

MyCiTi Bus Services in Cape Town

Capstone Project for Applied Data Science by IBM/Coursera

Background & Problem

Cape Town, one of the most beautiful cities in the world, also happens to be one of the most congested when it comes to road traffic. [1]

According to the INRIX 2018 Global Traffic Scorecard, people in Cape Town lose 162 hours per year to congestion. [2]

In recent years this issue has become even more acute due to a rapid population increase and a lack of investment into rail infrastructure.

Given that 80% of daily congestion in Cape Town consists of private vehicles [1], one solution would be to improve the shared public transport, which is currently represented by:

- Metrorail, a rail service that nowadays is considered by many too dilapidated and unsafe
- Minibus taxis, informally operated network of shared minibuses, with no set map or timetable [3]
- Golden Arrow, a privately-owned point-to-point bus service operating at morning and evening peak hours; the City of Cape Town is not able to control the extent and quality of this service
- MyCiTi, a bus rapid transit system provided by the City of Cape Town, operating on a regular time table throughout the day using dedicated road infrastructure and technology

While the need of investment into rail infrastructure is widely recognised, it appears that the City of Cape Town sees MyCiTi as the most viable response to public transportation needs. The MyCiTi network was launched in May 2010 and is still growing.

In this analysis we would like to focus on the MyCiTi network, first mapping and visualising its current extent and then suggesting ways for future growth, taking into account the following two facts [4]:

- 95% of public transport users are in the low to low-middle income brackets
- 95% of people in high-income communities own a car

So the real question that we are aiming to answer is: how should we expand the MyCiTi network so that we encourage private car users to switch over to public buses, and what areas currently served by Minibus taxis and Golden Arrow buses would benefit from MyCiTi?

We are confident that the Transport and Urban Development Authority of Cape Town has a team of data scientists, its own private data sources and a strategy for development, but we will still undertake this exercise from an outsider's perspective, using publicly available data. This analysis may be of interest to those working at the Transport and Urban Development Authority of Cape Town, or those with a general interest in public transport.

Data

From the administrative perspective the City of Cape Town is split into more than 700 communities called suburbs.

To understand the existing transportation needs and suggest how to alleviate congestion, we wanted to obtain the following 4 datasets:

- Coordinates of Cape Town's suburbs
- Locations of MyCiTi bus stops
- Number of residents per suburb
- Some indicator of suburb wealth, such as income tax paid by residents or their net income

Coordinates of Cape Town's suburbs are conveniently available from the City of Cape Town Open Data Portal [\[5\]](#) in shape of a GeoJSON file.

While initially we were planning to retrieve the coordinates of MyCiTi bus stops from a location service called Foursquare, during our analysis we realised that there is a dataset of superior quality provided by the City of Cape Town Open Data Portal [\[5\]](#), so we had to switch over to this newly discovered dataset, in order to improve the quality of our results.

The most recent Census took place in South Africa back in 2011, and we could not find more up-to-date information on the number of residents or their income. Instead we have found a 2018 dataset with the breakdown of suburb properties by value, also provided by the City of Cape Town Open Data Portal [\[5\]](#), so we used it to estimate both the number of residents per suburb and the suburb wealth.

Methodology

For our analysis we used the following approach to source and pre-process the data:

- First, we visualised the map of Cape Town's suburbs to better understand the geography we were dealing with.
- Then we obtained, cleansed and visualised the geographical locations of all MyCiTi bus stops using the Foursquare service. It turned out that the Foursquare data available was incomplete and there was a better data source available. We therefore obtained, cleansed and visualised this new dataset. The differences are explained later in this report.
- Finally, we obtained and cleansed the property valuations from the Open Data Portal of Cape Town.

Once we had a clean dataset, we visualised it to see which suburbs are more populous than others and what their existing MyCiTi coverage is.

Then we ran K-means clustering to classify Cape Town's suburbs by wealth. This allowed us to come up with suggestions on how to extend the MyCiTi services in the future.

