

How Does London Vote?

By Sofia Faqir

Capstone Project

IBM Data Science Professional Certificate Specialization

Table of contents

1. Problem and Background

2. Data gathering

3. Data Visualization & Exploratory Data Analysis (1)

4. Results: K-NN Classification (1)

4. Results: Decision Tree (2)

5. Discussion

6. Conclusion

7. References and Sources

Thank you!

1. Problem and Background

- British people have been called to the polling booths many times in recent years: Brexit Referendum, general elections, early general elections, mayoral elections, local elections etc.
- Voter fatigue has been increasing, which makes it even more important to understand where it is worth spending more energy (and money) canvassing and campaigning.
- **Understanding what drives a constituency to vote for a certain party** can be helpful in many ways:
 1. **The boundaries of constituencies** are regularly reviewed and amended to be fairer and more equal. This is a subjective measure, and should be tested against a range of innovative ways to make sure the governing party is not taking through boundaries more favourable to them.
 2. **Campaign optimization** can be empowered by data science. This is particularly relevant because of the rising costs of campaigns (and the multiplication of votes...).
- **Question: Using the results for the General Elections of 2017, can the data on venues in a specific constituency help predict how they will vote?**
- We will focus on the **Greater London area**: 73 parliamentary constituencies, while the UK as a whole has 650 constituencies

2. Data gathering

- The **British government** has clearly made a big effort in transparency and opened many sources of data that are readily available and easily accessible.
- The **constituencies and the party** that they have elected: The data on the General Elections of 2017 is available on **London Datastore**.
- The **wards belonging to each constituency**: Ward level increases the granularity when compiling the list of venues in a constituency. Constituencies are fairly large areas, and hence would require more work to drill down the data.
Scraping the website: <https://www.electoralcalculus.co.uk/> which had the list of wards for each constituency, and more data...
- **The coordinates for all the wards** in England are recorded in the Open Geography Portal, framework from the Office for National Statistics.
- Further **manual processing** was required.
- The **Venues** : calling the **FourSquare API** at ward level.

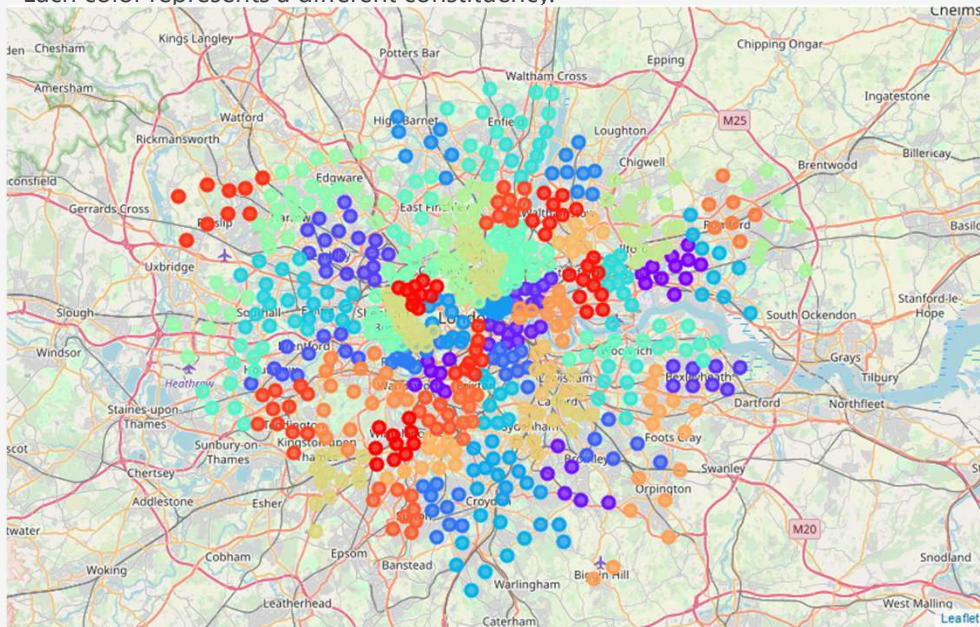
	Constit-Ward	Ward Latitude	Ward Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Barking-Abbey	51.539822	0.081291	Nando's	51.539655	0.081828	Portuguese Restaurant
1	Barking-Abbey	51.539822	0.081291	Cristina's	51.536523	0.076672	Steakhouse
2	Barking-Abbey	51.539822	0.081291	The Gym London Barking	51.536193	0.078601	Gym
3	Barking-Abbey	51.539822	0.081291	Subway	51.538688	0.080788	Sandwich Place
4	Barking-Abbey	51.539822	0.081291	Costa Coffee	51.539272	0.081341	Coffee Shop

3. Data Visualization & Exploratory Data Analysis (1)

Visualizing the features of the wards on the map helps get a first intuition.

