Aussie Pies

the great Australian bite. . .

. . .coming to Toronto

Prepared by: Andrew Huxham

Date: 28 Jun 2020

Commission: IBM Data Science – Coursera

the big idea

- Deliver Aussie pies as an additional sales opportunity for food outlets, offering a great product with a difference.

- Agile, scalable market entry for a new venture

the business profile

- Australian investor
- business to business distribution – not a shop front
- Toronto, Canada
- pre-start up concept

why aussie pies

- ✓ the great Australian bite
- ✓ convenient hand sized snack
- ✓ tasty pastry bowl and lid
- ✓ delicious dollop of meat fillings
- ✓ one for a snack – two for a meal – mix and match
- ✓ fair dinkum, true blue, real deal aussie meal

scope of analysis

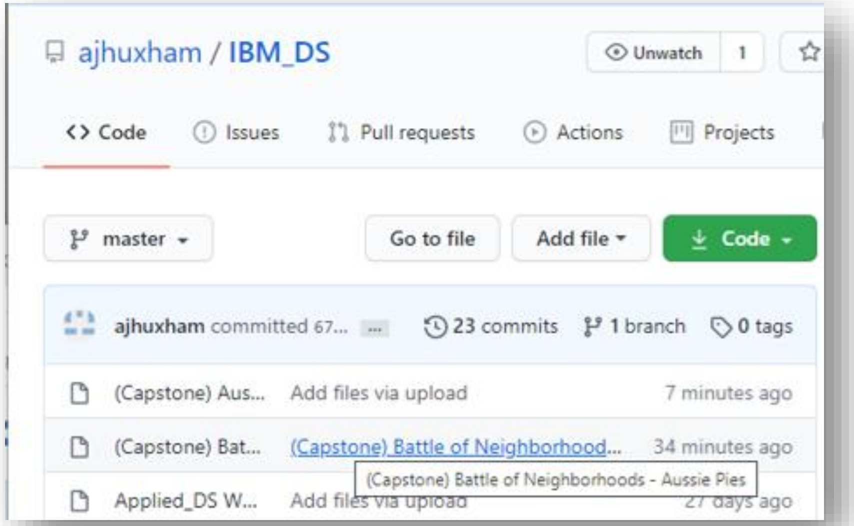
- preparatory research, may form part of future business case
- create profile of Toronto metropolitan venue
- understand potential venues likely relevance

(support, complement, challenge)

extra perspectives

references to follow

- Juypter notebook
  (code cell refences at page footer)



https://github.com/ajhuxham/IBM_DS/blob/master/(Capstone)%20Battle%20of%20Neighborhoods%20-%20Aussie%20Pies.ipynb

- Full report

Aussie Pies

the great Australian bite. . .

. . .coming to Toronto

Prepared by: Andrew Huxham
Date: 04 Jul 2020
Commission: IBM Data Science – Coursera

https://github.com/ajhuxham/IBM_DS/blob/master/(Capstone)%20Aussie%20Pies%20in%20Toronto%20-%20Report.pdf

data matters
venue profiles

supporters
- Most likely add aussie pies to product line

complementary
- adds to the overall area identity
- attracts potential customers

challengers
- most likely not to welcome the business

data matters

venue groups

supporters
- Casual eateries
- Bars and pubs

complementary
- Restaurants
- Specialty eateries
- Transportation
- Household supplies
- Lifestyle
- Other

challengers
- Bakeries

data matters

sources

target areas

- **Wikipedia:** Practical and accessible, fit for purpose
- **Postal codes**: mail delivery area for borough and neighborhoods around Toronto
- **Borough**: municipality
- **Neighborhood**: smaller community of borough

spatial coordinates

- **Cognitive lab:** permissions needed for access
- **Coordinates**: latitude and longitude
- **Postal codes**: mail delivery areas
- **Target area**: borough, neighborhood

venue information

- **Foursquare:** requires permissions for access
- **Venues**: name and category of venue
- **Coordinates**: latitude and longitude

data matters

sources - overlaps

Data type	Wikipedia - M Postal Codes - Canada	Cognitive - Geospatial Data	Foursquare	Total sources
Borough				1
Latitude				2
Longitude				2
Neighborhood				1
Postal Code				2
Venue Name				1
Venue Category				1
Total fields	3	3	4	10

data matters

solutions

database

- **Database:** refined to boroughs with 'Toronto' in name
- **Constraints**: venue details limited to 100 records per 500 metre radius from neighborhood centre

