

### TRABALHO DE CONCLUSÃO DE CURSO

# Análise exploratória de uma conversa do WhatsApp entre duas pessoas

Aluno 1

Aluno 2

Brasília, Setembro de 2022

UNIVERSIDADE DE BRASÍLIA

# UNIVERSIDADE DE BRASÍLIA Faculdade de Tecnologia

### TRABALHO DE CONCLUSÃO DE CURSO

# Análise exploratória de uma conversa do WhatsApp entre duas pessoas

### Aluno 1 Aluno 2

Trabalho de Conclusão de Curso submetido ao Departamento de Engenharia

Elétrica como requisito parcial para obtenção

do grau de Engenheiro Eletricista

#### Banca Examinadora

Prof. 1, ENE/UnB	
Orientador	
Prof. 2, ENE/UnB  Examinador interno	
Prof. 3, ENE/UnB	
Examinador interno	

#### FICHA CATALOGRÁFICA

OLIVEIRA, MATHEUS NOSCHANG DE; BRAZ, RUBENS SAITO MIRA

Análise exploratória de uma conversa do WhatsApp entre duas pessoas [Distrito Federal] 2022.

xvi, 15 p., 210 x 297 mm (ENE/FT/UnB, Engenheiro, Engenharia Elétrica, 2022).

Trabalho de Conclusão de Curso - Universidade de Brasília, Faculdade de Tecnologia.

Departamento de Engenharia Elétrica

Análise exploratória
 WhatsApp
 Visualização de dados
 Python

I. ENE/FT/UnB II. Título (série)

#### REFERÊNCIA BIBLIOGRÁFICA

OLIVEIRA, M.N.; BRAZ, R.S.M (2022). *Análise exploratória de uma conversa do WhatsApp entre duas pessoas*. Trabalho de Conclusão de Curso, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 15 p.

#### CESSÃO DE DIREITOS

AUTORES: Aluno 1; Aluno 2

TÍTULO: Análise exploratória de uma conversa do WhatsApp entre duas pessoas.

GRAU: Engenheiro Eletricista ANO: 2022

É concedida à Universidade de Brasília permissão para reproduzir cópias deste Trabalho de Conclusão de Curso e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. Os autores reservam outros direitos de publicação e nenhuma parte desse Trabalho de Conclusão de Curso pode ser reproduzida sem autorização por escrito dos autores.

#### Aluno 1

Departamento de Engenharia Elétrica (ENE) -

FΤ

Universidade de Brasília (UnB)

Campus Darcy Ribeiro

CEP 70919-970 - Brasília-DF - Brasil

#### Aluno 2

Departamento de Engenharia Elétrica (ENE) -

FT

Universidade de Brasília (UnB)

Campus Darcy Ribeiro

CEP 70919-970 - Brasília-DF - Brasil

Agradecimentos I	

Aluno 1

Agradecimentos 2

Aluno 2

### Agradecimentos

A Deu	ıs por ter me da	do saúde e	e sanidade	para ei	nfrentar (	os mais c	liversos	desafios
e percalços.								

Aluno 1

Gostaria de agradecer inicialmente ao meu grande amigo Matheus Noschang...

Aluno 2

#### **RESUMO**

A análise de dados vem se mostrando extremamente útil na identificação de padrões e informações pertinentes que permitem a tomada de decisões assertivas.

Palavras Chave: análise exploratória; whatsapp; visualização de dados; python.

#### **ABSTRACT**

Data analysis has been proving itself extremely useful for identifying patterns and pertinent information that allows assertive decision making.

Keywords: exploratory analysis; whatsapp; data visualization; python.

# SUMÁRIO

1	Intro	DUÇÃO	1
	1.1	DEFINIÇÃO DO PROBLEMA	1
		MENTAÇÃO TEÓRICA	
	2.1	INFRAESTRUTURA DO PROJETO	2
	2.1.1	Domínio	2
		OOLOGIA	
	3.1	Introdução	3
4	RESUL	TADO	4
5	Conci	_USÃO	14
RI	FEERÊ	NCIAS RIRI IOGRÁFICAS	15

