Market Segmentation

Analysing the respective market in India using Segmentation analysis for Ed-tech learning platforms

by

Group 2

Shardul

Madhurya

Mohit wani

Overview

EdTech may be a divisive subject. Because a big component of the education system is unionised, there are fears that EdTech is an attempt to save expenses by phasing out some in-class activities. EdTech proponents stress the software's and technology's ability to boost learning, allowing teachers to take on more of a facilitator role. With limited time, it's challenging for a teacher to follow the curriculum, catch up with lower-level students, and keep the rest of the class involved in their work. EdTech can possibly lead to improved outcomes for individual students and the class as a whole by automating the evaluation of aptitude and modification of difficulty.

There were two phases of deployment of technology in the classroom. The first step was to bring hardware into the classroom. Eventually, the discussion shifted to how to improve the software to better coordinate and use all of the hardware. EdTech is the term for these software solutions. Many of them are cloud-based and use educational research to determine how slowly or quickly a learner should go through various learning objectives.

Massive Open Online Courses (MOOCs) are a type of online course that uses technology to reach a large number of students all over the world. While these courses have their own set of problems, such as poor completion rates, they are an attempt to give knowledge in a manner that is convenient for the user.

Clustering implementation Overview:

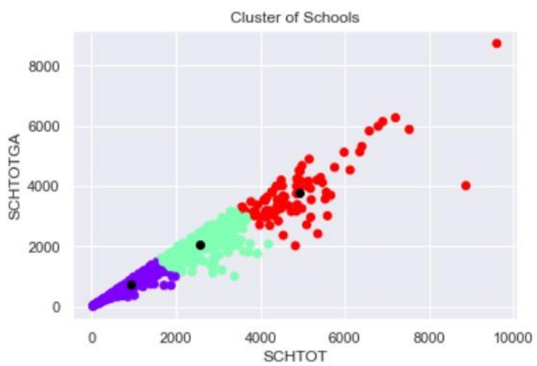
In [30]:
```
kmeans = KMeans(n_clusters= 3)
label = kmeans.fit_predict(X1)
print(label)
```

```
[1 1 0 0 0 1 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0 1 0 1 1 0
 1 1 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 1 1 1 0 1 0 1 1 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 2 1 0 1 1 1
 2 1 2 2 0 2 1 0 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 0 0 0 2 2 1 1 2
 1 1 1 1 2 1 2 2 2 2 2 2 1 1 1 1 1 1 2 1 1 1 0 0 1 1 2 2 1 2 2 1 1 1 2 0 1
 2 1 1 1 1 2 2 1 2 1 1 2 2 1 1 1 1 1 0 1 1 0 0 1 2 0 1 1 1 1 0 1 1 0 0 1 1
 1 1 1 1 1 1 0 1 1 0 0 0 1 2 1 0 0 1 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 1 0 0 0 0 0 1 2 1 0 1 1 0 0 1 2 1 1 1 1 1 1
 1 1 1 0 1 1 1 1 0 1 0 1 1 2 1 1 2 2 2 2 2 2 2 2 2 2 1 2 2 2 0 1 1 1 1 1 0
 2 1 1 1 0 1 1 1 1 0 1 1 1 0 1 0 0 0 0 1 0 1 0 1 1 2 1 1 1 1 1 1 1 1 0 1 1
 2 0 1 0 0 1 0 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1 1 0 1 1 0 1 0 0 0 0 0 0 1 1
 1 1 0 1 1 1 0 1 1 1 1 1 2 1 2 2 0 1 1 0 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1
 0 1 1 1 1 1 1 2 1 1 1 1 0 1 1 0 1 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0
 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 1 0 1 0 0 1 1 1 1 0 1 1 1 2 1 1
 0 2 2 2 1 1 0 2 2 1 1 1 1 1 2 1 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 0 0 0 0
 0 1 1 1 1 0 1 1 1 1 0 1 1 0 0 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
 0 0 0 0 1 0 1 1 0 1 1 1 0 1 0 1 1 0 1 0 1 0 0 1 1 0 1 0 1 0 1 1 0 1 0 0 0
 0 0 0 0 2 1 2 2 1 2 2 2 2 1]
```

