Estimating clusters of French people in Montreal based on top venues

Clustering culture expression is a valuable business

- People living abroad tend to cluster geographically and deeply change the landscape of their neighborhood bringing businesses (School, restaurant, religion). Chinatown in New York is a good example.
- Some clusters abroad are less obvious but can be estimated analyzing the neighborhoods' landscape.
- We can imagine developing new businesses targeting these clusters (for example French restaurant, wine shop if French cluster).
- Montreal in eastern Canada is both influenced by French and North American cultures. It makes it a good case study for estimating clusters of French population.

Data acquisition and cleaning

- The list of Paris districts and the list of Montreal boroughs on.
- Geocoder will give us the GPS coordinates of both cities' boroughs.
- Geojson file of Montreal for map observation.
- 2016 census report of French immigrants in Montreal.
- Popular venues of Paris boroughs and Montreal boroughs on Foursquare.
  - 2000 top venues from Paris + 1395 top venues from Montreal.
  - 205 categories to classify the 3395 top venues.
  - non-assigned categories were either reassigned or dropped.
 - Normalizing data.

Montreal identity : boroughs top venues

Horizontal bar plot of Montreal identity dispersion
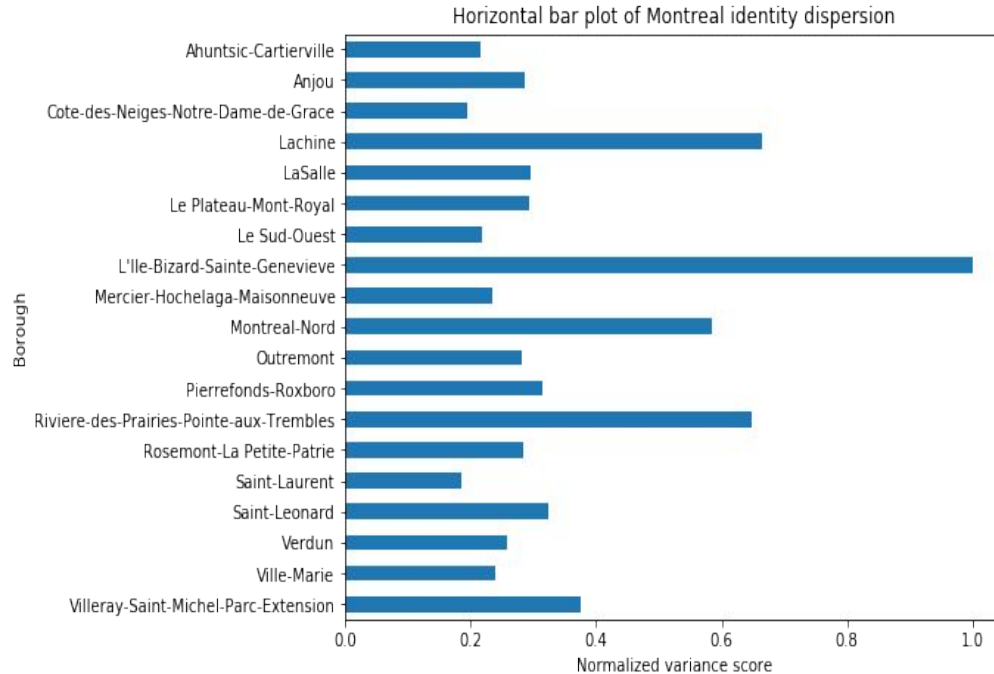
- Top 100 venues of a borough will define its identity.
- Venues are classified in categories.
- Less dense area → few nb of top venues & categories → higher weight → high score.
- More developed area → more top venues & categories → lighter weight → low score.

French identity : Paris top venues

- Paris is the capital of France, it contains 20 districts.
- French culture considered as top 100 venues of Paris boroughs.
- Normalizing occurences of categories data .
- Cat. high nb of hits → high weight → high score.
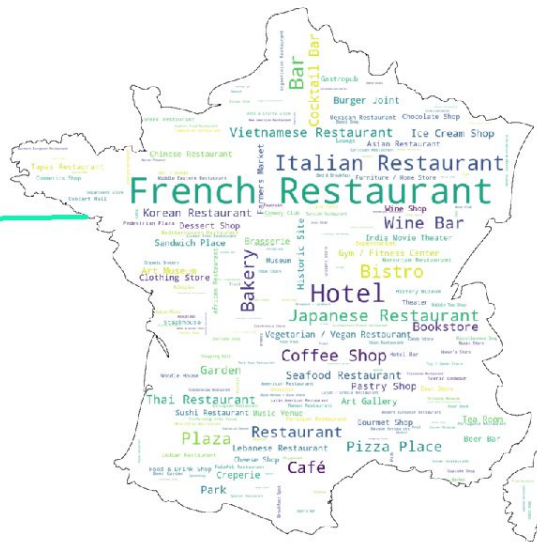- Cat. low nb of hits → low weight → low score.

Histogram plot of cleaned French identity distributed score
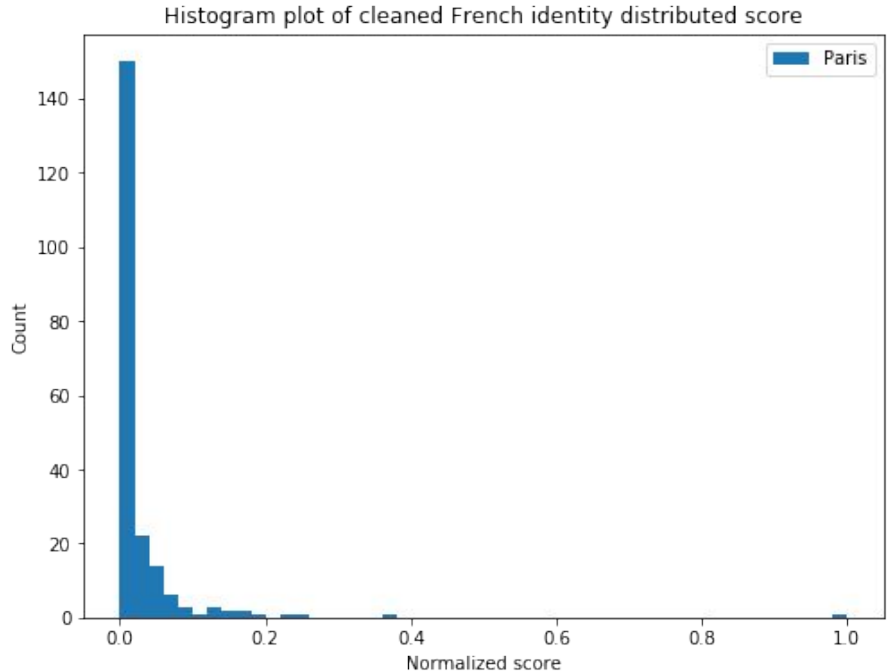
Applying French filter to Montreal boroughs

Venues variance in Montreal VS Filtered Montreal
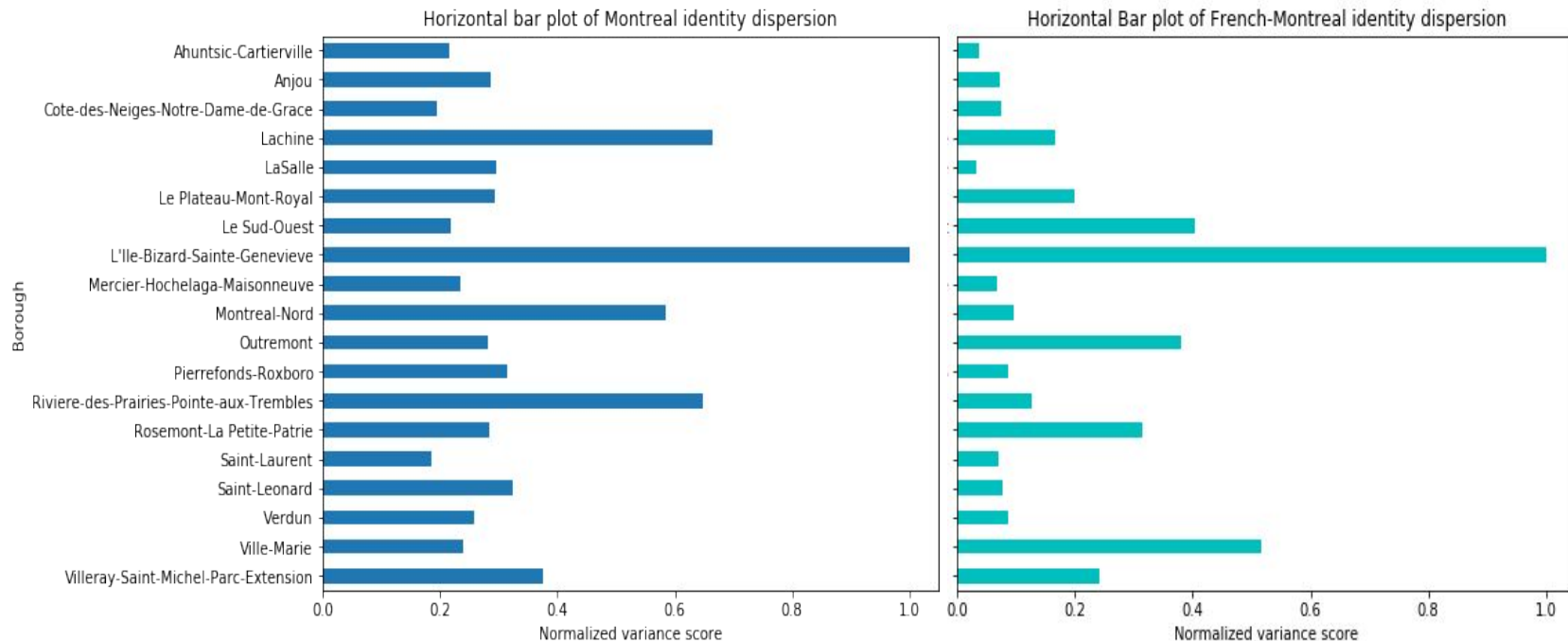
Ville-Marie best Montreal borough match

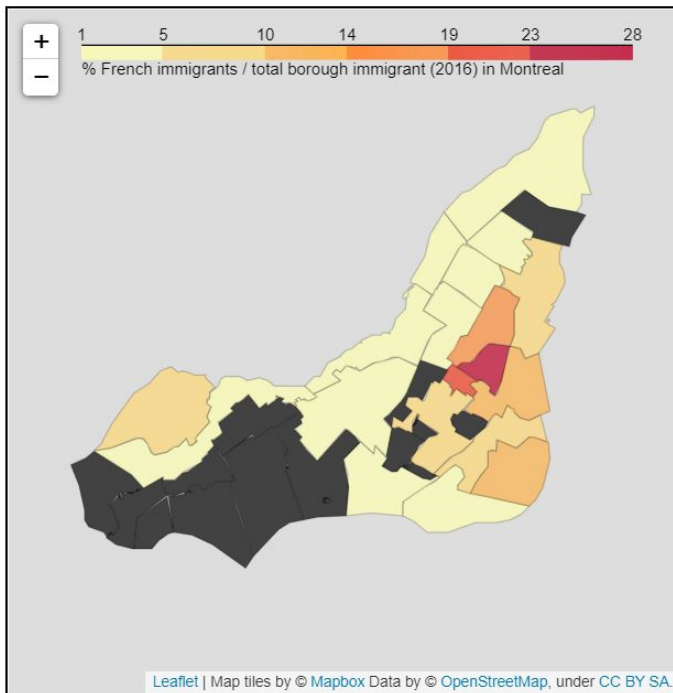Horizontal scatter plot of content-based algorithm prediction
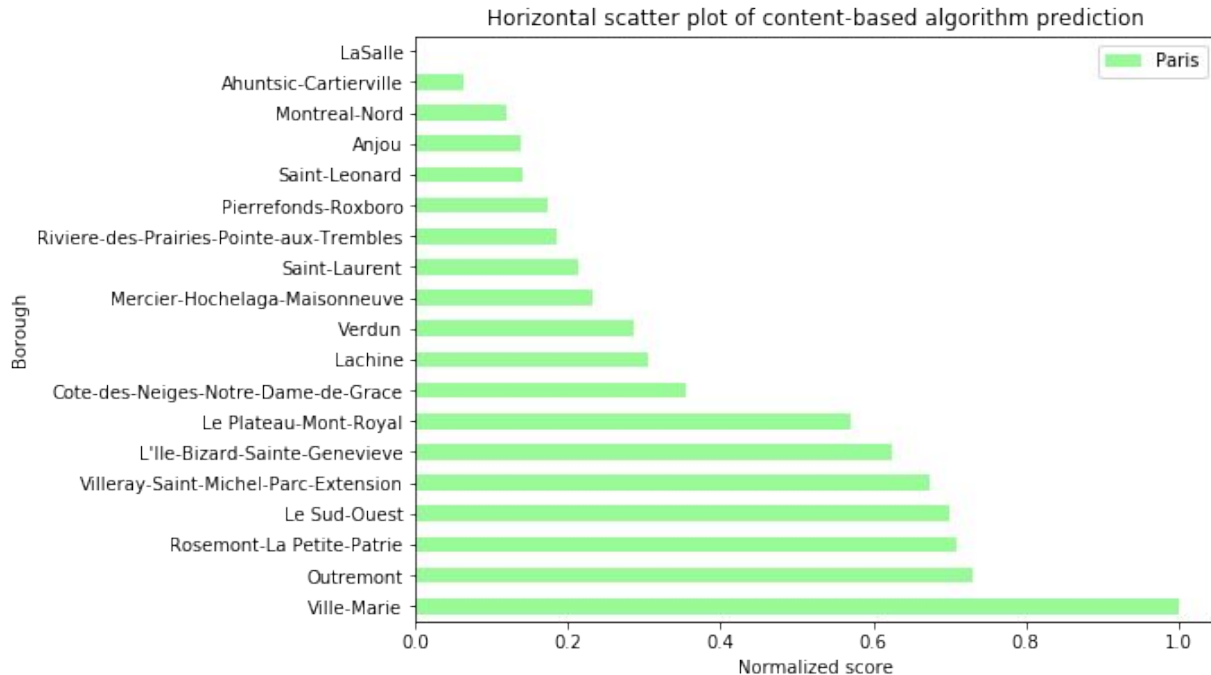
- The higher the score the better the match.
- Ville-Marie to Le Plateau Mont-Royale are the top scores.

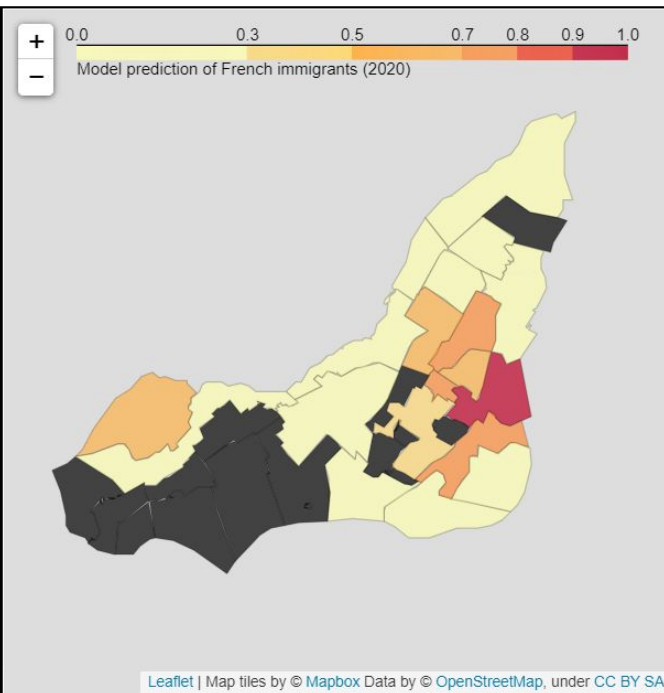# Official 2016 census data VS model prediction

2016 census data

%French / total borough immigrant

2020 model prediction

French top venues score

Clustering model to increase performance accuracy

- K Means clustering for K=2 clusters.
- The top score boroughs from model are sticking out (red dots).
- Red dots represent model clusters of French people.
- Blue dots represent model clusters of non-French people.
- Choropleth map represent actual clusters of French people.

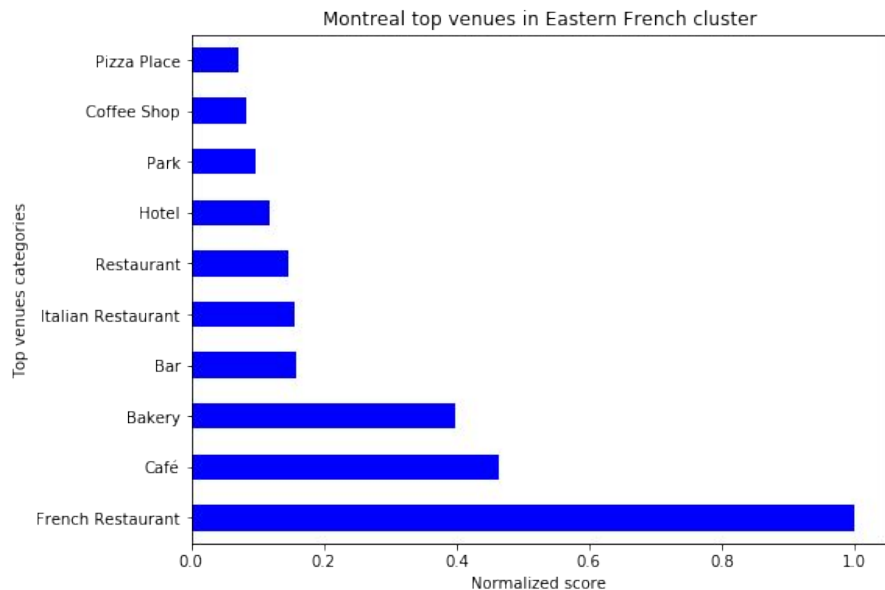% French immigrants / total borough immigrant (2016) in Montreal
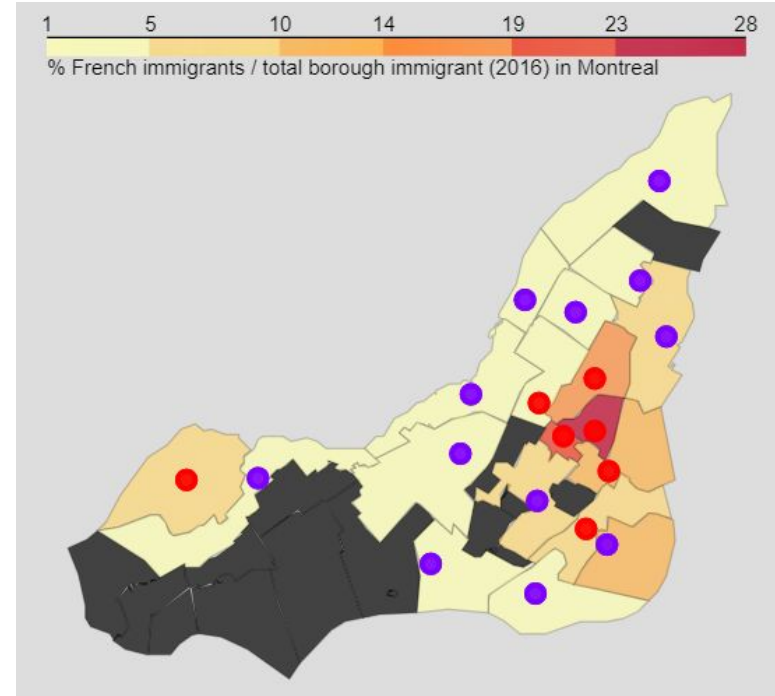
Census binary map to quantify performance

