NEIGHBORHOODS OF TORONTO

Sarah Beth Sanchini

OVERVIEW

PROBLEM

Need information based on neighborhoods

INTEREST

People want to know what services are in the area that they live and work

TARGET AUDIENCE

People new to the area of Toronto or business owners wanting to open a new business

METHOD

Python code used to extract data, explore and analyze neighborhoods

DATA

FOURSQUARE

The Foursquare data provide locational data

WIKIPEDIA

Information about different area codes in Toronto scraped from Wikipedia

GEOSPACIAL DATA

Postal Codes and thier corresponding Latitude and Longitude values

METHODOLOGY

DATA CLEANING

Data downloading and data scraping from multiple sources to supply the data necessary for analysis

EXPLORATORY ANALYSIS

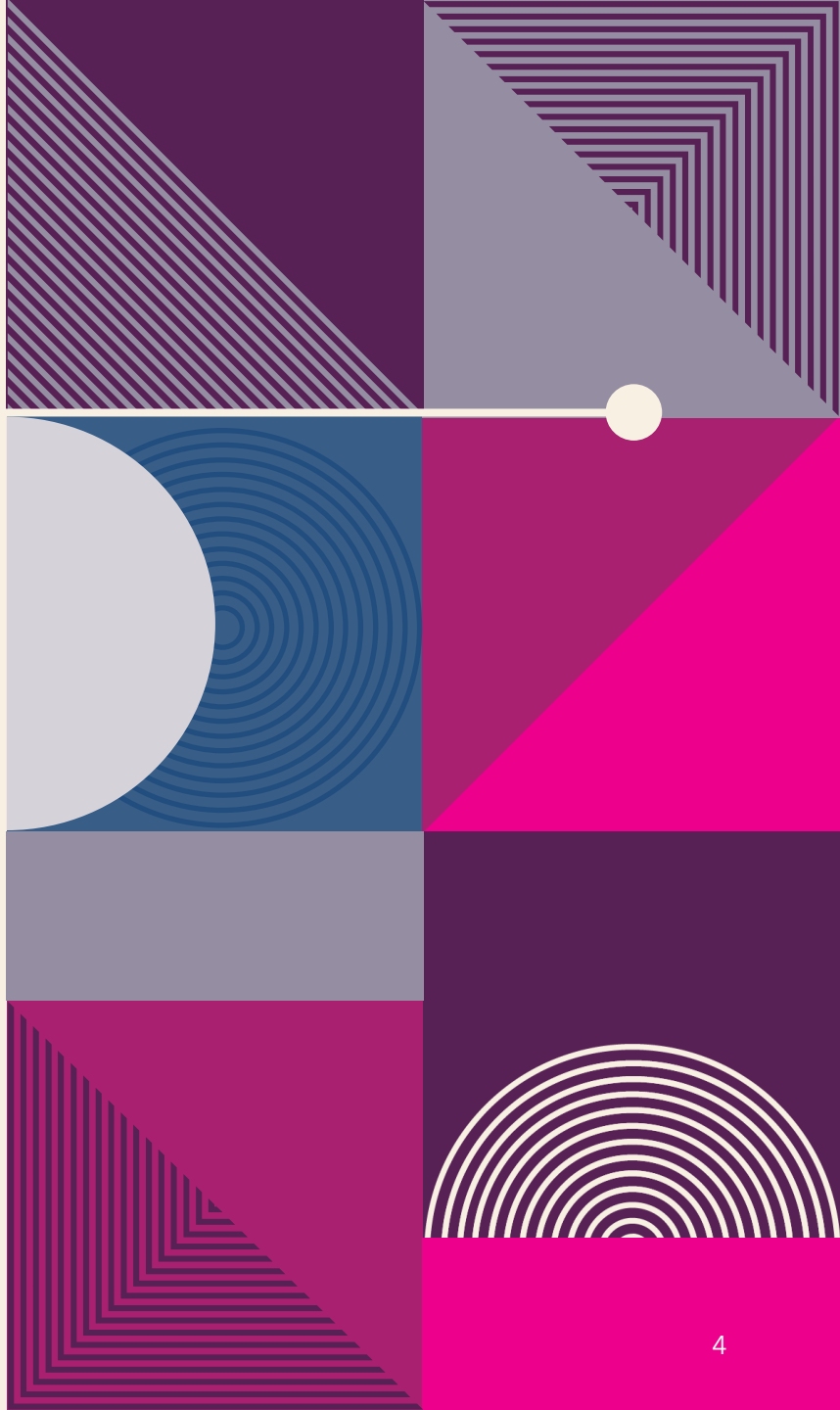
Exploratory Analysis on the Parkwoods

CLUSTERING

K-means clustering was performed to determine the top 10 most common venue type per neighborhood

DISCUSSION

Reviewing the outputs of the code and algorithms to glean insightful knowledge.



INTRODUCTION / BUSINESS PROBLEM

"THE BATTLE OF THE NEIGHBORHOODS"

Problem:

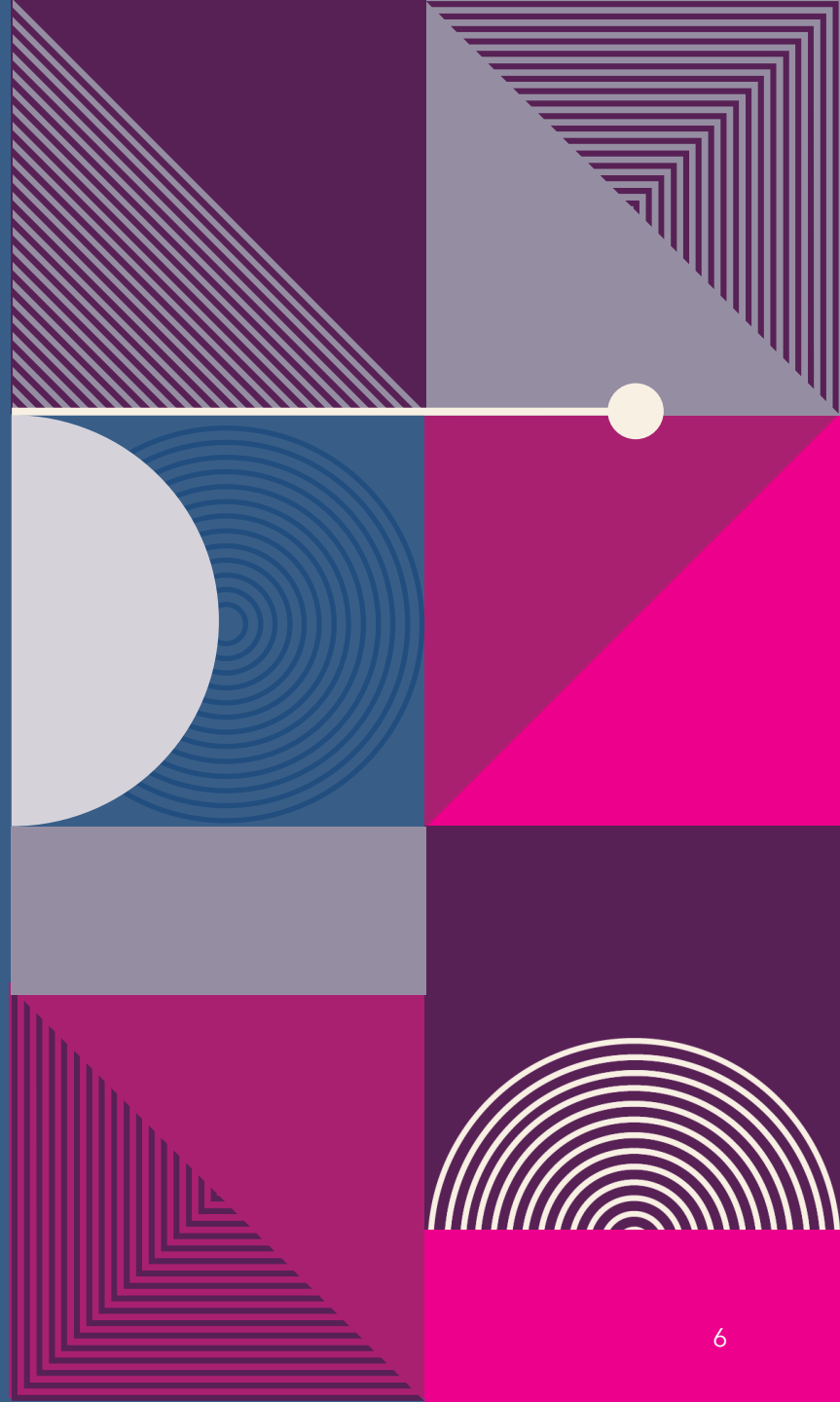
Different neighborhoods vary from each other in size, population and what services they offer. When a new person moves to a location, it is helpful to know details on each neighborhood so they can choose where to live according to their needs. This project clusters neighborhoods and then shows the top 5 venues in each neighborhood.

DATA – DETAILED VIEW

Foursquare API Data: The Foursquare data provide locational data - this includes venues, interesting locations, schools etc. given a specific location.

Information retrieved per Area Code:

Neighborhood
Neighborhood Latitude
Neighborhood Longitude
Venue
Name of Venue
Venue Latitude
Venue Longitude
Venue Category

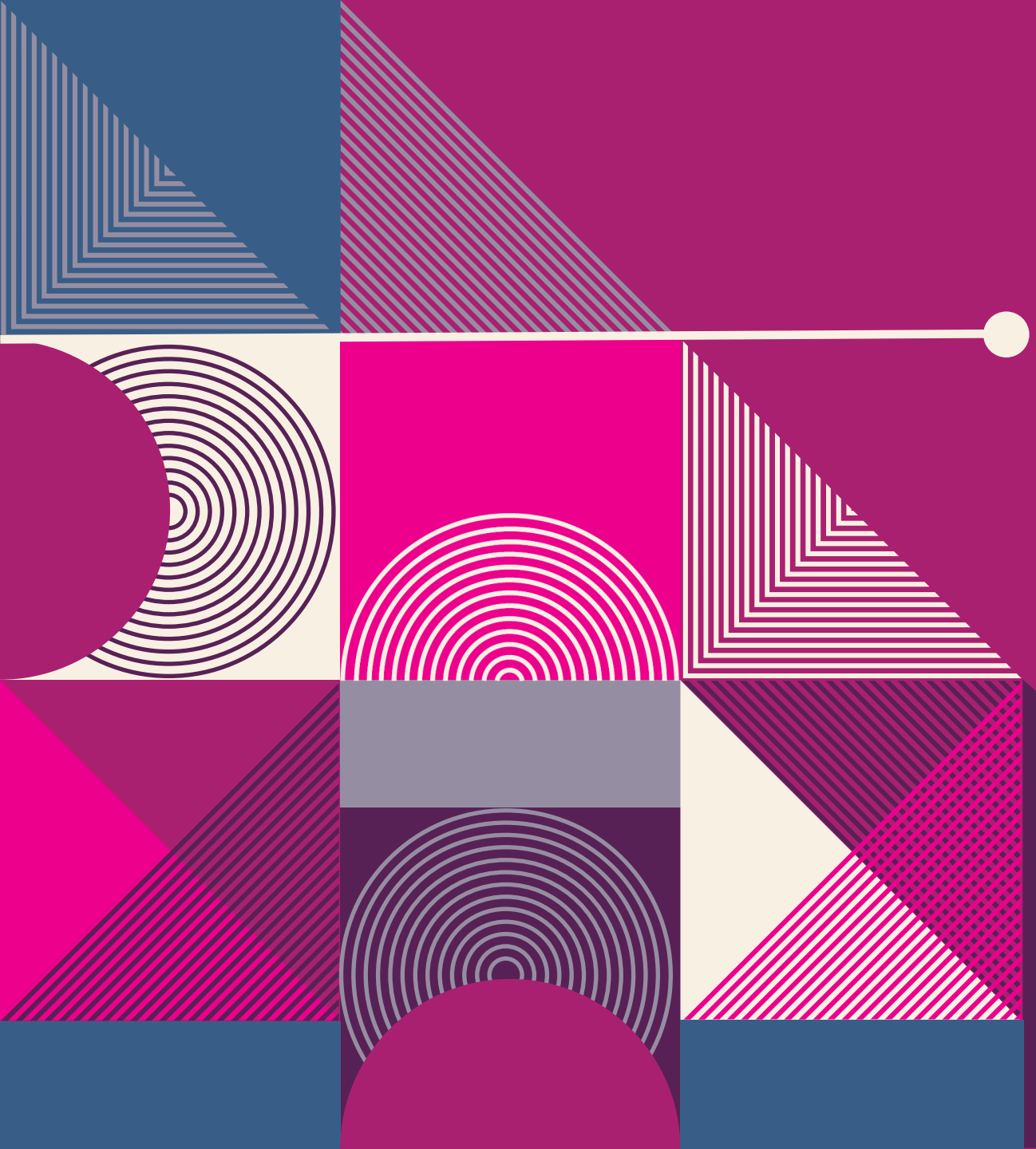


DATA – DETAILED VIEW

Wikipedia: Information about different area codes in Toronto scraped from Wikipedia Information retrived: Latitude, Longitude, Area/Zip Codes Data Link:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

3) **Geospacial Data**: Postal Codes and thier corresponding Latitude and Longitude values Data Link: https://cocl.us/Geospatial_data



DATA CLEANING

DATA CLEANING

Data downloading and data scraping from multiple sources to supply the data necessary for analysis. To get the data on Toronto, we first had to webscrape from Wikipedia. This data was cleaned so the columns would be in Postal Code, Borough and Neighborhood.

The postal code was the key to converting Latitude/Longitude values into Boroughs and Neighborhoods.

Out[8]:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Queen's Park	Ontario Provincial Government	43.662301	-79.389494
5	M9A	Etobicoke	Islington Avenue	43.667856	-79.532242

EXPLORATORY ANALYSIS

EXPLORATORY ANALYSIS PARKWOODS

To get and idea of what specifics to extract, we start with just one neighborhood = Parkwoods

Out[36]:

	name	categories	lat	lng
0	Brookbanks Park	Park	43.751976	-79.332140
1	KFC	Fast Food Restaurant	43.754387	-79.333021
2	Variety Store	Food & Drink Shop	43.751974	-79.333114

EXPLORATORY ANALYSIS

– NUMBER OF VENUES PER NEIGHBORHOOD

Foursquare is utilized to show which venues and categories are for each neighborhood in Toronto so we can compare and explore what results returned.

```
A total of 4260 venues have been returned
Grouped by neighborhood:
```

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Agincourt	8	8	8	8	8	8
Alderwood, Long Branch	18	18	18	18	18	18
Bathurst Manor, Wilson Heights, Downsview North	46	46	46	46	46	46
Bayview Village	8	8	8	8	8	8
Bedford Park, Lawrence Manor East	50	50	50	50	50	50

EXPLORATORY ANALYSIS TOP 5 VENUE TYPES

To have a more detailed look at what venues are in each neighborhood, we return the top 5 values which show the frequency of that particular category. Here are a few examples:

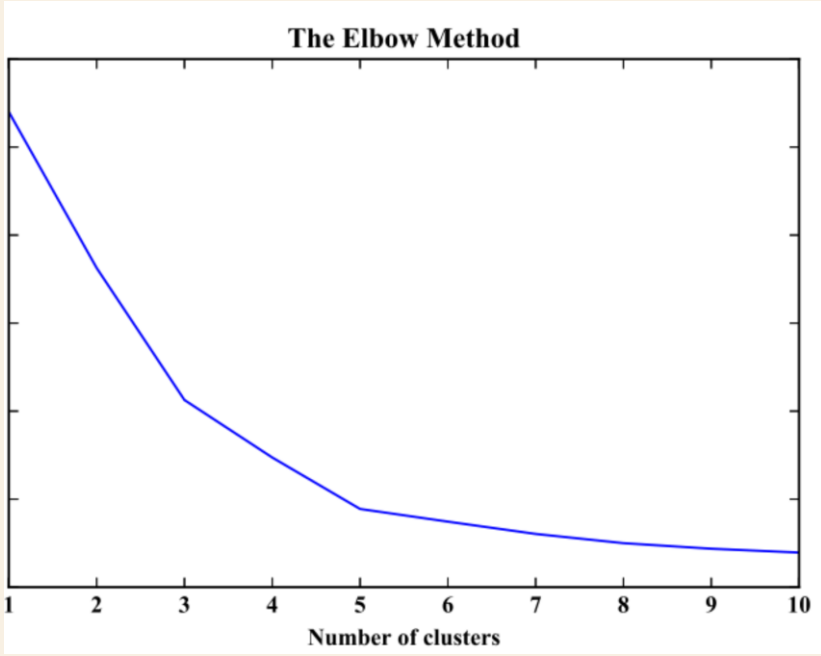
```
----Agincourt----
                       venue  freq
0             Breakfast Spot  0.25
1                     Lounge  0.25
2  Latin American Restaurant  0.25
3             Clothing Store  0.25
4          Accessories Store  0.00


----Alderwood, Long Branch----
                  venue  freq
0         Pizza Place  0.22
1  Athletics & Sports  0.11
2        Coffee Shop  0.11
3                Pub  0.11
4      Sandwich Place  0.11
```

CLUSTERING METHOD

K-MEANS ALGORITHM

- Used this clustering technique to determine what were the top 10 venue types/category for each neighborhood
- Used K=5 for the algorithm

ELBOW CURVE WITH ERROR



CLUSTER RESULTS

Algorithm

K-Means clustering algorithm with K = 5

Labels

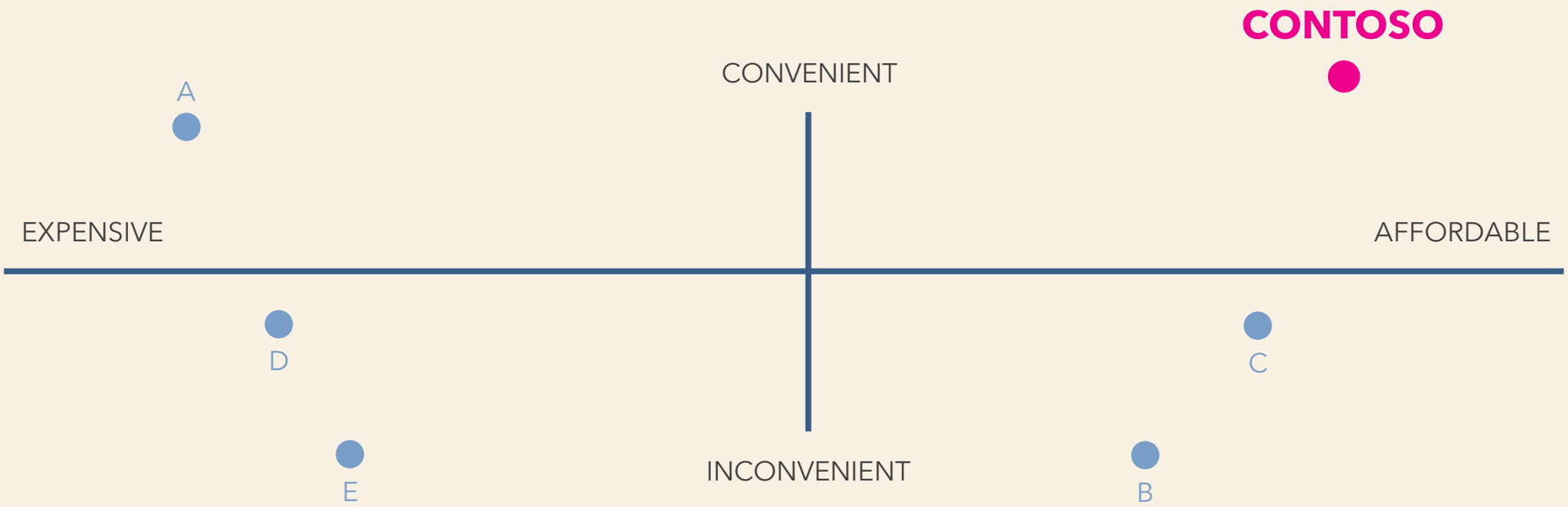
Latitude, Longitude and Neighborhood

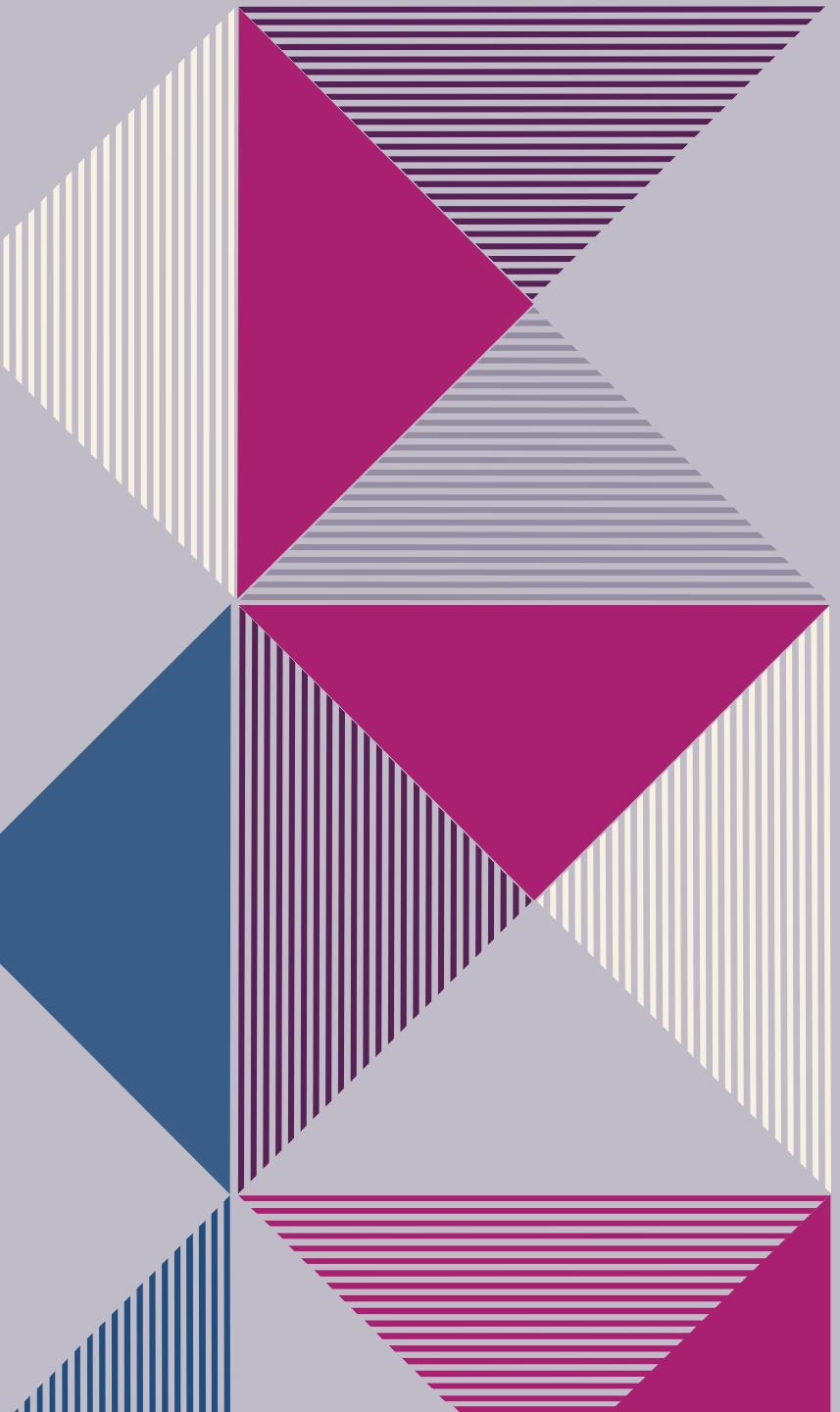
Output

Out[49]:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	North York	1.0	Food & Drink Shop	Fast Food Restaurant	Park	Dumpling Restaurant	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Drugstore
21	York	1.0	Park	Women's Store	Pool	Yoga Studio	Drugstore	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant

OUR COMPETITION GRAPHIC





GROWTH STRATEGY

How we'll scale in the future

PHASE 1: FEB 20XX

Roll out product to high profile or top-level participants to help establish the product

PHASE 2: MAY 20XX

Release of the product to the general public and monitor press release and social media accounts

PHASE 3: OCT 20XX

Gather feedback and adjust product design as necessary

DISCUSSION

PARKWOODS

This portion shows we can look in detail into each particular neighborhood

Out[36]:

	name	categories	lat	lng
0	Brookbanks Park	Park	43.751976	-79.332140
1	KFC	Fast Food Restaurant	43.754387	-79.333021
2	Variety Store	Food & Drink Shop	43.751974	-79.333114

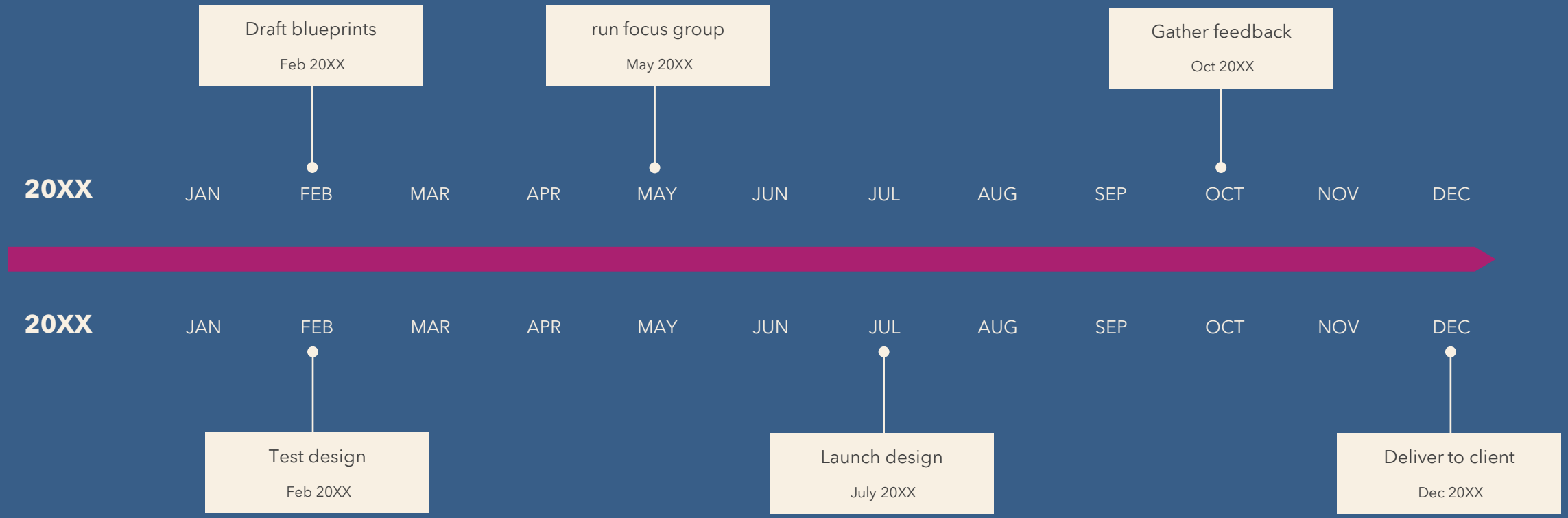
We can see that in Parkwoods, nearby there is a park called Brookbanks Park, a KFC and a Variety store. This level of detail can be useful for looking closely at a few neighborhoods

MOST COMMON VENUE CATEGORY

This portion shows what type of venue categories are most prominent in the neighborhood.

1st Most Common Venue > Food & Drink Shop
2nd Most Common Venue > Fast Food Restaurant
3rd Most Common Venue > Park
4th Most Common Venue > Dumpling Restaurant
5th Most Common Venue > Discount Store
6th Most Common Venue > Distribution Center
7th Most Common Venue > Dog Run
8th Most Common Venue > Doner Restaurant
9th Most Common Venue > Donut Shop
10th Most Common Venue > Drugstore

TWO YEAR ACTION PLAN



FINANCIALS

	Year 1	Year 2	Year 3	
INCOME				
Users	50,000	400,000	1,600,000	
Sales	500,000	4,000,000	16,000,000	
Average price per sale	75	80	90	
Revenue @ 15%	5,625,000	48,000,000	216,000,000	
Gross profit	**5,625,000**	**48,000,000**	**216,000,000**	
EXPENSES				
Sales & marketing	5,062,500	38,400,000	151,200,000	70%
Customer service	1,687,500	9,600,000	21,600,000	10%
Product development	562,500	2,400,000	10,800,000	5%
Research	281,250	2,400,000	4,320,000	2%
Total expenses	**7,593,750**	**52,800,000**	**187,920,000**	

MEET THE TEAM

Takuma Hayashi

President

Mirjam Nilsson

Chief Executive Officer

Flora Berggren

Chief Operations Officer

Rajesh Santoshi

VP Marketing

MEET THE TEAM

Takuma Hayashi
President

Mirjam Nilsson
Chief Executive Officer

Flora Berggren
Chief Operations Officer

Rajesh Santoshi
VP Marketing

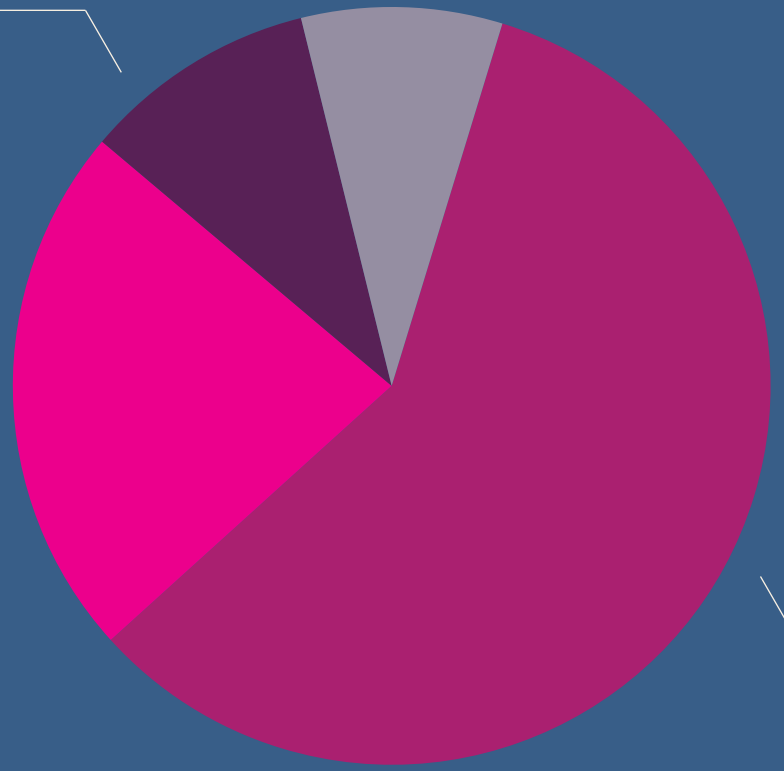
Graham Barnes
VP Product

Rowan Murphy
SEO Strategist

Elizabeth Moore
Product Designer

Robin Kline
Content Developer

FUNDING



$14,000

Angel Investments

Amount obtained through other investors

$12,000

Property

Revenue obtained from property rentals

$32,000

Cash

Liquid cash we have on hand

$82,000

Shares

Number of shares converted into USD

SUMMARY

Analyzing the neighborhoods of Toronto using only open data sources was surprisingly insightful. The grouping of Toronto into neighborhoods (using Latitude and Longitude values) transforms the data into information that is easy to digest for users.

Furthermore, by analyzing what are the most common types of venue categories in each neighborhood, a person is able to get an deeper understanding of each area.

This information can be used for various purposes, such as what is lacking in a neighborhood so you can open a business there.

Utilizing Fourquare and Wikipedia data lead to detailed information on the city in Toronto. **The battle of the neighborhoods is concluded!**

THANK YOU