Discovering Possible Business Opportunities Around Melbourne Train Stations

Antoine Maalouf

1. Introduction

Train stations can create business opportunities by leveraging the increased foot and road traffic that is created around train stations. By Analyzing different attributes of each train stations within the Melbourne metropolitan area, it's possible to locate underserved stations and evaluate the possibility of developing new businesses in its vicinity.

Melbourne is a fast-growing city with high levels of adoption of public transport among its residents. It's estimated that a total of 565 million trips are taken every year using at least one form of public transportation with 240.9 million that includes at least one leg by train [1].

Rents around popular stations such as Melbourne Central and Southern Cross stations have increased exponentially while competition have increased. It's possible to discover new underserved stations that can give a competitive advantage to property developers and certain businesses by filling the needs in underserved areas without paying the higher rents charged close to the CBD.

2. Data Sources

Several Data sources will be used to analyze the current situation of train stations along the lines serving the Melbourne CBD along with established businesses around them

- Train Station Geolocations
  - Provider: Department of Transport, Victoria, Australia
  - Format: Geojson file
  - License: Creative Commons Attribution (CC BY 4.0)
  - Usage: This dataset is used to determine the geographical coordinates of current train stations and use this data to discover nearby businesses and travel times.
- Train Station Patronage
  - Provider: Public Transport Victoria
  - Format: Excel sheet
  - License: Creative Commons Attribution (CC BY 4.0)
  - Usage: Extract the number of train lines serving each train station and the number passengers boarding.
- Existing establishments and their current ratings
  - Provider: Foursquare
  - Format: API
  - License: Proprietary, Licensed according to the Foursquare Terms and Conditions
  - Usage: Determine current establishment within walking distance of the stations and their current ratings

3. Methodology

Attributes from the above data sources are combined into a single data frame indexed by train station. Each attribute was separately evaluated to determine it's effect on the final data analysis. As a result, several attributes were dropped in the final analysis. This includes the types of bicycle storage, connecting modes of transportation, and rider intentions. These data were deemed unreliable due to the subjective nature in which they were collected and may cause errors in the result.

Some stations were dropped from the analysis due to several reasons:

- Occasional stations only used during special events or as a replacement for an out of service station
- Remote stations with very infrequent service
- Temporarily or permanently Decommissioned stations
- Lack of data from Australian government sources

Australian government data sources were supplied into two formats. GeoJSON files containing geographic locations of train stations along with parking and biking capacity and as Excel files containing patronage data along with surveys detailing customer intentions and connecting modes of transportation.

This data is compiled together into a data frame.

Table 1 Select Columns

Column Name	Description
STATION_NAME	Station name
LATITUDE	Geographic location of each station
LONGITUDE	
NUMBER_OF_ROUTES	Number of regular routes serving this station
BIKE_CAPACITY	Number of bikes that stored in or around the station
PARKING_CAPACITY	Amount of free parking facilities for rides
ANNUAL_PATRONAGE	Total number of people embarking at each particular station
NUMBER_OF_ROUTES	Number of regular routes serving this station
venues	Number of venues within walking distance of the station.

The Foursquare API is used to obtain information about existing businesses within the studied categories. The scope of the search was limited to a set number of categories within a radius of 150 m. The number of venues is then added as an attribute for each stations.

The categories were chosen based on the type of businesses that benefit from a large influx of commuters and pedestrian traffic within and around the train stations and reflect the types of businesses that are being considered in this report.

Stations with many venues around it will be considered as saturated and may not be an ideal location due to the availability of vacant properties and the fierce competition from existing establishment.

Table 2: Venue Categories

Establishment Type	Category
food truck	Food
café	Food
coffee shop	Food
fast food	Food
burger joint	Food
donut shop	Food
fish and chips	Food
food court	Food
fried chicken joint	Food
hot dog	Food
movies	Entertainment
museum	Education
stadium	Entertainment
university	Education

