ABSTRACT

In today's fast-moving world of marketing from product-orientation to customer-orientation, the management of customer treatment can be seen as a key to achieve revenue growth and profitability. The business marketers usually prefer to cooperate with fewer but larger buyers than the final consumer marketer. Most customer segmentation approaches based on customer value fail to account for the factor of time and the trend of value changes in their analysis. In this article, we classify customers based on their value using the RFM model and K-means clustering method. Then, an assessment of changes over several periods of time is carried out. When deciding in which segment to invest or how to distribute the marketing budget, managers generally take risks in making decisions without considering the real impact every client or segment has over organizational profits. In this paper, a segmentation framework is proposed that considers, firstly, the calculation of customer lifetime value, the current value, and client loyalty, and then the building of client segments by self-organized maps. The effectiveness of the proposed method is demonstrated with an empirical study in a cane sugar mill where a total of 9 segments of interest were identified for decision making.

TABLE OF CONTENTS

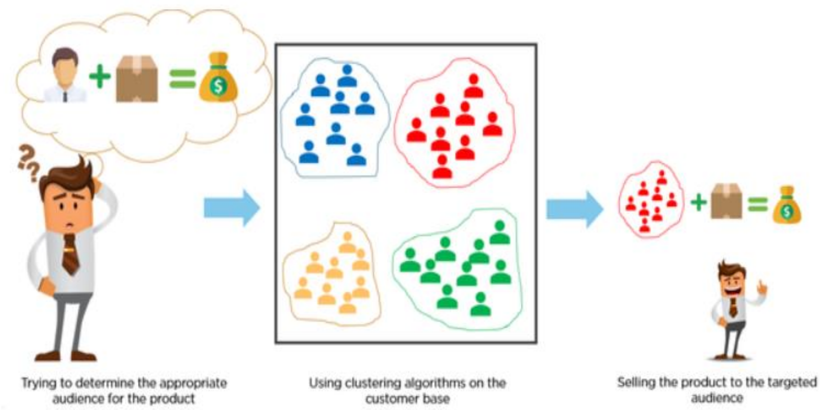
INTRODUCTION

Over the years, the increase in competition amongst businesses and the availability of large historical data repositories have prompted the widespread applications of data mining techniques in uncovering valuable and strategic information buried in organisations' databases. Data mining is the process of extracting meaningful information from a dataset and presenting it in a human understandable format for the purpose of decision support. The data mining techniques intersect areas such as statistics, artificial intelligence, machine learning and database systems. The applications of data mining include but not limited to bioinformatics, weather forecasting, fraud detection, financial analysis and customer segmentation. The thrust of this paper is to identify customer segments in a retail business using a data mining approach. Customer segmentation is the subdivision of a business customer base into groups called customer segments such that each customer segment consists of customers who share similar market characteristics. This segmentation is based on factors that can directly or indirectly influence market or business such as products preferences or expectations, locations, behaviours and so on. The importance of customer segmentation include, inter alia, the ability of a business to customise market programs that will be suitable for each of its customer segments; business decision support in terms of risky situation such as credit relationship with its customers; identification of products associated with each segments and how to manage the forces of demand and supply; unravelling some latent dependencies and associations amongst customers, amongst products, or between customers and products which the business may not be aware of; ability to predict customer defection, and which customers are most likely to defect; and raising further market research questions as well as providing directions to finding the solutions.

Clustering has proven efficient in discovering subtle but tactical patterns or relationships buried within a repository of unlabelled datasets. This form of learning is classified under unsupervised learning. Clustering algorithms include k-Means algorithm, k-Nearest Neighbour algorithm, Self-Organising Map (SOM) and so on. These algorithms, without any knowledge of the dataset beforehand, are capable of identifying clusters therein by repeated comparisons of the input patterns until the stable clusters in the training examples are achieved based on the clustering criterion or criteria. Each cluster contains data points that have very close similarities but differ considerably from data points of other clusters. Clustering has got immense applications in pattern recognition, image analysis, bioinformatics and so on. In this paper, the k-Means clustering algorithm has been applied in customer segmentation.

