Data Exploration Project

Commuting Distance to Work in Greater Melbourne

29.04.2020
Jinyoung (Travis) Hong

Contents

1. Introduction

The commute is a part of the lifestyle to more than 9 million Australians who go to work every weekday [1]. On average, an Australian worker travels 16.5 km to get to work and spends 5.75 hours a week taking commute trips [1][2]. And a study has shown that people who commute longer distances tend to have more days off while people who commute short distances are more likely to be happy and more productive [3]. Likewise, commuting distance has non-negligible impacts on our lives. Thus, this project aims to study factors that may influence the commuting distance through data exploration using Python and Tableau. This will allow one to examine the effects of one's method of travel, occupation and living arrangements such as property occupancy and family status on the commuting distance. The questions below are to be used as guidelines throughout the exploration. The scope of this project mainly looks upon the workers living in the Greater Melbourne area, yet regional Victoria is also taken into consideration for comparison.

1.1. Questions

- How does one's work affect commuting distance?
- How does the method of travel relate to commuting distance?
- How does one's living arrangements correlate with commuting distance?

2. Data Wrangling

2.1 Datasets

For this project, the following four datasets are used:
(Please note that links are provided in the Reference.)

- *'Commuting Distance by Personal Characteristics'* from Australian Bureau of Statistics – Census of Population and Housing 2016 [4]
- *'Commuting Distance from Place of Usual Residence'* from Australian Bureau of Statistics – Census of Population and Housing 2016 [5]
- *'Suburbs: Detailed information by suburb and segment of the community'* from Greater Dandenong – Statistical Data for Victorian Communities [6]
- *'Victoria Mesh Blocks ASGS Ed 2016 Digital Boundaries'* from Australian Bureau of Statistics – Australian Statistical Geography Standard (ASGS) [7]

All of these datasets are published in 2016. The first three datasets are tabular data and the last dataset is a shapefile. Before exploring and visualising the data, the tabular datasets were firstly inspected and underwent pre-processing to be reformatted and cleaned. Python was used during this wrangling process.

2.2. Data Inspection

Tabular data

In pre-processing the tabular datasets, they were firstly inspected using Python. The first dataset, 'Commuting Distance by Personal Characteristics' had 4 excel sheets, with row by column sizes of 1050 x 9, 2429 x 9, 1969 x 9 and 1855 x 9, containing information about the Victorian workers' average commuting distance by their occupations, industries, incomes and methods of travel with locations. The second dataset, 'Commuting Distance from Place of Usual Residence' had one excel sheet, with 2305 rows and 7 columns, containing information about average commuting distance travelled by workers from Victoria. And the last tabular dataset, 'Suburbs: Detailed information by suburb and segment of the community' contained an excel sheet of 2932 rows by 288 columns with

statistics about various personal, educational and community backgrounds of people living in different suburbs in Victoria.

Shapefile data
Tableau was used to inspect the shapefile dataset, 'Victoria Mesh Blocks ASGS Ed 2016 Digital Boundaries'. It contained shapefiles for the Victorian suburbs. The list of these suburbs was collected for reformatting purpose in the later process.

2.3. Data Reformatting and Cleaning

Tabular data
After inspecting the loaded data frames, irrelevant rows and columns were removed. Columns containing similar information were clustered, for example, columns depicting numbers of people renting dwellings from private owners and the government were combined as a single column representing a total number of people renting properties. Upon cleaning the columns, they were reordered and renamed to be more meaningful and practical.

For data entries for each column, inappropriate observations were either amended or deleted from the datasets. And referring to the list of suburbs obtained from the shapefile dataset during the inspection process, columns containing area names were filtered and edited to match the shapefile area names. Prior to matching the area names, the values of the suburb list from the shapefile dataset were firstly checked to ensure there were no anomalies. In filtering and amending the data entries, regular expressions were used.

For the method of travel dataset, percentages of people using different transportations to commute in Greater Melbourne suburbs were calculated and added to the dataset. Also, for the dataset containing information about people's living arrangements, proportions of people living with families to living alone and of people owning properties to renting properties were calculated and added to the dataset. And these ratios and commuting distances in different suburbs were normalised using min-max normalisation and added to the dataset.

Shapefile data
Using Tableau, variables of shapefile data were changed with more sensible and uniform names in accordance with the tabular datasets.

