



Band Tour Organizer

Coursera Capstone Project Report

Introduction/Business Case

- ▶ Band manager planning a worldwide/european/etc. tour
- ▶ Searching for promising venues in many different cities is difficult and time-consuming
- ▶ Assumption: similar neighborhoods are home to people with similar socioeconomic backgrounds, lifestyles and tastes
 - turnout for certain types of concerts (i.e. musical genres) will be higher
- ▶ Aim: to develop a tool to save time and effort on the part of the band manager organizing a tour, as well as increasing concert turnout and thereby revenue for the band.

Data

- ▶ Foursquare location data forms the basis
- ▶ Auxiliary data scraped from the web: a list of the top 20 cities in the European Union (https://en.wikipedia.org/wiki/List_of_cities_in_the_European_Union_by_population_within_city_limits)
- ▶ Geocoder data for geographical information (latitude and longitude)



Data Cleaning

- ▶ Assumption: Postal Code = Neighborhood
- ▶ Dropping final Postal Code digits for certain cities:
 - ▶ Birmingham
 - ▶ London
 - ▶ Warsaw
 - ▶ Stockholm
 - ▶ Bucharest
- ▶ 8414 venues in 583 neighborhoods/postal codes



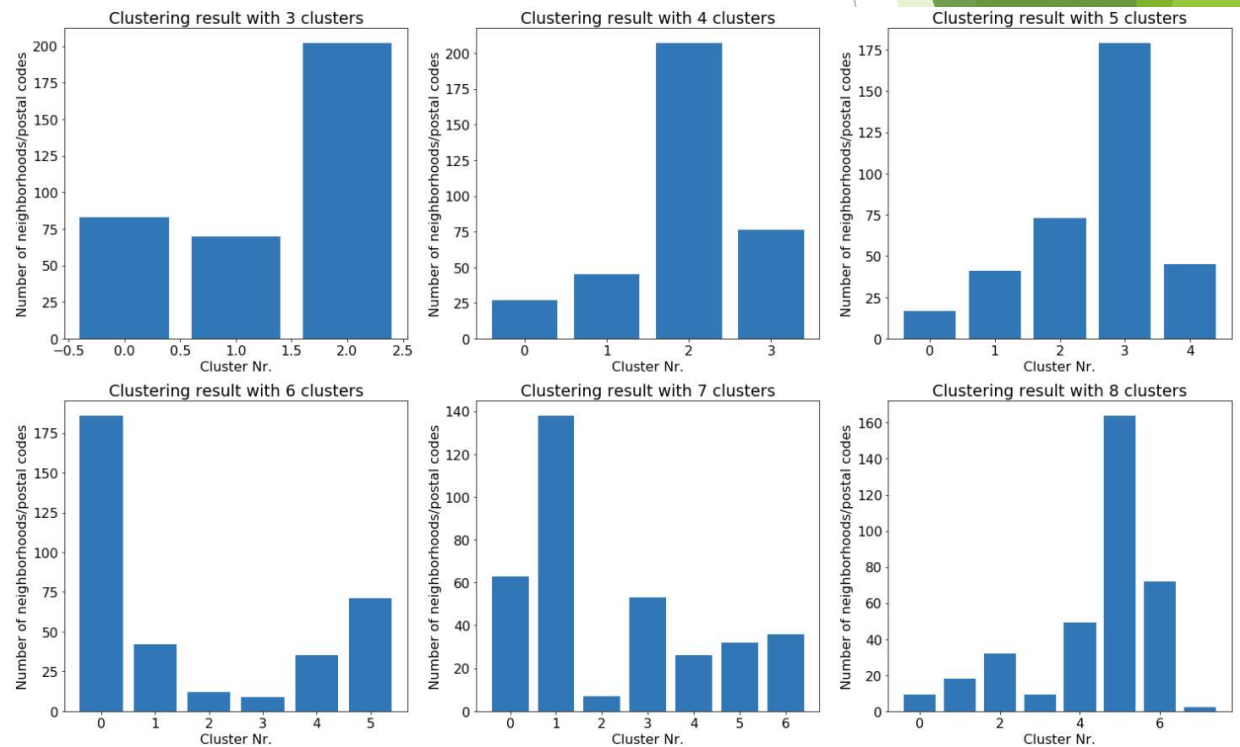
Data Cleaning

- ▶ Many neighborhoods with only one or two venues
 - ▶ Clustering will not be sensible
 - ▶ There is not a lot going on in these neighborhoods anyway
- ▶ Drop neighborhoods with only one or two venues
- ▶ This results in 8126 venues in 355 neighborhoods
- ▶ There are 289 unique venue categories in our test data set



Methodology

- ▶ K-means Clustering with different numbers of clusters
- ▶ Trade-off between a meaningful separation of different neighborhood types and the formation of tiny clusters
- ▶ We choose 5 clusters, since this results in relatively nicely distributed numbers of neighborhoods



