

20º Encontro Nacional de Pesquisadores em  
Jornalismo

12º Encontro de Jovens Pesquisadores em  
Jornalismo

**Pedro Jorge Chaves Mourão**

<https://lattes.cnpq.br/1705979956586966>

<https://www.linkedin.com/in/pedro-mourão/>



Doutorando em Sociologia (UECE)  
Analista e Desenvolvedor De Sistemas

Fortaleza-2022



## 1. Agradecimentos

Marcos e Edgard 😊

## 2. Quem sou eu?

- Analista de Dados, gerente TI, arquiteto de software (Assessor II - SUPESP/SSPDS-CE).
- Professor na faculdade de Sistemas de Informação (UNIFANOR).
- Professor Visitante na UNISECAL em Ponta Grossa, Paraná.
- **BLACKFLAG** Consultoria/Assessoria (Pesquisa de mercado, Data Science & Analytics)

## 3. Formação:

- Cientista Social (UECE)
- Analista desenvolvedor de Sistemas (Estácio)
- Mestre em Sociologia (Políticas Públicas) (UFC)
- Doutorando em Sociologia (método quali/quantitativo e técnicas computacionais) (UECE)

## 4. Clivagem conceitual: HUMANAS X EXATAS

“ No mundo real o problema é de humanas ou de exatas? ”



## 2. Como comecei: 2016 e conflitos nas redes sociais

<https://alicenews.ces.uc.pt/index.php?lang=1&id=30865>

# O Farol Digital <https://faroldigital.info/>



## O que é ETL?

# The ETL Process Explained



## Extract

Retrieves and verifies data  
from various sources



## Transform

Processes and organizes  
extracted data so it is usable

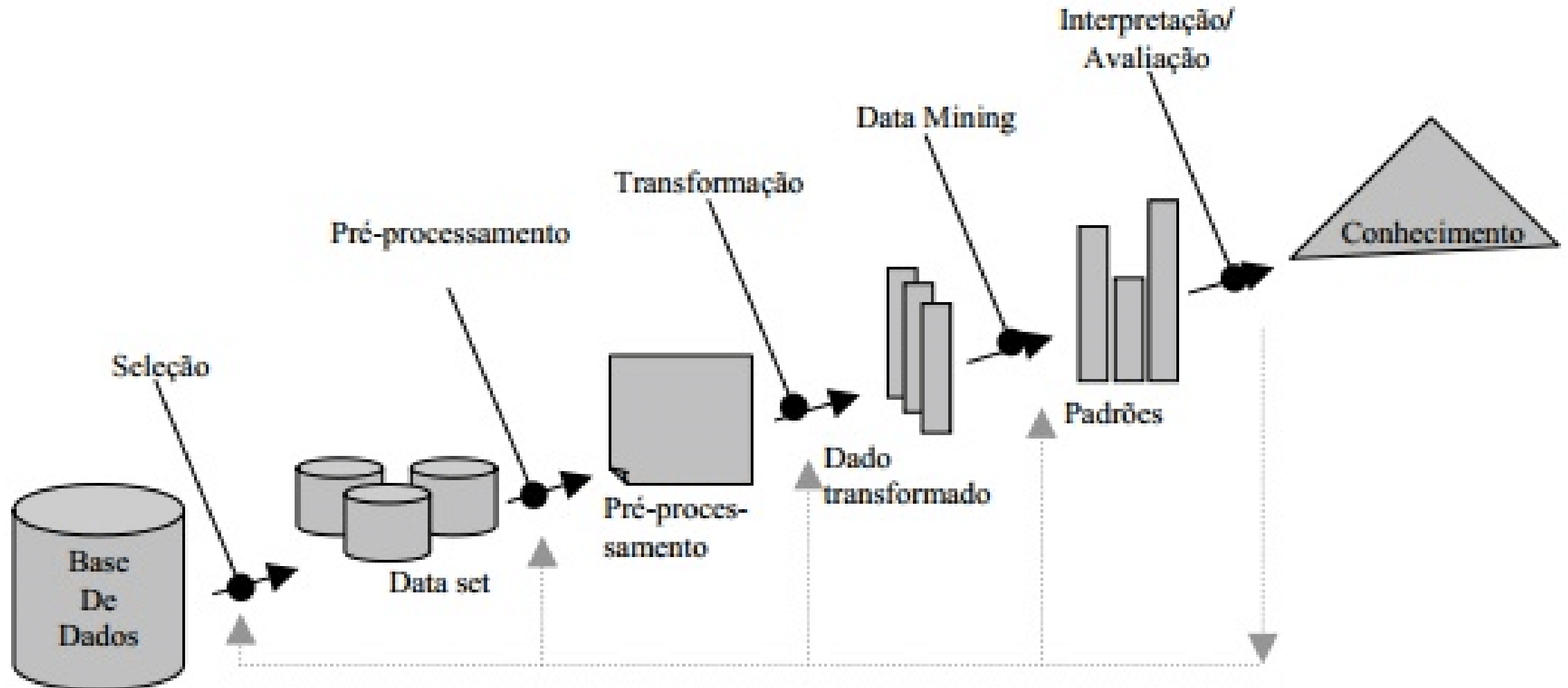


## Load

Moves transformed data  
to a data repository

**E= Recupera e verifica dados de várias fontes. T= processa e organiza os dados extraídos para que sejam utilizáveis. L= Move os dados transformados para um repositório de dados.**