2.4. Data Checking

Tabular data
Upon reformatting and cleaning the datasets, data checking was performed to check for any anomalies. Unique values and patterns in the entries were examined to find any anomalies. As all missing values and inappropriate values were either removed or amended in the wrangling process, no anomaly was detected. Some of the errors found in the earlier process include string entries existing in the integer column and irrelevant entries in the column. The first error had string entries such as 'Negative income', 'Not stated' and 'Nil income' in the income column consisted of integer entries. This error was corrected by removing the 'Negative income' and 'Not stated' entries while converting 'Nil income' to zero. And another error had entries such as 'Worked at home (e)', 'Did not go to work (e)' and 'Not stated' in the travel method column. As this project seeks to explore travel methods for commuting, these entries were regarded as irrelevant and were deleted from the dataset.

Besides, the datasets were visually plotted on Python to make sure there were no abnormal or missing data. Graphs were plotted without any errors and no abnormalities were found. After checking all the data, the indexes of the datasets were reset.

Shapefile data

To ensure the shapefiles were not corrupted and had no anomalies, they were plotted on a map using Tableau to see if they showed all of the Victorian suburbs correctly. And the area names were extracted from the dataset and examined on Python to identify any errors. No error was found in the dataset and just variable names were amended during the wrangling process.

2.5. Data Exporting

Once all datasets were pre-processed and error-free, the tabular datasets were exported as CSV files, consisted of 778 rows and 17 rows in total.

3. Data Exploration

In this section, data visualisations were performed to explore various factors that may influence commuting distance. Based on this exploration, the aforementioned questions are to be answered at the end of this project. Also, this will study commuters living in Greater Melbourne region and the rest of the Victoria region. These are also to be called metropolitan area and rural/regional area respectively.

3.1. Commuting Distance by Work

To examine effects of one's work on commuting distance, average commuting distances were plotted against industries, occupations and incomes for people living in Greater Melbourne and regional Victoria.

Effect of industry

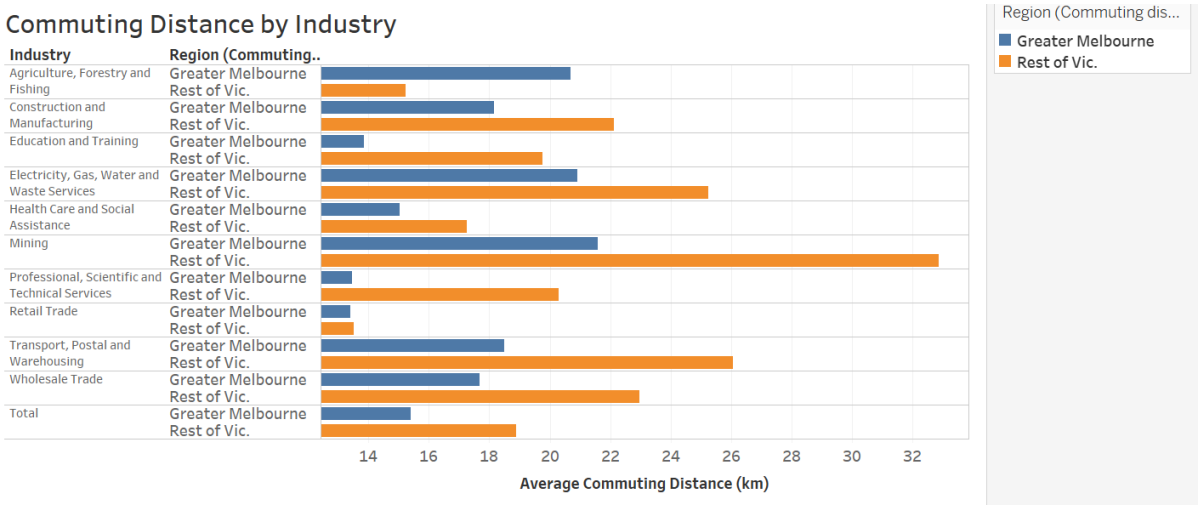

Figure 1 Commuting Distance by Industry in Greater Melbourne and Rest of Victoria

Figure 1 shows average commuting distances by industries for workers residing in Greater Melbourne and the rest of Victoria. This chart shows that in general, commuters living in the regional area travel further than commuters living in the metropolitan area. However, for agriculture, forestry and fishing industries, people from the metropolitan area commute longer distances than people from regional Victoria. This is likely because most businesses are located within the central business district or the metropolitan areas whereas most farms, forests and ports are located in the rural areas. Retail trade industry has the shortest average commuting distances of 13.39km and 13.50km from the Greater Melbourne and regional Victoria. Workers in retail trade travel almost the same distances from the metropolitan area and the regional area.

