La Trobe University
Department of Computer Science and Information Technology

BRP Data Analysis

By: Maninderpreet Singh Puri
Student ID: 20494381
To lecturer: Dr Dimetre Triadis

A report submitted in partial

fulfillment of the requirements

of the unit STM5IPL

in the Master of Data Science

La Trobe University, Melbourne Campus

February, 2021

Abstract

The BRP project (Business Resilience Program) is an initiate by the Melbourne Innovation Center (MIC) in collaboration with Australian Small Business and Advisory Services (ASBAS). The BRP program helps small businesses survive in Victoria, Tasmania and South Australia during the COVID-19 pandemic. It is an ongoing program which started in April. The data collected by MIC is from April till September. The services provided are in the form of workshops on topics illustrating the use of available online resources, managing human resource, finance, and stakeholders. This project is about the data analysis on data collected by MIC. The analysis aims to find new insights into the existing data by visuals to improve business strategy and enable stakeholders to decide on what measures to take while progressing in an ongoing program. We present the statistical analysis of client satisfaction and service hours using Power BI by other parameters. The results highlight that certain services lead to an improvement in satisfaction rating and an increased number of service hours. The steps carried out are described in this report: data preprocessing, transformation, visualization and analysis. The insights illustrate patterns, novel information, areas of improvement and future work in the Business Resilience Program.

Table of Contents

Glossary of terms

1. Statistical analysis- a scientific method of collecting data and finding patterns and trends
2. Data preprocessing- a part of the data mining process where data is cleaned before using it for analysis
3. Transformation- the process of converting one data form to another data form
4. Visualization- a way of creating visuals to communicate a message
5. Cloud services- on-demand services without the need to buy actual hardware or software
6. Big data- unstructured data in huge volumes
7. Natural language processing- field of computer science which deals with human and machine interaction
8. Correlation- dependency or any statistical relationship
9. Nominal- label variable without any qualitative value
10. Ordinal- variable with the natural order
11. Binary attributes- with only two values
12. Numerical attributes- attributes with the number where statistical methods can be applied
13. Parse- process of analyzing a string of words
14. Redundancies- repetitions of values
15. Outliers- abnormal values or of special interest
16. Patterns in data- change of values
17. Correlation matrix- the degree to which a pair of variables are related
18. Min/max and average- statistical measures
19. Q & A- Question and Answer
20. Word cloud- a collection of words in form of a cloud
21. Positive correlation- values of variables proportional to each other
22. Negative correlation- values of variables not proportional to each other
23. Parameter of interest- a variable that is the primary observation
24. Correlation hierarchy- hierarchy of correlated variables according to their index
25. Decision tree- used to make decisions based on the level of a variable in a tree.
26. Indigenous businesses- a business with more than 50% of indigenous employees

1.0 Introduction

Data analysis is a very important process for organizations with a large collection of data ,whether in physical storage space or cloud *(Foote, 2018)*. It has become extremely popular over the years due to the rise of big data and its benefits. Nowadays, there are many data analysis tools used like Qlik Sense, Tableau, Power BI, R Studio, SAS and programming languages like Python, R, Scala, SQL, etc. Each one of these has different capabilities and features suitable to specific needs. Because of Power BI's capabilities and ease of use, the BRP data collected by MIC and ASBAS digital solution is analyzed with Power BI. Power BI's open-source visuals, native R Integration, and advanced Excel feature using Data Analysis Expression (DAX) formula language, user access/security, natural language search interfaces and mobile interface feature makes it perfect for use in the BRP project for data analysis *(Xello, 2020).* Data Analysis is very essential for MIC, ASBAS digital solutions, and program partners (Adelaide Business Hub and Switch Tasmania) as it will present the overall impact of the BRP support services in Victoria, South Australia and Tasmania. Thus, this project aims to:

1. To identify and compare the success of services delivered by the number of hours and client satisfaction.
2. To compare divisions, services and their overall impacts.
3. To correlate parameters to predict future needs.
4. To find segments of services for increased client satisfaction using key influencers.
5. Analyze the overall impact of the BRP support services.
6. To get insights into small business challenges and needs.
7. To review the uptake in regional v/s metro or compare to population.
8. Provide recommendations for future support services.

The first section of the report will discuss the methodologies used in the project. It includes data type identification, preprocessing and visualization. Following this, the analysis will present an in-depth investigation of all the report findings. Finally, the scope of the project will be discussed in future scope section followed by a conclusion summarizing all the reports.

2.0 Methodologies

The data set collected by MIC had errors, missing values, miss-spelt values and wrong data types. To prepare the data for analysis, several steps were performed.

2.1 Data Types Identification

The data set was a mix of many data types and the process of identifying data involved understanding various data types, knowing differences and finally categorizing the columns. The data set consists of 44 columns in total. Most of the columns were nominal like: 'Legal name of business', 'Unique', 'ABN', 'Suburb', 'Postcode', 'First Name', 'Surname', 'Telephone', 'Email', 'State', 'Title', 'Prior experience with business advisors of any kind', 'Industry/ proposed industry of small business', 'How did the client hear about ASBAS Digital Solutions?', 'Has the client been referred to another program/advisory services?', 'Outcomes'. Out of all nominal attributes, most were insignificant for analysis like 'Legal name of business', 'Unique', 'ABN', 'First Name', 'Surname', 'Telephone', 'Email' and 'Title'. However, some attributes were essential, for instance 'Suburb', 'Postcode', 'State', as they could tell us about the most active locations in the program. Also, attributes like 'Prior experience with business advisors of any kind', 'Industry/ proposed

industry of small business', 'How did the client hear about ASBAS Digital Solutions?', 'Has the client been referred to another program/advisory services?' and 'Outcomes' can tell about the nature of the customer. There were a few attributes of the ordinal type like 'Date of service delivery', 'CURRENT Number of Employee', and 'Client satisfaction rating'. Date of service was used frequently to determine the patterns across months or how did the client requirements change during the entire timeline. The current number of employees can be related to other parameters to find customer behaviour according to the strength of employees, whereas client satisfaction is a critical parameter which measures the satisfaction level of customers. These attributes can be correlated with other parameters to find their impact on client satisfaction. There were many binary attributes with values 0 and 1, Yes and No, Y and N for instance ‘Consent to be surveyed’, ‘Metro’, ‘Regional’, ‘Indigenous in Business’, ‘Was this service delivered as part of the COVID-19 Support program?’, ‘Client satisfaction survey completed’, 'New or Returning Business' and ‘Is the client happy to recommend ASBAS services to another small business?’. Also, there were a few numerical attributes like ‘Number of service hours’, 'Services delivery modes', COVID-services, and non-COVID. These attributes are significant for analysis as they tell the number of attendees/number of hours consumed during the program. The success of the BRP program is determined by these parameters. Therefore the data type identification process helped in determining which methods should be used to preprocess data.

2.2 Preprocessing

The next crucial step was data processing. In any analysis, data preprocessing is the step where data gets transformed or encoded to bring it into a state where the machine can parse it easily *(Pandey, 2020).* In the Power BI desktop, this can be done in Power Query feature. All the preprocessing was executed in Power Query for transforming and querying data sets. Columns had to be renamed as they were long. For analysis, long definitions were not required like for example the column presenting the ‘Legal Business Name’ did not require details. Similarly, other columns were renamed to 'Current number of Employees', 'Indigenous in Business', 'How did client hear about ASBAS digital solutions', 'Has the client been referred to another program/advisory services?', 'Outcomes'. These names were descriptive enough to identify the purpose.

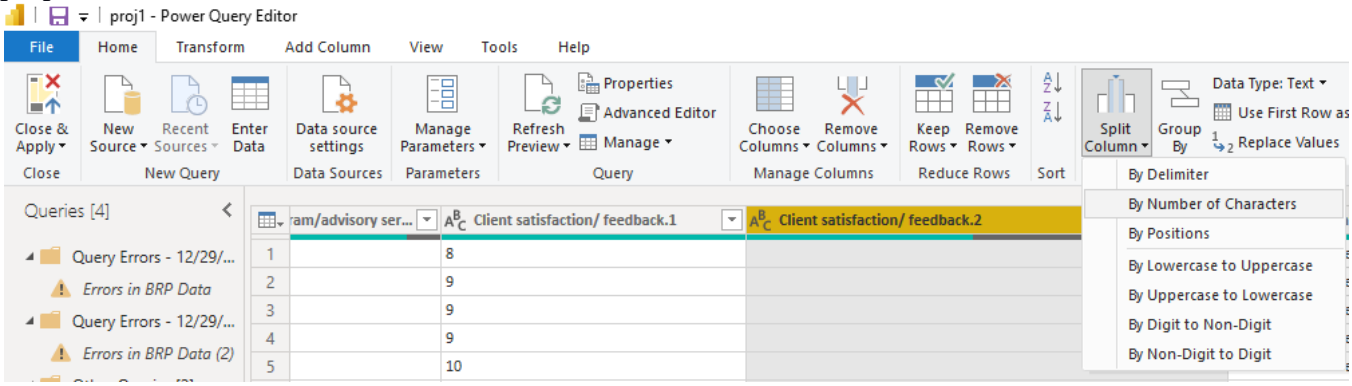

Figure 1 illustrates the Split Column function- By Number of Characters.

In 'Client satisfaction/ feedback Categories' the satisfaction rating was written with the feedback in the same column. Hence, the rating could not be used as a numeric value. Therefore, to make the client satisfaction rating into a numerical column and feedback into text attribute we split the column into two using split columns function, further splitting by the number of

character options as described in figure 1. The number of characters was defined: in this case 4, to split the column into two. Afterwards, transformations were applied like replacing the ‘No f’ values to ‘0’, to make client satisfaction rating purely numerical. Also, the feedback column had the remaining letters ‘eedback’ after the split. So they were replaced with ‘No feedback’.

Figure 2 percentage of empty values in the 'Unique' column.

As described in figure 2, the ‘Unique’ attribute was removed as 45% were empty values and ‘Legal Name of Business’ can be used instead to identify business names. The columns ‘Metro’ and ‘Regional’ were complements of each other so they were combined to create a single column where ‘0’ means the business is located in Regional area and ‘1’ means it is in Metro. This was done by dropping one column and renaming another column to 'Metro or Regional area'. There were redundancies in the ‘Titles’, ‘Divisions’ column which made these columns inconsistent. Using the Replace values option in Power Query they were made consistent. Also, the values in 'Was this service delivered as part of the COVID-19 Support program? ', 'Is the client happy to recommend ASBAS services to another small business?', 'Has the client been referred to another program/advisory services?' were replaced, where yes was denoted as Y and y. So using Replace values option columns were made consistent. Besides, ‘Outcome’ column had ‘null’ & ‘-’. We changed it to ‘no outcomes identified’ using the Replace values option in Power Query. Similarly, the '-' values to '0' in the ‘Client satisfaction rating' column were also replaced. Therefore data preprocessing made the data free of redundancies, errors and misspell values. This process prepared data for final step that is visualization.

2.3 Visualization

Next step in the analysis was data visualization. Data visualization provides an accessible way to notice and understand trends, outliers, and patterns in data. In the project, in total 12 reports were built representing different aspects of BRP data.