# Fases do processo de descoberta de conhecimento (knowledge discovery in databases KDD)





## Como funciona na prática?

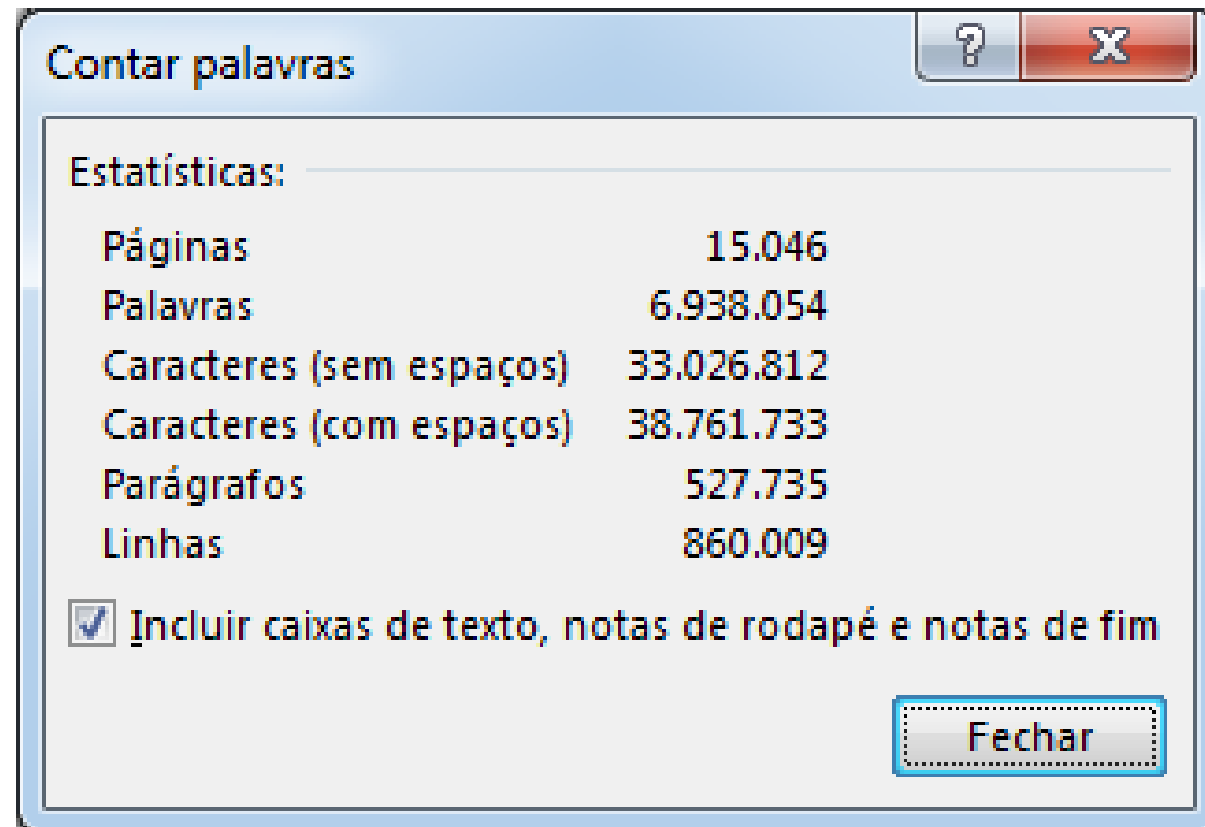
1. Escolher o que quer descobrir
2. Estudar quais as variáveis são relevantes para alcançar as conclusões esperadas
3. Verificar a viabilidade de recursos (dados, tempo, capital humano)
4. Vale a pena contratar um consultor externo?  
(possibilidade de cursos pelo:  
Farol Digital ou pela BlackFlag)

fx = Table.RenameColumns("#Colunas Removidas3",{{"Lista.2.2.1.1.2", "PAÍS"}, {"Lista.2.2.1.1.2", "REGIÃO"}})

	DATA	HORÁRIO	A <sup>B</sup> <sub>C</sub> PAÍS	A <sup>B</sup> <sub>C</sub> REGIÃO	A <sup>B</sup> <sub>C</sub> PESQUISADO	A <sup>B</sup> <sub>C</sub> MENSAGEM
1	01/08/2018	11:15:00	+55	61	9662-	<Arquivo de mídia oculto>
2	01/08/2018	11:16:00	+55	61	9936-	Eis a verdade.
3	01/08/2018	11:28:00	+55	61	9936-	<Arquivo de mídia oculto>
4	01/08/2018	11:43:00	+55	18	99691	<Arquivo de mídia oculto>
5	01/08/2018	11:45:00	+55	33	9136-	Tem muita gente que se diz apoiador de bolsonaro mas esta m
6	01/08/2018	12:00:00	+55	61	9936-	Não tenho dúvida disso. É a falência completa das nossas unive
7	01/08/2018	12:29:00	+55	11	99370	Ontem eu vi um vídeo de um travesti xingando aquele ser que l
8	01/08/2018	12:46:00	+55	61	9662-	Roda Viva Raiz
9	01/08/2018	12:46:00	+55	61	9662-	<Arquivo de mídia oculto>
10	01/08/2018	12:56:00	+55	98	8816-	Kkkkkkk
11	01/08/2018	12:56:00	+55	98	8816-	Massa demais
12	01/08/2018	12:56:00	+55	98	8816-	Quando foi?
13	01/08/2018	13:02:00	+55	11	99370	Kkk caraca nenm lembrava mais dele
14	01/08/2018	13:13:00	+55	11	99370	<Arquivo de mídia oculto>
15	01/08/2018	13:27:00	+55	66	9967-	Chok.vcf (arquivo anexado)
16	01/08/2018	13:28:00	+55	66	9967-	Mas um guerreiro ad ai
17	01/08/2018	13:28:00	+55	51	8544-	👉👉👉👉👉👉👉👉👉👉👉👉👉👉👉👉
18	01/08/2018	13:28:00	+55	66	9967-	Choque 📄
19	01/08/2018	13:36:00	+55	11	99370	Não estou conseguindo add
20	01/08/2018	13:39:00	+55	66	9967-	Choque.vcf (arquivo anexado)
21	01/08/2018	13:40:00	+55	11	99370	Agora deu
22	01/08/2018	13:41:00	+55	66	9967-	<Arquivo de mídia oculto>
23	01/08/2018	13:41:00	+55	11	99370	Bem vindo
24	01/08/2018	13:41:00	+55	66	9967-	<Arquivo de mídia oculto>
25	01/08/2018	13:41:00	+55	18	99691	<Arquivo de mídia oculto>
26						



