Voter's Assessment

Submitted by: Ashish Kumar Rout
Roll No.: 00411624417

Content

- Company profile
- Problem Statement
- Objective
- Description
- Technology used
- Data flow diagram
- Screenshots
- Conclusion
- Limitations
- Future scope
- References

Company profile

- Company name: **Aam Aadmi Party**

- Aam Aadmi Party (AAP), is an Indian political party, formally launched on 26 November 2012, and is currently the ruling party of the National Capital Territory of Delhi.

- It deals in various profiles like HR, Public Speaking, Data Analytics and Machine Learning.

Problem Statement:

- It has been seen that in 2013 Vidhan sabha election that aap and congress was almost toe to toe but suddenly in 2015 vidhan sabha election they won by a large margin and the margin was achieved by the votes acquired previously by congress.

- So we wanted to analyze the voters who are most likely to swing and which age category so we did this analysis and also we wanted to find the migrants as well as analyze different families to get better idea what kind of policy they might like.

- We find the swingable voters by category using a simple frequentist approach

- What we did was first, find the frequencies in different categories of age group through the data we extracted from electoral roll.

- Then, we also find the frequencies of different political parties through the data we extracted from form 20 of the same locality and merge the two to find the swingable voters by age group in different political parties.

Objectives:

- Find out how many male and female voters are there in the given electoral roll for the polling stations. Also segregate male and female voters separately in intervals of five years.

- Obtain form 20 data for five polling stations for the last assembly or parliament election. Find out the probabilities of different age groups voting for BJP.

- Segregate voters family wise and find out the name of the head of the family, house number and number of members in that family

- Which are the voters who are likely to be migrants.

- Try to study the epic id which is voter id, system. What information can you gather out of it? Given the structure of the EPIC ID, is it possible to find out when a specific person has become a voter in this locality or when was a person's name included in the electoral roll for this polling station. A range of years would also be okay.

Technology used:

- Python 3.7 with various libraries
- Libraries:
    - Numpy
    - Matplotlib and Seaborn
    - sklearn
    - Pdf2Image, Optical Character Recognition (OCR), Python Imaging Library (PIL), OpenCV and Tesseract
- Jupyter notebook
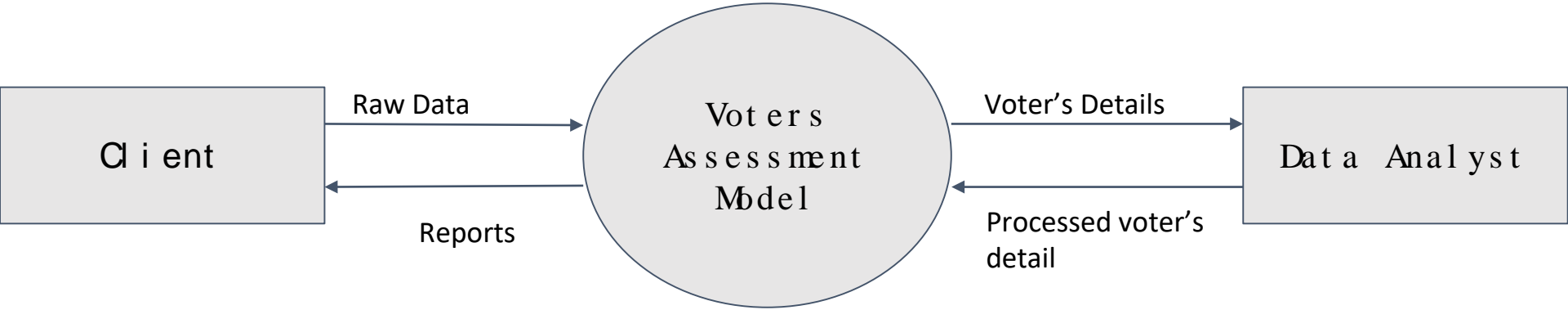- Anaconda 2018.12
- Excel 2019

Collection of data:

- We extracted electoral roll using image processing. Source of this data is election commission's website.

- Collection of features was done through web scraping.

