



# Aprendizado não supervisionado **Emissão de CO<sub>2</sub>**

## Membros

Cristian César, Rodrigo Coffani, Vitor Orsin, Vinícius Guimarães

# Agenda

- 1 Explicação do Dataset
- 2 Análise do Dataset
- 3 Pré-Processamento
- 4 Treinamento não supervisionado

# 1. Dataset: CO<sub>2</sub> Emissions



# Atributos do Dataset

Número de Linhas: 7385. Número de atributos: 12

Make	Model	Vehicle Class	Engine Size	Cylinders	Transmission
Fabricante	Modelo	Tipo do carro	Volume do motor	número de cilindros	tipo de câmbio
Ford, chevrolet, bmw, mercedes- benz, porsche	Valores únicos	SUV-SMALL MID-SIZE COMPACT SUV- STANDARD FULL-SIZE	Numérico: 0.9 até 8.4	Numérico discreto: 3 até 16	A - Automático AM - A. Manual AS - seleção de Shift AV - Variável M - Manual

# Atributos do Dataset

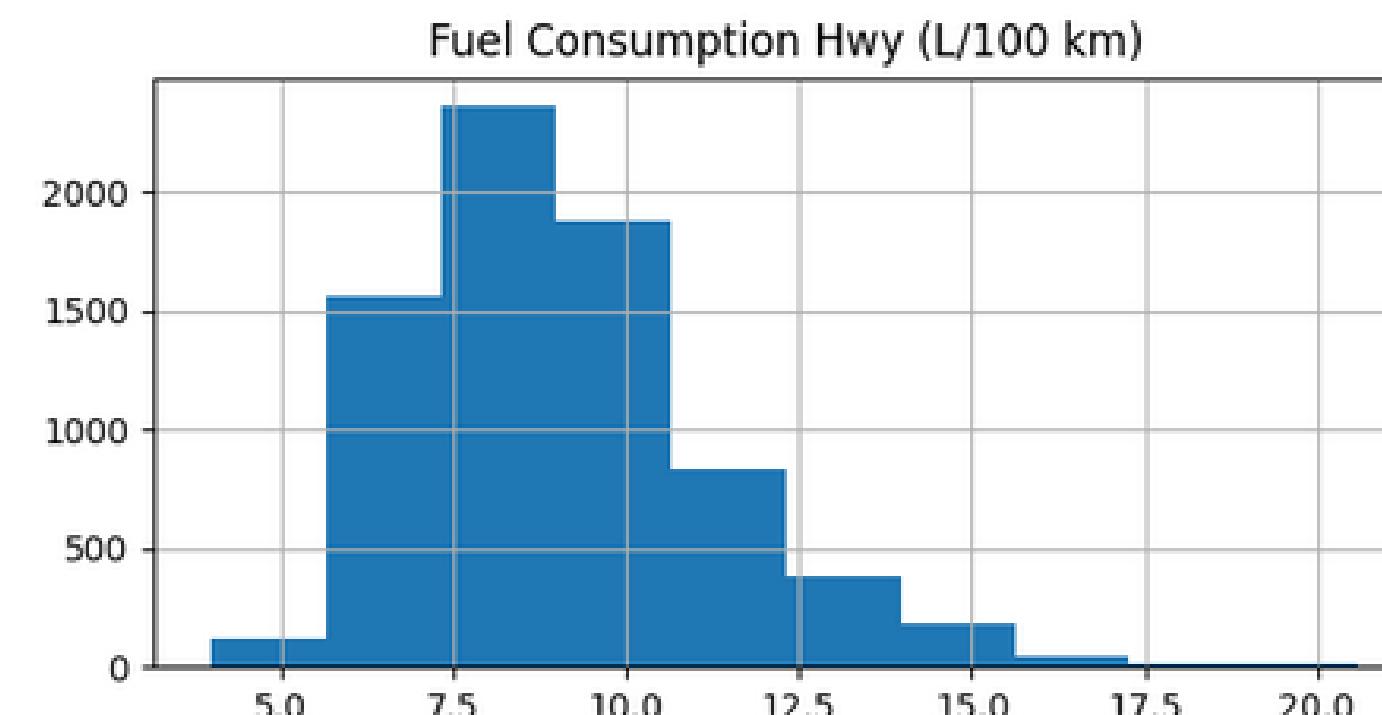
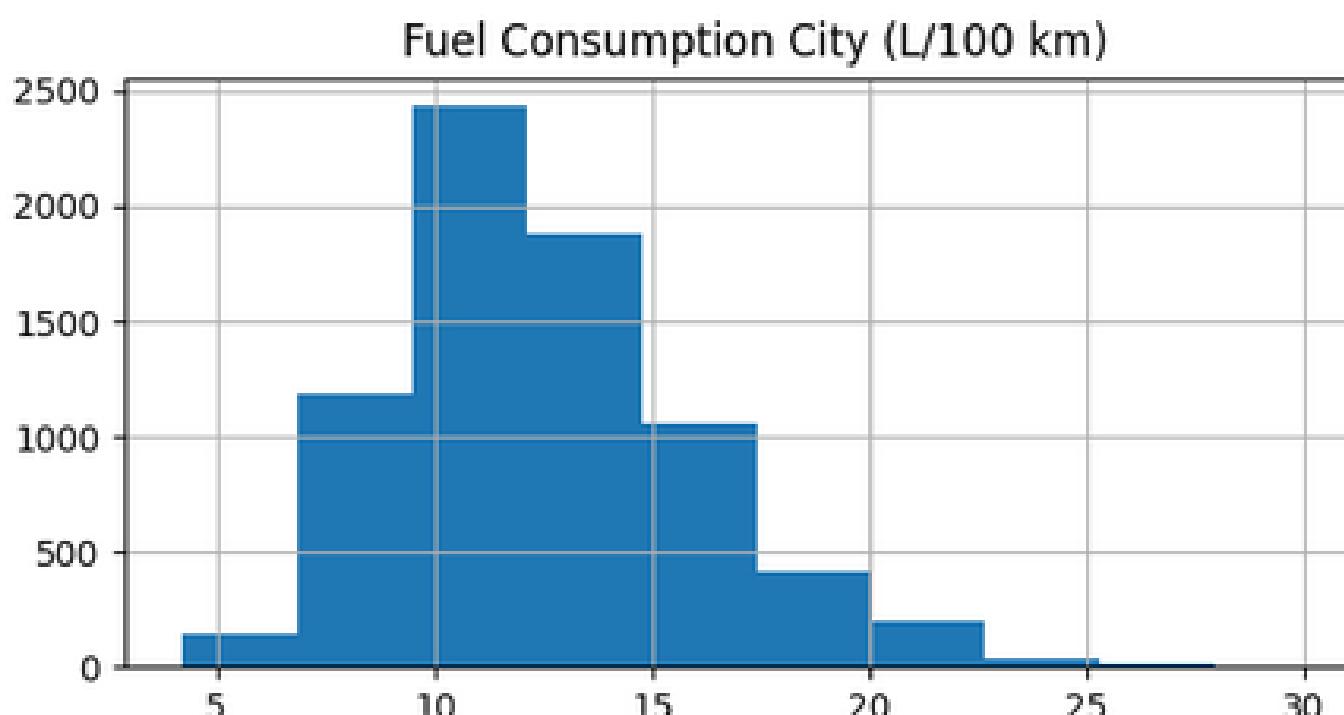
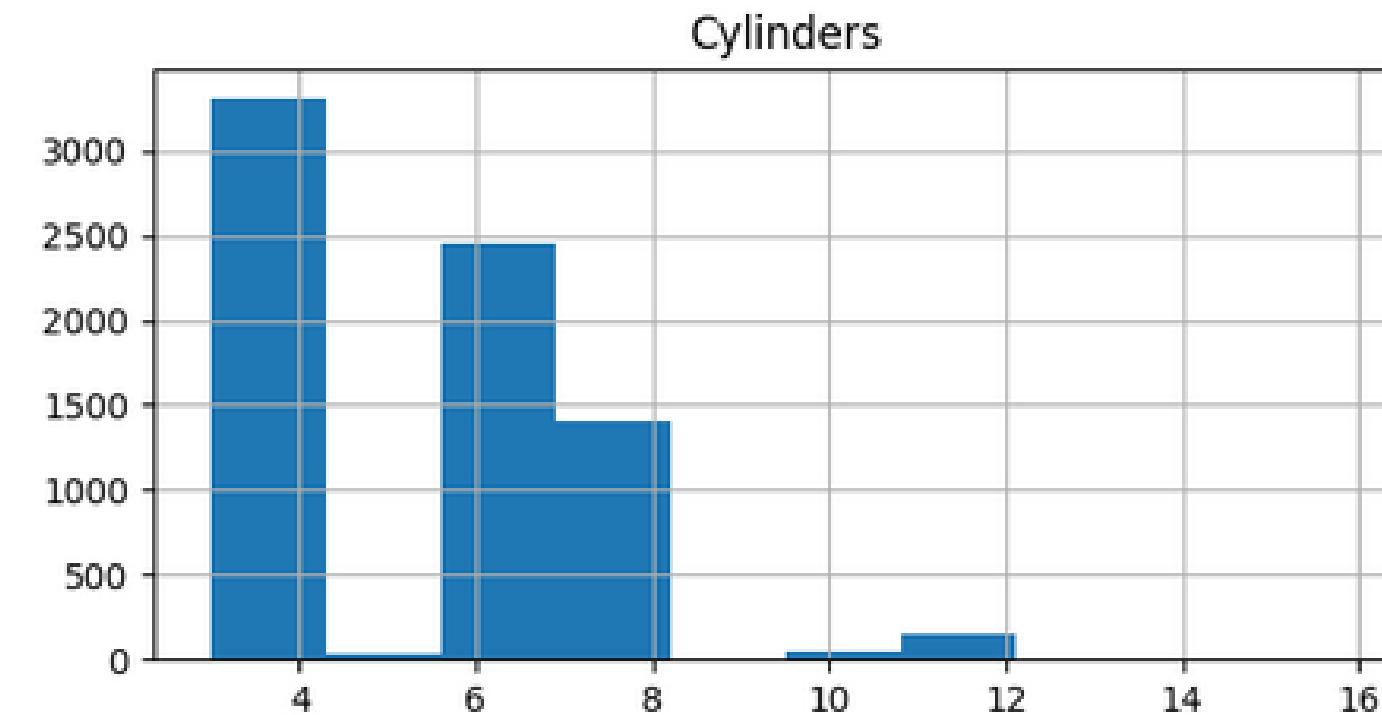
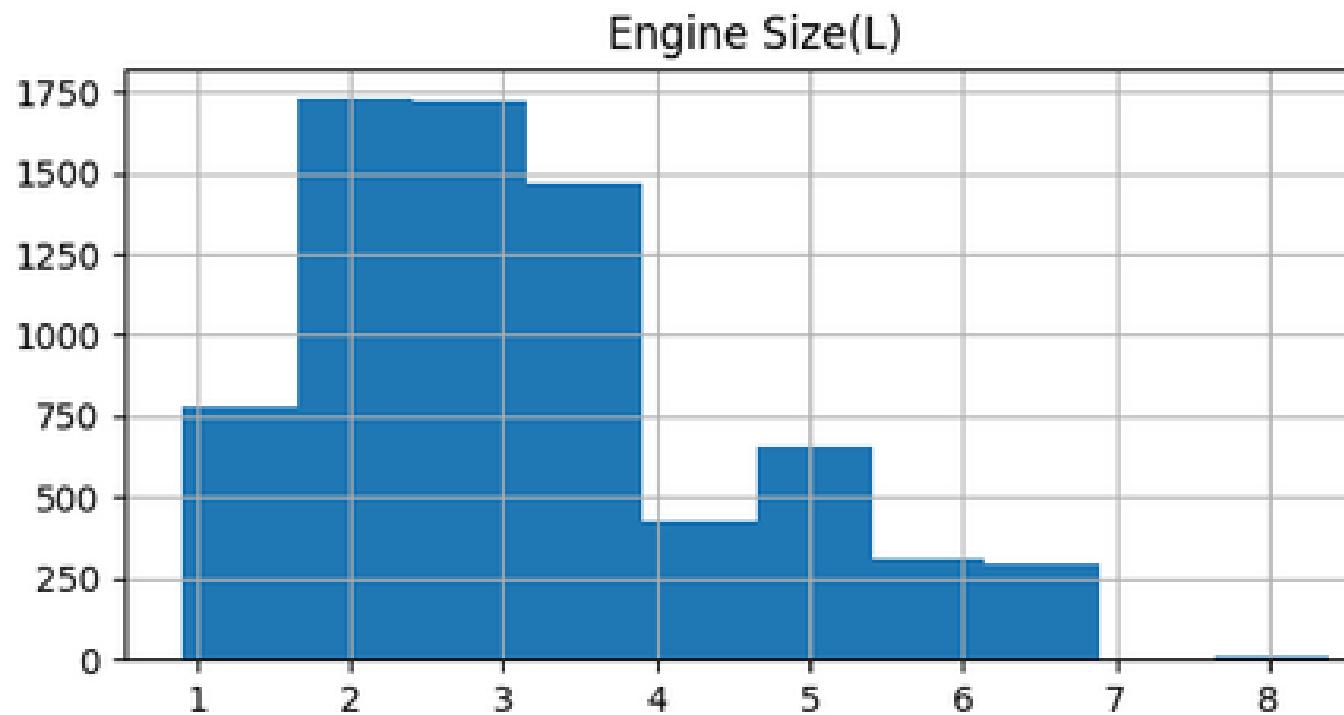
Número de Linhas: 7385. Número de atributos: 12