# MICRO DADOS DO BANCO DE DADOS

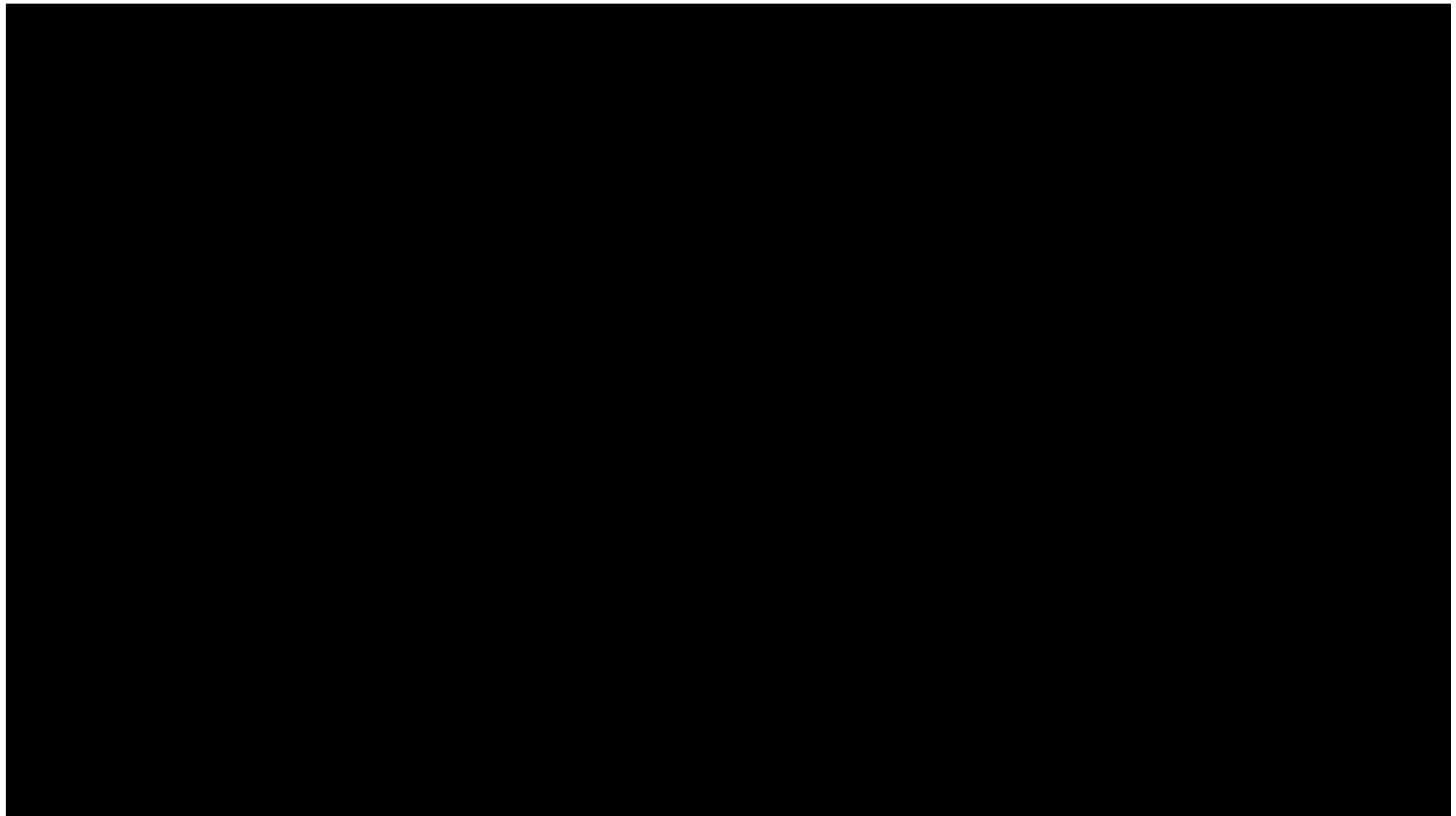


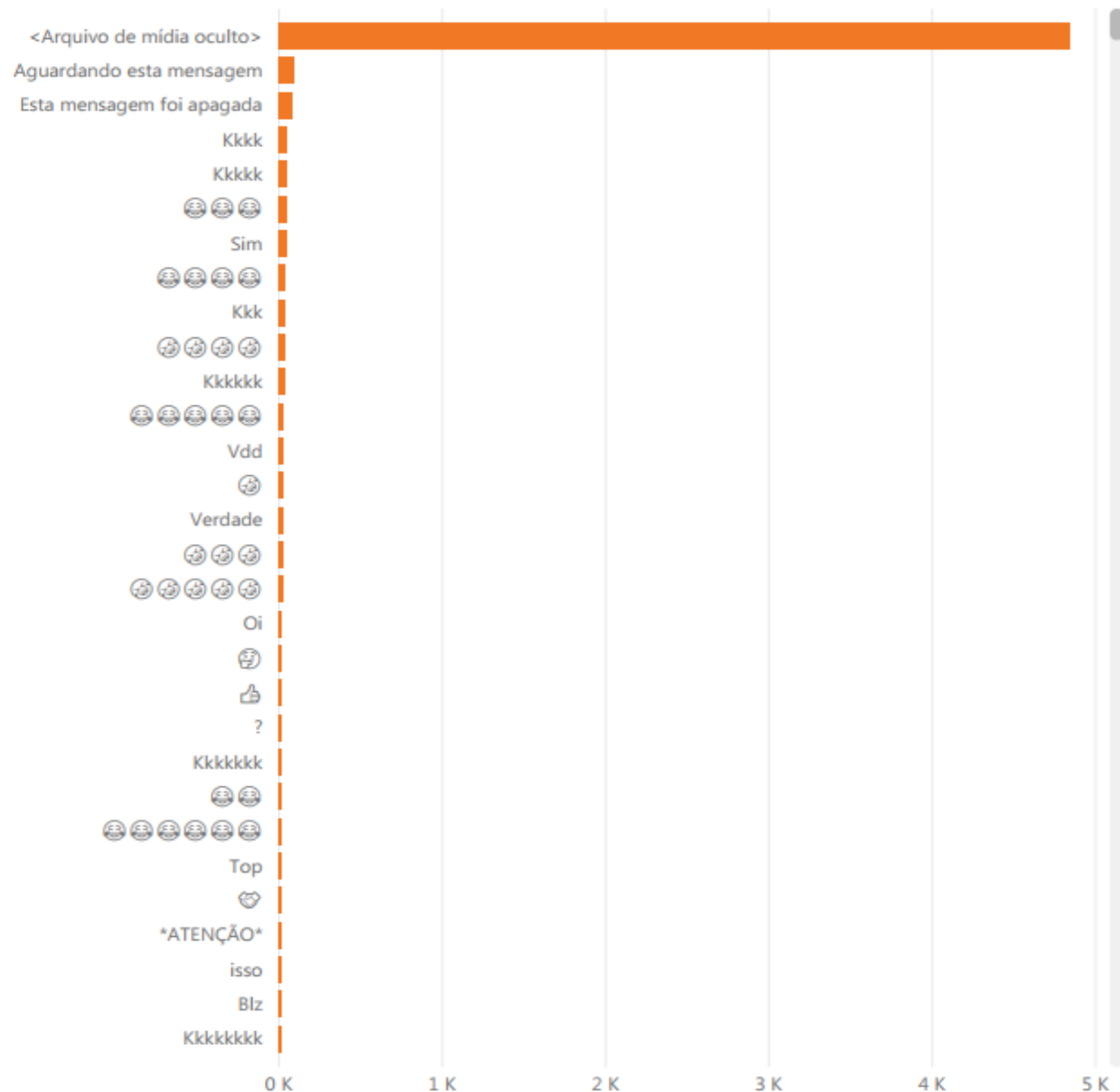
# O que dá para conseguir com Power BI



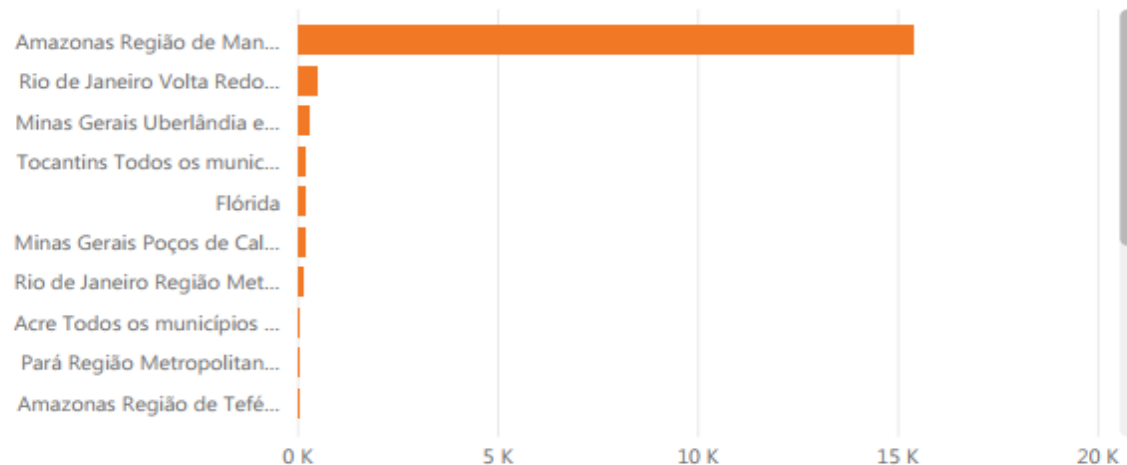
- Qual é o total de interações?
- Quais as mensagens mais enviadas?
- De qual (DDD/DDI) (país, estado, região) são os membros?
- Quem são os telefones que mais publicam no grupo?
- Quais são os links mais compartilhados?
- Quais horários mandam mensagens?
