

ANÁLISE DE DADOS DA GOOGLE PLAY STORE

Mateus Genaro

2 de março de 2020



INTRODUÇÃO

Identificação do problema e abordagens

01

METODOLOGIA

Ferramentas utilizadas e tratamento dos dados

02

RESULTADOS

Aplicações dos métodos e suas análises

03

CONCLUSÃO

A resposta para a pergunta central

04

01

INTRODUÇÃO

“O que faz os usuários gostarem dos aplicativos?”



A resposta está nos dados



BASES DE DADOS

Obtenção de informações qualitativas e quantitativas



MÉTODOS APROPRIADOS

Abordagens adequadas para uma análise efetiva



02

METODOLOGIA

METODOLOGIA

- **Principais Ferramentas:**

Python, Excel e Tableau



- **Tratamento e Limpeza dos dados**



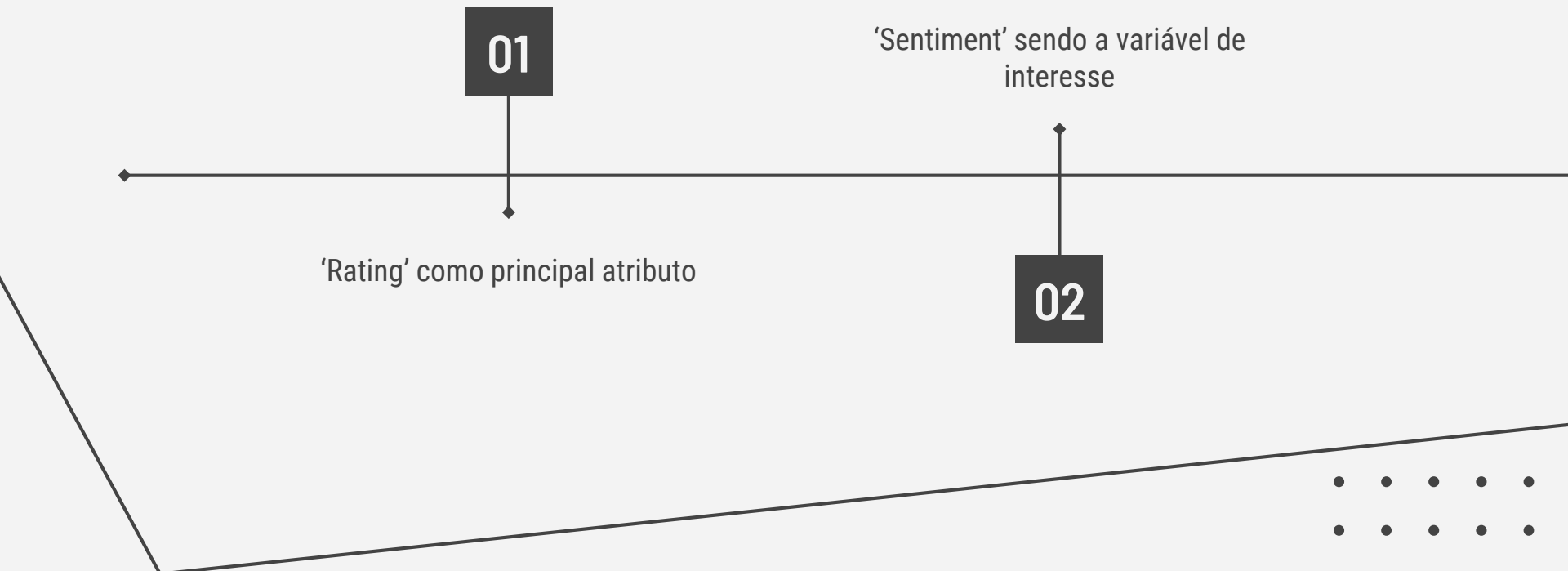
- **Modelagens:**

Naïve Bayes, Árvores de Decisão,

Florestas Aleatórias e K-Means



BASES DE DADOS

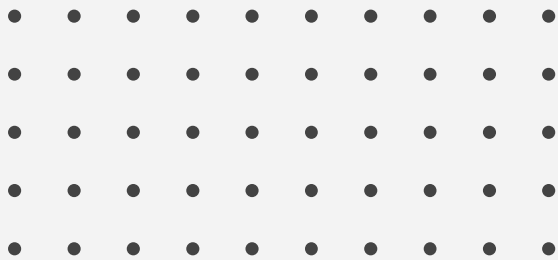




03

RESULTADOS





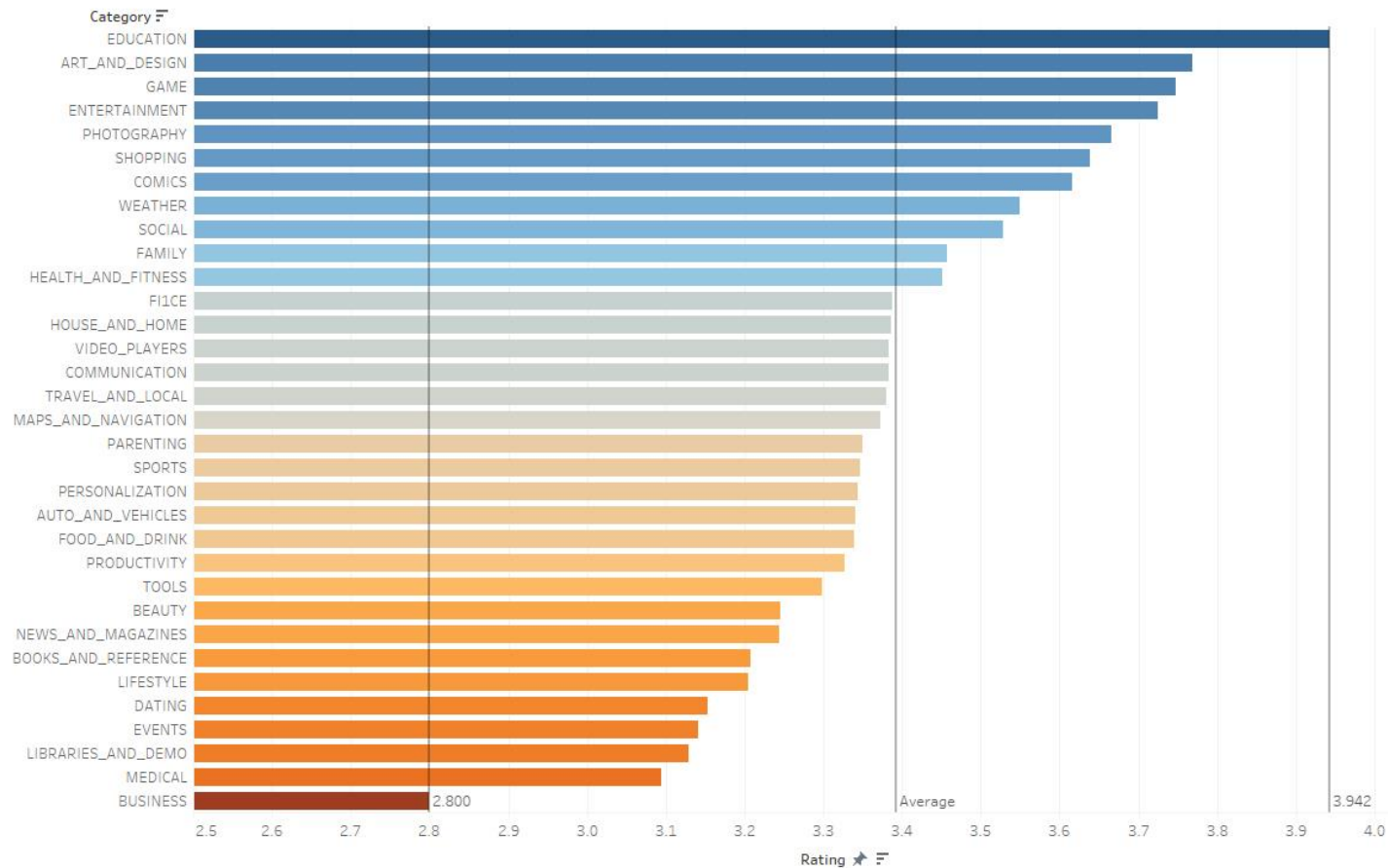
BASE DE DADOS 1

- A BASE DOS APLICATIVOS

Parâmetros estatísticos da base:

- Média total das notas médias por aplicativo: **3,40**
- Desvio padrão das notas médias: **1,09**
- Nota mínima: **1,00**
- Nota máxima: **5,00**
- Top 3 categorias com as maiores notas médias agrupadas por categoria:
Education (3,94); **Art and Design** (3,77); e **Game** (3,75)