- Optical character recognition techniques were used in image processing and geolocation API was used in web scraping.
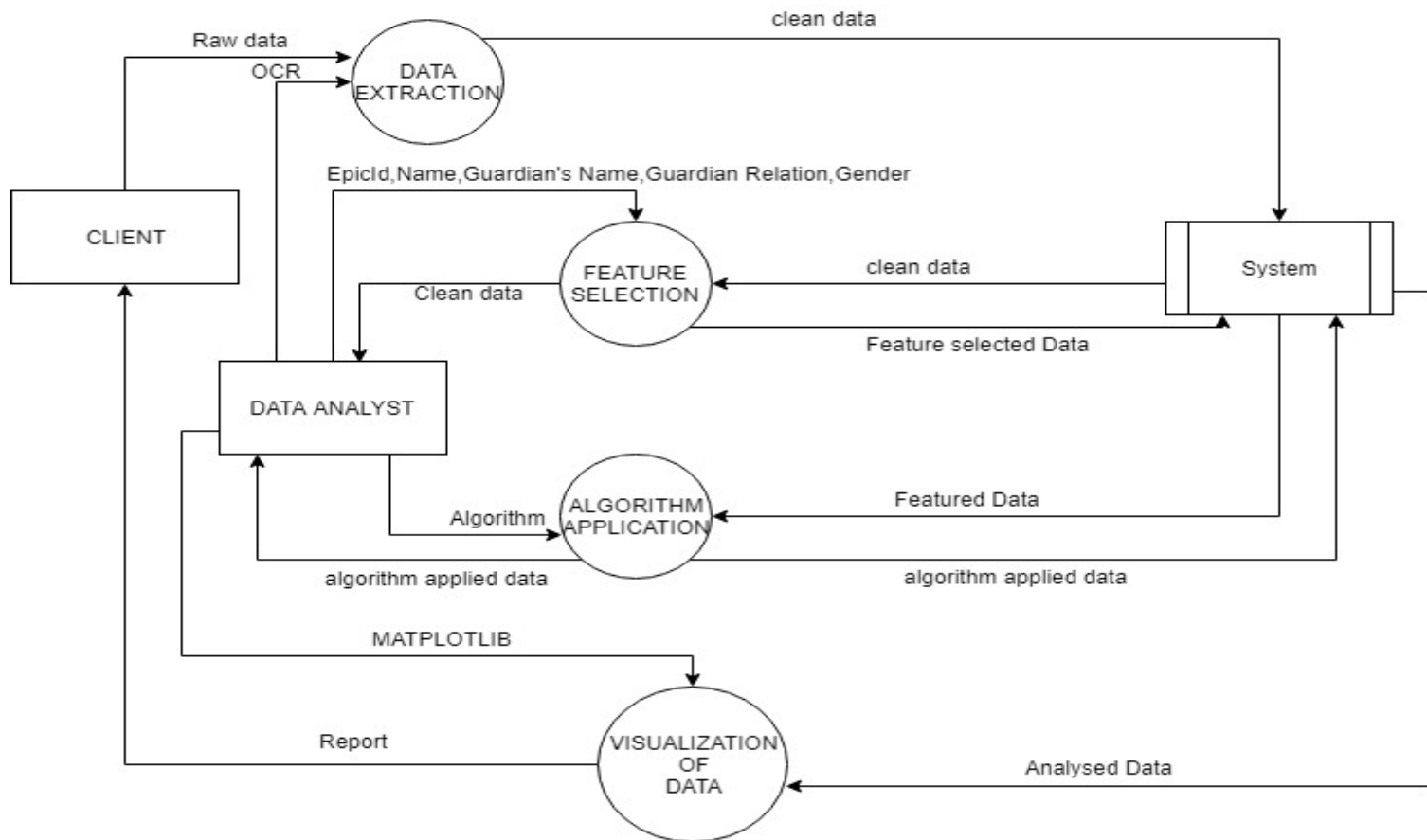
Use Case Diagram
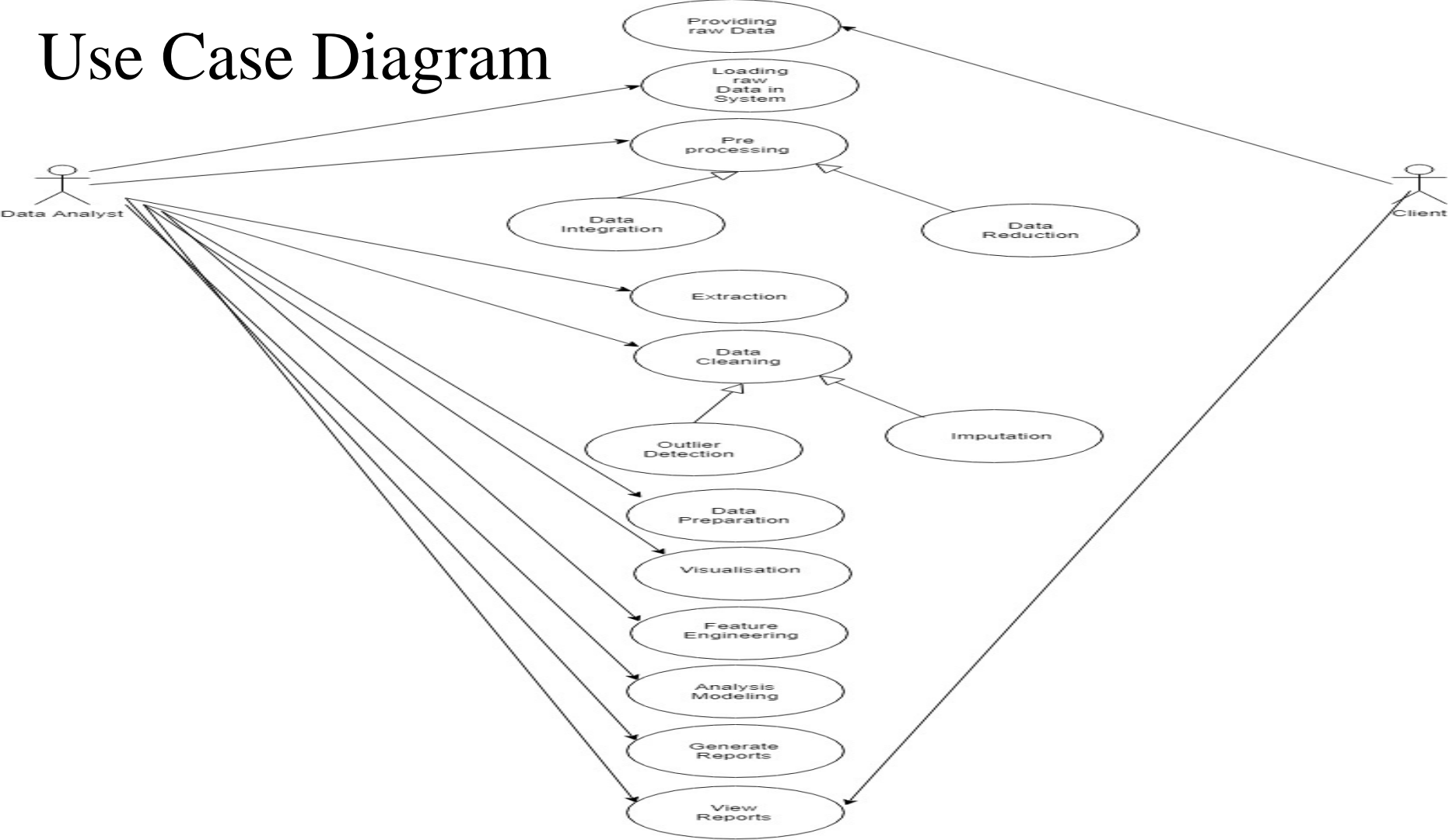
DFD: Level 0