Fuel Type	Fuel Consumption City(L/100)	Fuel Consumption Hwy(L/100)	Fuel Consumption Comb(L/100)	Fuel Consumption Comb(mpg)	CO2 Emissions (g/km)
Fabricante	Consumo na cidade	Consumo na estrada	Consumo em média	Consumo em média	Taxa de emissão de CO2
X - Gasolina Z - Gasolina Premium D - Diesel E - Ethanol N - Gás Natural	Numérico: 4.2 até 30.6	Numérico: 4 até 20.6	Numérico: 4 até 20.6	Numérico: 11 até 69	Numérico: 96 até 522

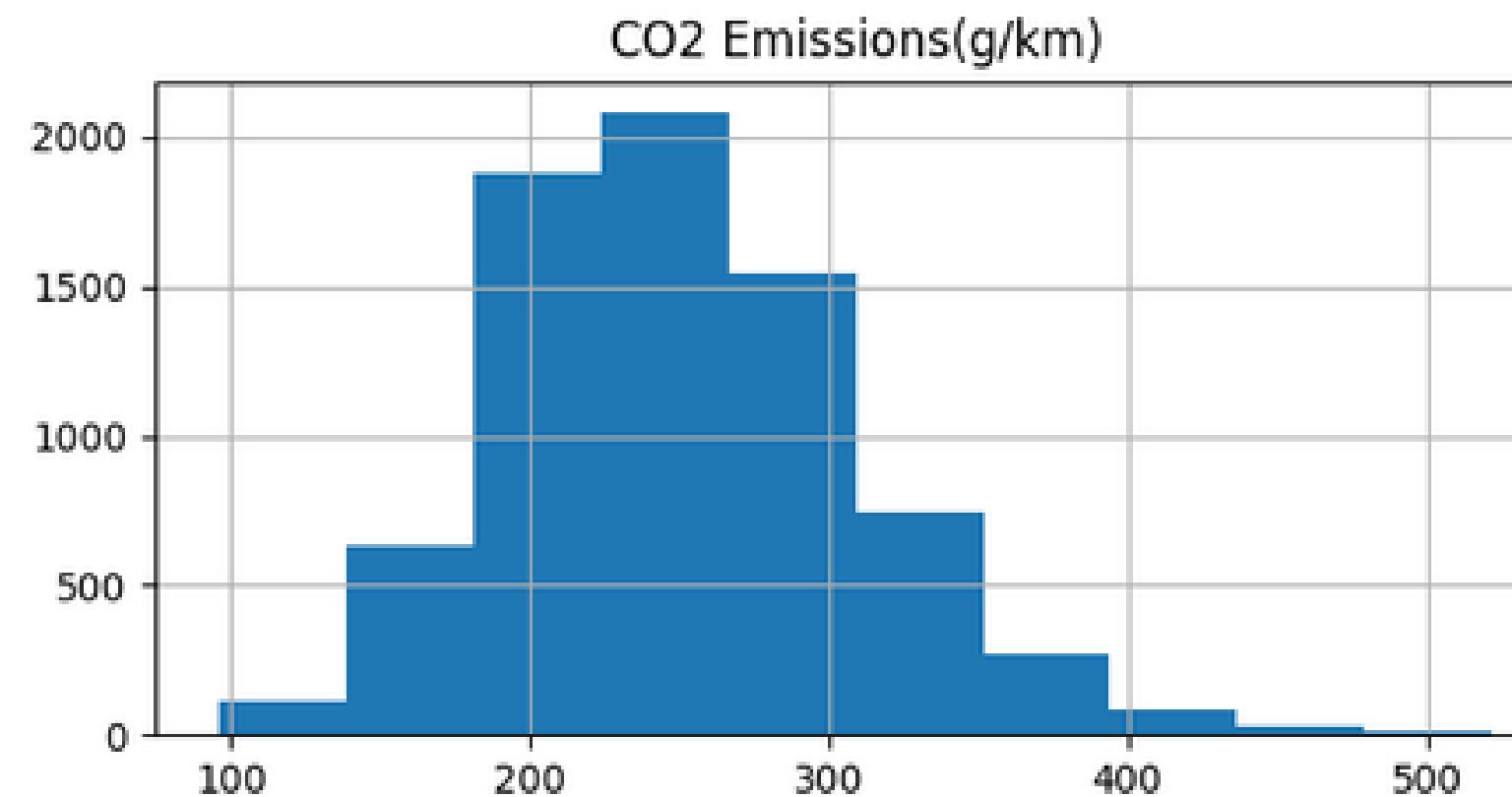
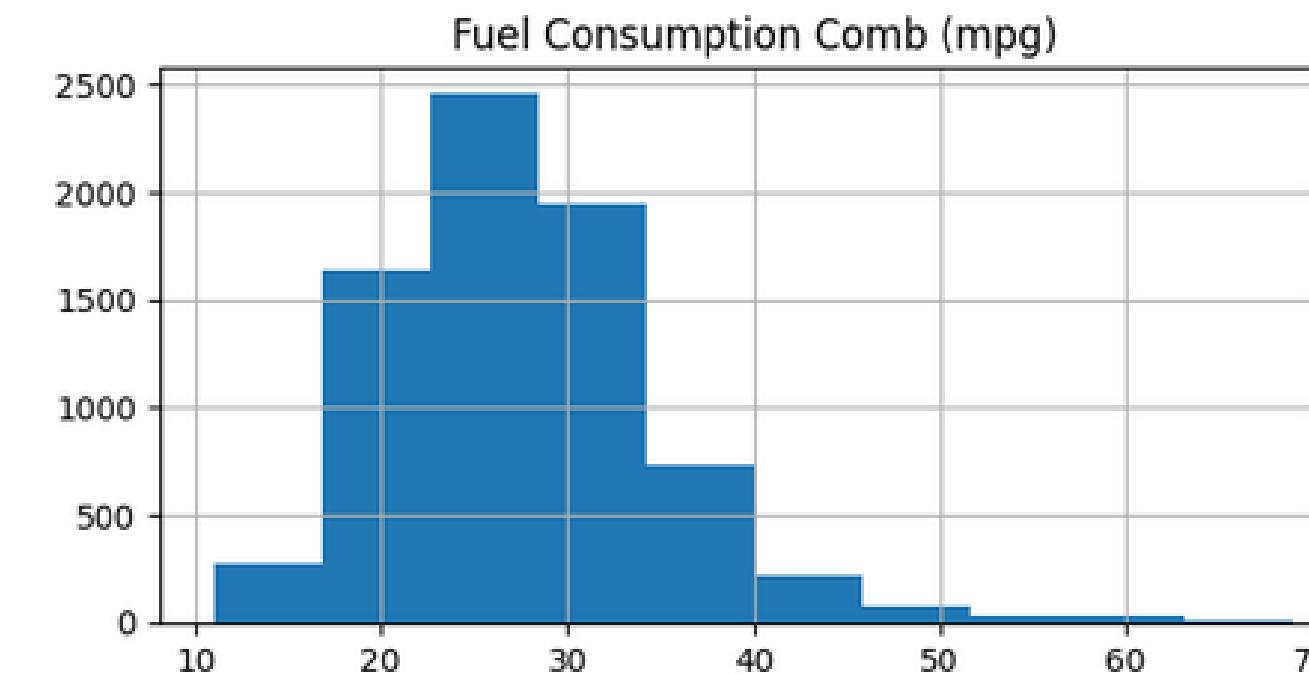
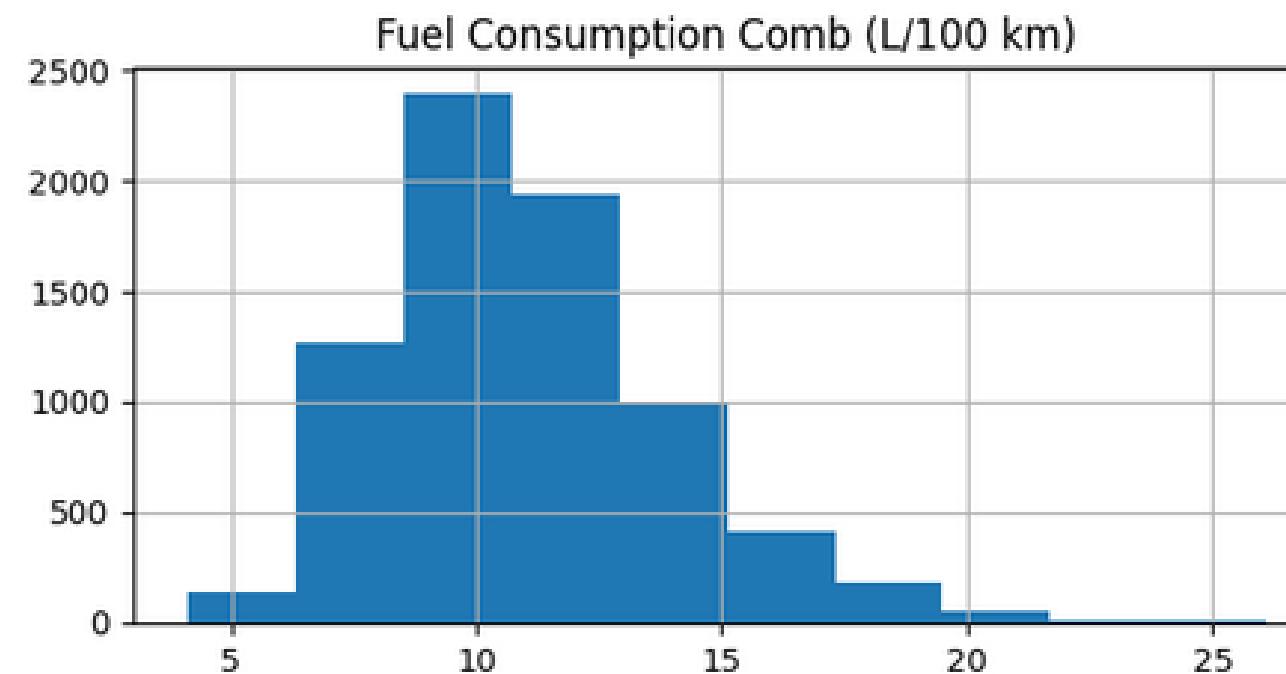
## 2. Análise do Dataset



