SMU
SINGAPORE MANAGEMENT UNIVERSITY

School of
Computing and
Information Systems

IS483: IS Project Experience (Business Analytics)

AY 2021/2022 Semester 1

Project CDG
Team CDG
Final Project Report

Team Members:

Name	ID	Email	Role
Choo Hui Xing	01378884	hxchoo.2018@scis.smu.edu.sg	Back-end Developer
Iris Har Jing Ru	01375564	iris.har.2018@scis.smu.edu.sg	Data Engineer
Ivy Hau Jia Yi	01350274	ivy.hau.2018@scis.smu.edu.sg	UI/UX Designer
Shazarifah Shawal	01373871	shazarifah.2018@scis.smu.edu.sg	Data Analyst
Shermin Tan	01374546	shermin.tan.2018@scis.smu.edu.sg	Project Manager
Sonia Johari	01376980	soniajohari.2018@scis.smu.edu.sg	Front-end Developer

Other Relevant Parties:

Name	Email	Organization	Department	Role
Gary How	garyhow@comfortdelgro.com	ComfortDelGro Corporation	Group Digital Office	Data Team Lead
Tan Poh Choo	pctan@smu.edu.sg	Singapore Management University		Project Supervisor

Table of Contents

Executive Summary

Singapore is a country with a well established transport infrastructure. From public buses to its Mass Rapid Transit (MRT) system, Singapore's public transport provides its services to its 5.8 million population on a daily basis. Other popular forms of transport amongst its population are taxi, and private hire vehicles (PHV) such as Grab and Gojek. ComfortDelgro (CDG), Singapore's largest taxi company, has recently launched their own ride hailing service early this year. With the fall in ridership during this pandemic throughout all services in the land commuter transportation sector, it is important for CDG to understand its consumer behaviour, the factors affecting their decision to get a good understanding of the transport demand. Thus, this project aims to discover the gaps and insights from the demand-influencing factors of the overall Singapore mobility market to provide CDG information on consumer behaviour. These insights from various analyses such as factor, cluster, and sentiment analysis will be displayed on PowerBI, a web-based dashboard to provide CDG better understanding of their customers.

1.0 Project Overview

1.1 Client's Background

ComfortDelgro (CDG) was established in 2003, from a merger between two Singapore transport companies, namely, the Comfort group and DelGro Corporation. Currently, it is one of the largest land transport companies in Singapore, as well as globally, with a shareholder base and footprint, operating in other countries such as Malaysia, China, and the United Kingdom. Known for its taxi services in Singapore, CDG's businesses also include, buses, rails, private hiring, car rental, leasing, driving centres, automotive engineering, inspections and testing services, non-emergency patient transport services, insurance broking services, and outdoor advertising.

1.2 Project Description

This project aims to provide CDG a better understanding of their consumers by identifying factors that influence transport demand to discover insights, gaps and strategies. To identify these factors, analysis such as factor, clustering, and correlation analysis will be conducted while taking into consideration various factors such as geographic, demographic, and interest to optimize strategies for different CDG business units. While such analysis does not specifically illustrate the exact transport demand, it is an important indicator of the factors that are influencing consumer's behaviour, which will in turn affect the transport demand. Machine learning algorithms will also be used for text analysis to discover general perception of the transportation landscape. The project will focus its analysis on tertiary students to provide for a more in-depth evaluation of this topic, while serving as a first-cut of the study on the population's transport demand. The business units included in the project

are namely, bus, train, taxi, private hire, and car rental. To provide a better understanding of the insights, a dashboard with the main insights from the project will be collated and visualized with the appropriate charts and explanations, with the target audience being the CDG data team. Any useful insights will then be determined by the data team and disseminated to the respective teams.

This project will utilize data from various external and primary sources, namely; Singapore's public transport and population data, primary survey and focus group, and data from Consumer Satisfaction Index of Singapore (CSISG). Further data collection will be scrapped from social media platforms and forums such as Reddit will be conducted as well.

1.2.1 Problem Statement

Transportation demand in Singapore has been unstable since the pandemic. With the changing restrictions from Covid-19, ridership had fallen across both public transport as well as taxis and private hires (TAN, 2021). In order for CDG to optimize their resources and marketing strategies during these times, it is necessary to understand the gaps in the market by conducting an analysis to understand the transportation demand in Singapore instead of estimating it. The team aims to do so by identifying demand-influencing factors that could affect commuters' decision making, which would ultimately affect the demand of transportation services. This will provide CDG a broader view of the transportation landscape in Singapore to help in designing their strategies.

Thus, the problem statement for this project is to:

" Discover the gaps and insights from the demand-influencing factors of the overall Singapore mobility market. "

To answer the problem statement, the team will run multiple analyses, namely, sentiment, factor, clustering, correlation, and geospatial analysis. The results and insights will be presented on a dashboard and delivered to CDG to provide an overview of all the analysis, as well as an easy to understand visualization. This dashboard will allow CDG to view the insights in a single platform.

With the dashboard, it will allow CDG to gain insights and gaps of the mobility transport market to tide through these unstable times.

1.3 Motivation

The motivation for this project is to help CDG gain a better understanding of the Singapore transportation landscape and discover the gaps in the current market.

CDG has several land transport business units in Singapore. However, there are still gaps in the transport market that would result in missed opportunities. With the dashboard, it can identify potential gaps in the market CDG could fill and leverage on, and better strategize market supply strategies based on various factors. Sentiments of their consumers and competitors will aid the company in further understanding areas of improvement.

1.4 Deliverables

The deliverables include the CSV files of all data from various sources, web-based dashboard, comprehensive report and the requested user guide by the client as well as google drive folder link consisting of poster, video and any other supporting documents. A data migration will be conducted to hand over the code to the Sponsor via Azure Data Lake.

2.0 Research Study and Preliminary Analysis

2.1 Focus Justification

The target audience are tertiary students. The client wanted to focus on tertiary students as it is the age group that the team is a part of as well, hence making the team experts amongst those in this generation. At the same time, the team's resources and connections entail those of a similar age group as them, hence making it easier to reach out to tertiary students to gather their responses as well. Due to the constraint in time and resources, it would be unfeasible to conduct an analysis of the entire Singapore population. In fact, the age group of those below 30 make up the highest age group that take both public bus and MRT (and individually) to work in Singapore, hence indicating that the same target audience are a significant age group in the overall mobility market that is worth studying on. (Data.gov.sg, 2015).

2.2 Literature Review

2.2.1 Clustering

Clustering Methods with Qualitative Data: A Mixed Methods Approach for Prevention Research with Small Samples

Based on a qualitative study done before in testing the accuracy of cluster assignment using three different clustering methods with binary data, results indicated that hierarchical, K-Means, and latent class clustering analysis produced similar levels of accuracy with binary data (Henry et al., 2015). However, the different clustering methods show that using KMeans to calculate distance for

categorical data points is not the most ideal. Hence, based on this research analysis, the team will be exploring other unsupervised algorithms such as Hierarchical Agglomerative Clustering.

2.2.2 Factor Analysis

The Demand Determinants for Urban Public Transport Services: A Review of the Literature

Based on a review article on the different demand determinants for public transport services, the study provided several empirical evidence of factors which create a public transport decision-maker. However, the review lacks analytical models in supporting how these factors should be considered in affecting the demand (Polat, 2012). Thus, there is a need to improve the analysis approach through the use of factor analysis to understand which factors are more important in affecting the demand.

2.2.3 Sentiment Analysis

Twitter Sentiment Analysis Using Machine Learning Techniques

In this paper, the author introduces sentiment analysis models based on Naive Bayes and Support Vector Machine that analyze sentiment analysis more effectively. (Le & Nguyen, 2015). Hence the team has decided to run these models as well to check the accuracy of the analysis.

2.2.4 Topic Modelling

Discovering themes and trends in transportation research using topic modeling

This study uses topic modeling and Latent Dirichlet allocation (LDA) and present measures to quantify topic distribution by aggregating the result by country, region, and time (Sun, L., & Yin, Y. (2017). Hence, LDA, Network of word co-presence algorithms would be suitable for the analysis.

2.2.5 Correlation Analysis

Measures of Association: How to Choose?

This article discusses the different correlation methods and the suitable correlation to run base in the types of the two variables and the aim of the analysis. (Khamis, 2008). Based on this article, the Spearman would be a better method to run for the correlation analysis as the variables are ordinal and a continuous data type. In addition, due to the multiple levels of rank, Spearman would be a better choice as compared to Kendall.

Public attitudes toward encouraging Sustainable Transportation

This study analyses the relationship between factors identified from factor analysis with various transportation variables. (Ting et al., 2017) Similar to the project the team is currently doing, this study ran a correlation analysis between ordinal and continuous variables using the Spearman method, which further proves that Spearman method is the most appropriate out of the three main methods of

correlation (Pearson, Kendall, Spearman). The study is also similar to the current project the team is running as it uses factor analysis to run a correlation analysis on the factors identified against other continuous variables.

2.3 Data Collection & Pre-Processing

For the data collection, the team acquired data from a variety of sources, ranging from primary survey data, public data such as LTA DataMall and data scrapped online from social media platforms.

For this project, the client lacked the data required to answer the problem statement. Thus, the team needed to conduct a primary survey. Although the team was limited by the lack of expertise in creating surveys, it is important to take this initiative to get a sensing of the market from the ground.

2.3.1 Scraped Data

Scraped data used libraries like Praw, Selenium, Igramscraper and Twint. Upon scrapping and initial evaluation of the chosen social media platforms (Facebook, Instagram, Twitter and Reddit), Reddit was deemed as the only appropriate platform with useful insights that could be used due to its forum-like page that would encourage discussion, which was the aim of data scraping. Other reasons Reddit was chosen was due to the tightened data security on other sites like Facebook, which made scraping difficult. Instagram is also not a suitable choice as it is largely based on image sharing instead of discussions that the team is looking for. Thus, it is the only social media the team proceeded in using for the dashboard.

<u>Twint</u>
Twint is used to scrape Twitter for their posts and comments. The results scraped from twitter did not provide data with regards to discussions. The main data scraped was the post created by CDG to inform users about relevant promotions ongoing. Hence, twitter was determined to be unsuitable for this project .

<u>Igramscraper</u>
Igramscraper was used to scrape Instagram for posts and comments. As Instagram is a platform that mainly shares its post via image, the post and comments are short and do not offer much insights in terms of discussion relating to transportation. Hence, Instagram was not suitable for this project.

<u>Selenium</u>

Selenium was used to scrape Facebook. However, with the tightened security on Facebook, it is difficult to scrape mentions or discussion on Facebook. Thus, the team decided to omit the scraping from Facebook.

Praw

Praw was used to crawl posts and comments in Reddit. As Reddit is a platform with a forum style, there were various discussions on various topics. Both comments and posts can be scraped as Praw is able to scrape those without the account being banned. As the team aims to find out additional information regarding the perception and thoughts of Singaporeans on transport, Reddit would be a good choice to extract that information. To extract this information, keywords such as 'transport', 'bus', and 'taxi' were used to scrape any post or comments with these keywords present. These keywords were scraped on specific pages like r/Singapore and r/AskSingapore to extract discussions or posts made by the Singapore community. This data will be used for further analysis in the later part of the project.

2.3.2 External Data

The team collected data from Singstat, Statista, Reddit and LTA DataMall. Population data for example, would allow the team to dissect the Singapore map by its respective planning area.

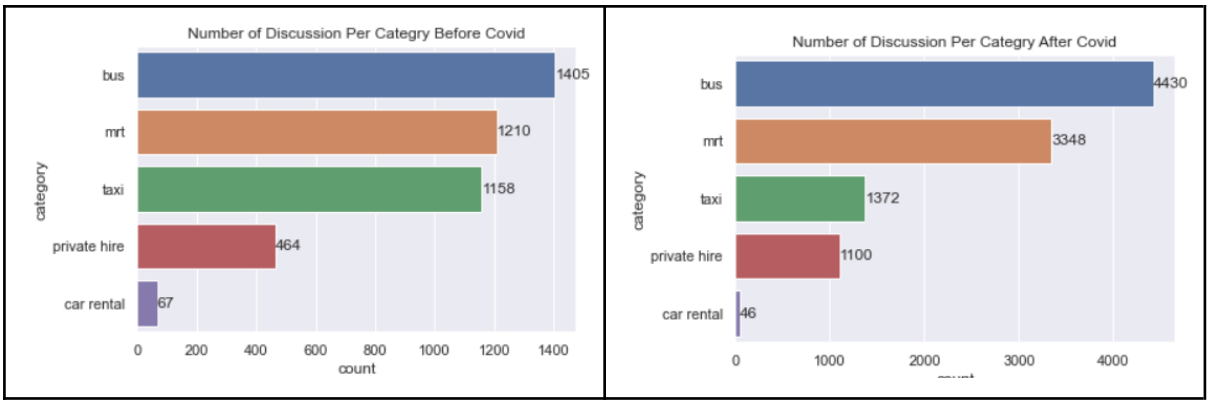


Figure 1: Bar Charts of Discussion Posts Per Category on Reddit

In Figure 1, there is an increase in discussion across all topics except car rental after Covid-19, especially for 'bus' and 'mrt'. Another external data source that the team looked into was the study on Customer Satisfaction Index of Singapore (CSISG) in Transport done by SMU's Institute of Service Excellence. They have kindly shared part of their data and some examples of data points are open ended questions on the perception of different transport modes and satisfaction on various

conditions like safety, cleanliness and ease. This data will allow the team to further explore new factors that contribute to the public's satisfaction of the current transport market.

2.3.3 Survey

Due to the lack of openly available data, the team decided to create a survey focusing on tertiary students to learn more about the sentiments on different transportation in Singapore. Using Microsoft Form, the team collected survey results from 402 respondents across Singapore by distributing the survey link through various Telegram channels. The team chose to use Telegram, a free messaging platform, to reach tertiary students by sending the survey into different universities and polytechnics channels, survey channels like SG Research Lobang as well as to the team's own social circles. Since the platform is free, it is easier for the team to disseminate the survey because Telegram has open group chats and channels that are accessible to the public.

In this survey, the team asked questions to find out the respondents their demographics, travel patterns, transport preferences, psychographics and their thoughts on the different business units. Refer to Appendix B for examples on how the survey was crafted.

Figure 2 below shows an overview of the demographics of the survey respondents (gender make-up, age group, location of residence), as well as the average scores of three psychographics - related questions to understand the cognitive factors that drive consumer behaviours of our respondents.

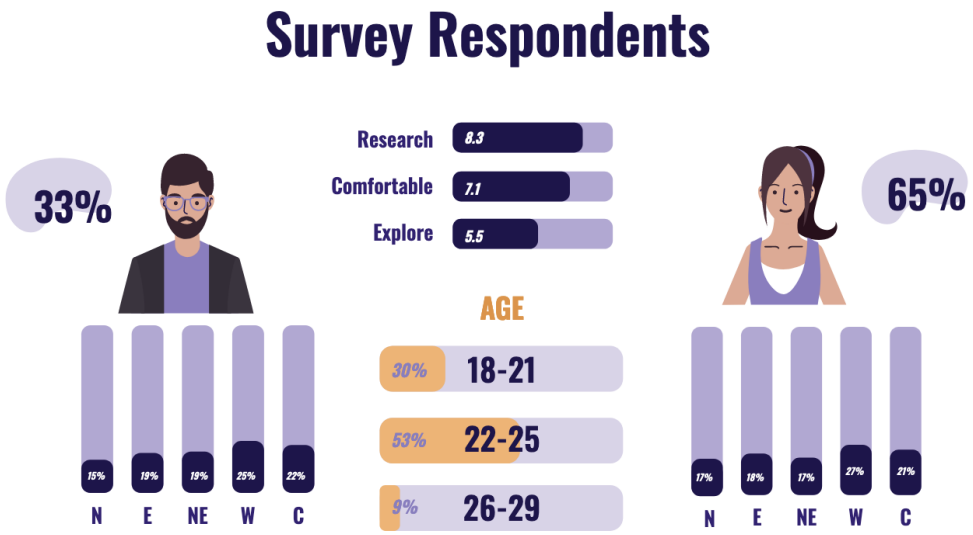


Figure 2: Demographics of Survey Respondents

The data extracted in excel file is formatted in a way that each question is a header and each row contains multiple answers chosen by each respondent. Below refers to the steps of how the data was cleaned and manipulated.

1. Renaming of columns	4. Removal of those not tertiary students
2. Manipulating of ranking questions to get most preferred, least preferred	5. Replacing missing categorical values with mode
3. Removal of unnecessary characters for example:‘ ; ‘, ‘ [] ‘	6. Label encoding for categorical data

Table 1: Steps of cleaning data

Given that the survey data mostly consisted of categorical variables, the team considered several encoding techniques that are suitable for the dataset. The team encoded the data with 2 different methods.

1) Label Encoding

For label encoding, the team replaced categorical variables with numerical values. While it seems like there aren't any big issues with this technique, label encoding induces a new problem since it uses number sequencing (Yadav, 2019). The issue comes when the technique introduces relations between the variables. This could affect the accuracy of finalised results when certain values are given higher importance than the other data values.

2) One-Hot Encoding

This technique refers to the creation of new columns with binary encoding for every data point for each categorical variable. An issue with this technique is that it will generate 1,000 additional new attributes the categorical attribute contains, for example, 1000 unique values. This is not desirable as it will further increase the dimension of the dataset that already contains 100 columns.

With that, the team decided to proceed with label encoding because if dummy or one hot encoding were to be implemented instead, the number of dimensions will be more than 1000 which is very undesirable to analyse. Furthermore, it does not make sense to proceed with these number of features when planning to do factor analysis to reduce dimensions.

Additionally, exploratory data analysis has been conducted to do initial investigation of the survey data. Below figures are some visualizations that help the team to spot patterns that can be analysed further.

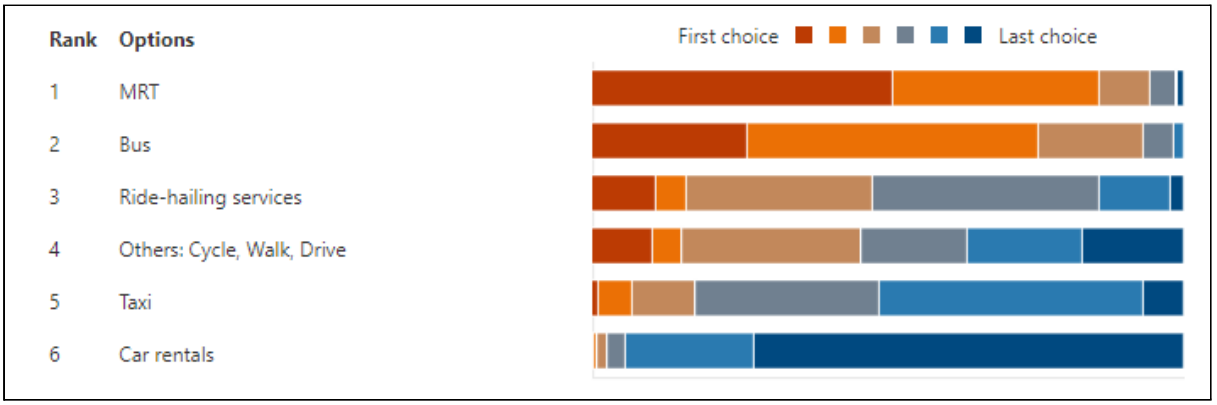


Figure 3: Respondents' preferred mode of transportation

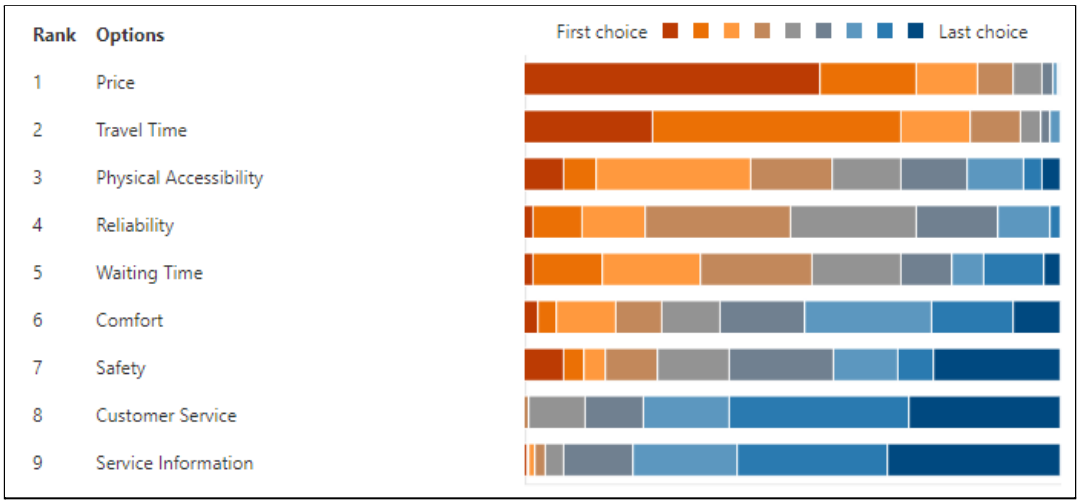


Figure 4: Respondents' ranking of factors

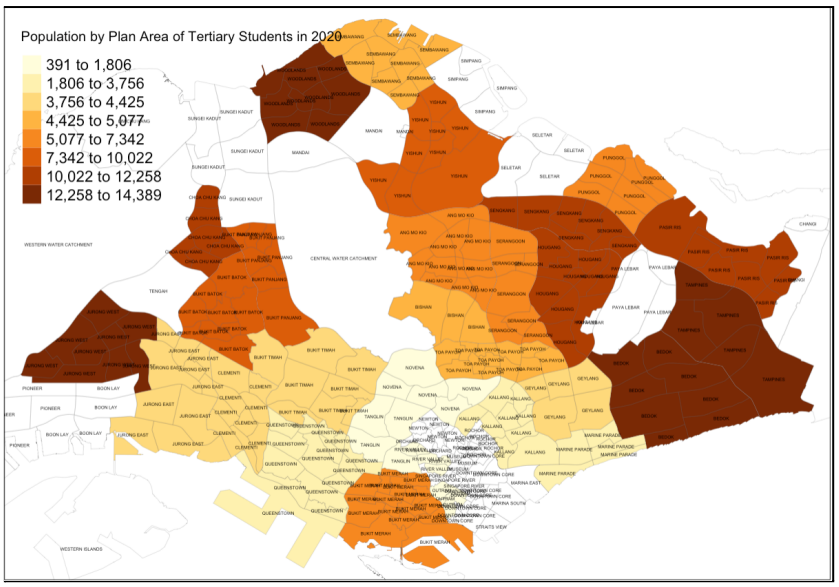


Figure 5: Choropleth Map of Tertiary Students Population by Planning Area (2020)

2.3.4 Focus Group

The team conducted a focus group to fill up the gaps that had yet to be answered. These gaps mainly cover car rental subsectors regarding tertiary consumers motivation, perceptions, and what they look up for in a rental car. The reason for doing so was due to the lack of responses from tertiary survey respondents that had rented before. The majority of the survey respondents had not rented a car before, thus, it was hard to get an accurate representation of the car rental market. The focus group allowed the team to engage individuals who had rented a car before and get a better gauge of their feelings and experiences.

2.4 In-Depth Analyses

2.4.1 Text Analysis

Text analysis allows the team to derive meaning from unstructured textual data such as open-ended responses and discussion posts and comments. The team utilized sentiment analysis and topic modelling to draw insights and measure consumer opinions on the transportation landscape.

2.4.1.1 Sentiment Analysis

The extraction of information from a piece of text is referred to as sentiment analysis (Tyagi & Sharma, 2018). Sentiment analysis is conducted to discover the overall sentiments of the target audience regarding the Singapore transportation market, the changes in their sentiment as well as the current gaps in the market that CDG can address.

Sentiment analysis uses natural language processing (NLP) to determine if a data is positive, negative, or neutral. Sentiment analysis helps to discover sentiments of comments that were collected via survey, reddit, and CSISG. The team trained the LSTM model on a pre-labelled sample Twitter dataset and to label the text as positive, negative or neutral (1,-1,0). Then, visualisations were created to examine the findings deeper.

2.4.1.1.1 Findings

From Reddit and CSISG, based on their inputs, people share a neutral to strongly positive sentiment towards public transportation in Singapore (Appendix D). From Reddit, the team analysed the topics and sentiments to identify changes before and after Covid-19. In general, there were no drastic changes in terms of sentiment patterns. There was an increase in discussion in all topics (excluding car rental) after Covid-19 on Reddit, especially for 'bus' and 'mrt'. In addition, the topics discussed in the combined negative word clouds are 'Time', 'Fare', 'Cost' and 'Service'.

In regards to the survey data, the limitation of the survey is that it was created by the team who lacked expertise in survey creation, thus the questions curated were not neutral and had an underlying bias in sentiment. For example, the question "Based off your ranking choices above, why do you prefer not to take the last 2 least preferred mode of transport?", it already has a negative connotation to it (Figure 6).

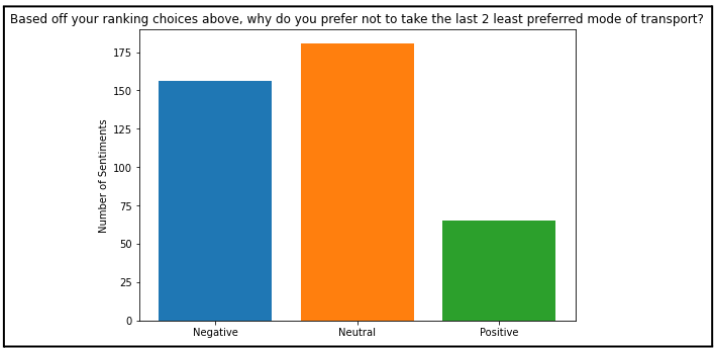


Figure 6: Sentiments of Survey Question

Thus, the surface level sentiments of sentiments from the survey are not very useful, but the word cloud visualisations allow a greater understanding to the various business units. For example, for the private hire business unit, the team can understand that people choose to ride hail from Grab mostly, and the reason people use these ride-hailing applications is because they are easy, fast and convenient (Figure 7).

Figure 7: Positive Word Cloud for Private Hire Preference Reasons (Survey)

2.4.1.1.2 Insights

Overall, people share a neutral to strongly positive sentiment towards public transportation in Singapore. This could imply that people are mostly indifferent or satisfied with the current transport landscape. However, there is still room for improvement regarding the various areas. For instance, the topics discussed across the negative word clouds from the various sources are 'time' and 'money' (Figure 8). This means that people are not happy with the waiting times, the price of transportation and the service they received using the different business units. Cost is a consistent factor considered by the general public, especially so to tertiary students. These are gaps in the market CDG can look into to improve the perception and satisfaction towards public transport.

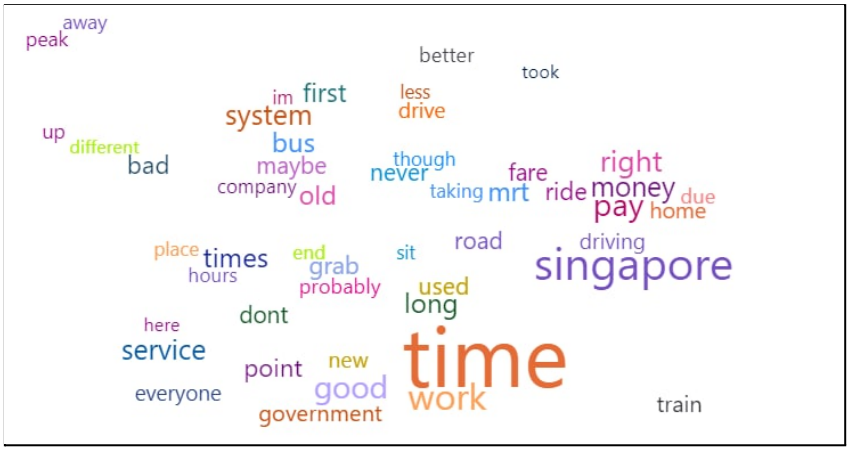


Figure 8: Overall Negative Word Cloud

2.4.1.1.3 Evaluation of model

As sentiment analysis is unsupervised learning, the team would choose the best text machine learning model to label the unseen data. The approach was to explore various machine learning models commonly used for sentiment analysis such as Logistic Regression (Tyagi & Sharma, 2018) on a fixed Twitter dataset. To measure the models based on quantifiable metrics, the team used accuracy, precision, recall and F1 score to determine the best model to utilise for the analysis.

Model	Accuracy	Precision	Recall	F1 Score
Long Short Term Memory (LTSM)	95.38 %	95.48 %	95.27 %	95.38 %
Naive Bayes	73.53 %	73.53 %	73.53 %	73.53 %
Logistic Regression	73.53 %	94.23 %	94.23 %	94.23 %
Valence Aware Dictionary and sEntiment Reasoner (Vader)	80.88 %	89.2 %	82.55 %	85.74 %

Table 2: Results of evaluation

Based on Table 2, LSTM achieved the highest accuracy, precision, recall and F1 Score among the various machine learning models based on the Twitter dataset. Thus, LSTM proved to be the best model in labelling sentiments of text. Since sentiment analysis is unsupervised, the evaluation of the final results is not possible unless the team manually labelled the textual data.

2.4.1.2 Topic Modelling

The team conducted topic modelling to help identify possible topics previously missed in the EDA and sentiment analysis word cloud to be applied into any further analysis. With topic modelling, the team can identify the major themes from the various data sources and uncover hidden topical patterns (KDnuggets, n.d.).

The team chose to use Latent Dirichlet Allocation (LDA). LDA is a Bayesian graphical model (Appendix C) for text document collections represented by bags-of-words. Each document in a collection of D documents is modelled as a multinomial distribution over T topics, with each topic being a multinomial distribution over W words in a topic model. Only a few words in each topic are typically important (have high likelihood), and only a few themes are found in each document. (Newman et al., 2010). In addition, the team combined the textual data from Reddit, the survey and CSISG to get an overview of top topics from all the sources.

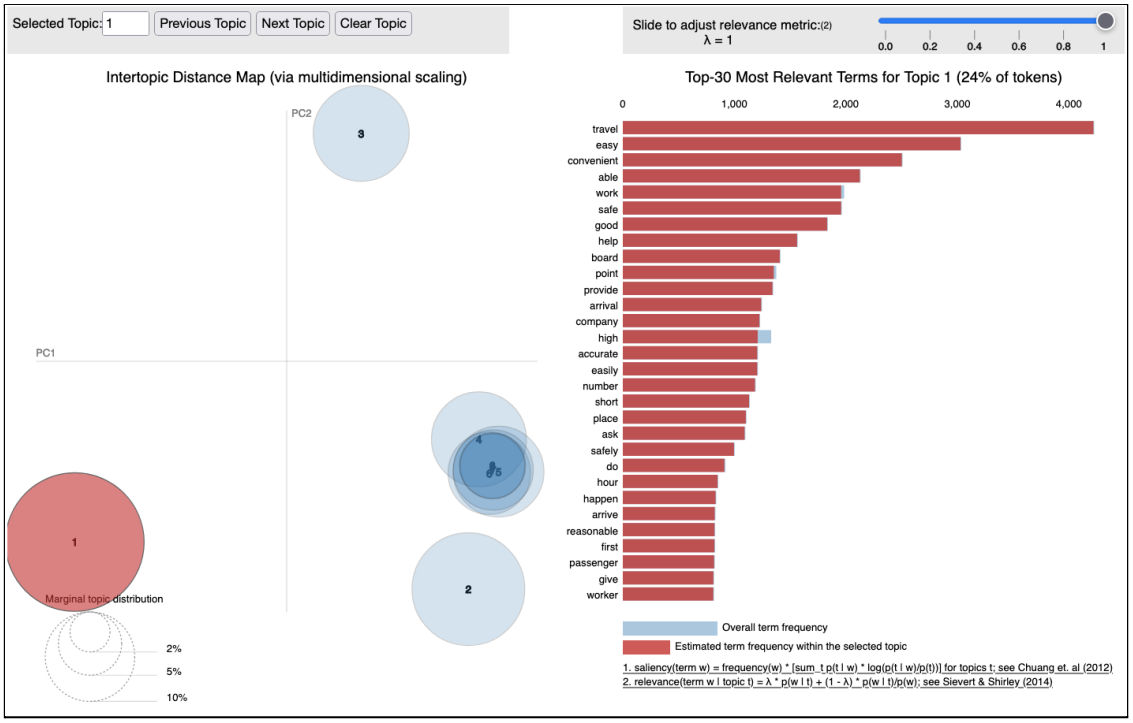


Figure 9: Combined Topic Modelling

2.4.1.2.1 Findings

After fine tuning the model to use the optimal number of topics and achieve its best coherence score (Appendix C), the team created a visualisation using pyLDAvis, a python package. Upon further inspection, the topics generally talked about users' experiences on public transport. The top 5 keywords from Topic 1 are ‘travel’, ‘easy’, ‘convenient’, ‘able’ and ‘work’. The top 5 key words in Topic 3 are ‘time’, ‘service’, ‘ride’, ‘wait’ and ‘destination’. The distance apart represents the relatedness of different topics. The greater the topic bubbles are apart, the greater the difference between the topics involved. For instance, Topic 1 talks about the ease of travelling whereas topic 3 discusses time taken while travelling.

2.4.1.2.2 Insights

As these topics such as time and fare are mentioned often, it can be inferred that these are common concerns for users when taking public transportation. Examining the various bubbles closer, the common keywords are ‘friendly’, ‘fast’ and ‘easy’ . It appears that users view the transportation in Singapore in a positive way, complimenting the various aspects of accessibility and service. Thus, cements the insight from sentiment analysis that there is a general satisfaction towards the current transportation facilities. Furthermore

A common limitation of real world data is that it may not always bring about useful insights. While the team managed to find out what people were talking about regarding public transport, these insights

are surface level since topic modelling can only derive the topics talked about. Furthermore, this information just reaffirms what CDG already knows. Thus, the need to explore other forms of analysis and use the topic modelling findings to complement the other insights.

2.4.1.2.3 Evaluation

To quantify the performance of the model, the team used topic coherence score. Topic coherence can be defined as the degree of significance between the words inside a topic in terms of how interpretable it is.The topic coherence measures employed in this study are designed to assess the quality of topics in a human-like manner. (Blair, Bi & Mulvenna, 2019).

A well performing model will generate topics that have the highest coherence score possible and a cohesiveness to each cluster based on human interpretation. After running the model multiple times, the final coherence score of the model is 0.364. Since the coherence score is not very high, the findings from topic modelling are not so definitive and substantiate the insight with other types of analysis.

2.4.2 Factor Analysis

Due to the high dimensionality of the survey dataset which was not easy to handle, the team decided to conduct factor analysis to reduce the number of columns the team originally had in the raw survey data. This data reduction technique helped to reduce the size of the dataset while maintaining the integrity of it (T, 2020). At the same time, it allowed the team to understand the extent to which each variable is associated with a common factor and uncover clusters of responses.

For data exploration, the team decided to split the dataset into 2 dataframes - Travel Patterns Dataset and Travel Perceptions Dataset as these 2 aspects answer the problem statement from different points of view. The team then conducted 2 statistical tests in evaluating the suitability of the dataset as derived from factor analysis. The team chose the few most recommended tests - Bartlett Test and KMO Test. The outputs from these tests are p-value and proportion of variance respectively. The table below shows the results of the adequacy test for the datasets. The Barlett's Test suggests that both datasets are statistically significant as correlation is present among the variables. Similarly, the results from the KMO test support the claim that the data is appropriate for factor analysis as the proportion of variance is considered adequate.

	Barlett's Test (p-value)	Kaiser-Meyer-Olkin (KMO) Test	Insights

Travel Pattern dataframe	3.1243425520312414e-189	0.5872836388962259	p-value indicated that it is statistically significant. KMO results is more than or equal to 0.6 which indicates that it is suitable to employ factor analysis
Travel Perception dataframe	0.0	0.7307595925352718	

Table 3: Adequacy test of dataset

The next step would be to initialize the factor analysis using a package called FactorAnalyzer. Methods such as visualising scree plot can help to decide on the number of factors. Table 3 shows a scree plot with the eigenvalues on the y-axis and the number of factors on the x-axis. The point where the slope of the curve is clearly leveling off indicates the number of factors that should be generated by the analysis (Rahn, n.d.). Additionally, one can look at the factors with >= 1 eigenvalues as the cut-off point for the number of factors. This is because a factor with an eigenvalue of 1 accounts for as much variance as a single variable. However a cut-off of 1 may result in leaving out a theoretically important factor whose eigenvalue is below 1. Hence, the team has decided to initialize the FactorAnalyzer with 20 factors based on the cut off value of 1.

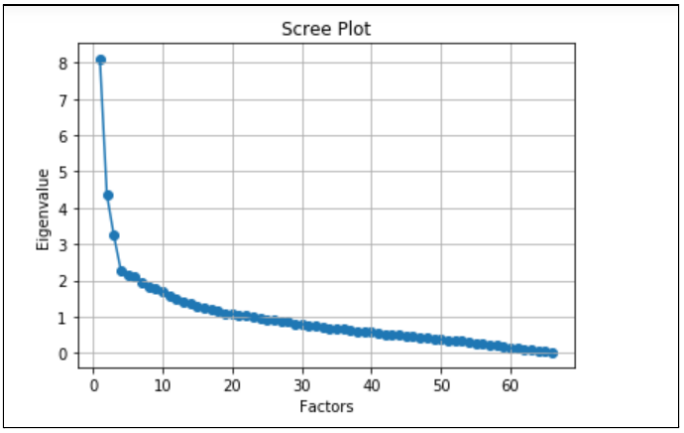


Figure 10: Scree plot of finding number of factors for travel perception dataframe

Next, the loading scores of the 20 factors are computed and exported into an excel file where the team will study the top few column variables that are most highly relevant to the factor identified. Loading scores can range from -1 to 1 with values close to -1 or 1 indicating that the variable strongly influences the factor and values close to 0 indicate that the variable has a weak influence on the factor. Hence, the team has identified 6 factors that are most interpretable despite choosing 20 as the initial

optimal number of factors as the team realised some of the factors did not have any high loading scores for any variables. The implementation of Cronbach Alpha's test checks the reliability of the factors (Howard, n.d.) and some of the alpha values of the factors are less than the acceptable range of alpha value more than 0.6 (Wati, Mahtari, Hartini & Amelia, 2019). With that, the team has identified the 6 different factors with alpha value more than 0.6 for the travel perception dataset as shown in Appendix D.

Similarly, factor analysis was conducted on the travel pattern dataset. However, the team could only identify 2 factors that were not as insightful and representative of the overall mobility demand as compared to the travel perception dataset. Therefore, the team concluded to proceed with travel perception only for clustering since travel patterns consist of two factors only.

2.4.3 Clustering Analysis

The aim of clustering is to understand if there are any present clusters amongst the population according to their demographics as well as perceptions as they have indicated in the survey. The team used 28 columns chosen under Factor Analysis to cluster the respondents into groups with similar characteristics. Three models were used to compare the results against one another, K-Modes (Cao), K-Modes (Huang) and Agglomerative Hierarchical Clustering.

K-Modes clustering refers to the unsupervised machine learning algorithm which clusters categorical variables. It uses modes instead of means and it calculates the similarities between the data points (Bonthu, 2021). There are 2 schools of method of K-modes clustering that were used in the analysis and they are Cao and Huang's methods.

Firstly, the team initialized the 2 elements separately and implemented the elbow method to find the optimal k number of clusters. While the elbow method can help to determine the optimal value of k, this cut off point may not be that obvious. Hence, the team decided to evaluate using silhouette score method as well. Silhouette Coefficient is useful in describing how similar the datapoint to other data points in its cluster relative to data points not in the cluster. Values closer to -1 will suggest that there is incorrect clustering, while values closer to 1 shows that each cluster is very dense, and hence a better clustering. Based on the implementation of these methods, the team decided to proceed with 3 as their final number of clusters to be used in the analysis.

Agglomerative hierarchical clustering refers to the unsupervised clustering method of assigning each point to an individual cluster and at each iteration. The algorithm merges the closest pair of clusters and this repeats until a single cluster is left (Sharma, 2019). To get the number of clusters for

hierarchical clustering, the team created a dendrogram, which is a tree-like diagram that tracks the sequence of splits. By observing the dendrograms, the team also plotted a vertical line with maximum distance and set the threshold in a way that it cuts the tallest vertical line whereby the number of vertical lines which are being intersected is the number of recommended clusters. In this case, the optimal number of clusters is 2. However, the team decided to evaluate both clustering models with 2 different numbers of clusters, 2 and 3 respectively in analysing how well the models performed.

Number of cluster	Type of algorithm	Silhouette Coefficient	Calinski Harabasz Index	Davies-Bouldin Index
2	Cao k-modes	-0.0128	2.94	8.57
	Huang k-modes	0.009164	5.87	6.62
	Agglomerative hierarchical clustering	0.684	1198.473	0.462
3	Agglomerative hierarchical clustering	0.714	1374.98	0.4000

Table 4: Evaluation of clustering models

The table above refers to the different evaluation methods for the clustering techniques. Methods include silhouette coefficient, Calinski Harabasz Index and Davies-Bouldin Index. While silhouette coefficient is used to measure the separation distance between clusters, Calinski Harabaz Index evaluates the models by finding the ratio of the sum of between-clusters dispersion and of inter-cluster dispersion for all clusters (Wei, 2020). On the other hand, Davies-Bouldin Index calculates the average similarity of each cluster with a cluster most similar to it (PyShark, n.d.). Hence, values with lower values indicating better clustering. Based on the results, the model using an agglomerative hierarchical clustering algorithm with 3 clusters gives the best clustering result. With that, the team will be analysing the insights using this model.

As most of the variables are in categorical values, it will be more suitable to visualize the differences in characteristics of each cluster in the form of a bar chart instead of a scatter plot. As a matter of fact, despite using the better performing model, the results of the clusters were not as significant. Nonetheless, there were some interesting insights on the factor of customer service as there are more

people who agree that customer service is important in Cluster B compared to those in other clusters who mostly feel neutral about it.

Hence, the team has decided to aggregate the data value of each factor for each row and compare the characteristics of each cluster using the aggregated values. Also, the team has decided to remove the factor of Convenience as the variables used under this factor mainly on rental cars and this was not representative of the overall mobility of transportation. Therefore, the team will be understanding the characteristics of clusters in terms of their perception using 6 factors and looking into each cluster's demographic. Table 5 refers to the overall insights while tables under Appendix E shows the breakdown percentage of respondents filtered by each factor and the extent of agreeability and consideration.

Factor	**Insights**
Affordability	83.2% of respondents have positive sentiment
Safety	99.1% positive sentiment
Customer Service	67.7% positive sentiment
Comfort	67.1% of respondent would not consider it as factor
Promotion	47.3% consider Promotion when taking mode of transportation
Accessibility	41.0% consider Accessibility

Table 5: Overall insights for all clusters

Hence, from the results and observations, the team has labelled the 3 different clusters according to the respective profile description. Those in Cluster A are labelled as comfort-focused respondents as they see importance in comfort when travelling and 58.2% of them are residing in the West area. Those in Cluster B are labelled as price-sensitive as 59.1% of them are staying in the North/North-east area and there are more users with negative sentiment on affordability suggesting that they feel more strongly about the cost of service. Those in Cluster C are experience-focused as promotion, customer service, accessibility are important factors to them based on the observations. The common description for all three clusters is that everyone feels positive about safety and almost half of respondents in each cluster have income less than $500.

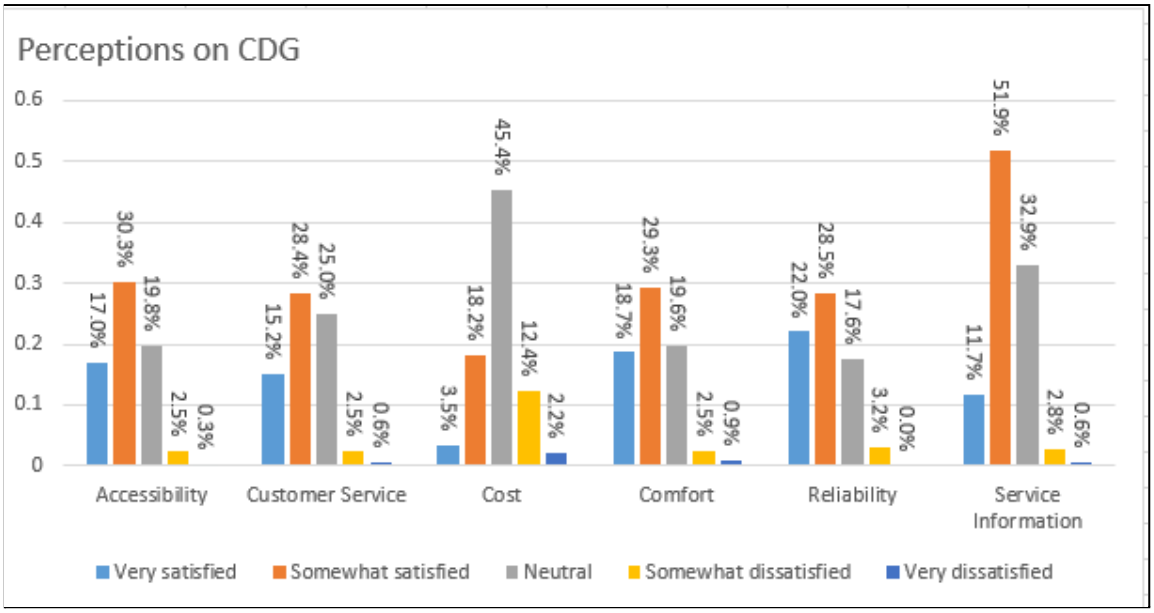


Figure 11: Table on percentage of respondents on CDG

Overall, the survey respondents' perception of CDG across all the above mentioned factors lie in the "satisfied" segment, except for the factor of Cost. It seems that 45.4% of the respondents actually indicated that they only feel neutral about this and the percentage of "dissatisfied" respondents is the highest recorded across all other factors.

Based on earlier mentioned analyses, the results actually indicate that the most popular factor when considering their mode of transport is actually price, while the survey respondents seem almost indifferent to the prices, or rather, do not have opinions on CDG's prices. There is still a gap for CDG to fill when it comes to sharing more about their costs so as to influence the target audience's perception on ComfortDelGro's cost.

2.4.3.1 Projection of Clusters

In order to understand the clusters further, the team performed a geospatial visualization of the respective clusters. As the evaluation of the clusters is based on the survey results, the implementation of applying raking to the data is the most appropriate method to project the clusters as opposed to the population distribution of the target audience.

Raking is the process of iterative proportional fitting where it adjusts the weight for each case until the sample distribution aligns with the population. Additionally, it is simple to implement as it only requires the marginal proportions for each variable used in weighting (Mercer et al., 2020). For the analysis, the team achieved the individual planning area weights by a ***division*** of the ***percentage of tertiary students in each planning area in Singapore*** and the ***percentage of the frequency***

distribution of the surveyors in each planning area. By applying the weights to the clustering result, the team will be able to attain the predicted number of tertiary students in Singapore in each respective cluster.

Through the following projection results as below, the team can possibly understand the planning areas that these clusters are situated in and the density of the population, which could eventually aid in evaluating targeted solutions for CDG in the future.

On a side note, based on the population dataset that are publicly available, the number of tertiary students in planning areas of *Downtown Core* and *Paya Lebar* have not been traced. Hence, these areas will be represented as 'Missing' in the maps. Although the data for tertiary students population in these areas were not collected this time round, CDG can potentially gain access to such information in the future and input it to the team's models for further results.

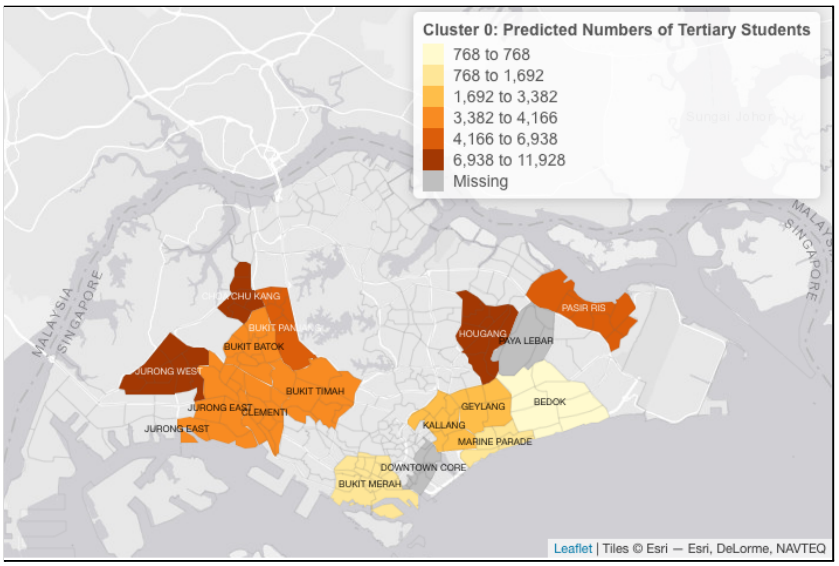


Figure 12: ClusterA: Comfort-focused

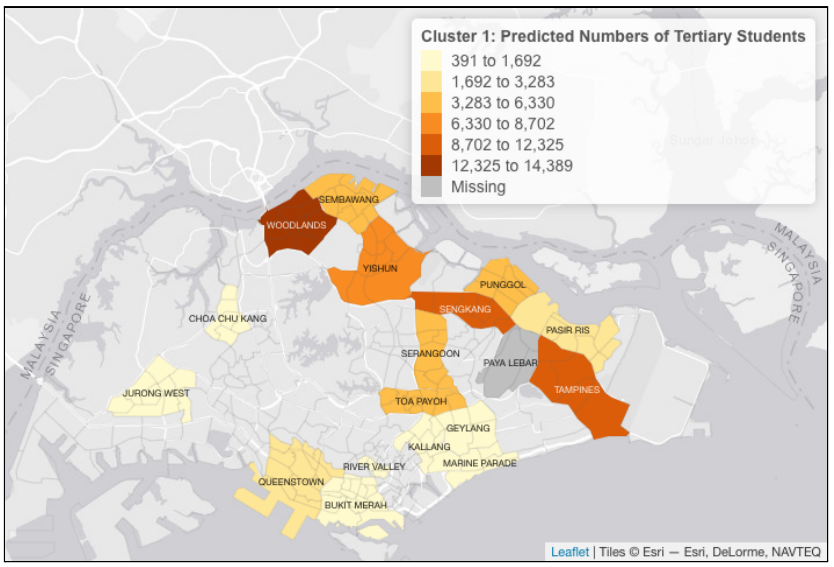


Figure 13: Cluster B: Price-sensitive

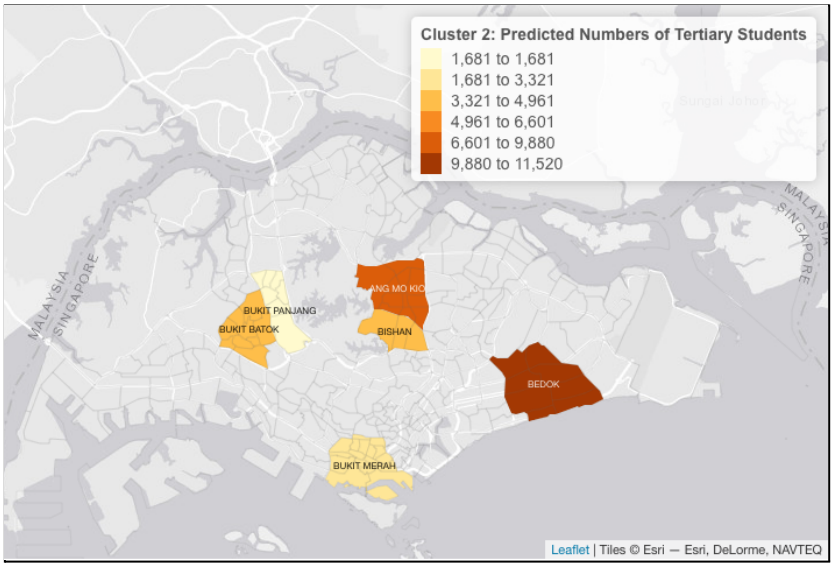


Figure 14: Cluster C: Experience-focused

2.4.4 Correlation Analysis

To identify the relationships between the factors and the different scoring from CSISG, the team conducted a correlation analysis with the aim of substantiating its findings to factor analysis.

Correlation analysis is the measure of strength and direction of relationship between two variables and is commonly used to identify relationships, patterns, or connections. Strength of relationship refers to the value of correlation coefficients ranging between +1 and -1, variables having stronger relationship the closer it is to +1 or -1 and weaker when it is closer to 0. The direction of relationship refers to the positive and negative sign of correlation coefficient. A positive sign indicates a positive relationship while a negative sign indicates a negative relationship.

2.4.4.1 Types of Correlation Analysis

There are three most common correlation analyses, namely, Pearson, Kendall, and Spearman correlation.

Pearson correlation analysis is a measure of the strength of linear relationship between two variables. The closer the correlation coefficient is to +1 or-1, the stronger the linear relationship. Correlation closer to 0 would indicate the variables having no linear relationship. The positive and negative of the correlation analysis also indicates the direction of the relationship. A positive correlation would represent an upward sloping line, while a negative correlation would represent a downward sloping line. Pearson correlations have five main assumptions. First, it assumes the variable to be normally distributed, which means variables should have a bell-shaped curve that is symmetric about the mean. Second, Pearson assumes the variables to have no significant outliers. As outliers are data points that do not follow the pattern of other data points, the presence of outliers would significantly affect the results from Pearson correlation. Third, it assumes that the variables are continuous data, which are numerical data types in the form of factions. Fourth, it assumes the variables to have a linear relationship, due to Pearson correlation being a measure of the strength of linear relationship. Lastly, Pearson correlation assumes homoscedasticity in the data. Homoscedasticity refers to how central the data is to the linear regression models, which means that the noise between independent and dependent variables are the same across all values of the independent variables.

Next, Kendall correlation analysis measures the strength of the dependence of two variables based on the concordance and discordance. Concordance refers to how similar data is being ordered the same way, and discordant is how differently the data are ordered in. It tests the similarities in how the data is being ordered when it is ranked. It is often used as an alternative to Pearson correlation when the data has failed one or more assumptions of Pearson correlation. It also serves as an alternative to Spearman when sample size is too small and has many ranks. Kendall analysis assumes the variables to be ordinal or continuous. Ordinal are ordered categorical data such as a satisfaction scale. It is also desirable for the data to have a monotonic relationship, which means to have both variables to be travelling in the same direction, when one variable increases, the other increases as well.

Lastly, Spearman correlation analysis measures the strength of the relationship between two variables. Spearman correlation assumes that the data must be ordinal and at least one variable to be monotonic to the other variable. Spearman correlation is also an option used for data that failed one or more assumptions of the Pearson Correlation.

