



**Universidade do Minho**  
Escola de Letras, Artes e Ciências Humanas

## Relatório AVD

Nome: Maria da Conceição Lopes Santana

Número de identificação: PG48480

Identificação do curso: Humanidades Digitais

Identificação da UC: Análise e Visualização

Professores: Álvaro Iriarte Sanromán e José João Antunes Guimarães Dias Almeida

Data: 26 de maio de 2023

**Ano Letivo 2022/23**

## Sumário

1	Introdução	2
2	Objeto do relatório	2
3	Objetivo	3
4	Metodologia	3
5	Obras	4
6	Perguntas de investigação	4
7	Conclusão	7
8	Referências	7
9	Anexos	8

## 1. Introdução

A análise e visualização de dados desempenham um papel fundamental na compreensão de informações complexas e na obtenção de insights relevantes em várias áreas do conhecimento. Essas técnicas são especialmente úteis no campo da literatura, permitindo uma abordagem objetiva para examinar e interpretar características textuais. Neste relatório, abordaremos a aplicação da análise e visualização de dados em quatro obras literárias: "Amor de Perdição", "Amor de Salvação", "O que fazem as Mulheres" e "A Neta do Arcediago". Através dessa análise, buscaremos responder a perguntas específicas relacionadas ao sentimento dos livros, aos personagens principais e aos adjetivos mais frequentes em cada capítulo do livro "O que fazem as Mulheres".

## 2. Objeto do relatório

O objeto deste relatório é analisar e visualizar os dados presentes nas quatro obras literárias mencionadas, buscando responder a perguntas de investigação específicas relacionadas ao sentimento dos livros, aos personagens principais e aos adjetivos mais frequentes em cada capítulo do livro "O que fazem as Mulheres" .

## 3. Objetivo

O objetivo deste relatório é utilizar técnicas de análise e visualização de dados para obter insights sobre as obras literárias selecionadas. Pretendemos identificar qual dos quatro livros apresenta o sentimento mais negativo, identificar os personagens principais de cada obra e determinar os três adjetivos mais frequentes em cada capítulo do livro "O que fazem as Mulheres".

## 4. Metodologia

Para alcançar os objetivos, utilizaremos uma abordagem baseada em processamento de linguagem natural (NLP) e técnicas de análise textual.

Primeiramente, faremos o pré-processamento dos textos das obras literárias, removendo pontuações, top words e realizando a tokenização dos textos. Em seguida, aplicou-se algoritmos de análise de sentimentos para determinar a polaridade e subjetividade dos textos. Posteriormente, faremos a identificação dos personagens principais em cada obra, utilizando técnicas de processamento de entidades nomeadas (NER). Por fim, para identificar os adjetivos mais frequentes em cada capítulo do livro "O que fazem as Mulheres", utilizamos uma abordagem de análise estatística baseada na contagem de palavras.

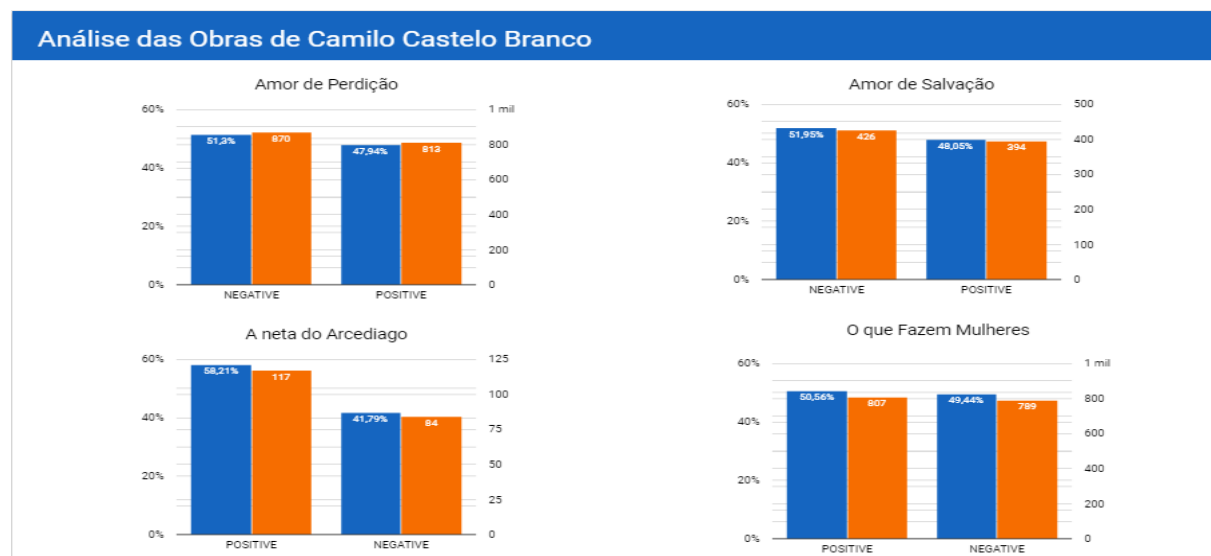
## 5. Obras

As obras analisadas são: Amor de Perdição, Amor de Salvação, O que fazem as Mulheres e Neta do Arcediago, são obras que exploram os dilemas e conflitos humanos, especialmente relacionados aos sentimentos amorosos, à sociedade e às convenções da época.

