

# Projeto Final

## Segmentação de Clientes



08/07/2020



# Pós-Graduação "Lato Sensu" Especialização em Análise de Big Data



## **Nome dos Alunos:**

Gustavo de Carvalho Ferreira e Vieira  
Rafael Polli Silva

## **Coordenadores:**

Profª Drª Alessandra de Ávila Montini  
Profª Dr. Adolpho Walter Pimazoni Canton



# Agenda

- 1. Objetivo do Trabalho
- 2. Contextualização do Problema
- 3. Base de Dados
  - i. Bases originais
  - ii. Transformações das bases
- 4. Análise Exploratória de Dados
- 5. Modelagem com Estatística Tradicional
- 6. Modelagem com Inteligência Artificial
- 7. Conclusões

# Metodologia de análise de dados



## Definição do problema

- Objetivos
- Conceitos
- Critérios
- Histórico de dados
- Variáveis

## Análise preliminar

- Medidas de posição
- Análise de frequências
- Gráficos
- Análise de *outliers*
- Análise de *missings*
- Validação sobre a consistência das informações

## Avaliação das técnicas

- K-means nativo do Spark

## Avaliação das técnicas

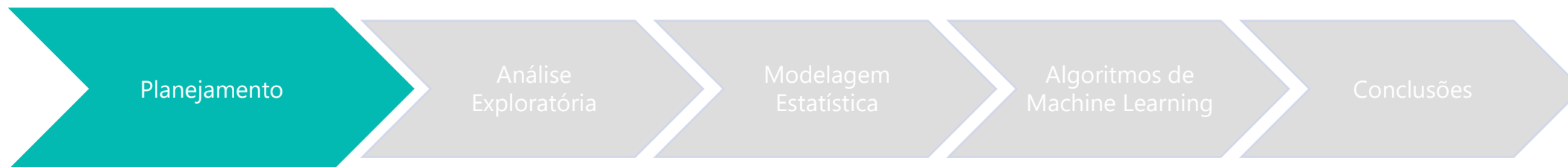
- biSecting K-means nativo do Spark
- Gaussian Mixture Model nativo do Spark
- Aglomeração do Scikit Learn
- DBSCAN do Scikit Learn
- MeanShift do Scikit Learn
- K-means do Scikit Learn

## Definição da técnica

- Validação dos resultados
- Escolha da técnica que melhor se adequa ao uso e estratégias



# Metodologia de análise de dados



## Definição do problema

- Objetivos
- Conceitos
- Critérios
- Histórico de dados
- Variáveis

## Análise preliminar

- Medidas de posição
- Análise de frequências
- Gráficos
- Análise de outliers
- Análise de missings
- Validação sobre a consistência das informações

## Avaliação das técnicas

- K-means nativo do Spark

## Avaliação das técnicas

- biSecting K-means nativo do Spark
- Gaussian Mixture Model nativo do Spark
- Aglomeração do Scikit Learn
- DBSCAN do Scikit Learn
- MeanShift do Scikit Learn
- K-means do Scikit Learn

## Definição da técnica

- Validação dos resultados
- Escolha da técnica que melhor se adequa ao uso e estratégias



# 1. Objetivo do Trabalho

6

## Segmentação de Clientes da Olist para definição de política comercial e de ações de marketing

O objetivo do trabalho é encontrar **a melhor forma de agrupamento dos clientes** da Olist (vendedores de marketplaces) para **estabelecer política comercial e ações de marketing diferenciadas** para cada grupo, visando **aumentar o faturamento e melhorar o resultado** da empresa.





## 2. Contextualização do Problema

7

A Olist é uma empresa que **auxilia vendedores a anunciar seus produtos nos principais marketplaces da internet** (Amazon, Mercado Livre, Americanas, Carrefour, Submarino, Via Varejo, Casas Bahia, B2W Digital, Extra, Shoptime, Ponto Frio, Madeira Madeira e Zoom), fornecendo de forma centralizada ferramentas para cadastro e gestão de produtos e estoques; gestão de vendas, finanças e estratégia; gestão da logística de entregas; e melhoria das posições nos sites de venda com a alta reputação da Olist.

Atualmente, a empresa possui 3 planos comerciais:

- Olist lite: faturamento de até R\$3.000,00, sem mensalidade, comissão de 21% por pedido.
- Olist pro: faturamento de R\$3.000,00 a R\$20.000,00, mensalidade de R\$79,90, comissão de 19% por pedido.
- Olist Premium: faturamento acima de R\$20.000,00, condições específicas para cada cliente.



## 2. Contextualização do Problema

8

A empresa gostaria de saber se essa divisão de clientes por faixas de faturamento é a mais adequada para estabelecer sua política comercial, além de obter insights para ações de marketing personalizadas de acordo com os perfis dos clientes, para aumentar seu faturamento e melhorar seu resultado.





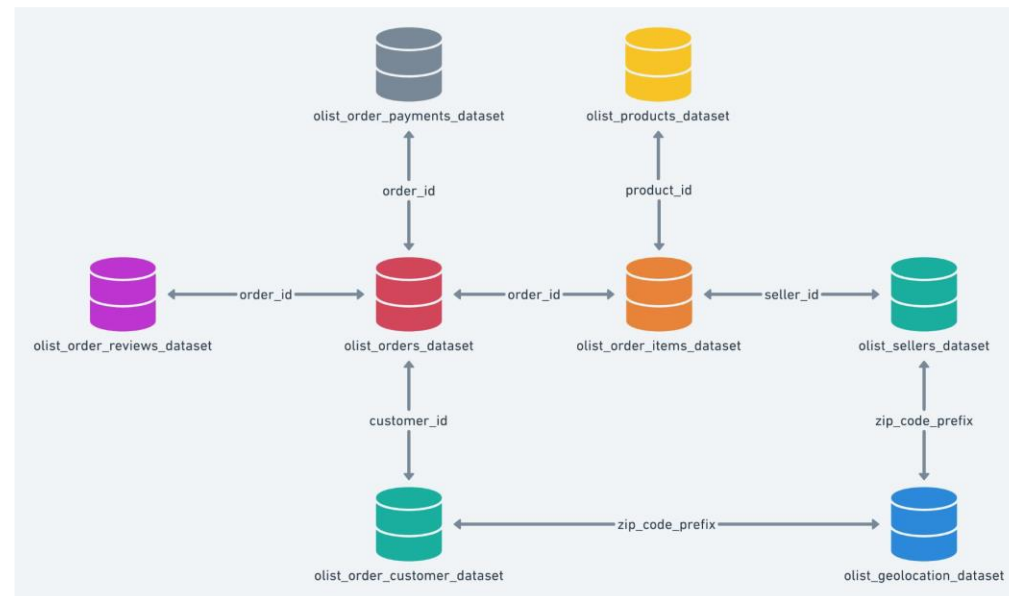
### 3. Bases de Dados

9

A Olist possui um conjunto de dados público disponível no kaggle . O conjunto de dados Brazilian E Commerce Public Dataset by Olist consiste de transações de clientes nos mais diversos marketplaces brasileiros.

A base principal é composta de uma série histórica de vendas com 99.441 pedidos emitidos entre setembro de 2016 e outubro de 2018.

Além desta, há bases auxiliares com detalhes sobre os produtos vendidos e os produtos de cada pedido, sobre os vendedores, sobre os compradores e os reviews dos compradores, sobre os pagamentos, e sobre dados de geolocalização. As duas últimas não serão utilizados em nosso estudo.



## 3.i. Bases Originais

Descrição da base olist orders dataset:

### Quantidade de registros

- 99.441

### Período analisado:

- 04/09/2016 a 16/10/2018

### Quantidade de variáveis

- 8

### Nomes das variáveis

- 1. order\_id: Id único da compra
- 2. customer\_id: Chave para o dataset de clientes, cada compra tem um customer\_id único
- 3. order\_status: Status da compra
- 4. order\_purchase\_timestamp: Momento da compra
- 5. order\_approved\_at: Momento de aprovação da compra
- 6. order\_delivered\_carrier\_date: Momento de postagem da compra
- 7. order\_delivered\_customer\_date: Momento de entrega da compra
- 8. order\_estimated\_delivery\_date: Tempo estimado de entrega informado ao comprador no momento da compra



## 3.i. Bases Originais

11

Descrição da base olist\_order\_items\_dataset:

### Quantidade de registros

- 112.650

### Quantidade de variáveis

- 7

### Nomes das variáveis

- 1. order\_id: Id único da compra
- 2. order\_item\_id: Número sequencial que identifica itens incluídos na mesma compra
- 3. product\_id: Id único do produto
- 4. seller\_id: Id único do vendedor
- 5. shipping\_limit\_date: Data limite de entrega do produto ao serviço de entrega
- 6. price: Preço do item na compra
- 7. freight\_value: Preço do frete correspondente ao item



## 3.i. Bases Originais

Descrição da base olist sellers dataset:

### Quantidade de registros

- 3.095

### Quantidade de variáveis

- 4

### Nomes das variáveis

- 1. seller\_id: Id único do vendedor
- 2. seller\_zip\_code\_prefix: Primeiros 5 dígitos do CEP do vendedor
- 3. seller\_city: Nome da cidade do vendedor
- 4. seller\_state: Estado do vendedor



## 3.i. Bases Originais

13

Descrição da base olist\_products dataset:

### Quantidade de registros

- 32.951

### Quantidade de variáveis

- 9

### Nomes das variáveis

- 1. product\_id: Id único do produto
- 2. product\_category\_name: Raiz da categoria do produto
- 3. product\_name\_lenght: Comprimento do nome do produto
- 4. product\_description\_lenght: Comprimento da descrição do produto
- 5. product\_photos\_qty: Número de fotos do produto
- 6. product\_weight\_g: Peso do produto em gramas
- 7. product\_lenght\_cm: Comprimento do produto em centímetros
- 8. product\_height\_cm: Altura do produto em centímetros
- 9. product\_width\_cm: Largura do produto em centímetros



## 3.i. Bases Originais

14

Descrição da base olist customers dataset:

### Quantidade de registros

- 99.441

### Quantidade de variáveis

- 5

### Nomes das variáveis

- 1. customer\_id: Chave para o dataset de pedidos, cada pedido tem um customer\_id único
- 2. customer\_unique\_id: Id único do comprador
- 3. customer\_zip\_code\_prefix: Cinco primeiros dígitos do CEP do comprador
- 4. customer\_city: Cidade do comprador
- 5. customer\_state: Estado do comprador





## 3.i. Bases Originais

15

Descrição da base olist\_order\_reviews\_dataset:

### Quantidade de registros

- 100.000

### Quantidade de variáveis

- 7

### Nomes das variáveis

- 1. review\_id: Id único do review
- 2. order\_id: Id único do pedido
- 3. review\_score: Nota de 1 a 5 dada pelo comprador
- 4. review\_comment\_title: Título do review
- 5. review\_comment\_message: Comentário do review
- 6. review\_creation\_date: Data que a pesquisa de satisfação foi enviada ao comprador
- 7. review\_answer\_timestamp: Data que o comprador respondeu à pesquisa de satisfação



## 3.ii. Transformações das Bases de Dados

16

### Bases Originais

- 6 tabelas diferentes
- 40 variáveis

### Transformações

- Joins das tabelas
- Remoção de 12 variáveis não utilizadas

### Base Final

- Analytical Base Table (ABT) para segmentação de clientes através de modelos de aprendizagem não supervisionados. Como os clientes da Olist são os vendedores (sellers), precisamos de uma tabela com as características de cada seller:
- Contagem de order\_id únicos (qtd\_pedidos);
- Contagem de product\_id vendidos pelo seller (qtd\_produtos);
- Divisão da contagem de product\_id pela contagem de order\_id únicos (produtos\_por\_pedido);
- Soma dos price e freight\_value do seller (valor\_de\_venda);
- Divisão de valor\_de\_venda por contagem de order\_id únicos (ticket\_medio);
- Contagem de customer\_unique\_id únicos do seller (qtd\_compradores);
- Contagem de order\_status únicos, para quantificar a qualidade de entrega dos vendedores;
- Recência = última data do dataset menos data do último pedido do vendedor;
- Frequência = média de número de vendas por mês;
- Média das diferenças entre order\_estimated\_delivery\_date e order\_delivered\_customer\_date do seller (dias\_atraso\_medio);
- Média de review\_score do seller;
- Categorias de produtos vendidas pelo seller;
- Coluna seller\_state.



# Metodologia de análise de dados



## Definição do problema

- Objetivos
- Conceitos
- Critérios
- Histórico de dados
- Variáveis

## Análise preliminar

- Medidas de posição
- Análise de frequências
- Gráficos
- Análise de *outliers*
- Análise de *missings*
- Validação sobre a consistência das informações

## Avaliação das técnicas

- K-means nativo do Spark

## Avaliação das técnicas

- biSecting K-means nativo do Spark
- Gaussian Mixture Model nativo do Spark
- Aglomeração do Scikit Learn
- DBSCAN do Scikit Learn
- MeanShift do Scikit Learn
- K-means do Scikit Learn

## Definição da técnica

- Validação dos resultados
- Escolha da técnica que melhor se adequa ao uso e estratégias

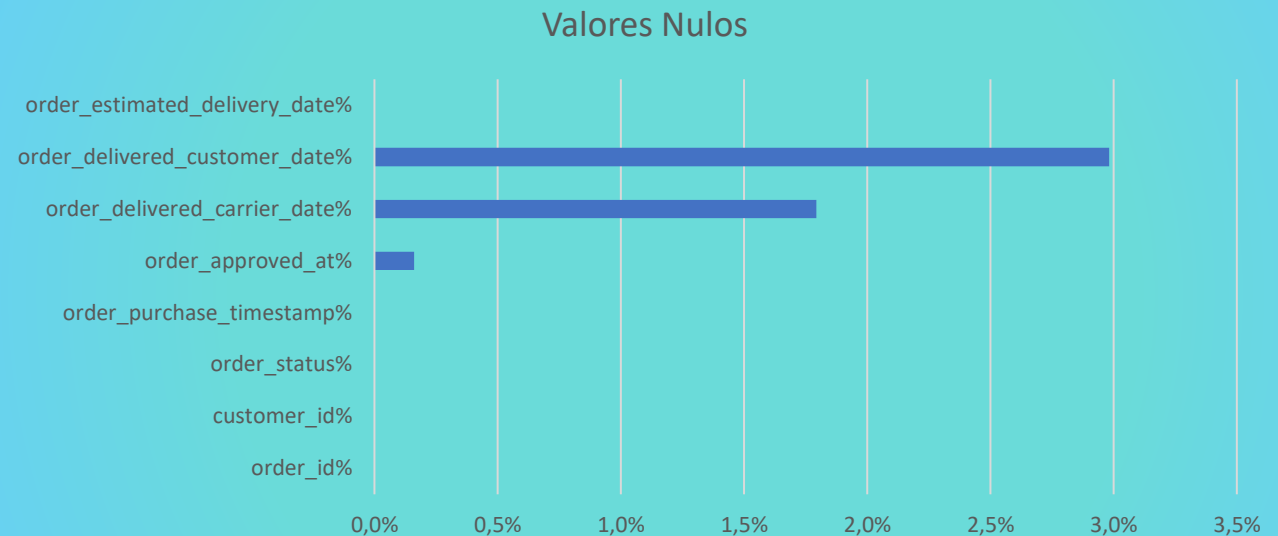




## 4. Análise Exploratória de Dados

18

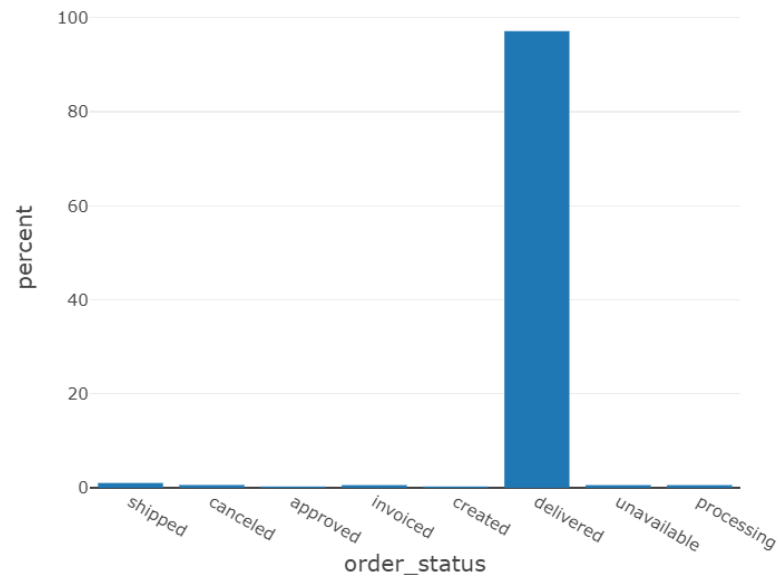
Através da exploração dos dados, definimos qual direcionamento daríamos para criação dos modelos e os tratamentos necessários para as bases de dados.



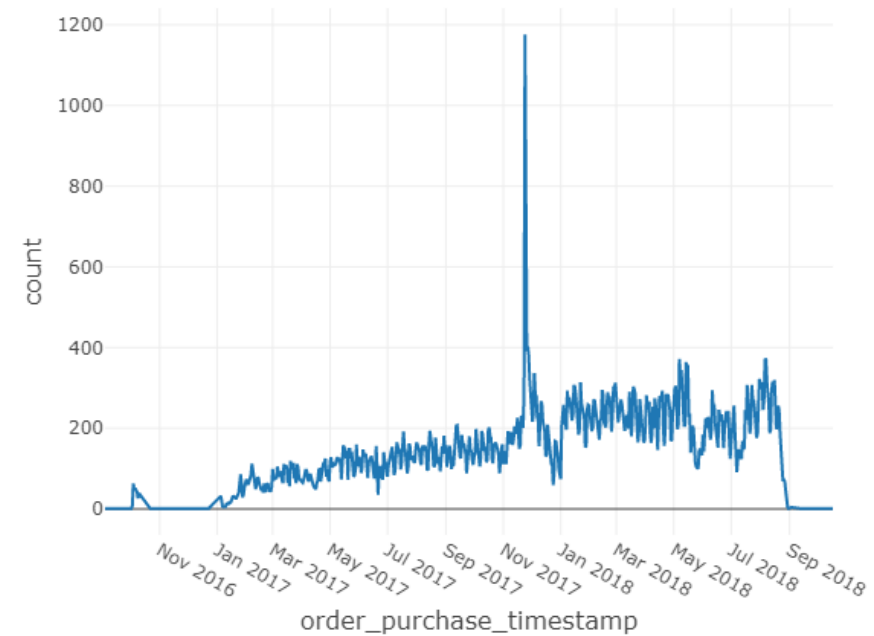
Como os valores nulos representam um volume pequeno nas bases utilizadas para a criação da ABT, decidimos por apenas removê-los.  
O gráfico de valores nulos da base **df\_orders** exemplifica isso.

## 4. Análise Exploratória de Dados

A distribuição da variável `order_status` mostra que a grande maioria das vendas já foram entregues.



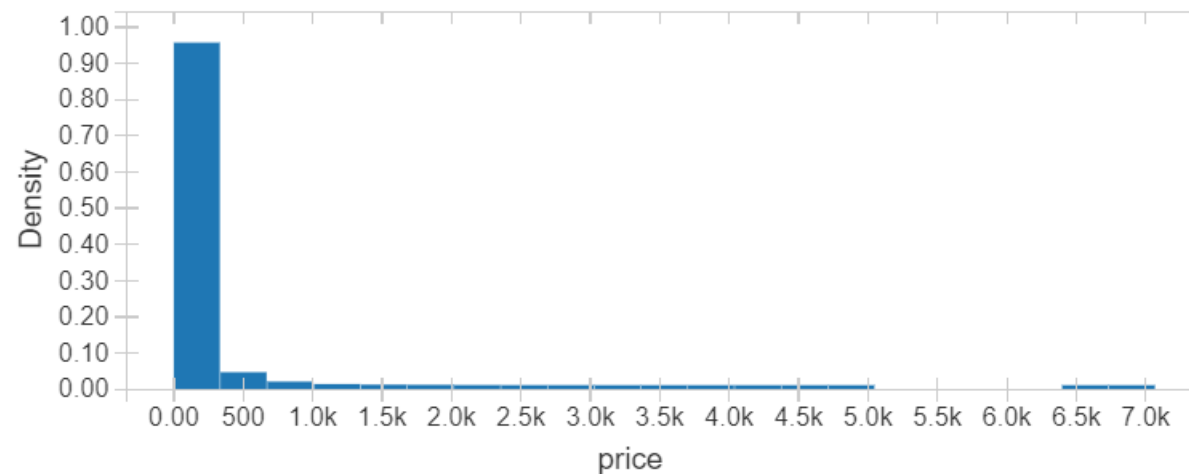
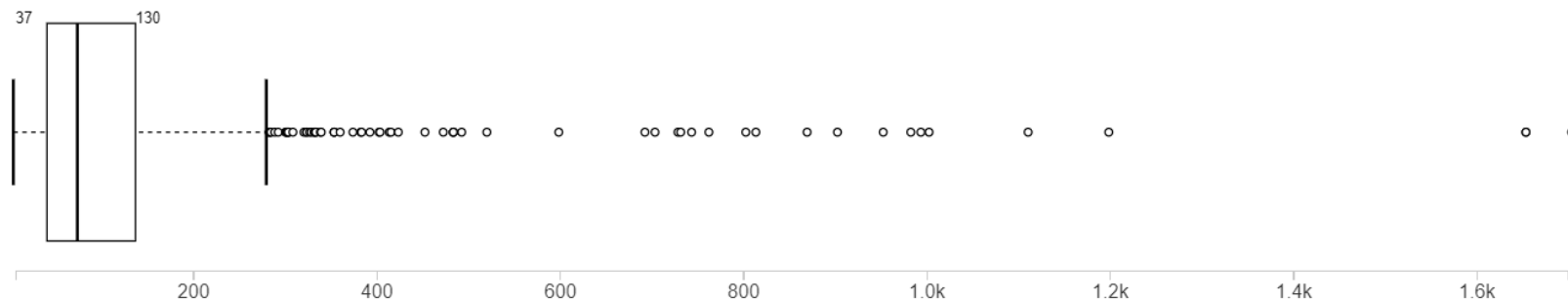
Notamos um pico no número de registro do dia 24 de Novembro de 2017, isso se dá pois a Black Friday ocorreu neste dia.



## 4. Análise Exploratória de Dados

Podemos ver através da distribuição da variável price que 25% dos produtos vendidos estão na faixa até R\$37 e 75% na faixa até R\$130.

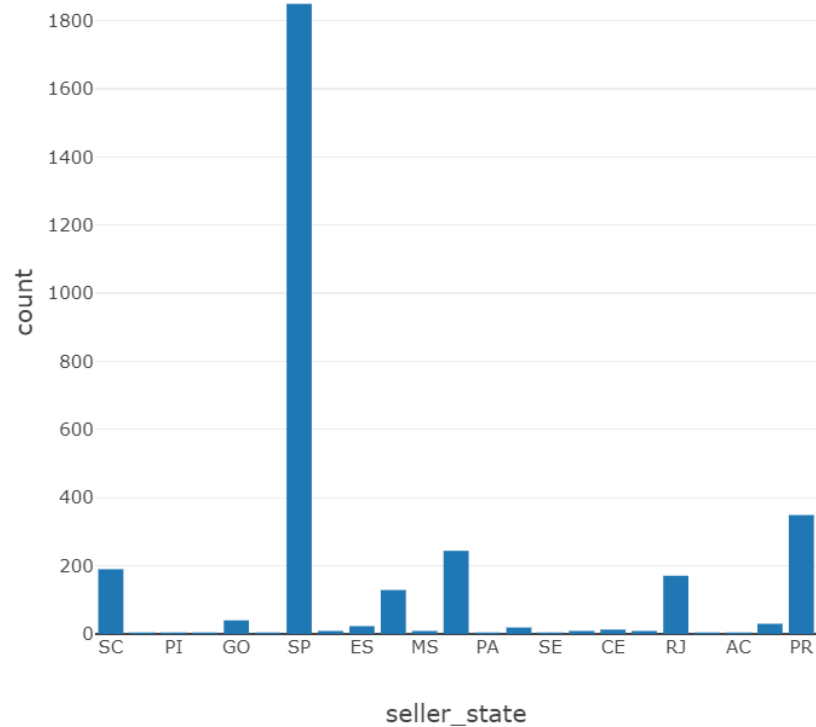
Existem registros de valores muito mais altos, mas são exceções.