Customer Segmentation over the years, the commercial world is becoming more competitive, as such organizations have to satisfy the needs and wants of their customers, attract new customers, and hence enhance their businesses. The task of identifying and satisfying the needs and wants of each customer in a business is a very complex task. This is because customers may be different in their needs, wants, demography, geography, tastes and preferences, behaviours and so on. As such, it is a wrong practice to treat all the customers equally in business. This challenge has motivated the adoption of the idea of customer segmentation or market segmentation, in which the customers are subdivided into smaller groups or segments wherein members of each segment show similar market behaviours or characteristics. According to, customer segmentation is a strategy of dividing the market into homogenous groups. posits that —the purpose of segmentation is the concentration of marketing energy and force on subdivision (or market segment) to gain a competitive advantage within the segment.

It's analogous to the military principle of concentration of force to overwhelm energy. Customer or Market segmentation includes geographic segmentation, demographic segmentation, media segmentation, price segmentation, psychographic or lifestyle segmentation, distribution segmentation and time segmentation.



K means

The algorithm is called k-means due to the fact that the letter k represents the number of clusters chosen. An observation is assigned to a particular cluster for which its

distance to the cluster mean is the smallest. The principal function of algorithm involves finding the k-means. First, an initial set of means is defined and then subsequent

classification is based on their distances to the centres. Next, the clusters' mean is computed again and then reclassification is done based on the new set of means. This is

repeated until cluster means don't change much between successive iterations [7]. Finally, the means of the clusters once again calculated and then all the cases are assigned to the

permanent clusters.

Given a set of observations (x^1, x^2,.., x^n), where each observation xi is a d-dimensional real vector.

Algorithm 1: K-Means

Input: Number of Clusters K, Distance D

Output: K clusters of the dataset

1. Select K clusters for the dataset

2. Assign the centroids randomly

3. For each data point, calculate the closest centroid.

4. Assign the point to the cluster

5. Set each cluster position to the mean of all data points assigned to the cluster.

6. Step 4 and 5 is repeated until no more changes.

FEATURES

Not only do companies strive to divide their customers into measurable segments according to their needs, behaviours or demographics but they also aim to determine the profit potential of each segment by analysing its revenue and cost impacts. Value-based segmentation evaluates groups of customers in terms of the revenue they generate and the costs of establishing and maintaining relationships with them. It also helps companies determine which segments are the most and least profitable so that they can adjust their marketing budgets accordingly.

Customer segmentation can have a great effect on customer management in that, by dividing customers into different groups that share similar needs, the company can market to each group differently and focus on what each kind of customer needs at any given moment. Large or small, niche customer segments can be targeted depending on the company's resources or needs.

In B2B marketing, companies are concerned with decision-makers' job titles, the industry sector, whether the company is public or private, its size, location, buying patterns and their technology at their disposal, for example.

In B2C marketing, companies are concerned with particular customers' profiles, attitudes and lifestyles. B2C companies may also be concerned with geographic location. B2C companies who segment customers based on their geographic location can tailor offers based on regional events and preferences. B2C companies can also customize offers based on the predominant languages spoken in each region.

Approaches to B2B customer segmentation include vertical or horizontal alignments. In vertical segmentation, companies select certain industries or job titles that would likely find their products appealing and then focus marketing efforts on those segments that they feel are most ready to buy. The benefit of vertical segmentation is that companies can offer services that are fine-tuned to particular industries. The needs of the financial services industry are different from those of the healthcare industry. If each segment was offered services customized to that industry, adoption and satisfaction might increase.

In horizontal segmentation, companies simply focus on one job title across a wide range of industries and organizations. The benefit of horizontal segmentation is a stronger focus on the needs of particular job titles or job roles. For example, a focus on Chief Financial Officers (CFO) can create product collateral, website messaging and email newsletters specifically tailored to that role.

