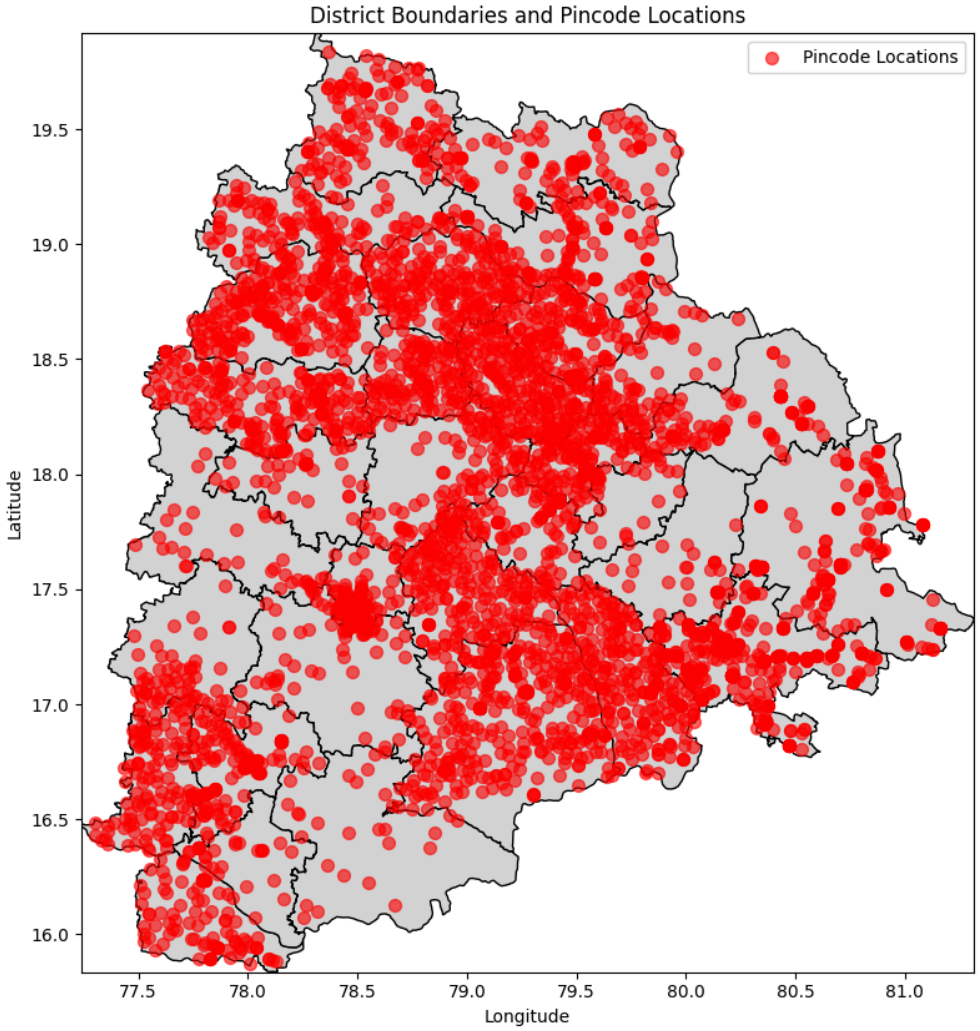
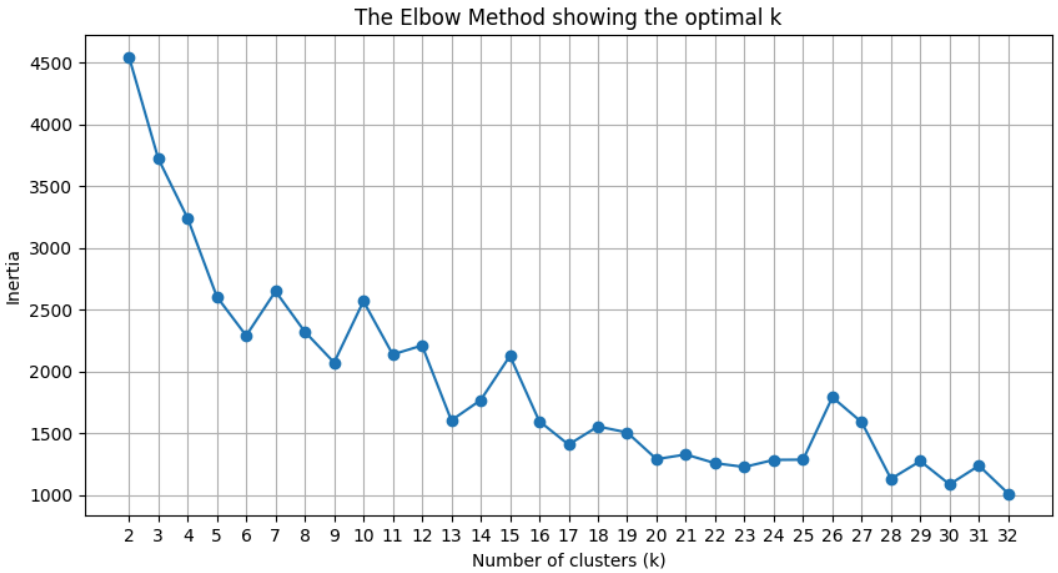
RESULTS :