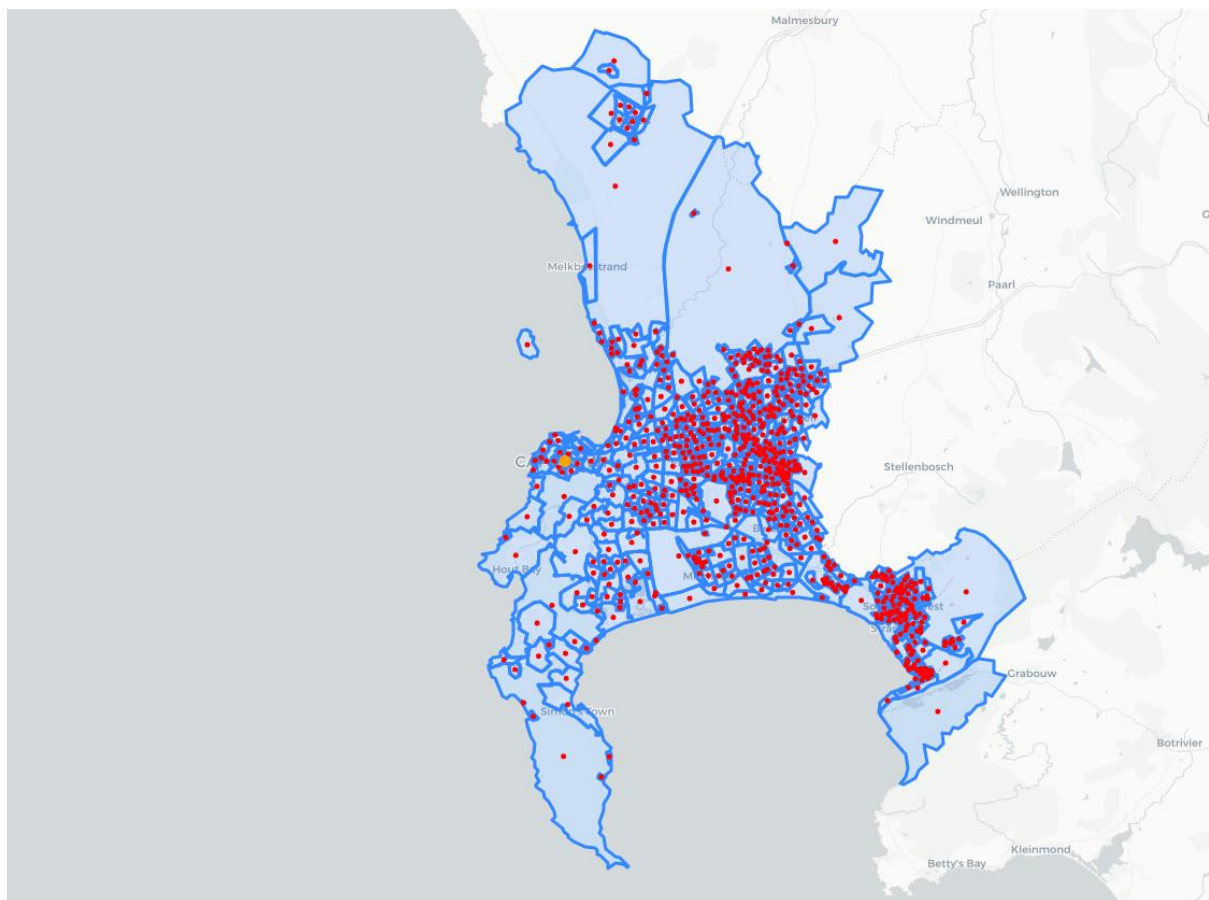
Data Acquisition and Analysis

A. Suburbs (Open Data Portal)

We loaded and parsed the GeoJSON file with Cape Town's suburbs, then calculated the centroid for each of the suburbs found.

Shown below is the resulting map, red dots representing centroids for each of the suburbs, while the slightly larger orange dot indicates the centre of Cape Town, as defined by the Nominatim service.

Map 1: The suburbs of Cape Town



B. MyCiTi Bus Stops (Foursquare)

It would have been easy to run one single Foursquare query to fetch all the bus stops, placing the centre of the search in the geographical point determined by Nominatim and supplying a sufficiently big radius to cover all of Cape Town.

Our practical experiment has shown that such an approach returns only a fraction of the data available in Foursquare, as there is a limit as to how many venues can be returned by a single query (only up to 50 venues returned).

We therefore had to employ a different approach:

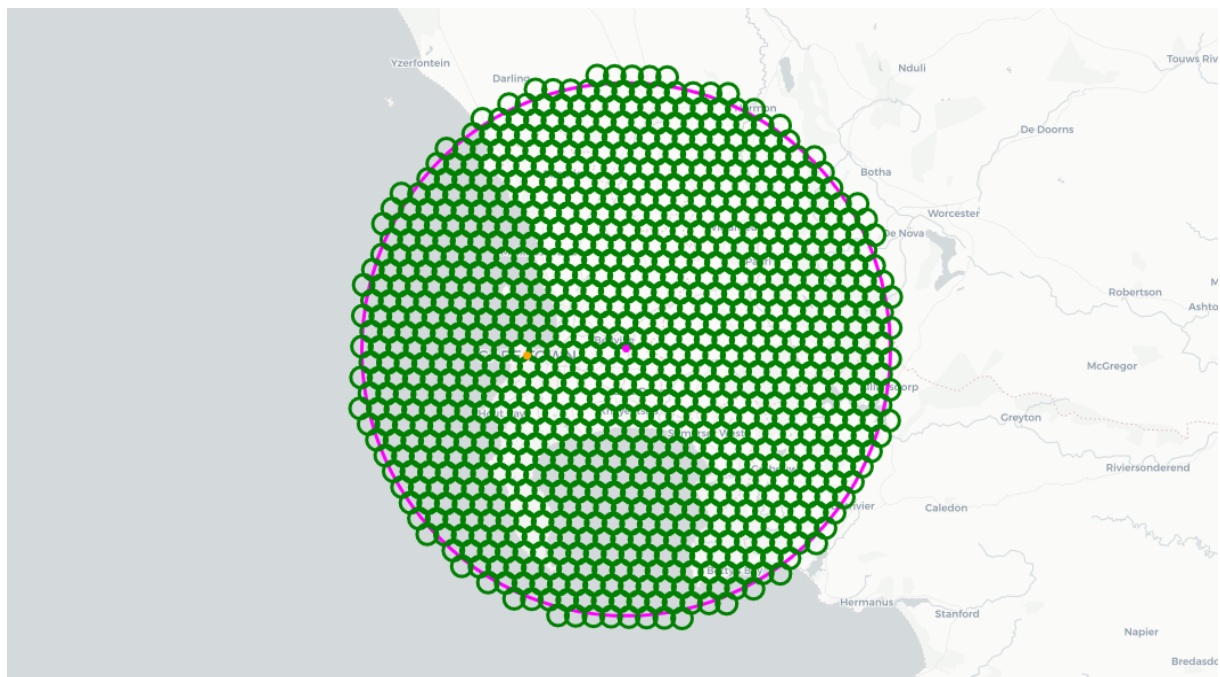
- Determine the minimum and maximum geographical coordinates of Cape Town's suburbs in order to determine the city's diameter.
- Re-position the city centre accordingly.
- Divide the resulting circle into smaller areas that can be searched by Foursquare without a data loss.

