



# BATTLE OF THE NEIGHBOURHOODS

COURSERA CAPSTONE PROJECT

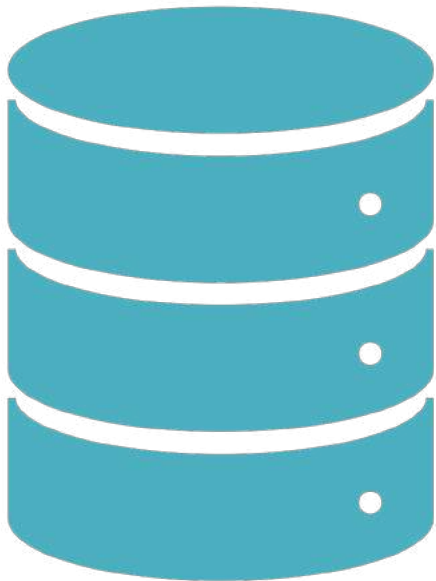
ANTON MÖLLER

# GOOD PURCHASE AB SUCCESSFUL RETAILER – LOOKING TO EXPAND INTO TORONTO



- Team of Data Scientists contracted to determine the best location for their grand opening
- Key requirements on location are
  - Low local competition
  - High customer base (i.e. high population)
- In addition, the location strategy should be easy to replicate

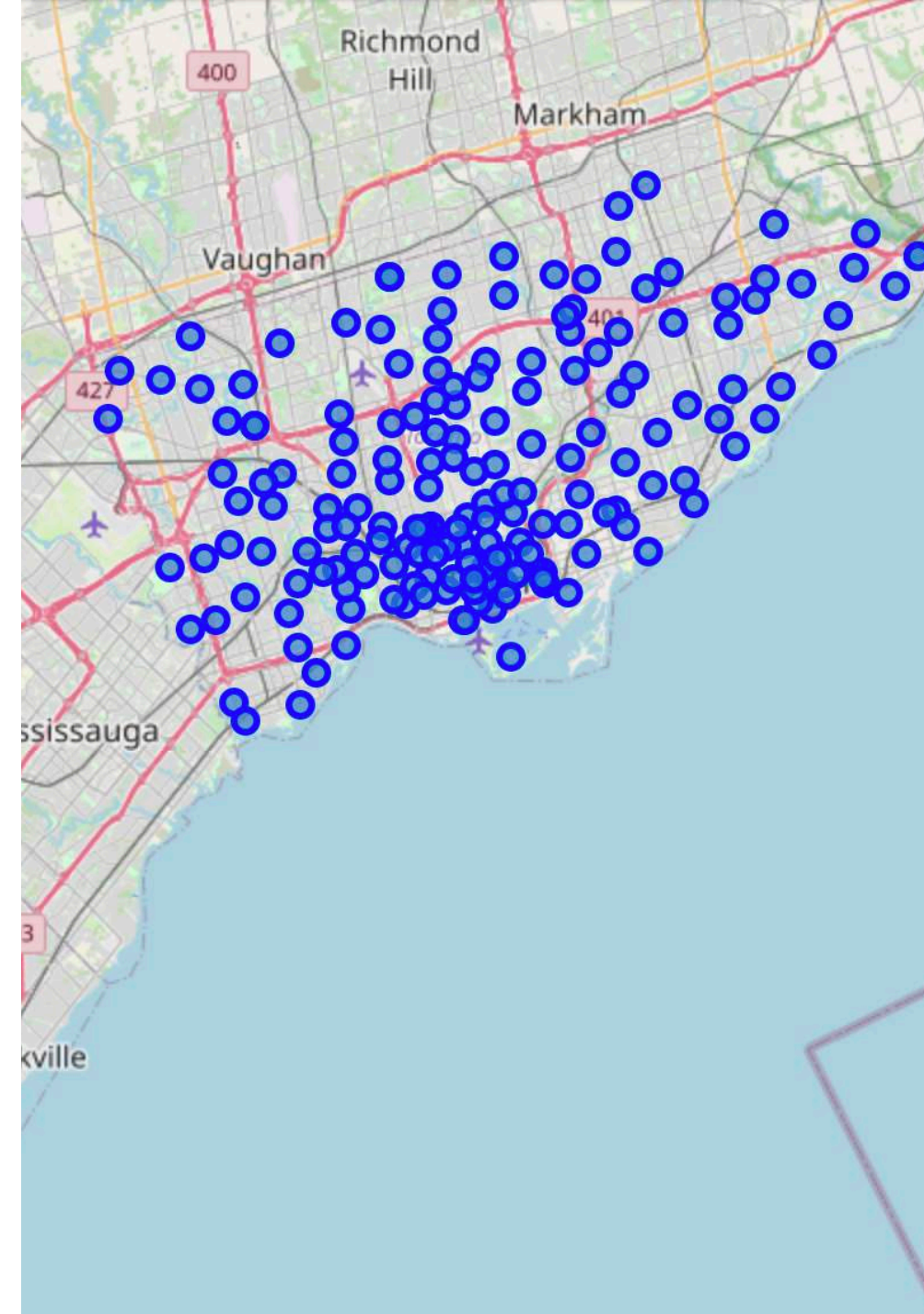
# DATA COLLECTION BY SEVERAL METHODS: WEB SCRAPING, GEO-API AND LOCATION DATA PROVIDERS