- Quais dias do mês mandam mas mensagens?
- Quantidade de mensagens por lugar de origem?

Vamos lá??





Número de mensagens por LUGAR



País	Número de mensagens
Brasil	16749
EUA	161
Argentina	1
<b>Total</b>	<b>16911</b>



17 K

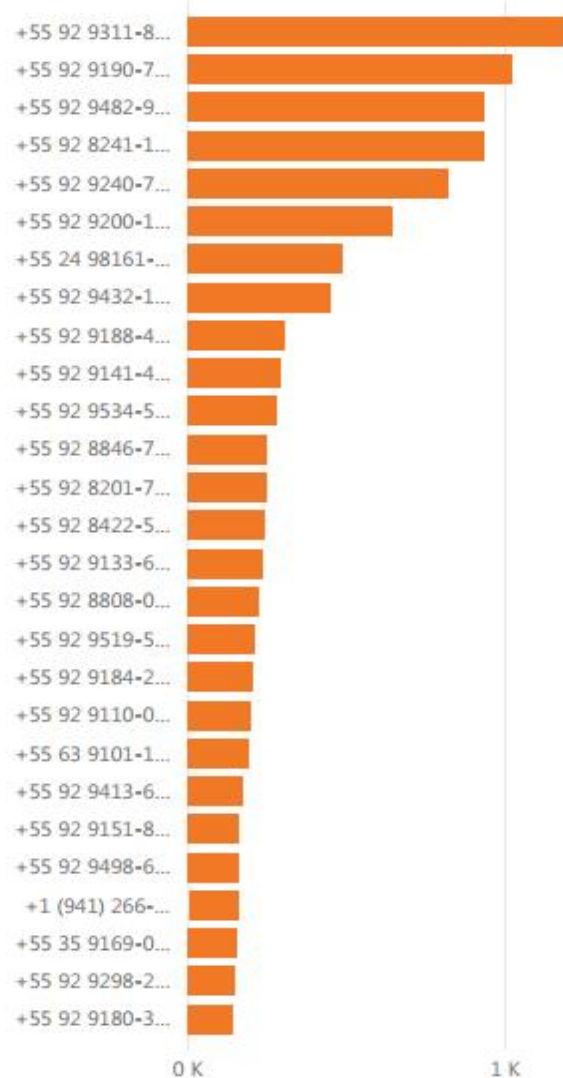
Número de mensagens



MENSAGEM	Número de interações
👉👉👉	8
Amém	7
Bem vindo	7
Boa	7
entendi	7
Malditos	7
👊👊👊👊👊👊	7
👊	7
😂😂😂😂😂😂😂	7
😂😂😂😂😂😂😂😂	7
alterou a imagem deste grupo	6
Biz	6
Exato	6