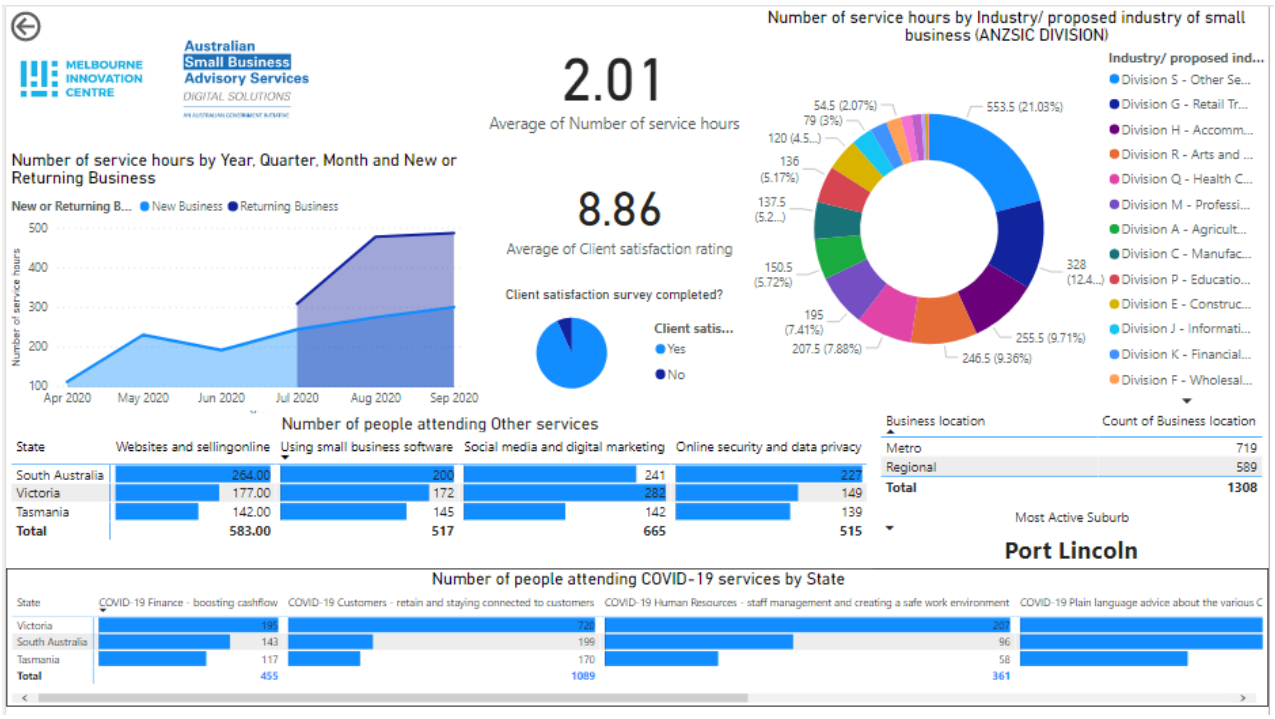

Figure 3 'BRP highlights' Report emphasizes the main attributes across the timeframe.

As described in figure 3, the first report tells the main areas of focus. It has a graph describing the number of service hours taken up by new or existing customers throughout the timeline from April to September. It also breaks down the number of service hours taken up by division, the average number of hours, average client satisfaction, has the client completed the satisfaction survey, number of people attending COVID-19 and other services, business location whether Metro or regional and most active suburb. The visualizations used are doughnut chart, pie chart, card, matrix and area chart. The average number of services hours and average client satisfaction were chosen as values for card visualizations. COVID-19 service and other services were combined in two matrixes to compare the uptake of services in both content types. All the selectors in the report can be used to filter data.

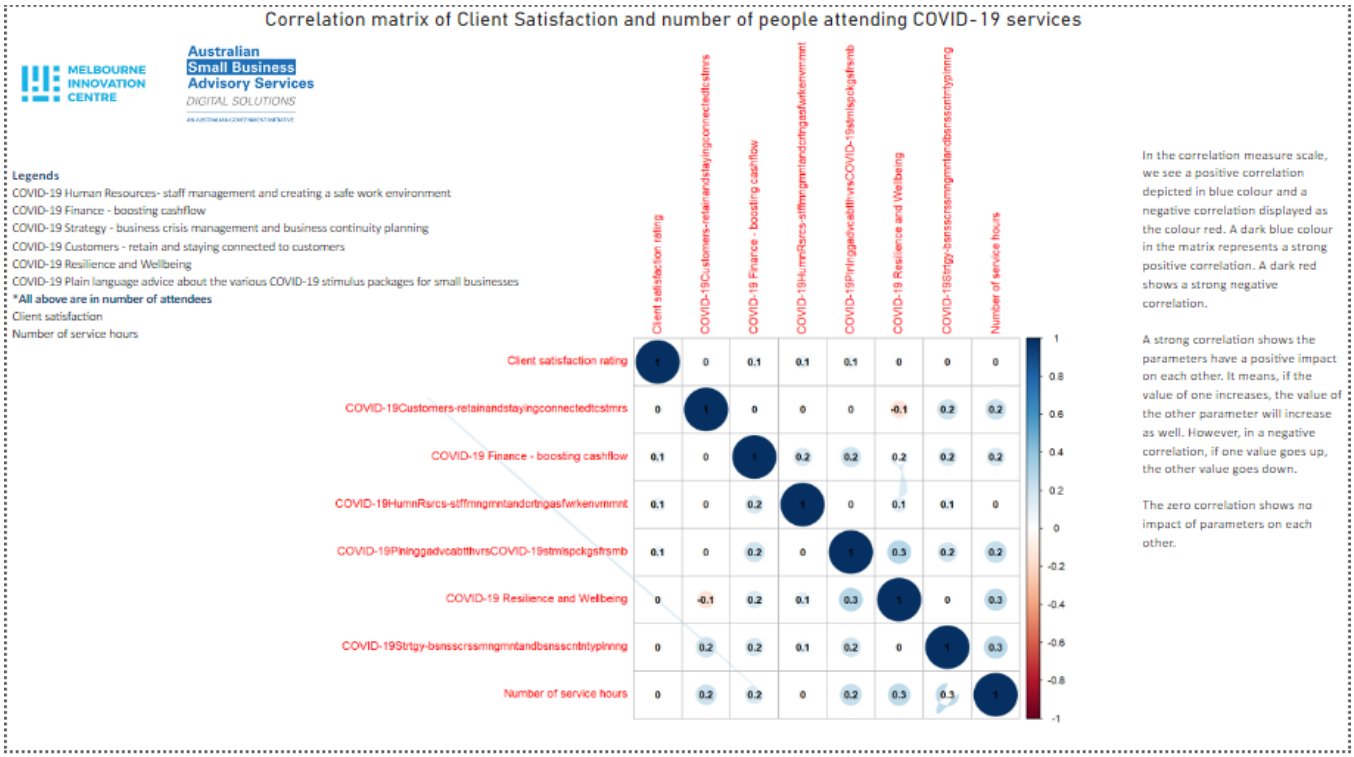

Figure 4 second report describes the correlation of client satisfaction, number of service hours and COVID -19 services.

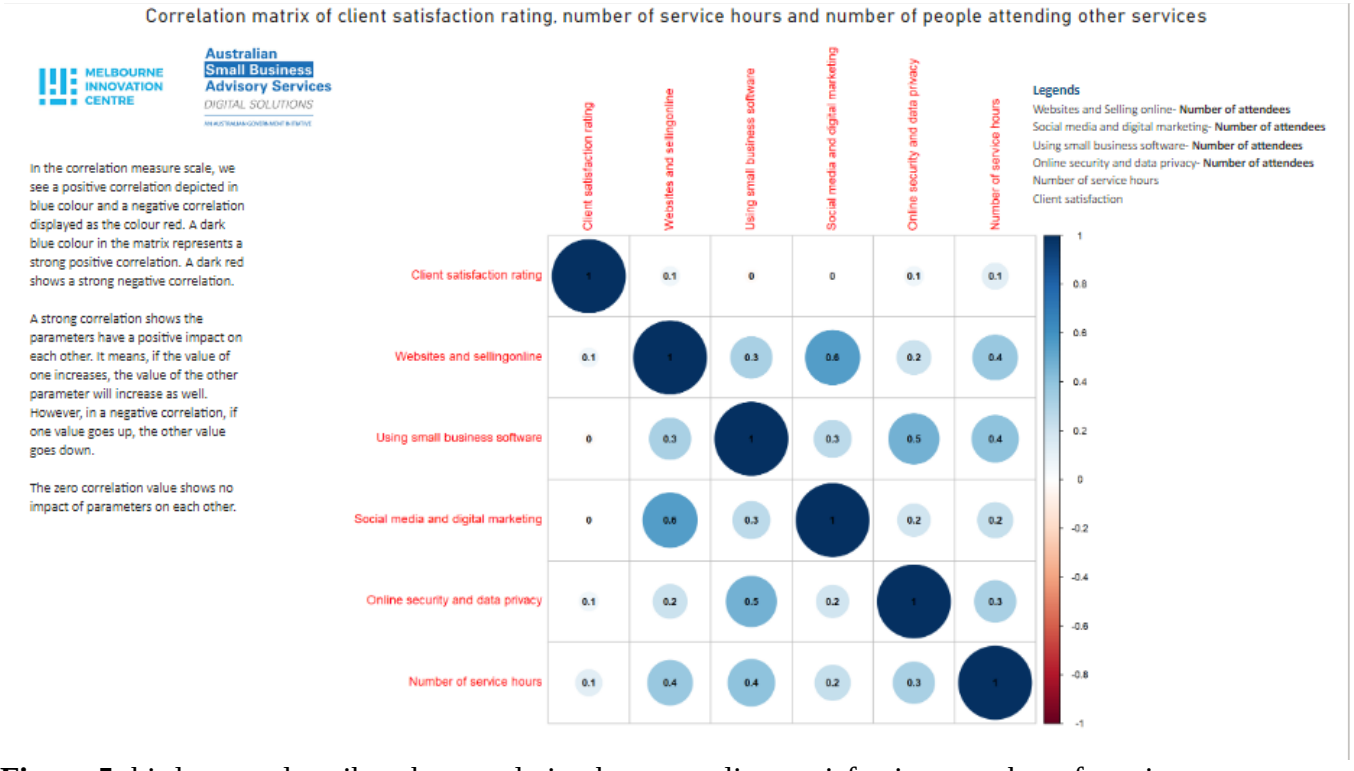

Figure 5 third report describes the correlation between client satisfaction, number of service hours and other services.

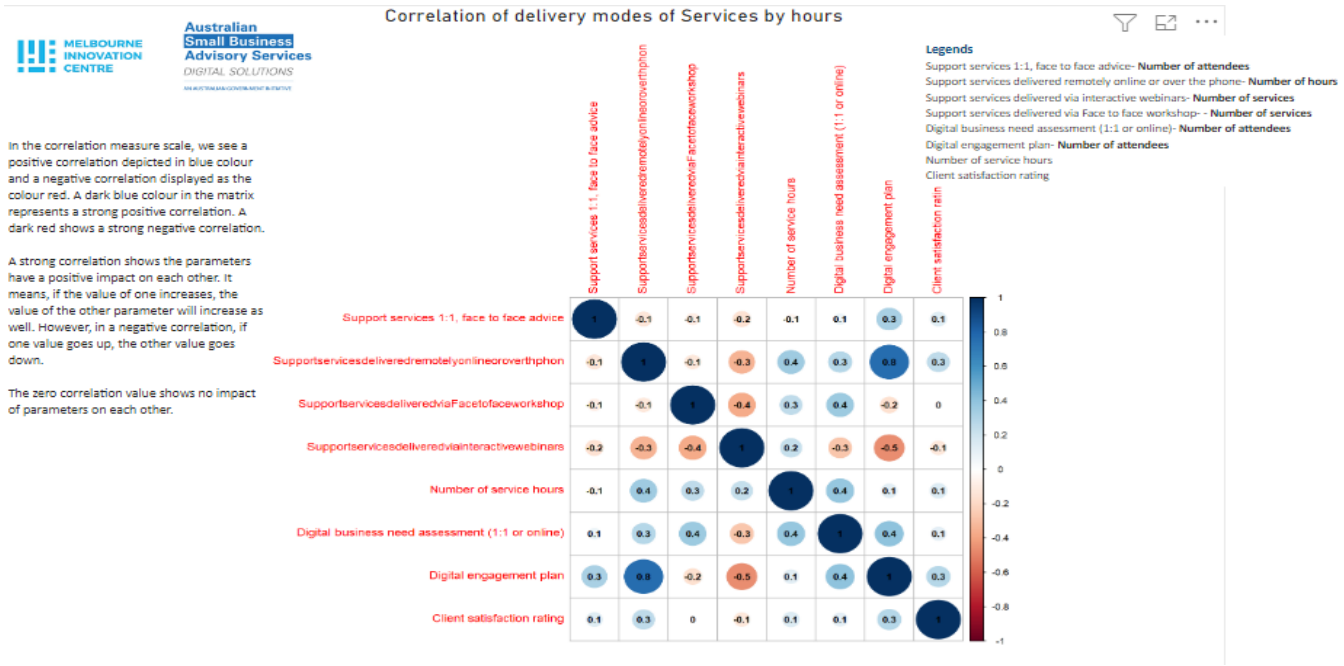

Figure 6 fourth report describes the correlation between client satisfaction, number of service hours and delivery modes.

The next three reports as depicted in figures 4, 5 and 6 are about the correlation of COVID-19 services, other services, and delivery modes of services with the number of service hours and client satisfaction rating. The definitions were provided using text boxes and legends were added. The three figures below depict the actual representation of the correlation reports.