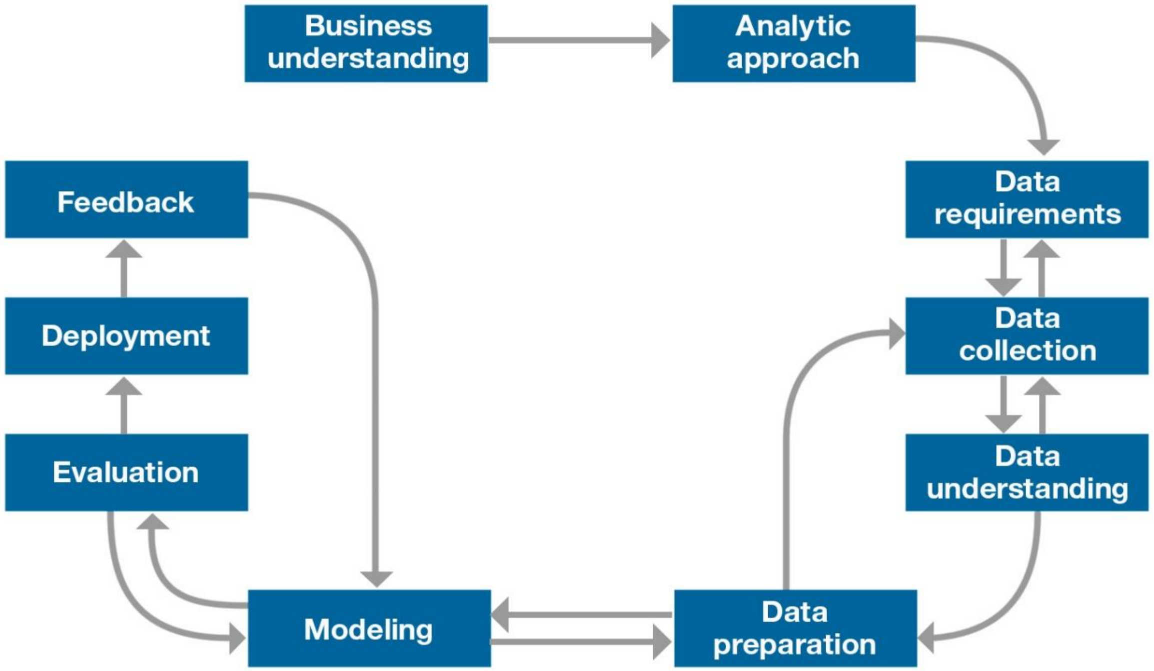
modelling

- **Classifications**: Numeric conversions of venue categories
- **Clusters**: identification of themes, clusters and outliers

visuals

- **dataframes:** matrix (table) of venues by category, area, clusters
- **Maps**: Spatial context, category clusters

analysis
guiding method



https://www.ibmbigdatahub.com/blog/why-we-need-methodology-data-science.

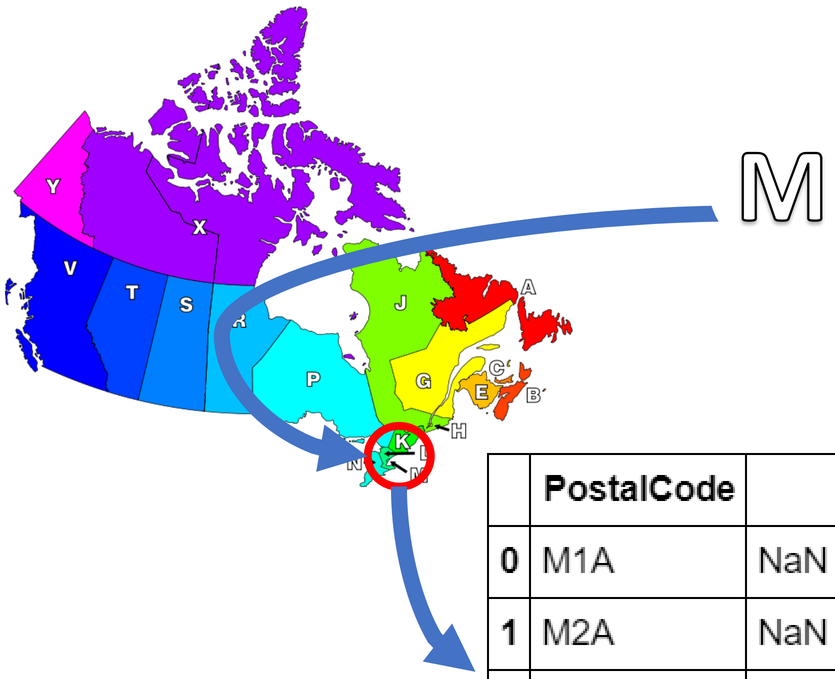
why a method

- **Value:** repeatability, reliability of results determines value
- **Confidence**: structured, tested process to generate results

defining target neighborhoods part 1

creating the first data frame

https://commons.wikimedia.org/wiki/File:Canadian_postal_district_map.svg



	PostalCode	Borough	Neighborhood
0	M1A	NaN	Not assigned
1	M2A	NaN	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

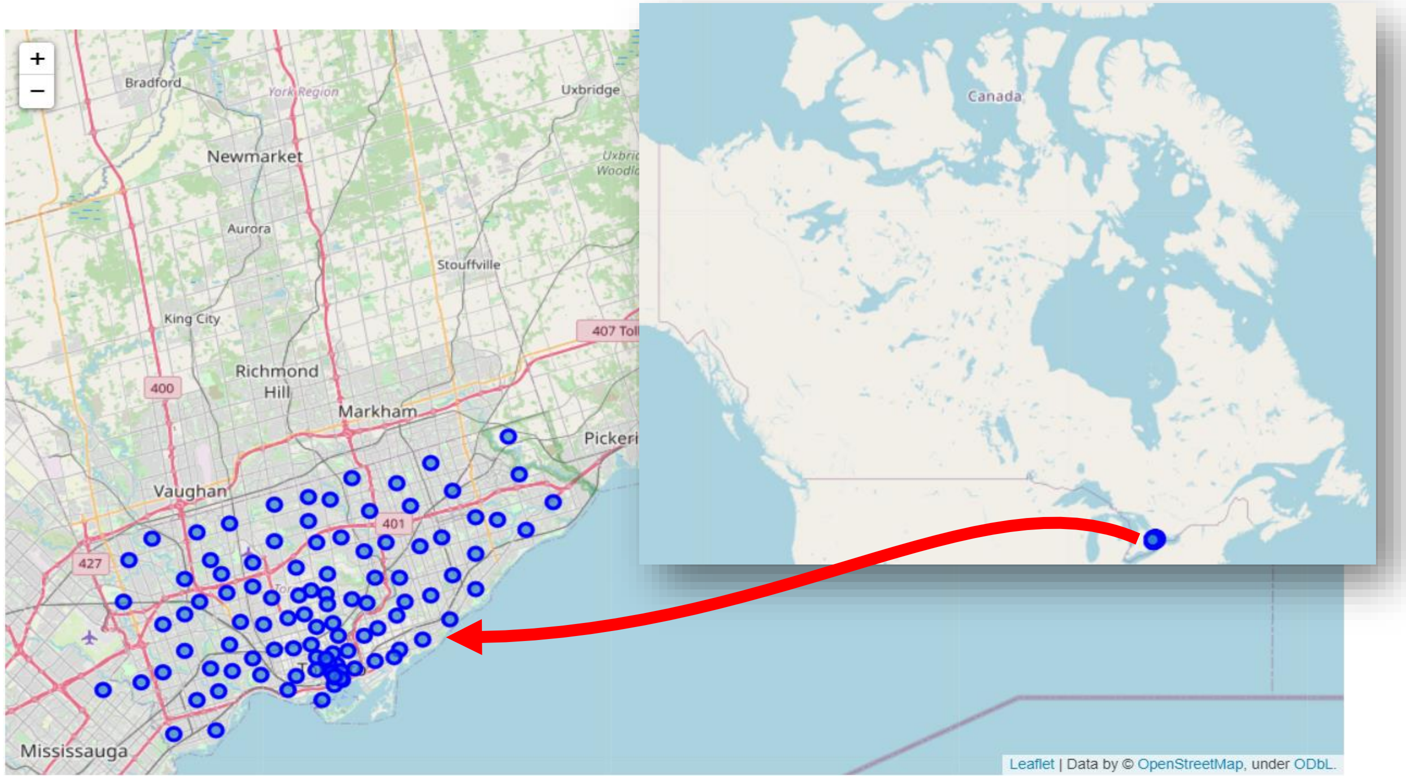
results

- **Source:** target areas, Canadian postal code 'M'
- **Start**: 180x rows of postal records; many records 'not assigned'
- **Preparation**: removed records 'not assigned'
- **Result**: 103x rows of boroughs and neighborhoods

Notebook cell references: ([0] – [5])

defining target neighborhoods part 1

integrating geospatial records



results

- **Source**: target areas dataframe + import geospatial coordinates
- **Preparation**: join dataframes on postal code; mapping algorithms
- **Result**: neighborhood centres overlaid on map of Toronto, Canada

Notebook cell references: ([6] – [10])

defining target neighborhoods part 2

refining the scope + integrating venue records

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	The Danforth West, Riverdale	43.679557	-79.352188	MenEssentials	43.677820	-79.351265	Cosmetics Shop

- Venue coordinates unique but very close to 'nearby' neighborhoods
- Limit 100 venues within 500 meters of neighborhood center

results

- **Source**: target areas dataframe + import Foursquare venue records
- **Preparation**: new dataframe limited to boroughs with 'Toronto' in name; append venue name, categories on nearby spatial coordinates
- **Result**: new dataframe report of 1,614 venues across target neighborhoods

Notebook cell references: ([11] – [16])

quantify venues by category
descriptive statistics

Top 5	Count
Coffee Shop	143
Café	89
Restaurant	54
Italian Restaurant	41
Hotel	37

Venue count / category	Value
Total categories	233.000000
Maximum	143.000000
Minimum	1.000000
Mean (average)	6.927039
Standard deviation	13.058515

results – indicating variety

- **Preparation**: group by venue category; descriptive statistics of venue category quantities
- **Result**: dataframe of venue categories (showing sample top 5 by count), summary statistics by count

Notebook cell references: ([17] – [19])

modelling

report most common venues

Neighborhood groups	1st Most Common Venue	10th Most Common Venue
Berczy Park	Coffee Shop	Shopping Mall
Brockton, Parkdale Village, Exhibition Place	Café	Furniture / Home Store
Business reply mail Processing Centre, South C...	Light Rail Station	Pizza Place
CN Tower, King and Spadina, Railway Lands, Har...	Airport Lounge	Airport Gate
Central Bay Street	Coffee Shop	Bubble Tea Shop

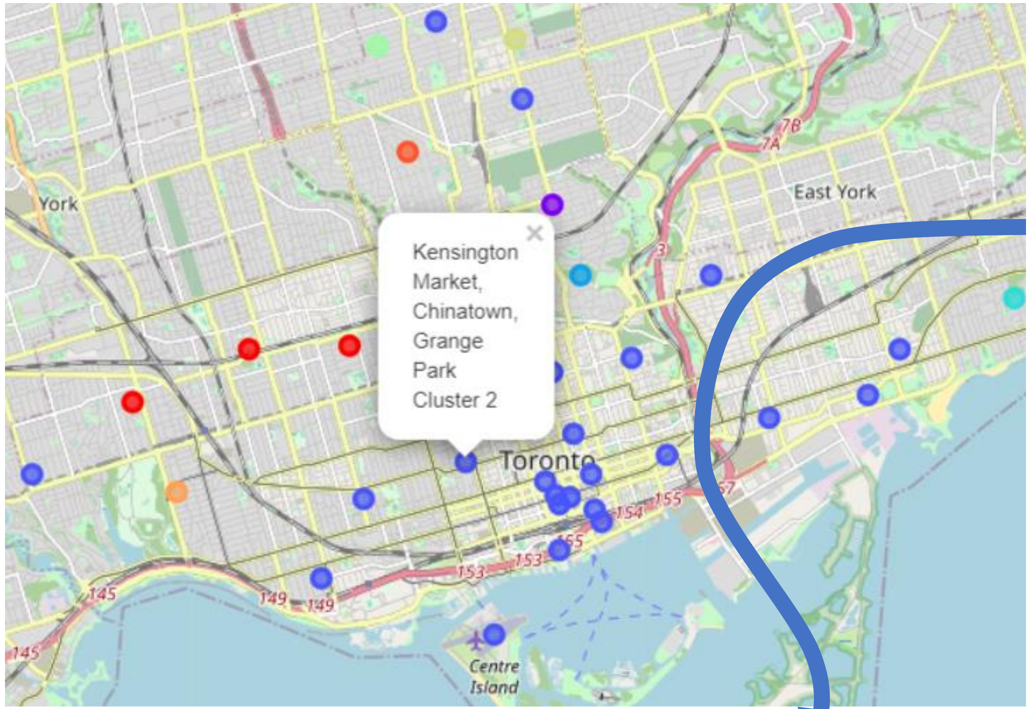
results – common venue report

- **Preparation**: numeric transformation of category text records, groupings of venues common to neighborhoods
- **Result**: dataframe of 10 most common venue categories across neighborhoods (showing sample of top 5 groups, 1st and 10th most common)

Notebook cell references: ([20] – [24])

modelling

map of most common venues



Cluster	Count
0	3
1	1
2	28
3	1
4	1
5	1
6	6
7	1
8	1
9	1

Cluster 2 sample

	Latitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
41	43.679557	2	Greek Restaurant	Coffee_Shop	Italian Restaurant	Bookstore	Ice Cream Shop
42	43.668999	2	Sandwich Place	Park	Fast Food Restaurant	Pizza Place	Gym

results – venue clusters

- **Preparation**: Kmeans cluster algorithm generating 10x clusters (0-9), assigned colours + mapping
- **Result**: 10x cluster groups overlaying metro Toronto, dominated by cluster 2 (blue)

Notebook cell references: ([25] – [39])

grouping by most common venues

map of most common venues
+ clusters

Group	Quantity	% of total Quantity
Restaurant	406	25.11%
Casual eatery	307	18.99%
Lifestyle	248	15.34%
Specialty eatery	237	14.66%
General retail	220	13.61%
Bar / Pub	82	5.07%
Bakery	41	2.54%
Hotel	37	2.29%
Transport	21	1.30%
Other	15	1.11%
Total	1614	

results – venue clusters

- **Preparation**: Venue categories manually allocated a group using a spreadsheet
- **Result**: report defining the profiles as per the business requirement

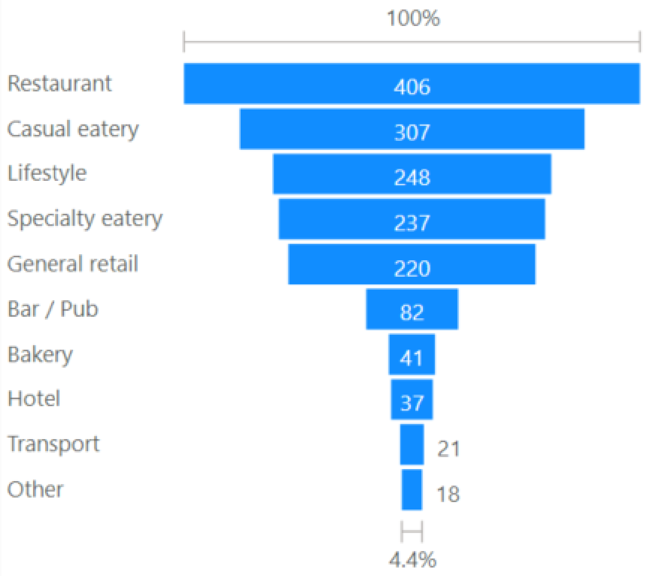
- *Prospective Supporters: Casual eateries 2nd largest group + Bar / Pubs*
- *Prospective Challengers: Bakeries ~<2.6%*

Note: performed separately on a spreadsheet

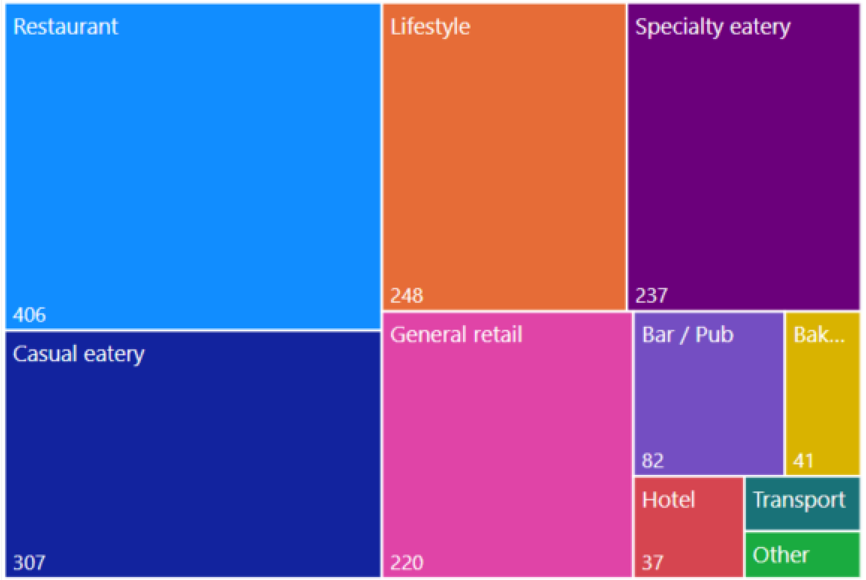
grouping venue categories

venue group profiles

'Funnel chart'



'Tree map'



results – group

- **Preparation**: Microsoft PowerBi used the Python data to create visuals
- **Result**: visuals showing comparative groups sizes across target neighborhoods

same information – shown different ways

Note: produced using Microsoft Power PI

grouping venue categories
international food themes

Nationality	Quantity
Japan	59
Italy	41
United States	39
Thailand	30
Mexico	22
France	21
Greece	17
China	11
Saudi Arabia	8
Brazil	7
India	7
Total	255

results – national foods

- **Preparation**: Grouped foods by nationality represented
- **Result**: visual indicating spread of national foods represented, with proportionate circle markers

No great Australian Bites!

Note: produced using Microsoft Power PI

analysis – quick review
additional analysis

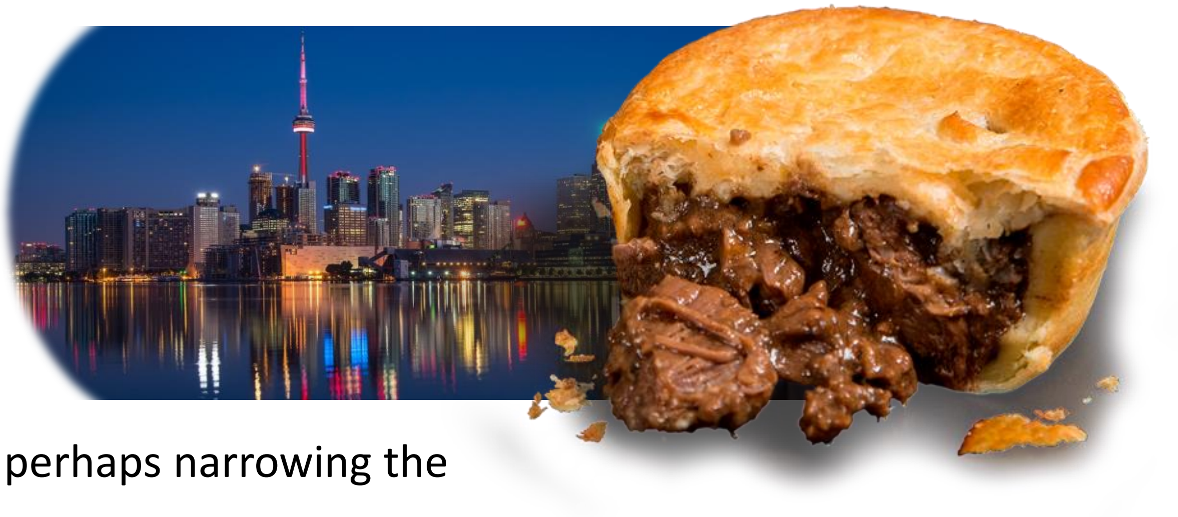
How methodology was applied though out the analysis steps

Methodology Step	Defining the Target Neighborhoods - part 1	Defining the Target Neighborhoods - part 2	Grouping venue categories by data priority	Modelling	Quantify venues by category	Grand Total
Step 1 - Business understanding	1					1
Step 2 - Analytic approach	1					1
Step 3 - Data requirements	1					1
Step 4 - Data collection	2	1				3
Step 5 - Data understanding	2	2	2		1	7
Step 6 - Data preparation	2	2	[illegible]	2	1	8
Step 7 - Modelling				2		2
Step 8 - Evaluation				3		3
Grand Total	**9**	**5**	**3**	**7**	**2**	**26**

Data preparation was continuous, requires most of the work

Steps 9 and 10 pending acceptance to progress by the business sponsor

analysis – quick review
additional analysis



suggestions for potential future analysis

- **refine clusters**: the cluster sets can continue to be refined, perhaps narrowing the areas of interest for this embryonic business proposal
- **demographics**: integrating socio-economic demographics with the venue categories may enhance the profile to understand potential customer base

the analysis mandate

- **status**: an initial profile of venue groups in metropolitan Toronto has been created
- **evaluate mandate**: the business sponsor needs to confirm if further analysis is warranted – no point if the decision is not to progress

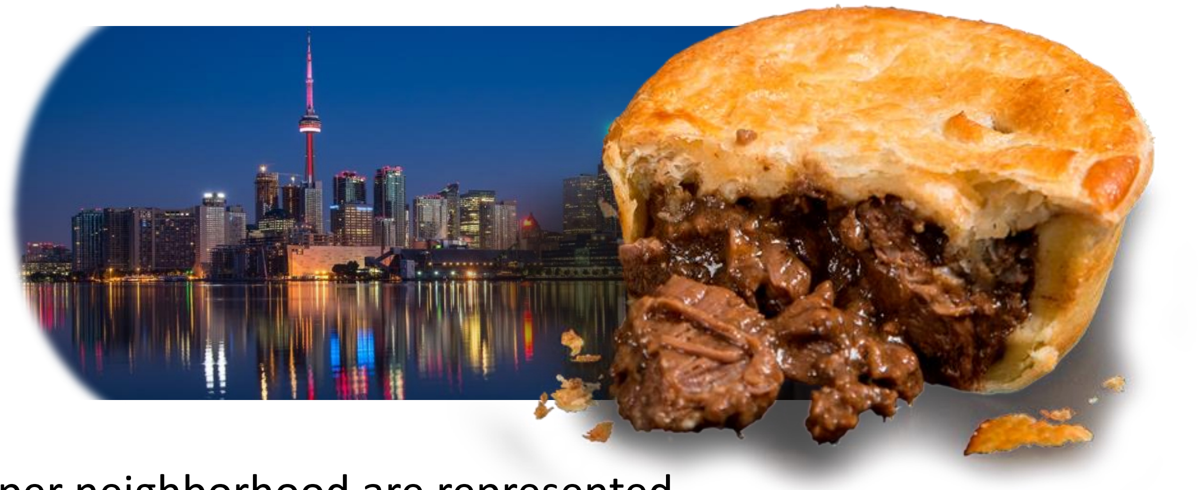
conclusion *basic metrics*

- **Target neighborhoods**: A shortlisted focal subset of **39** boroughs that include 'Toronto' in the name was systematically refined from an initial list of **180** postal codes.
- **Data priority:** The venue categories were grouped according to the data priorities, reshaping the perspective to show restaurants as the most prominent venue across the sample set.
- **Venue diversity:** The **1,617** venues across **233** unique categories identified, with coffee shops and café's as the most prominent.

conclusion *indications*

- **Opportunity:** Australian pies appear to be a relatively distinct product opportunity
- **Next steps:** reasonable prospects to 'have a go'

conclusion *themes*



- **Venue clusters:** The clusters of most common venues per neighborhood are represented by dataframe reports and street maps overlaid with markers indicating the clusters.
- **Venue grouping profile:** Approximately 24% of the 1,614 identified venues are either casual eateries or bar/ pubs, fitting the prospective market as potential distributors. Less than 3% are bakeries selling pastry products.
- **National food themes:** Venue categories indicate a diverse representation of international cuisines, of which Japan, Italy and United States dominate.

Australia in general and Australian pies in particular do not appear represented.