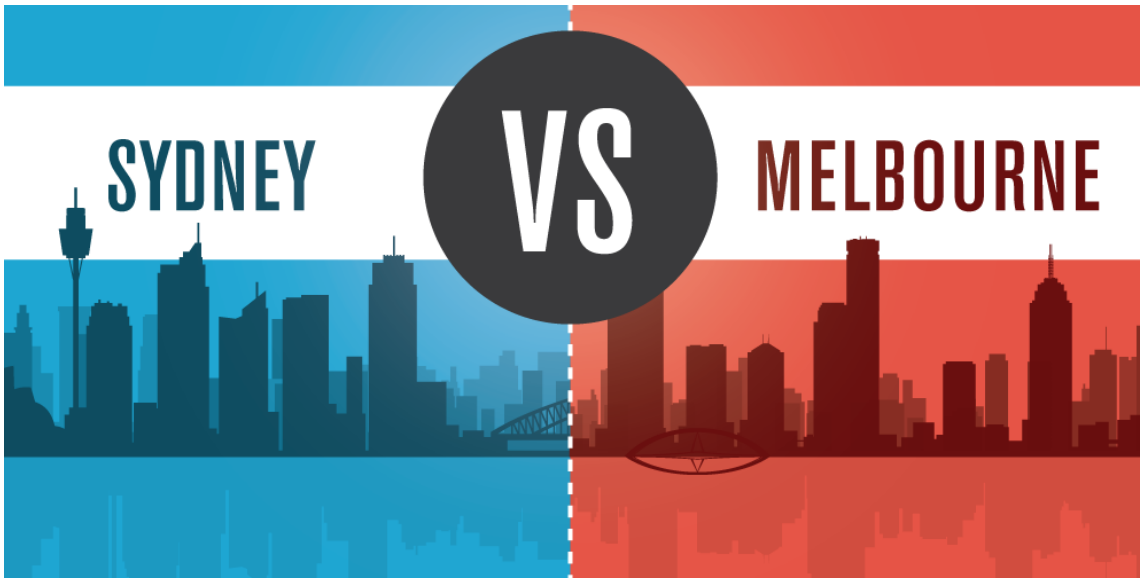

Settling the Great Australian Rivalry – Is Sydney Better than Melbourne?

Introduction

Since the mid 1800's Australia's two major two cities, Sydney and Melbourne, have battled it out over claims to be the best. Early protectionist policies naturally led Melbourne towards its future as a fashion and culture centre, while more liberal policies made Sydney the natural centre for finance and media.

As Australia fought for Federation, it was delayed as each city argued its case for being the capital of the newly formed country. Eventually a compromise was required, that would see a new capital city constructed between Melbourne and Sydney. But even then, there was no simple resolution: the site of the new capital, Canberra, was to be situated geographically closer to Sydney than Melbourne and thus Melbourne demanded to be home to the first federal parliament, while the new capital was built.

Even during World War I, Sydney complained that too much of the war effort was coming from Melbourne and the complaints and rivalry have not stopped since.

To add further division to this rivalry between two cities in a small country, each city adopted their own particular code of football, providing another area in which the cities could claim superiority, but in which these claims could not be tested.

Over the years the rivalry has been distilled to two claims: Melbourne claims to be the world's most liveable city, while Sydney contends that it's the world's best city. And while the rationale for holding these positions are justified in the historical strengths of the cities, do these claims still hold up?

Melbourne claims the most liveable title due to its culture and sport, while Sydney claims the best title due to its natural beauty and being an international city.

To settle the rivalry between the citizens of Australia's two largest cities, the aim of this analysis is to attempt to quantify the features and strengths of each city – does Melbourne have more culture than Sydney; is Sydney more international than Melbourne?

With some cold hard facts its hoped that finally it can be stated categorically whether Sydney is indeed better than Melbourne.