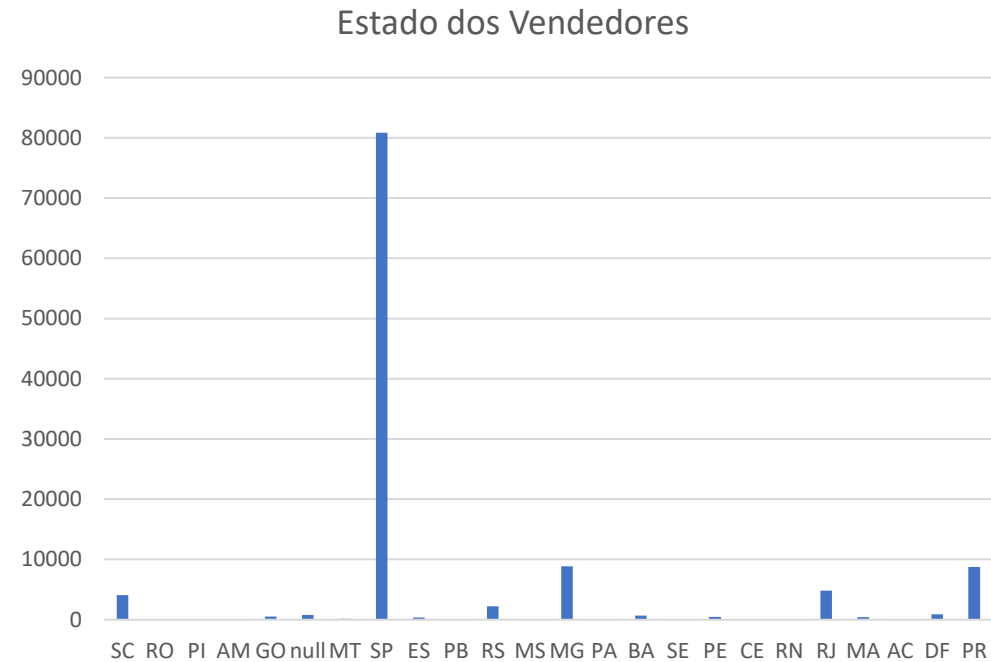


## 4. Análise Exploratória de Dados

Grande maioria dos vendedores registrados na base de vendedores está em São Paulo. Na sequência, vêm Paraná e Minas Gerais.

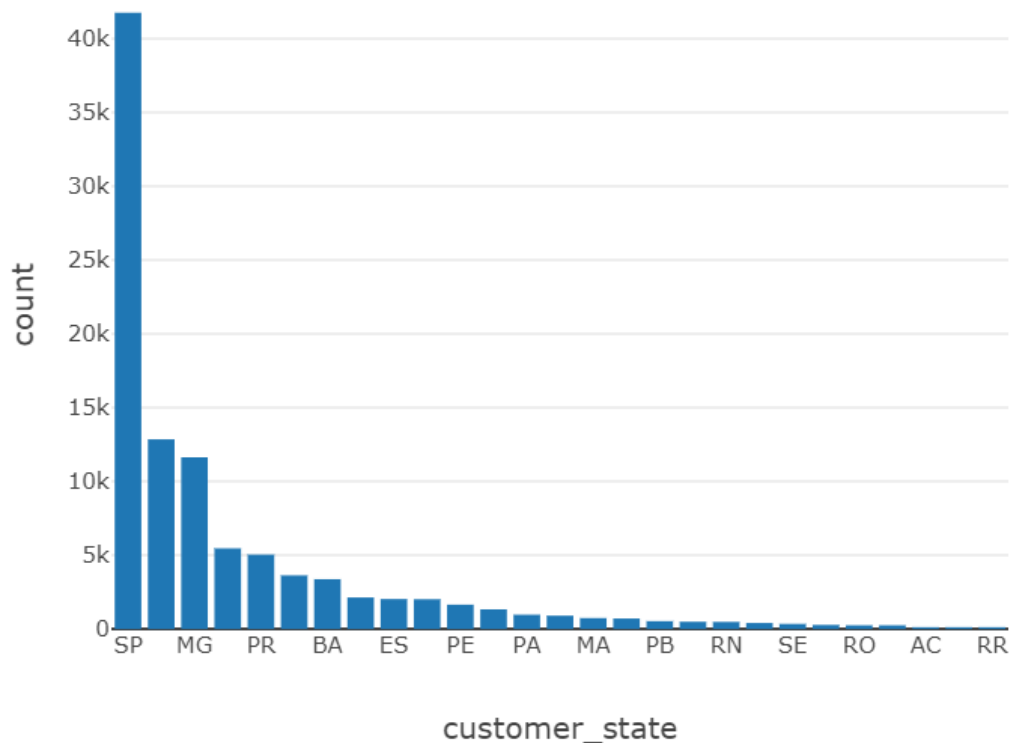


Após juntarmos as bases e comparar o número de vendas por estado do vendedor, podemos ver que os vendedores de São Paulo realizam muito mais vendas que os de outros estados.



## 4. Análise Exploratória de Dados

A exemplo da base de vendedores, os compradores do estado de São Paulo são maioria, porém os compradores do Rio de Janeiro passam ao segundo lugar.



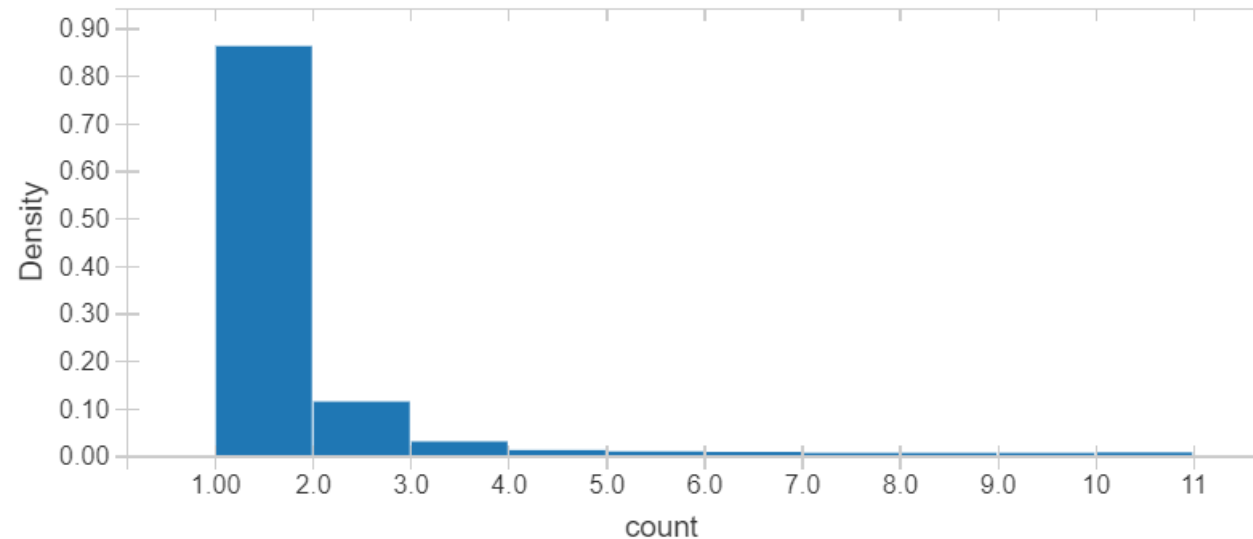
Após junção das bases, vemos que as vendas em sua maioria foram também para São Paulo.



## 4. Análise Exploratória de Dados

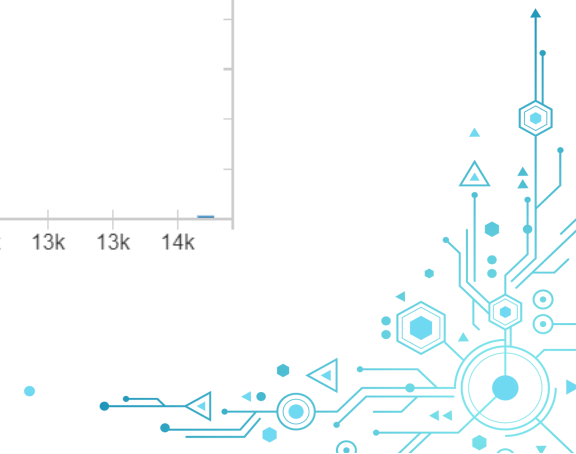
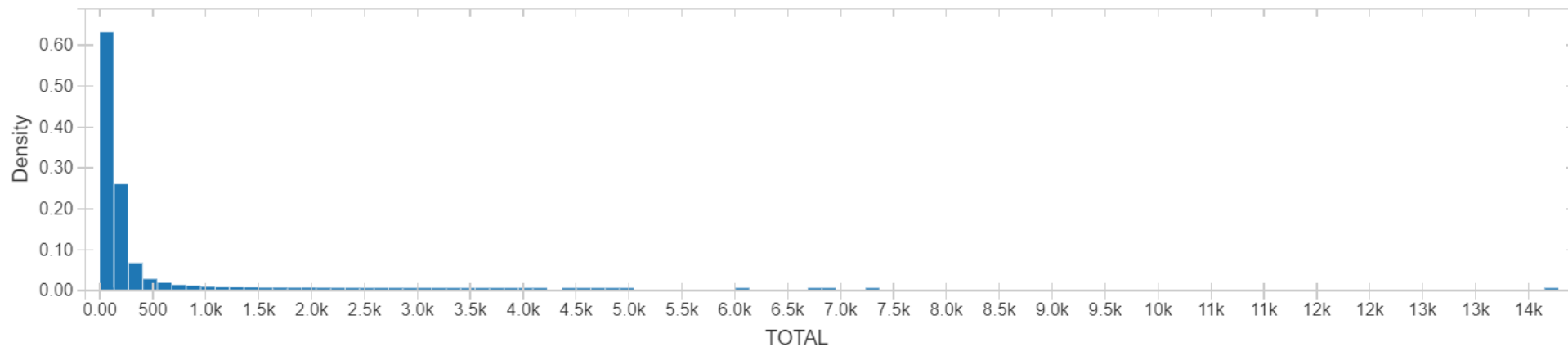
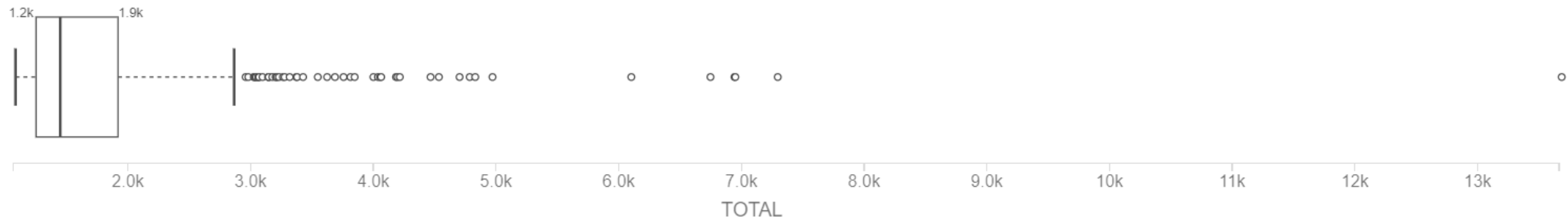
23

Pela distribuição da variável `customer_unique_id` nos pedidos após junção das bases, vemos que poucos clientes compraram mais de uma vez no período de extração da base.



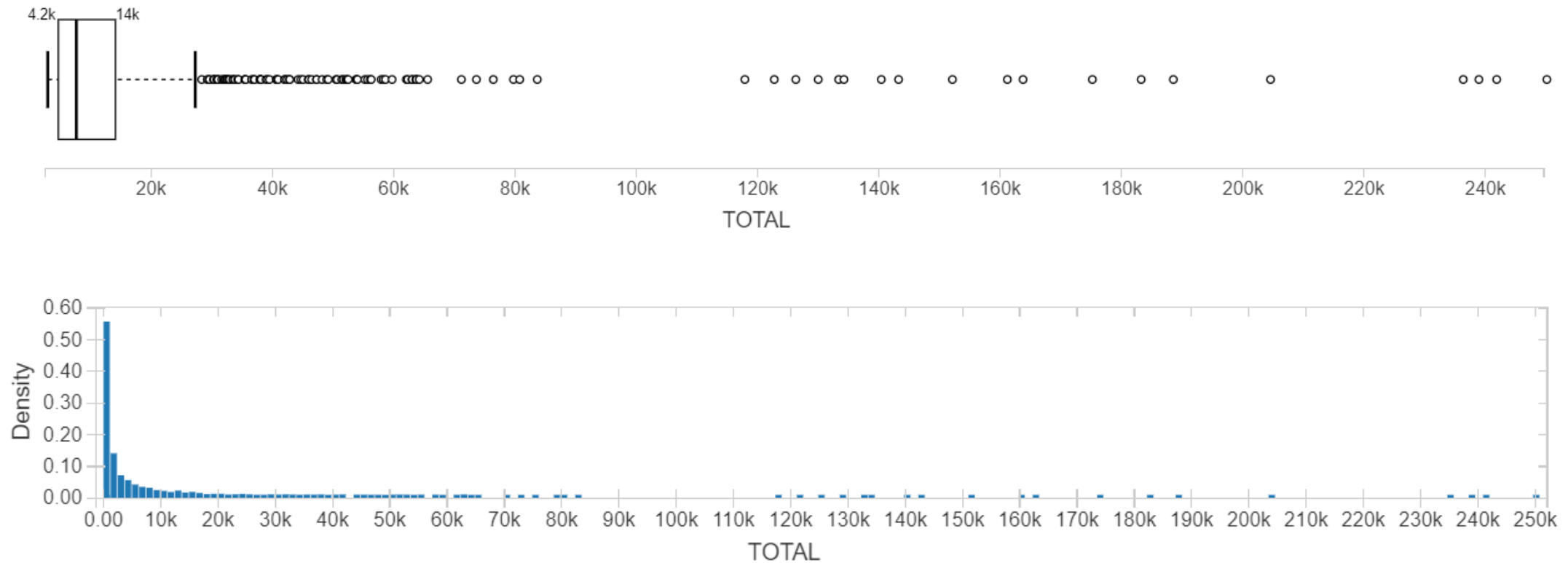
## 4. Análise Exploratória de Dados

Os pedidos tem média de R\$161,32, porém podemos ver que existem alguns pedidos de valores bem maiores na base.



## 4. Análise Exploratória de Dados

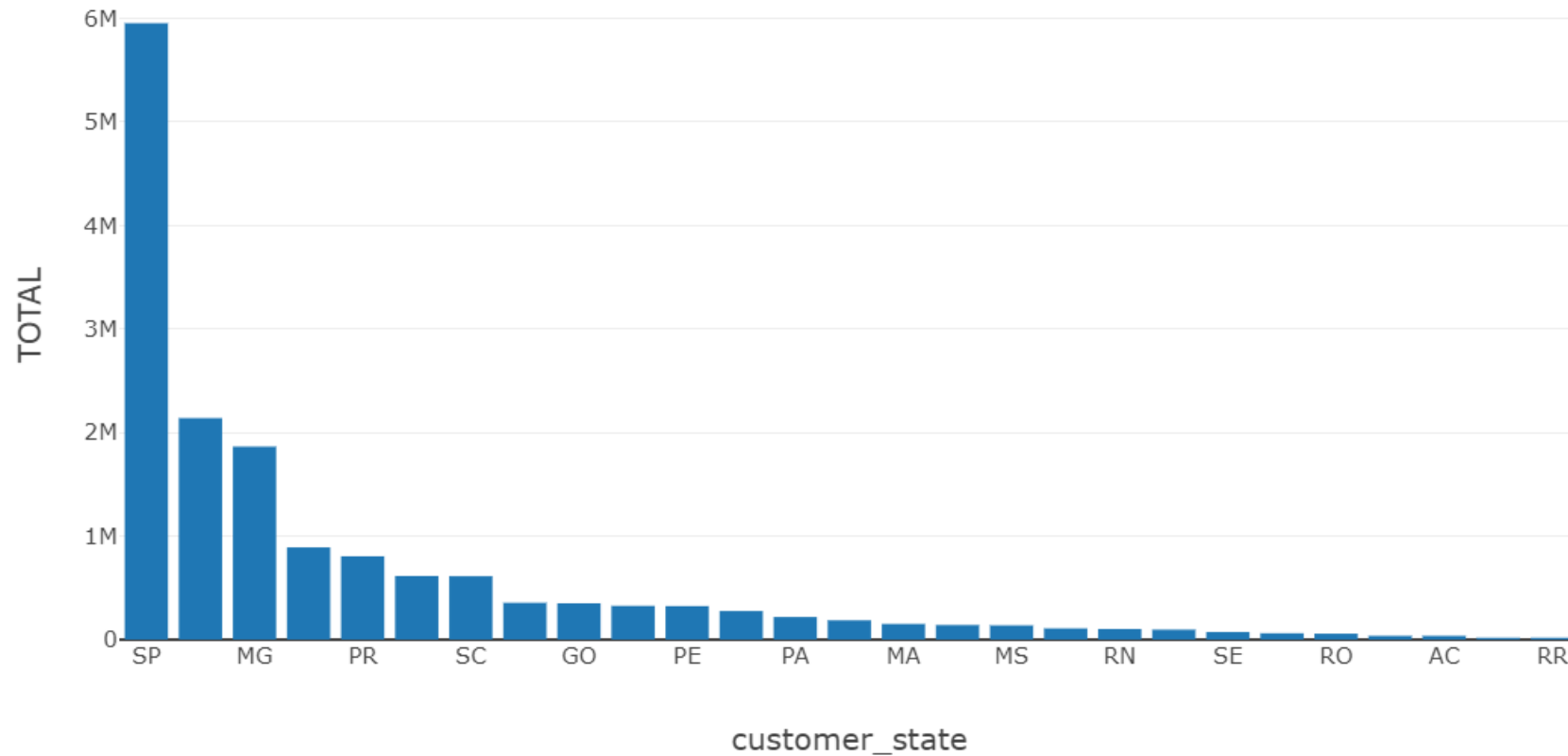
Os gráficos abaixo representam o faturamento total dos vendedores no período, é possível observar que há uma grande quantidade de vendedores que faturam pouco e poucos vendedores que faturam muito.



## 4. Análise Exploratória de Dados

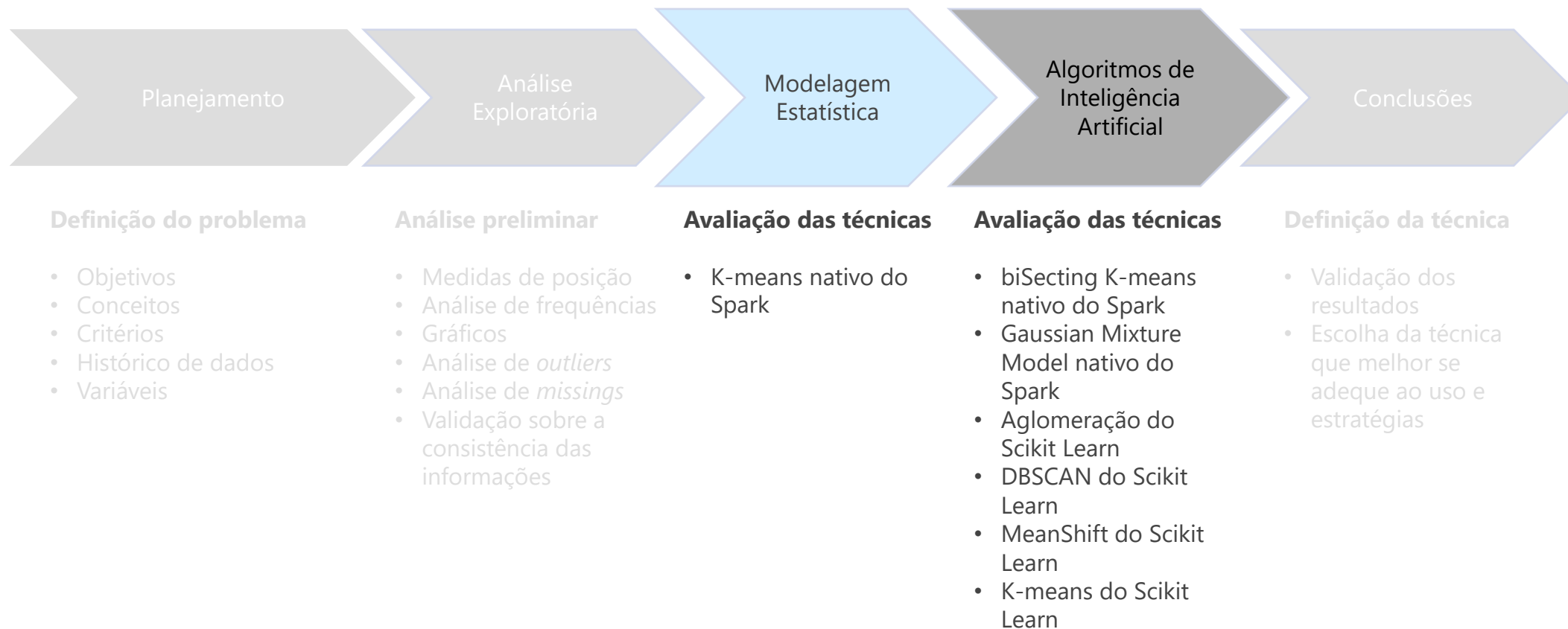
26

O valor vendido para o estado de São Paulo é praticamente o triplo do segundo estado com maior valor, Rio de Janeiro.