In [31]:
```
print(kmeans.cluster_centers_)
```

```
[[ 923.97306397  736.14141414]
 [2550.85049834 2050.26578073]
 [4921.7195122  3801.03658537]]
```

In [32]:
```
plt.scatter(X1[:,0], X1[:,1], c=kmeans.labels_,cmap= 'rainbow')
plt.scatter(kmeans.cluster_centers_[:,0], kmeans.cluster_centers_[:,1], color='black')
plt.title('Cluster of Schools')
plt.xlabel('SCHTOT')
plt.ylabel('SCHTOTGA')
plt.show()
```



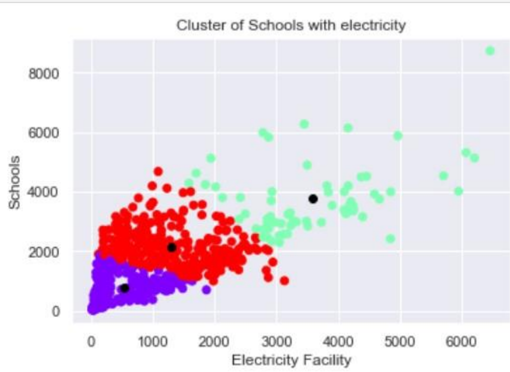
In [34]:
```
kmeans = KMeans(n_clusters= 3)
label = kmeans.fit_predict(X1)
print(label)
```

```
[0 2 0 0 0 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 2 1 0 0 2 0 0 0 0 2 2 0 2 2 0
 2 2 0 0 0 2 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 2 2 2 2 0 2 0 2 2 0 0 0 0 0
 0 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 2 2 2 2 1 2 0 0 0 2
 1 2 1 1 0 1 2 0 2 2 2 0 2 2 1 2 2 2 2 2 2 0 2 2 2 2 2 2 2 0 0 0 2 2 2 2 1
 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 0 0 2 0 2 2 2 1 2 2 2 2 2 0 2
 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 0 0 2 0 0 2 2 0 2 2 0 2 0 2 2 0 0 2 2
 2 2 2 2 2 0 0 2 2 0 0 0 2 2 2 0 0 2 0 2 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 0 0 0 0 0 2 2 2 0 2 2 0 0 0 2 2 2 0 0 2 2
 2 2 2 0 2 2 2 2 0 0 0 0 2 2 2 2 2 1 1 1 1 1 1 1 1 1 2 1 1 1 0 2 0 2 2 0 0
 2 2 0 0 0 2 2 2 2 0 0 2 2 0 0 0 0 0 0 2 0 0 0 2 2 2 2 2 2 2 2 2 2 2 0 2 2
 2 0 2 0 0 2 0 2 2 2 2 0 2 2 2 1 2 2 2 2 2 0 2 2 2 2 2 0 2 0 0 0 0 2 0 2 2
 2 2 0 2 2 2 0 2 2 2 2 2 2 2 2 1 0 2 2 0 2 2 2 0 2 2 2 2 2 2 0 2 2 2 2 2 2
 0 2 2 2 2 2 2 2 2 2 0 0 0 2 2 0 2 2 0 0 2 0 2 0 2 0 0 0 0 0 0 2 2 2 2 0 0
 2 0 0 0 0 2 0 0 0 2 2 0 0 0 0 2 2 1 2 2 0 2 2 1 0 2 2 2 1 1 0 2 2 1 1 1 2
 0 1 1 1 2 2 2 1 1 1 2 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 0 0 0 0
 2 2 2 2 2 0 2 2 2 2 0 2 1 2 0 2 0 0 2 2 2 2 0 0 0 0 0 0 2 0 2 2 0 0 0 0 0
 0 0 0 0 2 0 2 1 2 2 1 2 0 2 0 1 2 0 2 0 2 2 0 2 2 2 2 0 2 2 2 2 0 2 0 0 0
 0 0 0 0 1 1 1 1 2 1 1 1 1 1]
```