Data

To better quantify the features and strengths of Sydney and Melbourne, information about the respective cities was needed. Postcode data was available from Matthew Proctor and this data was used as a starting point of the analysis as it provided information on all postcodes and suburbs throughout Australia.

However, to get data representative of the greater cities of Sydney and Melbourne, the postcodes were limited to delivery areas (i.e. physical locations) for metro areas in NSW and Victoria – the states of which Sydney and Melbourne are capitals, respectively. This still provided a wide view of each city and so it was decided to take the four most central local government areas (LGAs) of each city, to be representative of the cities themselves.

In order to extract the information about each postcode using the Foursquare API, meant that a reasonable radius needed to be provided for each postcode. As the postcode sizes varied between cities and LGAs, some basic analysis was undertaken to estimate a suitable radius for each postcode in a given LGA. This resulted in a radius from 1km, for the central LGAs of Sydney and Melbourne, up to 3km, for postcodes in Melbourne – West LGA. Mapping of the radii showed that these simple estimates were suitable.

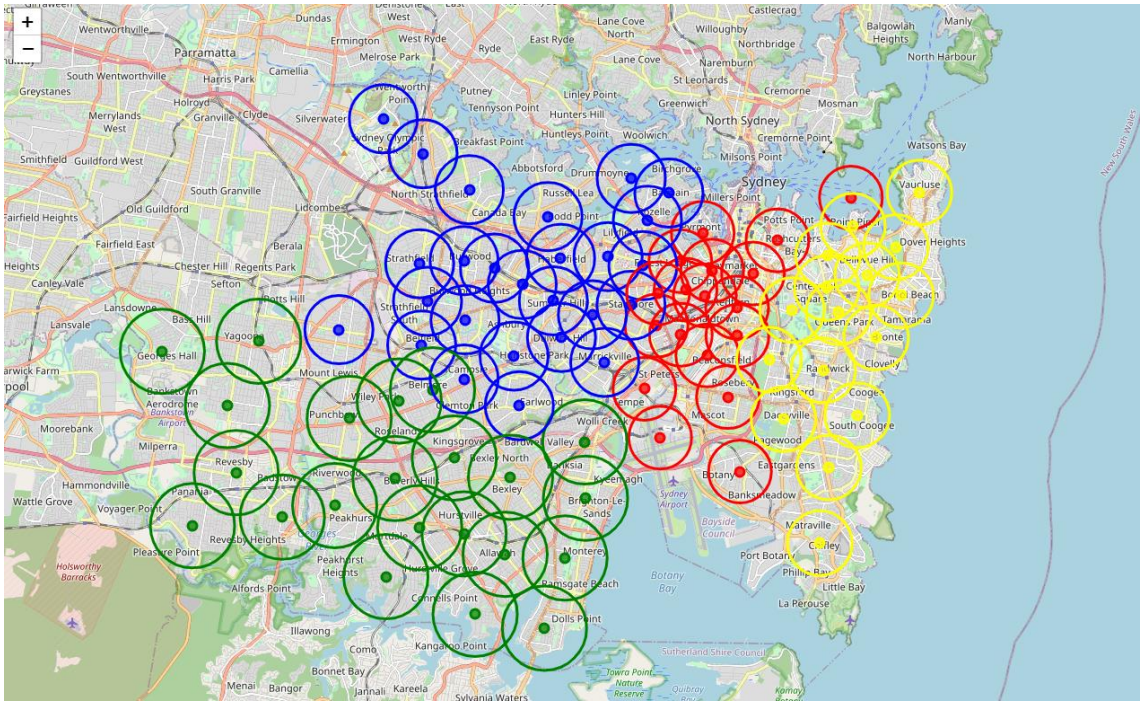

Figure 1: Map showing estimated radii for Sydney postcodes

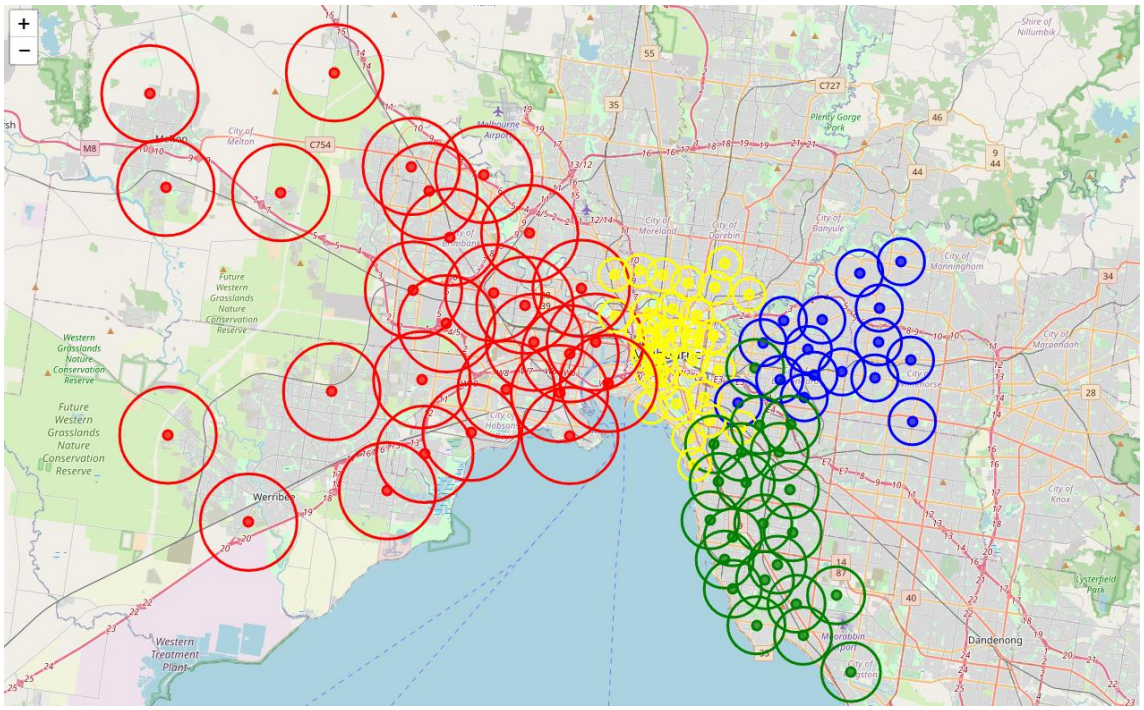

Figure 2: Map showing estimated radii for Melbourne postcodes

Using the calculated radius, along with the latitudes and longitudes provided in the postcodes data, all information for the given postcodes was extracted from Foursquare. This returned over 11,000 locations from the 182 postcodes. These locations were categorised by Foursquare into one of 403 location categories – such as restaurant, park, gym, etc.

These categories were too fine grained for the analysis, which looks to understand the cities from sporting, culture, nature and international diversity perspectives. To better improve the analysis, the classifications were further grouped into one of 17 groups, including international cuisine, sport, nature and cultural. These groups were manually developed by reviewing the provided categories and were appended to the Foursquare information.

It is this data that will be used to analyse and understand the features and strengths of Sydney and Melbourne, through their individual postcodes.

Methodology

Reviewing the extracted and grouped Foursquare data, it was noted that there were missing values in the Group column. Further investigation revealed that the missing values were related to the Category of Café and were due to an encoding issue due to the accent on the e. Given the limited nature of the issue, the Group was manually updated with the Category name as this was the expected Group name.

There was now a representation for each postcode based on only 17 dimensions, where previously there had been 403 dimensions. 17 dimensions was however still too many to be able to process without further analysis. To try to bring structure to the features of the 182 postcodes, it was decided to look for distinct segments of postcodes that would further reduce the dimensions and make interpretation easier.