2.4.4.2 Identifying the Right Correlation Methods

Before conducting the correlation analysis, it is first important to determine the right correlation methods to run based on the data. This would mean to evaluate each aspect of the data to determine which method is the best choice.

First, it is important to identify the types of data used in the analysis. The analysis uses two main variables, factors, and scoring. The factors used are questions asked in the CSISG survey, which was later manually classified by the team to fit similar factors identified in factor analysis. Examples of some factors are affordability, convenience, and comfort. These factors are the aggregation of the relevant questions, using the mean, which is based on a satisfaction scale from one to ten. The different scoring for CSISG are scores that were tabulated in its survey using weightage and other calculations. These scores are namely, customer expectations score, perceived overall quality score, customer satisfaction score, and customer loyalty score. The factors in this case can be identified as an ordinal data type while the scoring can be identified as a continuous data type.

Next, is to identify the distribution of data. This can be done by running a pair plot to view the scatter plot and histogram of each variable. Appendix G, Figure 48 shows the pair plot ran on factors and scoring for the train subsector. The histogram shows that while some data appears to be normally distributed, not all the data are. The scatter plot in Appendix G, Figure 49 shows a closer look at the relevant factors and the scoring. It is observed that the data distribution does not follow a linear pattern, nor does it follow any other pattern due to factors being an ordinal data.

After identifying its overall data type and distribution, the team can evaluate the appropriate correlation analysis to run. Firstly, Pearson would not be a suitable correlation to run as the data violated multiple assumptions. Pearson correlation requires the variable to be continuous data, however, only one of the variables is a continuous data type while the other is an ordinal data type. Pearson also assumes linearity and normal distribution of data, which the data both did not have. Since Pearson correlation is not suitable, alternatives must be considered, either Kendall or Spearman methods. While both methods are alternatives to Pearson, it is important to note what Kendall and Spearman are measuring. Kendall measures the strength of dependence between two variables while Spearman measures the strength of relationship between two variables. Since the aim of the analysis is to discover if there is any relationship between the two variables, the Spearman method would be more suitable in this aspect. Next, the type of data must be taken in consideration to choose the suitable method of analysis. As the data types are continuous and ordinal, the Spearman method would be better. According to a journal published by Harry Kharmis, if the variables are continuous and ordinal, Kendall would be a better choice. However, if the ordinal data has many ranking levels,

of six or more, Spearman should be considered instead (Khamis, 2008). Since the factors are an aggregation of the satisfaction scoring from one to ten, there are more than six factors. This would suggest that Spearman would be a better choice for correlation analysis.

2.4.4.3 Findings

As shown on Appendix G, Table 13, it depicts the overall correlation analysis results for each of the subsectors. Firstly, the main scoring the project should focus on is the customer satisfaction score. Overall, the correlation results across all subsectors are positive, which means that customer satisfaction scores tend to increase as average factor scores increase. However, the degree of association differs across different subsectors and factors.

For the bus subsector, customer satisfaction has an overall positive correlation result, with convenience having a higher correlation of 0.73 compared to the rest has a moderately high correlation around 0.48 to 0.56. It can be inferred that the convenience has a higher degree of association to customer satisfaction and a higher monotonic relationship.

For trains, the overall correlation results in this subsector are generally moderately high, ranging around 0.5 to 0.67, with convenience being the top, followed by safety and customer service. While the correlation results do not show higher relationships like the bus subsector, convenience is a common factor that has shown slightly higher correlation in both subsectors.

For taxis, while the results are all positive for customer satisfaction, the results are generally on the lower side, with the factors having the highest correlation being customer service with 0.51.

For booking apps, the results are also on the lower side despite it all having positive relationships, with customer service having a 0.43 correlation and convenience having a 0.41 correlation. As these two factors have a lower correlation result, these results are said to be a more neutral relationship to customer satisfaction.

2.4.4.4 Insights

In general, public transport such as buses and trains have a higher correlation to convenience. This would suggest that convenience has a higher relationship, hence higher possibility of convenience being one of the main factors consumers take into consideration. Looking back at the factor analysis, safety was one of the factors with a 99.1% positive sentiment, which is also the second highest factor with a higher correlation result. Thus, both safety and convenience are the main factors that affect consumer decisions when taking public transport.

Customer service has a moderately high correlation result, which means that it does not have a strong relationship to customer service. Further look at the factor analysis shows that customer service has a 67.7% positive sentiment, which is not high, but it is still high enough to be a factor to consider. It is also interesting to note, while 83.2% of respondents show positive sentiment in the factors analysis, it has the lowest correlation results against the customer satisfaction score CSISG. This might suggest that consumers do not think negatively of the current pricing for taxis, however, they are still not satisfied with it.

Booking app shows the lowest correlation results, with the top 3 factors of service information, convenience, and safety being in the 0.39 to 0.44 range. Furthermore, both service information and convenience are not a factor covered in the factor analysis due to the lack of these questions in the survey, there is not enough evidence to conclude that both factors are important to consumers in the taxi subsector. However, from these three factors, it can be concluded that consumers generally are slightly more anxious when using booking apps, as these three factors all point towards consumers trying to get as much information to feel safe while getting to their destination on time.

In summary, safety is a prevalent factor through the subsectors in correlation, factor analysis, and clustering. It is a factor prevalent throughout all analysis. While the majority of the survey respondents are indifferent to the safety of transportation, it is an important factor to take into consideration due to how prevalent it is.

2.4.5 Geospatial Analysis

To further comprehend the impact of factors and the varying scoring from CSISG, the team performed a geospatial analysis to visualize the differences and gaps based on a geographical segmentation. The main purpose of this analysis is to aid in identifying phenomenons or possible gaps due to changing spatial conditions or location-based events.

Geospatial analysis is an approach to applying statistical analysis and other analytical techniques to data which has a geographical or spatial aspect. In this aspect, the team will evaluate the analysis results based on the segmentation of the planning areas in Singapore.

The analysis will be split into 2 separate portions, an overview of the respective scoring and the comparison of the prevalent scoring relating to customer expectations and satisfaction.

2.4.5.1 Overview of CSISG scoring

The CSISG scoring is based on 5 components, namely Customer Expectation, Customer Satisfaction, Customer Loyalty User Trust, Perceived Overall Quality and Perceived Value. An in-depth definition of the respective scores could be referred to in Appendix I, 11.9.1.

To attain a clearer understanding of the various scores, the team will visualize the overall findings by the individual service sector, specifically Bus, MRT/LRT, Taxi and Taxi/Car booking application. Although the project covers the aspect of the car rental service in Singapore, the current CSISG dataset does not possess data relating to it. Hence, the team will rule out the aspect of that particular service for this analysis. The comprehensive breakdown and findings of the various service sectors could be referred to in Appendix I.

Overall, based on the results of the geographical outputs, the team can identify that the customer satisfaction score of a particular planning area is analogous to customer expectation, customer loyalty user trust, perceived overall quality and perceived values scores. This could be predominantly seen in the planning area, *Tanglin*. Taking into consideration the bus service sector as an example, when *Tanglin* has a high satisfaction score, the other scores would similarly be in the same scoring range as the former. Likewise, for those planning areas with a relatively low satisfaction score, the other scores would demonstrate a similar nature.

2.4.5.1.1 Insights

Firstly, the scores of each planning area could possibly be affected by the availability of a particular service in its respective planning area. This could be implied from the results of the individual visualizations of the Bus and MRT/LRT service sectors, where areas with the presence of at least 1 MRT/LRT station are seen with higher scoring as compared to other areas which do not. Additionally, the availability of buses is high and prevalent around the perimeters of a MRT/LRT station in Singapore, which could further imply that the effect of this phenomenon is viable.

Next, among changes of the planning areas over the years from 2016 to 2019, there has been a significant growth in their scoring, specifically areas like *Jurong West*, *Punggol*, *Yishun* and *Sengkang*. These planning areas are primarily in the perimeters of Build-To-Order (BTO) HDBs during that period (The world of Teoalida, 2021). With the influx of new residents, it would possibly affect their level of scoring due to the availability and accessibility of the mobility service. Taking *Jurong West* as an example, it could be seen that its scoring has increased over time after the launch of the extension of the North-South and East-West MRT lines in 2017 (Land Transport Authority, 2017). Moreover, when there are new launches of BTOs and MRT stations, new bus routes would be implemented for

the convenience of the public. Correspondingly, service sectors such as Taxi and Taxi/Car booking application, have seen a spike in increase of customer expectation scores in these planning areas, specifically *Sengkang* was identified as an outlier in 2019 with a score of 95.9. Therefore, this could possibly imply that the effect of new BTOs and MRT stations in a particular planning area can affect its respective scores.

2.4.5.2 Expectations vs Satisfaction

As customer satisfaction scores are pertinent to the others, the team could observe the differences between the target audience's level of expectations and satisfaction to further understand if there are any planning areas that are underperforming and why they are.

In order to obtain a difference between the expectation and satisfaction scores of each planning area, a simple cross reference was conducted to identify planning areas with higher expectation scores as compared to its opposing satisfaction scores.

With the positive difference of those affected planning areas, the team will be able to visualize and identify the top 3 areas that did not meet its expectations and gain a deeper understanding of the cause through a correlation analysis of its underlying factors.

2.4.5.2.1 Service Sector: Bus

As reference to Figure 15 below, the team can identify that planning areas such as *Ang Mo Kio*, *Yishun*, and *Bukit Merah*, are among those that did not meet the target audience's expectations of the Bus service sector. Due to the following, the team will delve into understanding why this phenomenon takes place by utilizing correlation scores of the planning areas and factors identified in the prior sections. Based on the correlation result, the factor of safety for these planning areas has the highest correlation suggesting that those who stay in that area place great importance in the safety of bus transportation.

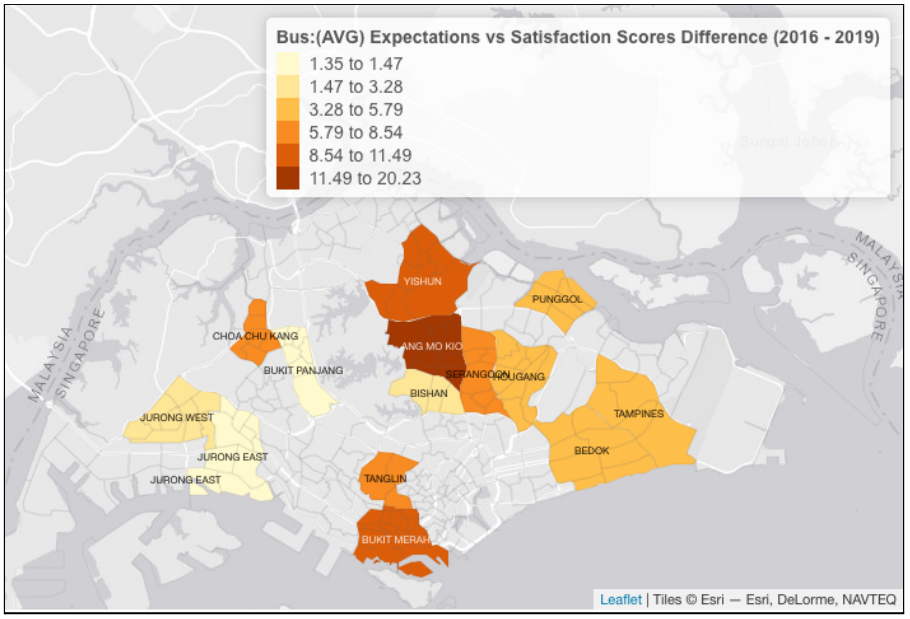


Figure 15: Bus: Average Difference of Expectations and Satisfaction scores

2.4.5.2.2 Service Sector: MRT/LRT

The team observed that planning areas like *Novena*, *Bishan*, *Kallang* are among those that did not meet the target audience's expectations of the MRT/LRT service sector based on Figure 16 and decided to understand deeper on the correlation of these factors in the planning areas. However, due to the insufficient data points, the team was unable to generate useful insights with regards to this sector.

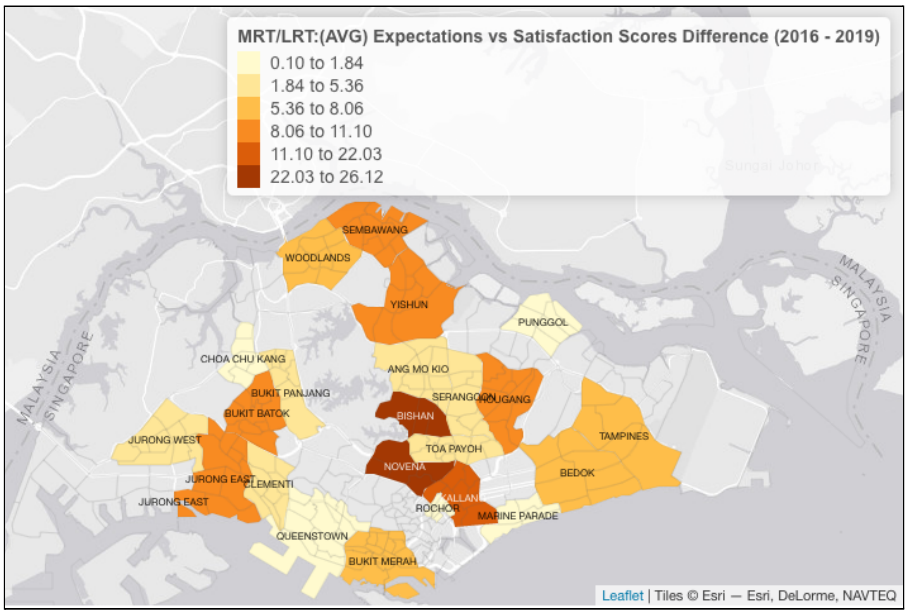


Figure 16: MRT/LRT: Average Difference of Expectations and Satisfaction scores

2.4.5.2.3 Service Sector: Taxi

Jurong East, *Bukit Merah*, *Sengkang* are the planning areas that the team looked into the factors deeper. The team observed that the top positively correlated factor with satisfaction score is safety, affordability and service information indicating that those living in these planning areas are concerned about these factors that can affect their demand and perception for the taxi service.

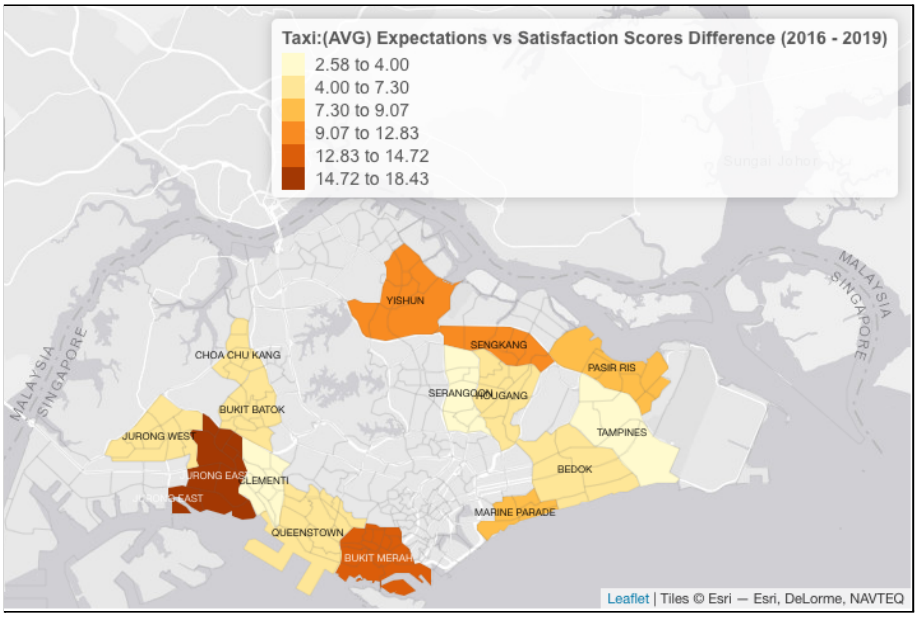


Figure 17: Taxi: Average Difference of Expectations and Satisfaction scores

2.4.5.2.4 Service Sector: Taxi/Car Booking Application

The most highly correlated factors with satisfaction score for those living in *Jurong East*, *Serangoon*, *Hougang* is mainly service information suggesting this factor would greatly affect their perception for this mode of transportation.

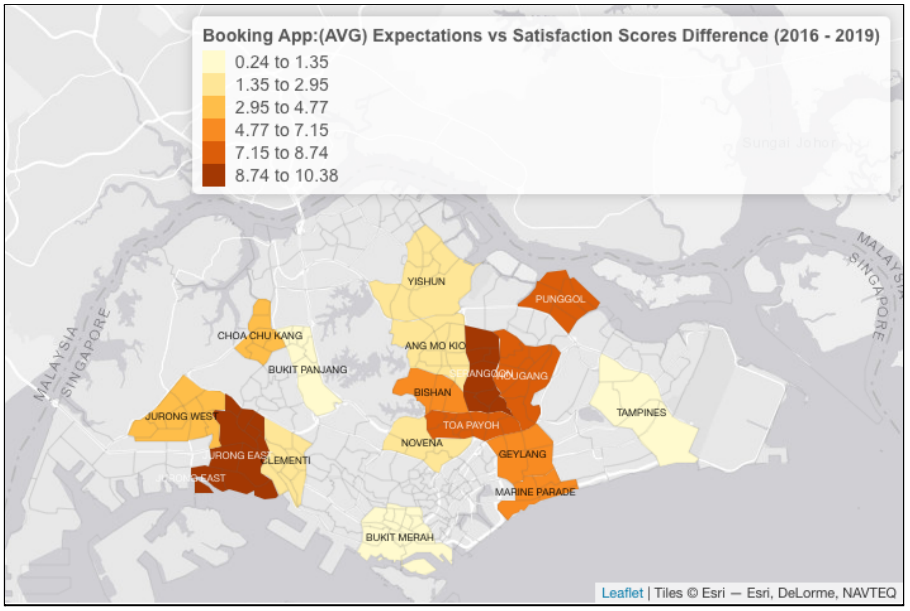


Figure 18: Taxi/Car Booking App: Average Difference of Expectations and Satisfaction scores

2.5 Overall Insights and Recommendations

2.5.1 Insight 1: Safety is a prevalent factor

As understood from the clustering analysis, the factor of "Safety" rarely appears on a respondent's top few factors for consideration of preferred mode of transport. However, Safety is a prevalent factor whereby it is commonly raised across the various analytics the team has done, indicating that it is still a very important factor to the target audience and it would still greatly impact their "demand" in the mobility market.

2.5.2 Insight 2: Positive sentiments but still sensitive to various factors

Through the text analysis conducted on the Reddit and Survey data, the target audience has shown that they generally have sentiments that fall in the range of happy to neutral on the topic of Singapore Mobility Market. However, they should not be mistaken to be inelastic to the various factors that might potentially impact their opinions or even their travelling behaviour. In fact, further analysis has also shown that the target audience is still relatively sensitive to cost of services which entails factors such as affordability and promotion. At the same time, each identified cluster, through clustering analysis, has a different relationship with the factors, and therefore has varying important factors that might affect their "demand".

2.5.3 Insight 3: Low pick-up rate of car rental service

Through the survey, the team observed that the majority of respondents are aware of car rental services in Singapore. However, out of all the respondents, only about 11% have rented a car before.

On top of that, the team conducted further focus group sessions with those who have rented cars before and realised yet again that the majority of the interviewees have in fact rented the car for only 0 to 5 times, indicating that they were very infrequent users. In fact, these focus group participants indicated that they only used these car rental services as they were just curious about it, and not because they had a strong need for a car rental service. This indicates that there is in fact a low pick-up rate, as well as a low retention rate although the existing companies have managed to raise awareness amongst the tertiary students.

2.5.4 Recommendations

Understanding the target audience preferences is an ever-changing answer and analysis has to be conducted regularly, so is social listening to understand their needs and wants. As there is no one size fits all solution, there is therefore a need to cater to each individual identified clusters' needs and priorities so as to maximise the results of their efforts.The current efforts into marketing car rental services have indeed achieved their goal of brand awareness, but more promotional efforts have to be put into ensuring that the target audience will follow through with their interest in the service, and actually consume it.

3.0 Solution Approach

3.1 Solution Design

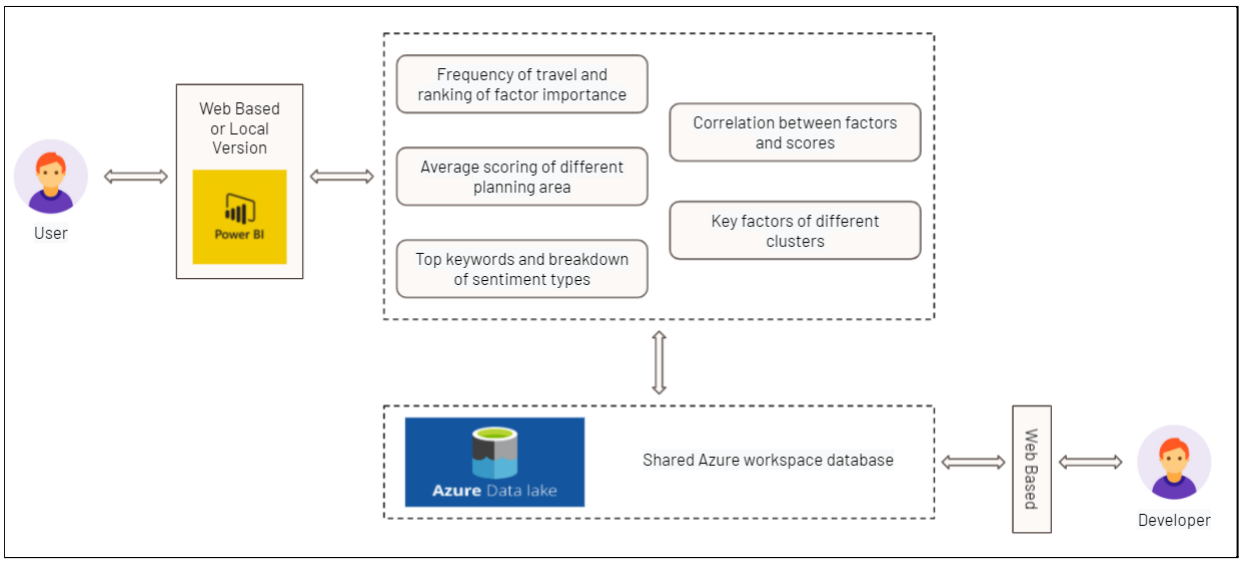


Figure 19: Solution Design Diagram

3.2 Product Features

The proposed solutions enables the users to:

1. Identify ratings of different planning areas
    - To allow users to view score ranking of different planning areas based on sub sector and different score metrics affecting demand (e.g. Quality, Value, Loyalty, Expectation and Satisfaction) based on CSISG Survey.
    - This will allow users to understand or potentially explore certain planning areas to work on company branding or look into service quality that scores lower.

2. Identify ranking and sentiments of keywords
    - To allow users to identify keywords ranking of different sub sectors and type of sentiments that people are currently talking about.
    - Based on the keywords identified, users could discover possible gaps by looking into negative sentiment types or areas for improvement that people have positive sentiments on.

3. Understand correlation between factors and scoring
    - The correlation scores allows users to find out the impact of different factors identified from the survey on score metrics in CSISG Survey.
    - Based on the ranking of correlation scores of different factors, the user would be able to learn the most important factors that affect different demand-influencing scores. From there, the user could look into different factors (e.g. Safety, Comfort, etc.) that could potentially increase demand due to its correlation.

4. Analyse key factors affecting different clusters
    - By filtering the survey factor breakdown chart, the user will be able to identify the cluster that has the strongest opinion on factors affecting demand (e.g. Affordability, Customer Service) of different sub sectors.
    - Based on the identified cluster, users could point out planning areas with greater population density and plan or improve the identified factors to increase demand.

3.3 Use Case Diagram

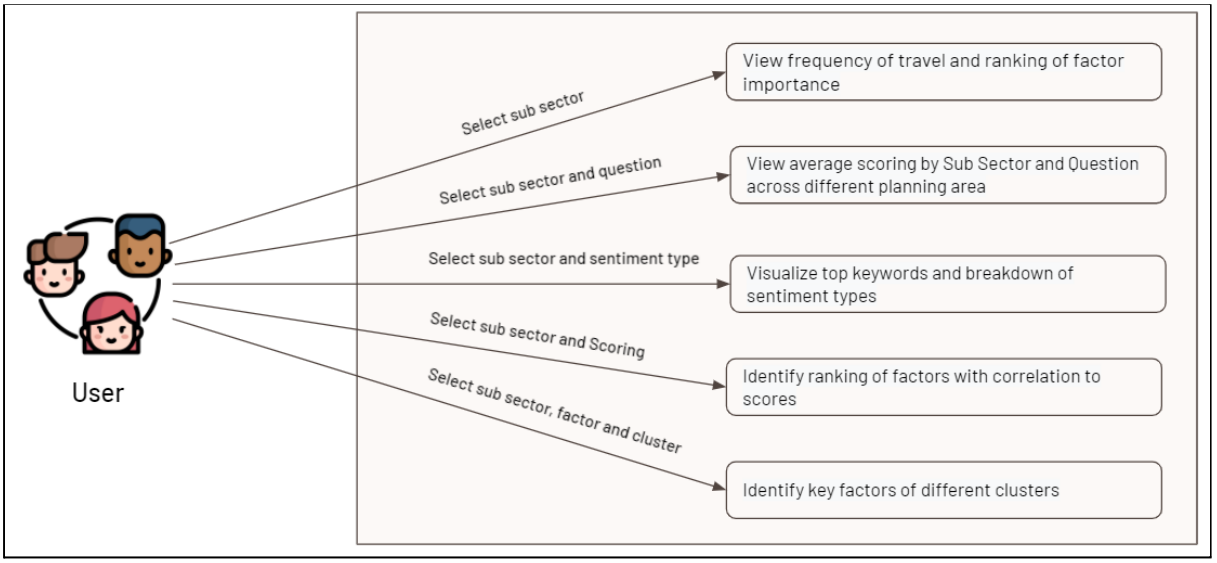


Figure 20: User Scenario 1

3.4 User Scenarios Diagram

Figure 21 below shows the user scenario of viewing frequency of travel and ranking of factor importance.

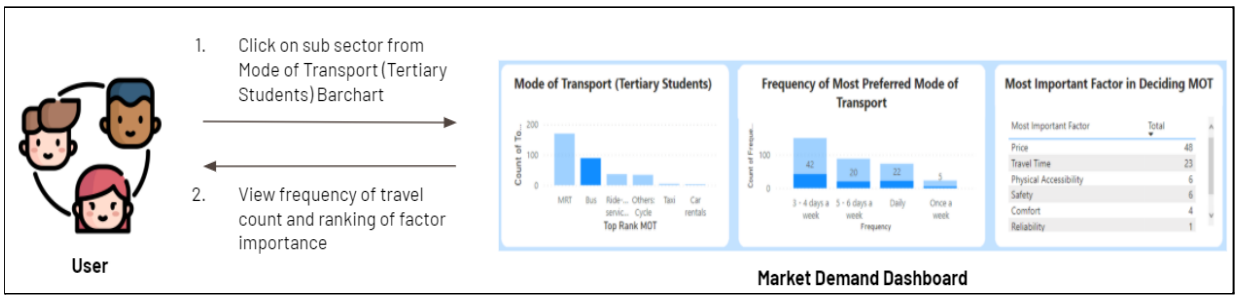


Figure 21: User Scenario 1

The second scenario refers to users viewing average scoring by sub sector and question across different planning areas.

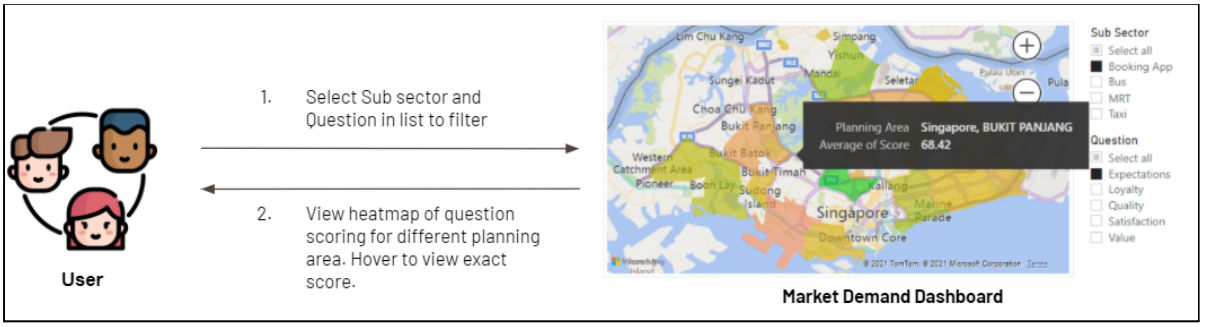


Figure 22: User Scenario 2

The third user scenario shows how they can visualize top keywords and breakdown of sentiment types.

Figure 23: User Scenario 3

Users can also identify ranking of factors with correlation to scores.

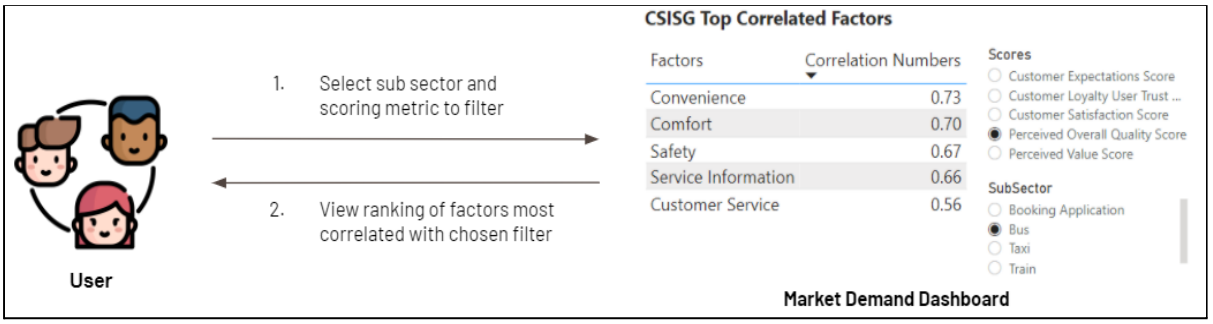


Figure 24: User Scenario 4

Lastly, they can interpret and understand how the different key factors affect each cluster.

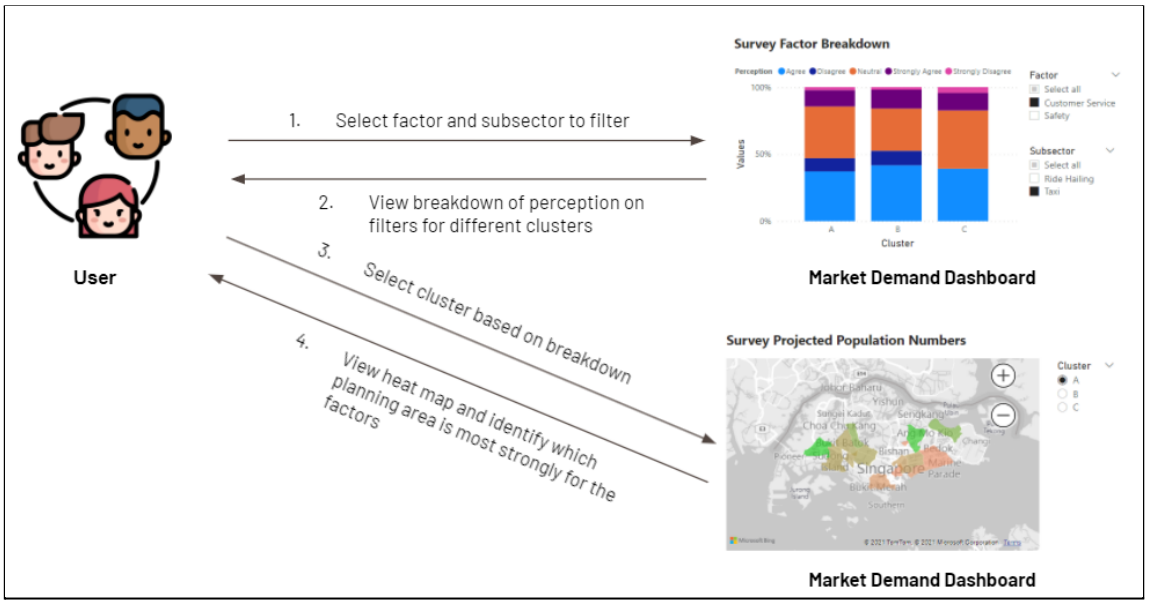


Figure 25: User Scenario 5

4.0 Technical Details

4.1 Software, Tools and Framework

Programming Languages	
Python Language 	Python was the main language used in the project. It is used for web scraping, data cleaning as well as for analysis.
R Language 	The team used R language for geospatial analysis.
Database	
CSV and XLSX files / Azure 	The data the team collected were saved in either CSV or XLSX file type and the files would be later transferred over to the Azure database set by the client.
Software and Tools	
Google Drive 	The team used Google Drive to facilitate collaboration for slides creation, report writing and recording minutes. It is also used to archive and save documents for easier sharing.
Github 	The team created a repository and was able to share and collaborate on different files (e.g. codes, data and dashboard).
PowerBI 	PowerBI is used to create dashboards for the clients to gain insights from the visualizations.
Frameworks and Libraries	
NLTK 	Natural Language Tool Kit was used in sentiment analysis to extract sentiments from text.
Pandas 	Pandas was used to create data frames and arrays to help structure the data.
Matplotlib 	Matplotlib was used to visualise plots such as histograms and matrix for better understanding of the datasets.

Scikit Learn	Scikit Learn was used to compute statistical formally foranalysos just as correlation analysis.
Keras	Keras was used for its artificial neural network for analysis such as sentiment analysis.
API & Other Tools	
Praw	To extract data from Reddit for analysis.
IGramscraper	To extract Instagram data like comments, captions and likes of accounts for analysis.
Twint	To extract data from Twitter.
Selenium	To extract data from Facebook and SGCarMart.
OneMap API	The team used the API to convert postal codes to Longitude, Latitudes and retrieving planning areas for geospatial analysis.

Table 6: Technical tools and details used

4.2 Application Architecture

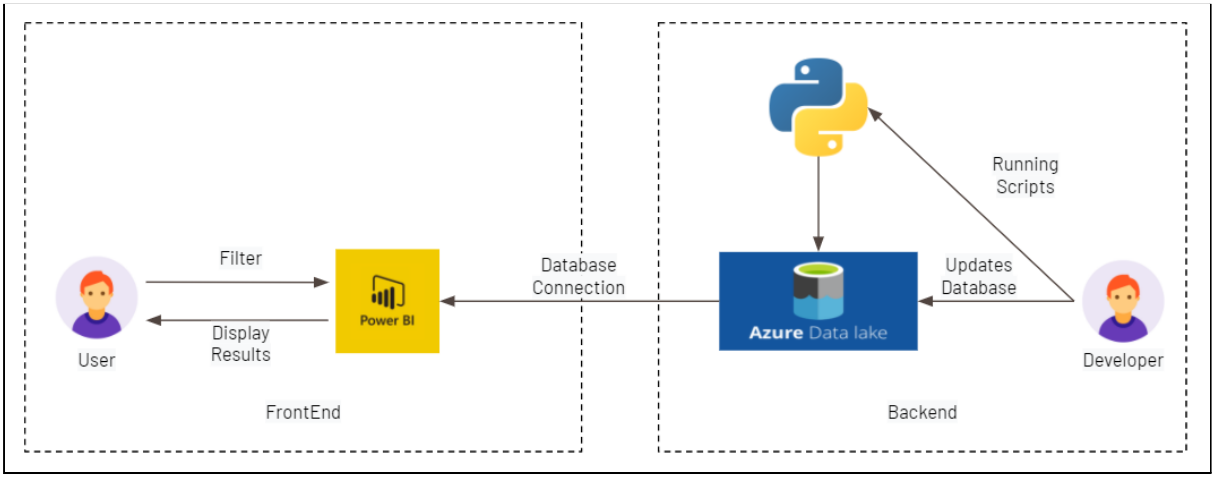


Figure 26: Application Architecture

4.2.1 Frontend

The user would need to refresh the data source that is connected to the Azure Database. Then, the user could view the latest updates visualization to find new insights.

4.2.2 Backend

The developer would be in-charge of running the scripts to generate new dataset and update the current database for the frontend user to be able to view updated visualizations.

4.3 Data Challenges

4.3.1 Insufficient or Irrelevant Data

The group encountered many irrelevant data and it was hard to find data that would be able to fulfil the project needs.

Data such as GPS data was insufficient to help identify transport demand in Singapore, it tells little more than the routes people took to travel around. Other data such as data scraped from social media (like Facebook, Twitter, Instagram, and Telegram) and publicly available data such as car park data was irrelevant and added little value to the project.

4.3.2 Readdressing the problem statement

The initial problem statement was not sufficiently defined thus, the team needed to relook at the available data that can be used for additional analysis needed to address the problem statement. To address this challenge, the team met up with the client and the team's project supervisor to expedite the change as soon as possible, with an internal meeting to discuss and decide on it.

4.4 Technical Challenges

The team had difficulties scraping Facebook due to limitations of Selenium API and stricter Facebook policies. Thus, the team was unable to scrape comments and had to look into other social media sites such as Reddit to collect data.

In addition, the initial analyses on scraped data from sources that were ultimately eliminated afterwards such as Twitter did not bring about relevant information or even insights. Therefore, the team had to refocus on obtaining data that will be applicable to the analysis and problem statement. At the same time, this resulted in the team having to re-evaluate the analysis methods to be used as well.

Lastly, the team did not have prior experience to scraping the different social media sites and had to research and experiment with different tools and libraries. This resulted in an increase of time spent on studying them at the initial phase of the project.

5.0 User Acceptance Testing (UAT)

5.1 UAT Goals

The aim of UAT is to ensure the dashboard is usable, intuitive and that the interface is user-friendly. The users should be able to complete all tasks independently and test out the various features of the PowerBI dashboard. Finally, UAT is also to identify any areas of bottleneck.

5.2 Methodology

The UAT was conducted with CDG Data Team that consisted of 4 members. Due to time constraints and Covid-19 measures, the UAT was conducted online and the team had a few days to complete the test on their own accord.

The data team had to record down any observations, additional feedback, problems they had faced and the time taken to perform each task.

5.3 UAT Tasklist

The team sent over this excel sheet shown in Figure 27 for each of the users to carry out and provide their input.

S/N	Test Case Type	Dashboard Tab Name	Description	Test Step	Observations	Time Taken
1	Functionality	Read Me	Read the information regarding the dashboard in the read me tab	Users to read the read me and to understand the information regarding each tab and what insight each chart can offer		
2	User Experience	All tabs	Tabs can be accessed	Switch between the various tabs (i.e. Overall, Perception) and note down if this function can be performed		
3	Usability	All tabs	Visualisations can be viewed with no errors	User should be able to view the visualisations without any errors from the dashboard. Please and note down if there are any errors identified. (i.e. data not found)		
4	Design	All tabs	Colours of visualisation are appropriate and font size of text are acceptable	User should be able to read the dashboard comfortably and note down if this function can be performed		
5	Functionality	All tabs	Users should be able to understand what each tabs is about	User to click onto each tab and summarise any key insight on the purpose of each tab in one sentence. Please note these down in the observations		
6	User Experience	Overall	Visualisations can be inspected	User to zoom in and out on the map under 'CSISG Scores' and note down if this function can be performed		
7	Functionality	Overall	Data can be filtered to update visualisations	User to filter the map of 'Average CSISG Scores' to view the perception scores of the Sub-Sector "MRT" on the question "Quality"		
8	Functionality	Overall	User can find out the frequency of transport with highest proportion of survey respondents	User to look at the overall to determine the frequency of transport with the most percentage of users		
9	Functionality	Perception	Data can be filtered, Information is clearly visualised	Users to click on negative sentiments to find the top 3 word for overall subsectors		
10	Functionality	Perception	Word cloud can be filtered by both filters	Users to click on Taxi in the subsector filter and Positive Sentiments in the sentiment filter to view the wordcloud and donut chart and note down if this function can be performed		
11	Functionality	Demand Influencing Factors Tab	Data can be filtered in terms of clusters and distribution of cluster visualised clearly on map	User to hover over any planning area and identify its corresponding population size for Cluster C		
12	Functionality	Demand Influencing Factors Tab	Users should be able to derive insights from the different charts	Users to click on the Affordability factor for Bus subsector to find out the cluster with the most Strongly Agree		
13	Design	Demand Influencing Factors Tab	Users should be able to see the difference in population at a glance	User to identify the planning area with highest population size in Cluster A		
14	Usability	Power BI Dashboard	Share dashboard with others	User to open and view the dashboard and note down if this function can be performed		
15	Usability	Power BI Dashboard	Download dashboard	User to clicks on '...' on top right hand corner and look for the 'Download' button and note down if this function can be performed		

Figure 27: UAT Task List

5.4 UAT Results & Insights

All the team members were able to manage to complete all tasks the team designed and completed the test cases in the stipulated time.

The team has consolidated all the feedback and observations by the data team.

1. All users spent the most time on the Readme tab and have given the feedback that it is too wordy.
2. The heatmaps' colour was misleading as a colour gradient scale of Red (highest) to Yellow (lowest) was used. The recommendation proposed by the clients was to follow a traffic light rule based scale of Green (highest) to Red (lowest).
3. Word cloud did not have the option of allowing Users to "Select All" and requires users to deselect all to view overall keywords.
4. Users had a hard time identifying the ranking of keywords as they are of similar sizes.

<u>Metrics</u>

Time Given : ~50 minutes

Average Time Taken of CDG Team: ~39 minutes

5.5 Improvements Made

The Readme tab was simplified and definitions were added to give context.

To make the heatmap more intuitive, the gradient scale of Red to Yellow is converted to a rule based scale of Green (Highest) to Yellow (Medium) to Red (Lowest).

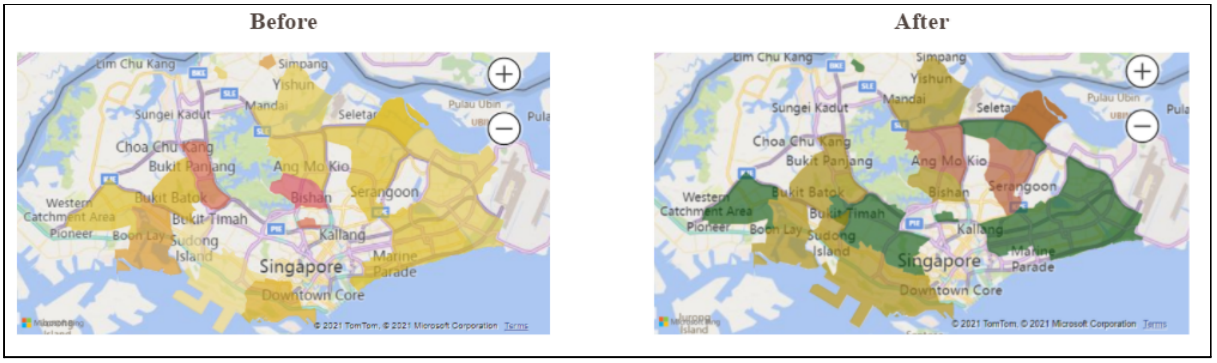


Figure 28: Changing gradient colour scale to rule based colour scale 1

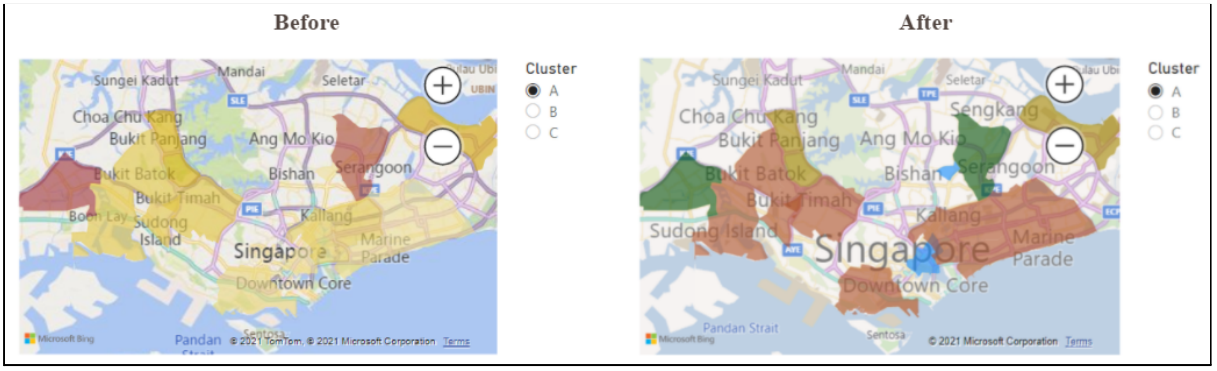


Figure 29: Changing gradient colour scale to rule based colour scale 2

Added option for users to "Select All" to view overall word cloud for all sub sectors.

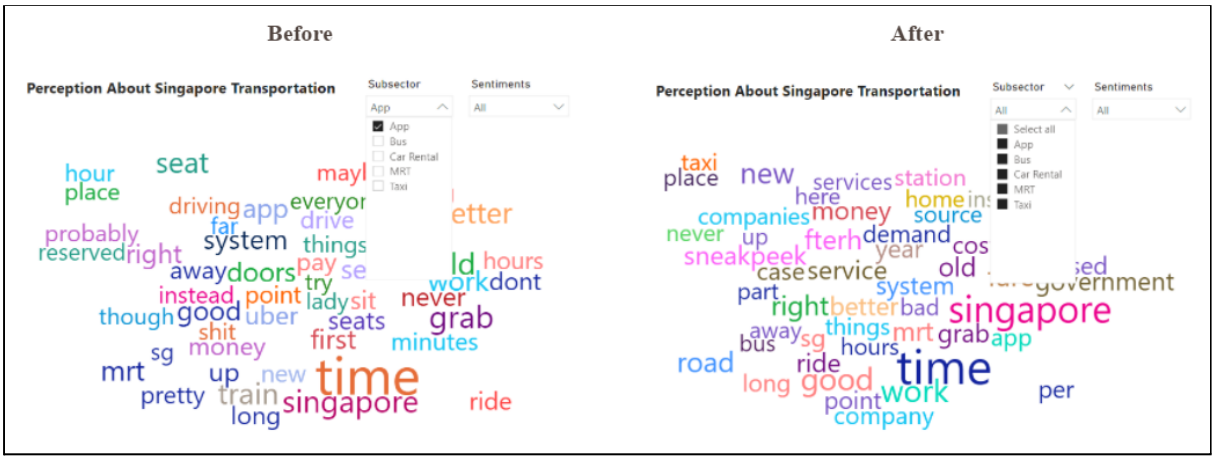


Figure 30: Before and after adding "Select All" Option

Changed the sizing scale so words have a greater distinction and users can better identify top keywords.

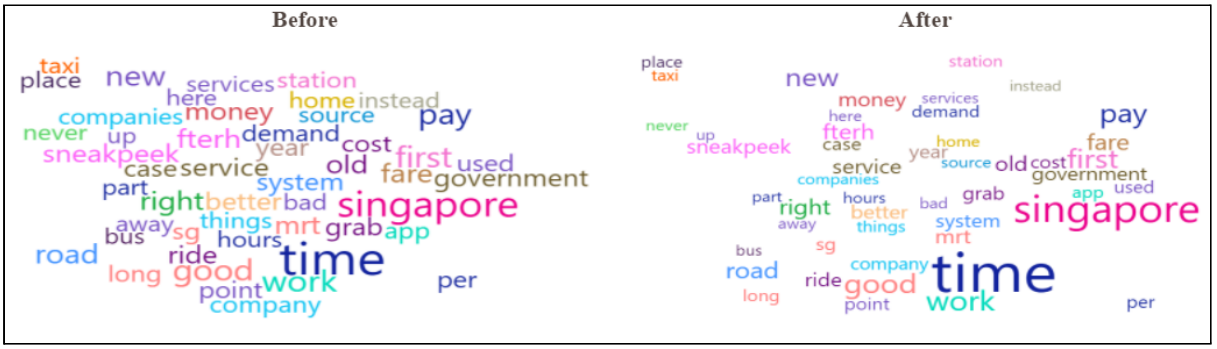


Figure 31: Resized word cloud

6.0 Final Product

6.1 User Interface

6.1.1 Read Me Tab

Users will be able to get a summary and explanation of the visualizations available in the different tabs.

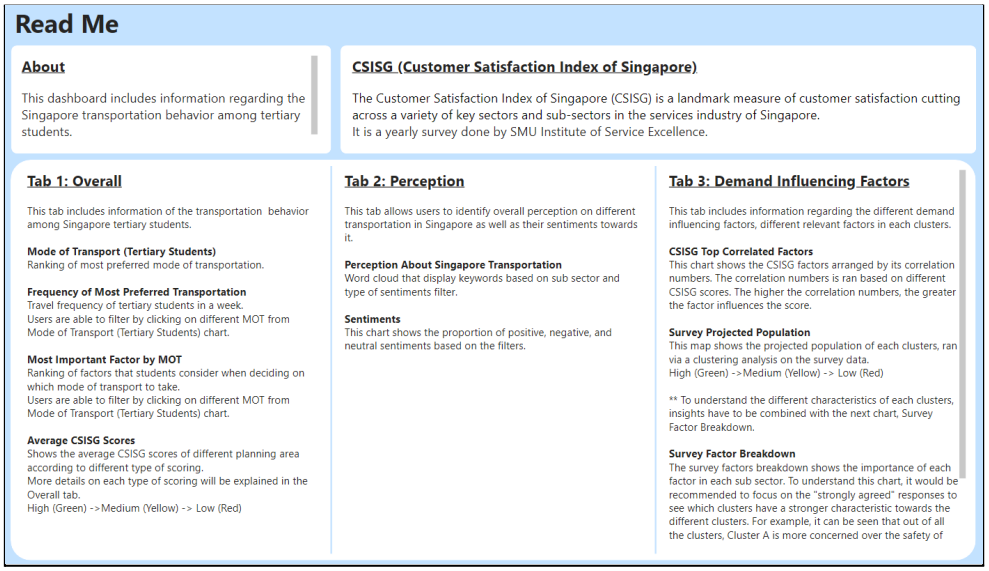


Figure 32: Dashboard Read Me Tab

6.1.2 *Overall Tab*

The user could view the overall ranking of Mode of Transport, frequency and ranking of factors in deciding Mode of Transport. By filtering sub-sectors and question types, users could identify ratings of different planning areas.

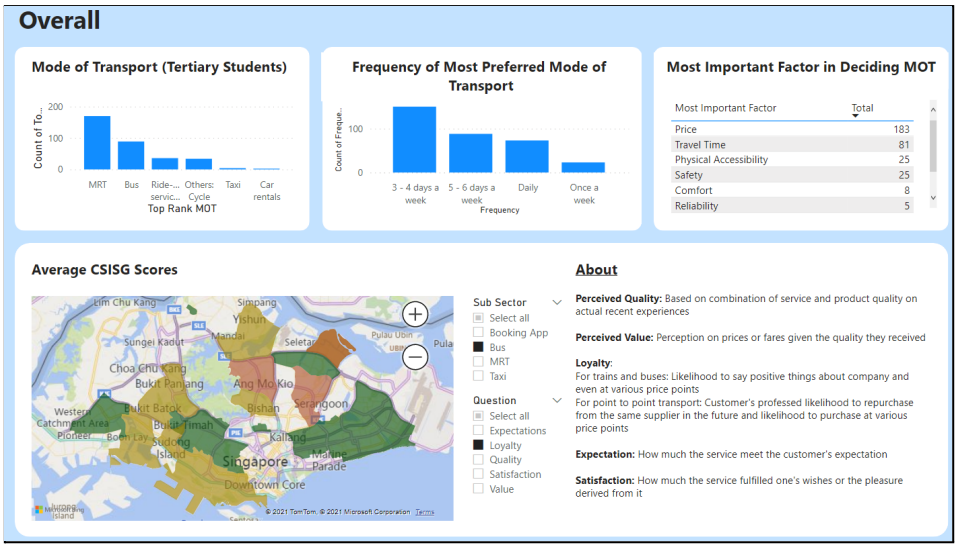


Figure 33: Dashboard Overall Tab

6.1.3 *Perception Tab*

By filtering subsectors and sentiments, the word cloud would display the top keywords and the donut chart would show the proportion of sentiment types.

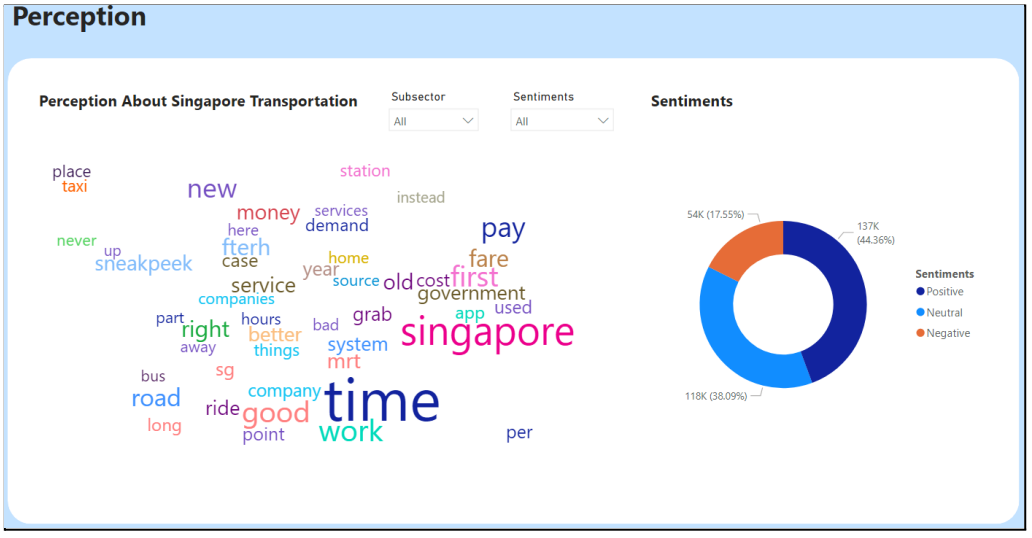


Figure 34: Dashboard Perception Tab

6.1.4 Demand Influencing Factors Tab

The user would be able to identify the impact of factors to each respective score based on the correlation results. By filtering the Survey Factor Breakdown graph, the user could identify factors that impact each cluster based on the extent of agreeability results and type of sub-sector. Survey Projected Population Numbers map can visualize the density of the clusters' population based on its respective planning areas. By comparing the results, the user can understand the behaviour of each cluster and point out the gaps of the market.