DFD LEVEL 1

Workflow Diagram
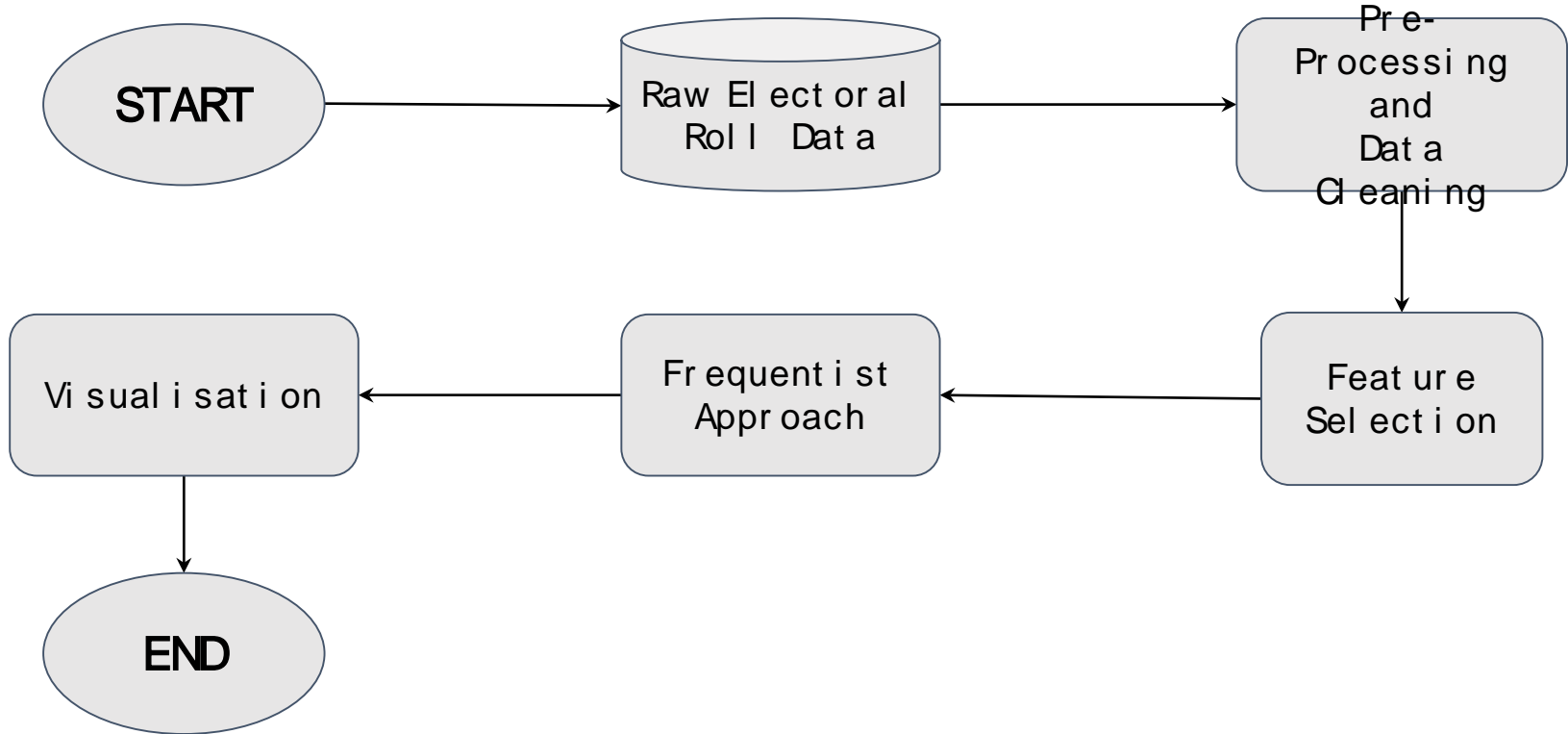
Extraction of Data

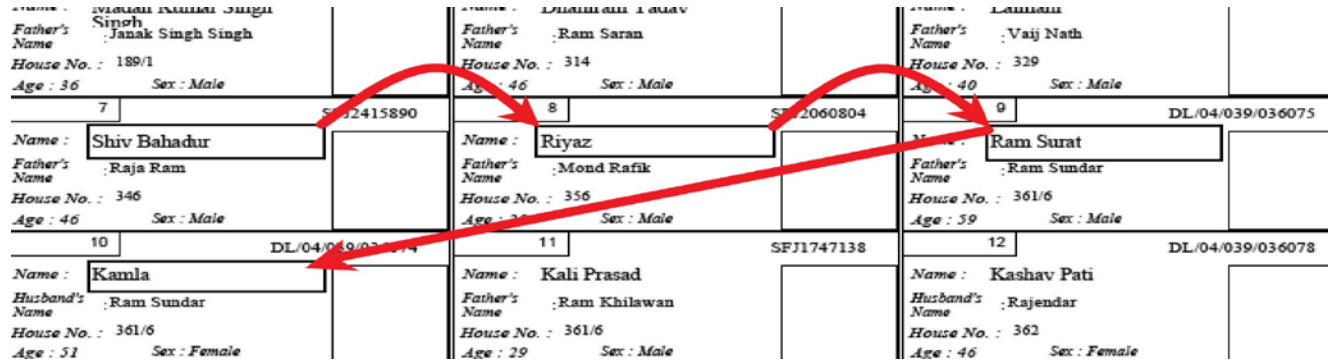- We crop part of image rectangularly and extract text of image.
- Additionally we have to extract same data 30 times. We know they have equal distance. So will hop one by one till we reach end of the page. We will do that for every feature we are extracting. Notice that in below image that we extend the size of box to max on right, that is because we are not sure what will be the length of name.

Combining Multiple files

- Used combine code to combine 500 files after extracting. We can combine tesseract 3 and 4 files easily using filenames

name	guardian	gen	g_relation	age
Pooja	Lalit Ram	F	F	20
Shanti Devi	Munna Lal	F	H	36
Rajesh Paswan	Sohan Paswan	M	F	35
Rajesh Kumar Kori	Algu Ram Kori	M	F	30
Prashuram Kimnar	Ram Dev Parsad	M	F	25
Manish Kumar	Ayodhya Prashad	M	F	32
Mazhar Ansan	Abbas Ansari	F	F	55
Bhagwati Devi	Pale Ram	F	H	70
Roshn Khatoon	Mustfa Ali	F	H	28
Sanjay Kumar	Fekan Prasad	F	F	26
Indrawati Devi	Roop Singh	F	H	51
Rupesh Sharma	Kanbaiya Lal Shamma	M	F	38
Arvind Kumar	Mangu	M	F	31
Sushila Devi	Rajeshmwar Ram	F	H	55