Customer segmentation vs. market segmentation

Companies can use marketing automation software to define and create customer segments. The customer segments can be based on demographic data, psychographic data and activity-based data such as actions that users took on a website. Companies use marketing automation software to configure, schedule and execute campaigns for particular customer segments.

Customer segmentation is different from market segmentation. An example of market segmentation is grouping customers by the products or services they purchase. A company may perform market segmentation based on distinct lines of business such as software, professional services and training. The company can then allocate resources to each market segment and employ separate marketing and advertising activities to each.

CUSTOMER SEGMENTATION MODEL

Customer Segmentation is the subdivision of a market into discrete customer groups that share similar characteristics. Customer Segmentation can be a powerful means to identify unsatisfied customer needs. Using the above data companies can then outperform the competition by developing uniquely appealing products and services.

Customer segmentation model

The proposed segmentation method is based on LTV calculation proposed by Kim et al. (2006) and Hwang et al. (2004) that considered three factors: current value, potential value, and customer loyalty. In addition, self-organizing maps are used as a tool for clustering the customer database and identifying the most valuable customers. The model considers the following steps.

Step 1: customer specification

First, it is necessary to define the scope of the analysis that will be done by defining business unit, geographical coverage, kind of product, customer aggregation level, the active or inactive status client, as well as the time that will be covered by the analysis. Making clear these parameters, the organization could perform better analysis and even plan the analysis in different levels.

Step 2: sales identification and payments done by customers

To continue with the process, some information about financial transactions has to be calculated for every customer. For revenues, the following has to be calculated:

Compilation of customers' historical purchase: information about transactions done by customer in the period of analysis should be recollected.

Compilation of customer arrears in payment: information about payment date of customer obligation in the period of analysis should be recollected, with the goal of identifying payments which was done after payment date. This information will be used in the next chapter for doing the calculation of customer earned value.

Assignation cost: variable costs and customer acquisition costs should be identified.

Step 3: representative costs identification in customer relationship

Identifying costs, which have been incurred in customer relationship, is really important. It should include direct and indirect acquisition, production, marketing and distribution costs.

2.4. Step 4: calculation of customer lifetime value

With revenues and costs per client, the customer life time value can be calculated, understanding it as the performance generated by each customer in time analysed. It is calculated using incomes and costs by following Eq. (1) of structural basic model.

$$CLV = \sum_{i=1}^{n} \frac{(Ri - Ct)}{(1+d)^{i-0.5}} \quad (1)$$

where Ri, Gross Income, is the value of amount bought per price paid, Ci is the relationship cost with customer that corresponds to the sum of all costs associated with the customer, and d, the discount rate, is the money discount rate established for reflecting cash flow risk.

2.5. Step 5: calculation of customer earned value

Before deciding to hold a customer, the expected effect on profit and portfolio risk should be determined. Glazer and Dhar (2003) affirmed that if there is better customer portfolio behaviour, there will be more contribution in the enterprise. In the model, customer earned value is the criteria which allowed comparing incomes and arrears in payment. This value corresponds to gross sales done by a customer minus arrears in time of the commercial relationship that is proposed by Kim et al. (2006).

$$Customerearnedvalue = \frac{grosssales - areas}{timeofanalysis} \quad (2)$$

2.6. Step 6: calculation of customer purchase rotation

In this model, purchase rotation is considered like brand customer preference, and it is a measure of customer loyalty. A high rotation means a high customer loyalty level. It is equivalent to sum of purchases number done by a customer in a period of specific time.

$$\sum_{j=1}^{n} TransationNumberof C_{ij} \quad (3)$$

where Cij corresponds to the customer i in the period of time j, and n is the last month of the period of analysis.

2.7. Step 7: design of self-organizing map

It is necessary to do the data normalization to prevent that bigger magnitudes in one criteria void lesser magnitudes from other criteria.

$$Xn = 2\left[\frac{X - Xmin}{Xmax - Xmin} - 1\right] \quad (4)$$

where Xn indicates the normalized value and X represents the original value. By the same way, Xmax and Xmin indicate the maximum value and the minimum value of the variables, respectively.