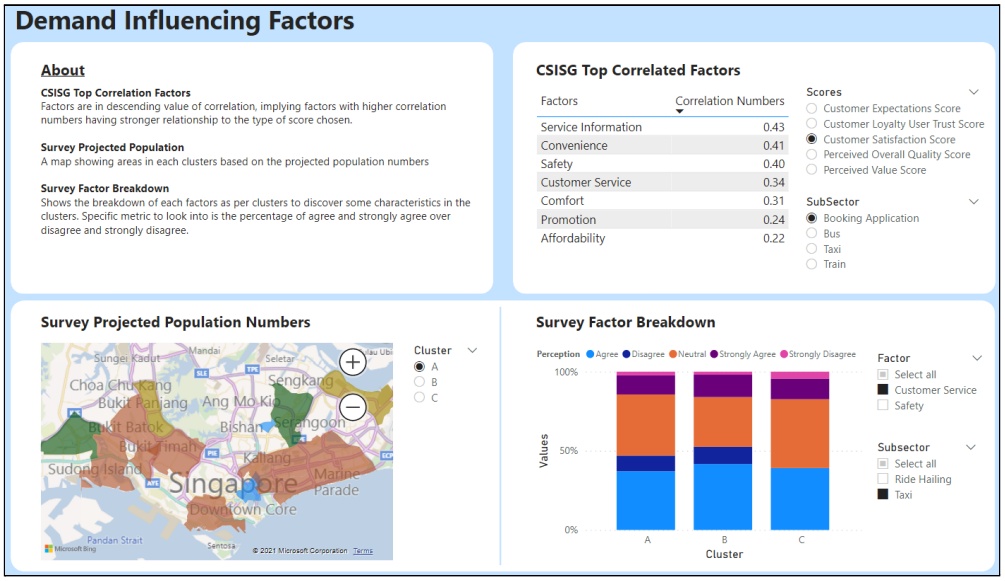


Figure 35: Dashboard Demand Influencing Factors Tab

6.2 User Guide

Refer to Appendix J for more details on the user guide.

6.3 Quality Attributes (KPI)

The current dashboard consists of tabs with interactive features such as dropdown and filters. Filters on the dashboard can help users to understand the data in different ways and select the filters that are most interesting to them. Despite the limited data sources and small dataset imported, the dashboard is also easily scalable if the dashboard were to be subjected to a larger dataset and data sources. While observations and insights will change accordingly, the dashboard will still remain interactive and scalable for future usage.

7.0 Project Management

7.1 Timeline

The team has planned out the project into phases of activities that entails its respective series of processes and time frames to complete. As seen on Figure 36 below, it is an overview of the actual timeline of this project. For a detailed breakdown of the live progress of the project, please refer to the following link for reference.

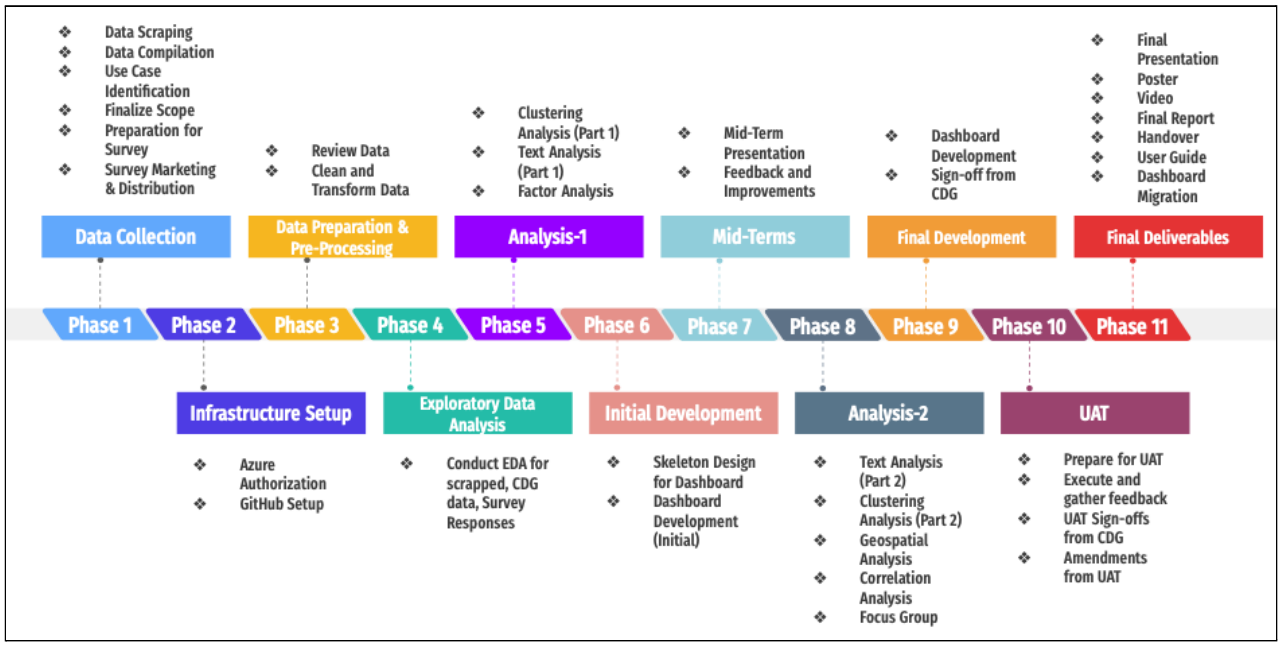


Figure 36: Actual Timeline of Project

Based on a comparison of the planned and actual timeline of the project, the team had a few minor changes in terms of the time period for specific phases, as seen on Figure 37 and 38 below. These changes are affected by the consideration to accommodate more time for specific analyses such as clustering and correlation, and scope changes.

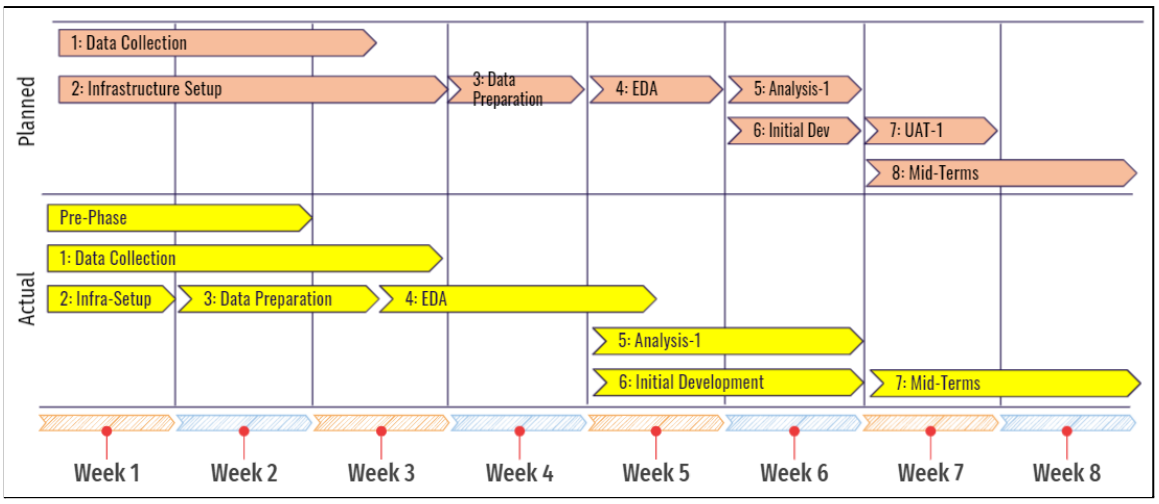


Figure 37: Changes in Timeline before Midterms

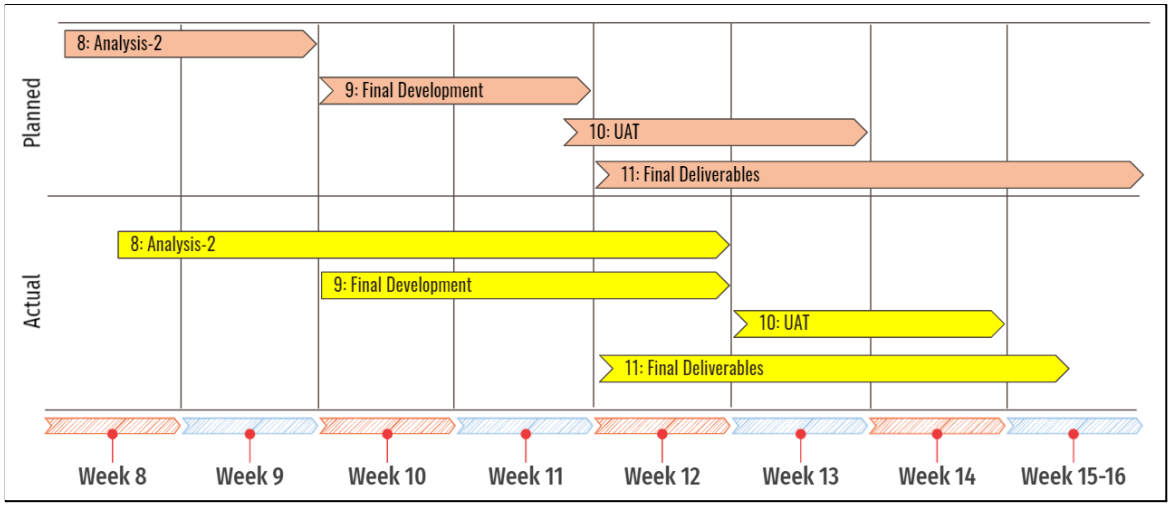


Figure 38: Changes in Timeline after Midterms

7.2 Approach

The team adopted the approach of a combination of elements from the waterfall and scrum project management methodology. The project was executed through a series of phases as seen in the previous section, and 3-day sprints were held in the initial half of the project due to scope changes, weekly sprints in the later half for rapid development and testing of the analysis models. Such approaches have made solving problems efficient and ensured that everyone is focused and is on the same page as one another.

Furthermore, buffer time has been planned in between each task which helps the team shift and set up more time for cases of scope changes and hiccups of the project. On a side note, prior to stricter Covid-19 restrictions, the team had to adopt meetings and sprints in an online format. However, the

team would still make time for occasional physical meets on the weekends when there are important deliverables due for either the client or school.

7.3 Scope Changes

The main scope change was the changes of problem statement and project aim during Week 6. This affected the progress and plans for all future phases of the project and due to these changes, additional tasks were added and weekly sprints were implemented to accelerate the progress and keep up with the timeline.

To ensure that the team is able to tackle the unpredictable and unplanned scope changes, proactive communication with stakeholders help facilitate the process of potential complications. Overall, the team managed to adapt quickly to changing circumstances and communicated frequently with stakeholders to ensure the changes made were according to the client's needs.

7.4 Stakeholder Management

Communication is key for good stakeholder management as it enables the team to clear doubts and understand the needs of the stakeholders more effectively. The team used a variety of tools to communicate with the stakeholders: Whatsapp, Telegram and Microsoft Teams.

The team held bi-weekly meetings with CDG's Team and ad-hoc meetings with the CDG's point of contact, Mr. Gary How for clarifications. Moreover, weekly meetings were held in the initial half of semester and bi-weekly meetings in the later half with the team's project supervisor, Professor Tan Poh Choo, to update on the team's progress, seek clarification and advice on how to move forward.

Lastly, internally within the team, meetings were held twice a week to discuss the team's progress, share findings and plan for the upcoming phases.

8.0 Gap Analysis & Future Work

8.1 Gap Analysis

A SWOT analysis was done by the team to better understand the limitations of the solution and look into how the project can be improved moving forward.

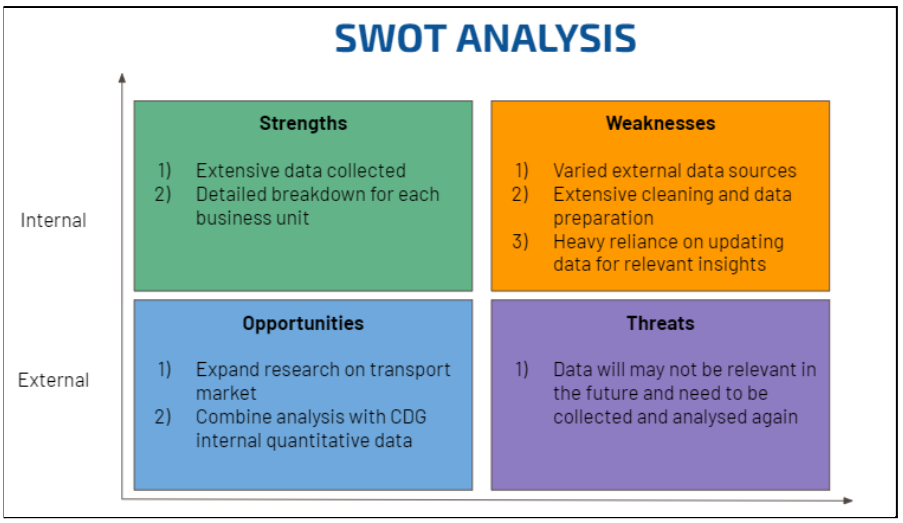


Figure 39: Gap analysis (SWOT)

Strengths

The team has performed extensive data collection from social media sites, surveys, focus groups and external sources for analysis. In addition, an in-depth breakdown of the analysis was done for each business unit to bring insights for CDG to understand the current market demand.

Weaknesses

The team heavily relies on external data sources like survey and CSISG, it may be difficult to replicate a similar survey and to get an updated CSISG dataset, the client would have to pay for it. Many of the datasets require extensive cleansing and data preparation for analysis which takes up a lot of time. As opinions and perceptions of consumers are constantly changing, there is a need to update the data frequently to get relevant insights.

Opportunities

The team could look into expanding its research on the current transportation market like competitor analysis. There is also an opportunity to improve the solution by combining analysis with CDG internal quantitative dataset to derive new insights.

Threats

Data that is currently used may no longer be relevant and require collection and re-analyzing again.

8.2 Future Work

8.2.1 Automation of manual processes

If there is a next phase for the project, the automation of the manual processes for data collection would value-add significantly to the client. Currently, for the data to be updated, CDG will be required to scrape data such as Reddit themselves by running the code and compiling the data. With

the use of automation, data can be generated more easily and without human input. Therefore, this provides them more time to examine the insights and strategies.

8.2.2 Direct Data Pipeline

Direct data pipeline of data into PowerBI dashboard would cut down the time required to collect data and allow CDG to optimise their resources and time. With applications like Microsoft Power Automate (Microsoft), survey data can be directly funneled into the dashboard, the dashboard will give real-time insights and enable CDG to make the most accurate business decisions.

8.2.3 Utilisation of more complex machine learning algorithms

In the future, the team can consider using more complex models for better results. For example, the use of modern neural network classifications such as Bidirectional Encoder Representations from Transformers (BERT) for the sentiment analysis of scrapped data and survey responses. Deep learning has shown significant progress in understanding text sentiment (Sharma et al., 2021) in order to label texts with higher accuracy and the insights drawn would be of greater business value.

9.0 Conclusion

9.1 Benefits for Sponsor

ComfortDelGro would be able to analyse the market demand of transportation across Singapore. The team conducted the first draft of research into the market, collating data from a variety of third party sources, cleaned the data and conducted the various analyses on it. From there, the team created a PowerBI dashboard for CDG to visualise the insights easily. On the dashboard, they can see the combination of various analyses to generate new insights. With the initial cut of research conducted by the team, CDG will be able to continue and expand their study further into the transport landscape in Singapore that is not covered in the project.

9.2 Team Effort

The team worked very well together and complemented one another's strengths and weaknesses. As some of the team's members were on full time internships, communication and teamwork was crucial to managing the workload. The team spent time outside of meetings to research and make sure deliverables were delivered on time. Strong understanding of individuals' responsibilities and roles enable the quick completion of required tasks. Furthermore, the team reached out to help one another when they were busy in school. At the time work was busier, the team would step up and take over their tasks.

9.3 Learning Takeaways

The team learnt that client management is quintessential when it comes to real-life projects. Challenges may arise unexpectedly and delays will occur, thus, being open and transparent with the client is the basis of a strong working relationship. In the event something does go wrong, the client will be able to understand the situation better. The team also learnt how to better gauge abilities of themselves and allocate sufficient time for tasks. Additionally, the application of waterfall project management allowed the team to deliver project deliverables promptly as well as adapting to scope changes in a real-world context. Finally, the team had the opportunity to learn software and technologies that extended beyond the school's curriculum.

10.0 References

Alfadda, A. (2014). Topic modeling for Wikipadia Pages. Topic Modeling For Wikipadia Pages. Retrieved November 27, 2021, from https://filebox.ece.vt.edu/~s14ece6504/projects/alfadda_topic/index.html

Bittrich, Kaden, M., Leberecht, C., Kaiser, F., Villmann, T., & Labudde, D. (2019). Application of an interpretable classification model on Early Folding Residues during protein folding. *BioData Mining*, *12*(1), 1–1. https://doi.org/10.1186/s13040-018-0188-2

Blair, S. J., Bi, Y., & Mulvenna, M. D. (2019). Aggregated topic models for increasing social media topic https://link.springer.com/content/pdf/10.1007%2Fs10489-019-01438-z.pdf

Bonthu, H. (2021). KModes Clustering Algorithm for Categorical data. Retrieved from https://www.analyticsvidhya.com/blog/2021/06/kmodes-clustering-algorithm-for-categorical-data/

Google. (n.d.). Classification: True vs. false and positive vs. negative. Google. https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative.

Harikrishnan N B. (2020). *Confusion matrix, accuracy, precision, recall, F1 score*. Medium. https://medium.com/analytics-vidhya/confusion-matrix-accuracy-precision-recall-f1-score-ade299cf63cd

Henry, D., Dymnicki, A. B., Mohatt, N., Allen, J., & Kelly, J. G. (2015) Clustering Methods with Qualitative Data: A Mixed Methods Approach for Prevention Research with Small Samples. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4939904/

Howard, C. n.d. Introduction to Cronbach's Alpha. Retrieved from https://mattchoward.com/introduction-to-cronbachs-alpha/

Koehrsen, W. (2021). When accuracy isn't enough, use precision and recall to evaluate your classification model. Built In. https://builtin.com/data-science/precision-and-recall

Land Transport Authority. (2017). Tuas West Extension Opens on 18 June 2017. Lta.gov.sg. https://www.lta.gov.sg/content/ltagov/en/newsroom/2017/4/2/tuas-west-extension-opens-on-18-june-2017.html

Le, & Nguyen, H. (2015). Twitter Sentiment Analysis Using Machine Learning Techniques. In Advanced Computational Methods for Knowledge Engineering (Vol. 358, pp. 279–289). Springer International Publishing.
https://doi.org/10.1007/978-3-319-17996-4_25

Khamis, H. (2008). Measures of Association: How to Choose?. *Journal Of Diagnostic Medical Sonography*, *24*(3), 155-162. https://doi.org/10.1177/8756479308317006

Mercer, A., Lau, A., & Kennedy, C. (2020). 1. how different weighting methods work. Pew Research Center Methods. https://www.pewresearch.org/methods/2018/01/26/how-different-weighting-methods-work/

Microsoft. (n.d.). Get started with power automate - power automate. Get started with Power Automate. Retrieved November 27, 2021, from https://docs.microsoft.com/en-us/power-automate/getting-started

Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010). Automatic evaluation of topic coherence. ACL Anthology. Retrieved November 27, 2021, from https://aclanthology.org/N10-1012.

Polat, C. (2012) Review Article: The Demand Determinants for Urban Public Transport Services: A Review of the Literature. Retrieved from https://docs.google.com/document/d/1xTICWcYZPxKGaf05EPxol8Goo0ZWc-gtLil1Hg2g8ho/edit#

PyShark. n.d. Davies-Bouldin Index for K-Means Clustering Evaluation in Python. Retrieved from https://pyshark.com/davies-bouldin-index-for-k-means-clustering-evaluation-in-python/

Rahn, M. (n.d.). Factor Analysis: A Short Introduction, Part 4–How many factors should I find? Retrieved from https://www.theanalysisfactor.com/factor-analysis-how-many-factors/

Resident working persons aged 15 years and over by usual mode of transport to work, age group and sex, 2015. Data.gov.sg. (n.d.). Retrieved November 27, 2021, from https://data.gov.sg/dataset/resident-working-persons-aged-15-yrs-over-by-usual-mode-of-transport-to-work-age-group-sex-2015?resource_id=34f4d6a6-a643-40b6-8b00-2528c7f95695

scikit-learn. (n.d.). Sklearn.metrics.f1_score. Scikit. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html

Sharma, P. (2019). A Beginner's Guide to Hierarchical Clustering and how to Perform it in Python. Analytics Vidhya. Retrieved from https://www.analyticsvidhya.com/blog/2019/05/beginners-guide-hierarchical-clustering/

Sharma, Kandasamy, I., & Vasantha, W. . (2021). Comparison of neutrosophic approach to various deep learning models for sentiment analysis. Knowledge-Based Systems, 223, 107058–. https://doi.org/10.1016/j.knosys.2021.107058

Subrahmannian, S. (2018). Learn to find topics in a text corpus. Medium. https://medium.com/@soorajsubrahmannian/extracting-hidden-topics-in-a-corpus-55b2214fc17d

Sun, L., & Yin, Y. (2017). Discovering themes and trends in transportation research using topic modeling. Transportation Research Part C: Emerging Technologies, 77, 49–66. https://doi.org/10.1016/j.trc.2017.01.013

Syed, S., & Spruit, M. (2017). Full-text or abstract? examining topic coherence scores using latent Dirichlet allocation. 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA). https://doi.org/10.1109/dsaa.2017.61

T, N. 2020. Data Reduction. Retrieved from https://binaryterms.com/data-reduction.html

TAN, C. (2021). Bus, train ridership in Singapore falls to 11-year low amid Covid-19 pandemic. Retrieved 26 November 2021, from https://www.straitstimes.com/singapore/transport/bus-train-ridership-in-singapore-falls-to-11-year-low-amid-covid-19-pandemic

Text mining 101: Topic modeling. KDnuggets. (n.d.). https://www.kdnuggets.com/2016/07/text-mining-101-topic-modeling.html

The world of Teoalida. (2021). *List of BTO projects and brochures*. The world of Teoalida. https://www.teoalida.com/singapore/btolist/

Ting, X., Ying, Z., Annette, B., & Shona, C. (2017). Public attitudes toward encouraging sustainable transportation: An Australian case study. International Journal of Sustainable Transportation, 11(8), 593-601. https://doi.org/10.1080/15568318.2017.1287316

Wati, M., Mahtari, S., Hartini, S., & Amelia, H. (2019). A Rasch model analysis on junior high school students' scientific reasoning ability. International Journal of Interactive Mobile Technologies (IJIM), 13(07), 141. https://doi.org/10.3991/ijim.v13i07.10760

Wei, H. (2020). How to measure clustering performances when there are no ground truth? Medium. Retrieved from https://medium.com/@haataa/how-to-measure-clustering-performances-when-there-are-no-ground-truth-db027e9a871c

Yadav D. (2019). Categorical encoding using Label-Encoding and One-Hot-Encoder. Towards Data Science. Retrieved from https://towardsdatascience.com/categorical-encoding-using-label-encoding-and-one-hot-encoder-911ef77fb5bd

11.0 Appendix

11.1 Appendix A: Data Scraped Libraries

```
c = twint.Config()
c.Username = "cdgtaxi_sg"
c.Store_json = True
c.Output = "data/cdgtaxi.json"
twint.run.Search(c)
```

```
1440845403637706765 2021-09-23 09:08:08 +0800 <cdgtaxi_sg> Heading out this wee
kend for some essential shopping?  The ComfortDelGro Taxi Booking App allows y
ou to go to two locations in one ride. And book it during non-peak period as ou
r fares may be cheaper.  Download our app now:  https://t.co/rwhwKduIoi  #RideW
ithComfort #taxi  https://t.co/Q6TF83206Y
1440605214516469770 2021-09-22 17:13:42 +0800 <cdgtaxi_sg> At ComfortDelGro Tax
i, we offer different vehicles, from hybrid to diesel, from regular sedan to lu
xurious ones - to suit different needs and preferences.  Come drive with us no
w:  https://t.co/D0vDillsf6  #drivewithcomfort #comfortunited #taxi #singapore
https://t.co/1eTTcSfirQ
```

Figure 40: Twint used in twitter data scraping

```
from igramscraper.instagram import Instagram
```

```
account = instagram.get_account('rydesharing')
```

```
print('Account info:')
print('Id: ', account.identifier)
print('Username: ', account.username)
print('Full name: ', account.full_name)
print('Biography: ', account.biography)
print('Profile pic url: ', account.get_profile_picture_url())
print('External Url: ', account.external_url)
print('Number of published posts: ', account.media_count)
print('Number of followers: ', account.followed_by_count)
print('Number of follows: ', account.follows_count)
print('Is private: ', account.is_private)
print('Is verified: ', account.is_verified)
```

```
Account info:
Id:  1529019801
Username:  rydesharing
Full name:  Ryde
Biography:  It's a better ride
Unlimited Referral Bonus - Refer your friends now!
Profile pic url:  https://instagram.fsin3-1.fna.fbcdn.net/v/t51.2885-19/s320x32
0/231672709_543471643471641_4915256645104833487_n.jpg?_nc_ht=instagram.fsin3-1.
fna.fbcdn.net&_nc_ohc=z2E7iO6ddKgAX8OiSKV&edm=ABfd0MgBAAAA&ccb=7-4&oh=ff50cf84d
cfe3dd37cff0ca614b09fc5&oe=611B877A&_nc_sid=7bff83
External Url:  https://bit.ly/3fkW36m
Number of published posts:  395
Number of followers:  2401
Number of follows:  32
Is private:  False
Is verified:  False
```

Figure 41: Igramscraper used to scrape Instagram

```
#imports here
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.common.by import By
from selenium.webdriver.support.wait import WebDriverWait
import time
import os
import wget
```

```
with open('tadasg_fb_utf8.txt', encoding='utf8', errors='ignore') as f:
    contents = f.read()
    print(contents)
```

```
[{'post_id': '4211271042146574', 'text': "#TADA3rdBirthday To those who have interacted with the video before this new uplo
ad: the original post was removed unexpectedly due to a Facebook glitch😠 but we guess it's alright, because...\n\nIt is TA
DA's actual 3rd Birthday today!\U0001f973 The video is dedicated to all TADA users - rider and drivers. We wouldn't have ma
de it this far without you believing in us. Zero-Commission for TADA drivers is the promise we have been keeping since 201
8. Until today, we still have faith in what Zero-Commission can do for building a happier and fairer ride-hailing communit
y.\n\nIf you like to support private-hire drivers during the P2HA, book zero-commission rides on TADA for your errands or u
se the On-Demand Delivery service which also remains zero-commission for drivers in your TADA app to deliver your parcels a
nd documents.\n\nLet's stay safe! We'll see you on TADA again 😊", 'post_text': "#TADA3rdBirthday To those who have interac
ted with the video before this new upload: the original post was removed unexpectedly due to a Facebook glitch😠 but we gue
ss it's alright, because...\n\nIt is TADA's actual 3rd Birthday today!\U0001f973 The video is dedicated to all TADA users -
 rider and drivers. We wouldn't have made it this far without you believing in us. Zero-Commission for TADA drivers is the p
romise we have been keeping since 2018. Until today, we still have faith in what Zero-Commission can do for building a happ
```

Figure 42: Selenium used to scrape Facebook

```
reddit = praw.Reddit(client_id='X7CNXZuB-dDBkiFMNRsLkg',
                     client_secret='oh1QHjnzAHAT6Zfk3v8OiuDhKkju8w',
                     user_agent='Scaping Example')
```

Create Function to scrape reddit

```
def scrape_comments(subreddit, keyword):
    sub = reddit.subreddit(subreddit)

    sub_dict = {}
    sub_lst = []
    sub_comments = []
    comment_dict = {}

    for submission in sub.search(keyword, limit = None):
        sub_dict['title'] = submission.title
        sub_dict['time created'] = pytz.utc.localize(datetime.utcfromtimestamp(submission.created_utc)).astim
        sub_dict['score']= submission.score
        sub_dict['id'] = submission.id
        sub_dict['url']= submission.url

        submission.comments.replace_more(limit = None)
        for comment in submission.comments.list():
            comment_dict['time created'] = pytz.utc.localize(datetime.utcfromtimestamp(comment.created_utc))
            comment_dict['author'] = str(comment.author)
            comment_dict['score'] = comment.score
            comment_dict['comment'] = comment.body
            sub_comments.append(comment_dict)
            comment_dict = {}
        sub_dict['comments'] = sub_comments

        sub_comments = []
        sub_lst.append(sub_dict)
        sub_dict = {}
        comment_dict = {}

    with open(f"../../Data Collection and Preprocessing/Scrapping/Reddit/{subreddit}_{keyword}.json", 'w', e
        file.write(json.dumps(sub_lst, indent=4))
```

Figure 43: Praw used to scrape Reddit

11.2 Appendix B: Sample of Survey Questions

Type of questions	Question
Demographics	What is your age? What is your gender? What is your monthly income range? What is your current employment status? Are you currently studying? If you are, where? / Are you a tertiary student? In which part of Singapore are you residing?
Travel Patterns	How often do you take the public transport per week? Which modes of transportation do you often use? On average, how much in total do you spend on public transport per month?
Transport Preference	Please rank your preferred mode of transport Please rank these factors based on importance when you are deciding your mode of transport
Psychographics	On a scale of 0 - 10, how likely are you to research on the most efficient route to an unfamiliar destination before embarking on it? On a scale of 0 -10, while travelling on the public transport, how comfortable are you with being around other commuters? On a scale of 0 - 10, how likely are you to explore a new route to your destination?

Table 7: List of some survey questions

11.3 Appendix C: Text Analysis

Sentiment Analysis metrics

True Positive = an outcome where the model *correctly* predicts the *positive* class.

True Negative = an outcome where the model *correctly* predicts the *negative* class.

False Positive = an outcome where the model *incorrectly* predicts the *positive* class.

False Negative = an outcome where the model *incorrectly* predicts the *negative* class.

(Google, n.d.)

Precision(PR) = the ratio tp / (tp + fp) where tp is the number of true positives and fp the number of false positives.

Recall(RE) = the ratio tp / (tp + fn) where tp is the number of true positives and fn the number of false negatives. (scikit-learn, n.d.)

Accuracy(CA) = the number of correctly classified data instances over the total number of data instances. (Harikrishnan N B, 2020)

F1 Score(F1) = the harmonic mean of the precision and recall (Koehrsen, 2021)

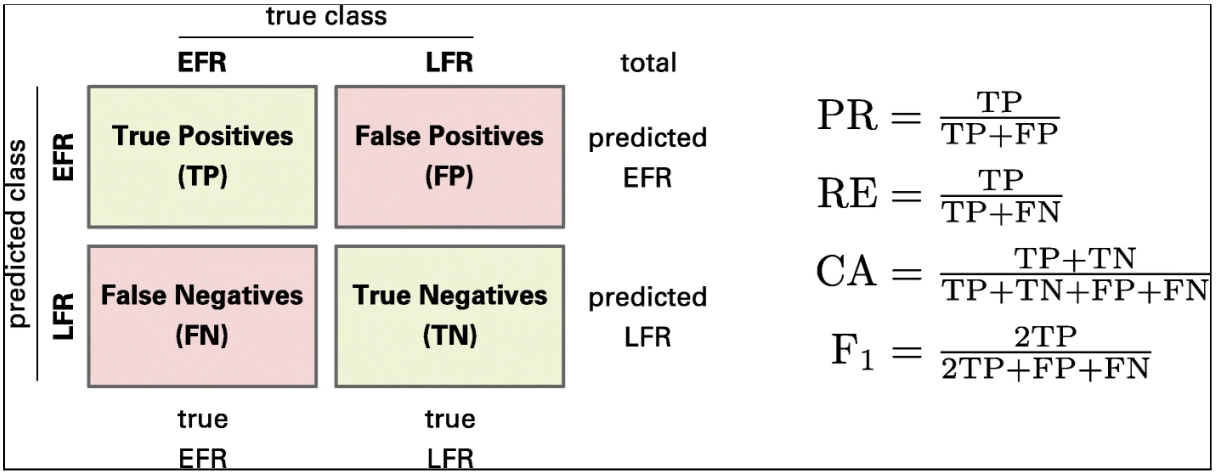


Figure 44: Confusion matrix and simplified equations for metrics (Bittrich et al., 2019)

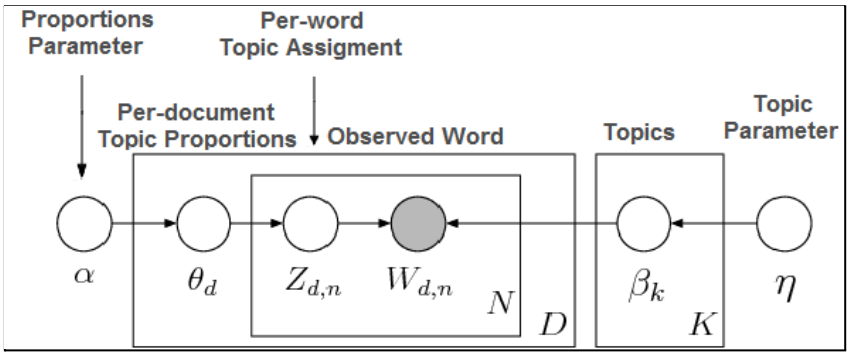


Figure 45: Visualization of LDA model (Alfadda, 2014)

$$\vec{v}(W') = \left\{ \sum_{w_i \in W'} \mathbf{NPMI}(w_i, w_j)^\gamma \right\}_{j=1,\ldots,|W|} \quad (3)$$

$$\mathbf{NPMI}(w_i, w_j)^\gamma = \left(\frac{\log \frac{P(w_i, w_j)+\epsilon}{P(w_i) \cdot P(w_j)}}{-\log(P(w_i, w_j)+\epsilon)} \right)^\gamma \quad (4)$$

$$\phi_{S_i}(\vec{u}, \vec{w}) = \frac{\sum_{i=1}^{|W|} u_i \cdot w_i}{\|\vec{u}\|_2 \cdot \|\vec{w}\|_2} \quad (5)$$

Figure 46: Coherence score breakdown (c_v) (Syed & Spruit, 2017)

$$\text{Coherence} = \sum_{i<j} \text{score}(w_i, w_j)$$

Figure 47: Coherence score formula (Subrahmannian, 2018)

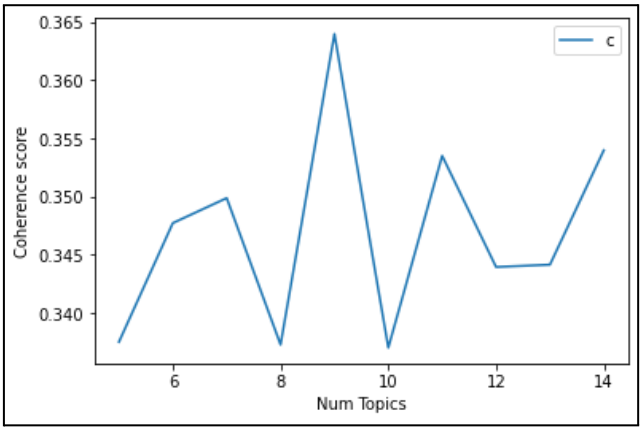
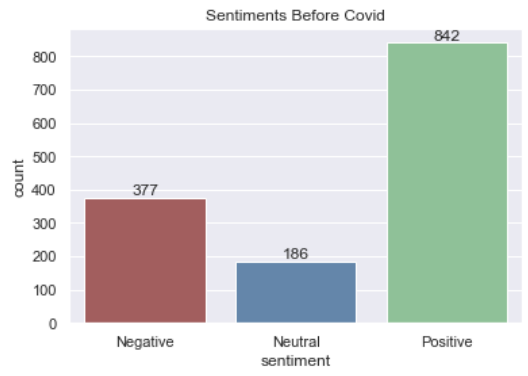
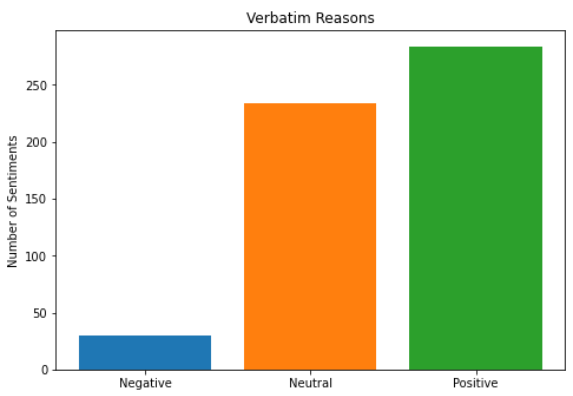


Figure 48: Elbow Model for Optimal Number of Topics for Topic Modelling

11.4 Appendix D: Sentiment Analysis Charts



Reddit Sentiments Before Covid-19



CSISG 2015 to 2019 sentiments

Reddit Combined Negative Word Cloud

CSISG 2020 Areas of Improvement Negative Word Cloud

11.5 Appendix E: Factor Analysis

Factors	Variables	Cronbach Alpha Value
Convenience	convenient_rental_car difficulties_rental_car reason_chosen_rental_car_company_convenient reason_rental_car_convenience reason_rental_car_occasions	0.931
Affordability	affordable_bus affordable_mrt	0.758
Safety	safe_taxi afe_ride_hailing safe_bus safe_mrt	0.736
Customer Service	customer_service_taxi customer_service_ride_hailing	0.645
Comfort	reason_taxi_comfort_of_privacy reason_ride_hailing_ comfort_of_privacy	0.728
Promotion	reason_ride_hailing_promotion reason_taxi_promotion	0.606
Accessibility	reason_ride_hailing_accessible reason_taxi_more_accessible reason_ride_hailing_easy_booking	0.579

Table 8: List of factors identified for travel perception dataset

11.6 Appendix F: Clustering Analysis

# of respondents	**Sentiment on factor**				
Factors	Strongly Agree	Agree	Neutral	Disgaree	Strongly Disagree
Affordability	48.5%	35.1%	14.2%	2.2%	0.0%
Safety	60.4%	37.3%	2.2%	0.0%	0.0%
Customer Service	19.4%	47.8%	28.4%	4.5%	0.0%
	Whether they will consider factor				
	Yes		Neutral	No	
Comfort	21.6%		14.2%	64.2%	
Promotion	47.0%		29.1%	23.9%	
Accessible	41.8%		23.9%	34.3%	

Table 9: Results of Cluster A based on the aggregated values

# of respondents	**Sentiment on factor**				
Factors	Strongly Agree	Agree	Neutral	Disgaree	Strongly Disagree
Affordability	34.4%	47.4%	15.6%	2.6%	0.0%
Safety	52.6%	47.4%	0.0%	0.0%	0.0%
Customer	16.2%	51.3%	27.3%	5.2%	0.0%

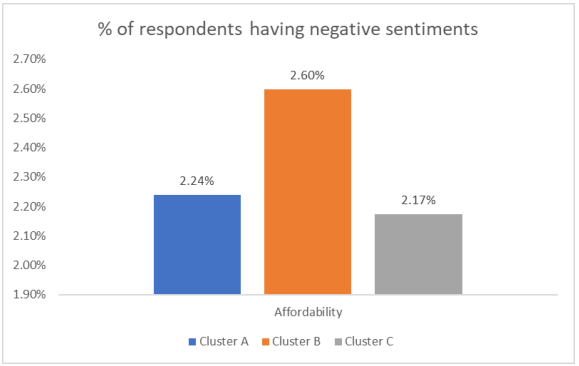
Service					
	Whether they will consider factor				
	Yes		Neutral	No	
Comfort	12.3%		16.9%	70.8%	
Promotion	44.2%		27.9%	27.9%	
Accessible	39.0%		27.9%	33.1%	

Table 10: Results of Cluster B based on the aggregated values

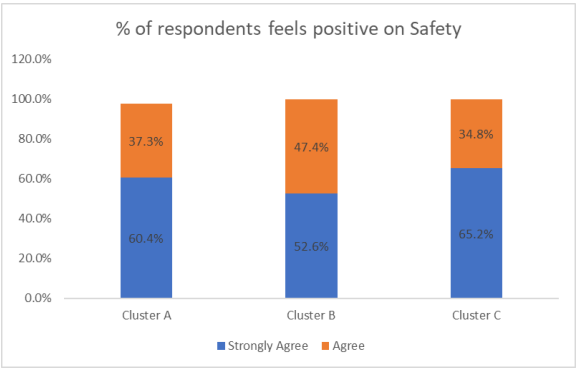
# of respondents	**Sentiment on factor**				
Factors	Strongly Agree	Agree	Neutral	Disgaree	Strongly Disagree
Affordability	41.3%	45.7%	10.9%	2.2%	0.0%
Safety	65.2%	34.8%	0.0%	0.0%	0.0%
Customer Service	26.1%	43.5%	28.3%	2.2%	0.0%
	Whether they will consider factor				
	Yes		Neutral	No	
Comfort	17.4% 19.6% 63.0%		19.6%	63.0%	

Promotion	58.7%	26.1%	15.2%
Accessible	45.7%	23.9%	30.4%

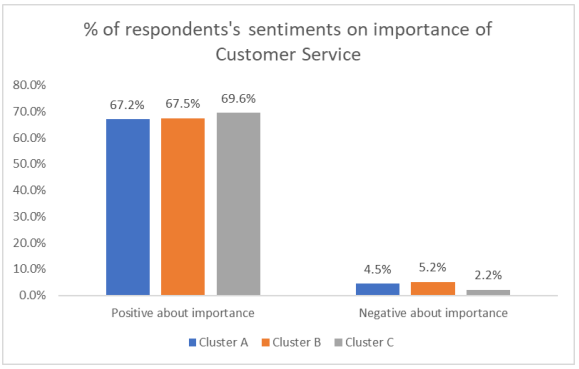
Table 11: Results of Cluster C based on the aggregated values



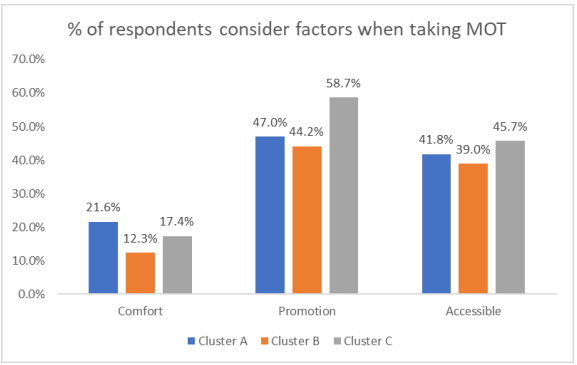
More respondents in Cluster B has negative sentiments on Affordability



Safety is not a current concern for all respondents as almost 100% feels positive about the factor



- 5.2% in Cluster B don't see importance in customer service when travelling → higher than other clusters
- 69.6% respondents in Cluster C agree/strongly agree about the importance of customer service



- 21.6% of those in Cluster A consider comfort when travelling
- 58.7% of those in Cluster C consider promotion when choosing their mode of transport
- Accessibility is important to 45.7% respondents in Cluster C

Table 12: Observations and insights for all cluster

11.7 Appendix G: Correlation Analysis

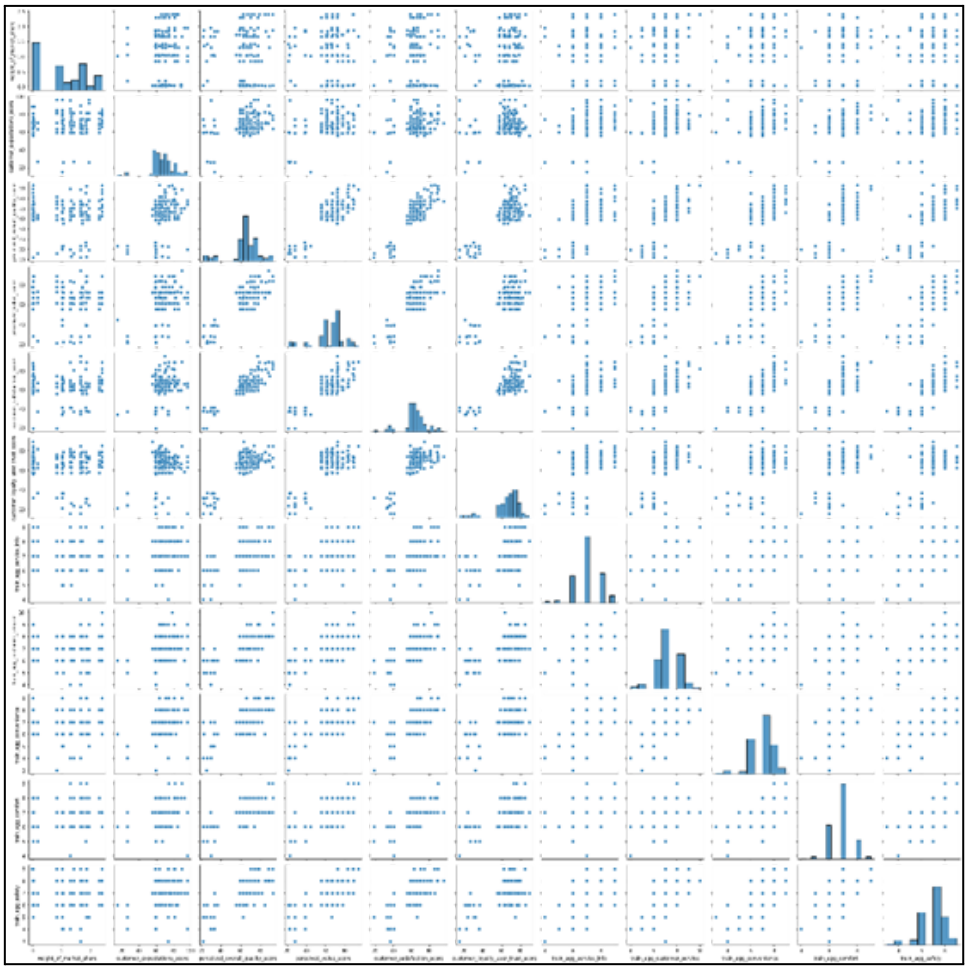


Figure 48: Pair plot of correlation dataset for train subsector

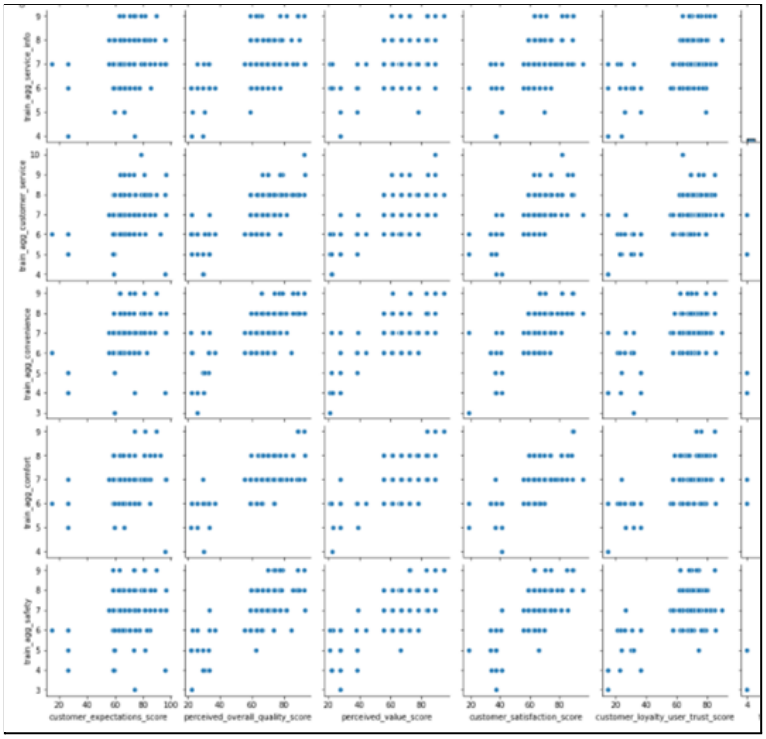


Figure 49: Scatter plot of ordinal and continuous variables

	Customer Expectation Score	Perceived Overall Quality Score	Perceived Value Score	Customer Satisfaction Score	Customer Loyalty User Trust Score
Bus					
Service Information	0.426121	0.546361	0.522351	0.56051	0.525716
Customer Service	0.421063	0.435145	0.414906	0.480028	0.507758
Convenience	0.382501	0.667392	0.542232	0.731708	0.523447
Safety	0.435537	0.553685	0.513515	0.562911	0.437296
Train					
Service Information	0.278647	0.415863	0.373003	0.504673	0.393348
Customer	0.285749	0.588372	0.404657	0.609747	0.411225

Service					
Convenience	0.287681	0.561155	0.344823	0.669141	0.452431
Comfort	0.205009	0.590847	0.481743	0.562942	0.390374
Safety	0.282822	0.635019	0.512204	0.636153	0.359933
Taxi					
Service information	-0.00281	0.291913	0.078617	0.396293	0.18568
Customer Service	0.109519	0.401835	0.458972	0.510043	0.377111
Comfort	-0.18026	0.426257	0.029255	0.431426	0.074499
Safety	-0.02	0.54697	0.25021	0.382544	0.255435
Affordability	-0.12984	0.244269	0.13393	0.257244	0.271664
Booking Application					
Service information	0.287837	0.405456	0.310722	0.432319	0.320355
Customer Service	0.059335	0.515026	0.29334	0.342521	0.296854
Convenience	0.13308	0.501001	0.332131	0.411377	0.389001
Comfort	0.196863	0.452914	0.414116	0.314976	0.244542
Safety	0.23973	0.499342	0.354348	0.399217	0.281305
Affordability	0.230853	0.182892	0.238064	0.21903	0.252001
Promotion	0.146487	0.127885	0.264051	0.241833	0.1677

Table 13: Correlation results Spearman correlation

11.8 Appendix H: Geospatial Analysis

Sub-sector	Top Positively Correlated Factor	Correlation Value
Bus	Safety	0.87
MRT/LRT	Insufficient data to support observation	-
Taxi	Safety Affordability Service Information	0.87 0.87 0.75
Taxi/Car Booking Application	Service Information	0.69

Table 14: Top positive correlation

11.9 Appendix I: Comparison of Respective Scoring from Year 2016 to 2019

This comparison is performed based on the aggregated scores from the Customer Satisfaction Index of Singapore (CSISG) survey dataset. The comparison of respective scoring will be dissected into sectors namely Bus, MRT/LRT, Taxi, and Booking Application in the following sections below.

Disclaimer 1: Due to the format of how the CSISG survey was conducted, they would only circulate the survey in a number of regions in Singapore yearly. Hence, there would be missing data in certain regions when comparing from year to year.; Scores are evaluated with a max score of 100.

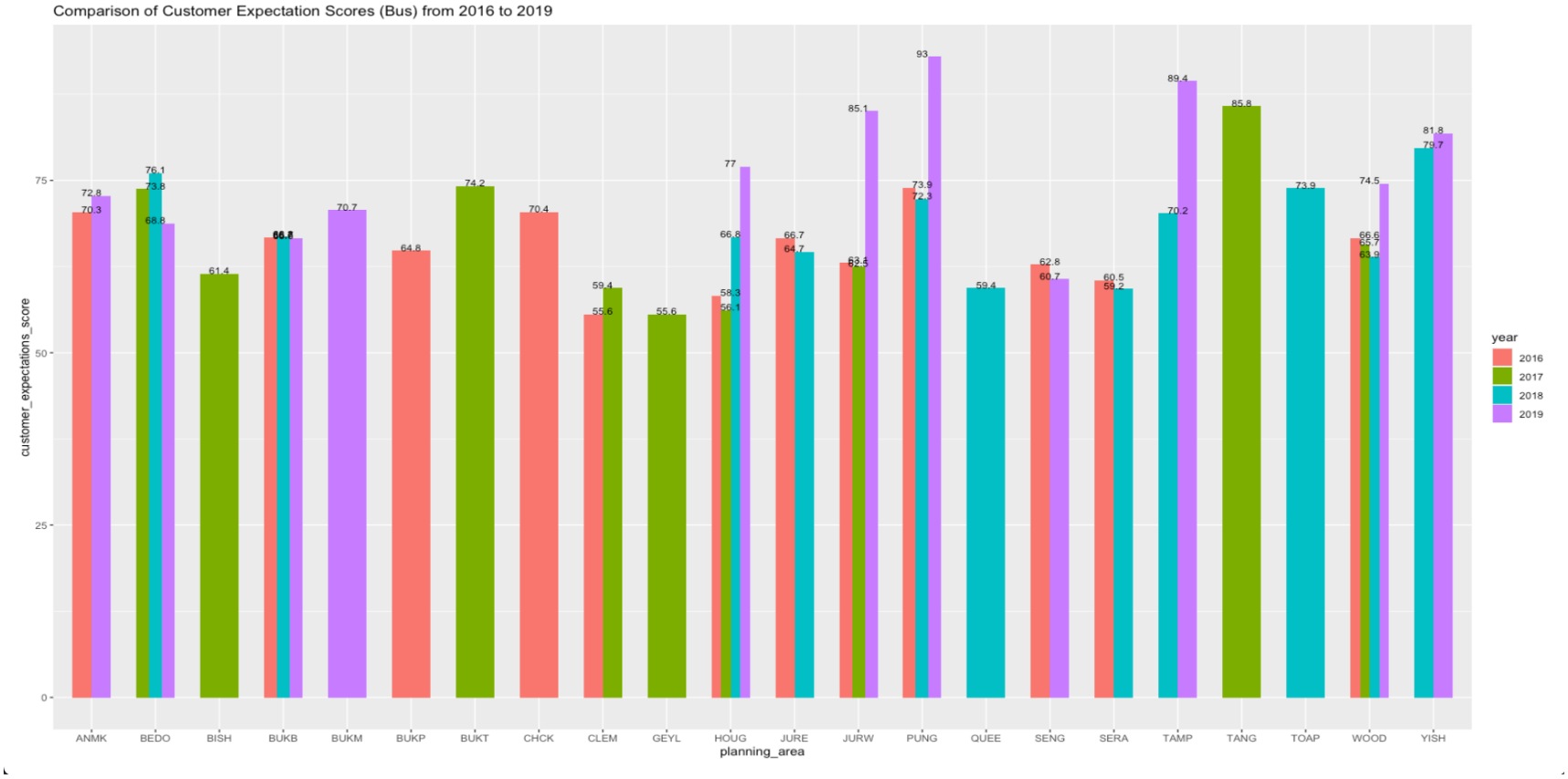
11.9.1 Definition of Respective Scores

Type of Score(s)	Definition
Customer Expectation	Customer Expectations are any set of behaviours or actions that individuals anticipate when interacting with a particular service.
Customer Satisfaction	Customer Satisfaction is determined by customer expectations, perceived quality and perceived value. In general, customer satisfaction is a measurement to determine how happy customers are with the mobility market.
Customer Loyalty User Trust	(for *Trains and Buses*) The final desired outcome of Customer Satisfaction for most public service providers is a combination of (1) the customer's likelihood to say positive things about the company, and (2) the likelihood to still say positive things about the brand at various price points (price tolerance). This component is usually a good indicator of the commuters' confidence in the Public Transportation. (for *Taxis and Booking Application*) The final desired outcome of Customer Satisfaction for most competitive market industries is a combination of (1) the customer's professed likelihood to repurchase from the same supplier in the future, and (2) the likelihood to purchase a company's products or services at various price points (price tolerance). Customer loyalty is the critical component of the model as it stands as a proxy for profitability for the company.
Perceived Overall Quality	Based on the combination of service and product quality as perceived by customers based on their actual recent experience. This is affected by the companies' investments in training, infrastructure, and development. This is shaped by customers' perception of (1) overall quality they felt they received, (2) Ability of the company to meet their personal requirements, and (3) the reliability of their products and services.
Perceived Value	Influenced by customers' perceptions of the quality and pricing structure. This is shaped by the customers' (1) perception on the price or fares given the quality they received, and then (2) on their perception of quality received given the price or fares they paid.