The diameter of Cape Town was determined to be 117,830 metres, and the centre of our search was re-positioned to the east of the original Nominatim city centre.

We then created a honeycomb-like pattern of circles for the Foursquare search. Initially we set the diameter of each cell to be one-tenth of the diameter of Cape Town, however later we had to reduce the size to 1/30th to ensure there is no data loss (Foursquare restricts the number of items returned by a single query).

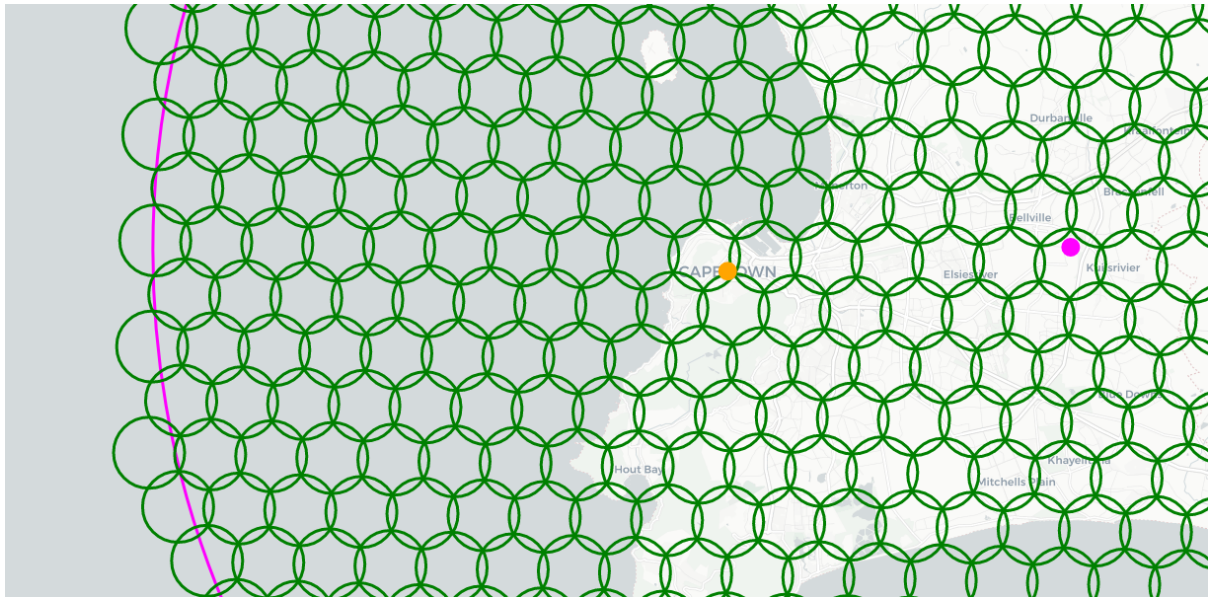
The map below shows the Foursquare search zones in green, the re-positioned city centre and the new city boundaries in magenta, and the original Nominatim centre of Cape Town in orange.

Map 2: Foursquare Search Zones



A close-up of the same map:

Map 3: Foursquare Search Zones (Close-Up)



While the distance between the search zone centres is 1/30th of the city diameter (i.e. approximately 4 kilometres), we increased the diameter of search zones to 4.5 kilometres to ensure there is some overlap so that we do not miss a single bus stop in our search.

We could have eliminated those search zones that are entirely located in the ocean, as we do not expect to find any bus stops there. But we did not.

To search for bus stops in Foursquare, we first identified which Foursquare categories should be in our scope. We found three categories containing the word 'bus':

	level	Name	id
902	1	Bus Station	4bf58dd8d48988d1fe931735
903	2	Bus Line	4bf58dd8d48988d12b951735
904	1	Bus Stop	52f2ab2ebcbc57f1066b8b4f

It was not entirely clear what the distinction between the three categories was, so we decided to pull all of them from Foursquare.

The resulting dataset contained 165 items, and it still was unclear what the distinction between the categories was. Also there seemed to be more 'bus stations' in Cape Town than 'bus stops', which did not make sense:

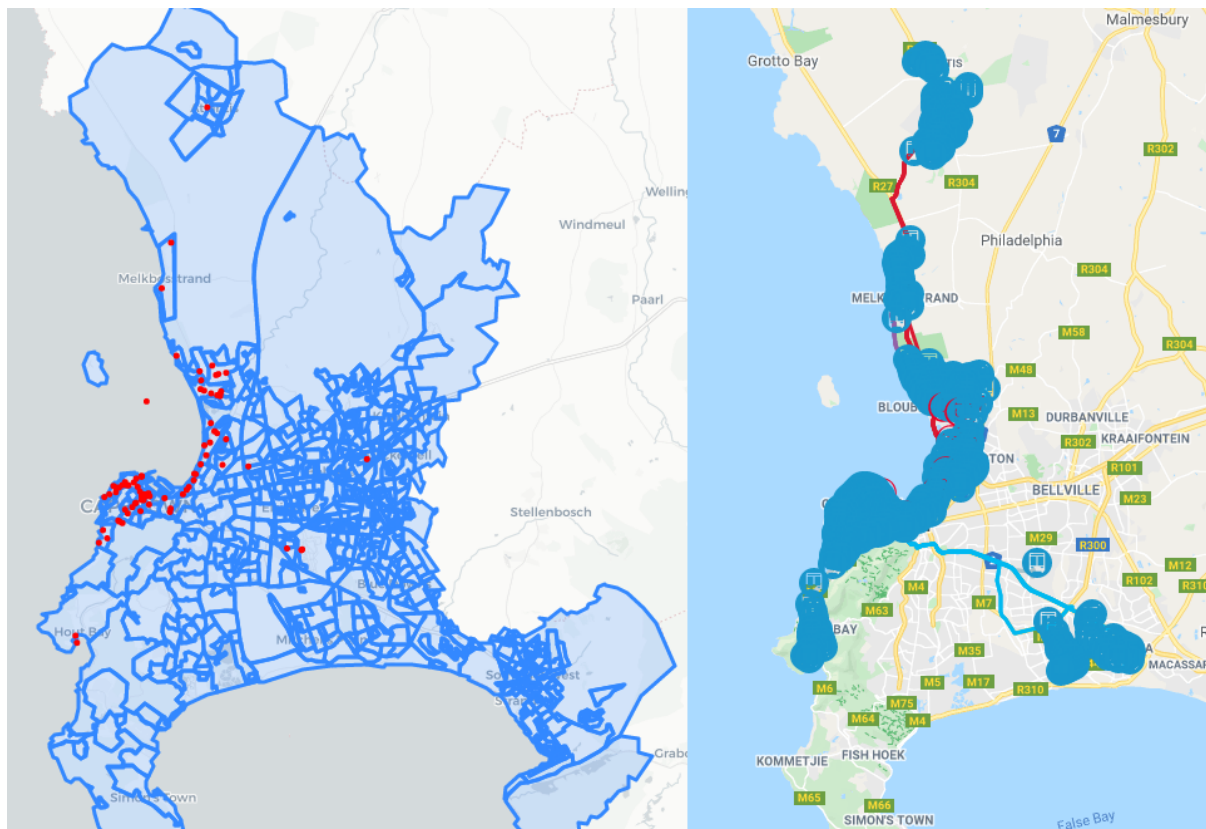
Name	lat	lon	category
Fish Hoek Taxi Rank / Bus Station	-34.136455	18.432575	Bus Station
Partybus SA	-34.156093	18.971288	Bus Station
Bushalte 'Koekamerstraat' (Lijn 61)	-34.083333	18.666667	Bus Station
Bus Stop - Somerset To CPT	-34.077321	18.846245	Bus Station
MyCiti Bus - Northsjore	-34.046962	18.348096	Bus Station

Name	lat	lon	category
Kronendal Park & Ride	-34.040592	18.359938	Bus Line
MyCiti Bus - Atlantic Skipper (108)	-34.053463	18.349653	Bus Stop
Salfred's Coach And Minibus Tours	-34.060235	18.480399	Bus Station
Mitchell's plain town center(bus stop)	-34.042322	18.622367	Bus Station
Golden Arrow Town Centre	-34.033809	18.612445	Bus Station

We also noted that the dataset contained other items that we were not interested in, such as Partybus, Park & Ride, Coach Tours and Golden Arrow bus stops, which we discarded, keeping only those items that contained both 'MY' and 'CIT' (case-insensitive search).

This cleansing left us with 83 bus stops, which are shown on the map below (left), against the actual MyCiti bus map from the official MyCiti web site [6] (right).

Map 4: Foursquare Bus Stops vs. Official Bus Map



Our observations:

- Foursquare did not return any MyCiti bus stops in Khayelitsha and Mitchells Plain (the clusters in the south-east), even though services have been operating there for more than 5 years.
- The overall density of bus stops is far from what we see on the official bus map, indicating that the Foursquare dataset is far from complete (e.g. see the Hout Bay cluster in the south-west).
- Despite our hopes, one of the Foursquare bus stops is located in the Atlantic Ocean.

This prompted us to think if there was an alternative data provider that we could use instead of Foursquare. Conveniently, we found a similar dataset on the Open Data Portal [5], however at that stage we were not confident if quality of the new dataset was any better, so had to repeat the data acquisition, cleansing and visualisation.

C. MyCiTi Bus Stops (Open Data Portal)

The initial dataset that we loaded looked as shown below:

name	lat	lon
Atlantis	-33.563850	18.491811
Civic Centre	-33.920718	18.429722
Pontiac	-34.045131	18.621764
Pontiac	-34.045131	18.621764
Civic Centre	-33.920718	18.429722

We removed all duplicates and all bus stops located close to each other and having the same name, so that the name of bus stop became the primary key. The resulting dataset contains 518 entries.