In [35]:
```
print(kmeans.cluster_centers_)
```

```
[[ 528.330033    776.77557756]
 [3574.671875   3782.640625  ]
 [1286.98083067 2140.56549521]]
```

In [36]:
```
plt.scatter(X1[:,0], X1[:,1], c=kmeans.labels_,cmap= 'rainbow')
plt.scatter(kmeans.cluster_centers_[:,0], kmeans.cluster_centers_[:,1], color='black')
plt.title('Cluster of Schools with electricity')
plt.xlabel('Electricity Facility')
plt.ylabel('Schools')
plt.show()
```



Market Segmentation

The market is divided into two segments based on hardware: interactive displays and interactive projectors. The interactive displays segment is expected to have the largest share in the global education technology (EdTech) market because these displays help to increase learner engagement and encourage student enthusiasm for learning. Furthermore, by the end of 2030, the increased use of interactive displays to exhibit material such as photographs, movies, and video calls is expected to boost growth in this market sector. Furthermore, the K-12 segment is expected to have the greatest share, owing to the growing trend of game-based learning for children in kindergarten through grade 12, as well as the increased usage of new technologies are used.

The continuous expansion of internet connectivity throughout the world, as well as various technical breakthroughs like as 5G, blockchain, cloud services, Internet of Things (IoT), and Artificial Intelligence (AI), have bolstered economic growth tremendously in the previous two decades. Globally, more than 4.5 billion people were actively using the internet as of April 2021. Furthermore, the rise of the ICT industry has contributed greatly to GDP growth, worker productivity, and R&D investment, among other economic shifts in many countries throughout the world.

Furthermore, the ICT sector's creation of goods and services contributes to economic growth and development. ICT good exports (percentage of total good exports) increased from 10.816 in 2015 to 11.536 in 2019, according to data from the United Nations Conference on Trade and Development's database. In 2019, these exports in Hong Kong SAR, China were 56.65%, 25.23 percent in East Asia & Pacific, 26.50 percent in China, 25.77 percent in Korea, Rep., 8.74 percent in the US, and 35.01 percent in Vietnam. These are some of the key aspects that are propelling the industry forward.

Situational Analysis

Environment

The development and production of digital gadgets has a negative impact on the environment. Between 70% and 80% of the energy consumed over the lifespan of a digital item happens during their initial fabrication, regardless of how they are actually utilised in a classroom. The manufacture of any digital technology "needs a large global network," as Crawford and Joler's forensic "anatomy" of Amazon's Echo device demonstrates, to permit the smelting, processing, and mixing of raw ingredients that are sent halfway around the world to be constructed. Each of these processes entails the accumulation of hazardous waste, hazardous substances, and the disposal of toxic waste.

Current Trend in EDTECH:

According to a new research by PwC, virtual reality technology allows pupils to learn more content faster. It will become simpler to bring learning experiences to life as edtech solutions such as video conferencing software grow more flexible and include components of VR and AR.
Bite-size learning modules, live training, and the use of AR (augmented reality) and gamification to make learning interactive are all current educational trends. Institutions are automating their enrollment/admission and student engagement procedures in terms of operations.
The K-12 sector dominated the EdTech market with a 41.2 percent share. In the K-12 sector, game-based learning is expected to become more popular.

TARGET MARKET

Marketing teams may adapt their message to a specific set of clients via targeted marketing. The moment at which the marketing mix comes together to determine the optimum offer and marketing approach for each target market is the targeting strategy.

Following the selection of target groups, whether in EdTech or any other industry, the suitable targeting technique is chosen. Your customer and consumer are largely the same when it comes to degree and post grad students, which is rare in the K-12 market. As a result, the communication and approach in both circumstances will be different.