11.9.2 Service Sector: ***Bus***

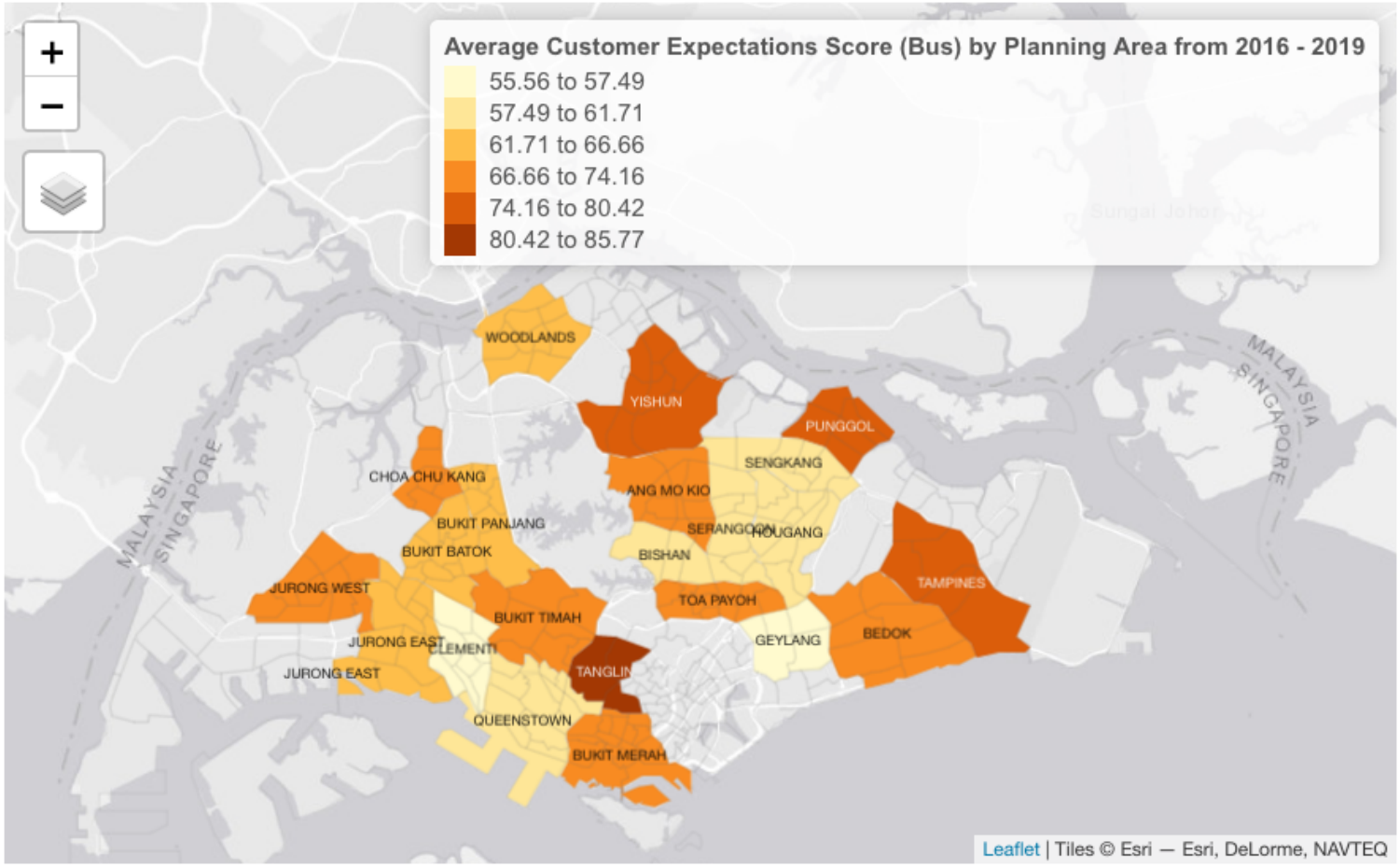
In the following comparison charts, the team will be looking into details of each of the respective scoring and how it changes from Year 2016 to 2019.

11.9.2.1 Customer Expectation



Findings 1: There is a positive increase of the expectation scores from Year 2016 to 2019, across the respective planning areas.

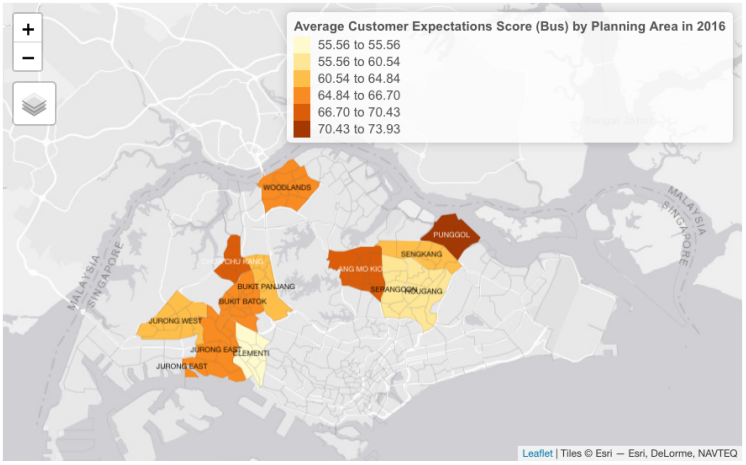
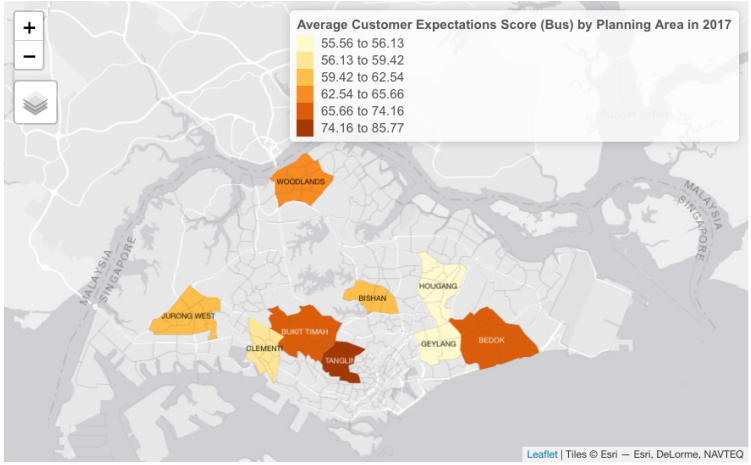
Findings 2: *Jurong West*, *Punggol* and *Tampines* can be seen with a drastic increase in expectations in Year 2019.

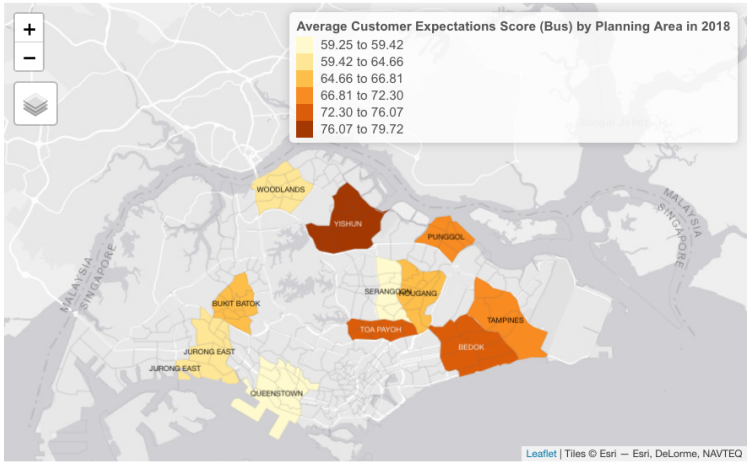


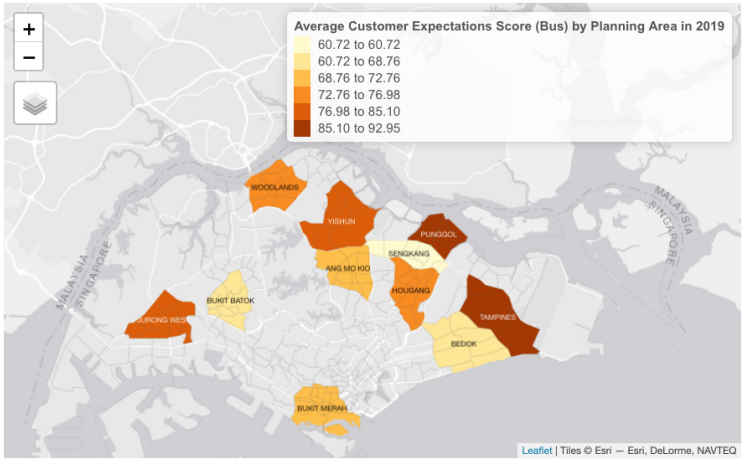
Findings 3: When the team aggregates the score to get an average across the years, *Tanglin (85.77)* is seen as the planning area with the highest expectation score.

Findings 4: Based on the chart above, planning areas with a higher expectation scores are predominantly those with 1 or no MRT lines available at their area, for example, Tanglin (No MRT Line), Yishun (North-South Line), Punggol (North-East Line) and Tampines (East-West Line). Hence, there could be a possibility that due to that, they would expect more from the current bus services.

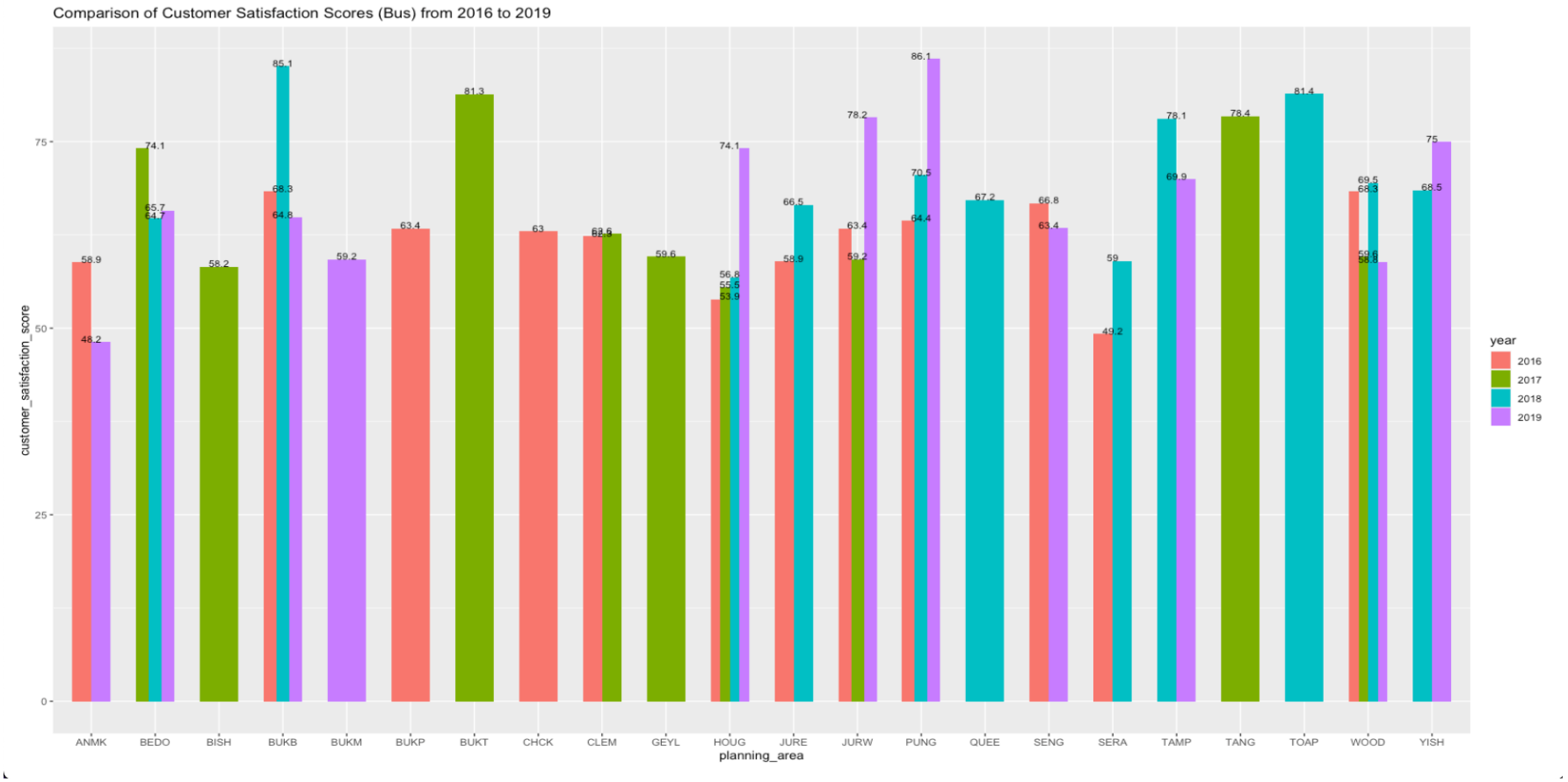
*Breakdown of Customer **Expectation** Scores by Year and Planning Area*

<table>
<tr><td>2016</td><td>2017</td></tr>
<tr><td> 
</td><td> 
</td></tr>
<tr><td>2018</td><td>2019</td></tr>
</table>


Average Customer Expectations Score (Bus) by Planning Area in 2018
59.25 to 59.42
59.42 to 64.66
64.66 to 66.81
66.81 to 72.30
72.30 to 76.07
76.07 to 79.72
WOODLANDS
YISHUN
PUNGGOL
SERANGOON
HOUGANG
BUKIT BATOK
TOA PAYOH
TAMPINES
BEDOK
JURONG EAST
JURONG EAST
QUEENSTOWN
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

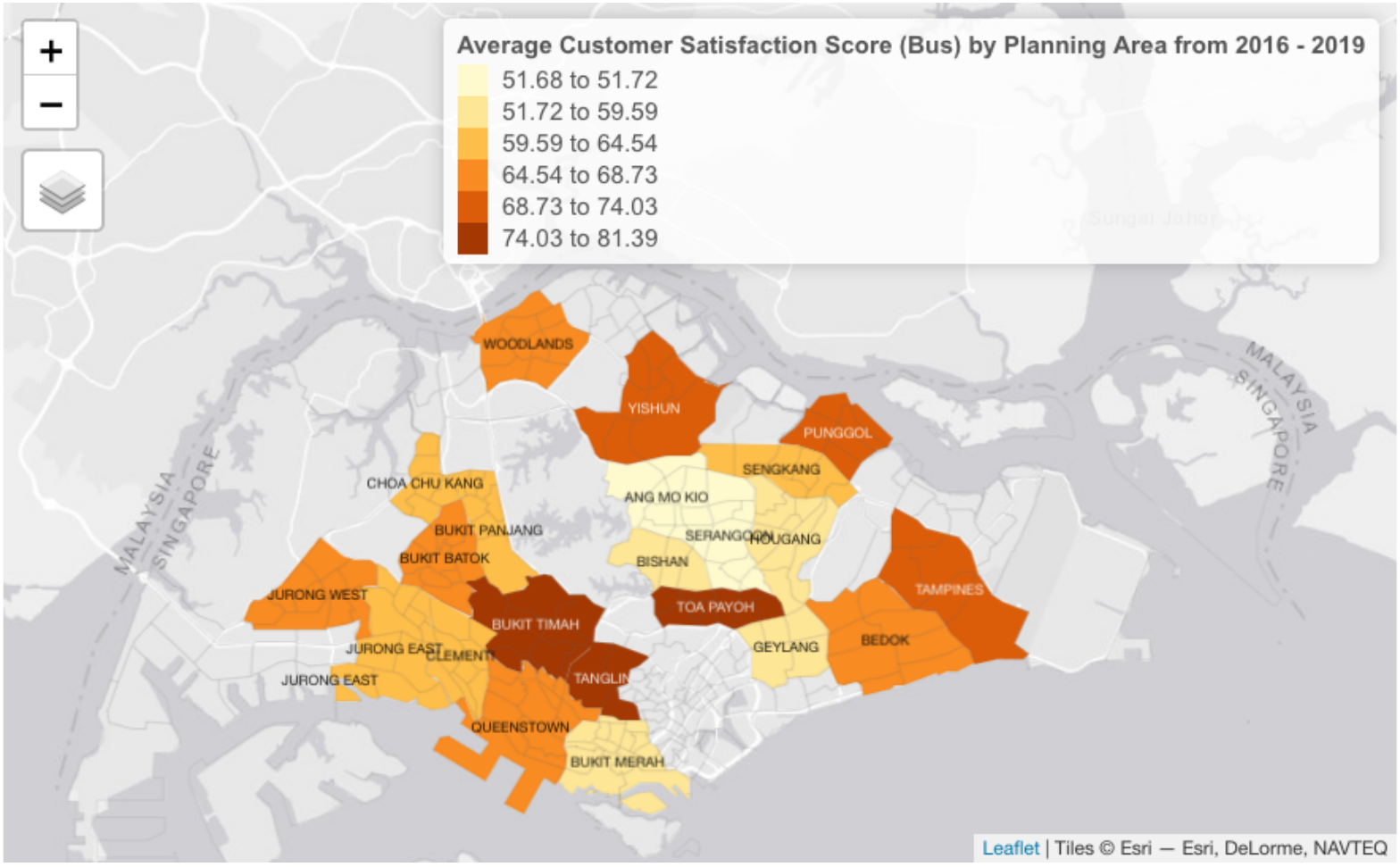

Average Customer Expectations Score (Bus) by Planning Area in 2019
60.72 to 60.72
60.72 to 68.76
68.76 to 72.76
72.76 to 76.98
76.98 to 85.10
85.10 to 92.95
WOODLANDS
YISHUN
PUNGGOL
SENGKANG
ANG MO KIO
HOUGANG
BUKIT BATOK
TAMPINES
BEDOK
BUKIT MERAH
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

11.9.2.2 Customer Satisfaction



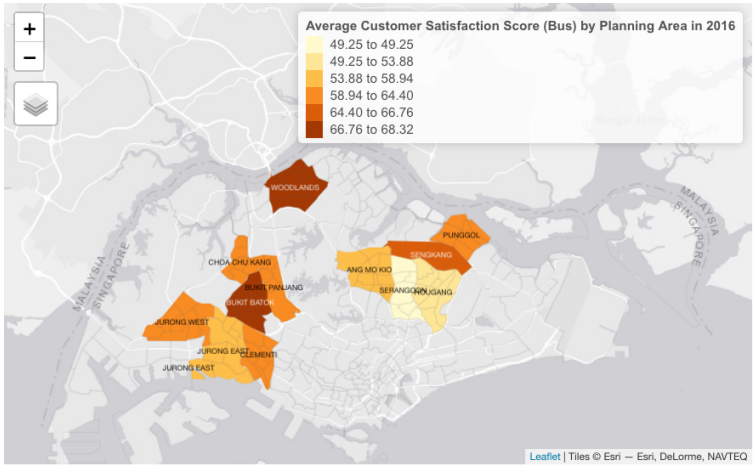
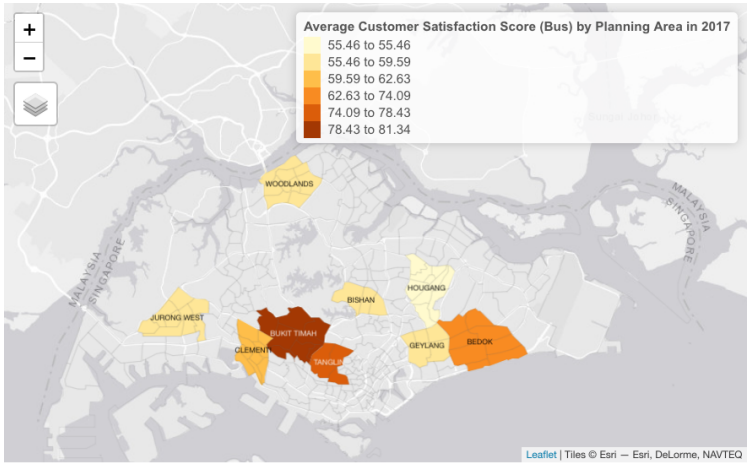
Findings 1: There are drastic differences in terms of growth of satisfaction scores from year to year and planning areas. Positive increase and growth could be seen in areas like *Hougang*, *Jurong West*, *Punggol and Yishun*. However, areas like *Bedok*, *Bukit Batok*, *Sengkang*, and *Woodlands* have been significantly seen with extreme drops over the years.

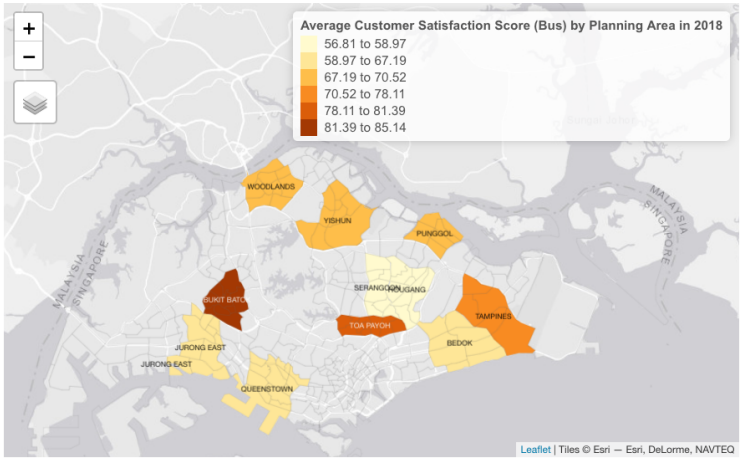
Findings 2: Planning areas that are seen with positive increase could be a possibility of the effect of the availability of new MRT stations, routes and Build-to-Order HDBs. Due to that, new bus routes have been implemented for the convenience of the public.

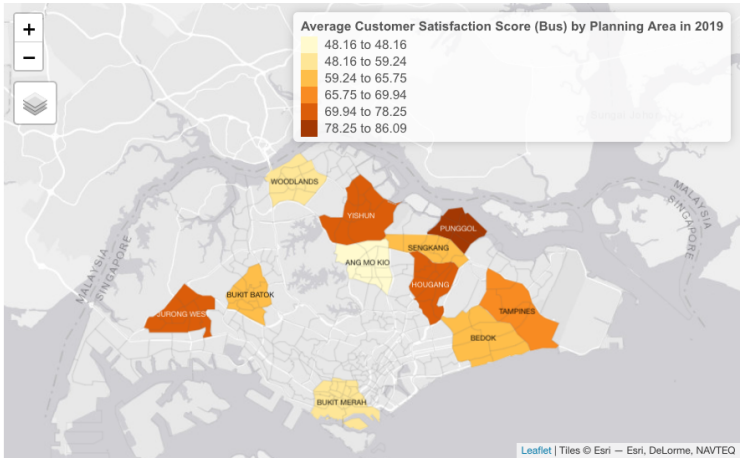


Findings 3: Areas with higher average satisfaction scores over the years could be predominantly seen in the Central area of Singapore (*Bukit Timah*, *Tanglin*, *Toa Payoh*).

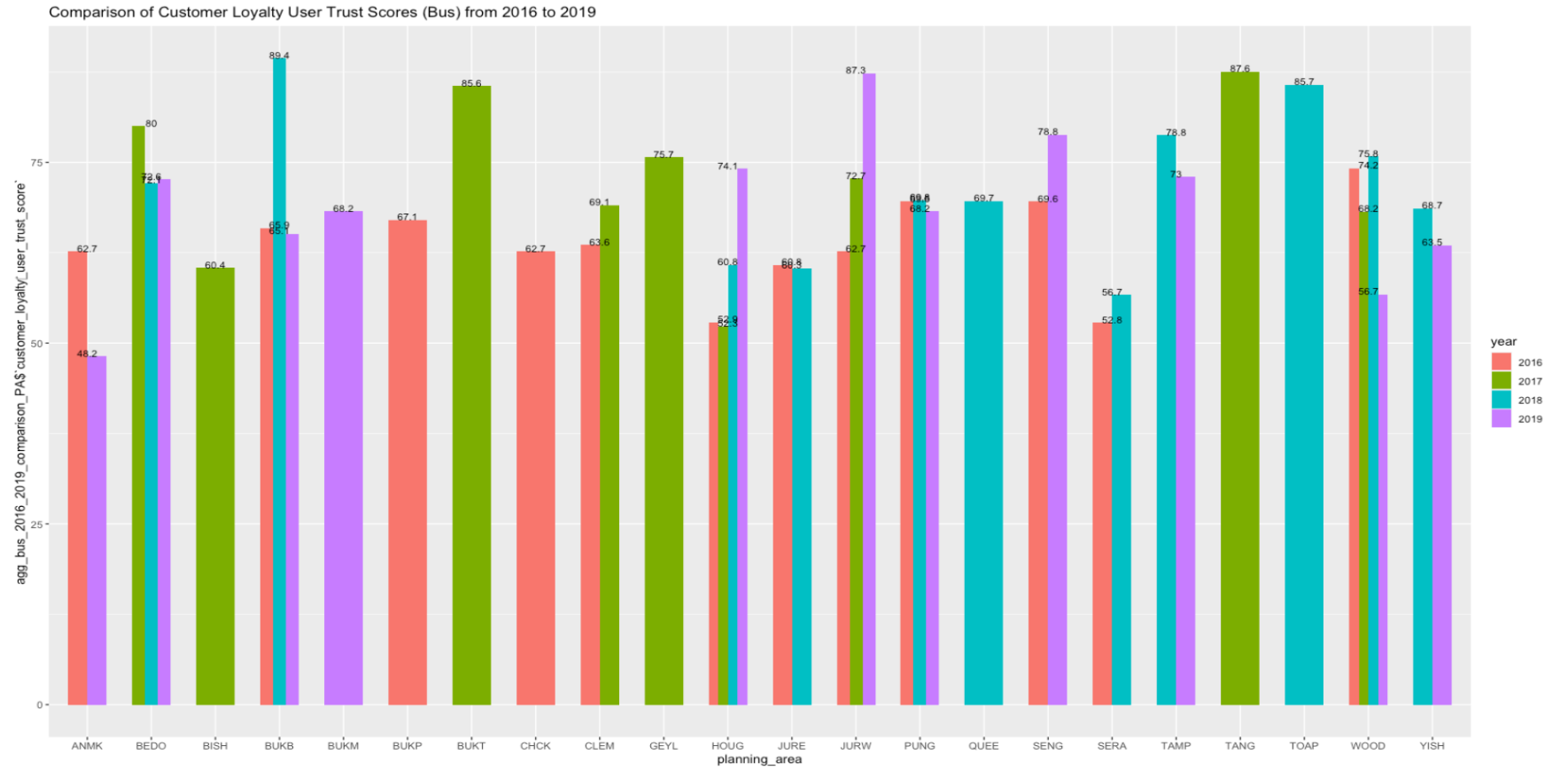
*Breakdown of Customer **Satisfaction** Scores by Year and Planning Area*

2016	2017
	
2018	**2019**


Average Customer Satisfaction Score (Bus) by Planning Area in 2018
56.81 to 58.97
58.97 to 67.19
67.19 to 70.52
70.52 to 78.11
78.11 to 81.39
81.39 to 85.14
WOODLANDS
YISHUN
PUNGGOL
SERANGOON
HOUGANG
BUKIT BATOK
TOA PAYOH
TAMPINES
BEDOK
JURONG EAST
JURONG EAST
QUEENSTOWN
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

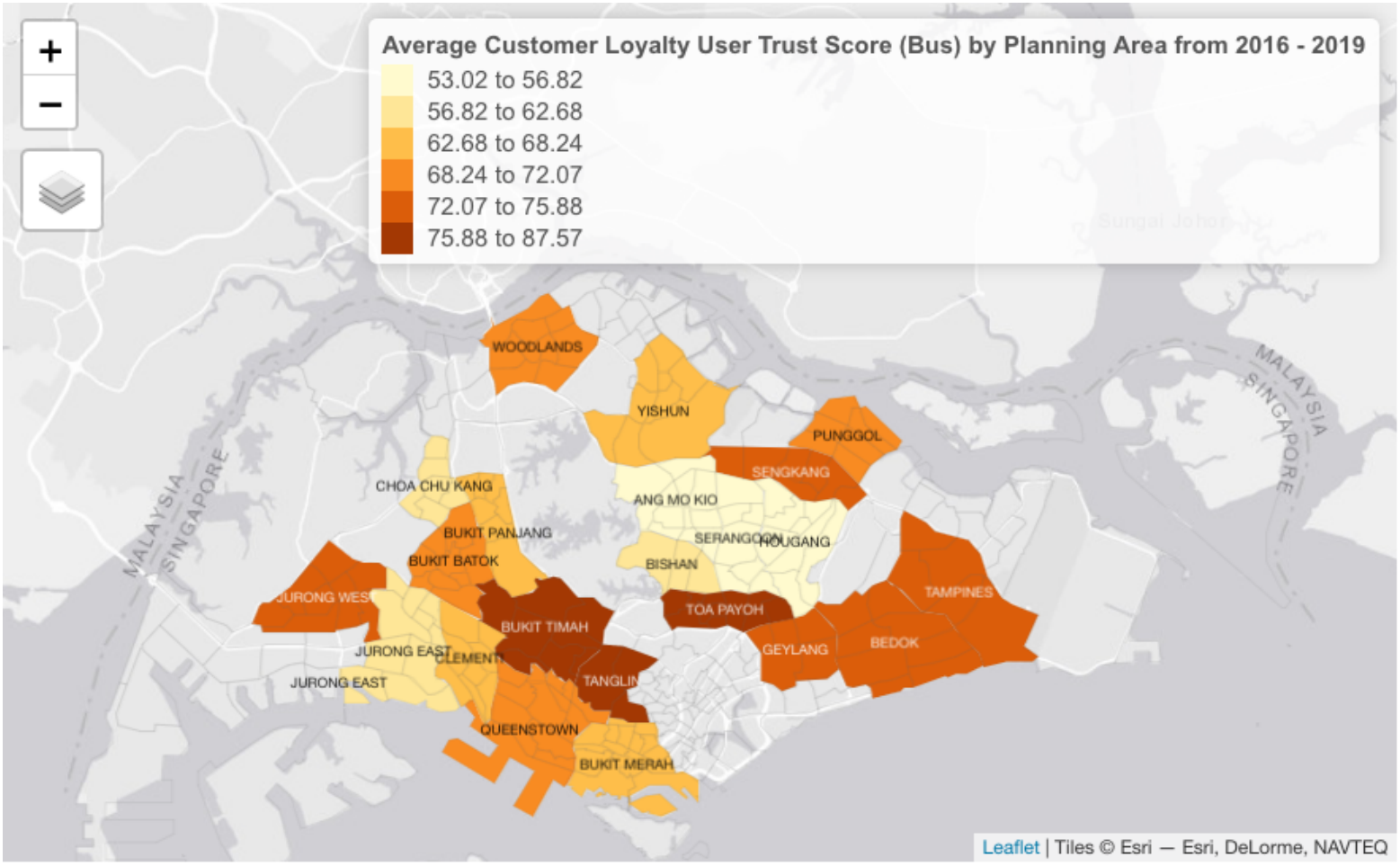

Average Customer Satisfaction Score (Bus) by Planning Area in 2019
48.16 to 48.16
48.16 to 59.24
59.24 to 65.75
65.75 to 69.94
69.94 to 78.25
78.25 to 86.09
WOODLANDS
YISHUN
PUNGGOL
SENGKANG
ANG MO KIO
HOUGANG
BUKIT BATOK
TAMPINES
BEDOK
BUKIT MERAH
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

11.9.2.3 Customer Loyalty User Trust



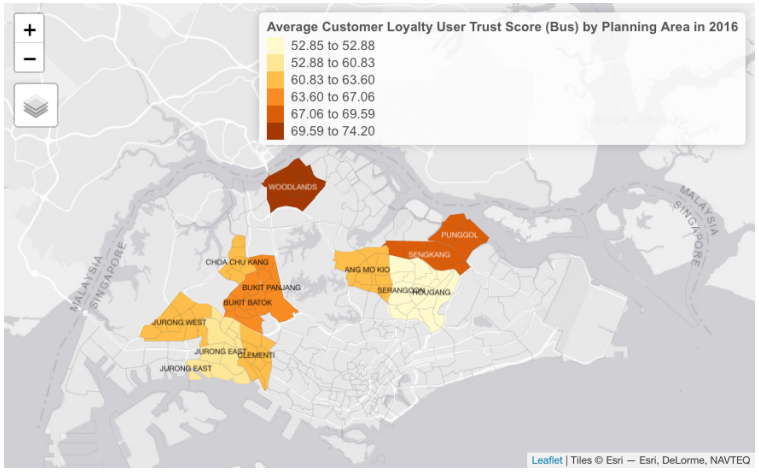
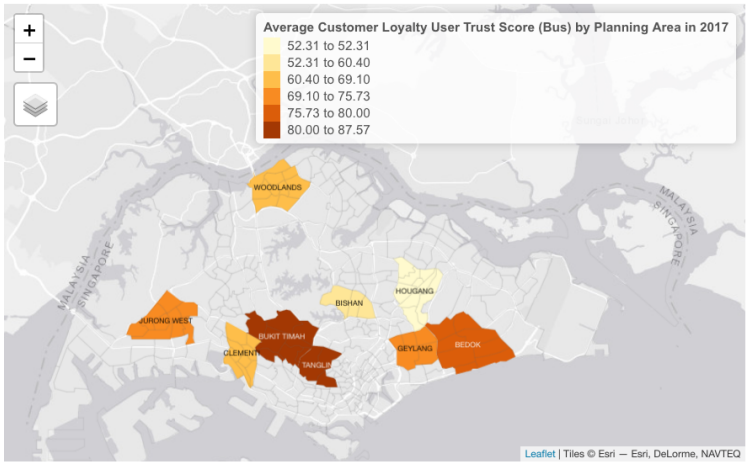
Findings 1: Positive increase over the varying years can be seen mainly in areas like *Hougang*, *Jurong West*, and *Sengkang*. However, areas like *Ang Mo Kio*, *Bedok*, *Bukit Batok*, *Tampines*, and *Yishun* can be seen with drastic drops in 2019.

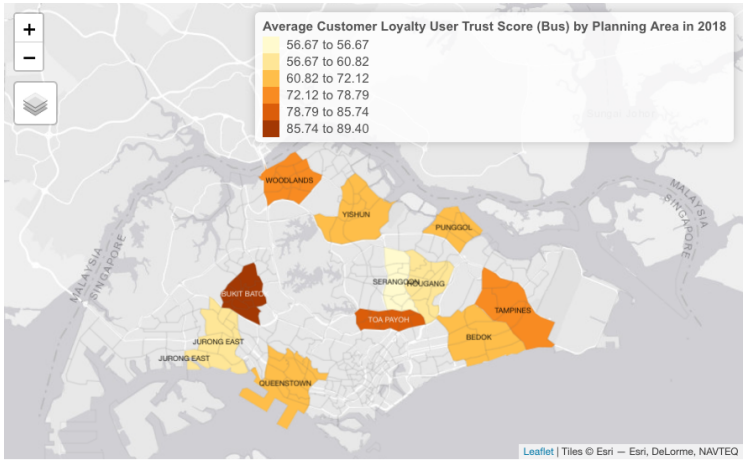
Findings 2: In general, the target audience has a scoring of at least 50, and that they would say neutral to positive things about the bus services in Singapore, excluding *Ang Mo Kio* with a score of 48.2 in 2019.

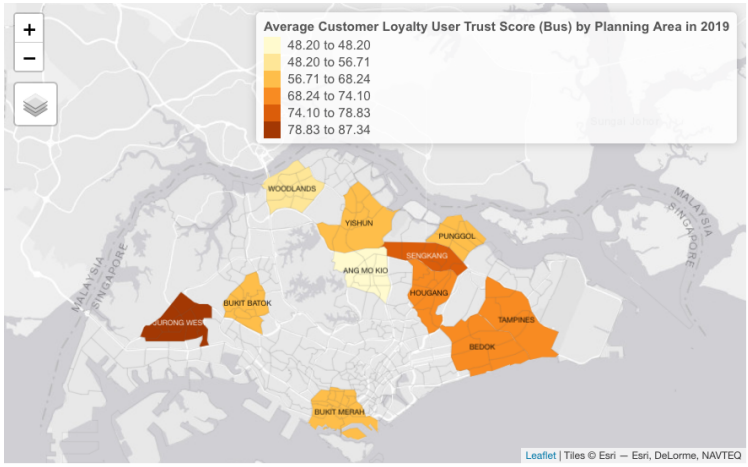


Findings 3: Areas with higher average customer loyalty user trust scores could be predominantly seen in the Central area of Singapore (*Bukit Timah*, *Tanglin*, *Toa Payoh*). This phenomenon could be seen as opposed to the average satisfaction scores of bus services and this could infer that customer loyalty user trust scores are immensely affecting their satisfaction scores in these areas.

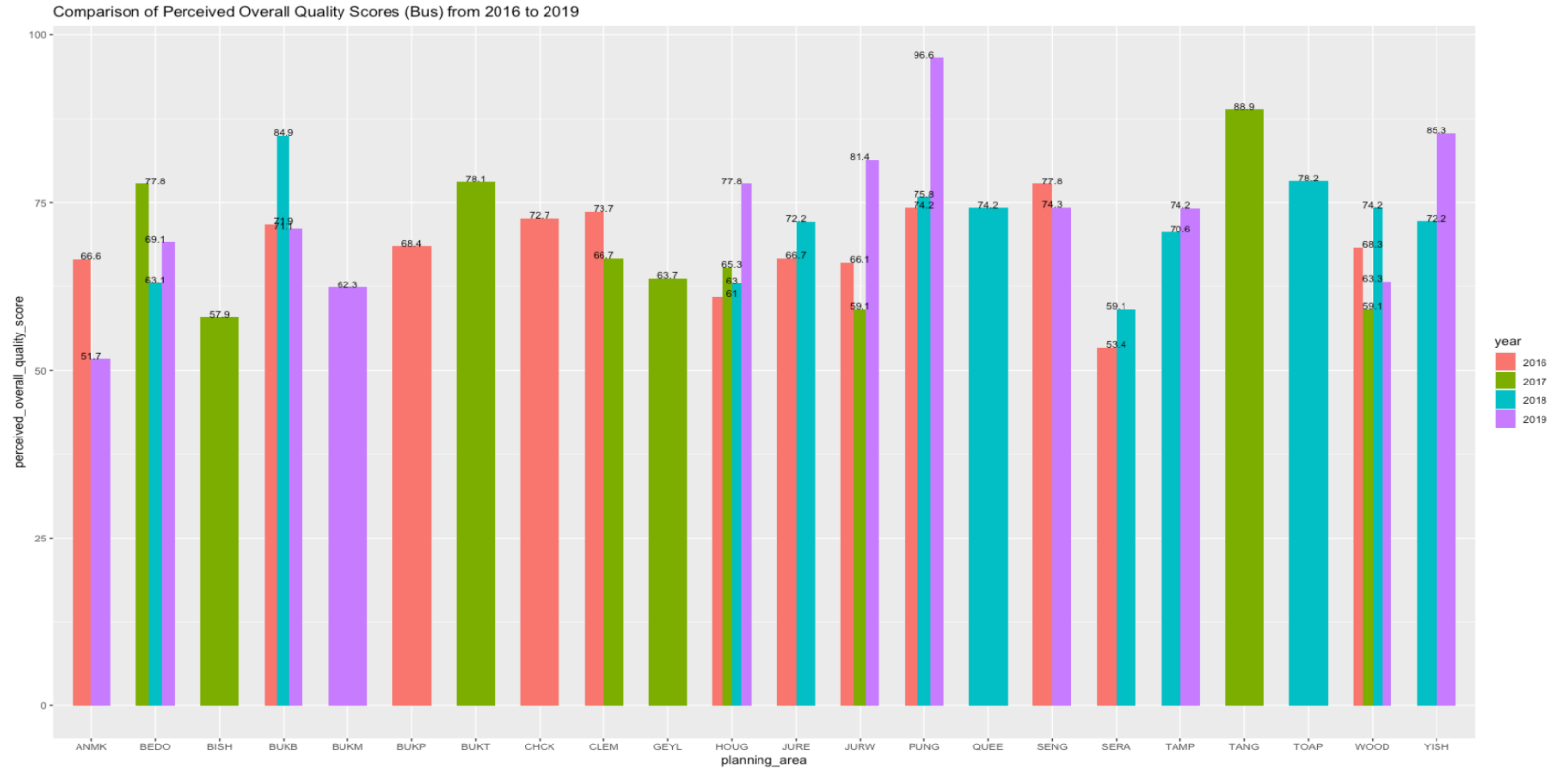
*Breakdown of Customer **Loyalty User Trust** Scores by Year and Planning Area*

2016	2017
	
2018	2019


Average Customer Loyalty User Trust Score (Bus) by Planning Area in 2018
56.67 to 56.67
56.67 to 60.82
60.82 to 72.12
72.12 to 78.79
78.79 to 85.74
85.74 to 89.40
WOODLANDS
YISHUN
PUNGGOL
SERANGOON
HOUGANG
BUKIT BATO
TOA PAYOH
TAMPINES
BEDOK
JURONG EAST
JURONG EAST
QUEENSTOWN
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

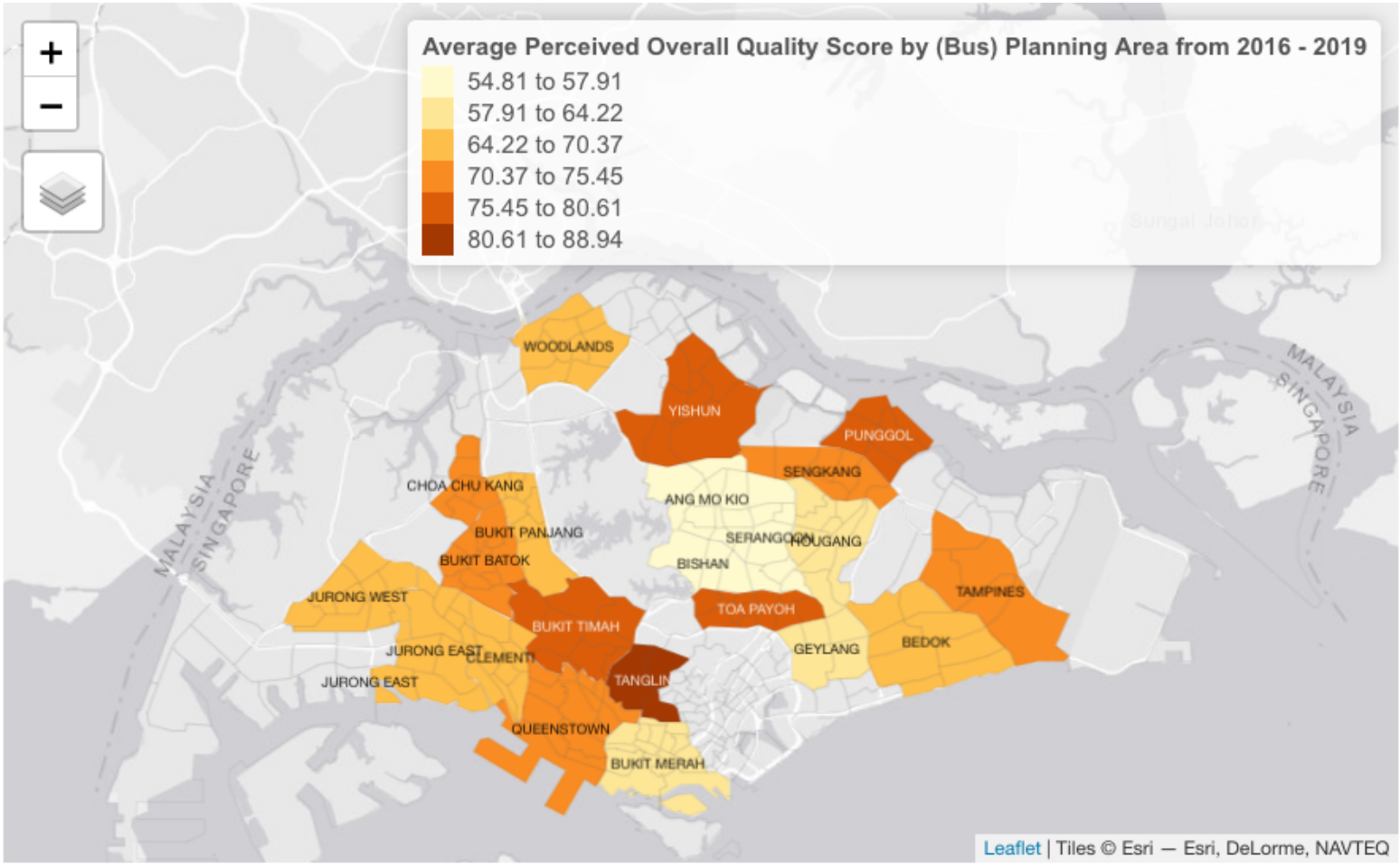

Average Customer Loyalty User Trust Score (Bus) by Planning Area in 2019
48.20 to 48.20
48.20 to 56.71
56.71 to 68.24
68.24 to 74.10
74.10 to 78.83
78.83 to 87.34
WOODLANDS
YISHUN
PUNGGOL
SENGKANG
ANG MO KIO
HOUGANG
BUKIT BATOK
JURONG WEST
TAMPINES
BEDOK
BUKIT MERAH
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

11.9.2.4 Perceived Overall Quality



Findings 1: Generally, the target audience felt positive about the overall quality they received from bus services yearly.

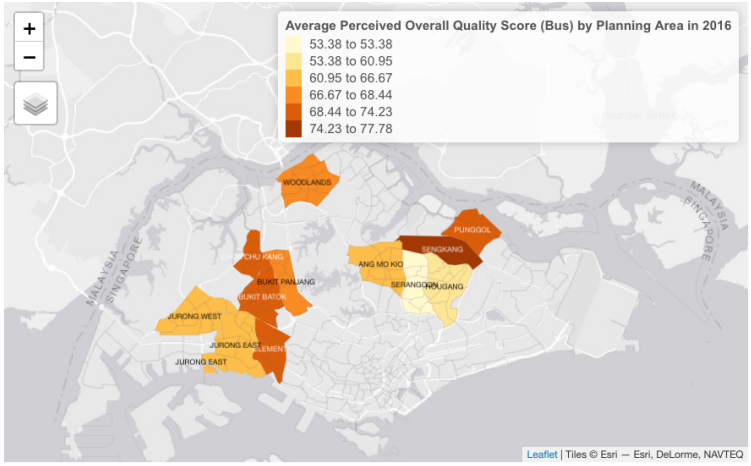
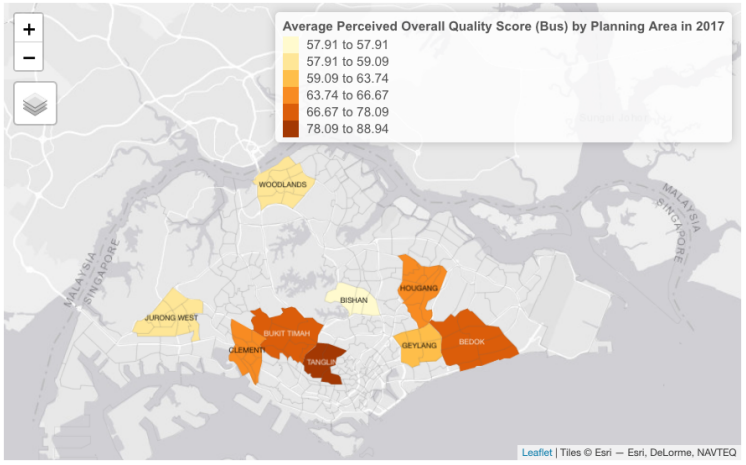
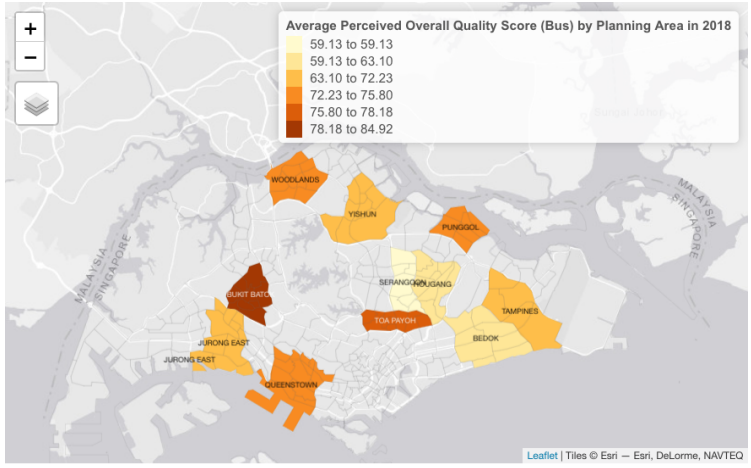
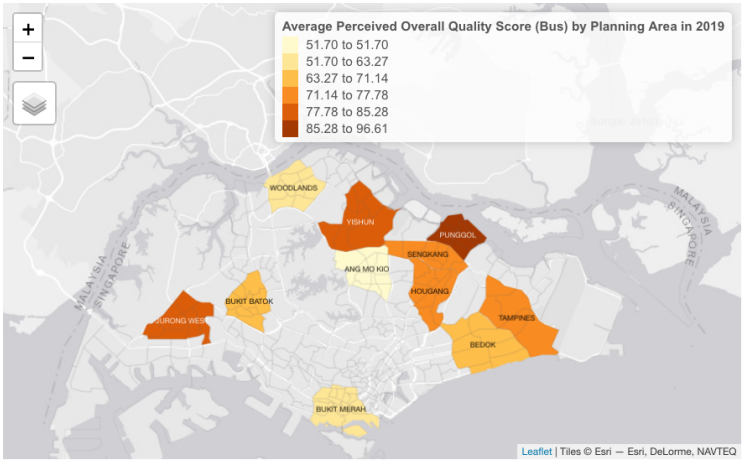
Findings 2: Areas like *Bedok*, *Bukit Batok* and *Woodlands* could be seen with varying perceptions in the quality they received, and if the service provider did meet their personal requirements.



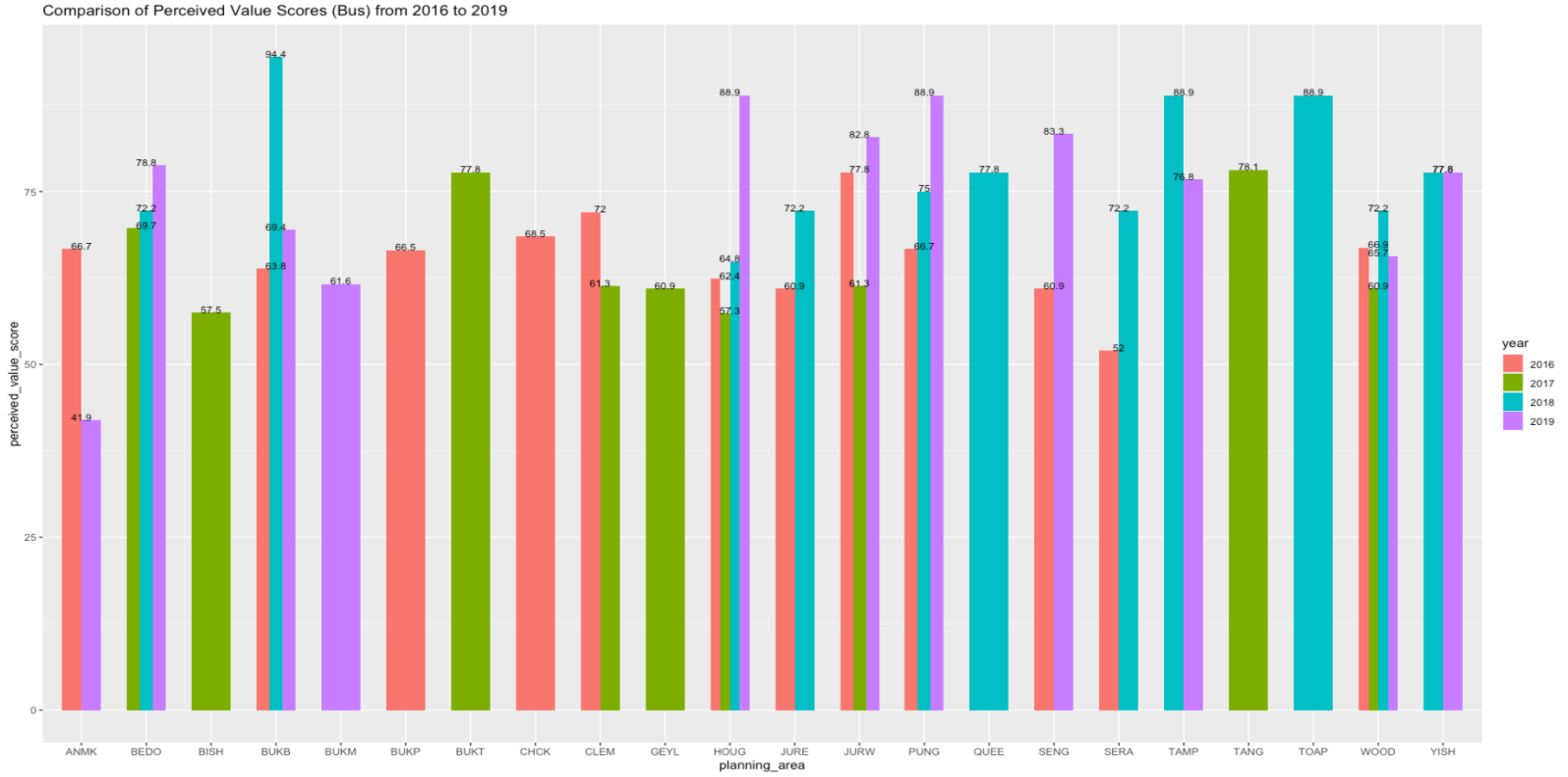
Findings 3: *Tanglin* (88.94) could be seen with the highest perceived overall quality score from 2016 to 2019.