After having identified and normalized the data, some parameters have to be defined in order to run self-organizing maps. Those parameters are: network architecture, initialization procedure, and training algorithm. Final simulation was run using Matlab 7.12.0 (R2011a) which solves clustering problems by SOM with its Neural Network Clustering Tool (nctool).

First, the normalized data can be considered as the input matrix to be organized. The neurons in the entrance layer belong to the customer number and the neurons in the exit layer will be the segments number obtained. This entrance vector has dimension three X = [X1, X2, X3], because the three segmentation criteria are considered for the model; X1 represents customer lifetime value, X2 represents customer earned value and X3 represents customer purchase rotation.

On the other hand, the network architecture has to be defined. This network has one layer, with neurons organized in a grid which is defined by the number of rows and columns in the grid. In addition, initial values are given to the prototype vectors. Among random sample and linear initialization procedures, random method was chosen. Finally, training the network following to the minimum distance rule for finding the winner neuron of data matrix and an entrance vector is selected randomly for checking the red.

EXISTING METHOD

The most common ways in which businesses segment their customer base are:

Demographic information, such as gender, age, familial and marital status, income, education, and occupation.

Geographical information, which differs depending on the scope of the company. For localized businesses, this info might pertain to specific towns or counties. For larger companies, it might mean a customer's city, state, or even country of residence.

Psychographics, such as social class, lifestyle, and personality traits.

Behavioural data, such as spending and consumption habits, product/service usage, and desired benefits.

There are three main approaches to market segmentation:

A priori segmentation, the simplest approach, uses a classification scheme based on publicly available characteristics — such as industry and company size — to create distinct groups of customers within a market. However, a priori market segmentation may not always be valid, since companies in the same industry and of the same size may have very different needs.

Needs-based segmentation is based on differentiated, validated drivers (needs) that customers express for a specific product or service being offered. The needs are discovered and verified through primary market research, and segments are demarcated based on those different needs rather than characteristics such as industry or company size.

Value-based segmentation differentiates customers by their economic value, grouping customers with the same value level into individual segments that can be distinctly targeted.

Techniques for making segmented customer lists

Research your possible customer segments first. When you start collecting and looking at the data on customers in general, you'll find group divisions will start to rise to the surface. Those divisions–while probably not well-defined at this point–can help you start to group your customers together. To get started, think about where you store data about your customers.

Mine your CRM

Your demographic data, decision-making status, and a lot of your customer history can be found right in your customer relationship management tool—if your team has been using it correctly. If you’re not able to pull complete customer records by age, location, or another demographic piece, it might be time to do a little CRM data cleanse.

Pull example customer lists from these fields in your CRM. Remember, you know your customers best, so these are example ideas, not a full list. Feel free to jump on any segments you find that fit your particular customers.

B2B segments

- Company location
- Job title
- Lead status
- Vendors
- Partners
- Industry verticals
- Estimated customer lifetime value

B2C segments

- Age
- Location (by zip code, state, country)
- Language
- Recent purchases
- Gender (although this can be an unreliable signal/segment)
- Estimated customer lifetime value

PROPOSED METHOD

Try marketing automation and email tools for more advanced customer segments

Marketing automation tools often have built-in email segmentation tools that divide your lists based on customer interactions with emails and content you've sent them, product interests (if your MA is connected to your ecommerce tool, inventory management tool, or ERP), and interaction with forms and pages on your website. Because marketing automation and CRM tools have a Venn diagram of overlapping capabilities, you may also use some of the segments you looked for in your CRM here as well.

Try segmenting based on this data from your marketing automation or email marketing tool:

- Highly engaged customers who engage with your emails/content
- Customers who respond to coupons/promotional content
- Customers who view and interact with your content on mobile, desktop, or tablet
- Leads by lead score, or funnel stage

Leads by last content type or product they engaged with (downloads, blog post, social media post)

Whew! Once you're done pulling the segment, you're ready to get started building your campaign. You'll also want to upload that list of customer emails into your marketing automation or email marketing tool so they're loaded and waiting for when you're ready to pull the trigger on your campaign.