Then we enriched the dataset, associating each bus stop with the suburb where it is located:

name	lat	lon	suburb
10th Avenue	-34.041992	18.617469	MITCHELLS PLAIN CBD
6th Avenue	-34.052644	18.625384	EASTRIDGE
Abalone	-33.831848	18.546718	RICHWOOD
Acacia	-33.579636	18.502975	PROTEA PARK - WESFLEUR
Adderley	-33.920845	18.424676	FORESHORE

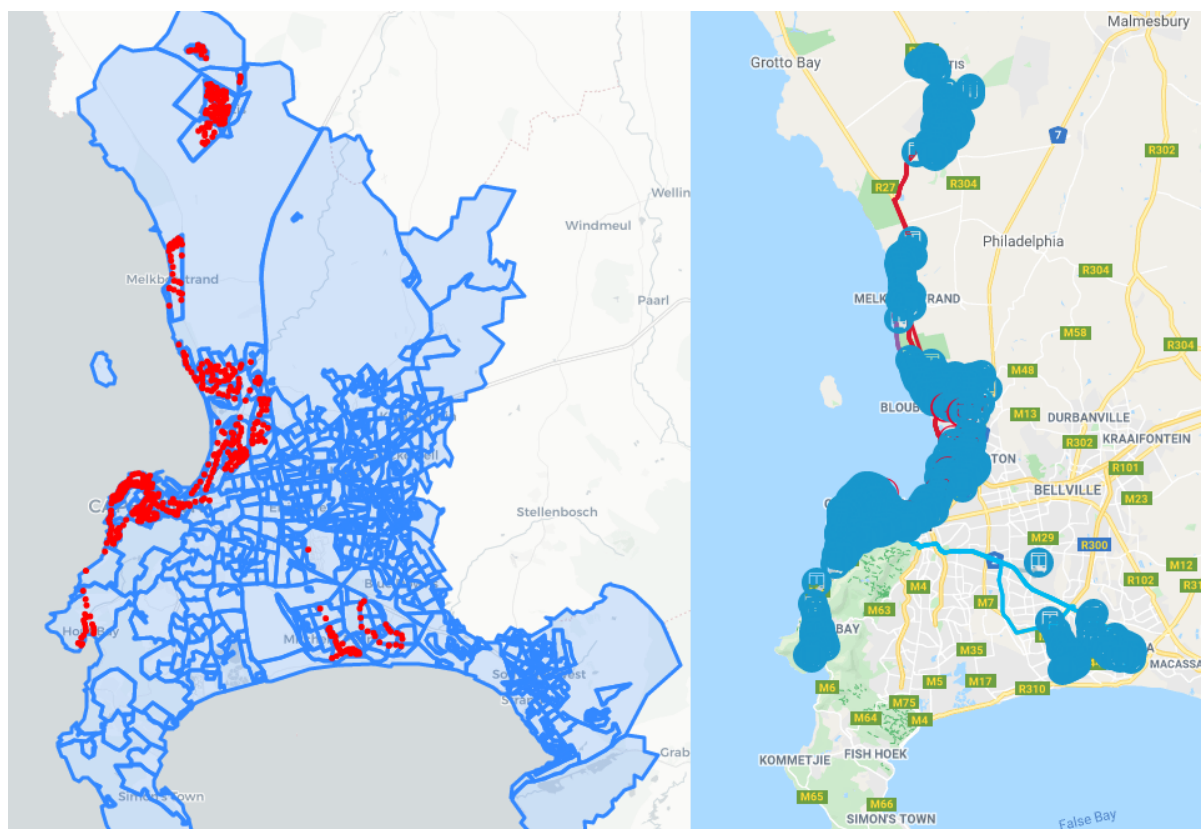
Only one bus stop, Perlemoen, could not be allocated to a suburb. A closer look at the map revealed that it probably belonged to BLAAUWBERGSTRAND:

Map 5: Perlemoen in BLAAUWBERGSTRAND



We then visualised the 518 bus stops found (left), against the actual MyCiTi bus map from the official MyCiTi web site [6] (right):

Map 6: Open Data Portal Bus Stops vs. Official Bus Map



Given this was a much better result than what we obtained from Foursquare, we decided to stick with this dataset.

At this stage we already see the current extent of the MyCiTi network, and how much there is still to do to cover most of the city.

D. Property Valuations (Open Data Portal)

We then loaded the property valuations from the Open Data Portal, aiming to understand how many people live in a suburb and how wealthy the suburb is. The initial dataset looked as follows:

Suburbs	0-200k	2-400k	4-600k	...	300-400M	400-500M	>500M	Unnamed: 37
ZEVENZICHT	5	NaN	NaN	...	NaN	NaN	NaN	164
ZONNEBLOEM	605	84	191	...	NaN	NaN	1	2638
ZOO PARK	2	1	NaN	...	NaN	NaN	NaN	382
Grand Total	183567	142189	106097	...	36.0	18	44	843454

This dataset groups properties by suburb and value band.

Given that we needed to analyse and visualise this data at some point, it was essential that suburb names in the property valuations dataset matched those in our suburbs dataset we extracted earlier.

We have therefore applied a number of adjustments to the properties dataset, as described below.

First, there were a number of misspellings and 1-to-1 mis-mappings:

Old Name	New Name
BRENTWOOD PARK - Durbanville	BRENTWOOD PARK - DURBANVILLE
DIEPRIVER	DIEP RIVER
EKUPHUMULENI	EKUPHUMLENI
ENKANINI INFORMAL SETTLEMENT	ENKANINI
KATAMUDA	KALAMUNDA
ONVERWACHT (Gordons Bay)	ONVERWACHT - GORDONS BAY
ONVERWACHT	ONVERWACHT - THE STRAND
PINATI	PINATI ESTATE
PROTEAVALLEY	PROTEA VALLEY
SCHOTSCHEKLOOF	BO-KAAP
SEA BREEZ PARK	SEA BREEZE PARK
SILWERTOWN	SILVERTOWN - KHAYELITSHA
SIR LOWRYS PASS VILLAGE CENTRE	SIR LOWRYS PASS VILLAGE CENTRE
SIR LOWRYS PASS	SIR LOWRYS PASS
SUNVALLEY	SUN VALLEY
SURREY	SURREY ESTATE
VANGUARD	VANGUARD ESTATE
WELCOME	WELCOME ESTATE
ZEEKOEIVLEI	ZEEKOEVLEI
ZONNEBLOEM	DISTRICT SIX

