Predicting the best suburbs to live in

Finding the right suburb is important for young families

◈ Families with kids need a safe environment to raise their kids

◈ Access to educational services such as schools and museums are important

◈ Access to medical services such as hospitals and pharmacies are important

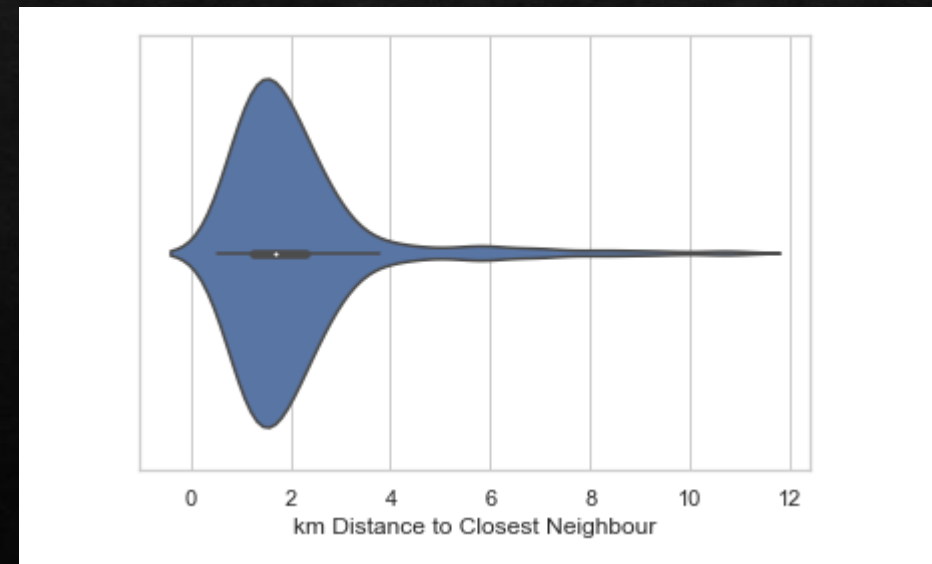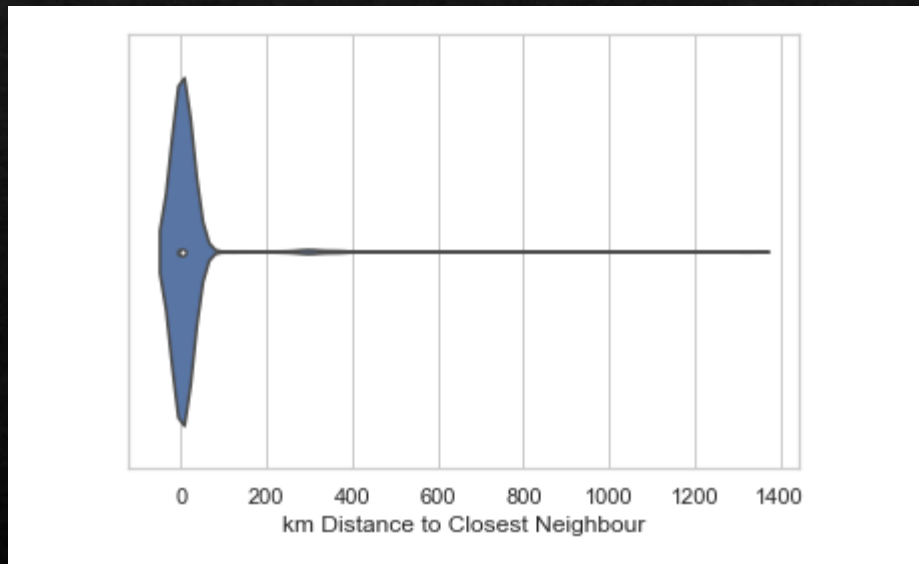◈ Access to recreational and outdoor facilities such as parks

Data acquisition and cleaning

◈ Suburbs names for the greater Perth region was web scraped from a Wikipedia page