# Metodologia de análise de dados

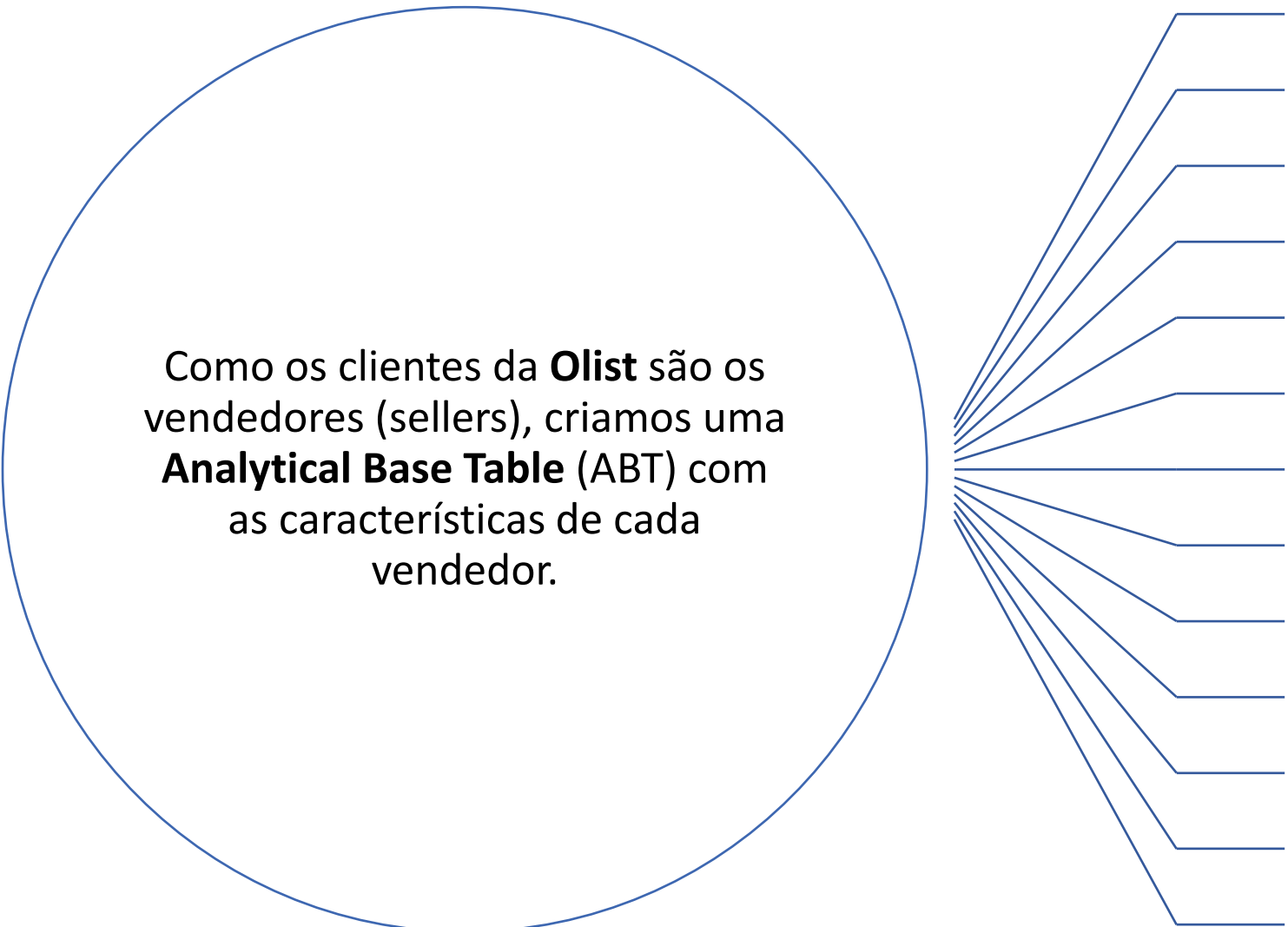


# 5. Modelagem

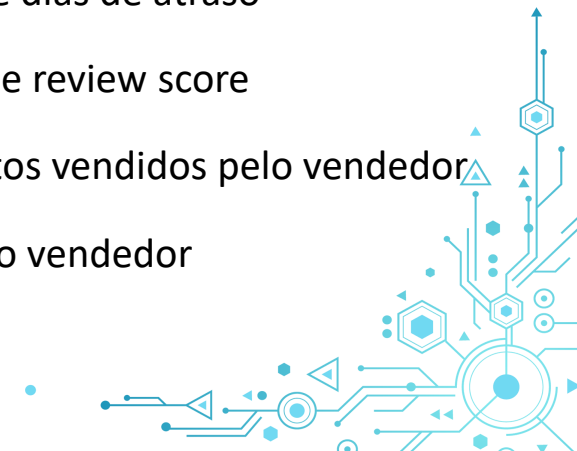
ANALYTICAL BASE TABLE | SELEÇÃO DE VARIÁVEIS

28

Como os clientes da **Olist** são os vendedores (sellers), criamos uma **Analytical Base Table (ABT)** com as características de cada vendedor.



- # Pedidos
- # Produtos
- # Produtos por Pedido
- \$ Valor de Venda
- \$ Ticket Médio
- # Compradores
- # Pedidos por Status
- Recência (dias desde a última venda)
- Frequência (vendas por mês)
- Média de dias de atraso
- Média de review score
- Categorias de produtos vendidos pelo vendedor
- UF do vendedor

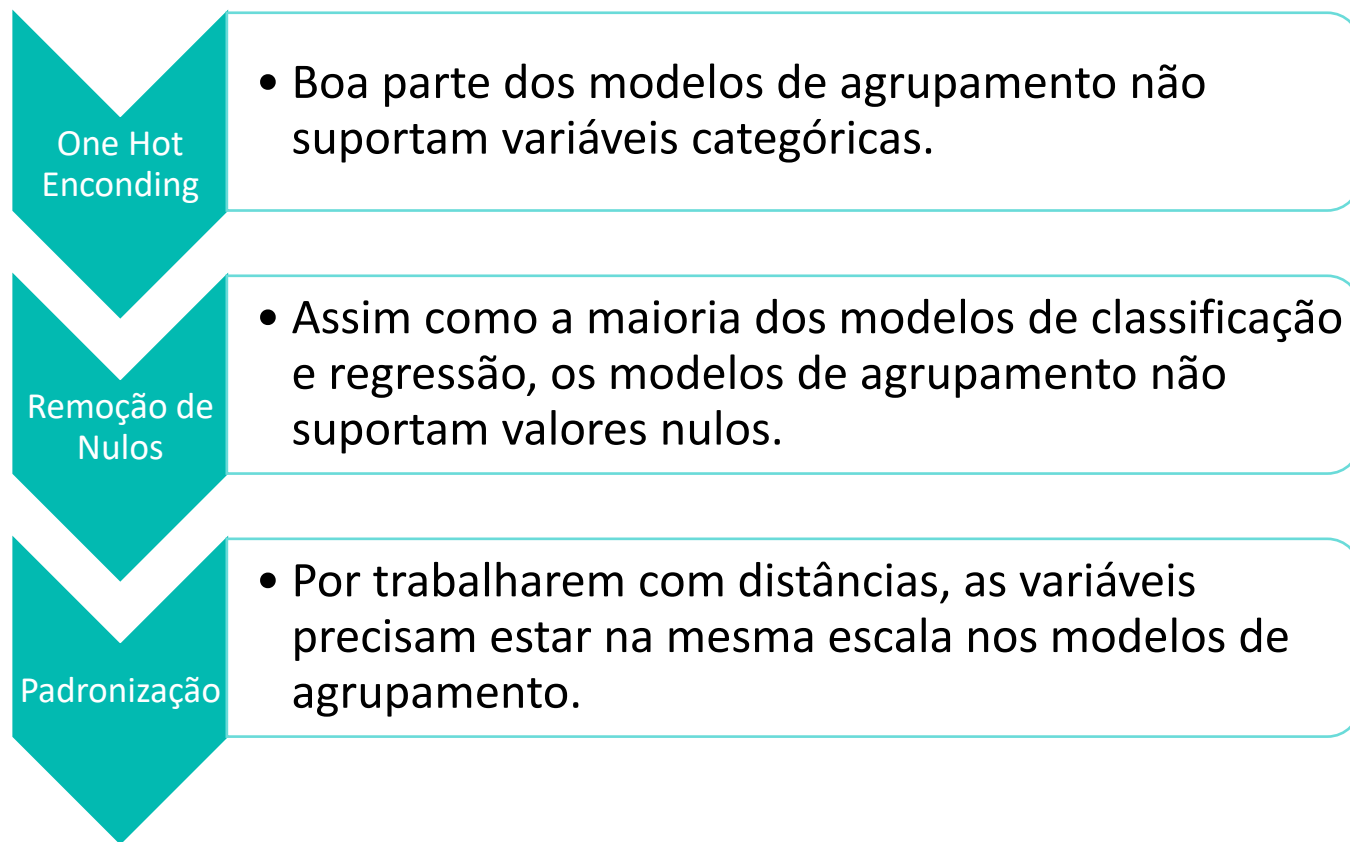


# 5. Modelagem

ANALYTICAL BASE TABLE | FEATURE ENGINEERING

29

Devido à aplicação de modelos de clustering, alguns tratamentos foram aplicados na base para que o treinamento fosse permitido.

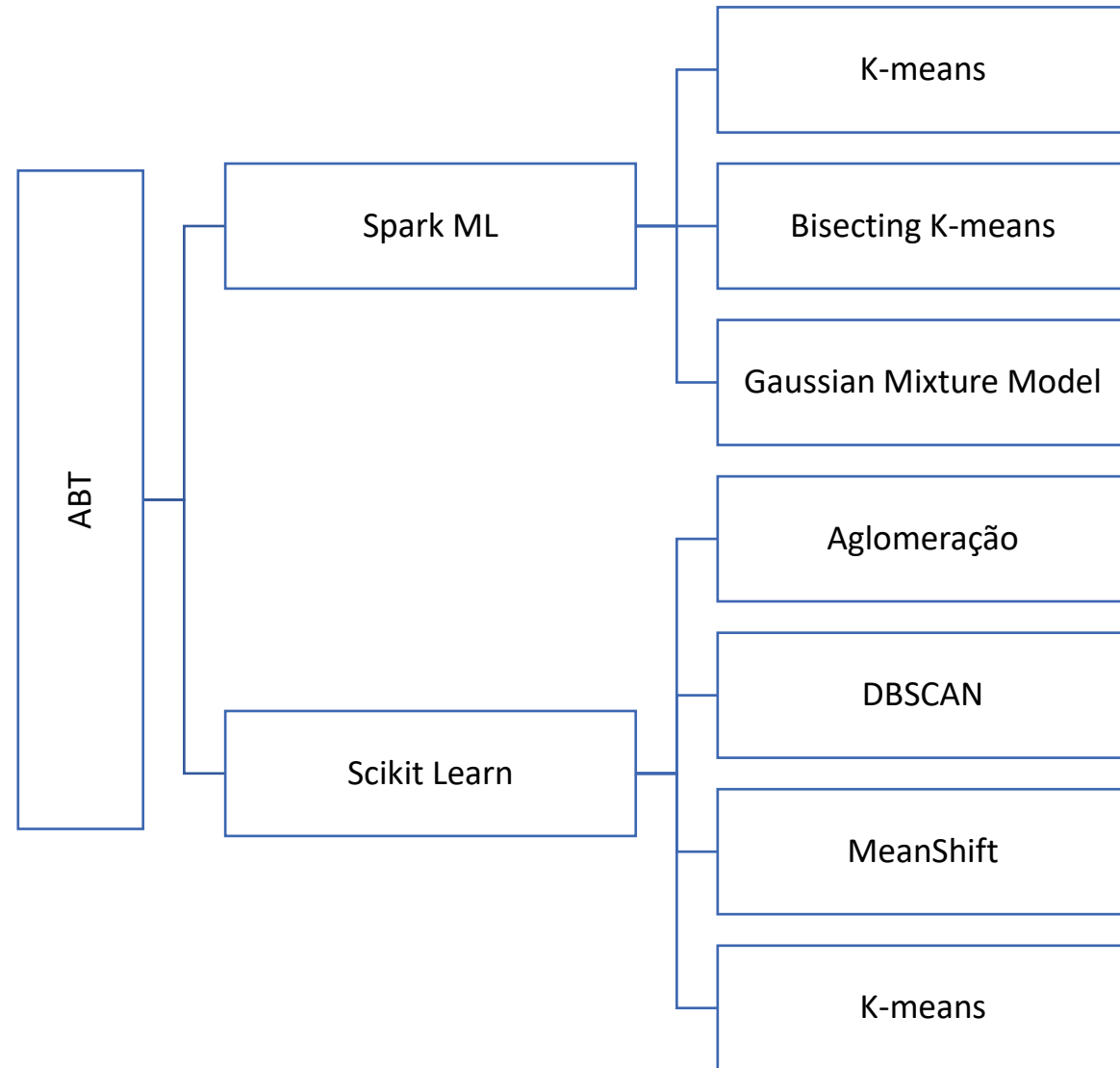


# 5. Modelagem

SEGMENTAÇÃO | APRENDIZADO DE MÁQUINA

Na sequência, fizemos a segmentação de vendedores através de modelos de aprendizagem de máquina não supervisionados: quando não temos a variável target, desenvolvemos modelos matemáticos complexos que testam a correlação entre as variáveis explicativas até encontrar padrões que possibilitem o agrupamento.

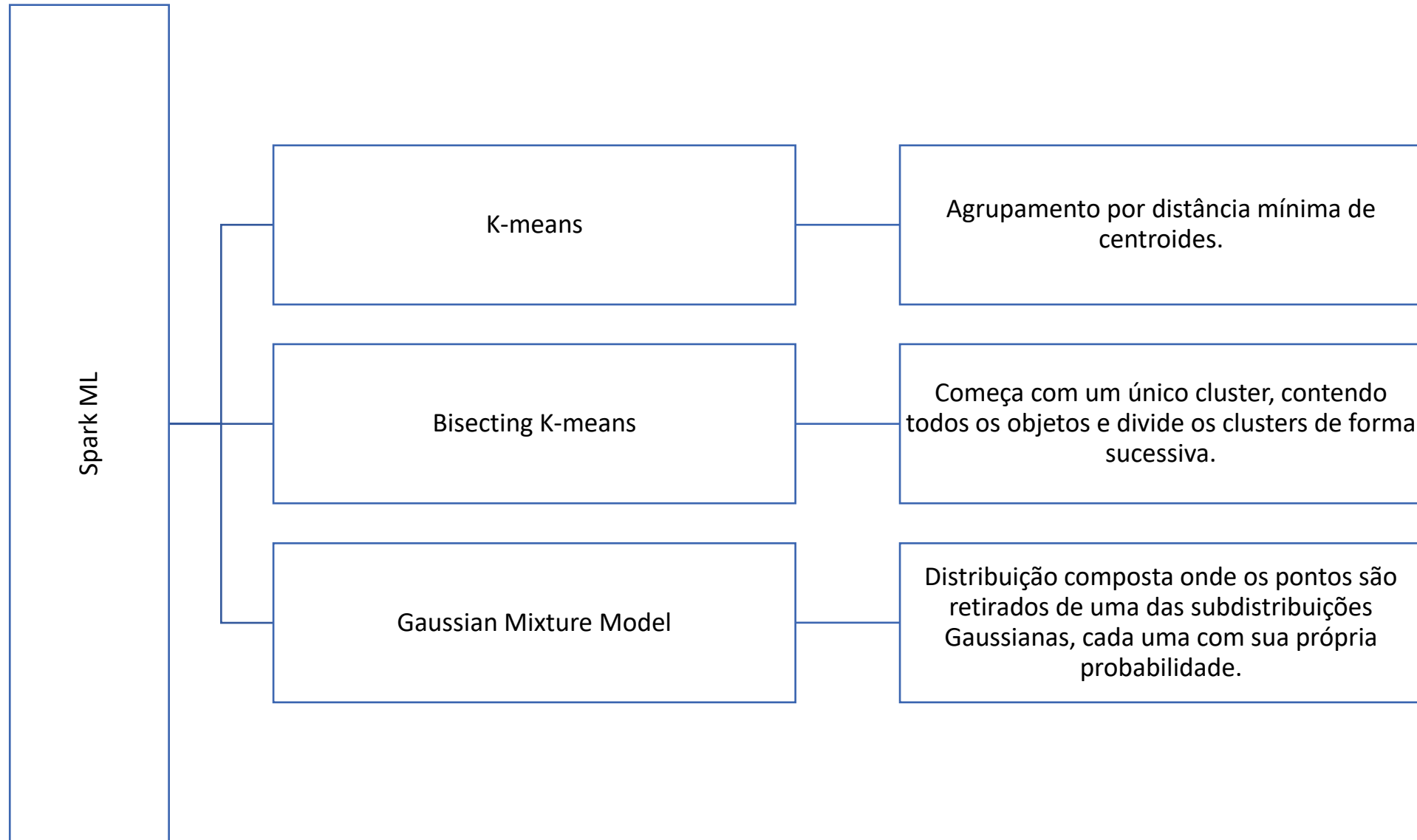
Para definir a melhor forma de agrupamento, utilizamos técnicas de modelagem nativas do Spark e do Scikit Learn, com diferentes graus de sucesso.



# 5. Modelagem

TÉCNICAS DE AGRUPAMENTO | SPARK ML

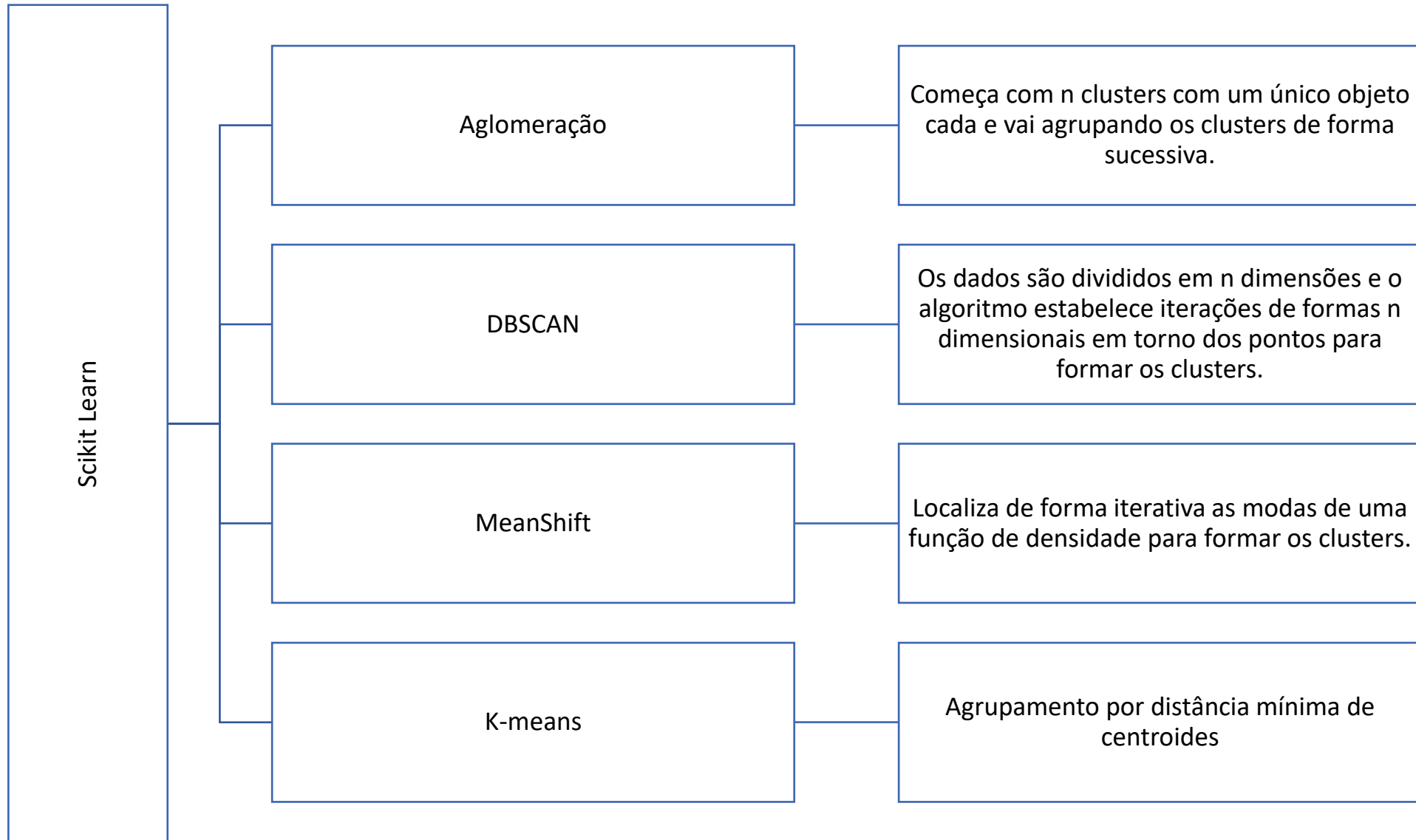
31



# 5. Modelagem

TÉCNICAS DE AGRUPAMENTO | SCIKIT LEARN

32





# 5. Modelagem

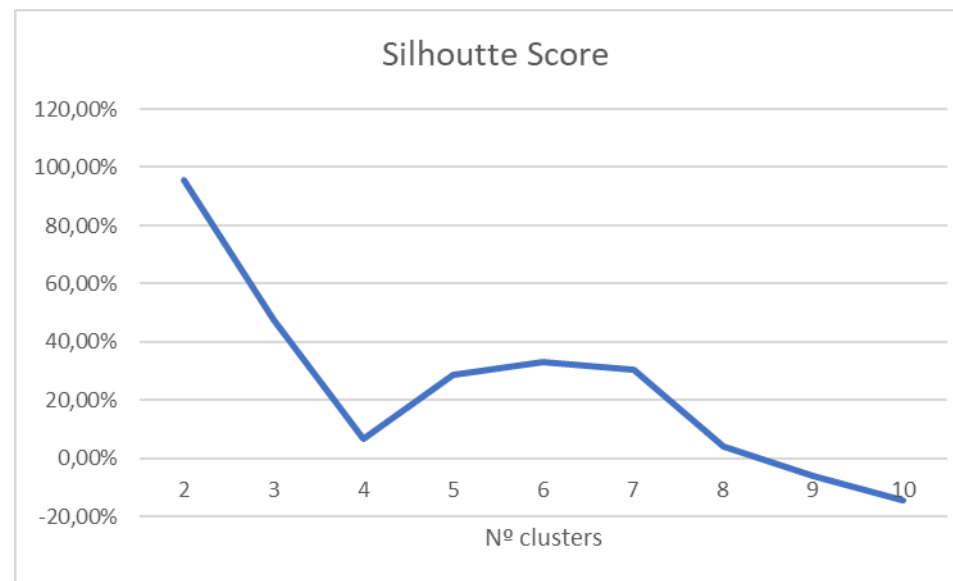
RESULTADOS | SPARK ML K-MEANS

33

## Spark ML K-means

Chegamos à possibilidade de 3 diferentes agrupamentos: modelos com 2, 3 ou 6 grupos.

Em todos eles o fator mais relevante aparenta ser o número de vendas, logo podemos inferir que a segmentação ocorre principalmente entre vendedores grandes (provavelmente grandes lojas e varejistas) e vendedores menores.



# 5. Modelagem

RESULTADOS | SPARK ML K-MEANS

## Spark ML K-means

Chegamos à possibilidade de 3 diferentes agrupamentos: modelos com 2, 3 ou 6 grupos.

Em todos eles o fator mais relevante aparenta ser o número de vendas, logo podemos inferir que a segmentação ocorre principalmente entre vendedores grandes (provavelmente grandes lojas e varejistas) e vendedores menores.

K = 2
<ul style="list-style-type: none"><li>• Silhoutte score: 95%</li><li>• Grupo 0: 3.023 vendedores</li><li>• Grupo 1: 72 vendedores</li></ul>

K = 3
<ul style="list-style-type: none"><li>• Silhoutte score: 47%</li><li>• Grupo 0: 2.999 vendedores</li><li>• Grupo 1: 33 vendedores</li><li>• Grupo 2: 72 vendedores</li></ul>

K = 6
<ul style="list-style-type: none"><li>• Silhoutte score: 33%</li><li>• Grupo 0: 2.732 vendedores</li><li>• Grupo 1: 2 vendedores</li><li>• Grupo 2: 23 vendedores</li><li>• Grupo 3: 331 vendedores</li><li>• Grupo 4: 6 vendedores</li><li>• Grupo 5: 1 vendedor</li></ul>

Modelo separou os vendedores entre os grandes vendedores (minoridade) e os pequenos vendedores (maioria). Não faz muito sentido do ponto de vista de negócios pois a divisão é muito radical: modelo descartado



# 5. Modelagem

RESULTADOS | SPARK ML K-MEANS

## Spark ML K-means

Chegamos à possibilidade de 3 diferentes agrupamentos: modelos com 2, 3 ou 6 grupos.

