LOCATING A NEW VENUE IN AN UNKNOWN CITY

USING OPENING HOURS TO DETERMINE FOOT TRAFFIC HOT SPOTS

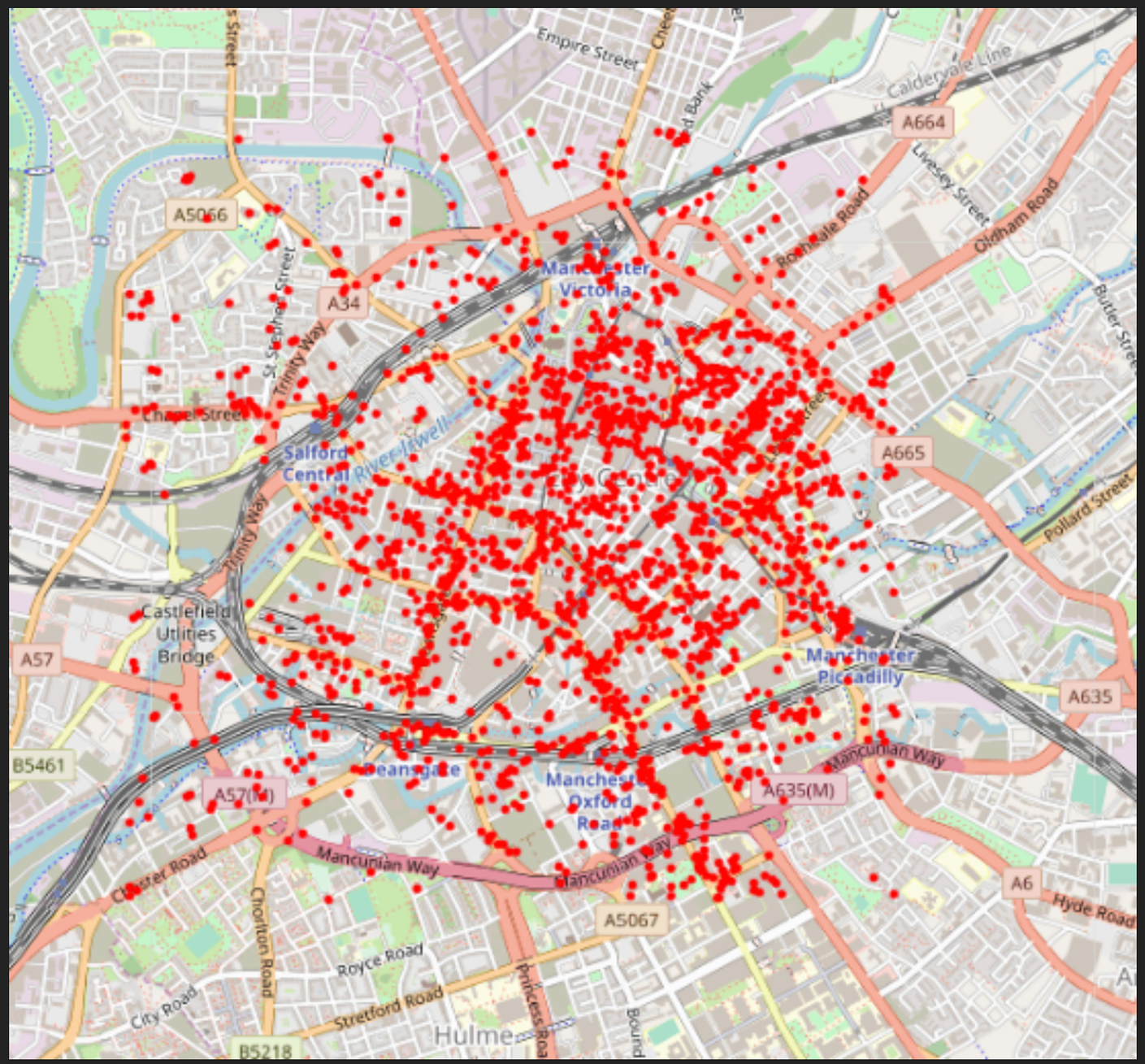
DAVE CK

APPLIED DATA SCIENCE CAPSTONE

WHERE ARE THE CUSTOMERS, AND WHEN?