- Top 3 categorias com as menores notas médias agrupadas por categoria:
Business (2,80); **Medical** (3,09); e **Libraries and Demo** (3,13)

Gráfico de Category por Rating



BASE DE DADOS 1



BASE DE DADOS 1

MODELOS E PRECISÕES

- Separação dos dados e aplicação dos modelos

■ **Naïve Bayes:** Precisão = 19,43%

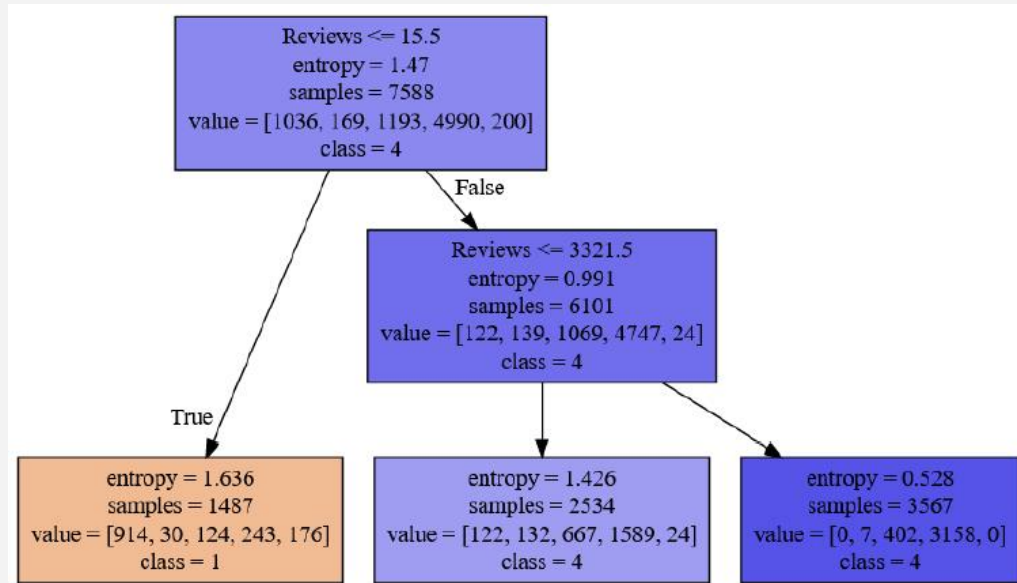
■ **Árvore de Decisão:** Precisão = 74,35%

■ **Florestas Aleatórias:** Precisão = 75,52%



BASE DE DADOS 1

Árvore de decisão



BASE DE DADOS 2

- A BASE DOS 'USER REVIEWS'

Média em/ Sentimento	Positivo	Neutro	Negativo
Polaridade	0,37	0	-0,26
Subjetividade	0,57	0,08	0,53