Em todos eles o fator mais relevante aparenta ser o número de vendas, logo podemos inferir que a segmentação ocorre principalmente entre vendedores grandes (provavelmente grandes lojas e varejistas) e vendedores menores.

K = 2
<ul style="list-style-type: none"><li>• Silhoutte score: 95%</li><li>• Grupo 0: 3.023 vendedores</li><li>• Grupo 1: 72 vendedores</li></ul>

K = 3
<ul style="list-style-type: none"><li>• Silhoutte score: 47%</li><li>• Grupo 0: 2.999 vendedores</li><li>• Grupo 1: 33 vendedores</li><li>• Grupo 2: 72 vendedores</li></ul>

K = 6
<ul style="list-style-type: none"><li>• Silhoutte score: 33%</li><li>• Grupo 0: 2.732 vendedores</li><li>• Grupo 1: 2 vendedores</li><li>• Grupo 2: 23 vendedores</li><li>• Grupo 3: 331 vendedores</li><li>• Grupo 4: 6 vendedores</li><li>• Grupo 5: 1 vendedor</li></ul>

Grupo 0 faz poucas vendas e só é interessante para a Olist no volume. Grupo 1 faz mais vendas do que a maioria dos vendedores, mas ainda num patamar bem abaixo do que os vendedores premium. Para compensar, vende produtos mais caros. Grupo 2 faz muitas vendas e garante bom faturamento. São perfis distintos o suficiente para aplicar na política comercial: modelo selecionado para análise final.



# 5. Modelagem

RESULTADOS | SPARK ML K-MEANS

## Spark ML K-means

Chegamos à possibilidade de 3 diferentes agrupamentos: modelos com 2, 3 ou 6 grupos.

Em todos eles o fator mais relevante aparenta ser o número de vendas, logo podemos inferir que a segmentação ocorre principalmente entre vendedores grandes (provavelmente grandes lojas e varejistas) e vendedores menores.

K = 2	K = 3	K = 6
<ul style="list-style-type: none"><li>• Silhoutte score: 95%</li><li>• Grupo 0: 3.023 vendedores</li><li>• Grupo 1: 72 vendedores</li></ul>	<ul style="list-style-type: none"><li>• Silhoutte score: 47%</li><li>• Grupo 0: 2.999 vendedores</li><li>• Grupo 1: 33 vendedores</li><li>• Grupo 2: 72 vendedores</li></ul>	<ul style="list-style-type: none"><li>• Silhoutte score: 33%</li><li>• Grupo 0: 2.732 vendedores</li><li>• Grupo 1: 2 vendedores</li><li>• Grupo 2: 23 vendedores</li><li>• Grupo 3: 331 vendedores</li><li>• Grupo 4: 6 vendedores</li><li>• Grupo 5: 1 vendedor</li></ul>

Grupo 1 tem apenas 2 vendedores, de perfis completamente distintos: um tem 3 pedidos, o outro 396; um vendeu 4 produtos, o outro vendeu 430; um faturou 945 e o outro 18.470. Grupo 5 tem apenas 1 vendedor, e Grupo 4 tem apenas 6 vendedores. Concluimos que não faz sentido do ponto de vista de negócios, pois qualquer decisão a ser tomada para afetar menos de 1% dos vendedores da base, mesmo que traga benefício, será muito custosa considerando horas gastas para análise, desenho de solução e implementação.



# 5. Modelagem

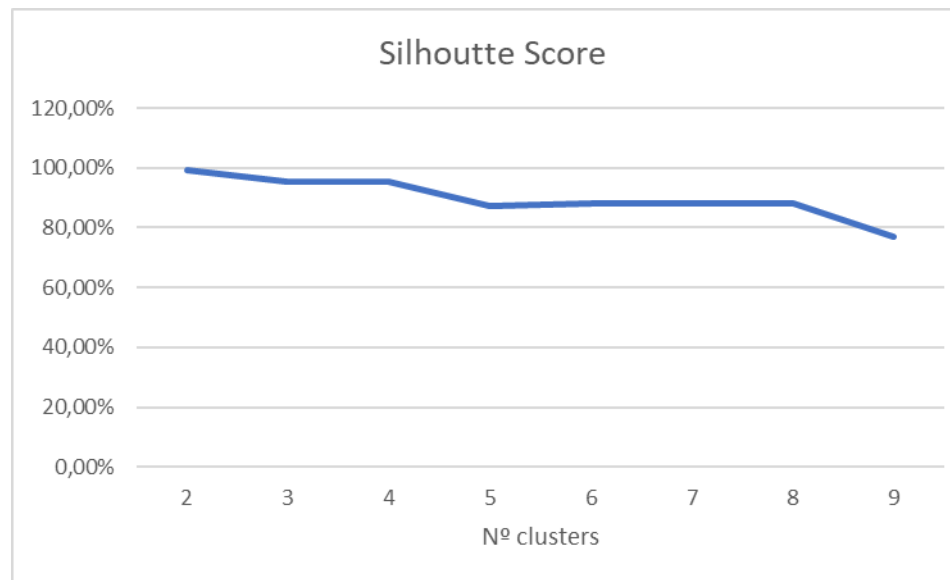
RESULTADOS | SPARK ML BISECTING K-MEANS

37

## Spark ML Bisecting K-means

Chegamos à possibilidade de 3 diferentes agrupamentos: modelos com 2, 3 ou 4 grupos.

É possível perceber que as variáveis quantidade de pedidos e valor de venda tem grande influência na segmentação dos clientes, classificando-os em vendedores pequenos, médios e grandes.



# 5. Modelagem

RESULTADOS | SPARK ML BISECTING K-MEANS

38

## Spark ML Bisecting K-means

Chegamos à possibilidade de 3 diferentes agrupamentos: modelos com 2, 3 ou 4 grupos.

É possível perceber que as variáveis quantidade de pedidos e valor de venda tem grande influência na segmentação dos clientes, classificando-os em vendedores pequenos, médios e grandes.

K = 2

- Silhoutte score: 99%
- Grupo 0: 3.076 vendedores
- Grupo 1: 19 vendedores

K = 3

- Silhoutte score: 95%
- Grupo 0: 2.957 vendedores
- Grupo 1: 119 vendedores
- Grupo 2: 19 vendedores

K = 4

- Silhoutte score: 95%
- Grupo 0: 2957 vendedores
- Grupo 1: 119 vendedores
- Grupo 2: 12 vendedores
- Grupo 3: 7 vendedores

O resultado do biSec K-means com 2 clusters ficou semelhante ao K-means nativo do Spark, mas com desempenho inferior: lá, o segundo grupo ficou com 72 vendedores. Sendo assim, vamos descartar a análise de perfil desta técnica.



# 5. Modelagem

RESULTADOS | SPARK ML BISECTING K-MEANS

39

## Spark ML Bisecting K-means

Chegamos à possibilidade de 3 diferentes agrupamentos: modelos com 2, 3 ou 4 grupos.

É possível perceber que as variáveis quantidade de pedidos e valor de venda tem grande influência na segmentação dos clientes, classificando-os em vendedores pequenos, médios e grandes.

K = 2

- Silhoutte score: 99%
- Grupo 0: 3.076 vendedores
- Grupo 1: 19 vendedores

K = 3

- Silhoutte score: 95%
- Grupo 0: 2.957 vendedores
- Grupo 1: 119 vendedores
- Grupo 2: 19 vendedores

K = 4

- Silhoutte score: 95%
- Grupo 0: 2957 vendedores
- Grupo 1: 119 vendedores
- Grupo 2: 12 vendedores
- Grupo 3: 7 vendedores

O resultado do biSec K-means com 3 clusters ficou tão interessante quanto o K-means nativo do Spark: lá, o 1º grupo ficou com 2990 vendedores, o 2º grupo ficou com 33 e o 3º com 72. Vamos realizar a análise de perfil desta técnica e comparar com os perfis gerados através do K-means nativo do Spark com 3 clusters para definir qual o melhor modelo.



# 5. Modelagem

RESULTADOS | SPARK ML BISECTING K-MEANS

40

## Spark ML Bisecting K-means

Chegamos à possibilidade de 3 diferentes agrupamentos: modelos com 2, 3 ou 4 grupos.

É possível perceber que as variáveis quantidade de pedidos e valor de venda tem grande influência na segmentação dos clientes, classificando-os em vendedores pequenos, médios e grandes.

K = 2

- Silhoutte score: 99%
- Grupo 0: 3.076 vendedores
- Grupo 1: 19 vendedores

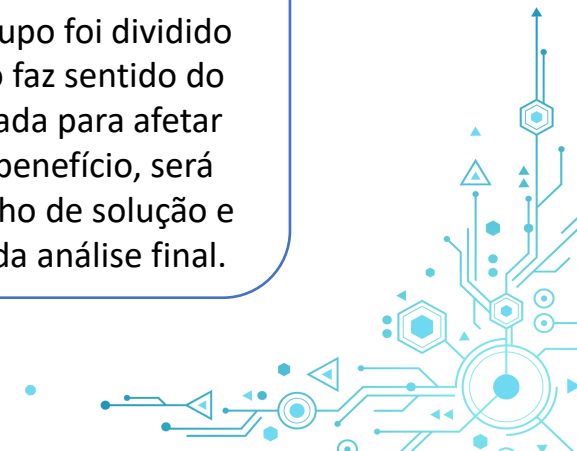
K = 3

- Silhoutte score: 95%
- Grupo 0: 2.957 vendedores
- Grupo 1: 119 vendedores
- Grupo 2: 19 vendedores

K = 4

- Silhoutte score: 95%
- Grupo 0: 2957 vendedores
- Grupo 1: 119 vendedores
- Grupo 2: 12 vendedores
- Grupo 3: 7 vendedores

O resultado do biSec K-means com 4 clusters não parece agregar muito em comparação com o agrupamento com 3 clusters visto acima. Não houve diferença no nº de vendedores do 1º e do 2º grupo, e o 3º grupo foi dividido em 2, com pouca diferenciação entre si. Concluimos que não faz sentido do ponto de vista de negócios, pois qualquer decisão a ser tomada para afetar menos de 0,1% dos vendedores da base, mesmo que traga benefício, será muito custosa considerando horas gastas para análise, desenho de solução e implementação. Portanto, vamos excluir esse agrupamento da análise final.





## 5. Modelagem

RESULTADOS | SPARK ML GAUSSIAN MIXTURE MODEL

41

### Spark ML Gaussian Mixture Model

Não conseguimos entender o funcionamento do Gaussian Mixture Model devido à pobreza da documentação. Rodamos com 2 e com 3 clusters e o resultado foi o mesmo: predição = 0 para todas as entradas. Portanto, optamos por abordar o problema utilizando o Pandas e o Scikit Learn, onde aplicaremos 4 técnicas de agrupamento não supervisionado: Aglomeração, DBSCAN, MeanShift e K-means.



# 5. Modelagem

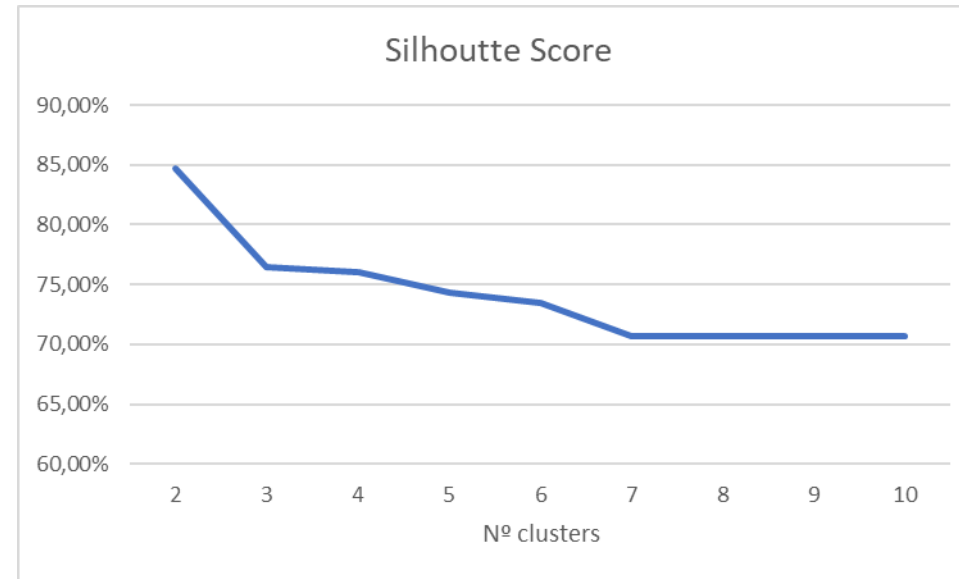
RESULTADOS | SCIKIT LEARN AGLOMERAÇÃO

42

## Scikit Learn Aglomeração

Chegamos à possibilidade de 3 diferentes agrupamentos: modelos com 2, 3 ou 8 grupos.

Infelizmente, apesar do alto silhouette score, as divisões não fizeram sentido do ponto de vista de negócio e decidimos descartar esta técnica.



# 5. Modelagem

RESULTADOS | SCIKIT LEARN AGLOMERAÇÃO

43

## Scikit Learn Aglomeração

Chegamos à possibilidade de 3 diferentes agrupamentos: modelos com 2, 3 ou 8 grupos.

Infelizmente, apesar do alto silhouette score, as divisões não fizeram sentido do ponto de vista de negócio e decidimos descartar esta técnica.

K = 2

- Silhouette score: 84%
- Grupo 0: 3.094 vendedores
- Grupo 1: 1 vendedor

K = 3

- Silhouette score: 76%
- Grupo 0: 3.089 vendedores
- Grupo 1: 1 vendedor
- Grupo 2: 5 vendedores

K = 8

- Silhouette score: 70%
- Grupo 0: 5 vendedores
- Grupo 1: 2 vendedores
- Grupo 2: 3.080 vendedores
- Grupo 3: 1 vendedor
- Grupo 4: 1 vendedor
- Grupo 5: 2 vendedores
- Grupo 6: 3 vendedores
- Grupo 7: 1 vendedor

A divisão por aglomeração com 2 clusters não faz sentido do ponto de vista de negócios, pois qualquer decisão a ser tomada para afetar apenas 1 vendedor da base, mesmo que traga benefício, será muito custosa considerando horas gastas para análise, desenho de solução e implementação. Como já temos bons resultados com a divisão por K-means nativa do Spark, vamos descartar esta divisão.



# 5. Modelagem

RESULTADOS | SCIKIT LEARN AGLOMERAÇÃO

## Scikit Learn Aglomeração

Chegamos à possibilidade de 3 diferentes agrupamentos: modelos com 2, 3 ou 8 grupos.

Infelizmente, apesar do alto silhouette score, as divisões não fizeram sentido do ponto de vista de negócio e decidimos descartar esta técnica.

K = 2

- Silhouette score: 84%
- Grupo 0: 3.094 vendedores
- Grupo 1: 1 vendedor

K = 3

- Silhouette score: 76%
- Grupo 0: 3.089 vendedores
- Grupo 1: 1 vendedor
- Grupo 2: 5 vendedores

K = 8

- Silhouette score: 70%
- Grupo 0: 5 vendedores
- Grupo 1: 2 vendedores
- Grupo 2: 3.080 vendedores
- Grupo 3: 1 vendedor
- Grupo 4: 1 vendedor
- Grupo 5: 2 vendedores
- Grupo 6: 3 vendedores
- Grupo 7: 1 vendedor

A divisão por aglomeração com 3 clusters não faz sentido do ponto de vista de negócios, pois qualquer decisão a ser tomada para afetar menos de 0,5% da base de vendedores, mesmo que traga benefício, será muito custosa considerando horas gastas para análise, desenho de solução e implementação. Como já temos bons resultados com a divisão por K-means nativa do Spark, vamos descartar esta divisão.



# 5. Modelagem

RESULTADOS | SCIKIT LEARN AGLOMERAÇÃO

45

## Scikit Learn Aglomeração

Chegamos à possibilidade de 3 diferentes agrupamentos: modelos com 2, 3 ou 8 grupos.

Infelizmente, apesar do alto silhouette score, as divisões não fizeram sentido do ponto de vista de negócio e decidimos descartar esta técnica.

K = 2

- Silhouette score: 84%
- Grupo 0: 3.094 vendedores
- Grupo 1: 1 vendedor

K = 3

- Silhouette score: 76%
- Grupo 0: 3.089 vendedores
- Grupo 1: 1 vendedor
- Grupo 2: 5 vendedores

K = 8

- Silhouette score: 70%
- Grupo 0: 5 vendedores
- Grupo 1: 2 vendedores
- Grupo 2: 3.080 vendedores
- Grupo 3: 1 vendedor
- Grupo 4: 1 vendedor
- Grupo 5: 2 vendedores
- Grupo 6: 3 vendedores
- Grupo 7: 1 vendedor

A divisão por aglomeração com 8 clusters não faz sentido do ponto de vista de negócios, pois qualquer decisão a ser tomada para afetar menos de 0,5% da base de vendedores e com a maioria dos grupos com 5 ou menos vendedores, mesmo que traga benefício, será muito custosa considerando horas gastas para análise, desenho de solução e implementação. Como já temos bons resultados com a divisão por K-means nativa do Spark, vamos descartar esta divisão e, consequentemente, a técnica de aglomeração como um todo.



# 5. Modelagem

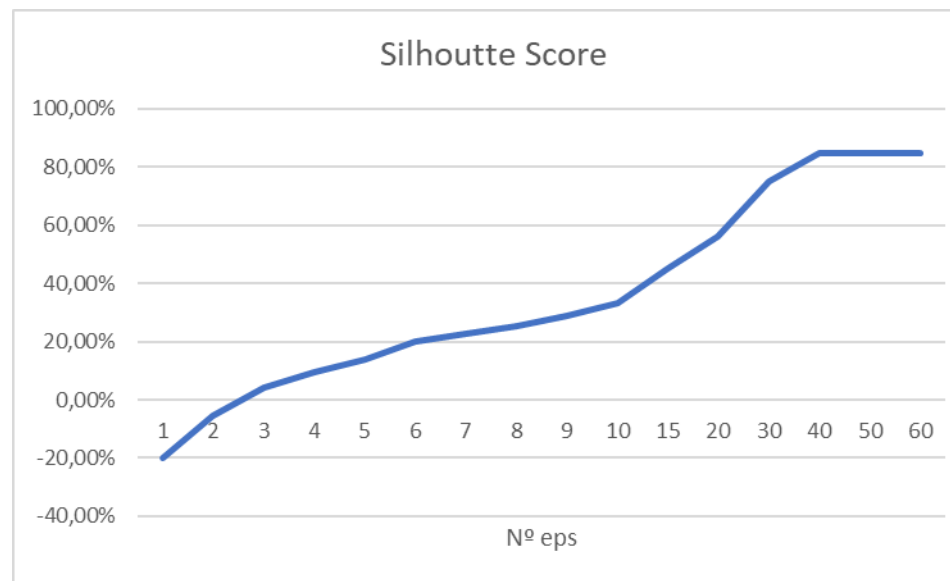
RESULTADOS | SCIKIT LEARN DBSCAN

46

## Scikit Learn DBSCAN

Selecionamos DBSCAN com  $\text{eps} = 40, 20$  e  $18$ , equivalente a  $k = 2, 3$  e  $7$ .

Infelizmente, apesar do alto silhouette score, as divisões não fizeram sentido do ponto de vista de negócio e decidimos descartar esta técnica.



# 5. Modelagem

RESULTADOS | SCIKIT LEARN DBSCAN

## Scikit Learn DBSCAN

Selecionamos DBSCAN com  $\text{eps} = 40, 20 \text{ e } 18$ , equivalente a  $k = 2, 3 \text{ e } 7$ .

Infelizmente, apesar do alto silhouette score, as divisões não fizeram sentido do ponto de vista de negócio e decidimos descartar esta técnica.

$K = 2$

- Silhouette score: 84%
- Grupo 0: 3.094 vendedores
- Grupo -1: 1 vendedor

$K = 3$

- Silhouette score: 56%
- Grupo 0: 3.037 vendedores
- Grupo -1: 51 vendedores
- Grupo 1: 7 vendedores

$K = 7$

- Silhouette score: 45%
- Grupo 0: 2.099 vendedores
- Grupo 1: 7 vendedores
- Grupo 2: 6 vendedores
- Grupo 3: 6 vendedores
- Grupo 4: 8 vendedores
- Grupo 5: 5 vendedores
- Grupo -1: 64 vendedores

A divisão por DBSCAN com 2 clusters não faz sentido do ponto de vista de negócios, pois qualquer decisão a ser tomada para afetar apenas 1 vendedor da base, mesmo que traga benefício, será muito custosa considerando horas gastas para análise, desenho de solução e implementação. Como já temos bons resultados com a divisão por K-means nativa do Spark, vamos descartar esta divisão.



# 5. Modelagem

RESULTADOS | SCIKIT LEARN DBSCAN

## Scikit Learn DBSCAN

Selecionamos DBSCAN com  $\text{eps} = 40, 20 \text{ e } 18$ , equivalente a  $k = 2, 3 \text{ e } 7$ .

Infelizmente, apesar do alto silhouette score, as divisões não fizeram sentido do ponto de vista de negócio e decidimos descartar esta técnica.

$K = 2$

- Silhouette score: 84%
- Grupo 0: 3.094 vendedores
- Grupo -1: 1 vendedor

$K = 3$

- Silhouette score: 56%
- Grupo 0: 3.037 vendedores
- Grupo -1: 51 vendedores
- Grupo 1: 7 vendedores

$K = 7$

- Silhouette score: 45%
- Grupo 0: 2.099 vendedores
- Grupo 1: 7 vendedores
- Grupo 2: 6 vendedores
- Grupo 3: 6 vendedores
- Grupo 4: 8 vendedores
- Grupo 5: 5 vendedores
- Grupo -1: 64 vendedores

A divisão por DBSCAN com 3 clusters não faz sentido do ponto de vista de negócios, pois qualquer decisão a ser tomada para afetar menos de 2% da base de vendedores, mesmo que traga benefício, será muito custosa considerando horas gastas para análise, desenho de solução e implementação. Como já temos bons resultados com a divisão por K-means nativa do Spark, vamos descartar esta divisão.





# 5. Modelagem

RESULTADOS | SCIKIT LEARN DBSCAN

49

## Scikit Learn DBSCAN

Selecionamos DBSCAN com  $\text{eps} = 40, 20 \text{ e } 18$ , equivalente a  $k = 2, 3 \text{ e } 7$ .

Infelizmente, apesar do alto silhouette score, as divisões não fizeram sentido do ponto de vista de negócio e decidimos descartar esta técnica.

$K = 2$

- Silhouette score: 84%
- Grupo 0: 3.094 vendedores
- Grupo -1: 1 vendedor

$K = 3$

- Silhouette score: 56%
- Grupo 0: 3.037 vendedores
- Grupo -1: 51 vendedores
- Grupo 1: 7 vendedores

$K = 7$

- Silhouette score: 45%
- Grupo 0: 2.099 vendedores
- Grupo 1: 7 vendedores
- Grupo 2: 6 vendedores
- Grupo 3: 6 vendedores
- Grupo 4: 8 vendedores
- Grupo 5: 5 vendedores
- Grupo -1: 64 vendedores

A divisão por DBSCAN com 7 clusters não faz sentido do ponto de vista de negócios, pois qualquer decisão a ser tomada para afetar apenas 3% da base de vendedores e com a maioria dos grupos com 8 ou menos vendedores, mesmo que traga benefício, será muito custosa considerando horas gastas para análise, desenho de solução e implementação. Como já temos bons resultados com a divisão por K-means nativa do Spark, vamos descartar esta divisão e, consequentemente, a técnica de DBSCAN como um todo.



