Topic Modeling for HR Survey Responses

Desi Pilla
AMS: Advanced Advertising Data

VIACOMCBS

Introduction: Desi Pilla

❖ Current town: Bethany Beach, Delaware

   o Hometown: Brooklyn, New York

❖ Education:

   o University of Delaware

   o Master of Science in Data Science, December 2020 *(hopefully)*

   o Bachelor's of Mechanical Engineering, December 2019

      – Minors in Mathematics, Physics

❖ Previous experiences:

   o DuPont de Nemours, Inc. – Software / Mechanical Engineering Intern

   o W.L. Gore & Associates – R&D Lab Technician

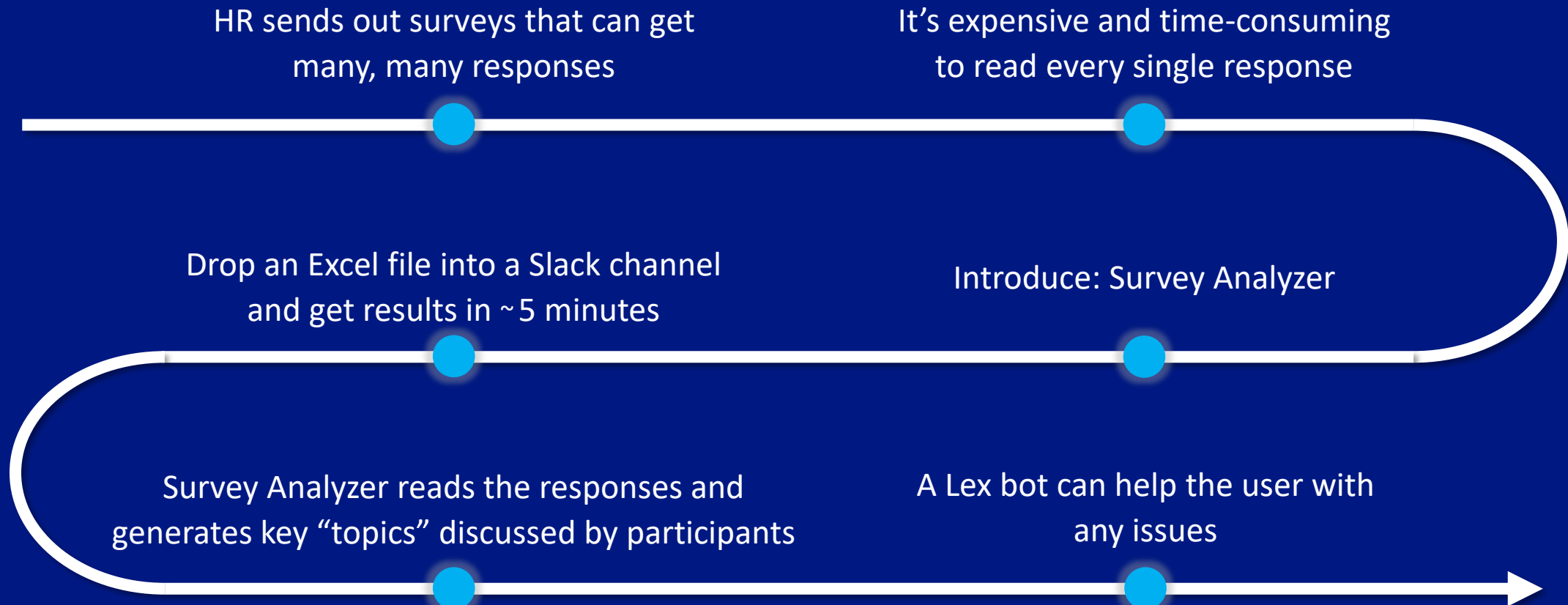   o University of Delaware – Undergraduate Teaching Assistant

Contributors

I would like to recognize the valuable input made by

Matt Moocarme

Problem Statement

HR sends out surveys that can get many, many responses

It's expensive and time-consuming to read every single response

Drop an Excel file into a Slack channel and get results in ~5 minutes

Introduce: Survey Analyzer

Survey Analyzer reads the responses and generates key "topics" discussed by participants