Then there were pairs of valuations related to the same suburb, we grouped the valuation data and summed it up:

Old Name	New Name
AMANDA GLEN - Bellville	AMANDA GLEN
AMANDA GLEN - Durbanville	AMANDA GLEN
EVERSDAL - Bellville	EVERSDAL
EVERSDAL - Durbanville	EVERSDAL
KALKFONTEIN 1	KALKFONTEIN
KALKFONTEIN II	KALKFONTEIN
KENRIDGE - Bellville	KENRIDGE
KENRIDGE - Durbanville	KENRIDGE
CAPRI	SUNNYDALE

Finally, there were 6 suburbs in the valuation data that could not be unambiguously mapped to a geographical suburb. There was also 1 suburb that we could not locate on the map at all, DEO GRACIA. We therefore had to exclude all these from our analysis:

Suburb Name
AVONDALE
PROTEA PARK
ROSEDALE
VREDENBERG
WESTRIDGE
WORLDS VIEW
DEO GRACIA

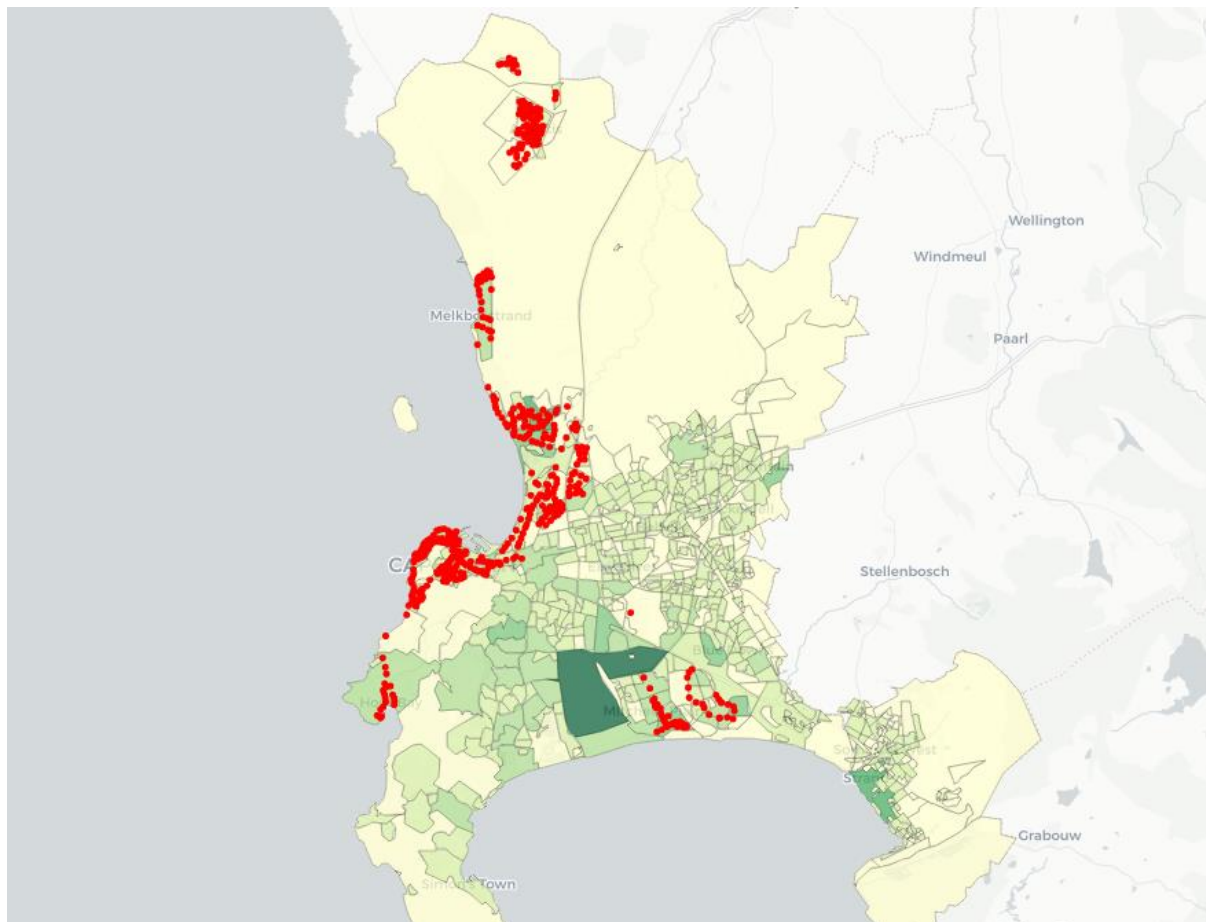
Following some further investigation we came to the conclusion that the dataset included both residential and industrial properties such as warehouses, which ideally we wanted to filter out. For this reason we applied the following further changes:

- Excluded all properties valued at more than ZAR 200M, as they are likely to be industrial
- Excluded suburbs with less than 300 properties (while they may have a need in public transportation, they are likely to be of a lower priority comparing to more populous suburbs)
- Excluded suburbs containing FARM in their name (except SPANISH FARM, which is a residential suburb)
- Excluded suburbs containing INDUSTRIA in their name, and a few others known to be an industrial area or a shopping mall

We also re-calculated the total number of properties per suburb.

