

# Modelo para o processamento de Linguagem Natural

Chat-BTG

# Objetivo da Sprint



Pré-processamento e  
limpeza da Base de dados



Modelo Bag of Words

# AGENDA



**Retrospectiva  
Sprint 1**



**Entendimento da  
base**



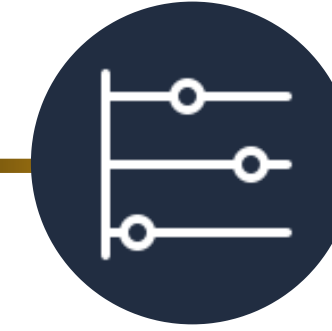
**Análise descritiva**



**Pré-  
processamento**



**Bag of Words**



**Cronograma**



**Próximos passos**

01

# Retrospectiva Sprint 1

# Retrospectiva da Sprint 1

---



Persona



User Stories



Canvas Proposta de Valor



Solução

02

# Entendimento da Base de Dados

# Entendimento da Base de Dados

---



Anomalia



Probabilidade de Anomalia



Processado

03

# Análise descritiva



# Análise Quantitativa

---



Quantidade de linhas em cada colunas



Quantidade de *emojis* no *dataset*



Quantidade de comentários por autores que mais aparecem

# Análise Qualitativa

---

## *Wordcloud*



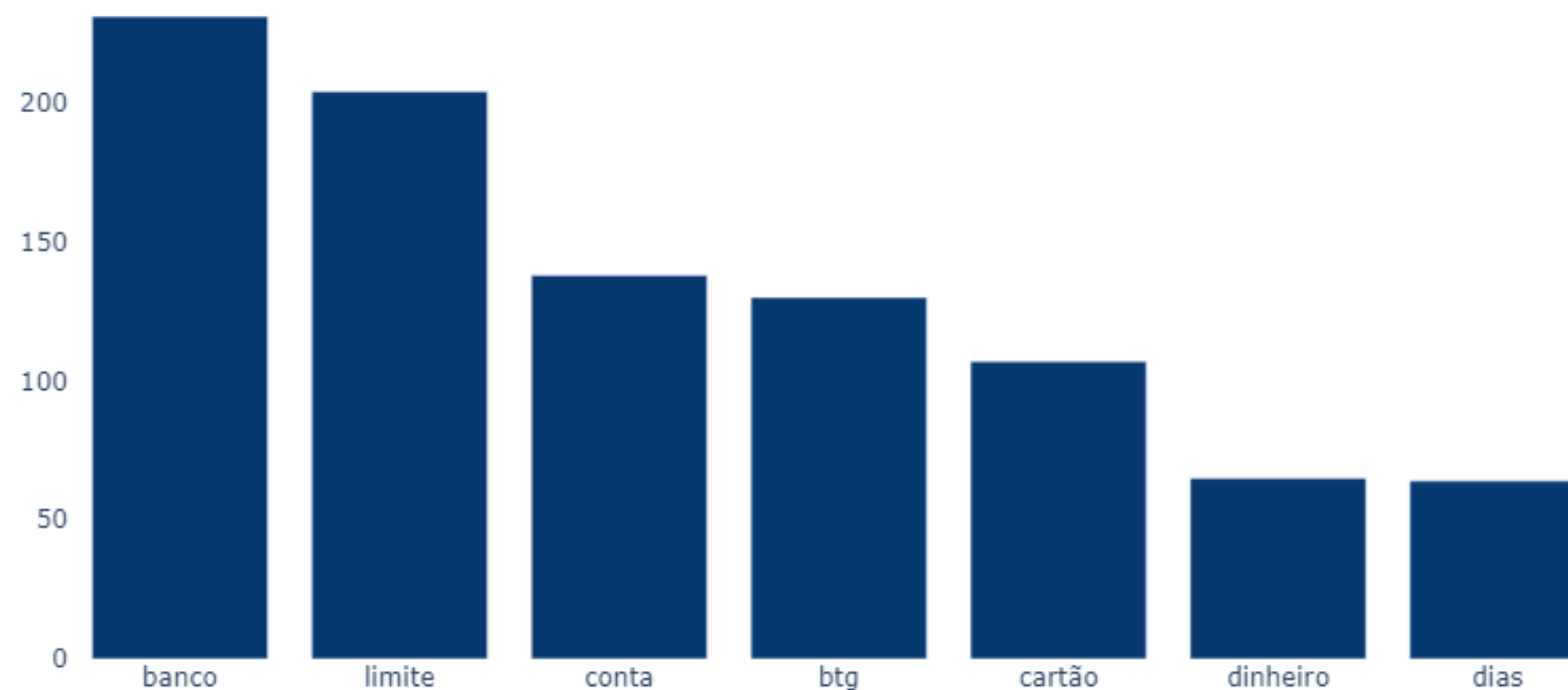
# Análise Qualitativa

---



Palavras que mais aparecem por tipo de *post*

Frequência de Palavras



REELS

1525

# Análise Qualitativa



Relação determinística entre as colunas Anomalia e Comentário

		texto	anomalia
anomalia			
-1.0	1799	👊	-1.0
	687	Olá, Lucas! Tudo bem? \n\nSentimos muito pelo ...	-1.0
	1347	Fica em casa	-1.0
	1301	Olá, Vitor! Tudo bem?\n\nFicamos felizes com s...	-1.0
	3205	O canal de atendimento do app é péssimo!! Fiqu...	-1.0
0.0	2092	mention	0.0
	1557	reply	0.0
	3216	mention	0.0
	2648	comment	0.0
	1961	mention	0.0

# Análise Qualitativa



## Uso de *Emoji* na Base de dados

6025 emojis.

emoji percentage

👏	0.289761
💙	0.114259
🚀	0.100174
😊	0.059187
🔥	0.034024
😍	0.032284
❓	0.027219
❤️	0.021522

🙌	0.020889
😂	0.018199
❓	0.014243
😞	0.008862
🙏	0.008546
🎄	0.008546
✓	0.008229

04

# Pré-processamento

# Pré-processamento

---

- ✓ Colocar letras minúsculas e *tokenização*
- ✓ Retirada de *stop words*
- ✓ *Stemming*

05

# *Bag of Words*



# Modelo bag of words

---

- ✓ Técnica de processamento
- ✓ Representação de um texto como um vetor de frequência de palavras
- ✓ Uso em classificação de texto, análise de sentimento e recuperação de informações.

# *Bag of Words* e Resultados

```
[ [1 0 1 0 0 0 0 0 0 0]
  [0 0 1 0 0 1 0 0 0 0]
  [0 0 1 0 0 0 0 1 0 0]
  [0 0 1 0 0 0 1 0 0 0]
  [0 0 1 1 0 0 0 0 0 0]
  [0 0 1 0 0 0 0 0 1 0]
  [0 1 1 0 0 0 0 0 0 0]
  [0 0 1 0 1 0 0 0 0 0]
  [0 0 1 0 0 0 0 0 0 1]]
```

contém a palavra: 1

não contém a palavra: 0

top 10 words:

- btg
- conta
- \$
- sempre
- dias
- enviamos
- ano
- bom
- ter
- R\$

06

# Cronograma

Etapa	Status	Sprint
Entendimento de Negócios	Feito	1
Mapeamento de Persona	Feito	1
Modelo de Bag of Words	Feito	2
Modelo utilizando Word2Vec e Análise de sentimento	Em progresso	3
Pré-processamento e desenvolvimento da proposta	A iniciar	4
Solução Final em formato MVP	A iniciar	5

# Muito obrigado pela atenção!

Estamos disponíveis para maiores esclarecimentos.