Número de mensagens por APOIADORES



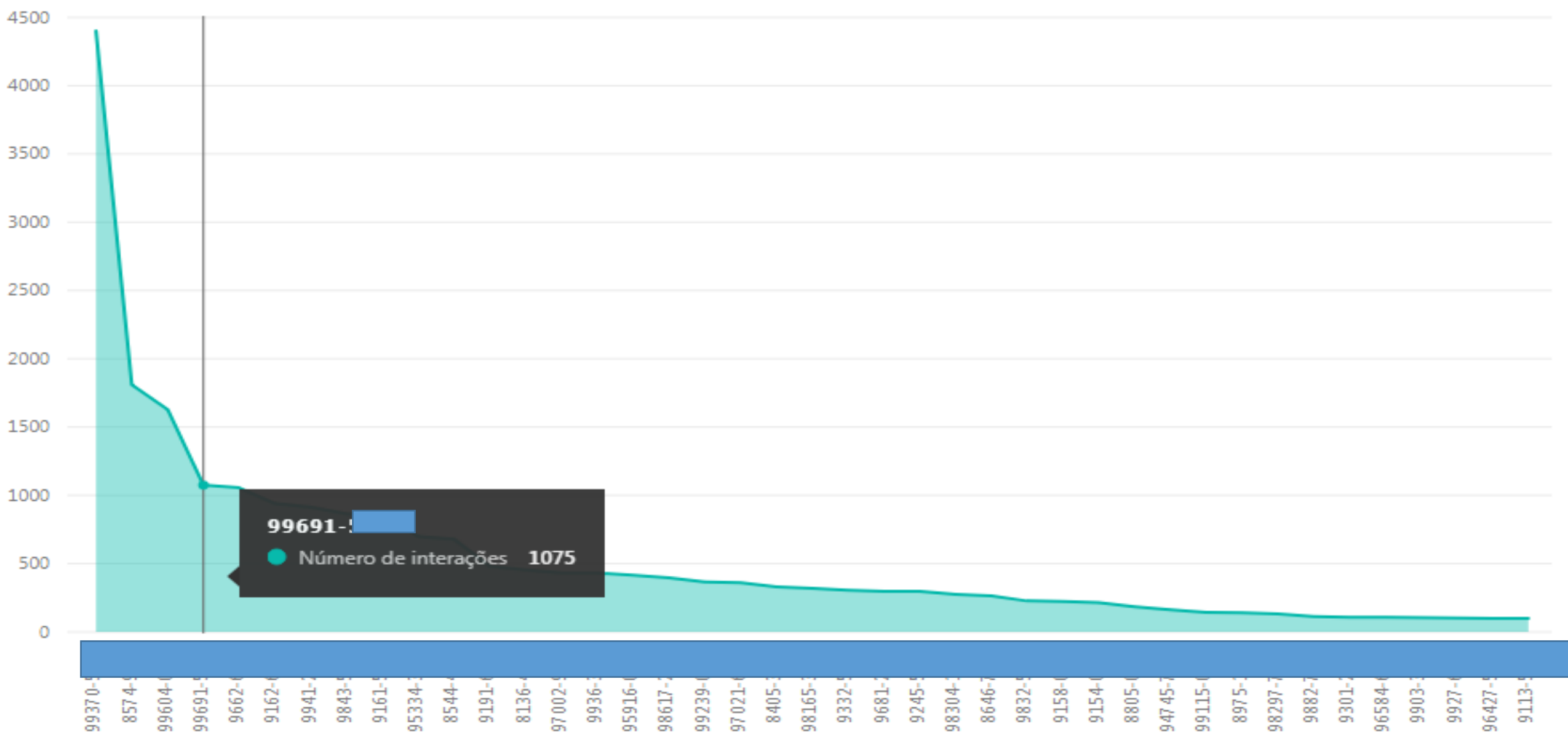
LUGAR



< Voltar ao Relatório

NÚMERO DE INTERAÇÕES

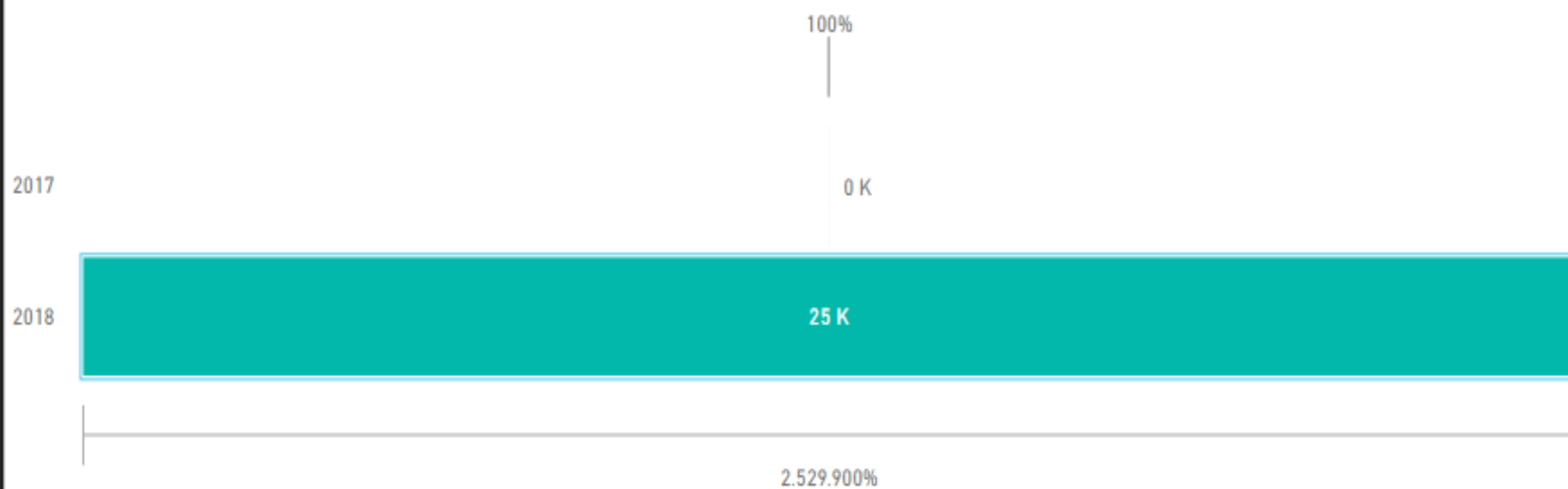
POR PARTICIPANTE





[Voltar ao Relatório](#)

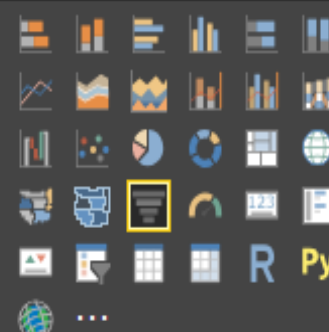
NÚMERO DE INTERAÇÕES POR ANO



Ano	Número de interações
2017	1
2018	25299



VISUALIZAÇÕES



Grupo

- DATA ▾ ×
- Ano ×
- Trimestre ×
- Mês ×
- Dia ×

Valores

Número de interações ▾ ×

Descrições

Adicionar campos de dados...

FILTROS

Filtros de nível visual

DATA - Ano ▾  
é (Todos)

CAMPOS

Procurar

PRESIDENTE JAIR BO

DATA

☐ HORA

☐ MENSAGEM

☒ Número de intera

☐ PÁIS

☐ PARTICIPANTE

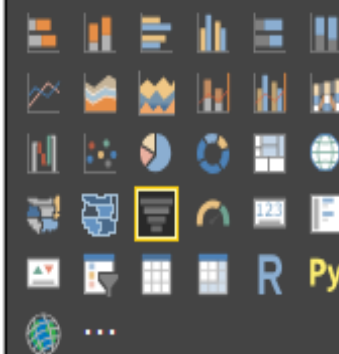
☐ REGIÃO

[Voltar ao Relatório](#)

NÚMERO DE INTERAÇÕES POR MÊS



## VISUALIZAÇÕES



### Grupo

- DATA ▾ ×
- Ano ×
- Trimestre ×
- Mês ×
- Dia ×

### Valores

Número de interações ▾ ×

### Descrições

Adicionar campos de dados...

## FILTROS

### Filtros de nível visual

DATA - Ano ▾

## CAMPOS

Procurar

PRESIDENTE JAIR BOLSONARO

DATA

HORA

MENSAGEM

Número de interações

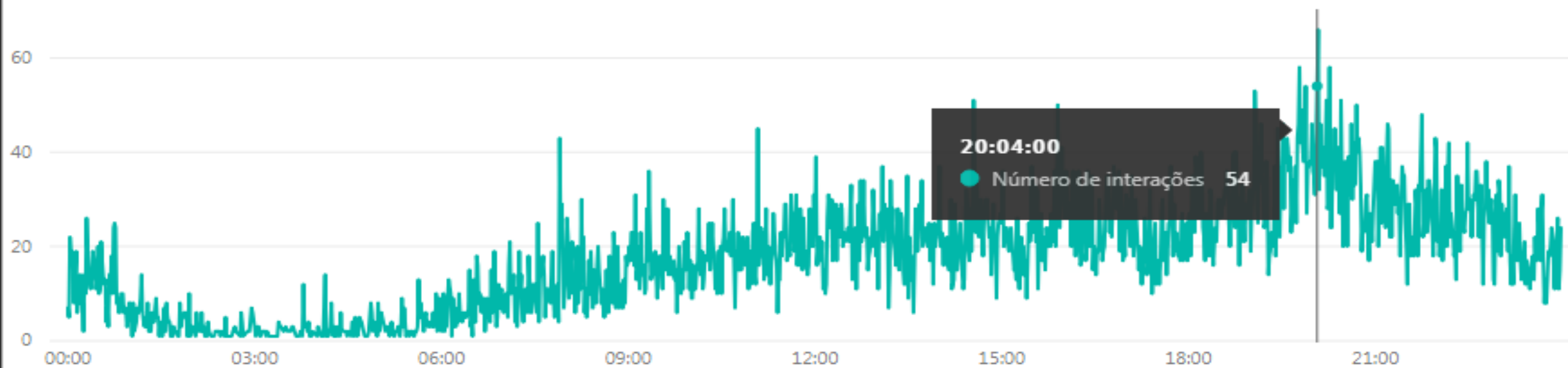
PÁIS

PARTICIPANTE

REGIÃO

< Voltar ao Relatório

NÚMERO DE INTERAÇÕES POR HORA



HORA Número de interações

00:00:00	7
00:01:00	5
00:02:00	22
00:03:00	19
00:04:00	19
00:05:00	17
00:06:00	8
00:07:00	15
00:08:00	19
00:09:00	6
00:10:00	9
00:11:00	8
00:12:00	4