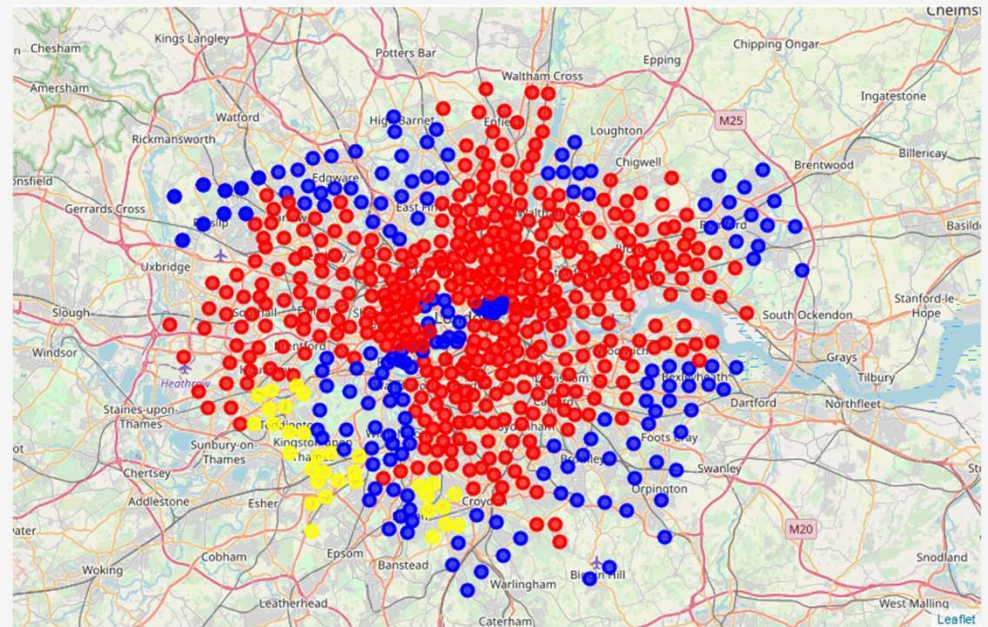
Map of constituencies:

Each color represents a different constituency.



Map of parties:

Red is Labour, Blue is Tory, Yellow is Lib Dem

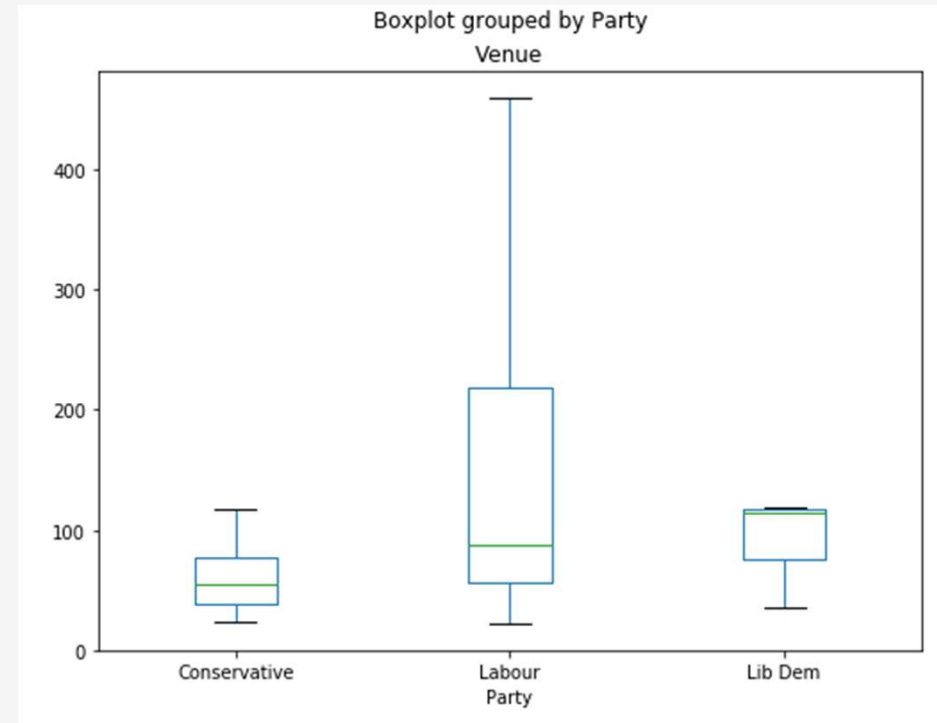


3. Data Visualization & Exploratory Data Analysis (2)

Number of Venues :

After removing the **outliers** to make the graph more readable, we get the below box graph :

Large number of venues does strongly indicate a **Labour** constituency, while a smaller number of venue can be found in both parties.