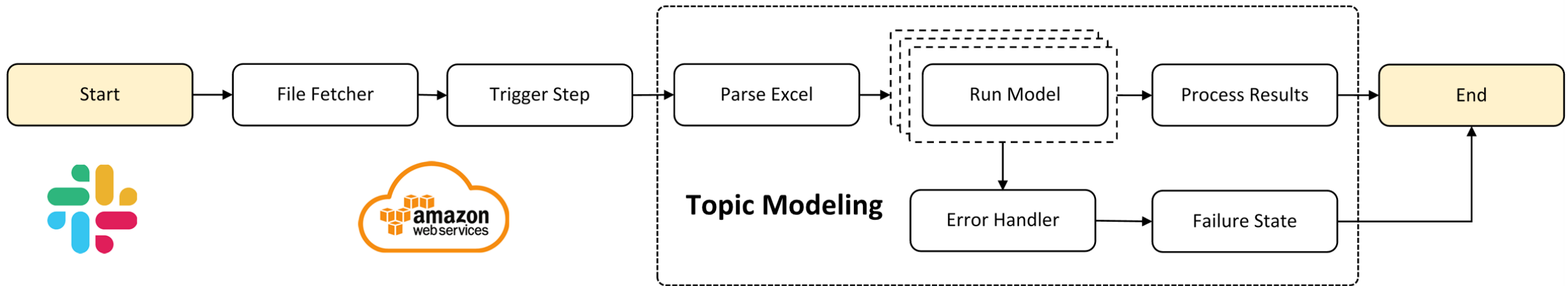A Lex bot can help the user with any issues

VIACOMCBS

Goal

❖ Create an application that

- Is easily accessible / usable by ViacomCBS employees

- Does not require advanced technical knowledge

- Learns key "topics" or "themes" present in the survey responses

- Produces easily interpretable results for the user

VIACOMCBS

Workflow

Topic Modeling

❖ Each survey is split by question

❖ The responses to each question are collected and cleaned
- Lowercase
- Strip numbers / symbols
- Split responses into sentences

❖ Latent Dirichlet Allocation (LDA)
- Implemented with Mallet

❖ Chooses the number of topics by optimizing coherence score
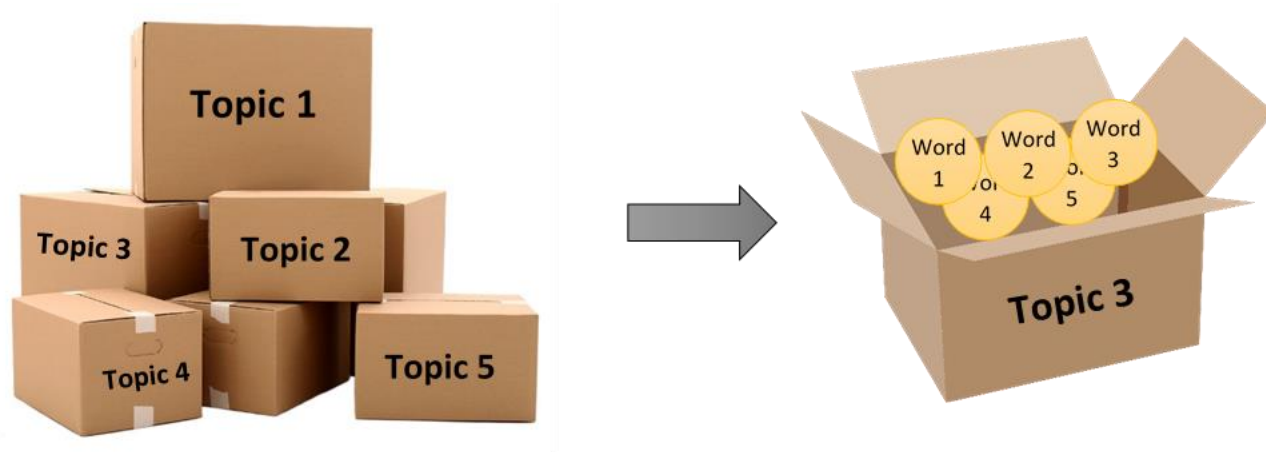
❖ Creates visualizations using pyLDAvis

spaCy

GENSIM
topic modelling for humans

MALLET
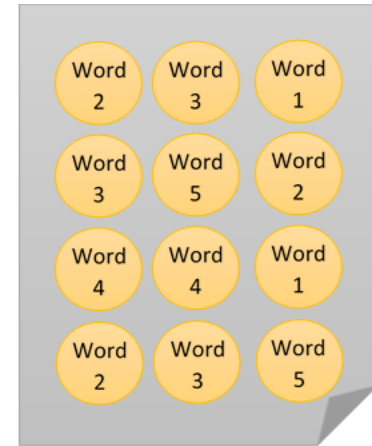machine learning for language toolkit
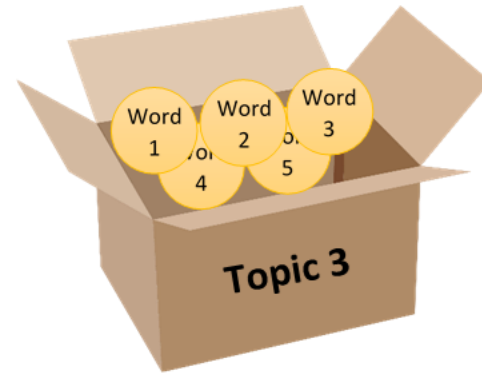
VIACOMCBS