   ◇ 353 suburbs were found

◈ The geospatial data for each suburb was retrieved using geopy libraries

◈ Venues within a radius of each suburb were obtained using the Foursquare API

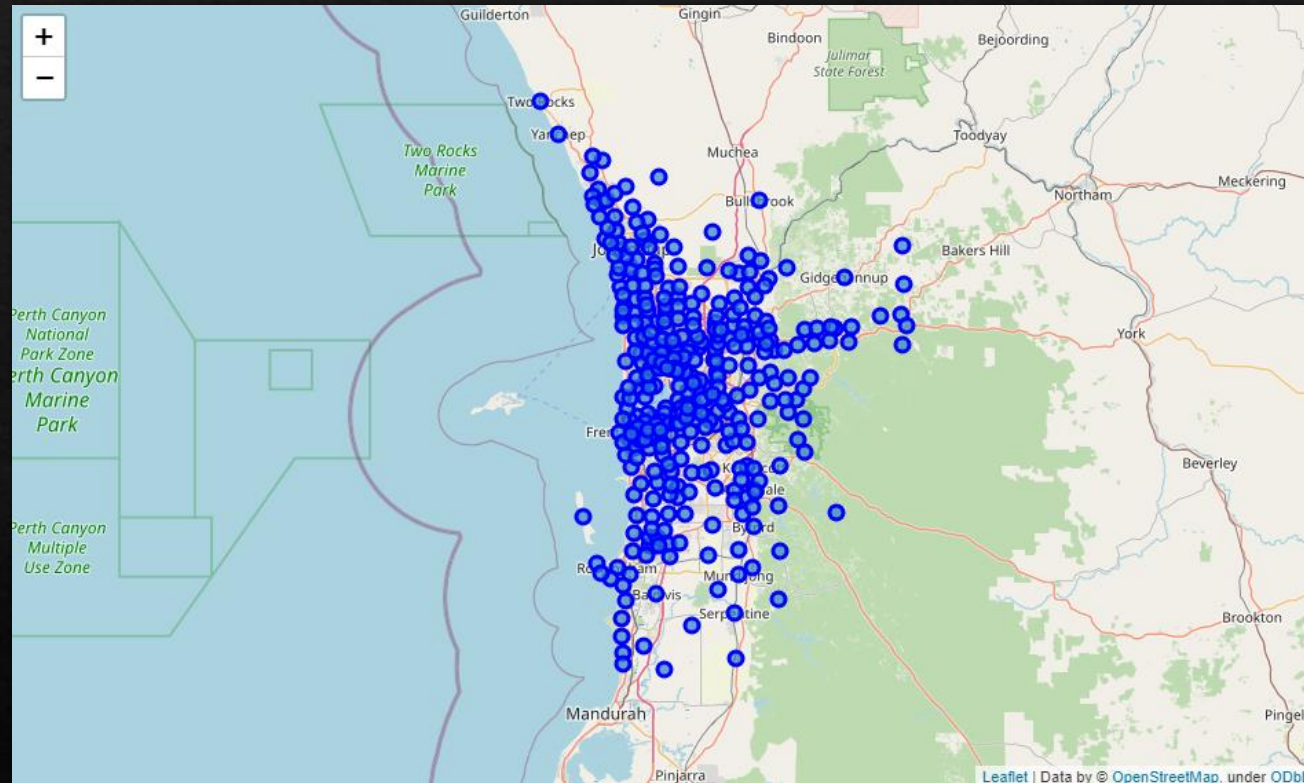   ◇ Only a subset of 8 venue categories were conspired as applicable venues

Data acquisition and cleaning

◈ A few outliers in the dataset was detected. These suburbs were outside of the area considered for this model.

◈ Suburbs with neighbouring suburb further then 50km were removed from the dataset

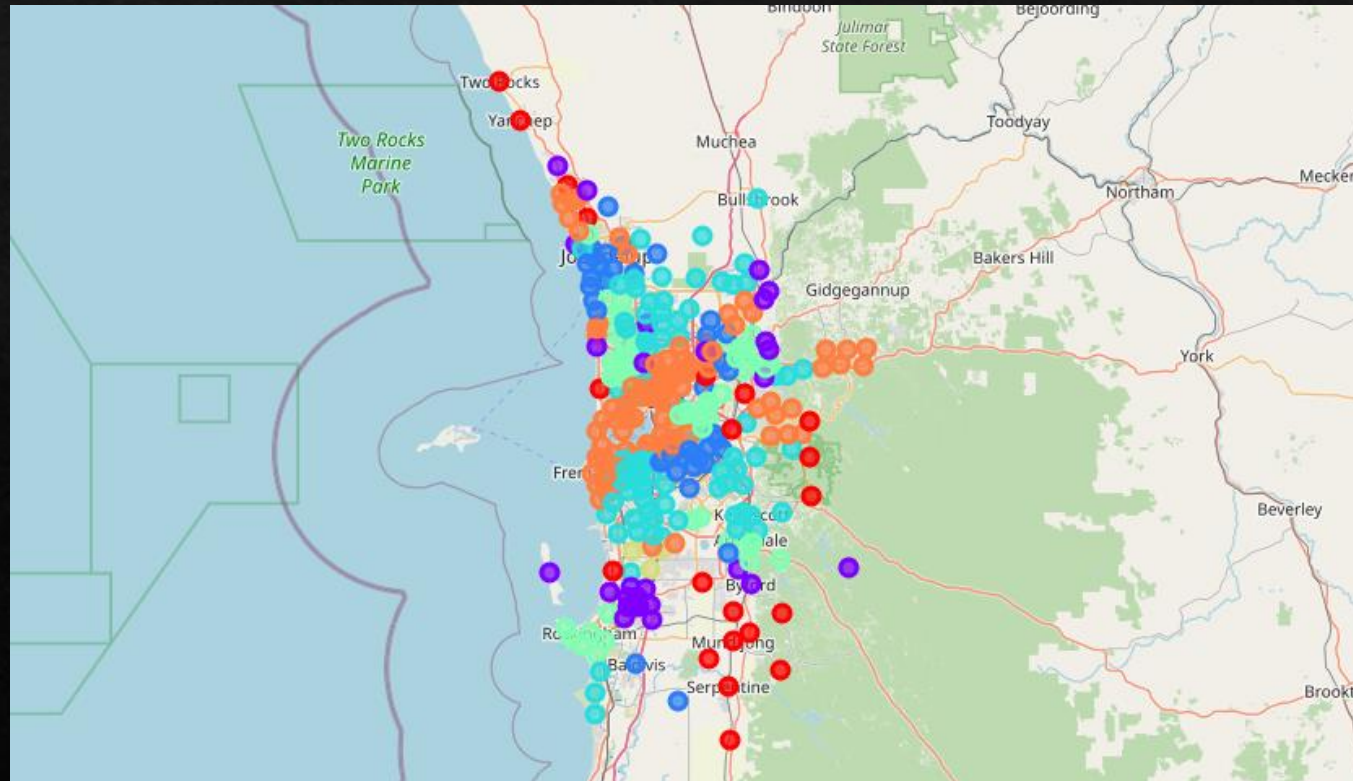# Visualising the suburb dataset on a map