- Every city is different, and familiarity with local norms is hard-won.
- Investors and planners would benefit from knowing the distribution of foot traffic throughout the day in a given city.
- The time-distributed foot traffic can be inferred from the opening times of the existing venues, helping to avoid pitfalls like:
  - investing in a nightclub on a street filled mostly with daytime stores
  - investing in a hardware store in the middle of a clubbing area
  - opening a venue at the top of a steep hill with little foot traffic

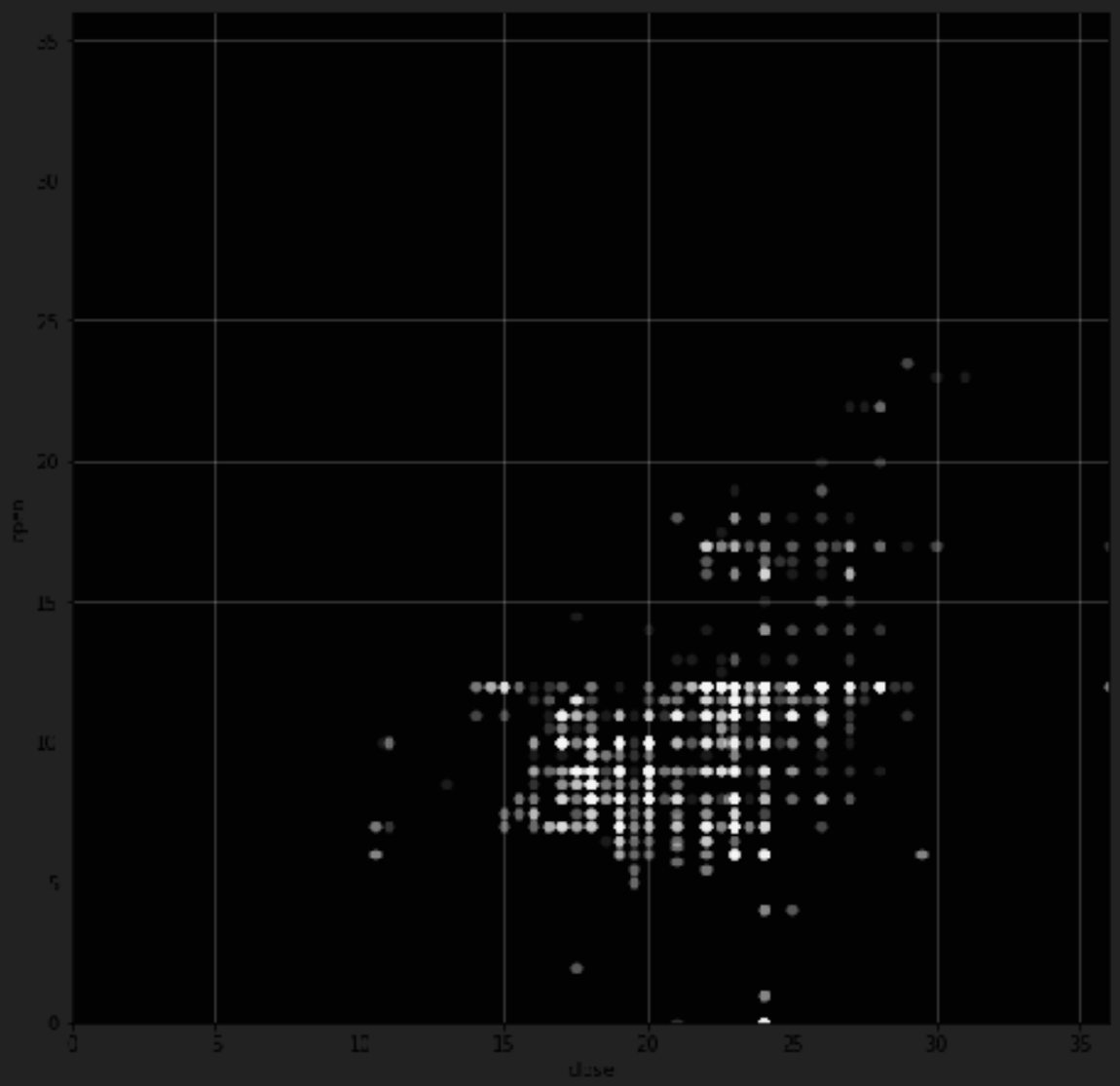
HOW TO GET FAMILIAR WITH A REMOTE LOCATION