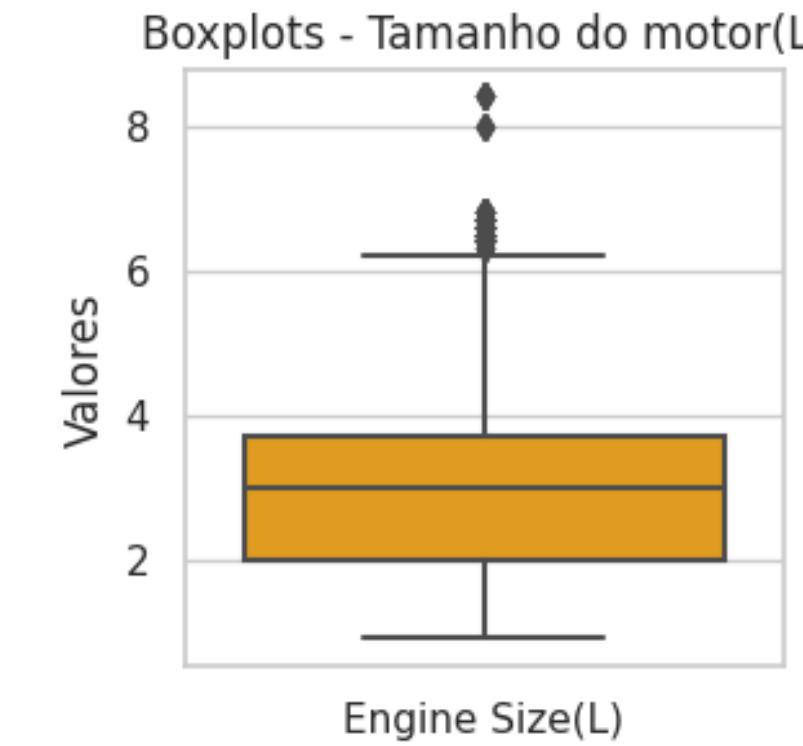
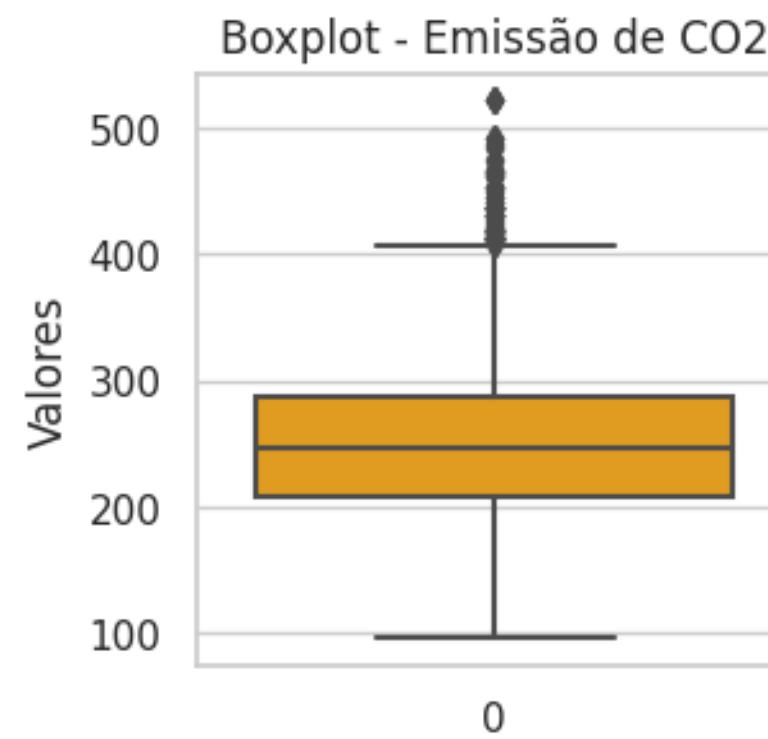
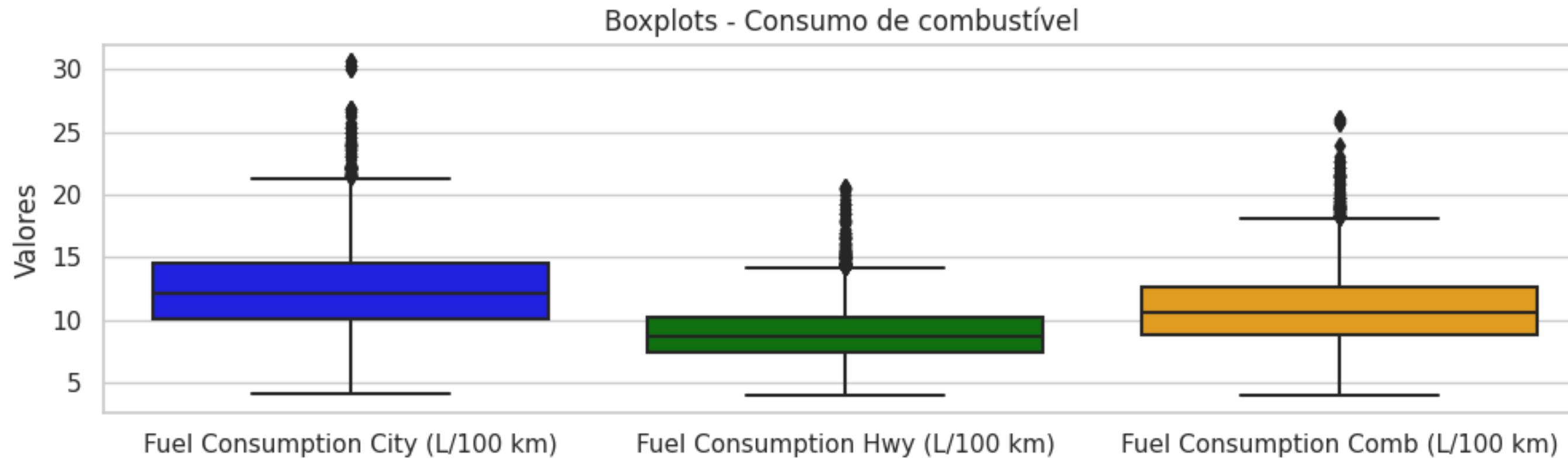
# Visão dos Dados: Histograma