# LISTA DE FIGURAS

4.1	Formulário de envio da conversa a ser analisada no site Analizap	4
4.2	Estatísticas descritivas de namorados.	5
4.3	Gráficos de pizza de namorados.	5
4.4	Série temporal de namorados.	6
4.5	Gráfico mensal de namorados.	6
4.6	Gráficos polares de namorados.	6
4.7	Mapa de calor de namorados.	7
4.8	Estatísticas descritivas de colegas de trabalho.	7
4.9	Gráficos de pizza de colegas de trabalho	8
4.10	Série temporal de colegas de trabalho.	8
4.11	Gráfico mensal de colegas de trabalho.	8
4.12	Gráficos polares de colegas de trabalho.	9
4.13	Mapa de calor de colegas de trabalho	9
4.14	Estatísticas descritivas de familiares.	10
4.15	Gráficos de pizza de familiares.	10
4.16	Série temporal de familiares.	10
4.17	Gráfico mensal de familiares	11
4.18	Gráficos polares de familiares.	11
4 19	Mapa de calor de familiares	11

# LISTA DE TABELAS

# LISTA DE ALGORITMOS

# LISTA DE ACRÔNIMOS

CSS Cascading Style Sheets - Folhas de Estilo em Cascatas

DNS Domain Name System - Sistema de nome de domínio

HTML HiperText Markup Language - Linguagem de Marcação de Hipertexto

# 1 INTRODUÇÃO

As inúmeras transformações digitais e a pandemia de Covid-19 fizeram com que as dinâmicas de comunicação interpessoal sofressem profundas alterações e adaptações para enfrentarmos os novos desafios. Diferentes mensageiros vêm ganhando cada vez mais usuários e se consolidando no mercado, sendo o atual líder o WhatsApp. De acordo com pesquisa realizada pela Opinion Box em parceria com o site de notícias Mobile Time, 99% dos brasileiros que possuem *smartphones* têm o referido aplicativo instalado [1].

### 1.1 DEFINIÇÃO DO PROBLEMA

Apesar do WhatsApp ser o mensageiro mais utilizado no Brasil, o aplicativo não supre as necessidades de informações estatísticas sobre uma conversa pessoal, emergindo daí a motivação para a elaboração deste trabalho. Além disso, as ferramentas existentes que propõem-se a resolver este problema apresentam visualizações simples e estatísticas tanto quanto rasas como resultado de suas análises, como por exemplo: gráficos mal formatados e com cores destoantes, poucas informações oferecidas, ausência de estatística de tempo gasto na conversa, etc. Portanto, detectou-se espaço para melhoria.

# 2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta um resumo sobre os fundamentos teóricos utilizados como referência para este trabalho. Também são apresentados trabalhos relacionados e um descritivo da infraestrutura desenvolvida para a criação do site Analizap.

#### 2.1 INFRAESTRUTURA DO PROJETO

#### 2.1.1 Domínio

Todo site precisa de uma infraestrutura básica para estar online e um dos aspectos mais importantes e lembrados pelos usuários é o nome do domínio, que nada mais é do que o endereço digitado no navegador de internet para acessar a página desejada. O domínio do site aqui apresentado é <a href="https://analizap.tk">https://analizap.tk</a>>.

### 3 METODOLOGIA

Este capítulo versa sobre as metodologias aplicadas para o desenvolvimento do projeto. Primeiramente implementou-se um protótipo utilizando o Jupyter Notebook para facilitar a visualização e desenvolvimento dos códigos Python. Em seguida, segmentou-se o código em quatro grandes arquivos - parsers, visualizations, dataframe-generator e statistics - que conectam-se diretamente ao front-end (desenvolvido na mesma linguagem), onde são apresentados os resultados numa página web.

### 3.1 INTRODUÇÃO

Inicialmente realizou-se uma pesquisa de mercado para levantar as ferramentas existentes que fazem análise de conversas do WhatsApp. Para a busca de sites foi utilizado o Google, para aplicativos, a Google Play Store (Android) e App Store (iOS) e os trabalhos acadêmicos foram encontrados utilizando o Google Scholar. Foi feito um apanhado das principais e mais relevantes informações apresentadas pelos artigos e serviços relacionados, compiladas na lista a seguir:

### **4 RESULTADO**

Aqui expõem-se os resultados obtidos pelos algoritmos desenvolvidos neste trabalho a fim de exemplificar a proposta do produto desenvolvido.

O resultado obtido ao final do projeto foi um produto que objetiva informar o usuário sobre curiosidades e fatos de uma conversa do WhatsApp entre duas pessoas com total segurança, anonimidade e de forma gratuita. O tempo de processamento para gerar as estatísticas em todos os casos testados foi inferior a 1 minuto (conversas com até 200 mil mensagens), sendo todos os gráficos interativos, permitindo a função *zoom*, exportação do gráfico como imagem e, nos casos em que se aplica, a filtragem dos dados exibidos no gráfico.