Another consideration is the geographical location. In EdTech, however, every parent, regardless of location, wants their kid to flourish and have the greatest tools to do so. So, in marketing a product, disposable money, access to great education, regional/cultural influences, and content consumption are some crucial differentiators to consider.

Analysis and Approaches used for Segmentation

Clustering

Clustering is a popular exploratory data analysis tool for gaining an understanding of the data's structure. It is the challenge of finding subgroups in data so that data points within the same subgroup (cluster) are extremely similar while data points within other clusters are quite dissimilar. To put it another way, we strive to find homogeneous subgroups within the data so that data points in each cluster are as comparable as feasible based on a similarity metric like euclidean-based distance or correlation-based distance. The choice of the similarity metric to utilise depends on the application.Clustering analysis may be done based on features (identifying subgroups of samples based on features) or samples (finding subgroups of samples based on samples) (finding subgroups of features based on samples).

K Means Algorithm

The K Means method is an iterative technique that attempts to divide a dataset into pre-defined separate non-overlapping subgroups (clusters), with each data point belonging to just one of these groups. It attempts to make intra-cluster data points as comparable as feasible while maintaining clusters as distinct (far) as possible. It distributes data points to clusters in such a way that the sum of the squared distances between them and the cluster's centroid (arithmetic mean of all the data points in that cluster) is as little as possible. Within clusters, the less variance there is, the more homogenous (similar) the data points are.

The following is how the k means algorithm works:
1. Determine the number of clusters (K).
2. Initialize the centroids by shuffling the dataset first and then picking K data points at random for the centroids without replacing them.

3. Continue iterating until the centroids do not change. i.e. the clustering of data points does not change.

Expectation maximisation is the method used by k-means to solve the issue.

The data points are assigned to the nearest cluster in the E-step. The centroid of each cluster is computed in the M-step. A breakdown of how we can solve it mathematically may be seen below.

$$J = \sum_{i=1}^{m} \sum_{k=1}^{K} w_{ik} \|x^i - \mu_k\|^2$$

And M-step is:

$$\frac{\partial J}{\partial \mu_k} = 2 \sum_{i=1}^{m} w_{ik}(x^i - \mu_k) = 0$$

$$\Rightarrow \mu_k = \frac{\sum_{i=1}^{m} w_{ik} x^i}{\sum_{i=1}^{m} w_{ik}}$$

Applications

The K means technique is widely utilised in a range of applications, including market segmentation, document clustering, picture segmentation, and compression, among others. When we do a cluster analysis, we normally want to achieve one of two things:

1. Gain a clear understanding of the structure of the data we're working with.
2. Cluster-then-predict, in which various models are developed for distinct subgroups if we feel there is a lot of diversity in their behaviour.

Implementation

In [11]:
```
data.head()
```

Out[11]:

State Name	DISTNAME	SCH1	SCH2	SCH3	SCH4	SCH5	SCH6	SCH7	SCHTOT	SCH1GA	...	KITSTOT	ENR501	ENR502	ENR503	ENR504	ENR505	ENR5
JAMMU & KASHMIR	KUPWARA	1099	811	3	14	2	84	47	2060	982	...	1752	938	236	0	13	1	
JAMMU & KASHMIR	BARAMULA	1205	829	8	11	5	170	72	2300	1107	...	1430	1024	234	0	10	4	
JAMMU & KASHMIR	SRINAGAR	253	405	46	2	2	257	7	972	212	...	464	227	105	2	2	2	
JAMMU & KASHMIR	BADGAM	750	535	9	10	1	101	58	1464	699	...	872	651	129	0	7	0	
JAMMU & KASHMIR	PULWAMA	544	347	4	13	2	88	45	1043	511	...	691	515	121	0	10	1	