# 5. Modelagem

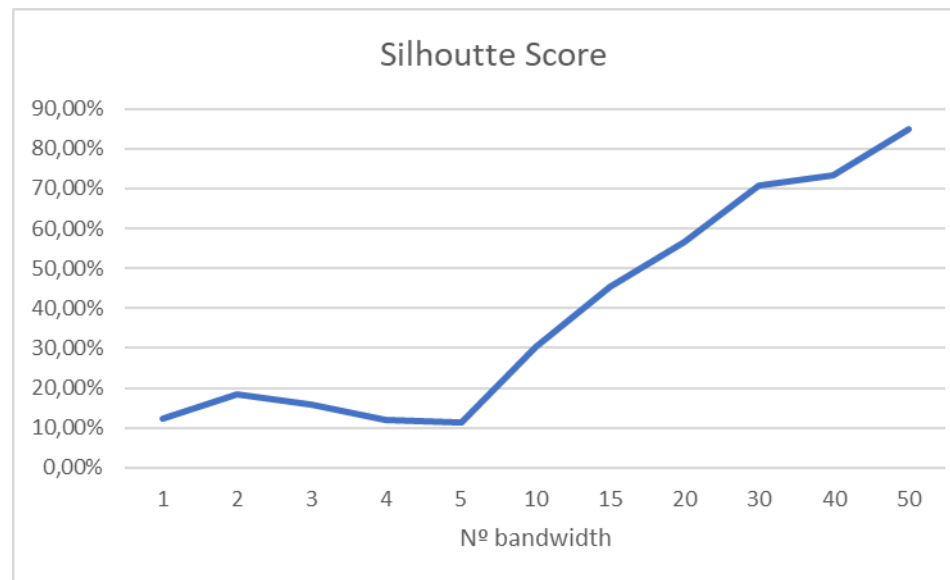
RESULTADOS | SCIKIT LEARN MEANSHIFT

50

## Scikit Learn MeanShift

Selecionamos MeanShift com bandwidth = 50, 40 e 35, equivalente a  $k = 2, 4$  e  $8$ .

Infelizmente, apesar do alto silhouette score, as divisões não fizeram sentido do ponto de vista de negócio e decidimos descartar esta técnica.



# 5. Modelagem

RESULTADOS | SCIKIT LEARN MEANSHIFT

51

## Scikit Learn MeanShift

Selecionamos MeanShift com bandwidth = 50, 40 e 35, equivalente a  $k = 2, 4$  e  $8$ .

Infelizmente, apesar do alto silhouette score, as divisões não fizeram sentido do ponto de vista de negócio e decidimos descartar esta técnica.

$K = 2$

- Silhouette score: 84%
- Grupo 0: 3.094 vendedores
- Grupo 1: 1 vendedor

$K = 4$

- Silhouette score: 73%
- Grupo 0: 3.090 vendedores
- Grupo 1: 2 vendedores
- Grupo 2: 2 vendedores
- Grupo 3: 1 vendedor

$K = 8$

- Silhouette score: 70%
- Grupo 0: 3.083 vendedores
- Grupo 1: 2 vendedores
- Grupo 2: 2 vendedores
- Grupo 3: 2 vendedores
- Grupo 4: 3 vendedores
- Grupo 5: 1 vendedor
- Grupo 6: 1 vendedor
- Grupo 7: 1 vendedor

A divisão por MeanShift com 2 clusters não faz sentido do ponto de vista de negócios, pois qualquer decisão a ser tomada para afetar apenas 1 vendedor da base, mesmo que traga benefício, será muito custosa considerando horas gastas para análise, desenho de solução e implementação. Como já temos bons resultados com a divisão por K-means nativa do Spark, vamos descartar esta divisão..



# 5. Modelagem

RESULTADOS | SCIKIT LEARN MEANSHIFT

52

## Scikit Learn MeanShift

Selecionamos MeanShift com bandwidth = 50, 40 e 35, equivalente a  $k = 2, 4$  e  $8$ .

Infelizmente, apesar do alto silhoutte score, as divisões não fizeram sentido do ponto de vista de negócio e decidimos descartar esta técnica.

$K = 2$

- Silhoutte score: 84%
- Grupo 0: 3.094 vendedores
- Grupo 1: 1 vendedor

$K = 4$

- Silhoutte score: 73%
- Grupo 0: 3.090 vendedores
- Grupo 1: 2 vendedores
- Grupo 2: 2 vendedores
- Grupo 3: 1 vendedor

$K = 8$

- Silhoutte score: 70%
- Grupo 0: 3.083 vendedores
- Grupo 1: 2 vendedores
- Grupo 2: 2 vendedores
- Grupo 3: 2 vendedores
- Grupo 4: 3 vendedores
- Grupo 5: 1 vendedor
- Grupo 6: 1 vendedor
- Grupo 7: 1 vendedor

A divisão por MeanShift com 4 clusters não faz sentido do ponto de vista de negócios, pois qualquer decisão a ser tomada para afetar apenas 0,2% da base de vendedores, mesmo que traga benefício, será muito custosa considerando horas gastas para análise, desenho de solução e implementação. Como já temos bons resultados com a divisão por K-means nativa do Spark, vamos descartar esta divisão.



# 5. Modelagem

RESULTADOS | SCIKIT LEARN MEANSHIFT

53

## Scikit Learn MeanShift

Selecionamos MeanShift com bandwidth = 50, 40 e 35, equivalente a  $k = 2, 4$  e  $8$ .

Infelizmente, apesar do alto silhouette score, as divisões não fizeram sentido do ponto de vista de negócio e decidimos descartar esta técnica.

$K = 2$

- Silhouette score: 84%
- Grupo 0: 3.094 vendedores
- Grupo 1: 1 vendedor

$K = 4$

- Silhouette score: 73%
- Grupo 0: 3.090 vendedores
- Grupo 1: 2 vendedores
- Grupo 2: 2 vendedores
- Grupo 3: 1 vendedor

$K = 8$

- Silhouette score: 70%
- Grupo 0: 3.083 vendedores
- Grupo 1: 2 vendedores
- Grupo 2: 2 vendedores
- Grupo 3: 2 vendedores
- Grupo 4: 3 vendedores
- Grupo 5: 1 vendedor
- Grupo 6: 1 vendedor
- Grupo 7: 1 vendedor

A divisão por MeanShift com 8 clusters não faz sentido do ponto de vista de negócios, pois qualquer decisão a ser tomada para afetar apenas 0,4% da base de vendedores e com a maioria dos grupos com 3 ou menos vendedores, mesmo que traga benefício, será muito custosa considerando horas gastas para análise, desenho de solução e implementação. Como já temos bons resultados com a divisão por K-means nativa do Spark, vamos descartar esta divisão e, consequentemente, a técnica de MeanShift como um todo.



# 5. Modelagem

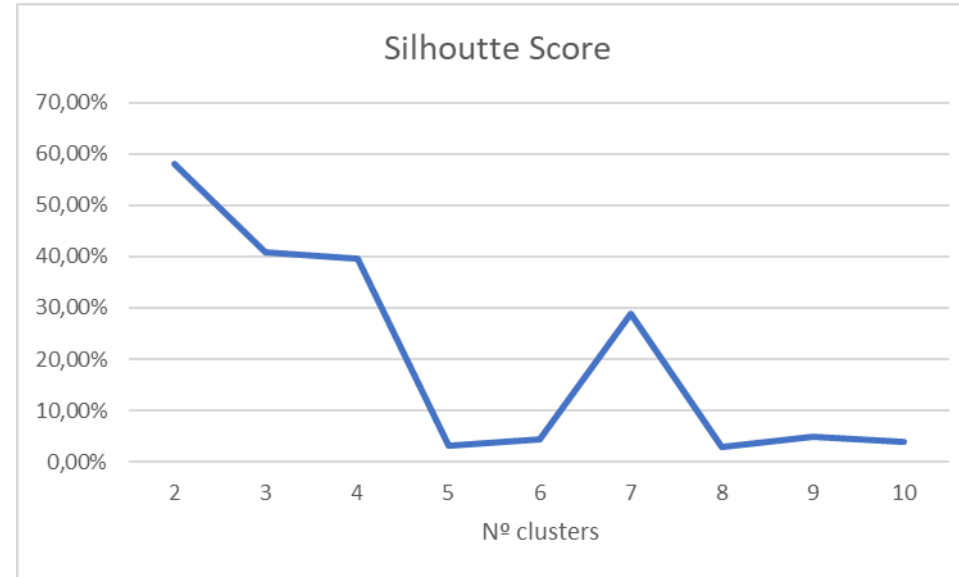
RESULTADOS | SCIKIT LEARN K-MEANS

54

## Scikit Learn K-means

Selecionamos os modelos com 2, 3 e 4 clusters, por possuírem melhor silhoutte score.

Destes, o agrupamento com 3 clusters se mostrou bastante promissor.



# 5. Modelagem

RESULTADOS | SCIKIT LEARN K-MEANS

55

## Scikit Learn K-means

Selecionamos os modelos com 2, 3 e 4 clusters, por possuírem melhor silhoutte score.

Destes, o agrupamento com 3 clusters se mostrou bastante promissor.

K = 2

- Silhoutte score: 58%
- Grupo 0: 3.034 vendedores
- Grupo 1: 61 vendedores

K = 3

- Silhoutte score: 40%
- Grupo 0: 2.814 vendedores
- Grupo 1: 263 vendedores
- Grupo 2: 18 vendedores

K = 4

- Silhoutte score: 39%
- Grupo 0: 2.806 vendedores
- Grupo 1: 18 vendedores
- Grupo 2: 261 vendedores
- Grupo 3: 10 vendedores

O resultado do K-means com 2 clusters do Scikit Learn ficou muito semelhante ao K-means nativo do Spark, onde o segundo grupo de vendedores premium ficou com 72 vendedores. Sendo assim, vamos descartar a análise de perfil desta técnica.



# 5. Modelagem

RESULTADOS | SCIKIT LEARN K-MEANS

56

## Scikit Learn K-means

Selecionamos os modelos com 2, 3 e 4 clusters, por possuírem melhor silhoutte score.

Destes, o agrupamento com 3 clusters se mostrou bastante promissor.

K = 2

- Silhoutte score: 58%
- Grupo 0: 3.034 vendedores
- Grupo 1: 61 vendedores

K = 3

- Silhoutte score: 40%
- Grupo 0: 2.814 vendedores
- Grupo 1: 263 vendedores
- Grupo 2: 18 vendedores

K = 4

- Silhoutte score: 39%
- Grupo 0: 2.806 vendedores
- Grupo 1: 18 vendedores
- Grupo 2: 261 vendedores
- Grupo 3: 10 vendedores

O resultado do K-means com 3 clusters do Scikit Learn ficou aparentemente melhor do que o do K-means nativo do Spark, onde o 1º grupo ficou com 2999 vendedores, o segundo grupo com 33 e o terceiro com 72. Vale a pena nos debruçarmos sobre este agrupamento e gerar os perfis na análise final.





# 5. Modelagem

RESULTADOS | SCIKIT LEARN K-MEANS

57

## Scikit Learn K-means

Selecionamos os modelos com 2, 3 e 4 clusters, por possuírem melhor silhoutte score.

Destes, o agrupamento com 3 clusters se mostrou bastante promissor.

K = 2

- Silhoutte score: 58%
- Grupo 0: 3.034 vendedores
- Grupo 1: 61 vendedores

K = 3

- Silhoutte score: 40%
- Grupo 0: 2.814 vendedores
- Grupo 1: 263 vendedores
- Grupo 2: 18 vendedores

K = 4

- Silhoutte score: 39%
- Grupo 0: 2.806 vendedores
- Grupo 1: 18 vendedores
- Grupo 2: 261 vendedores
- Grupo 3: 10 vendedores

O resultado do K-means com 4 clusters do Scikit Learn não parece agregar muito em comparação com o agrupamento com 3 clusters visto acima. O 3º grupo continua com 18 vendedores e o 4º grupo foi formado com apenas 10 vendedores, retirados do 1º e do 2º grupos. Concluimos que não faz sentido do ponto de vista de negócios, pois qualquer decisão a ser tomada para afetar menos de 0,3% dos vendedores da base, mesmo que traga benefício, será muito custosa considerando horas gastas para análise, desenho de solução e implementação. Portanto, vamos excluir esse agrupamento da análise final.



# 5. Modelagem

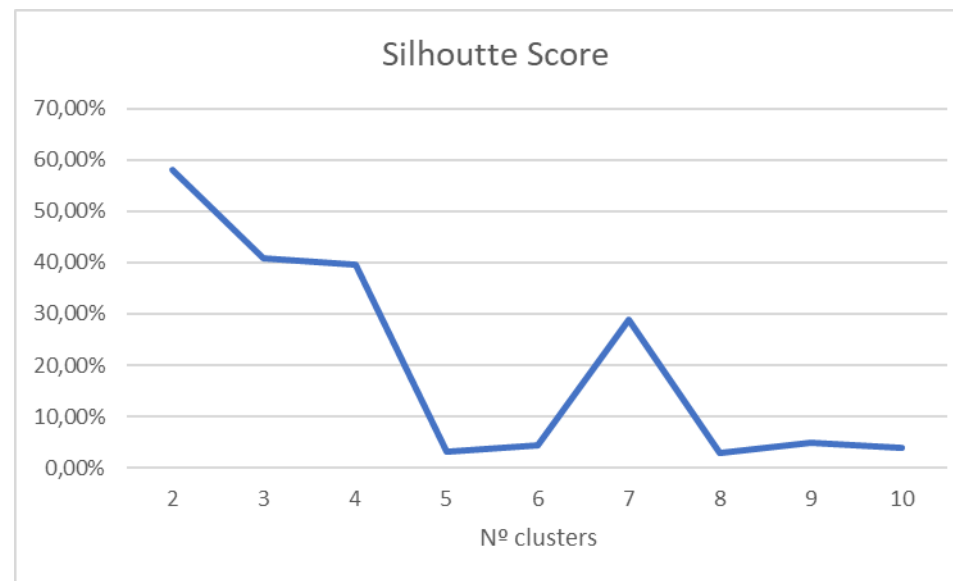
RESULTADOS | SCIKIT LEARN K-MEANS

58

## Scikit Learn K-means

Ainda nesta técnica, decidimos testar o modelo com 6 clusters (mesmo com silhouette score muito baixo), para poder compará-lo com o resultado do K-means nativo do Spark, e o modelo com 7 clusters, que tem silhouette score razoável.

Curiosamente, o agrupamento com 6 clusters, apesar de ter métrica estatística baixa, parece interessante da ótica de negócios.



# 5. Modelagem

RESULTADOS | SCIKIT LEARN K-MEANS

59

## Scikit Learn K-means

Ainda nesta técnica, decidimos testar o modelo com 6 clusters (mesmo com silhouette score muito baixo), para poder compará-lo com o resultado do K-means nativo do Spark, e o modelo com 7 clusters, que tem silhouette score razoável.

Curiosamente, o agrupamento com 6 clusters, apesar de ter métrica estatística baixa, parece interessante da ótica de negócios.

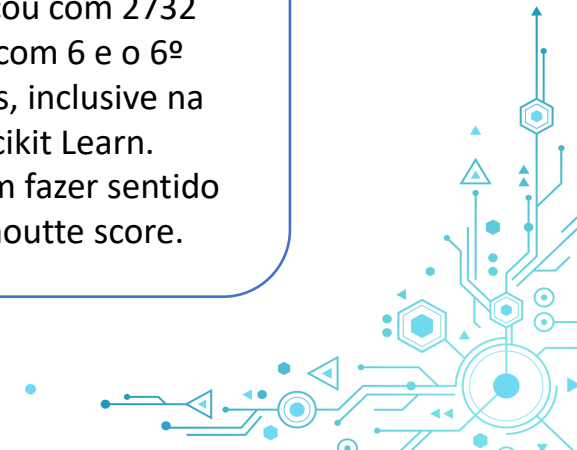
K = 6

- Silhoutte score: 4%
- Grupo 0: 194 vendedores
- Grupo 1: 542 vendedores
- Grupo 2: 156 vendedores
- Grupo 3: 14 vendedores
- Grupo 4: 2.170 vendedores
- Grupo 5: 19 vendedores

K = 7

- Silhoutte score: 28%
- Grupo 0: 34 vendedores
- Grupo 1: 270 vendedores
- Grupo 2: 1 vendedor
- Grupo 3: 35 vendedores
- Grupo 4: 19 vendedores
- Grupo 5: 55 vendedores
- Grupo 6: 2.681 vendedores

O resultado do K-means com 6 clusters do Scikit Learn ficou muito superior em comparação ao do K-means nativo do Spark, onde o 1º grupo ficou com 2732 vendedores, o 2º grupo com 2, o 3º com 23, o 4º com 331, o 5º com 6 e o 6º com 1. Aqui, temos grupos mais populosos e melhor distribuídos, inclusive na comparação com o agrupamento de 4 clusters do K-means do Scikit Learn. Faremos o perfil destes 6 grupos na análise final porque parecem fazer sentido do ponto de vista de negócios, à despeito do baixo índice no silhouette score.



# 5. Modelagem

RESULTADOS | SCIKIT LEARN K-MEANS

60

## Scikit Learn K-means

Ainda nesta técnica, decidimos testar o modelo com 6 clusters (mesmo com silhouette score muito baixo), para poder compará-lo com o resultado do K-means nativo do Spark, e o modelo com 7 clusters, que tem silhouette score razoável.

Curiosamente, o agrupamento com 6 clusters, apesar de ter métrica estatística baixa, parece interessante da ótica de negócios.

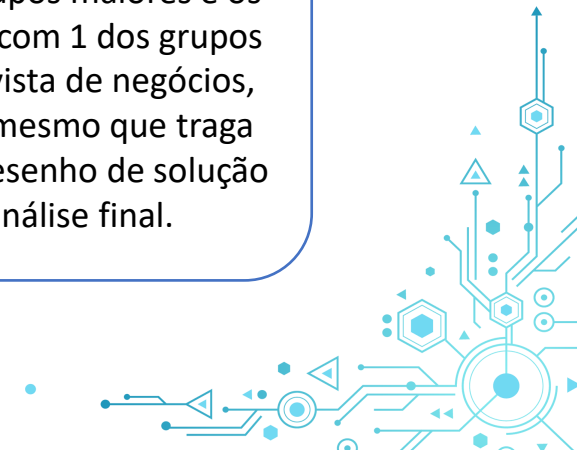
K = 6

- Silhoutte score: 4%
- Grupo 0: 194 vendedores
- Grupo 1: 542 vendedores
- Grupo 2: 156 vendedores
- Grupo 3: 14 vendedores
- Grupo 4: 2.170 vendedores
- Grupo 5: 19 vendedores

K = 7

- Silhoutte score: 28%
- Grupo 0: 34 vendedores
- Grupo 1: 270 vendedores
- Grupo 2: 1 vendedor
- Grupo 3: 35 vendedores
- Grupo 4: 19 vendedores
- Grupo 5: 55 vendedores
- Grupo 6: 2.681 vendedores

O resultado do K-means com 7 clusters do Scikit Learn não parece agregar muito em comparação com o agrupamento com 6 clusters visto acima. Há os 2 grupos maiores e os outros 5 grupos correspondem a menos de 5% da base de vendedores, com 1 dos grupos com apenas 1 indivíduo. Concluimos que não faz sentido do ponto de vista de negócios, pois qualquer decisão a ser tomada para afetar grupos tão pequenos, mesmo que traga benefício, será muito custosa considerando horas gastas para análise, desenho de solução e implementação. Portanto, vamos excluir esse agrupamento da análise final.



## 5. Modelagem

RESULTADOS | SELEÇÃO DE MODELOS

61

Dentre todas as técnicas e números de grupos utilizados, escolhemos as que aparentam agregar mais valor ao negócio.