PROCEDURE OF ANALYSIS

- We did use Frequentist approach:
- First, we found the age group wise population through electoral data we Extracted.
- Then, using form 20 data we find the likelihood of major parties.
- After that, we combine the age group wise likelihood and major parties likelihood to come up with likelihood of different age group voting for different parties.
- We also found out about the Religion of every person using their surnames.
- Finding about which person is likely to be a Migrant.

```
Age_group
18-23      85
23-28     140
28-33     158
33-38     142
38-43     115
43-48     112
48-53      96
53-58      82
58-63      38
63-68      26
68-73      22
73-78      20
78-83      10
83-88       4
93-98       1
```
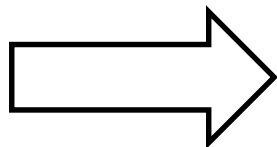


```
Age_group
18-23     0.080875
23-28     0.133206
28-33     0.150333
33-38     0.135109
38-43     0.109420
43-48     0.106565
48-53     0.091342
53-58     0.078021
58-63     0.036156
63-68     0.024738
68-73     0.020932
73-78     0.019029
78-83     0.009515
83-88     0.003806
93-98     0.000951
```
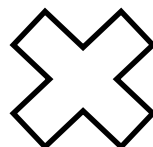


```
Congress    0.416546
BJP         0.194485
AAP         0.364296
```
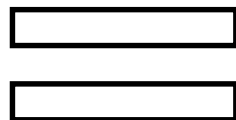

	CONGRESS	BJP	AAP
18-23	0.680838	1.57831	5.74586
23-28	1.12138	2.59957	9.46378
28-33	1.26556	2.93379	10.6805
33-38	1.1374	2.6367	9.59897
38-43	0.921134	2.13536	7.77382
43-48	0.897105	2.07965	7.57102
48-53	0.768947	1.78256	6.48945
53-58	0.656809	1.5226	5.54307
58-63	0.304375	0.705596	2.56874
63-68	0.208256	0.482776	1.75756
68-73	0.176217	0.408503	1.48716
73-78	0.160197	0.371366	1.35197
78-83	0.0800986	0.185683	0.675984
83-88	0.0320395	0.0742733	0.270394
93-98	0.00800986	0.0185683	0.0675984