## Número de mensagens por hora (média)

00:00

03:00

06:00

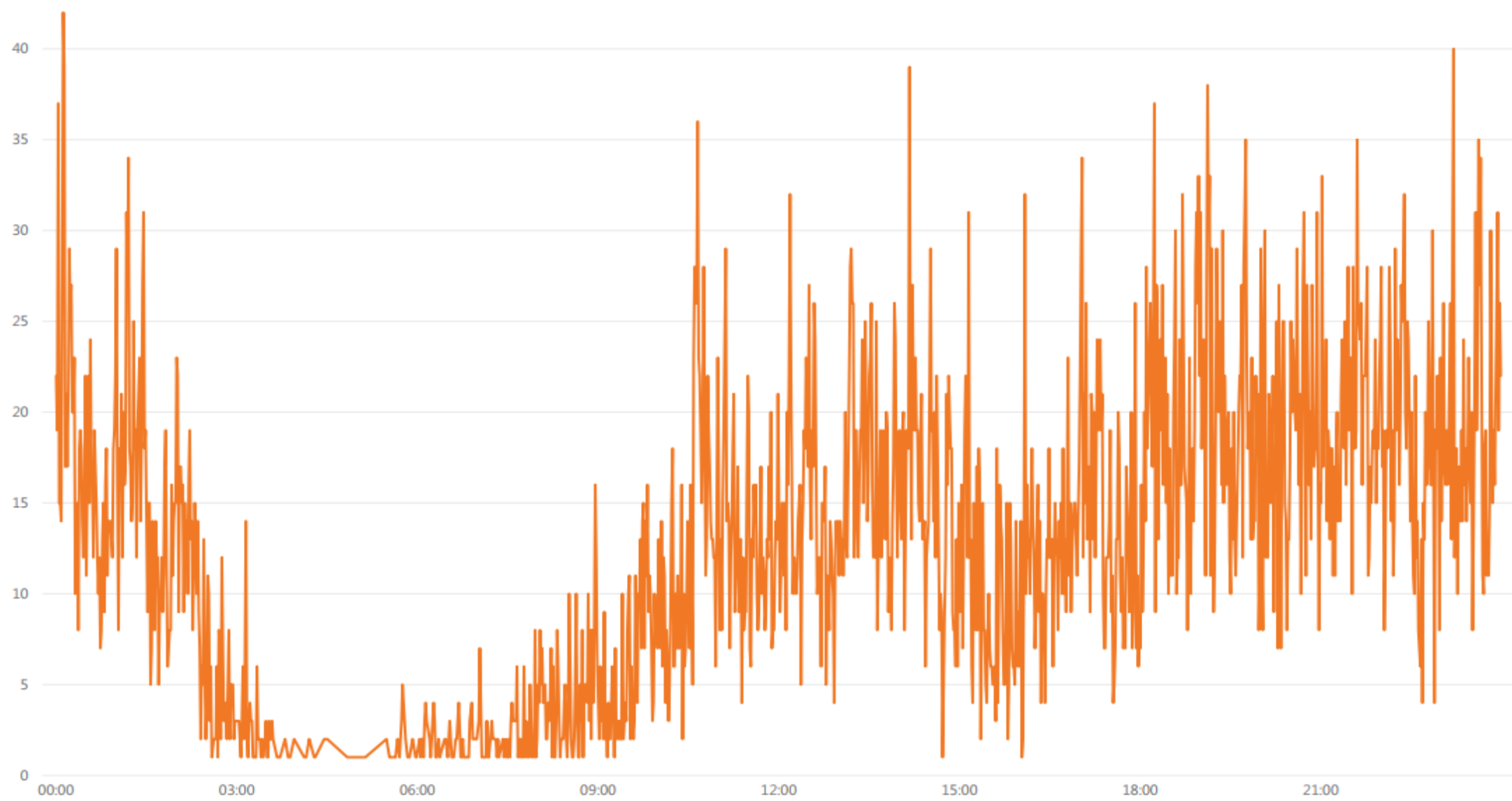
09:00

12:00

15:00

18:00

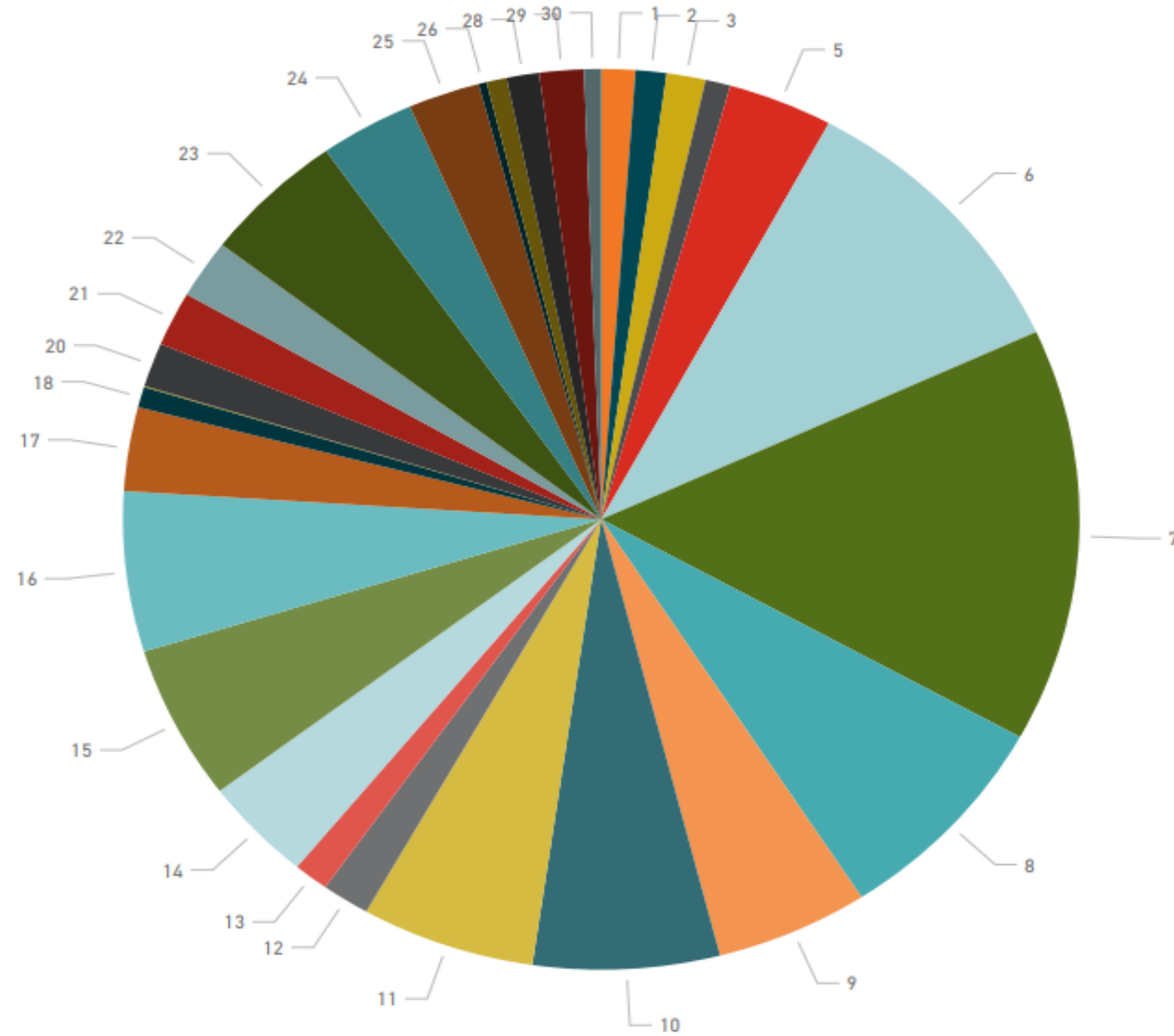
21:00



Número de mensagens por LUGAR

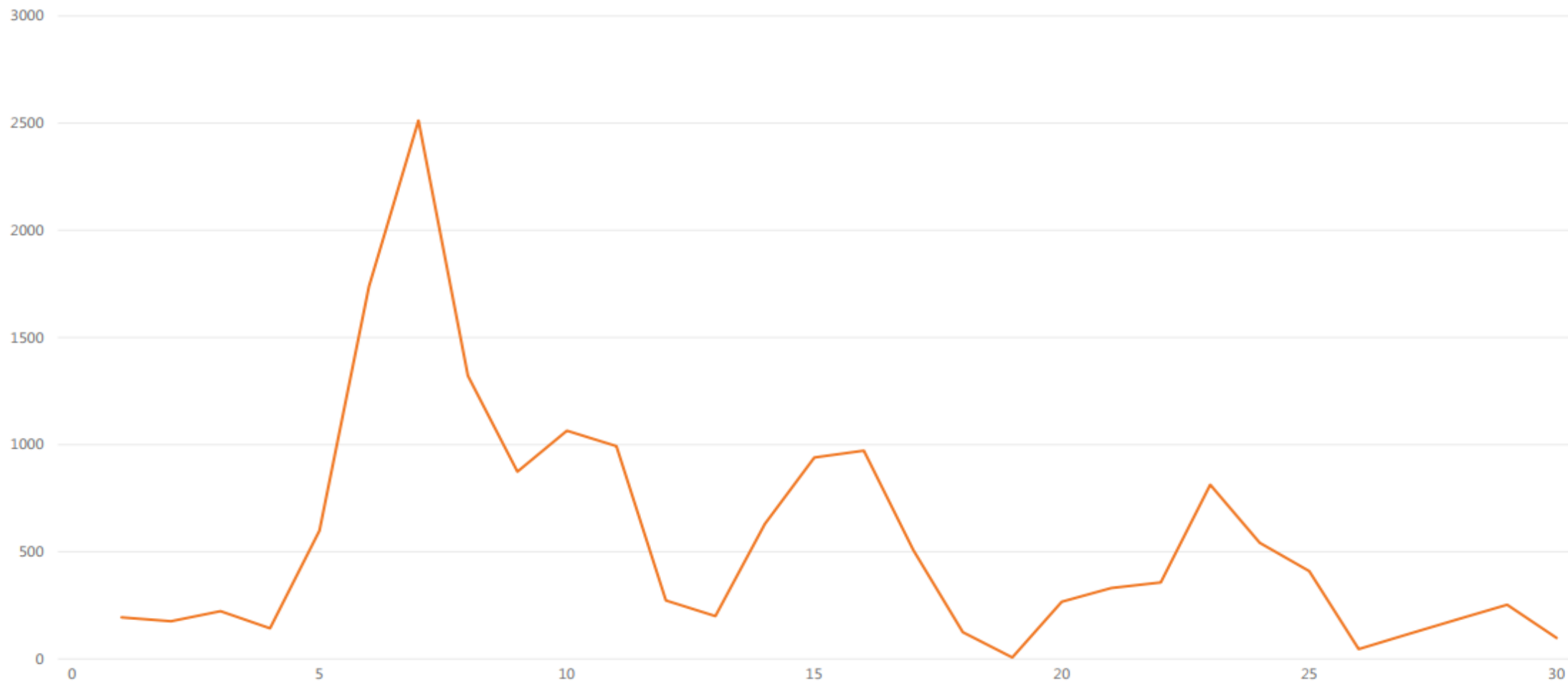


# Número de mensagens por dia do mês





Número de mensagens por Dia



Numero de mensagens por dia do mês

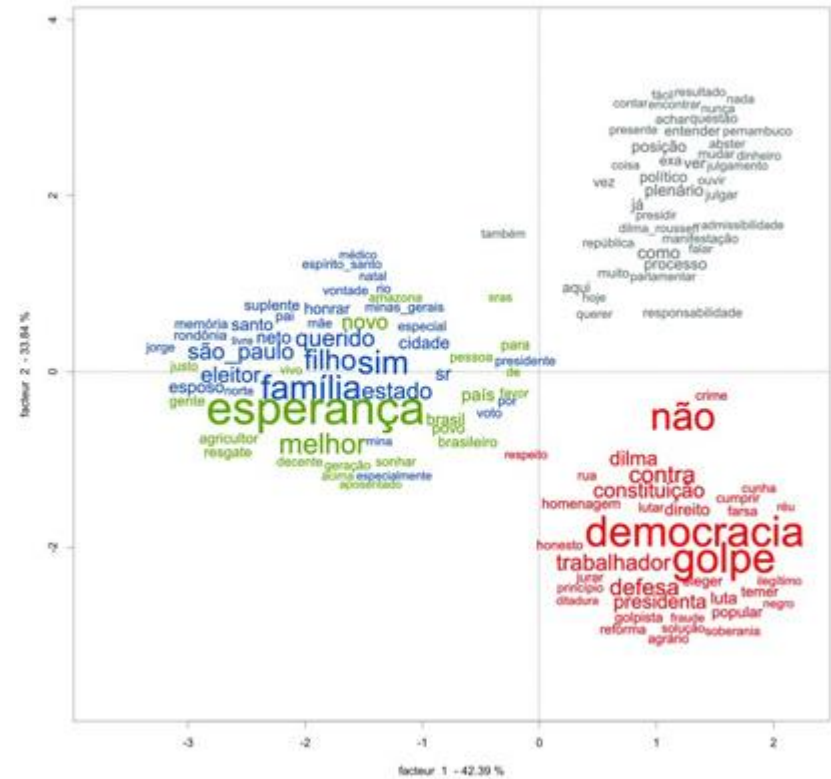
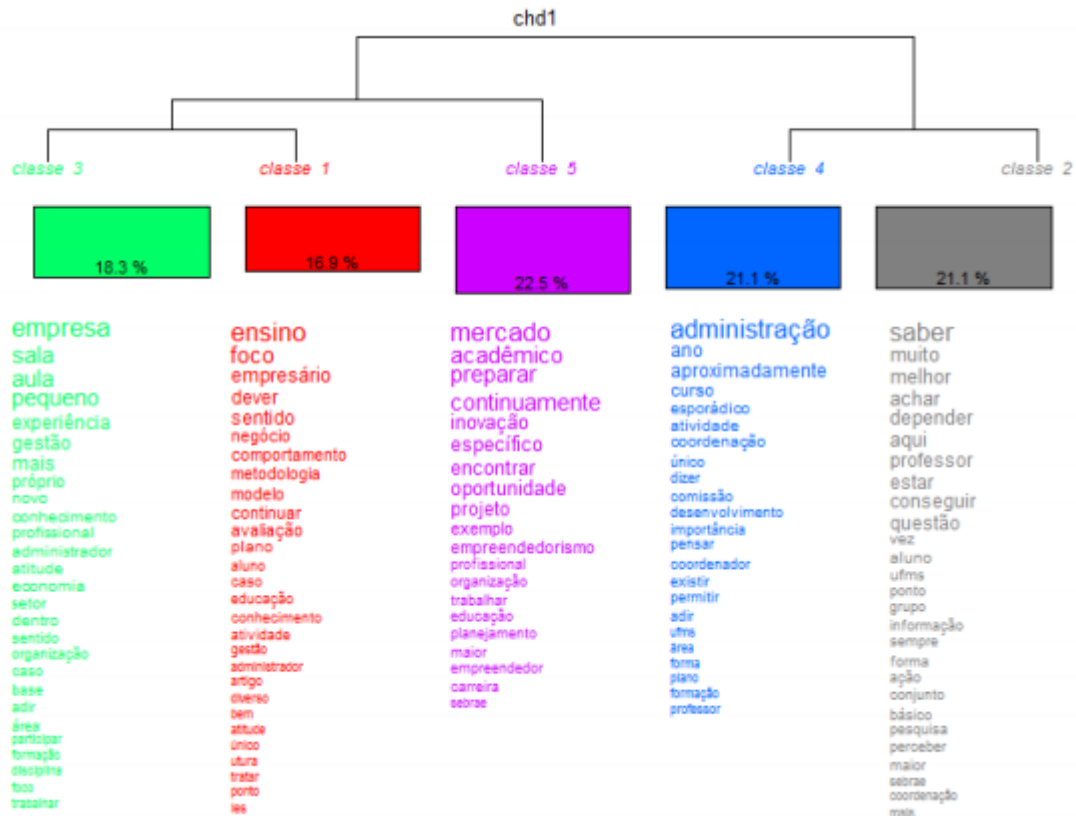
- **MÉTRICAS PARA ANÁLISE DOS GRUPOS**
- 
- Variáveis  
DATA; HORÁRIO; PAÍS; LOCALIDADE; USUÁRIO; MENSAGEM
- Análise de sentimento do total dos nomes do grupos.
- Número de mensagens por grupo (interações)
- Polaridade dos emojis (análise de sentimentos)
- Número de mensagens total de todos grupos
- Número de mensagens por DDD (LUGAR)
- Números de mensagens por DDI (PÁIS)
- Número de caracteres por mensagem
- Número de mensagens por termo
- Número de links enviado por usuário
- Mensagem mais reproduzidas em grupos (período do dia, dia do mês, dia da semana, semana do mês, mês)
- Número de mensagens por apoiador (ID)
- Mapa de locais de onde vieram as mensagens por grupo
- Número de mensagens por hora do dia
- Número de mensagens por dia da semana (por mês e pelo total)
