**CSCI 620/Section 03/Mior, Introduction to Big Data, Spring 2215 Assignment 8 – Clustering**

**Submitted by – Prakhar Gupta(pg9349)**

**Ans1**

(Refer to query.txt and q1.py)

**Ans2**

(Refer to q2.py)

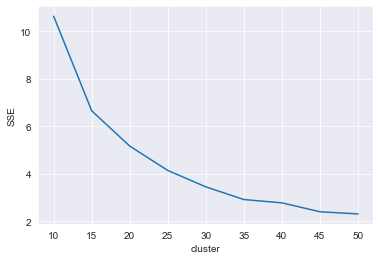
**Ans3**

(Refer to q3.py)

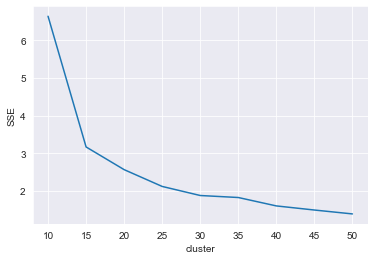
**Ans4**

(Refer to q4.py and (q4\_2.ipynb or q4\_2.py) ) Also have added CSV files containing the SSE of the code run

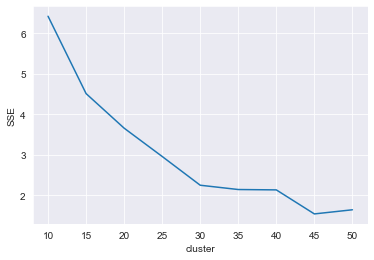
Action



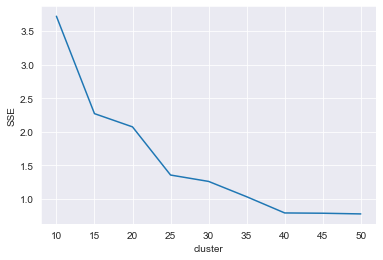
Horror



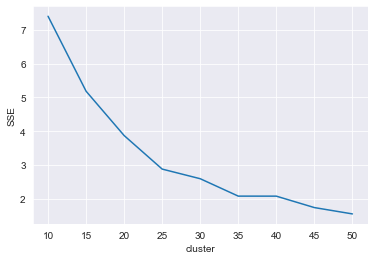
Romance



SciFi



Thriller



**Ans5**

(Refer to q5.py and (q5\_2.ipynb or q5\_2.py))

Optimal clusters size as per elbow plots

Action - 25

Horror - 15

Romance - 30

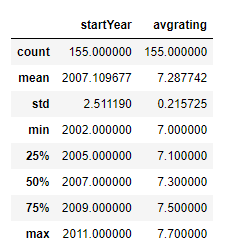
Sci-Fi - 25

Thriller - 35

***Action***

Example cluster No: 20

Descriptive Stats of cluster 20



We can see for this cluster the start year is between 2002 and 2011 and average rating are 7 to 7.7.

It has 155 movies in total

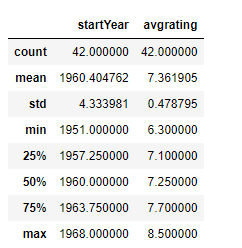
This cluster include movies like –

**'300', 'The Adventures of Tintin', "Pirates of the Caribbean: Dead Man's Chest", "Pirates of the Caribbean: At World's End"**

***Horror***

Example cluster No: 1

Descriptive Stats of cluster 1



We can see for this cluster the start year is between 1951 and 1968 and average rating are 6.3 to 8.5. It has 42 movies in total

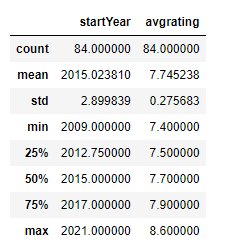
This cluster include movies like –

**'Carnival of Souls', 'What Ever Happened to Baby Jane?', 'The Birds', 'The Haunting', 'The Raven', 'Black Sabbath'**

***Romance***

Example cluster No: 27

Descriptive Stats of cluster 27



We can see for this cluster the start year is between 2009 and 2021 and average rating are 7.4 to 8.6.

It has 84 movies in total

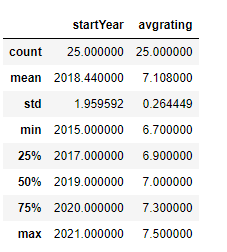
This cluster include movies like –

**'Gully Boy', 'Arjun Reddy', 'Sanam Teri Kasam', 'Shaadi Mein Zaroor Aana', 'Victoria', 'Bareilly Ki Barfi'**

***Sci-Fi***

Example cluster No: 3

Descriptive Stats of cluster 3



We can see for this cluster the start year is between 2015 and 2021 and average rating are 6.7 to 7.5.

It has 25 movies in total

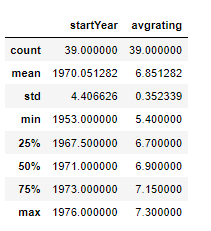
This cluster include movies like –

**'Passengers', 'Upgrade', 'Ready Player One', 'Tenet', 'Finch', "Don't Look Up"**

***Thriller***

Example cluster No: 32

Descriptive Stats of cluster 32



We can see for this cluster the start year is between 1953 and 1976 and average rating are 5.4 to 7.3.

It has 39 movies in total

This cluster include movies like –

**'The Odessa File', 'The Eagle Has Landed', 'Cul-de-sac', 'Mean Streets', 'THX 1138', 'The Chase', 'Vanishing Point', "On Her Majesty's Secret Service", 'Diamonds Are Forever'**