1   INTRODUCTION

Pandemic has severly limited the available leisure opportunities which negatively affected the holidays of the people. However, once a country successfully flatten the curve, businessess will start to reopen. Though the borders of states may remain closed, it is possible to alleviate the feeling of enclosure to the people by providing diverse venues. By identifying popular venue categories using Foursquare Location Data in a state and suggesting these to a different state, we can create a sense of different land or place. For example, in Sydney, theatres are popular but rare in Brisbane, therefore open a theatre to create a Sydneyesque feeling.

2   DATA DESCRIPTION

The data for this problem uses Foursquare and the location of Australian cities. I will be investigating major cities in Australia namely, Sydney, Melbourne, Brisbane therefore their geolocation will be taken through google maps by finding their center longitude and latitude. As for Foursquare, I will be utilizing its /Venues data and this will return the top 100 venues in the city.

3   METHODOLOGY

The approach to this problem is simple. Use Foursquare Venue option to return the top 100 venues. Group the venues by categories to have the frequency values. Identify which categories popular and exists in the state.

Firstly, I would create the cities datasets. Then combine them all so that the categories are all consistent throughout all datasets in order for one hot encoding to work. Then groupby and find their frequency.

To give out a list of suggested categories to open in a state. Find which states are disimilar together using K-NN and measure their purity index. Kmeans groups the data through a 'bubble' whereas data points closer to the kmean cluster centroid are labeled as the same. This means that the kmean cluster centroids distance can indicate similarity between them. Euclidean distances between the centroids is used to identify their 'similarity'

4 RESULTS

Here is the frequency of venue categories of the cities.

```
----Brisbane----
                  venue  freq
0                  Café  0.11
1                 Hotel  0.07
2                  Park  0.06
3                   Bar  0.06
4              Beer Bar  0.05
5           Coffee Shop  0.03
6   Japanese Restaurant  0.03
7                Bakery  0.02
8          Burger Joint  0.02
9      Pedestrian Plaza  0.02


----Melbourne----
                  venue  freq
0          Cocktail Bar  0.09
1           Coffee Shop  0.07
2                   Bar  0.07
3      Asian Restaurant  0.04
4                  Café  0.04
5          Dessert Shop  0.04
6     Korean Restaurant  0.03
7                 Plaza  0.03
8              Wine Bar  0.02
9             Bookstore  0.02


----Sydney----
                  venue  freq
0                  Café  0.10
1          Cocktail Bar  0.05
2                 Hotel  0.05
3       Thai Restaurant  0.05
4            Restaurant  0.04
5           Coffee Shop  0.04
6                  Park  0.04
7               Theater  0.04
8    Chinese Restaurant  0.03
9  Australian Restaurant  0.03
```

Here is the Kmeans result:

	Cluster Labels	City	Argentinian Restaurant	Art Gallery	Art Museum	Asian Restaurant	Athletics & Sports	Australian Restaurant	BBQ Joint	Bakery	...	Tapas Restaurant	R(
0	0	Brisbane	0.00	0.02	0.00	0.02	0.01	0.01	0.00	0.02	...	0.00	
1	1	Melbourne	0.01	0.01	0.00	0.04	0.00	0.01	0.01	0.01	...	0.00	
2	2	Sydney	0.00	0.01	0.01	0.00	0.00	0.03	0.00	0.02	...	0.01	

Here we can see that the cities are well separated from each other; No city is alike in their venue categories which can mean that Sydney does not have a brisbanesque feeling in venue categories. However, this does not give a metric for dissimilarity which I aim for therefore it is needed to look into their centroid distances as mentioned in the methodology.

Here is the Euclidean distances between centroids (Cities):

```
Brisbane to Melbourne
 [[0.         0.17944358]
 [0.17944358 0.        ]]

Brisbane to Sydney
 [[0.         0.12727922]
 [0.12727922 0.        ]]

Melbourne to Sydney
 [[0.         0.16673332]
 [0.16673332 0.        ]]
```

We can see that Brisbane and Melbourne are both close to similarity to Sydney. We can also see that Brisbane and Melbourne are most dissimilar among the group while Sydney and Melbourne are dissimilar; Both cities Brisbane and Sydney are dissimilar to Melbourne.

5 DISCUSSION

As we have seen from the results, Sydney and Brisbane are both dissimilar to Melbourne among all the cities while the most dissimilarity is Brisbane and Melbourne. With this knowledge, popular venues in Melbourne are recommended to be introduced to Brisbane and Sydney to create a more diverse feel of the city.

Top 10 Melbourne Popular Venues to open on Sydney and Brisbane:

```
----Melbourne----
                 venue  freq
0          Cocktail Bar  0.09
1           Coffee Shop  0.07
2                   Bar  0.07
3      Asian Restaurant  0.04
4                  Café  0.04
5          Dessert Shop  0.04
6     Korean Restaurant  0.03
7                 Plaza  0.03
8              Wine Bar  0.02
9             Bookstore  0.02
```

Top 10 Brisbane Popular Venues to open on Melbourne:

```
----Brisbane----
                  venue  freq
0                  Café  0.11
1                 Hotel  0.07
2                  Park  0.06
3                   Bar  0.06
4              Beer Bar  0.05
5           Coffee Shop  0.03
6   Japanese Restaurant  0.03
7                Bakery  0.02
8          Burger Joint  0.02
9      Pedestrian Plaza  0.02
```

6 CONCLUSION

It is suggested that to open venues based on Melbourne's popular venues while for Melbourne would be to open Brisbane's popular venues. Although the conventional focus of venues is its success and popularity in an area, this suggestion goes against the norm for a reason: To enhance a sense of wonder of finding a new place. It is like opening more café in a city since it is popular. It does not introduce originality and creates problem of over saturation.