- Generate a grid of coordinates around a central point of interest.
- Query Foursquare for venues around each point, remove duplicates and any outside the edges of the grid.
- Query Foursquare for opening hours of each venue and extract.
- Final dataframe is 13930 sets of opening hours and the venue they relate to.

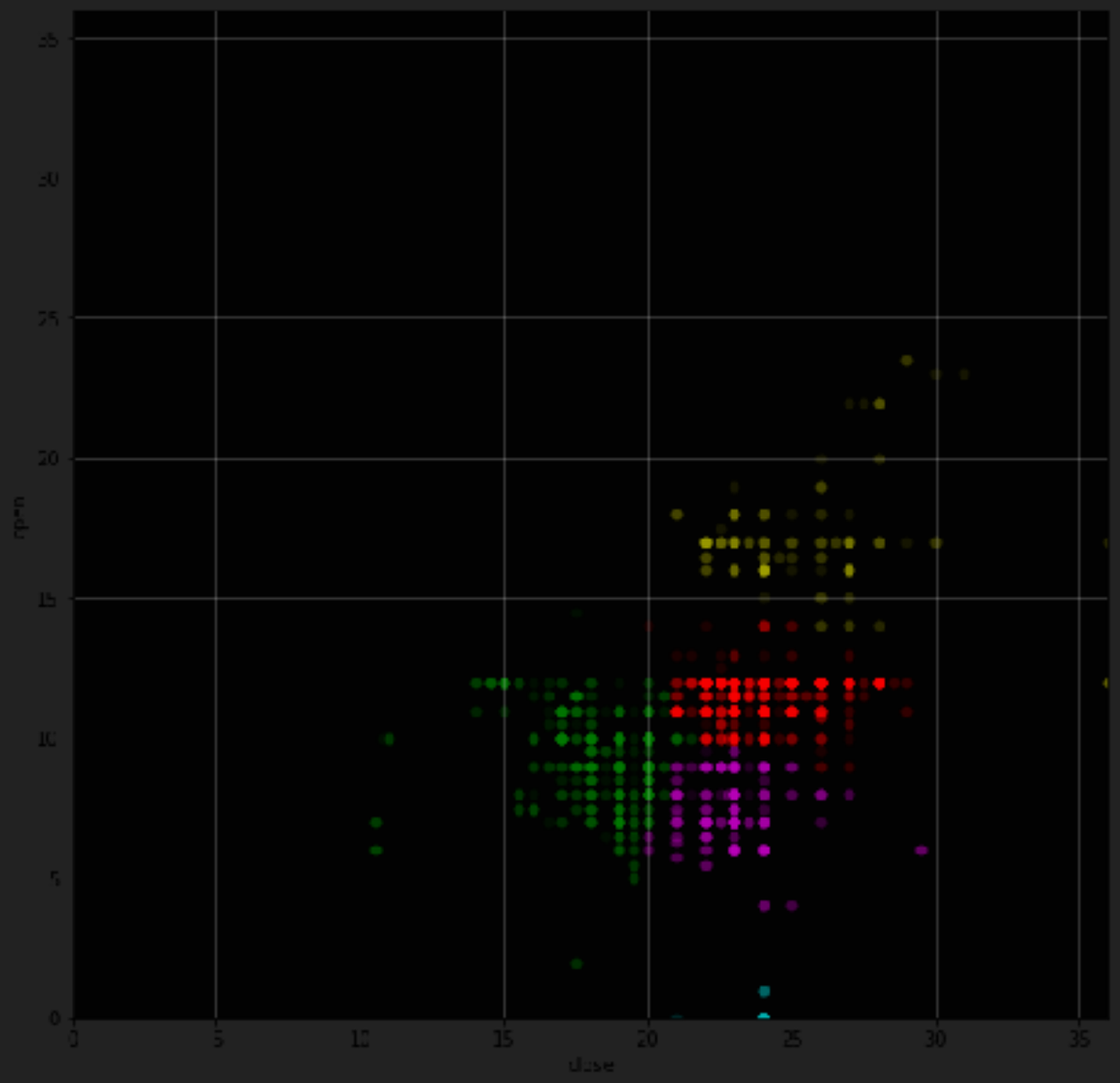
CLUSTER THE OPENING HOURS

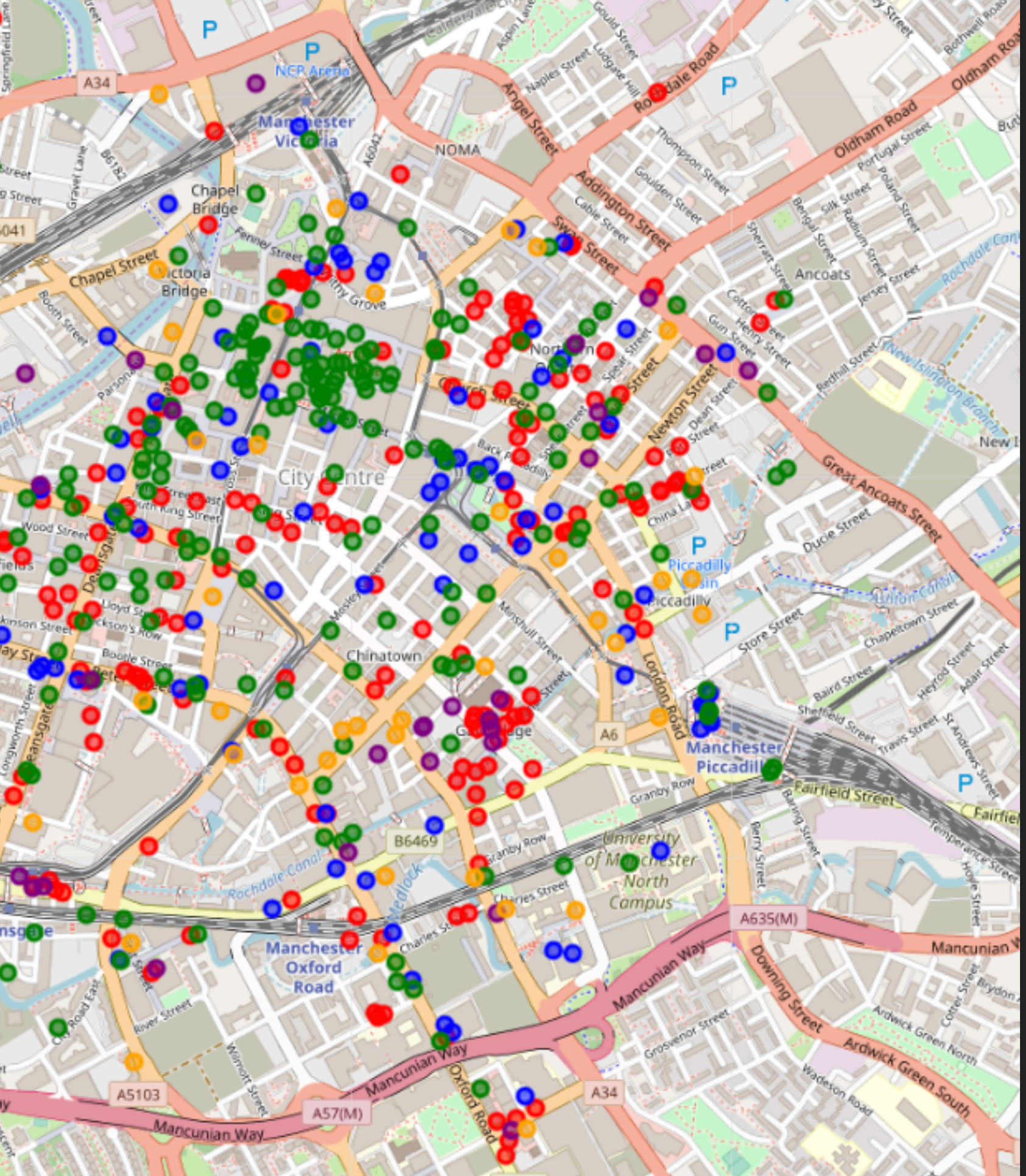
- The opening hours spread quite a range.
- Closing times that went past midnight are represented as decimals over 24.
- Very little discernible density grouping to datapoints.