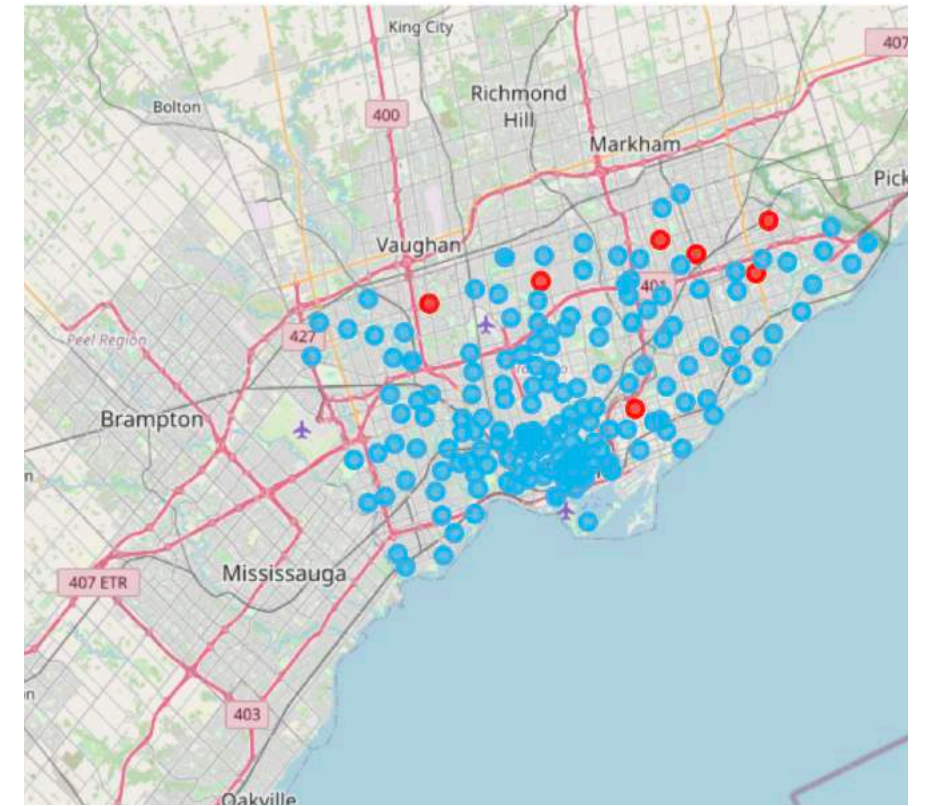
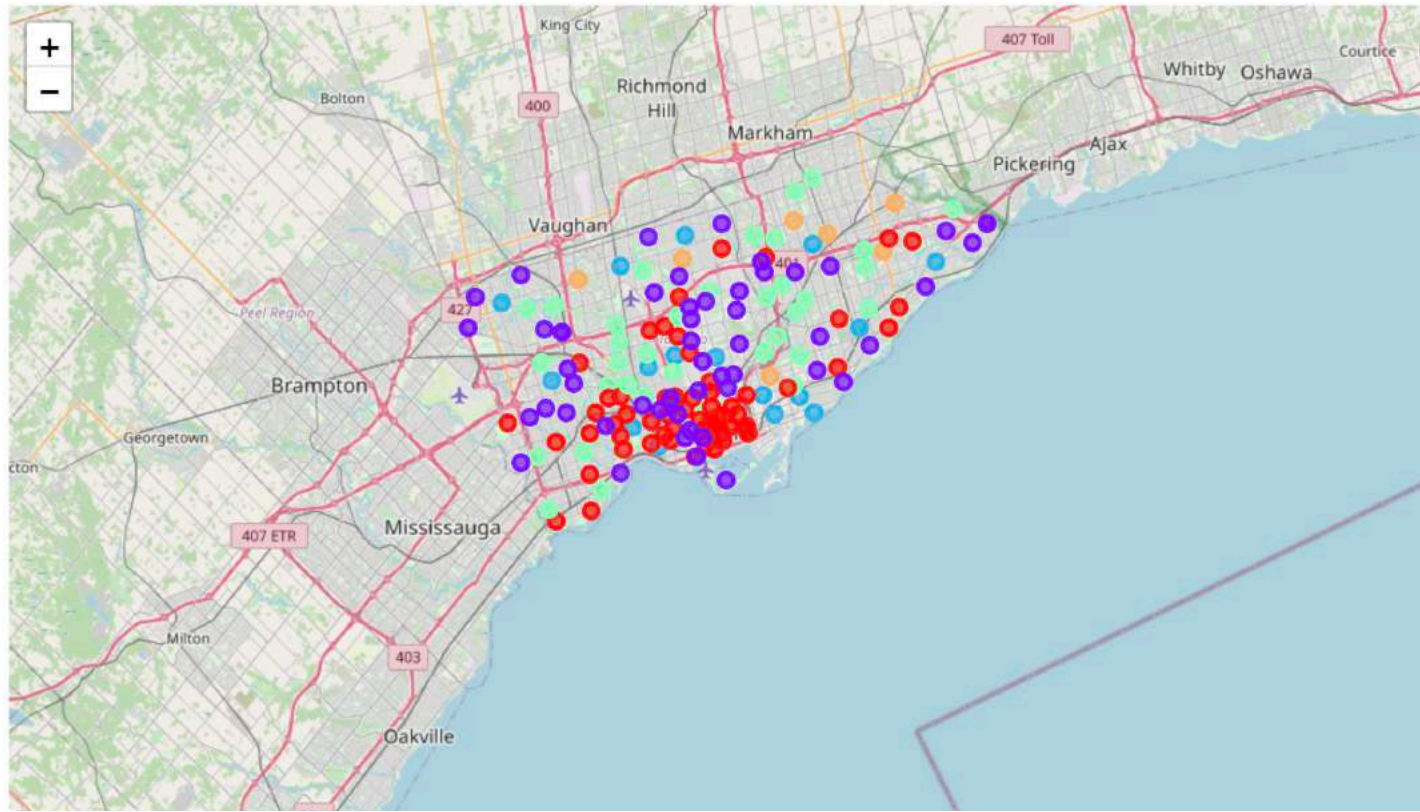