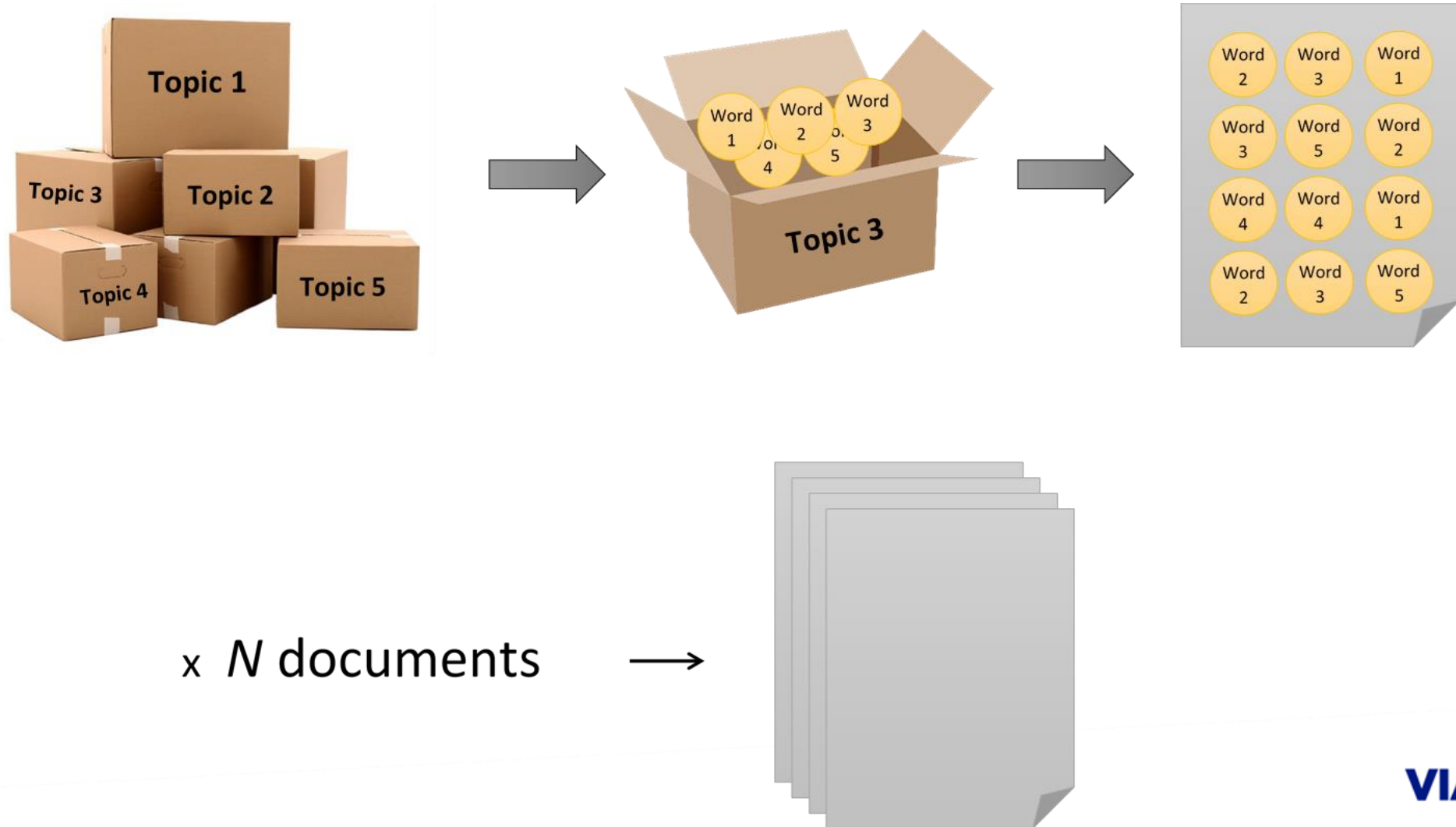# Latent Dirichlet Allocation (LDA)
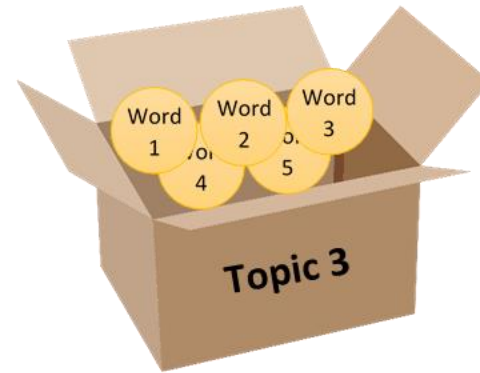
LDA

LDA

LDA

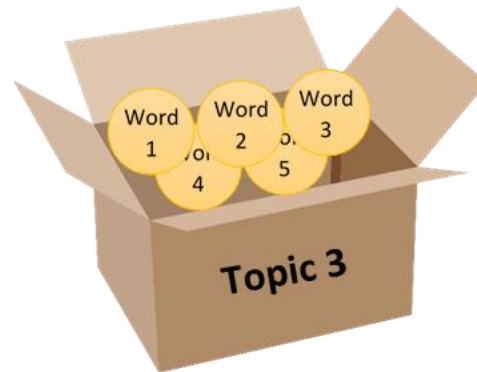# LDA

LDA

What words make up each topic?

x *N* documents →

LDA

What words make up each topic?

What is the probability of selecting each word?

x *N documents*

LDA
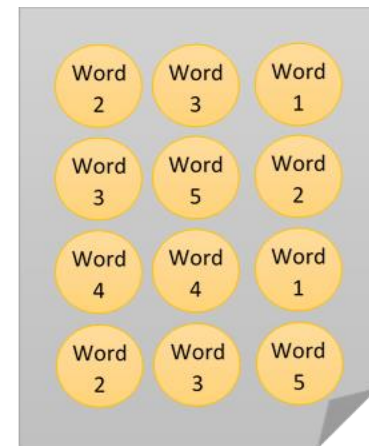
What is the probability of selecting a topic?

What words make up each topic?
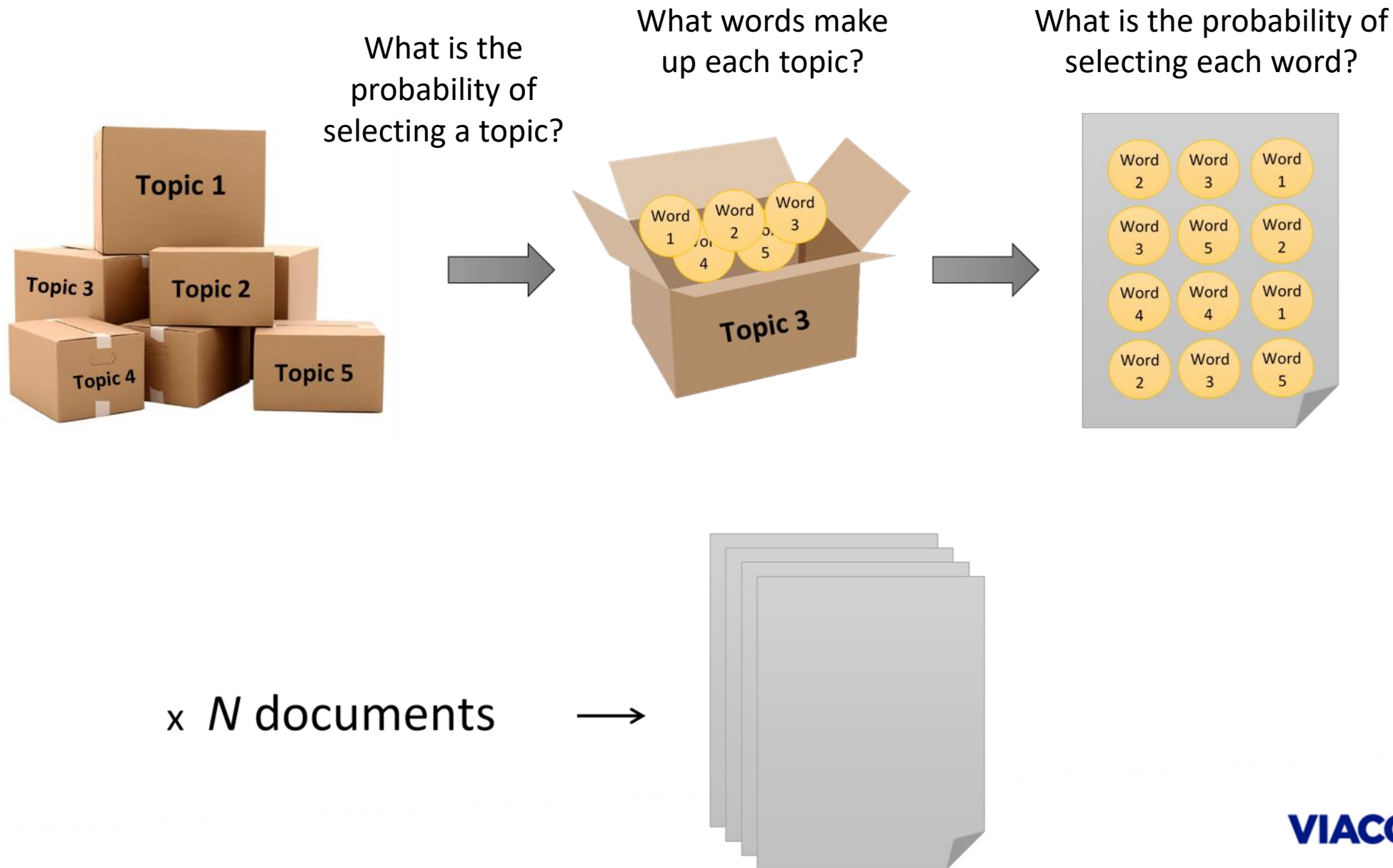
What is the probability of selecting each word?

x *N documents* →

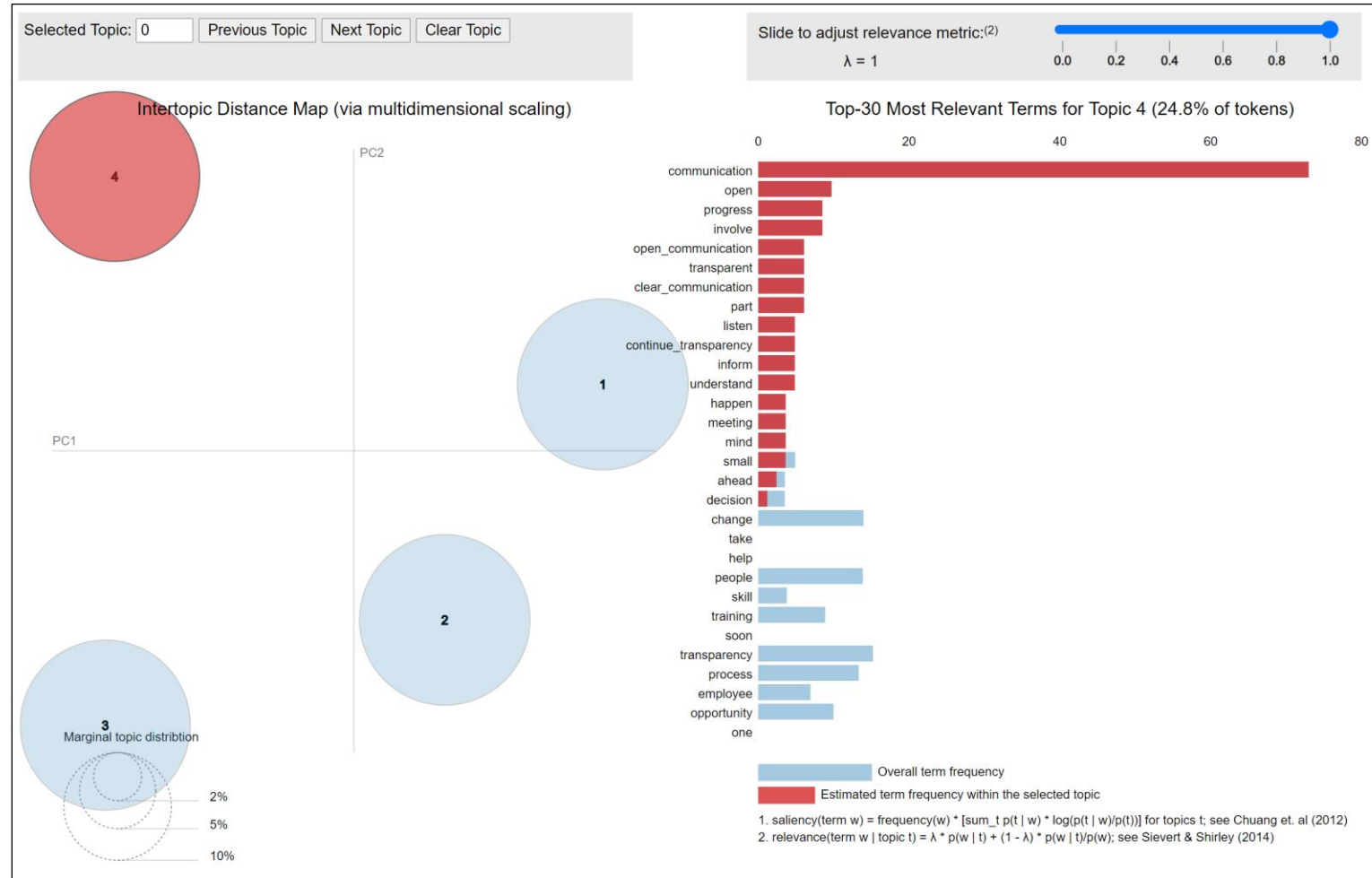# Result Files

Results – HTML File

Results – HTML File

Show [10 ⌄] entries   Search: [   ]