The compiled data is clustered to perform the needed analysis. Two clustering techniques are first considered with the first being K-means while the second being DBSCAN. The first is K-means clustering where several values for k have been evaluated.
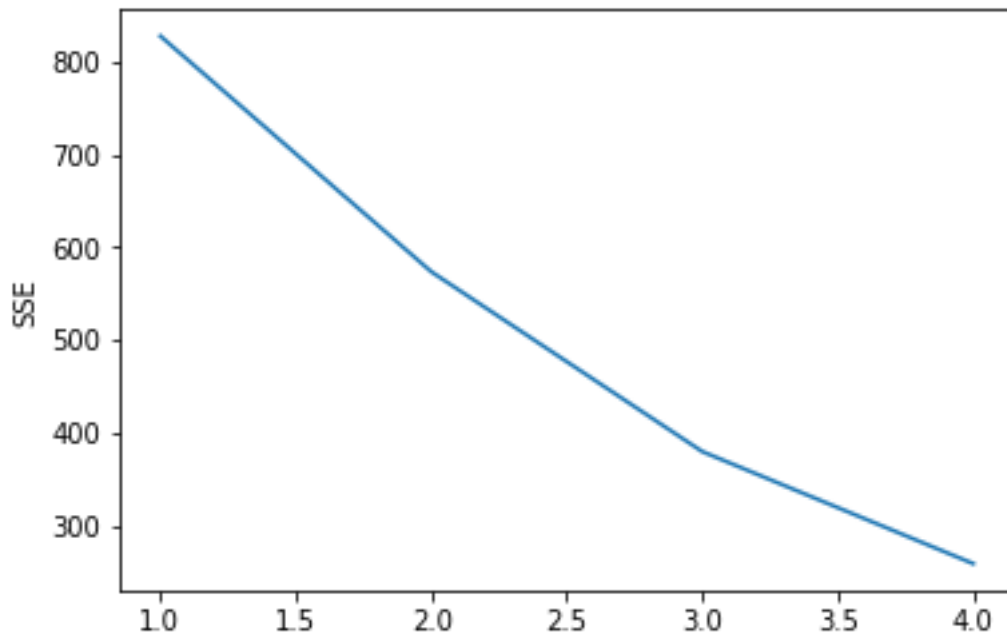

Figure 1K means clustering

From the above graph, it was determined that k=3 is the best value that renders smaller error. Stations were then mapped by cluster to get an understanding of the geographic distribution of each cluster
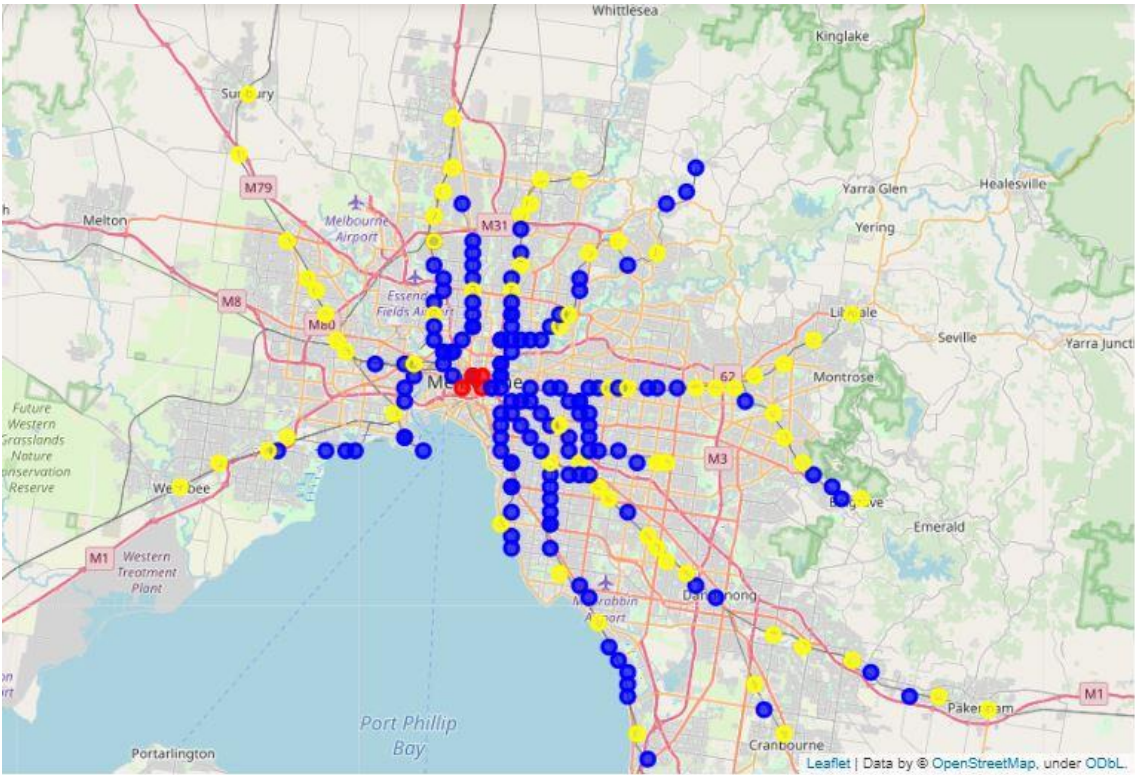
After studying the above data, it was determined that K-means clustering doesn't offer any significant information about the current status of each train station. The stations were distributed between very high volume in red, urban in blue, and suburban yellow.

