Ask Questions to Make Data-Driven Decisions

Step 1: Define the Problem and Business Objective

Scenario

You are working for a nondescript organization within the Triad. It is January 2020. Unbeknownst to you at the time there will be a pandemic worldwide about two months later. Everyone will be meeting remotely for the near future and that includes advising/leader meetings. The organization you are working for is wondering whether they should continue to do only online meetings for advising, should they revert to in person meetings or use a hybrid model once the pandemic is over and people return to site again.

There are Six main different types of problems that we can investigate for the project. However, the problem type that would fit best would be to Categorize things. We are categorizing what members think of the evolving situation.

### **Categorizing things**

**An example of a problem requiring analysts to categorize things is a company's goal to improve customer satisfaction. Analysts might classify customer service calls based on certain keywords or scores. This could help identify top-performing methods of meetings for our organization.**

Ask Three Questions that the organization you are working for on their decision about Remote Meetings:

How do certain demographics affect how people feel about Virtual Advising/meetings?

Does the member’s location affect how members feel about Virtual Advising/meetings?

What kind of levels of difficulty do our members experience when using online services?

Characters/Team Members

It is just you with some help from Jim Pierson.

You will produce a report with the following deliverables:

1. A clear statement of the business task

2. A description of all data sources used

3. Documentation of any cleaning or manipulation of data

4. A summary of your analysis

5. Supporting visualizations and key findings

6. Your top three recommendations based on your analysis

### **Business Objective**

You will help the organization you are working for what to decide what the best course of action is for future advising and meetings (in person only, online only or hybrid).

Prepare Data for Exploration:

Step 2: Prepare the Data

You acquired that data from a survey that you prepared with Jim Pierson’s help. You could not make the survey too long or there will be the possibility that people would not finish it. We need to make sure that our questions follow the SMART outline that is in our notes from the Google Data Analytics Course. SMART Questions include questions that are:

**Specific**- Does our question address the problem in step 1? Does the question have context?

**Measurable**- The question needs to have a measurable answer so that we can compare our results, the results are not only quantitative data, however, quantitative data is more easily measurable than qualitative data and I personally prefer quantitative data to qualitative data for this reason.

**Action-Oriented**- Our questions should inspire action, if our questions do not inspire action but are supplementary and cover topics that we would not need to solve then they would not be extremely helpful to us.

**Relevant**- The question needs to be about the problem we need to solve.

**Time-Bound**- Are our questions relevant to the specific time being studied? It does us not good to answer a question that is out of the scope of the time we want. What good would it do us to solve a problem that is not currently an issue?

What is your sex?

1. Male
2. Female

This is supposed to be specific. It is sex, not gender, therefore there are only two answers for a person to answer. The question’s relevancy could potentially be called into question; however, it is possible that males and females might prefer unique styles of communication on average, so it seems relevant to include it in our survey.

What is your status at the Organization?

1. Full Time All On Site
2. Full Time Hybrid Schedule
3. Full Time All Online
4. Part Time All On Site
5. Part Time Hybrid Schedule
6. Part Time All Online

This question is specific, as includes all types of members that attend the organization they are working for on their location. To make this question more time relevant we should have included an additional question to see if some of these members were leaving soon.

Which Web Conferencing Program do you Prefer the Most?

1. Organization In House Software
2. Microsoft Teams
3. Zoom
4. Skype
5. WebEx

This question is Action-Oriented because it will encourage our organization to seek out a specific platform based on the data. It is relevant because we want to know which platform will be the most preferred or easiest for our members.

Roughly How Often Do You Need to talk to your Leader(s)/Advisor outside of meetings?

1. 3-4 Times a Week
2. 0-2 Times a Week
3. 3-4 Times a Month
4. 3-4 Times a Quarter

This question is Specific as it shows how often our members need to meet with their leaders and advisors outside of meetings. It could be a big clue to whether online meetings would be preferred (the more a member meets with others the more likely online would be better).

In What Way Would You like to be Notified About Meeting Sessions the Most?

1. Smartphone Notifications
2. Emails
3. In Application/Program Announcements

This question is Relevant, being that we can figure out how to contact our members about future updates to their meetings. It can also help inspire the organization to pick one of these services to notify members in the future. However, we should have asked whether the member owns a smartphone in another question just to see if smartphone notifications are limited by the number of non-smartphones owning members.

On a Scale of 1 to 10 (1 Being the Least and 10 Being the Most) How Difficult would it be for You to use Web Conferencing Software to Meet with Your Leader?

This question is the most measurable question in the entire survey. It uses base 10 just so we can get out of 100%.

Would You Rather Meet Your Leader/Adviser Face to Face on Site or Online?

1. Face to Face
2. Online

This question is Specific (as it answers the Business Objective Directly). It is Relevant since it is a binary simple question. Maybe we should have asked this question first?

Do You Agree or Disagree that Web Conferencing is the Future for Our Organization?

1. Strongly Agree
2. Agree
3. Neither Agree nor Disagree
4. Disagree
5. Strongly Disagree

This question could potentially be very measurable. It scalable to a 100% model (Strongly Agree could be 100%, Agree 75%, Neither Agree nor Disagree 50%, Disagree 25%, Strongly Disagree 0%).

What is Your Biggest Concern About using Web Conferencing Software for Meetings and For One-on-One Conferencing with Your Leader?

Free Response

This question is relevant. However, I admit it is a little open ended. Personally, it is good to include at least one open ended question on a survey. We will have to categorize this data using filtering with SQL in Google’s Big Query.

● Where is your data located?

This data is stored on a Microsoft’s OneDrive Server. Data being backed up to the cloud negates the risks of storing on my own desktop.

● How is the data organized?

The Data is organized into a single table. The sample size is small enough and the questions are organized in such a way that there is no need to create a database and normalize the data beyond a single table.

● Are there issues with bias or credibility in this data? Does your data ROCCC? R- Reliable, O- Original, C- Comprehensive, C- Current, C- Cited

The Data is reliable. The data was collected by our organization itself, so it is First-Party Data. We used third party tools to collect the data, but the survey was still conducted our organization. The Data is Original. The data was not copied from another source. The data may or may not be current (we should have asked if the member was planning to be there next quarter). The data does not need to be cited because it is our data. It did not come from a second party or third-party source. The data is owned by our organization.

● How are you addressing licensing, privacy, security, and accessibility?