Findings 4: The target audience in all the respective planning areas could be seen with a positive score of more than 50 from 2016 to 2019.

*Breakdown of **Perceived Overall Quality** Scores by Year and Planning Area*

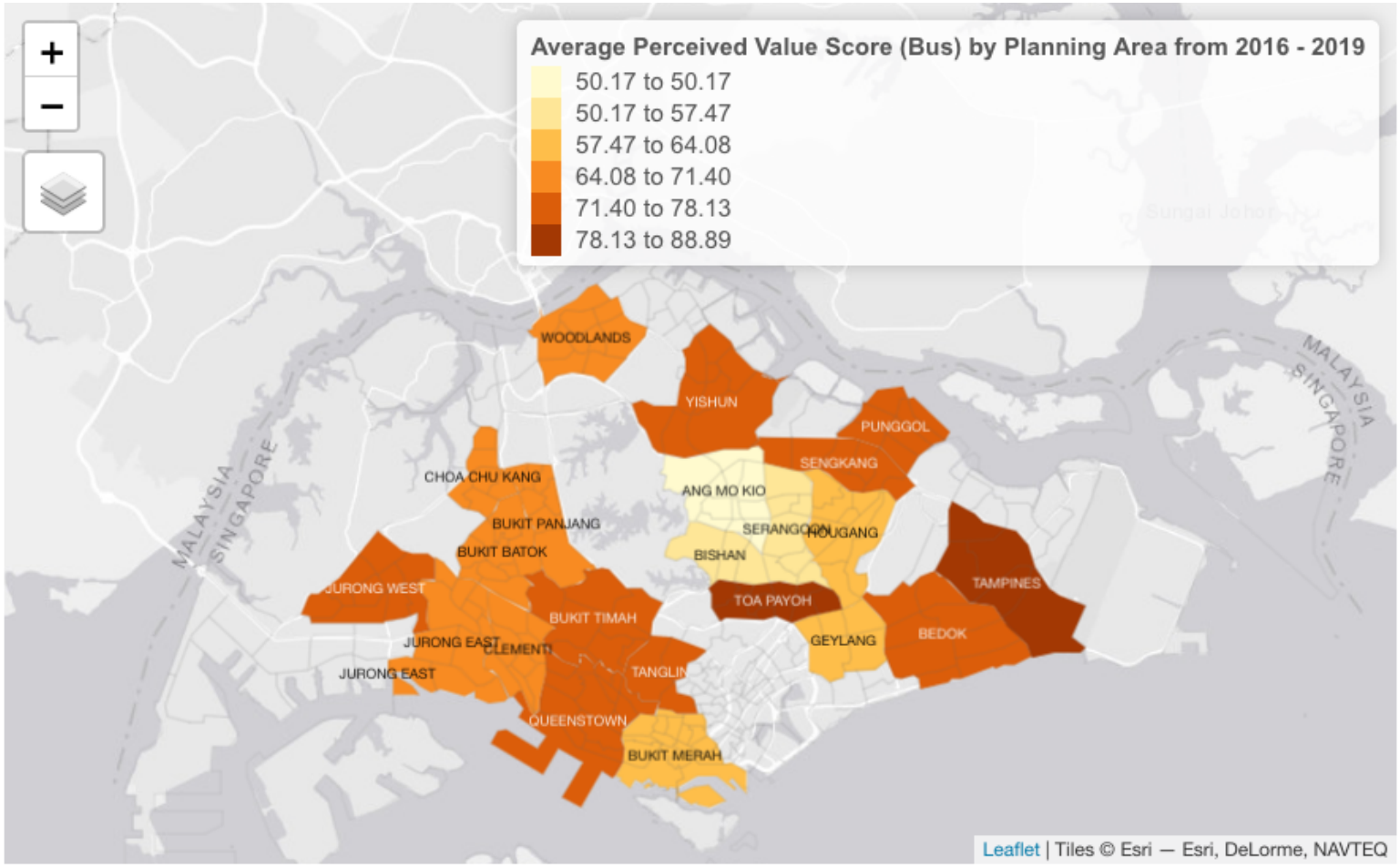
2016	2017
	
2018	**2019**
	

11.9.2.5 Perceived Value



Findings 1: Areas like *Bedok*, *Hougang*, Jurong East*, Punggol*, *Sengkang* and *Serangoon* have a positive increase in the perceptions of quality and pricing of the bus services.

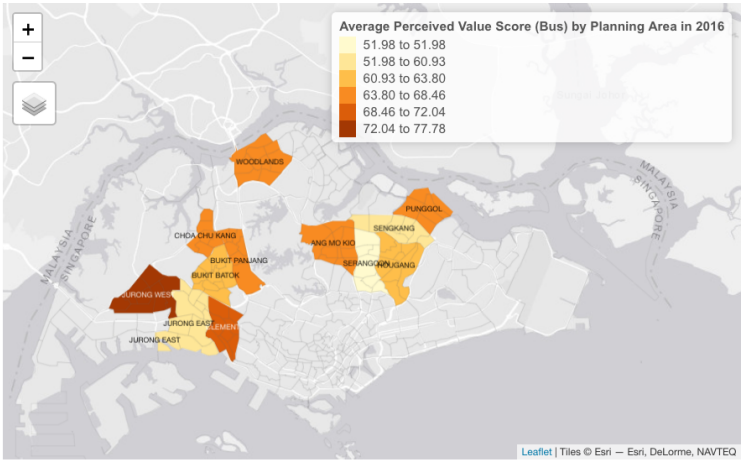
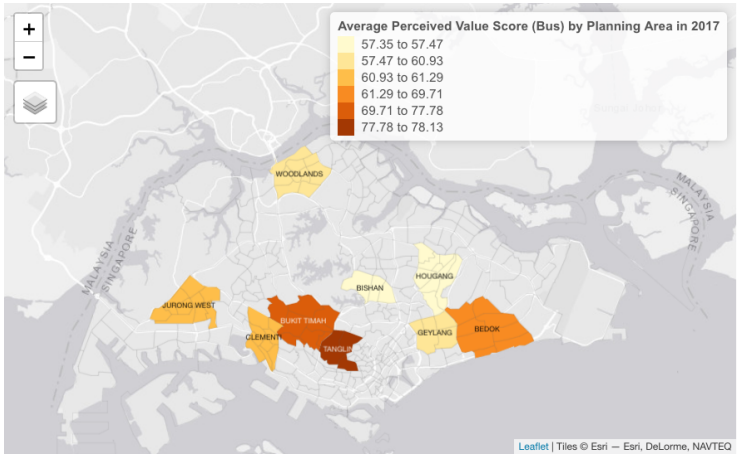
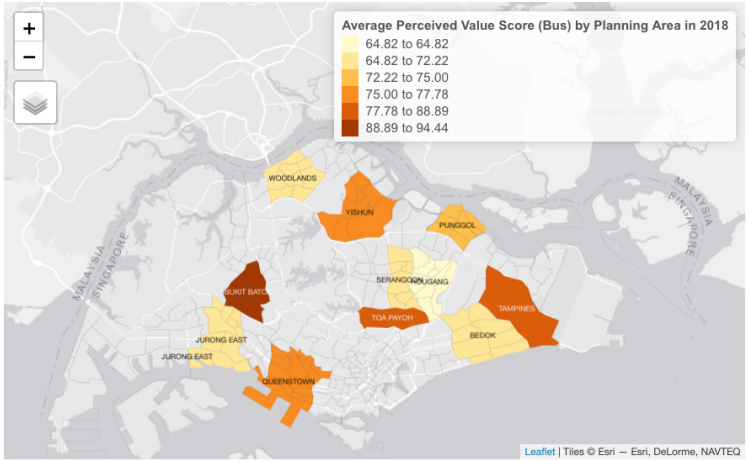
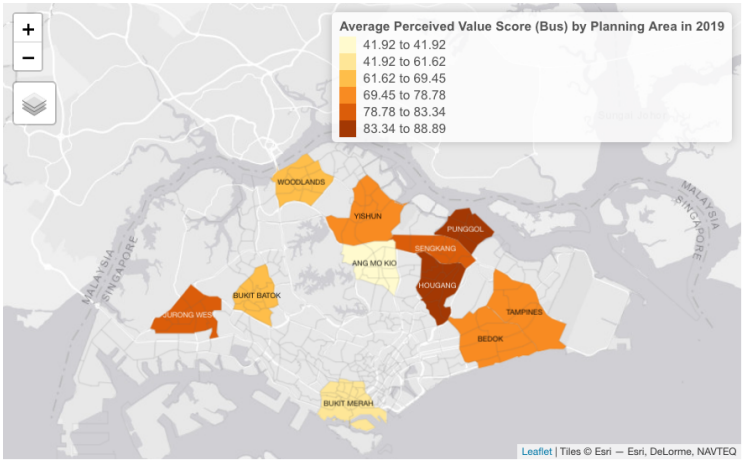
Findings 2: It could be seen that there has been varying differences in the perceived value scores of the target audience over the years. In 2018, there have been areas with a huge spike and sudden drops in their perceived value scores. For example, *Bukit Batok*, increased from 63.8 in 2016 to 94.4 in 2018 and dropped back to 69.4 in 2019. And, *Jurong West*, from 77.8 in 2016, to a drop of 61.3 in 2017, and growth to 82.8 in 2019.



Findings 3: An outlier, *Ang Mo Kio*, could be spotted, with a low score of 50.17 of their average perceived value score from 2016 to 2019.

Findings 4: Generally, the target audience felt positive towards the perceptions of quality and pricing of the bus services in Singapore.

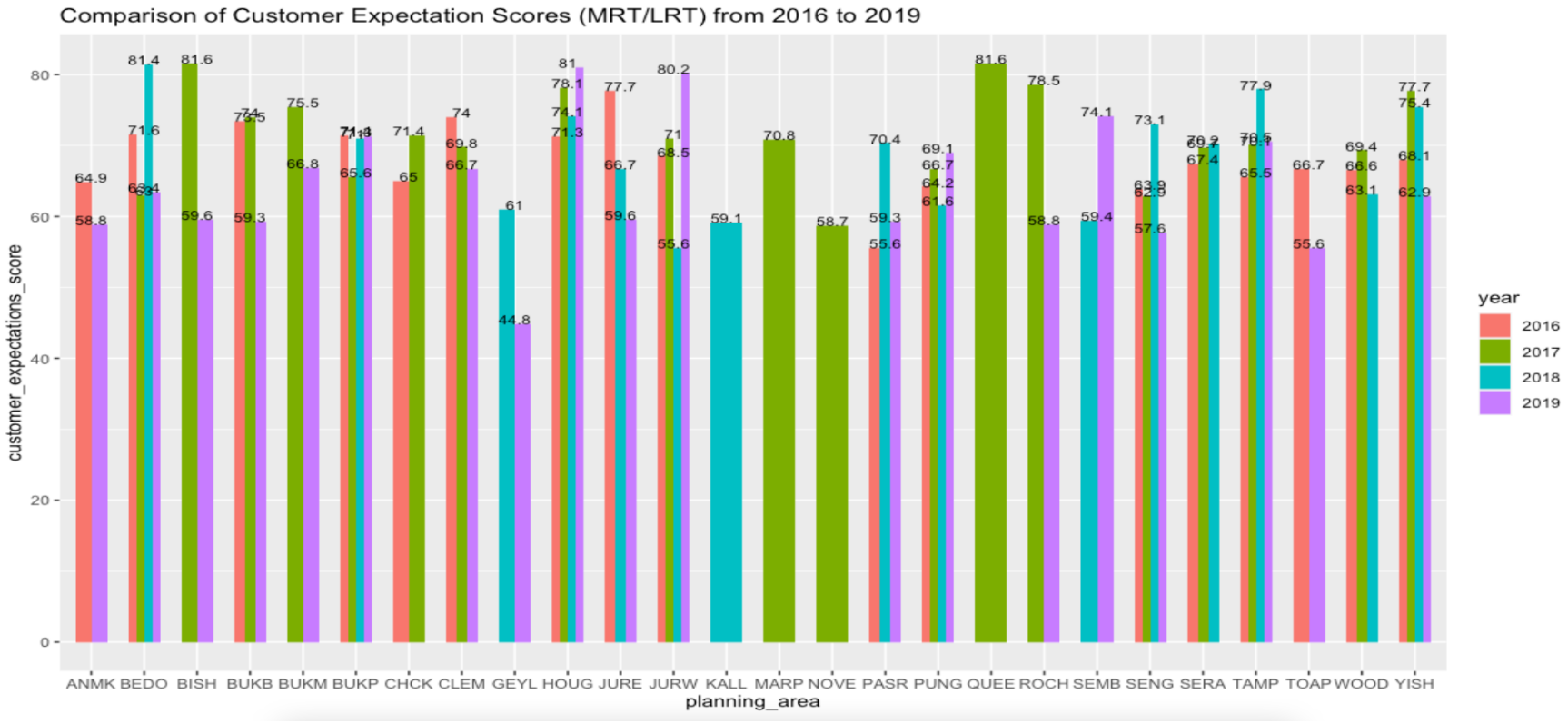
*Breakdown of **Perceived Value** Scores by Year and Planning Area*

2016	2017
	
2018	**2019**
	

11.9.3 Service Sector: ***MRT/LRT***

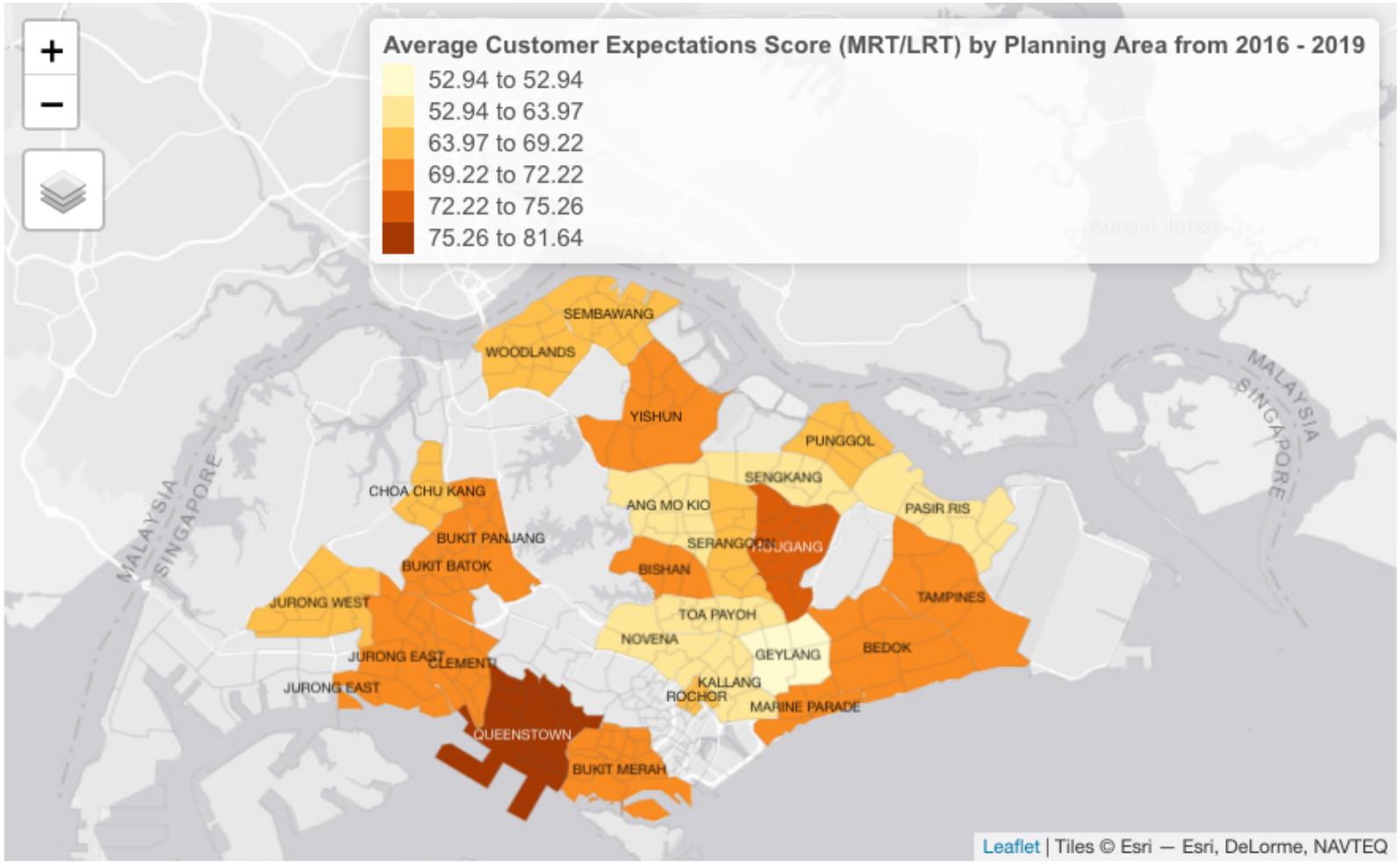
In the following comparison charts, the team will be looking into details of each of the respective scoring and how it changes from Year 2016 to 2019.

11.9.3.1 Customer Expectation



Findings 1: It could be seen that the target audience has a decrease in their expectations scores for the MRT/LRT service in 2019. However, there are areas that could be seen with an increase in their expectation scores, for example, *Hougang*, *Jurong West*, *Punggol* and *Sembawang*.

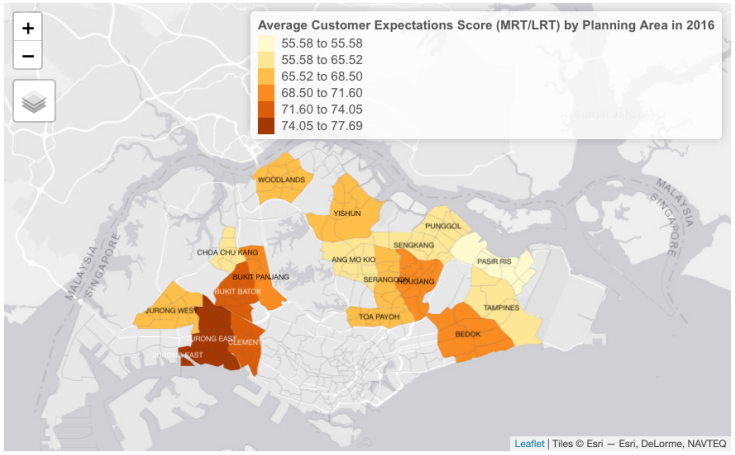
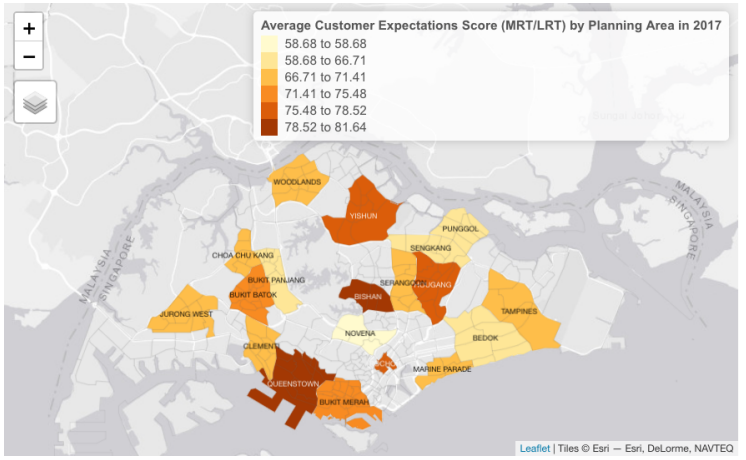
Findings 2: *Geylang* could be an outlier as it has the lowest expectation score of 44.8 in 2019. This could be due to the fact that there are no MRT/LRT services in this area.

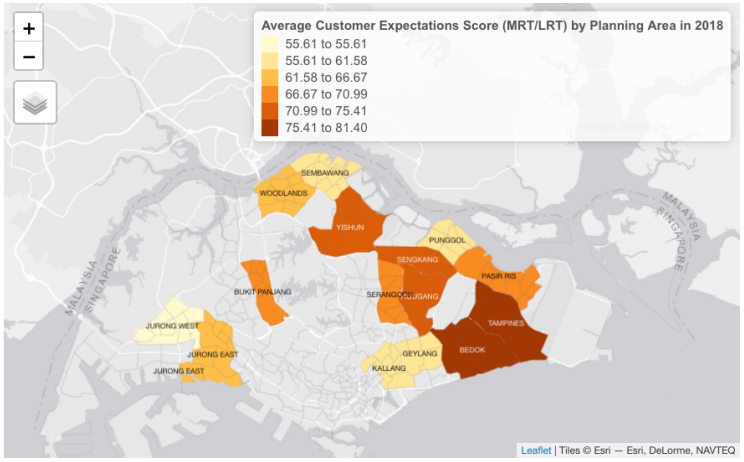


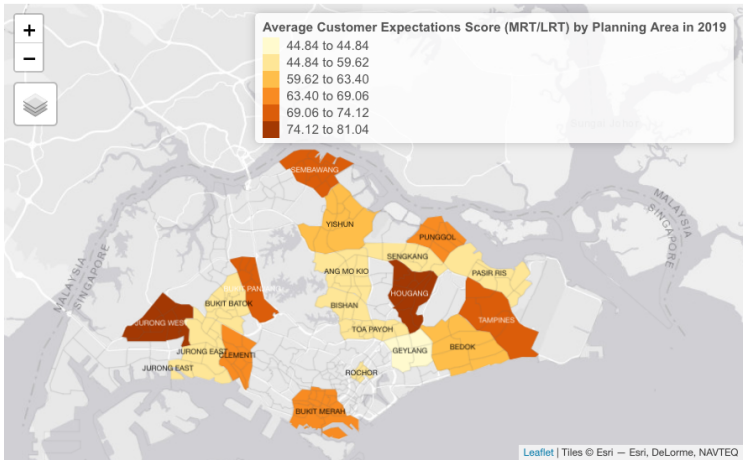
Findings 3: *Queenstown* could be seen with the highest average expectation score (81.64) towards the MRT/LRT services in Singapore, from 2016 to 2019.

Findings 4: Juxtaposed to the previous bar chart, *Geylang* is still the area with the lowest average expectation score over the years.

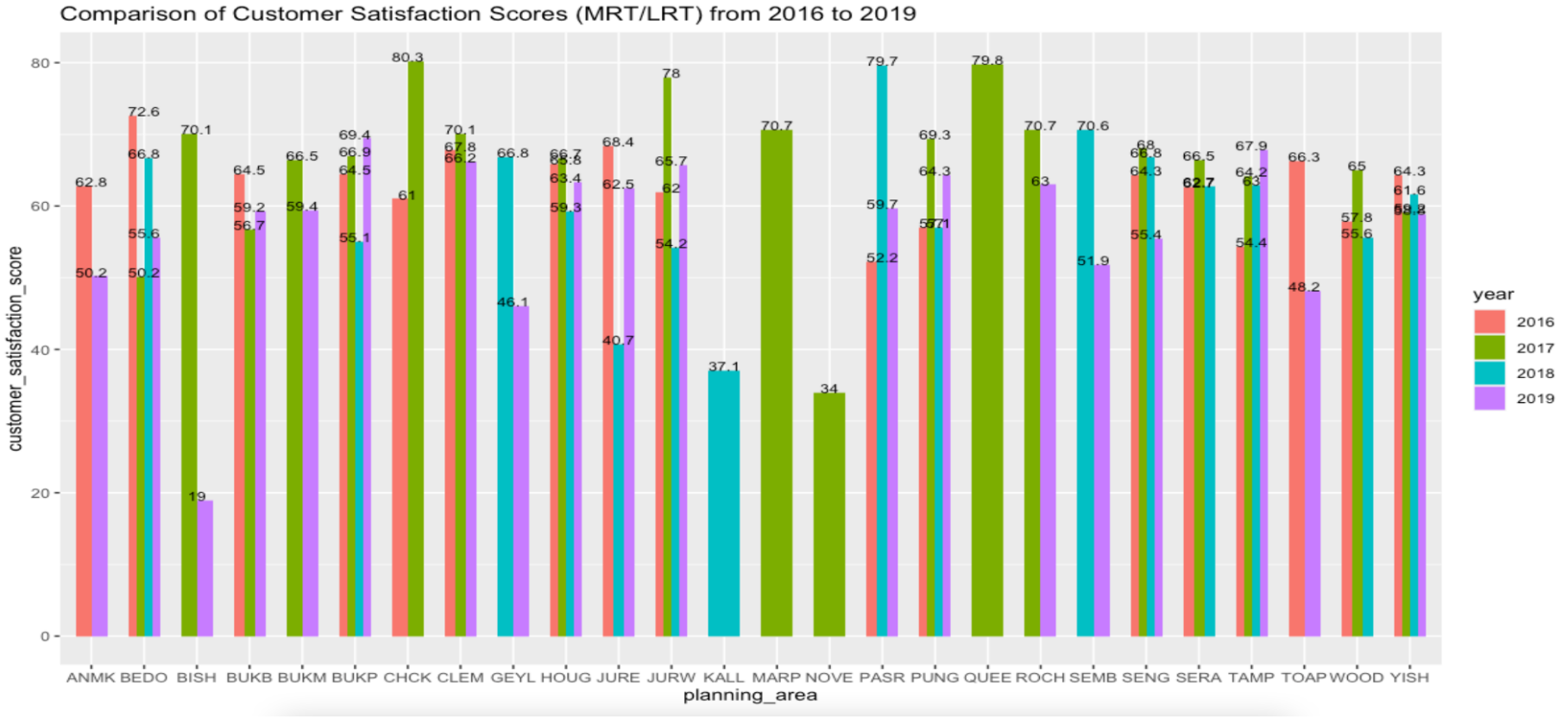
*Breakdown of Customer **Expectation** Scores by Year and Planning Area*

2016	2017
	
2018	2019


Average Customer Expectations Score (MRT/LRT) by Planning Area in 2018
55.61 to 55.61
55.61 to 61.58
61.58 to 66.67
66.67 to 70.99
70.99 to 75.41
75.41 to 81.40
SEMBAWANG
WOODLANDS
YISHUN
PUNGGOL
SENGKANG
PASIR RIS
BUKIT PANJANG
SERANGOON
HOUGANG
TAMPINES
JURONG WEST
JURONG EAST
GEYLANG
BEDOK
KALLANG
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

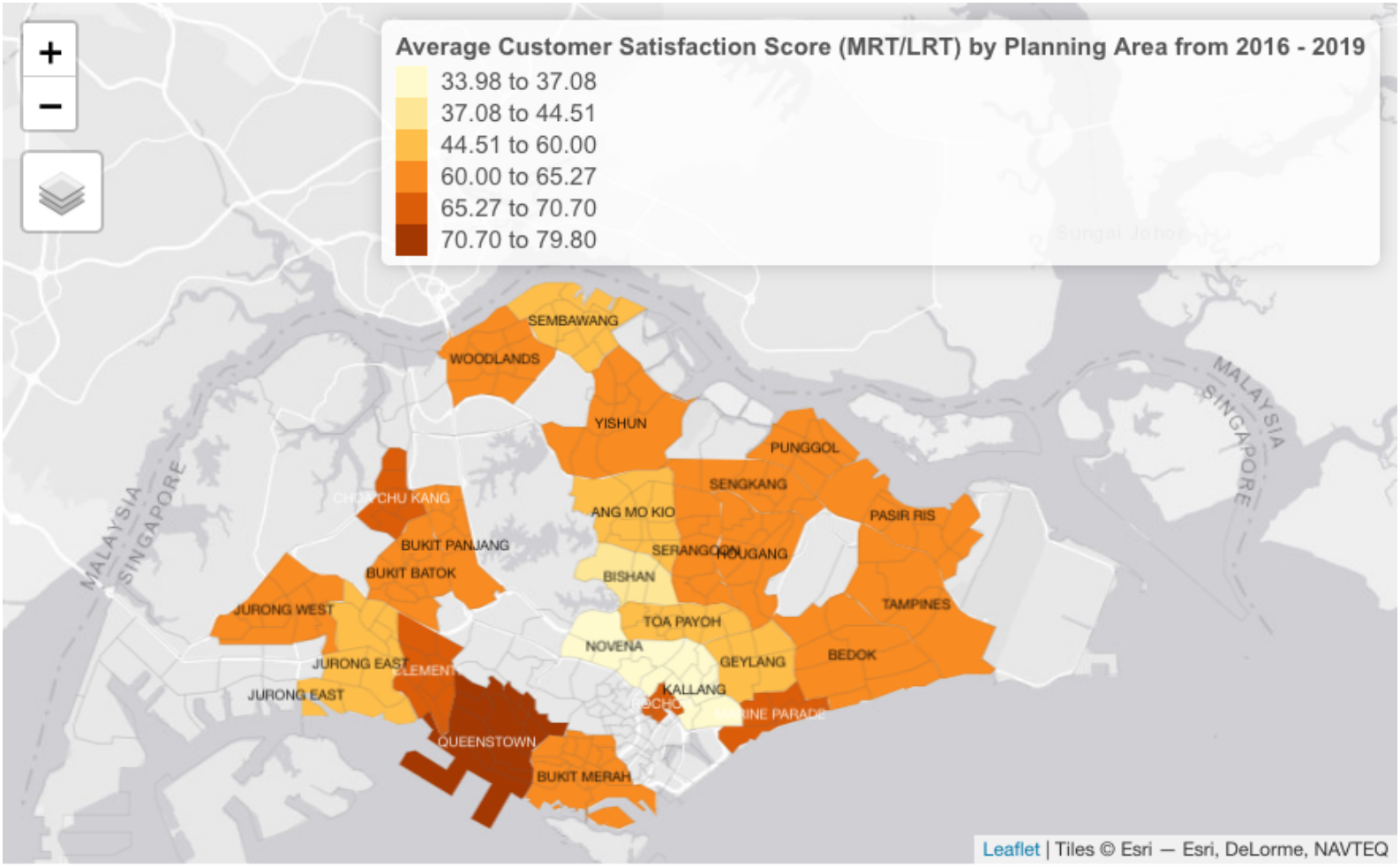

Average Customer Expectations Score (MRT/LRT) by Planning Area in 2019
44.84 to 44.84
44.84 to 59.62
59.62 to 63.40
63.40 to 69.06
69.06 to 74.12
74.12 to 81.04
SEMBAWANG
YISHUN
PUNGGOL
SENGKANG
ANG MO KIO
PASIR RIS
HOUGANG
BISHAN
BUKIT BATOK
TAMPINES
TOA PAYOH
JURONG WEST
JURONG EAST
GEYLANG
BEDOK
CLEMENTI
ROCHOR
BUKIT MERAH
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

11.9.3.2 Customer Satisfaction



Findings 1: An outlier, *Bishan*, could be spotted with the all-time lowest satisfaction score of 19 in 2019.

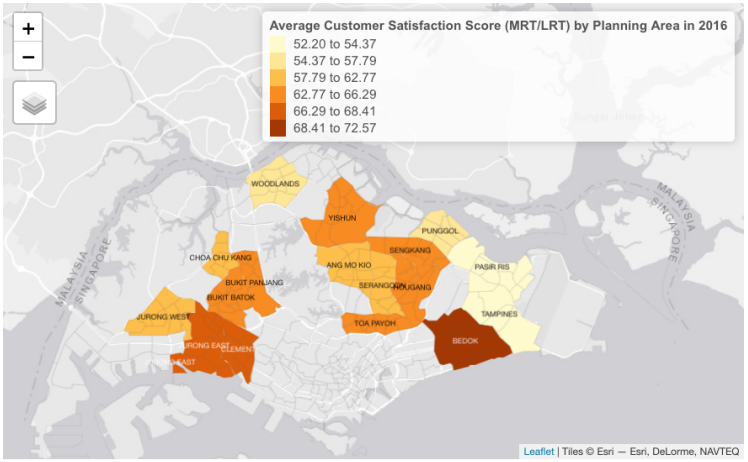
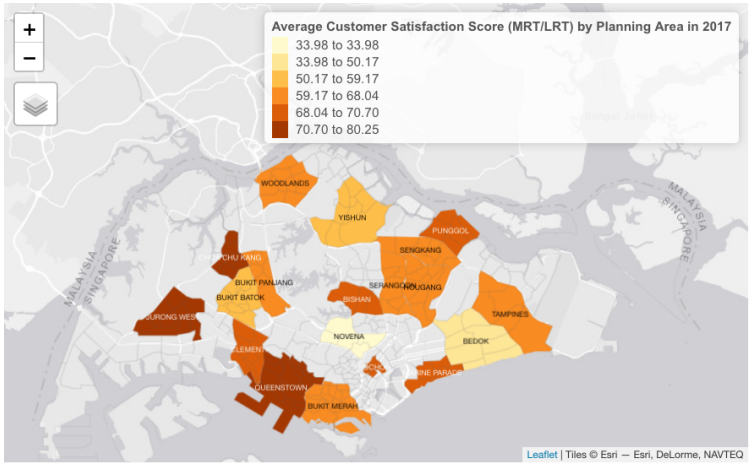
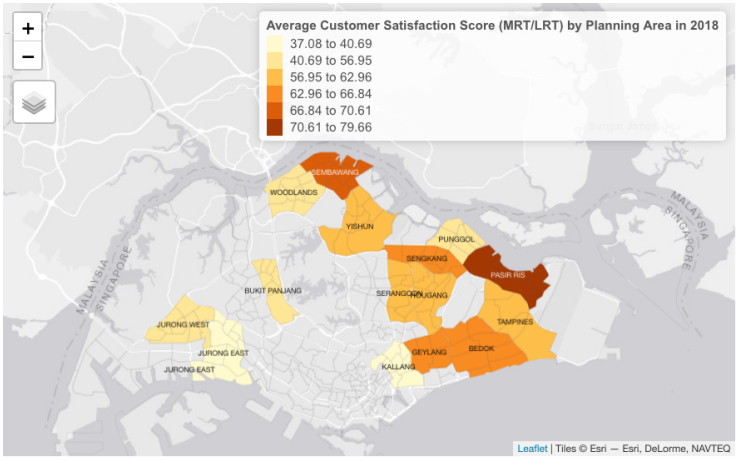
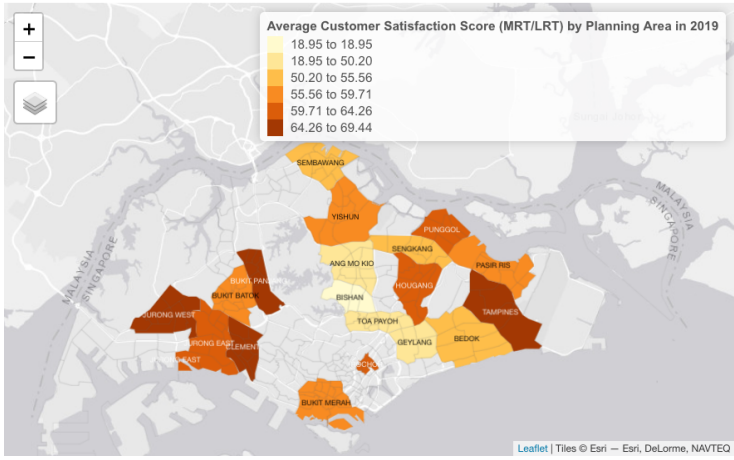
Findings 2: The overall satisfaction levels in the various planning areas have extreme differences of their satisfaction scores from year to year. Furthermore, it could be seen that there is a predominance of a major decrease in their satisfaction levels in 2019. This phenomenon could be a possible effect of the numerous MRT/LRT breakdowns in 2019.



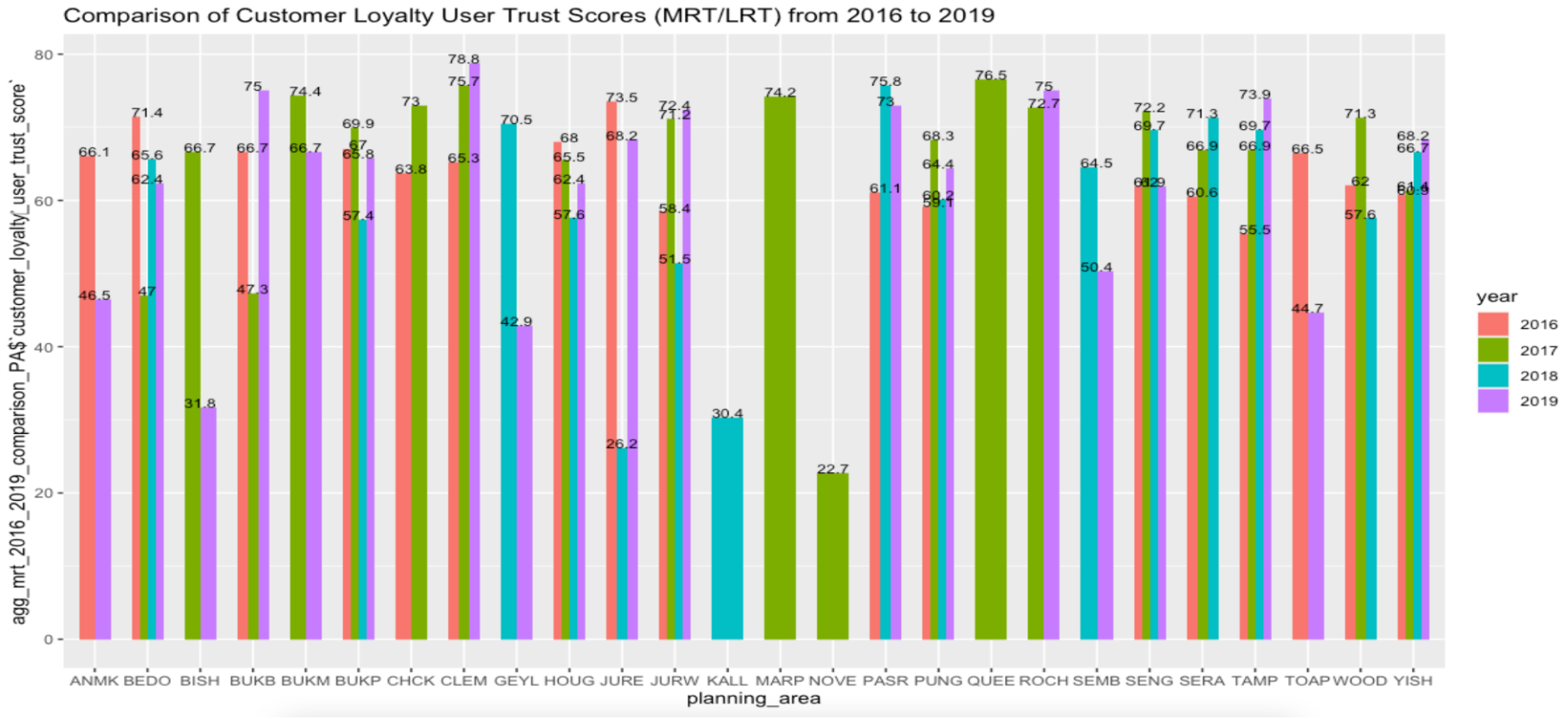
Findings 3: *Queenstown* could be seen with the highest average satisfaction score of 79.8 from 2016 to 2019.

Findings 4: Areas like *Novena*, *Kallang* and *Bishan* are the main few that are highly dissatisfied with the MRT/LRT services over the years.

*Breakdown of Customer **Satisfaction** Scores by Year and Planning Area*

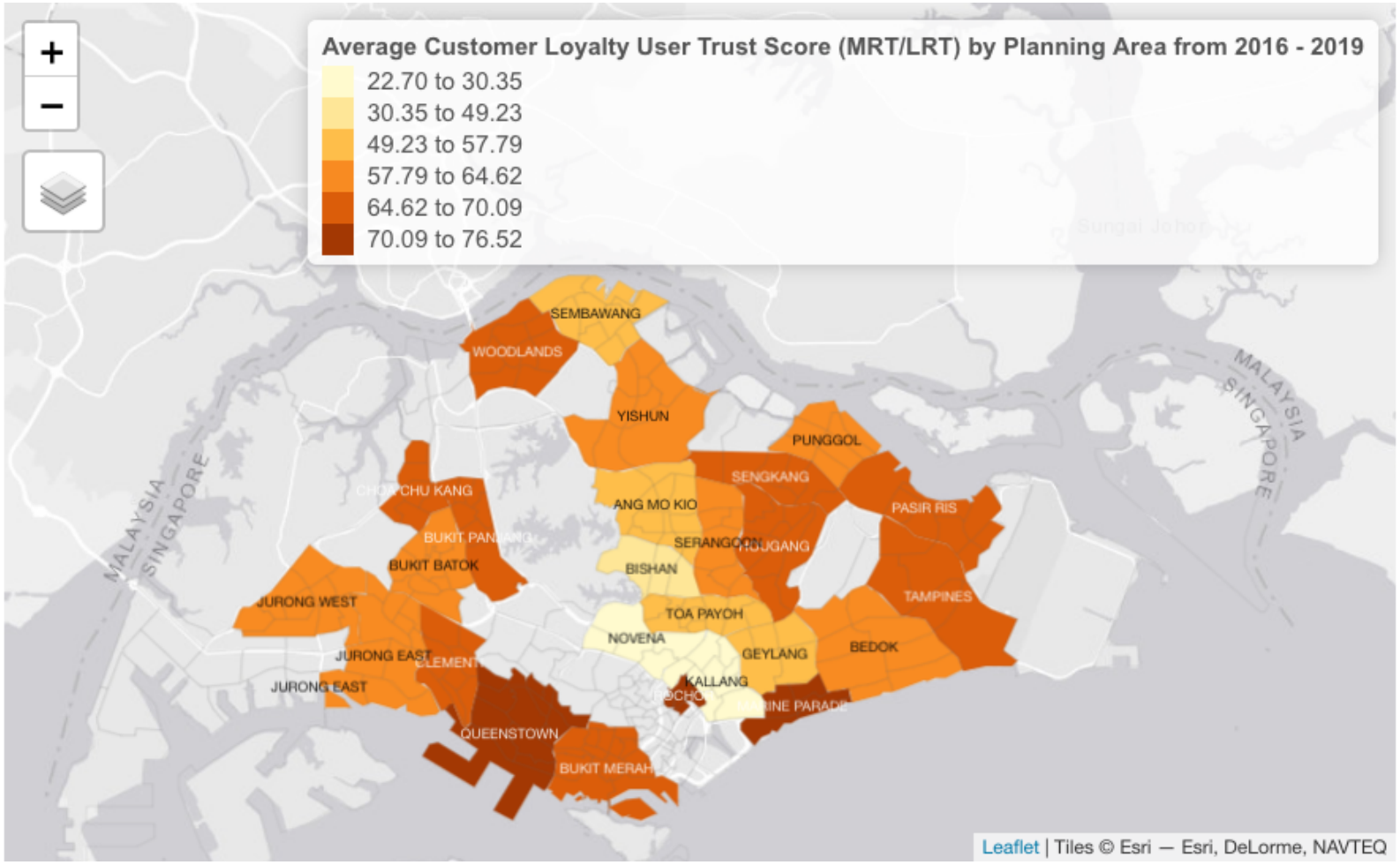
2016	2017
	
2018	**2019**
	

11.9.3.3 Customer Loyalty User Trust



Findings 1: There has been a decrease of its customer loyalty user trust scores in numerous areas over the respective years. However, areas like *Clementi*, *Serangoon*, *Tampines*, and *Yishun* are seen with a positive growth.

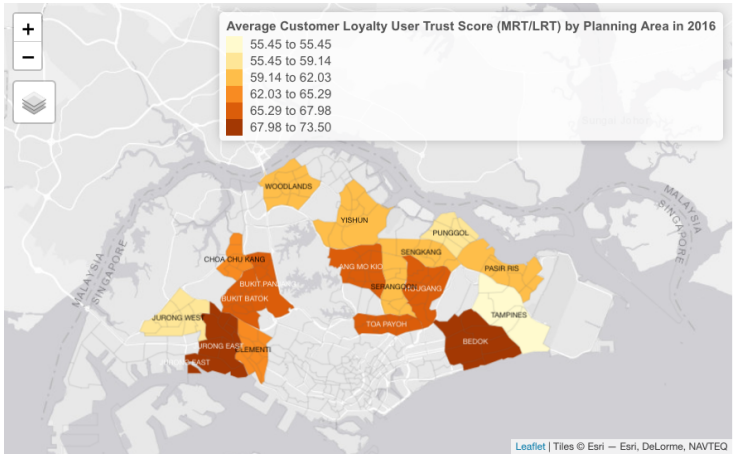
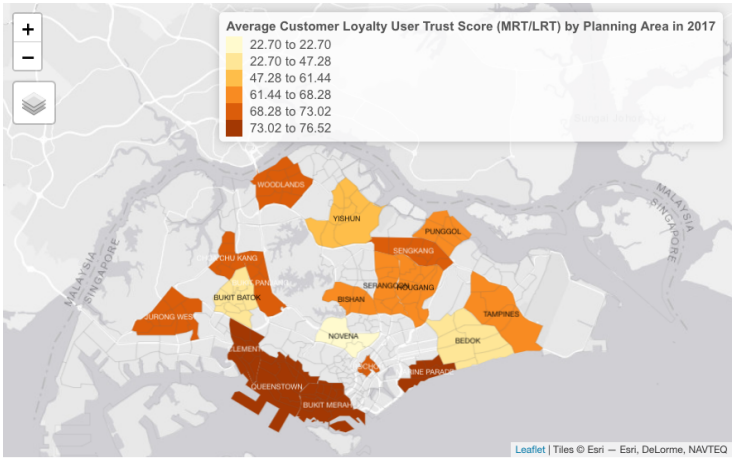
Findings 2: Overall, most would have negative opinions about the MRT/LRT services in Singapore.

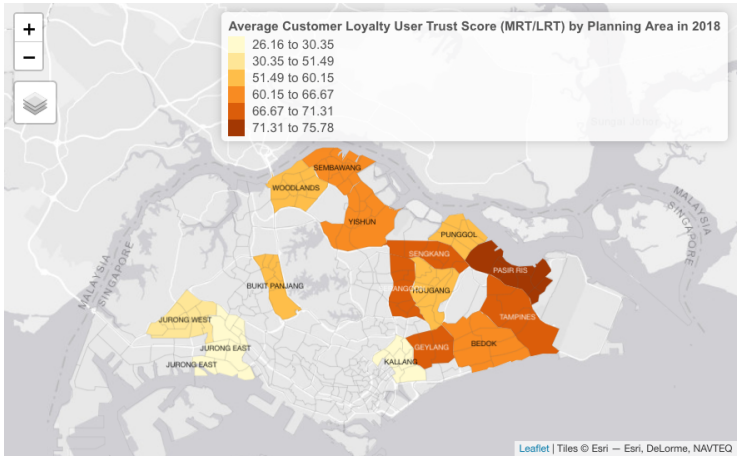


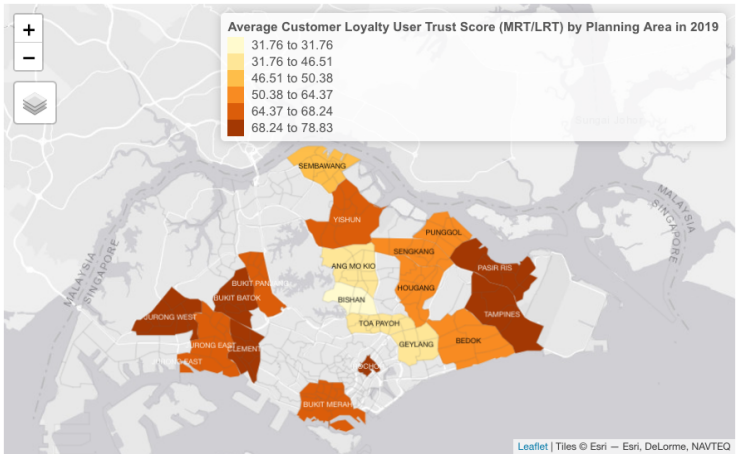
Findings 3: *Queenstown*, *Rochor* and *Marine Parade* are among those that have more positive opinions of the MRT/LRT services in Singapore.

Findings 4: More than half of the planning areas have negative things to say about the MRT/LRT services in Singapore from 2016 to 2019.

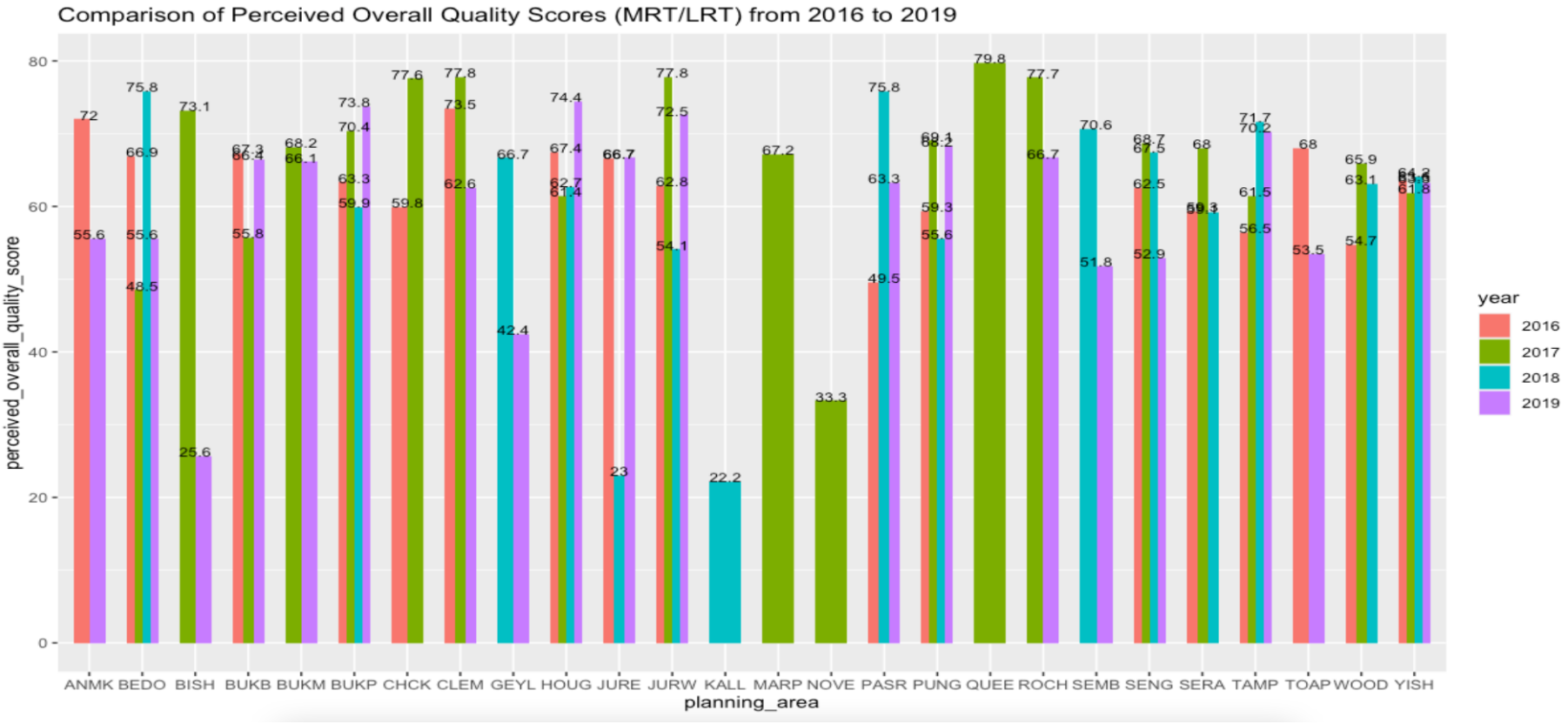
*Breakdown of Customer **Loyalty User Trust** Scores by Year and Planning Area*

2016	2017
	
2018	2019


Average Customer Loyalty User Trust Score (MRT/LRT) by Planning Area in 2018
26.16 to 30.35
30.35 to 51.49
51.49 to 60.15
60.15 to 66.67
66.67 to 71.31
71.31 to 75.78
SEMBAWANG
WOODLANDS
YISHUN
PUNGGOL
SENGKANG
PASIR RIS
BUKIT PANJANG
HOUGANG
TAMPINES
JURONG WEST
JURONG EAST
GEYLANG
BEDOK
KALLANG
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

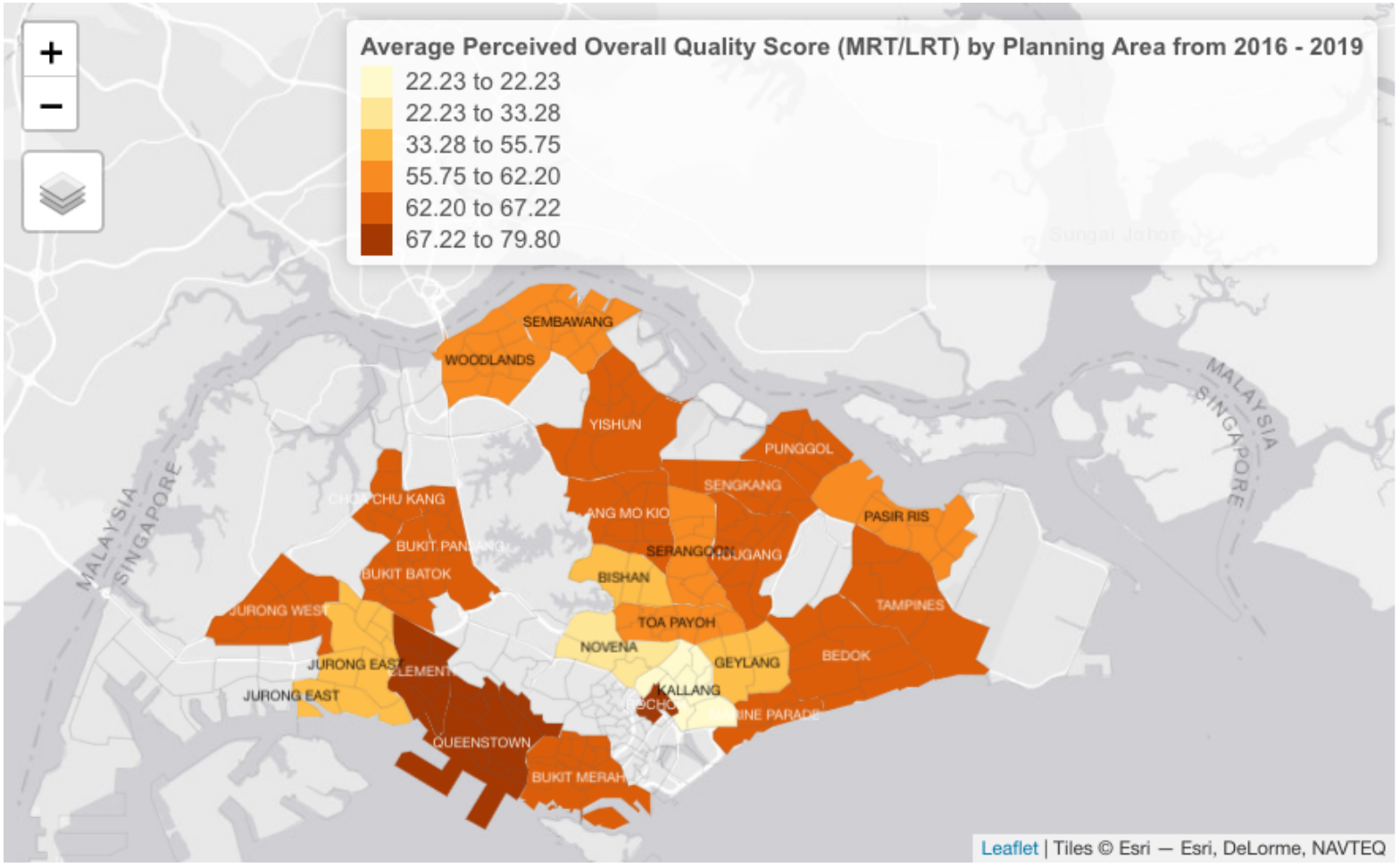

Average Customer Loyalty User Trust Score (MRT/LRT) by Planning Area in 2019
31.76 to 31.76
31.76 to 46.51
46.51 to 50.38
50.38 to 64.37
64.37 to 68.24
68.24 to 78.83
SEMBAWANG
YISHUN
PUNGGOL
SENGKANG
ANG MO KIO
PASIR RIS
HOUGANG
BISHAN
TAMPINES
TOA PAYOH
GEYLANG
BEDOK
BUKIT MERAH
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

11.9.3.4 Perceived Overall Quality



Findings 1: Generally, the target audience felt negative about the overall quality they received from the MRT/LRT services yearly.