5 rows × 188 columns

In [12]:
```
df_1=data.iloc[:,129:179]
```

In [13]:
```
data.drop(columns=df_1,inplace=True)
print(data.shape)
```

```
(680, 138)
```

In [14]:
```
df_3=data.iloc[:,73:89]
```

In [9]:
```
data.drop(columns=df,inplace = True)
data.head()
```

Out[9]:

DISTCD	State Name	DISTNAME	SCH1	SCH2	SCH3	SCH4	SCH5	SCH6	SCH7	SCHTOT	...	KITSTOT	ENR501	ENR502	ENR503	ENR504	ENR505	ENR50
101	JAMMU & KASHMIR	KUPWARA	1099	811	3	14	2	84	47	2060	...	1752	938	236	0	13	1	
102	JAMMU & KASHMIR	BARAMULA	1205	829	8	11	5	170	72	2300	...	1430	1024	234	0	10	4	2
103	JAMMU & KASHMIR	SRINAGAR	253	405	46	2	2	257	7	972	...	464	227	105	2	2	2	3
104	JAMMU & KASHMIR	BADGAM	750	535	9	10	1	101	58	1464	...	872	651	129	0	7	0	1
105	JAMMU & KASHMIR	PULWAMA	544	347	4	13	2	88	45	1043	...	691	515	121	0	10	1	

5 rows × 189 columns

In [10]:
```
data.set_index('State Name',inplace= True)
```

In [5]:
```
data.describe().transpose()
```

Out[5]:

	count	mean	std	min	25%	50%	75%	max
SCH1	680.0	1245.761765	991.465739	0.0	527.75	1048.0	1721.75	7982.0
SCH2	680.0	408.391176	422.718678	0.0	75.00	241.5	677.25	3177.0
SCH3	680.0	55.088235	109.377370	0.0	5.00	19.0	60.00	1555.0
SCH4	680.0	216.747059	318.356296	0.0	4.00	26.0	358.00	1590.0
SCH5	680.0	56.961765	88.032754	0.0	3.00	13.0	80.00	630.0
...	...	...	...	...	...	...	...	...
C5_G	680.0	18108.420588	14811.543998	49.0	7448.50	14363.5	24722.75	81234.0
C6_G	680.0	16350.511765	13144.336785	71.0	7071.00	13398.5	22558.50	83678.0
C7_G	680.0	16046.535294	13114.081901	51.0	6840.50	13199.0	21982.00	89490.0
C8_G	680.0	15638.133824	12752.086416	61.0	6703.50	12426.0	21477.25	79484.0
C9_G	680.0	14156.525000	12004.150433	62.0	6025.75	11542.5	18949.25	79444.0

252 rows × 8 columns

In [6]:
```
data.drop(columns=['Unnamed: 0','SCH9','SCH9GA','SCHBOY9','SCHGIR9','ENR9','TCH9'],inplace=True)
```