### KMeans MLSpark

- $K = 3$

### Kmeans SKlearn

- $K = 3$
- $K = 6$

### Bisecting Kmeans MLSpark

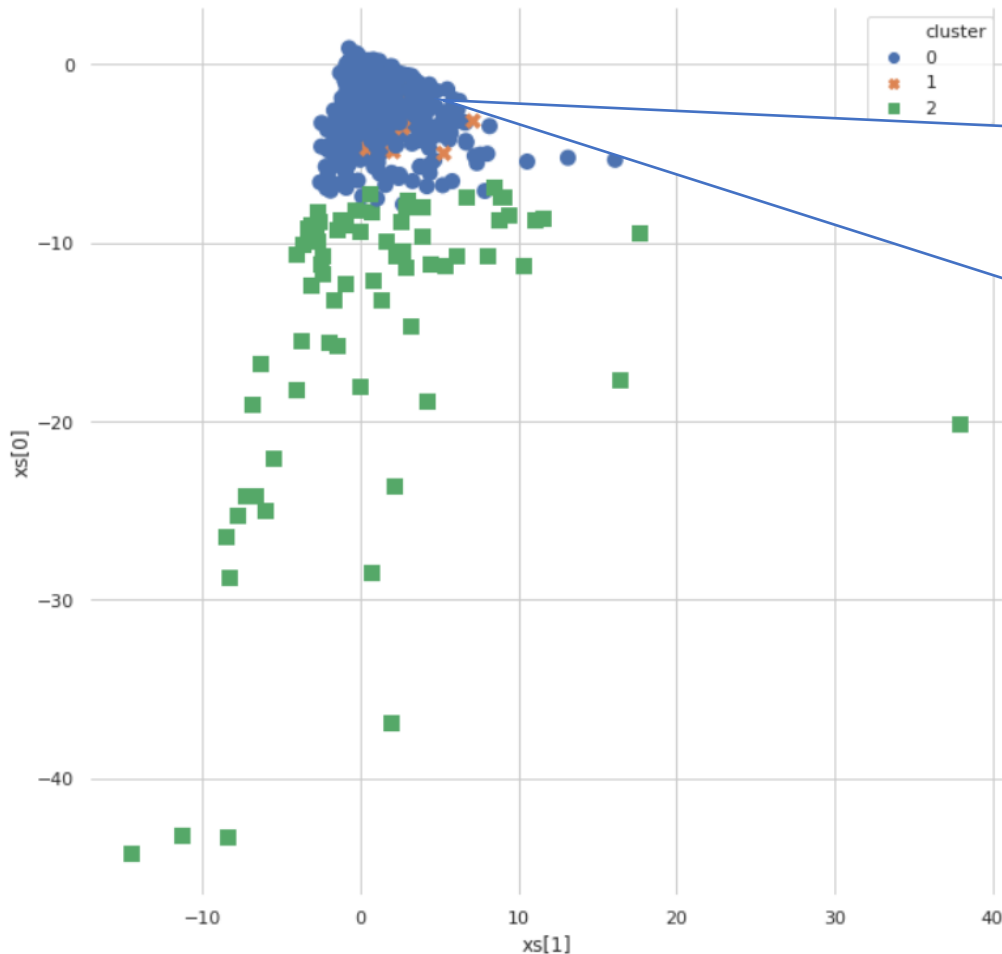
- $K = 3$



# 5. Modelagem

KMeans MLSpark K = 3 | PERFIL DOS GRUPOS

62



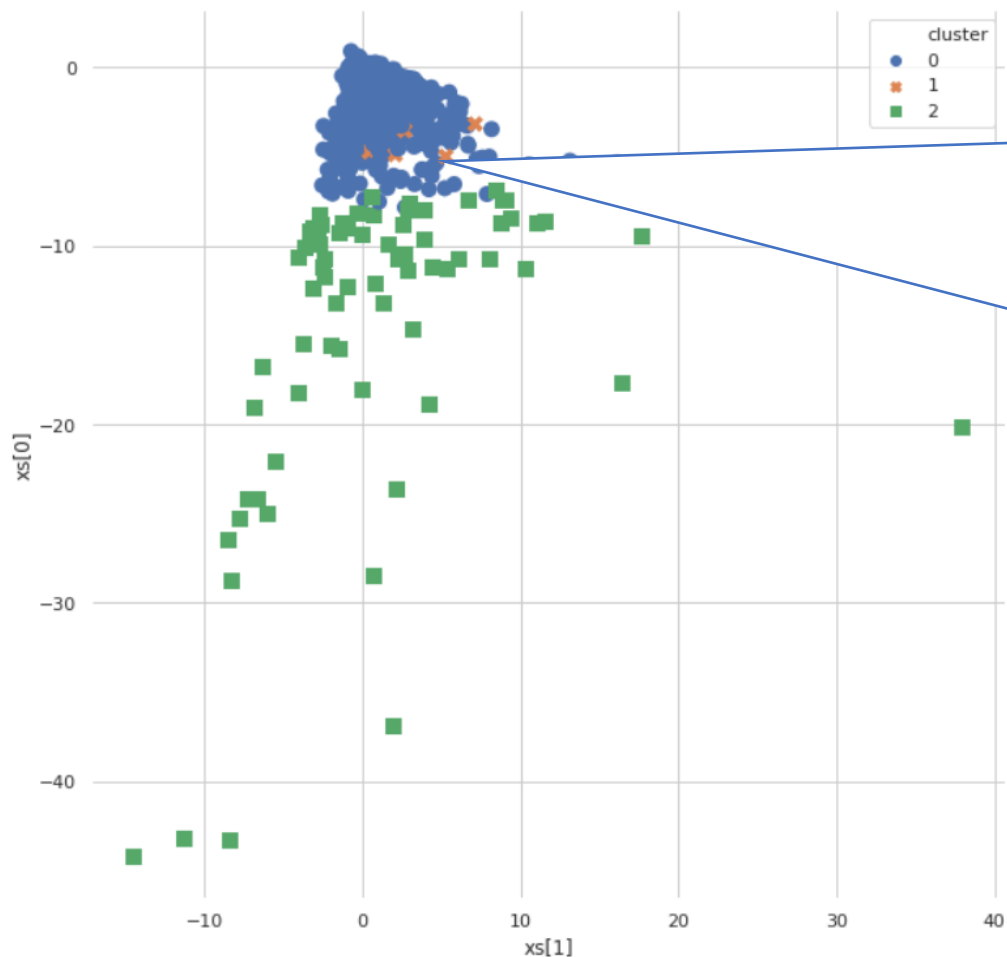
## Prediction=0:

- Qtd. Vendedores: 2.990
- Média de pedidos: 20
- Média de compradores diferentes: 20
- Média de produtos vendidos: 23
- Média de faturamento: pouco menos de R\$3.500,00
- Ticket médio: R\$171,44
- Preço médio: R\$151,03
- Recência: quase 5 meses desde a última venda
- Frequência: 3 vendas a cada 2 meses
- Atraso na entrega: alguns vendedores costumam atrasar
- Esse é o perfil da maioria dos vendedores da Olist: faz poucas vendas e só é interessante para a Olist no volume. Seus produtos são de valor intermediário, se comparados aos outros grupos.

# 5. Modelagem

KMeans MLSpark K = 3 | PERFIL DOS GRUPOS

63



## Prediction=1:

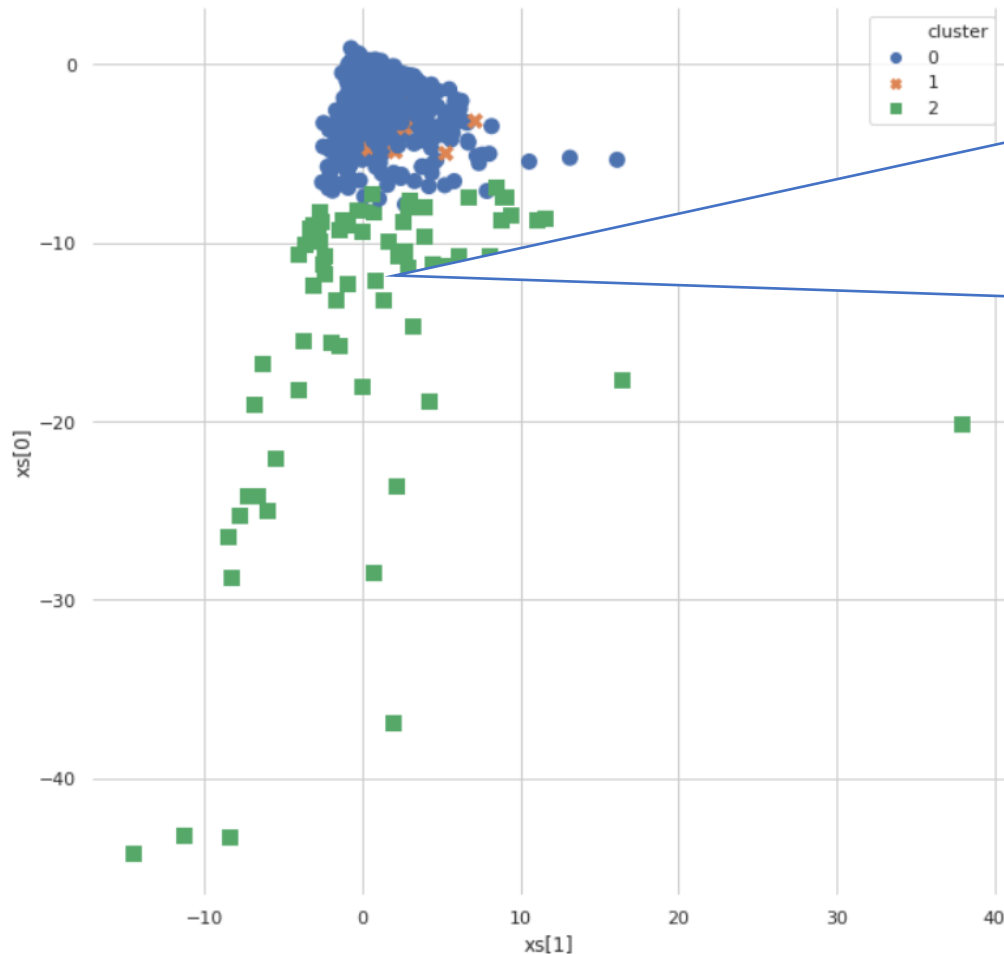
- Qtd. Vendedores: 33
- Média de pedidos: 26
- Média de compradores diferentes: 25
- Média de produtos vendidos: 31
- Média de faturamento: pouco mais de R\$6.000,00
- Ticket médio: R\$228,41
- Preço médio: R\$188,64
- Recência: pouco mais de 1 mês desde a última venda
- Frequência: quase 5 vendas a cada 2 meses
- Atraso na entrega: nenhum vendedor desse grupo costuma atrasar
- Esse é o perfil de vendedores aspirantes a premium da Olist: faz mais vendas do que a maioria dos vendedores, mas ainda num patamar bem abaixo do que os vendedores premium. Para compensar, vende produtos mais caros.



# 5. Modelagem

KMeans MLSpark K = 3 | PERFIL DOS GRUPOS

64



## Prediction=2:

- Qtd. Vendedores: 72
- Média de pedidos: 531
- Média de compradores diferentes: 525
- Média de produtos vendidos: 599
- Média de faturamento: pouco mais de R\$73.300
- Ticket médio: R\$138,05
- Preço médio: R\$122,37
- Recência: pouco mais de 16 dias desde a última venda
- Frequência: quase 30 vendas por mês
- Atraso na entrega: nenhum vendedor desse grupo costuma atrasar
- Esse é o perfil de vendedores premium da Olist: faz muitas vendas e garante bom faturamento, mesmo vendendo produtos mais baratos que os vendedores dos outros grupos.

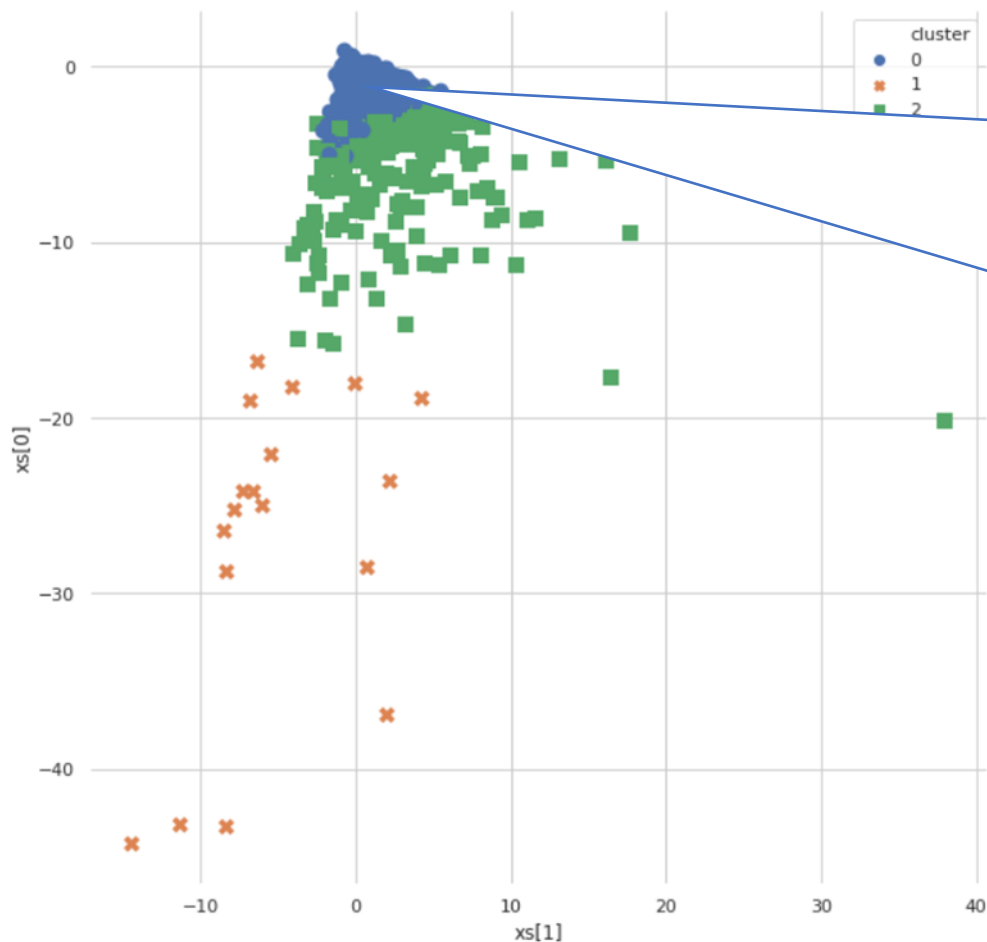




# 5. Modelagem

KMeans SKLearn K = 3 | PERFIL DOS GRUPOS

65



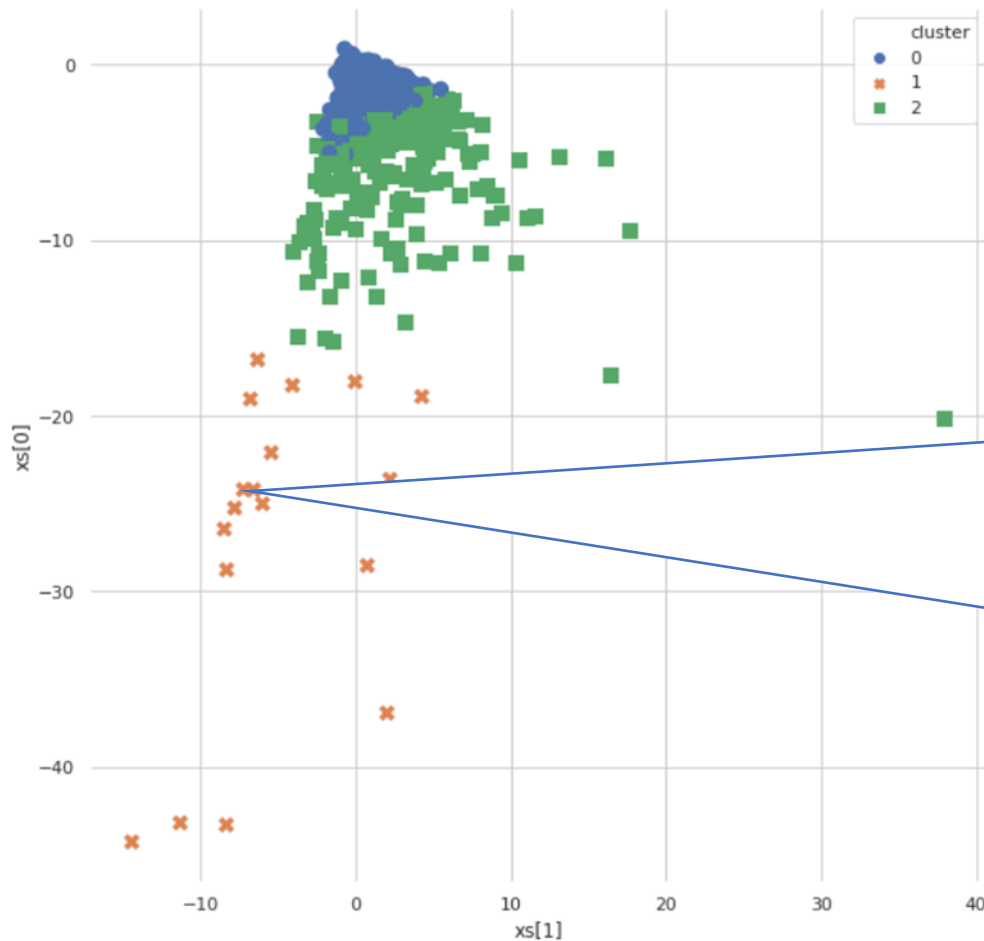
## Prediction=0:

- Qtd. Vendedores: 2.814
- Média de pedidos: 14
- Média de compradores diferentes: 14
- Média de produtos vendidos: 16
- Média de faturamento: quase R\$2.700
- Ticket médio: R\$182,52
- Preço médio: R\$159,18
- Recência: 5 meses desde a última venda
- Frequência: 4 vendas a cada 3 meses
- Atraso na entrega: alguns vendedores costumam atrasar
- Esse é o perfil da maioria dos vendedores da Olist: faz poucas vendas e só é interessante para a Olist no volume, mesmo que seus produtos sejam de maior valor agregado em relação aos outros grupos.

# 5. Modelagem

KMeans SKLearn K = 3 | PERFIL DOS GRUPOS

66



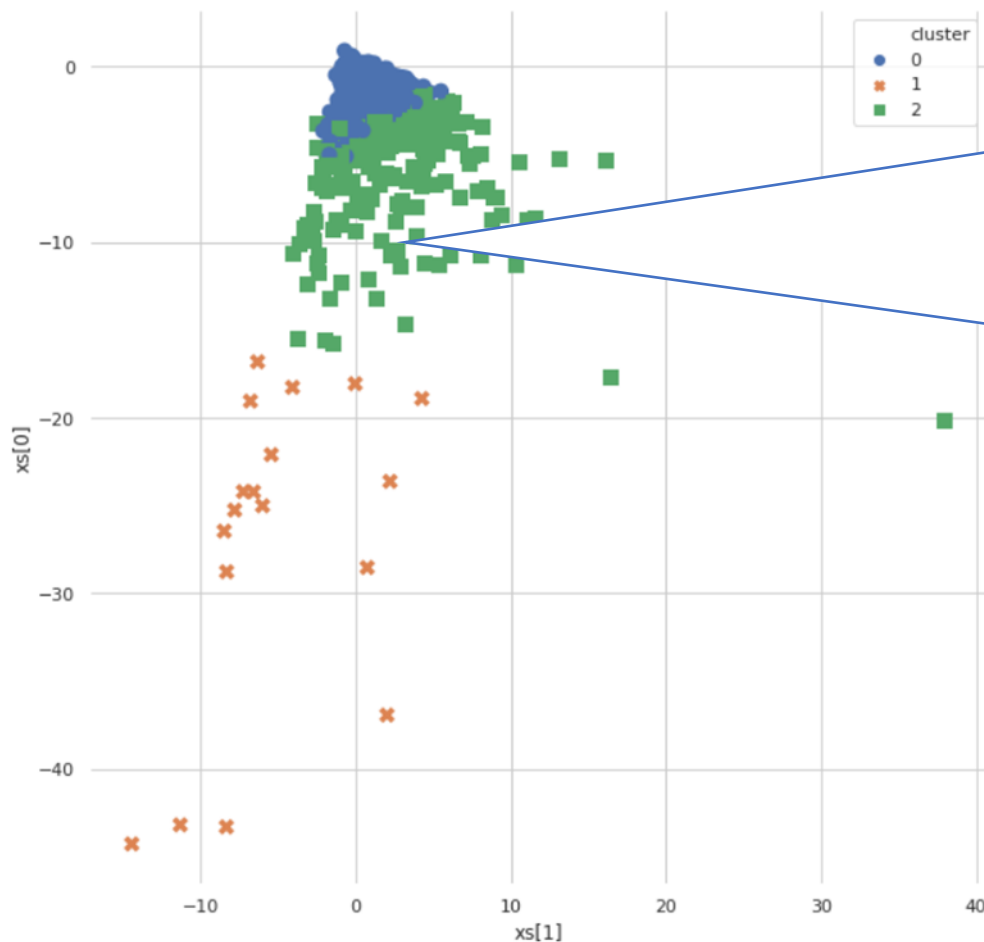
## Prediction=1:

- Qtd. Vendedores: 18
- Média de pedidos: 1.118
- Média de compradores diferentes: 1.104
- Média de produtos vendidos: 1.284
- Média de faturamento: pouco mais de R\$143.000
- Ticket médio: R\$127,98
- Preço médio: R\$111,39
- Recência: quase 10 dias desde a última venda
- Frequência: 61 vendas por mês
- Atraso na entrega: nenhum vendedor desse grupo costuma atrasar
- Esse é o perfil de vendedores premium da Olist: faz muitas vendas e garante bom faturamento, mesmo vendendo produtos de menor valor agregado em relação aos outros grupos.

# 5. Modelagem

KMeans SKLearn K = 3 | PERFIL DOS GRUPOS

67



## Prediction=2:

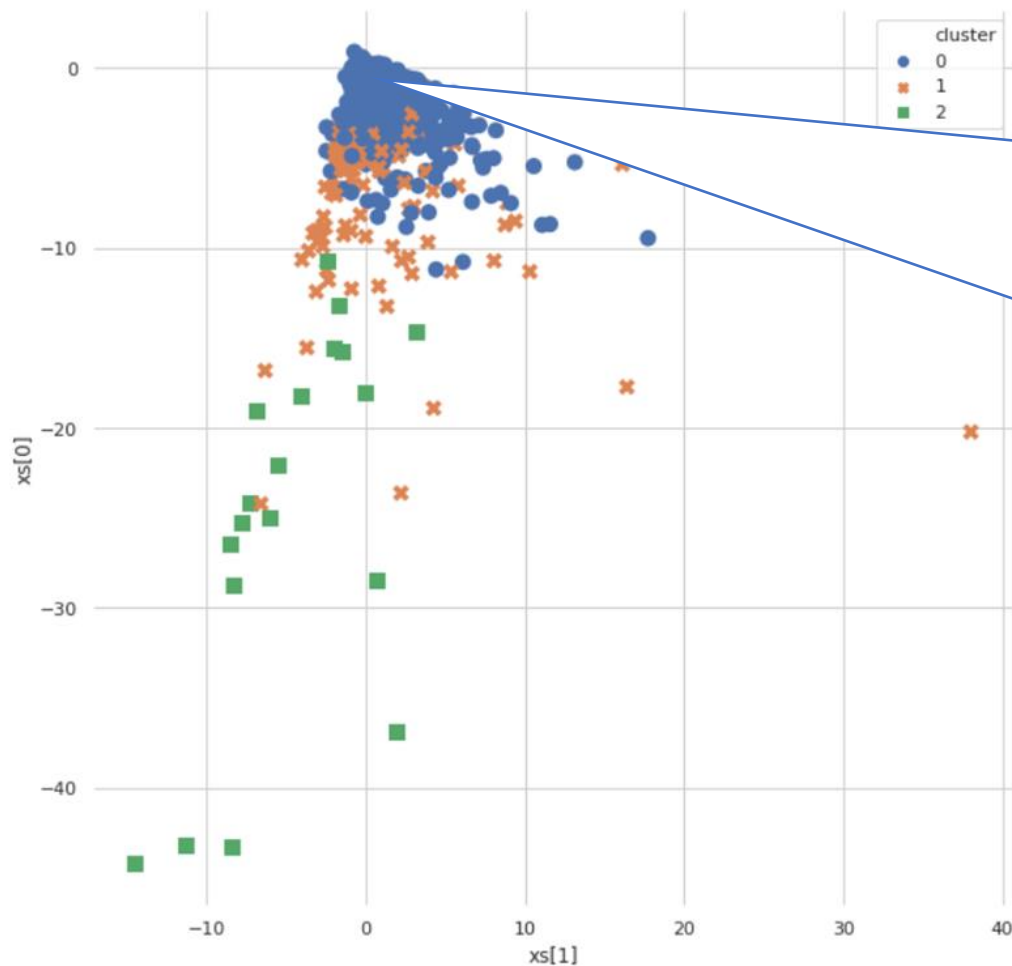
- Qtd. Vendedores: 263
- Média de pedidos: 146
- Média de compradores diferentes: 144
- Média de produtos vendidos: 162
- Média de faturamento: pouco mais de R\$22.000
- Ticket médio: R\$150,36
- Preço médio: R\$135,41
- Recência: quase 2 meses desde a última venda
- Frequência: 9 vendas por mês
- Atraso na entrega: pouquíssimos vendedores desse grupo costuma atrasar
- Esse é o perfil de vendedores aspirantes a premium da Olist: faz mais vendas do que a maioria dos vendedores, mas ainda num patamar bem abaixo do que os vendedores premium. Seus produtos são de valor intermediário em relação aos outros 2 grupos.