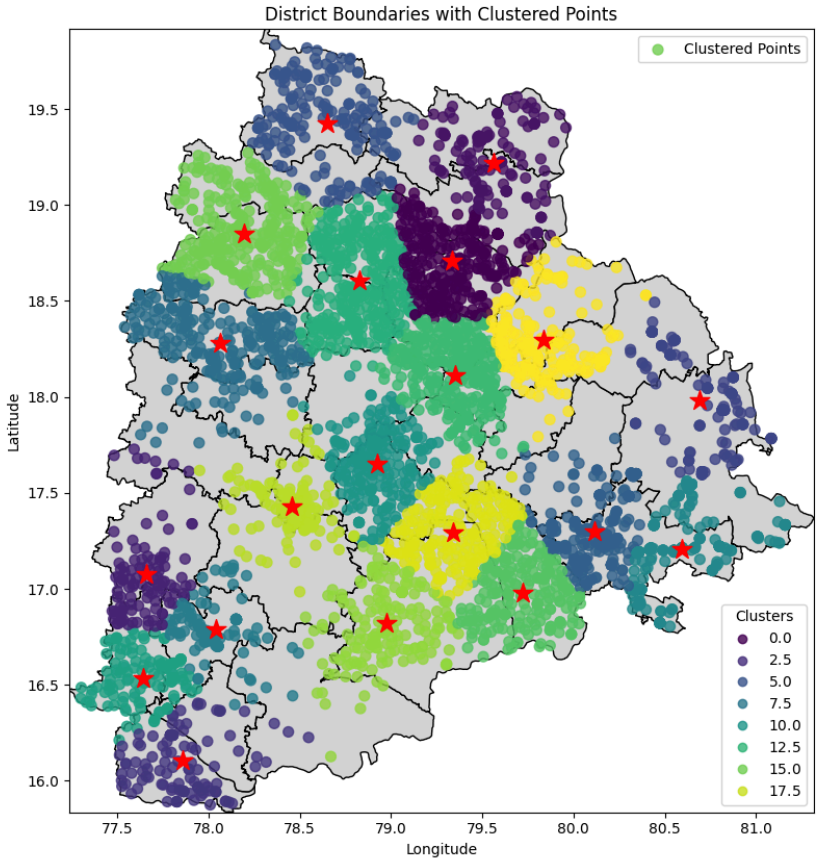
DATA VISUALISATION :