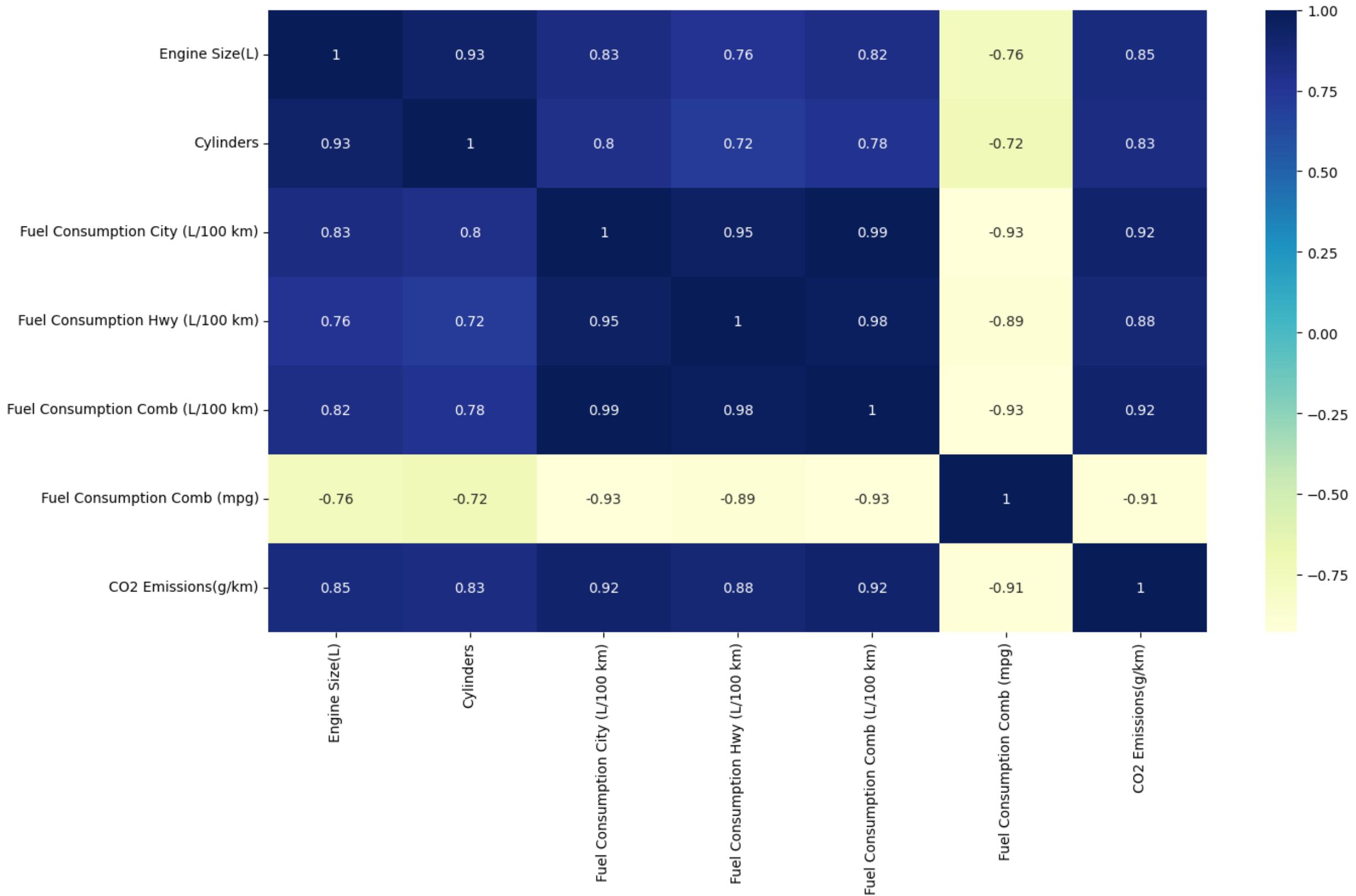
# Visão dos Dados: Histograma



# Visão dos Dados: Boxplots



# Visão dos Dados:Heatmap



# 3. Pré Processamento



# Seleção de atributos

## **Remover coluna "Fuel consumption Comb (mpg)"**

é redundante, visto que se trata da conversão métrica de outro campo.

## **Remover coluna "Fuel Consumption City (L/100 km)"**

é redundante, visto que se trata da conversão métrica de outro campo.

## **Remover coluna "Fuel Consumption Highway (L/100 km)"**

é redundante, visto que se trata da conversão métrica de outro campo.

# Pré Processamento

**Remoção de duplicatas e embaralhamento da amostra**

Quantidade de linhas duplicadas: 1394

## **Transformação de atributos**

Convertendo strings para objetos usando o Label Encoder

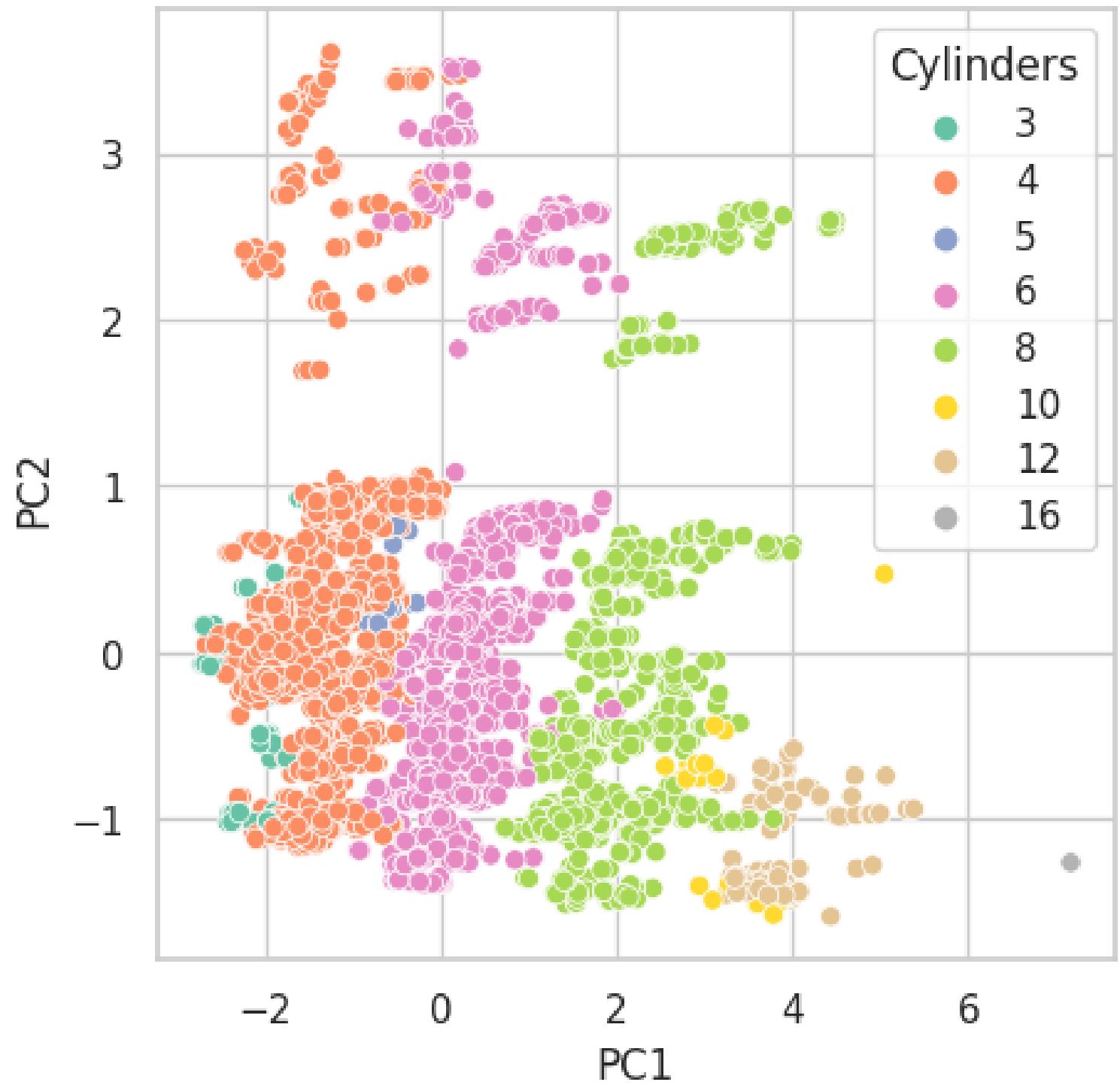
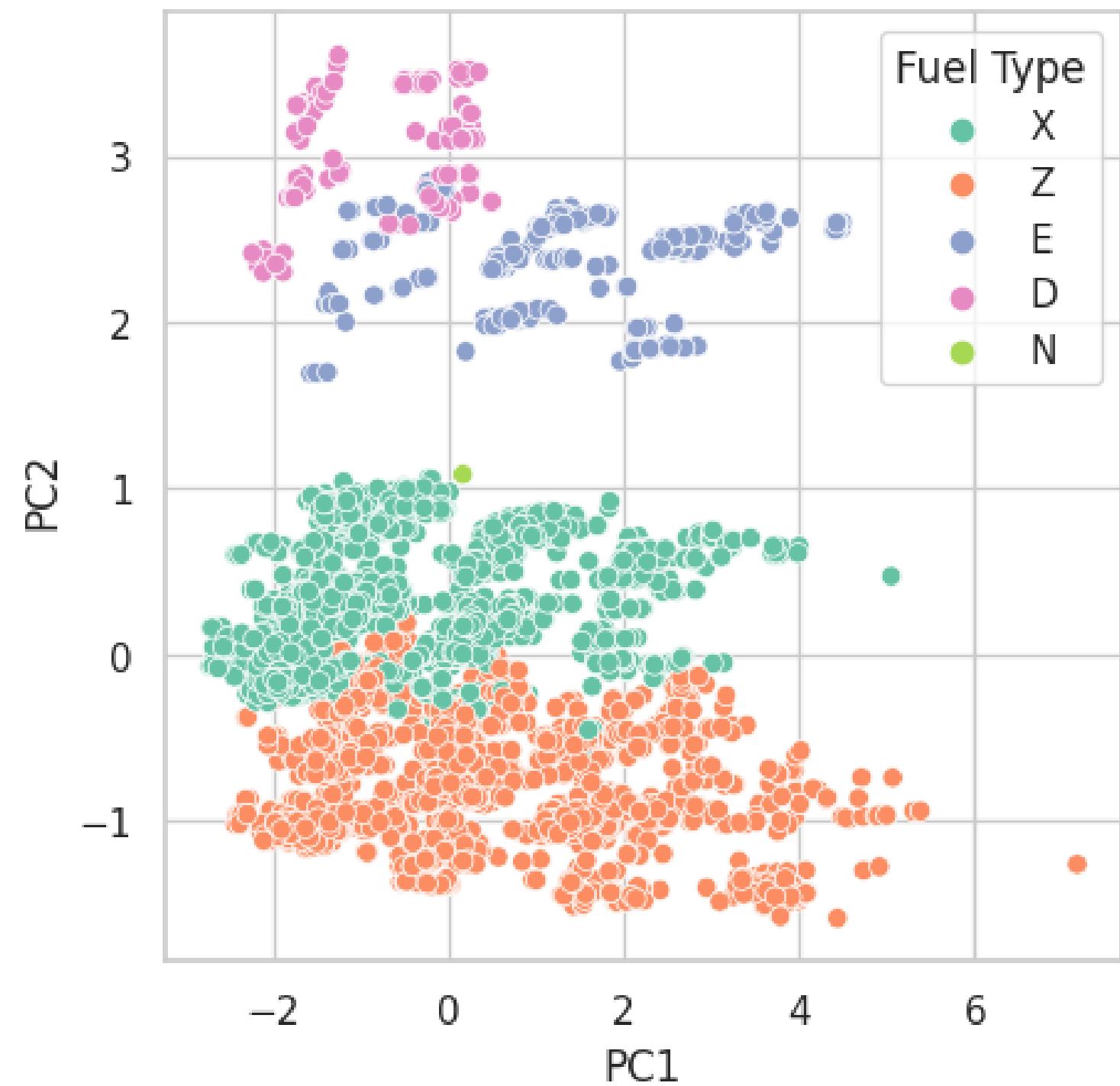
## **Normalização**