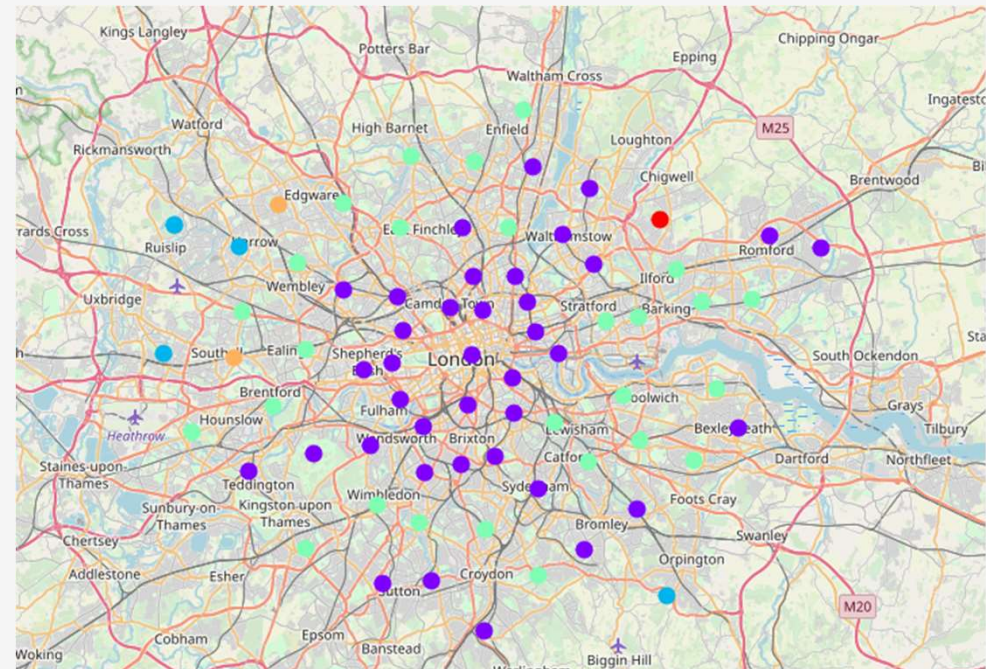


3. Data Visualization & Exploratory Data Analysis (3)

K-Means Clustering :

- The map shows the results for $k = 5$.
- Tory constituencies seem to be more frequent in cluster 0.
- There is no particular insight on Labour constituencies.

Cluster	Tory	Labour	Lib Dem	Colour
0	12	24	2	Purple
1	3	2	0	Green
2	5	21	1	Red
3	1	1	0	Blue
4	0	1	0	Orange



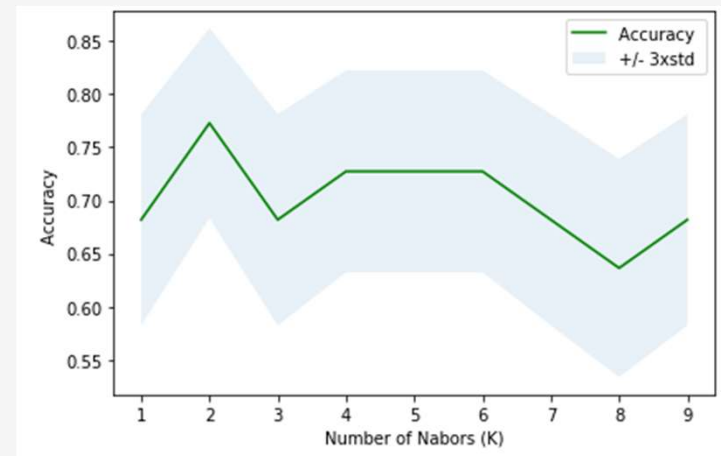
Cluster	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Pub	Coffee Shop	Café	Italian Restaurant	Hotel
2	Grocery Store	Park	Pub	Coffee Shop	Café

4. Results: K-NN Classification (1)

- After dividing the set into a train set and a test set, and run the K-NN algorithm for different values of K, we get this graph.
- **K=2** gave the best result, as per below graph of accuracy score for different levels of K:

Train set Accuracy: 0.90

Test set Accuracy: 0.77



4. Results: Decision Tree (2)

➤ I also run the Decision Tree algorithm on the data set. I only used as features the number of: Café, Coffee Shop, Grocery Store, Park, Pub, Total number of venues