The data (in the GitHub) Version is anonymous. The person’s email and name will not be published with the final set of data. The data will, however, not completely be randomized and is in fact real. Another thing I did to increase privacy was not publish the name of the organization I worked with. Forsyth Tech wanted us to go out into the community and work for another organization, so I kept their name from being published with the data.

● How did you verify the data’s integrity?

I ran the following code and I got zero results. The thing that would be the likeliest to threaten data integrity would be getting results outside the time I had to collect them (March 16th through April 30th). This shows that all our results fall within the specific time that we need them to.

SELECT \*

FROM `forsyth-capstone.Survey\_Data.Capstone`

WHERE Completion\_Time NOT BETWEEN '0020-03-16 00:00:00' AND '0020-04-30 23:59:59';

● How does it help you answer your question?

The scale of this project is exceedingly small and primary sourced so data integrity should not take long to figure out. I know that all the results were in the right time and all from our organization.

● Are there any problems with the data? Key tasks

1. Download data and store it appropriately.

2. Identify how it is organized.

3. Sort and filter the data.

Process Data from Dirty to Clean

Step 3: Process the Data

● What tools are you choosing and why?

The scale of this project is incredibly small. I could use Excel, but I would rather use Big Query and R Programming to find the visual solutions to my problem.

● Have you ensured your data’s integrity?

I made sure the data is Reliable and Current with [my code](#Verify_Data) in the previous step. This along with the fact the data is from a primary source shows that the data has integrity.

● What steps have you taken to ensure that your data is clean?

I ran queries in the previous step to make sure that the data lined up as part of ROCCC. The only character in the acronym that the data has not been checked against is Cited. There is no need to Cite the data since it was all taken from a primary source.

● How can you verify that your data is clean and ready to analyze?

Running queries to make sure that the data was current and that only members filled out the survey helped in making sure that the data is ready to analyze. There were not many different columns that needed filtering with this project. I could not include the email query in the capstone because it would reveal which organization I was working with (with the organization header in the email). I also added the Biggest\_Category\_Concern column manually to clean up the data. It is a label based on the Biggest Concern Colum and the label had to be inferred from the Biggest Concern Column contents since the Biggest\_Concern Column is unstructured data.

● Have you documented your cleaning process so you can review and share those results

[See my code above.](#Verify_Data)

Figuring Out Which Columns are Categories and Which are Subcategories:

[This is our table of “Categories” and “Subcategories”.](#Categories_And_Subcategories) The Categories are measurements that can be measured independently of subcategories, and you can use aggregate functions with them (Sum, Average, Min, Max, Median or Mode). They still may need to be split up based on things like time (however, time is not supposed to be a subcategory) The Subcategories are the measurements that being measured independently of categories or other subcategories can become a problem, is not descriptive enough or presents a noticeably clear picture of the data.

Subcategories always must be paired with a category or another subcategory. Categories can be independent of subcategories, or they can be paired with other Categories or Subcategories (but it is not necessary to pair them to get a chart/measurement from them or use an aggregate function with them).

Things such as map coordinates, geographical locations or id columns are usually subcategories. IDs cannot be added or subtracted (using aggregate functions on them is meaningless) and max, min, median, mode and mean would be meaningless to get from ID columns too. You can use count or count distinct, but not much more than that. This is also true for latitude and longitude values.

Calculating the Sample Size:

To keep the organization I collaborated with private I am not going to list how many people are in the organization. However, we can still calculate the accuracy with the sample size we have and know up to how many people we can have in our organization with the confidence level and margin of error we want. It is possible to calculate this manually, but we will opt for using the Sample Size Calculator offered by Google.

Our Sample Size is 57 people we want a Confidence Level of 90% and a Margin of Error of 10%. When we plug these numbers into our Sample Size Calculator, we put in 320 people for the population size to get a sample size of 320 people. So, with our collected sample we can say with a 90% level of confidence within a 10% margin of error for a Population Size of 320 people. If we want a Confidence Level of 95% with a 10% margin of error with our Sample Size, the organization will have 135 people. If we want a Confidence Level of 99% with a 10% Margin of Error, we can use this Sample (57 People) for an organization with a Population Size of 85 people.

Analyze Data to Answer Questions

Step 4: Analyze the Data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Categories | Category Notes | Subcategories |  | Unstructured Data |
| Perceived Difficulty of Software | Perceived Difficulty of the Software can have a meaningful average, median and mode taken from the data without using another category or subcategory as a reference. This data is scalar. |  |  | Biggest Concern |
| Agreement on Web Future | Agreement on the Web's Future can have a meaningful average, median and mode taken from the data without using another category or subcategory as a reference. This data is scalar once it has been paired with equivalent values (100%, 75%, 50%, 25%, 0%). |  |  |  |
|  |  | Sex |  |  |
|  |  | Status (at Organization) |  |  |
|  |  | Preferred Software |  |  |
|  |  | Visit Status |  |  |
|  |  | Notification Type |  |  |
|  |  | Leader Interaction |  |  |
|  |  | Biggest Concern Category |  |  |

|  |  |  |
| --- | --- | --- |
| Top Questions | Measurements |  |
| How does Sex of the Member Affect the Biggest Concern of the Member? | Count (Sex) | Group By (Biggest Concern Category) |
| How does the Sex of the Member Affect the Average Perceived Difficulty of using the Software? | Group By (Sex) | Average(Difficulty) |
| How does the Status of the Member Affect the Biggest Concern of the Member? | Count (Status) | Filter (Status) |
| How does the Status of the Member Affect the Perceived Difficulty of using the Software? | Group By (Status) | Average (Perceived  Difficulty) |
| How does the Location of the Member Affect the Biggest Concern of the Member? | Group By (Status) | Biggest Concern Category |
| How does the Location of the Member Affect the Perceived Difficulty of using the Software? | Group By (Status) | Average(Perceived Difficulty) |
| Perceived Difficulty of the Software Average | Average (Perceived Difficulty) |  |
| Perceived Difficulty Median | Median (Perceived Difficulty) |  |
| Perceived Difficulty Mode | Mode (Perceived Difficulty) |  |
| How does the Meeting Frequency Affect the Perceived Difficulty | Group By (Meeting Frequency) | Average(Perceived Difficulty) |

● How should you organize your data to perform analysis on it?

I organized my data based on the [chart above](#Top_Questions_For_Analysis).