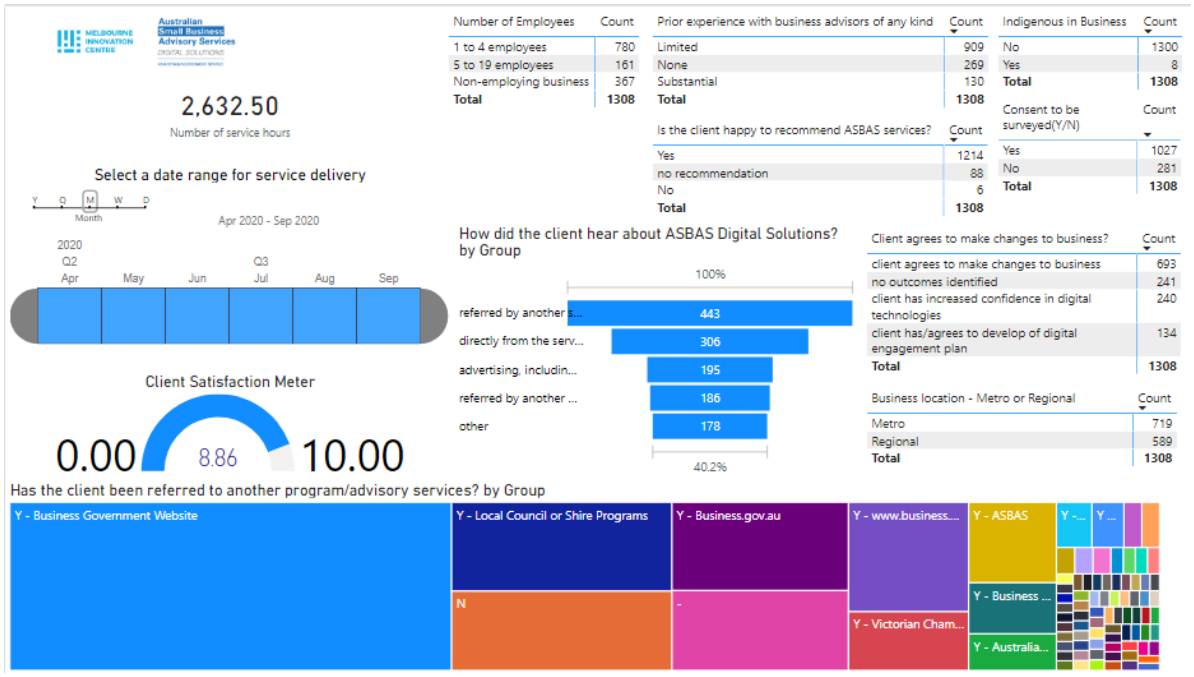

Figure 7 fifth report describes the key numbers in attributes across the timeframe.

In figure 7 we can notice that the report highlights the parameters in form of key numbers according to the selection of time range in the timeline selector. It highlights the mix/max and average client satisfaction during the selected timeframe, how client heard about ASBAS digital Solutions, the number of employees, prior experience with business advisory, Indigenous in business, consent to be surveyed, is the client happy to recommend, the client agrees to make changes, business location (Metro or Regional), to which program advisory has the client been referred. Every graph selection can be used as a filter. This is one of the best features of Power BI which enables users to cross filter on other graphs.

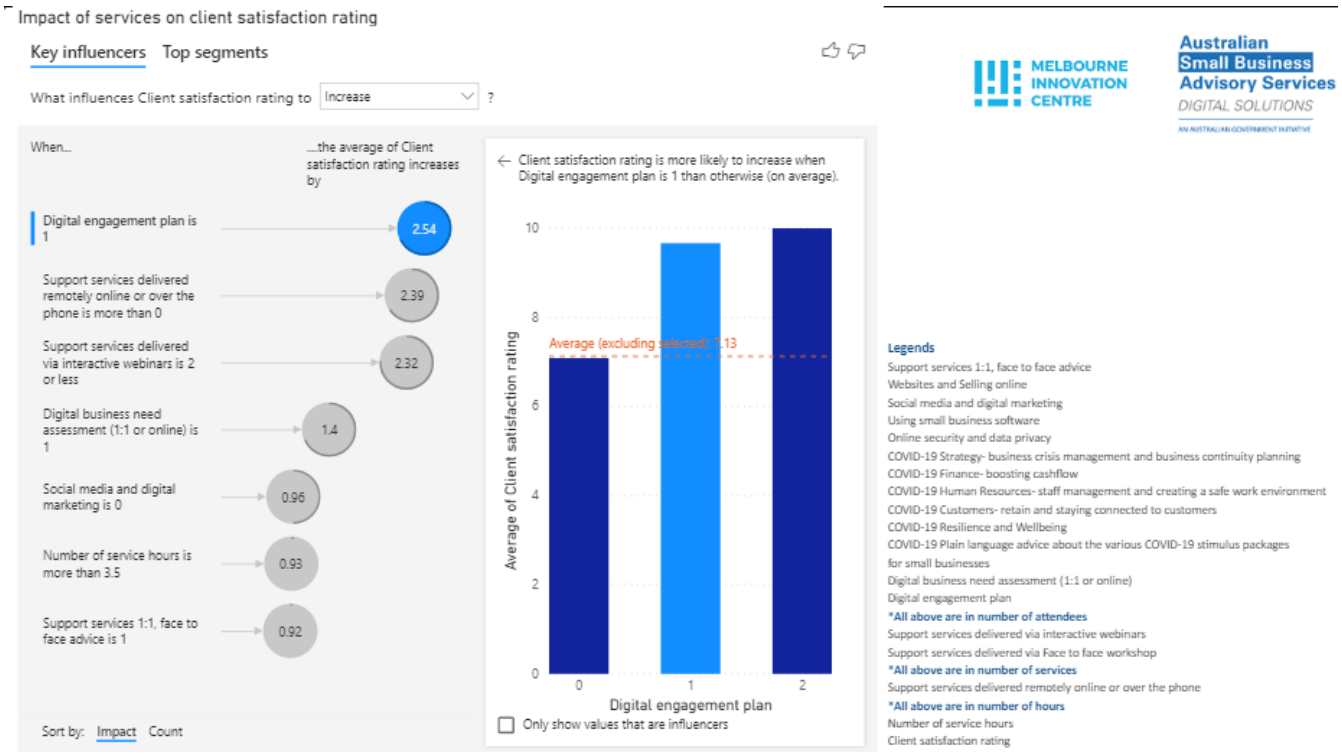

Figure 8 sixth report highlights the impact of services on client satisfaction rating.

In figure 8, the visualization 'key influencer' tells about the impact of all services on satisfaction rating. It also tells how changing values in services can affect the ratings.

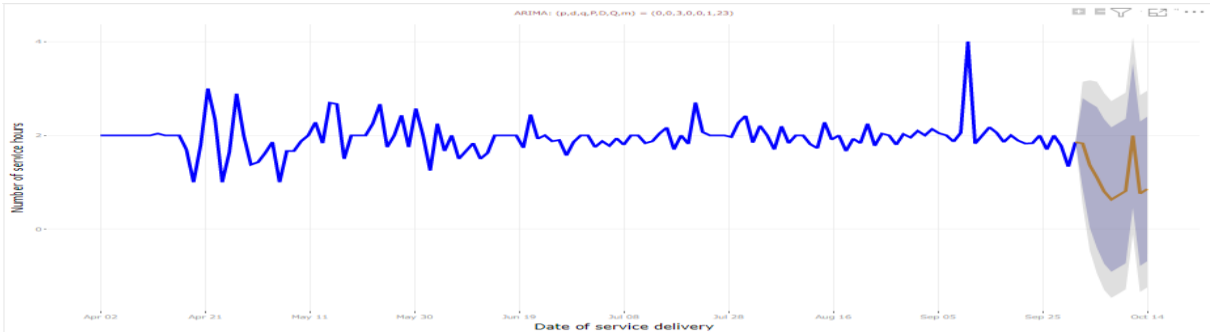

Figure 9 seventh report highlights the prediction of the number of service hours.

Figure 9 highlights the prediction of the number of service hours. The visualization is called ‘Forecast with ARIMA’.

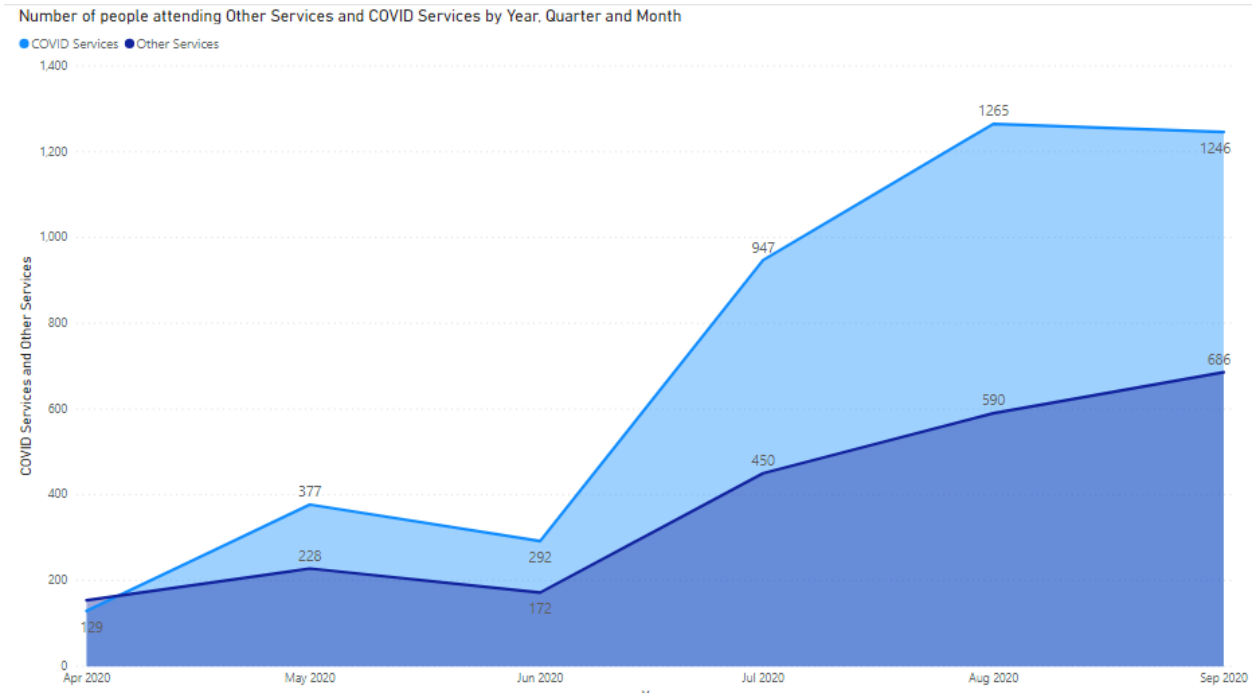

Figure 10 report highlights the number of people attending COVID-19 and other services.

Figure 10 highlights the report is an area chart which tells about the number of people attending COVID-19 services v/s other services.

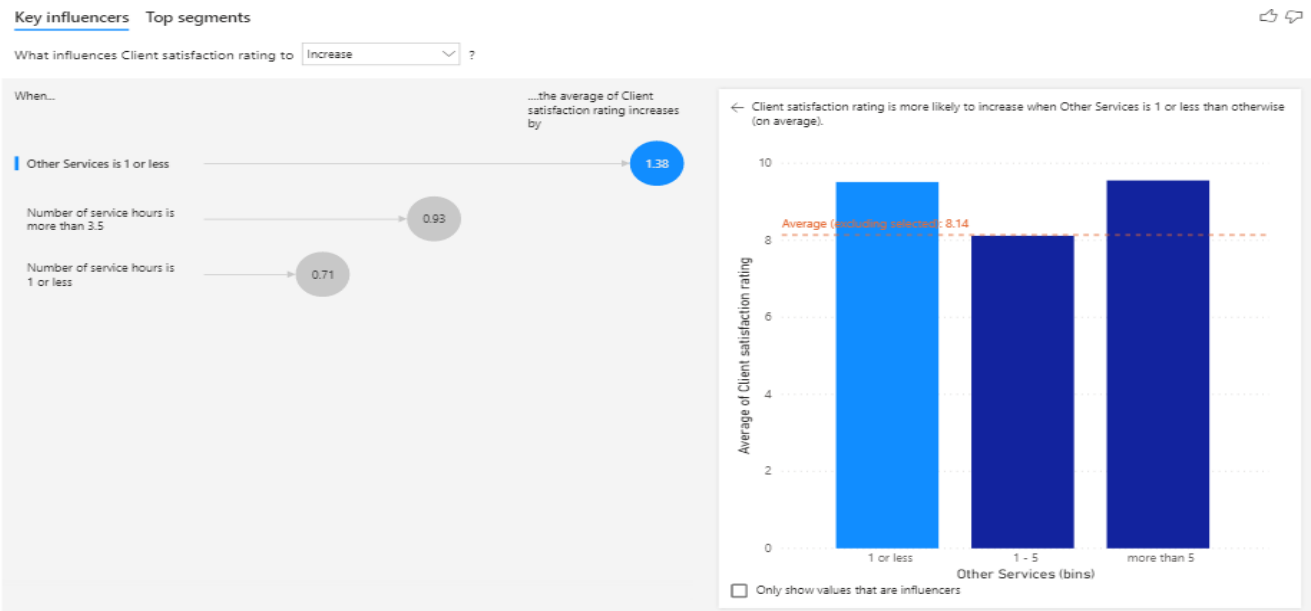

Figure 11 ninth report highlights the influences of content type and the number of service hours on client satisfaction rating.