Whereas, the mining industry depicts the greatest commuting distances of 21.60km and 32.88km for workers living in Greater Melbourne and regional Victoria. And the difference in commuting distances between these areas is the greatest at 11.28km amongst industries.

Effect of occupation

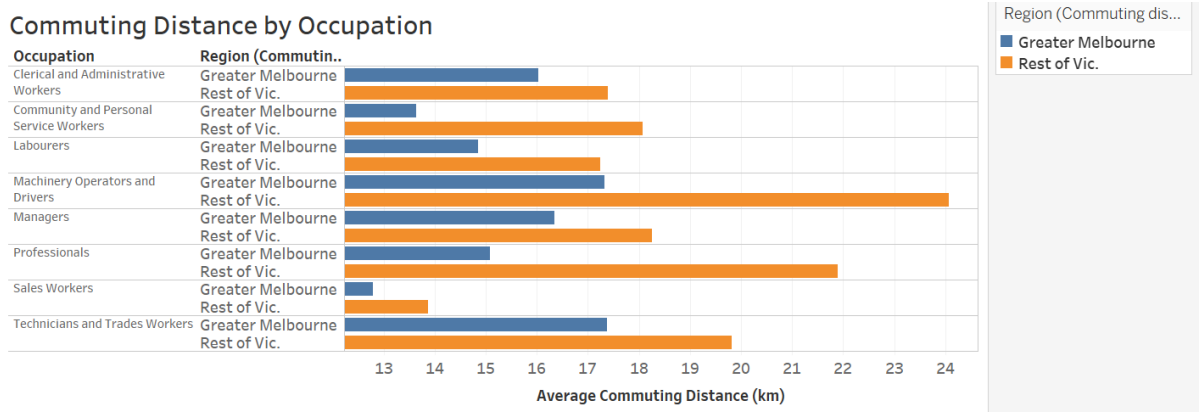

Figure 2 Commuting Distance by Occupation in Greater Melbourne and Rest of Victoria

Likewise, for commuting distance by occupation, workers in rural Victoria show greater average commuting distances than workers in the metropolitan area. Sales workers travel the least distances of 12.79km and 13.87km for both Greater Melbourne and regional Victoria. On the other side, machinery operators and drivers travel the furthest for 17.33km and 24.07km from the metropolitan region and the rural region. Sales workers have the smallest gap in commuting distances with 1.08km between the metropolitan region and regional Victoria, while machinery operators and drivers have the largest gap in commuting distances with 6.73km. On average, workers travel 15.38km and 18.82km from the metropolitan area and the regional area respectively.

Effect of income

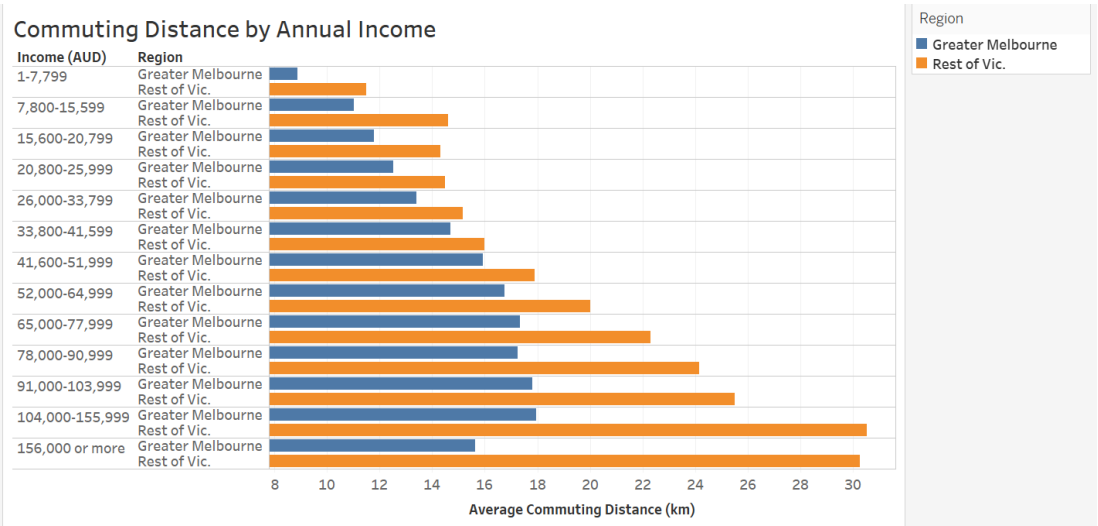

Figure 3 Commuting Distance by Income in Greater Melbourne and Rest of Victoria

Figure 3 shows a gradual increase in average commuting distances from 8.88km to 17.94km and from 11.50km to 30.52km as annual income increases for workers living in Greater Melbourne and regional Victoria respectively. Despite there are drops in commuting distances at an income range of $156,000 or more for workers from Greater Melbourne and regional Victoria, in general, people commuting from regional Victoria travel longer distances than people commuting from the metropolitan region for all income ranges. The difference in commuting distances becomes larger as income gets higher.