K-MEANS CLUSTERS OPENING HOURS

- K-means with k=5 clusters into sensible groups.
  - Red group open 12:00 - 24:00.
    Common venue: pubs
  - Pink group open 7:00 - 23:00.
    Common venue: grocery stores
  - Green group open 9:00 - 18:00.
    Common venue: coffee shops
  - Yellow group open 17:00 - 25:00.
    Common venue: gay bars.
  - Blue group open 0:00 - 24:00.
    Common venue: hotels.



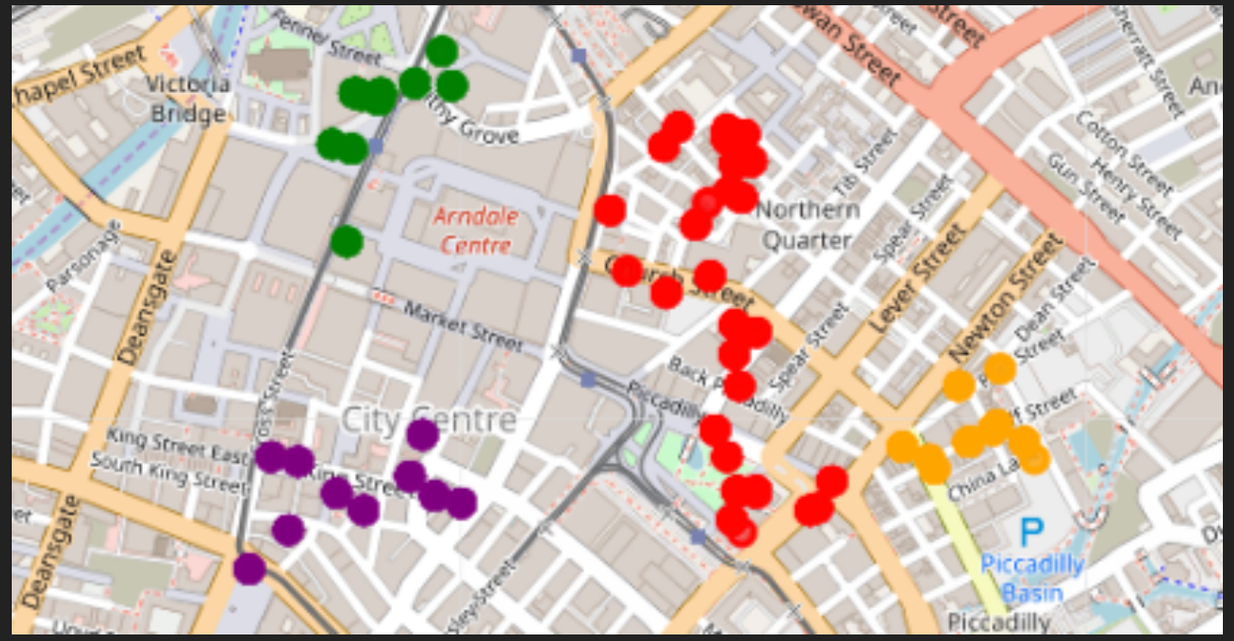


Venues coloured by their opening hour group

CLUSTERING EACH GROUP BY LOCATION

- Filter venues to only one opening-hour group.
- DBSCAN finds dense areas of venues matching that group.