# 5. Modelagem

Bisecting KMeans MLSpark K = 3 | PERFIL DOS GRUPOS

68



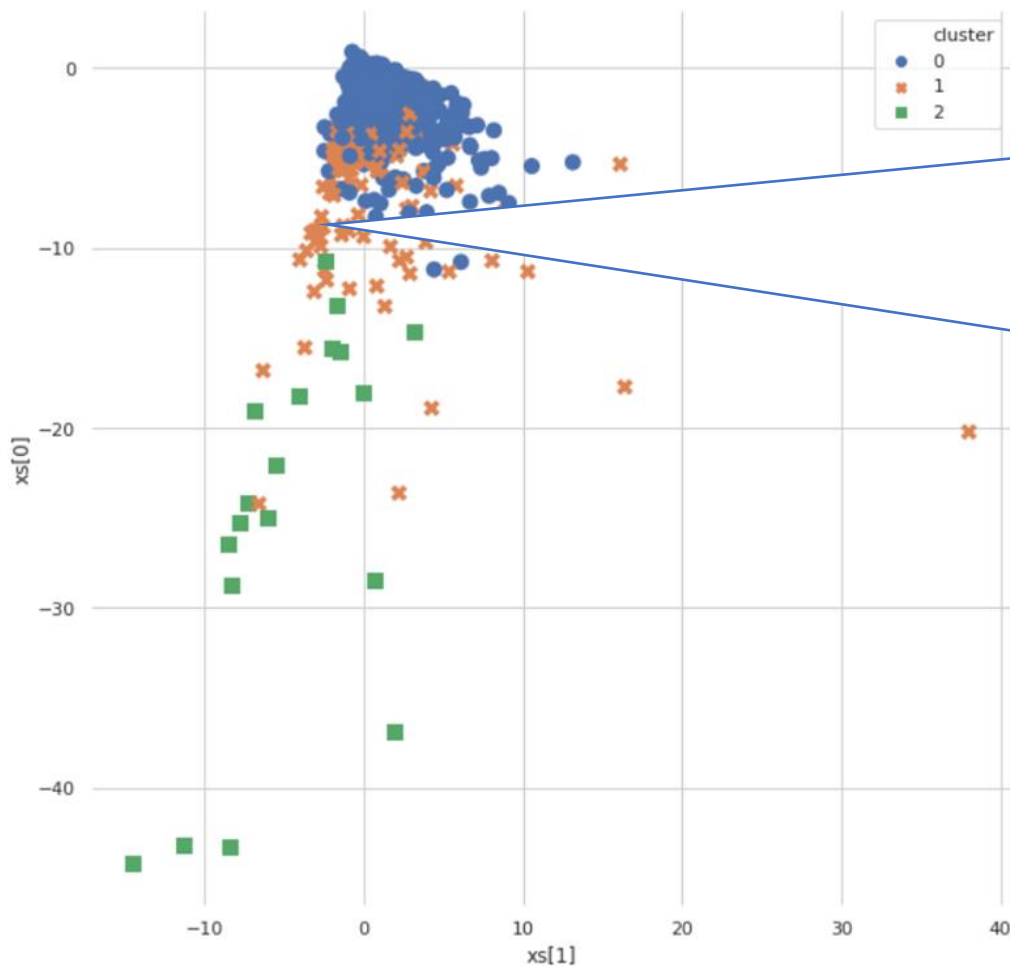
## Prediction=0:

- Qtd. Vendedores: 2.957
- Média de pedidos: 18
- Média de compradores diferentes: 18
- Média de produtos vendidos: 20
- Média de faturamento: pouco menos de R\$2.700
- Ticket médio: R\$145,36
- Preço médio: R\$127,90
- Recência: quase 5 meses desde a última venda
- Frequência: 3 vendas a cada 2 meses
- Atraso na entrega: alguns vendedores costumam atrasar
- Esse é o perfil da maioria dos vendedores da Olist: faz poucas vendas, seus produtos tem baixo valor agregado em relação aos outros grupos e só é interessante para a Olist no volume.

# 5. Modelagem

Bisecting KMeans MLSpark K = 3 | PERFIL DOS GRUPOS

69



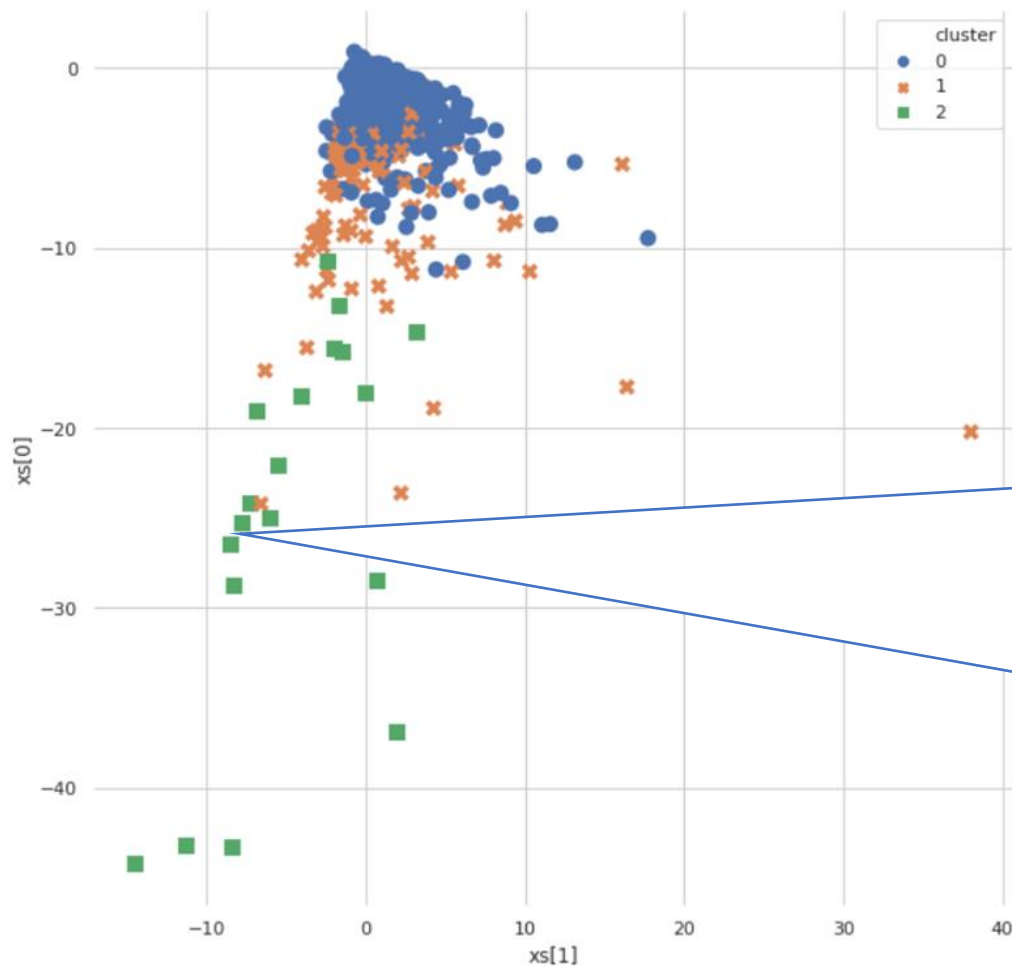
## Prediction=1:

- Qtd. Vendedores: 119
- Média de pedidos: 226
- Média de compradores diferentes: 223
- Média de produtos vendidos: 252
- Média de faturamento: pouco mais de R\$40.200
- Ticket médio: R\$387,74
- Preço médio: R\$177,91
- Recência: pouco mais de 1 mês desde a última venda
- Frequência: quase 14 vendas por mês
- Atraso na entrega: poucos vendedores desse grupo costumam atrasar
- Esse é o perfil dos vendedores intermediários da Olist: tem número considerável de vendas mais qualificadas, pois vende produtos de maior valor agregado (preço médio elevado) e mais produtos por venda (maior ticket médio).

# 5. Modelagem

Bisecting KMeans MLSpark K = 3 | PERFIL DOS GRUPOS

70



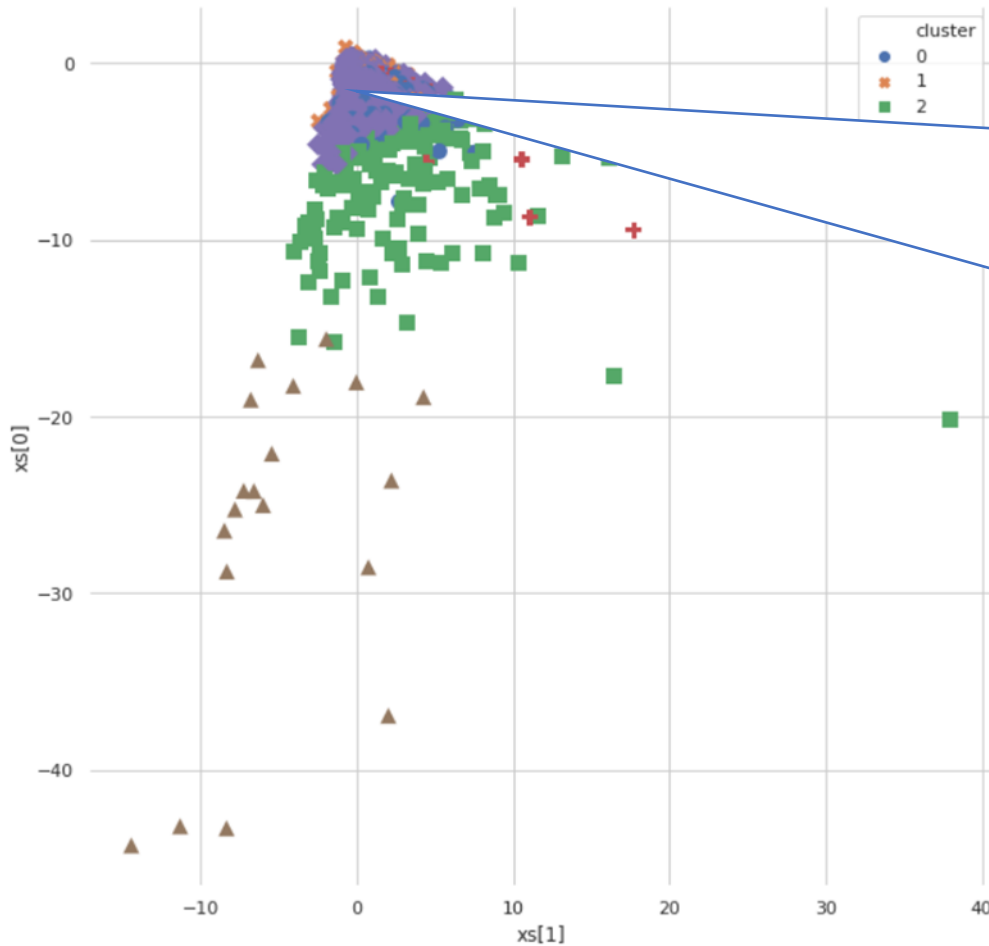
## Prediction=2:

- Qtd. Vendedores: 19
- Média de pedidos: 988
- Média de compradores diferentes: 976
- Média de produtos vendidos: 1134
- Média de faturamento: pouco mais de R\$170.200
- Ticket médio: R\$172,17
- Preço médio: R\$150,07
- Recência: 15 dias desde a última venda
- Frequência: 52 vendas por mês
- Atraso na entrega: nenhum vendedor desse grupo costuma atrasar
- Esse é o perfil dos vendedores de elite da Olist: fazem muitas vendas e garantem bom faturamento. Seus produtos são de valor intermediário em relação aos outros 2 grupos.

# 5. Modelagem

KMeans SKLearn K = 6 | PERFIL DOS GRUPOS

71



**Prediction=0:**

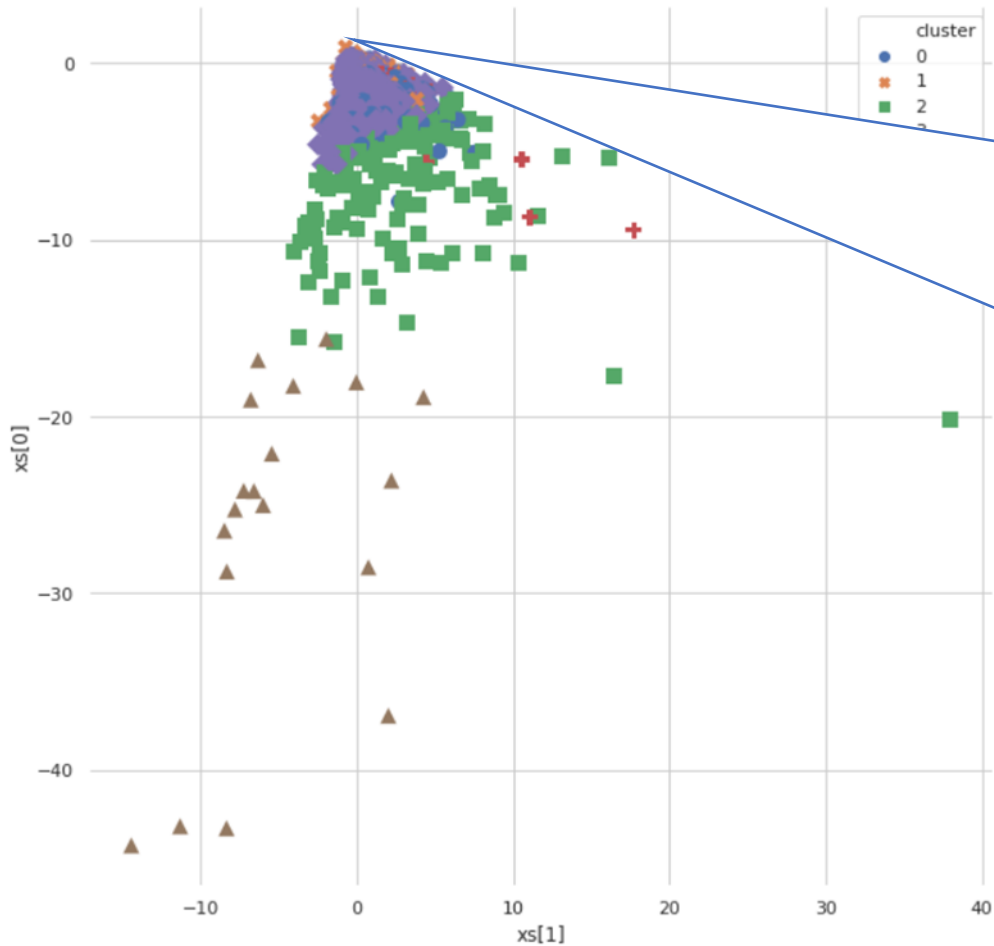
- Qtd. Vendedores: 194
- Média de pedidos: 26
- Média de compradores diferentes: 26
- Média de produtos vendidos: 29
- Média de faturamento: pouco mais de R\$7.100
- Ticket médio: R\$271,18
- Preço médio: R\$246,18
- Recência: mais de 3 meses desde a última venda
- Frequência: 2 vendas por mês
- Atraso na entrega: alguns vendedores costumam atrasar
- Esse é o perfil de vendedores de produtos mais caros



# 5. Modelagem

KMeans SKLearn K = 6 | PERFIL DOS GRUPOS

72



## Prediction=1:

- Qtd. Vendedores: 542
- Média de pedidos: 5
- Média de compradores diferentes: 5
- Média de produtos vendidos: 6
- Média de faturamento: pouco mais de R\$1.100
- Ticket médio: R\$212,36
- Preço médio: R\$175,57
- Recência: mais de 11 meses desde a última venda
- Frequência: 1 venda a cada 3 meses
- Atraso na entrega: alguns vendedores desse grupo costumam atrasar
- Esse é o perfil dos vendedores que não deram certo e desistiram da plataforma, com média de review score de 2,5.

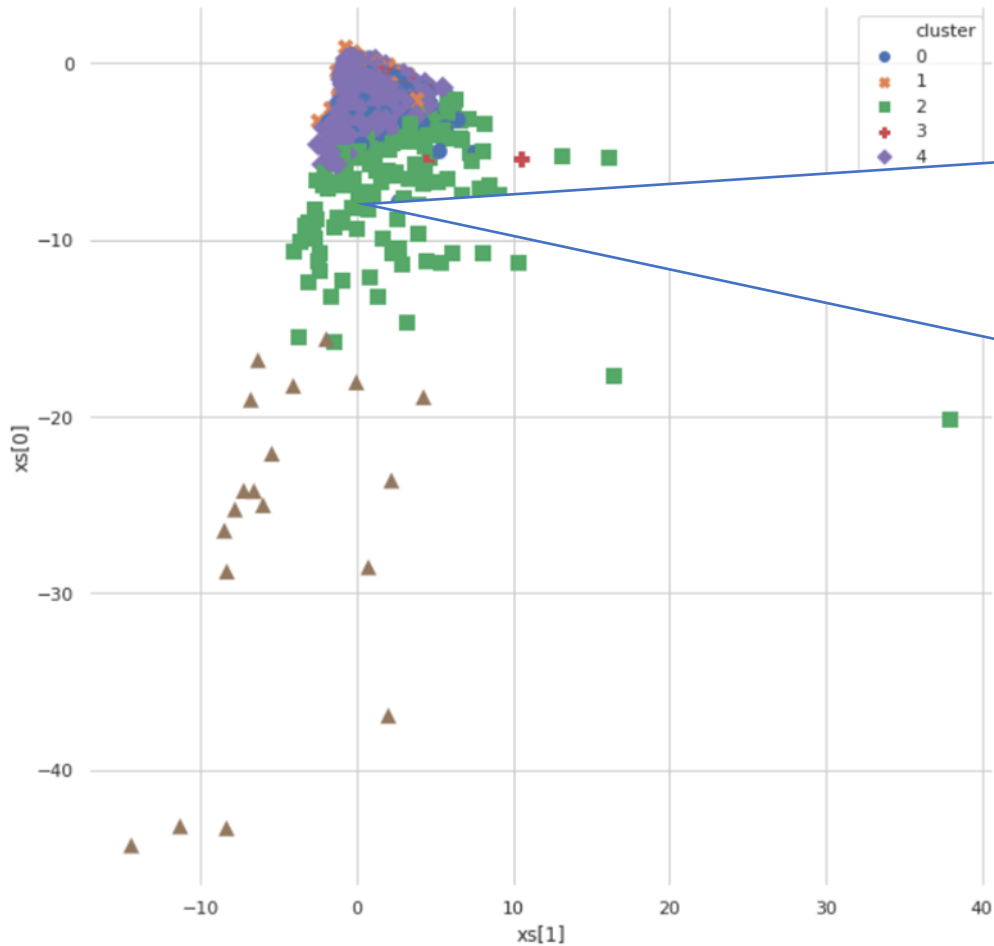




# 5. Modelagem

KMeans SKLearn K = 6 | PERFIL DOS GRUPOS

73



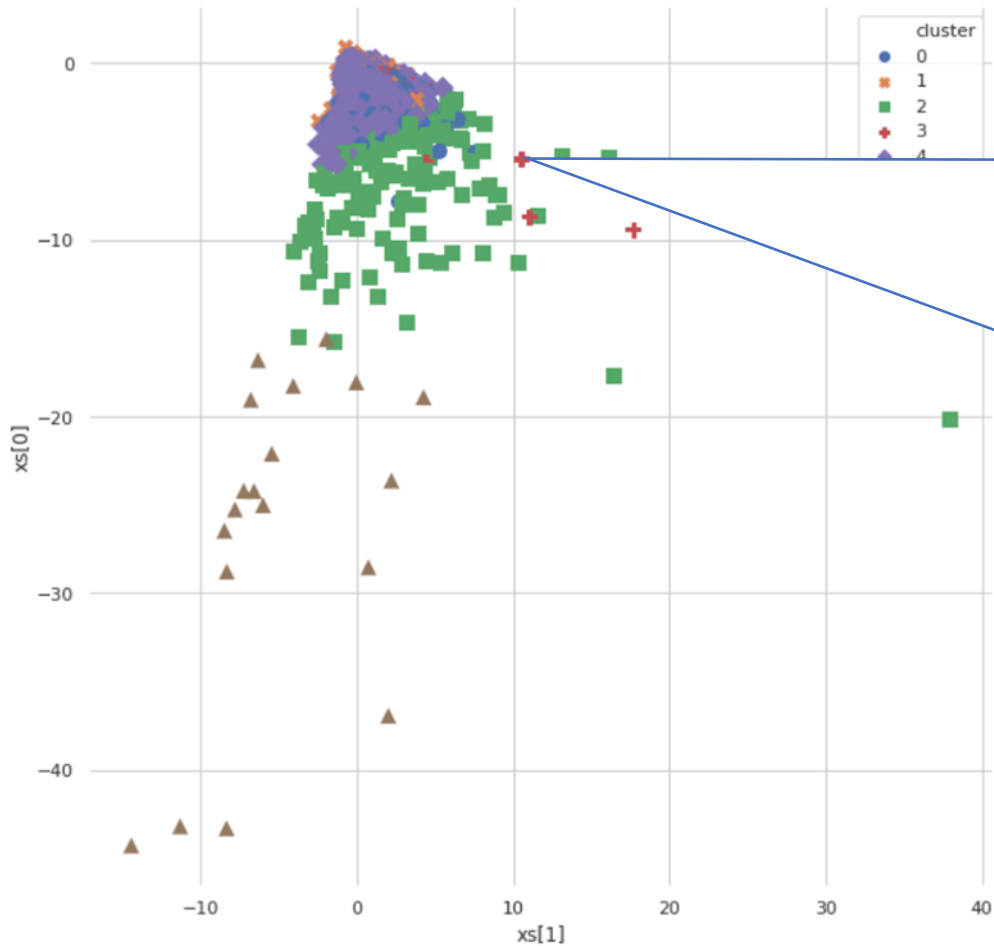
## Prediction=2:

- Qtd. Vendedores: 156
- Média de pedidos: 188
- Média de compradores diferentes: 186
- Média de produtos vendidos: 208
- Média de faturamento: em torno de R\$26.600
- Ticket médio: R\$141,71
- Preço médio: R\$127,78
- Recência: quase 1 mês e meio desde a última venda
- Frequência: 11 vendas por mês
- Atraso na entrega: pouquíssimos vendedores desse grupo costumam atrasar
- Esse é o perfil de vendedores aspirantes a premium da Olist: faz mais vendas do que a maioria dos vendedores, mas ainda num patamar bem abaixo do que os vendedores premium. Seus produtos são de valor intermediário em relação aos outros grupos.

# 5. Modelagem

KMeans SKLearn K = 6 | PERFIL DOS GRUPOS

74



## Prediction=3:

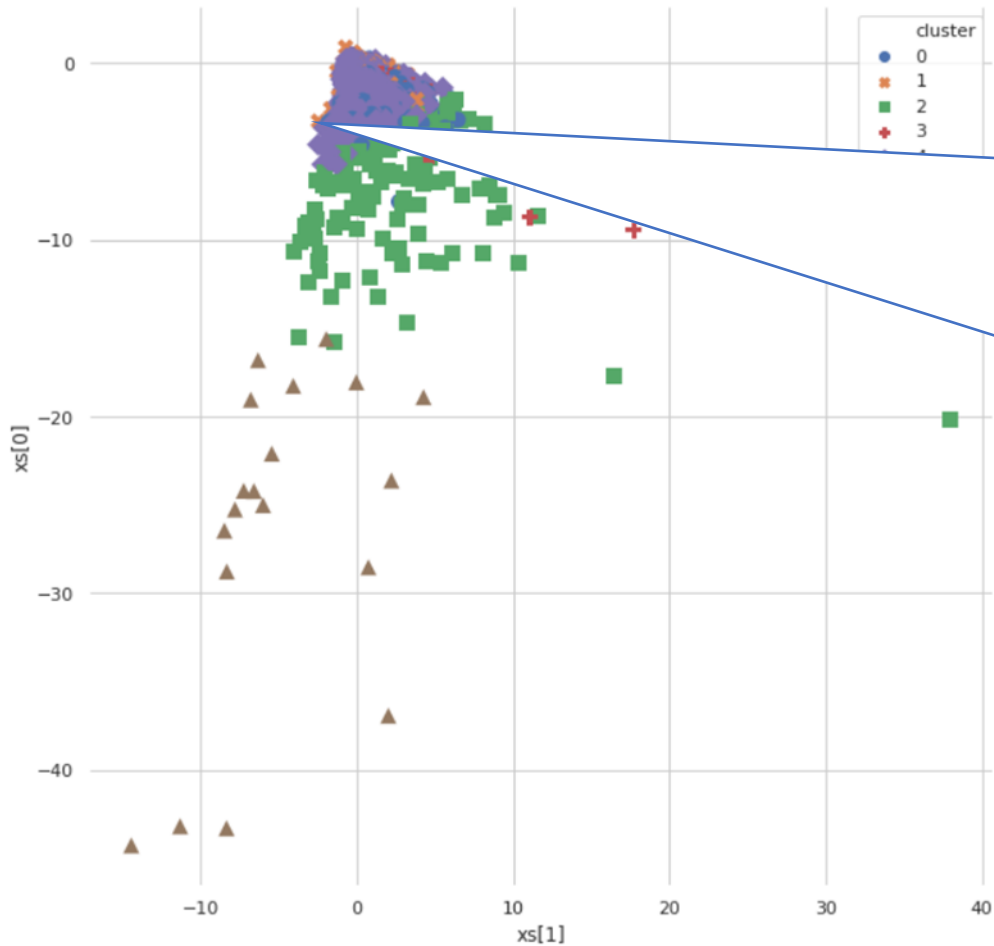
- Qtd. Vendedores: 14
- Média de pedidos: 45
- Média de compradores diferentes: 45
- Média de produtos vendidos: 50
- Média de faturamento: em torno de R\$4.500
- Ticket médio: R\$100,04
- Preço médio: R\$90,26
- Recência: mais de 1 mês desde a última venda
- Frequência: quase 5 vendas por mês
- Atraso na entrega: nenhum vendedor desse grupo costuma atrasar
- Esse é o perfil de um subgrupo de vendedores do baixo clero: número de vendas baixo, trabalha com produtos mais baratos, mas compensa um pouco no volume.