Figure 11 reports key influencer illustrating the impact of content type on satisfaction rating.

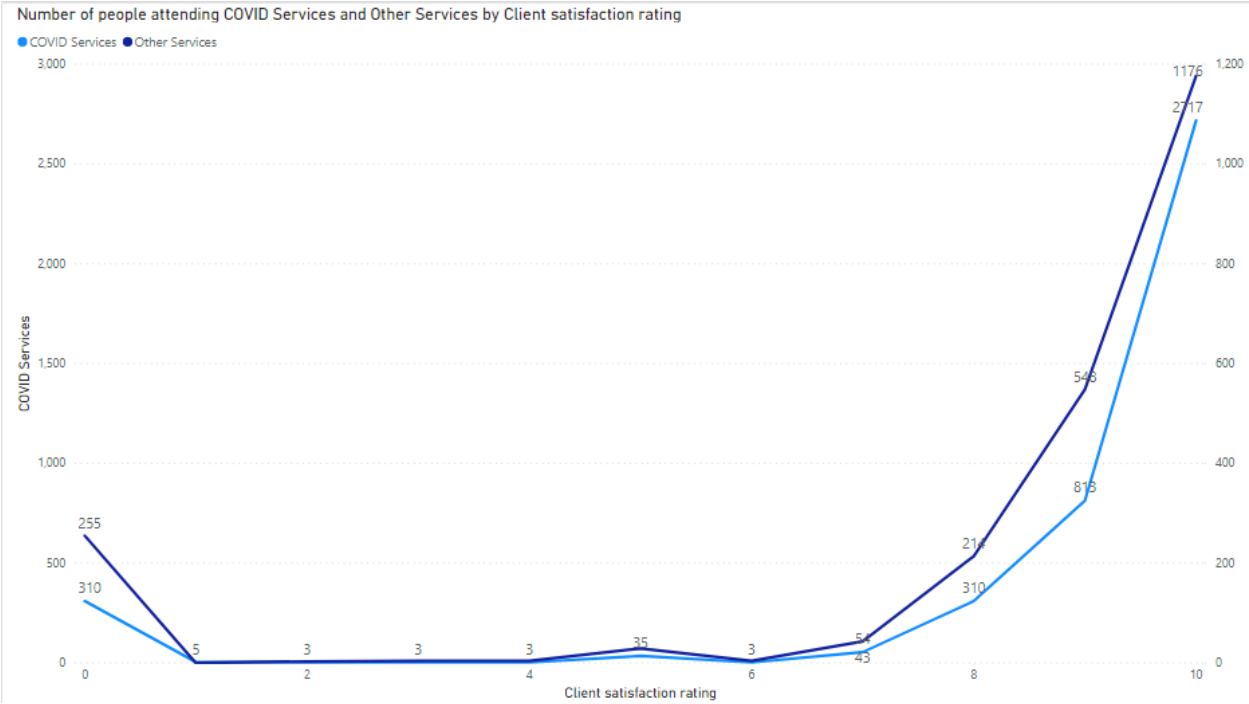

Figure 12 tenth report highlights the number of people attending COVID and other services by client satisfaction rating.

Figure 12 highlights a line chart comparing client satisfaction by COVID services v/s other services (content-type).

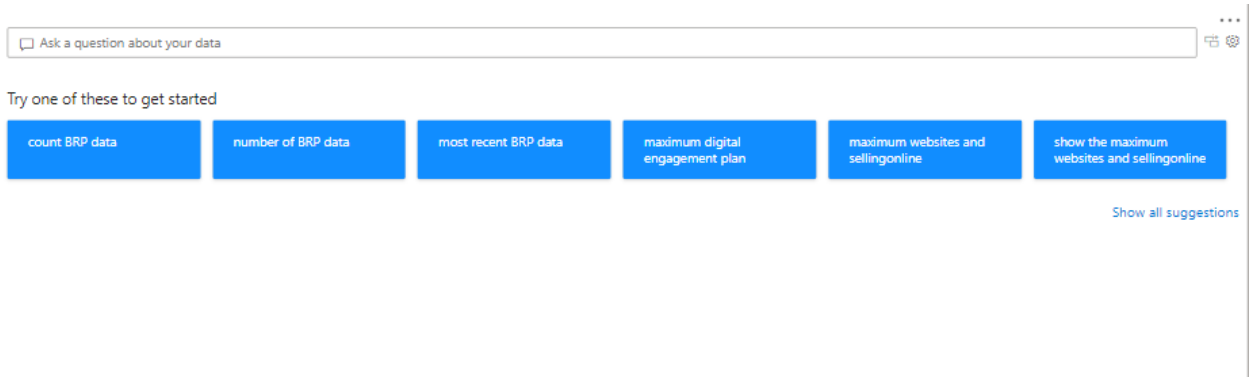

Figure 12 eleventh report highlights the Q & A section.

In figure 12, the report is visual a Q & A where user can type questions to get answers directly. For example, average client satisfaction rating by states.

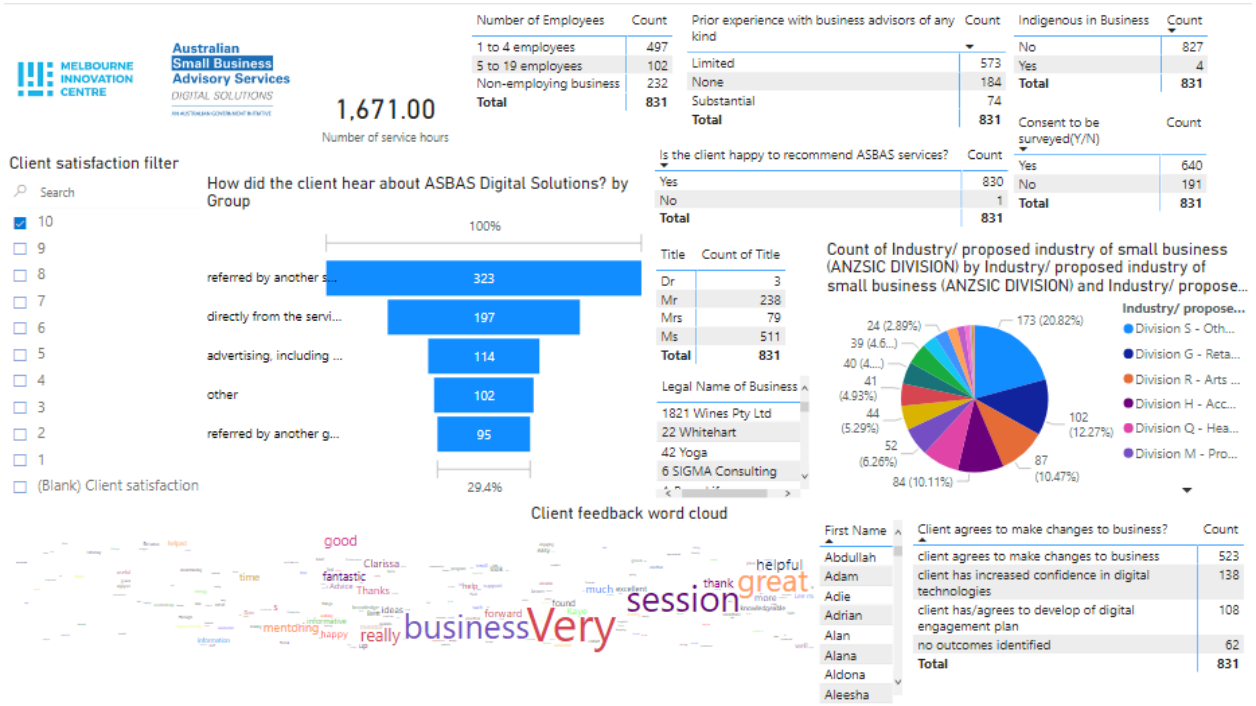

Figure 13 report highlights the word cloud of the client's feedback.

The last report is a fun visual that tells about the client's reaction by presenting a word cloud of the client's feedback. In figure 13, we can notice data is filtered using the client rating filter. Also, we can check which business was happy and unhappy with the service, the name of the review writer and gender. We can check other information like number of employees, division, prior experience with advisory, Indigenous in business, is the client happy to recommend and consent to be surveyed for selected client ratings.

3.0 Analysis

The final stage of a data science project is drawing out an analysis. The first report says that during the year 2020, average client satisfaction was 8.86 out of 10 and an average number of service hours were 2.01. The doughnut chart highlights the distribution of Divisions by the number of hours. We can notice in figure 3 that 'Division S- other services' was the highest consumer of the number of service hours and 'Division O- Public Administration and Safety' is the lowest consumer. We can notice by the total count in matrixes that the most popular service is 'COVID-19 Customers' and the least popular service is 'COVID-19 Human Resources' in all the Covid-19 services. This is relevant to the objective of MIC to find the success of services delivered by the number of hours and client satisfaction. Also, we meet the objective of comparing Divisions, services and their overall impacts. We can also notice that Victoria was the state with the most number of people attending the services**.**

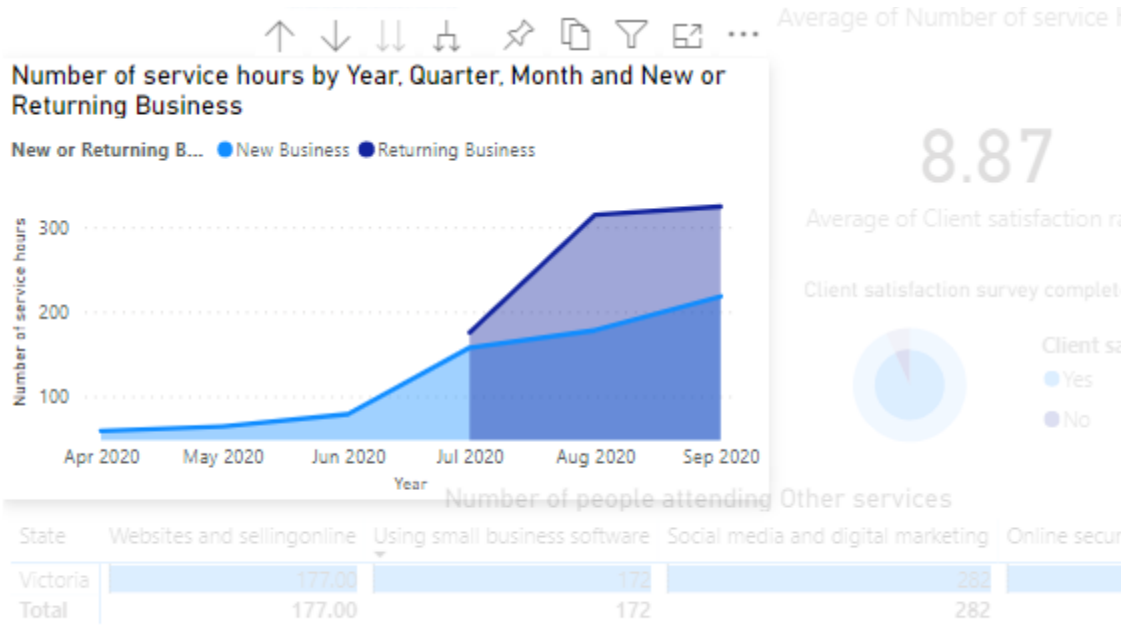

Figure 14 filtered timeline by Victoria.