Here is the map showing the number of properties per suburb. The greener the area, the more residents it has. MyCiTi bus stops are show in red:

Map 7: Most Populous Suburbs vs. MyCiTi Bus Stops



Looking at the map we immediately see that something interesting could be done about PHILIPPI and STRAND, which are currently not served by MyCiTi. On the other hand, SEA POINT and PARKLANDS are among the most populous suburbs, but both already have a MyCiTi service.

For the same of our analysis, however, we are not only in the number of people living in a suburb, but also their wealth and the likelihood of owning a private vehicle.

We therefore attempted to cluster the suburbs to get an idea of the residents' wealth. Following some analysis (including the elbow method, which we ultimately rejected) we decided to split the suburbs into four clusters.

Number of suburbs in each resulting cluster:

cluster	Suburbs
0	344
1	80
2	16
3	4

Looking at the counts, it is logical to assume that the wealthiest people live in suburbs in cluster 3.

We then enriched the dataset with the count of bus stops in each suburb that we found earlier, the table below shows the top 5 by the number of properties:

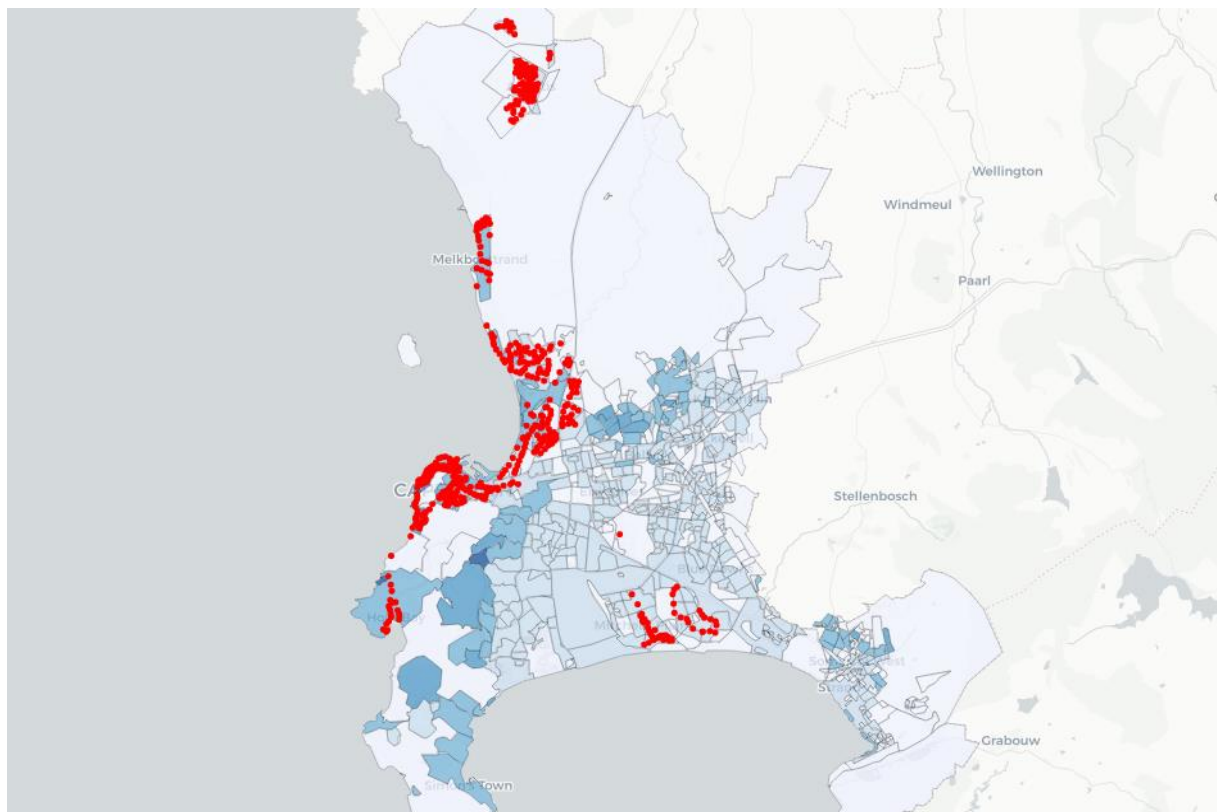
Suburbs	total	cluster	count
PHILIPPI	21,318	0	0
STRAND	13,300	0	0
SEA POINT	12,326	1	20
PARKLANDS	11,225	0	21
CAPE TOWN CITY CENTRE	10,325	1	21

Results

First of all, let's visualise the suburbs clustered into four categories by the value of real estate. The darker the blue colour, the more expensive the real estate is. Our logic is that people in wealthier suburbs are more likely to own a car, and therefore it is these suburbs that need a public bus service to encourage these people to use buses for their daily commute to work.

Existing MyCiTi bus stops are still shown with red dots. Note that even though some of the wealthier (darker) suburbs are almost invisible as they are covered over with red dots, this does not bother us because we are looking for wealthier suburbs not yet covered by the MyCiTi service.