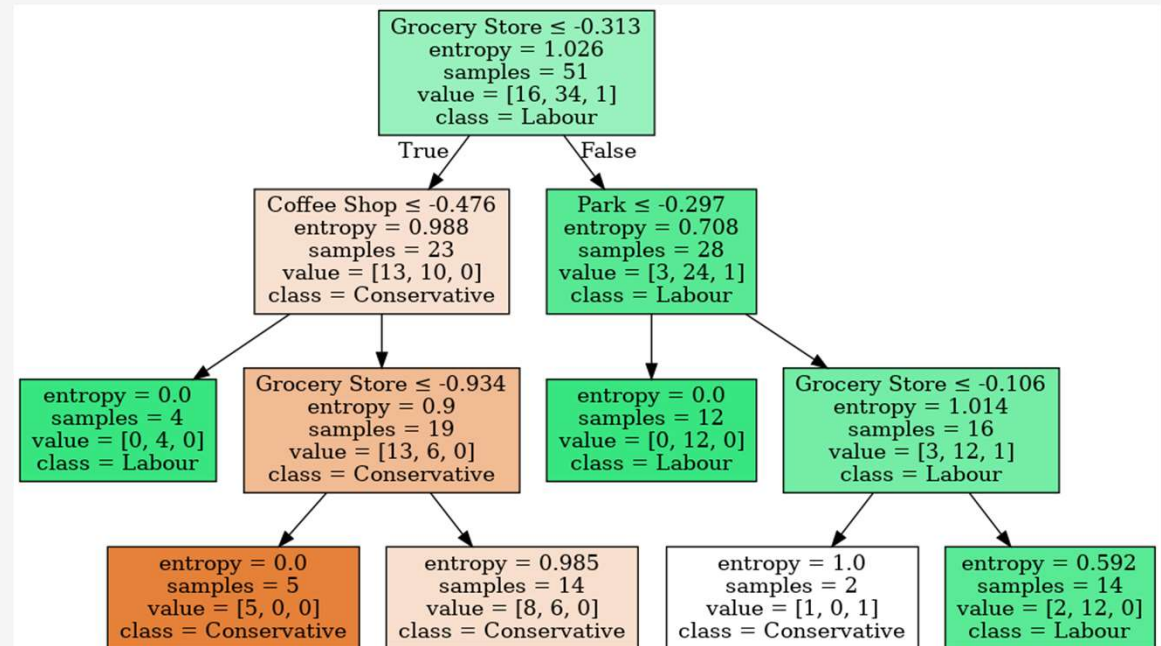
➤ Using a tree with a max depth of 4, I got an accuracy of:

Train set Accuracy: 0.85

Test set Accuracy: 0.64

➤ Increasing the depth did not improve the accuracy.

➤ Interestingly, we can see that the **Total Number of Venues** is not being used by the Decision Tree.



5. Discussion

- The venues in a constituency provide **some keys** to understand the **voting patterns** in the Greater London area. However, there was **no conclusive method** that would allow to predict the vote.
- The 2017 General Elections were notoriously bad for the Conservative party in the Greater London area because of **Brexit**.
- We should run this analysis on the **2015 General Elections** instead to see if there is a stronger predictive power.
- This study could be made more granular by running it at ward level, for local or parliamentary elections, as well as for other big questions like Brexit. It can also be extended to the rest of the UK, since we have seen patterns in how people have voted:
 - England Vs. Scotland
 - Big Cities Vs. smaller cities & rural areas
 - University degree
 - income level

6. Conclusion

- The question being asked is very relevant and topical: **How to better predict how people are going to vote?**
- Here we have tried to give elements of answer based on the venues in a constituency: their numbers, the frequency of a given venue category...
- The analysis did bring some answers to this question. For example, the larger the number of venues, the more likely the constituency was to vote Labour. However, the accuracy of the results didn't seem to be good enough to be used by itself.
- We would probably gain in increasing the granularity of the study, and going to the ward level, and country level.

7. References and Sources

London Data Store: <https://data.london.gov.uk/>

Electoral Calculus, by Martin Baxter: <https://www.electoralcalculus.co.uk/>

Open Geography Portal, Office for National Statistics : <http://geoportal.statistics.gov.uk/>

The London Datastore, by the Greater London Authority (GLA) : <https://data.london.gov.uk/>

Boundary Commission for England: <https://boundarycommissionforengland.independent.gov.uk/>

Thank you!

Many thanks for having taken the time to read my presentation.