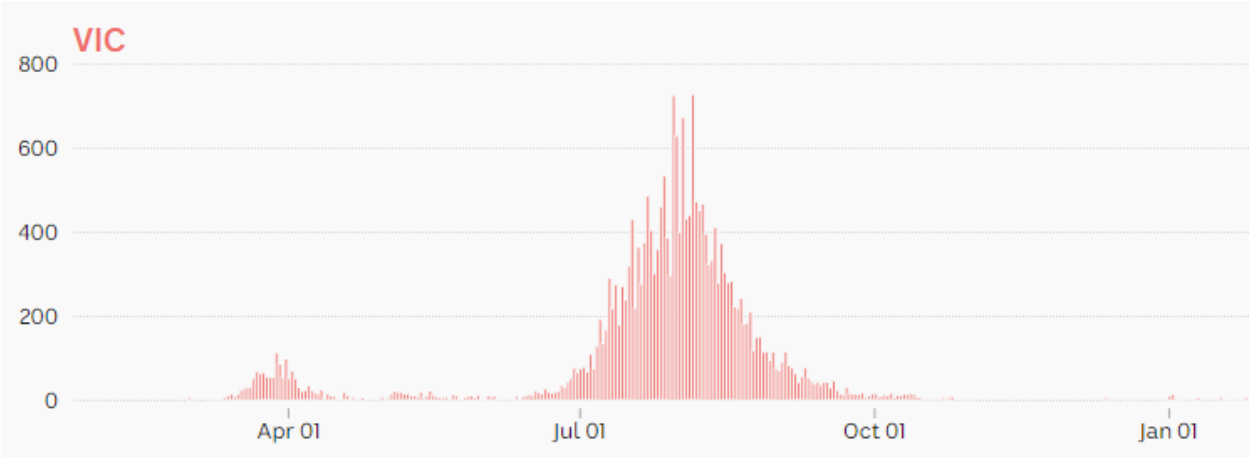

Figure 15 number of cases in Victoria in 2020 *(Ting, Scott, Workman & Hutcheon, 2021).*

In figure 14 we filtered the report by selecting Victoria. We can notice many interesting insights. The timeline highlights an increase in the number of service hours from Jun 2020 onwards. If we compare it to actual COVID cases in figure 15, it highlights that the number of COVID cases increased in Victoria at the same time *(Ting, Scott, Workman & Hutcheon, 2021).* Similarly, other interesting patterns can be discovered in other states.

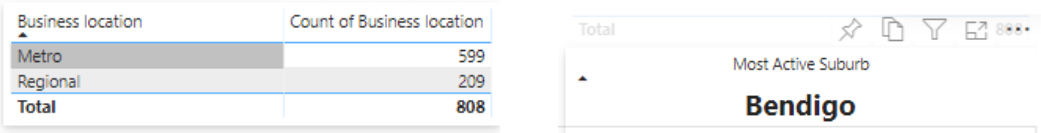

Figure 16 illustrating the most active suburb.

The report also tells the most active suburb during a timeline. We can notice in figure 16 that in 2020 Bendigo was most actively involved in the program and most of the suburbs who participated were in the Metro area.

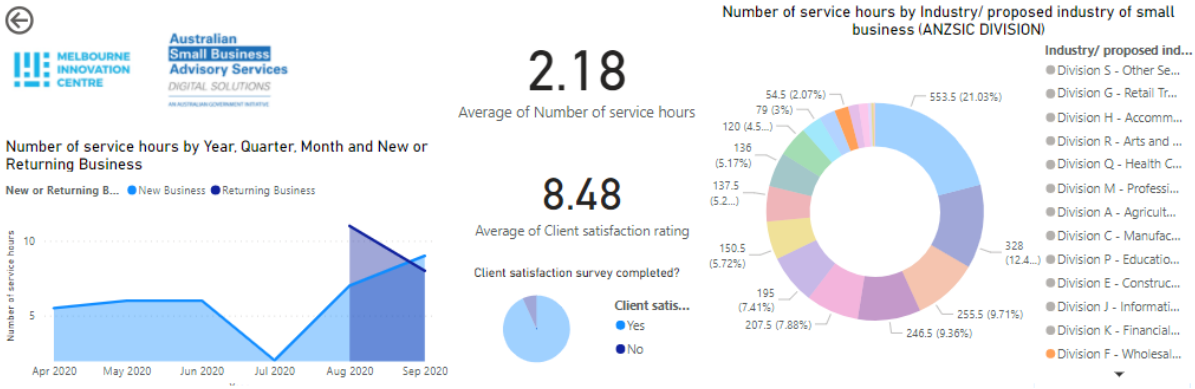

Figure 17 filtered timeline by 'Division F-Wholesale Trade'.

Similarly finding statistics is easy and many important patterns can be identified using this method. For example in figure 17 we can notice for 'Division F-Wholesale Trade' the number of service hours decreased in July and spiked again in August with returning customer.

Table 1 illustrating the positive correlation hierarchy of client satisfaction, the number of hours and COVID services.

Parameter of interest (for positive correlation)	Correlation Hierarchy
Number of service hours	1. COVID-19 Resilience and Wellbeing → COVID-19 Plain Language advice → COVID-19 Strategy or COVID-19 Finance 2. COVID-19 Strategy → COVID-19 Customers or COVID-19 Finance or COVID-19 Plain Language advice
Client Satisfaction	1. COVID-19 Finance → COVID-19 Human Resource or COVID-19 Plain Language advice or COVID-19 Resilience and Wellbeing or COVID-19 Strategy or Number of service hours 2. COVID-19 Human Resource → COVID-19 Finance 3. COVID-19 Plain Language advice → COVID-19 Resilience and Wellbeing
COVID-19 Plain Language advice	1. COVID-19 Resilience and Wellbeing → Number of service hours
COVID-19 Resilience and Wellbeing	1. COVID-19 Plain Language advice→ COVID-19 Finance or Number of service hours or COVID-19 Strategy 2. Number of service hours → COVID-19 Strategy
COVID-19 Customers	1. COVID-19 Strategy→ Number of service hours 2. Number of service hours→ COVID-19 Resilience and Wellbeing or COVID-19 Strategy
COVID-19 Strategy	1. Number of service hours → COVID-19 Resilience and Wellbeing
COVID-19 Human Resource	1. COVID-19 Finance→ Number of service hours or COVID-19 Resilience and Wellbeing or COVID-19 Strategy or COVID-19 Plain Language advice or COVID-19 Human Resource
COVID-19 Finance	1. Number of service hours

	2. COVID-19 Resilience and Wellbeing 3. COVID-19 Strategy 4. COVID-19 Plain Language advice 5. COVID-19 Human Resource

In the second report, the correlation matrix highlights positive correlations between services, number of services hours and client satisfaction. We can notice that client satisfaction rating has a 0.1 positive correlation with COVID-19 Finance, Human Resource and Plain language advice. This means that there is a 10 per cent chance that if someone attends these services, client satisfaction will increase. Besides, the matrix reveals which service to prioritize first. For instance, if the organization targets client satisfaction they should market COVID-19 Finance, Human Resource and Plain language advice.

Table 2 illustrating the positive correlation hierarchy of the number of service hours, client satisfaction, and other services.

Parameter of interest (for positive correlation)	Correlation Hierarchy
Number of service hours	1. Websites and selling online→ Social media and digital marketing→ Using small business software→ Online security and data privacy 2. Using small business software→ Online security and data privacy
Client Satisfaction	1. Websites and selling online → Social media and digital marketing→ Using small business software→ Online security and data privacy→ Number of service hours 2. Online security and data privacy→ Using small business software 3. Websites and selling online or Using small business software
Websites and selling online	1. Social media and digital marketing → Using small business software 2. Number of service hours → Using small business software or Online security and data privacy
Social media and digital marketing	1. Websites and selling online → Number of service hours
Using small business software	1. Online security and data privacy→ Number of service hours 2. Number of service hours→ Websites and selling online
Online security and data privacy	1. Using small business software→ Number of service hours

The second correlation matrix highlights the relation between non-COVID services, client satisfaction rating and the number of service hours. The matrix reveals most of the parameters are positively related to each other. Table 2 illustrates the positive correlation hierarchy.

Table 3 illustrating the positive correlation hierarchy of the number of service hours, client satisfaction and delivery modes.

Parameter of interest (for positive correlation)	Correlation Hierarchy

Number of service hours	1. Support service delivered remotely online or over the phone→ Digital engagement plan→ Digital business needs assessment (1:1 or online) 2. Digital business need assessment (1:1 or online) → Support service delivered via face to face workshop or → Digital engagement plan
Client Satisfaction	1. Support service delivered remotely online or over the phone→ Digital engagement plan→ Digital business need assessment (1:1 or online) 2. Digital engagement plan→ Digital business need assessment (1:1 or online)
Support service 1:1 face to face advice	1. Digital engagement plan→ Digital business needs assessment (1:1 or online) 2. Client Satisfaction→ Digital engagement plan or Support service delivered remotely online or over the phone
Support service delivered remotely online or over the phone	1. Digital engagement plan→ Digital business need assessment (1:1 or online)
Support service delivered via interactive webinars	1. Number of service hours→ Support service delivered remotely online or over the phone, or Digital business needs assessment (1:1 or online)
Support service delivered via face to face workshop	1. Digital business need assessment (1:1 or online) → Support service delivered via face to face workshop or Number of service hours or Digital engagement plan 2. Number of service hours→ Support service delivered remotely online or over the phone, or Digital business needs assessment (1:1 or online)
Digital engagement plan	1. Support service delivered remotely online or over the phone→ Digital engagement plan→ Digital business need assessment (1:1 or online)
Digital business needs assessment (1:1 or online)	1. Support service delivered remotely online or over the phone→ Digital engagement plan→ Digital business need assessment (1:1 or online) 2. Number of service hours→ Support service delivered remotely online or over the phone, or Digital business needs assessment (1:1 or online) 3. Digital engagement plan→ Digital business need assessment (1:1 or online)

The third correlation matrix highlights the correlation between the delivery modes of service, client satisfaction rating and the number of service hours. It reveals that parameters have both positive and negative correlations with each other. Table 3 highlights the positive correlation hierarchy. The above tables highlight one of the many relationships which can be drawn from the correlation matrix. A decision tree can be created using these relationships. The main benefit of these tables for MIC is that when offering services they can decide on which service to prioritize for better performance in targeted services. This meets the objective of MIC to correlate parameters to predict future needs.

Figure 18 illustrates the gradual increase in satisfaction rating over the months.

The fifth report tells about key numbers in the project. It explains more about the success story of the program. In figure 18, we can notice that from April to September there is a gradual increase in satisfaction ratings.

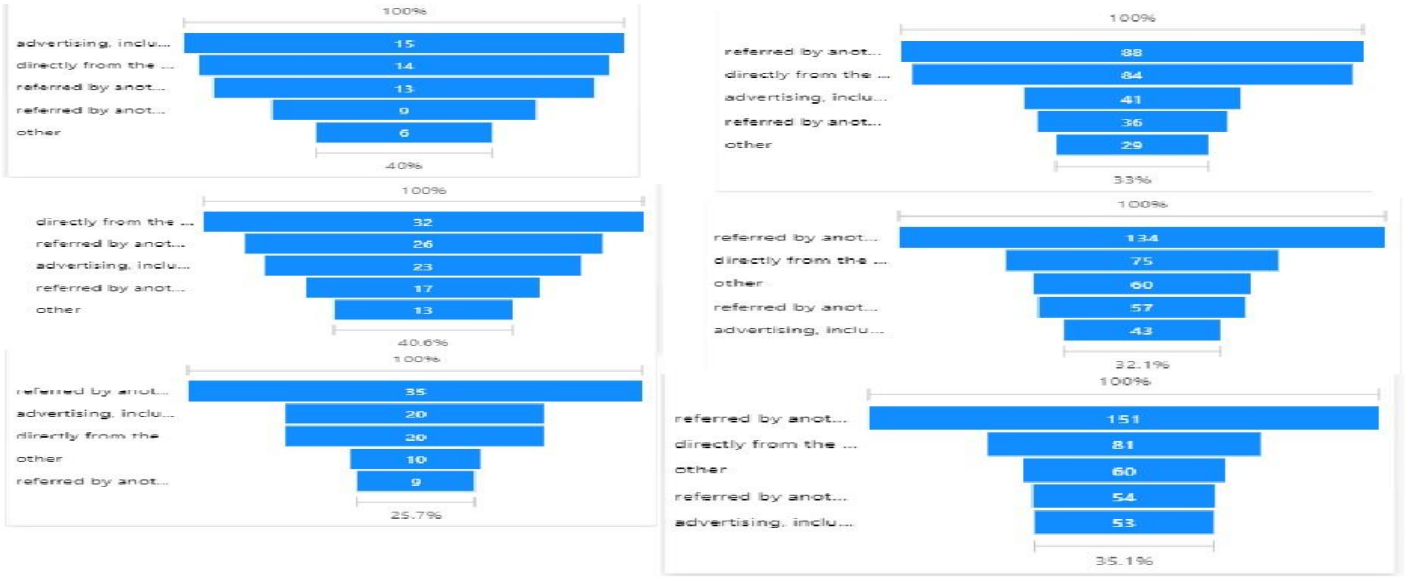

Figure 19 illustrating the change in how clients heard about ASBAS digital solution (displayed by column).