[How does Sex of the Member Affect the Biggest Concern of the Member?](#Males_And_Females_Difference)

SELECT COUNT(M\_Or\_F), Biggest\_Concern\_Category

FROM `forsyth-capstone.Survey\_Data.Capstone`

WHERE M\_Or\_F = "Male"

GROUP BY Biggest\_Concern\_Category;

SELECT COUNT(M\_Or\_F), Biggest\_Concern\_Category

FROM `forsyth-capstone.Survey\_Data.Capstone`

WHERE M\_Or\_F = "Female"

GROUP BY Biggest\_Concern\_Category.

This gives us this chart in Google Sheets.

|  |  |  |
| --- | --- | --- |
| Sex vs. Biggest Concern | Male | Female |
| Privacy | 1 | 0 |
| Schedule | 1 | 0 |
| None | 6 | 12 |
| Technical Difficulties | 4 | 7 |
| Materials | 2 | 1 |
| Speed | 3 | 0 |
| Security | 1 | 0 |
| Learning Style | 2 | 7 |
| Availability | 4 | 4 |
| Costs | 1 | 0 |
| Null | 0 | 1 |
| Total | 25 | 32 |

This is R code to get our visuals for this question:

install.packages("tidyverse")

install.packages("ggplot2")

install.packages("LessR")

library(ggplot2)

# Pie Chart with Percentages

slices <- c(4, 4, 24, 16, 8, 12, 4, 8, 16, 4)

lbls <- c("Privacy", "Schedule", "None", "Technical Difficulties", "Materials", "Speed", "Security", "Learning Style", "Availability", "Costs")

pct <- round(slices/sum(slices)\*100)

lbls <- paste(lbls, pct) # add percents to labels

lbls <- paste( lbls, "%", sep="") # ad % to labels

pie(slices,labels = lbls, col=rainbow(length(lbls)),

main="Male Biggest Concerns")

# Pie Chart with Percentages

slices <- c(12, 7, 1, 7, 4, 1)

labels <- c("None", "Technical Difficulties", "Materials", "Learning Style", "Availability", "Null")

pct <- round(slices/sum(slices)\*100)

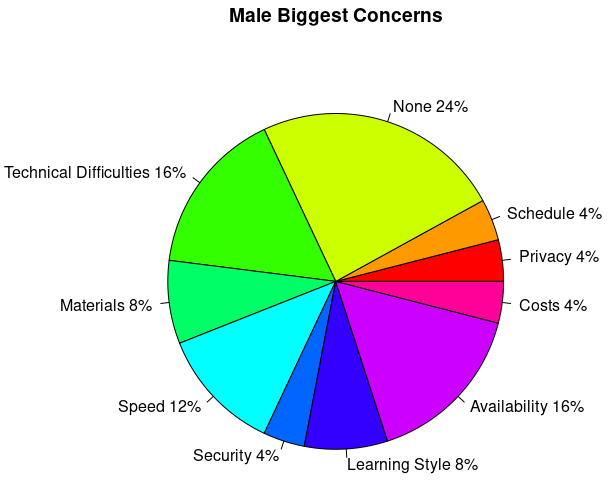
lbls <- paste(labels, pct) # add percents to labels

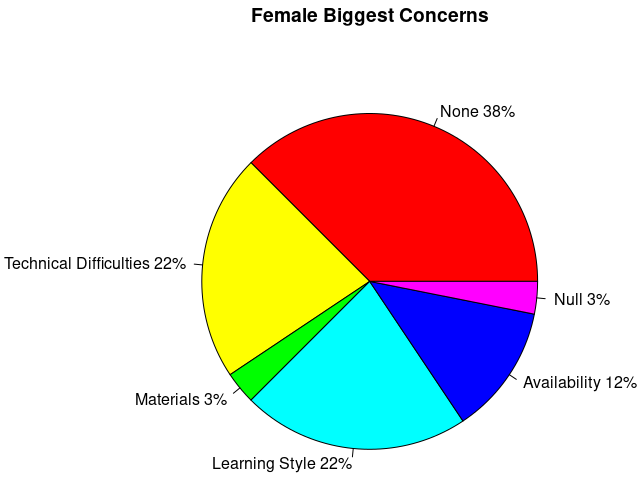
lbls <- paste(labels, "%", sep="") # ad % to labels

pie(slices, labels = lbls, col=rainbow(length(lbls)),

main="Female Biggest Concerns")

This is visuals created from [the R Code](#R_Code_Male_And_Female):





According to the Data, Technical Difficulties, Availability and Learning Style were the towards the top of the concerns for Both Males and Females. Things such as costs, privacy, security, materials, speed and schedule were minor concerns. This shows that the environment and ease of use is considered the most important thing about connecting remotely for the organization.

If members do not feel like they can understand how the software works, the software is inaccessible or if they feel disconnected from their leaders because they are not in the room with them, they will not go along with meetings with their leaders remotely. These are major concerns for both Male sand Females.

The only part where Males and Females differed in a large way was about learning styles. Females were much more concerned about learning styles than Males. Perhaps Men are not as concerned about being in the room with their leader or other members. However, regardless, the environment feel of the software plays a significant role in how people feel about online meetings.

[How does the Sex of the Member Affect the Average Perceived Difficulty of using the Software?](#Males_vs_Females_Statistical_Difference)

SELECT AVG(Difficulty)

FROM `forsyth-capstone.Survey\_Data.Capstone`

WHERE M\_Or\_F = "Male"

GROUP BY M\_Or\_F;

= 3.08

This code gives us the Average perceived difficulties of Males.

SELECT Difficulty

FROM `forsyth-capstone.Survey\_Data.Capstone`

WHERE M\_Or\_F = "Male"

ORDER BY Difficulty;

|  |  |
| --- | --- |
| Row | Difficulty |
| 1 | 1 |
| 2 | 1 |
| 3 | 1 |
| 4 | 1 |
| 5 | 1 |
| 6 | 1 |
| 7 | 1 |
| 8 | 1 |
| 9 | 1 |
| 10 | 1 |
| 11 | 1 |
| 12 | 1 |
| 13 | 1 |
| 14 | 1 |
| 15 | 2 |
| 16 | 3 |
| 17 | 3 |
| 18 | 4 |
| 19 | 4 |
| 20 | 5 |
| 21 | 7 |
| 22 | 8 |
| 23 | 8 |
| 24 | 9 |
| 25 | 10 |

= 1

The Median of the Difficulty for Males is an average between the two middle rows (since there is an odd number of rows). The Median is a value of 1.

