

The background of the slide is a black and white photograph of the New York Stock Exchange building. The image shows the classical architecture with large columns and the words "NEW YORK STOCK EXCHANGE" inscribed on the pediment. The text is overlaid on this image.

AGRUPAMENTO: ***BANK CUSTOMER SEGMENTATION***

[Dataset](#)

[Códigos](#)

Vitor Lopes Fabris, 769822
Jayme Sakae dos Reis Furuyama, 761044

Sumário

01

**Escolha do
conjunto de
dados**

02

**Estudo do
conjunto de
dados**

03

**Algoritmos de
agrupamento**

04

**Interpretação
dos resultados
obtidos**



01

Escolha do
conjunto de
dados

Escolha do conjunto de dados



Unsupervised Learning on Country Data

Poucos dados, atributos
bem comportados



Trending YouTube Video Statistics and Comments

Conjunto voltado para
tarefas de NLP



Data Science Job Salaries

Muitos atributos categóricos,
preferimos outro domínio



Spaceship Titanic

Agrupamento resultou em
clusters não significativos



Bank Customer segmentation

Atributos

- **TransactionID**
 - **CustomerID**
 - **CustomerDOB**
 - **CustGender**
 - **CustLocation**
 - **CustAccountBalance**
 - **TransactionDate**
 - **TransactionTime**
 - **TransactionAmount**
-

Alguns dos problemas no Dataset

- **Valores Nulos**
- **Datas Inválidas**



Estudo do
conjunto de
dados

02

MAS ANTES...

Atributos Tratados:

- **CustomerDOB**
- **TransactionDate**
- **TransactionTime**
- **CustLocation**

Atributos Não Modificados:

- **CustAccountBalance**
- **TransactionAmount**
- **CustGender**

Atributos descartados:

- **CustomerID**
- **TransactionID**

MAS ANTES...

TransactionDate e TransactionTime:

Estes dois atributos foram combinados de forma que se transformasse em um único atributo TimeStamp

Formato TransactionDate

DD/MM/YY

Formato TransactionTime

HourMinSec

Formato TimeStamp

Quantidade de segundos a partir do dia

01/01/1970 até a transação

MAS ANTES...

CustLocation

Esse atributo levanta um problema, existem 9275 cidades diferentes, assim para uma melhor visualização e não cair na maldição das dimensões

Consideramos apenas as 3 cidades com maiores transações:

- Mumbai
- Bangalore
- New Delhi

CustomerDOB

Utilizamos apenas o ano de nascimento do cliente

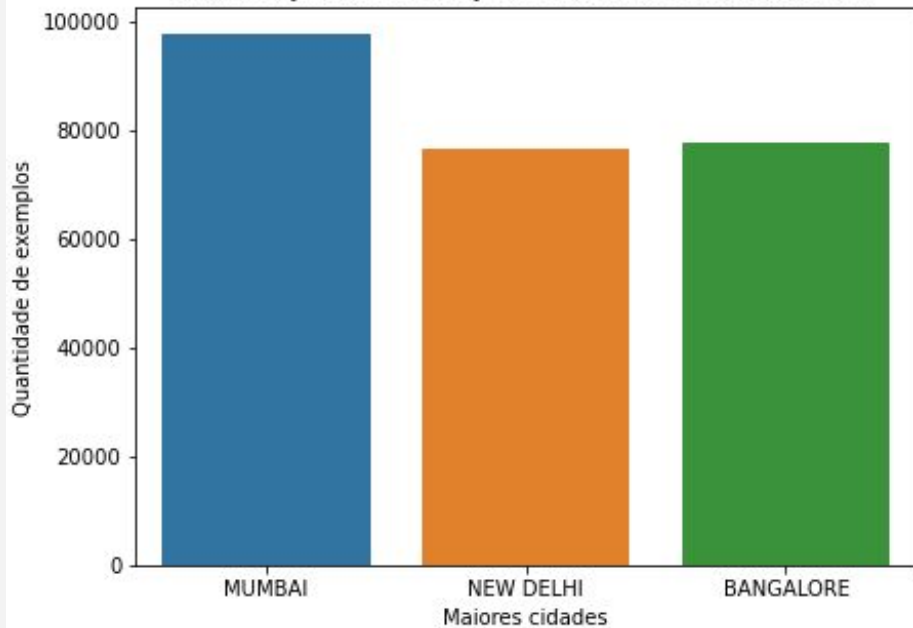
Pessoas que nasceram antes de 1900 foram descartadas

A low-angle, black and white photograph of several tall skyscrapers reaching towards a cloudy sky. The perspective creates a sense of height and scale. A white rectangular border is centered over the image, framing the text.

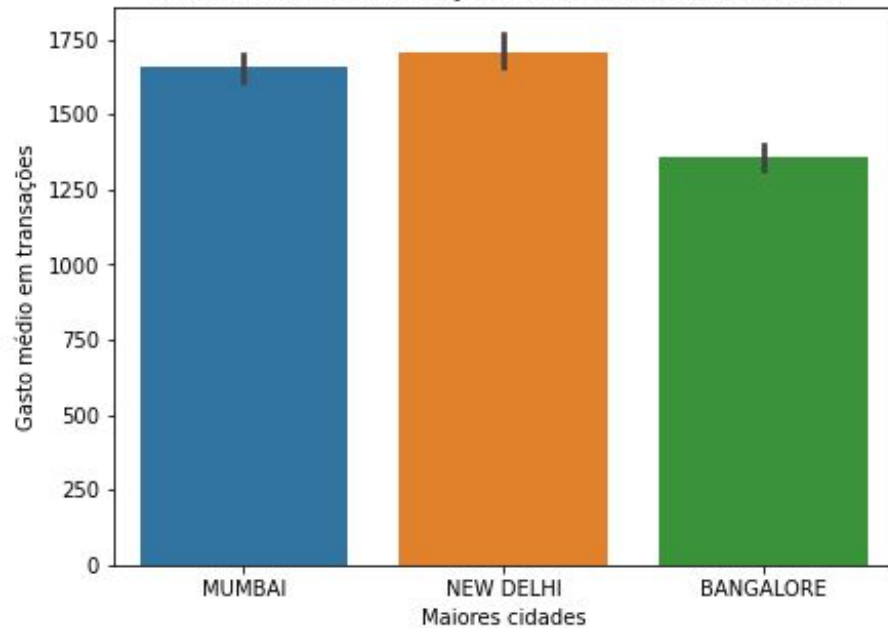
Visualização Gráfica

Observando as cidades

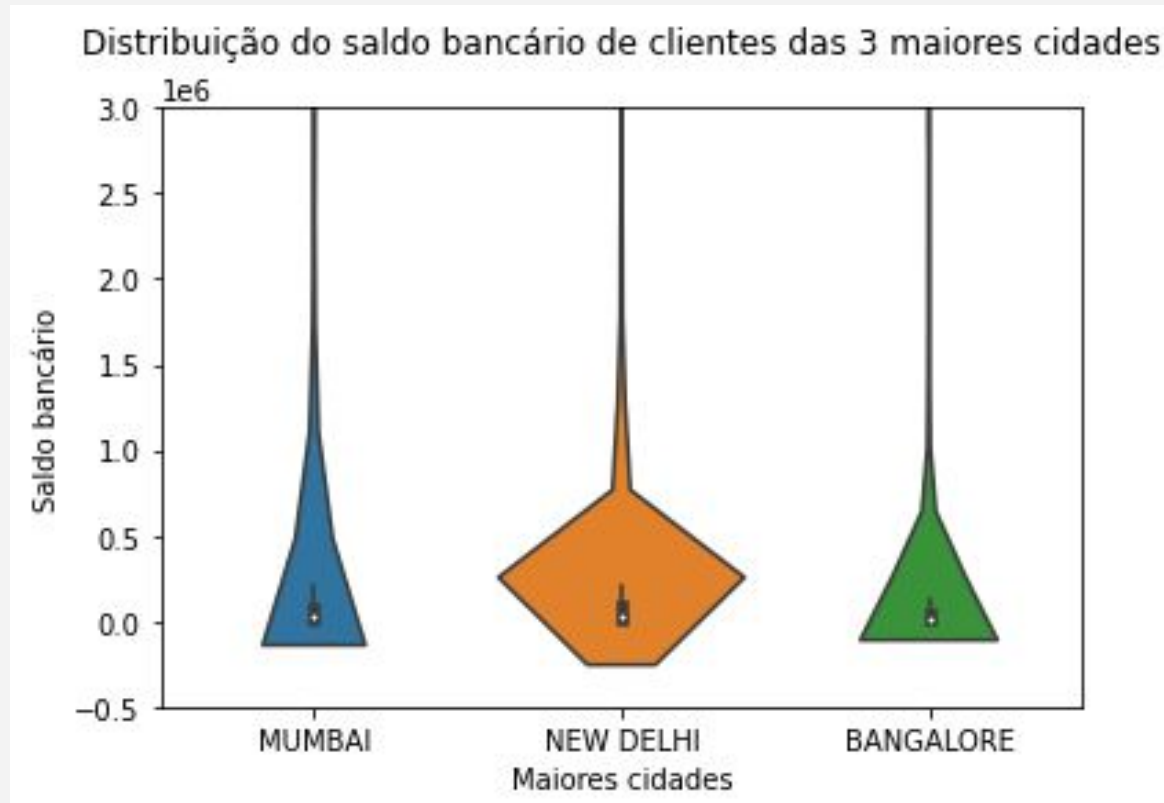
Distribuição das transações entre as 3 maiores cidades