To facilitate the clustering of the postcodes, one hot encoding was used on the Groups to produce 17 new variables. These in turn were averaged across each postcode to provide each postcode with a vector of the percent of locations within it that fell within each Group.

This data was then explored for the two main city areas to show the “penetration” or percent of all locations that fit into the given Group. On first glance, this showed Melbourne city centre to be more retail and cuisine while Sydney city centre was more functional as a transport hub. But clearly these were only one of many postcodes that make up the respective cities, so further analysis was required. All 182 postcodes were profiled to show the top five Groups, by order of penetration.

While clustering was to be used to identify the distinct postcodes in terms of features, the number of clusters was not known in advance and the unsupervised technique to be used, K Means, requires that the number of clusters is provided.

To get a better view on potential clusters, the postcodes were clustered using from one to ten clusters and the Sum of Squared Error (SSE) calculated. This allowed the reduction in SSE as well as the distribution of cluster sizes to be seen for each possible number of clusters. From the SSE chart, while not entirely clear, it can be seen that the SSE begins to reduce less once the number of clusters goes above four.

Picking four as the number of clusters, it can be seen in the distribution that four reasonable size clusters are obtained, which is important as at times clusters can pick up outliers only, leaving some clusters with only a handful of cases and often limiting their usefulness in understanding the different segments.

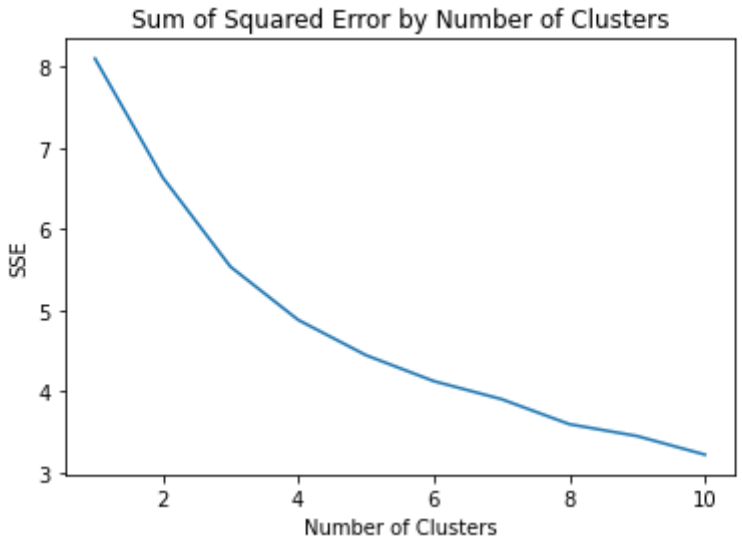

Figure 3: Change in SSE for changing cluster numbers

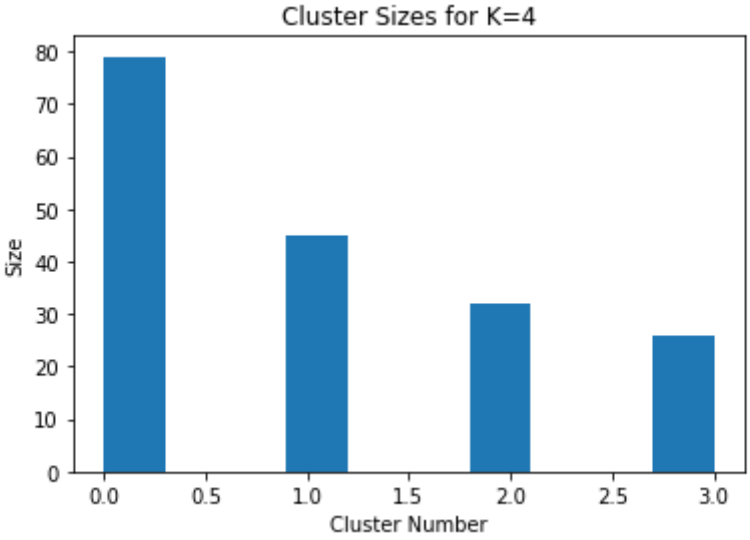

Figure 4: Distribution of cluster sizes for four clusters

Once four was identified as the number of clusters, the identified segments were included with the 17 Group percentages for each postcode. This allowed an "average" segment to be calculated by taking the average for each Group in the Segment. The average segments, with their profiles, can be seen below and these were used to interpret the segments and assign a label to each.