Workers commuting from Greater Melbourne and regional Victoria commuted the shorted distances of 8.88km and 11.50km in an income range of $1-$7,799. And workers in an income range of $104,000-$155,999 travelled furthest with distances of 11.50km and 30.52km. As shown from the chart in figure 1, this large commuting distance at high incomes may have been contributed by workers in the mining industry where they show similar commuting distances around 30km and according to iMINCO, an average salary in the mining industry is as $123,844 [8], which falls within the same income range. In terms of commuting distance difference, an annual salary range of $1-$7,799 shows the smallest difference of 2.62km and an annual salary range of $156,000 or more shows the greatest difference of 14.61km between the metropolitan area and the regional area.

3.2. Commuting Distance by Method of Travel

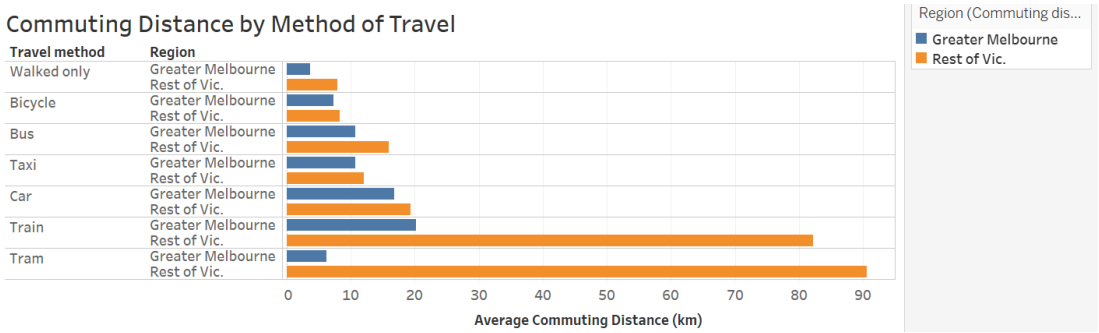

Figure 4 Commuting Distance by Method of Travel in Greater Melbourne and Rest of Victoria

Figure 4 shows a distribution of average commuting distances by different methods of travel in Greater Melbourne and regional Victoria. In overall, people living in regional Victoria commute greater distances compared with people living in Greater Melbourne. For people commuting from the metropolitan area, travelling by walking has the shortest commuting distance of 3.64km and travelling by train has the longest commuting distance of 20.19km. Meanwhile, for people commuting from the regional area, travelling by walking also has the shortest commuting distance of 7.95km but travelling by tram has the longest commuting distance of 90.62km. This longest commuting distance from regional Victoria is followed by train which has a commuting distance of 82.17km.

People who walk to work have the least difference of 4.31km in commuting distances and people who take a tram to work have the biggest difference of 84.43km in commuting distances.

In Victoria, regional bus and train operate in regional areas and this may contribute to the large gap in commuting distances between the metropolitan area and the regional area. However, according to Public Transport of Victoria, trams do not operate in regional zones, hence having an average commuting distance of 82.17km by tram is unrealistic [9]. This could be due to human errors in writing or reading the word, ‘tram’ instead of ‘train’ while processing Census form in the paper. Or people may have added commuting distances that were travelled by other transport to the commuting distances taken by tram.