SELECT Difficulty, COUNT(Difficulty)

FROM `forsyth-capstone.Survey\_Data.Capstone`

WHERE M\_Or\_F = "Male"

Group By Difficulty

ORDER BY Difficulty;

|  |  |
| --- | --- |
| Difficulty | Count |
| 1 | 14 |
| 2 | 1 |
| 3 | 2 |
| 4 | 2 |
| 5 | 1 |
| 7 | 1 |
| 8 | 2 |
| 9 | 1 |
| 10 | 1 |

= 1

The Mode is the value that occurs the most in the data. The mode of the Male’s Difficulty is 1.

The Mean of Males Difficulty is 3.08. The Median is 1. The Mode is 1. This shows that the data is skewed towards the lower end of the difficulty. Males seem to not have major problems with the software and would adapt well to using it (if they are understanding matches how they perceive it).

SELECT AVG(Difficulty)

FROM `forsyth-capstone.Survey\_Data.Capstone`

WHERE M\_Or\_F = "Female"

GROUP BY M\_Or\_F;

= 3.4374999

This code gives us the Average perceived difficulties of Females.

SELECT Difficulty

FROM `forsyth-capstone.Survey\_Data.Capstone`

WHERE M\_Or\_F = "Female"

ORDER BY Difficulty;

|  |  |
| --- | --- |
| Row | Difficulty |
| 1 | 1 |
| 2 | 1 |
| 3 | 1 |
| 4 | 1 |
| 5 | 1 |
| 6 | 1 |
| 7 | 1 |
| 8 | 1 |
| 9 | 1 |
| 10 | 1 |
| 11 | 1 |
| 12 | 1 |
| 13 | 1 |
| 14 | 1 |
| 15 | 1 |
| 16 | 2 |
| 17 | 2 |
| 18 | 2 |
| 19 | 3 |
| 20 | 3 |
| 21 | 3 |
| 22 | 3 |
| 23 | 3 |
| 24 | 4 |
| 25 | 6 |
| 26 | 8 |
| 27 | 8 |
| 28 | 8 |
| 29 | 10 |
| 30 | 10 |
| 31 | 10 |
| 32 | 10 |

= 2

The Median of difficulty for Females is 2.

SELECT Difficulty, COUNT(Difficulty)

FROM `forsyth-capstone.Survey\_Data.Capstone`

WHERE M\_Or\_F = "Female"

Group By Difficulty

ORDER BY Difficulty;

|  |  |
| --- | --- |
| Difficulty | Count |
| 1 | 15 |
| 2 | 3 |
| 3 | 5 |
| 4 | 1 |
| 6 | 1 |
| 8 | 3 |
| 10 | 4 |

= 1

The Mode is the value that occurs the most in the data. The Mode of the Female’s Difficulty is 1.

The Mean of Female’s Difficulty is 3.4374999. The Median is 2. The Mode is 1. This shows that the data is skewed towards the lower end of the difficulty for Females.

We will run a Statistical Hypothesis Test on the two means from our samples to determine whether there is a statistical difference between the average difficulties of Males and Females. We’ll calculate our Standard Deviations for our Samples and we’ll plug them into an online calculator at <https://infrrr.com/means/difference-in-means-hypothesis-test-calculator>.

A white paper with black writing

Description automatically generated with low confidence

A picture containing table

Description automatically generated

Table

Description automatically generated

A white paper with black writing

Description automatically generated with low confidence

= 194.6749

A white paper with black writing

Description automatically generated with low confidence

Text, letter

Description automatically generated

This gives us the Standard Deviation for the Male Sample. We will plug this into the Calculator online. Next, We will calculate the Standard Deviation for the Female Sample.

A white paper with black writing

Description automatically generated with low confidence

Table

Description automatically generated

Text

Description automatically generated

A white paper with black writing

Description automatically generated with low confidence

A picture containing text

Description automatically generated

Text, letter

Description automatically generated

Text

Description automatically generated with medium confidence

We plug these Standard Deviations into the Online Calculator at <https://infrrr.com/means/difference-in-means-hypothesis-test-calculator>. We get this as our answer.

**DO NOT REJECT NULL HYPOTHESIS**

**Null Hypothesis:** μ1 - μ2 = 0

**Alternative Hypothesis:** μ1 - μ2 ≠ 0

You would expect to see a difference between the sample means as extreme as -0.3575 66.28% of the time under the null hypothesis.

Since 66.28% is greater than the significance level of 1.00%, you do not have sufficient evidence to reject the null hypothesis.

This shows that based on the two samples we have of the entire population at our organization there is no statistical difference between the populations’ averages. This means that sex does not affect the populations’ average difficulty at a level that can be measured statistically. This is true at a significance level of 1%.

[How Does the Status of the Member Affect the Biggest Concern of the Member?](#Biggest_Concerns_Of_Demographics)

We will measure this question by proportion instead of pure numbers because there is a significant difference between totals of different statuses of our members.

This SQL code gives us the count of the total number of Full Time Hybrid Schedule members

SELECT COUNT(Status)

FROM `forsyth-capstone.Survey\_Data.Capstone`

WHERE Status = "Full Time All On Site";

= 11

SELECT COUNT(Status)

FROM `forsyth-capstone.Survey\_Data.Capstone`

WHERE Status = "Full Time Hybrid Schedule";

= 27

SELECT COUNT(Status)

FROM `forsyth-capstone.Survey\_Data.Capstone`

WHERE Status = "Full Time All Online";

= 6

SELECT COUNT(Status)

FROM `forsyth-capstone.Survey\_Data.Capstone`

WHERE Status = "Part Time All On Site";

= 1

(Hybrid is misspelled as “hyrid”)

SELECT COUNT(Status)

FROM `forsyth-capstone.Survey\_Data.Capstone`

WHERE Status = "Part Time Hyrid Schedule";

= 5

SELECT COUNT(Status)

FROM `forsyth-capstone.Survey\_Data.Capstone`

WHERE Status = "Part Time All Online";

= 7

|  |  |  |
| --- | --- | --- |
| **Status** | **Biggest\_Concern\_Category** | **Percentage** |
| Part Time All On Site | None | 100 |
| Full Time All Online | None | 50 |
| Part Time All Online | None | 42.86 |
| Full Time All On Site | Learning Style | 27.27 |
| Full Time All On Site | Technical Difficulties | 27.27 |
| Part Time Hyrid Schedule | None | 27.27 |
| Full Time Hybrid Schedule | None | 25.93 |
| Full Time Hybrid Schedule | Technical Difficulties | 22.22 |
| Full Time Hybrid Schedule | Learning Style | 22.22 |
| Full Time Hybrid Schedule | Availability | 18.52 |
| Full Time All Online | Security | 16.67 |
| Full Time All Online | Availability | 16.67 |
| Full Time All Online | Technical Difficulties | 16.67 |
| Part Time All Online | Schedule | 14.29 |
| Part Time All Online | Availability | 14.29 |
| Part Time All Online | Costs | 14.29 |
| Part Time All Online | Materials | 14.29 |
| Full Time All On Site | Privacy | 9.09 |
| Full Time All On Site | None | 9.09 |
| Full Time All On Site | Materials | 9.09 |
| Full Time All On Site | Null | 9.09 |
| Full Time All On Site | Availability | 9.09 |
| Part Time Hyrid Schedule | Technical Difficulties | 9.09 |
| Part Time Hyrid Schedule | Speed | 9.09 |
| Full Time Hybrid Schedule | Speed | 7.41 |
| Full Time Hybrid Schedule | Materials | 3.7 |

Based on proportions, the top 3 results are None. All statuses except ‘Full Time All On Site’ had a none result within the top 10 results by proportion. Technical Difficulties and Learning Styles were both twice in the Top Ten Results. Availability was also in the top 10. None of the other problems were in the top 10. This ties in with what I stated in the first question. General Problems or Over Aching Problems, such as Availability, Technical Difficulties and Learning Styles were the highest concerns. More specific problems such as Speed, Materials, Costs, Privacy and Schedule were not the highest concerns of any of the statuses.

|  |  |  |  |
| --- | --- | --- | --- |
| **Status** | **Biggest\_Concern\_Category** | **Count** | **Percentage** |
| Full Time All Online | None | 3 | 50 |
| Full Time All On Site | Learning Style | 3 | 27.27 |
| Full Time All On Site | Technical Difficulties | 3 | 27.27 |
| Full Time Hybrid Schedule | None | 7 | 25.93 |
| Full Time Hybrid Schedule | Technical Difficulties | 6 | 22.22 |
| Full Time Hybrid Schedule | Learning Style | 6 | 22.22 |
| Full Time Hybrid Schedule | Availability | 5 | 18.52 |
| Full Time All Online | Security | 1 | 16.67 |
| Full Time All Online | Availability | 1 | 16.67 |
| Full Time All Online | Technical Difficulties | 1 | 16.67 |
| Full Time All On Site | Privacy | 1 | 9.09 |
| Full Time All On Site | None | 1 | 9.09 |
| Full Time All On Site | Materials | 1 | 9.09 |
| Full Time All On Site | Null | 1 | 9.09 |
| Full Time All On Site | Availability | 1 | 9.09 |
| Full Time Hybrid Schedule | Speed | 2 | 7.41 |
| Full Time Hybrid Schedule | Materials | 1 | 3.7 |

Full Time members were most concerned with Learning Styles and Technical Difficulties or not at all. Once again, overarching problems, not extremely specific problems.

|  |  |  |  |
| --- | --- | --- | --- |
| **Status** | **Biggest\_Concern\_Category** | **Count** | **Percentage** |
| Part Time All On Site | None | 1 | 100 |
| Part Time All Online | None | 3 | 42.86 |
| Part Time Hyrid Schedule | None | 3 | 27.27 |
| Part Time All Online | Schedule | 1 | 14.29 |
| Part Time All Online | Availability | 1 | 14.29 |
| Part Time All Online | Costs | 1 | 14.29 |
| Part Time All Online | Materials | 1 | 14.29 |
| Part Time Hyrid Schedule | Technical Difficulties | 1 | 9.09 |
| Part Time Hyrid Schedule | Speed | 1 | 9.09 |

Part Time Members seem to have less issues with the software than Full Time Members. All three of the top concerns were None.

|  |  |  |  |
| --- | --- | --- | --- |
| **Status** | **Biggest\_Concern\_Category** | **Count** | **Percentage** |
| Part Time All On Site | None | 1 | 100 |
| Full Time All On Site | Learning Style | 3 | 27.27 |
| Full Time All On Site | Technical Difficulties | 3 | 27.27 |
| Full Time All On Site | Privacy | 1 | 9.09 |
| Full Time All On Site | None | 1 | 9.09 |
| Full Time All On Site | Materials | 1 | 9.09 |
| Full Time All On Site | Null | 1 | 9.09 |
| Full Time All On Site | Availability | 1 | 9.09 |

All On Site Members were concerned with Technical Difficulties, Learning Styles or not at all.

|  |  |  |  |
| --- | --- | --- | --- |
| **Status** | **Biggest\_Concern\_Category** | **Count** | **Percentage** |
| Part Time Hyrid Schedule | None | 3 | 27.27 |
| Full Time Hybrid Schedule | None | 7 | 25.93 |
| Full Time Hybrid Schedule | Technical Difficulties | 6 | 22.22 |
| Full Time Hybrid Schedule | Learning Style | 6 | 22.22 |
| Full Time Hybrid Schedule | Availability | 5 | 18.52 |
| Part Time Hyrid Schedule | Technical Difficulties | 1 | 9.09 |
| Part Time Hyrid Schedule | Speed | 1 | 9.09 |
| Full Time Hybrid Schedule | Speed | 2 | 7.41 |
| Full Time Hybrid Schedule | Materials | 1 | 3.7 |

Hybrid Members were concerned with Technical Difficulties, Learning Styles or not at all.

|  |  |  |  |
| --- | --- | --- | --- |
| **Status** | **Biggest\_Concern\_Category** | **Count** | **Percentage** |
| Full Time All Online | None | 3 | 50 |
| Part Time All Online | None | 3 | 42.86 |
| Full Time All Online | Security | 1 | 16.67 |
| Full Time All Online | Availability | 1 | 16.67 |
| Full Time All Online | Technical Difficulties | 1 | 16.67 |
| Part Time All Online | Schedule | 1 | 14.29 |
| Part Time All Online | Availability | 1 | 14.29 |
| Part Time All Online | Costs | 1 | 14.29 |
| Part Time All Online | Materials | 1 | 14.29 |

Online only Members had no major concern.

Common Patterns for all the different ways to look at the data shows that people are either not concerned or they are concerned about Technical Difficulties or the Learning Style that the meeting takes on when people are learning online. Members that are currently all online do not have any major concerns about the software in their top results. Part Time Members also do not have major concerns about software as well. This shows that people who have transitioned to an online presence only have gotten acquainted with the software well. It will be difficult for some members to get adjusted, but it does seem that transitioning is possible for most members. There is, however, the possibility that members who were already technologically savvy decided to learn online.

[How Does the Status of the Member Affect the Perceived Difficulty of using the Software?](#Status_Of_Member_Average_Difficulty)

SELECT Status, AVG(Difficulty) AS Average\_Difficulty

FROM `forsyth-capstone.Survey\_Data.Capstone`

GROUP BY Status;

|  |  |
| --- | --- |
| **Status** | **Average Difficulty** |
| Part Time All On Site | 10 |
| Full Time All On Site | 4.272727273 |
| Full Time Hybrid Schedule | 3.407407407 |
| Full Time All Online | 3 |
| Part Time All Online | 2 |
| Part Time Hybrid Schedule | 1.2 |

The lowest difficulty averages belonged to members who were already using online meetings as either as a hybrid system or are already fully online. On Site members scored the highest average for difficulty. Part Time All On Site scored especially high.

[How does the Location of the Member Affect the Biggest Concern of the Member?](#Location_And_Biggest_Concern)

|  |  |  |  |
| --- | --- | --- | --- |
| **Status** | **Biggest\_Concern\_Category** | **Percentage** | **Count** |
| Part Time All On Site | None | 100 | 1 |
| Full Time All On Site | Learning Style | 27.27 | 3 |
| Full Time All On Site | Technical Difficulties | 27.27 | 3 |
| Full Time All On Site | Privacy | 9.09 | 1 |
| Full Time All On Site | None | 9.09 | 1 |
| Full Time All On Site | Materials | 9.09 | 1 |
| Full Time All On Site | Null | 9.09 | 1 |
| Full Time All On Site | Availability | 9.09 | 1 |

On site Members seemed to be most concerned about Learning Styles and Technical Difficulties or not at all.

|  |  |  |  |
| --- | --- | --- | --- |
| **Status** | **Biggest\_Concern\_Category** | **Percentage** | **Count** |
| Full Time All Online | None | 50 | 3 |
| Part Time All Online | None | 42.86 | 3 |
| Full Time All Online | Security | 16.67 | 1 |
| Full Time All Online | Availability | 16.67 | 1 |
| Full Time All Online | Technical Difficulties | 16.67 | 1 |
| Part Time All Online | Schedule | 14.29 | 1 |
| Part Time All Online | Availability | 14.29 | 1 |
| Part Time All Online | Costs | 14.29 | 1 |
| Part Time All Online | Materials | 14.29 | 1 |

Online Members seemed to mostly not be concerned about anything.

|  |  |  |  |
| --- | --- | --- | --- |
| **Status** | **Biggest\_Concern\_Category** | **Percentage** | **Count** |
| Part Time Hyrid Schedule | None | 27.27 | 3 |
| Full Time Hybrid Schedule | None | 25.93 | 7 |
| Full Time Hybrid Schedule | Technical Difficulties | 22.22 | 6 |
| Full Time Hybrid Schedule | Learning Style | 22.22 | 6 |
| Full Time Hybrid Schedule | Availability | 18.52 | 5 |
| Part Time Hyrid Schedule | Technical Difficulties | 9.09 | 1 |
| Part Time Hyrid Schedule | Speed | 9.09 | 1 |
| Full Time Hybrid Schedule | Speed | 7.41 | 2 |
| Full Time Hybrid Schedule | Materials | 3.7 | 1 |

Hybrid Scheduled Members were either not concerned or were concerned with Technical Difficulties, Learning Style or Availability.

On Site Members seemed to be concerned mostly either by Learning Styles, Technical Difficulties or Not at All. Online Members did not seem to be extremely concerned about anything and their top results were all None. Hybrid Members seemed to also be concerned with Learning Styles and Technical Difficulties or Not at All. The two areas that need to be tackled the most are Technical Issues and Learning Styles. These answers are based on proportions and consider the fact that there are a lot of differing numbers of members at separate locations (for example, there are 27 Full Time Hybrid Members while there is only 1 Part Time All On Site Member)

[How does the Location of the Member Affect the Perceived Difficulty of using the Software](#Location_Affect_Perceived_Difficulty)?

SELECT Status, AVG(Difficulty) AS Average\_Difficulty

FROM `forsyth-capstone.Survey\_Data.Capstone`

GROUP BY Status

ORDER BY Average\_Difficulty DESC;

|  |  |
| --- | --- |
| Status | Average\_Difficulty |
| Part Time All On Site | 10 |
| Full Time All On Site | 4.272727273 |
| Full Time Hybrid Schedule | 3.407407407 |
| Full Time All Online | 3 |
| Part Time All Online | 2 |
| Part Time Hybrid Schedule | 1.2 |

Members that are On Site have the largest difficulty with the software. Hybrid and Online Members tend to score lower on average perceived difficulty.

[Perceived Difficulty of the Software Average](#Perceived_Mean)

SELECT AVG(Difficulty)

FROM `forsyth-capstone.Survey\_Data.Capstone`;

= 3.2807017543859645

[Perceived Difficulty Median](#Perceived_Median)

|  |  |
| --- | --- |
| Row | Difficulty |
| 1 | 1 |
| 2 | 1 |
| 3 | 1 |
| 4 | 1 |
| 5 | 1 |
| 6 | 1 |
| 7 | 1 |
| 8 | 1 |
| 9 | 1 |
| 10 | 1 |
| 11 | 1 |
| 12 | 1 |
| 13 | 1 |
| 14 | 1 |
| 15 | 1 |
| 16 | 1 |
| 17 | 1 |
| 18 | 1 |
| 19 | 1 |
| 20 | 1 |
| 21 | 1 |
| 22 | 1 |
| 23 | 1 |
| 24 | 1 |
| 25 | 1 |
| 26 | 1 |
| 27 | 1 |
| 28 | 1 |
| 29 | 1 |
| 30 | 2 |
| 31 | 2 |
| 32 | 2 |
| 33 | 2 |
| 34 | 3 |
| 35 | 3 |
| 36 | 3 |
| 37 | 3 |
| 38 | 3 |
| 39 | 3 |
| 40 | 3 |
| 41 | 4 |
| 42 | 4 |
| 43 | 4 |
| 44 | 5 |
| 45 | 6 |
| 46 | 7 |
| 47 | 8 |
| 48 | 8 |
| 49 | 8 |
| 50 | 8 |
| 51 | 8 |
| 52 | 9 |
| 53 | 10 |
| 54 | 10 |
| 55 | 10 |
| 56 | 10 |
| 57 | 10 |

The Median of All Perceived Difficulties is the average of row 28 and 29 (since there is an odd number of rows). The Median Value is 1

[Perceived Difficulty Mode](#Perceived_Mode)

SELECT Difficulty, COUNT(\*)

FROM `forsyth-capstone.Survey\_Data.Capstone`

GROUP BY Difficulty

ORDER BY Difficulty;

|  |  |
| --- | --- |
| Difficulty | Count |
| 1 | 29 |
| 2 | 4 |
| 3 | 7 |
| 4 | 3 |
| 5 | 1 |
| 6 | 1 |
| 7 | 1 |
| 8 | 5 |
| 9 | 1 |
| 10 | 5 |

The mode is the value that repeats the most. This mode is 1.

