Getting to know Austin, TX

Capstone Project - The Battle of Neighborhoods

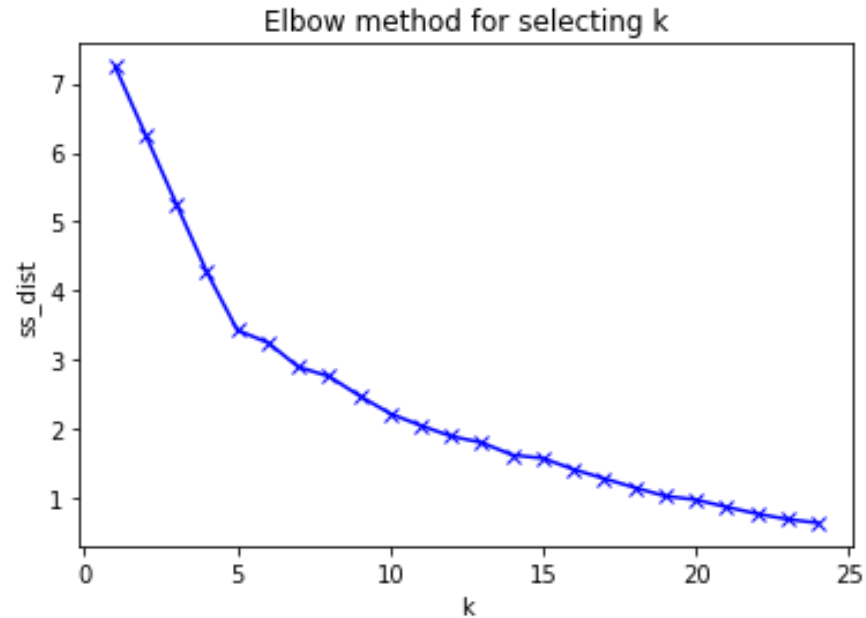Coursera Applied Data Science Capstone

Prepared by Dimitar Kunev

The issue

- Newcomers to Austin, TX need objective data to help them narrow down their search for a new home based on the type of activities most important to them
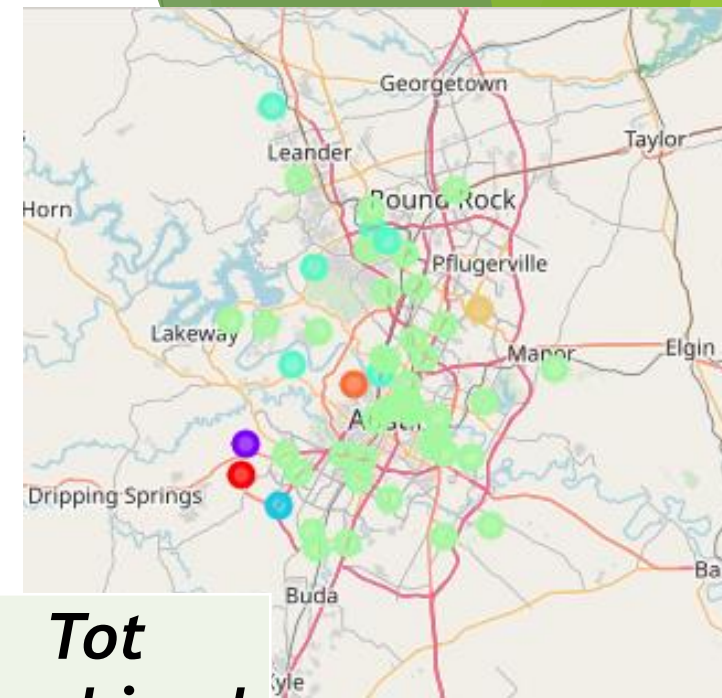
The solution - data

- 88 ZIP codes for the greater Austin area
  - Source – city-data.com
- 1,309 venues across 246 categories
  - Source – FourSquare
- Crime statistics for 48 ZIP codes
  - Source – Austin PD

The solution - methodology

▶ Import and prep/clean data

▶ Run descriptive statistics

▶ Cluster analysis

  ▶ K-means clustering

  ▶ Elbow method for determining the optimal number of clusters

Elbow method for selecting k

The solution - results

Cluster Labels	ZIP code	Violent	Property	Tot indexed	Tot Non Indexed	Tot Combined
0	1	0	2	2	4	6
1	1	9	99	108	166	274
2	1	NaN	NaN	NaN	NaN	NaN
3	1	5	97	102	211	313
4	6	19.75	265.5	285.25	564	849.25
5	35	81.03	893.42	974.45	2,121.67	3,096.15
6	1	28	466	494	901	1,395.00
7	1	9	797	806	619	1,425.00

Conclusion

- Different activity profiles throughout the city are associated with different levels and types of crime

- Activity profiles can be a useful tool in determining which areas of town to focus house hunting on

- For people primarily driven by a need for safety and security, an alternative approach (segmenting the city based on criminal activity) may be more appropriate