Four clusters of venues with typical opening times of 11:00 to 24:00



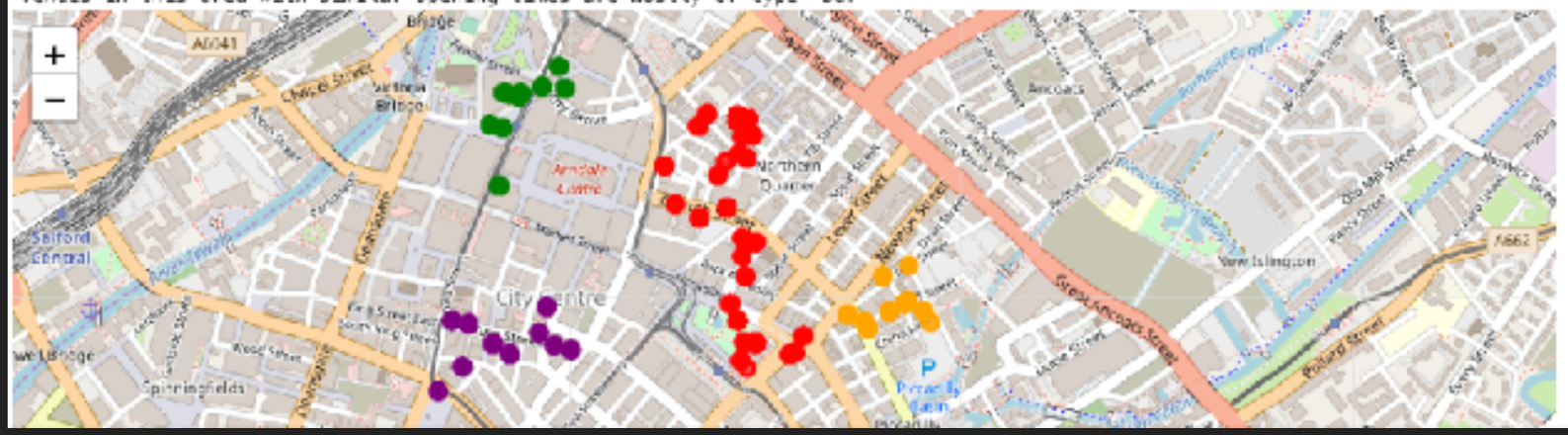
PROVIDE SUMMARY INFORMATION ON THE CLUSTERS

A function can produce suggestions on the basis of an intended opening and closing time.

- Colour-coded map of all venues with closely matching opening times in their clusters.
- Lat/long coordinates of the central point of each cluster.
- The address at that location.
- Most common venue type in that area.

```
Group 0 is centred around 53.48074463282498, -2.2431091685226434
----------------------------------------------------------------
Suggested address to investigate: Pizza Express, 84-86, King Street, City Centre, Manchester, Greater Manchester, North West Engla
nd, England, M2 4WQ, UK
Venues in this area with similar opening times are mostly of type 'Italian Restaurant'


Group 1 is centred around 53.48136020201502, -2.2322593405733465
----------------------------------------------------------------
Suggested address to investigate: Northern Monk Refectory, 10, Tariff Street, Northern Quarter, City Centre, Manchester, Greater M
anchester, North West England, England, M1 2FF, UK
Venues in this area with similar opening times are mostly of type 'Bar'
```



JOB DONE BUT ALWAYS ROOM FOR IMPROVEMENT

- The final function produces useful results at no cost.
- Improvements could be made to automate selection of parameters at each stage of the machine learning process.
- Switching to XY coordinate systems could give a more flexible set of input option.
- The end result is imperfect but correctly identifies hot spots of foot traffic in the area purely on the basis of intended opening hours.