- ✓ What to do with segmented customer lists?

Divide & Conquer

Pull a test segment of customers from your vast stores of data and plan 3-10 weeks of content for these customers. Send them special emails with content that speaks directly to this group. Place offers and content (articles, videos, blog posts, images) targeted to this audience on the social media channels where they spend their time.

Identifying right customer and providing right service at right time and treating different types of customers differently is the key to success in business.

So, a predictive model will be used to segregate customers into different groups based on their transactional data. Once the customers are segregated then their associative buying pattern are identified to enhance the profit for the organization future coming customer.

METHODOLOGIES

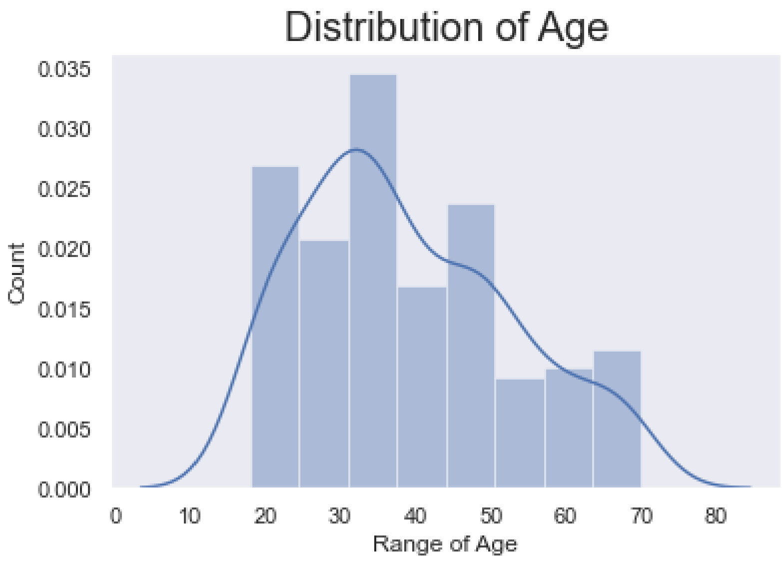
In order to identify the target customers Clustering technique can be used for cluster analysis. Clustering is defined as to group data in clusters/segments so that data within segment are similar while data across the segments are dissimilar. Various techniques can be used for clustering like k-means, hierarchical, grid-based model-based technique. In this paper we proposed to use K-means technique for customer segmentation due its following advantages: This technique suits for the data with numeric features and often terminates at local optimum. It is highly scalable and efficient for large data sets. It is fast in modelling and its result is more understandable.

K Means Clustering is an unsupervised machine learning algorithm. The output is not predicted in unsupervised learning algorithm. It finds similar patterns. The clustering [10] [12] of data is based on the similarity that occurs in the dataset. The k means algorithm finds number of clusters the dataset may be grouped into. For each row, the cluster number will be assigned randomly by the algorithm. The centroid of each cluster is determined. The following two steps are performed repeatedly until the within cluster sum of squares is minimized.

- Reassign data points to the cluster whose centroid point is very nearer.
- Determine new centroid for each cluster obtained

The within cluster variation is calculated by between_SS / total_SS. X1,X2,…..Xj is a set of observations. The cluster variance is defined as the sum of the squared variations of mean of the cluster of all the rows in the dataset. The goal of clustering algorithm is minimizing the within cluster variance. Minimum value of cluster variance is more preferable.

Customer-centric companies have long understood the need to manage their customer portfolio, rather than just their portfolio of products or services. These firms know that the customer portfolio must be the fundamental factor guiding how a company is organized, what it manages, and what it measures.



A segmentation strategy provides companies with the insight they need to manage their businesses profitably and with a customer focus. Segmentation delivers that insight by subdividing a customer portfolio into multiple categories, based on such attributes as behaviours, value, and needs. Additionally, segmentation asks questions to identify such groups as the highest revenue-generating customers, value-destroyers, which consumers should be targeted for acquisition, and who should be earmarked for cross-selling. This, in turn, enables organizations to better address customers' needs and requirements, as well as predict customers' future behaviour, including the likelihood of churn.