Voter's trend in 2013
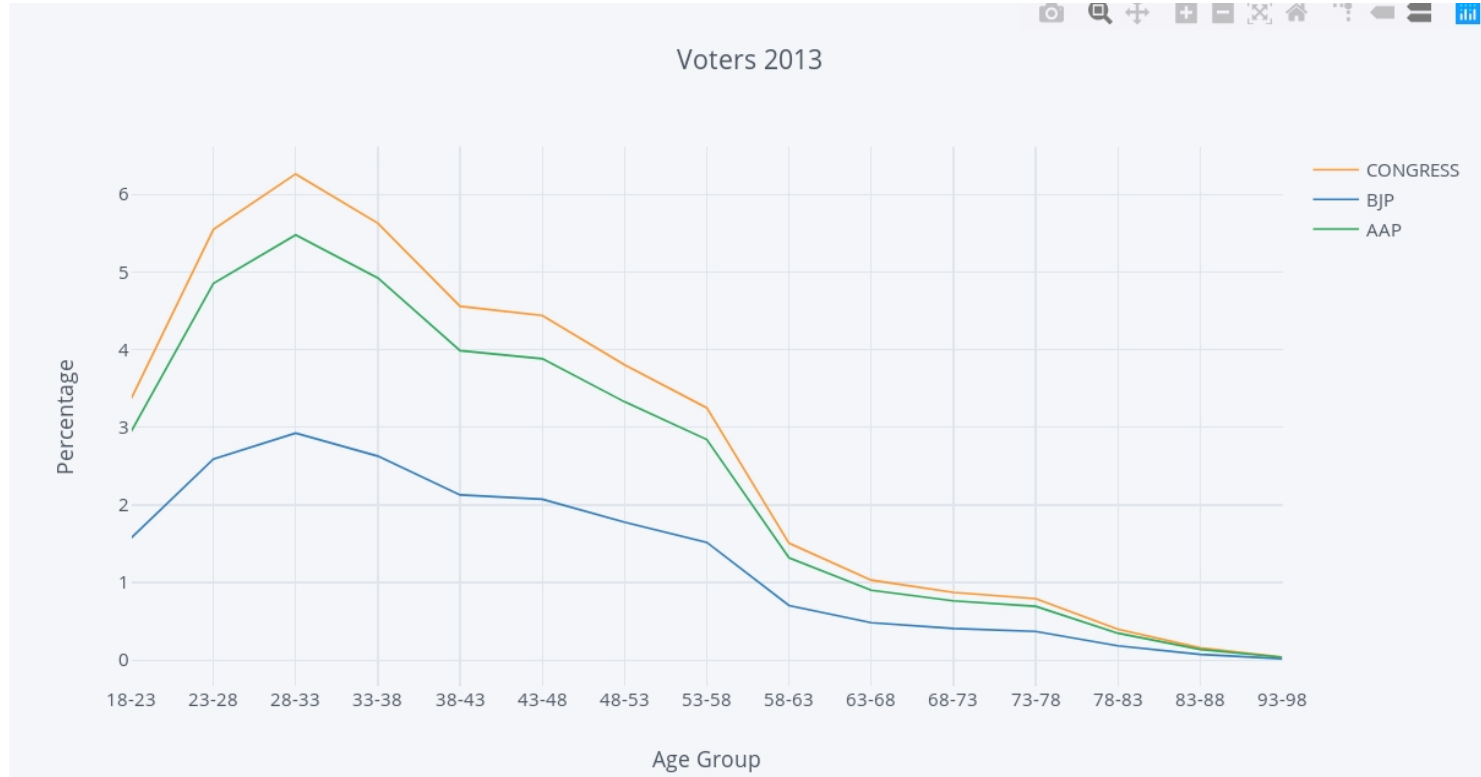
Voter's trend in 2015
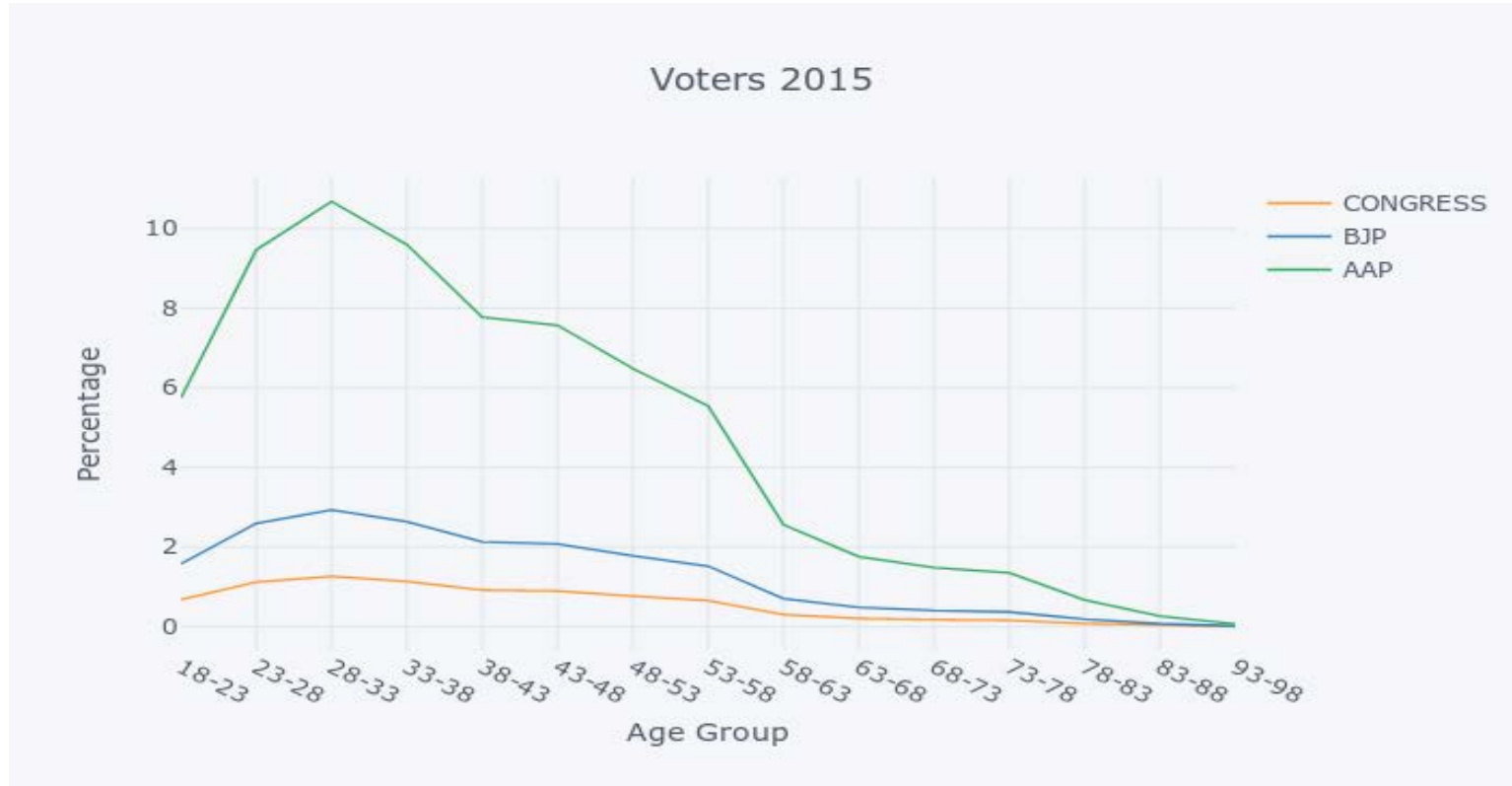
Head of the family