Map 8: Suburb Wealth vs. MyCiTi Bus Stops



Top 10 suburbs in cluster 0 – shown in pale blue on the map:

suburb	total number of properties	cluster	number of bus stops
PHILIPPI	21,318	0	0
STRAND	13,300	0	0
PARKLANDS	11,225	0	21
TABLE VIEW	10,025	0	14
UMRHABULO TRIANGLE	9,288	0	2
GUGULETU	8,932	0	0
WALLACEDENE	8,131	0	0
TAFELSIG	7,994	0	19
MFULENI	7,603	0	0
NYANGA	7,056	0	0

Top 10 suburbs in cluster 1:

suburb	total number of properties	cluster	number of bus stops
SEA POINT	12,326	1	20
CAPE TOWN CITY CENTRE	10,325	1	21
CLAREMONT	9,695	1	0
RONDEBOSCH	7,056	1	0
HOUT BAY	6,963	1	19
GARDENS	5,348	1	16
FISH HOEK	5,302	1	0
DURBANVILLE	4,947	1	0
PINELANDS	4,738	1	0
GREEN POINT	4,451	1	14

Top 10 suburbs in cluster 2:

suburb	total number of properties	cluster	number of bus stops
CONSTANTIA	4,327	2	0
NEWLANDS	3,168	2	0
CAMPS BAY / BAKOVEN	2,933	2	20
NOORDHOEK	2,158	2	0
WESTLAKE	1,724	2	0
FRESNAYE	1,083	2	2
WELGEMOED	957	2	0
SUNSET BEACH	876	2	1
BARONETCY ESTATE	662	2	0
WELGEDACHT	591	2	0

Top 10 suburbs in cluster 3 (oops, there are only 4) – shown in dark blue on the map:

suburb	total number of properties	cluster	number of bus stops
BANTRY BAY	822	3	2
CLIFTON	683	3	5
BISHOPSCOURT	580	3	0
LLANDUDNO	331	3	0

Discussion

Let's begin with the richest suburbs of **cluster 3**. There are only 4: **Bantry Bay**, **Clifton**, **Bishopscourt** and **Llandudno**. Two of them already have a MyCiTi service, the other two do not. These are traditionally the most expensive areas of Cape Town to live in, with a limited number of properties. A reliable bus service is highly unlikely to attract these suburbs' residents, and even if it was, the impact on congestion relief would be insignificant given the number of properties and people in these suburbs. We therefore recommend no changes to the MyCiTi service as far as this cluster is concerned.

Moving on to **cluster 2**, we see that the most populous suburbs in this cluster are **Constantia** and **Newlands**. These are relatively prosperous areas where highly skilled individuals choose to live, and we believe that at least some 90% of them are likely to own a car. It is exactly these neighbourhoods that need a MyCiTi service, along with a large-scale publicity campaign explaining why switching to safe and reliable public transport is good for the overall environment.

Looking at **cluster 1**, **Claremont** and **Rondebosch** are the two most populous suburbs that still have no MyCiTi service. Most people living here are also likely to own a car, so we would recommend the same strategy as for Constantia and Newlands in cluster 2. Noting that Claremont and Rondebosch lie next to Newlands, it may be possible to serve all these suburbs with the same bus route. **Fish Hoek** is the next most populous suburb with no MyCiTi service, however we note that it does have a working train service to Cape Town's centre, therefore it is questionable whether a new MyCiTi line would be the best answer - perhaps it would be wiser to invest into the rail infrastructure instead.

Finally, **cluster 0** is a somewhat different story. Many of the people in these neighbourhoods do not own a car and are having to rely on whatever public transport is available, be it shared minibus taxis or the rush-hour-only Golden Arrow buses. **Strand** is probably an exception as it does have a working rail service, albeit it takes almost 2 hours to reach the centre of Cape Town, which would be unacceptable for daily commute. Similar to Fish Hoek above, we would recommend investing into the rail infrastructure instead of creating a MyCiTi link to Strand. The rest of the suburbs, especially **Philippi** and **Guguletu**, seem to be in desperate need of a reliable bus service, and this is where the Transport and Urban Development Authority of Cape Town should focus its attention. This is, however, more a question of equal opportunities rather than traffic congestion relief.

Conclusion

In this analysis we have looked at the existing MyCiTi bus network, attempted to group Cape Town's suburbs by wealth in order to understand where people are more likely to use a private car, and offered two suggestions as to how to grow the network:

1. To alleviate traffic congestion, we recommend establishing a new, safe and reliable MyCiTi service from Cape Town's centre to the suburbs of **Constantia**, **Newlands**, **Claremont** and **Rondebosch**. Given the existing culture and lifestyle of these suburbs, some effort would be required to motivate people to use buses instead of cars - this may include advertisements or other incentives.
2. We also recommend to link **Philippi** and **Guguletu** to the city centre, not so much for the sake of traffic congestion, but to provide reliable public transport to very populous suburbs that need it most.

References

- [1] <https://www.tct.gov.za/en/transport/transport-network>
- [2] <https://www.news24.com/SouthAfrica/News/multi-pronged-approach-needed-for-traffic-congestion-city-of-cape-town-20190227>
- [3] <https://www.whereismytransport.com/case-studies/integrating-the-informal-mapping-cape-towns-minibus-taxi-network>
- [4] <https://tdacontenthubstore.blob.core.windows.net/resources/756c7e76-1ab5-45a9-9e97-ecee4e150e79.pdf>
- [5] <https://odp-cctegis.opendata.arcgis.com>
- [6] <https://www.myciti.org.za/en/routes-stops/route-stop-station-map>