Segmentation methodologies that focus on behaviour, life stage, and needs help companies identify the characteristics of a product, service, or channel that appeal to specific customer groups. Moreover, looking at consumers' needs and behaviours without a specific product, service, or channel in mind can lead to innovative ideas that can help to meet those needs.

Ultimately, segmentation provides companies with insight into actionable consumer segments, along with the best ways to meet their needs and increase their value. However, to maximize the benefits of segmentation, companies need to understand their customers from multiple dimensions: demographics, life stage, financial contribution, potential value, profitability figures, behaviours, and needs. An ideal customer segmentation strategy should consider all these dimensions in order to provide a holistic view of the consumer.

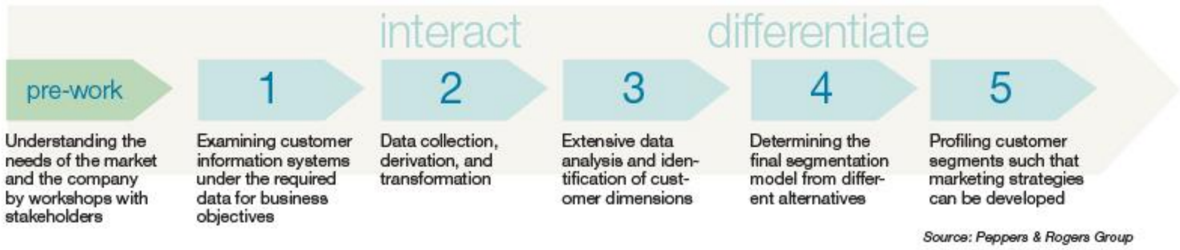
- Added insight and growth potential

Once a segmentation framework is in place companies will gain new insight into their customers' value and potential value. For example, in every customer portfolio there is a group that does not generate financial value. It is critical to identify this group and then decide on the best strategy for handling these customers. Segmentation not only clearly identifies lower-value consumers, but also tracks their changes.

Our recommendation is to not ignore the lower-value segments, but rather to optimize the company's resources to match customer value and improve the bottom line. Using a multidimensional segmentation framework allows companies to decide on the optimum allocation of resources, including not wasting scarce and expensive resources on lower-value segments.

Figure 2: A Step-by-Step Approach to Segmentation

For each level of segmentation there is an end-to-end process starting from data analysis and ending with customer profile determination.



Furthermore, it helps companies to identify ways to improve the value-generation capacity of those lower-value segments. This is when multidimensional segmentation becomes critical because it provides companies with a deeper understanding of the behaviours, life stages, and needs of lower-value consumers and enables the development of new, less costly products, services, and channels that match these criteria. Thus, companies can increase the financial value generated from this segment.

IMPLEMENTATION

First of all, we will be importing the datasets of a customer which consists of 200 rows and 5 columns. Columns consist of CustomerID, Gender, Age, Annual Income, Spending Score.

Training the Customer Segmentation model consists of

- Data Exploration
- Data Visualisation
- Selecting Number of clusters for K-means algorithm
- Clustering
- Plotting