O site Analizap conta com uma página de exemplo interativa de resultados que demonstra o produto final entregue aos usuários do serviço. Para obter os resultados para uma conversa pessoal, é preciso exportar o *chat* escolhido e submeter o documento de texto no formulário apresentado na Figura 4.1, disponível na página inicial do site.



Figura 4.1: Formulário de envio da conversa a ser analisada no site Analizap.

Para ilustrar o resultado final com casos reais, foram selecionadas três conversas previamente autorizadas com participantes que configuram classes distintas de conversas: relacionamento amoroso, profissional e familiar. As capturas de tela do site podem ser vistas abaixo para cada uma das classes de conversa. Por fim, é apresentada uma análise sobre as características intrínsecas de cada tipo de conversa e algumas diferenças entre elas.

Um detalhe importante é que o gráfico de palavras mais utilizadas não é apresentado por privacidade. Além disso, no gráfico intitulado "Quantidade de mensagens e tempo de conversa por dia", nos casos de uma conversa profissional e familiar, os exemplos utilizados são de conversas de muitos meses e, por isso, apresentamos apenas uma estatística para facilitar a visualização

das informações: a quantidade de mensagens enviadas (barras laranjadas). Não são exibidas as estatísticas "Minutos" (tempo gasto na conversa, barras azuis) e a "Mensagens por minuto" (velocidade das mensagens trocadas, linha cinza escuro).

#### 1. Relacionamento amoroso:



Figura 4.2: Estatísticas descritivas de namorados.

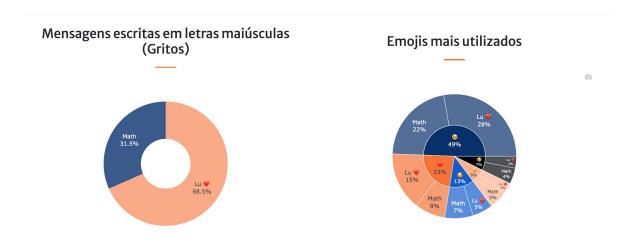


Figura 4.3: Gráficos de pizza de namorados.

#### Quantidade de mensagens e tempo de conversa por dia



Figura 4.4: Série temporal de namorados.



Figura 4.5: Gráfico mensal de namorados.

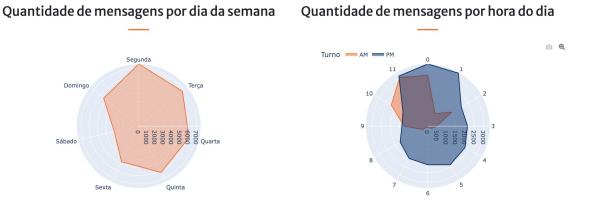


Figura 4.6: Gráficos polares de namorados.

#### Mapa de calor de dia da semana x hora do dia



Figura 4.7: Mapa de calor de namorados.

#### 2. Relacionamento profissional:



Figura 4.8: Estatísticas descritivas de colegas de trabalho.

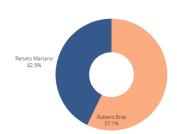




Figura 4.9: Gráficos de pizza de colegas de trabalho.

Quantidade de mensagens e tempo de conversa por dia \_\_\_\_

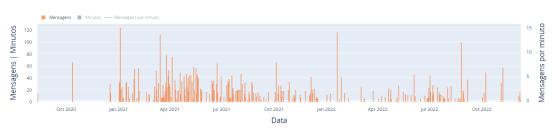


Figura 4.10: Série temporal de colegas de trabalho.

Quantidade de mensagens por mês

—

Septentification de la mensagens por mês

Describination de la mensagens por mensagen

Figura 4.11: Gráfico mensal de colegas de trabalho.

#### Quantidade de mensagens por dia da semana

#### Quantidade de mensagens por hora do dia



Figura 4.12: Gráficos polares de colegas de trabalho.



Figura 4.13: Mapa de calor de colegas de trabalho.

#### 3. Relacionamento familiar:



Figura 4.14: Estatísticas descritivas de familiares.

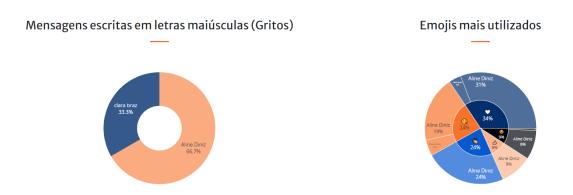


Figura 4.15: Gráficos de pizza de familiares.