- Segment 0: "Retail and Entertainment" → high levels of retail and entertainment
- Segment 1: "International Cuisine" → high levels of international cuisine
- Segment 2: "High Volume Retail" → high level of retails and fast food
- Segment 3: "Nature and Sport" → high level of nature and sport

Segment	1st Most Common Category	2nd Most Common Category	3rd Most Common Category	4th Most Common Category	5th Most Common Category
0	Retail	International Cuisine	Café	Entertainment	Sport
1	International Cuisine	Retail	Café	Sport	Fast Food
2	Retail	International Cuisine	Fast Food	Sport	Café
3	Café	Nature	Sport	Retail	International Cuisine

Segment	Accomodation	Café	Commercial	Cultural	Education	Entertainment	Fast Food	Finance	Government	International Cuisine	Medical	Nature	Religious	Restaurant	Retail	Sport	Transport
0	0.010485	0.203800	0.006801	0.018478	0.000380	0.108564	0.043618	0.000452	0.003906	0.205960	0.001794	0.047469	0.000215	0.047949	0.220082	0.056318	0.023731
1	0.007900	0.124047	0.003262	0.006447	0.001379	0.051326	0.060740	0.000000	0.001059	0.330051	0.001045	0.054514	0.000000	0.028132	0.225747	0.073750	0.030602
2	0.007843	0.073309	0.016923	0.004849	0.000000	0.042028	0.132696	0.000000	0.000000	0.137110	0.000460	0.043433	0.000332	0.023969	0.361499	0.098585	0.056963
3	0.010125	0.177123	0.000000	0.023001	0.000000	0.049112	0.051093	0.000000	0.002410	0.108899	0.001642	0.145233	0.000574	0.059541	0.133966	0.139779	0.097501

Figure 5: Profile of final four segments along with the penetration for each group

Having identified each postcode as one of the four segments, which have been labelled, the analysis is ready for interpretation.

Results

With each postcode now classified as Retail and Entertainment, International Cuisine, High Volume Retail or Nature and Sport, a radar plot was produced that displays the percent of postcodes in the city that fell into each category. The radar plot can be seen below.