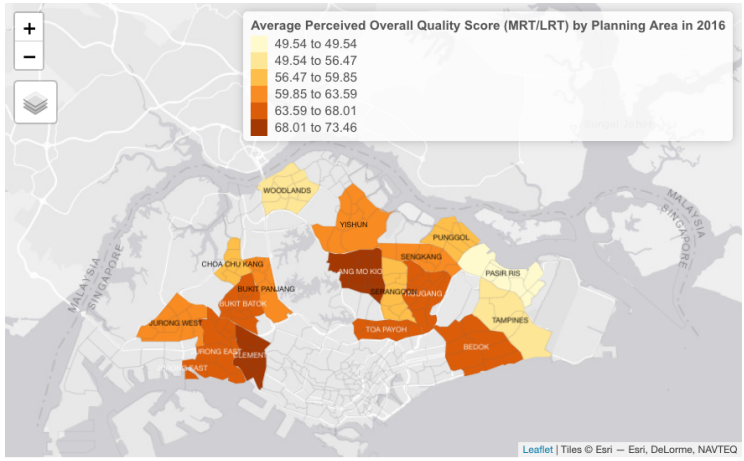
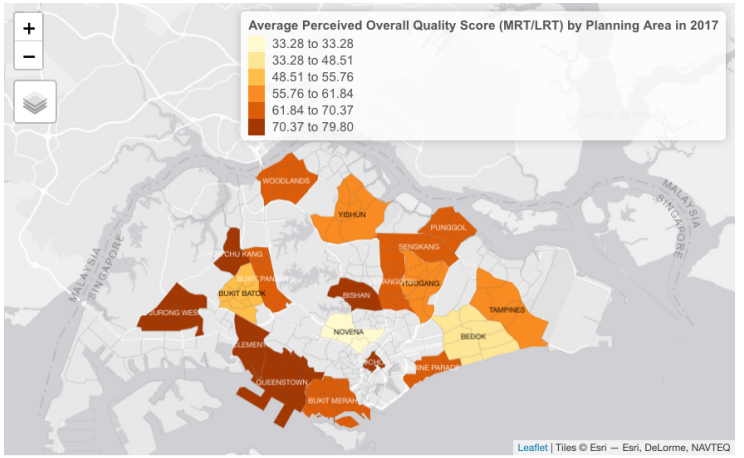
Findings 2: There is no planning area that could be seen with a positive growth of its perceived overall quality scores over the years.

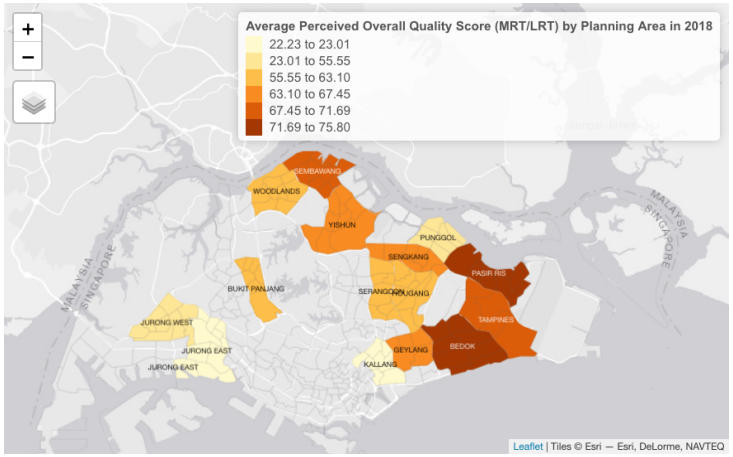


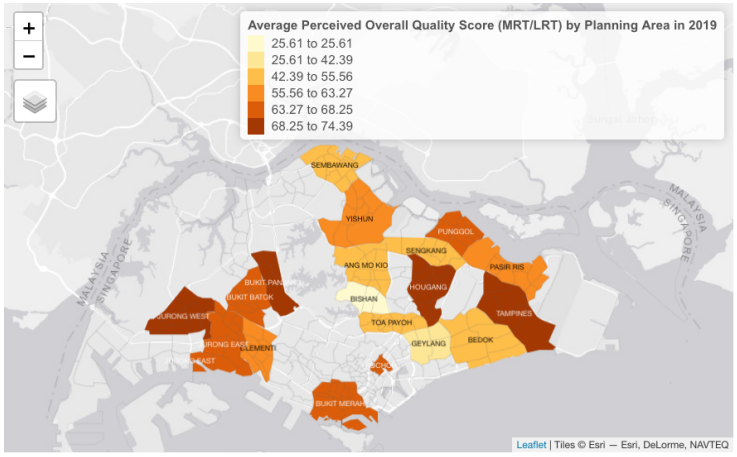
Findings 3: *Clementi*, *Queenstown* and *Rochor* are those that have the highest perceived overall quality scores from 2016 to 2019.

Findings 4: Majority of the planning areas have an average of 55.75 to 62.20 perceived overall quality scores from 2016 to 2019. This phenomenon has shown that they have neutral to negative perceptions in the quality they received, as well as that the service provider barely meets their personal requirements.

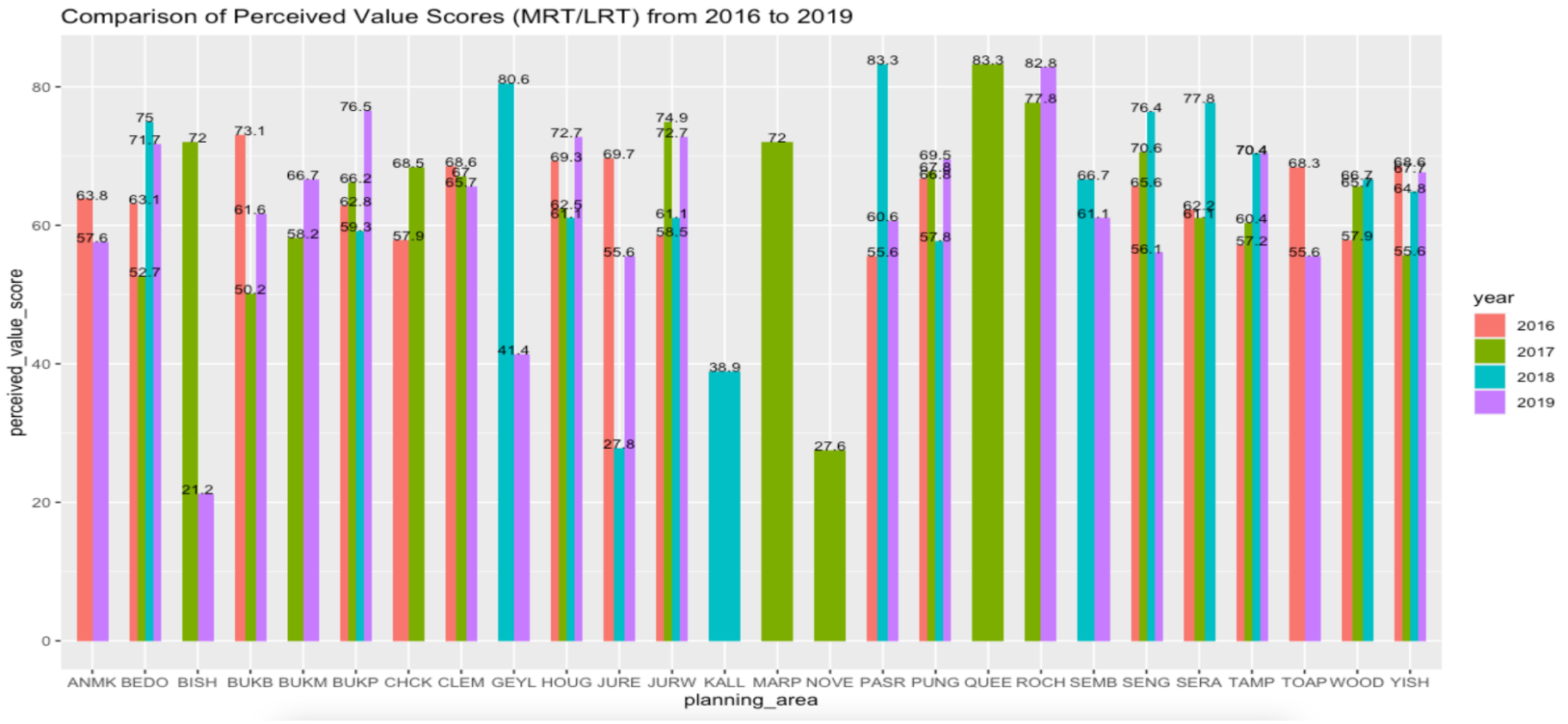
*Breakdown of **Perceived Overall Quality** Scores by Year and Planning Area*

2016	2017
	
2018	2019


Average Perceived Overall Quality Score (MRT/LRT) by Planning Area in 2018
22.23 to 23.01
23.01 to 55.55
55.55 to 63.10
63.10 to 67.45
67.45 to 71.69
71.69 to 75.80
SEMBAWANG
WOODLANDS
YISHUN
PUNGGOL
SENGKANG
PASIR RIS
BUKIT PANJANG
SERANGOON
HOUGANG
TAMPINES
JURONG WEST
JURONG EAST
GEYLANG
BEDOK
KALLANG
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

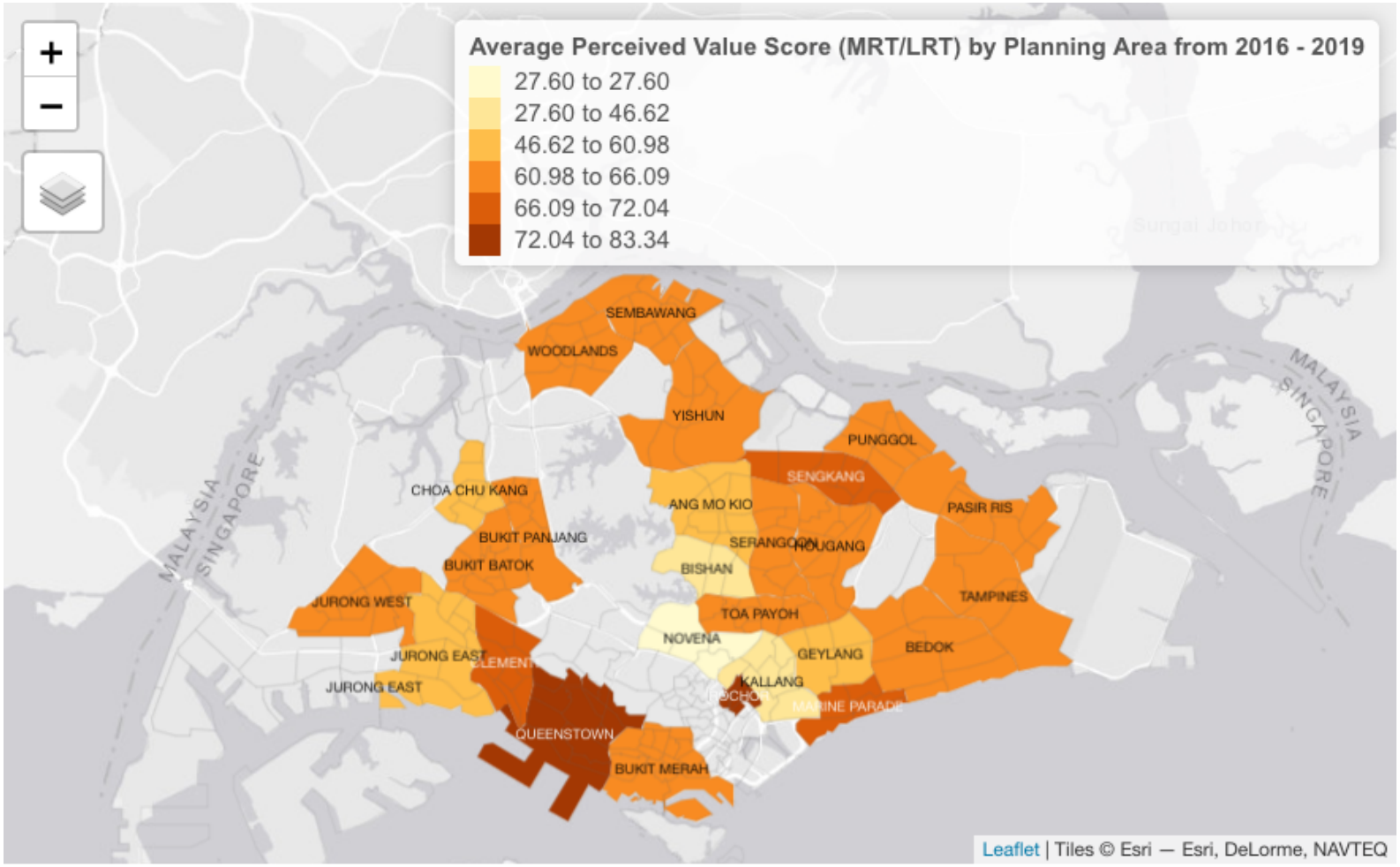

Average Perceived Overall Quality Score (MRT/LRT) by Planning Area in 2019
25.61 to 25.61
25.61 to 42.39
42.39 to 55.56
55.56 to 63.27
63.27 to 68.25
68.25 to 74.39
SEMBAWANG
YISHUN
PUNGGOL
SENGKANG
ANG MO KIO
PASIR RIS
BUKIT BATOK
HOUGANG
BISHAN
TAMPINES
JURONG WEST
TOA PAYOH
GEYLANG
BEDOK
BUKIT MERAH
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

11.9.3.5 Perceived Value



Findings 1: There has been a positive increase in the perceived value scores across the various planning areas in 2019.

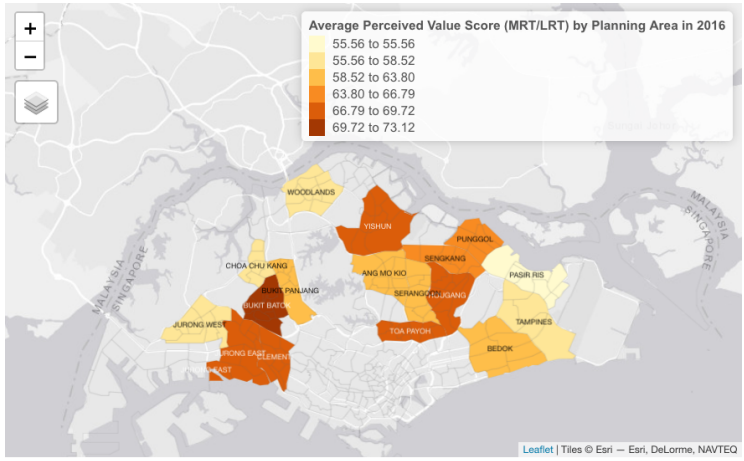
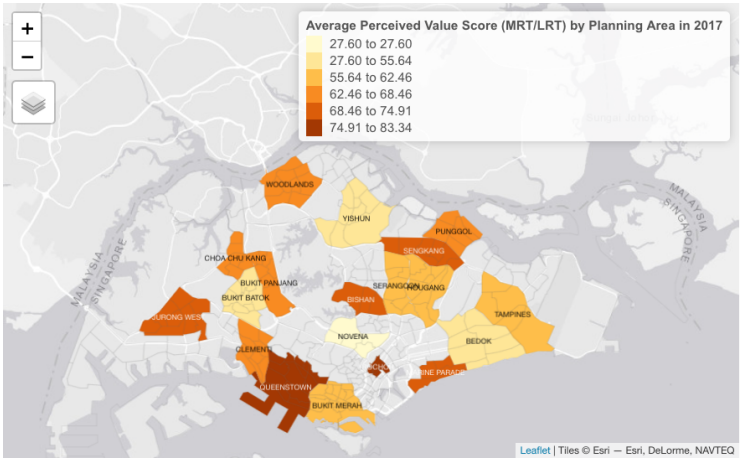
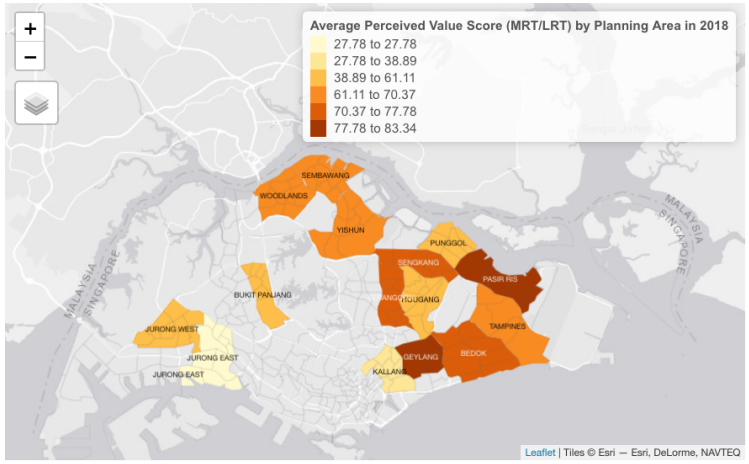
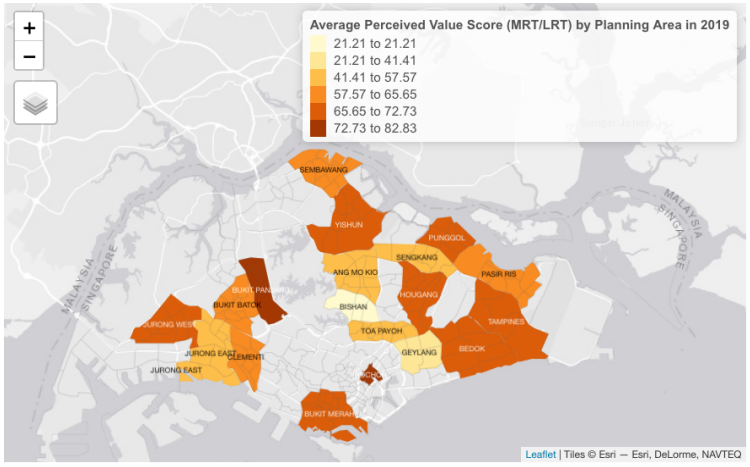
Findings 2: *Ang Mo Kio*, *Bishan*, *Geylang*, *Sembawang*, *Serangoon*, and *Toa Payoh*, are those few that have a decrease in their scoring over the years.



Findings 3: *Queenstown* and *Rochor* are the main few that have the highest average perceived value scores for the MRT/LRT service from 2016 to 2019.

Findings 4: Majority of the planning areas have an average of less than 66.09 average perceived value scores for this particular service over the years. This could probably infer that the perceptions of quality and pricing of this service are generally neutral to negative.

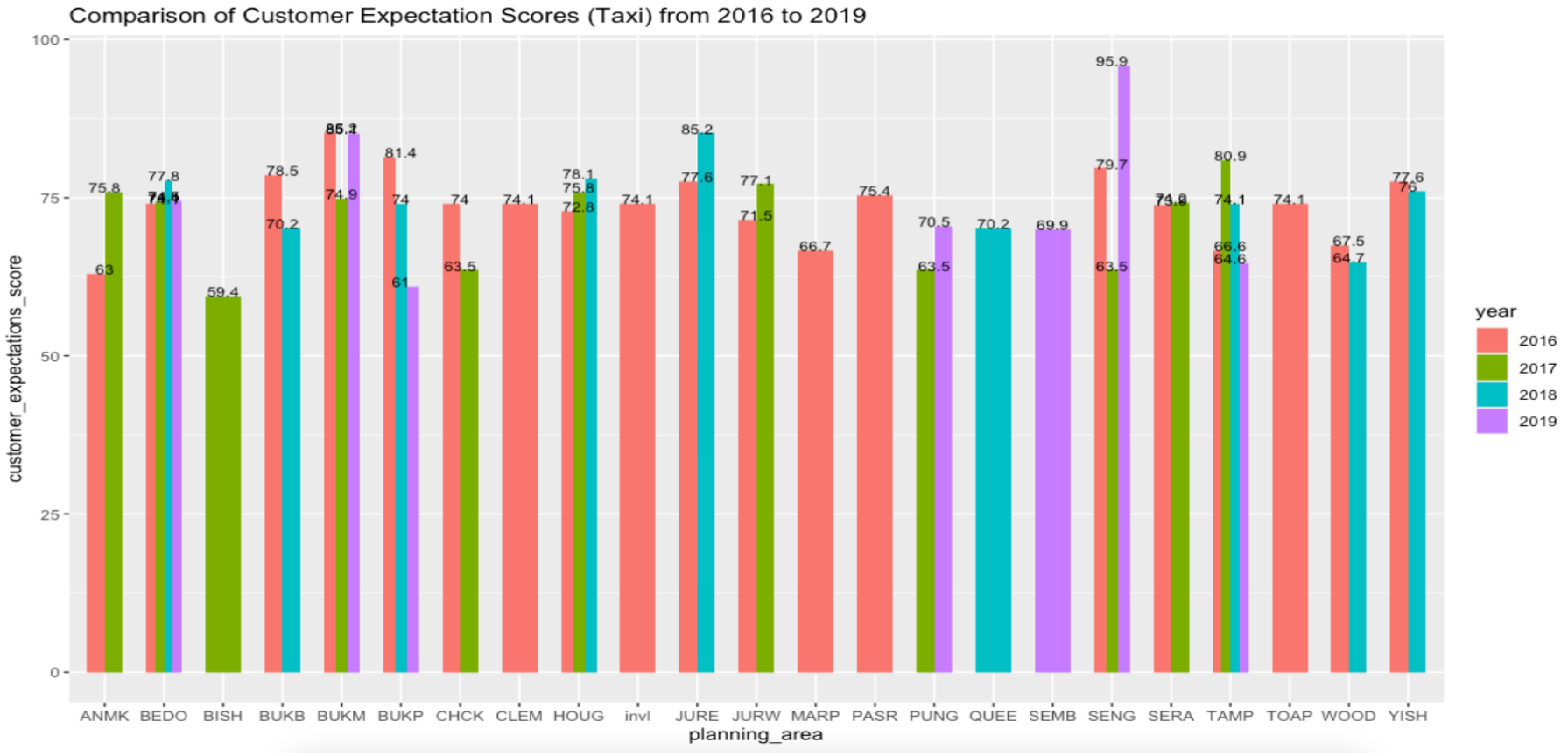
*Breakdown of **Perceived Value** Scores by Year and Planning Area*

2016	2017
	
2018	**2019**
	

11.9.4 Service Sector: ***Taxi***

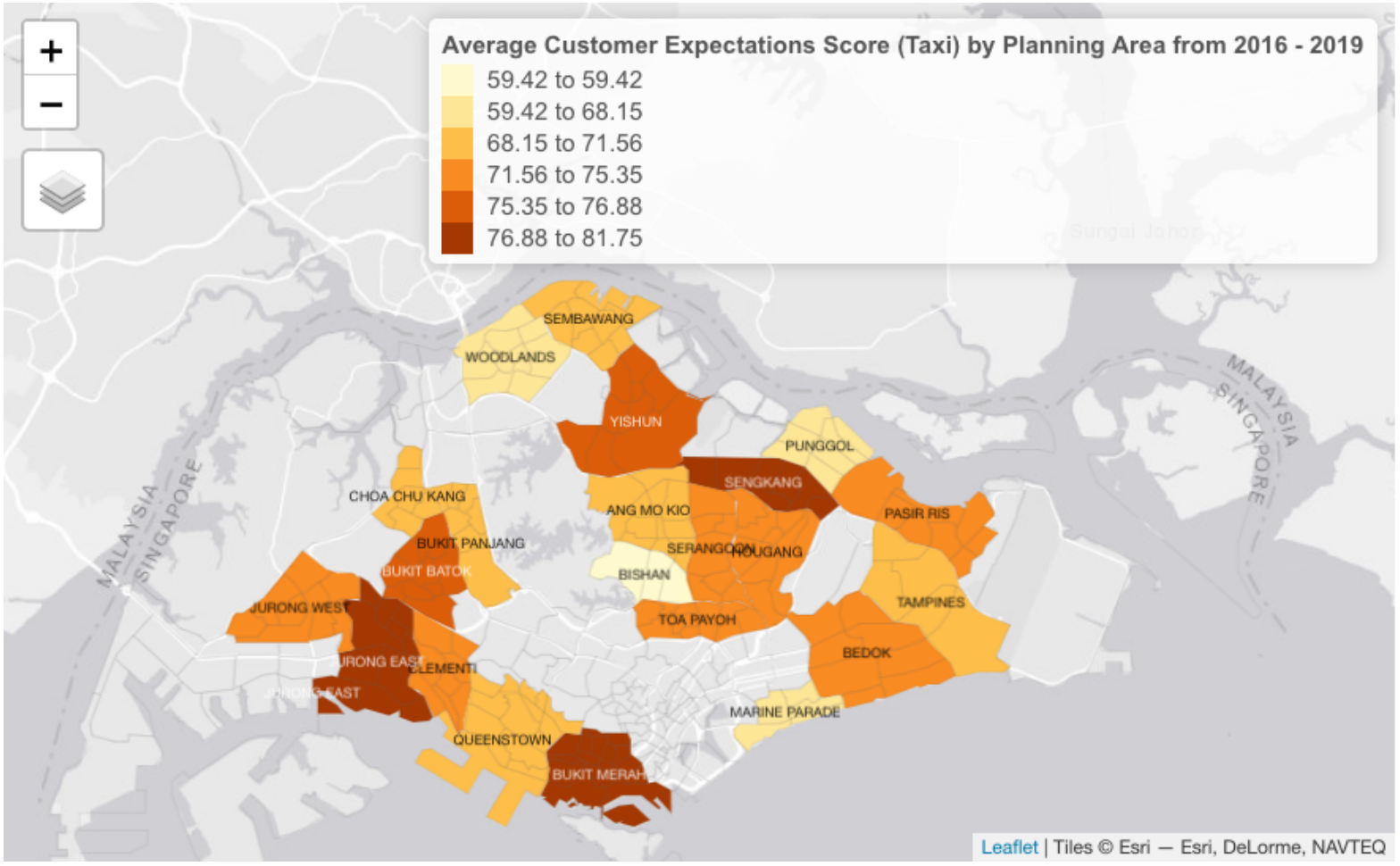
In the following comparison charts, the team will be looking into details of each of the respective scoring and how it changes from Year 2016 to 2019.

11.9.4.1 Customer Expectation



Findings 1: It can be seen that there has been a positive increase in the target audience's expectations over the years. Furthermore, *Sengkang* could be seen with an extremely high expectation score of 95.9 in 2019.

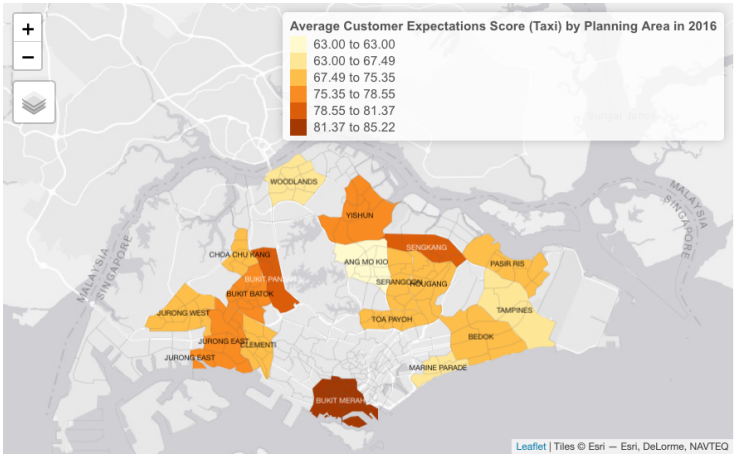
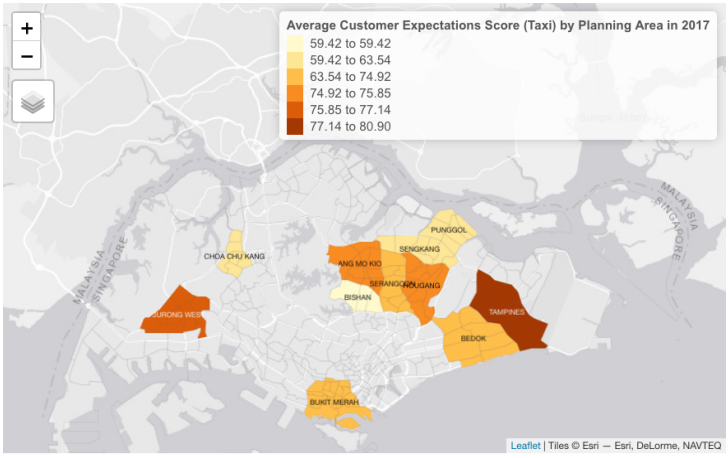
Findings 2: *Bukit Batok*, *Bukit Panjang*, *Choa Chu Kang* and *Tampines* have a decrease in their expectation scores in 2019.

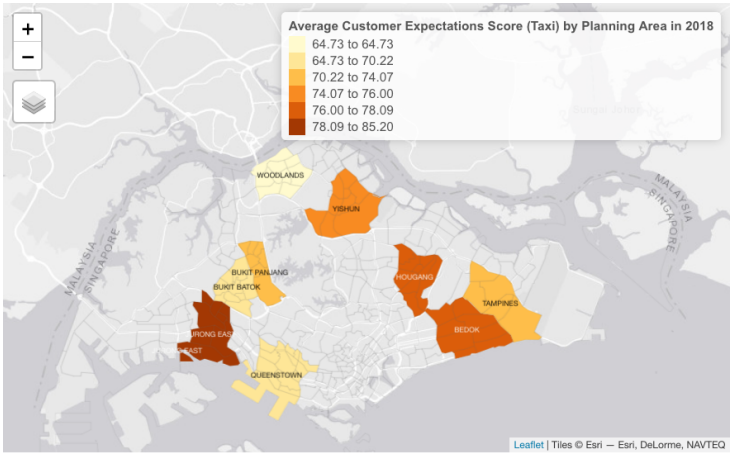


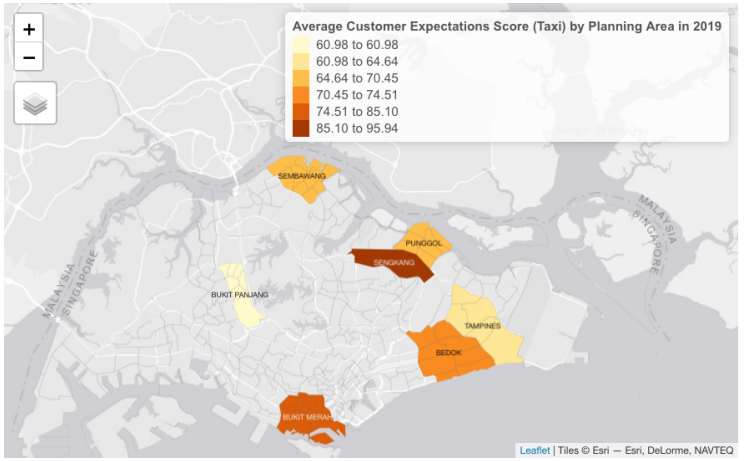
Findings 3: *Jurong East*, *Bukit Merah* and *Sengkang* are the top few with the highest average customer expectation scores for the taxi services in Singapore from 2016 to 2019. This could possibly infer that they have a higher anticipation of a particular set of behaviours or actions from the respective service sector.

Findings 4: As an overall, majority of the planning areas have a neutral expectation level towards the taxi services in Singapore.

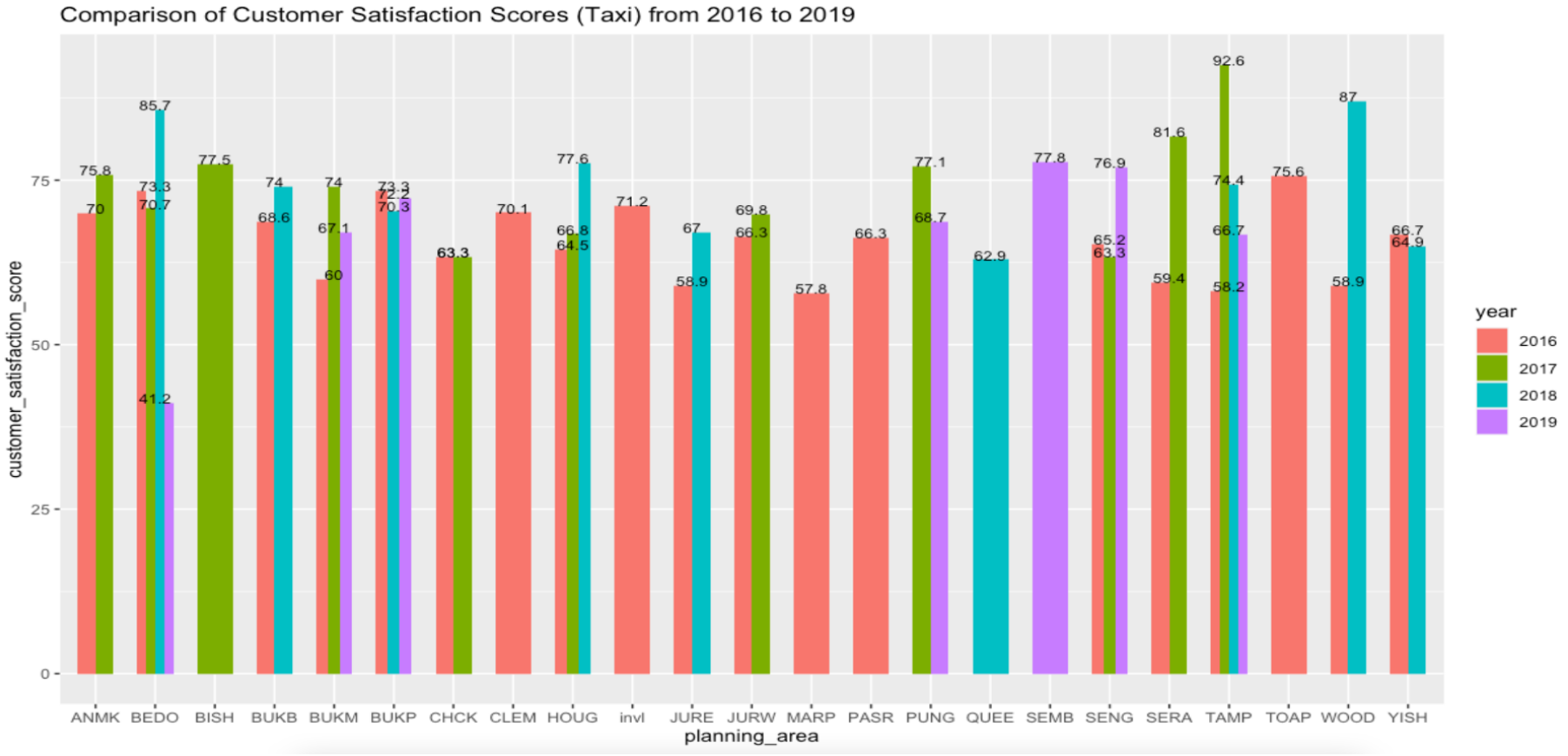
*Breakdown of Customer **Expectation** Scores by Year and Planning Area*

2016	2017
	
2018	2019


Average Customer Expectations Score (Taxi) by Planning Area in 2018
64.73 to 64.73
64.73 to 70.22
70.22 to 74.07
74.07 to 76.00
76.00 to 78.09
78.09 to 85.20
WOODLANDS
YISHUN
BUKIT PANJANG
BUKIT BATOK
HOUGANG
TAMPINES
BEDOK
QUEENSTOWN
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

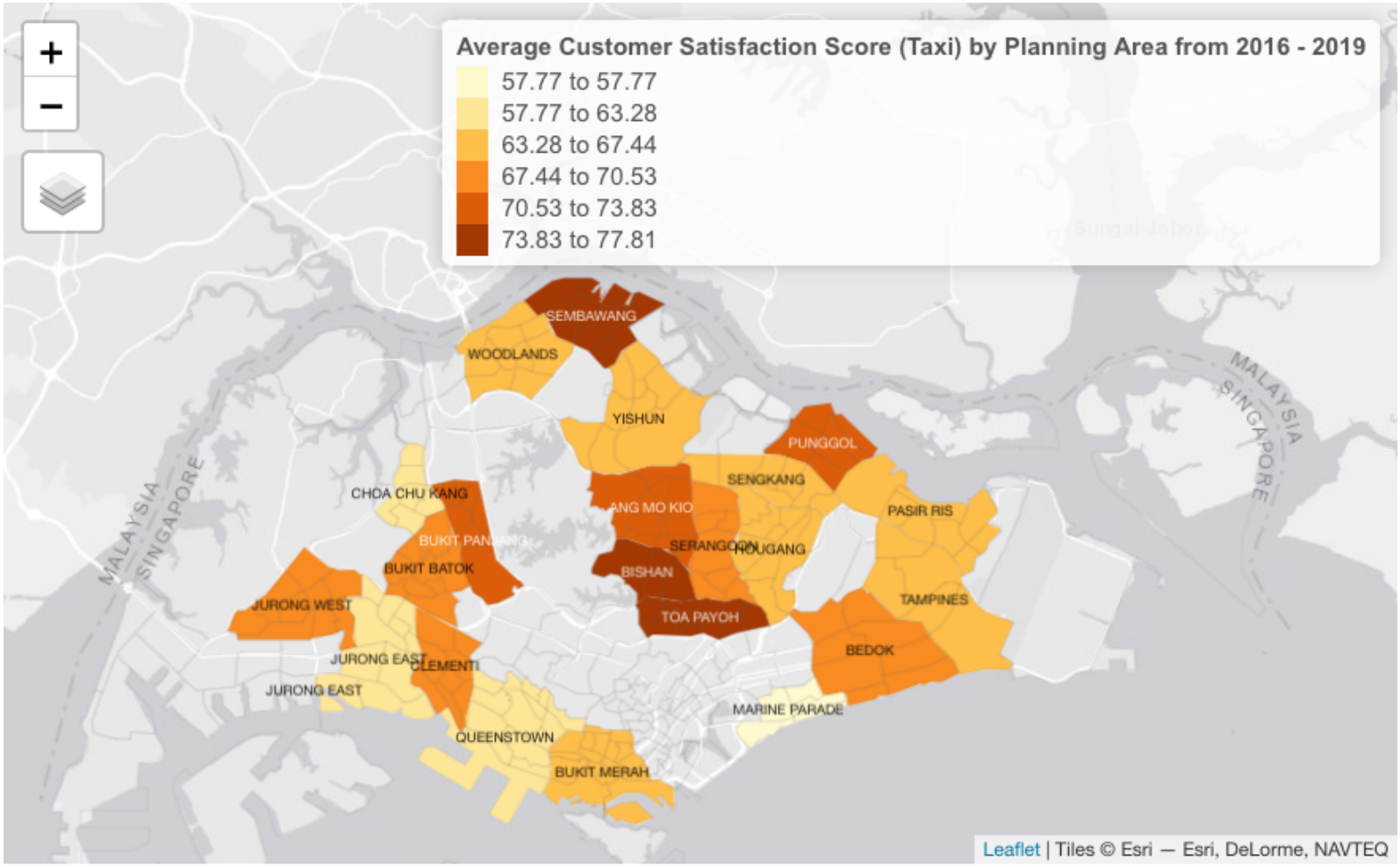

Average Customer Expectations Score (Taxi) by Planning Area in 2019
60.98 to 60.98
60.98 to 64.64
64.64 to 70.45
70.45 to 74.51
74.51 to 85.10
85.10 to 95.94
SEMBAWANG
PUNGGOL
SENGKANG
BUKIT PANJANG
TAMPINES
BEDOK
BUKIT MERAH
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

11.9.4.2 Customer Satisfaction



Findings 1: An outlier, *Bedok*, could be spotted with the lowest satisfaction score of 41.2 in 2019.

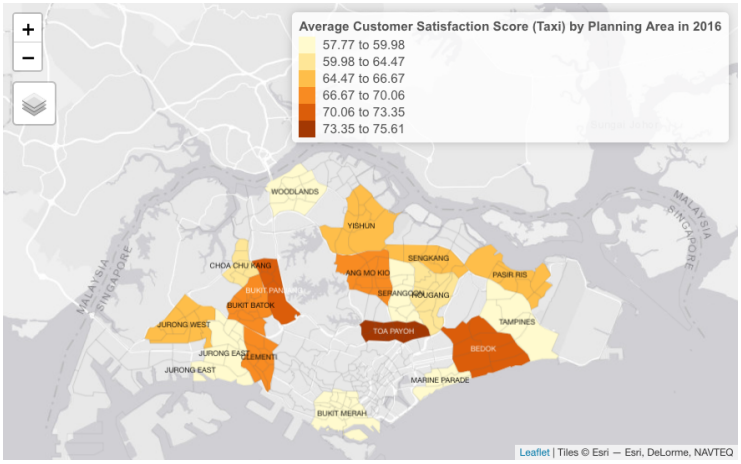
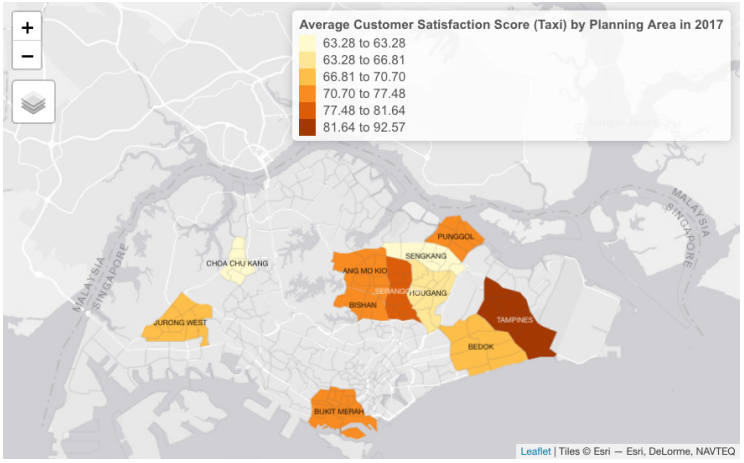
Findings 2: There has been a neutral to positive increase in customer satisfaction scores across the varying planning areas yearly. However, areas like *Bedok*, *Punggol* and *Tampines* have been seen decreasing significantly over the years.

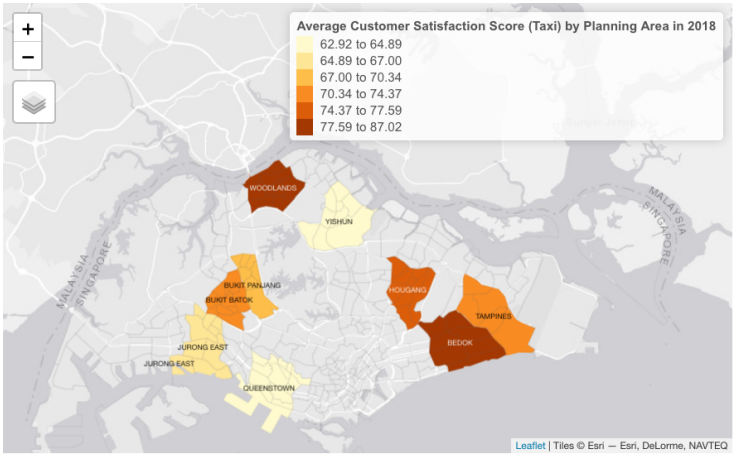


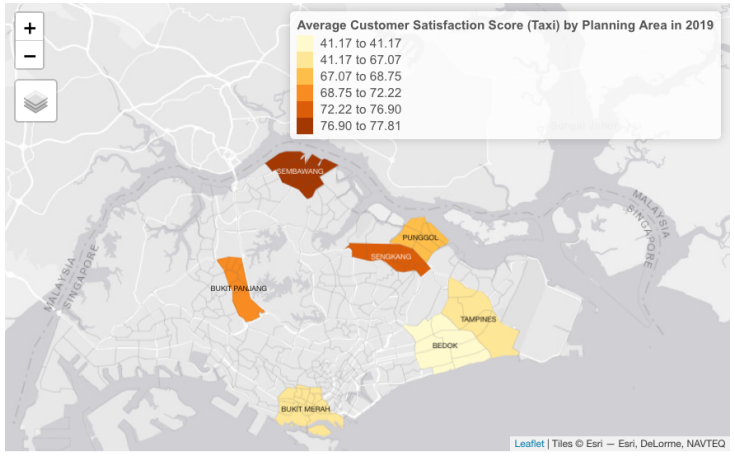
Findings 3: *Bishan*, *Toa Payoh* and *Sembawang* are the top few with the highest average customer satisfaction scores from 2016 to 2019.

Findings 4: *Marine Parade, Jurong East*, *Queenstown* and *Choa Chu Kang* are the few with the lower average customer satisfaction scores and this could probably imply that they have neutral to negative satisfaction levels towards the Taxi services in Singapore.

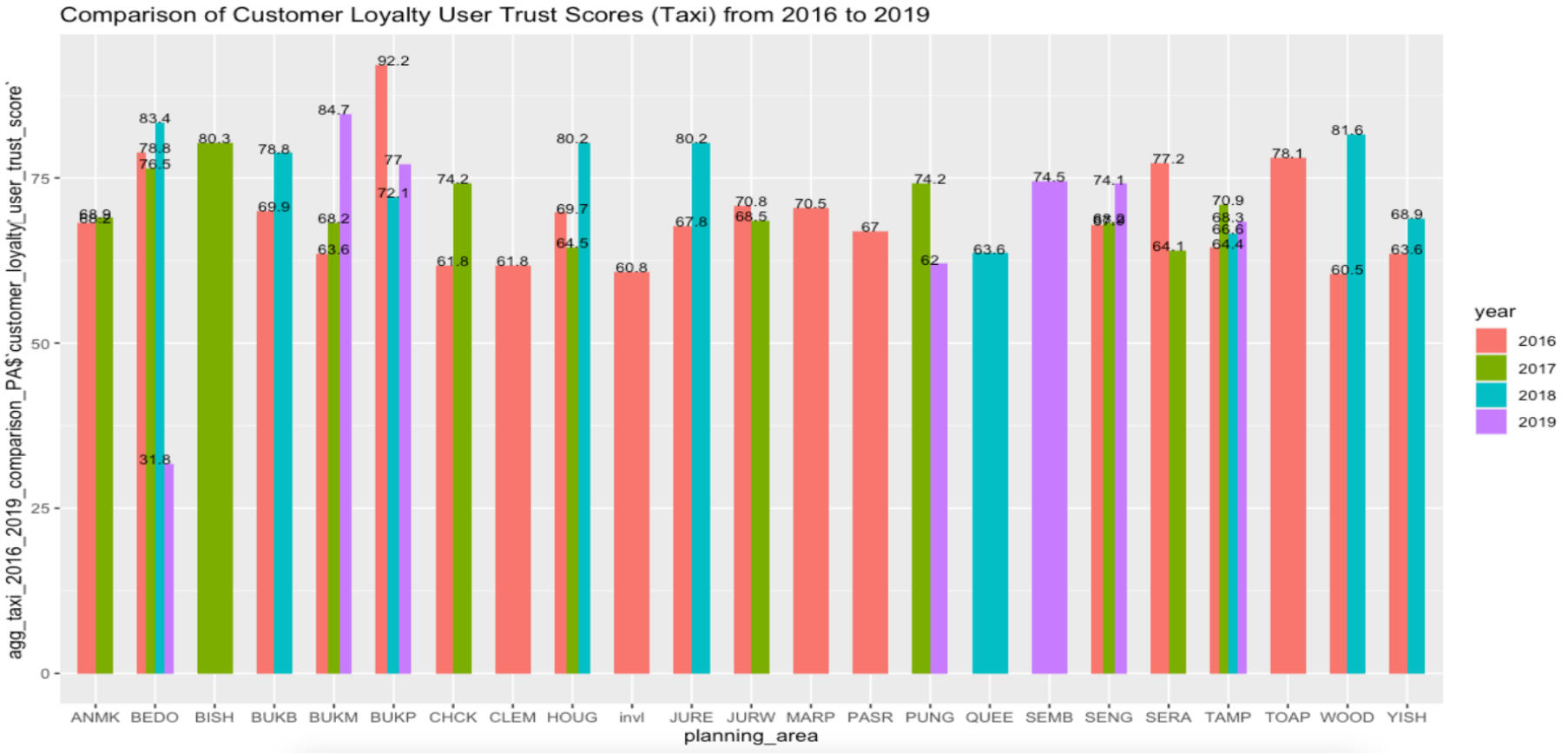
*Breakdown of Customer **Satisfaction** Scores by Year and Planning Area*

2016	2017
	
2018	2019


Average Customer Satisfaction Score (Taxi) by Planning Area in 2018
62.92 to 64.89
64.89 to 67.00
67.00 to 70.34
70.34 to 74.37
74.37 to 77.59
77.59 to 87.02
WOODLANDS
YISHUN
BUKIT PANJANG
BUKIT BATOK
HOUGANG
TAMPINES
BEDOK
JURONG EAST
JURONG EAST
QUEENSTOWN
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

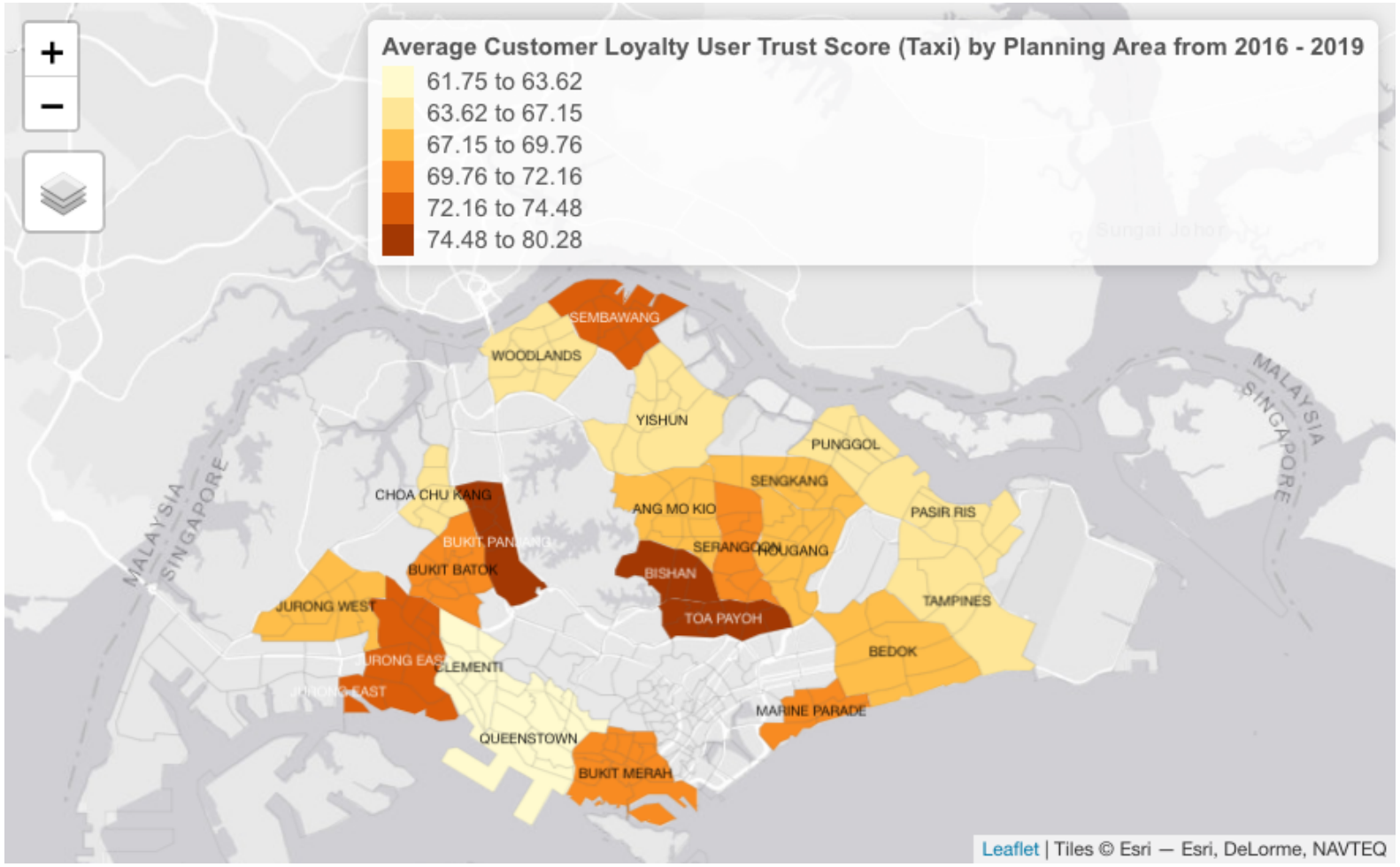

Average Customer Satisfaction Score (Taxi) by Planning Area in 2019
41.17 to 41.17
41.17 to 67.07
67.07 to 68.75
68.75 to 72.22
72.22 to 76.90
76.90 to 77.81
SEMBAWANG
PUNGGOL
SENGKANG
BUKIT PANJANG
TAMPINES
BEDOK
BUKIT MERAH
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

11.9.4.3 Customer Loyalty User Trust



Findings 1: There has been a neutral to positive increase in the customer loyalty user trust scores in numerous areas over the respective years. However, *Bedok* can be seen with an all-time lowest score of 31.8 in 2019.