[Perceived Mean, Median and Mode of Difficulty](#Perceived_Mean_Median_And_Mode)

The mode is 1 and the Mean is 1. The average for the difficulty is 3.28. This shows that the data is skewed towards the bottom values and is not quite an evenly bell-shaped curve. This shows that the perceived difficulty is fairly low and that a lot of members do already understand the software.

[How does the Meeting Frequency Affect the Perceived Difficulty?](#Meeting_Frequency)

SELECT Meeting\_Frequency, AVG(Difficulty) AS Average\_Difficulty

FROM `forsyth-capstone.Survey\_Data.Capstone`

GROUP BY Meeting\_Frequency

ORDER BY Average\_Difficulty DESC;

|  |  |
| --- | --- |
| Meeting\_Frequency | Average\_Difficulty |
| 3-4 Times a Week | 7.5 |
| 3-4 Times a Quarter | 3.4 |
| 3-4 Times a Month | 2.8 |
| 0-2 Times a Week | 2.707317073 |

Meeting time is not a large factor unless the member is meeting the Leader/Advisor 3 to 4 times a week. This is the greatest number of times a member can meet with their Leader/Advisor outside of organization meetings. If the member meets between 0-2 Times a Week and 3-4 Times a quarter the perceived difficulty increases only slightly.

● What surprises did you discover in the data?

The biggest surprise I had from the data is that a lot of people had no concerns whatsoever about using the software. In every location None was in the top results (On Site, Hybrid and All Online). This shows that a transition may not be as difficult as we think it would be for the members overall.

● What trends or relationships did you find in the data?

The biggest trends in the data showed that Learning Styles and Technical Difficulties were among the highest concerns of members. Current All Online Members stated the most that they did not have any major concerns about the Software.

● How will these insights help answer your business questions?

We need to ask the question “How Tech Savvy are you?” to know whether the All-Online Members opted to be solely online because they are already tech savvy or if they adapted to being online and do not have major concerns because they have gotten used to it. This is the largest question that I have about the data. If we get results that show that Members have become more accustomed to online instead of already being good with technology, then we will know that members are able to adapt to online and it would make the stronger case for implementing an online solution. Technical Difficulties can be addressed by having resources for members, learning styles and availability will be harder to tackle.

Sharing Data Through the Art of Visualization

Step 5: Share

● What story does your data tell?

How do certain demographics affect how people feel about Virtual Advising/meetings?

How Does Sex of the Member Affect the Biggest Concern of the Member?

The only major difference in concern between Males and Females was the Learning Style. There were almost three times as many Females (proportionally) were concerned about Learning Styles than Males were. The rest of the categories were not quite different (proportionally).

How Does Sex of the Member Affect the Average Perceived Difficulty of using the Software?

There is no statistically measurable difference between average Perceived Difficulty of the population of Males and Females based on their sample Averages. The average perceived difficulty is the same for males and females. Sex does not play a role in average perceived difficulty.

How Does the Status of the Member Affect the Biggest Concern of the Member?

The largest concerns, based on status, revolved around Learning Styles and Technical Difficulties for almost all groups. This was consistent across the board.

How Does the Status of the Member Affect the Perceived Difficulty of Using the Software?

The Lowest Difficulty Averages belonged to members who are already using online software for their meetings or are on a hybrid system. On Site members scored the highest on average perceived difficulty. Part Time On Site scored the highest possible average difficulty score that was possible. The real question is, do tech savvy people use online meetings or do online meetings make people more tech savvy?

Does the member’s location affect how members feel about Virtual Advising/meetings?

How Does the Location of the Member Affect the Biggest Concern of the Member?

Online Members do not seem to be extremely concerned about anything while On site members are most concerned with Learning Styles and Technical Difficulties of the software. Hybrid Members seem to be concerned with Learning Style and Technical Difficulties too; however, they are less concerned than On Site Members are about Learning Styles and Technical Difficulties (proportionally).

How Does the Location of the Member Affect the Perceived Difficulty of using the Software?

On Site members have the highest Average Perceived Difficulty of the web software. Hybrids and Online Members have lower Average Perceived Difficulties.

What kind of levels of difficulty do our members experience when using online services?

The average perceived difficulty of all members is 3.28. This is less than half of the scale. The Median is 1 and the Mode is also 1. This shows that the Perceived difficulty is skewed towards the lower side. The curve is not exactly a bell curve, it is weighted towards the lower side (or the 1 on the scale). This shows that the difficulty is not that high for a lot of people.

● How do your findings relate to your original question?