- Setting 12% as optimized binary value for classifying actual clusters of French people.
- Performance measures:
  - F-1 score : 0.727
  - Jaccard index : 0.429
  - Accuracy : 0.842
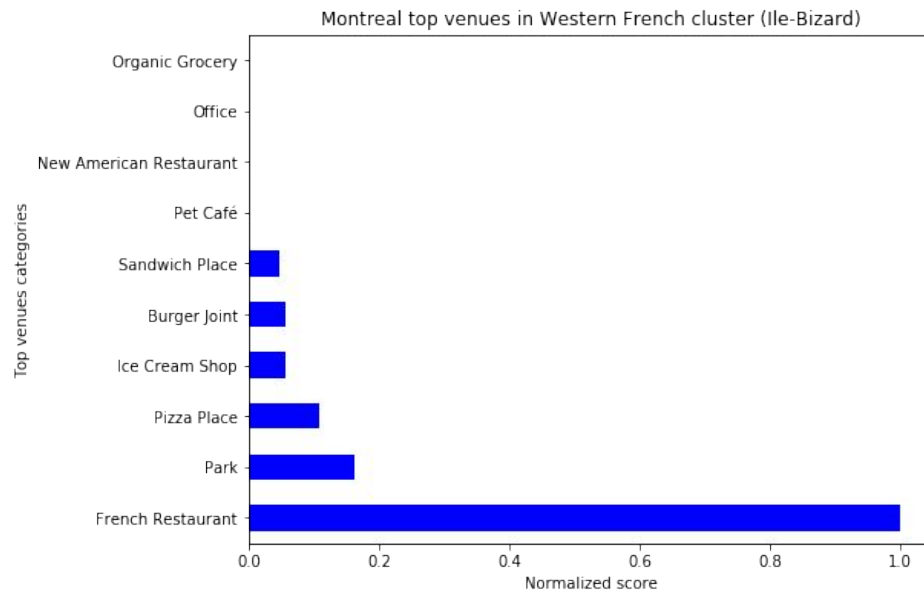 - RMSE : 0.158

- True east cluster and False west cluster.

0   12   28

% French immigrants / total borough immigrant (2016) in Montreal
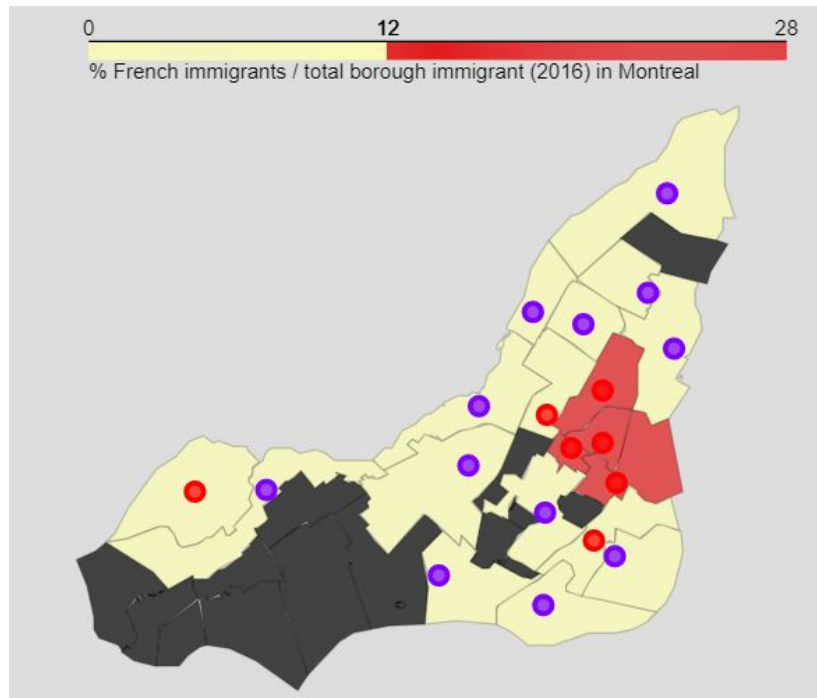
Clusters top venues

East cluster: French culture at its best.   West cluster: French Restaurants&Nature

Conclusion and what next

- Low investment model was able to spot actual French cluster at Montreal.
- Can be enhanced reconsidering some established simplifications:
    - Defining French culture as Paris venues can not be enough. A culture is more complex; we may need every cities of France to obtain a more reliable identity matrix.
    - Searching top 100 venues by borough is not enough. I may consider searching top 100 in multiple smaller areas to populate the database and/ or changing the way to weight data.
    - We stated cultural expression is changing the landscape of the city. But it also depends on the city/ country's immigration policy so it would not necessary work everywhere.
- An immediate step could be to test and check performance on many cities and using different cultures to validate the model.