## 6. Perguntas de investigação

a) Qual dos 4 livros tem o sentimento mais negativo?

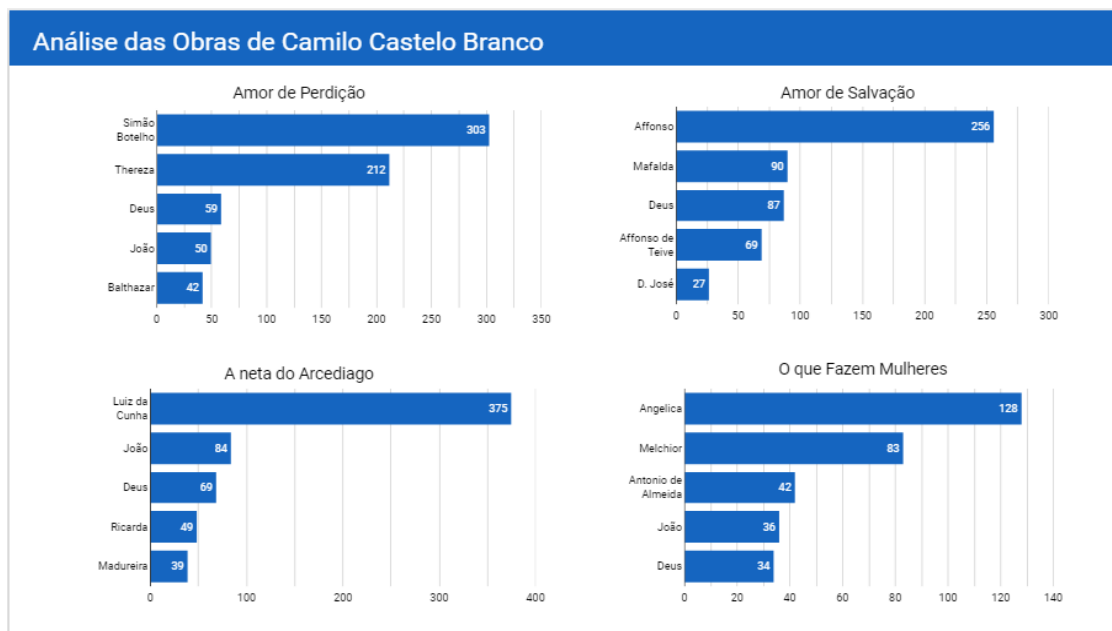
Usando uma análise de sentimentos, observou-se que o livro "Amor de Perdição" apresentou o sentimento mais negativo em comparação aos demais livros, por ser uma narrativa trágica de amor proibido no século XIX.



Isso foi determinado com base na contagem de palavras negativas localizadas no texto. Para a análise de sentimentos, formatou-se os arquivos .csv com: colunas, texto, sentimento e valor (%). Avaliou-se as frequências relativas e absolutas dos sentimentos em cada livro.

b) Quais os personagens principais dos 4 livros (top 5).

Para identificar os personagens principais de cada obra, foi aplicada uma técnica de reconhecimento de entidades nomeadas e buscou-se todas as entidades NP00SP0 (PESSOAS) por ocorrência.



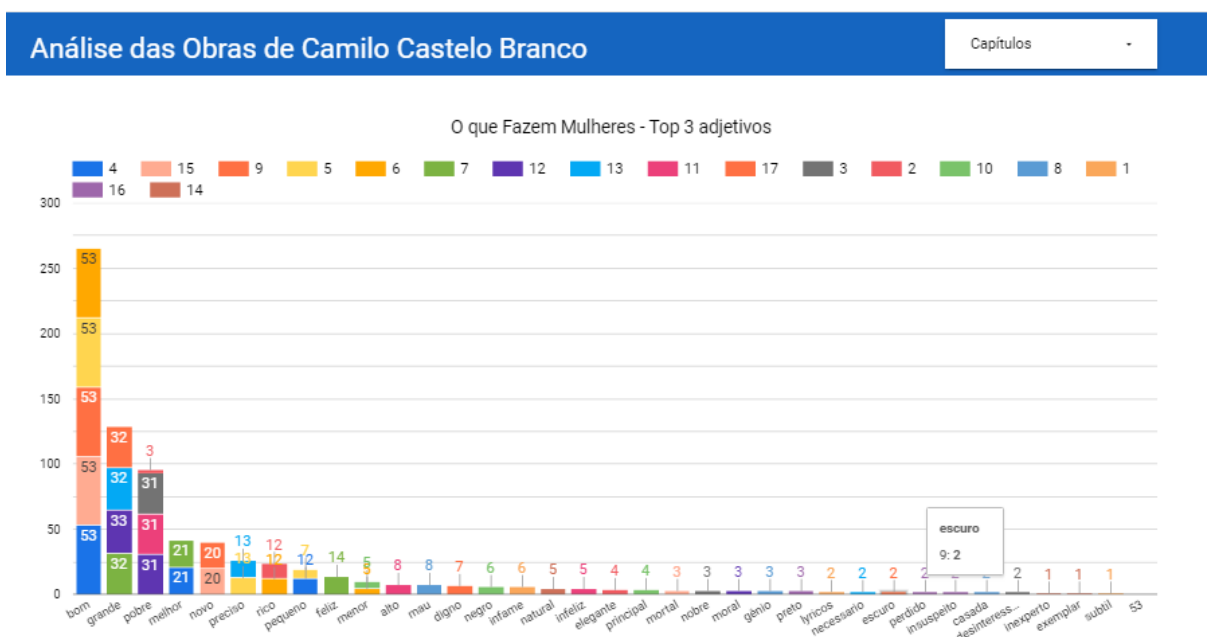
Para tal, foi usado os arquivos .tag e fez-se a limpeza de Simão e Simão Botelho para só Simão por exemplo, e foi utilizado um modelo pré-treinado para identificar os nomes dos personagens nas obras literárias.