Name	Gender	Age	ardian	elation	amily_H	Addre	st_guar	Religic	Age_gro	EID
Mahesh K	Male	43	Rati Ram	Father's	Head	15/793	Ram	Hindu	38-43	199123(
Mohan Ku	Male	33	Rajender	Father's	Head	16/41-E	Yadav	Hindu	28-33	1644268
Jitnder Ku	Male	36	Ashish Kur	Father's	Head	16/147E	Kumar	Hindu	33-38	1805951
Gaytri	Female	39	Rajesh Ku	Husband's		16/164E	Kumar	Hindu	38-43	1162544
Rajesh Ku	Male	39	Om Praka	Father's	Head	16/164E	Prakash	Hindu	38-43	814566
Prateek W	Male	25	Harish Kur	Father's		16/164E	Wadhwa	Hindu	23-28	1610062
Imran	Male	24	Mehandi I	Father's	Head	16/172-E	Hassan	Hindu	23-28	1645928
Ilyas	Male	22	M Mehan(	Father's		16/172-E	Hassan	Hindu	18-23	160602!

Family Analysis

Part Number	SrN	EPIC ID	Gender	Age	Relationship	Address	First Name	Last Name	R_First Name	R_Last Name	Religion	Family S
AC0230001	9	RJN1816742	Male	24	Father's	13/5	Mohd	Nasim	Mohd	Azij	Others	1
AC0230001	14	RJN1673656	Male	25	Father's	16/107	Md	Jabir	Md	Maksud	Others	1
AC0230001	16	RJN1648500	Male	60	Father's	16/152-E	Ainul	Haque	Abdul	Shakoor	Others	1
AC0230001	19	RJN1673630	Female	32	Husband's	16/237	Zinat		Md	Ashfak	Others	1
AC0230001	24	RJN1815025	Female	32	Husband's	16/337	Yasmin		Rais	Ahmad	Others	1
AC0230001	32	RJN1554731	Male	47	Father's	16/436/E	Mohamad	Moti	Md	Jalil	Others	1
AC0230001	40	RJN1769247	Female	28	Husband's	16/446	Shebbun		Kuddush		Others	2
AC0230001	41	RJN1730209	Male	26	Father's	16/446	Kuddush	Mohd	Sokat		Others	
AC0230001	61	RJN1989431	Male	26	Father's	16/466-E	Ranjit	Norbert	Norbert		Others	1
AC0230001	66	RJN1765064	Female	24	Father's	16/487	Jyoti		Lakhpat		Others	1
AC0230001	100	RJN1970010	Male	39	Father's	16/1001	Md	Ahmad	Md	Muslim	Others	1
AC0230001	102	RJN1954063	Female	29	Husband's	16/1109	Bibi	Rehana	Mohd	Afak	Others	1
AC0230001	128	RJN1807394	Female	32	Husband's	446	Asgari		Mustak	Ahmad	Others	2
AC0230001	132	RJN2019486	Male	46	Father's	456	Mahesh		Sidhiram		Others	1
AC0230001	136	RJN1788859	Male	49	Father's	477	Md	Sajid	Md	Isha	Muslim	3
AC0230001	137	RJN1545722	Female	47	Husband's	477	Razia	Khatoon	Sajid	Khan	Muslim	
AC0230001	139	RJN1727205	Male	35	Father's	477	Md	Ansarul	Md	Makbul	Muslim	
AC0230001	146	RJN1993581	Female	35	Husband's	504	Bibi	Safina	Riyazuddin		Others	1
AC0230001	167	RJN2036597	Male	20	Father's	1770	Hanan	Shahid	Mohd	Shahid	Muslim	1
AC0230001	170	RJN1711951	Male	33	Father's	4457/55	Sheikh	Jahangir	Sheikh	Siraj	Others	1
AC0230001	174	RJN1954568	Male	40	Father's	6319	Md	Ismail	Md	Johrab	Others	1
AC0230001	177	RJN1997046	Male	28	Father's	9562/47	Shokat		Nanne		Others	1
AC0230001	180	RJN2025948	Female	29	Husband's	10328	Rubi	Bano	Rabi	Ahmad	Others	1
AC0230001	181	RJN1821841	Male	31	Father's	11660	Mohd	Junaid	Mohd	Sarfuddin	Others	1
AC0230001	255	RJN1185388	Male	26	Father's	F16/428						

