DATA SCIENCE K-MEANS LAB

Roll no: 20K-0409

Screen Shots

DATAset#1

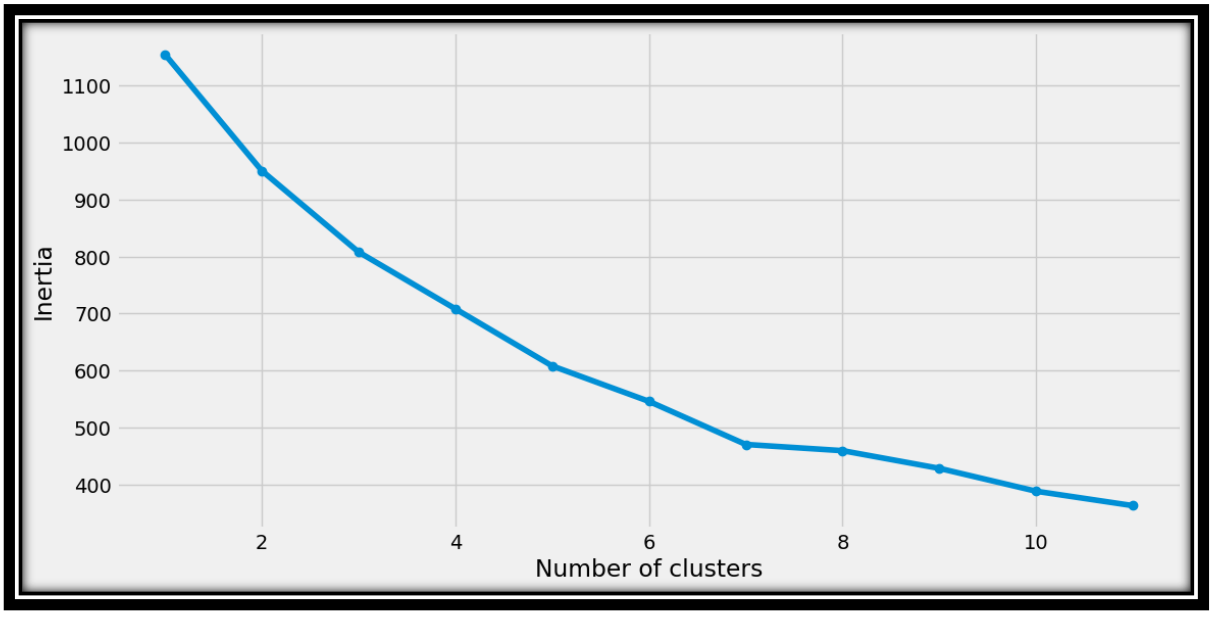
1st dataset: 80 cereal

https://www.kaggle.com/datasets/crawford/80-cereals

```
[133] 1 # reading the data and looking at the first five rows of the data
      2 data1=pd.read_csv("cereal.csv")
      3 data1.head()
```

	name	mfr	type	calories	protein	fat	sodium	fiber	carbo	sugars	potass	vitamins	shelf	weight	cups	rating
0	100% Bran	N	C	70	4	1	130	10.0	5.0	6	280	25	3	1.0	0.33	68.402973
1	100% Natural Bran	Q	C	120	3	5	15	2.0	8.0	8	135	0	3	1.0	1.00	33.983679
2	All-Bran	K	C	70	4	1	260	9.0	7.0	5	320	25	3	1.0	0.33	59.425505
3	All-Bran with Extra Fiber	K	C	50	4	0	140	14.0	8.0	0	330	25	3	1.0	0.50	93.704912
4	Almond Delight	R	C	110	2	2	200	1.0	14.0	8	-1	25	3	1.0	0.75	34.384843