Also, we can notice in figure 19 how time clients heard about ASBAS Digital solutions changes over time. The majority changed from ‘advertising’ to ‘referred by another small business’ during the timeline.

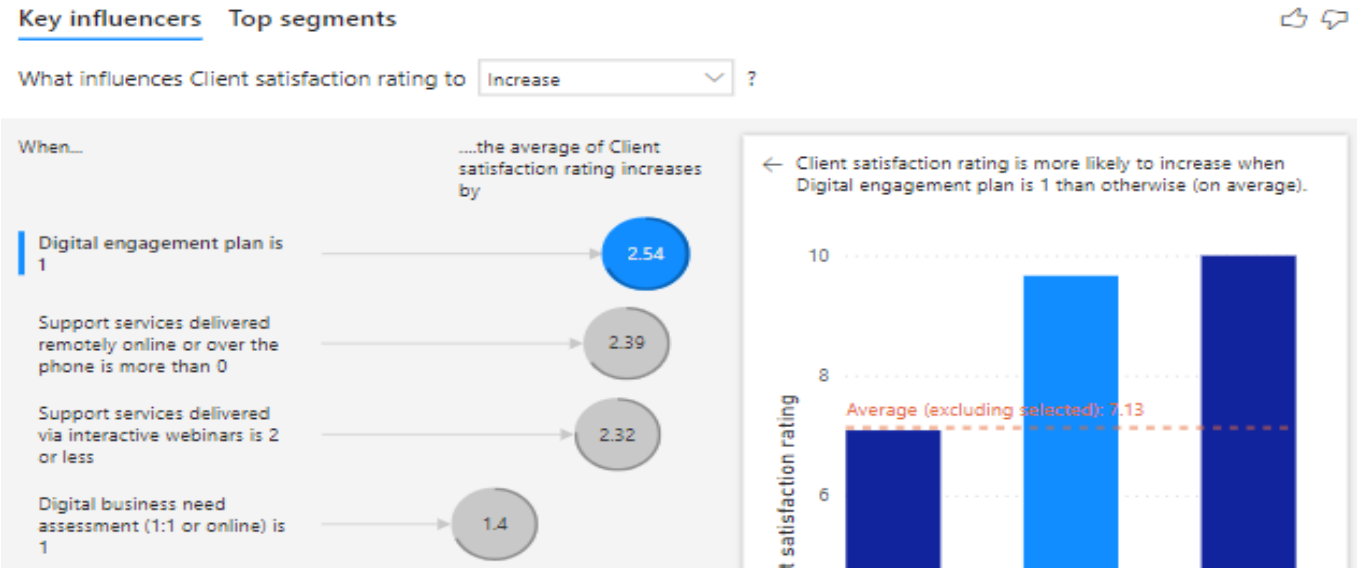

Figure 20 key influencer illustrating increased client satisfaction with other parameters.

The sixth report tells us about the key influences of the parameters on the client satisfaction rating. In figure 20 we can notice that the key influencer tells us that ‘Digital Engagement Plan’ has the biggest impact followed by ‘Support service delivered online or over the phone’, followed by ‘Support services delivered via interactive webinars’ and so on.

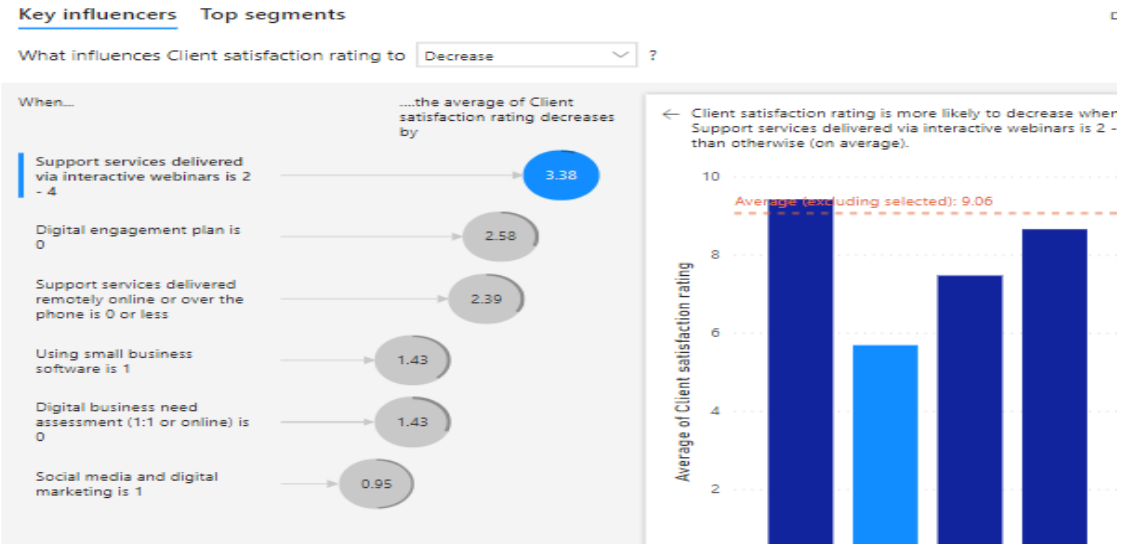

Figure 21 key influencer illustrating decreased client satisfaction with other parameters.

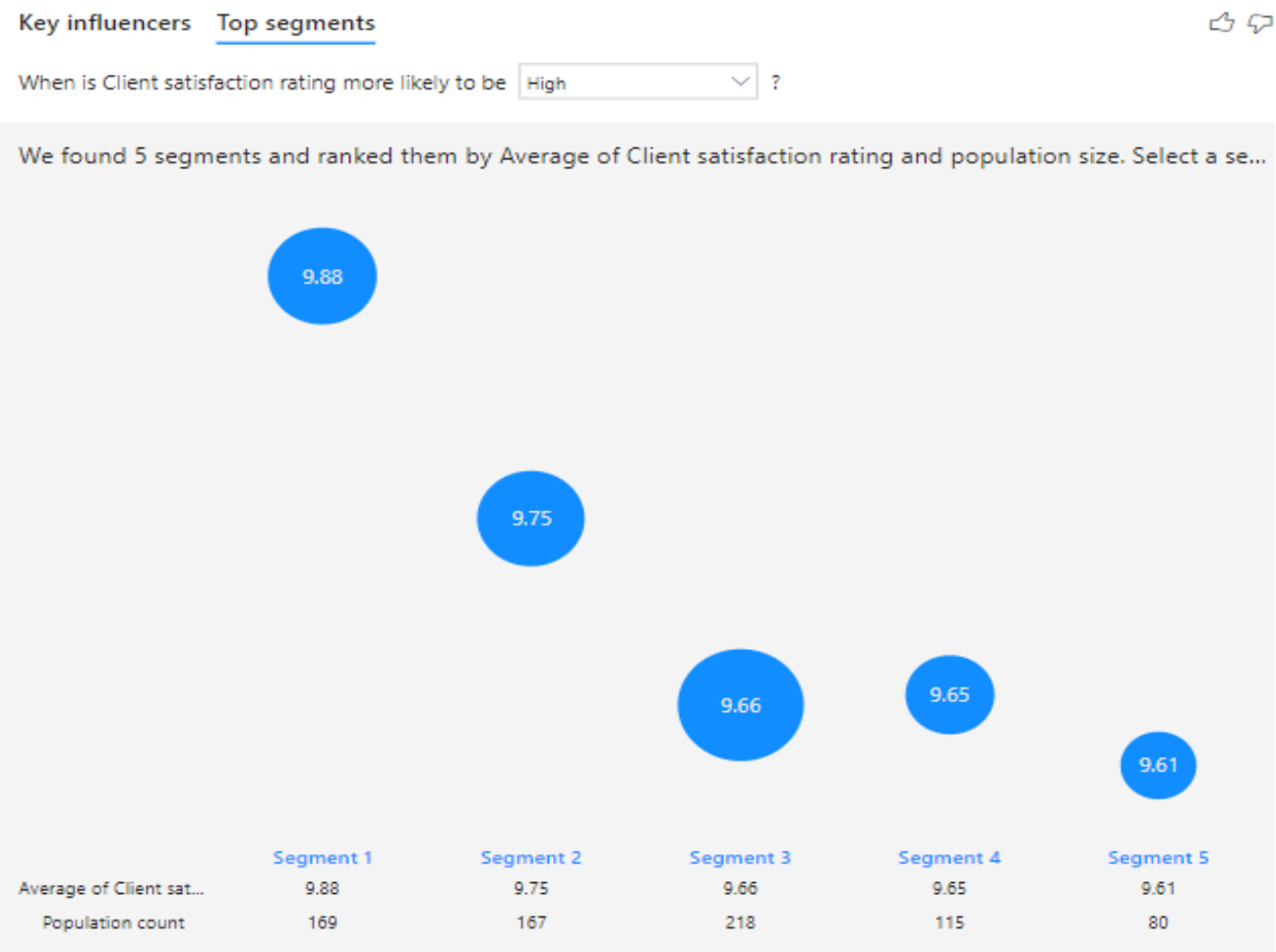

Figure 22 highlights the segments in parameters impacting client satisfaction.

We can also notice negative influencing parameters in figure 21. Also, segments can be checked in figure 22 to notice the combined effect of two or more parameters. This meets the objective of finding the segments for increased client satisfaction.

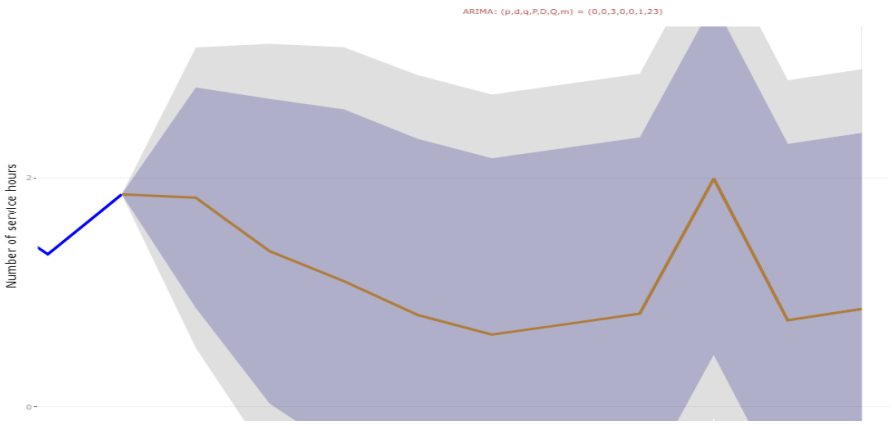

Figure 23 highlights the predicted number of hours in October.

The seventh report predicts the number of service hours for the next 13 days after the end of the timeline. In figure 23, it highlights that the number of service hours can drop in the future.