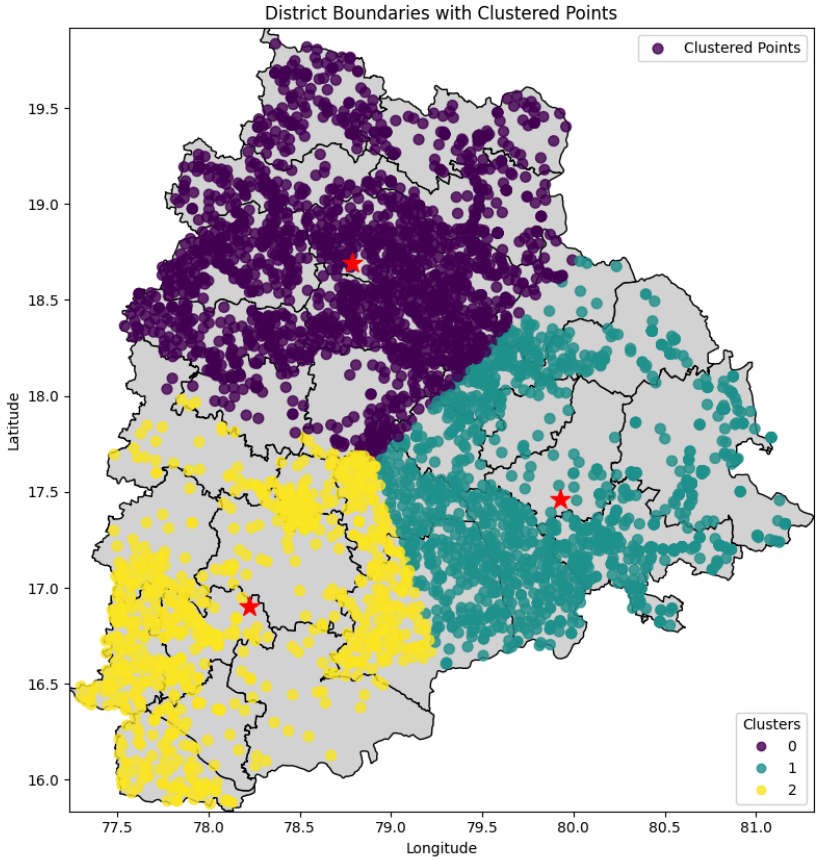
CLUSTERING ANALYSIS:



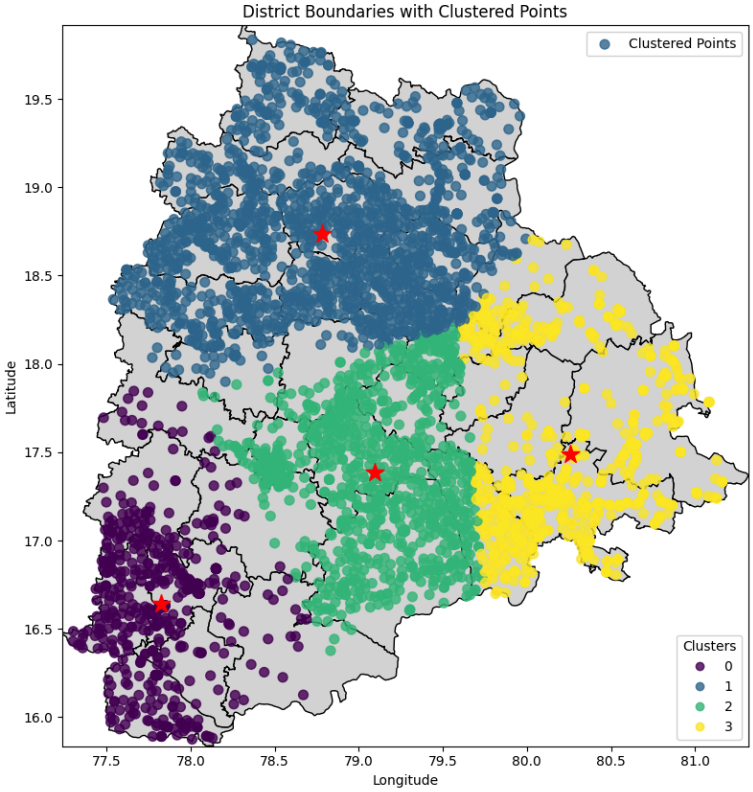
K = 20 looked best



K = 3


District Boundaries with Clustered Points
Clustered Points
Clusters
0
1
2
Latitude
Longitude
19.5
19.0
18.5
18.0
17.5
17.0
16.5
16.0
77.5
78.0
78.5
79.0
79.5
80.0
80.5
81.0

K=4



After comparing the map with a telangana map with districts :

I chose k = 20 from the elbow plot,

When I plotted with 3 centroids (k=3):

I found it looked like a cake being correctly divided into 3 parts. 😅

It was evenly divided, where each centroid had 1/3 rd of the points, giving an idea that the population of telangana is distributed evenly in those 3 regions.

When I plotted with 4 centroids (k=4):

After comparing the maps with k = 3 and k = 4, the northernmost centroid in the map(k=3) did not change much in comparison to that of the map with k = 4.

the rest 2/3 rds of the map(k = 3) was divided into 3 parts.

