

The background features a series of concentric circles in light gray, some solid and some dashed, creating a ripple effect. A large red speech bubble is centered on the page, containing the title and subtitle. The speech bubble has a small tail pointing downwards.

Identifying Potential Cities for Expansions

IBM Coursera Datascience – Capstone Project

Aidan Donoghue

The background of the slide features several thin, curved lines in a light gray color, some solid and some dashed, creating a sense of motion or a stylized globe. On the left side, there is a large red rectangular area that serves as a container for the title. This area is divided into two horizontal sections: a smaller top section and a larger bottom section. The word "Introduction" is written in white text within the larger bottom section. To the right of this red area, there is a bulleted list of three items, each starting with a small red square. The text in the list is black and describes a business expansion scenario.

Introduction

- A client has a successful business in Auckland, New Zealand
- They are looking to expand, but want to know the best cities to consider.
- Business success can be determined by comparing how similar other cities are to the clients current city.

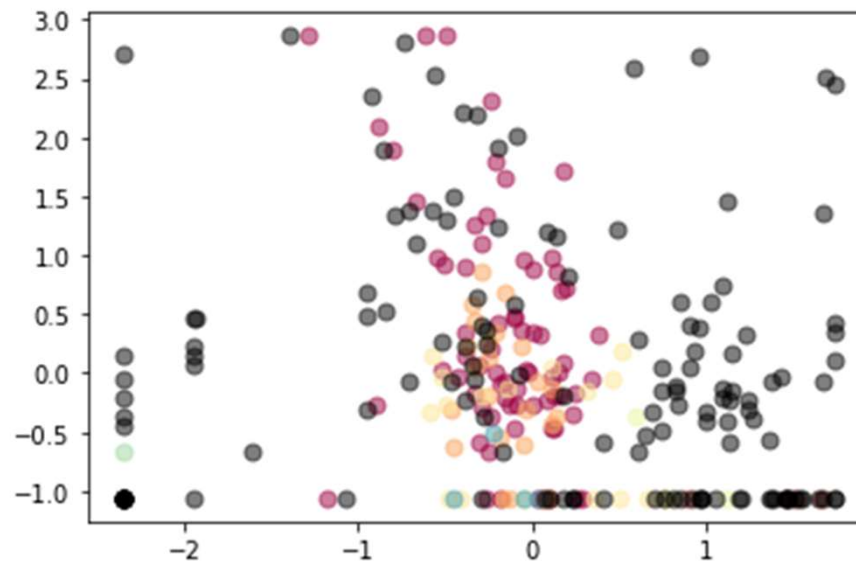
The background of the slide features several sets of thin, curved lines in light gray, some solid and some dashed, creating a sense of motion or data flow. A large red speech bubble is positioned on the left side, containing the word 'Data'.

Data

- Two data sources are used for this analysis:
 - Quality-of-life database (Kaggle)
 - FourSquare Location Data
- Both data sources were modified to help with the machine learning investigations

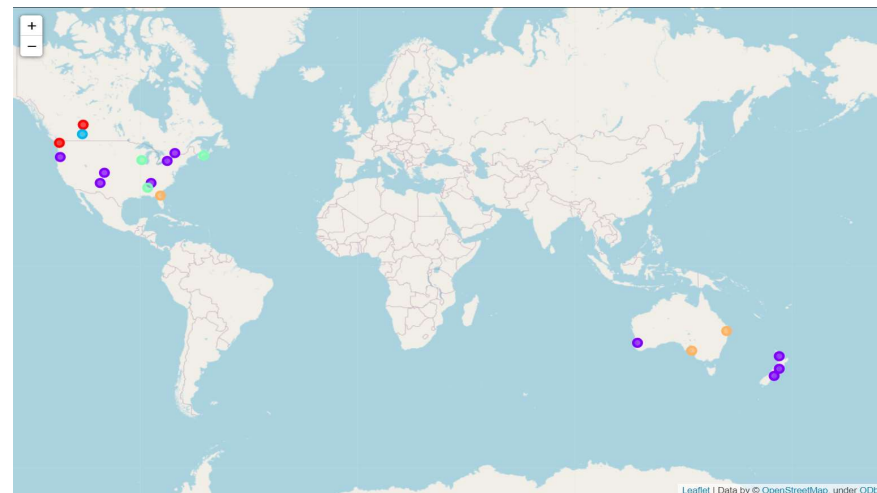
Data Relationships

- The relationships between cities have been investigated by unsupervised classification / clustering
- The quality-of-life results were assessed first, using DBSCAN unsupervised clustering.



Data Relationships

- The results of the DBSCAN analysis, were passed into FourSquare to gather location venue data.
- Finally a k-means method was used to further separate/group the cities.



Conclusion

- The final output is a list of cities that have both:
 - Similar quality-of-life indices
 - Similar city-centre venues
- For this example it appears the coffee shops, bars and food are the leading similarities in the selected cities.

	UA_Name	UA_Country	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
12	Auckland	New Zealand	Theater	Japanese Restaurant	Steakhouse
260	Wellington	New Zealand	Café	Coffee Shop	Bar
185	Perth	Australia	Coffee Shop	Sushi Restaurant	Korean Restaurant
67	Colorado Springs	Colorado	Coffee Shop	Hotel	Mexican Restaurant
57	Chattanooga	Tennessee	Bar	Coffee Shop	Music Venue
47	Buffalo	New York	Hotel	Coffee Shop	Pizza Place

The background of the slide features several thin, curved lines in a light gray color, some solid and some dashed, creating a sense of motion or flow. On the left side, there is a large red rectangular area with a white triangle pointing downwards from its bottom center, resembling a speech bubble or a callout box. The text 'Further Considerations' is written in white within this red area.

Further Considerations

- As this is a fairly highlevel investigation, it is recommended more effort is put into determining quality-of-life data sets.
- The location data should be refined to represent the clients business (i.e. Pizza shops)
- The cities have also been chosen very coarsely, we would recommend refining the search to boroughs/neighbourhoods in each city.