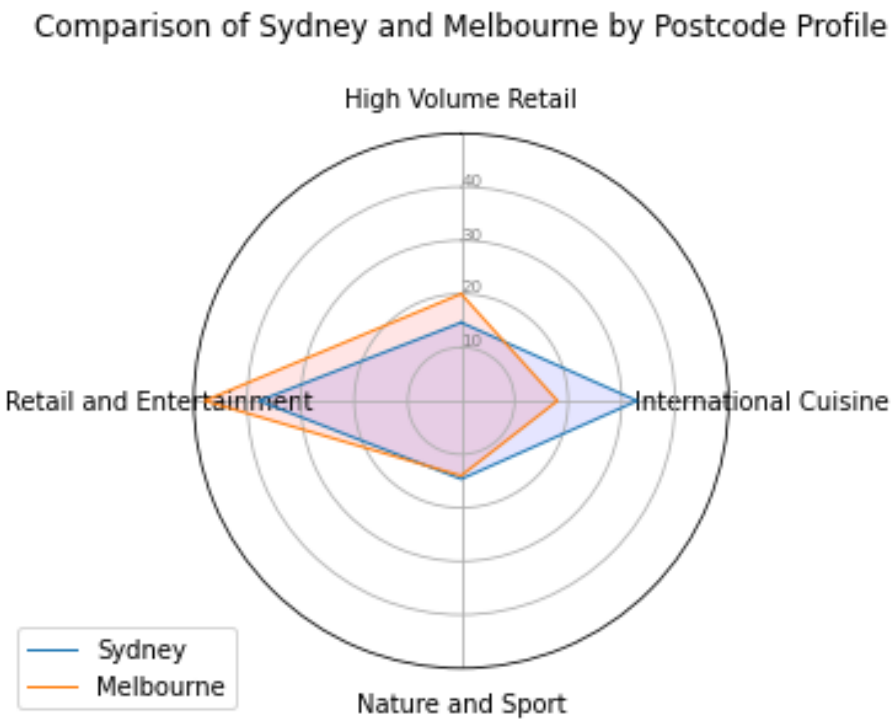

Figure 6: Radar plot comparing Sydney and Melbourne

From the plot it can be seen that Melbourne significantly exceeds Sydney in postcodes that are High Volume Retail and Retail and Entertainment, which aligns with Melbourne being the perceived capital of fashion in Australia. Sydney has double the number of postcodes dedicated to International Cuisine and this too supports a general perception of Sydney as an international city.

More controversially, there is little separating Sydney and Melbourne in Nature and Sport, with Sydney only just ahead. The fact that sport and nature were often found in the same postcodes is not surprising given the association between sport and the outdoors, but it does little to answer the questions about Melbourne's claim to be the sporting capital and Sydney's renown as a city gifted with natural beauty.

To better understand what might be drivers of these strengths and features, the postcodes were again plotted and coloured by the Segment into which they fell.