Methods of travel in Greater Melbourne

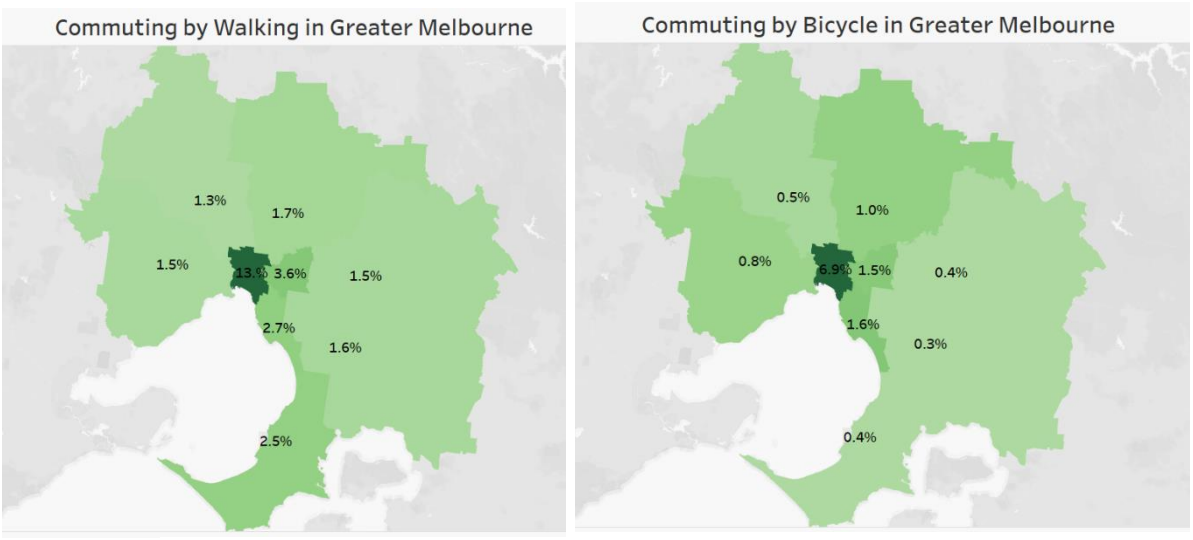

Figure 5 Commuting Methods in Greater Melbourne (Left: Walking, Right: Bicycle)

Figure 5 shows the percentage of people living in Greater Melbourne commuting by walking on left and by bicycle on right. It shows that people living in the central Melbourne commute by walking the most with 13%, followed by Inner East Melbourne, Inner South Melbourne and Mornington Peninsula with 3.6%, 2.6% and 2.5% respectively. The rest of the outer regions show percentages around 1.5%. For commuting by bicycle, it shows a similar trend where people living in the central Melbourne show the highest percentage of 6.9%, followed by Inner East Melbourne and Inner South Melbourne with 1.5% and 1.6%. The other suburbs show 1% or lower percentages.

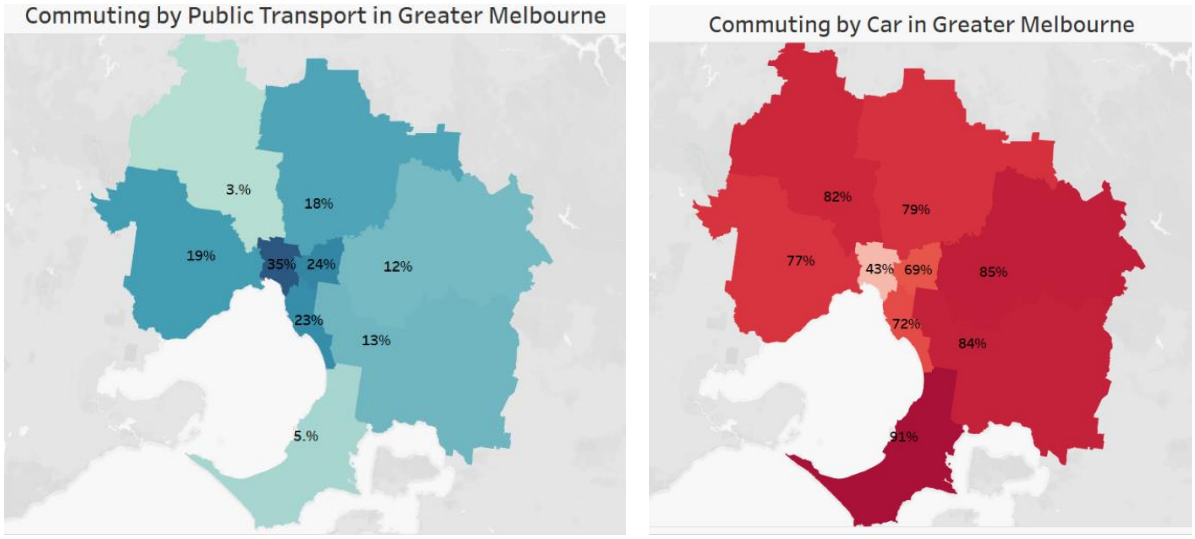

Figure 6 Commuting Methods in Greater Melbourne (Left: Public Transport, Right: Car)

Figure 6 illustrates the percentage of people living in the metropolitan area commuting by public transport on left and by car on right. In terms of commuting by public transport, people residing in Inner Melbourne show the largest percentage of 35%, followed by people living in Inner East Melbourne and Inner South Melbourne with 24% and 23%. People from North West Melbourne commuted by public transport the least with a percentage of 3%. Within Greater Melbourne, people living in Mornington Peninsula commuted the most by car with a percentage of 91%. And people living in Inner Melbourne commuted the least by car with a percentage of 43%.

3.3. Commuting Distance by Family Status and Property Occupancy

Commuting distance in Greater Melbourne on map

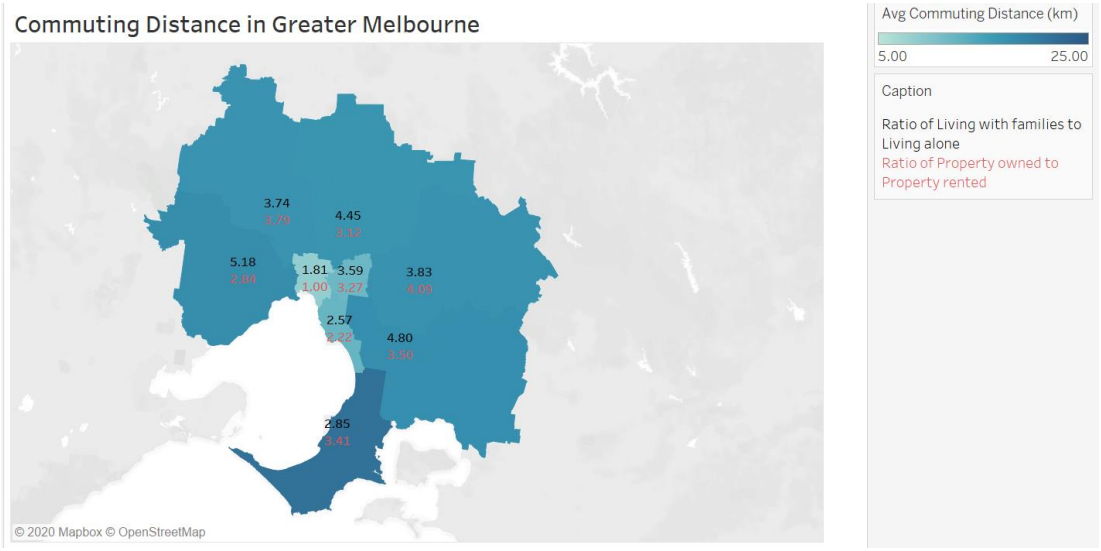

Figure 7 Commuting Distance in Greater Melbourne on Map

Figure 7 shows a choropleth map of Greater Melbourne representing the average commuting distance from each region by colour density. This map also contains labels showing a ratio of people living with families to living alone in black labels and a ratio of people owning properties to renting properties in red labels. This map describes that the average commuting distance is at its lowest in Inner Melbourne with 8.29km, while it is at its highest in the Mornington Peninsula with 21.69km. Inner East Melbourne and Inner South Melbourne have an average commuting distance of 11.76km and 12.31km respectively, while rest of Greater Melbourne areas show around 17km for their average commuting distances. As the colour density suggests, the average commuting distance in an area grows as it gets further away from Inner Melbourne.

Additionally, proportions of people living with families to living alone and of people owning dwellings to renting dwellings are the lowest in Inner Melbourne and they increase in outer areas. To examine correlations between these living arrangements and commuting distance, a scatter plot matrix is plotted in figure 6.