	response_id	text	dominant_topic ▼	keywords
60	Q6-127-0	Open and clear communication on goals and progress	4	communication, open, involve, progress, clear_communication, part, open_communication, transparent, understand, continue_transparency
61	Q6-129-0	Open communication is key	4	communication, open, involve, progress, clear_communication, part, open_communication, transparent, understand, continue_transparency
62	Q6-55-0	More open communications	4	communication, open, involve, progress, clear_communication, part, open_communication, transparent, understand, continue_transparency
63	Q6-0-0	Listen all parts involved in changes before take any decision	4	communication, open, involve, progress, clear_communication, part, open_communication, transparent, understand, continue_transparency
64	Q6-77-0	Open communication throughout the process	4	communication, open, involve, progress, clear_communication, part, open_communication, transparent, understand, continue_transparency
65	Q6-91-0	clear communication in smaller groups focused in each cluster	4	communication, open, involve, progress, clear_communication, part, open_communication, transparent, understand, continue_transparency
66	Q6-50-0	Transparent communication	4	communication, open, involve, progress, clear_communication, part, open_communication, transparent, understand, continue_transparency
67	Q6-49-0	Transparent communication	4	communication, open, involve, progress, clear_communication, part, open_communication, transparent, understand, continue_transparency
68	Q6-11-0	Have an open mind and look for ways to improve the deliverable needs	4	communication, open, involve, progress, clear_communication, part, open_communication, transparent, understand, continue_transparency
69	Q6-146-0	Open communication and understanding of deadlines with integration in a way it doesn't affect day to day operations	4	communication, open, involve, progress, clear_communication, part, open_communication, transparent, understand, continue_transparency

Showing 1 to 10 of 80 entries   Previous 1 2 3 4 5 … 8 Next

VIACOMCBS

Results – Excel File

dominant_topic	keywords	num_documents	perc_documents
1	team, integration, people, update, clear, email, good, plan, regard, employee	76	0.3207
2	system, continue, process, work, finance, opportunity, company, constant, clarity, group	62	0.2616
3	transparency, change, communicate, information, share, training, business, weekly, unit, town	41	0.173
4	communication, open, involve, progress, clear_communication, part, open_communication, transparent, understand, continue_transparency	58	0.2447

Results – Excel File

question_id	response_id	text	dominant_topic	topic_perc_contrib	keywords
Q6	Q6-0-0	Listen all parts involved in changes before take any decision	4	0.2877	communication, open, involve, progress, clear_communication, part, open_communication, transparent, understand, continue_transparency
Q6	Q6-1-0	Offer training that will help people gain the skills that they will require as part of	3	0.2685	transparency, change, communicate, information, share, training, business, weekly, unit, town
Q6	Q6-2-0	Communication as soon as possible	4	0.2647	communication, open, involve, progress, clear_communication, part, open_communication, transparent, understand, continue_transparency
Q6	Q6-4-0	transparency	3	0.2647	transparency, change, communicate, information, share, training, business, weekly, unit, town
Q6	Q6-5-0	Understand what it takes to do what we do before accelerating all our deliverable deadlines	4	0.2647	communication, open, involve, progress, clear_communication, part, open_communication, transparent, understand, continue_transparency
Q6	Q6-6-0	Communication in the process ahead	4	0.269	communication, open, involve, progress, clear_communication, part, open_communication, transparent, understand, continue_transparency
Q6	Q6-7-0	Develop a cross-training strategy for employees	1	0.2647	team, integration, people, update, clear, email, good, plan, regard, employee

Interacting with the user

Survey Analyzer

- ❖ Survey Analyzer is a Slack app that
  - Can be used by anyone in the AA workspace
  - Has an easy drag-and-drop interface to upload files
  - Provides an instructions guide available on the home tab
  - Utilizes Amazon Lex as a back-end to process user requests
  - Provides summary usage statistics

Amazon Lex

VIACOMCBS

Demo

Messages   About

Survey Analyzer ☆

Messages   About

Desiderio Pilla 8:49 AM

Hello Survey Analyzer. Here is my survey responses file. (edited)

Excel Spreadsheet ▾

📗 **MarComms Town Hall - June 11.xlsx**
17 kB Excel Spreadsheet

Did this Town Hall help you feel more supported?

I appreciated Julia's openness at the beginning of the Town Hall, and I

I personally did not, however I belieive it was very helpful for many.
Great speakers!

I'm glad that leadership addressed the issues. Although, I will say that i

Thought the doctors were great. Think it would have been better in sm

I would have like to see more of my white colleagues speak up and sha
Incredibly important to come together as a group for this conversation

Survey Analyzer ⭐

Messages   About

📗 **MarComms Town Hall - June 11.xlsx**
17 kB Excel Spreadsheet

Did this Town Hall help you feel more supported?

I appreciated Julia's openness at the beginning of the Town Hall, and I

I personally did not, however I belieive it was very helpful for many.
Great speakers!

I'm glad that leadership addressed the issues. Although, I will say that i

Thought the doctors were great. Think it would have been better in sm

I would have like to see more of my white colleagues speak up and sha
Incredibly important to come together as a group for this conversation

---

📊 **Survey Analyzer** `APP` 8:49 AM
Please say "Analyze my file" to begin.