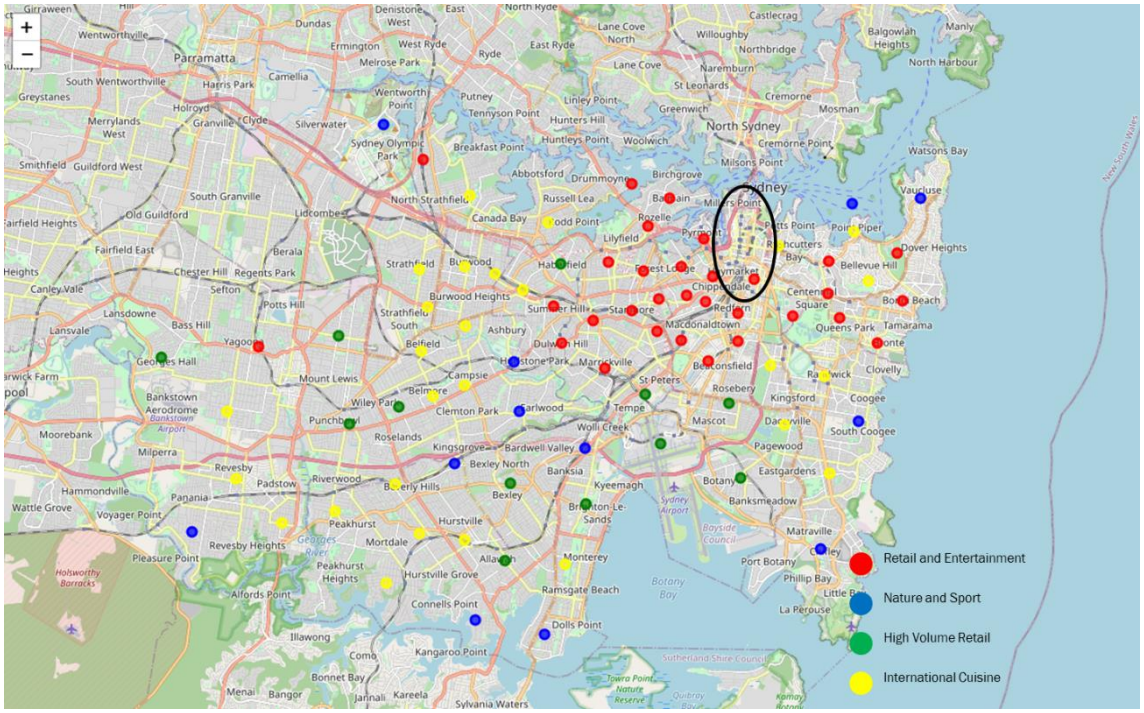

Figure 7: Sydney postcodes as classified by clustering locations found within

In Sydney, Nature and Sport postcodes tend to be located by the water, either ocean or river. Retail and Entertainment is clustered around the CBD (marked in black), while High Volume Retail are dotted further out in more suburban areas. International Cuisine can be found both out west and in the east, with very little in the south – aligning with the more multicultural areas of Sydney.

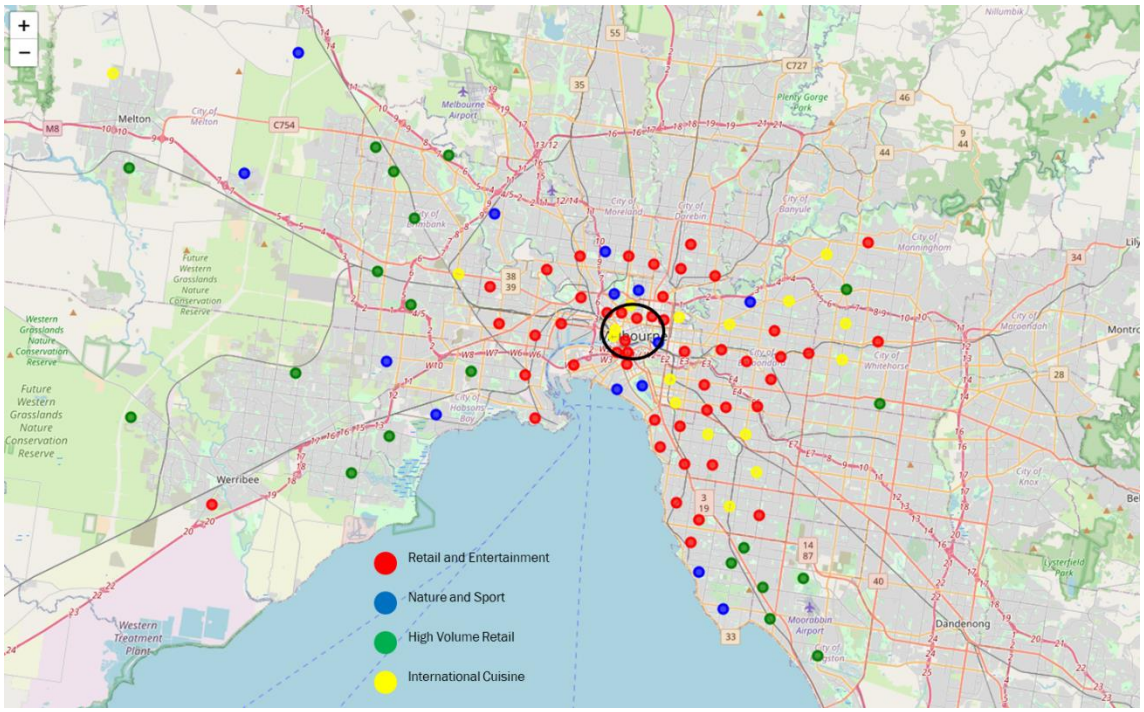

Figure 8: Melbourne postcodes as classified by clustering locations found within

In Melbourne, Nature and Sport has less reliance on water and are also found in more suburban areas as well as parklands. Again, Retail and Entertainment are clustered around the CBD, with High Volume Retail being found in the more outlying suburban areas. International Cuisine is found in the south and east and is almost non-existent in the north and west, which is dominated by High Volume Retail.