# 5. Modelagem

KMeans SKLearn K = 6 | PERFIL DOS GRUPOS

75



## Prediction=4:

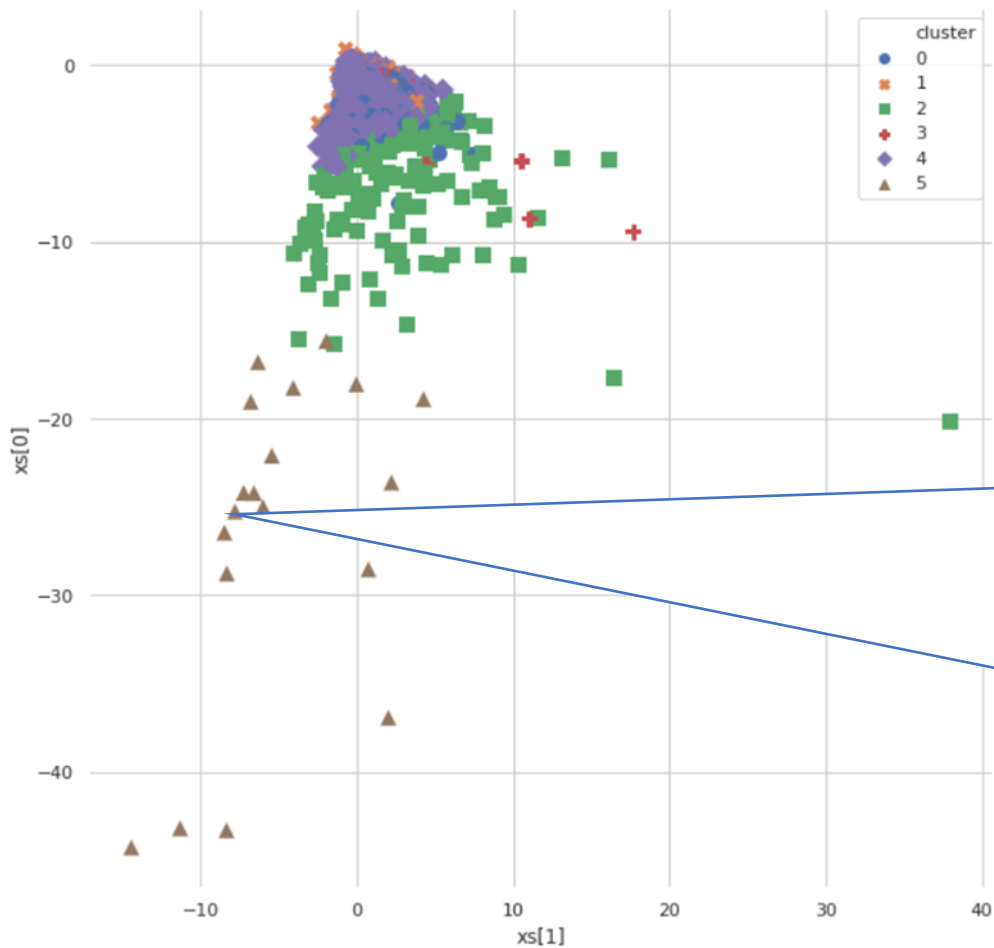
- Qtd. Vendedores: 2.170
- Média de pedidos: 19
- Média de compradores diferentes: 18
- Média de produtos vendidos: 21
- Média de faturamento: quase R\$3.200
- Ticket médio: R\$167,43
- Preço médio: R\$146,40
- Recência: pouco mais de 3 meses desde a última venda
- Frequência: 5 vendas a cada 3 meses
- Atraso na entrega: poucos vendedores desse grupo costumam atrasar
- Esse é o perfil da maioria dos vendedores da Olist: faz poucas vendas e só é interessante para a Olist no volume.



# 5. Modelagem

KMeans SKLearn K = 6 | PERFIL DOS GRUPOS

76



## Prediction=5:

- Qtd. Vendedores: 19
- Média de pedidos: 1.090
- Média de compradores diferentes: 1.076
- Média de produtos vendidos: 1.248
- Média de faturamento: em torno de R\$146.300
- Ticket médio: R\$134,22
- Preço médio: R\$117,25
- Recência: quase 10 dias desde a última venda
- Frequência: quase 60 vendas por mês
- Atraso na entrega: nenhum vendedor desse grupo costuma atrasar
- Esse é o perfil de vendedores premium da Olist: faz muitas vendas e garante bom faturamento, mesmo vendendo produtos de valor agregado mais baixo.



# Metodologia de análise de dados



## Definição do problema

- Objetivos
- Conceitos
- Critérios
- Histórico de dados
- Variáveis

## Análise preliminar

- Medidas de posição
- Análise de frequências
- Gráficos
- Análise de *outliers*
- Análise de *missings*
- Validação sobre a consistência das informações

## Avaliação das técnicas

- K-means nativo do Spark

## Avaliação das técnicas

- biSecting K-means nativo do Spark
- Gaussian Mixture Model nativo do Spark
- Aglomeração do Scikit Learn
- DBSCAN do Scikit Learn
- MeanShift do Scikit Learn
- K-means do Scikit Learn

## Definição da técnica

- Validação dos resultados
- Escolha da técnica que melhor se adequa ao uso e estratégias



## 6. Avaliação Final das Técnicas de Agrupamento

Comparando os resultados dos agrupamentos, temos os seguintes destaques:

### Grupos

- Agrupamentos com 3 clusters fazem bastante sentido do ponto de vista de negócio

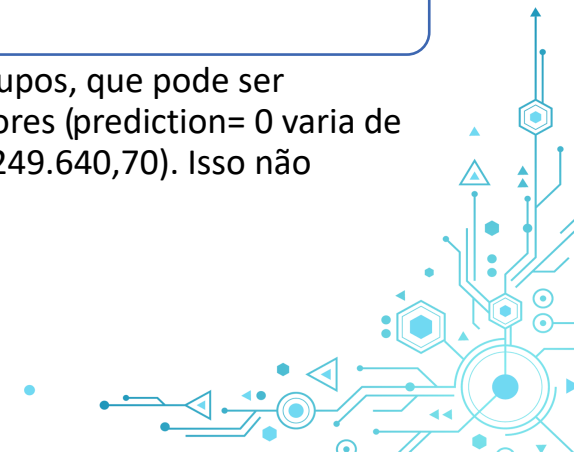
### Métricas:

Agrupamentos com 3 clusters possuem melhor silhoutte score em comparação com o agrupamento de 6 clusters do k-means do Scikit Learn

- K-means nativo do Spark com 3 clusters = 47,5%;
- K-means Scikit Learn com 3 clusters = 40,89%;
- biSec K-means nativo do Spark com 3 cluster = 95,5%;
- K-means Scikit Learn com 6 clusters = 4,29%.

### Modelo escolhido

Como o próprio silhoutte score da técnica biSec K-means dá a entender, esta foi a melhor técnica de separação dos grupos, que pode ser evidenciada quando se analisa a distribuição do valor de venda: são 3 faixas bem distintas, sem sobreposição de valores (prediction= 0 varia de R\$12,22 a R\$20.777,19; prediction=1 varia de R\$21.591,68 a R\$83.189,65; prediction=2 varia de R\$117.409,50 a R\$249.640,70). Isso não acontece nas demais formas de agrupamento, com sobreposição das faixas.



## 6. Avaliação Final das Técnicas de Agrupamento

Comparando os resultados dos agrupamentos, temos os seguintes destaques:

A separação em 6 grupos utilizando o K-means do Scikit Learn, apesar de ter o mais baixo silhoutte score dentre as técnicas selecionadas, traz 6 perfis muito interessantes do ponto de vista de negócios:

- vendedores de produtos mais caros (194 vendedores, 6,3% da base);
- vendedores que não deram certo e desistiram da plataforma (542, 17,5%);
- vendedores aspirantes a premium da Olist: faz mais vendas do que a maioria dos vendedores, mas ainda num patamar bem abaixo do que os vendedores premium (156, 5%);
- subgrupo de vendedores do baixo clero: número de vendas baixo, trabalha com produtos mais baratos, mas compensa um pouco no volume (14, 0,5%);
- maioria dos vendedores da Olist: faz poucas vendas e só é interessante para a Olist no volume (2170, 70,1%);
- vendedores premium da Olist: faz muitas vendas e garante bom faturamento (19, 0,6%).





### Recomendações para a Olist:

- 1) Utilizar a técnica de agrupamento **Bisecting K-means nativa do Spark** para classificar os vendedores numa curva ABC aprimorada pela inteligência artificial, de forma a definir uma política comercial diferenciada para cada um dos 3 grupos de vendedores, levando em conta o valor que cada grupo traz para a empresa
- 2) Utilizar a técnica de agrupamento **K-means com 6 clusters do Scikit Learn** para classificar os vendedores pelo perfil de uso da plataforma Olist e, assim, criar ações de marketing direto diferenciadas para cada perfil (ex.: resgate de vendedores desistentes, bonificação para vendedores aspirantes a premium se esforçarem para atingir o perfil premium)





Avaliação de Ganho – Política Comercial

Plano	Limite de Venda	Qtd. Vendedores	Valor de Venda	Comissão OLIST	Mensalidade OLIST	Faturamento Comissão OLIST	Faturamento Mensalidade OLIST	Total Faturamento OLIST
Lite	R\$ 3.000	2.189	R\$ 1.679.286	21%	R\$ -	R\$ 352.650	R\$ -	R\$ 352.650
Pro	R\$ 20.000	756	R\$ 5.968.876	19%	R\$ 79,90	R\$ 1.134.086	R\$ 60.404	R\$ 1.194.491
Premium	Sem limite	150	R\$ 8.269.154	15%	R\$ -	R\$ 1.240.373	R\$ -	R\$ 1.240.373
Total Faturamento OLIST						R\$ 2.787.514		

Plano	Limite de Venda	Qtd. Vendedores	Valor de Venda	Comissão OLIST	Mensalidade OLIST	Faturamento Comissão OLIST	Faturamento Mensalidade OLIST	Total Faturamento OLIST
Lite	R\$ 20.000	2.945	R\$ 7.648.163	21%	R\$ -	R\$ 1.606.114	R\$ -	R\$ 1.606.114
Pro	R\$ 100.000	131	R\$ 5.034.334	19%	R\$ 79,90	R\$ 956.523	R\$ 10.467	R\$ 966.990
Premium	Sem limite	19	R\$ 3.234.820	15%	R\$ -	R\$ 485.223	R\$ -	R\$ 485.223
Total Faturamento OLIST						R\$ 3.058.328		

Plano	Limite de Venda	Qtd. Vendedores	Valor de Venda	Comissão OLIST	Mensalidade OLIST	Faturamento Comissão OLIST	Faturamento Mensalidade OLIST	Total Faturamento OLIST
Lite	R\$ 20.000	2.945	R\$ 7.648.163	19%	R\$ 74,99	R\$ 1.453.151	R\$ 220.846	R\$ 1.673.997
Pro	R\$ 100.000	131	R\$ 5.034.334	15%	R\$ -	R\$ 755.150	R\$ -	R\$ 755.150
Premium	Sem limite	19	R\$ 3.234.820	14%	R\$ -	R\$ 452.875	R\$ -	R\$ 452.875
Total Faturamento OLIST						R\$ 2.882.021		

# 7. Conclusões

## Avaliação de Ganho – Política Comercial

### Política comercial atual

- Adotamos valor de comissão de 15% para o grupo premium, pois não temos conhecimento do valor real.

			Valor de Venda	Comissão OLIST	Mensalidade OLIST	Faturamento Comissão OLIST	Faturamento Mensalidade OLIST	Total Faturamento OLIST
			R\$ 8.286	21%	R\$ -	R\$ 352.650	R\$ -	R\$ 352.650
			R\$ 5.968.876	19%	R\$ 79,90	R\$ 1.134.086	R\$ 60.404	R\$ 1.194.491
Premium	Sem limite	150	R\$ 8.269.154	15%	R\$ -	R\$ 1.240.373	R\$ -	R\$ 1.240.373
Total Faturamento OLIST						R\$ 2.787.514		

Plano	Limite de Venda	Qtd. Vendedores	Valor de Venda	Comissão OLIST	Mensalidade OLIST	Faturamento Comissão OLIST	Faturamento Mensalidade OLIST	Total Faturamento OLIST
Lite	R\$ 20.000	2.945	R\$ 7.648.163	21%	R\$ -	R\$ 1.606.114	R\$ -	R\$ 1.606.114
Pro	R\$ 100.000	131	R\$ 5.034.334	19%	R\$ 79,90	R\$ 956.523	R\$ 10.467	R\$ 966.990
Premium	Sem limite	19	R\$ 3.234.820	15%	R\$ -	R\$ 485.223	R\$ -	R\$ 485.223
Total Faturamento OLIST						R\$ 3.058.328		

Plano	Limite de Venda	Qtd. Vendedores	Valor de Venda	Comissão OLIST	Mensalidade OLIST	Faturamento Comissão OLIST	Faturamento Mensalidade OLIST	Total Faturamento OLIST
Lite	R\$ 20.000	2.945	R\$ 7.648.163	19%	R\$ 74,99	R\$ 1.453.151	R\$ 220.846	R\$ 1.673.997
Pro	R\$ 100.000	131	R\$ 5.034.334	15%	R\$ -	R\$ 755.150	R\$ -	R\$ 755.150
Premium	Sem limite	19	R\$ 3.234.820	14%	R\$ -	R\$ 452.875	R\$ -	R\$ 452.875
Total Faturamento OLIST						R\$ 2.882.021		

Avaliação de Ganho – Política Comercial

Plano	Limite de Venda	Qtd. Vendedores	Valor de Venda	Comissão OLIST	Mensalidade OLIST	Faturamento Comissão OLIST	Faturamento Mensalidade OLIST	Total Faturamento OLIST
Lite	R\$ 3.000	2.189	R\$ 1.679.286	21%	R\$ -	R\$ 352.650	R\$ -	R\$ 352.650
Pro	R\$ 20.000	756	R\$ 5.968.876	19%	R\$ 79,90	R\$ 1.134.086	R\$ 60.404	R\$ 1.194.491
Premium	Sem limite	150	R\$ 8.269.154	15%	R\$ -	R\$ 1.240.373	R\$ -	R\$ 1.240.373
Total Faturamento OLIST						R\$ 2.787.514		

Proposta 1

- Apenas reclassificando as linhas de corte de acordo com os valores obtidos com a técnica de Bisecting K-means, haveria um ganho de R\$270.813,59.

Plano	Limite de Venda	Qtd. Vendedores	Valor de Venda	Comissão OLIST	Mensalidade OLIST	Faturamento Comissão OLIST	Faturamento Mensalidade OLIST	Total Faturamento OLIST
Lite	R\$ 20.000	2.945	R\$ 7.648.163	19%	R\$ -	R\$ 1.606.114	R\$ -	R\$ 1.606.114
Pro	R\$ 100.000	131	R\$ 5.034.334	19%	R\$ 79,90	R\$ 956.523	R\$ 10.467	R\$ 966.990
Premium	Sem limite	19	R\$ 3.234.820	15%	R\$ -	R\$ 485.223	R\$ -	R\$ 485.223
Total Faturamento OLIST						R\$ 3.058.328		

Plano	Limite de Venda	Qtd. Vendedores	Valor de Venda	Comissão OLIST	Mensalidade OLIST	Faturamento Comissão OLIST	Faturamento Mensalidade OLIST	Total Faturamento OLIST
Lite	R\$ 20.000	2.945	R\$ 7.648.163	19%	R\$ 74,99	R\$ 1.453.151	R\$ 220.846	R\$ 1.673.997
Pro	R\$ 100.000	131	R\$ 5.034.334	15%	R\$ -	R\$ 755.150	R\$ -	R\$ 755.150
Premium	Sem limite	19	R\$ 3.234.820	14%	R\$ -	R\$ 452.875	R\$ -	R\$ 452.875
Total Faturamento OLIST						R\$ 2.882.021		



Avaliação de Ganho – Política Comercial

Plano	Limite de Venda	Qtd. Vendedores	Valor de Venda	Comissão OLIST	Mensalidade OLIST	Faturamento Comissão OLIST	Faturamento Mensalidade OLIST	Total Faturamento OLIST
Lite	R\$ 3.000	2.189	R\$ 1.679.286	21%	R\$ -	R\$ 352.650	R\$ -	R\$ 352.650
Pro	R\$ 20.000	756	R\$ 5.968.876	19%	R\$ 79,90	R\$ 1.134.086	R\$ 60.404	R\$ 1.194.491
Premium	Sem limite	150	R\$ 8.269.154	15%	R\$ -	R\$ 1.240.373	R\$ -	R\$ 1.240.373
Total Faturamento OLIST						R\$ 2.787.514		

Plano	Limite de Venda	Qtd. Vendedores	Valor de Venda	Comissão OLIST	Mensalidade OLIST	Faturamento Comissão OLIST	Faturamento Mensalidade OLIST	Total Faturamento OLIST
Lite	R\$ 20.000	2.945	R\$ 7.648.163	21%	R\$ -	R\$ 1.606.114	R\$ -	R\$ 1.606.114
Pro	R\$ 100.000	131	R\$ 683.234	19%	R\$ 79,90	R\$ 956.523	R\$ 10.467	R\$ 966.990
Premium	Sem limite	19	R\$ 485.223	15%	R\$ -	R\$ 485.223	R\$ -	R\$ 485.223

Proposta 1

- Há um porém: na nova classificação, todos os antigos vendedores Lite e Pro foram agregados no novo Lite, e os antigos vendedores Premium foram subdivididos nos novos Pro e Premium. Com isso, as comissões subiriam de 19% para 21% para quem era Pro, e de 15% para 19% + mensalidade para parte dos antigos Premium. Essa piora nas condições comerciais pode gerar uma fuga de vendedores da carteira atual da OLIST, matando o ganho no faturamento com a mudança de política.

Plano	Limite de Venda	Qtd.
Lite	R\$ 20.000	
Pro	R\$ 100.000	
Premium	Sem limite	

Total Faturamento OLIST
R\$ 1.673.997
R\$ 755.150
R\$ 152.875



# 7. Conclusões



## Avaliação de Ganho –

Plano	Limite de Venda	Qtd. Vendedores
Lite	R\$ 3.000	
Pro	R\$ 20.000	
Premium	Sem limite	

Plano	Limite de Venda	Qtd. Vendedores
Lite	R\$ 20.000	
Pro	R\$ 100.000	
Premium	Sem limite	

Plano	Limite de Venda	Qtd. Vendedores	Valor de Venda	Comissão OLIST
Lite	R\$ 20.000	2.945	R\$ 7.648.163	19%
Pro	R\$ 100.000	131	R\$ 5.034.334	15%
Premium	Sem limite	19	R\$ 3.234.820	14%

## Proposta 2

- Aqui, além de adotar a nova classificação de limite de venda para definição do plano, sugerimos alterar as condições comerciais:
  - O plano Lite passa a ter mensalidade, num valor menor do que o cobrado no antigo plano Pro, e a comissão baixa de 21% para 19% (ou seja, quem era Lite ganha diminuição da comissão e quem era Pro ganha diminuição da mensalidade).
  - O plano Pro deixa de ter mensalidade, e a comissão baixa de 19% para 15% (na verdade, não mudamos as condições para quem estava no antigo Premium e foi reclassificado para o novo Pro).
  - O plano Premium baixa a comissão de 15% para 14%.
- Ou seja, é possível fazer ações de marketing para mostrar a todos os públicos que as condições comerciais melhoraram para os vendedores.

R\$ -	R\$ 485.223	R\$ -	R\$ 485.223
Faturamento OLIST	R\$ 3.058.328		

Mensalidade OLIST	Faturamento Comissão OLIST	Faturamento Mensalidade OLIST	Total Faturamento OLIST
R\$ 74,99	R\$ 1.453.151	R\$ 220.846	R\$ 1.673.997
R\$ -	R\$ 755.150	R\$ -	R\$ 755.150
R\$ -	R\$ 452.875	R\$ -	R\$ 452.875
Total Faturamento OLIST		R\$ 2.882.021	

Avaliação de Ganho – Política Comercial

						Faturamento Comissão OLIST	Faturamento Mensalidade OLIST	Total Faturamento OLIST
						R\$ 352.650	R\$ -	R\$ 352.650
						R\$ 1.134.086	R\$ 60.404	R\$ 1.194.491
						R\$ 1.240.373	R\$ -	R\$ 1.240.373
						2.787.514		
						Faturamento Comissão OLIST	Faturamento Mensalidade OLIST	Total Faturamento OLIST
						R\$ 1.606.114	R\$ -	R\$ 1.606.114
						R\$ 956.523	R\$ 10.467	R\$ 966.990
						R\$ 5.223	R\$ -	R\$ 485.223
						3.058.328		
Plano	Limite de Venda	Qtd. Vendedores	Valor de Venda	Comissão OLIST	OLIST	Faturamento Comissão OLIST	Faturamento Mensalidade OLIST	Total Faturamento OLIST
Lite	R\$ 20.000	2.945	R\$ 7.648.163	19%	R\$ 74,99	R\$ 1.453.151	R\$ 10.846	R\$ 1.673.997
Pro	R\$ 100.000	131	R\$ 5.034.334	15%	R\$ -	R\$ 755.150	R\$ -	R\$ 755.150
Premium	Sem limite	19	R\$ 3.234.820	14%	R\$ -	R\$ 452.875	R\$ -	R\$ 452.875
Total Faturamento OLIST						R\$ 2.882.021		

**Proposta 2**

- Do ponto de vista da OLIST, conseguimos um aumento de faturamento de R\$94.507,42, sem risco de fuga dos vendedores atuais, já que todos ganharam melhores condições comerciais.
- Mais do que isso, invertemos completamente a curva de faturamento por plano, com o plano Lite passando de 12% do faturamento para 58%, diminuindo drasticamente a grande dependência que a OLIST tinha dos 150 clientes do antigo plano Premium, que era responsável por 44% do faturamento. Ou seja, se antes perder clientes Premium trazia grande impacto para o faturamento da OLIST, com a nova proposta esse risco é muito mais diluído.