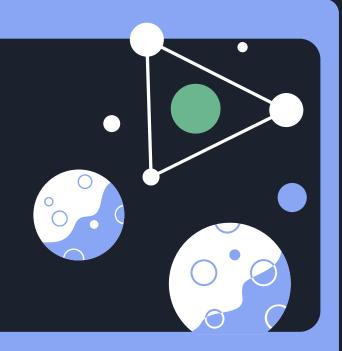
COMPASS Análise de NPS



GUIA DE CONTEÚDO

01

EDA

Obtenção e Preparação dos Dados

03

NLP

Processos de linguagem natural e modelos

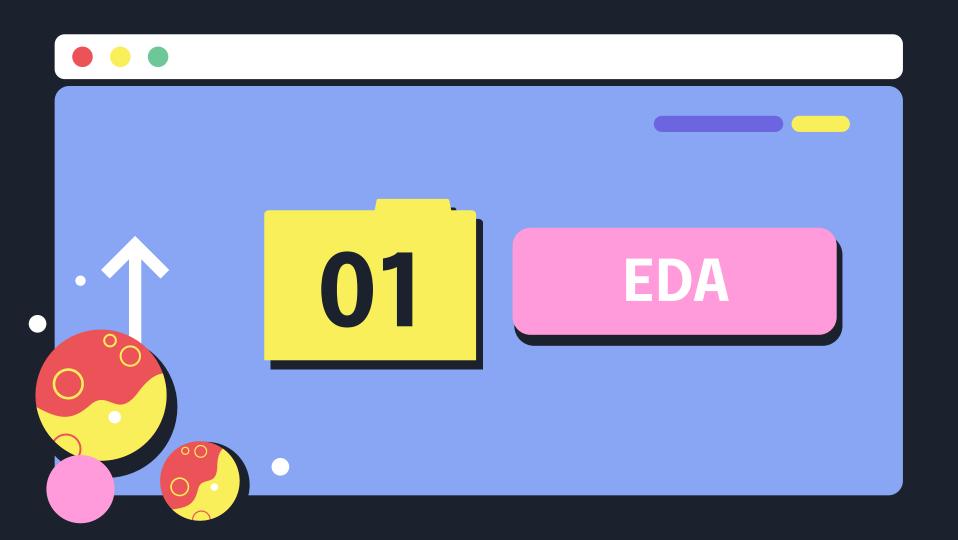
02

INSIGHTS

Desnvolvimento e compreensão das váriaveis e causalidades 04

FINALE

Conclusão das observações e conhecimentos adquiridos no processo







Juntamos diversas tabelas de ligações com a tabela NPS, assim tendo a avaliação de cada interação

Visualization

Disposição de uma variedade de gráficos e ilustrações sobre o conjunto de dados

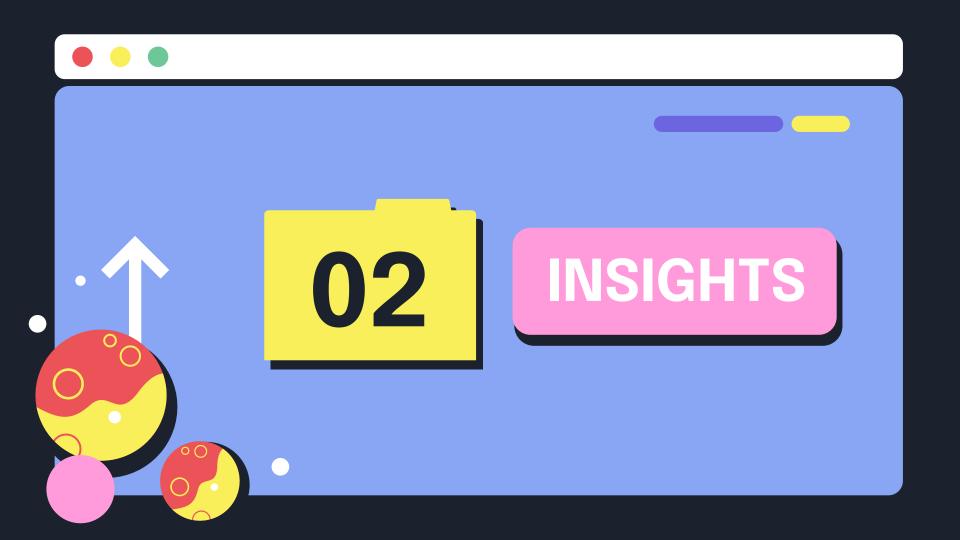
Extraction

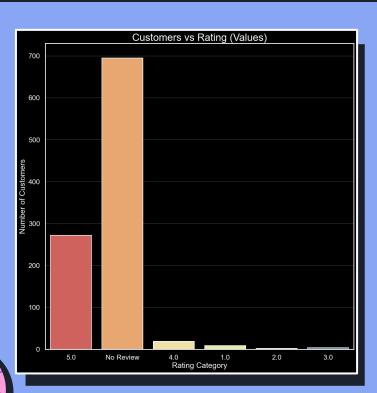
Dados obtidos através de arquivos .csv/.json contendo a transcrição das ligações