Discussion

Almost 70% of suburbs are classed as Retail (Retail and Entertainment or High Volume Retail) in Melbourne, while Sydney is more diverse with only 55% being classified Retail. While this supports Melbourne's claim as the fashion capital of Australia, it is not backed up by any significant level of cultural locations. This may be due to the biased use of Foursquare, which is mainly used in relation to shopping and eating and less so for cultural attractions. A next step in the analysis may be to provide more emphasis when identifying and classifying Cultural locations to better ensure they contribute to the identified segments.

Sydney has double the rate of International Cuisine to that seen in Melbourne and these postcodes are located across a more geographically diverse range of postcodes in Sydney. This very much supports Sydney's claim to be an International City – drawing influences from around the world.

While Sydney and Melbourne are about equal in terms of Nature and Sport, it can be seen that although sport and nature are correlated in the classifications, Sydney is very much dependent on the water for both aspects. This may be indicative of a sporting culture that is driven more by favourable nature and weather than an outright love of sport for sport's sake. Melbourne on the other hand has a more even geographic spread of Nature and Sport and this may be indicative of a more rounded appreciation of sport, less tied to the natural surrounds.

Overall, Sydney demonstrably proves its international status and more than holds its own in the areas of sports and nature with Melbourne. Melbourne on the other hand, while holding its own in nature and sport, fails to prove any true strength in Cultural aspects, a key point in their claim to be Australia's best city. Given this weakness in Melbourne's argument, along with Sydney meeting all its own (self-appointed) criteria, the conclusion is unavoidable: Sydney is indeed better than Melbourne.

Conclusion

In this analysis it has been shown that Sydney is indeed better than Melbourne, by virtue of their claimed strengths being supported by observed strengths, using crowd sourced information from Foursquare.

Of course, as robust as the analysis may be, the results are still heavily dependent on human bias in two distinct ways.

The first is in the crowd sourcing of the information in Foursquare. The demographics and the usage patterns tend to focus on retail and eating, with other location types lower in their level of capture. This can of course lead to biased results, where the weight of comparison is highest in these areas. Likewise, the differences in demographics and usages between cities has the potential to bias the results. Sydneysiders, who may be more interested in nature, may capture these locations in greater numbers than Melburnians, due to personal preferences rather than the existence or non-existence of the locations.

The second potential source of bias is in the classification of the locations and in the execution of the clustering. Both are open to interpretation and so a given point of view may alter the trajectory of the analysis in either a minor or major way. With the conclusions available after the first iteration there are steps that could be taken in a further iteration that could help focus the analysis on answering the question more completely. Principle among these would be the individual focus on culture, sport and nature – data permitting; and perhaps entertaining more clusters to provide a finer definition between postcodes.

In the end, one can only quote George Box in that "*all models are wrong, but some are useful*" and hope that this model has at least been useful in answering the question "*Is Sydney better than Melbourne?*" in the affirmative.

References

Campbell, M., Schultz, A., Pobjie, B. and Pobjie, B., 2019. *A History Of The Sydney V Melbourne Rivalry*. [online] Crikey. Available at: <https://www.crikey.com.au/2019/07/25/melbourne-sydney-rivalry-history/> [Accessed 15 November 2020].

Enterprise.foursquare.com. 2020. *Foursquare Location Intelligence For Enterprise*. [online] Available at: <https://enterprise.foursquare.com/products/places> [Accessed 15 November 2020].

Holtz, Y., 2020. *#391 Radar Chart With Several Individuals*. [online] The Python Graph Gallery. Available at: <https://python-graph-gallery.com/391-radar-chart-with-several-individuals/> [Accessed 22 November 2020].

Matthew Proctor, A., 2020. *Matthew Proctor*. [online] Matthew Proctor's Blog. Available at: <https://www.matthewproctor.com/> [Accessed 15 November 2020].