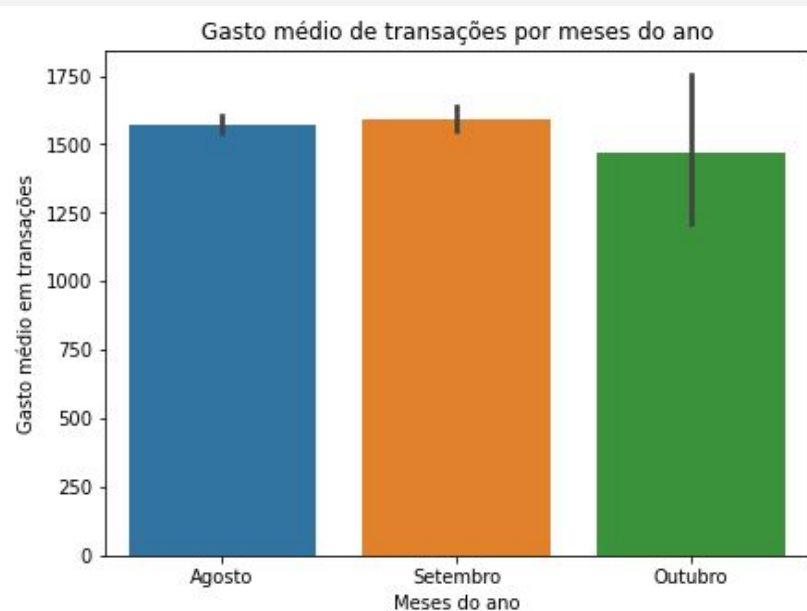
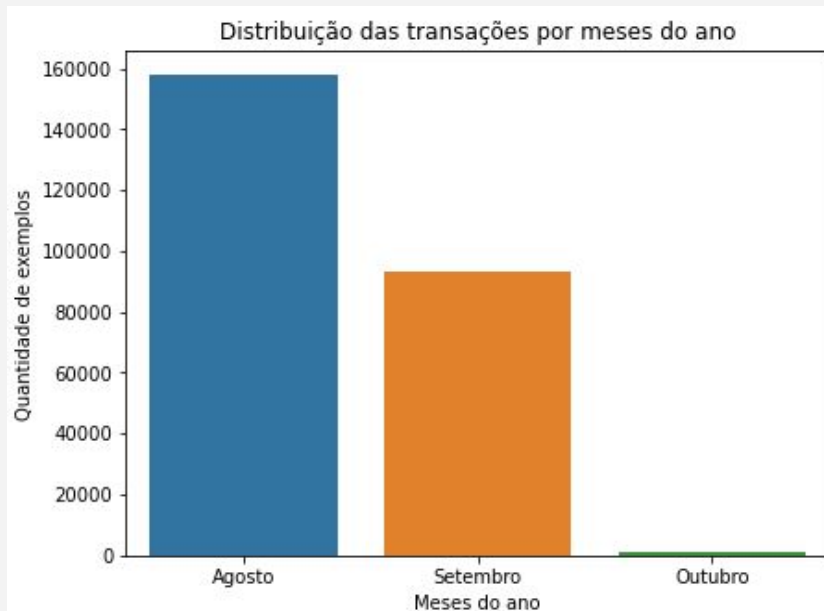
Gasto médio em transações entre as 3 maiores cidades



Observando as cidades

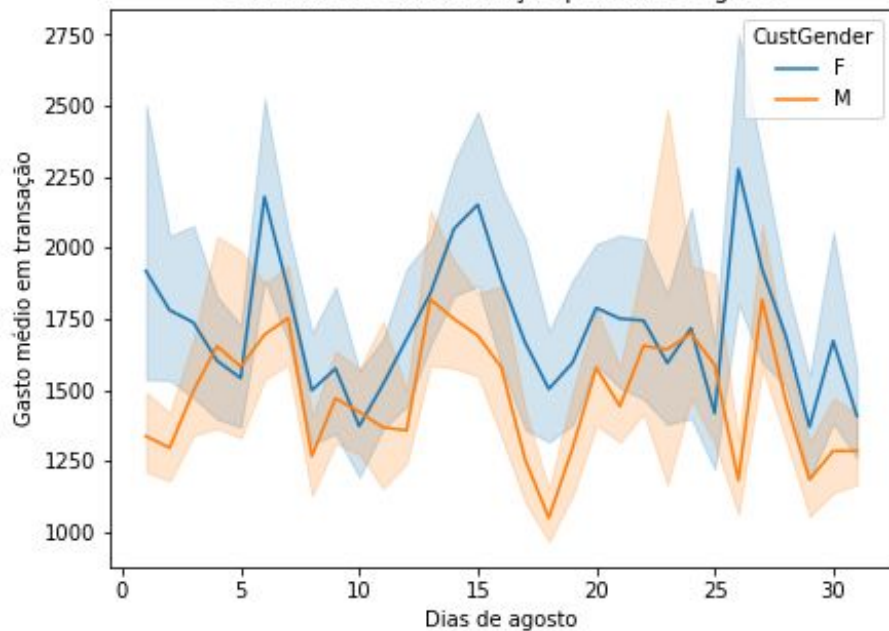


Observando o comportamento no ano

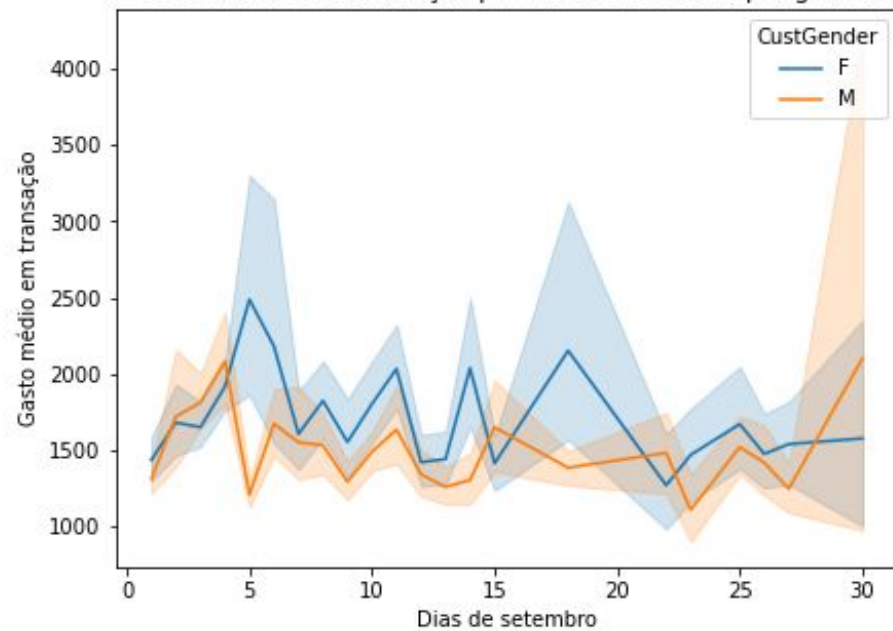


Observando o comportamento no ano

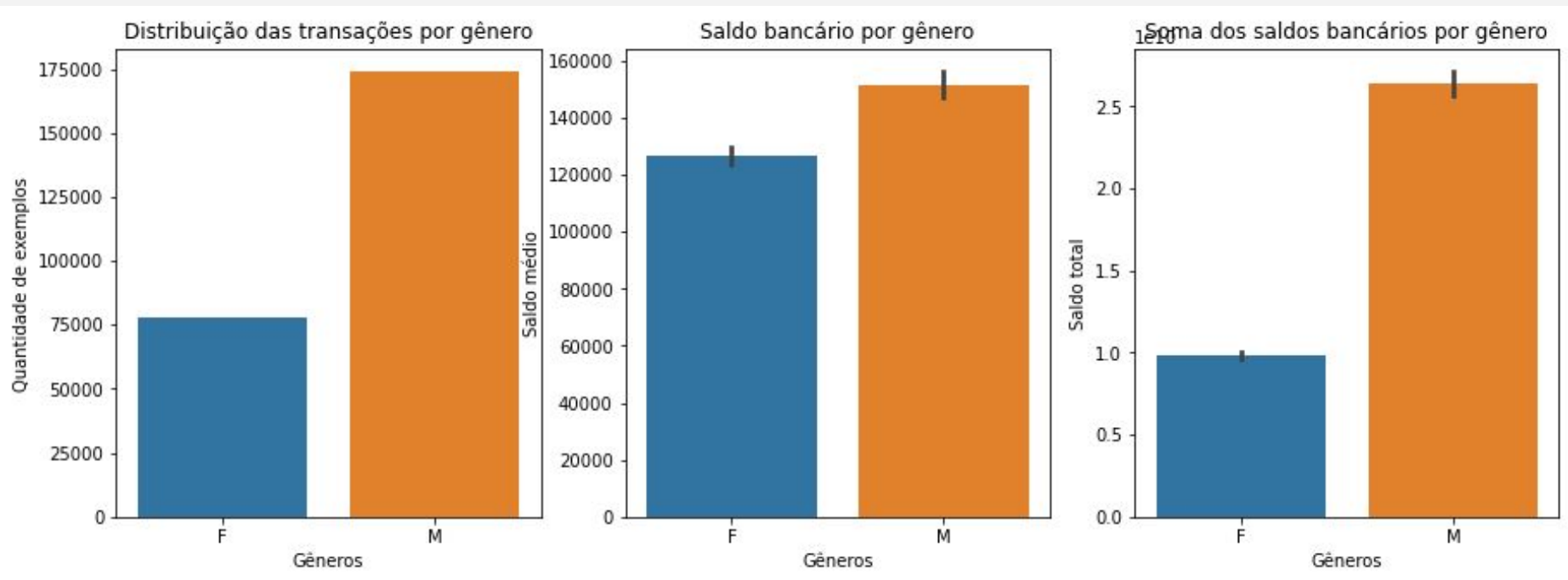
Gasto médio em transação por dia de agosto



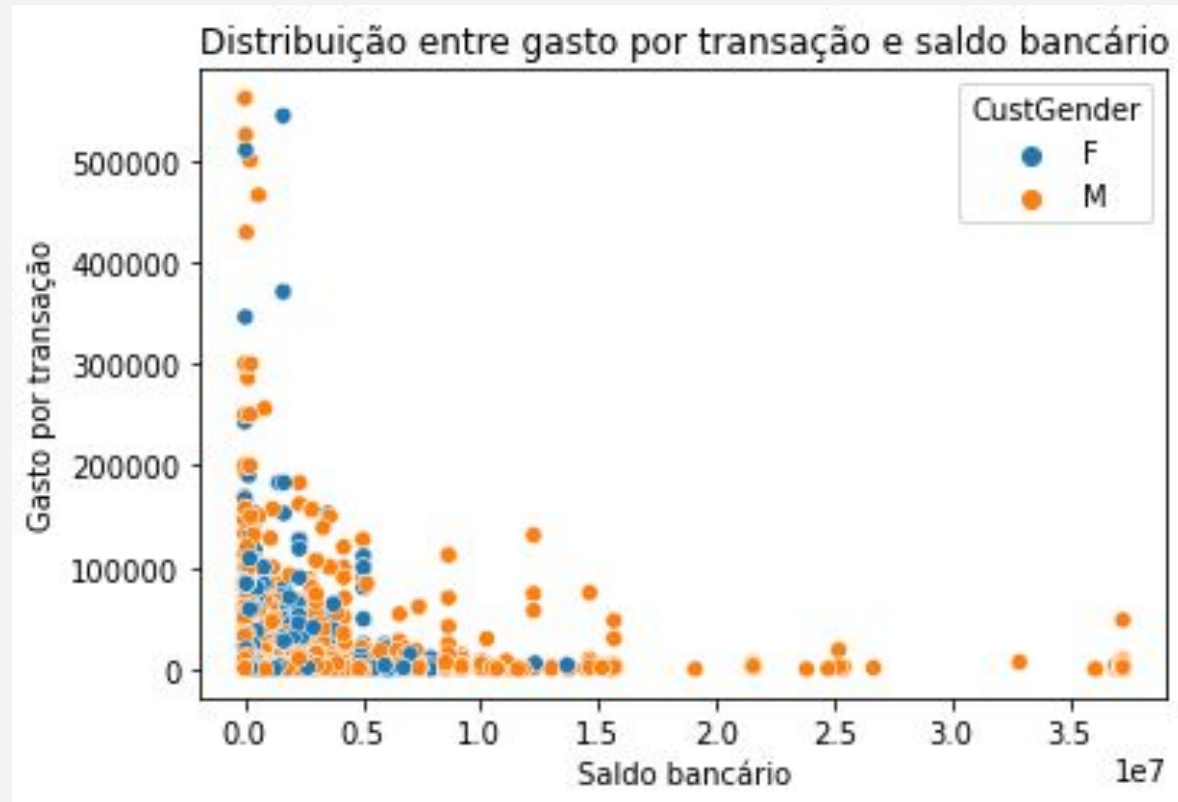
Gasto médio em transação por dia de setembro, por gênero



Observando o comportamento por gênero



Observando o comportamento por gênero



03

Agrupadores
e
Avaliações





K-Means

Quais avaliadores foram usados?