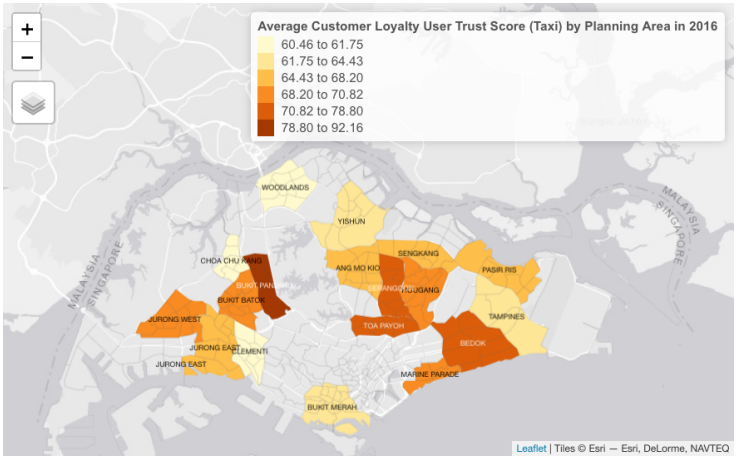
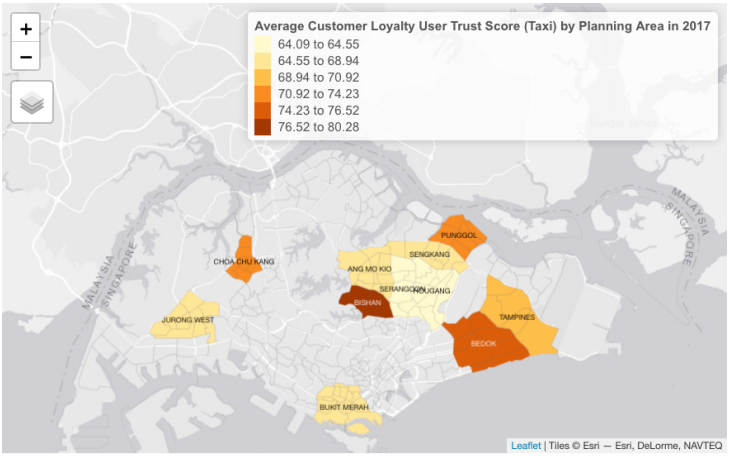
Findings 2: Overall, most would have neutral to positive opinions about the Taxi services in Singapore.

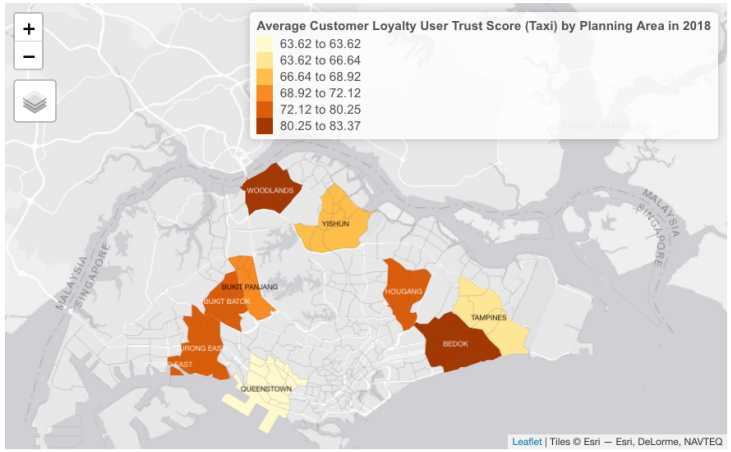


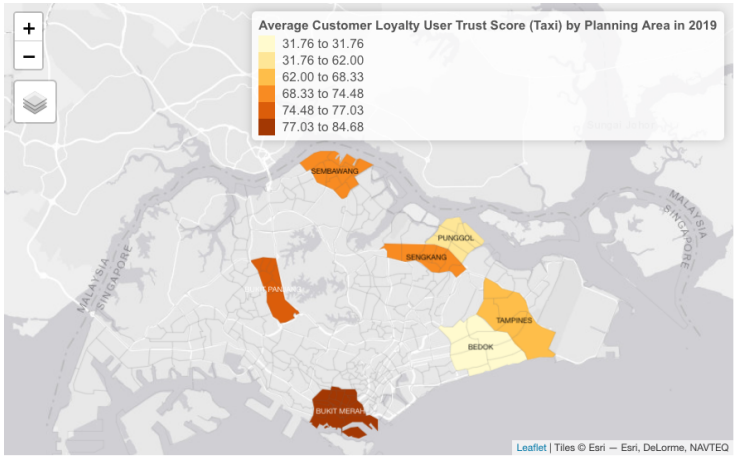
Findings 3: *Bukit Panjang*, *Bishan* and *Toa Payoh* are among those that have more positive opinions of the Taxi services in Singapore.

Findings 4: Overall, the majority of the planning areas have neutral opinions about this particular mobility sector.

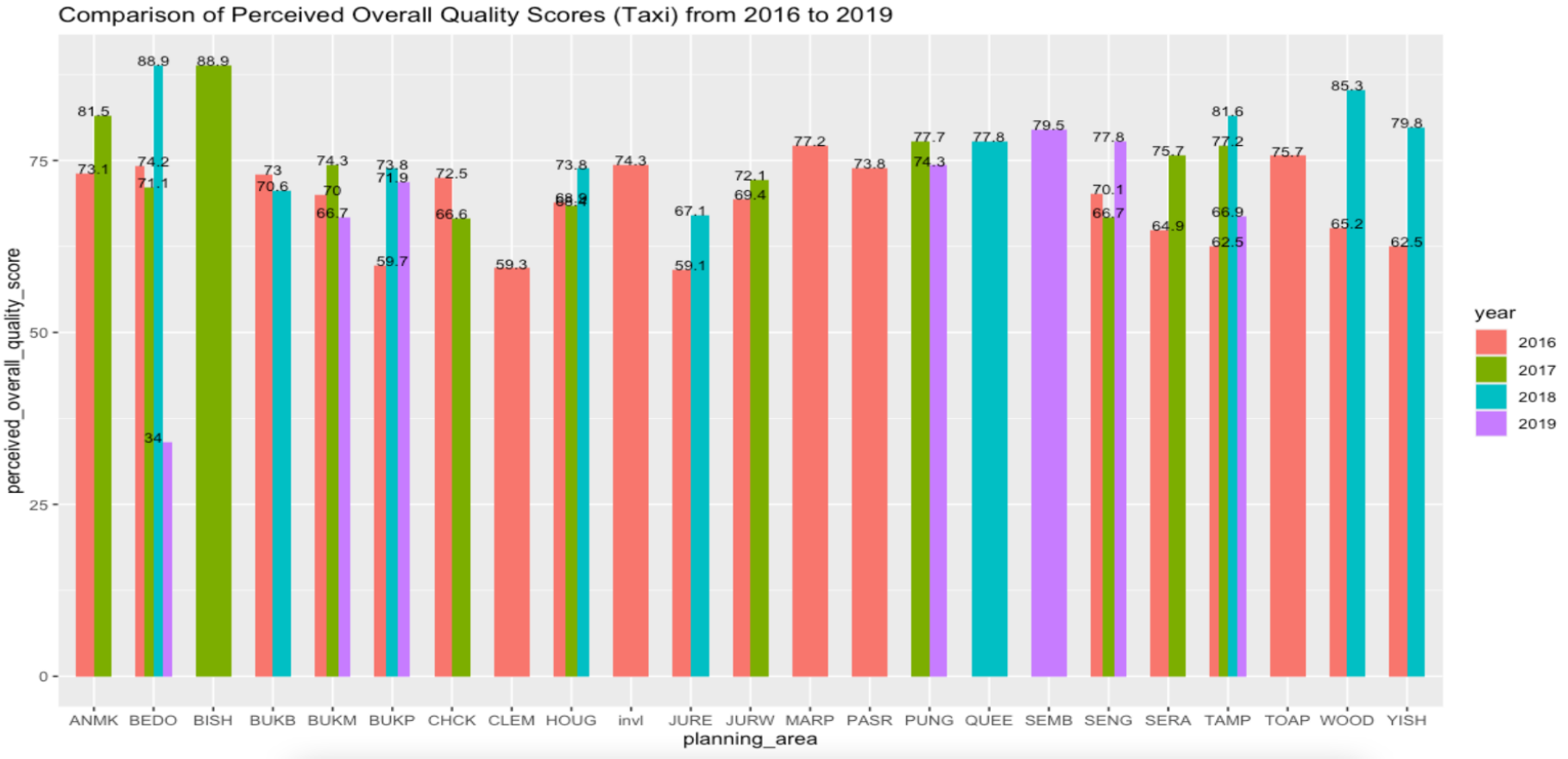
*Breakdown of Customer **Loyalty User Trust** Scores by Year and Planning Area*

2016	2017
	
2018	2019


Average Customer Loyalty User Trust Score (Taxi) by Planning Area in 2018
63.62 to 63.62
63.62 to 66.64
66.64 to 68.92
68.92 to 72.12
72.12 to 80.25
80.25 to 83.37
WOODLANDS
YISHUN
BUKIT PANJANG
BUKIT BATOK
HOUGANG
TAMPINES
BEDOK
QUEENSTOWN
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

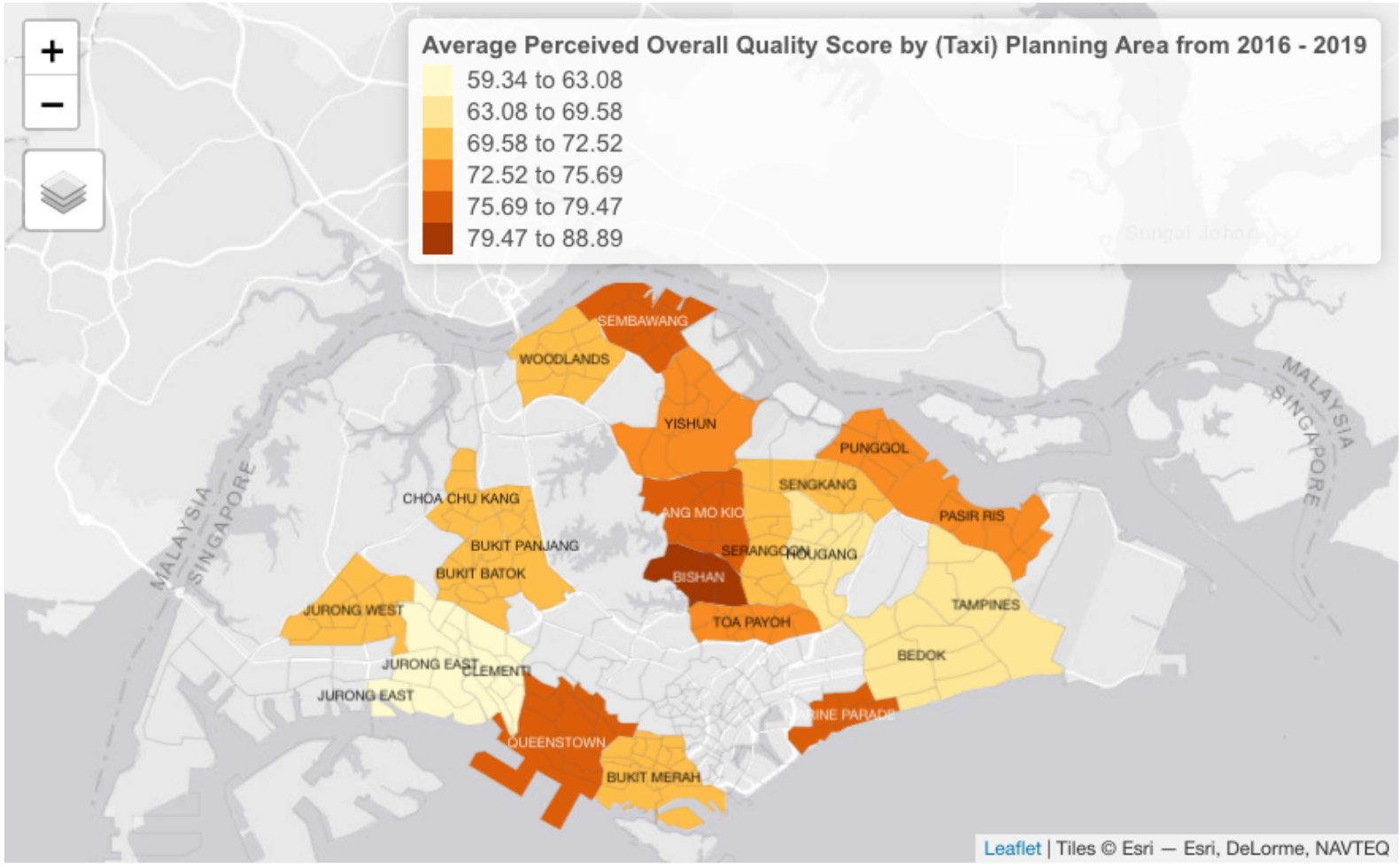

Average Customer Loyalty User Trust Score (Taxi) by Planning Area in 2019
31.76 to 31.76
31.76 to 62.00
62.00 to 68.33
68.33 to 74.48
74.48 to 77.03
77.03 to 84.68
SEMBAWANG
PUNGGOL
SENGKANG
TAMPINES
BEDOK
BUKIT MERAH
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

11.9.4.4 Perceived Overall Quality



Findings 1: Generally, the target audience felt positive about the overall quality they received from the Taxi services yearly.

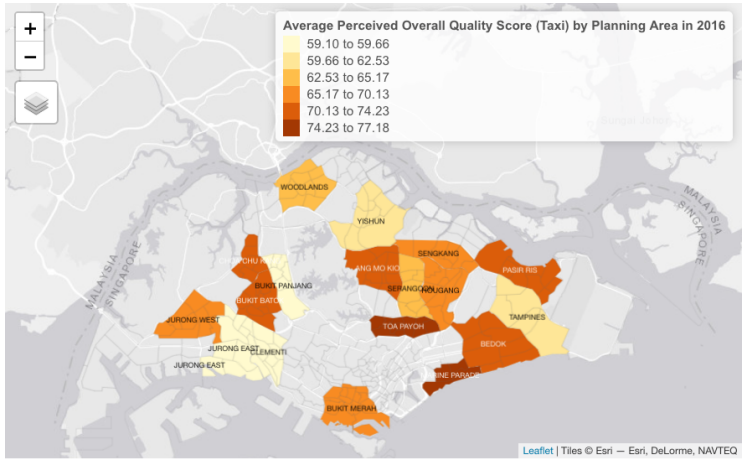
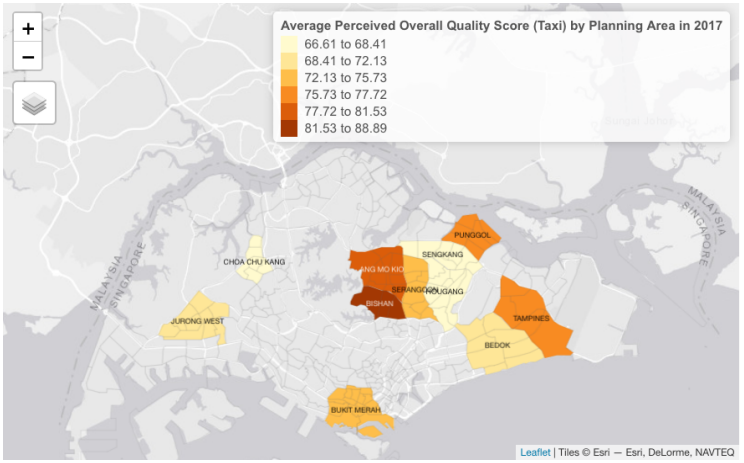
Findings 2: An outlier, *Bedok*, could be seen with an extreme drop of perceived overall quality score (34) in 2019.

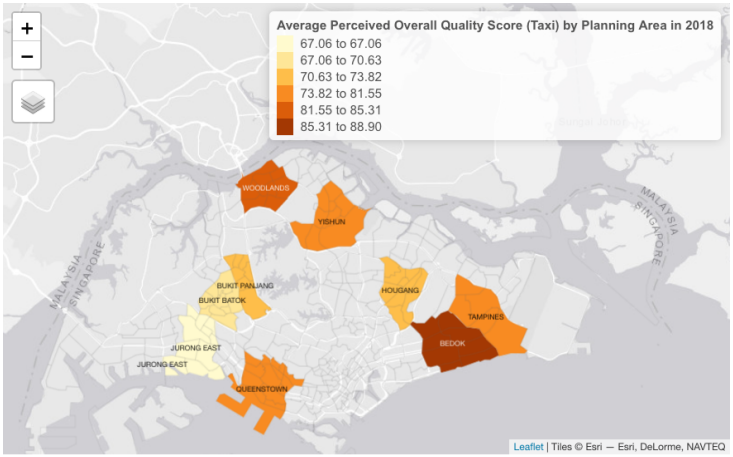


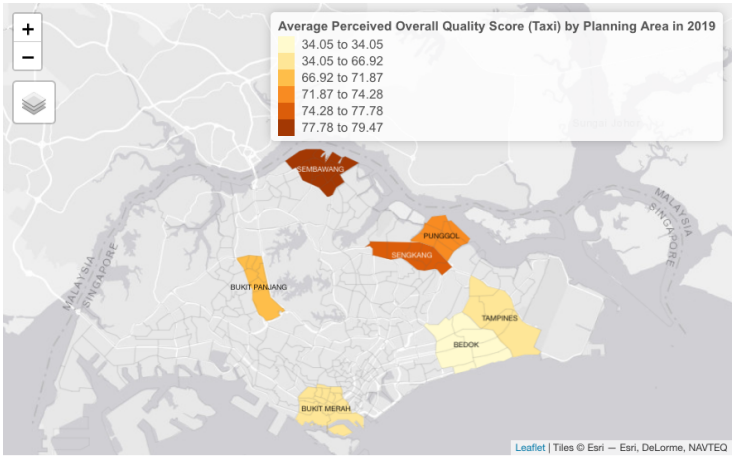
Findings 3: *Bishan* is the one with the highest average perceived overall quality score from 2016 to 2019.

Findings 4: Planning areas in the West and East of Singapore could be seen with a general neutral perspective in the quality they received, as well as if the service provider meets their personal requirements.

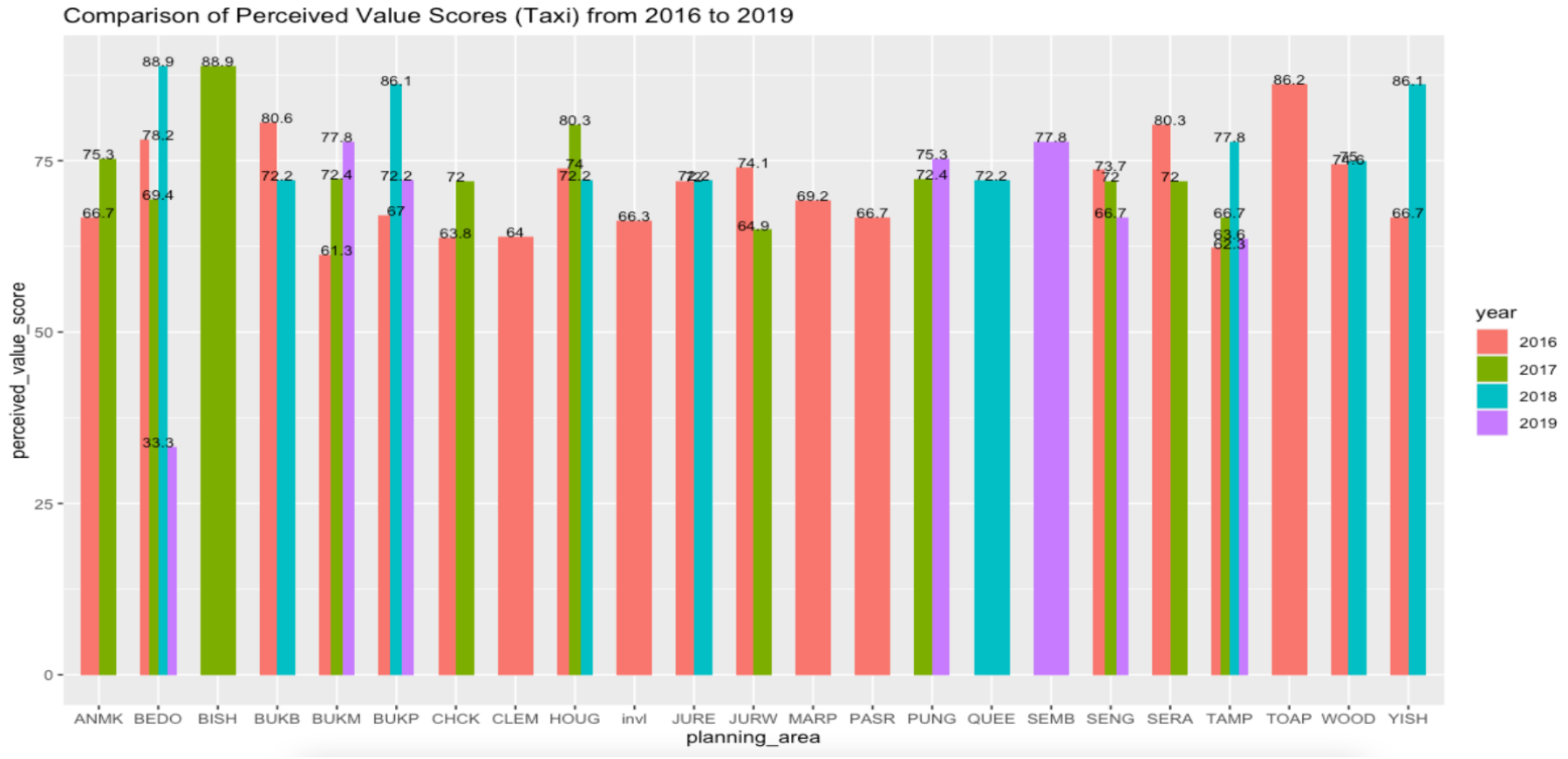
*Breakdown of **Perceived Overall Quality** Scores by Year and Planning Area*

2016	2017
	
2018	2019


Average Perceived Overall Quality Score (Taxi) by Planning Area in 2018
67.06 to 67.06
67.06 to 70.63
70.63 to 73.82
73.82 to 81.55
81.55 to 85.31
85.31 to 88.90
WOODLANDS
YISHUN
BUKIT PANJANG
BUKIT BATOK
HOUGANG
TAMPINES
BEDOK
JURONG EAST
JURONG EAST
QUEENSTOWN
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

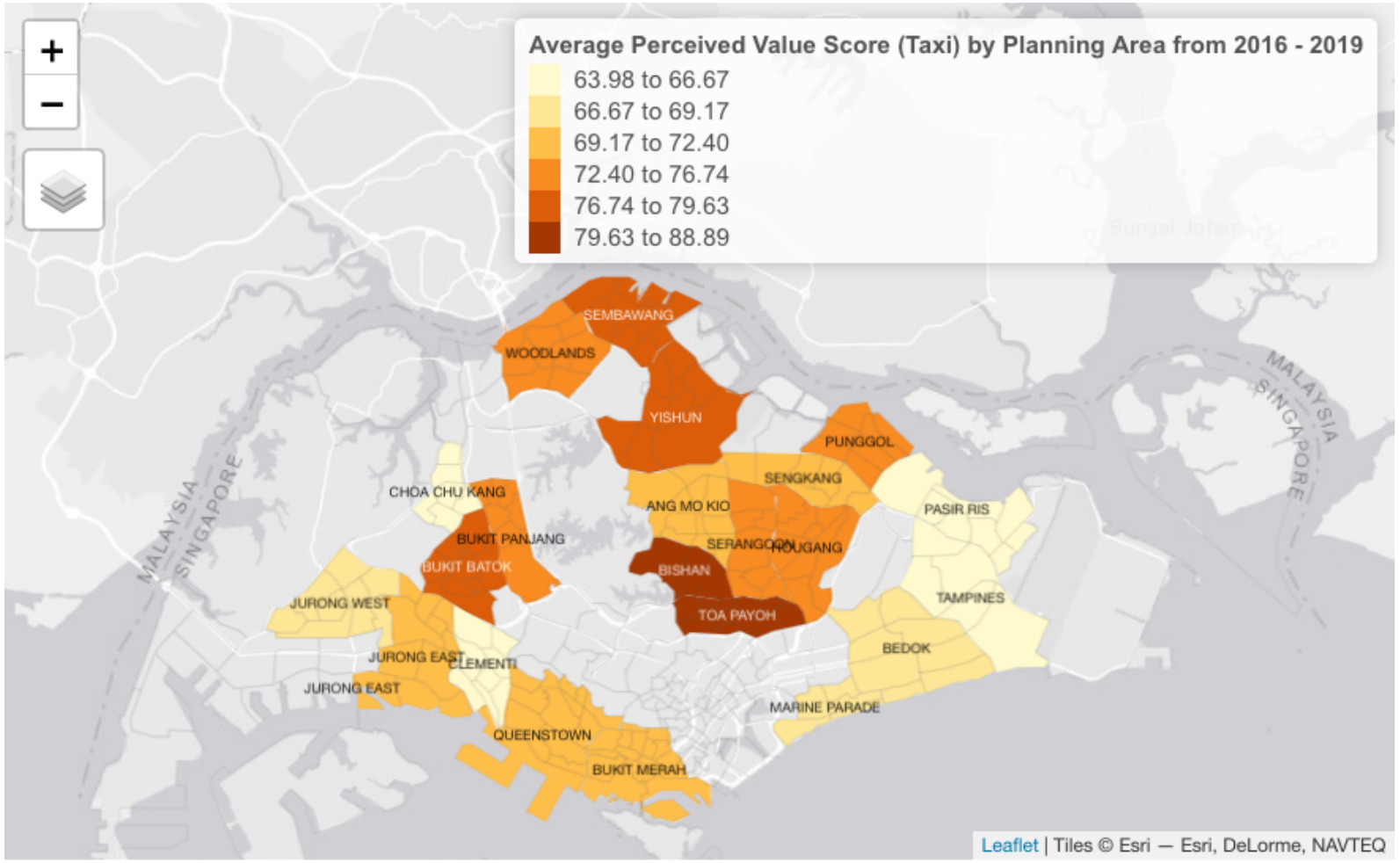

Average Perceived Overall Quality Score (Taxi) by Planning Area in 2019
34.05 to 34.05
34.05 to 66.92
66.92 to 71.87
71.87 to 74.28
74.28 to 77.78
77.78 to 79.47
SEMBAWANG
PUNGGOL
SENGKANG
BUKIT PANJANG
TAMPINES
BEDOK
BUKIT MERAH
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

11.9.4.5 Perceived Value



Findings 1: There has been a neutral to positive increase in the perceived value scores across the respective planning areas yearly.

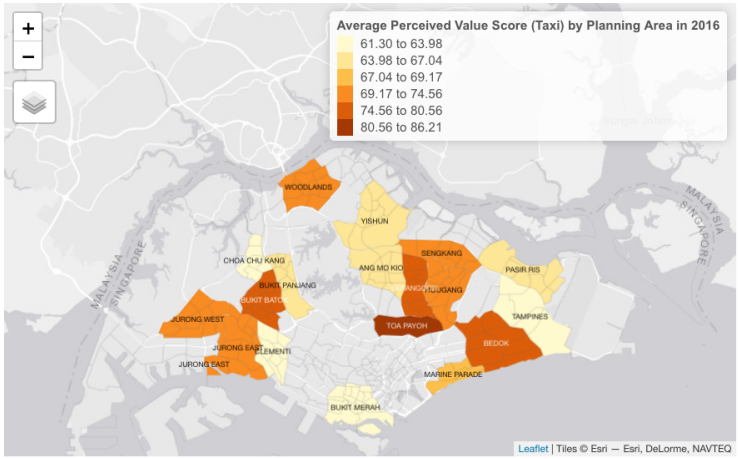
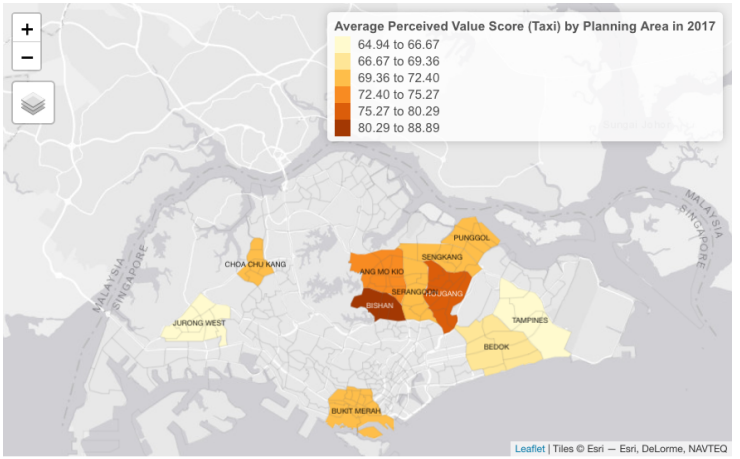
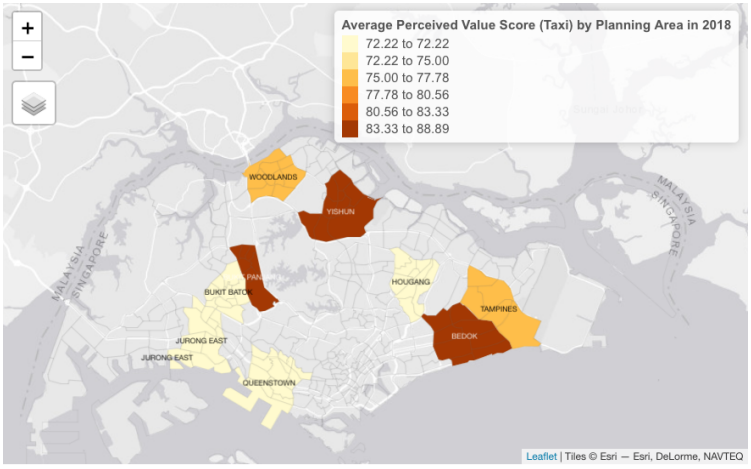
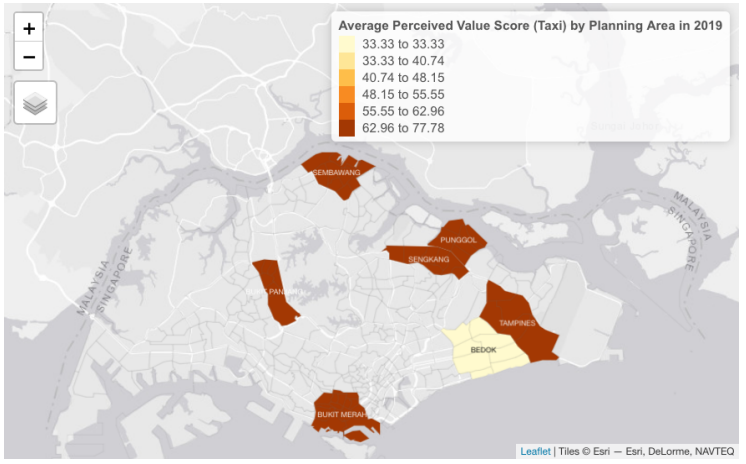
Findings 2: *Bedok* could be seen with the sudden decrease to 33.3 of its score in 2019.



Findings 3: *Bishan* and *Toa Payoh* are the main few that have the highest average perceived value score for Taxi services in Singapore.

Findings 4: Planning areas in the bottom half of the map could be seen with an average ranging from 63.98 to 72.40 of the average perceived value scores for this particular service. This could possibly imply that the perceptions of quality and pricing of this service are generally neutral.

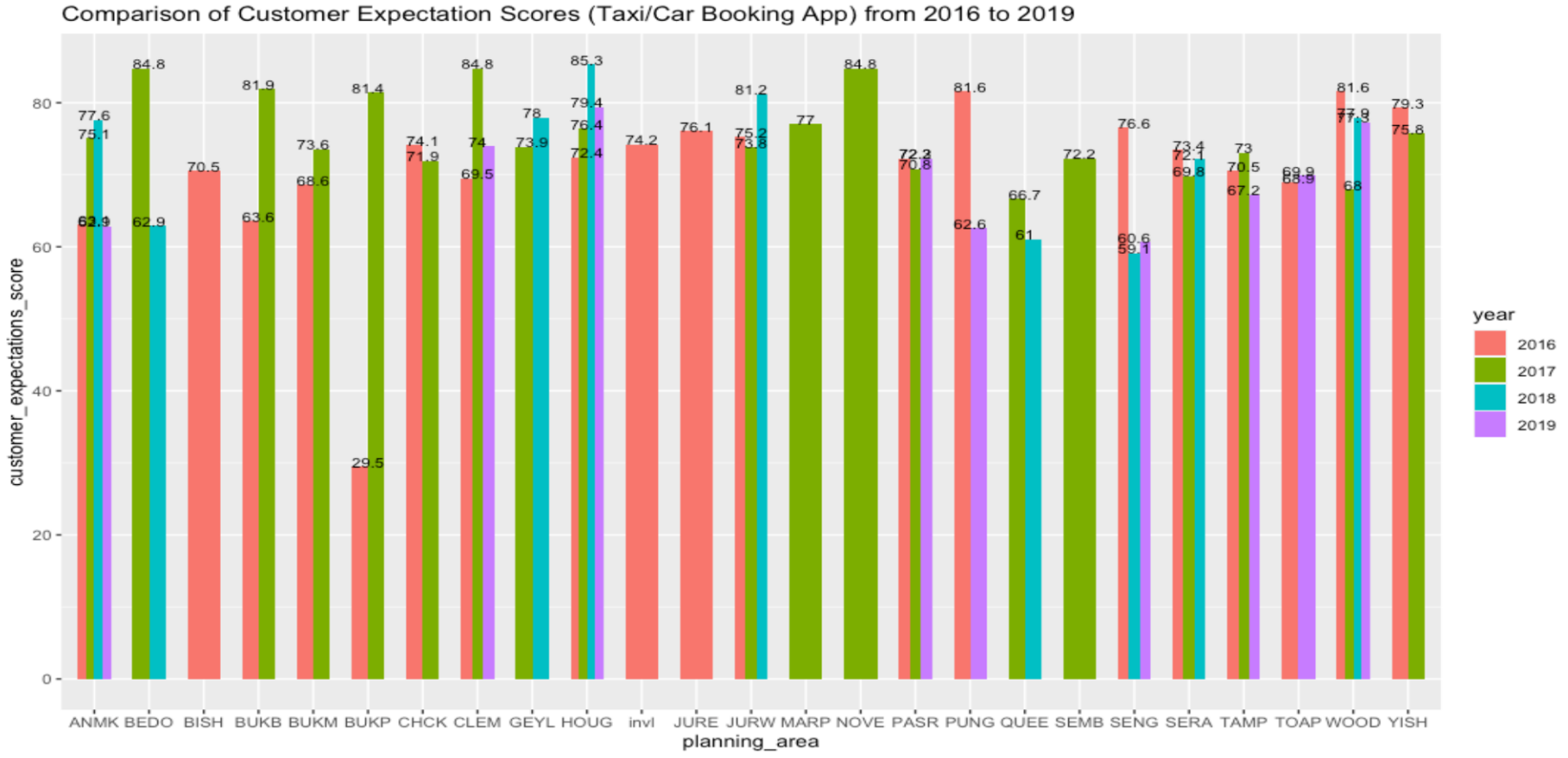
*Breakdown of **Perceived Value** Scores by Year and Planning Area*

<table>
<tr><th>2016</th><th>2017</th></tr>
<tr><td> 
</td><td> 
</td></tr>
<tr><th>2018</th><th>2019</th></tr>
<tr><td> 
</td><td> 
</td></tr>
</table>

11.9.3 Service Sector: ***Booking Application***

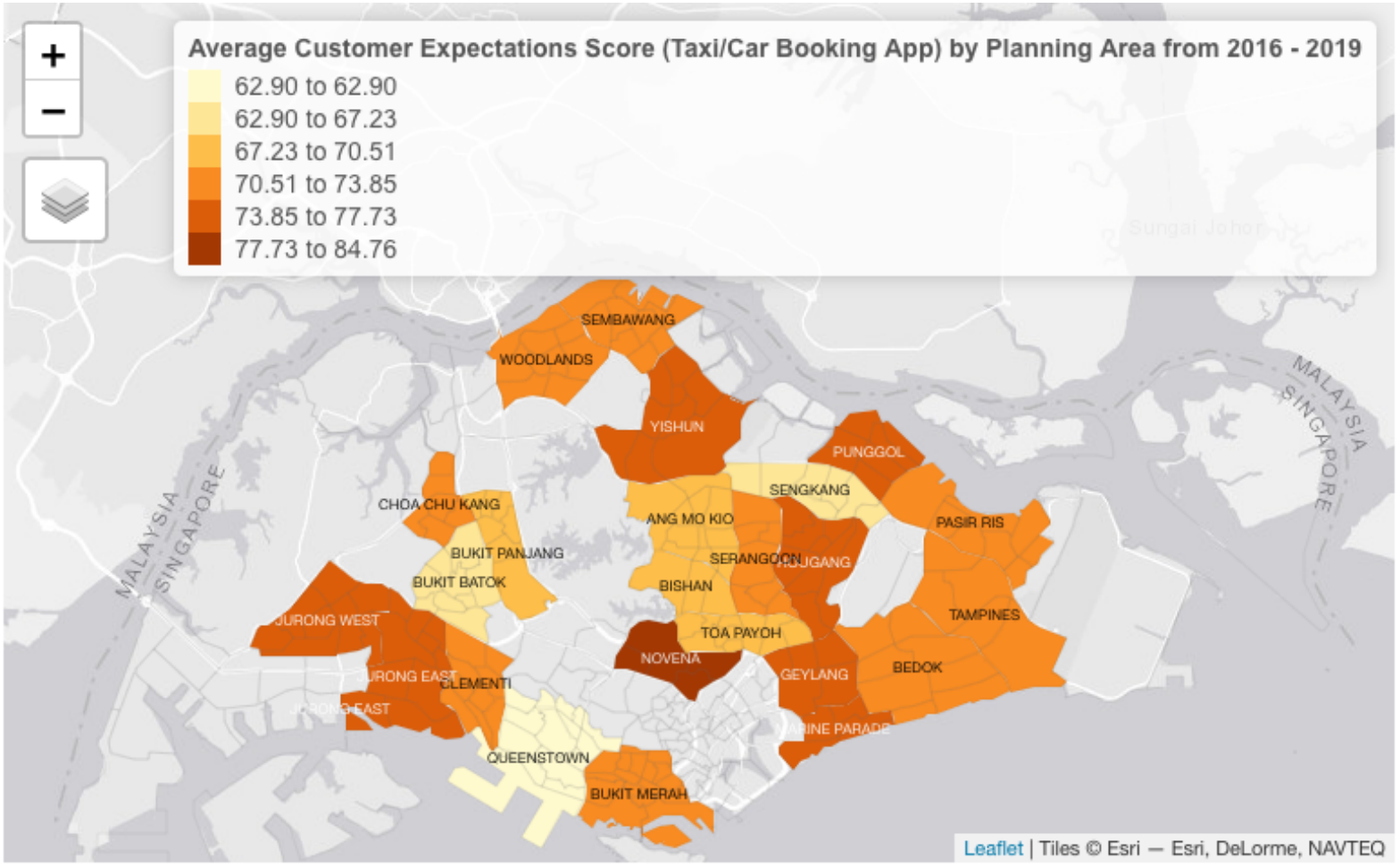
In the following comparison charts, the team will be looking into details of each of the respective scoring and how it changes from Year 2016 to 2019.

11.9.3.1 Customer Expectation



Findings 1: Numerous planning areas could be seen with drastic increase of the target audience's expectations over the years.

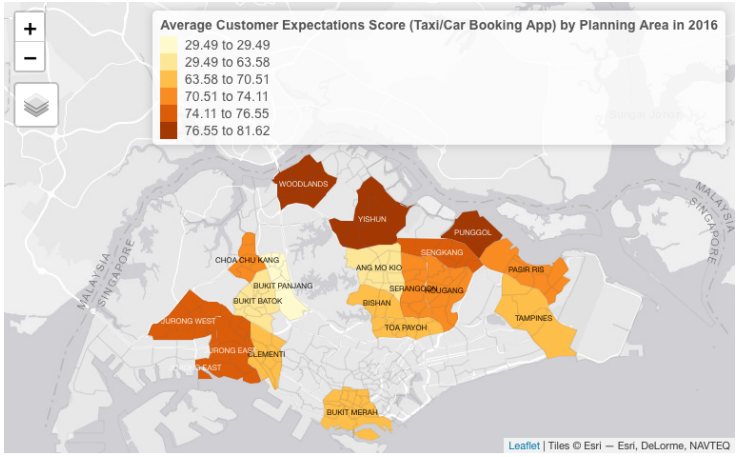
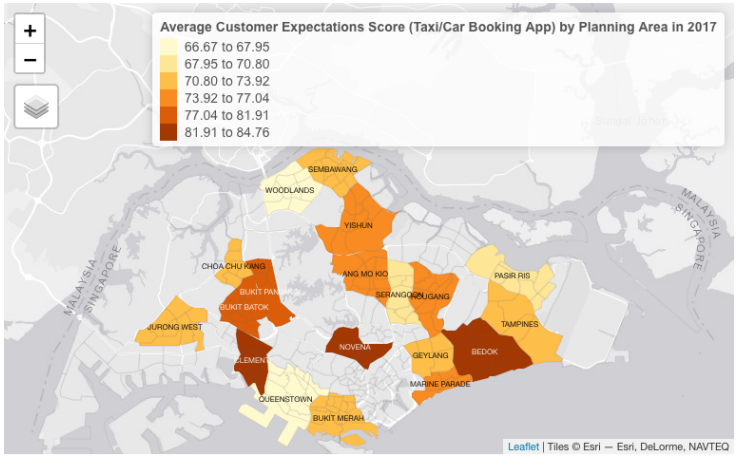
Findings 2: Particularly, *Ang Mo Kio*, *Bedok*, *Punggol* and *Sengkang* have progressively decreased their expectations over the years.

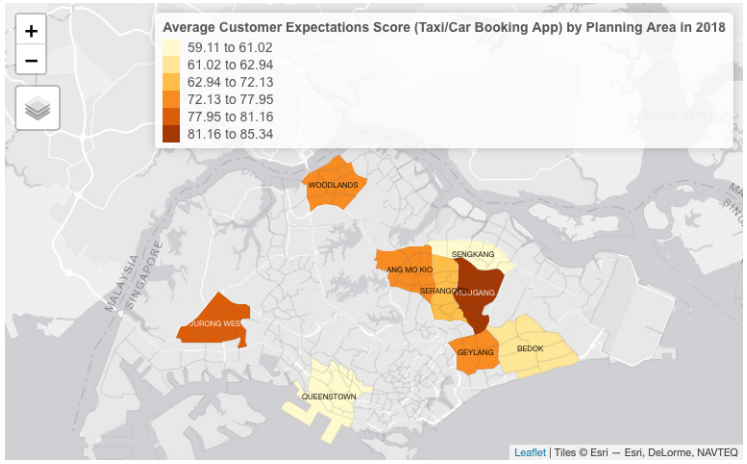


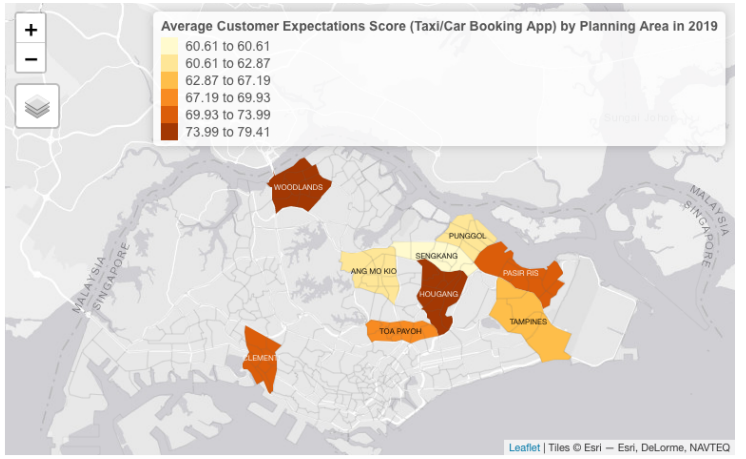
Findings 3: *Novena* could be seen with the highest average customer expectation score from 2016 to 2019. This could possibly infer that they have a higher anticipation of a particular set of behaviours or actions from the respective service sector.

Findings 4: Overall, the majority of the planning areas have a generally neutral to high expectations level towards the Taxi/Car Booking Application service in Singapore.

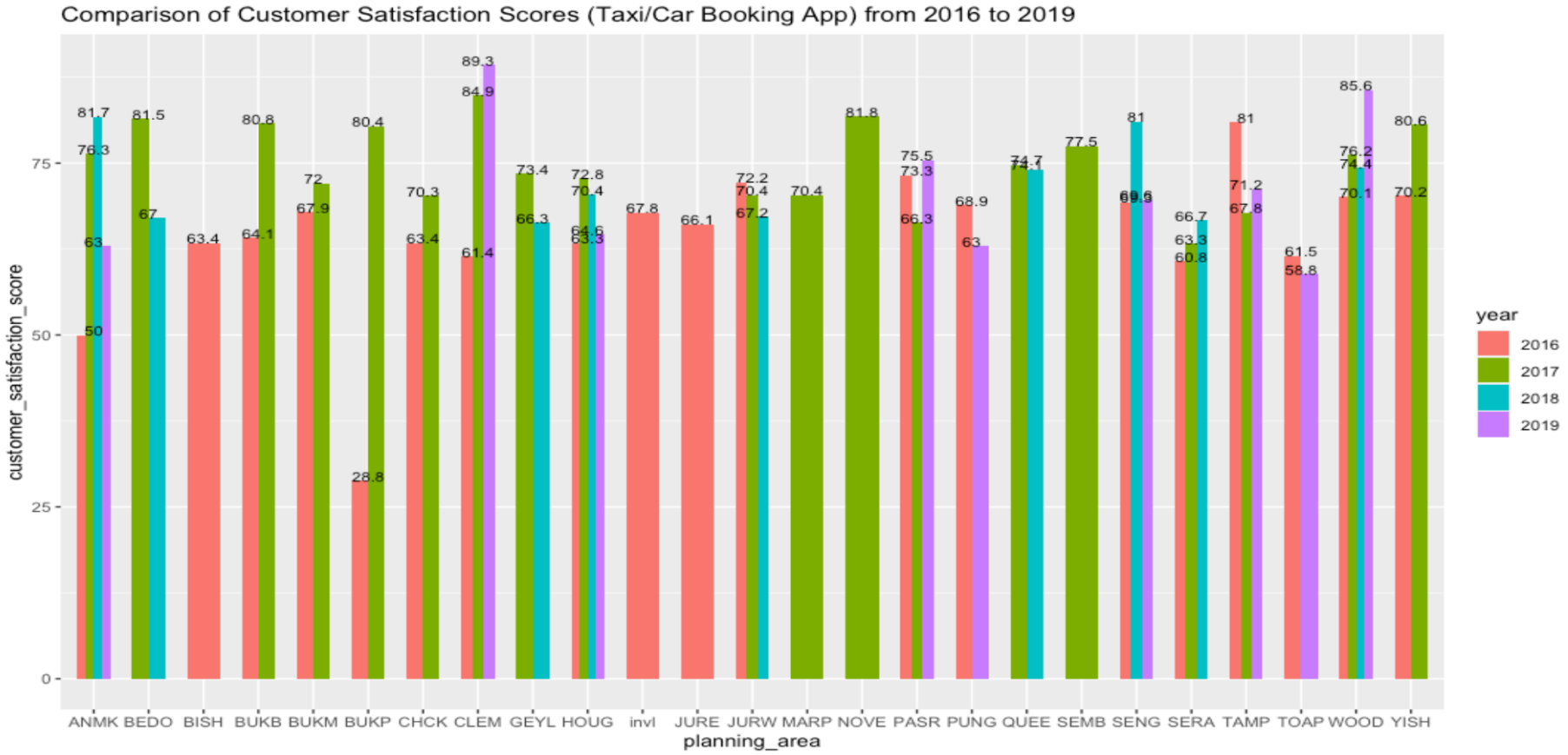
*Breakdown of Customer **Expectation** Scores by Year and Planning Area*

2016	2017
	
2018	2019


Average Customer Expectations Score (Taxi/Car Booking App) by Planning Area in 2018
59.11 to 61.02
61.02 to 62.94
62.94 to 72.13
72.13 to 77.95
77.95 to 81.16
81.16 to 85.34
WOODLANDS
SENGKANG
ANG MO KIO
SERANGOON
PUNGGOL
JURONG WEST
GEYLANG
BEDOK
QUEENSTOWN
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

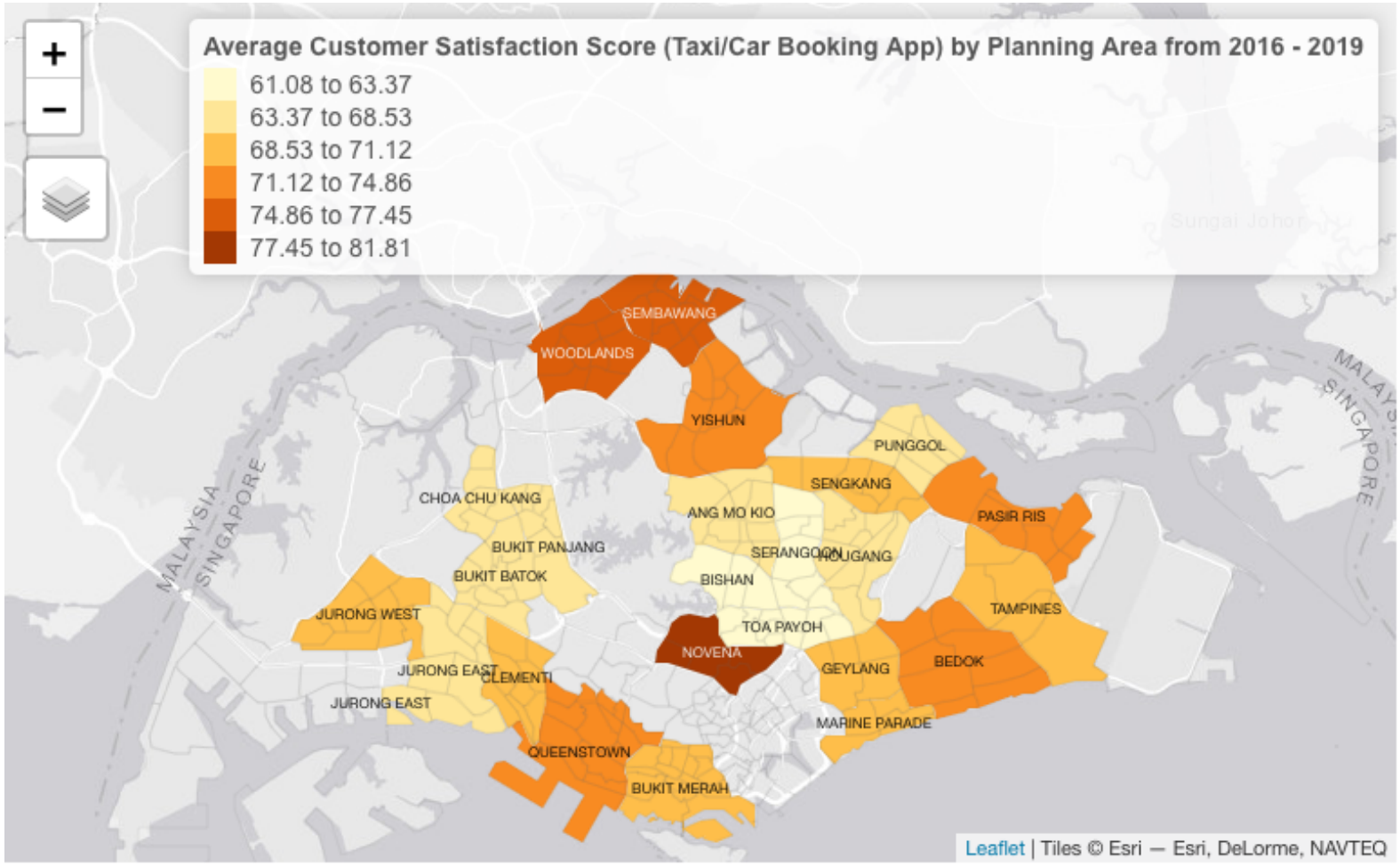

Average Customer Expectations Score (Taxi/Car Booking App) by Planning Area in 2019
60.61 to 60.61
60.61 to 62.87
62.87 to 67.19
67.19 to 69.93
69.93 to 73.99
73.99 to 79.41
WOODLANDS
PUNGGOL
SENGKANG
ANG MO KIO
PASIR RIS
HOUGANG
TOA PAYOH
TAMPINES
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

11.9.3.2 Customer Satisfaction



Findings 1: On a yearly basis, the customer satisfaction scores across the respective planning areas have been pretty neutral.