BASE DE DADOS 2

MODELOS E PRECISÕES

■ **Naïve Bayes:** Precisão = 96,64%

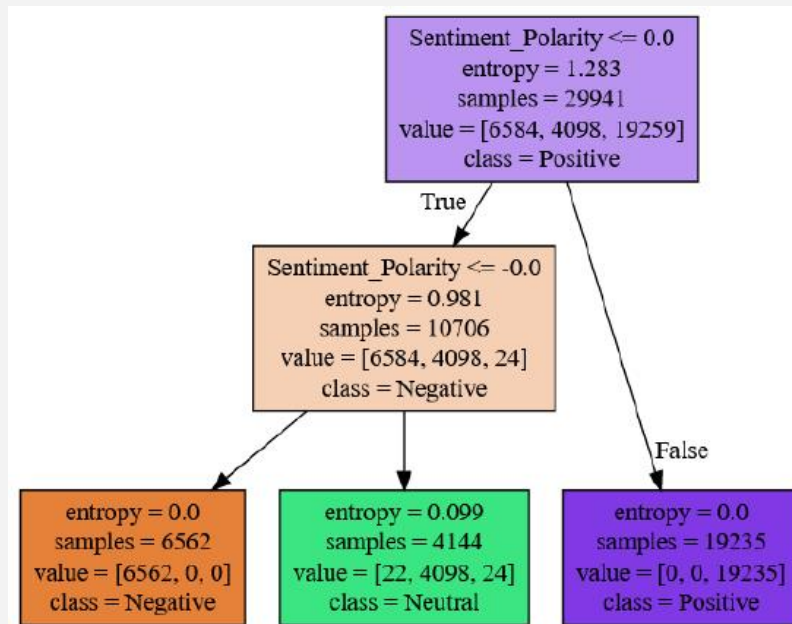
■ **Árvore de Decisão:** Precisão = 99,88%

■ **Florestas Aleatórias:** Precisão = 99,86%



BASE DE DADOS 2

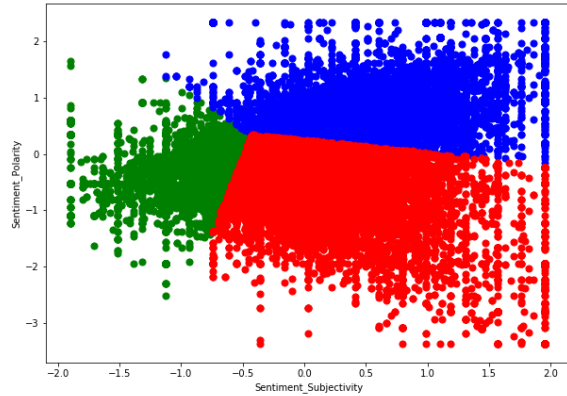
Árvore de decisão



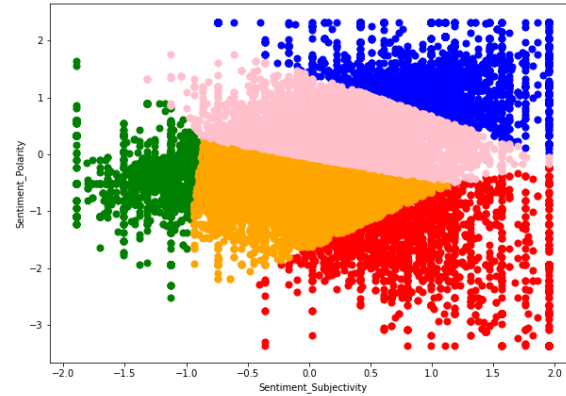
BASE DE DADOS 2

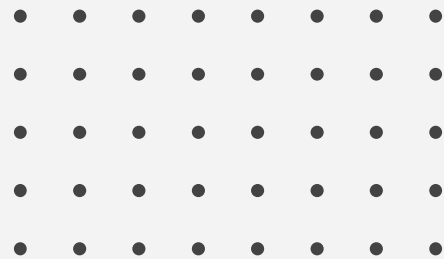
K-Means

3 clusters



5 clusters





A pergunta continua: “O que faz os usuários gostarem dos aplicativos?”