Did this Town Hall help you feel more supported?

I appreciated Julia's openness at the beginning of the Town Hall, and I

I personally did not, however I belieive it was very helpful for many.
Great speakers!

I'm glad that leadership addressed the issues. Although, I will say that i

Thought the doctors were great. Think it would have been better in sm

I would have like to see more of my white colleagues speak up and sha
Incredibly important to come together as a group for this conversation

Survey Analyzer APP 8:49 AM
Please say "Analyze my file" to begin.

Desiderio Pilla 8:49 AM
Analyze my file

Survey Analyzer ☆

Messages   About

> Did this Town Hall help you feel more supported?
>
> I appreciated Julia's openness at the beginning of the Town Hall, and I
>
> I personally did not, however I belieive it was very helpful for many. Great speakers!
>
> I'm glad that leadership addressed the issues. Although, I will say that i
>
> Thought the doctors were great. Think it would have been better in sm
>
> I would have like to see more of my white colleagues speak up and sha Incredibly important to come together as a group for this conversation

Survey Analyzer APP 8:49 AM
Please say "Analyze my file" to begin.

Desiderio Pilla 8:49 AM
Analyze my file

AWS calls the Slack API and grabs all files that have been uploaded to Survey Analyzer by this user

Survey Analyzer ⭐

Did this Town Hall help you feel more supported?

I appreciated Julia's openness at the beginning of the Town Hall, and I

I personally did not, however I belieive it was very helpful for many. Great speakers!

I'm glad that leadership addressed the issues. Although, I will say that i

Thought the doctors were great. Think it would have been better in sm

I would have like to see more of my white colleagues speak up and sha Incredibly important to come together as a group for this conversation

Survey Analyzer `APP` 8:49 AM
Please say "Analyze my file" to begin.

Desiderio Pilla 8:49 AM
Analyze my file

AWS calls the Slack API and grabs all files that have been uploaded to Survey Analyzer by this user

List of files:
  [MarComms Town Hall –June 11.xlsx]

Survey Analyzer ☆

Messages   About

> I appreciated Julia's openness at the beginning of the Town Hall, and I
>
> I personally did not, however I belieive it was very helpful for many. Great speakers!
>
> I'm glad that leadership addressed the issues. Although, I will say that i
>
> Thought the doctors were great. Think it would have been better in sm
>
> I would have like to see more of my white colleagues speak up and sha Incredibly important to come together as a group for this conversation

Survey Analyzer `APP` 8:49 AM
Please say "Analyze my file" to begin.

Desiderio Pilla 8:49 AM
Analyze my file

Survey Analyzer `APP` 8:49 AM
Is your file MarComms Town Hall - June 11.xlsx ?

AWS calls the Slack API and grabs all files that have been uploaded to Survey Analyzer by this user

List of files:
 [MarComms Town Hall –June 11.xlsx]

Great speakers!

I'm glad that leadership addressed the issues. Although, I will say that i

Thought the doctors were great. Think it would have been better in sm

I would have like to see more of my white colleagues speak up and sha
Incredibly important to come together as a group for this conversation

Survey Analyzer APP 8:49 AM
Please say "Analyze my file" to begin.

Desiderio Pilla 8:49 AM
Analyze my file

Survey Analyzer APP 8:49 AM
Is your file MarComms Town Hall - June 11.xlsx ?

Desiderio Pilla 8:49 AM
yes

List of files:
  [MarComms Town Hall –June 11.xlsx]

Great speakers!

I'm glad that leadership addressed the issues. Although, I will say that i

Thought the doctors were great. Think it would have been better in sm

I would have like to see more of my white colleagues speak up and sha
Incredibly important to come together as a group for this conversation

Survey Analyzer `APP` 8:49 AM
Please say "Analyze my file" to begin.

Desiderio Pilla 8:49 AM
Analyze my file

Survey Analyzer `APP` 8:49 AM
Is your file MarComms Town Hall - June 11.xlsx ?

Desiderio Pilla 8:49 AM
yes

AWS downloads
MarComms Town Hall –June 11.xlsx
from the Slack API and uploads it to s3

Survey Analyzer `APP` 8:49 AM
Please say "Analyze my file" to begin.

Desiderio Pilla 8:49 AM
Analyze my file

Survey Analyzer `APP` 8:49 AM
Is your file MarComms Town Hall - June 11.xlsx ?

Desiderio Pilla 8:49 AM
yes

Survey Analyzer `APP` 2:06 PM
MarComms Town Hall - June 11.xlsx has been sent to the server for processing.
If you do not receive a confirmation message within a few minutes, please try again.
REF_ID: F017893QCKZ

Survey Analyzer ☆

Messages   **About**

Survey Analyzer `APP` 8:49 AM
Please say "Analyze my file" to begin.