Acquiring venue data

◇ Foursquare API was used to retrieve all venues based on predetermined venue category list

◇ the dataset was grouped by suburb by taking the mean of the frequency of occurrence of each category

◇ the top 5 most common and most uncommon venues were calculated

	Suburb	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	1st Most Uncommon Venue	2nd Most Uncommon Venue	3rd Most Uncommon Venue	4th Most Uncommon Venue	5th Most Uncommon Venue
0	Alexander Heights	Dance Studio	Zoo Exhibit	Movie Theater	Arts & Entertainment	Circus	Art Gallery	Theater	Science Museum	Salsa Club	Roller Rink
1	Alfred Cove	Dance Studio	Movie Theater	Zoo Exhibit	Arts & Entertainment	Circus	Art Gallery	Theater	Science Museum	Salsa Club	Roller Rink
2	Alkimos	Arts & Entertainment	Music Venue	Zoo Exhibit	Movie Theater	Circus	Art Gallery	Theater	Science Museum	Salsa Club	Roller Rink
3	Applecross	Art Gallery	Movie Theater	Dance Studio	Indie Movie Theater	Arts & Entertainment	Multiplex	Theater	Science Museum	Salsa Club	Roller Rink
4	Ardross	Art Gallery	Movie Theater	Dance Studio	History Museum	Arts & Entertainment	Multiplex	Theater	Science Museum	Salsa Club	Roller Rink

Clustering

◈ Using k-means clustering method the suburbs were clustered into 7 clusters

◈ Clusters plotted on a map using folium

Clustering

⬥ The top 5 most common and uncommon venues grouped by cluster label to determine theme

Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	1st Most Uncommon Venue	2nd Most Uncommon Venue	3rd Most Uncommon Venue	4th Most Uncommon Venue	5th Most Uncommon Venue
3	Dance Studio	Zoo Exhibit	Movie Theater	Arts & Entertainment	Circus	Art Gallery	Theater	Science Museum	Salsa Club	Roller Rink
3	Dance Studio	Movie Theater	Zoo Exhibit	Arts & Entertainment	Circus	Art Gallery	Theater	Science Museum	Salsa Club	Roller Rink
1	Arts & Entertainment	Music Venue	Zoo Exhibit	Movie Theater	Circus	Art Gallery	Theater	Science Museum	Salsa Club	Roller Rink
6	Art Gallery	Movie Theater	Dance Studio	Indie Movie Theater	Arts & Entertainment	Multiplex	Theater	Science Museum	Salsa Club	Roller Rink
6	Art Gallery	Movie Theater	Dance Studio	History Museum	Arts & Entertainment	Multiplex	Theater	Science Museum	Salsa Club	Roller Rink

Recommendations

◈ Incorporate more data into model such as house pricing, population and crime statistics

◈ Add weighted venue categories as some venues might be more important then others

◈ Use better defined suburb boundaries for more accurate suburb analysis

◈ Accuracy of the model has room for improvement

Conclusion

◈ Clustering models such as k-means work well when working with geospatial data

◈ Clustering models can be build to aid families in determining best location to live in based on proximity to desired venues

◈ This model can be leveraged for many more applications