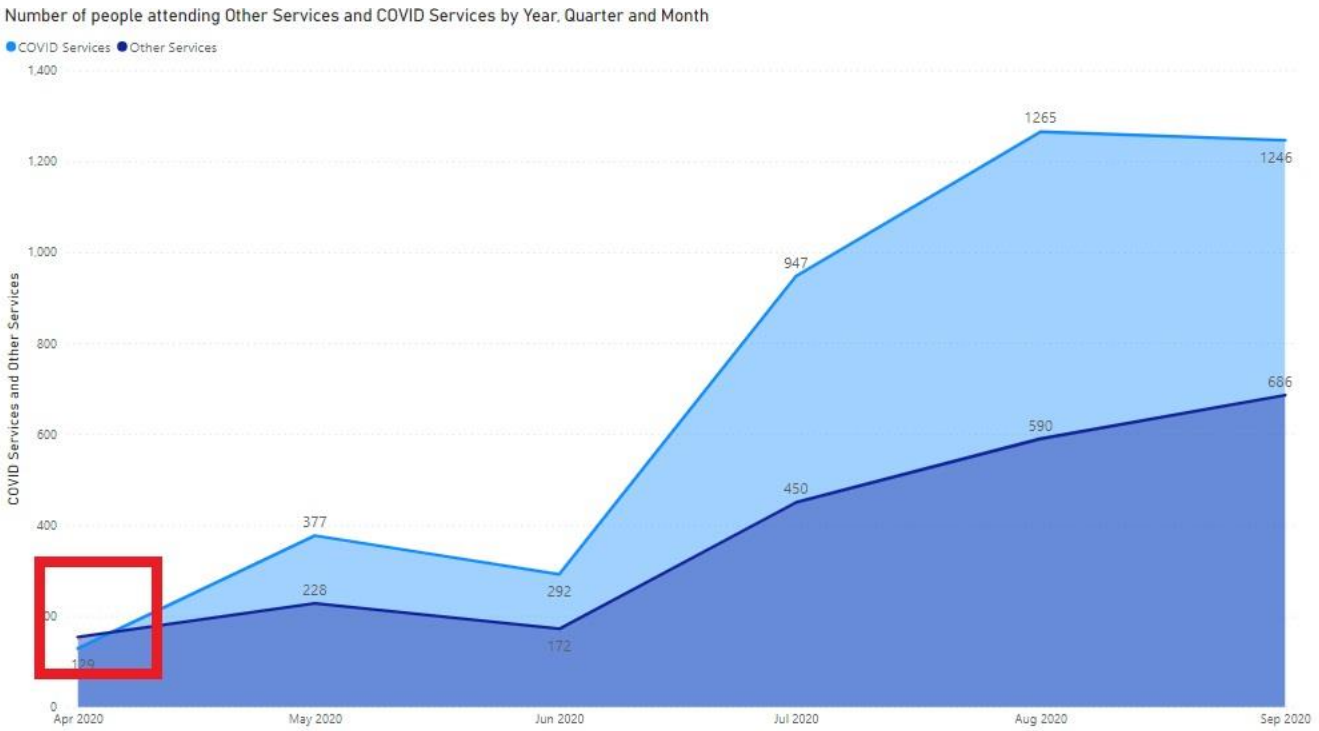

Figure 24 highlights the number of hours consumed by content type.

In the eighth report, we can notice in figure 24**,** initially, the uptake in other services and COVID-19 services was the same. But over time, the number of people attending COVID-19 services grew drastically. By the end of September 2020, twice the number of people was attending COVID-19 services.

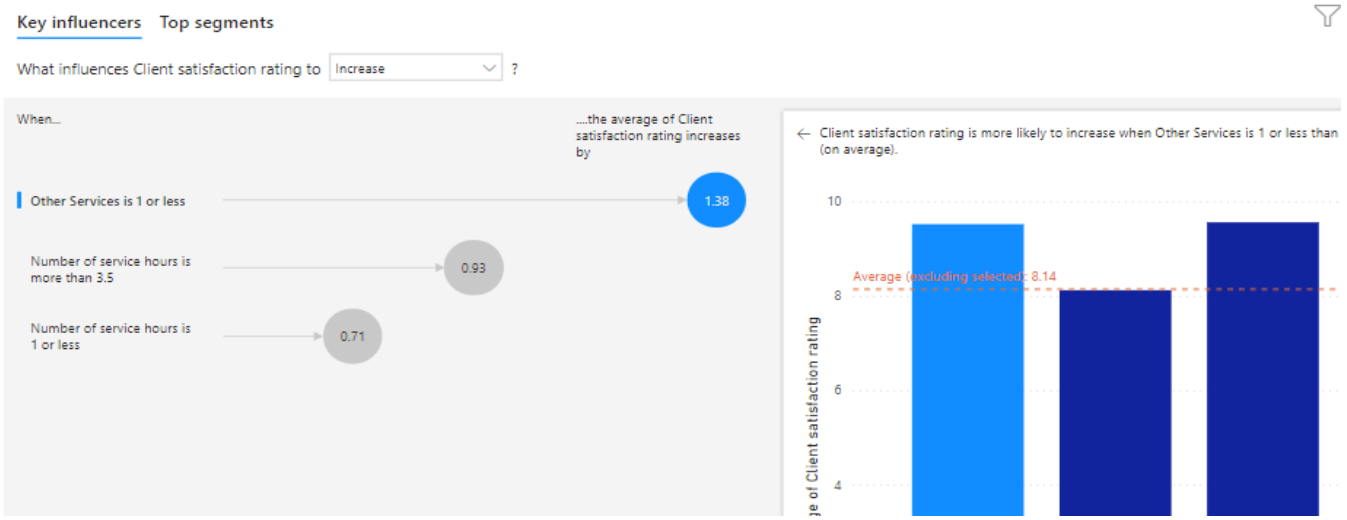

Figure 25 highlights the influences of content type on client satisfaction rating.

In the ninth report, figure 25 highlights the influence of content type on client satisfaction rating. It uncovers the fact that client satisfaction rating increases when other services are 1 or less and 5 or more. Similarly, the overall impact of the BRP support services can be observed.

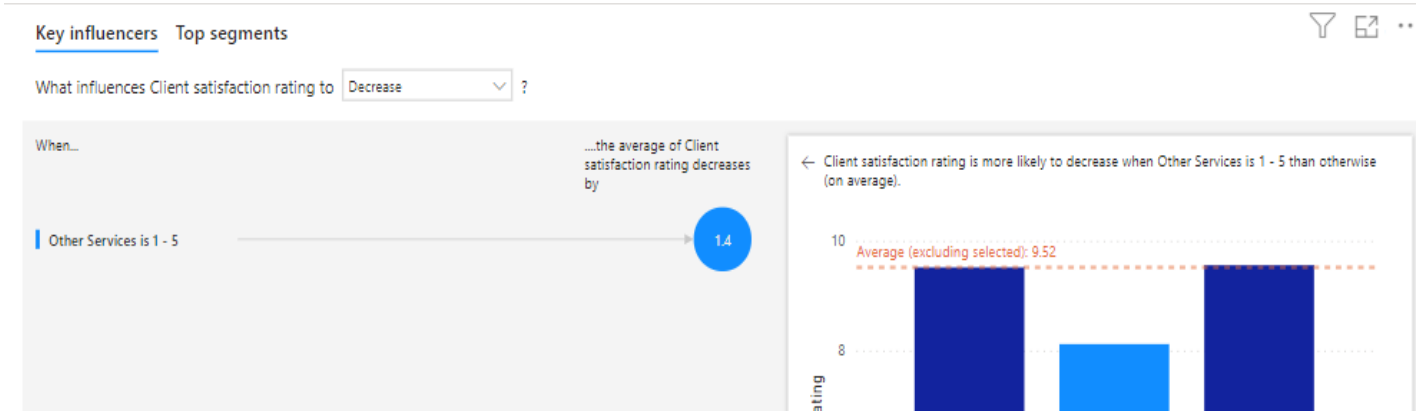

Figure 26 highlights the negative impact of content type on client satisfaction rating.

We can also check which parameters can negatively impact the ratings. As we can in notice figure 26 if other services are between 1-5, rating decreases.

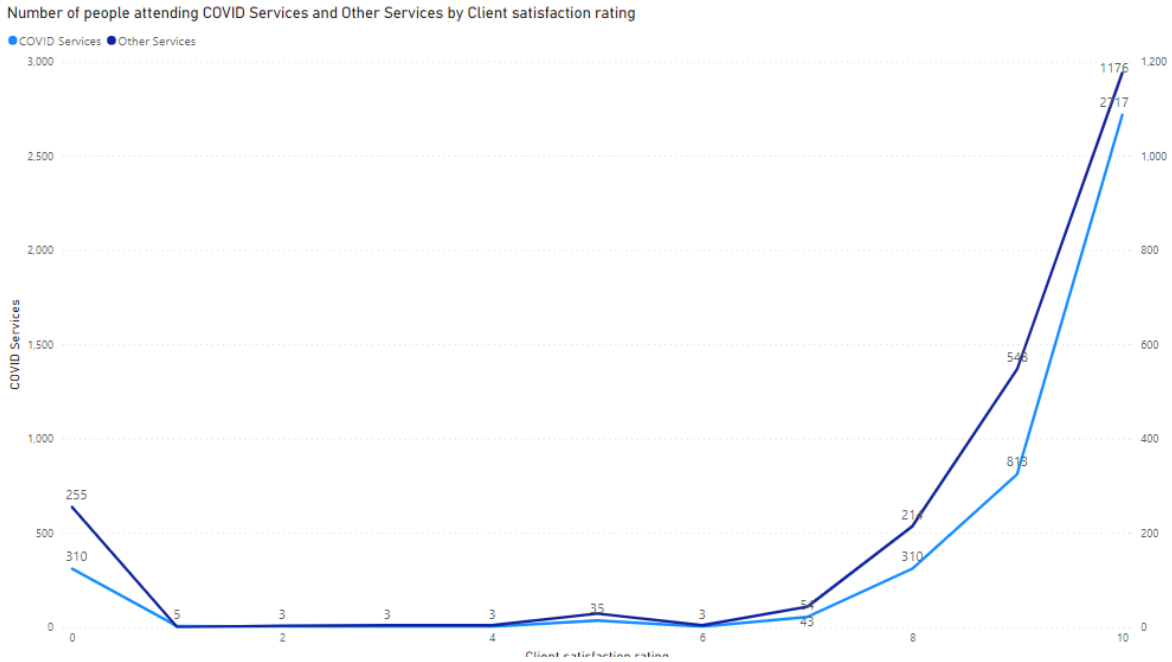

Figure 27 highlights the number of people attending COVID-19 and Other services by client satisfaction rating.

In the tenth report, in figure 27 we can notice the number of service hours by client satisfaction rating. We can notice in figure 27 that only a few people attended services with rating 1, 2, 3, 4,

5 and 6. Most people gave no rating or stayed in the range of 7-10. A few people gave 1-6 ratings. So these are the outliers and values of interest as they can provide useful information on how to improve the services and get insights into small business challenges and needs.

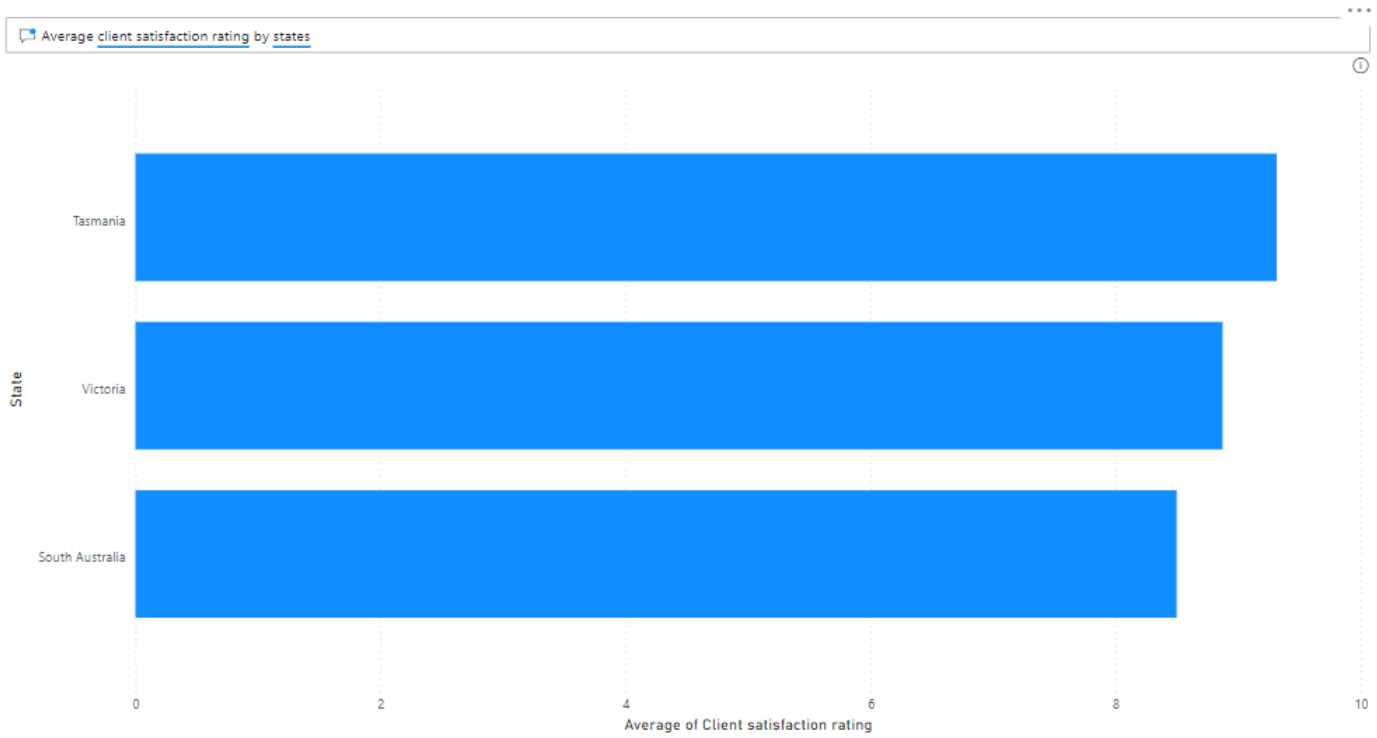

Figure 28 highlights the Q&A chatbox which is answering an average of ratings by states.