Desiderio Pilla 8:49 AM
Analyze my file

Survey Analyzer `APP` 8:49 AM
Is your file MarComms Town Hall - June 11.xlsx ?

Desiderio Pilla 8:49 AM
yes

Survey Analyzer `APP` 2:06 PM
MarComms Town Hall - June 11.xlsx has been sent to the server for processing.
If you do not receive a confirmation message within a few minutes, please try again.
REF_ID: F017893QCKZ

A Lambda function is triggered by the file upload and initiates the Step Function

Survey Analyzer `APP` 8:49 AM
Please say "Analyze my file" to begin.

Desiderio Pilla 8:49 AM
Analyze my file

Survey Analyzer `APP` 8:49 AM
Is your file MarComms Town Hall - June 11.xlsx ?

Desiderio Pilla 8:49 AM
yes

Survey Analyzer `APP` 2:06 PM
MarComms Town Hall - June 11.xlsx has been sent to the server for processing.
If you do not receive a confirmation message within a few minutes, please try again.
REF_ID: F017893QCKZ

A Lambda function is triggered by the file upload and initiates the Step Function

The step function grabs the data from s3

Survey Analyzer ☆

Messages   About

Please say "Analyze my file" to begin.

Desiderio Pilla 8:49 AM
Analyze my file

Survey Analyzer APP 8:49 AM
Is your file MarComms Town Hall - June 11.xlsx ?

Desiderio Pilla 8:49 AM
yes

Survey Analyzer APP 2:06 PM
MarComms Town Hall - June 11.xlsx has been sent to the server for processing.
If you do not receive a confirmation message within a few minutes, please try again.
REF_ID: F017893QCKZ

Your file has been received. Please allow up to 15 minutes for processing.

> A Lambda function is triggered by the file upload and initiates the Step Function

> The step function grabs the data from s3

Messages   About

Please say "Analyze my file" to begin.

The model runs. All temporary files are stored in s3

Desiderio Pilla 8:49 AM
Analyze my file

Survey Analyzer `APP` 8:49 AM
Is your file MarComms Town Hall - June 11.xlsx ?

Desiderio Pilla 8:49 AM
yes

Survey Analyzer `APP` 2:06 PM
MarComms Town Hall - June 11.xlsx has been sent to the server for processing.
If you do not receive a confirmation message within a few minutes, please try again.
REF_ID: F017893QCKZ

Your file has been received. Please allow up to 15 minutes for processing.

Please say "Analyze my file" to begin.

> The model runs. All temporary files are stored in s3

Desiderio Pilla 8:49 AM

Analyze my file

> The results are uploaded to s3

Survey Analyzer `APP` 8:49 AM

Is your file MarComms Town Hall - June 11.xlsx ?

Desiderio Pilla 8:49 AM

yes

Survey Analyzer `APP` 2:06 PM

MarComms Town Hall - June 11.xlsx has been sent to the server for processing.
If you do not receive a confirmation message within a few minutes, please try again.
REF_ID: F017893QCKZ

Your file has been received. Please allow up to 15 minutes for processing.

Survey Analyzer ⭐

ⓘ

Messages　　About

🔍 **Survey Analyzer** `APP`　2:06 PM

MarComms Town Hall - June 11.xlsx has been sent to the server for processing.
If you do not receive a confirmation message within a few minutes, please try again.
REF_ID: F017893QCKZ

Your file has been received. Please allow up to 15 minutes for processing.

> **The results are uploaded to s3**

Here are your result files:

Excel Spreadsheet ▾

📗	**MarComms Town Hall - June 11** 28 kB Excel Spreadsheet

MarComms Town Hall - June 11 Q1 ▾

```
1   <meta charset="UTF-8">
2   <html><head>
3    <link rel="stylesheet" type="text/css"
    href="https://cdn.datatables.net/1.10.20/css/jquery.dataTables.min.css">
4
5
```

MarComms Town Hall - June 11 Q2 ▾

```
1   <meta charset="UTF-8">
2   <html><head>
3    <link rel="stylesheet" type="text/css"
    href="https://cdn.datatables.net/1.10.20/css/jquery.dataTables.min.css">
4
5
```

MarComms Town Hall - June 11 Q3 ▾

```
1   <meta charset="UTF-8">
2   <html><head>
3    <link rel="stylesheet" type="text/css"
    href="https://cdn.datatables.net/1.10.20/css/jquery.dataTables.min.css">
4
5
```

Summary

VIACOMCBS

Messages   About

Messages   About

Desiderio Pilla 12:03 PM

summary

Survey Analyzer ☆

Desiderio Pilla 12:03 PM
summary

Survey Analyzer `APP` 12:03 PM
3 surveys have been analyzed in the last week.
3 surveys have been analyzed in the last month.
----

3 surveys have been analyzed all time.

Thank you!

VIACOMCBS