ELBOW graph for cereal dataset , cluster 4



DataSet #2

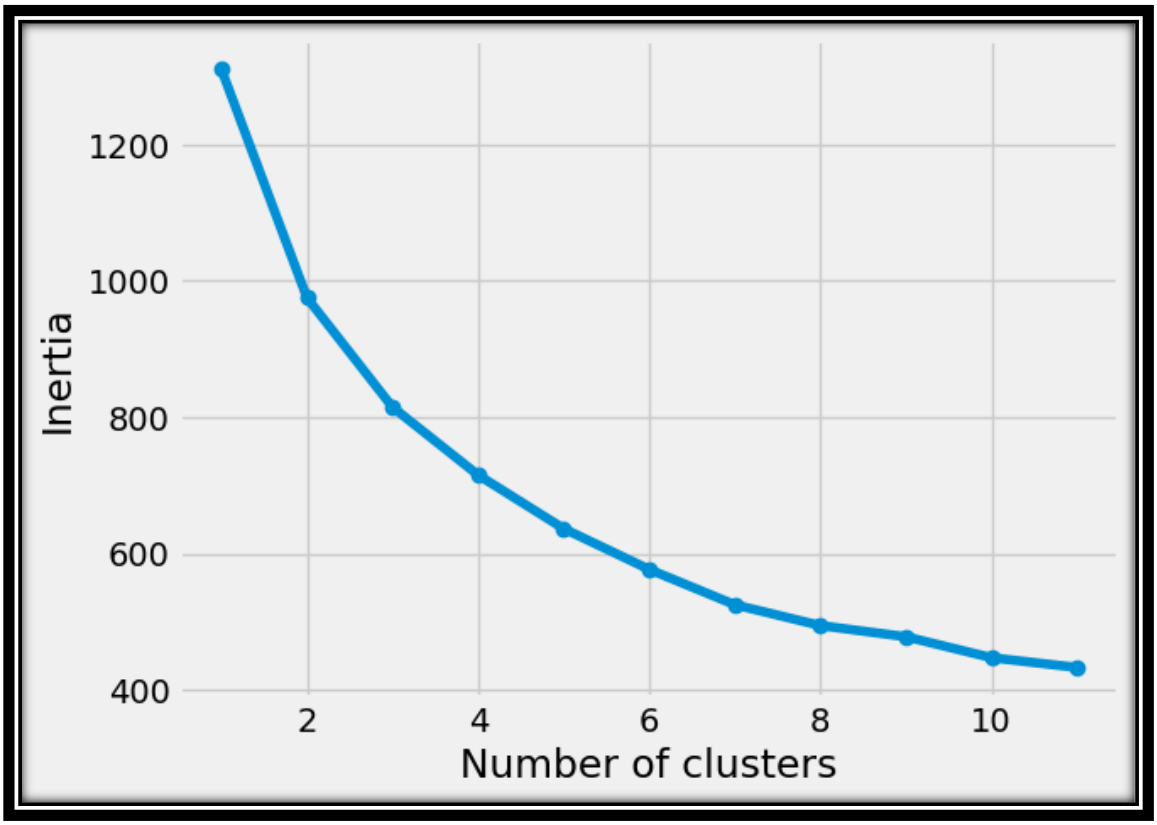
2nd data set : using groundhog day forecast temperature dataset

https://www.kaggle.com/datasets/groundhogclub/groundhog-day

```
[54] 1 # reading the data and looking at the first five rows of the data
     2 data2=pd.read_csv("ground_day.csv")
     3 data2.head()
```

	Year	Punxsutawney Phil	February Average Temperature	February Average Temperature (Northeast)	February Average Temperature (Midwest)	February Average Temperature (Pennsylvania)	March Averag Temperatur
0	1886	No Record	NaN	NaN	NaN	NaN	Na
1	1887	Full Shadow	NaN	NaN	NaN	NaN	Na
2	1888	Full Shadow	NaN	NaN	NaN	NaN	Na
3	1889	No Record	NaN	NaN	NaN	NaN	Na
4	1890	No Shadow	NaN	NaN	NaN	NaN	Na

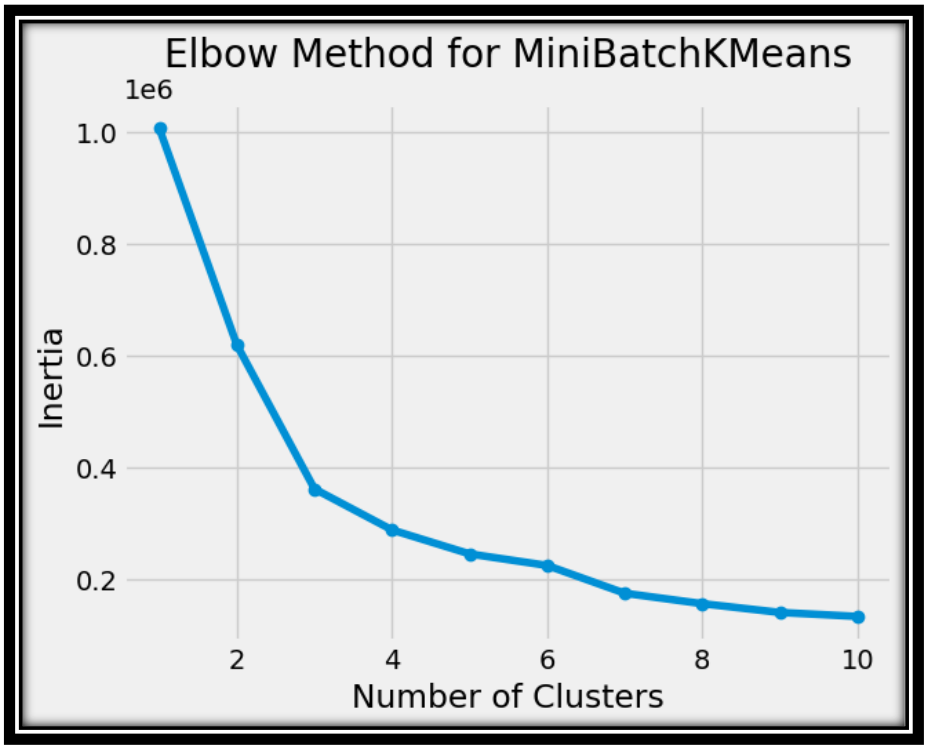
Elbow Graph for ground day hog: cluster : 4



How will you incorporate the streaming data into your already existing clusters?

Using MiniBatchKMeans

For cereal:



For ground day hog: