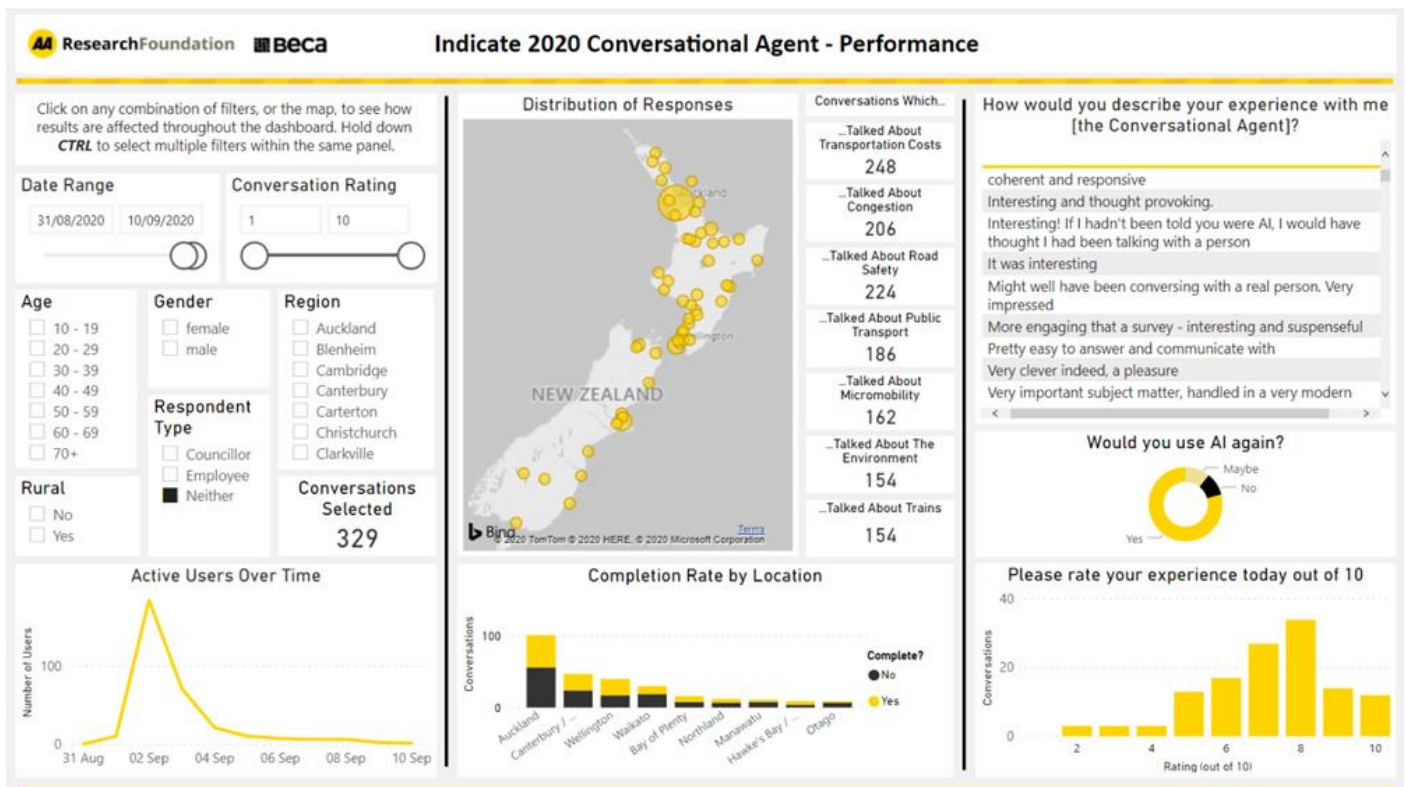


Indicate 2020 – Artificial Intelligence for Member Engagement

Prepared for the Automobile Association Research Foundation (AARF)

by Beca Limited

September 2020



Contents

1	Introduction	1
1.1	Background	1
1.2	Research Approach	1
2	Research Objectives	2
2.1	Research Outcomes	2
3	The Artificial Intelligence Technology	3
3.1	Research Tool	3
3.2	AI, Natural Language Processing and IBM Watson	3
3.3	Design, Testing and Training	3
4	Indicate 2020	5
4.1	Functionality	5
4.2	Sentiment	5
4.3	Topics Indicate collects information on	6
4.4	Reporting & Dashboarding	6
5	Trial Results	11
5.1	Introduction	11
5.2	District Councillor Conversations	11
5.3	Feedback on using Indicate 2020	20
6	Interim Recommendations	21
7	Members Release Results	22
7.1	Introduction	22
7.2	AA Member Conversations	23
7.3	Feedback on using Indicate 2020	32
8	Conclusions on the use of Artificial Intelligence	37
9	Potential Next Steps	39
10	Appendix – Summary of Testing Results	40

Table of Figures

Figure 4-1 Summary of Engagement page (from the Indicate dashboard http://feedback.indicate.chat , p. 1) .	7
Figure 4-2 Summary of Demographics page (from the Indicate dashboard http://feedback.indicate.chat , p. 2)	7
Figure 4-3 'Issue-Specific' responses page (from the Indicate dashboard http://feedback.indicate.chat , p.3) .	8
Figure 4-4 'Issue-Specific' responses page (from the Indicate dashboard http://feedback.indicate.chat , p.4) .	8
Figure 4-5 Issue Specific Focus page (Congestion) (from the Indicate dashboard http://feedback.indicate.chat , p. 5).....	9
Figure 4-6 General Transport Questions page (Congestion) (from the Indicate dashboard http://feedback.indicate.chat , p. 6).....	9
Figure 4-7 AI Experience Feedback page (from the Indicate dashboard http://feedback.indicate.chat , p. 7)	10
Figure 4-8 – Leader board dash (see http://leaderboard.indicate.chat)	10
Figure 5-1 - Demographic information on respondents (from the Indicate dashboard http://feedback.indicate.chat , p. 2).....	11
Figure 5-2 - Demographic information on respondents (from the Indicate dashboard http://feedback.indicate.chat , p. 6).....	12
Figure 5-3 - Summary of responses to questions on the topic of congestion (from dashboard)	13
Figure 5-4 Summary of responses to questions on the topic of public transport (from dashboard)	14
Figure 5-5 Summary of responses to questions on the topic of environment (from dashboard)	15
Figure 5-6 Summary of responses to questions on the topic of transportation costs (from dashboard).....	16
Figure 5-7 Summary of responses to questions on the topic of road safety (from dashboard)	17
Figure 5-8 Summary of responses to questions on the topic of micromobility (from dashboard)	18
Figure 5-9 Summary of responses to questions on the topic of trains (from dashboard)	19
Figure 5-10 – Respondents’ engagement with, and feedback on, the conversational agent experience (from the Indicate http://feedback.indicate.chat , p. 7)	20
Figure 8-1 - Demographic information on respondents (from the Indicate dashboard http://feedback.indicate.chat , pp. 1 and 2).....	23
Figure 8-2 - Demographic information on respondents (from the Indicate dashboard http://feedback.indicate.chat , p. 6).....	24
Figure 8-3 - Summary of responses to questions on the topic of congestion (from dashboard)	25
Figure 8-4 Summary of responses to questions on the topic of public transport (from dashboard)	26
Figure 8-5 Summary of responses to questions on the topic of environment (from dashboard)	27
Figure 8-6 Summary of responses to questions on the topic of transportation costs (from dashboard).....	28
Figure 8-7 Summary of responses to questions on the topic of road safety (from dashboard)	29
Figure 8-8 Summary of responses to questions on the topic of micromobility (from dashboard)	30

| INDICATE 2020 |

Figure 8-9 Summary of responses to questions on the topic of trains (from dashboard) 31

Figure 8-10 – Respondents’ engagement with, and feedback on, the conversational agent experience (from the Indicate <http://feedback.indicate.chat>, p. 7 and conversation length from p. 1) 32

Figure 11-1 Comprehensive Testing Suite as of the 7th of February 2020 40

1 Introduction

1.1 Background

The New Zealand Automobile Association (the AA) has 1.7 million members across New Zealand. The Motoring Affairs team of the AA undertakes advocacy on behalf of AA members to promote improved road safety and investment in the transport system, and acts as the motorists' champion on issues including costs and charges.

Underpinning the advocacy of the AA is a comprehensive Member survey system. Members are surveyed as part of both a regular quarterly survey to track attitudes and trends, and by one-off surveys on specific issues. This survey approach provides the Motoring Affairs team with insight into members' views on a wide range of topics.

This current system provides opportunities for members to express opinions on issues they are being surveyed on through free-form text boxes. On some issues a huge volume of material is collected, but its analysis is time consuming and difficult. The use of artificial intelligence (AI) offers the opportunity to systematically collect member opinions in a way that is accessible for analysis, enabling the depth and breadth of member sentiment on important transport issues to be more easily understood.

1.2 Research Approach

This research project through the AA Research Foundation examined the use of artificial intelligence to fill a gap in knowledge around AA Members' views.

The research programme was as follows:

Research Task	Programme
Initial Meeting with Automobile Association Communications & Policy Staff	June 2019
Initial Prototype of AI Tool, called "Indicate 2020", including Chatbot interface (http://indicate.chat) and dashboard (http://dashboard.indicate.chat).	November 2019
Testing and Training of the tool amongst Beca team and AA Policy & Media personnel.	December 2019 through March 2020
Creation of a staff engagement dashboard (http://leaderboard.indicate.chat) and release of Indicate 2020 to AA Staff during Covid-19 lockdown period.	April 2020
Release of Indicate 2020 to Automobile Association District Councillors to collect their views on a range of topics, and their feedback	May – June 2020

on using the Indicate 2020 AI Tool (http://feedback.indicate.chat)	
Meeting with AA Team to discuss research findings	June 2020
Release of Indicate 2020 to AA Panel	July 2020
Reporting on AA Councillor Trial	July 2020
Release of Indicate 2020 to AA Membership	September 2020
Final Reporting on Research Project	October 2020

2 Research Objectives

2.1 Research Outcomes

The research outcomes identified in June 2019 were:

- Analytics identifying the issues discussed, their suggestions, and sentiment towards issues / initiatives.
- Analytics on conversations held by the AI tool (locations, issues discussed, etc.)
- Analysis of how the use of AI has altered / improved the ability to collect AA membership sentiment towards road safety initiatives in NZ.

3 The Artificial Intelligence Technology

3.1 Research Tool

The AI tool used in this research was initially developed by Beca in early 2019 to enable a new and innovative method for engaging with people on transport issues. A key feature of the tool is the ability to use Artificial Intelligence to process conversation transcripts and identify key issues.

For this research project, the tool was branded “Indicate” and given a “personality” and “tone” appropriate for conversations with AA members.

3.2 AI, Natural Language Processing and IBM Watson

Artificial Intelligence, or AI, is a term which appears with ever-increasing frequency. Specific definitions of AI vary from context to context but, in the most general terms, it refers to systems capable of simulating human thought processes and intelligence. This includes processes such as Machine Learning, Image Processing, Natural Language Processing and Speech Recognition. The common thread across each of these processes is the ability of AI systems to learn from ‘experience’ – that is, to take onboard new information and adapt the process accordingly – rather than following a static set of rules.

For the purposes of this research, the branch of AI of most interest is Natural Language Processing (NLP), which involves extracting information such as concepts, entities (people, places, things), keywords, sentiment, emotion and semantic roles from ordinary speech or text. NLP is the primary form of AI employed by Watson Assistant, IBM’s platform for conversational AI.

Watson Assistant is one component within the broader Watson suite of applications and tools, which collectively aim to enable the building, running and managing of AI, with a particular view to making AI accessible for business applications. The Indicate tool used in this research is a Conversational Assistant developed using the Watson Assistant platform.

A note on Conversational Assistants: Conversational Assistants are often referred to colloquially as “Chatbots” - whilst their origins are similar, and both employ Natural Language Processing, it is important to understand from a technical perspective that Conversational Agents and Chatbots are not the same. Whereas a chatbot aims simply to converse in a ‘natural manner’, a conversational agent has the specific goal of engaging with people: either by establishing what help they require and providing information when appropriate (‘Customer Service’ type agent), or by listening to their thoughts and guiding the conversation / asking relevant questions so as to draw out detailed opinions on a desired topic (‘Consultation’ type agent). The Indicate tool developed for this research falls into the latter category.

3.3 Design, Testing and Training

The core design process for Indicate was the development of a conversation tree, such that the ‘branch’ within the tree that Indicate will select to pursue next depends on the user’s response to the question at hand. Indicate is programmed with specific phrases and responses, and uses Watson AI to determine the intent, sentiment or subject matter of the user’s response in order to determine where to move to within the conversation tree. If the user’s response does not contain sufficient information to determine a next step, Indicate has the ability to seek more information.

The conversation tree was designed such that the core dialogue is structured into distinct 'premises', or sections of the conversation, for instance: Opening, Issue-Specific Discussion, General Transport Discussion, Feedback Discussion, etc. This is supported by various handler components, which deal with off-topic or unexpected responses, known as digressions, distractions and clarifications.

Once the core structure of Indicate was in place, development then consisted of a combination of two processes: testing and training.

3.3.1 Testing

In terms of formal testing, a set list of testable 'abilities' was prepared in accordance with standard industry expectations of Conversational Agent capability. Each premise within the conversation was tested against each of these abilities as applicable. A full summary of the performance of each premise against each ability is provided in the Appendix (Section 8); but broadly speaking, testing covered abilities such as the ability to ask, repeat and vary the dialogue of questions, to detect relevant responses, and so forth.

3.3.2 Training

Both before and after the formal testing process was undertaken, training of the Conversational Agent took place on an ongoing basis. This involved having conversations with the Conversational Agent, then flagging and correcting at a manual level if a response was not as expected, or if an 'intent' (user direction, request or intention) or 'entity' (thing, place, concept, etc) was incorrectly identified. This manual identification and correction enabled the underlying NLP model to retrain itself accordingly, and therefore correctly deal with similar situations or user responses in future.

Although less structured than the quantitative testing process described above, this training process nevertheless comprised a core part of the development of Indicate, in particular, because it is the most effective way of identifying and accounting for edge cases.

4 Indicate 2020

4.1 Functionality

The Indicate Conversational Agent has a wide range of functionality. Its core functionality includes the ability to:

- Ask open ended and yes/no questions and repeat, clarify or rephrase those questions as appropriate;
- Interpret responses to questions and continue the conversation or press for further detail as appropriate;
- Identify the sentiment of free-form responses to open-ended questions, and classify responses to closed questions (e.g. numerical ratings, yes/no questions, demographic indicator questions);
- Follow user instructions such as skipping a question, or restarting the conversation, or escalating the conversation to a real person.

Indicate also has the ability to handle a broad variety of digressions or distractions, such as:

- Responding appropriately if asked about itself, then guiding the conversation back toward its core purpose;
- Telling a joke when humour seems appropriate, then guiding the conversation back toward its core purpose;
- Answering 'FAQ' type questions if asked, then guiding the conversation back toward its core purpose.

4.2 Sentiment

Alongside collecting face-value responses (e.g. yes/no answers, freeform text answers to open questions, quantitative ratings), Indicate also records 'sentiment' information for response to open-ended questions. Sentiment is recorded as a number between -1 and 1, with 1 indicating that the sentiment of the (open-ended) response was wholly positive, -1 indicating that the sentiment of the response was wholly negative, and 0 indicating neutral content.

This functionality enables respondents' comments to be unpacked instantly and scalably, without requiring substantial manual effort or restricting discussion to 'closed' questions only. Trends, including those by region or by demographic factors, can be easily identified, as can trends in the relationship between responses to different questions. Where sentiment results on any given topic are unexpected or of interest, dashboarding provides the ability to 'deep dive' into the relevant comments in full.

As with any attempt to draw quantitative conclusions from qualitative data, there are limitations on the usefulness of sentiment for a small number of conversations. The real advantage lies in the speed and scalability of analysis of open-form responses, which, of course, is less of a concern when the number of conversations is small.

4.3 Topics Indicate collects information on

Indicate collects information across three broad areas: basic demographic information, issue-specific information, and general transport-based questions.

In terms of demographic information, Indicate collects the age, gender, and location of users, including whether that location is rural or urban. More targeted information such as usual mode of transport and number of vehicles owned is also gathered. Collecting this information enables insights about the core issue-based questions to be drawn at a demographic-specific level.

In the central 'body' of its conversations, Indicate collects information on seven key issues. These include: Transportation Costs, Congestion, Road Safety, Public Transport, Micromobility (e.g. e-scooters), the Environment and Trains.

Indicate also collects information on a small number of general transport topics, including views on the government's current approach to transport in each respondent's area, and whether or not political parties' transport policies will affect how respondents choose to vote in the general election.

4.4 Reporting & Dashboarding

In order to get the most value out of this Conversational Agent's NLP capabilities, a live dashboard was prepared. The live dashboard has eight pages, as follows:

Page 1: Summary of Engagement – usage trends, conversations had, geographic distribution of responses. Selecting a particular location 'bubble' enables usage in that location to be viewed.

Page 2: Summary of Demographics – location, age, gender, urban or rural, vehicle ownership, usual mode of transport. Selecting any particular response (e.g. the 'ages 50-59' bar) enables demographics and average conversation sentiment to be viewed for the selected group only.

Pages 3 and 4: Summary of Responses to Issue-Specific questions. For each set of free-form text responses, the full set of responses can be viewed or exported by clicking the 'focus mode' (hover-over) button in the upper right corner of that visual. Selecting one (or more) sentiment bars enables a filtered 'deep dive' – for instance, selecting the most negative sentiment bar for congestion would enable all issue-specific responses to be viewed only for people who responded particularly negatively on the topic of congestion – therein enabling insights to be drawn about how people who felt a certain way about one issue may feel about another issue.

Page 5: Focus Page – in depth view of responses to a single specific issue, with visibility of change in average sentiment over time. The sentiment slider enables specific groups of responses (e.g. all those who responded positively to any degree) to be examined as a group.

Page 6: Summary of General Transport Responses – ease of transport rating, suggestions for government action on transport, views on road safety and road maintenance. As for pages 3-5, for each set of free-form text responses, the full set of responses can be viewed or exported by clicking the 'focus mode' (hover-over) button in the upper right corner of that visual. Selecting a particular location or transport rating from the visuals on the left enables responses to open questions to be viewed filtered by that selection.

Page 8: Summary of Feedback on AI Experience – rating of the Conversational Agent experience and free-form feedback. This page also provides Filter Selection for the full dashboard – enabling filtering on aspects such as demographics or region which are applied to all content-based pages dashboard (pages 3 through 6).

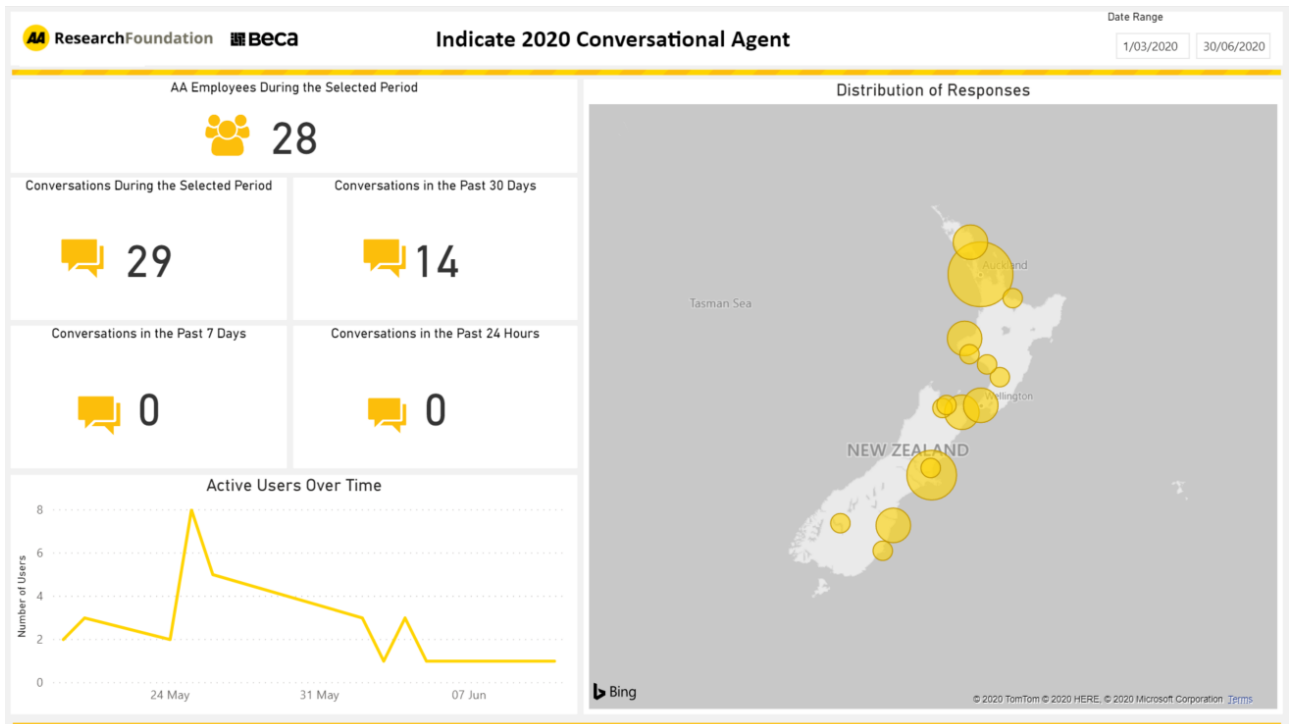


Figure 4-1 Summary of Engagement page (from the Indicate dashboard <http://feedback.indicate.chat>, p. 1)

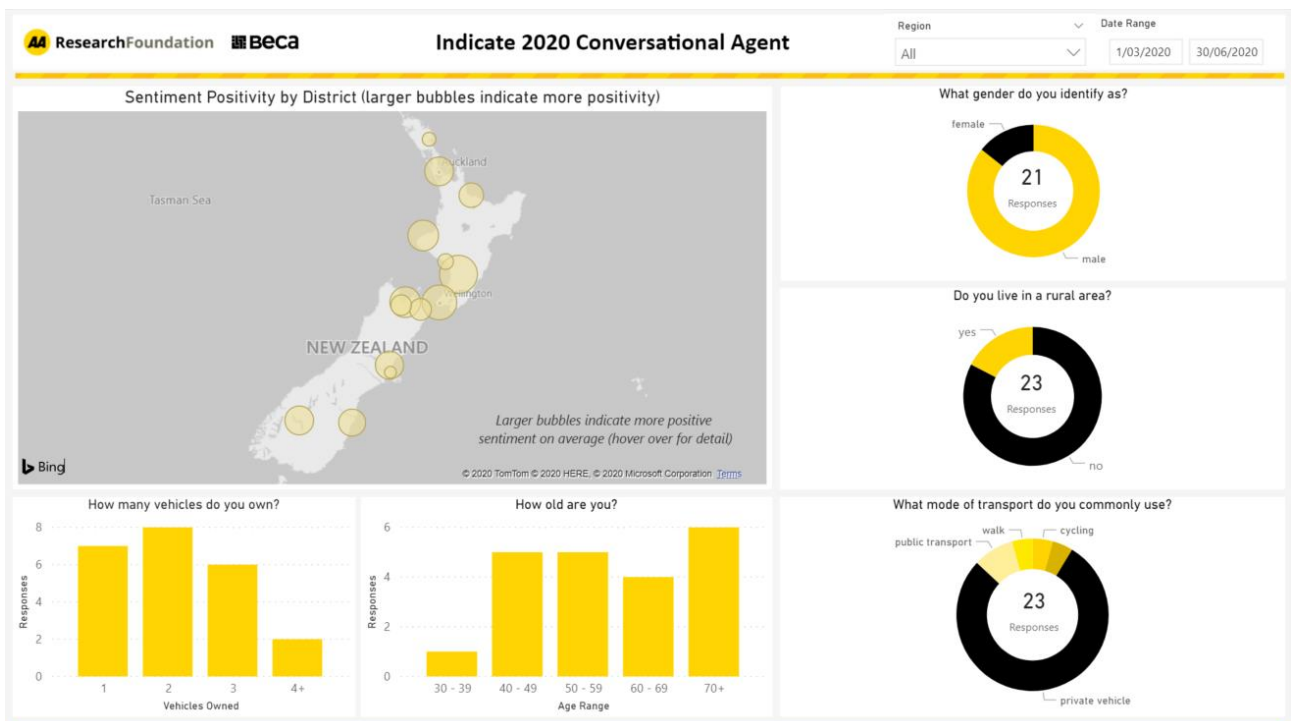


Figure 4-2 Summary of Demographics page (from the Indicate dashboard <http://feedback.indicate.chat>, p. 2)

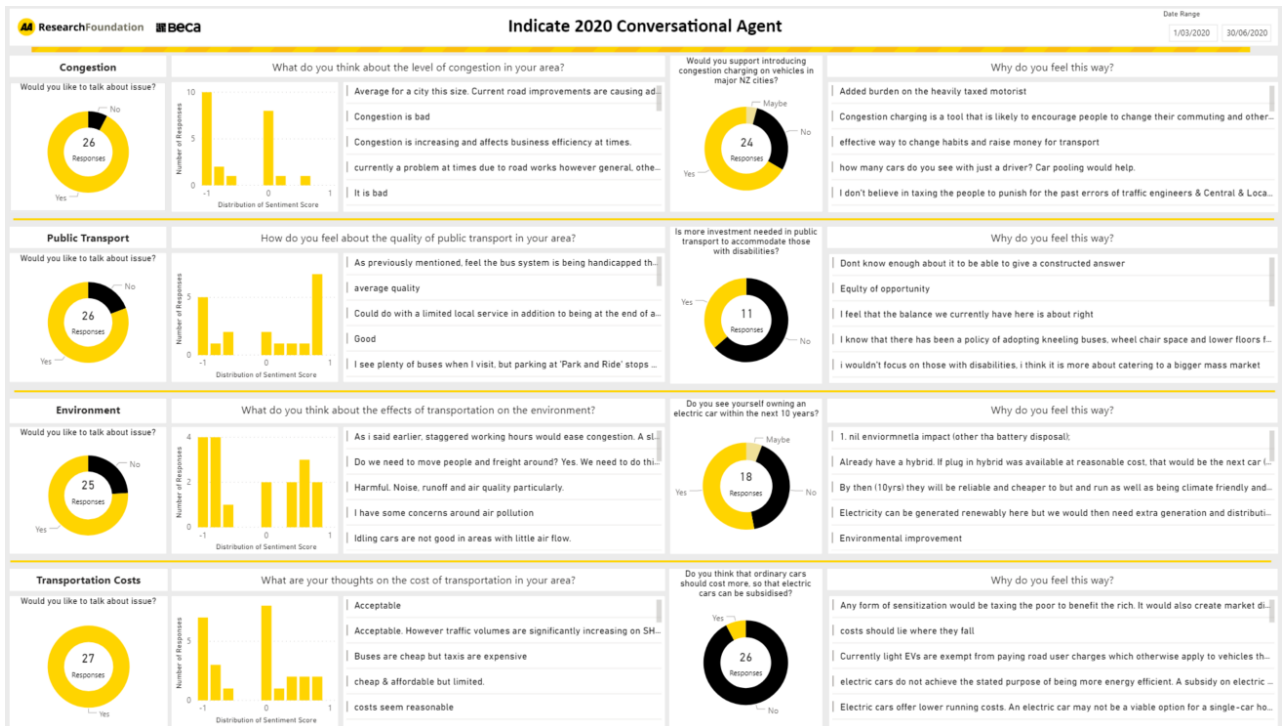


Figure 4-3 'Issue-Specific' responses page (from the Indicate dashboard <http://feedback.indicate.chat>, p.3)

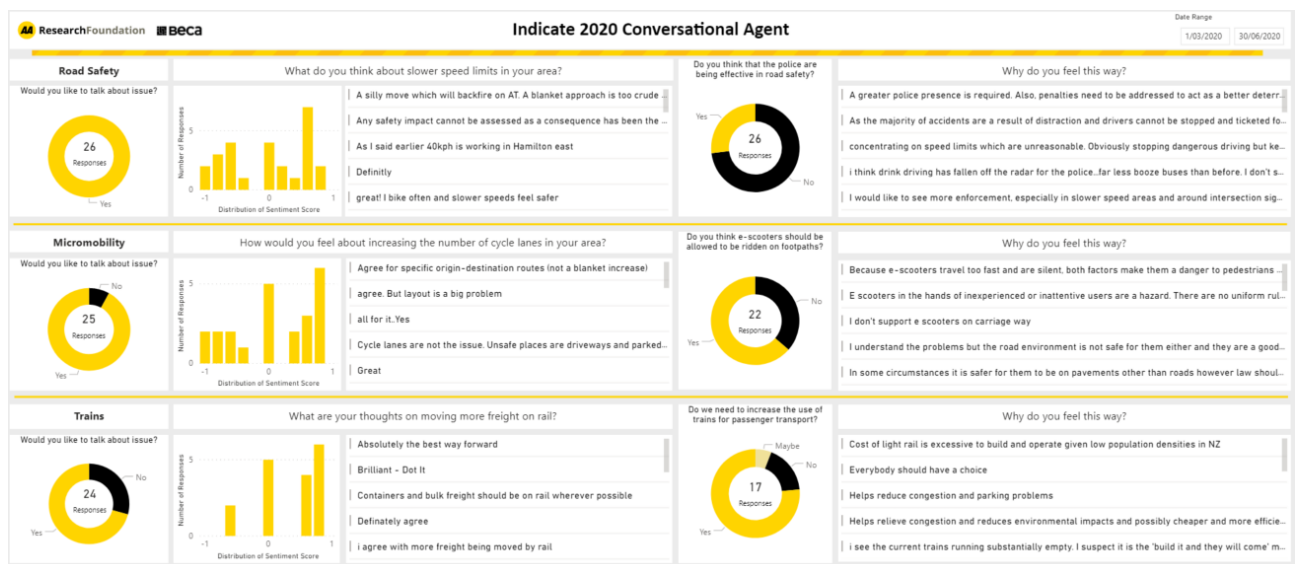


Figure 4-4 'Issue-Specific' responses page (from the Indicate dashboard <http://feedback.indicate.chat>, p.4)

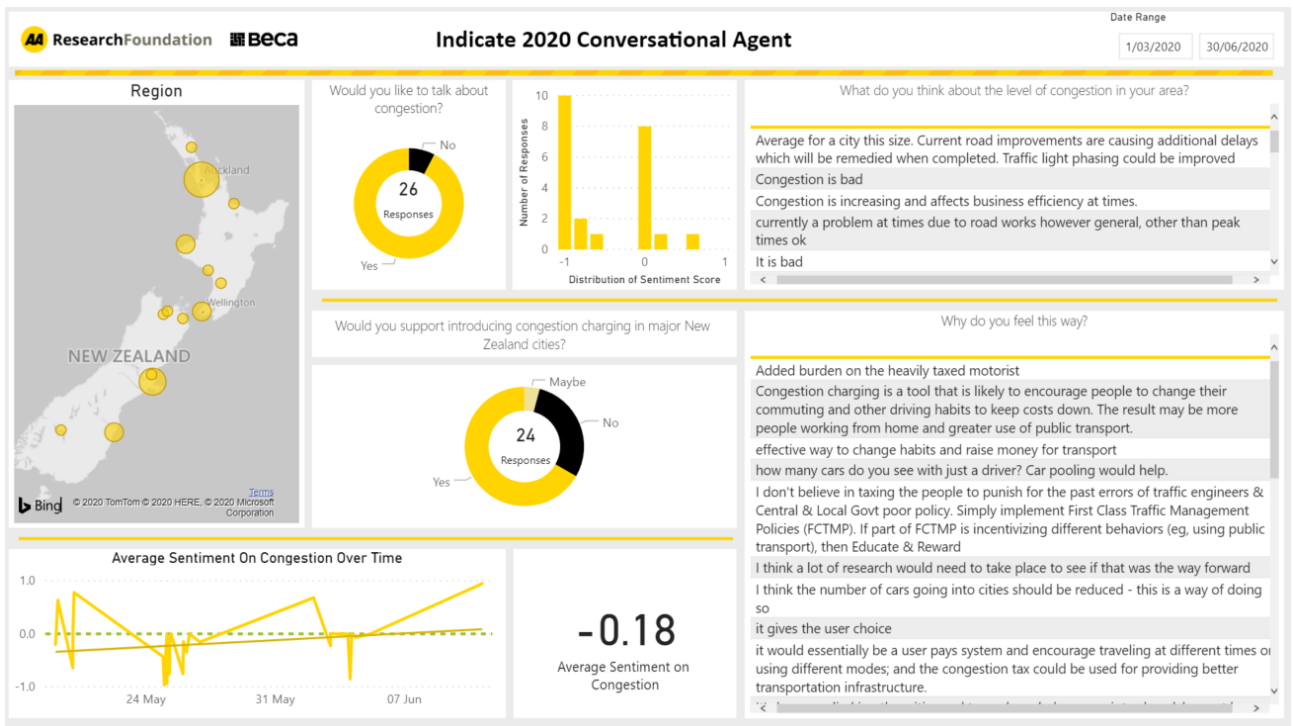


Figure 4-5 Issue Specific Focus page (Congestion) (from the Indicate dashboard <http://feedback.indicate.chat>, p. 5)

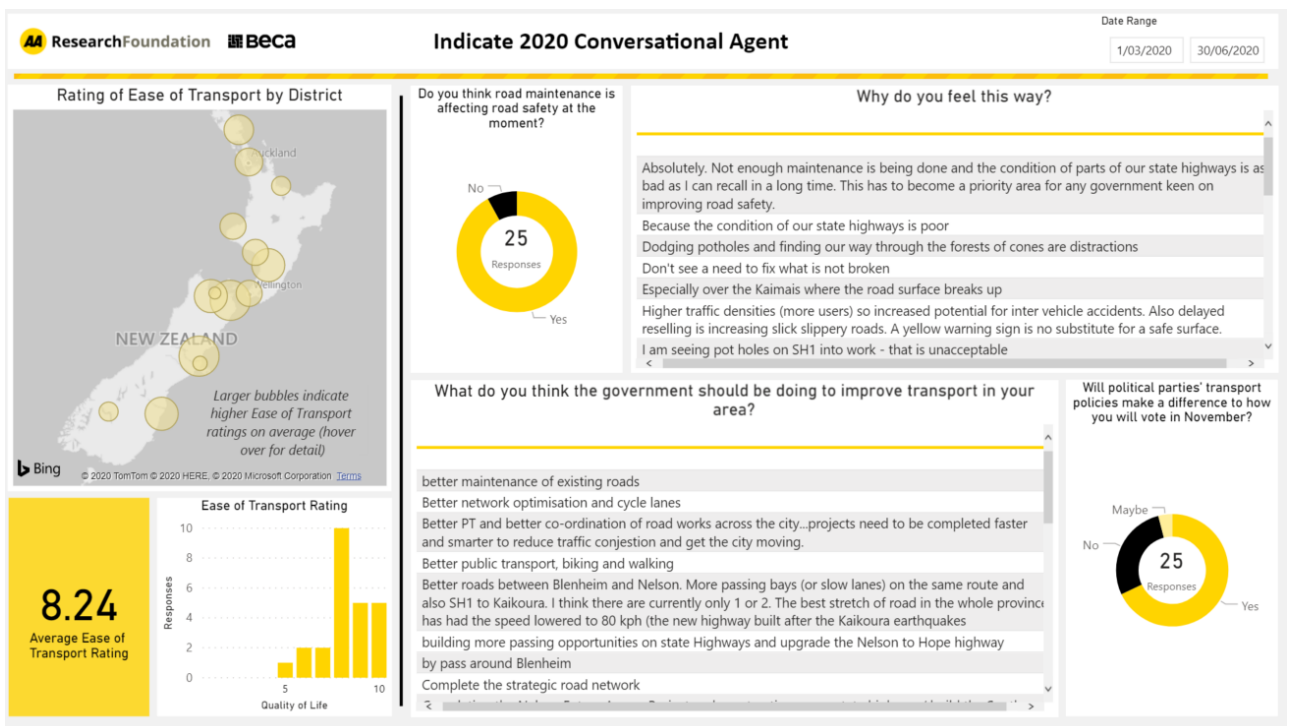


Figure 4-6 General Transport Questions page (Congestion) (from the Indicate dashboard <http://feedback.indicate.chat>, p. 6)

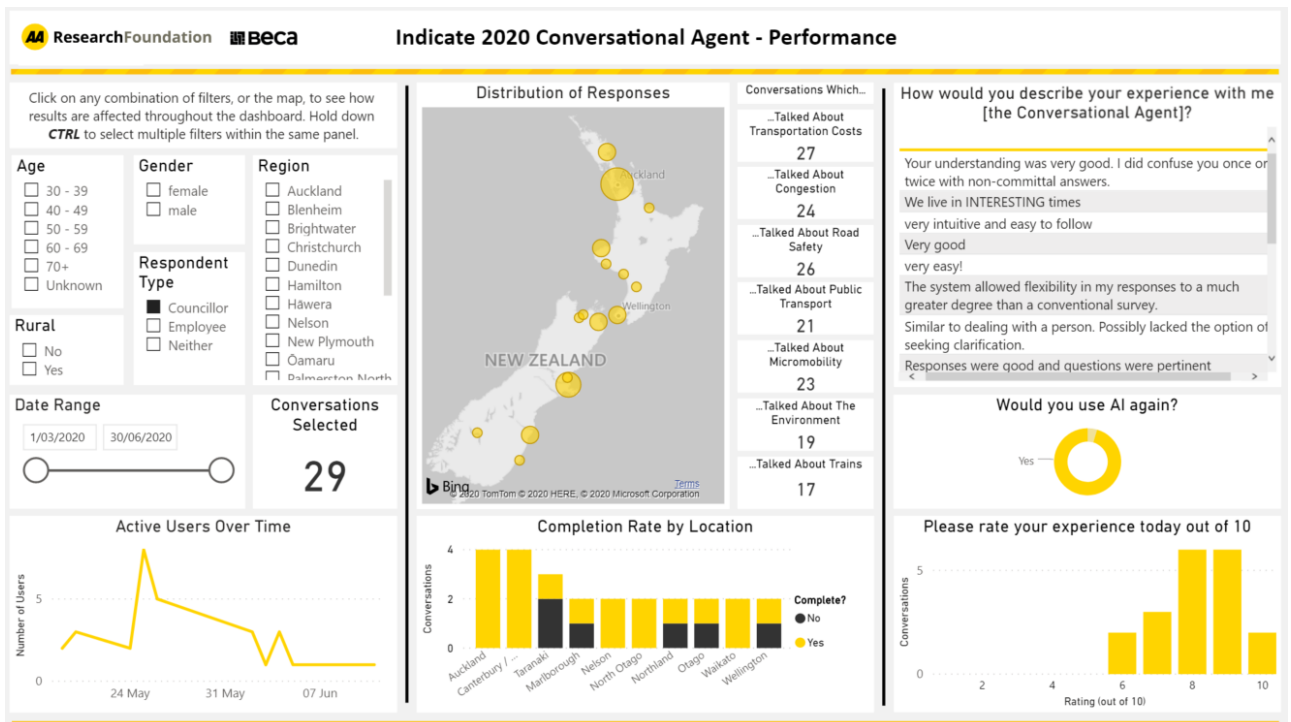


Figure 4-7 AI Experience Feedback page (from the Indicate dashboard <http://feedback.indicate.chat>, p. 7)

For the AA staff trial period, a further one page 'leader board' was also prepared, with a view to incentivising participation in the trial through gamification (competition) – charity vote, highest/lowest sentiment scores, longest conversation.

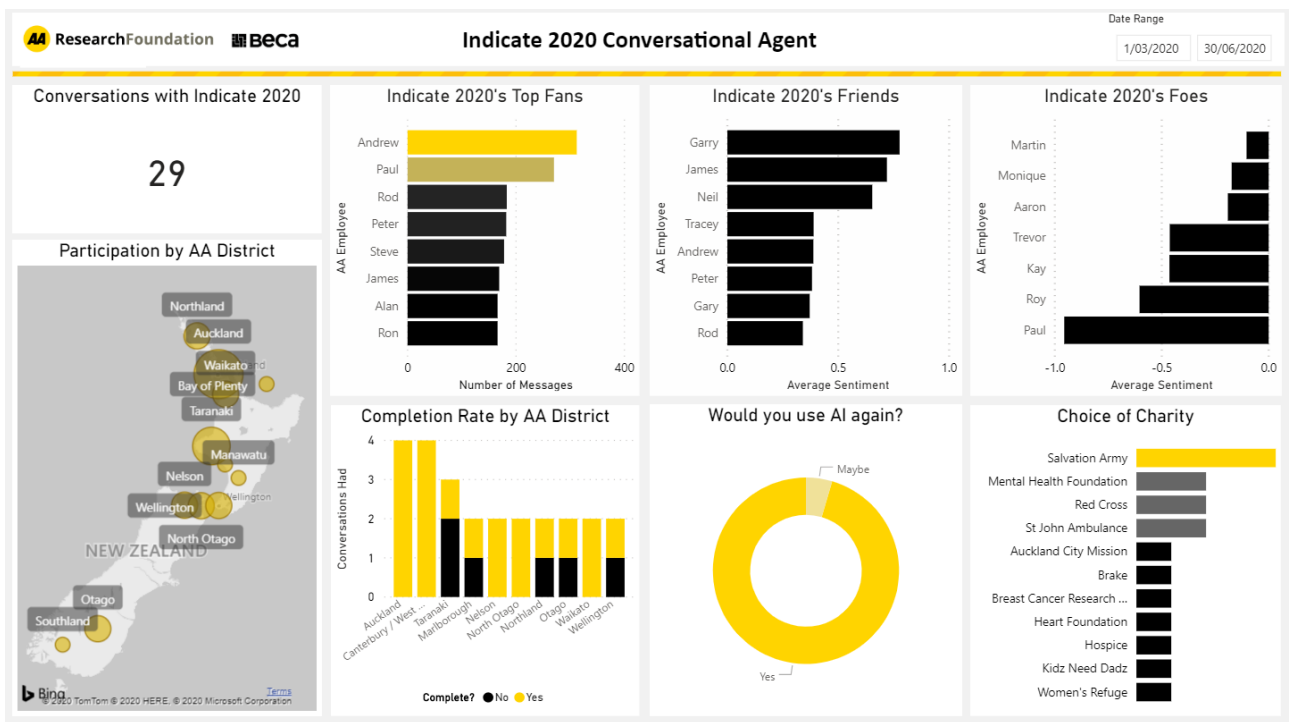


Figure 4-8 – Leader board dash (see <http://leaderboard.indicate.chat>)

5 Trial Results

5.1 Introduction

After initial testing and development was undertaken, Indicate was made available to AA staff including members of the communications team, before being trialled by AA District Councillors.

29 AA District Councillors took part in this trial; the results presented in the sections below comprise a summary of their demographics, the opinions they shared on transport issues, and their feedback on the experience of using Indicate.

5.2 District Councillor Conversations

5.2.1 Demographics / Locations

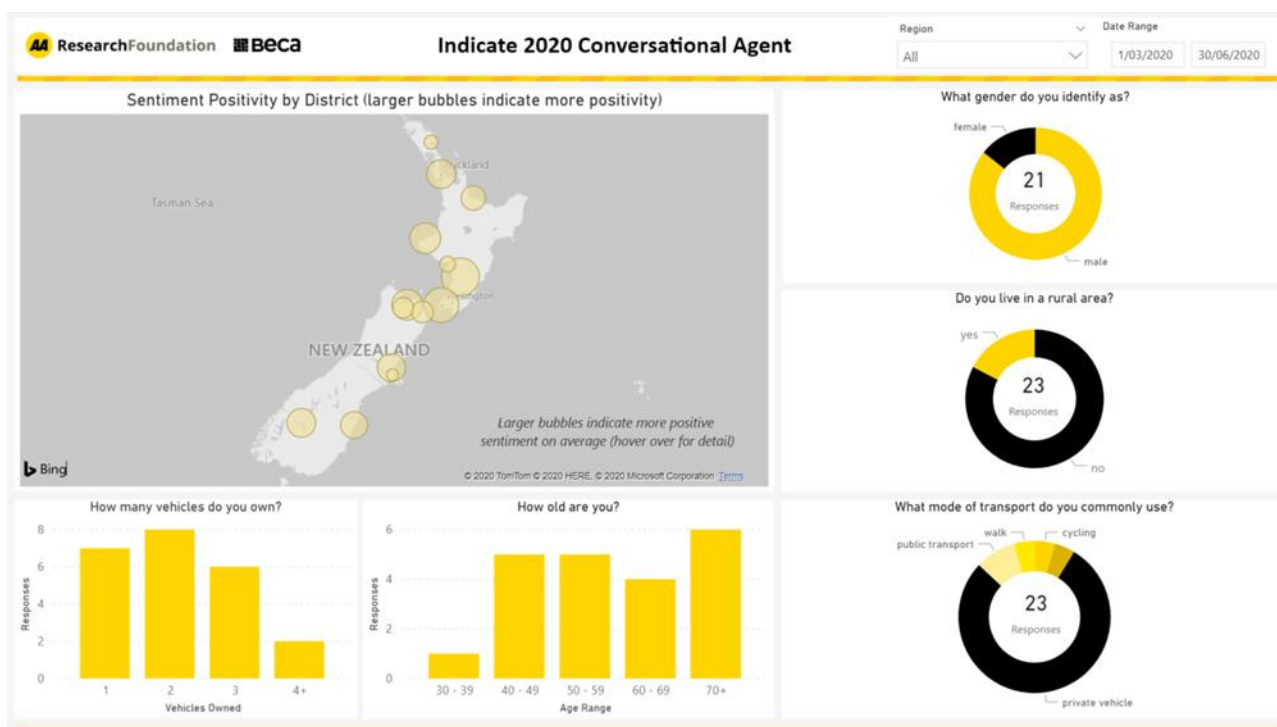


Figure 5-1 - Demographic information on respondents (from the Indicate dashboard <http://feedback.indicate.chat>, p. 2)

AA District Councillors who participated in the trial of Indicate were spread across the country, although the majority were not from rural areas. Over 80% of respondents were male, with ages spread reasonably evenly across all age groups from 40 and above. One respondent amongst the AA District Councillors was under 40. Private vehicles were the most commonly used mode of transport; with respondents owning an average of two vehicles per household.

5.2.2 General Transport-Based Questions

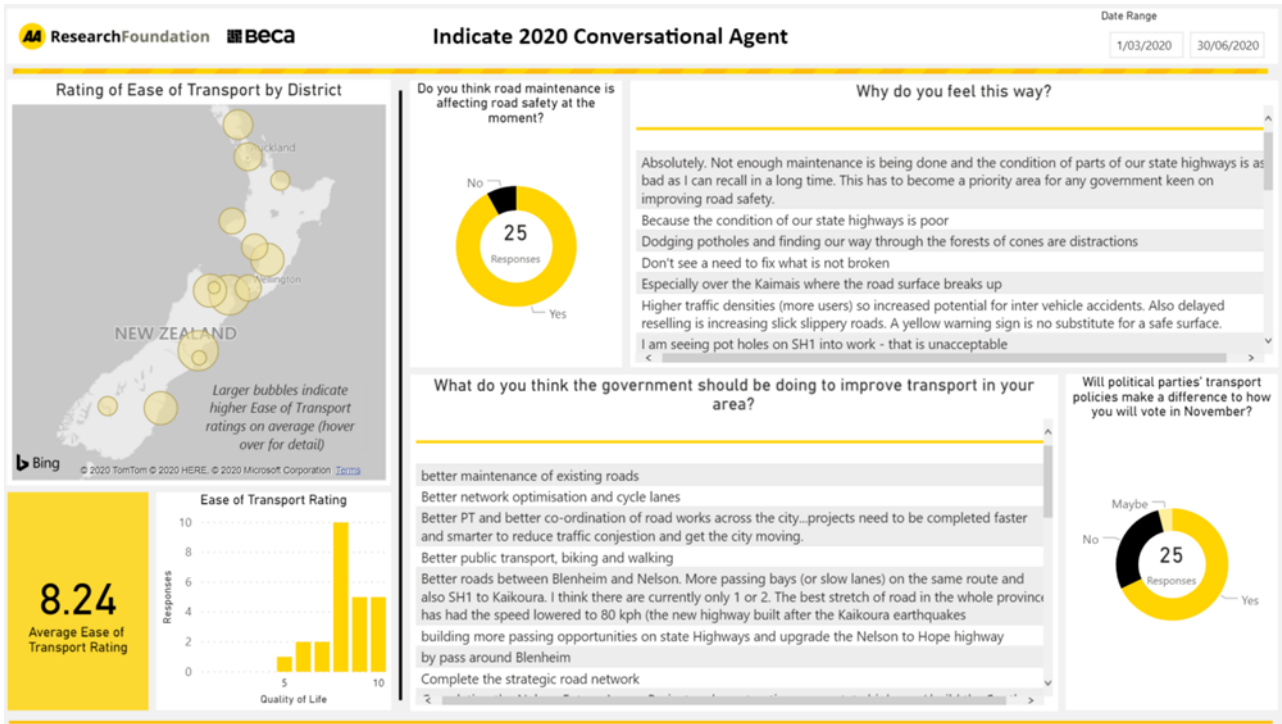


Figure 5-2 - Demographic information on respondents (from the Indicate dashboard <http://feedback.indicate.chat>, p. 6)

When asked to rate the ‘ease of transport’ in their area out of 10, all respondents chose a score of at least 5, with an average score of 8.24 overall – the highest scores came from respondents outside of the main centres.

This positive trend notwithstanding, the vast majority of AA District Councillors indicated that they believe road maintenance to be having an effect on road safety at the moment, with around two thirds of respondents suggesting that the nature of political parties’ transport policies would make a difference to how they choose to vote in the general election.

Region-specific concerns emerged as a key theme in responses to these general transport questions.

5.2.3 Issue-Specific Questions

Respondents were given the opportunity to share their opinions on each of seven core topics. A summary and selection of their responses for each topic is given below. It is possible from the dashboard to select to see all the responses verbatim.

Congestion



Figure 5-3 - Summary of responses to questions on the topic of congestion (from dashboard)

26 of 29 councillors chose to discuss the topic of congestion.

A significant cluster of negative sentiment responses indicates alignment on respondents' views on congestion in their area.

Sentiment on the issue of congestion charging was less consistent - around two thirds of respondents indicating support for the concept, with the concept of 'user pays' emerging as a key theme.

Public Transport

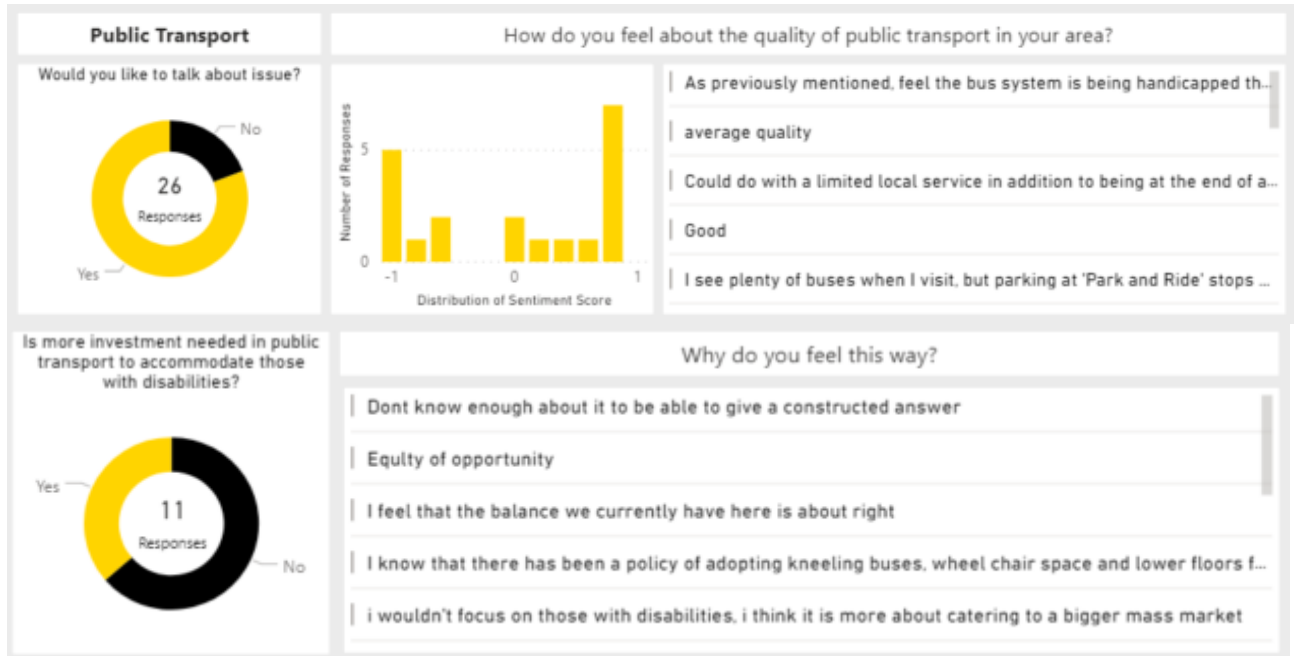


Figure 5-4 Summary of responses to questions on the topic of public transport (from dashboard)

26 out of 29 councillors chose to discuss the topic of public transport.

When asked about quality of public transport in their area, responses were reasonably split between having positive and negative sentiment, with bunching at either extreme.

A comparatively low response rate on the issue of accommodating those with disabilities suggests a reluctance to broach this topic.

Environment

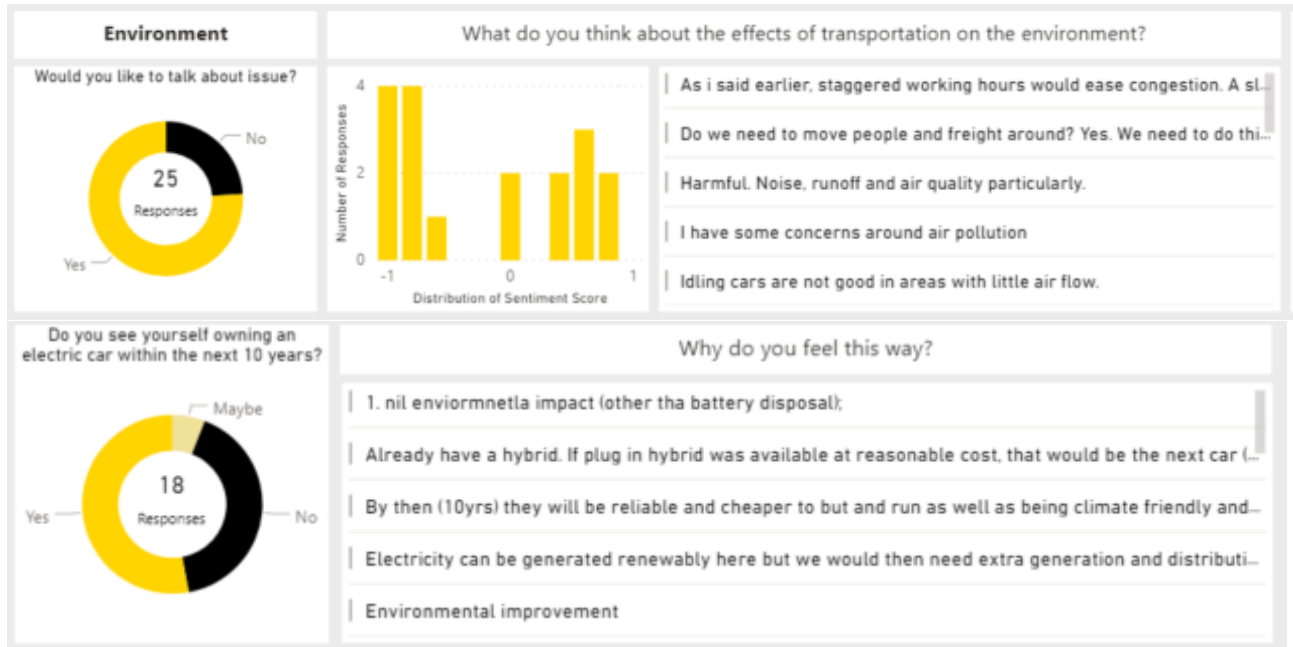


Figure 5-5 Summary of responses to questions on the topic of environment (from dashboard)

25 out of 29 councillors chose to discuss the topic of the environment.

A cluster of negative sentiment responses indicates alignment on respondents' views on the effects of transport on the environment.

Responses on the topic of likely future ownership of electric vehicles was more evenly spread across the board, with a high number of neutral responses.

Transportation Costs

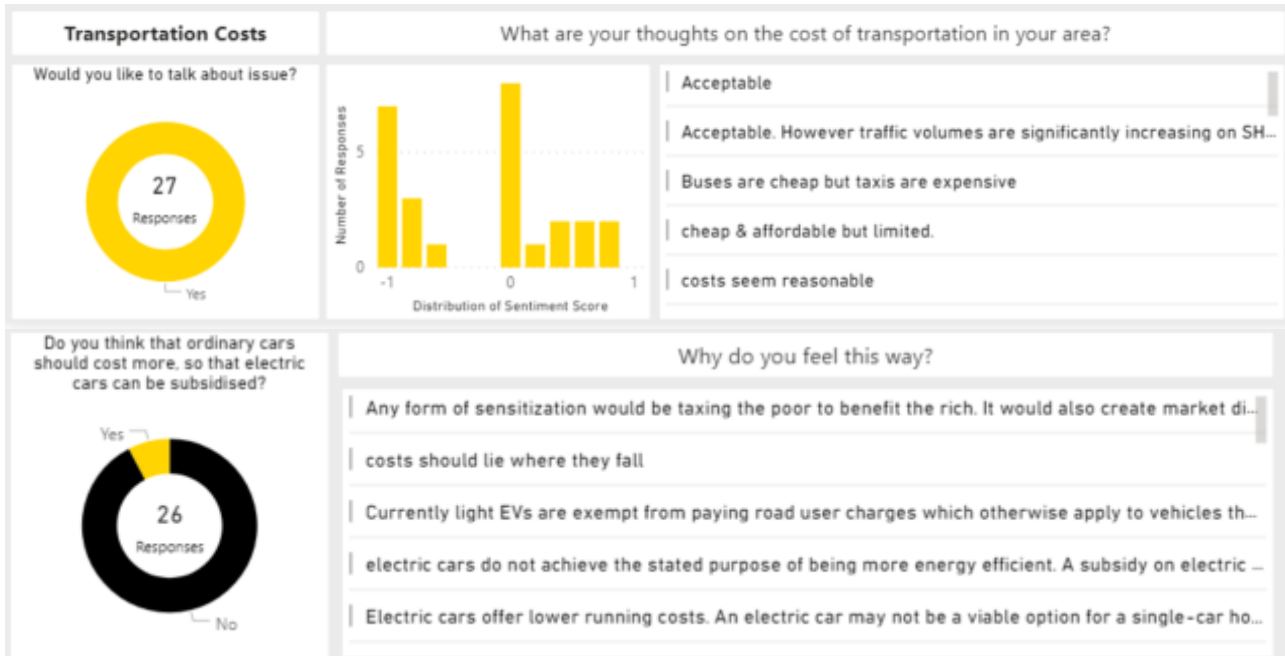


Figure 5-6 Summary of responses to questions on the topic of transportation costs (from dashboard)

27 out of 29 councillors chose to discuss the topic of transportation costs.

When asked about the cost of transportation in their area, responses were reasonably split between having positive and negative sentiment, with a high number of neutral responses.

Responses toward the possibility of increasing the cost of ordinary cars to enable electric car subsidies showed a significant trend of negative sentiment.

Road Safety

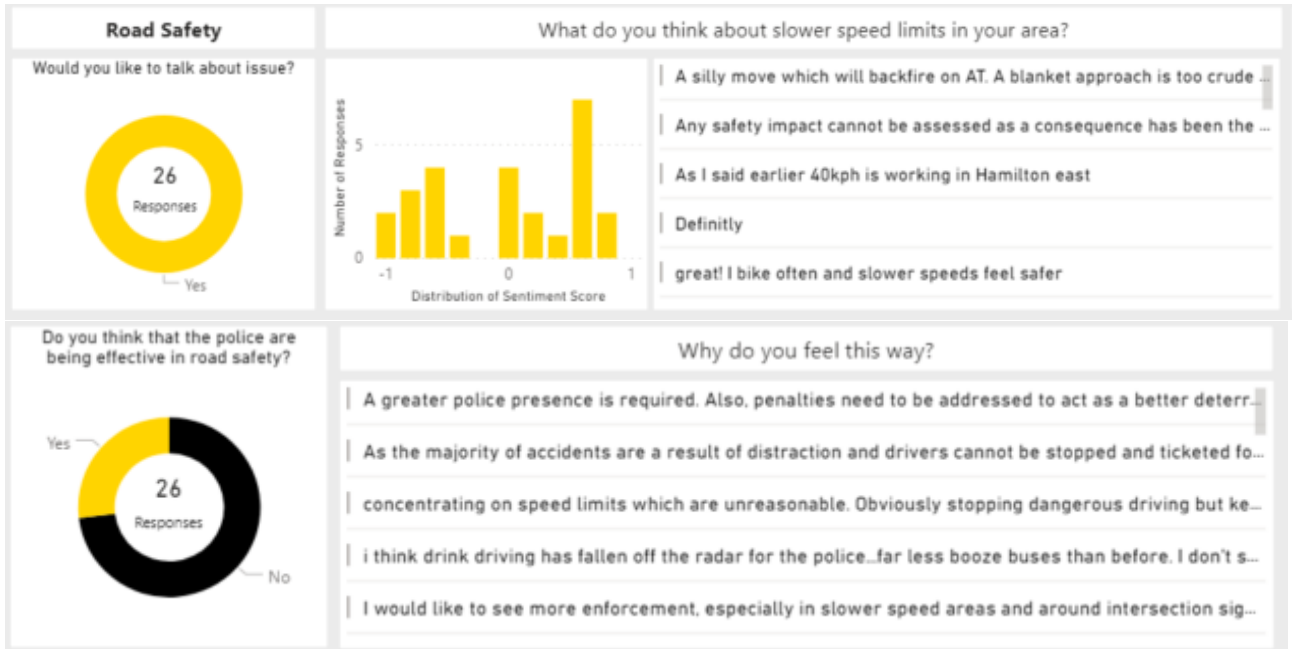


Figure 5-7 Summary of responses to questions on the topic of road safety (from dashboard)

26 out of 29 councillors chose to discuss the topic of road safety.

When asked about the possibility of slower speed limits in their area, the sentiment of responses was spread across the spectrum, with filtering indicating that region-specific experiences were a factor.

Sentiment on the issue of the effectiveness of police in road safety showed a strong negative response, with visibility, speeding, drink-driving and distracted driving all featuring as key themes across responses.

Micromobility



Figure 5-8 Summary of responses to questions on the topic of micromobility (from dashboard)

25 out of 29 councillors chose to discuss the topic of micromobility.

When asked about the possibility of increasing the number of cycle lanes in their area, responses were reasonably split between having positive and negative sentiment, with only a slight trend towards a positive view of the subject.

Sentiment toward allowing e-scooters on footpaths showed a much stronger negative trend, with safety concerns as a key theme across responses.

Trains

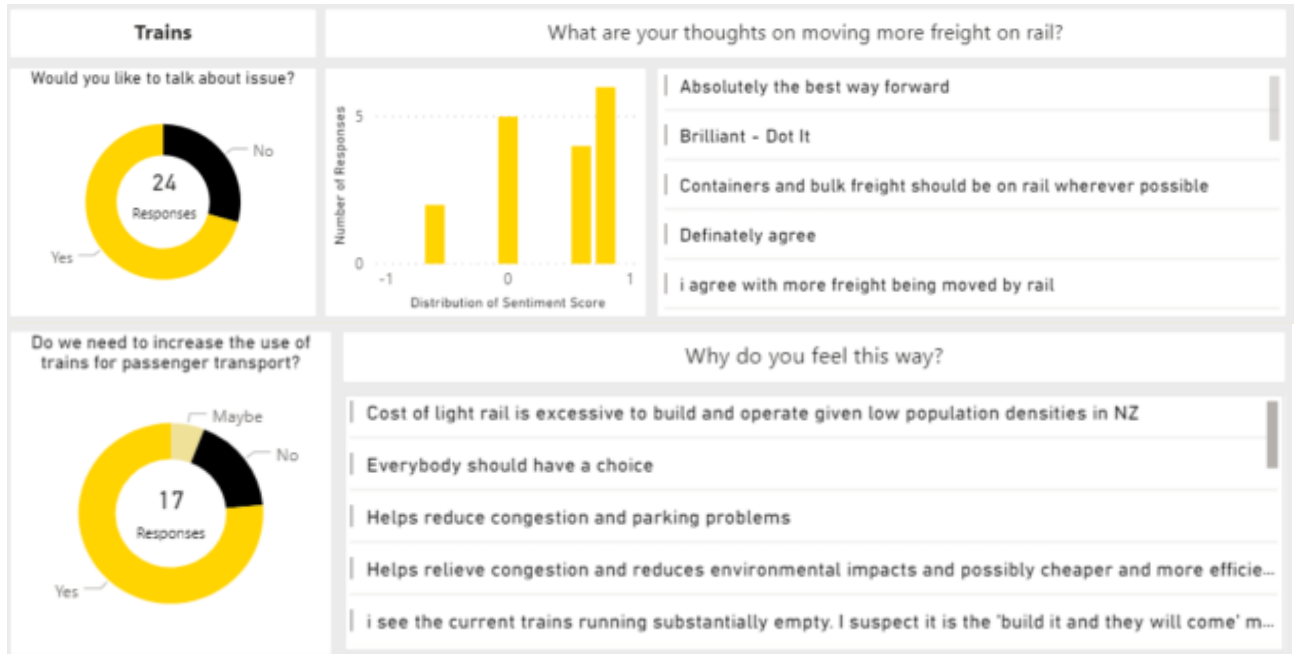


Figure 5-9 Summary of responses to questions on the topic of trains (from dashboard)

24 out of 29 councillors chose to discuss the topic of trains.

When asked about the possibility of moving more freight on rail, the majority of responses showed a positive sentiment.

However, views on the increased use of trains for passenger transport were more polarised, with high costs and reducing congestion impacts emerging as competing themes.

5.3 Feedback on using Indicate 2020

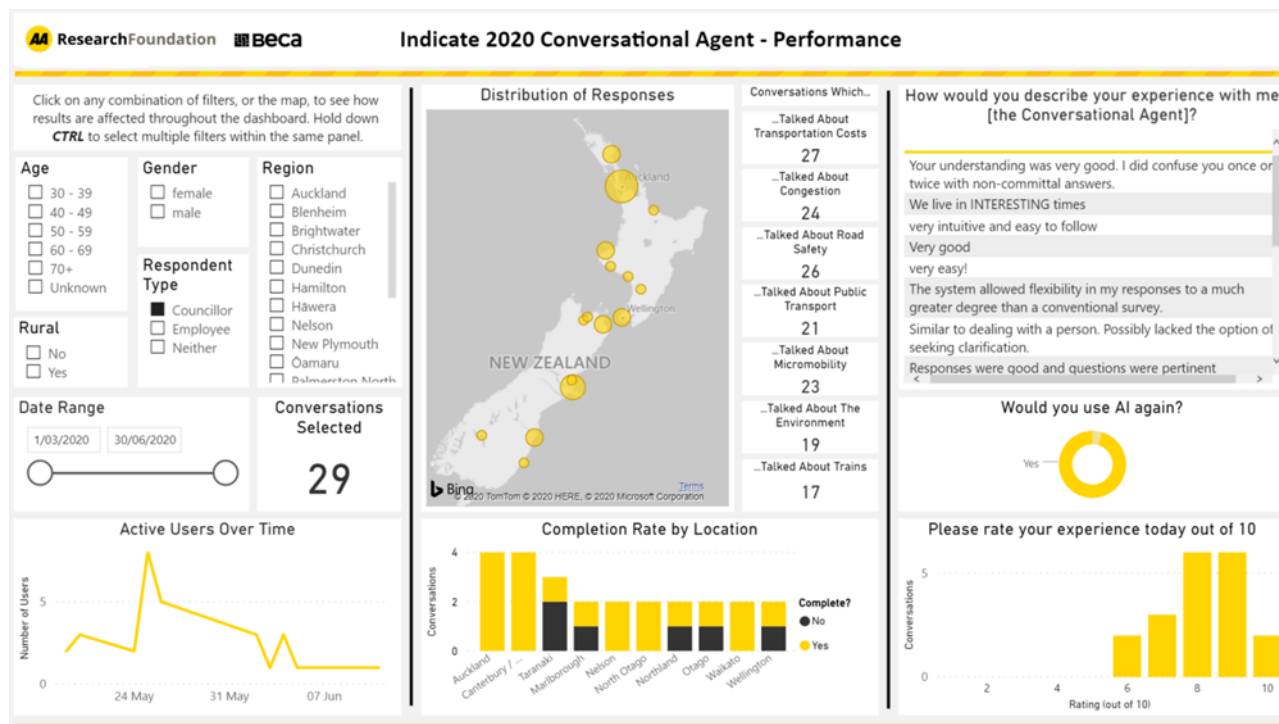


Figure 5-10 – Respondents’ engagement with, and feedback on, the conversational agent experience (from the Indicate <http://feedback.indicate.chat>, p. 7)

Alongside content-based questions, Indicate users were also given the opportunity to provide feedback on their experience with the conversational agent itself. When asked if they would use AI again, 95% of respondents indicated that yes, they would, with the remainder settling on ‘maybe’.

Notably, no respondents indicated that they would not use AI again if given the opportunity.

When asked to rate their experience with the conversational agent out of 10, all AA District Councillors who responded gave at least a 6, with the average score being 8.16.

Given the opportunity to offer free-form feedback on their experience, respondents noted that they found it to be intuitive, flexible and easy to follow, with responses that felt appropriate. Some users did note that they found the questions a little repetitive, and that they would like to have seen the conversational agent probe a little further into their initial responses, rather than just accepting them at face value. Overall, though, feedback on the experience was broadly positive, in line with the quantitative rating results above.

6 Interim Recommendations

Following the trial of Indicate with AA District Councillors, the following recommendations were made to the AARF in July 2020:

- The implementation of Indicate has demonstrated that with appropriate training it is possible to collect feedback from a cross section of people on transport issues and provide them with a highly satisfactory experience.
- A gamification / charitable module is recommended to increase the number of people who will use Indicate. If this is included, combined with email / social media contact and reminders, it is likely that a substantial proportion of a target audience may respond.
- The recommendation from this research project is that Indicate be examined for engagement with a large proportion of the Automobile Association's membership, to identify sentiment and views towards a range of transport issues, with the potential to identify areas of particular interest for the upcoming 2020 parliamentary elections.

Based on these recommendations, a rollout of Indicate to the wider AA Membership was planned, and occurred in September 2020.

7 Members Release Results

7.1 Introduction

Following the trial of Indicate by AA District Councillors, in September 2020 Indicate was released to the wider AA membership, through a link in an e-newsletter-type communication.

329 AA Members had a conversation with Indicate; the results presented in the sections below comprise a summary of their demographics, the opinions they shared on transport issues, and their feedback on the experience of using Indicate.

7.2 AA Member Conversations

7.2.1 Demographics / Locations

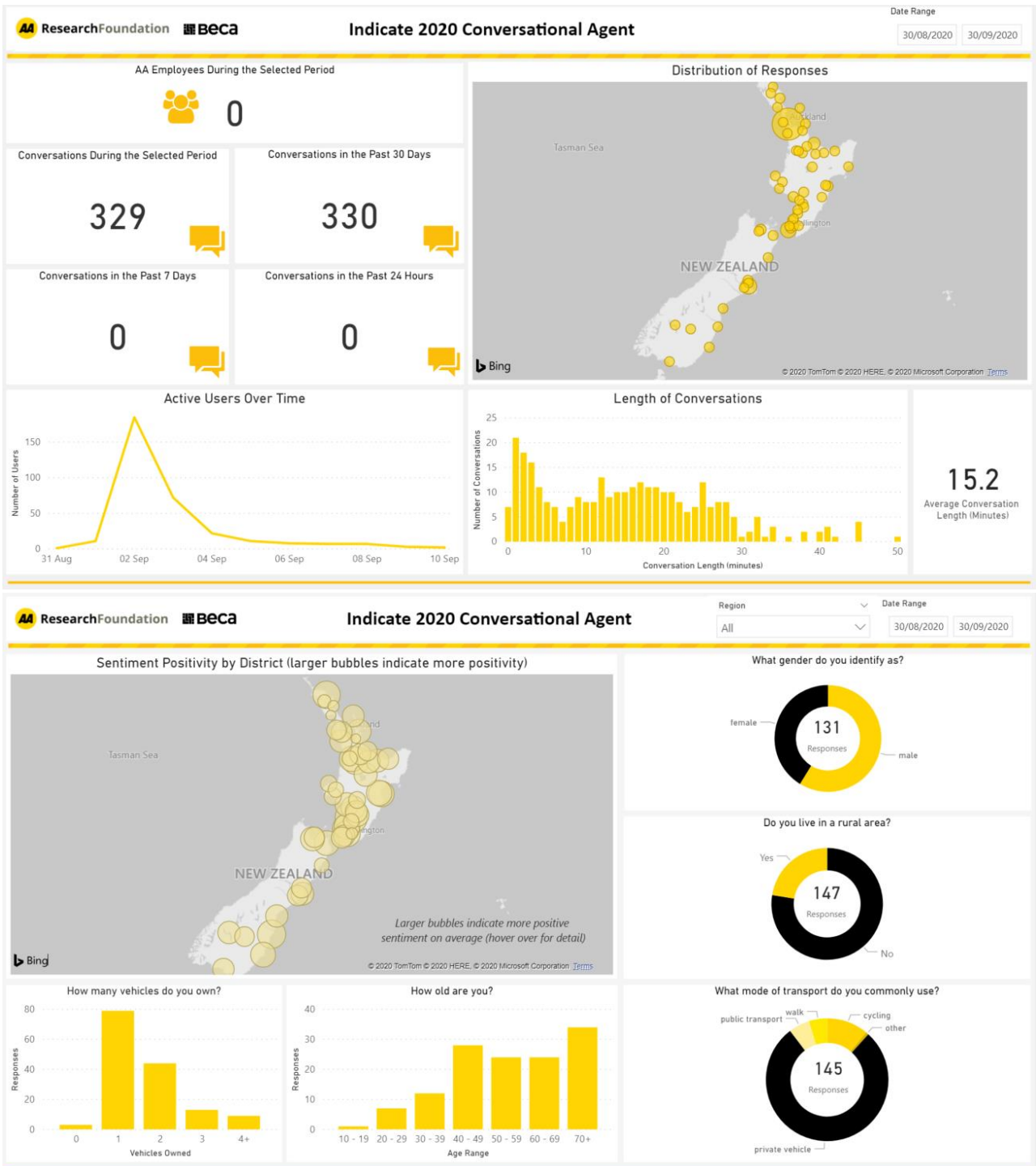


Figure 7-1 - Demographic information on respondents (from the Indicate dashboard <http://feedback.indicate.chat>, pp. 1 and 2)

The 329 AA Members who participated were spread across the country, although the majority were from urban areas. Around 60% of respondents were male, with ages spread reasonably evenly

across all age groups from 40 and above; 20 respondents identified themselves as being under 40. Private vehicles were the most commonly used mode of transport (78%); with respondents owning an average of 1.7 vehicles per household. The average conversation with Indicate lasted approximately 15 minutes.

7.2.2 General Transport-Based Questions

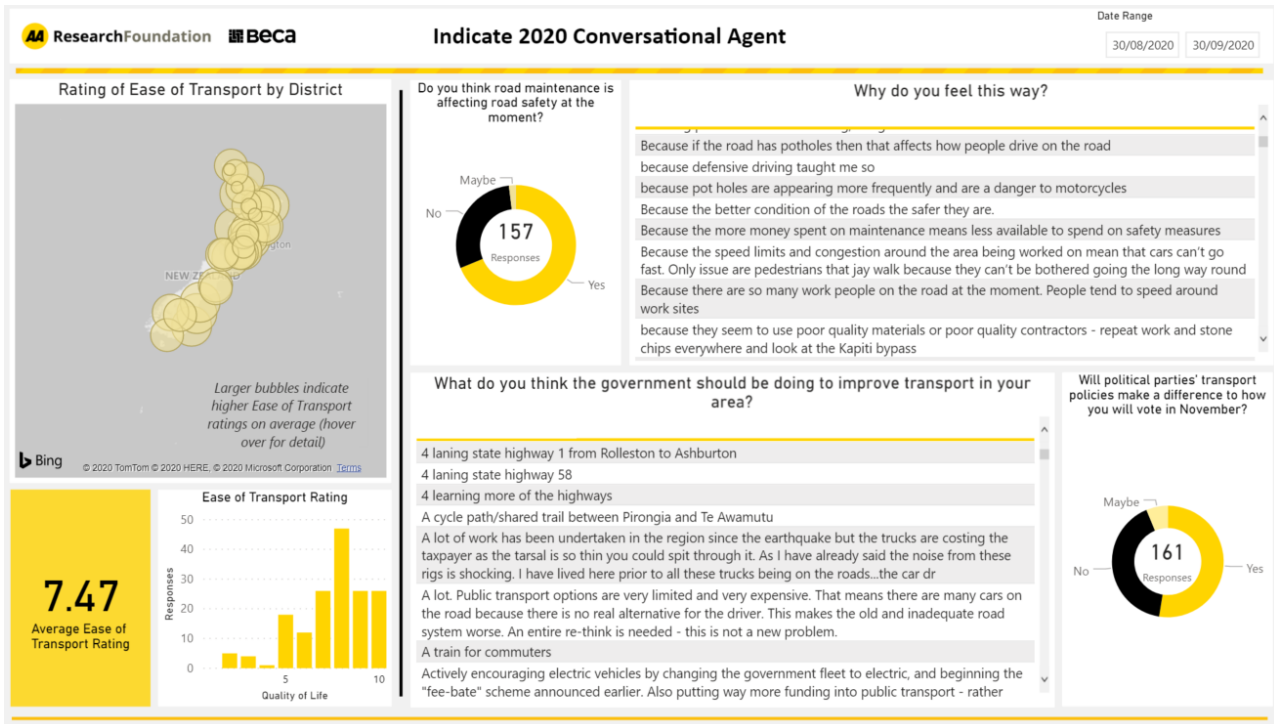


Figure 7-2 - Demographic information on respondents (from the Indicate dashboard <http://feedback.indicate.chat>, p. 6)

When asked to rate the ‘ease of transport’ in their area out of 10, the vast majority of respondents chose a score of at least 5, with an average score of 7.47 overall.

This positive trend notwithstanding, the 69% of AA Members who responded indicated that they believe road maintenance to be having an effect on road safety at the moment, with around over half of respondents suggesting that the nature of political parties’ transport policies would make a difference to how they choose to vote in the general election.

Region-specific concerns emerged as a key theme in responses to these general transport questions, with public transport, cycle facilities, and the completion of major roading projects all appearing as common touchpoints.

It is possible from the dashboard to view all verbatim responses to these open-ended general questions.

7.2.3 Issue-Specific Questions

Respondents were given the opportunity to share their opinions on each of seven core topics. A summary and selection of their responses for each topic is given below. It is possible from the dashboard to view all of the responses verbatim.

Congestion

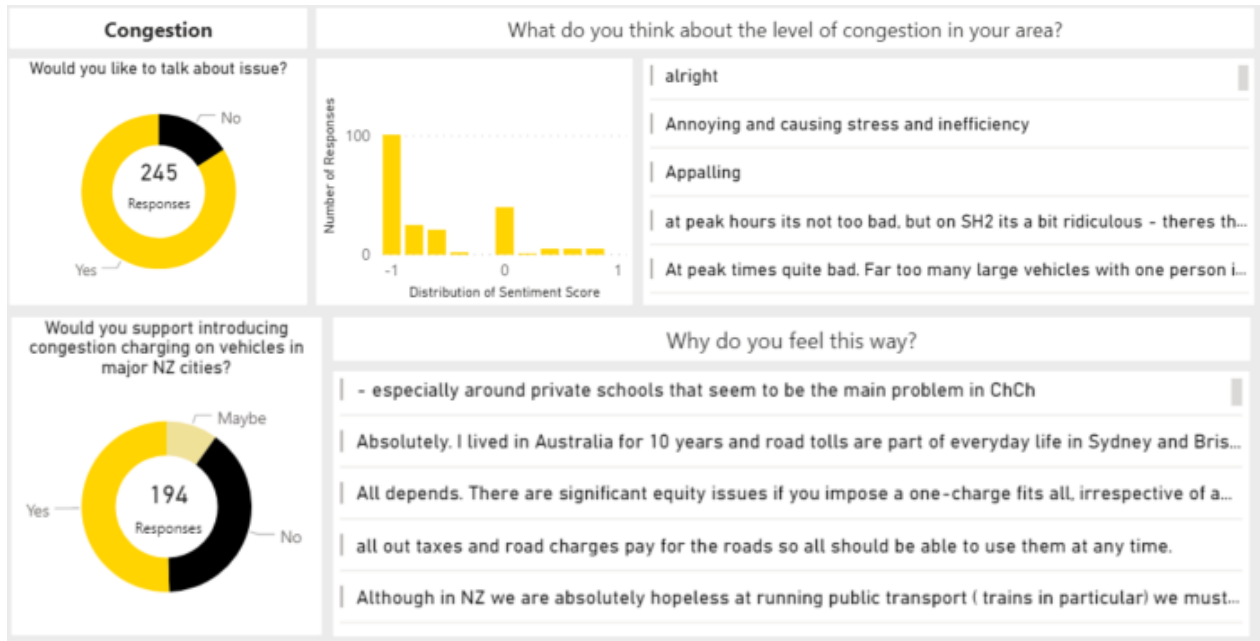


Figure 7-3 - Summary of responses to questions on the topic of congestion (from dashboard)

206 of 329 respondents chose to discuss the topic of congestion.

A significant cluster of negative sentiment responses indicates alignment on respondents' views on congestion in their area.

Sentiment on the issue of congestion charging was less consistent - around half of respondents indicated support for the concept, with the idea of 'user pays' emerging as a key theme.

Public Transport

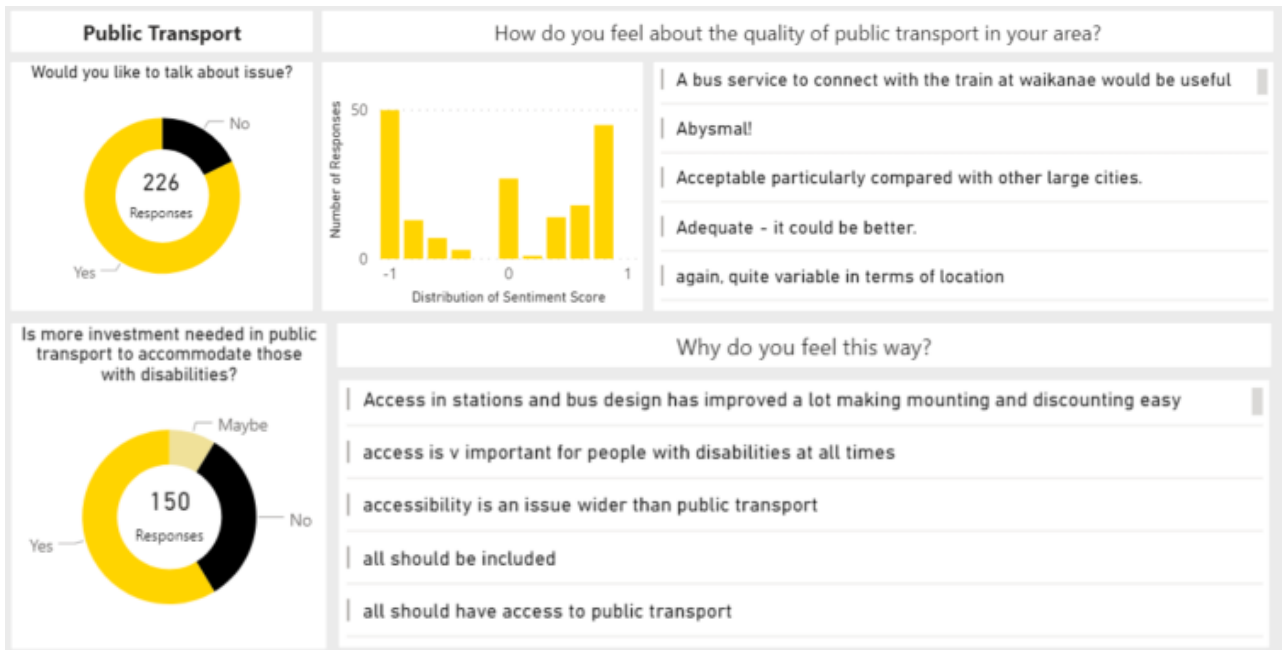


Figure 7-4 Summary of responses to questions on the topic of public transport (from dashboard)

186 of 329 respondents chose to discuss the topic of public transport.

When asked about quality of public transport in their area, responses were reasonably split between having positive and negative sentiment, with bunching at either extreme.

The vast majority of respondents noted the importance of accessibility for public transport; opinions varied on whether public transport is sufficiently accessible/accommodating already.

Environment

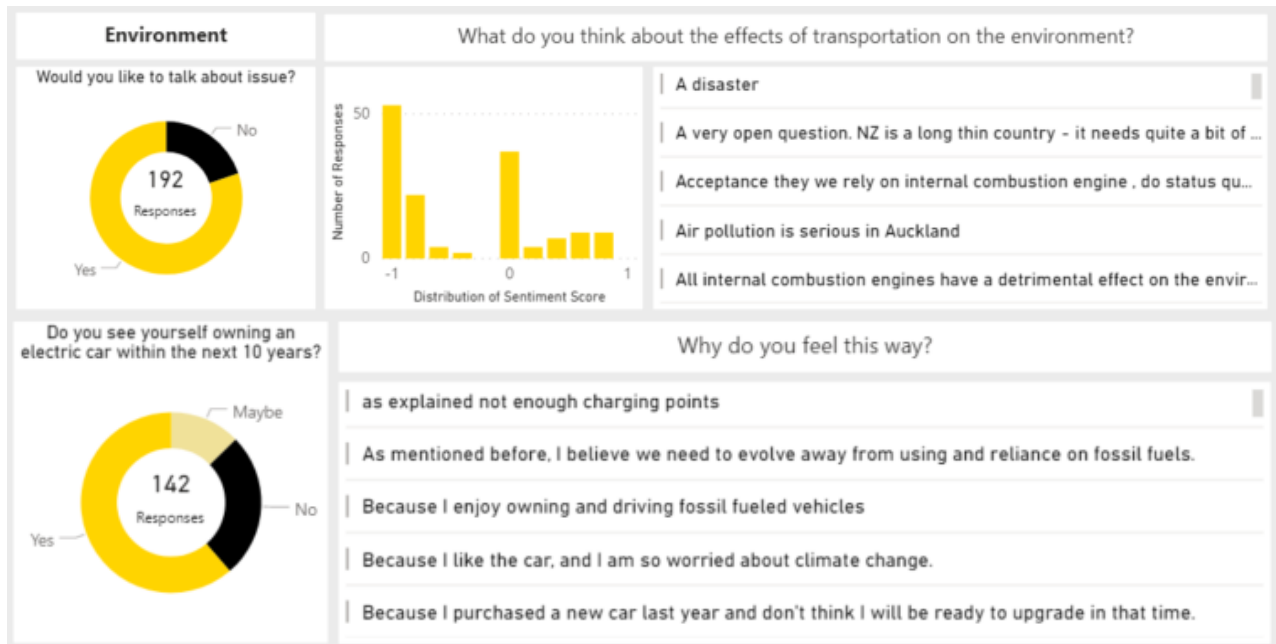


Figure 7-5 Summary of responses to questions on the topic of environment (from dashboard)

154 of 329 respondents chose to discuss the topic of the environment.

A cluster of negative sentiment responses indicates alignment on respondents' views on the effects of transport on the environment.

Responses on the topic of likely future ownership of electric vehicles were more evenly spread, with a comparatively high number of neutral responses.

Transportation Costs

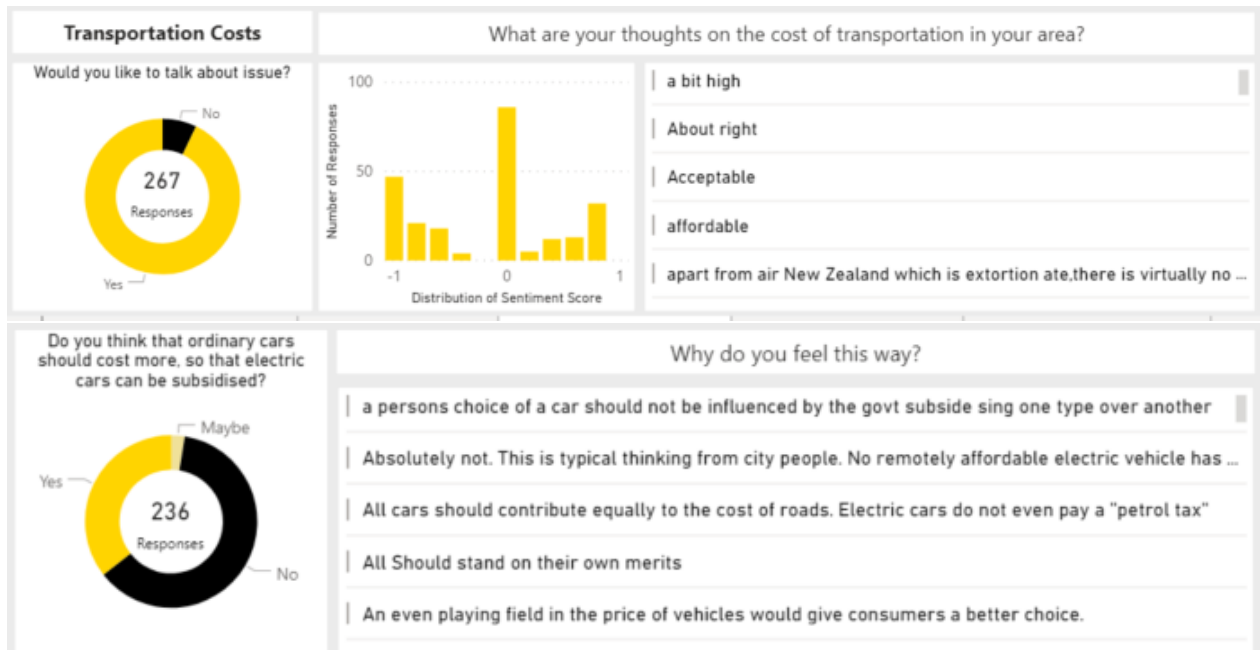


Figure 7-6 Summary of responses to questions on the topic of transportation costs (from dashboard)

248 of 329 respondents chose to discuss the topic of transportation costs.

When asked about the cost of transportation in their area, responses were reasonably split between having positive and negative sentiment, with a high number of neutral responses.

Responses toward the possibility of increasing the cost of ordinary cars to enable electric car subsidies showed a significant trend of negative sentiment.

Road Safety

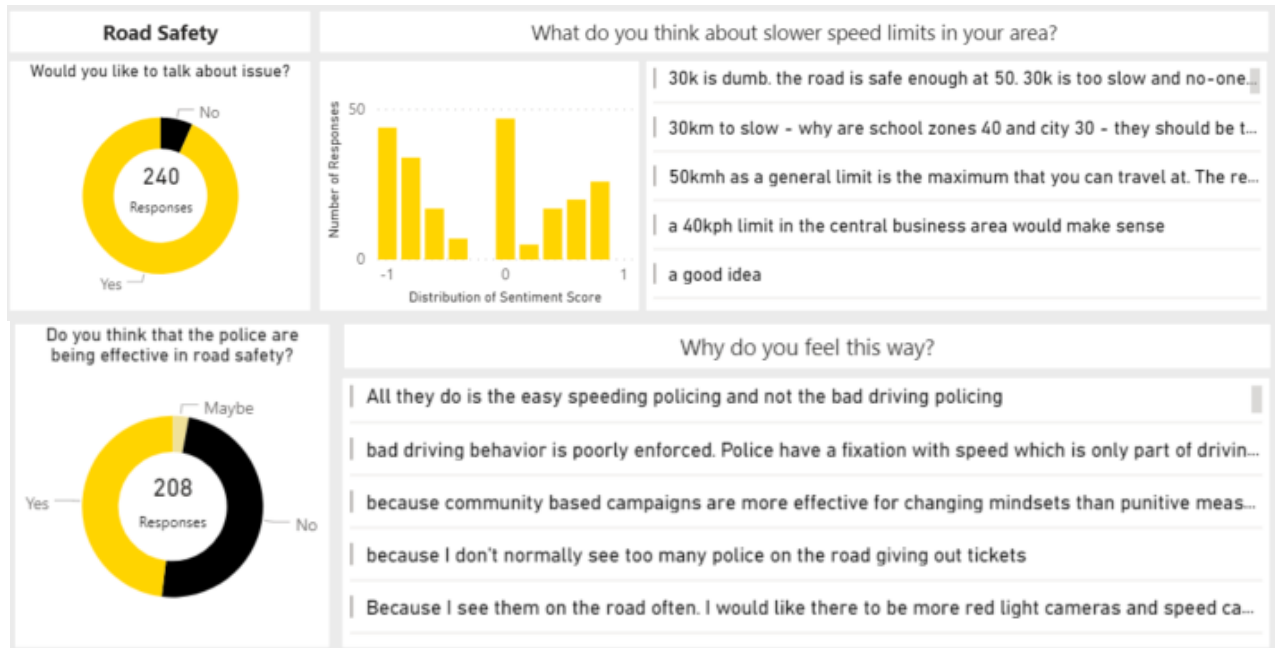


Figure 7-7 Summary of responses to questions on the topic of road safety (from dashboard)

224 of 329 respondents chose to discuss the topic of road safety.

When asked about the possibility of slower speed limits in their area, the sentiment of responses was spread across the spectrum, with filtering indicating that region-specific experiences were a factor.

Likewise, views on the issue of the effectiveness of police in road safety were polarised, with visibility, speeding, drink-driving and distracted driving all featuring as key themes across responses.

Micromobility

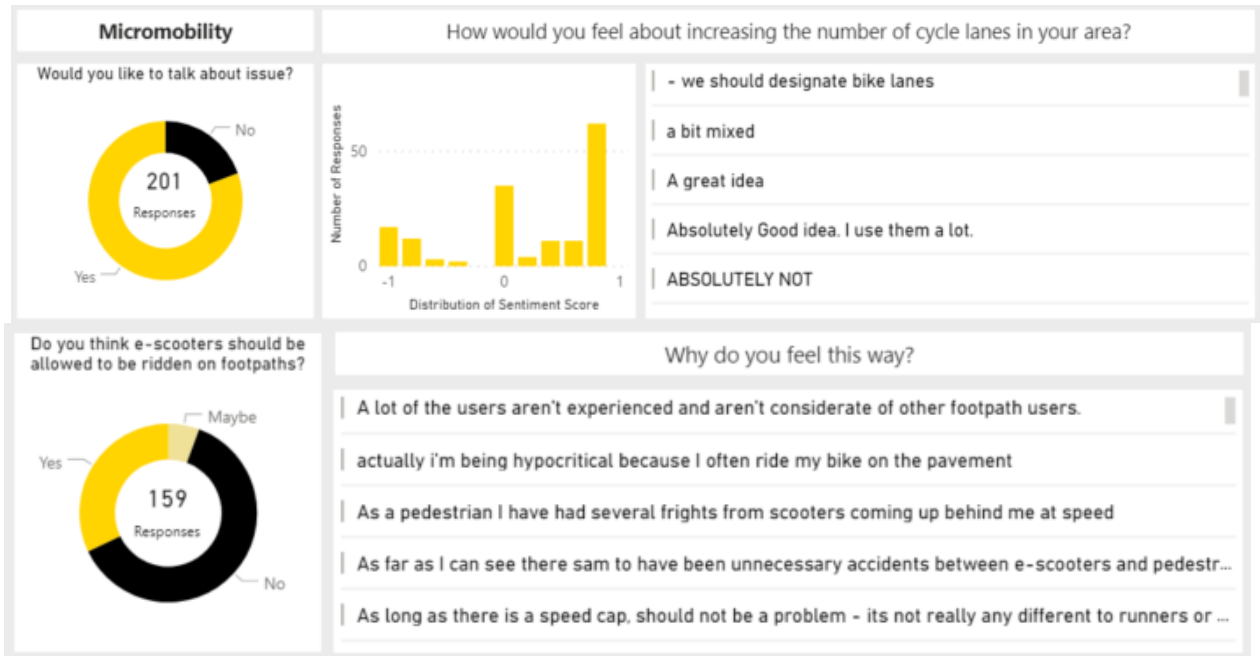


Figure 7-8 Summary of responses to questions on the topic of micromobility (from dashboard)

162 of 329 respondents chose to discuss the topic of micromobility. Y

When asked about the possibility of increasing the number of cycle lanes in their area, responses were diverse: most respondents indicated some degree of support for the idea but, where opposition occurred, it was with strongly negative sentiment.

Sentiment toward allowing e-scooters on footpaths showed a much broader negative trend, with safety concerns as a key theme across responses.

Trains

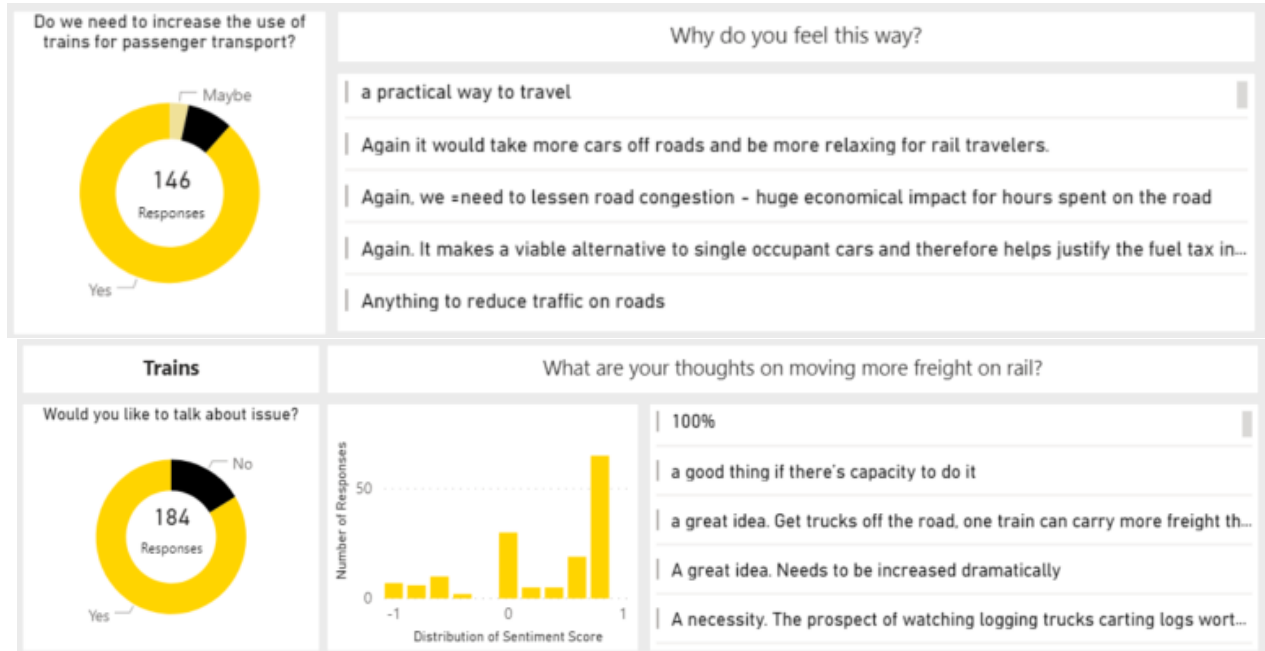


Figure 7-9 Summary of responses to questions on the topic of trains (from dashboard)

154 of 329 respondents chose to discuss the topic of trains.

When asked about the possibility of moving more freight on rail, the majority of responses showed a positive sentiment.

Views on the increased use of trains for passenger transport were similarly positively skewed, with reducing congestion emerging as a key theme.

7.3 Feedback on using Indicate 2020

Feedback on Indicate was collected at the midpoint and then at the end of the conversation. Users could opt out from providing feedback.

Key elements of the feedback are:

- Conversations on Indicate lasted on average more than 15 minutes.
- Almost all users would use the AI system again
- Satisfaction rates were highest amongst the 70+ age group, lowest amongst the 40 to 49-year age group, and slightly higher amongst rural as compared to urban users.

There were no identifiable differences in satisfaction scores by gender or location.

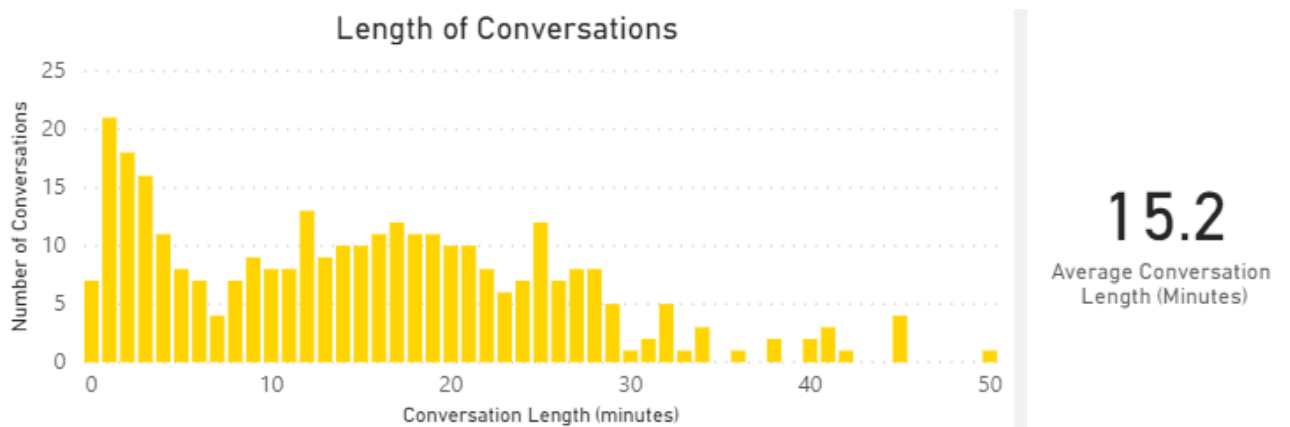
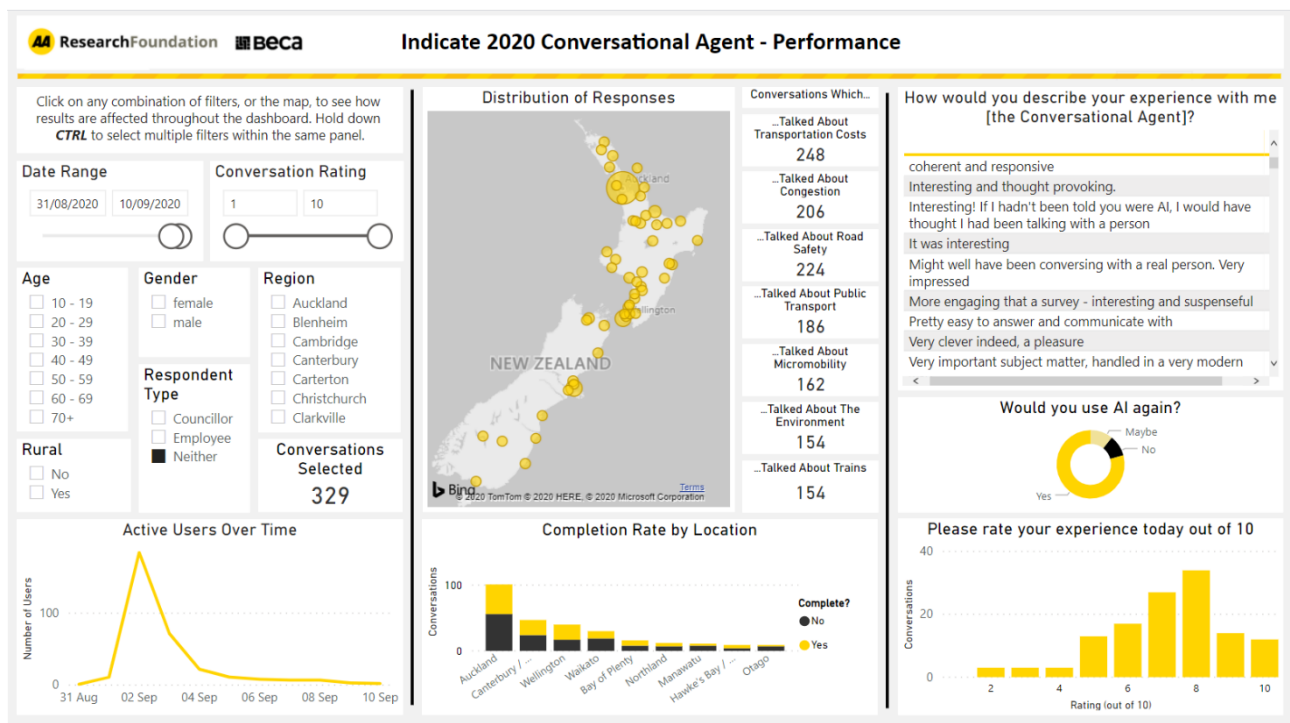


Figure 7-10 – Respondents’ engagement with, and feedback on, the conversational agent experience (from the Indicate <http://feedback.indicate.chat>, p. 7 and conversation length from p. 1)

When asked if they would use AI again, 90% of respondents indicated that they would consider doing so.

When asked to rate their experience with the conversational agent out of 10, the vast majority of AA members who responded gave at least a 5, with the average score being 7.08. Free-form feedback on the experience was broadly positive, in line with the quantitative rating results above.

Typical areas of dissatisfaction were:

- Indicate was not enabled to answer questions / provide information on issues
- Indicate asked for yes / no answers for complex questions.

The second dissatisfaction was due to the need to understand whether a person supported or otherwise a particular policy or solution. Indicate can be easily re-scripted to ask about support, rather than force a yes / no response.

Full verbatim feedback is shown in the table below.

Table 7.1 Member feedback on the Conversational Agent experience.

How would you describe your experience with [the Conversational Agent]?
coherent and responsive
Interesting and thought provoking.
Interesting! If I hadn't been told you were AI, I would have thought I had been talking with a person
It was interesting
Might well have been conversing with a real person. Very impressed
More engaging than a survey - interesting and suspenseful
Pretty easy to answer and communicate with
Very clever indeed, a pleasure
Very important subject matter, handled in a very modern way.
A bit long and I realise because need for info for plan formulation to government for infrastructure projects!
Fine. Questions fair and easy enough to understand. Time is about right. any longer would not be so good.
Good
Hiccup. Good-ish
I THOROUGHLY enjoyed it. Thanx
Indicate is the most responsive bot I've ever tried to chat to. I did get asked the same questions more than once, which was frustrating, but aside from that everything was good!
interesting and engaging
It has been fast and efficient, however I would have liked to have one answer just slightly longer than allowed
it was fun and straightforward.
practical, quick,
pretty good
really awesome, easy to give feedback and engage with. Some of the wording could have been better and the chat UI is a bit boring but otherwise all good
Very easy.
Better than filling in the usual survey

Conversational model is easy though I find some of the positive reinforcement a bit contrived and insincere

cool

Different but very easy

everything was great, there was one answer to a question which you kept trying to skip which got annoying - you're welcome to just accept my answer lol

fine

Generally excellent, but asking for a yes/no answer to a complex issue doesn't make sense.

Good

Good except my email address is .com and you wouldn't accept it.

Good to try a new thing that felt OK

good, at the beginning i entered a word which you took as a command to return to the beginning - apart from that it was very easy

good, need to recognise some slang a bit more such as "sure" or ":"ok" as affirmative answers

Good. Really easy. Makes a survey better.

Good. Sometimes questions repeated because you did not fully comprehend my answer. But it still gave me more flexibility than Yes/No surveys.

hopefully a good way of you getting lots of data

I think there is a big risk of superficiality in this approach. Nuanced answers are not welcome clearly.

Interesting

Interesting and for most part on point. A couple of my responses seemed to baffle you a bit

Interesting and impressed.

Interesting task. But how are your owners going to digest the answers?

interesting!

Interesting, engaging, semi-useful

It has been an interesting exercise. You responded better than I was expecting and, in many cases, as good as a "real" person would have done.

It was easier and more interactive than a traditional survey.

It was interesting with a range of topics to talk about. It would have been nicer to have had more space as sometimes I had more to say

mostly good but 1 or 2 times response didn't seem optimal

Not as intuitive as I'd hoped. A couple of pretty standard follow up questions and the pattern of questions on a topic became obvious after the first few

Pretty good, I hope you can learn a few more tricks overtime so that you can provide me with more dynamic responses

Quite good but occasionally had trouble understanding some of my replies.

Really interesting, enjoyable

Repetitive, but smooth

Straightforward to answer. Just need the opportunities to add further comment on the topics e.g. support rail for freight but it needs to be on a competitive basis.

This programme was not very good and dealing with non- yes no answers. Most questions asked were of this type

very easy to do

You didn't quite get my meaning when answering a couple of questions. At the beginning you asked a question but didn't wait for my response before moving ahead. A digital human would have been a better experience than a chatbot.

A little bit weird, being in a virtual conversation with an entity I don't know. Also, feeling a bit untapped not knowing what is involved & time required.

better than others i've used

fine

Glitchy

Good

Good but would be helpful to know how long the conversation will be or how many topics because don't feel in control of the conversation. Unlike a survey when you can see a measure of your progress through it.

Good, but the first responses are a little too repetitive. More variation.

Good. I wish I could type new paragraphs in answers. And you didn't understand oks.

I enjoyed it although you clearly can't cope with complex answers

Interactive

Interesting

It was interesting

its ok - a bit frustrating not being able to nuance answers

Nice and fast. Better than rote questions. Qualitative!

Ok

ok but very generalist

overall good. response time was good. a couple of replies I gave to answers weren't acknowledged by the system.

Quick and sensible

Simple logical flow

some Obvious generic responses But that is to be expected but overall does the job for a survey

The AI bot often couldn't handle more complex or nuanced answers. It wanted you to break your answer down to yes or no where often the answer is yes and no.

The experience depends on the questions not your replies. Most were okay but some are really stupid.

thought provoking

You don't handle 'grey' or 'in between' answers well. If I say e.g. "not very much" you ask me to redefine as "Yes" or "No", when it isn't that binary

your answers are patronizing

A bit repetitive

AI has lots of rooms for improvement

generally good but some of the questions I didn't understand. perhaps you should provide an option to offer to ask the question in a different way.

I needed guidance on answering questions. e.g. was it ok to put several topics in one reply. Or should have I stuck to single topic answers.

I touched on it above. 'I found this conversation quite stilted and annoying. Would have been simpler to fill in a form.'

Interesting

Interesting.

it was alright but you had a few standoff-ish responses

it was mostly just questions

like a choose your own adventure story. Acceptable but quite limited in scope.

Not all questions work with yes/no.

Not to different than a fixed survey but easier to answer with more detail

Not too bad for a machine!

Ok, but frustrating when I've already answered a question to later be asked the same question again.

Also frustrating in relation to not accepting answers unless they are in a specific format

Speedy responses.clear dialogue

You didn't really understand that many of my answers. Also some questions can't be answered properly with a strict 'yes' or 'no'.

You didn't understand one of the answers I gave - other than that okay

Conversation was fine. Technically really bad experience on my Samsung galaxy. The keyboard pops up over the chat so I can't see what I'm typing and I have to minimize my keyboard every single time I want to see the bits response

Fairly boring

I have just answered your question

I'd like to have been able to share all my thoughts on road safety strategy, and not specifically about where I live

interesting and not too slow, better than I thought it would be

It feels like I'm doing the heavy lifting. I could have filled out a written questionnaire faster.

Ok

ok - some of the responses were idiotic since they came from computer generation - these include I'll keep that in mind, I'll note that down. Why give responses as if it a human? No need

Rangitikei is a province in NZ - you only appear to recognise cities (are you a BECA java product) look out beyond the bombs there are BECAs out there everywhere !

Repetitive, not as enjoyable as other AI bots I've talked with, unable to parse full sentences, required one-word answers = less of a pleasant experience

Too 'robotic' and impersonal

Frustrating

I thought the questions were superficial and the AI system didn't pick up on the clues I gave you to respond more intelligently

not impressed with your understanding of what I type

You are very polite, but you are limited in understanding the possible answers your respondents might give you, such as not recognising where people live, or when they don't understand your question

Did not comprehend answers. Trapped me into yes an no answers that did not fully cover the situation. Not intuitive. If this is the future of AI we are doomed.

First half fine, it went downhill from there. Happy for anyone in your team to ring me and discuss 0272487289

Like txtng answers to a survey

poor

Slightly frustrating

doesnt understand basic phrases - not human enough to be conversational, so if your one going to accept certain answers, just give me a form it will be a lot faster for me

largely a waste of time.You are not very smart

This AI seems to have serious flaws in its capability to processing written conversational English, and a complete inability to extract the desired bits of information from a larger context. Frankly, you might as well have used a static survey. It would have been less of a pain.

8 Conclusions on the use of Artificial Intelligence

Potential use cases for a tool like Indicate include:

- Carrying out periodic ‘annual survey’-type activities, and comparing between years;
- Carrying out one-off ‘survey’-type activities, e.g. to gather views on a specific transport project;
- Gathering information/views on general topics on an ongoing (rolling) basis, and identifying any trends in responses over time;
- Gathering information/views on specific issues where a fast turnaround time is required e.g. on responses to the Covid-19 pandemic;
- Asking further questions of specific demographics only, e.g. members under 25, or members whose primary mode of transport is cycling;
- Asking further questions of those who express a specific view on an issue only, e.g. those who express significant concern about road safety in their area.

The key difference between traditional online survey methods and Indicate, was that Indicate has the ability to accept and process answers to open-ended questions and display them in a secure interactive dashboard.

The depth and breadth of opinions collected from AA members who had a conversation with Indicate demonstrates the potential value of the tool for understanding attitudes to, and experiences with, transport.

The member participation rate was similar to that expected for a standard survey sent out through the same channels. This suggests that using a novel method of engagement did not, in and of itself, have any significant impact on the number of people who chose to participate.

A potential barrier for getting people to use Indicate is from their personal poor experiences with ‘chatbots’. Two features were introduced to act as incentives to mitigate this possible reluctance. Firstly, Indicate was adapted to be able to collect votes on people’s preferred charity to receive a donation (i.e. the most-voted charity would receive a donation at the end of the ‘survey’ process). Secondly, a leader-board was created for AA staff, allowing participants to see who had the longest, most positive or most negative conversations with Indicate (<http://leaderboard.indicate.chat>), therein incentivising participation through gamification. A leader-board was also created for Councillors (but not shared).

Nonetheless, following a conversation with Indicate, over 90% of respondents indicated they would consider using AI again, which supports continuing to spend time investigating how Indicate and other new technology could be integrated into member organisations.

Demographic statistics indicate that the profile of members who participated was diverse, covering a range of age groups and locations, with a good balance of gender. Satisfaction was highest among the 70+ age group.

One of the advantages of using AI in this context is that the conversational format encourages users to spend more time than they perhaps might on a standard survey, therefore, enabling a greater breadth and depth of feedback to be collected from each individual user.

The average conversation had with Indicate by participating AA members was 15 minutes, with half of all conversations lasting between 15 and 30 minutes. For the small sample of 300+ members, this equates to more than 80 hours of ‘face time’, with essentially unlimited capacity to scale up to

thousands or millions of responses. This demonstrates the ability of AI to enable genuine discussions to take place on a scale that would simply not be possible 'in person'.

A number of conclusions can be drawn from the feedback on the experience itself which was provided by participants towards the end of their conversations with Indicate. The average satisfaction rating of seven, with only a handful of participants scoring the experience a four or below, demonstrates that the tool was largely meeting or exceeding the expectations and preconceptions of users. The open-ended feedback on the conversational agent experience offers insight into how Indicate could be further developed in order to lift the user satisfaction scores further:

- There is an opportunity to rework some of the 'yes/no' dialogue to alleviate a sense of repetitiveness noted by some participants;
- Some participants felt that the format resembled a survey too strongly – there is an opportunity here to rework the dialogue structure, both to make users aware that the AI *can* discuss topics beyond the set questions, and to actively *encourage* people to raise any issues that interest or concern them;
- As with any AI, there is room for growth in Indicate's language choices and conversational style – as more conversations are had, more clarity can be gained on what language style users are most comfortable with and responsive to.

Sentiment analysis is another area where AI offers fertile ground, as it enables open-ended responses to be sliced easily and strongly held views to be identified without significant administrative effort. AI can also show trends in sentiment for a particular issue / location / demographics over time.

A review of the AA members' verbatim responses and the associated sentiment scores suggests that sentiment is not a perfect metric, particularly where users speak sarcastically or in double negatives, but offers a good indicative picture especially across large numbers of responses.

The culmination of this research in the September rollout to members represents a significant and promising milestone in using Artificial Intelligence to better understand how AI can be used to gain insights on people's views.

9 Potential Next Steps

Based on the observations and findings of this research, we have identified the following potential next steps:

1. Examine methods to increase participation rates, including the use of a range of languages, interfaces, digital avatars.
2. Repeat the survey to determine any change in sentiment or support for particular transport issues since the General Election.
3. Publicise the potential for the use of AI for member engagement to other Automobile Associations, equivalent organisations, and through professional conferences and events.
4. Examine using AI to process and incorporate existing open-text feedback from emails, letters and surveys.
5. Examine the potential for using Indicate as an always-on feedback tool for AA members.

Beca would like to thank the Trustees of the Automobile Association Research Foundation for their support of this project.

10 Appendix – Summary of Testing Results

Testable agent abilities	Head & Tail		P0 - Name/Location/TopicIntro			P1 - Issues				P2 - General Questions				P3 - Suggestions for improvements	P4 - Background Questions						P5 - Feedback			
	Welcome	End	Q1	Q2	Q3	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q1	Q2	Q3	Q4	Q5	Q6	Q1	Q2	Q3	Q4
A - About the Agent 1 - Greets the user. 2 - States its name. 3 - States that it's AI (not human). 4 - States its purpose.																								
B - Question Delivery 1 - Question are asked using simple language. 2 - Able to repeat the question again. 3 - Dialogue variations for same question.																								
C - Response to Input 1 - Detects relevant answer (validity of input). 2 - Detects additional information. 3 - Responds to empty input. 4 - Digressions activated (if appropriate). 5 - Dialogue when nothing is understood. 6 - Dialogue when AI understands relevant inputs. 7 - Moves on if invalid input recognised many times. 8 - Terminates conversation if bot is not functioning properly. 9 - Responds respectfully to unfamiliar names. 10 - Can handle timeouts on both website and phone interface.																								
D - Recognition of Common Commands 1 - Skip question. 2 - Go back to previous question. 3 - I'm confused about this question. 4 - Why are you asking me this. 5 - Can stop entire conversation. 6 - Responds to asking to speak to human/manager. 7 - Time until end/progress through conversation. 8 - Able to restart from the beginning.																								
E - Child Node Questions (Questions which aren't a separate node) 1 - Skip question. 2 - Go back to previous question. 3 - Digressions activated (if appropriate).																								
F - Digressions 1 - Scolds user's profanity. 2 - Can tell one or two line jokes. 3 - Can tell user information about the itself (ethnicity, birth, language). 4 - Can respond to questions on how AI is doing. 5 - Responds to user saying farewell. 6 - Respond to topics it wants to avoid talking about. 7 - Can define terms such as congestion charging, public transport etc. 8 - All distractions work (Harry Potter, Rugby etc.) 9 - Respond to FAQ's. 10 - Can handle phone interface. 11 - Can respond to the user asking the exact same question back to the AI. 12 - Allows user to ask the AI a question i.e. Can I ask you a question? 13 - Can tell user which transport issues the AI has questions on. 14 - AI can tell the user the time and date (NZST).																								



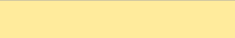

Legend	
Implemented and tested correctly	
Not Applicable	
Minor Issue	
Major Issue	

Figure 10-1 Comprehensive Testing Suite as of the 7th of February 2020