Overall idea

		Residents Size	No. of Households		Age-Groups	Total	Female		Male
		14+	4						
		13	3		Age-Groups	Total	Female		Male
		12	5		18-23	111	41		70
		11	1		24-29	230	95		135
		10	2		30-35	238	115		123
		9	7		36-41	215	104		111
		8	13		42-47	179	67		112
		7	11		48-53	134	67		67
		6	14		54-59	85	41		44
		5	24		60-65	63	31		32
		4	28		66-71	60	33		27
		3	31		72-77	34	19		15
		2	76		78-83	16	11		5
		1	285		84-89	6	3		3
					Total	1371	627		744

Religion	Percent of Booth residents	Number of booth residents			Gender	Percent of Booth residents		Number of booth residents
Hindu	91%	1252			Male	54.26%		744
Muslim	8.50%	117			Female	45.73%		627
Christian	0.15%	2			Total	100%		1371
Total	100%	1371						

Migrants Analysis

art Numbe	SrNo	EPIC ID	Name	Gender	Age	ardian Na	elationshi	Address	EID	MIGRANTS
AC023000	1	RJN04294	Sandeep G	Male	44	Jaibhagwa	Father's	3	429449	
AC023000	2	RJN04294	Anju Gupta	Female	43	Sandeep G	Husband's	3	429464	
AC023000	3	RJN04294	Rajnish Gu	Male	42	Jaibhagwa	Father's	3	429472	
AC023000	4	RJN04294	Nisha Gup	Female	40	Rajnish Gu	Husband's	3	429480	
AC023000	5	RJN19726	Parinita Gu	Female	20	Sandeep G	Father's	3	1972669	
AC023000	6	RJN19349	Tanushree	Female	33	Anand Kun	Father's	5-A/11005	1934909	MIGRANT
AC023000	7	XFF16466	Saurabh D	Male	35	Rameshwa	Father's	8A/71G	1646661	MIGRANT
AC023000	8	RJN18138	Janak Man	Female	54	Baboo Aro	Husband's	9A/6	1813863	MIGRANT
AC023000	9	RJN18167	Mohd Nas	Male	24	Mohd Azij	Father's	13/5	1816742	MIGRANT
AC023000	10	RJN18789	Ganki Pras	Male	38	Narayan D	Father's	16/4	1878932	MIGRANT
AC023000	11	RJN18242	Manak Cha	Male	65	Baldev Ram	Father's	16/88	1824217	MIGRANT
AC023000	12	RJN18202	Leela Wati	Female	64	Manak Cha	Husband's	16/88	1820280	MIGRANT
AC023000	13	RJN18205	Sailander 1	Male	41	Haridev Ti	Father's	16/103-I	1820520	MIGRANT
AC023000	14	RJN16736	Md Jabir	Male	25	Md Maksu	Father's	16/107	1673656	
AC023000	15	RJN15958	Santosh K	Male	25	Ram Newa	Father's	16/152	1595875	
AC023000	16	RJN16485	Ainul Haqu	Male	60	Abdul Shak	Father's	16/152-E	1648500	MIGRANT
AC023000	17	RJN17907	Dimple	Female	24	Hukum Cha	Father's	16/179-H	1790798	MIGRANT

Conclusion

- **2823** out of **6173** are migrants in part 4 of Karol Bagh.

- The most Swingable voters are in the age group 18-23, 23-28, 28-33.

- Many peoples living alone so they became head of their own family.

Limitations

- The data transfer undergoes encryption methodology that is conventionally complex and costly.

- The requirement of high processing systems and upgraded browsers capable of handling trillions

  of records is another issue.

FUTURE PROSPECTS

- We can further extend this project by conducting ***Frequentist Aprroach, Family analysis***(between two or more different constituencies, for eg: R.K. Puram and Rajinder Nagar).

- Currently we have only conducted ***intra-cluster analysis*** (within one constituency).

- We can predict features of other constituencies by determining the similarity of their regions based on the affluence of that region.

References

- https://www.tutorialspoint.com/python/
- https://www.scrapehero.com/web-scraping-tutorials/
- https://www.stackoverflow.com/
- https://www.youtube.com/watch?v=mKxFfjNyj3c
- Christopher M. Bishop, "Pattern Recognition and Machine Learning", Springer, 2006
- Jiawei Han, Micheline Kamber and Jian Pei, "Data Mining Techniques and Concepts", Morgan Kaufmann Publishers, 3rd Ed., 2012
- Allen B. Downey, "Think Stats: Probability and Statistics for Programmers", Green Tea Press, 2011

ACKNOWLEDGEMENT

This project has been successfully completed under the mentorship of:

- *Mr. Akhilesh Ajith Kumar*
- *Mr. Suyash Gulati*