RobustScaler, menos sensível a valores extremos.

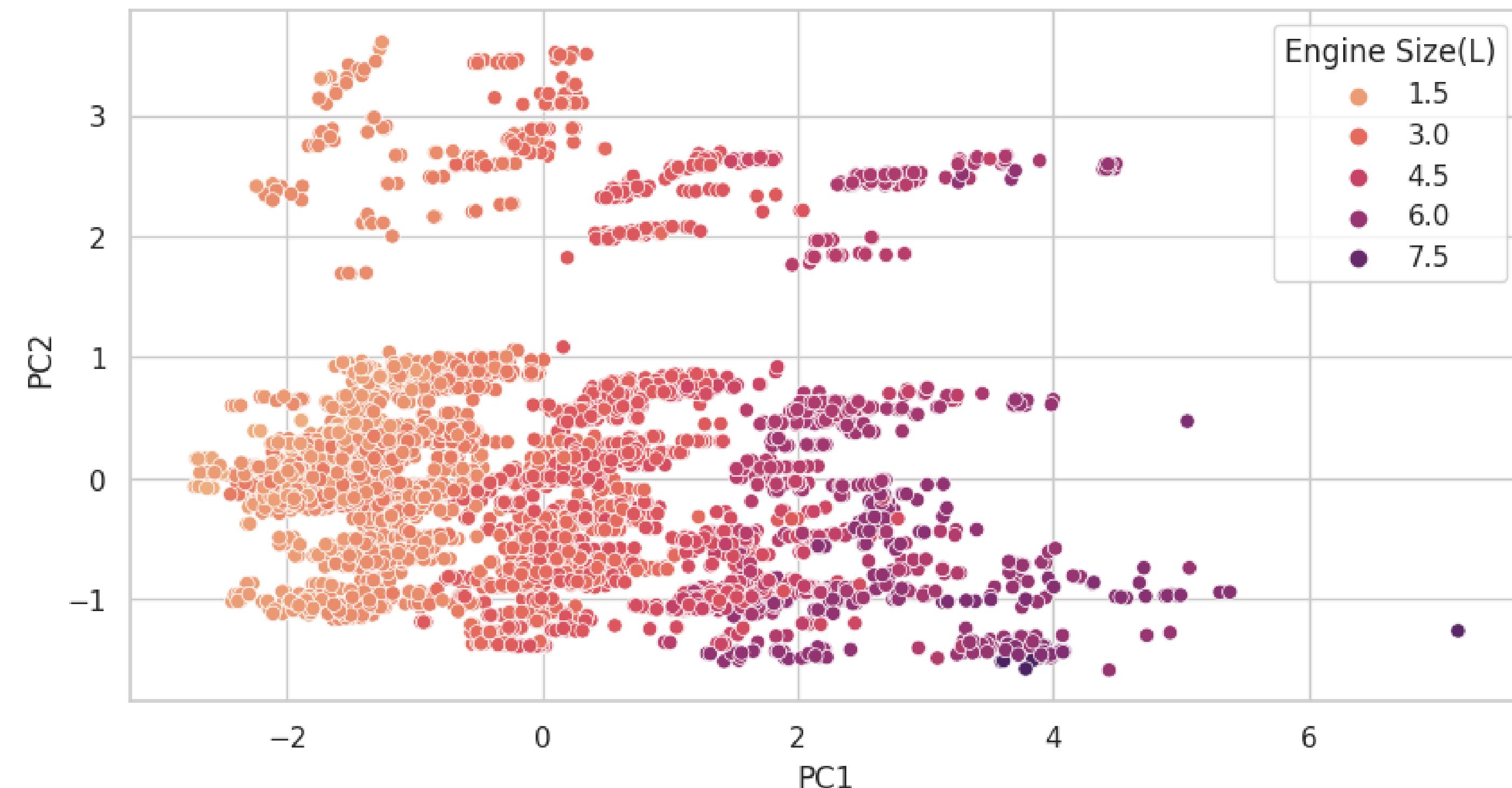
## **Separação do atributo "CO2 Emission"**