Methodology

- ▶ Algorithm:
 - ▶ Choose a reference venue (home base of our band)
 - ▶ Find cluster number for reference venue
 - ▶ Find other venues from same cluster with venue category ==
 - ▶ Rock Club
 - ▶ Concert Hall
 - ▶ Piano Bar
 - ▶ Irish Pub
 - ▶ Jazz Club
 - ▶ Performing Arts Venue
 - ▶ Theater
 - ▶ Karaoke Bar
 - ▶ Music Venue
 - ▶ Opera House
 - ▶ Amphitheater
 - ▶ Salsa Club
 - ▶ Country Dance Club
 - ▶ Cultural Center

Results

- ▶ We will take a look at three test cases (Reference locations)
 - ▶ "Ampere" in Munich, Germany
 - ▶ "Costello Club" in Madrid, Spain
 - ▶ "VK Concerts" in Brussels, Belgium

Results - Case 1

- ▶ The "Ampere" is a venue of the category "rock club"
- ▶ belongs to cluster 2
- ▶ 80 proposed venues

```
proposed_venues.shape
```

```
(80, 5)
```

```
proposed_venues.head(10)
```

	City	Postal Code	Venue	Venue Category	Cluster Labels
1015	Munich	80336	Kennedy's Irish Bar & Restaurant	Irish Pub	2
2370	Vienna	1090	Charlie P's	Irish Pub	2
2508	Warsaw	00-6	miejsce Chwila	Rock Club	2
2526	Warsaw	00-6	12on14 Jazz Club	Jazz Club	2
2749	Munich	80333	Rote Sonne	Music Venue	2
2763	Munich	81667	Muffathalle	Music Venue	2
2771	Munich	81667	Ampere	Rock Club	2
3377	Naples	80134	Alter Ego	Music Venue	2
4246	Warsaw	00-3	Chopin Point Warsaw	Concert Hall	2
4894	Birmingham	B12 9	mac birmingham	Performing Arts Venue	2

Results - Case 2

- ▶ The "Costello Club" is a venue of the category "concert hall"
- ▶ clustered to neighborhood type 3
- ▶ one of the larger clusters
- ▶ very large number (411) of suggested venues

```
proposed_venues.shape
```

```
(411, 5)
```

```
proposed_venues.head(10)
```

	City	Postal Code	Venue	Venue Category	Cluster Labels
501	Bucharest	0300	St. Patrick	Irish Pub	3
587	Vienna	1010	Pickwick's	Irish Pub	3
956	Munich	80331	Kilians	Irish Pub	3
1755	London	W1D 4	Ronnie Scott's Jazz Club	Jazz Club	3
1859	Berlin	10117	Vincent	Piano Bar	3
1952	Madrid	28004	Microteatro por dinero	Performing Arts Venue	3
1972	Madrid	28013	Costello Club	Concert Hall	3
2011	Madrid	28004	Toni 2	Piano Bar	3
2053	Rome	00187	Gregory's Jazz Club	Jazz Club	3
2254	Bucharest	0300	MOJO	Karaoke Bar	3

Results - Case 3

- ▶ the "VK Concerts" is a venue of the category "music venue"
- ▶ clustered in neighborhood type 4
- ▶ there are 37 proposed venues

```
proposed_venues.shape
```

```
(37, 5)
```

```
proposed_venues.head(10)
```

	City	Postal Code	Venue	Venue Category	Cluster Labels
3313	Cologne	51063	Gebäude 9	Music Venue	4
6836	London	WC2E 7	Lyceum Theatre	Theater	4
6859	London	WC2H 9	The Hospital Club	Performing Arts Venue	4
6867	London	WC2H 9	Ambassadors Theatre	Theater	4
6870	London	WC2H 9	St Martin's Theatre	Theater	4
6883	London	S	Scotch of St James	Music Venue	4
6957	Berlin	10115	Schokoladen	Rock Club	4
6986	Berlin	10115	Kunstfabrik Schlot	Jazz Club	4
7067	Madrid	28010	Honky Tonk Bar	Concert Hall	4
7305	Bucharest	0301	Music Club	Music Venue	4

Discussion/Conclusions

- ▶ Algorithm works and generates usable results
- ▶ The number of suggested venues depends greatly upon the size of the clusters
- ▶ For test case 2, list of over 400 proposed venues is too unspecific
- ▶ For test cases 1 and 3, list is useful
- ▶ Option: filter results for venue category
 - ▶ For example, if the reference venue is a "rock club", experience might tell us that "jazz club" or "piano bar" are less promising due to the venue type
- ▶ Better: add more specific data to our dataframe, for example socioeconomic data, population average age, ethnicity, mean income, etc.