In [4]:
```
data.info
```

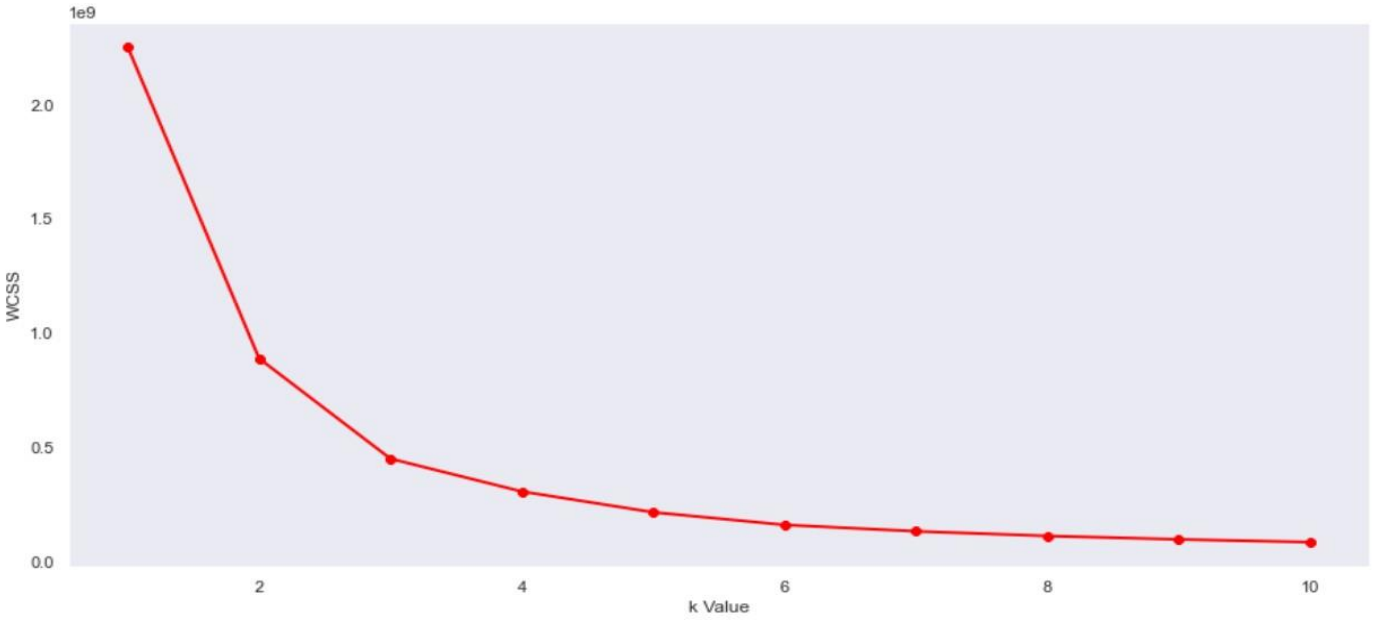
Out[4]:
```
<bound method DataFrame.info of        Unnamed: 0      State Name     DISTNAME  SCH1  SCH2  SCH3  SCH4  SCH5  \
DISTCD
101       2014-15  JAMMU & KASHMIR      KUPWARA  1099   811     3    14     2
102       2014-15  JAMMU & KASHMIR     BARAMULA  1205   829     8    11     5
103       2014-15  JAMMU & KASHMIR     SRINAGAR   253   405    46     2     2
104       2014-15  JAMMU & KASHMIR       BADGAM   750   535     9    10     1
105       2014-15  JAMMU & KASHMIR      PULWAMA   544   347     4    13     2
...           ...              ...          ...   ...   ...   ...   ...   ...
3606      2014-15        TELANGANA  RANGA REDDY  2771   876    21     5    17
3607      2014-15        TELANGANA  MAHBUBNAGAR  3091   986    11     0    10
3608      2014-15        TELANGANA     NALGONDA  2817   628    13     1    34
3609      2014-15        TELANGANA     WARANGAL  2847   735    15     0    31
3610      2014-15        TELANGANA     KHAMMAM   2319   630    11     0     4

          SCH6  SCH7  ...   C9_B   C1_G   C2_G   C3_G   C4_G   C5_G   C6_G  \
DISTCD                ...
101         84    47  ...   6376  10368   8665   8543   7807   7229   6324
102        170    72  ...   8029  18126  11291  10873  10104  10015  10163
103        257     7  ...   6433   8422   7984   8032   7957   7717   7315
104        101    58  ...   5070   7077   6186   6051   5758   5439   5051
105         88    45  ...   3753   4438   3849   3891   3750   3567   3476
...        ...   ...  ...    ...    ...    ...    ...    ...    ...    ...
3606       147  1525  ...  44552  59588  55394  54084  52205  51487  50850
3607        96   994  ...  28335  43914  37932  37795  37222  35751  33076
3608        42  1133  ...  25734  30289  26462  26385  26295  26204  25479
3609       116  1132  ...  27602  33246  25163  25192  25111  25402  25846
3610        83   610  ...  19076  20131  19117  19843  19543  20254  19818

          C7_G   C8_G   C9_G
```