Guardamos o valor para poder ser utilizado como métrica

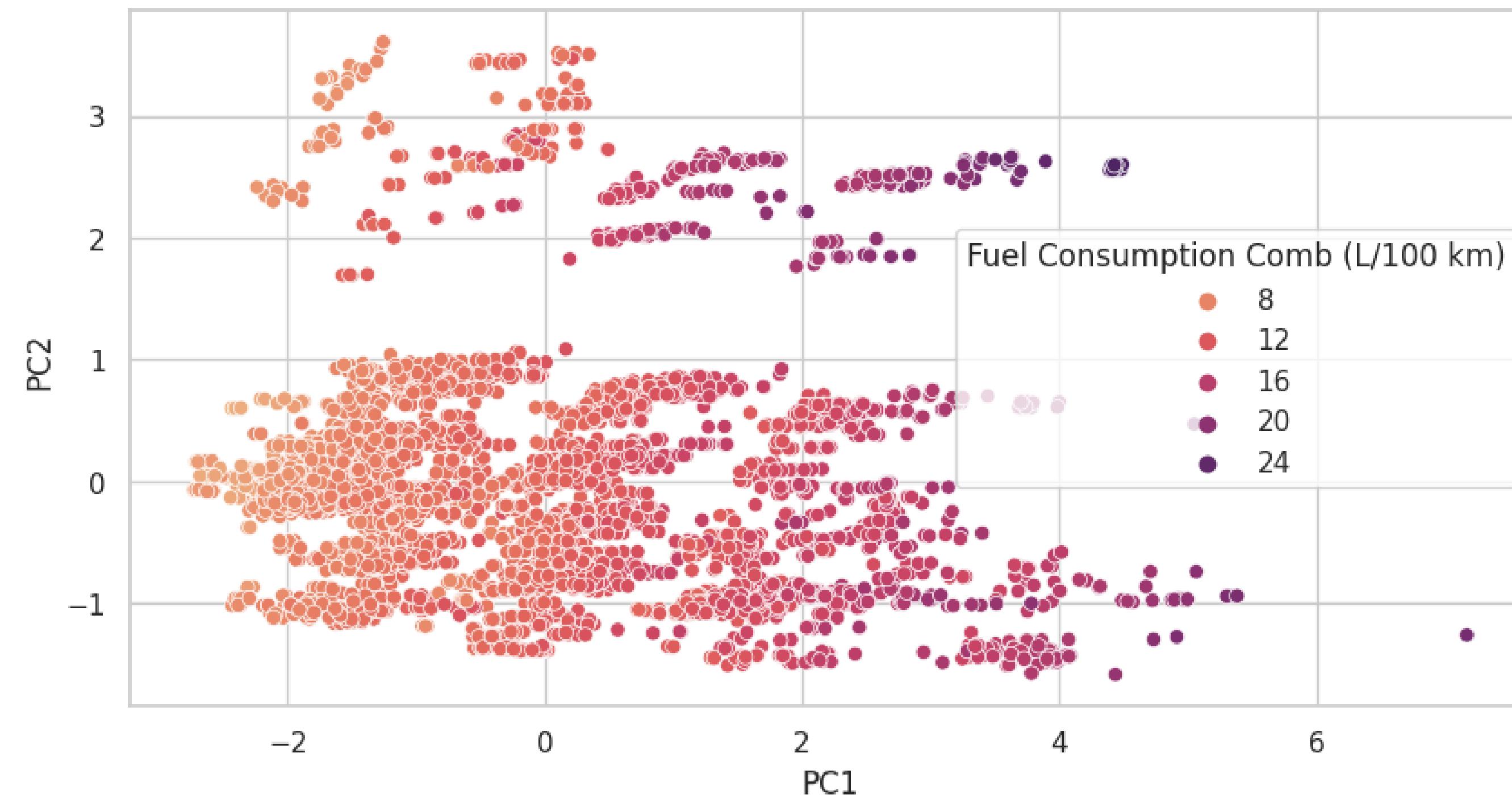
# Pré Processamento: PCA



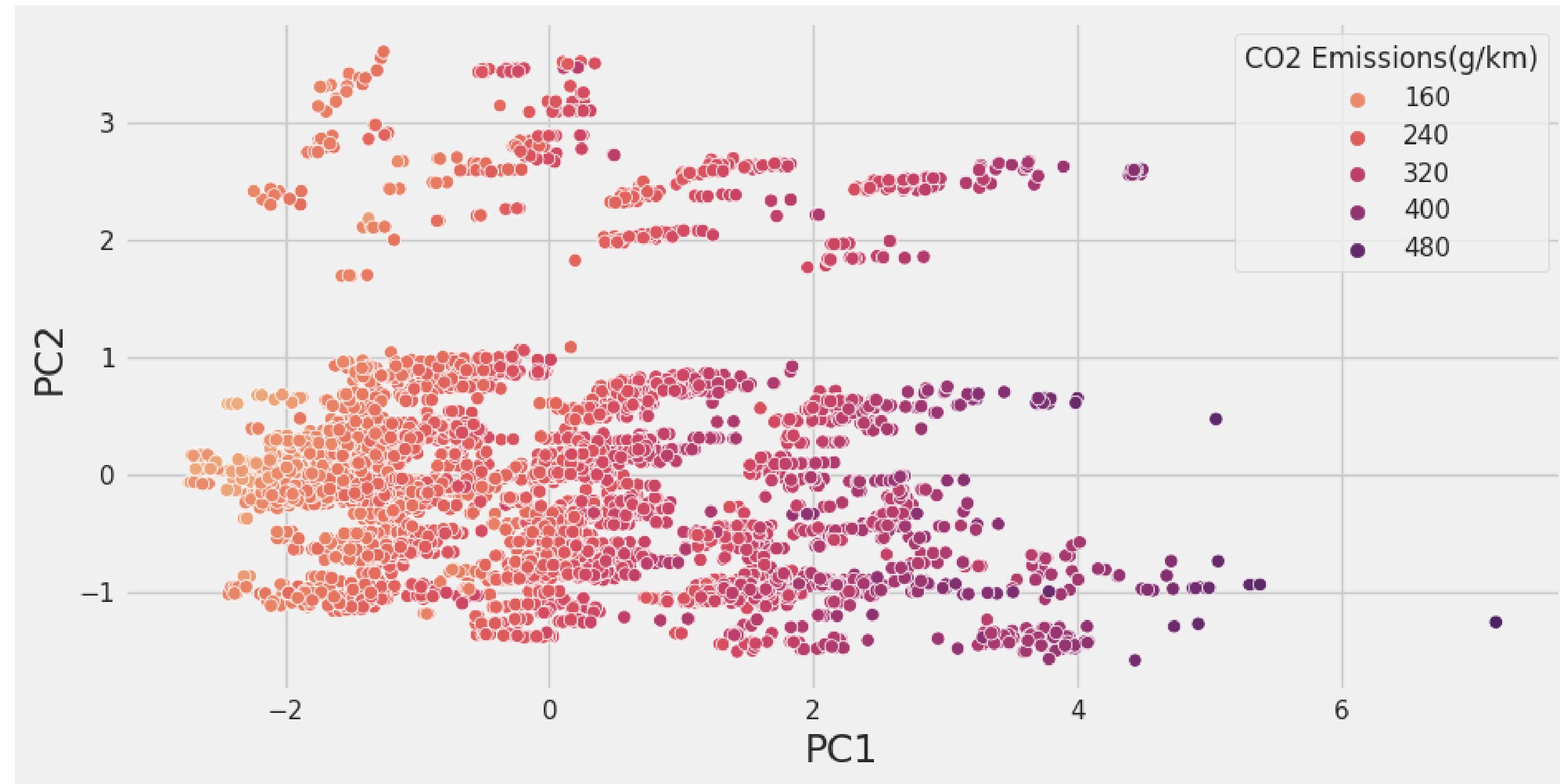
# Pré Processamento: PCA



# Pré Processamento: PCA



# Pré Processamento: PCA



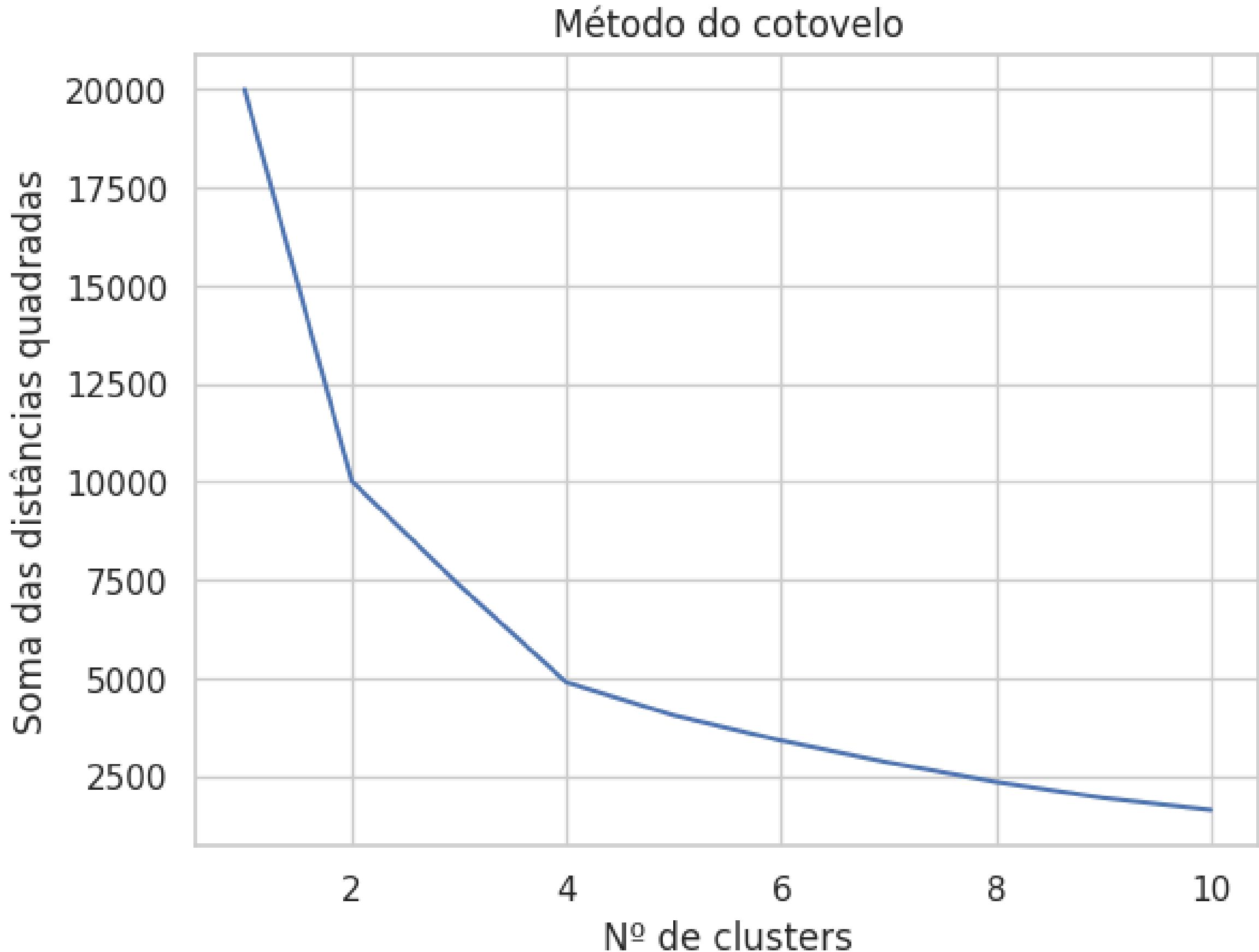


# 1. Testes não supervisionados

# K-means

Utilização do  
método do  
cotovelo

Número de clusters  
considerado: **4**