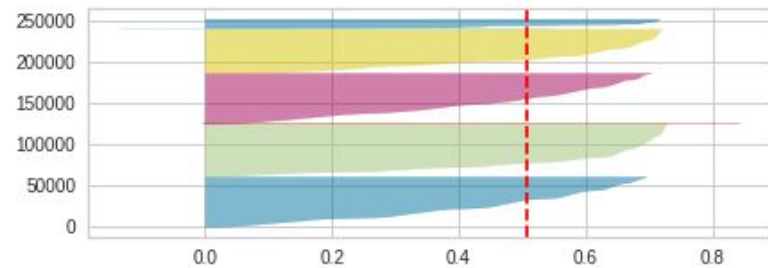
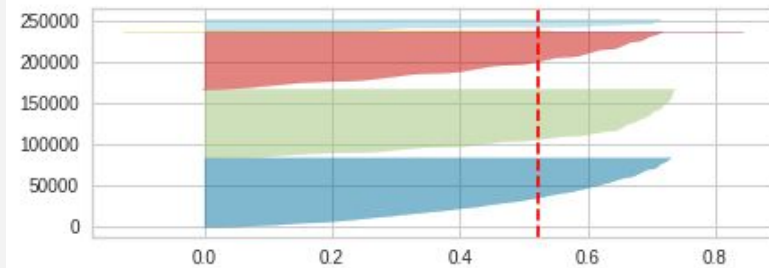
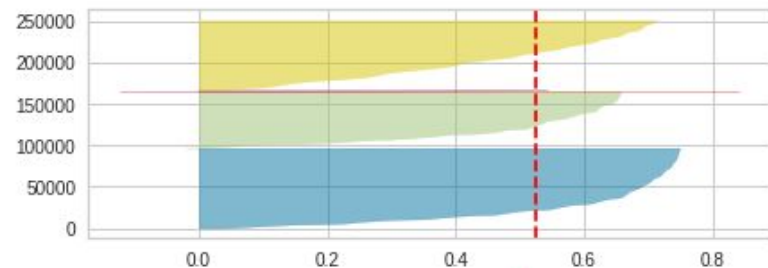
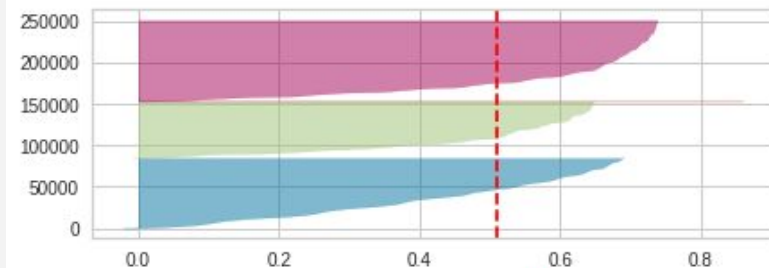
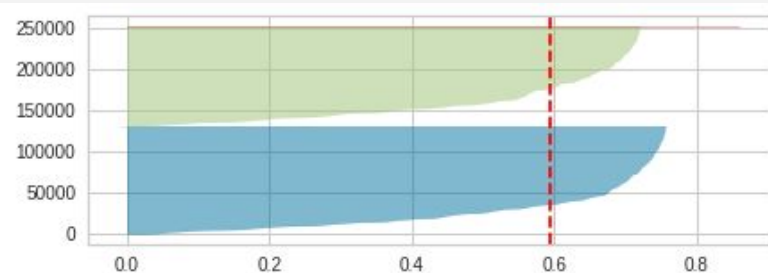
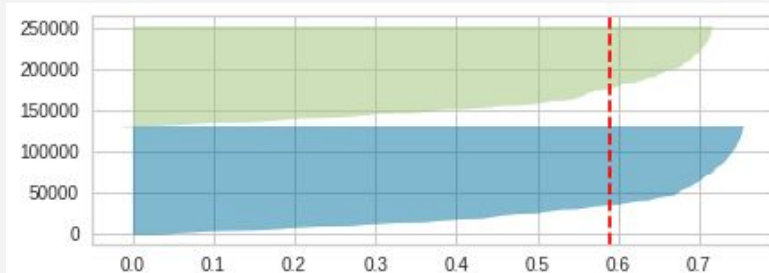
Usamos para o K-Means dois métodos de avaliação, sendo eles:

- **Método do cotovelo (ou joelho);**
- **Silhouette Score**

Método do cotovelo no K-Means



Silhouette Score no K-Means





HDBSCAN

Como o HDBSCAN tem um próprio avaliador, não fizemos uma comparação de avaliação.

Os resultados dos clusters do HDBSCAN com os parâmetros de $\alpha = 1$ e o tamanho mínimo dos clusters de 600, retornou:

Índice do HDBSCAN	Quantidade de dados
-1	25.341
0	77.692
1	4.956
2	1.016
3	142.805



**Interpretação
dos resulta-
dos obtidos**

04

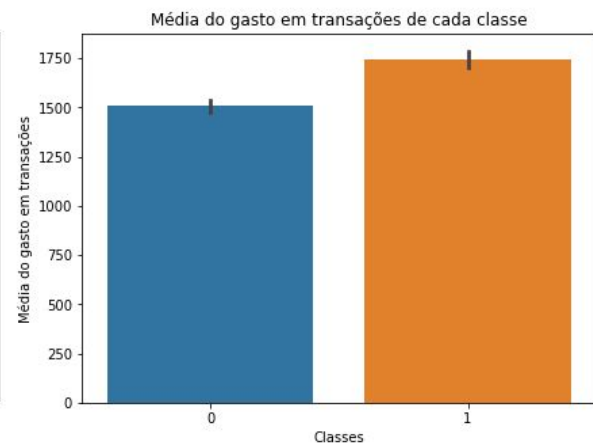
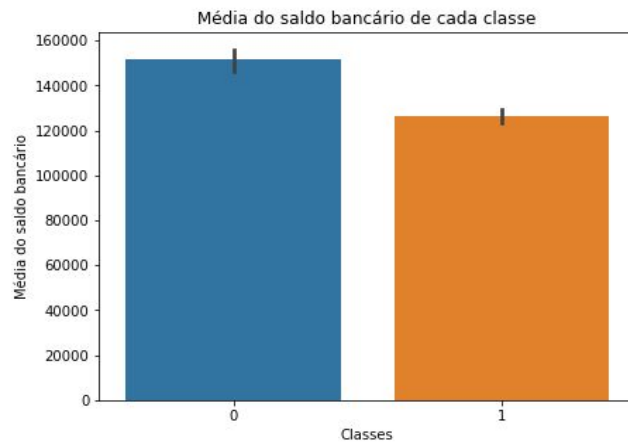
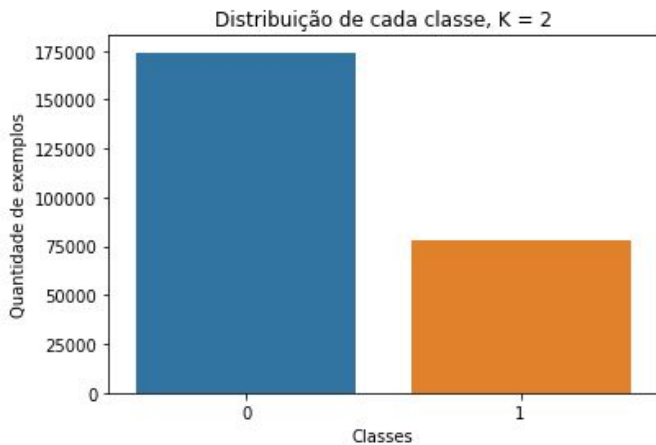


K-Means

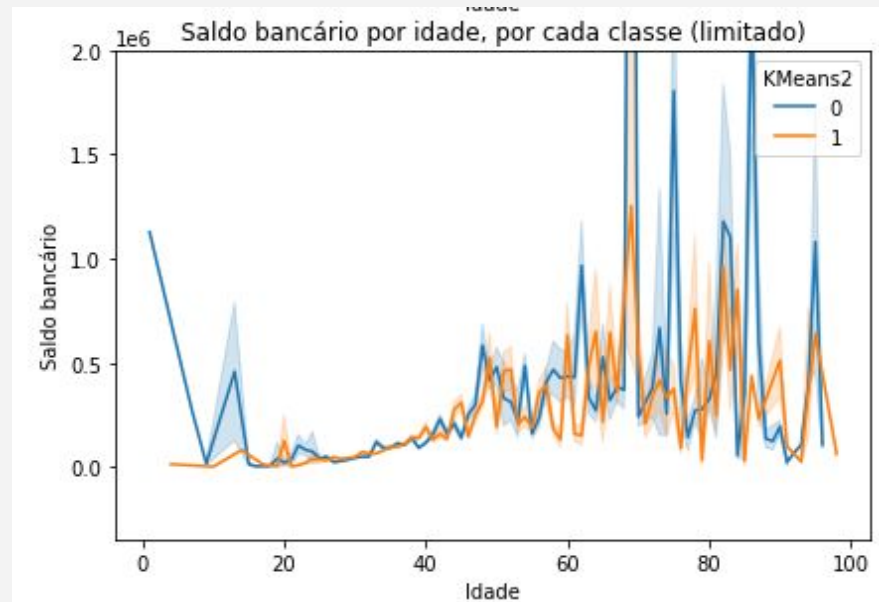
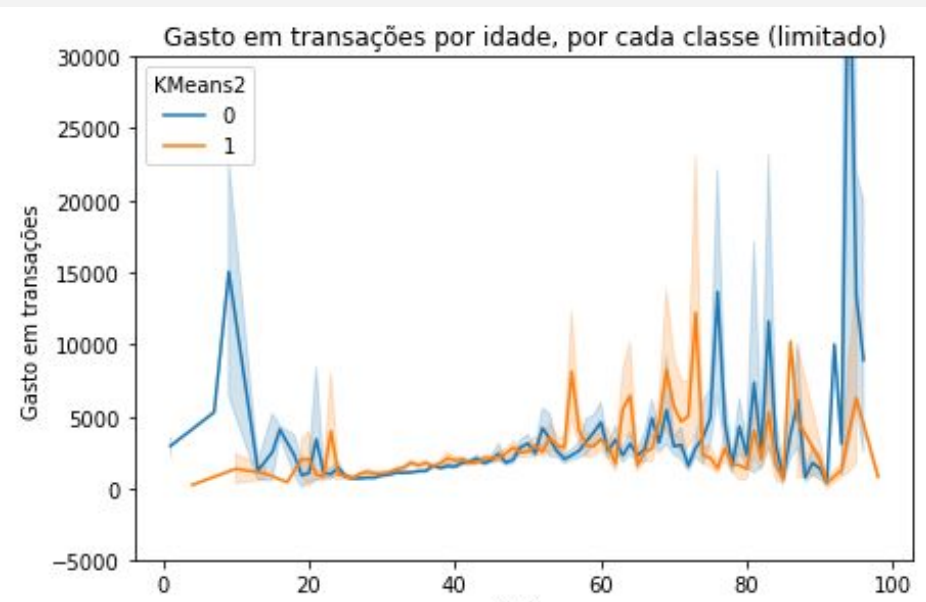


