Problem statement:

ABC is an online content sharing platform that enables users to create, upload and share the content in the form of videos. It includes videos from different genres like entertainment, education, sports, technology and so on. The maximum duration of video is 10 minutes.

Users can like, comment and share the videos on the platform.

Based on the user's interaction with the videos, engagement score is assigned to the video with respect to each user. Engagement score defines how engaging the content of the video is.

Understanding the engagement score of the video improves the user's interaction with the platform. It defines the type of content that is appealing to the user and engages the larger audience.

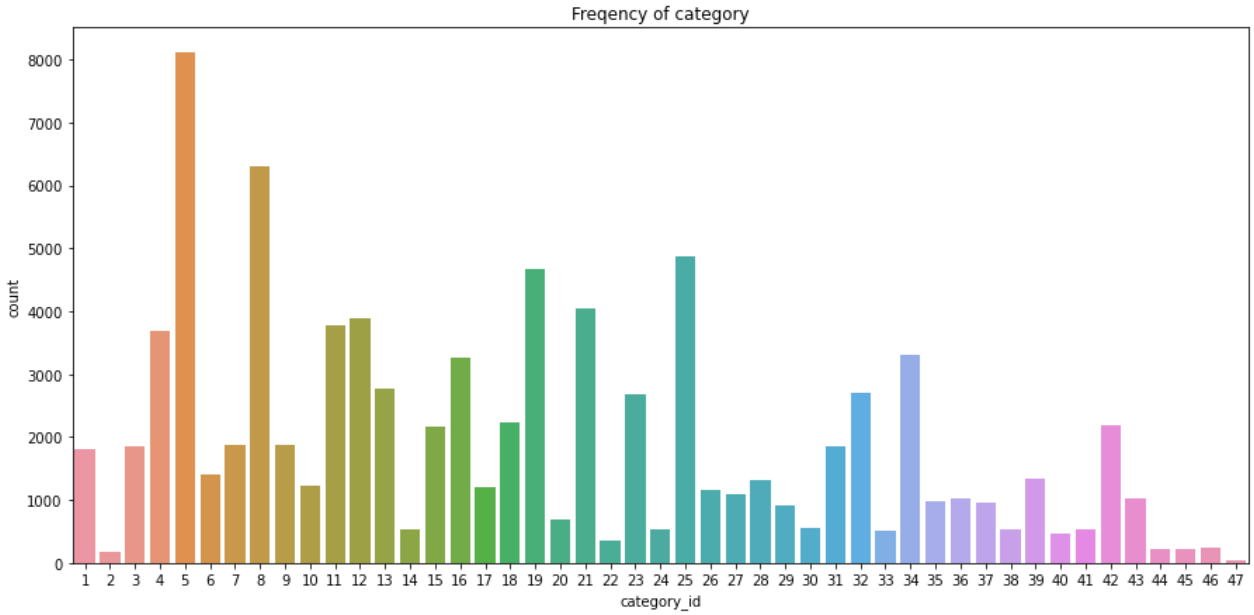
Objective

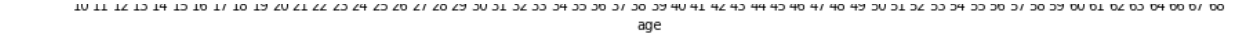
The main objective of the problem is to develop the machine learning approach to predict the engagement score of the video on the user level.