DBSCAN was then used to cluster the data. The clusters were separated into three categories separated as follows.

DB_LABEL	NUMBER_OF_ROUTES	BIKE_CAPACITY	PARKING_CAPACITY	ANNUAL_PATRONAGE	venues
0	1.769231	13.241026	206.133333	1.081385	11.435897
1	1.000000	0.000000	0.000000	0.284286	0.000000
2	1.400000	0.000000	0.000000	0.426000	1.000000

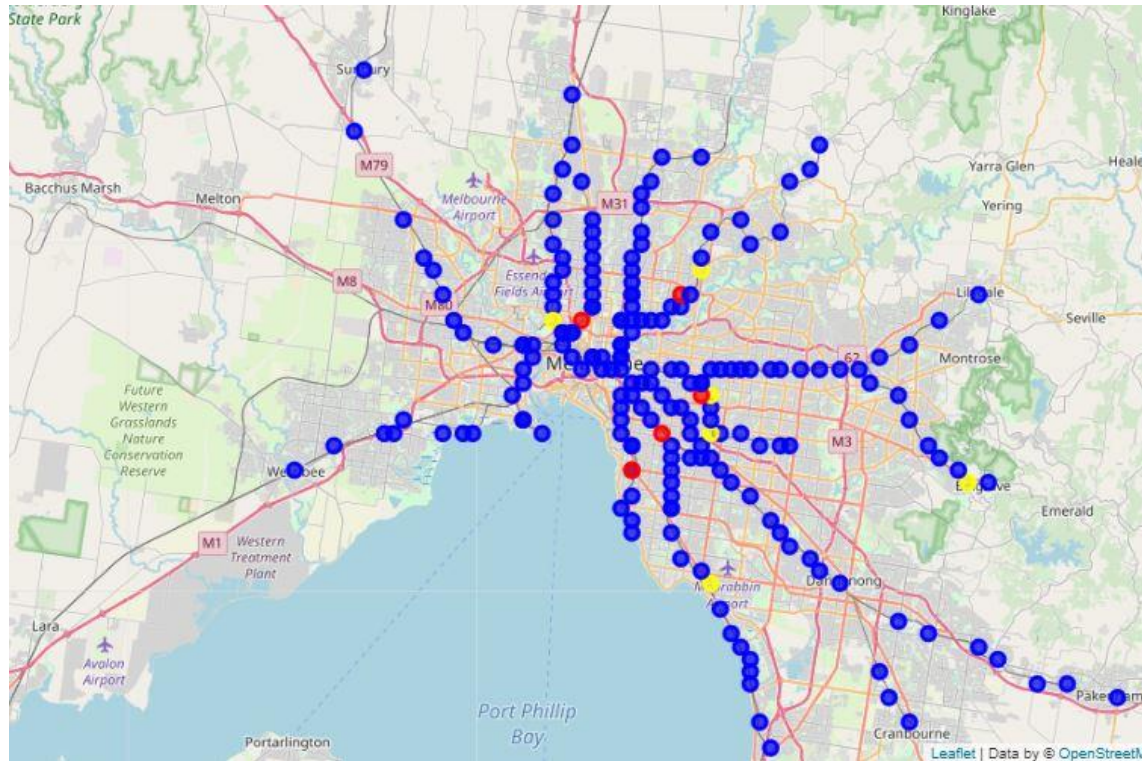Clusters are mapped to better understand the geographical distribution of the clusters.

Figure 3 DBSCAN Results

4. Results

Clustering have showed that there are generally three clusters of stations. Large popular stations served by multiple trainline and is already served by multiple train stations. This includes Southern Cross, Flinders, Footscray, and Melbourne Central and other stations within the city look and inside inner suburbs. Those station present a great business opportunity but where real estate values are much higher than average and competition is very strong due to many well-established venues within walking distance.

The second cluster of stations are those served by a single train line and have low rider patronage and very little businesses around it. These stations don't offer much of an opportunity due to the lack in the number and frequency of foot traffic around the station. The benefit of being next to a train station will be negligible. Examples of these stations include Aircraft and Altona.

The third cluster of stations are stations that have a medium number of riders, are served with multiple train lines, and a relatively low number of existing venues. These stations offer the best opportunity due to them having a good number of patronage while having a low number of existing venues around them. Such stations can be individually analyzed to pinpoint the specific opportunities.

5. Discussion

The analysis has showed many business opportunities around the less known stations where real estate prices are reasonable while station patronage is high enough to offer a net benefit.