Analysis

Realizado os estudos em cima dos dados brutos (missing data, imbalance, etc)

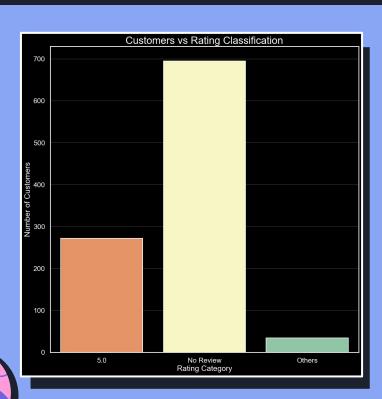






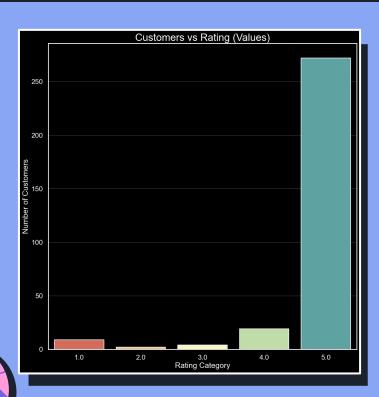
A problemática da não-avaliação ou avaliação extremamente positiva

A maior parte dos dados disponíveis não possui sequer uma avaliação por parte do usuário, e a maior parte daqueles que de fato avaliaram são nota 5



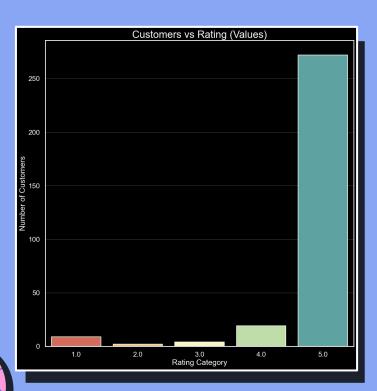
A problemática da não-avaliação ou avaliação extremamente positiva

A maior parte dos dados disponíveis não possui sequer uma avaliação por parte do usuário, e a maior parte daqueles que de fato avaliaram são nota 5



Ter muitas avaliações positivas por si só não é um problema.

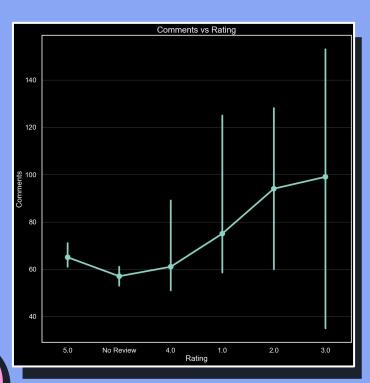
É bom que tenhamos um bom número de avaliações positivas, mas isso pode nos dizer que os clientes procuram responder apenas em determinadas ocasiões.



Os números podem indicar que as avaliações podem não ser feitas de acordo com a experiência do usuário

Algumas possíveis situações:

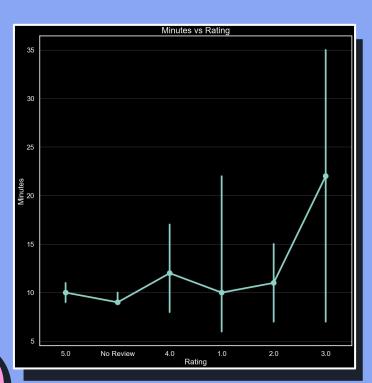
- Cliente fica muito feliz com o a resolução de seu problema e avalia positivamente o atendimento
- Cliente fica muito bravo por seu problema não ter sido resolvido e desliga o telefone sem realizar a avaliação



Ainda utilizando as avalições individuais, podemos traçar algumas características

Com base na mediana do número de comentários por Nota de avaliação temos:

1. Clientes com baixas notas tendem a conversarem mais ou trocarem mais informação nos atendimentos, pode estar relacionado a tentativas de resolução de problemas



Ainda utilizando as avalições individuais, podemos traçar algumas características

Com base na mediana do número de de minutos gastos na chamada por Nota de avaliação temos:

1. Seguindo o mesmo padrão de comentários, temos que o cliente tende a se tornar um passivo quanto mais minutos de chamada temos, as outras avaliações são mais constantes em seu tempo.