$K = 2$

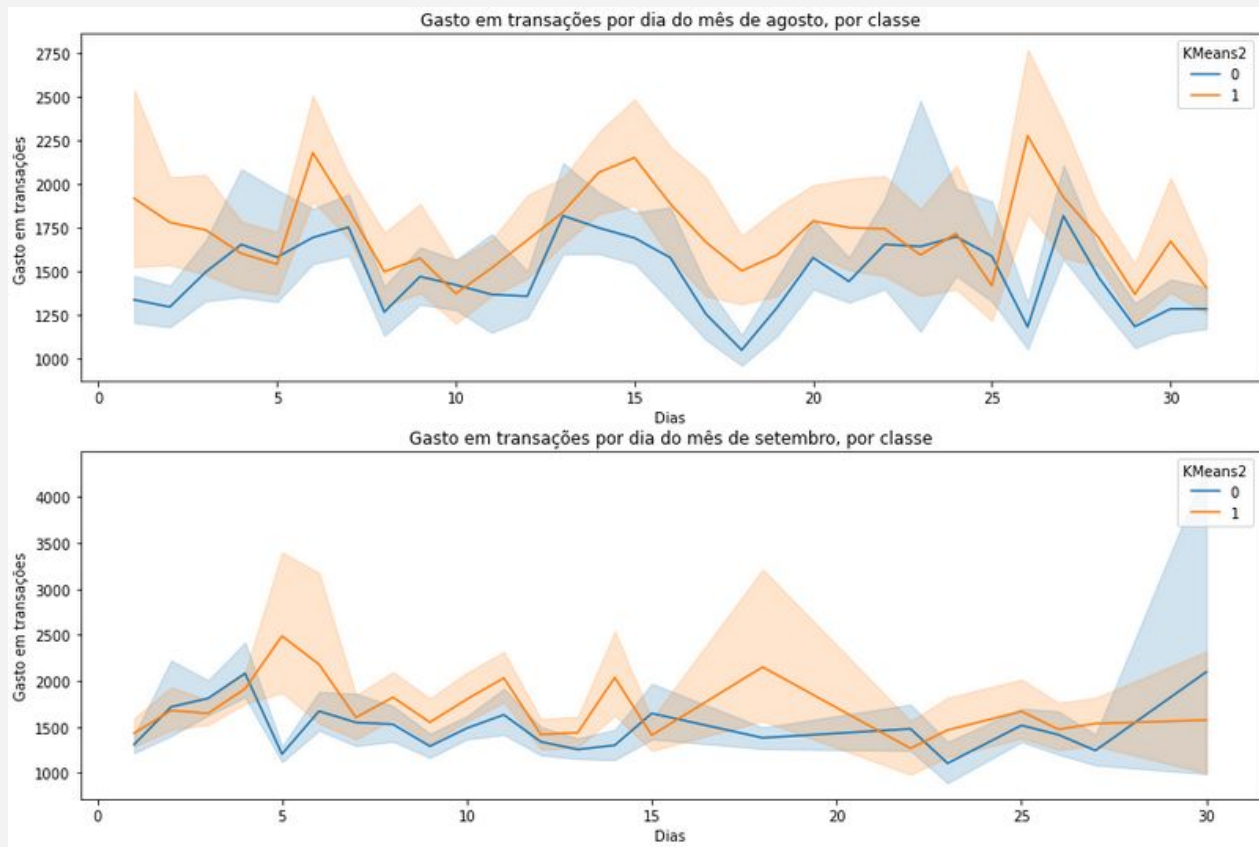
K = 2



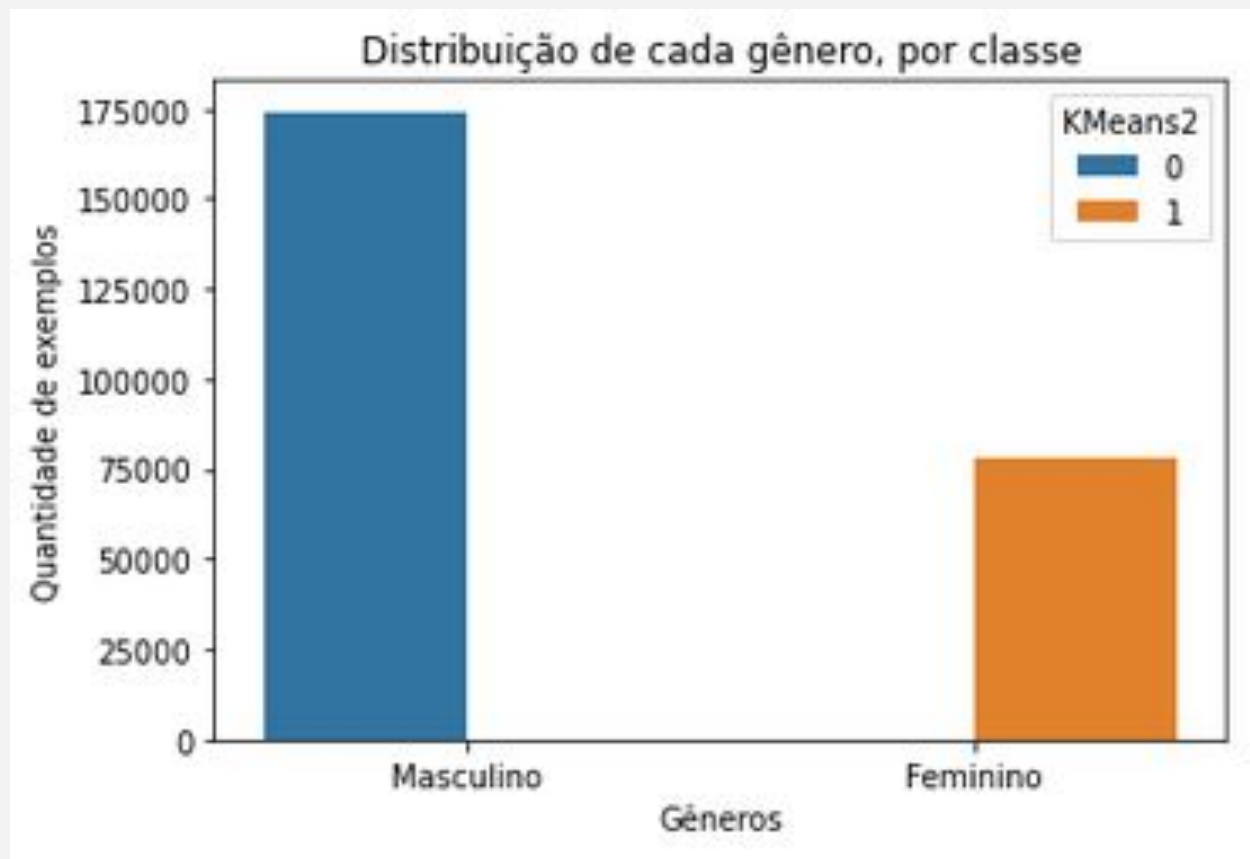
$K = 2$



$K = 2$



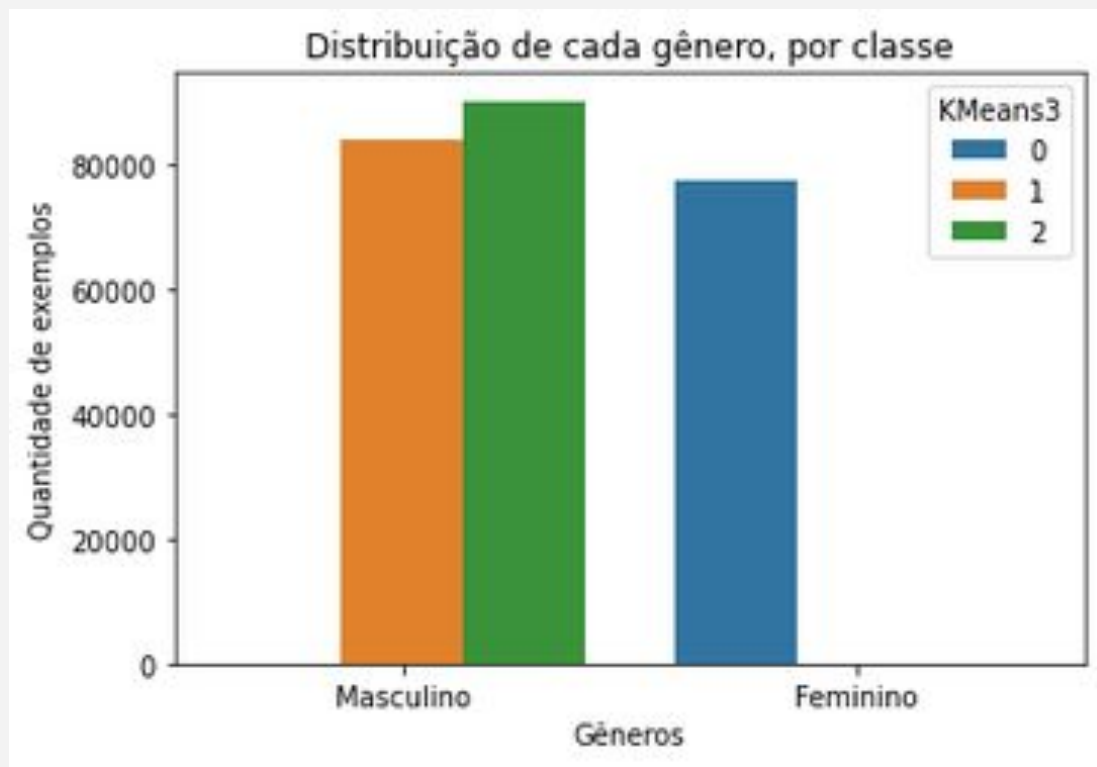
$K = 2$





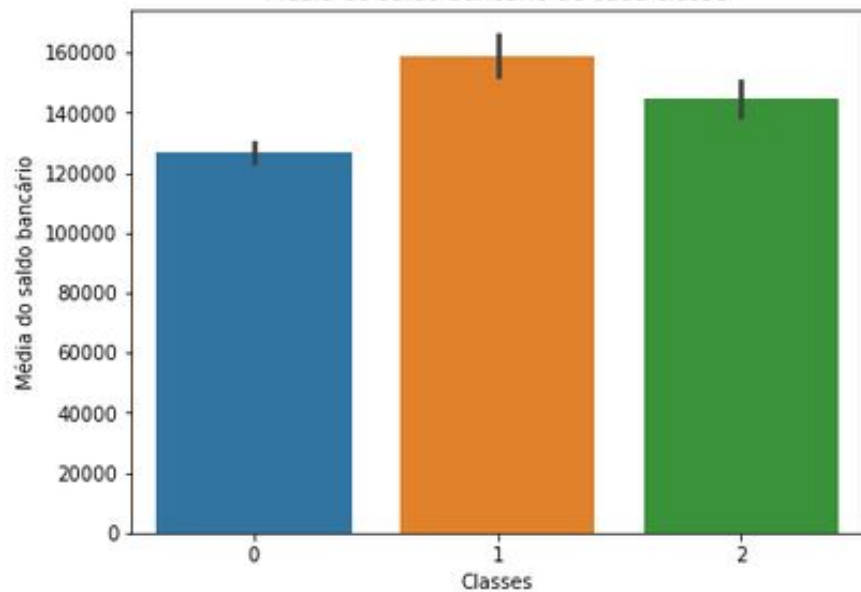
$K = 3$

$K = 3$