- Número de mensagens por dia do mês (30º do mês)
- mensagens por Dia (11/12/2018)

- Número de mensagens por mês
- Média de mensagens por hora
- Média de mensagens por Dia
- Média de mensagens por Semana
- Médias de mensagens por mês
- Mensagens apagadas por apoiador (ID)
- Tamanho da mensagem por apoiador (ID) (por ID de um só estado, por ID estrangeiro e nacional)
- Mensagens compartilhadas por: período do dia (manhã, tarde, noite; dia; semana, mês.
- Média de links enviados (enviados + encaminhados) id nacional e id internacional
- Total de links enviados por grupo e por ID (por grupo e por total) id nacional e id internacional
- Média de mídia enviadas (enviados + encaminhados) id nacional e id internacional
- Total de mídias enviadas por ID (por grupo e por total) id nacional e id internacional
- Fontes mais compartilhadas por apoiador (ID)
- Polaridade da notícia compartilhada no grupo
- Textos em inglês
- Wordcloud de cada grupo e do todo.



# O que dá para conseguir com Iramuteq

- Auxilia tanto em análise de conteúdo como em análise de discurso
- Estatísticas textuais
- Classificação hierárquica descendente.
- Análises de Similitude
- Nuvem de Palavras
- Outros gráficos da Análise Textual