This tells us that the population density in the north of Telangana is high and there is high density of pincodes. In Urban area a pin may cover smaller region. This tells us that Northern Telangana is more urbanized compared to Southern Telangana.

While in rural areas pin can cover larger areas. This tells us that Middle and eastern parts of Telangana are rural.

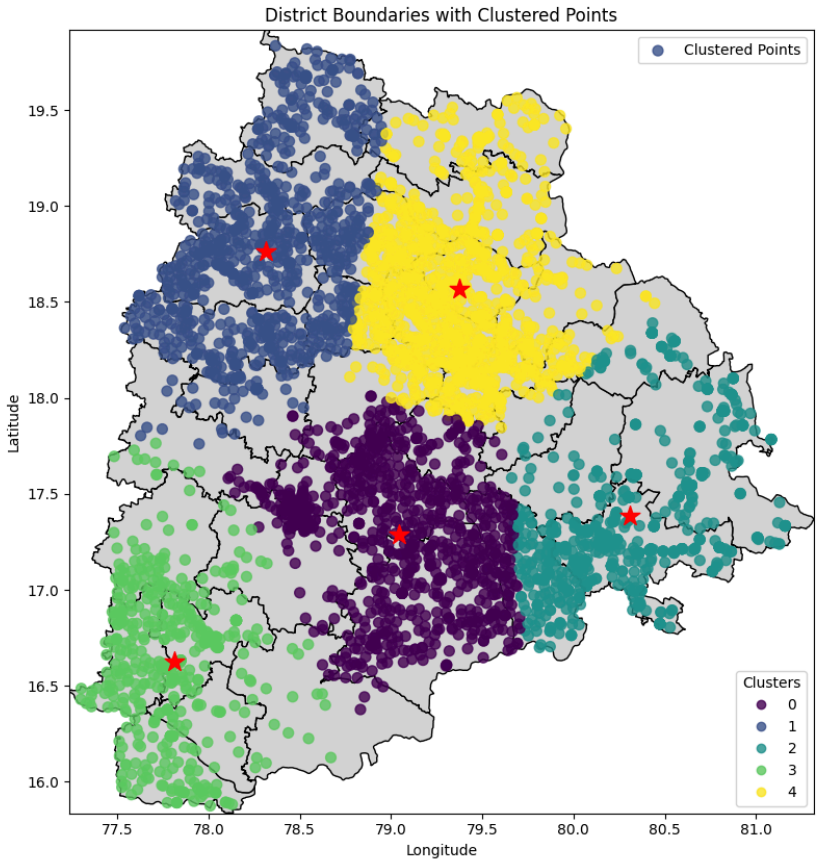
When I plotted k = 5, The northern region which finally split into 2 regions.

From k = 6, it became harder to analyze this data.

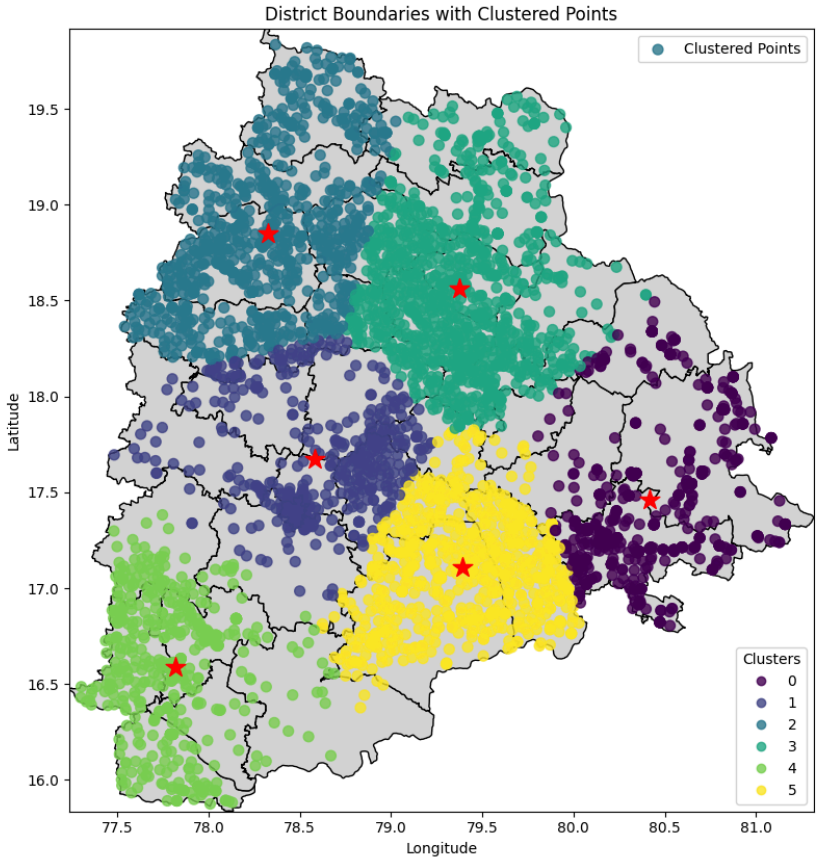
So, I used the elbow method and came to the conclusion that k = 20, fit best as all other values after that did not show any good change.