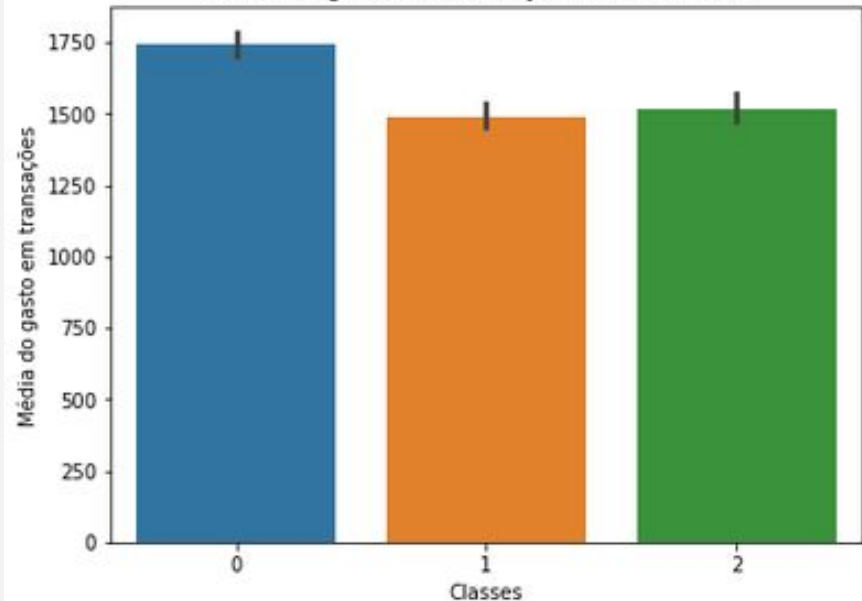


$K = 3$

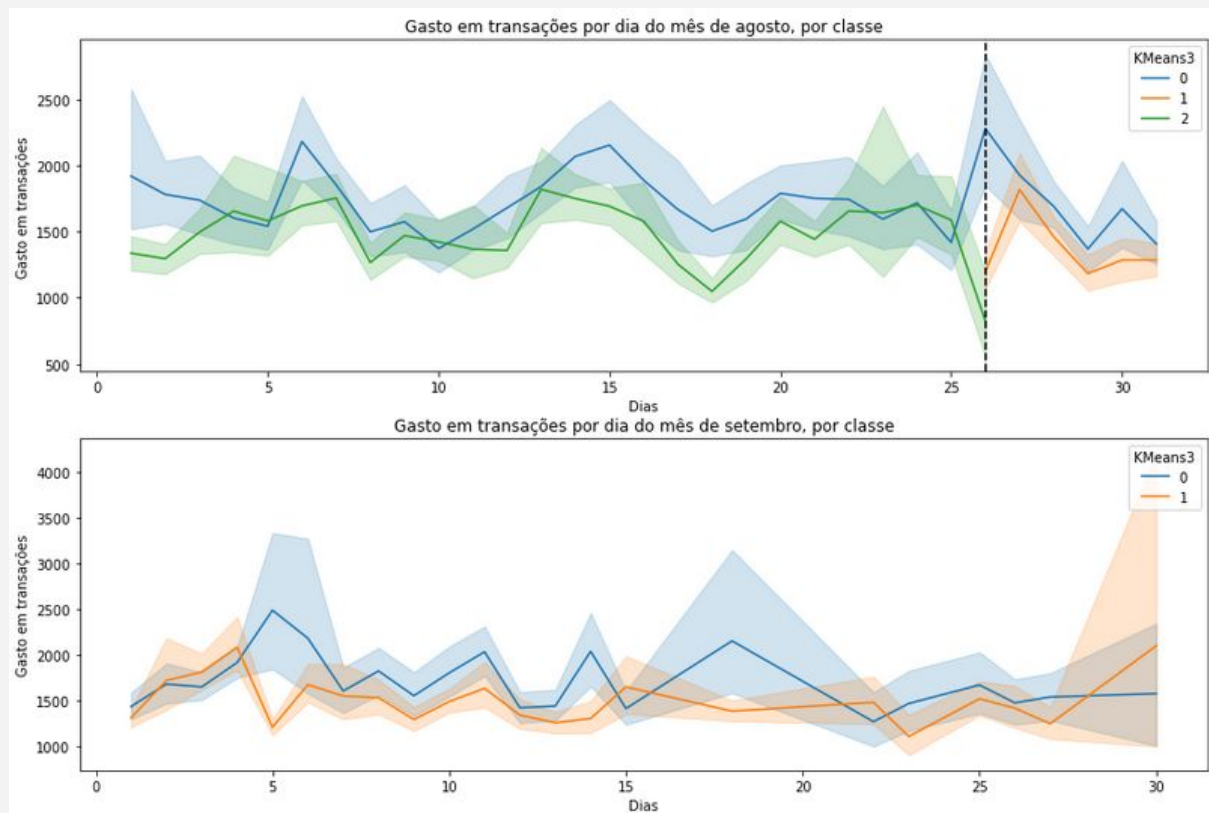
Média do saldo bancário de cada classe



Média do gasto em transações de cada classe



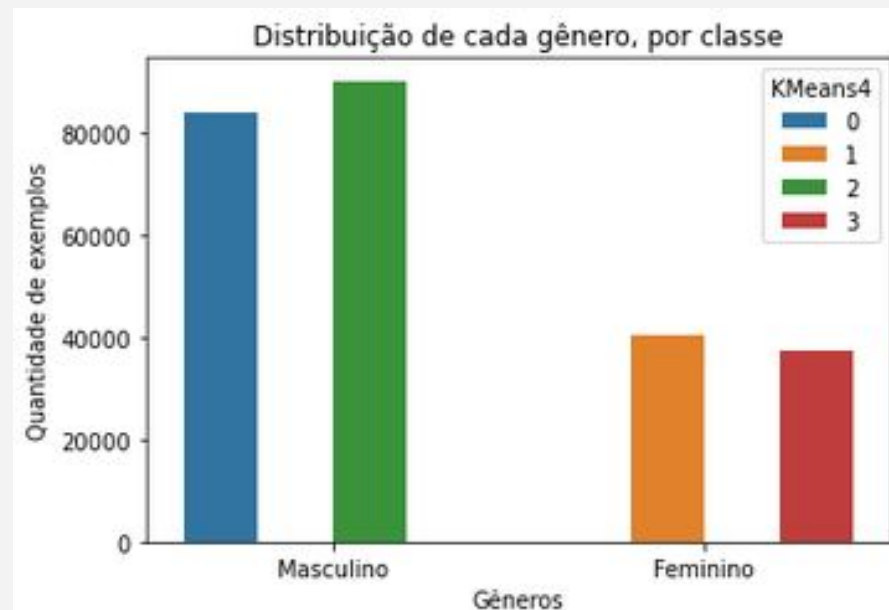
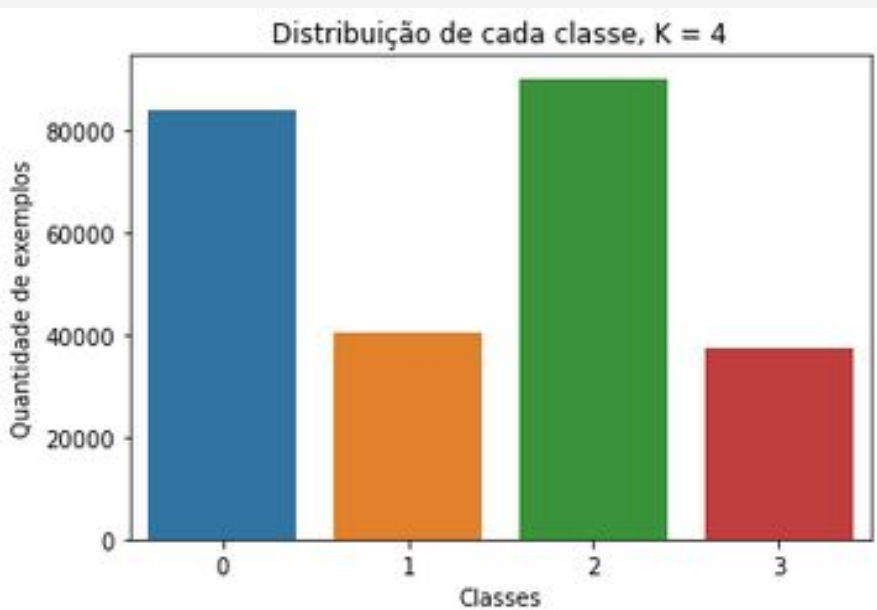
$K = 3$