Segmentation

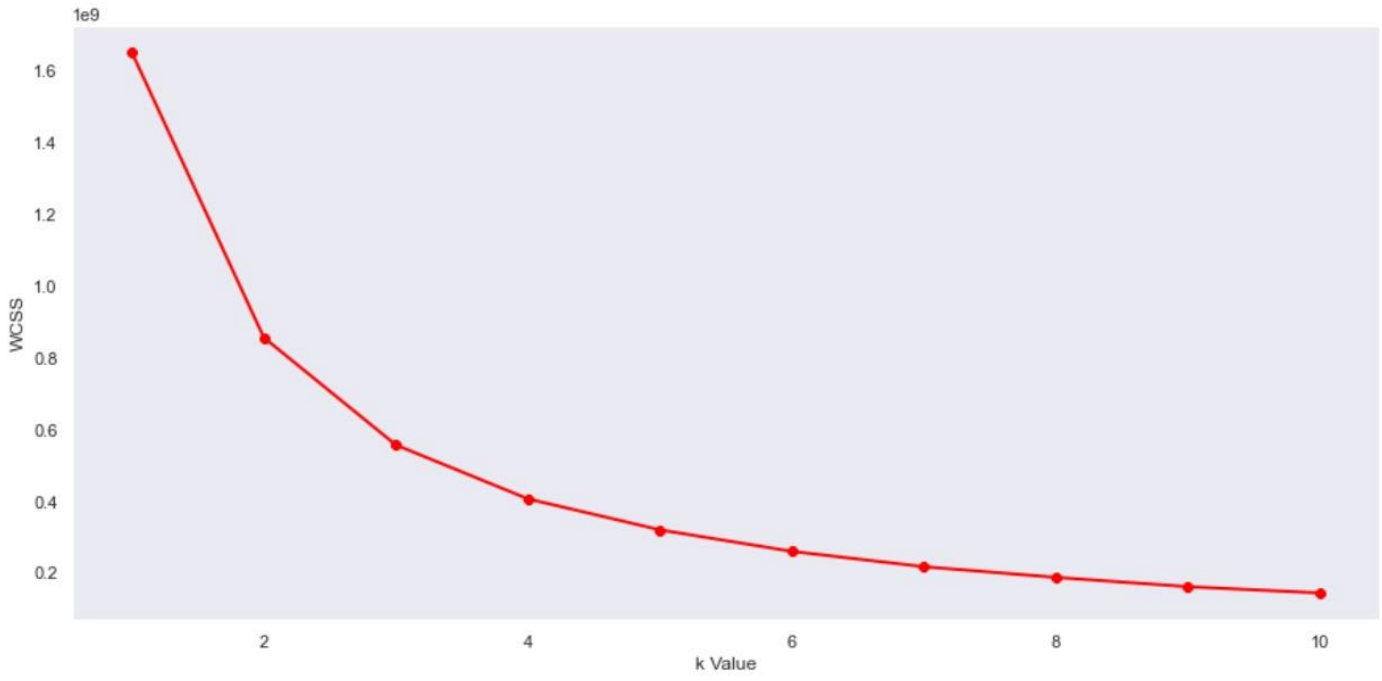
Using K Means

And we can use the Elbow method to find the optimum K value. For this our plot is something like this.