These stations have been identified in the middle suburbs that are experiencing high growth and have yet to see businesses opening around them. They offer a unique and time sensitive opportunity to potential business owners where they can benefit from the first mover advantage and capture a large but mostly unknown source of traffic. Cluster number 2, that is represented in red, have included these types of stations.

Table 4 Cluster number 2

STATION_CODE	STATION_NAME	NUMBER_OF_ROUTES	BIKE_CAPACITY	ANNUAL_PATRONAGE	PRE_PEAK	AM_PEAK	INTERPEAK	PM_PEAK
WSN	willison	1.0	0.0	0.09	26.93	211.22	54.61	31.79
EGT	eaglemont	1.0	0.0	0.18	43.84	389.63	133.89	56.20
MAL	malvern	3.0	0.0	1.08	149.14	1373.26	849.80	1009.23
GVE	gardenvale	1.0	0.0	0.47	62.42	857.01	342.84	335.45
RPK	royal park	1.0	0.0	0.31	23.06	165.57	267.97	380.64

After analyzing stations in cluster 2, Malvern train station was shown to present a good combination of patronage and low number of nearby venues. Served by three separate train routes and with a larger than average peek and off peek foot traffic, Malvern station represent the type of business opportunity that was sought after in this report.
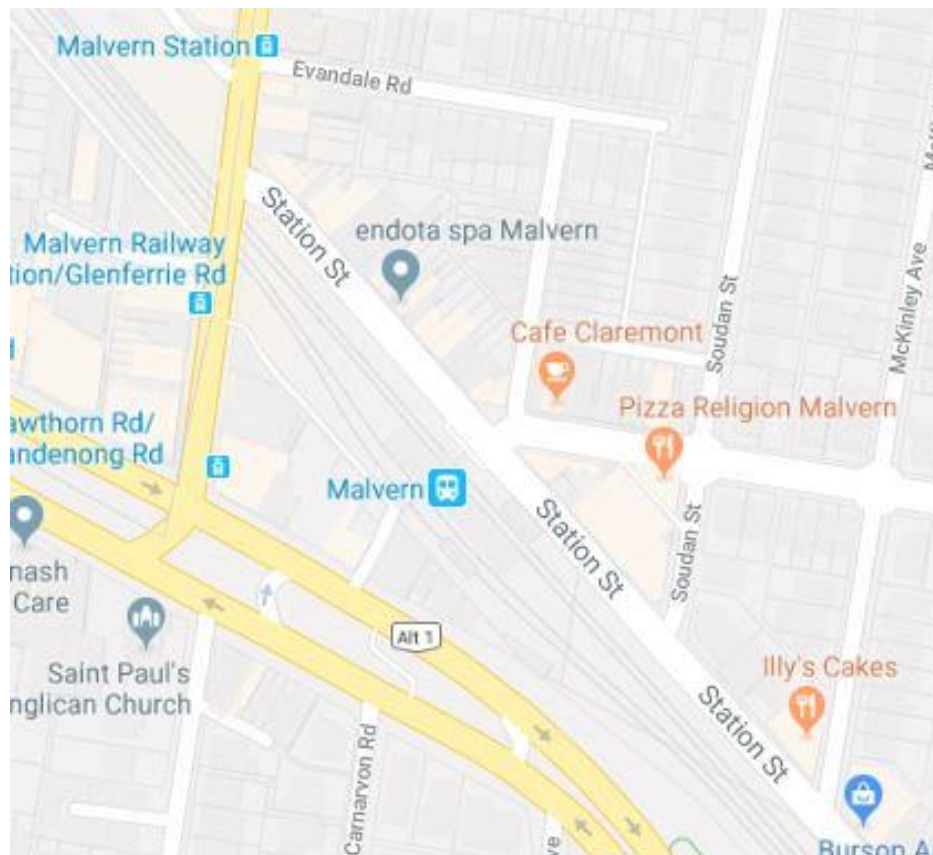

Figure 4 Malvern Station

By having a health combination of residential and commercial establishments around the station and a low number of venues in the required categories, it's clear that such station will present an opportunity for a business owner interested in setting up convenience store, coffee shop, or a food truck. Rents in this area are lower than the city's average but it's recommended for a business to purchase the property to benefit from the increasing real estate prices that are inevitable once the popularity of the area have increase.

6. Conclusion

After analyzing train stations for their patronage levels, train service levels, and number of existing venues, it was discovered that several stations present a unique opportunity for potential business owners to take advantage of by capturing the high levels of foot traffic caused by the presence of the station. The analysis has shown the case of Malvern station to have the exact type of attributes that would constitute a good fit for setting up a fast food joint or a convenience store. Real estate costs around this station are relatively low. These underserved stations present the opportunity to capture the growth of the area prior to it becoming at par with already large areas in the city.