Correlation between commuting distance and living arrangements

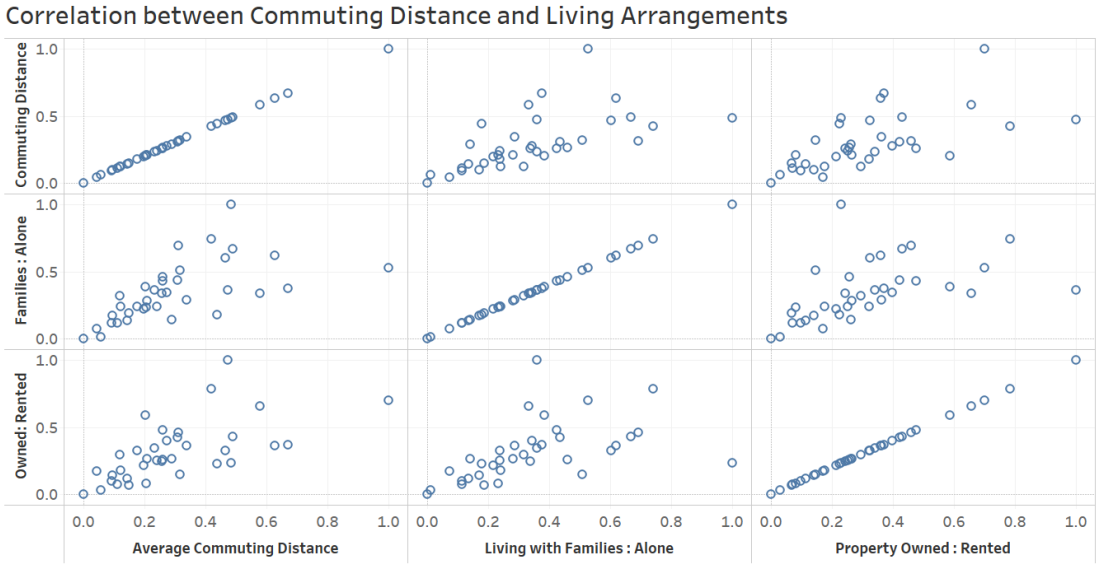

Figure 8 Correlation between Commuting Distance and Living Arrangements

According to the scatter matrix plot, there seem to have positive correlations between commuting distance and proportions of living with families to living alone and of owning properties to renting properties. By using an Excel function, correlation coefficients were calculated. Commuting distance and the ratio of living with families to living alone have a correlation coefficient of 0.62, meanwhile, commuting distance and the ratio of owning properties to renting properties have a correlation coefficient of 0.64. According to (Frost, 2018), a correlation coefficient of 0.6 implies a moderate positive relationship. Hence, commuting distance is in moderate correlation with these living arrangements.
On the other hand, a ratio of living with families to living alone and a ratio of owning properties to renting properties have a correlation coefficient of 0.51, which proves that they do not have a strong correlation with each other.

4. Conclusion

This project aims to evaluate various factors that may impact commuting distance for people residing in Greater Melbourne. Appropriate and reliable data sources were obtained from Australian government websites and through data wrangling and data checking processes using Python and Tableau, the tabular and shapefile datasets were reformatted and cleaned for efficient data exploration. Tableau was used to visualise the pre-processed datasets and the following questions were studied to determine their correlations with commuting distance.

4.1. How does one's work affect commuting distance?

Industry

It was shown that an individual's work such as industry, occupation and income have correlations to commuting distance. For industry, people working in retail trade commute the shorted distance of 13.39km and 13.50km from Greater Melbourne and regional Victoria, meanwhile, people working in mining commute the longest distance of 21.60km and 32.88km from Greater Melbourne and regional Victoria. And in general, people living in regional Victoria travel further compared with people living in Greater Melbourne. However, for agriculture, forestry and fishing industry, people from Greater Melbourne travel larger distance.

Occupation

For occupation, sales workers commute the least distance of 12.79km and 13.87km from the metropolitan area and the regional area. On the other side, machinery operators and drivers commute the greatest distance of 17.33km and 24.07km from the metropolitan and the regional area. And similar to the industry trend, people residing in the regional Victoria show greater average commuting distances than people residing in the metropolitan area.

Income

Average commuting distances show a gradual increase with growing incomes except for the high-end income range of $156,000 or more. People earning from $1 to $7,799 per annum travelled the shorted average commuting distance of 8.88km and 11.50km from the metropolitan area and the regional area. On the other side, people earning annual incomes from $104,000 to $155,999 travelled the biggest average commuting distances of 17.94km and 30.52km from the metropolitan area and the regional area. Workers living in the regional area travel further distances than workers living in the metropolitan area. And this gap increases as income increases.

In overall, it was shown that certain industries and occupations have considerable effects on commuting distance and people earning higher income tend to travel further from their residences.

Also, on average, people residing in regional Victoria commute longer distances in comparison with people residing in Greater Melbourne.

4.2. How does the method of travel relate to commuting distance?

Amongst various methods of travel for commuting, people living in Greater Melbourne and regional Victoria commute the least distance by walking with 3.64km and 7.95km. And people living in Greater Melbourne commute the largest distance of 20.19km by train, while people living in regional Victoria commute the greatest distance of 90.62km by tram, followed by a commuting distance of 82.17km by train. People residing in the regional area commute further than people residing in the metropolitan area. Especially, for train and tram, average commuting distances show significant differences of 61.98km and 84.43km between travelling from Greater Melbourne and regional Victoria respectively. However, as trams do not operate in the regional zones, large commuting distance by tram from the regional area may have been an error.