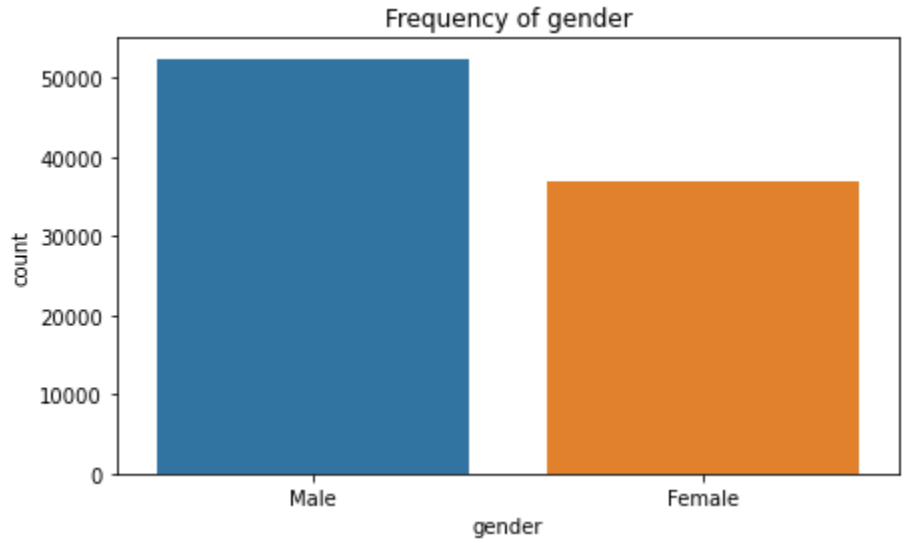
Solution

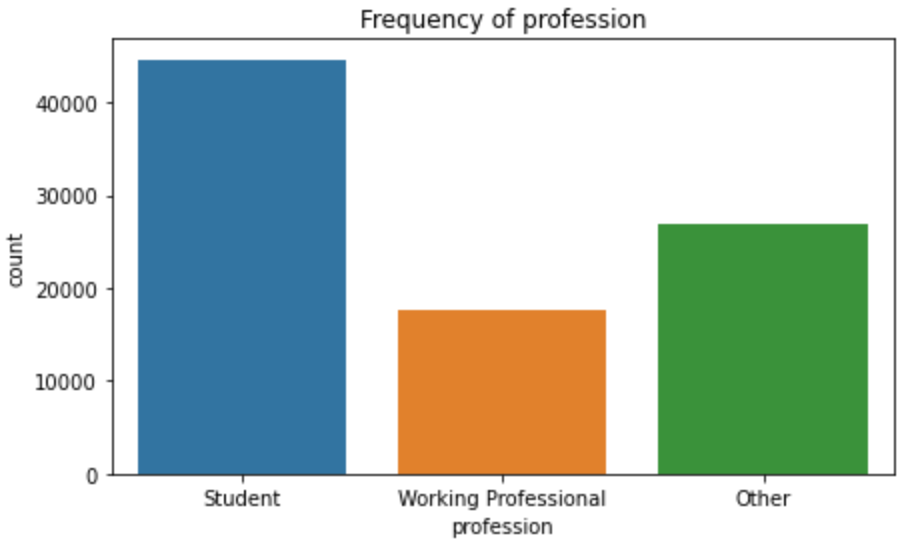
About dataset:

Univariate analysis

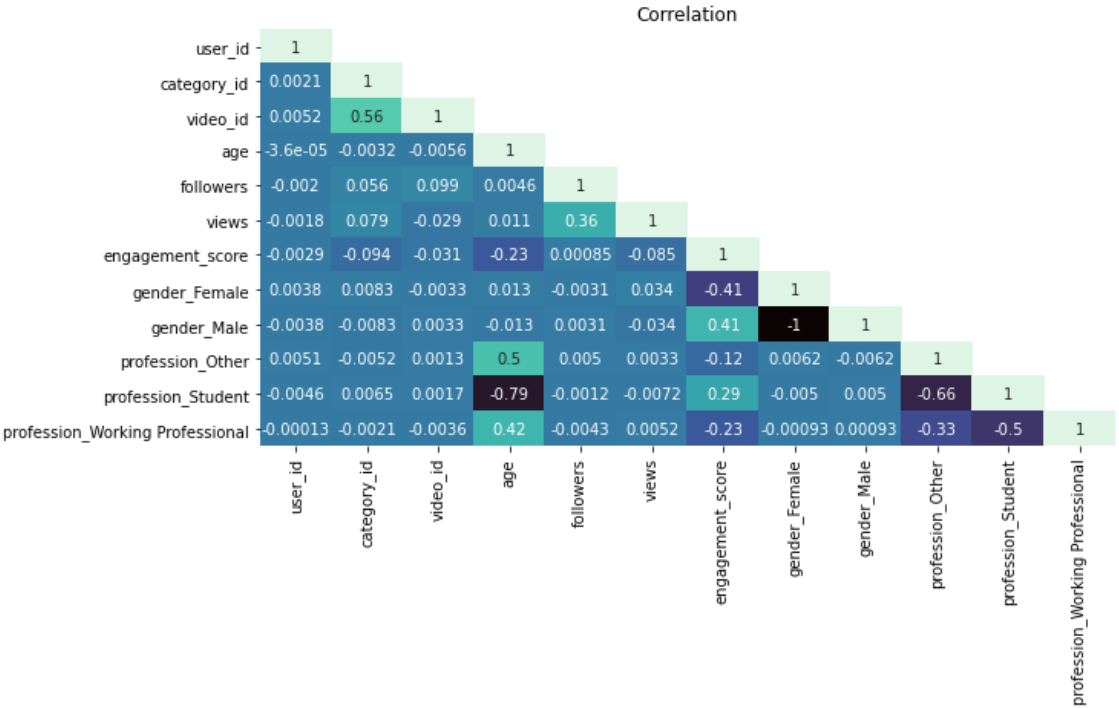

Freqency of category
count
category_id


age


Frequency of gender
count
Male
Female
gender



Multivariate analysis



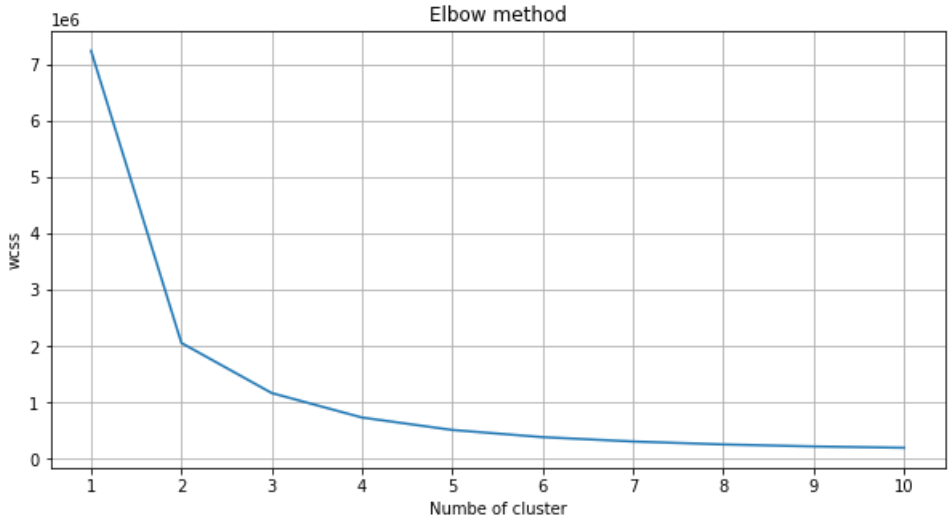
Checking the VIF to check if there is multicollinearity

	Variable	vif
0	user_id	1.000113
1	category_id	1.471325
2	video_id	1.482423
3	age	2.675225
4	followers	1.168328
5	views	1.174796
6	gender_Female	inf
7	gender_Male	inf
8	profession_Other	inf
9	profession_Student	inf
10	profession_Working Professional	inf

I have group the users by creating clusters based on age, gender and profession.

K-means algorithm is used to create clusters.

After evaluating the elbow curve 3 numbers of cluster is selected



Also, age is further grouped in below bins.

['4-13','10-18','18-25','25-35','35-45','45-70']