Classificação NPS utilizada



Imagem de:

https://bicsp.min-saude.pt/pt/biufs/3/30032/3141473/O%2OQUE%2OFAZEMOS/Relat%C3%B3rio%2Ode%2OAvalia%C3%A7%C3%A3o%2Odo%2OInqu%C3%A9rito%2Ode%2OSatisfa%C3%A7%C3%A3o%2OUtentes%2O2021.pdf

Classificação NPS (Promotores, passivos e detratores) de acordo com as notas

95%

Promotores

01%

Passivos

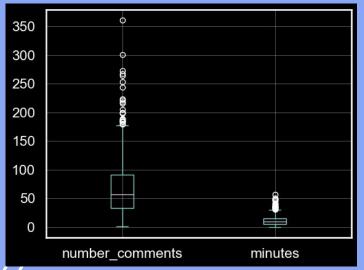
03%

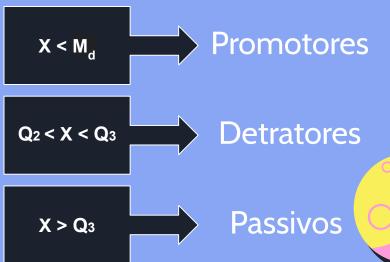
Detratores



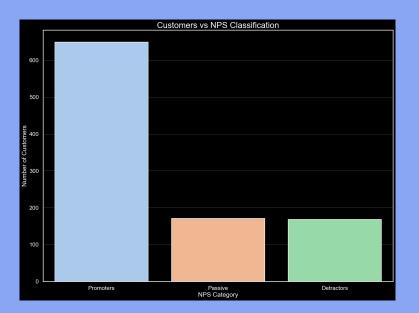
Com isso temos um Índice de NPS Geral de: 91,5 Que é Excelente!

Realizando um experimento para balancear nossos dados, foi feito uma classificação nos dados não classificados com base no número de comentários, sendo assim:





Com isso, teremos um conjunto de dados com valores com maior divergência.

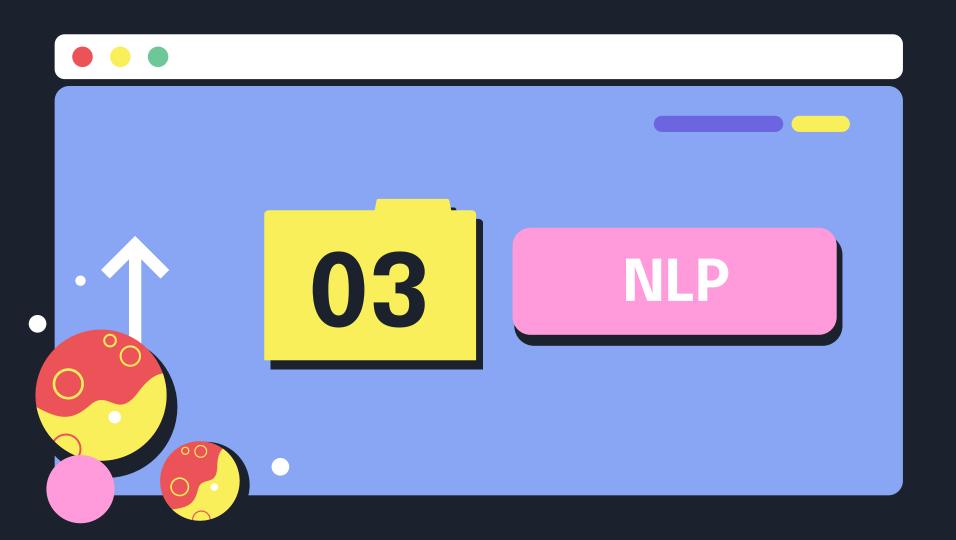


Nesse caso teremos um Índice de NPS Geral de: 48,6 Que ainda é muito bom!

Aproximação

É válido relembrar que esse é um valor estimado utilizando dados de comentário para preencher os dados faltantes no nosso conjunto de dados, isso pode não refletir exatamente o comportamento de clientes e suas avaliações





Nesta seção iremos trabalhar com NLP.