- Successfully ranking neighbourhoods depends on two critical data sources:
  - Neighbourhood dataset (Names, locations, demographics)
  - Competition dataset (Names, types, locations)
- Neighbourhood data collected by web scraping (Wikipedia collection of Canada Census data), appended by Geocoding data (Google Geocoding API)
- Competition data collected by location data provider Foursquare Venue API
- Data cleaned and post-processed

# STEP-BY-STEP ANALYSIS

1. Data collected (web scraping, location venue providers)
2. Data cleaned and post-processed
3. Neighbourhoods clustered together by k-means machine learning to determine set of suitable store locations
  - Neighbourhoods evaluated on population density (50% weight) and competition score (50% weight)





5 CLUSTERS OF SIMILAR NEIGHBOURHOODS  
FOUND

CLUSTER #4 FOUND TO HAVE IDEAL  
COMBINATION OF HIGH  
POPULATION (DEMAND) AND LOW  
COMPETITION (SUPPLY)

## FINAL RECOMMENDATION – CLUSTER #4



Cluster 4 gives >5 suitable locations for the grand opening of Good Purchase in Toronto, and offers a potential future roadmap for more openings, all of which meets the set out criteria

Woburn  
Old East York  
Agincourt  
Elia  
L'Amoureux  
...



Our algorithm is scalable and future-proof, and can be used again and again to respond to shifts in the market dynamics