# K-means

Eixos considerados foram PC1 e PC2, resultantes da aplicação do PCA

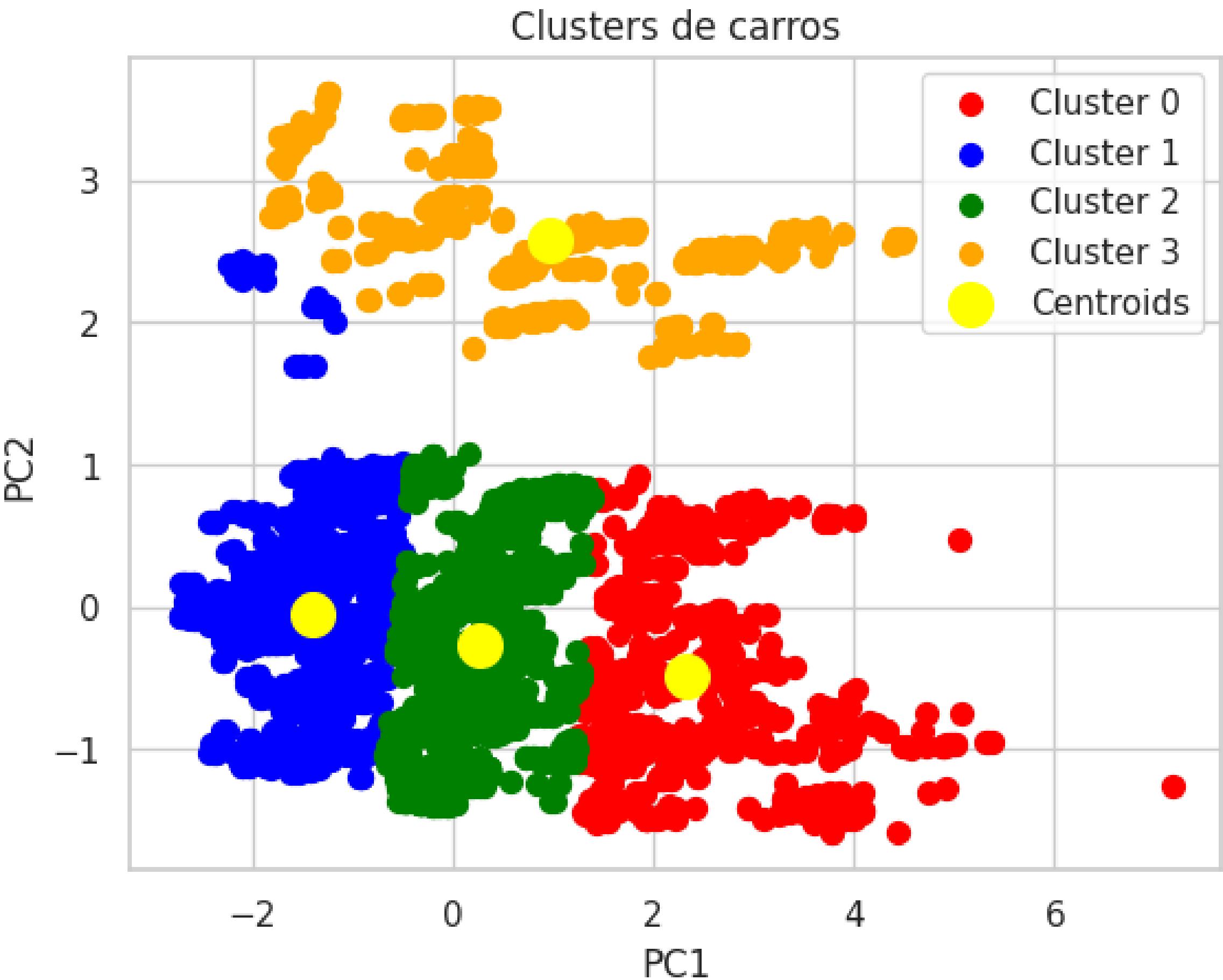
Número de clusters: **4**

Cluster 0: 1136

Cluster 1: 2542

Cluster 2: 1863

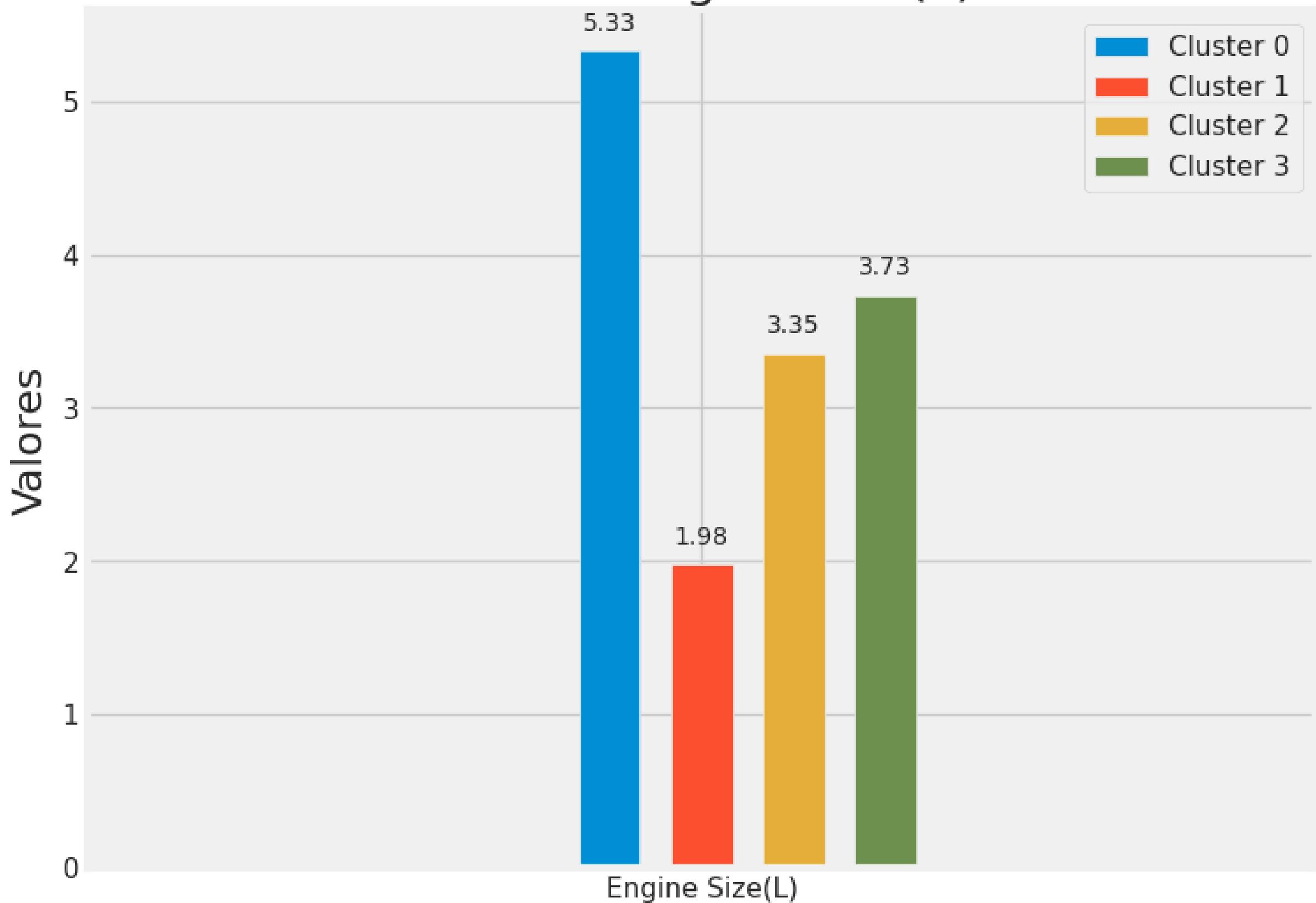
Cluster 3: 450



# K-means

Comparação do tamanho médio dos motores de carros entre os 4 clusters

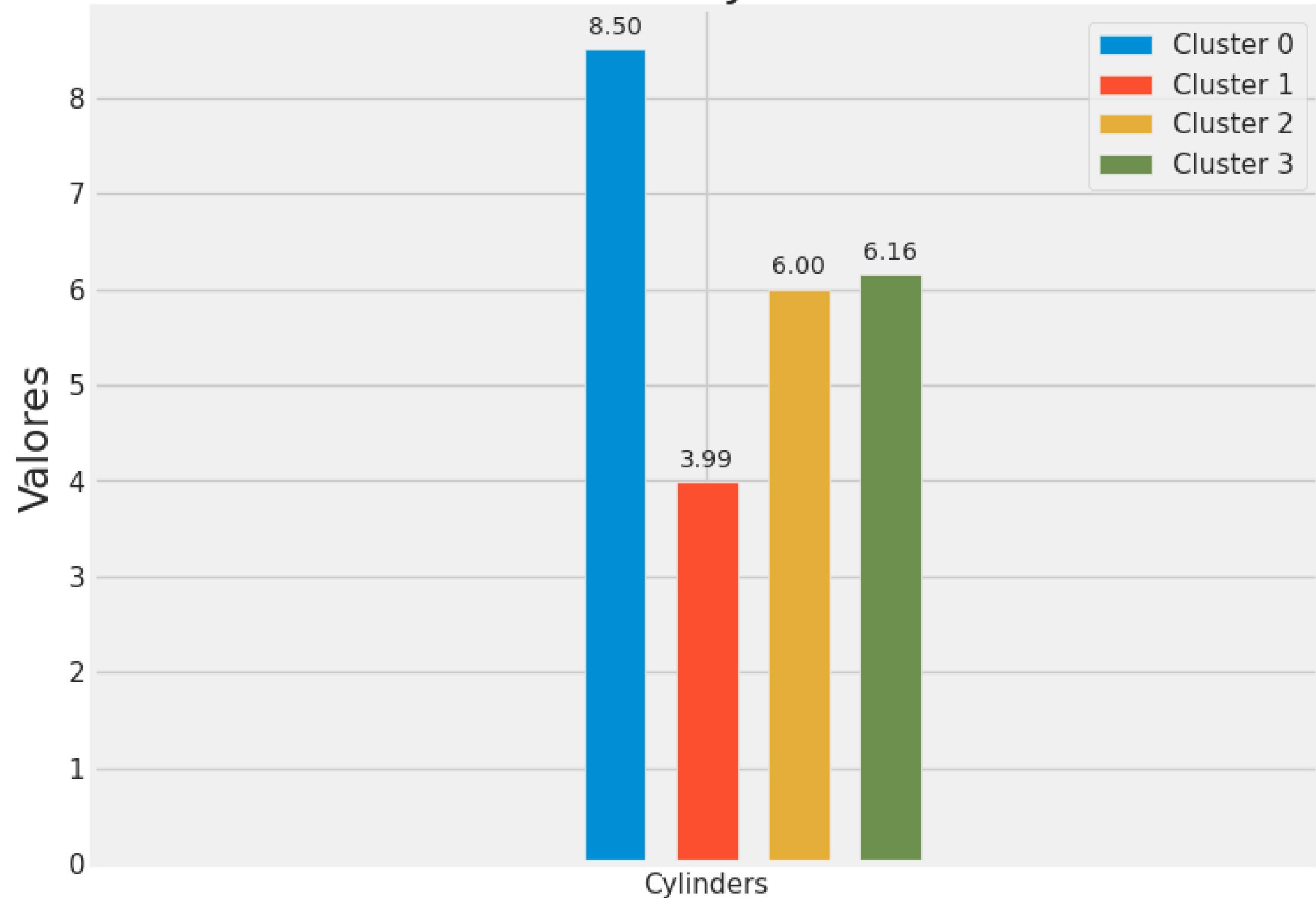
MEAN - Engine Size(L)