Estaremos trabalhando com os comentários feitos pelos clientes durante o atendimento





00

Extração

Extraímos os Dados textuais contidos nos arquivos csv/json

Corretor

Foi utilizado um corretor para reparar as frases dos comentários disponíveis

Vetorização

Utilizados o TF-IDF para vetorização do texto contido no corpus (comentários)

Predição

Tentaremos prever, com base no texto, a nota do usuário

Quais são os problemas?

Frases incoerentes

Muitas das frases contidas no nosso corpus possuem frases incompletas ou incoerentes, complicando todo o processo

Erros de digitação

MUITAS palavras incorretas e com letras faltando (por isso foi utilizado o corretor)

Dados desbalanceados

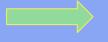
Mais de 90% dos dados são da categoria 5, o que prejudica bastante o treinamento do modelo pois temos pouquíssimos dados de algumas classes

Corrigindo Frases

Utilizamos a biblioteca Hunspell e PySpellChecker para a correção de frases com erros de digitação

Exemplo de Utilização:

"esta frae esta erada, por favr crrija caso for possive"



"esta frase esta errada por favor corrija caso for possivel"



Vetorização

Utilizamos a técnia do TF-IDF para vetorizar nosso texto com base na frequência de palavras e sua posição



Nosso Modelo de classificação ML no momento não atingiu bon resultados devido aos grandes problemas do dataset e falta de dados para as classes...

Testes de Input

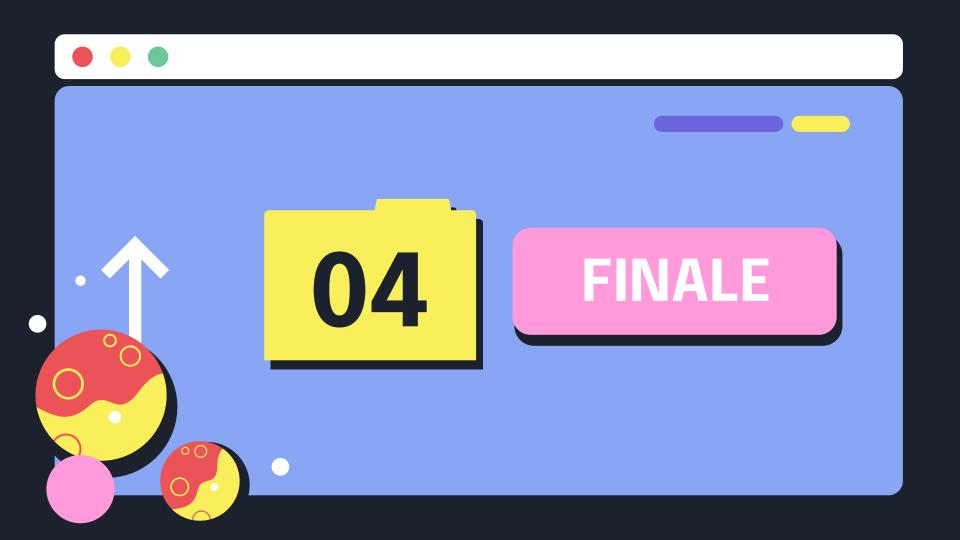
- "vou na agencia na sexta feira falar com o gerente"
- 2. "muito obrigado por resolver meu problema, amei seu atendimento"



Nota 4

Nota 5

Métricas	Valores
F1-Score	0.32
Precision	0.33
Recall	0.31



Apesar dos pesares...

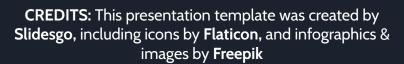
Ainda que o modelo não tenha tido um resultado muito bom, aqui estão propostas de melhorias futuras:

- Mais dados disponíveis para cada classe de avaliação (1-5) para um treino menos enviesado
- Localizar o CORE ou a ideia principal do sentimento do cliente na ligação e assim reduzir o texto necessário para o modelo.
- Usar técnicas como Stemming para transformar cada palavra em sua forma raiz
- Utilizar alguma técnica semelhante àquelas utilizadas pelo WordNet para encontrar sinônimos das palavras e assim aplicar Data Augmentation no nosso dataset, o problema aqui é a língua portuguesa.
- Mudar alguma regra de avaliação, talvez fazer com que o funcionário possa avaliar por outro meio além da pós-chamada





Alguma pergunta? victorgsbarbosa@gmail.com (41) 99176-7941



Please keep this slide as attribution