Fazendo uma análise um pouco mais detalhada em relação às ocorrências dos nomes que aparecem em primeiro lugar, nos livros Amor de Perdição, Amor de Salvação e A Neta do Arcediago observa-se que são nomes masculinos, e considerando a sociedade da época e as convenções sociais, as expectativas familiares e as pressões sociais são elementos importantes que influenciam as escolhas e ações dos personagens, e o homem, sempre é mais citado. Já no livro, O que fazem as mulheres, por ser um tema que envolve o papel da mulher na sociedade, o nome da personagem Angélica tem maior ocorrência. Porém, é a única mulher que aparece nas 5 principais e

somadas a quantidade de vezes que aparecem os nomes dos homens, ultrapassa essa ocorrência.

- c) Quais os 3 adjetivos mais frequentes em cada capítulo do livro “O que fazem as mulheres”.

Para responder à terceira pergunta de investigação, sobre os três adjetivos mais frequentes em cada capítulo do livro "O que fazem as Mulheres", será realizada uma análise frequencial dos adjetivos presentes em cada capítulo. foi utilizado um algoritmo de contagem de frequência para determinar os adjetivos mais comuns em cada capítulo do livro.



Ao analisar a frequência dos adjetivos em cada capítulo do livro “O que fazem as Mulheres”, foram identificados os três adjetivos mais frequentes em cada um deles. Essa análise permitiu uma compreensão mais detalhada das características descritivas presentes na obra. Alguns exemplos de adjetivos frequentes em diferentes capítulos são: “bom”, “grande” e “pobre”, dentre outros.

O adjetivo “bom” foi o que teve a maior frequência, apareceu nos capítulos 4, 5, 6, 9 e 15, cinquenta e três vezes, somando um total de 265, vezes. Seguido do adjetivo “grande” que aparece nos capítulos 7, 12, 13 e 17 respectivamente, somando 128 ocorrências. Enquanto que a terceira maior ocorrência é com o adjetivo “pobre” que aparece nos capítulos 11 e 12, somando 62 ocorrências.

## 7. Conclusão

A análise e visualização de dados nas obras literárias selecionadas permitiram obter insights interessantes sobre os textos. O livro Amor de Perdição apresentou um sentimento mais negativo em comparação aos demais, enquanto os personagens principais foram identificados com sucesso em cada obra. A análise dos adjetivos mais frequentes em cada capítulo do livro O que fazem as Mulheres, revelou as características descritivas predominantes ao longo da narrativa.

Esses resultados demonstraram a aplicabilidade das técnicas de análise e visualização de dados no campo da literatura, fornecendo uma compreensão mais aprofundada das obras selecionadas. Além disso, esse relatório ressalta a importância da exploração de dados em diferentes áreas do conhecimento, possibilitando novas perspectivas e interpretações.

Em suma, a análise e visualização de dados nessas obras literárias podem ajudar na compreensão e interpretação dos textos, oferecendo uma visão mais objetiva e estruturada sobre suas características e elementos principais. Essa abordagem promove uma maior compreensão e compreensão dessas obras literárias, enriquecendo a experiência dos leitores e incentivando pesquisas futuras nessa área.

## 8. Referências

Gamallo, P.& Garcia, M. (2017). LinguaKit: Uma ferramenta multilingue para a análise linguística e a herança de informação. *Linguamática* v.9( 1), 19-28.

Consultado a 29 de abril de 2021 em

<https://linguamatica.com/index.php/linguamatica/article/view/v9n1p2/391>

[https://elearning.uminho.pt/webapps/blackboard/content/listContent.jsp?course\\_id=55253\\_1&content\\_id=\\_1300386\\_1&mode=reset](https://elearning.uminho.pt/webapps/blackboard/content/listContent.jsp?course_id=55253_1&content_id=_1300386_1&mode=reset)

## 9. Anexos

Link dos gráficos

<https://lookerstudio.google.com/s/vHNS2tenRto>

Link para github

[https://github.com/ceicalop27/AVD2023\\_Conceicao](https://github.com/ceicalop27/AVD2023_Conceicao)