analysis

Four ways to define your target audience

There are a few considerations that will assist you in identifying the appropriate target audience for your marketing efforts. However, the objective should be to be as exact as possible when selecting a target demographic so that your EdTech marketing initiatives get the intended outcomes. Let's take a closer look at the specifics.



Geo-location: When you're doing research on your target market, you'll need to pick where you'll focus your marketing efforts. This selection will be made based on your establishment's physical location and the distance from which you expect students to travel.

Demographics: Marketing campaigns should be built based on the demographics of your target audience. When determining which demographics to target, for example, you'll need to consider gender, age, education, family size, occupation, household income, and other factors.

Psychographics: The lifestyle, behavior, and personality of your audience base are all things to consider while doing psychographic analysis. You should also consider whether the target audience is receptive to fresh ideas.

Behavior: Analyze the demands of our target audience and assess their degree of understanding. In addition, we must assess their responsiveness to specific health-care services and goods.

Marketing Mix

Setting prices for our products is both an art and a science. Most importantly, you must know and understand your cost of production. From there you can adjust Based on product characteristics, a specific pricing strategy, customer price Sensitivity, customer values, and other factors. Price contributes to the perception of your product, that is, when consumers see a product price it sends signals to them about quality, match with the market outlet, expectations for assistance, etc. Keeping Accurate and complete records accounting for all steps – production, packaging, Storage, promotion, transportation/distribution, and sales – will assist you in setting a price and making adjustments as necessary.

4Ps of Marketing Mix

The 4Ps helps companies to review and define key issues that affect the marketing of its products and services and is often now referred to as the 7Ps framework for the digital marketing mix.

Marketing as a whole relies on all seven Ps.It is essential to consider them as a whole, and not in isolation. Customers must experience a coherent view of your company and your product, and that can only come from viewing the customer experience from end-to-end across all seven Ps.

Importance of Marketing Mix

It helps understand what our product or service can offer to our customers and helps plan a successful product offering. Helps with planning, developing and executing effective marketing strategies. Help determine whether your product or service is suitable for your customers.

4Ps of Marketing Mix

Customizing the Market Mix

The marketing mix refers to the set of actions, or tactics, that a company uses to promote its brand or product in the market. The 4Ps make up a typical marketing mix -

Price, Product, Promotion and Place.

1**. Price**: refers to the value that is put for a product. It depends on costs of production, segment targeted, ability of the market to pay, supply - demand and a host of other direct and indirect factors. There can be several types of pricing strategies, each tied in with an overall business plan.

2. **Product**: IT refers to the item actually being sold. The product must deliver a minimum level of performance; otherwise even the best work on the other elements of the marketing mix won't do any good.

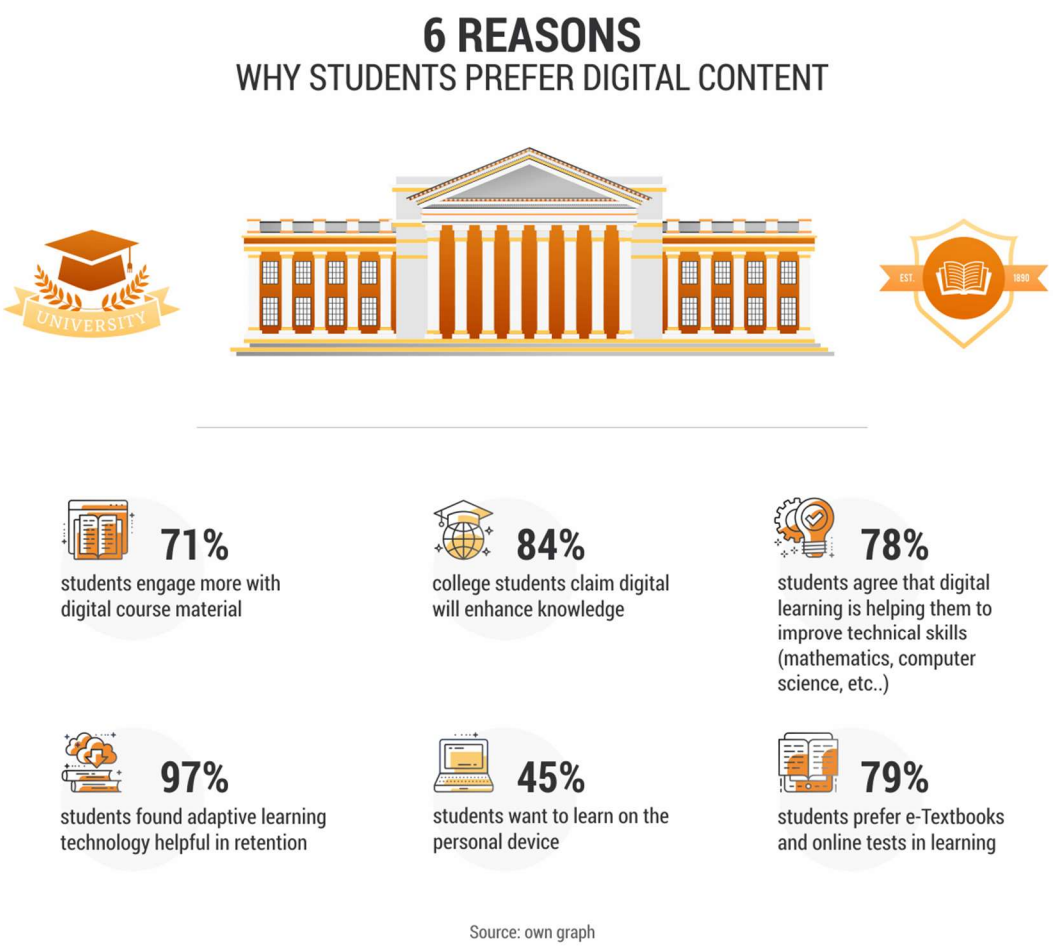
3**. Place**: It refers to the point of sale. In every industry, catching the eye of the consumer and making it easy for them to buy it is the main aim of a good distribution or 'place' strategy. Retailers pay a premium for the right location. In fact, the mantra of a successful retail business is 'location, location, location'.

4. **Promotion**: It refers to all the activities undertaken to make the product or service known to the user and trade. This can include advertising, word of mouth, press reports, incentives, commissions and awards to the trade. It can also include consumer schemes, direct marketing, contests and prizes.

All the elements of the marketing mix influence each other. They make up the business plan for a company and handle it right, and can give it great success. The marketing mix needs a lot of understanding, market research and consultation with several people, from users to trade to manufacturing and several others.

Potential Customer base:

Students from higher literacy rates states and with some specific income are going to be our potential customer.



Codes

All the codes used in this project can be found on

GitHub Link: https://github.com/Rishiverma1993/India-Student-Education-EDA.git