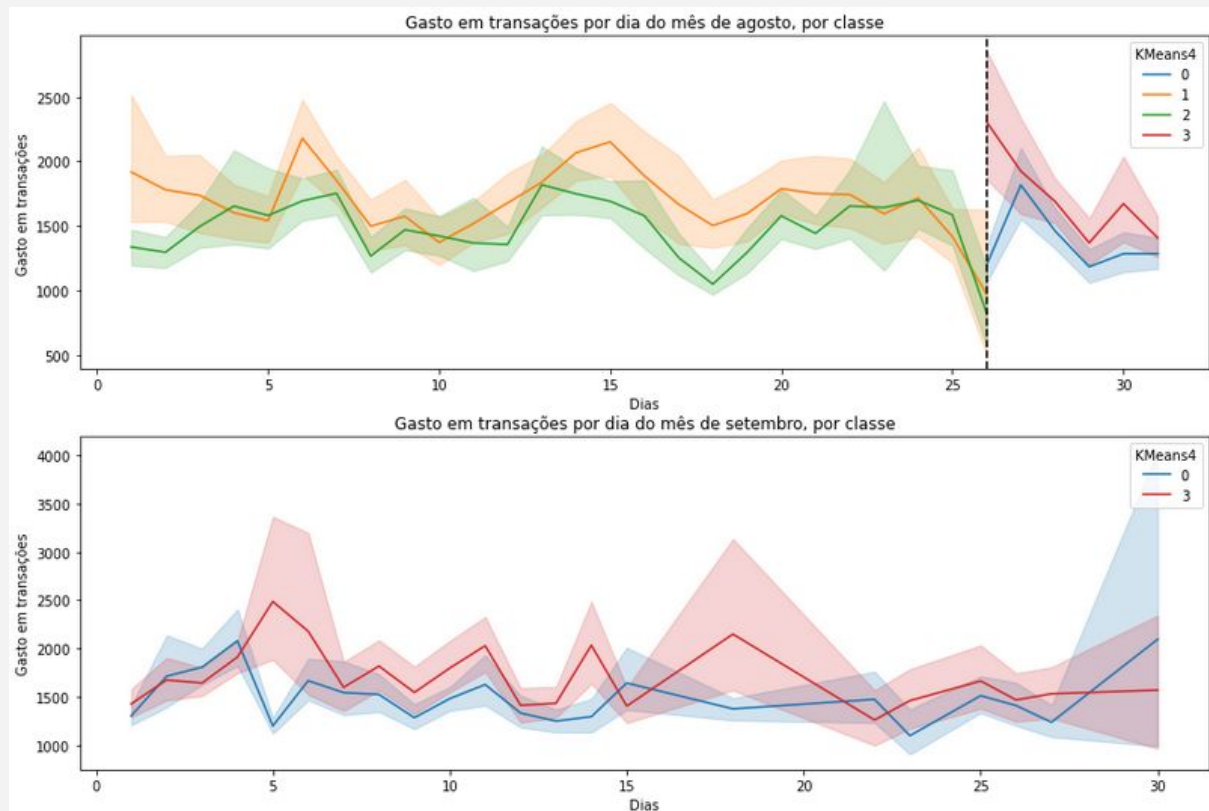


$K = 4$

K = 4



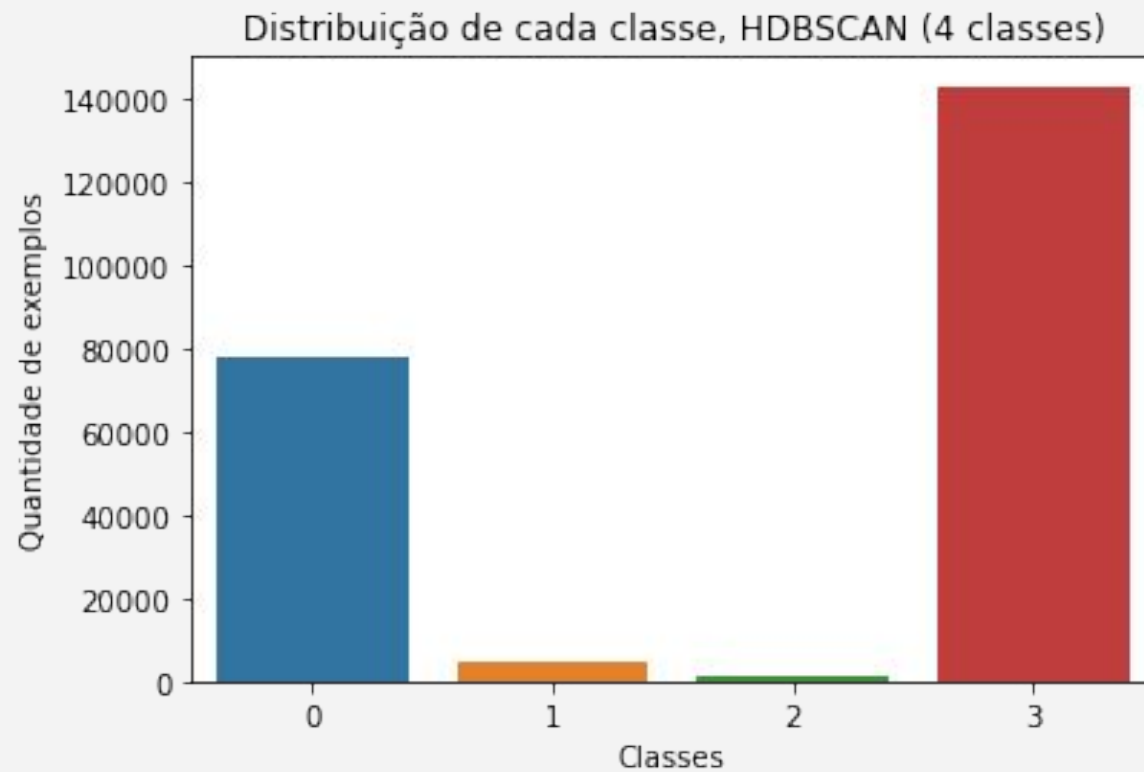
K = 4





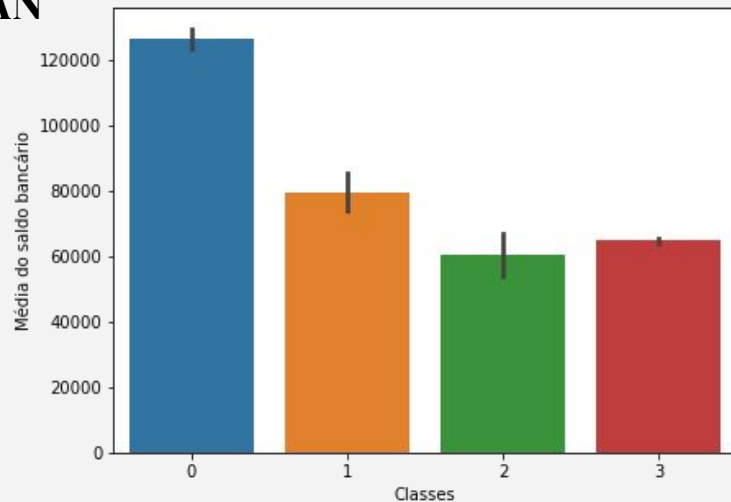
HDBSCAN

HDBSCAN

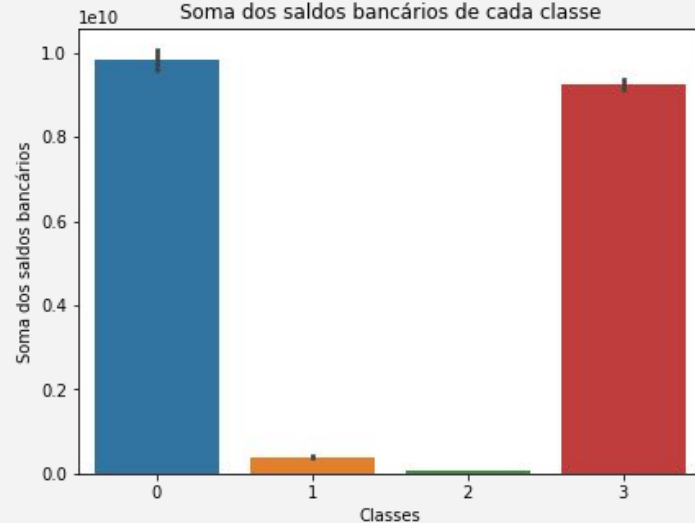


HDBSCAN

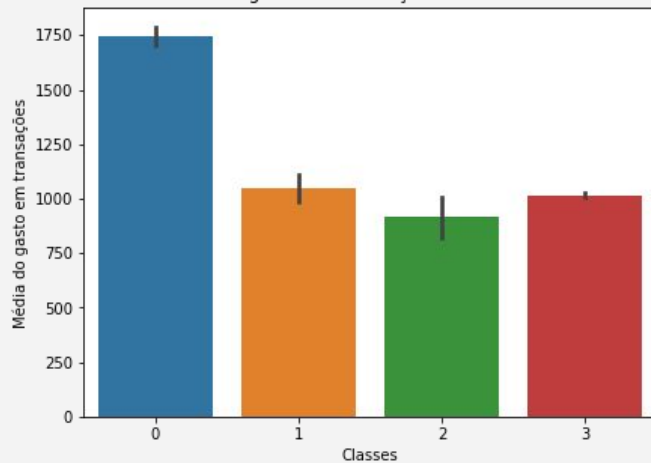
Média do saldo bancário de cada classe



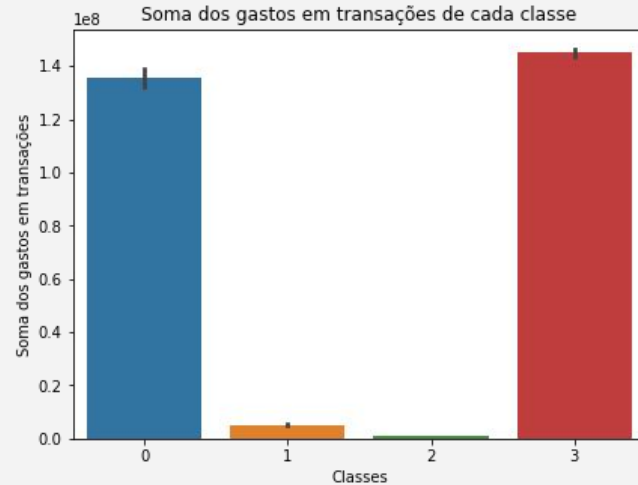
Soma dos saldos bancários de cada classe