# Prós

- 2 ou 3 podem fazer o trabalho de uma equipe inteira de pesquisa. Ex: Só eu no WhatsApp consegui analisar dados de 60 grupos com em média 200 pessoas (256 max - 100 min) o que dá um teto de 15360 pessoas estudadas por um pesquisador = uma pequena cidade do interior.
- Tempo reduzido para pesquisas grandes
- Ótimo para sondagens exploratórias e monitoramento
- Método Quali-Quanti que pode ser interdependente ou não (modelo modular)
- Gasto com material é quase zero (notebook, energia elétrica, software)
- Método altamente escalável (tanto serve para o Unilever como para o dono de mercearia do Barroso que tenha pelo menos 5 a 6 funcionários e relevância no bairro)

# Contras

- Exige conhecimento de médio a avançado em duas áreas de conhecimento opostas (antropologia e computação)
- Se uma das etapas do processo der errada afeta toda a sequência dele (suscetível ao efeito cascata)
- Método é experimental
- Softwares e notebooks podem ser caros.
- Podem surgir problemas relacionados a questão ética com o dados e as iterações com os usuários (caso Cambridge Analytics) o que obriga ter qualquer projeto ter um Data Protection Manager
- O que está offline dificilmente será captado

**Obrigado**

**Pedro Mourão**

**pjmourao\_cs@hotmail.com**