The Q & A visual uses natural language processing to make reports instantly. In figure 28, we can notice that it highlights average client satisfaction rating by states.

Figure 29 suggestions by Q & A chat box

As describes in figure 29, it also provides suggestions of similar visuals that can be used with the words written in the chatbox.

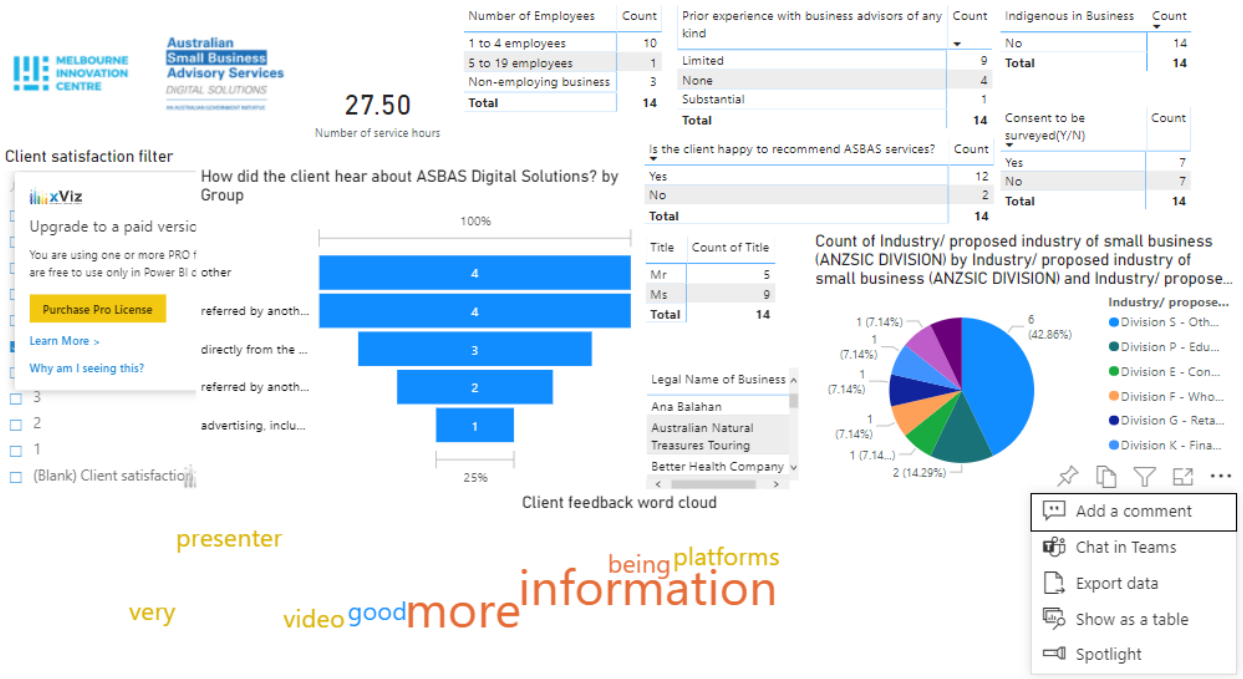

Figure 30 selecting client rating 5 and highlighting data as a table option.

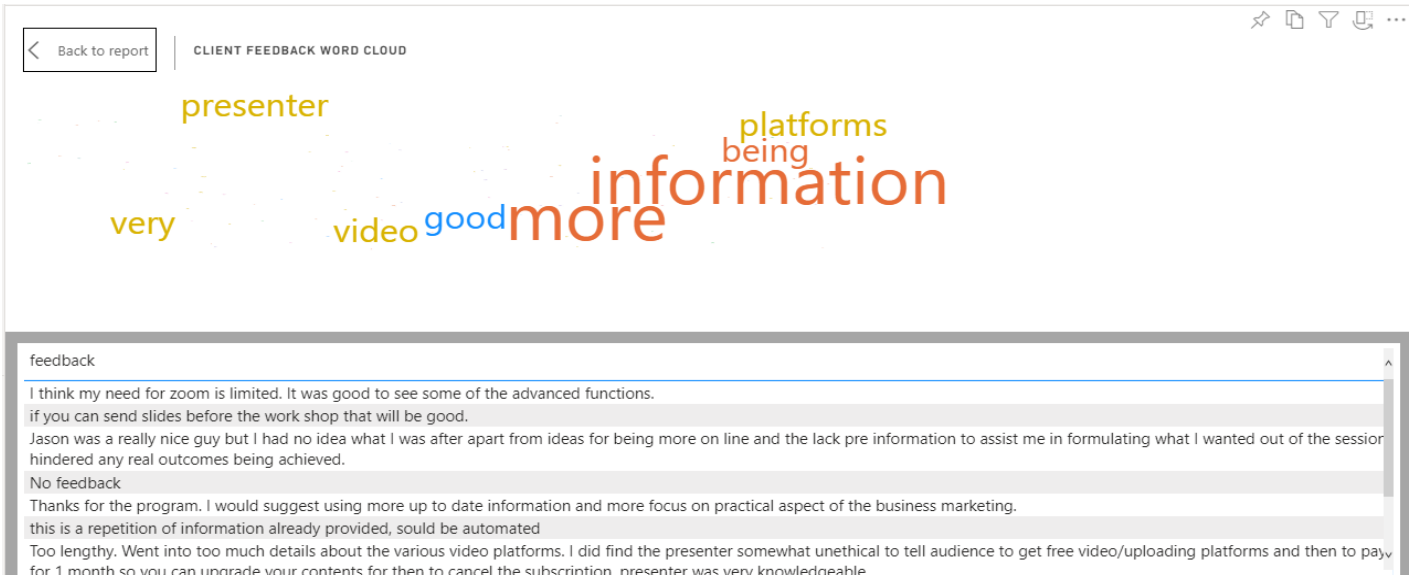

Figure 31 table highlighting feedback in detail.

The last report is a fun visual. We can find the feedback provided by clients by a cloud visual. The visual highlights a cloud of words used in the feedback note. In figure 30 and 31 we can notice that if we filter by 5 ratings, the names of clients as well as their business names can be filtered. We can check the feedback in detail by looking at it in a tabular form.

These reports can help MIC understand the uptake in regional v/s metro and decide future support services. There are some limitations like there are more than 1000 indigenous businesses in Victoria, South Australia and Tasmania (*Supply Nation, 2021*). How to reach all of them? The project covered only 8 indigenous businesses. Also, there is a gap in the market reached by MIC and the actual market in a state. For example, only 8 indigenous businesses participated in the BRP program. In actuality, there were many more to cover. Similarly were all businesses in Construction, education, agriculture, retail, etc reached?

4.0 Further work

The success of the BRP project depends on two main parameters: client satisfaction and the number of service hours. AI or machine learning models can be used further to predict the consumption is services in demand and help the organization manage its resources *(Rajesh,, Bharadwaj, Alam & Tahernezhadi, 2021)*. It can also be used to predict the need for certain services in specific regions and how will it benefit in terms of increased profits for participating organization. Using the results, further work can be done to reach more customers and impact more small businesses.

Also, results suggest that some divisions need more focus as they have a big market share in certain areas. These results can be used to handle the future delivery of services. A new marketing strategy can be developed to perform better in terms of people attending workshops and being satisfied with services by MIC.

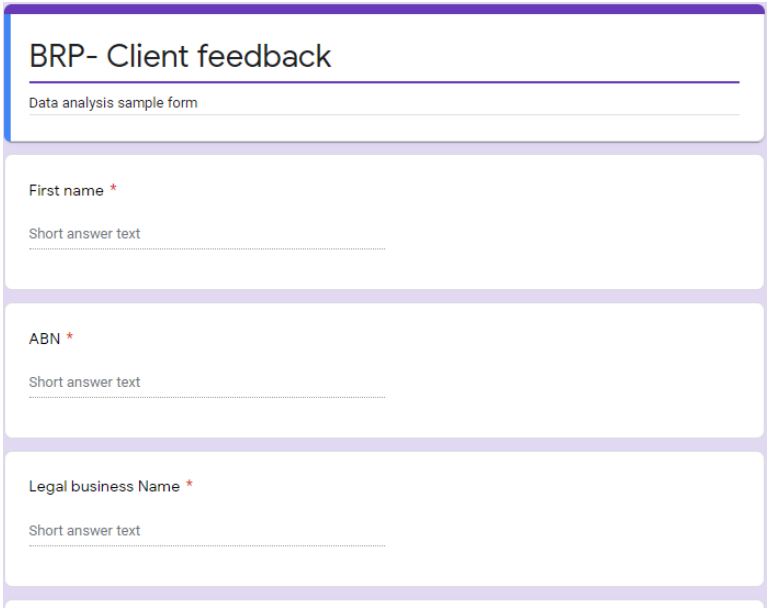

Figure 32 BRP questionnaire using Google forms

During the preprocessing phase, we noticed a lot of problems in the consistency of data. We recommended using Google forms for data collecting in future as it restricts the user filling forms to selecting only available values. In figure 32 we can notice that this can reduce inconsistencies like ‘Y’ and ‘y’. A demo was illustrated to MIC for future process.

5.0 Conclusion

The project’s work revolves around preprocessing, creating visualization and consulting about data with stakeholders. Finding insights is a very detailed and time-consuming process. This is up to MIC to explore the facts in reports and make use of them. Therefore, only the main aspects have been touched in this project and explained. Indeed, there is a scope of finding more insights in certain areas and reflect on the results.

6.0 References

1. K.D.F. (2018, September 25). *A Brief History of Analytics*. Www.Dataversity.Net. https://www.dataversity.net/brief-history-analytics/#
2. Xello. (2020, June 10). *9 reasons why you should use Power BI for business analytics*. Https://Xo.Xello.Com.Au/. https://xo.xello.com.au/blog/reasons-why-you-should-use-power-bi#:%7E:text=Power%20BI%20offers%20data%20preparation,it%20into%20insights%20more%20easily.
3. Pandey, P. (2020, August 5). *Data Preprocessing : Concepts - Towards Data Science*. Https://Towardsdatascience.Com/. https://towardsdatascience.com/data-preprocessing-concepts-fa946d11c825
4. I.T., N.S., M.W., & S.H. (2021, February 12). *Charting the pandemic: Latest data on the NSW and Victoria outbreaks*. ABC News. https://www.abc.net.au/news/2020-03-17/coronavirus-cases-data-reveals-how-covid-19-spreads-in-australia/12060704?nw=0
5. *Supply Nation | Australia’s largest national directory of Indigenous businesses*. (2021, February 7). Supply Nation. https://supplynation.org.au/
6. D.P.R., B., D.M.A., & D.M.T. (2020). A Data Science Approach to Football Team Player Selection. *2020 IEEE International Conference on Electro Information Technology (EIT)*, 175–183. https://doi.org/10.1109/EIT48999.2020.9208331

7.0 Appendices

Following questions can be answered by the reports in Power BI

- Which divisions are not marketed very well? Which needs focus?
- Which divisions were interested to take up more hours in a certain time frame? Or trends for a division? Which needs to be focused on?

- The number of people attending a service tells which division to focus on and what needs not to be focused on?
- Tells what type of business is popular in each state?
- What to do? Where to focus? Based on the facts how to do marketing can be decided.
- Does the correlation matrix tell which order to follow to get the highest satisfaction rating? Which services to prioritize while offering to customers for better satisfaction rates or increased number of hours?
- Which delivery mode is best and worst? What needs to be done to improve it?
- Highlights how involved were Indigenous businesses?
- We can compare the market reached and actual market in a state by researching online: For example, only 8 indigenous businesses participated in the BRP program. In actual, there were many more to cover. Similarly were all businesses in construction, education, agriculture, retail, etc reached? If not how to reach them?
- Can identify the space for improvement?
- What was client feedback? What unhappy clients said?
- Predict the average number of service hours?
- Impact of different services and content type (COVID or non COVID) on satisfaction rating.