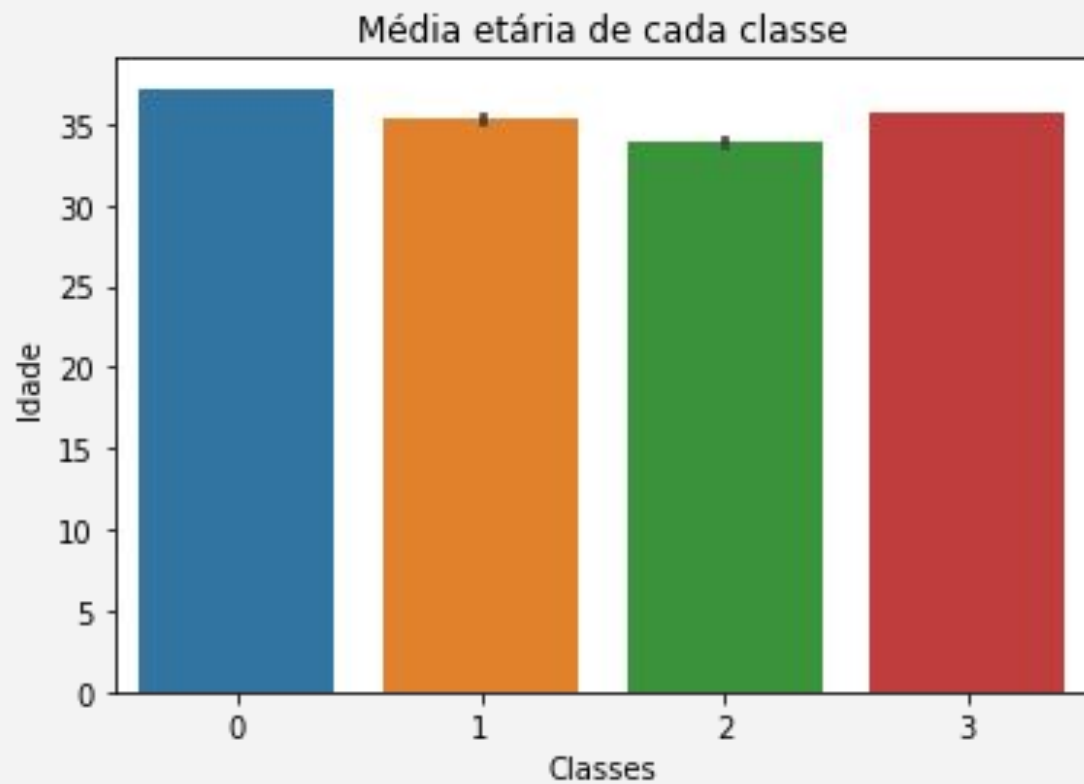
Média do gasto em transações de cada classe

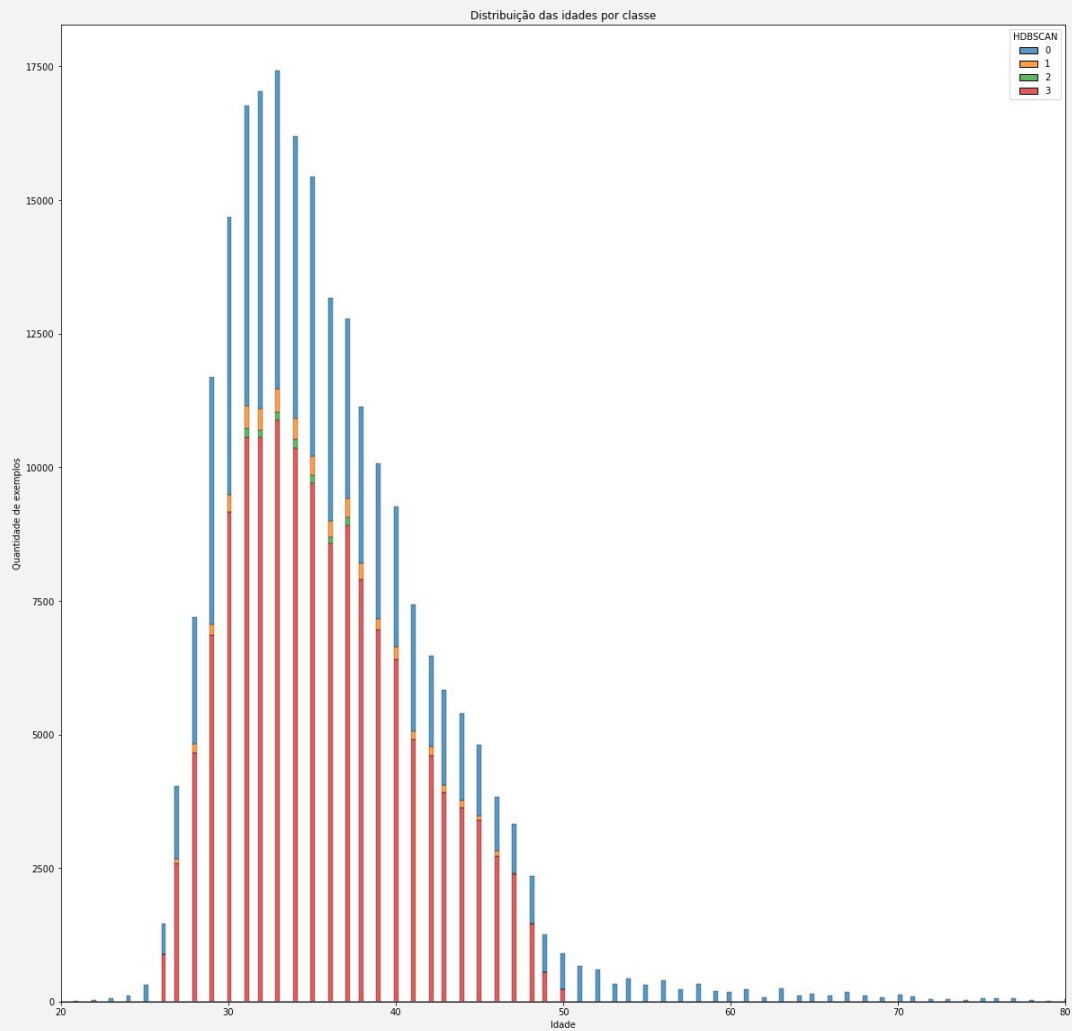


Soma dos gastos em transações de cada classe



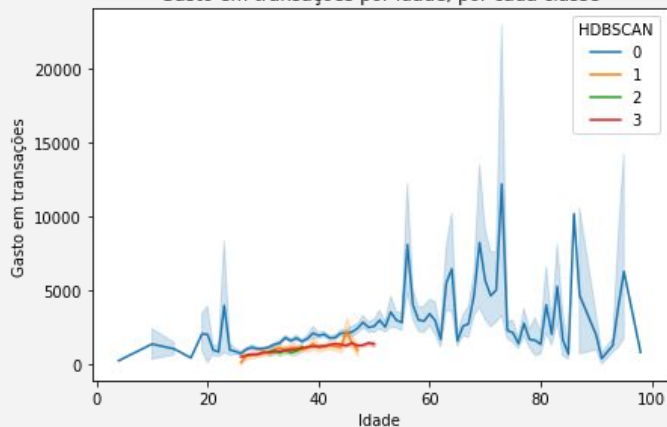
HDBSCAN



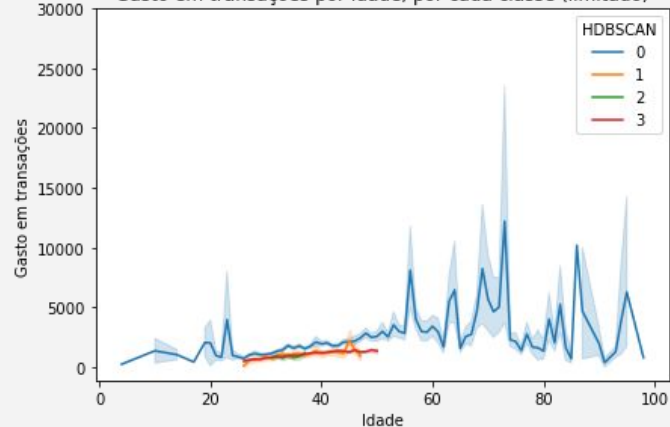


HDBSCAN

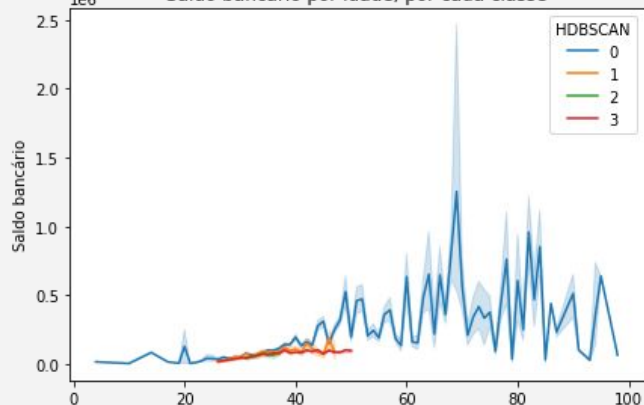
Gasto em transações por idade, por cada classe



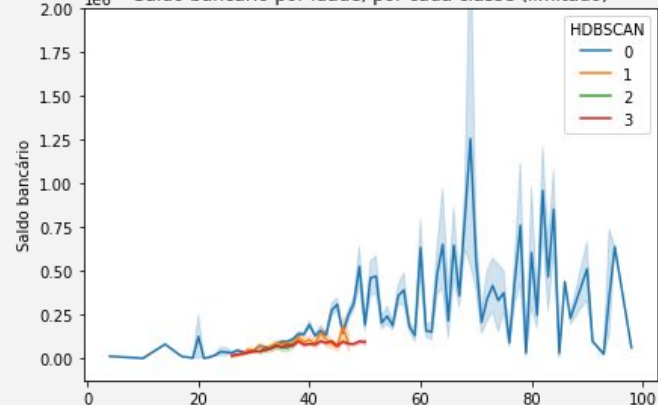
Gasto em transações por idade, por cada classe (limitado)