These are my inferences.

K = 5



K = 6



CV + NLP

Data interpreted from the images stored in a csv file:

F10

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	I AM REALLY ANNOYED BY YOUR CONSTANT COMPLAINING AND YOU NEVER OFFER ANY SOLUTIONS WHICH IS VERY UNHELPFUL AND NEGATIVE															
2	IT IS FRUSTRATINS THAT YOU NEVER PAY ATTENTION DURING DISCUSSIONS AND YOUR LACK OF FOCUS IS REALLY AFFECTING OUR PROGREOS															
3	I AM DELIGHTED BY YOUR FRIENDLINESS AND YDU ALWAYS MAKE EVERYONE FEEL WELCDME WHICH FOSTERS A SENSE OF COMMUNITY															
4	IT IS WONDERFUL THAT YOU ALWAYS SHOW KINDNESS AND YOUR EMPATHY TOWARDS OTHERS IS TRULY HEARTWARMINJ AND APPRECIATED															
5	YOUR ANALYSIS OF THE DATA WAS ACCURATE AND WELL PRESENTED PROVIDING A CLEAR UNDERSTANDING OF THE TRENDS AND PATTERNS															
6	THE MEETING MINUTES YOU PREPARED WERE DETAILED AND WELL ORGANIZED ACCURATELY REFLECTING THE DISCUSSIONS AND DECISIONS MADE															

Sentiment Analysis:

	A	B
	line	sentiment
	I AM REALLY ANNOYED BY YOUR CONSTANT COMPLAINING AND YOU NEVER OFFER ANY SOLUTIONS WHICH IS VERY UNHELPFUL AND NEGATIVE	Angry
	IT IS FRUSTRATIWG THAT YOU NEVER PAY ATTENTION DURING DISCUSSIONS AND YOUR LACK OF FOCUS IS REALLY AFFECTING OUR PROGREOS	Angry
	I AM DELIAHTED BY DOUR FRIENDLINESS AND YOU ALWAYS MAKE EVERYONE FEEL WELCOME WHICH FOSTERS A SENSE OF COMMUNITY	Happy
	IT IS WONDERFUL THAT YOU ALWAYS SHOW KINDNESS AND YOUR EMPATHY TOWARDS OTHERS IS TRULY HEARTWARMINT AND APPRECIATED	Happy
	YOUR ANALYSIS OF THE DATA WAS ACCURATE AND WELL YRESENTED PROVIDING A CLEAR UNOERSTANDING OF THE TRENDS AND PATTERNS	Neutral
	THE MEETING MINUTES YOU PREPARED WERE DETAILED AND WELL ORGANIZED ACCURATELY REFLECTING THE DISCUSSIONS AND DECISIONS MADE	Neutral

Methods :

I used the Elbow method (k_best_elbow.py) to come up with the best k value. The plot is in the k values folder.

For image processing, I cut the given image into 28*28 pixels to obtain the result.

I used Naive Bayes theorem to do sentiment analysis.

Reference links:

https://www.youtube.com/watch?v=5w5iUbTlpMQ
https://www.analyticsvidhya.com/blog/2022/03/building-naive-bayes-classifier-from-scratch-to-perform-sentiment-analysis/
https://www.youtube.com/watch?v=MV6dgAGTv-k&t=4s