Within Greater Melbourne, the least people walk to commute, followed by bicycle. Inner Melbourne shows the highest percentages of 13% and 6.9% for people walking and taking a bicycle to commute, whereas other suburbs show much smaller percentages. For public transport, workers residing in Inner Melbourne show the greatest percentage of 35% of people using public transport to commute, followed by Inner East Melbourne and Inner South Melbourne with percentages of 24% and 23%. Having said that, people living in outer Melbourne suburbs show a greater percentage of using a car to commute. 43% of the people living in Inner Melbourne commute by car, yet 91% of the people living in Mornington Peninsula commute by car.

Thus, people who commute by walking and a bicycle tend to commute shorter distances, while people who commute by train tend to commute longer distances. And people from regional Victoria likely to travel further than people from Greater Melbourne. Besides, more people commute by walking, bicycle and public transport in Inner Melbourne than outer Melbourne suburbs. In comparison, more people commute by a car in outer Melbourne suburbs than Inner Melbourne.

4.3. How does one’s living arrangements correlate with commuting distance?

Living arrangements such as family status and property occupancy have a reasonable correlation with commuting distance. For family status, a correlation coefficient of commuting distance and a ratio of people living with family to living alone is 0.62. And for property occupancy, a correlation coefficient of commuting distance and a ratio of people owning properties to renting properties is 0.64. Hence, commuting distance is moderately correlated with these living arrangement factors. However, family status and property occupancy do not show a strong correlation with each other.

5. Reflection

Throughout this project, it was learnt that there are various factors influencing commuting distance such as work, method of travel and even living arrangements. The project would have been more versatile and in-depth with more comprehensive datasets such as job availabilities in different suburbs in Victoria and limitation of transportations in regional areas.

And as the datasets are based on the Census conducted in 2016, some insights may be outdated and different from the current trends. And if the datasets had multiple yearly data, a time-series analysis could have been done to visualise how commuting changes over time.

As the next Census is to be conducted in 2021, a continuation of this exploration project would give a great understanding of how commuting has changed over 5 years.

6. Reference

1. Australian Bureau of Statistics. (2018). 2071.0.55.001 - Census of Population and Housing: Commuting to Work - More Stories from the Census, 2016, 2016. Retrieved from: https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/2071.0.55.001~2016~Main%20Features~Feature%20Article:%20Journey%20to%20Work%20in%20Australia~40
2. RMIT ABC. (2015). Fact check: Do nine in 10 Australians spend more than 90 minutes a day commuting?. ABC News. Retrieved from: https://www.abc.net.au/news/2015-07-09/commuting-times-travel-shorten/6592510
3. Liang, Ma., Runing, Ye. (2019). Does daily commuting behaviour matter to employee productivity?. Journal of Transport Geography, 76 (2019), 130-141.
4. Australian Bureau of Statistics. (2016). Commuting Distance by Personal Characteristics. Australian Bureau of Statistics. Retrieved from: https://www.abs.gov.au/ausstats/subscriber.nsf/log?openagent&2071055001%20-%20commuting%20distance%20by%20personal%20characteristics.xls&2071.0.55.001&Data%20Cubes&1AF69AD91F2CA7EFCA25829400126358&0&2016&22.05.2018&Latest
5. Australian Bureau of Statistics. (2016). Commuting Distance from Place of Usual Residence. Australian Bureau of Statistics. Retrieved from: https://www.abs.gov.au/ausstats/subscriber.nsf/log?openagent&2071055001%20-%20commuting%20distance%20from%20usual%20residence.xls&2071.0.55.001&Data%20Cubes&E16B01F03743421BCA258294001262CB&0&2016&22.05.2018&Latest
6. Greater Dandenong. (2016). Suburbs: Detailed information by suburb and segment of the communities. Greater Dandenong. Retrieved from: https://greaterdandenong.com/document/28875/detailed-information-by-suburb-and-segment-of-commuity
7. Australian Bureau of Statistics. (2016). Victoria Mesh Blocks ASGS Ed 2016 Digital Boundaries. Australian Bureau of Statistics. Retrieved from: https://www.abs.gov.au/AUSSTATS/subscriber.nsf/log?openagent&1270055001_mb_2016_vic_shape.zip&1270.0.55.001&Data%20Cubes&04F12B9E465AE765CA257FED0013B20F&0&July%202016&12.07.2016&Latest
8. iMINCO. (n.d.). Mining Jobs Salaries – How Much Can you Earn?. iMINCO. Retrieved from: http://iminco.net/mining-job-salaries-how-much-can-you-earn-mining-careers/
9. Public Transport of Victoria. (n.d.). Zones. Public Transport of Victoria. Retrieved from: https://www.ptv.vic.gov.au/tickets/fares/zones/