In [29]:
```
X1 = data.loc[:, ["SCHTOT","SCHTOTGA"]].values

from sklearn.cluster import KMeans
wcss = []
for k in range(1, 11):
    kmeans = KMeans(n_clusters=k, init='k-means++')
    kmeans.fit(X1)
    wcss.append(kmeans.inertia_)
plt.figure(figsize=(15,7))
plt.grid()
plt.plot(range(1,11),wcss, linewidth=2, color='red', marker="8")
plt.xlabel('k Value')
plt.ylabel('WCSS')

plt.show()
```



In [33]:
```
X1 = data.loc[:, ["SELETOT","SCHTOTGA"]].values

from sklearn.cluster import KMeans
wcss = []
for k in range(1, 11):
    kmeans = KMeans(n_clusters=k, init='k-means++')
    kmeans.fit(X1)
    wcss.append(kmeans.inertia_)
plt.figure(figsize=(15,7))
plt.grid()
plt.plot(range(1,11),wcss, linewidth=2, color='red', marker="8")
plt.xlabel('k Value')
plt.ylabel('WCSS')

plt.show()
```

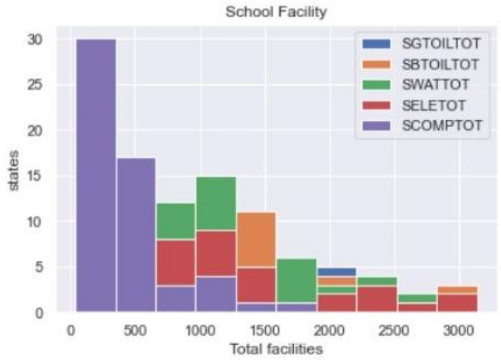


In [26]:
```
import numpy as np
```

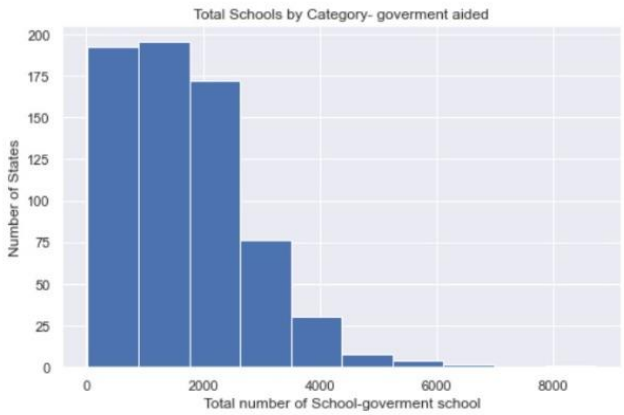
In [27]:
```
Total=['SGTOILTOT','SBTOILTOT','SWATTOT','SELETOT','SCOMPTOT']
```

In [28]:
```
data.loc[['JAMMU & KASHMIR', 'HIMACHAL PRADESH', 'PUNJAB'],Total].plot.hist()
plt.title('School Facility')
plt.xlabel('Total facilities')
plt.ylabel('states')
```

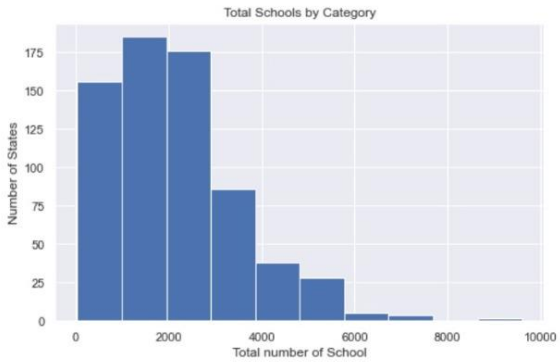
Out[28]:
```
Text(0, 0.5, 'states')
```



In [25]:
```
data['SCHTOTGA'].plot(kind='hist',figsize=(8,5))
plt.title('Total Schools by Category- goverment aided')
plt.ylabel('Number of States')
plt.xlabel('Total number of School-goverment school')
plt.show()
```



In [24]:
```
data['SCHTOT'].plot(kind='hist',figsize=(8,5))
plt.title('Total Schools by Category')
plt.ylabel('Number of States')
plt.xlabel('Total number of School')
plt.show()
```



Codes

All the codes used in this project can be found on

https://github.com/SasukeUchiha7/market-segmentation

References

Datasets that has been used in this project are taken from

https://www.kaggle.com/datasets