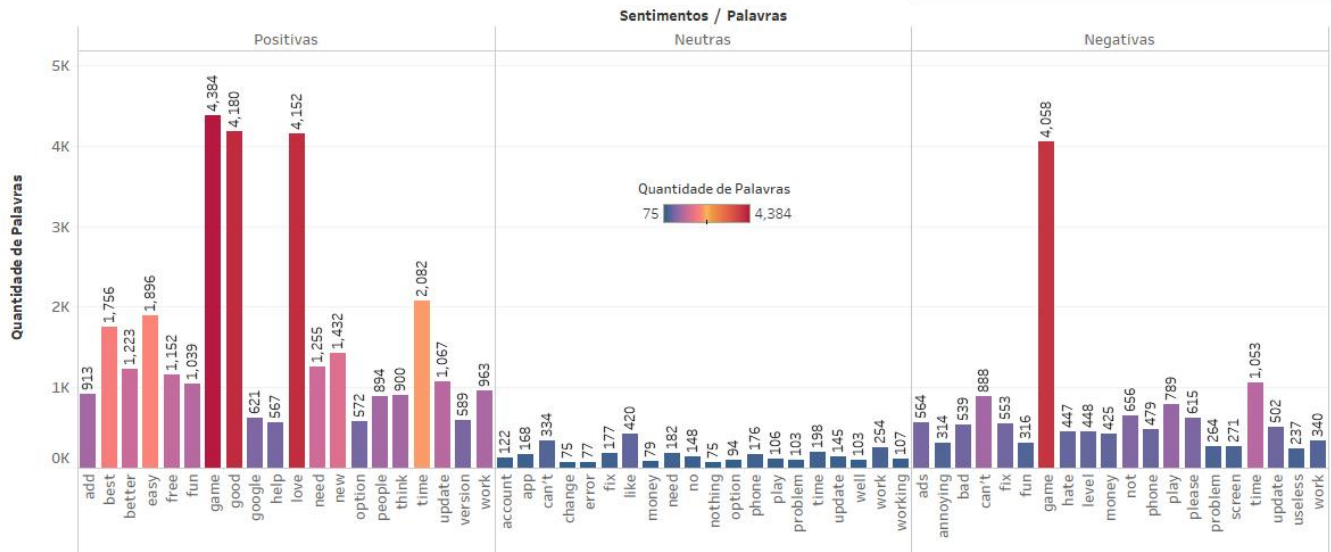
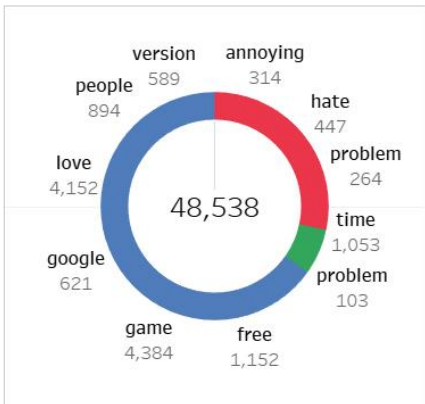


Palavras-chave

Análise qualitativa e tratamento da base de dados para decidir as palavras mais representativas

Legenda

Word cloud for 'Positivas' (Positive) featuring words such as: love, game, good, time, people, better, fun, free, update, work, problem, bad, fun, like, fix, phone, level, time, work, work, useless, account, problem, update, time, people, help, error, working, app, ads, play, money, new, best, play, need, can't, not, option, annoying, fix, add, please, hate, screen, option, change, nothing, need, think, money, can't, well, money.





04

CONCLUSÃO



"O que faz os usuários gostarem dos aplicativos?"

BASE DE DADOS 1

- Diferença de nota entre as categorias
- Relação direta do número de avaliações por aplicativo e a sua nota média respectiva

BASE DE DADOS 2

- Base de dados auto consistente
- Possibilidade de novas classificações
- Palavras-chave nos **comentários**



Obrigado!

Alguma pergunta?



genaro.mateus@gmail.com

+55 41 99890-2678