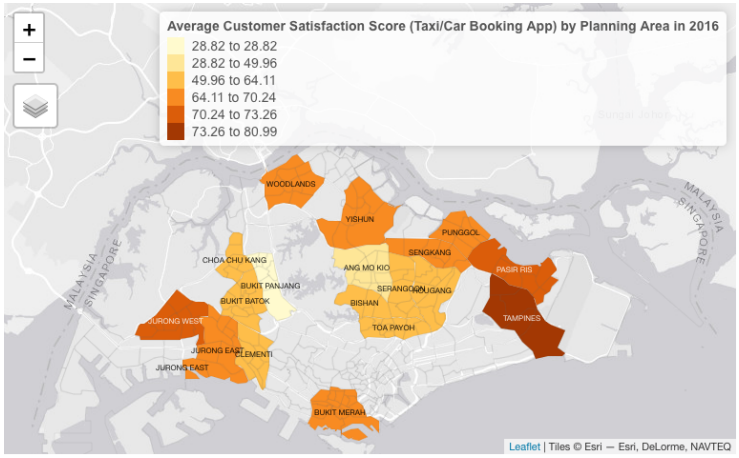
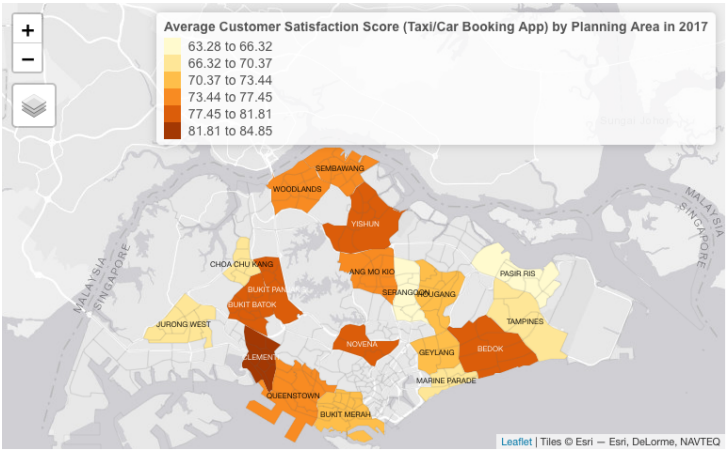
Findings 2: Particularly, areas like *Ang Mo Kio* and *Bedok* could be seen with a slight decrease from 2018 to 2019 and 2017 to 2018 respectively.

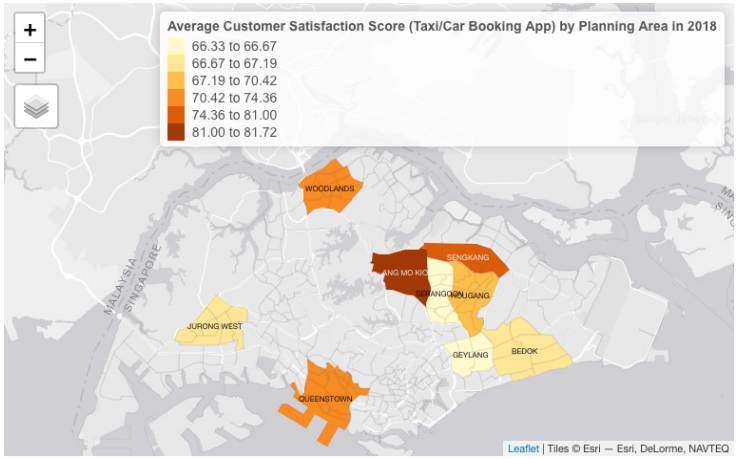


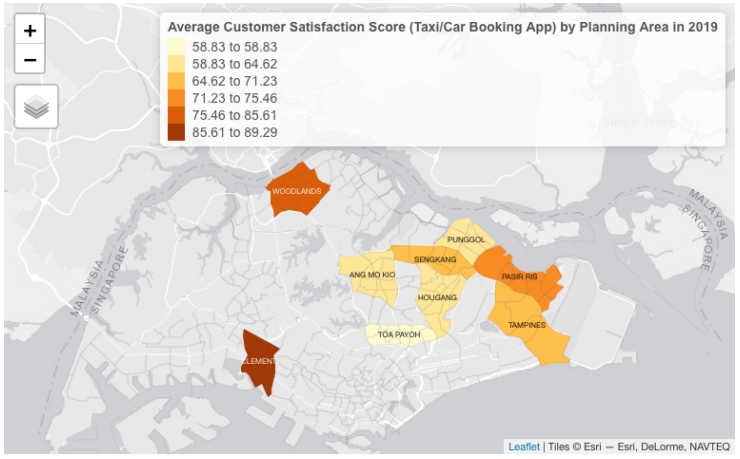
Findings 3: *Novena* could be seen with the highest average customer satisfaction score from 2016 to 2019. Furthermore, planning areas in the North could also be seen as those that have a higher score as compared to the other areas.

Findings 4: Majority of the planning areas have a neutral satisfaction level towards the Taxi/Car Booking Application services in Singapore.

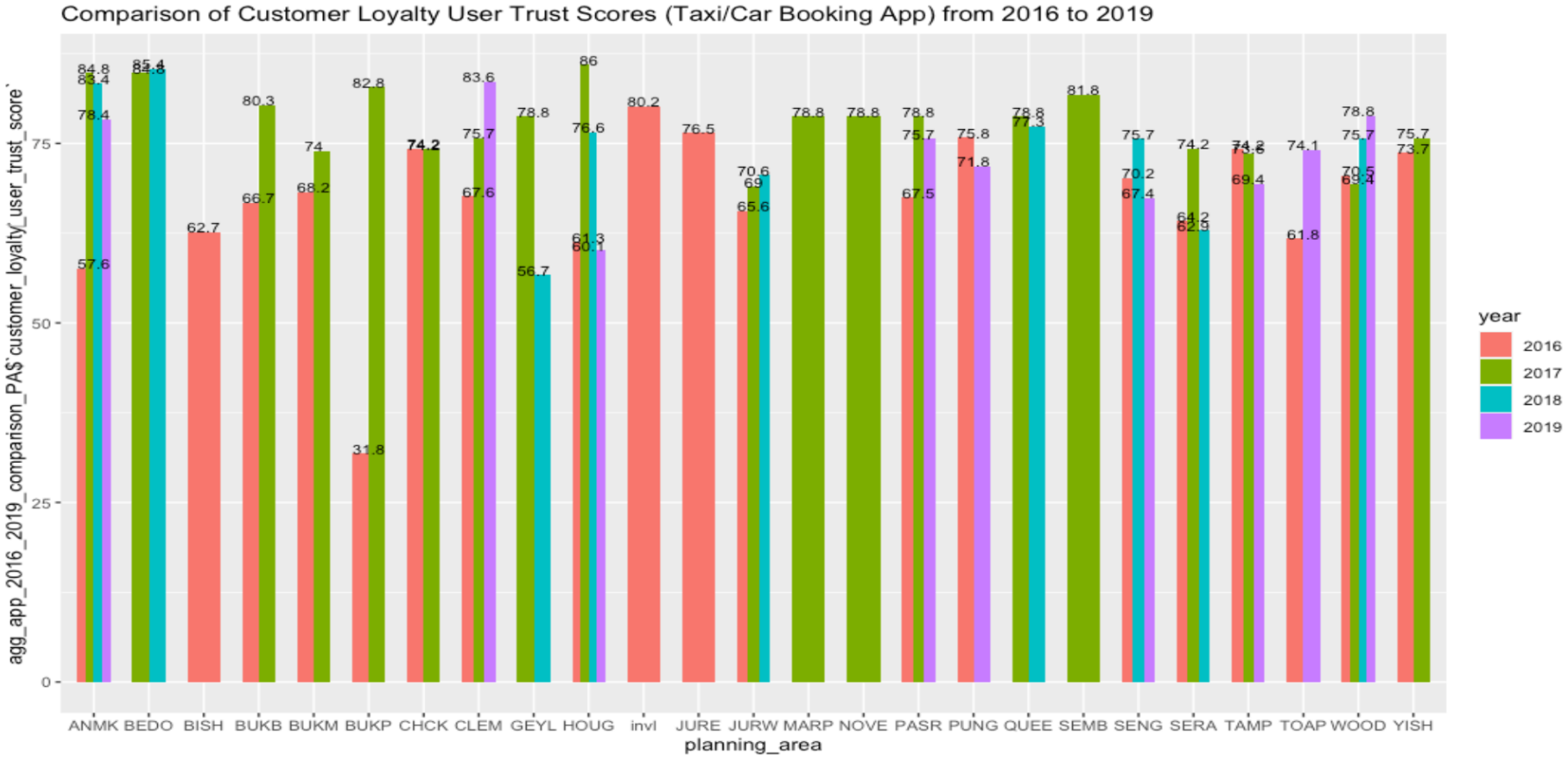
*Breakdown of Customer **Satisfaction** Scores by Year and Planning Area*

2016	2017
	
2018	2019


Average Customer Satisfaction Score (Taxi/Car Booking App) by Planning Area in 2018
66.33 to 66.67
66.67 to 67.19
67.19 to 70.42
70.42 to 74.36
74.36 to 81.00
81.00 to 81.72
WOODLANDS
SENGKANG
ANG MO KIO
SERANGOON
HOUGANG
JURONG WEST
GEYLANG
BEDOK
QUEENSTOWN
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

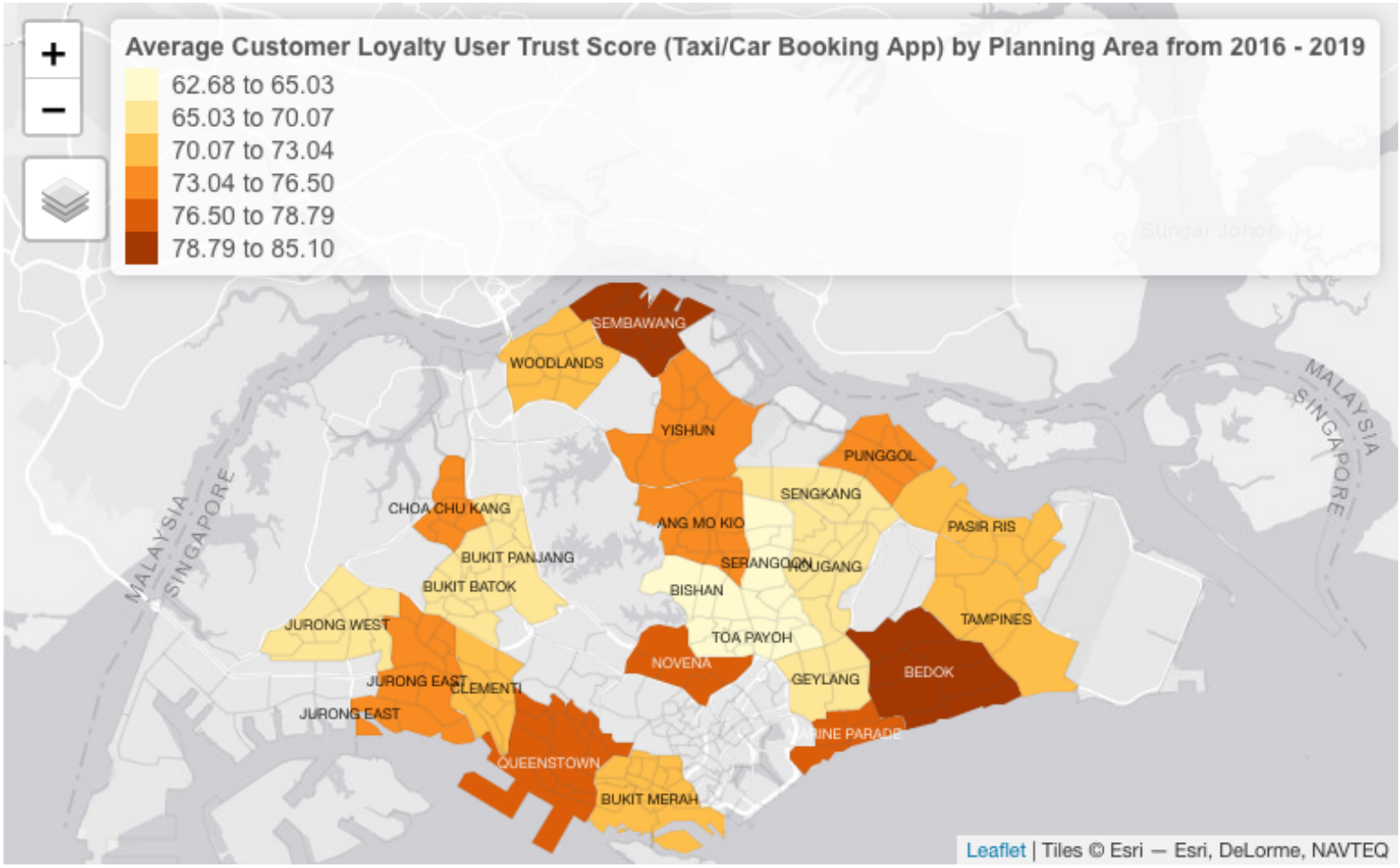

Average Customer Satisfaction Score (Taxi/Car Booking App) by Planning Area in 2019
58.83 to 58.83
58.83 to 64.62
64.62 to 71.23
71.23 to 75.46
75.46 to 85.61
85.61 to 89.29
WOODLANDS
PUNGGOL
SENGKANG
ANG MO KIO
PASIR RIS
HOUGANG
TOA PAYOH
TAMPINES
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

11.9.3.3 Customer Loyalty User Trust



Findings 1: Majority of the planning areas have a neutral to positive increase in the customer loyalty user trust scores over the respective years. However. *Geylang* and *Hougang* are seen with a decreasing behaviour over the years.

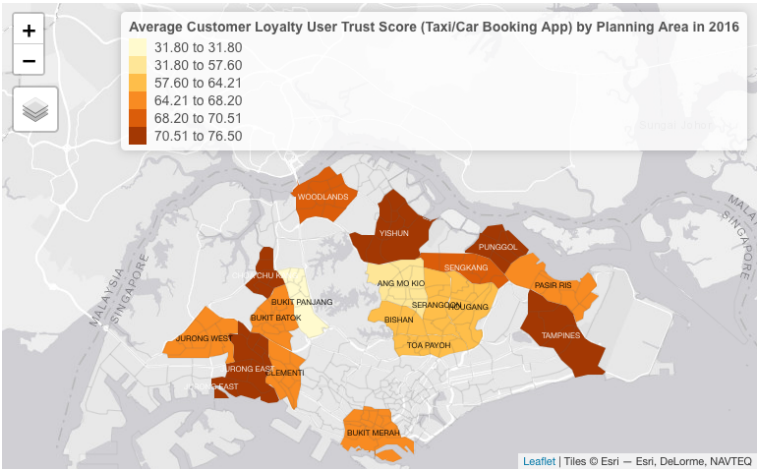
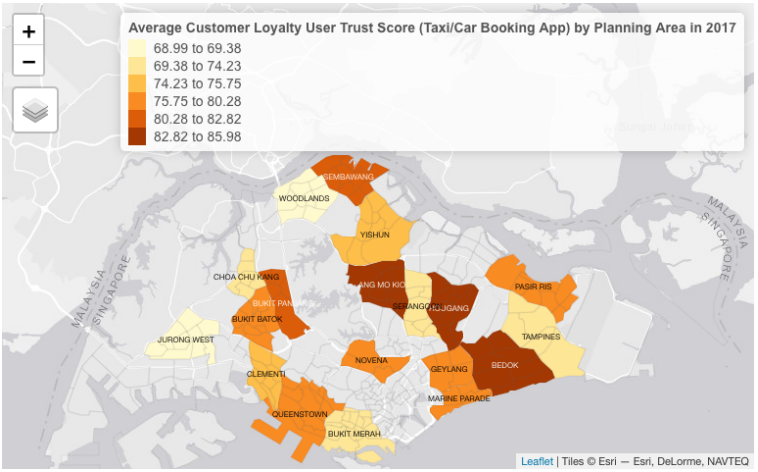
Findings 2: Overall, most would have neutral to positive opinions about the Taxi/Car Booking Application services in Singapore.

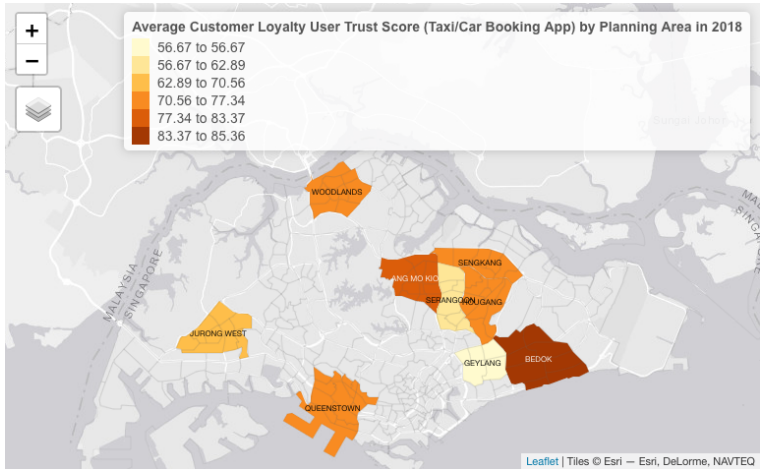


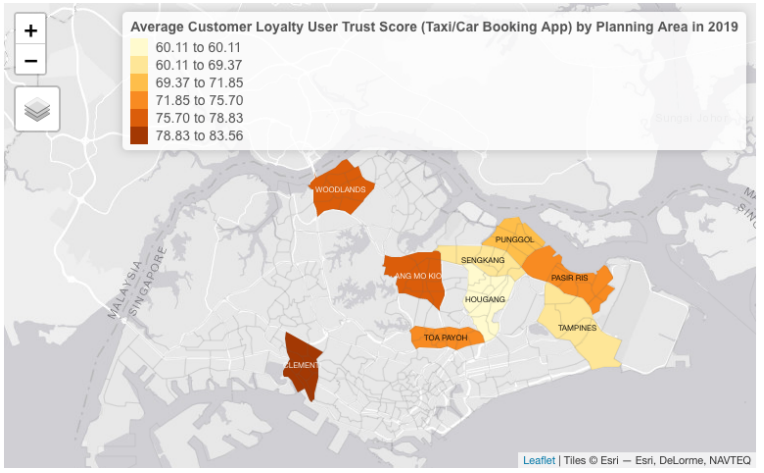
Findings 3: *Sembawang* and *Bedok* are among those that have more positive opinions of the Taxi/Car Booking Application services in Singapore.

Findings 4: Although *Bishan*, *Toa Payoh* and *Serangoon* are in the group with the lowest average customer loyalty user trust scores, they would lean towards a neutral perspective of their opinions on this particular service based on their scores.

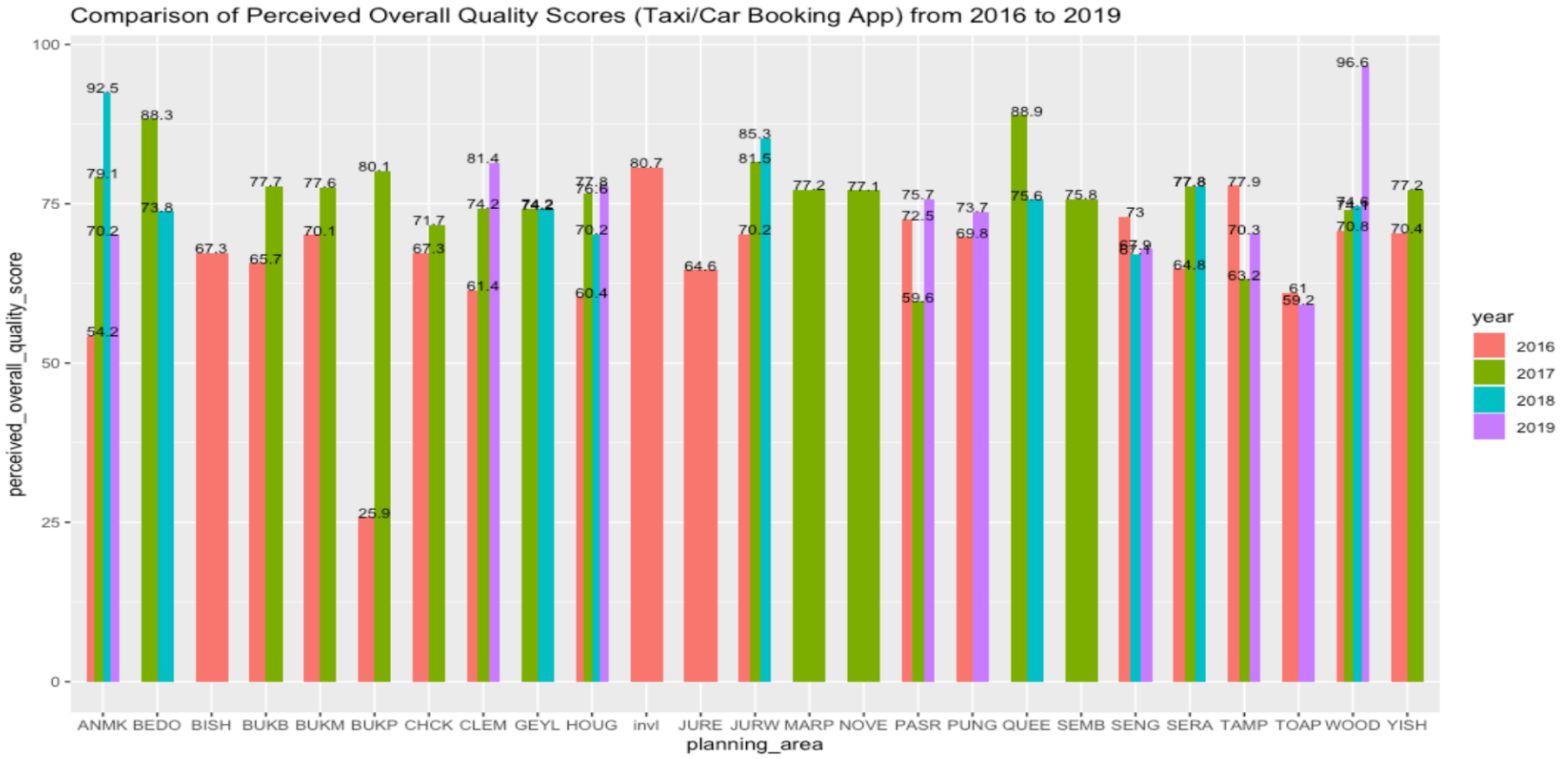
*Breakdown of Customer **Loyalty User Trust** Scores by Year and Planning Area*

2016	2017
	
2018	2019


Average Customer Loyalty User Trust Score (Taxi/Car Booking App) by Planning Area in 2018
56.67 to 56.67
56.67 to 62.89
62.89 to 70.56
70.56 to 77.34
77.34 to 83.37
83.37 to 85.36
WOODLANDS
SENGKANG
HOUGANG
JURONG WEST
GEYLANG
BEDOK
QUEENSTOWN
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

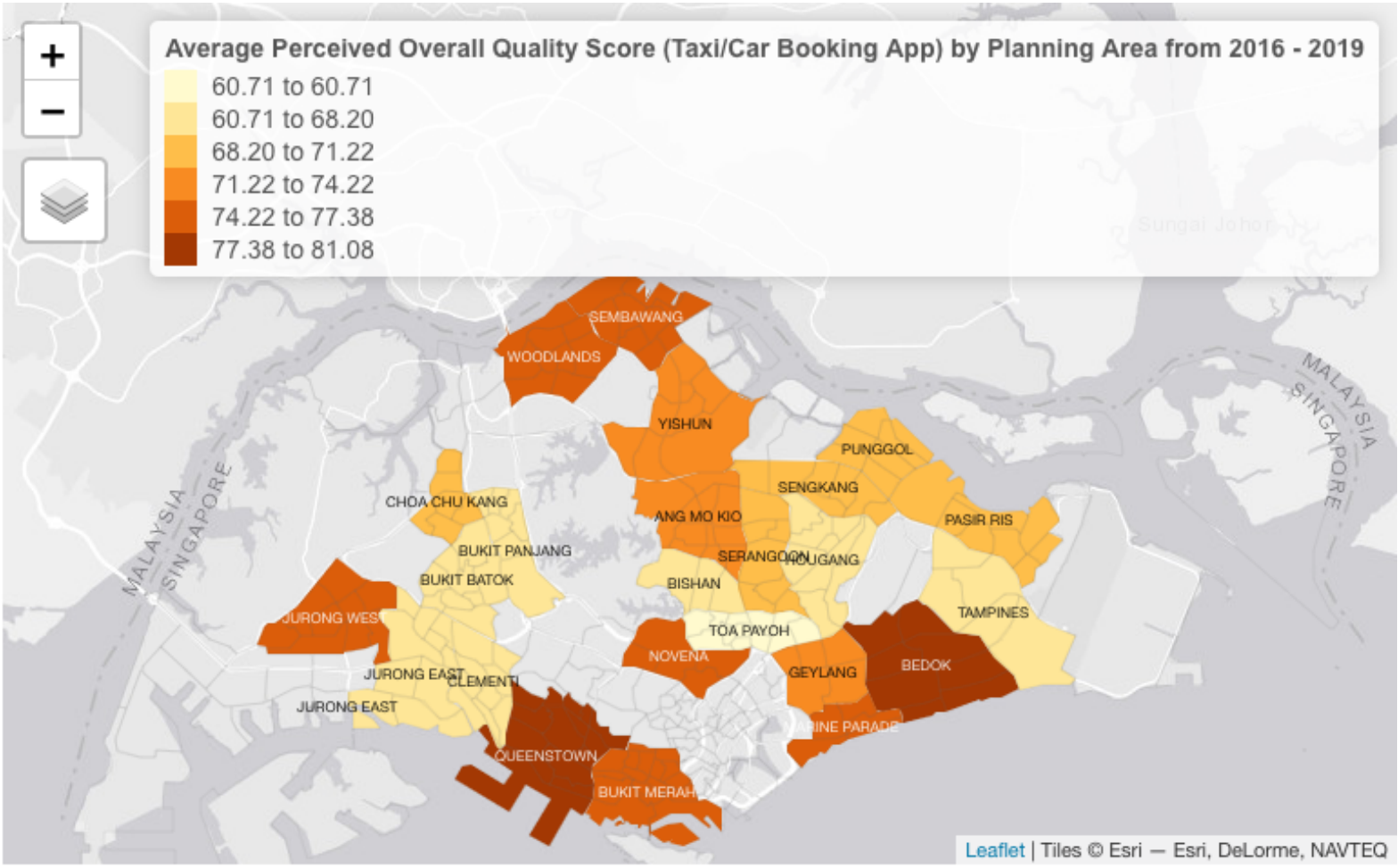

Average Customer Loyalty User Trust Score (Taxi/Car Booking App) by Planning Area in 2019
60.11 to 60.11
60.11 to 69.37
69.37 to 71.85
71.85 to 75.70
75.70 to 78.83
78.83 to 83.56
WOODLANDS
PUNGGOL
SENGKANG
PASIR RIS
HOUGANG
TAMPINES
TOA PAYOH
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

11.9.3.4 Perceived Overall Quality



Findings 1: Generally, the target audience felt positive about the overall quality they received from the Taxi/Car Booking Application services yearly. Particularly, *Woodlands* achieved an all-time high of 96.6 perceived overall quality score in 2019.

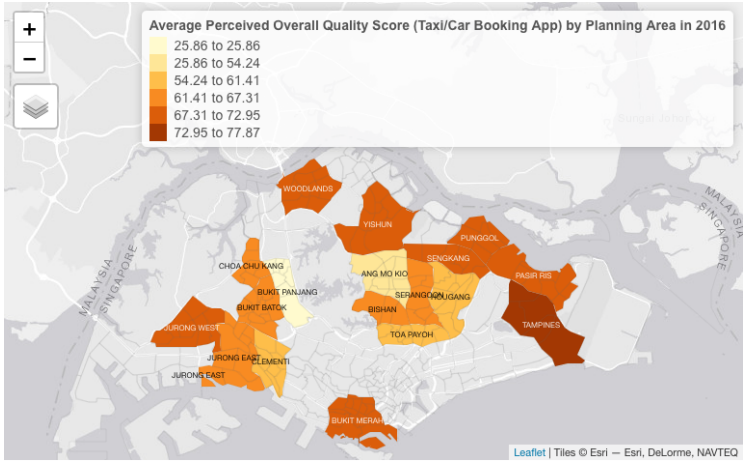
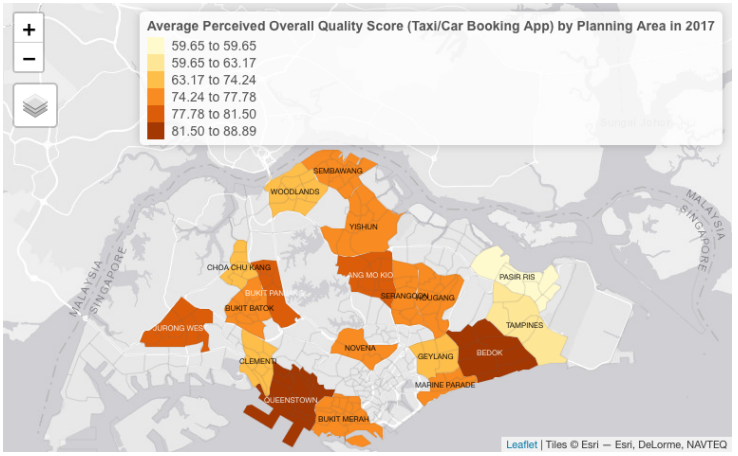
Findings 2: *Ang Mo Kio* and *Bedok* can be seen with a slight decrease in their scores from 2018 to 2019 and 2017 to 2018 respectively.

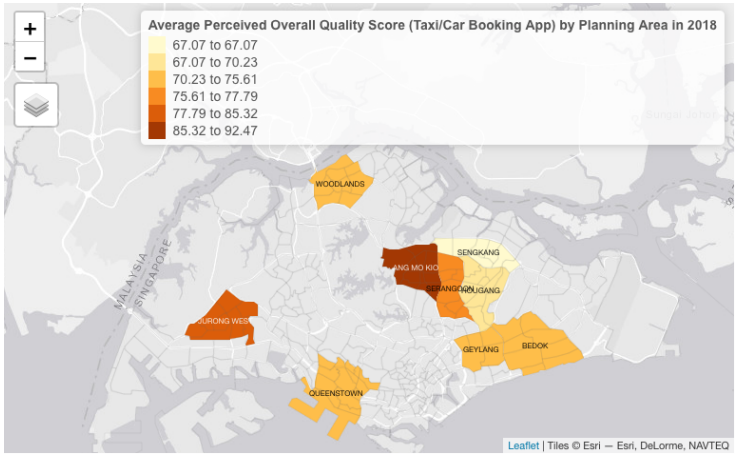


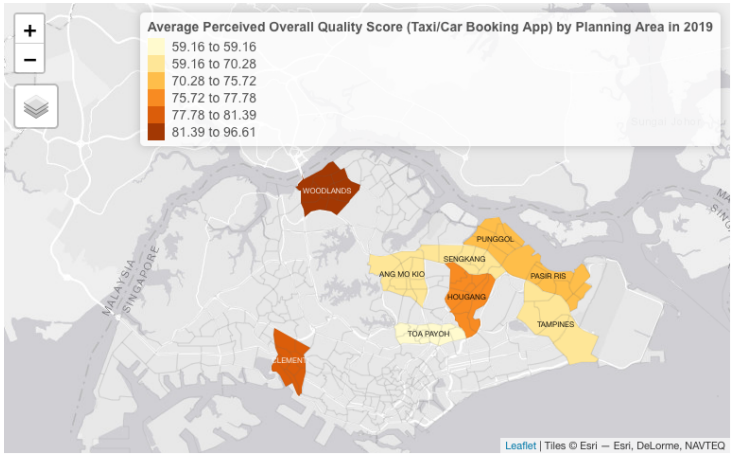
Findings 3: *Queenstown* and *Bedok* are among those with the highest average perceived overall quality scores from 2016 to 2019.

Findings 4: Various areas in the West and North-East regions could be seen with a general neutral perspective in the quality they received, as well as if the service provider meets their personal requirements.

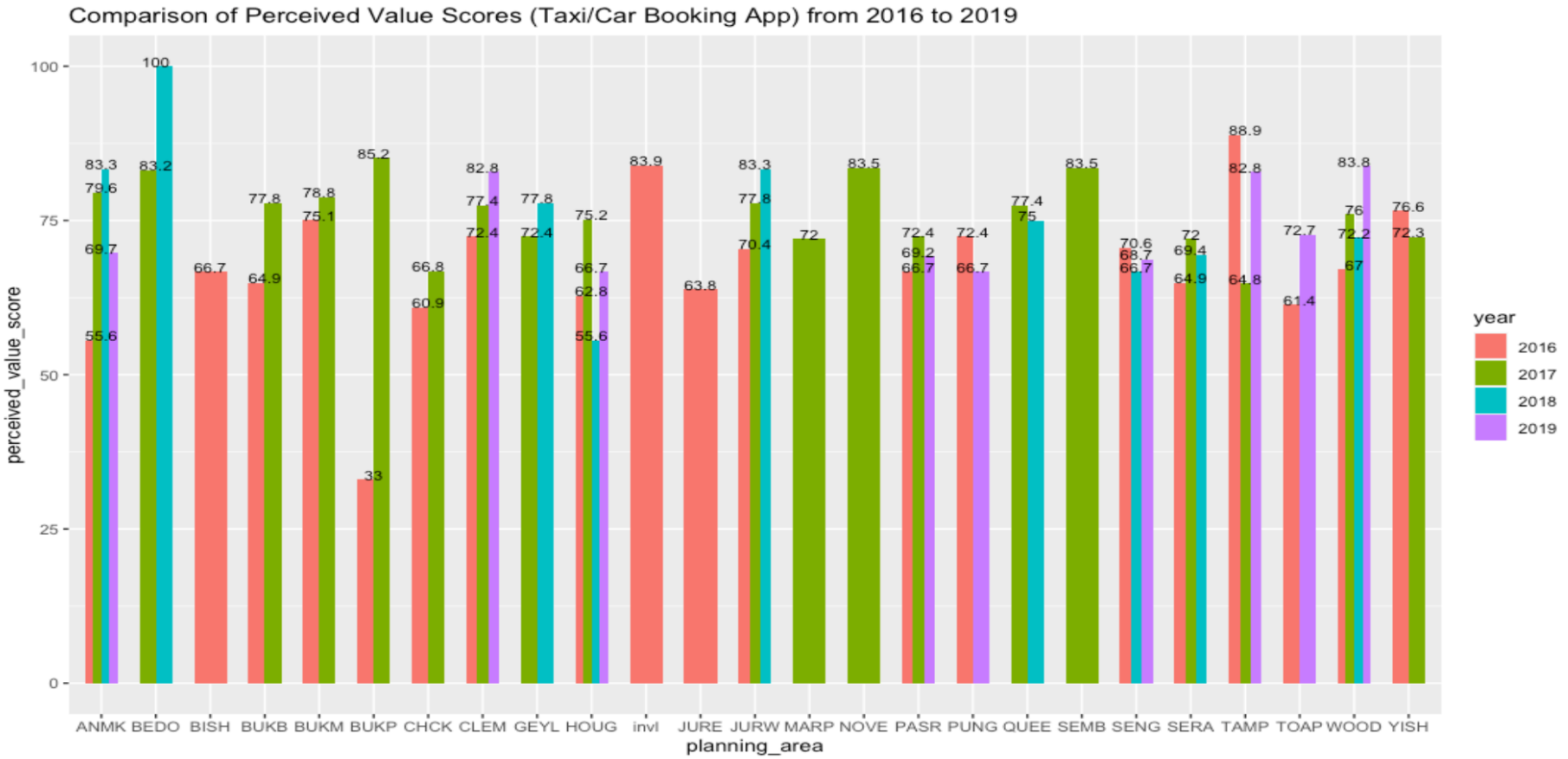
*Breakdown of **Perceived Overall Quality** Scores by Year and Planning Area*

2016	2017
	
2018	**2019**


Average Perceived Overall Quality Score (Taxi/Car Booking App) by Planning Area in 2018
67.07 to 67.07
67.07 to 70.23
70.23 to 75.61
75.61 to 77.79
77.79 to 85.32
85.32 to 92.47
WOODLANDS
SENGKANG
HOUGANG
GEYLANG
BEDOK
QUEENSTOWN
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

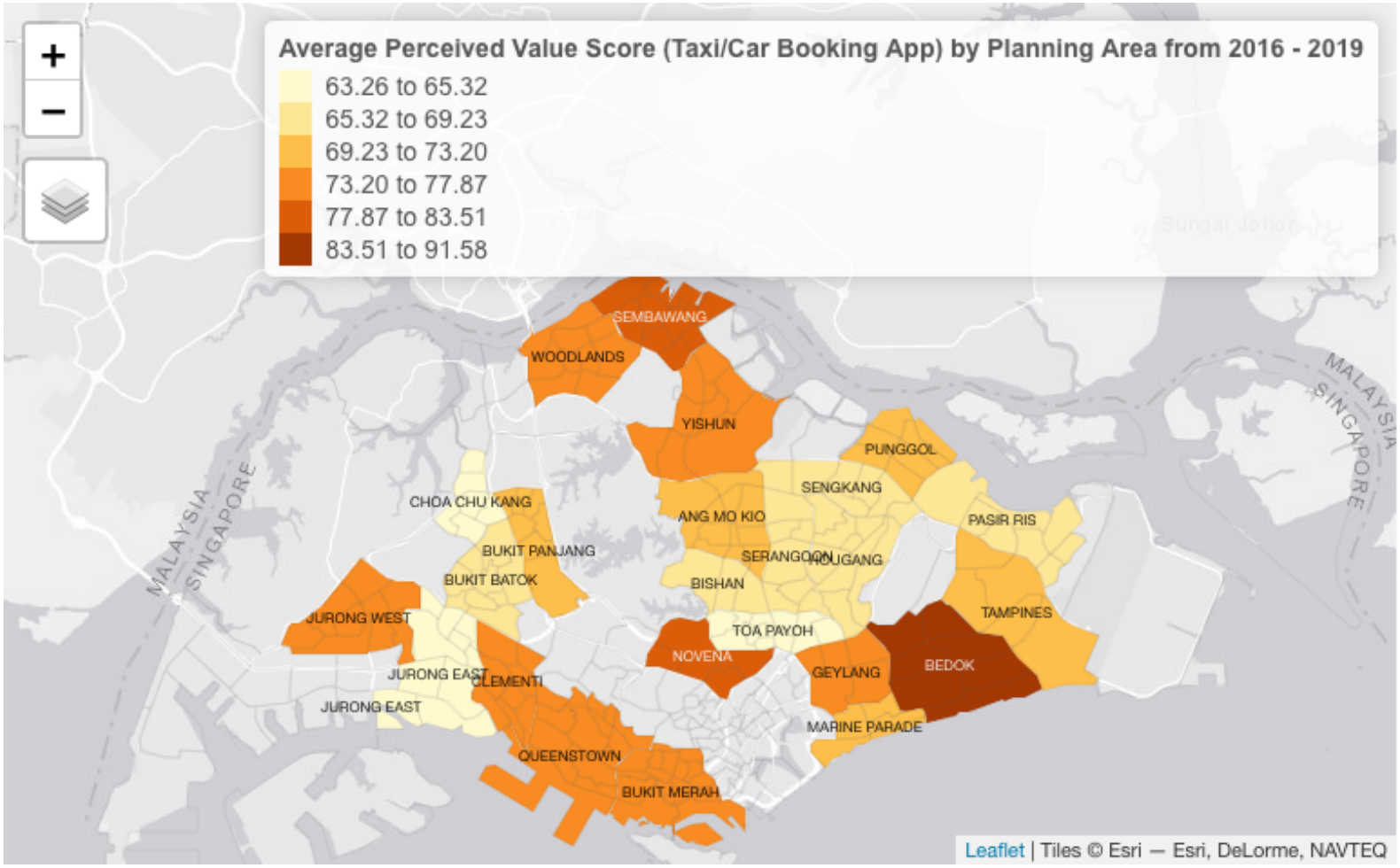

Average Perceived Overall Quality Score (Taxi/Car Booking App) by Planning Area in 2019
59.16 to 59.16
59.16 to 70.28
70.28 to 75.72
75.72 to 77.78
77.78 to 81.39
81.39 to 96.61
WOODLANDS
PUNGGOL
SENGKANG
ANG MO KIO
PASIR RIS
HOUGANG
TOA PAYOH
TAMPINES
MALAYSIA
SINGAPORE
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

11.9.3.5 Perceived Value



Findings 1: There has been a neutral to positive increase in the perceived value scores across the respective planning areas yearly.

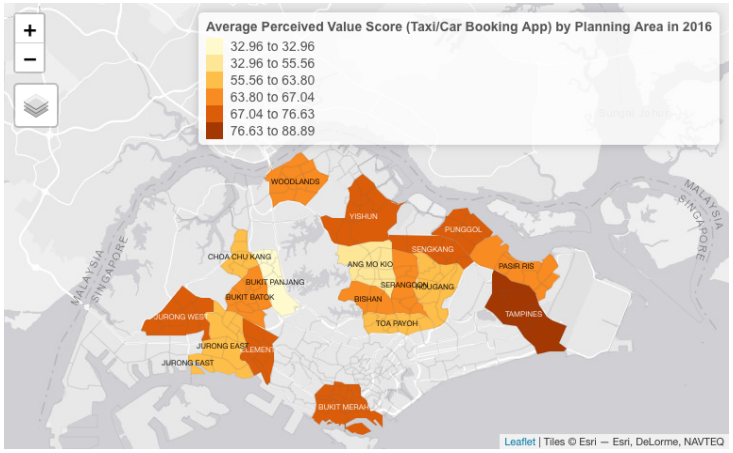
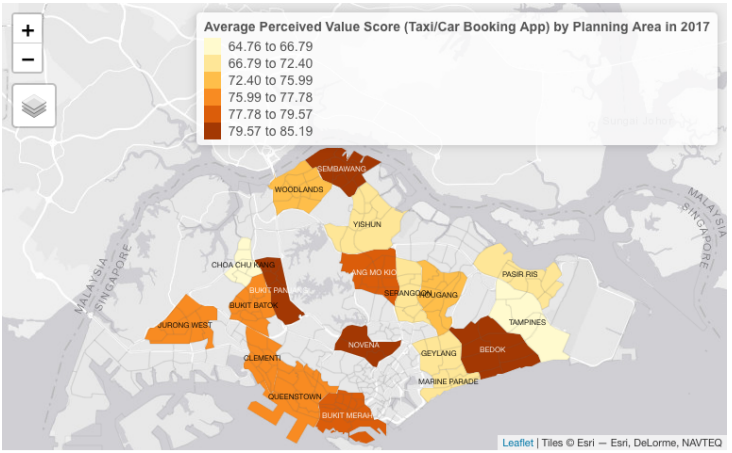
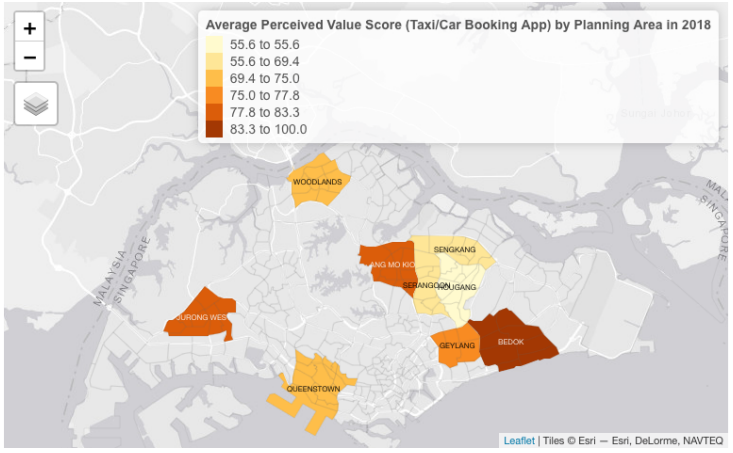
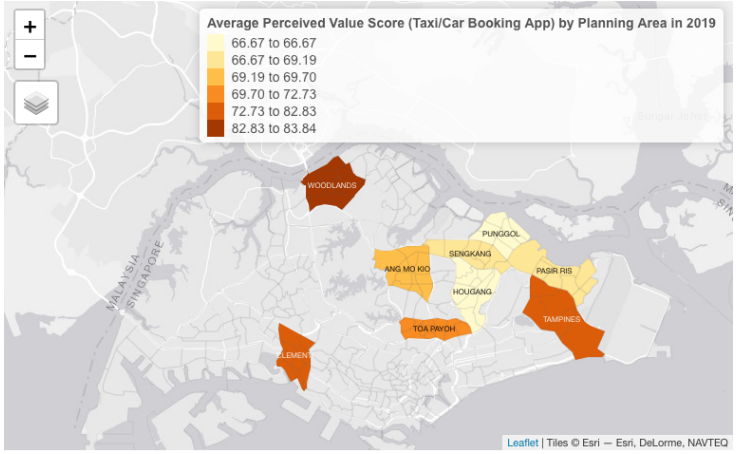
Findings 2: Particularly, *Ang Mo Kio* could be seen with a slight decrease from 83.3 to 69.7 of its scores from 2018 to 2019.



Findings 3: *Bedok* is the one with the highest average perceived value score for Taxi/Car Booking Application services in Singapore.

Findings 4: Various planning areas in the center of the map could be seen with an average ranging from 63.26 to 73.20 of the average perceived value scores for this particular service. This could possibly imply that the perceptions of quality and pricing of this service are generally neutral.

*Breakdown of **Perceived Value** Scores by Year and Planning Area*

2016	2017
	
2018	**2019**
	

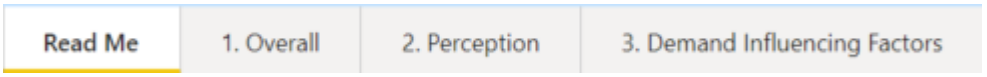
11.10 Appendix J: User Guide

About the Dashboard

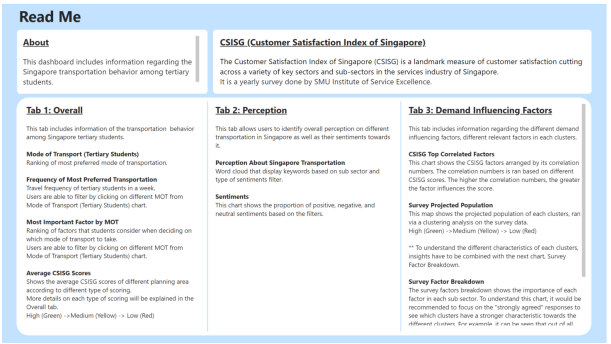
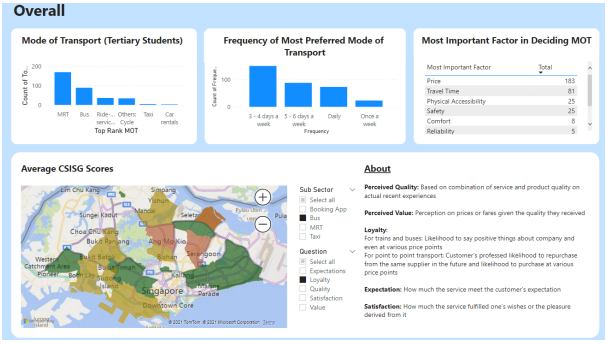
This dashboard is designed to show the overall point-to-point transportation landscape of Singapore by providing tabular and graphical visualization. The dashboard has 4 tabs in total, namely, Read me, Overall, Perception, Demand-influencing factor.

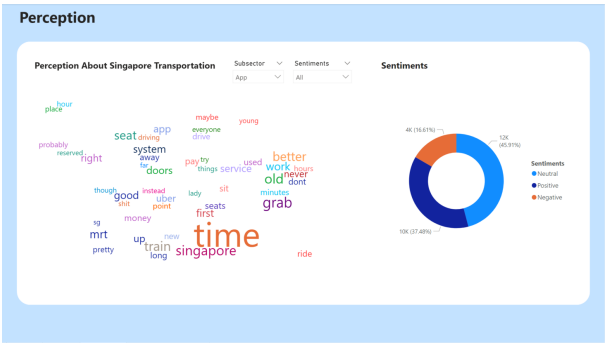
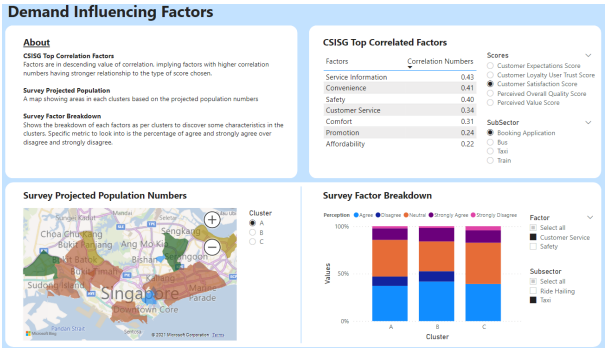
Navigating the Dashboard Tabs

To navigate to the different dashboard, click on the tabs located at the bottom of the dashboard.

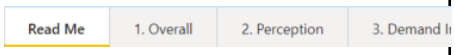
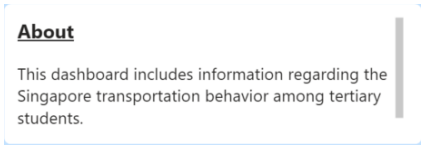
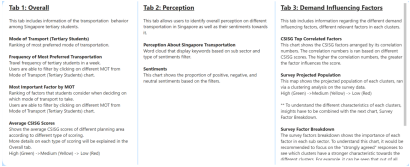


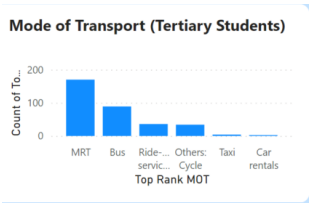
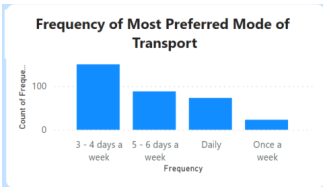
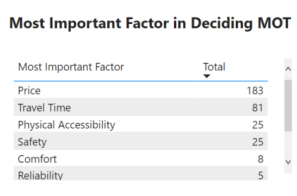
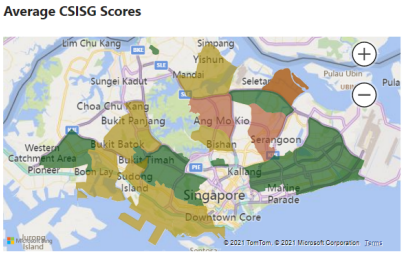
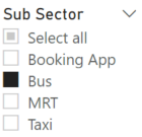
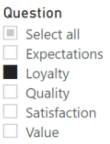
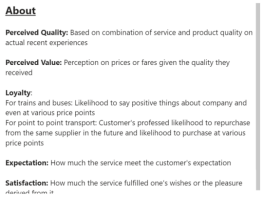
Available Dashboard Tabs

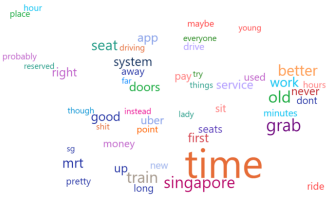
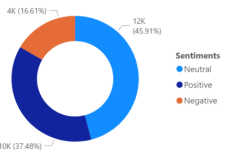
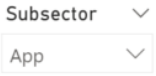
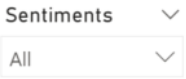
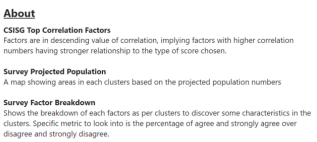
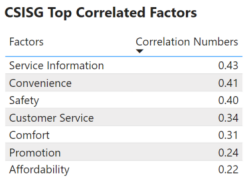
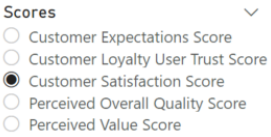
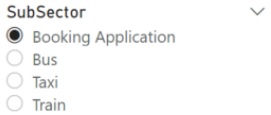
Tabs	Screenshots	Details
Read Me		This tab provides users with the necessary information on each tab and charts. Users will be able to get a summary and explanation of the visualizations available.
1. Overall		The overall tab shows the overall ranking of Mode of Transport, frequency and ranking of factors in deciding Mode of Transport. By filtering sub-sectors and question types, users could identify ratings of different planning areas.

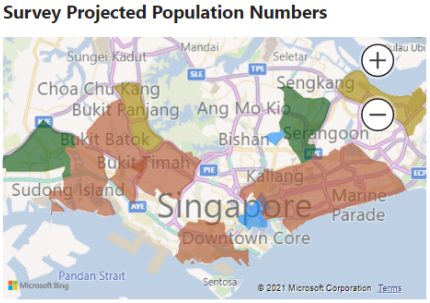
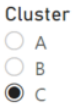
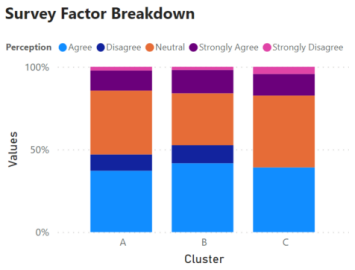
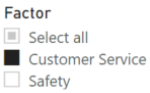
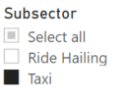
2. Perception		This tab shows the overall ground sentiments and discussions of Singapore transportation. The word cloud would visualize the top keywords of discussion and the donut chart would show the proportion of sentiment types.
3. Demand-Influencing Factor		The demand-influencing tab includes information regarding the different demand influencing factors, and the relevant factors in each cluster.

Available Dashboard Components

Components	Screenshots	Details
Tabs		Allows users to navigate each tab by clicking on the respective tabs.
Read Me Tab		
About Text Box		This is a text box to information users about the purpose of the dashboard
CSISG Text Box		This is a text box about who CSISG is and their relevance to the dashboard
Tabs About Text Box		This is a text box describing the different tabs, its purpose, and what each chart is about.
1. Overall		

Mode of Transport Chart		This chart shows the mode of transport preferred by tertiary students, showing the count of each preferred transport.
Frequency of Most Preferred Mode of Transport Chart		This chart shows the travel frequency of tertiary students base on their most preferred mode of transport.
Most Important Factor in Deciding Mode of Transport Table		This chart shows the most important factor survey respondents choose when deciding which mode of transport to take.
Average CSISG Scores Map		This map shows the average CSISG Scores based on each sub sector and questions filters.
Sub Sector Filter		The subsector filter can be used to filter the Average CSISG Scores map.
Question Filter		The questions filter can be used to filter the Average CSISG Scores map.
About Text Box		This About text box explains the different scoring and what it means.
2. Perception		

Perception About Singapore Transportation WordCloud		This word cloud shows the keywords discussed by Singaporeans about transportation.
Sentiment Donut Chart		This chart shows the different sentiment breakdown.
Subsector Filter		This filter can be used to filter out the specific sub sector for perception and sentiments.
Sentiments Filter		This filter can be used to filter out the specific sentiments for the perception word cloud and sentiments donut chart.
3. Demand-Influencing Factors		
About Text Box		This text box tells users about what the individual charts are.
CSISG Top Correlated Factors Table		This table shows the correlation numbers of each factor in descending order.
Scores Filter		This filter can be used to filter the specific scoring for the correlation factors.
SubSector Filter		This filter can be used to filter the specific sub sector for the correlation factors.

Survey Projected Population Numbers Map		This map shows the different clusters and this projected population numbers in each planning area.
Cluster Filter		This cluster filter will be used on the projected population numbers map to view the different clusters.
Survey Factor Breakdown Chart		This chart shows the different perception breakdown per factor to identify characteristics of each cluster.
Factor Filter		This filter can be used to filter the survey factor breakdown.
SubSector Filter		This filter can be used to filter the survey factor breakdown.