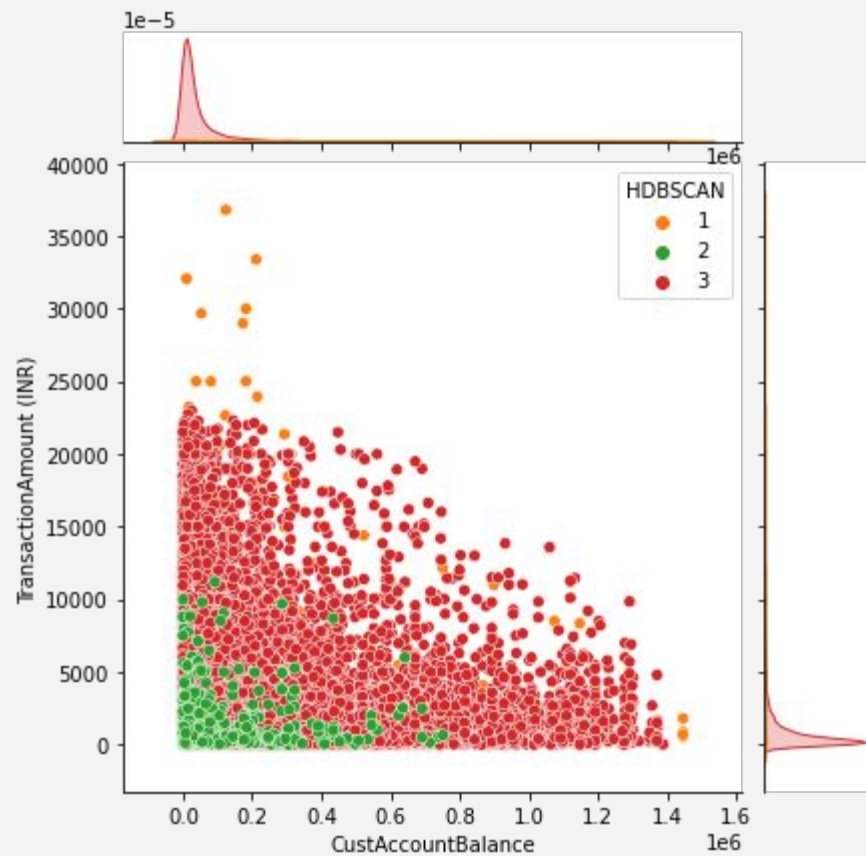
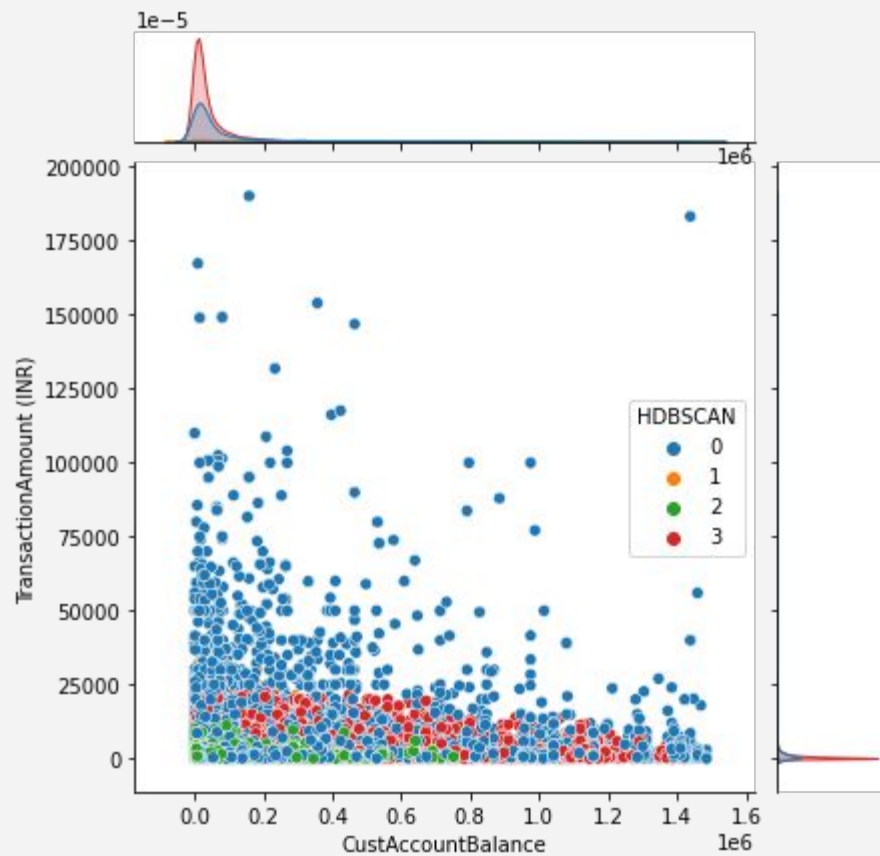
Saldo bancário por idade, por cada classe



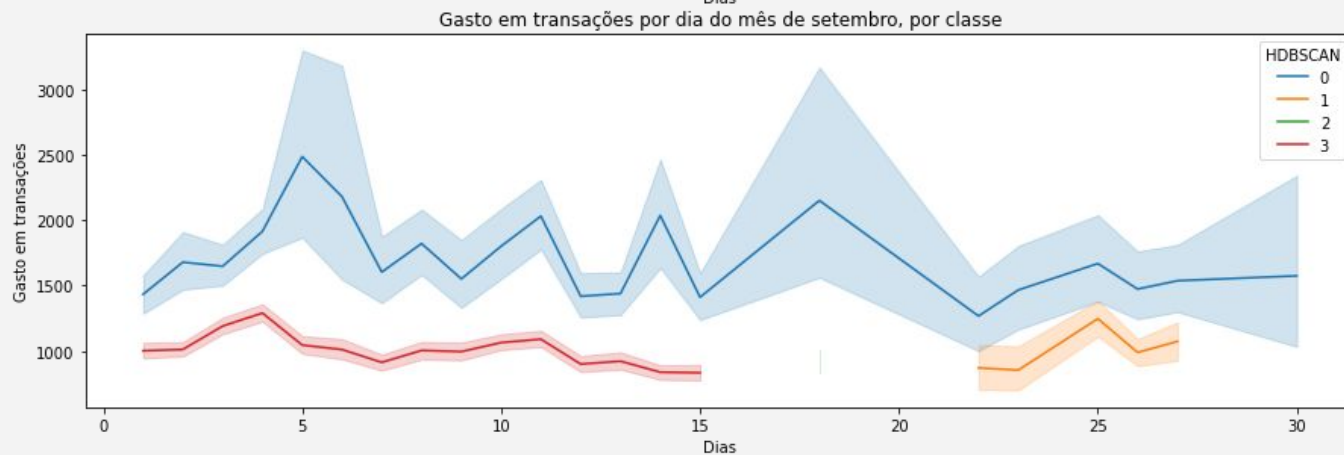
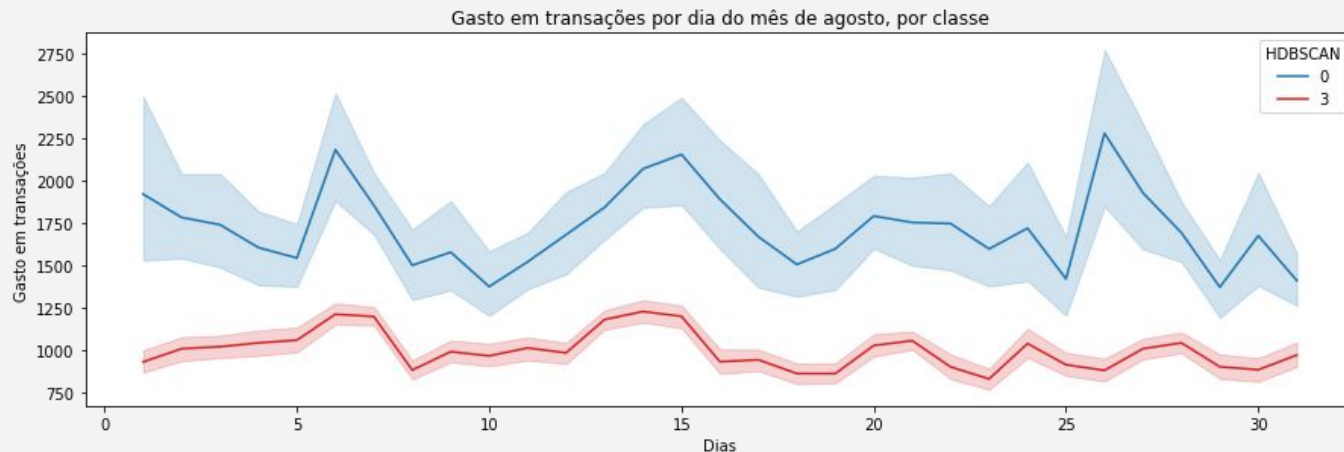
Saldo bancário por idade, por cada classe (limitado)



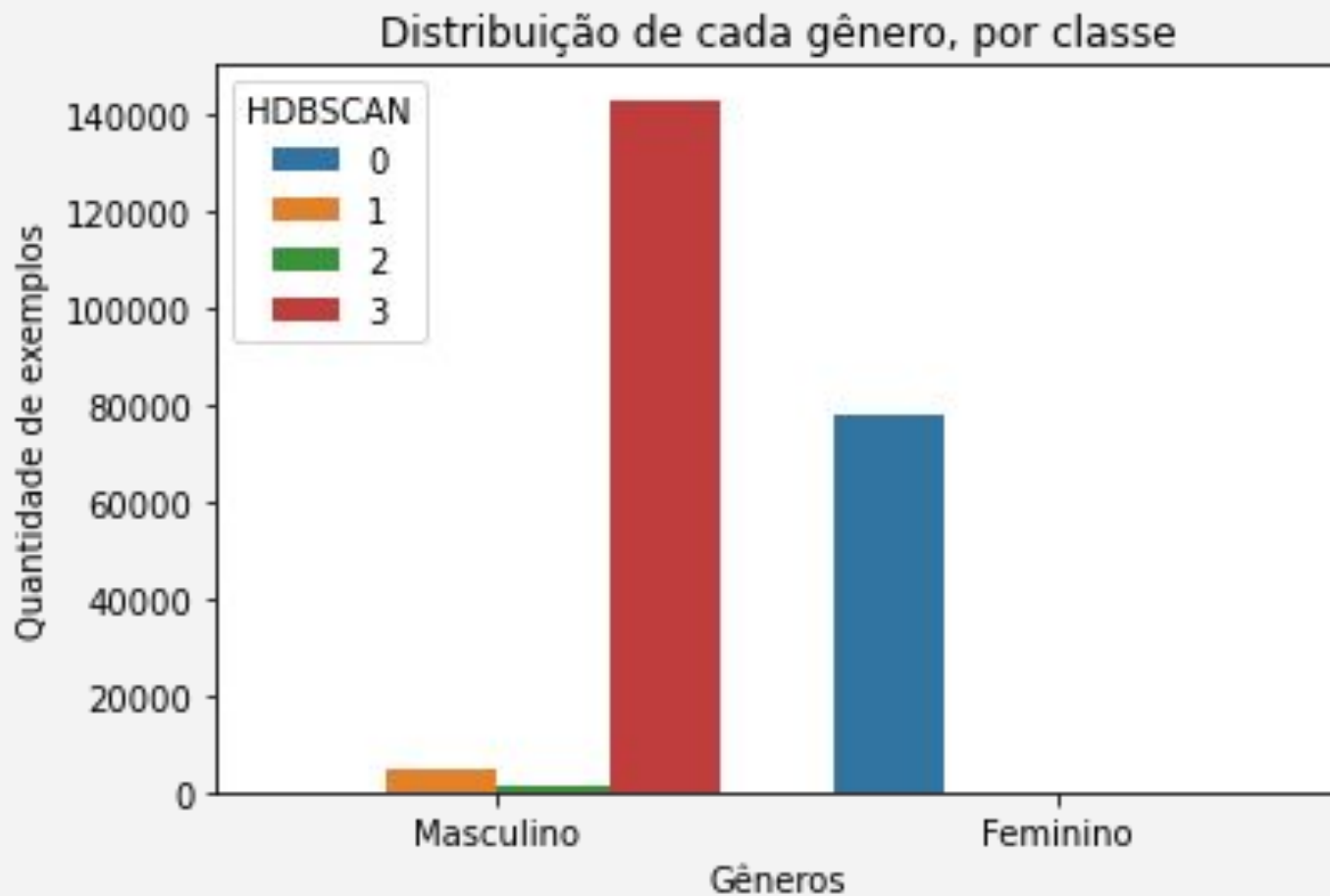
HDBSCAN



HDBSCAN



HDBSCAN





Obrigado pela atenção

Alunos:

Jayme Sakae dos Reis Furuyama - 761044

Vitor Lopes Fabris - 769822