CODE:-

```
#!/usr/bin/env python
# coding: utf-8

# # <center>Customer Segmentation</center>

# In[1]:



import pandas as pd
import matplotlib.pyplot as plt
get_ipython().run_line_magic('matplotlib', 'inline')
import seaborn as sns
print(pd.__version__)



# # Data Exploration

# In[2]:
```

```
df = pd.read_csv("Mall_Customers.csv")
df.head()


# In[20]:


df.shape


# In[3]:


df.isna().sum()


# In[4]:


df.dtypes


# In[5]:


df1 = df.drop(['CustomerID','Gender','Age'],axis='columns')
df1


# In[6]:
```

```
df['Gender'].replace(['Female','Male'], [0,1],inplace=True)
df.Gender


# # Data Visualization

# In[7]:


plt.ylabel('Annual Income (k$)')
plt.xlabel('Spending Score (1-100)')
plt.scatter(df['Spending Score (1-100)'],df['Annual Income (k$)'])


# In[8]:


sns.set(style = 'whitegrid')
sns.distplot(df['Annual Income (k$)'])
plt.title('Distribution of Annual Income', fontsize = 20)
plt.xlabel('Range of Annual Income')
plt.ylabel('Count')
plt.show()


# In[9]:


sns.set(style = 'dark')
sns.distplot(df['Age'])
plt.title('Distribution of Age', fontsize = 20)
```

```
plt.xlabel('Range of Age')
plt.ylabel('Count')
plt.show()


# In[10]:


sns.set(style = 'dark')
sns.distplot(df['Spending Score (1-100)'])
plt.title('Distribution of Spending Score', fontsize = 20)
plt.xlabel('Range of Spending Score')
plt.ylabel('Count')
plt.show()


# In[11]:


#Count and plot gender
sns.countplot(y = 'Gender', data = df, palette="husl", hue = "Gender")
df["Gender"].value_counts()


# In[12]:


#Pairplot with variables we want to study
sns.pairplot(df, vars=["Age", "Annual Income (k$)", "Spending Score (1-100)"],  kind ="reg",
hue = "Gender", palette="husl", markers = ['o','D'])
```

```
# # Spending Score According to Age

# In[13]:


sns.lmplot(x = "Age", y = "Annual Income (k$)", data = df, hue = "Gender")

# In[14]:


sns.lmplot(x = "Annual Income (k$)", y = "Spending Score (1-100)", data = df, hue = "Gender")


# In[15]:


sns.lmplot(x = "Age", y = "Spending Score (1-100)", data = df, hue = "Gender")


# # Selecting Number of Clusters

# In[16]:


from sklearn.cluster import KMeans

#Creating values for the elbow
X = df.loc[:,["Age", "Annual Income (k$)", "Spending Score (1-100)"]]
inertia = []
k = range(1,20)
for i in k:
    means_k = KMeans(n_clusters=i, random_state=0)
```

```
    means_k.fit(X)
    inertia.append(means_k.inertia_)


# In[17]:


#Plotting the elbow
plt.title("Elbow Method",fontsize=20)
plt.plot(k , inertia , 'bo-')
plt.xlabel('Number of Clusters') , plt.ylabel('Inertia')
plt.show()


# # Clustering

# In[18]:


#Creating and fitting the model kmeans with 5 clusters
model = KMeans(n_clusters=5, random_state=0)
model.fit(X)
labels = model.labels_
centroids = model.cluster_centers_


# In[19]:


import plotly as py
import plotly.graph_objs as go
```

```
#Create a 3d plot to view the data sepparation made by Kmeans
trace1 = go.Scatter3d(
    x= X['Spending Score (1-100)'],
    y= X['Annual Income (k$)'],
    z= X['Age'],
    mode='markers',
     marker=dict(
        color = labels,
        size= 10,
        line=dict(
            color= labels,
        ),
        opacity = 0.9
     )
)
layout = go.Layout(
    title= 'Clusters',
    scene = dict(
            xaxis = dict(title  = 'Spending Score (1-100)'),
            yaxis = dict(title  = 'Annual Income (k$)'),
            zaxis = dict(title  = 'Age')
        )
)
fig = go.Figure(data=trace1, layout=layout)
py.offline.iplot(fig)


# In[ ]:
```

CONCLUSION

K means Clustering implies grouping is one of the most famous bunching calculations and as a rule the principal thing specialist apply when tackling grouping assignments to get a thought of the structure of the dataset. The objective of K means Clustering implies is to bunch information focuses into particular non-covering subgroups. One of the significant utilizations of K means clustering implies grouping is division of clients to improve comprehension of them which thus could be utilized to build the income of the organization. On line shopping does not allow the customers to touch the product. They have to attract the customers by showing images, photos and other displays. Data Mining is a most powerful tool to discover knowledge from the database. In this work, the data set is normalized to produce improved results.