Figura 4.16: Série temporal de familiares.

#### Quantidade de mensagens por mês



Figura 4.17: Gráfico mensal de familiares.



Figura 4.18: Gráficos polares de familiares.

Mapa de calor de dia da semana x hora do dia



Figura 4.19: Mapa de calor de familiares.

Conforme é possível observar nas Figuras 4.1 até a Figura 4.18, existem diferenças consideráveis entre os gráficos e estatísticas. Essas diferenças serão examinadas nos tópicos abaixo:

Total de mensagens da conversa:
 A conversa de exemplo de namorados possui uma quantidade maior de mensagens enviadas quando comparada com as demais classes, apesar de ser a mais recente de todas. Isso

pode ser explicado pela intensidade de um relacionamento amoroso, que costumeiramente envolve troca de mensagens diárias. Isso também é refletido na média de mensagens trocadas por dia, que é de 211 mensagens para o caso dos namorados, 11 para o relacionamento profissional e 5 no caso do relacionamento familiar.

#### • Quantidade de dias conversados:

Percentual de dias em que houve conversa desde a primeira mensagem enviada até a data de exportação:

Relacionamento amoroso: 84.8%
Relacionamento profissional: 29.5%
Relacionamento familiar: 28.51%

Conforme citado no item acima, pela dinâmica intrínseca de um relacionamento amoroso, a proporção de dias conversados é bem maior do que os outros casos analisados.

#### • Dia mais ativo:

As estatísticas de cada classe de conversa utilizada como exemplo são:

Relacionamento amoroso: 1.511 mensagensRelacionamento profissional: 124 mensagens

- Relacionamento familiar: 145 mensagens

Essa quantidade destoante da quantidade de mensagens trocadas em um único dia no caso da conversa entre namorados pode ser explicada pelo fato de no início do relacionamento haver um período em que o casal se conhece melhor, passando muitas horas conversando seguidamente.

#### • *Emojis* utilizados:

No caso do relacionamento amoroso e do familiar, os *emojis* enviados normalmente envolvem expressões de amor e de carinho. Para o caso do relacionamento entre colegas de trabalho, os *emojis* utilizados são mais descontraídos, um fato interessante é que apenas um dos participantes desta conversa envia *emojis*.

#### • Série temporal:

Pode-se observar com facilidade os *outliers*: para o caso do relacionamento amoroso, notase um pico evidente na metade do mês de agosto. Essa anomalia pode ser explicada por uma viagem que ocorreu no dia 7 de agosto em que os namorados ficaram separados por dez dias. Percebe-se que o mês de agosto foi marcado por essa distância extraindo essa informação do gráfico, pois fica nítida a quebra do padrão dos meses anteriores.

#### • Quantidade de mensagens por dia da semana:

O *chat* entre colegas de trabalho apresenta um pico de mensagens trocadas na segunda-feira, comportamento esperado já que não há expediente durante o final de semana. A conversa entre namorados tem uma queda de mensagens trocadas na sexta e no sábado, dias que normalmente eles estão juntos. Já para o caso da conversa entre mãe e filha (relacionamento

familiar), há uma queda considerável das mensagens trocadas no final de semana, já que ambas estão juntas pois residem na mesma casa.

• Quantidade de mensagens por hora do dia:

O gráfico destoante dos outros neste caso é o da conversa entre colegas de trabalho, é possível perceber claramente a concentração das mensagens durante o horário de expediente, entre 9h e 17h. Os outros exemplos possuem mensagens distribuídas ao longo de todo o dia.

### 5 CONCLUSÃO

Aqui analisam-se os resultados obtidos pelos algoritmos desenvolvidos neste trabalho, a fim de explorar o potencial da análise sugerida.

O Analizap foi concebido com o propósito de extrair *insights* e curiosidades sobre conversas entre duas pessoas no aplicativo WhatsApp.

# REFERÊNCIAS BIBLIOGRÁFICAS

1 COSTA, L. M. A pirataria do nome de domínio na internet. *Revista de Administração de Empresas*, v. 41, n. 1, p. 45–53, mar. 2001. ISSN 0034-7590. Disponível em: <a href="http://www.scielo.br/scielo.php?script=sci\_arttext&pid=S0034-75902001000100006&lng=pt&tlng=pt">http://www.scielo.br/scielo.php?script=sci\_arttext&pid=S0034-75902001000100006&lng=pt&tlng=pt</a>.