**Teen Marketing Classification**

**Problem Statement**

The many millions of teenage consumers using social networking sites have attracted the attention of marketers struggling to find an edge in an increasingly competitive market. One way to gain this edge is to identify segments of teenagers who share similar tastes, so that clients can avoid targeting advertisements to teens with no interest in the product being sold.

Build K-means clustering model to classify the interest of teenagers by using various attributes.

**Data Description**

The SNS dataset contains 30000 observations (rows) each represents a high school student and 40 features (columns) that provides information for the student.

1 - **Gradyear:** Graduation year of the student (2006, 2007, 2008, 2009)

2 - **Gender:** Gender of the student (male, female)

3 - **Age:** Age of the student

4 - **Friends:** No of friends

For clustering, 36 words were chosen to represent five categories of interests: namely extracurricular activities, fashion, religion, romance, and antisocial behaviour. The 36 words include terms such as football, sexy, kissed, bible, shopping, death, and drugs. The final dataset indicates, for each person, how many times each word appeared in the person’s SNS profile.

**Word list**

5 - basketball

6 - football

7 - soccer

8 - softball

9 - volleyball

10 - swimming

11 - cheerleading

12 - baseball

13 - tennis

14 - sports

15 - cute

16 - sex

17 - sexy

18 - hot

19 - kissed

20 - dance

21 - band

22 - marching

23 - music

24 - rock

25 - god

26 - church

27 - jesus

28 - bible

29 - hair

30 - dress

31 - Blonde

32 - mall

33 - shopping

34 - clothes

35 - hollister

36 - abercrombie

37 - die

38 - death

39 - drunk

40 - drugs

**Evaluation**

Evaluation will be based on:

* Feature Selection
* Model Comparison
* Cluster Selection

**Feature Selection**

Select the right features based on importance and significance.

**Model Comparison**

Use k-means to build multiple cluster.

**Cluster Selection**

Select the cluster using the Elbow method.

**Expected Output**

Model should be able to predict the cluster an observation belongs to.