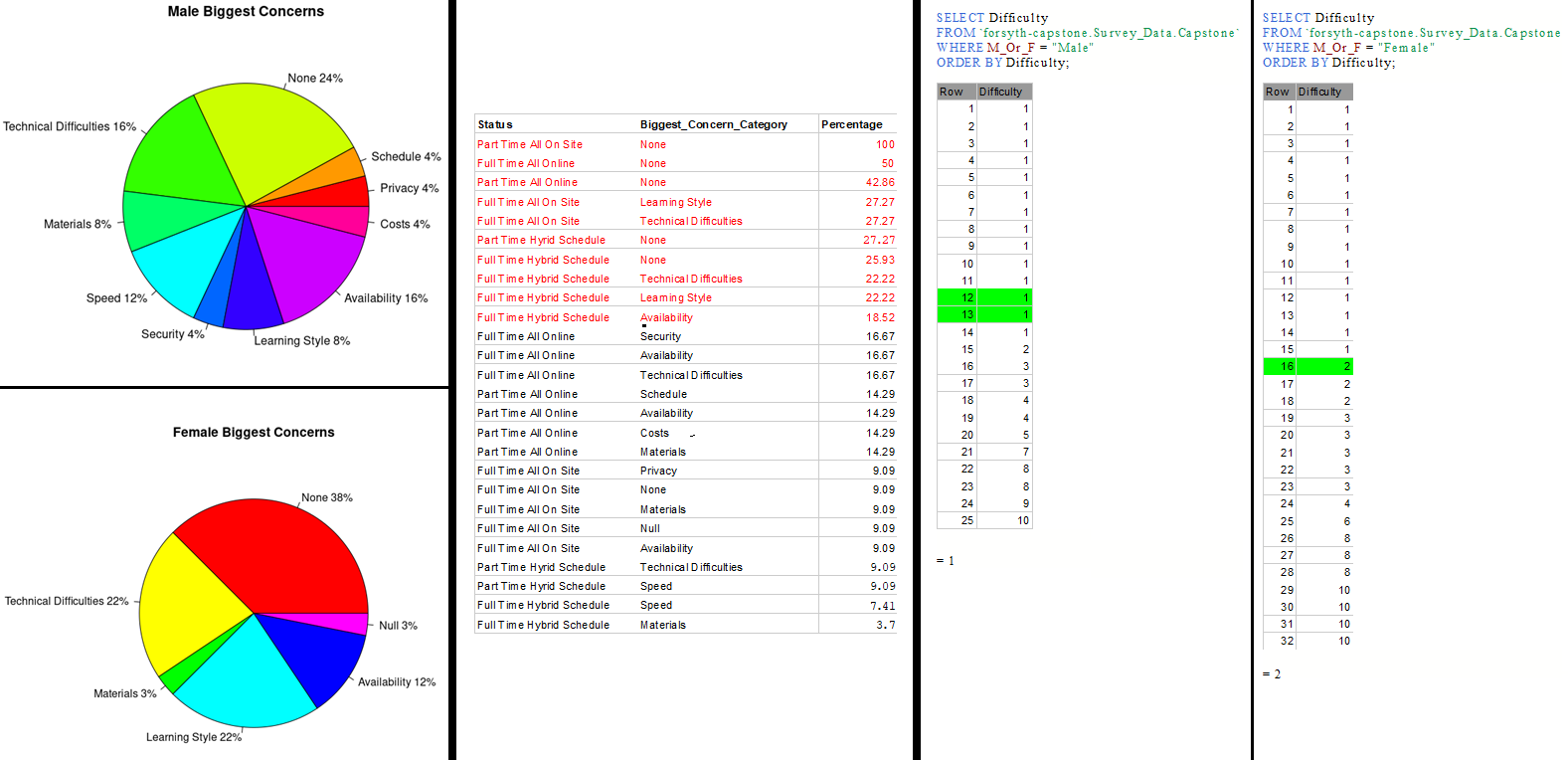
They show how the members perceive the web conferencing software. About half the concerns could be managed with Tech Support while the other half, or the Learning Style, would be harder to tackle.

● Who is your audience? What is the best way to communicate with them?

The audience is Members of the organization I am working with. Some of the higher ups will hear the results, but the target demographic will relate to members who will be affected by the changes.

● Can data visualization help you share your findings?

There are not many visualizations that could help with this report. Most calculations involved tables and not graphic visuals. The Mode and Median are taken from tables. The Biggest Concern by Categories is shown by tables as well. There is not a lot of visually appealing data for this project. The red highlights the top 10 categories. The green highlighted cells are the median. You can deduce the mode from those two images as well (the value that repeats the most).



● Is your presentation accessible to your audience?

The presentation is not as audience friendly as my other projects (it is more based on tables). However, with color coding and filtering in descending order I believe that the audience can understand what the top results are easily enough.

Step 6: Act

1. How do certain demographics affect how people feel about Virtual Advising/meetings?

2. Does the member’s location affect how members feel about Virtual Advising/meetings?

3. What kind of levels of difficulty do our members experience when using online services?

● What is your final conclusion based on your analysis?

We need to survey our members again and determine whether A) Repeated Online Use makes Online Use Easier or B) More Tech Savvy people use Online because they are Tech Savvy. We know that Learning Styles and Technical Issues are the top concerns across Online, Hybrid and On Site. We can tackle the issue of Technical Issues, but Learning Style will still be a large concern for members online. One way to help with the learning style is to limit the size of meetings heavily. A smaller meeting will increase the interaction between all members and because everyone is meeting asynchronously, we can have more meetings at different times. Therefore, more meetings mean smaller groups and more interpersonal communication between all participating members. Meetings with 6-8 people would feel affable while meetings with 30 people would be too large for everyone to talk or have their voices heard.

● How could your team and business apply your insights?

First Suggestion- Survey for A) Repeated Online Use makes Online Use Easier or B) More Tech Savvy people use Online because they are Tech Savvy. Second Suggestion- Also, they need to figure out how small they can make the meetings so that the communication feels more personable. Third Suggestion- They also need to provide more support for technical issues, even going as far as limiting the types of devices that people can connect with (like limiting it to certain computers/tablets and certain brands) so that support will be more widespread and easier for people to use. I would not do away with meeting in person completely, but if the organization makes meeting remotely easier and more convenient, I image more people would use it.

● What next steps would you or your stakeholders take based on your findings?

Figure out how to subsidize the cost of devices for members who want to be remote only. Figure out what types of devices our members need and what brand/devices we should limit them to. This will make support easier (as supporting 5 different devices with technical support is much more efficient and easier then supporting dozens of devices). Figure out how small we can make the meetings so that people have more interpersonal communications with their leaders and other members. I would not do away with meeting in person but make meeting remotely easier and more appealing.

● Is there additional data you could use to expand on your findings?

SELECT Future\_Web\_At\_Organization, COUNT(\*) AS Number

FROM `forsyth-capstone.Survey\_Data.Capstone`

GROUP BY Future\_Web\_At\_Organization

ORDER BY COUNT(\*) DESC;

Using this SQL Code, I got this Chart:

|  |  |
| --- | --- |
| Future\_Web\_At\_Organization | Number |
| Strongly Agree | 18 |
| Agree | 12 |
| Neither Agree nor Disagree | 16 |
| Disagree | 5 |
| Strongly Disagree | 6 |

46 Out of 57 Members in our Sample (80.7%), have a neutral to favorable opinion on Web Conferencing being the future of our organization. With our confidence interval and margin of error based on our sample we can be 90% sure that between 70% and 90% of people in our organization feel this way. We should do what we can to make sure that people who are on the fence are willing to decide to connect remotely in the future. Increasing Technical Support, shrinking meeting sizes and subsidizing devices would be a step in the right direction.