# K-means

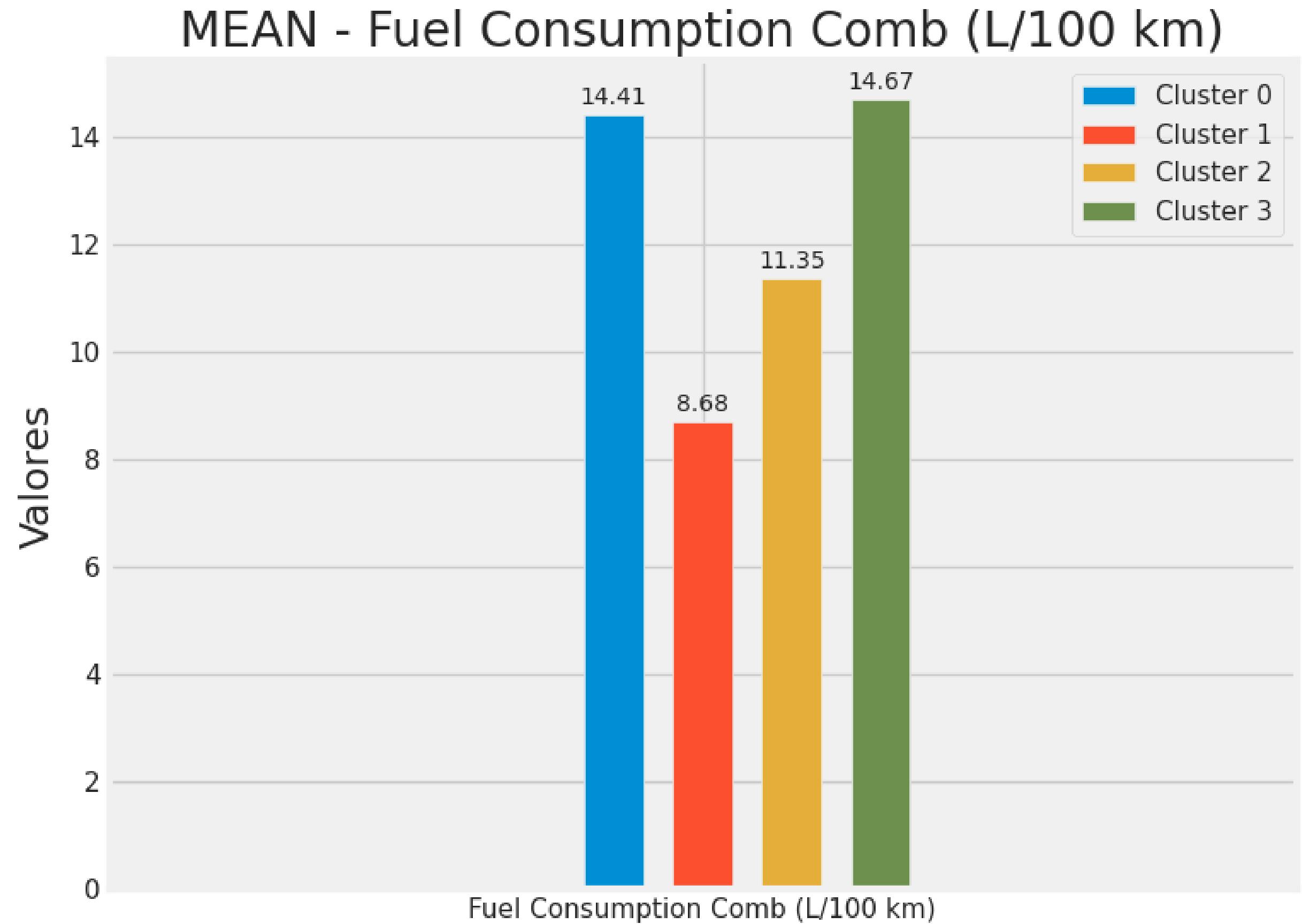
Comparação da  
quantidade média  
de cilindros dos  
carros entre os 4  
clusters

MEAN - Cylinders



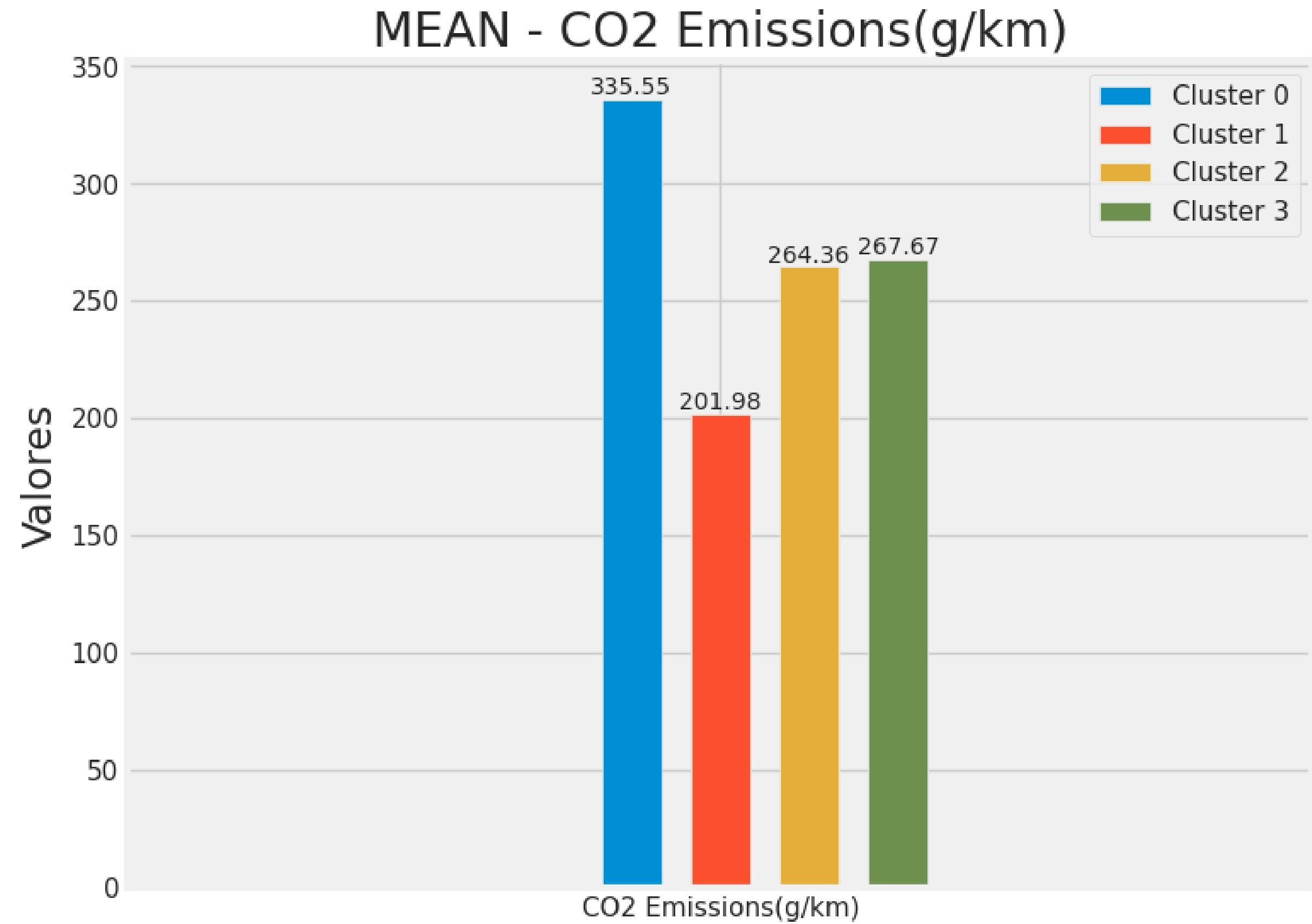
# K-means

Comparação do consumo médio de combustível dos carros entre os 4 clusters



# K-means

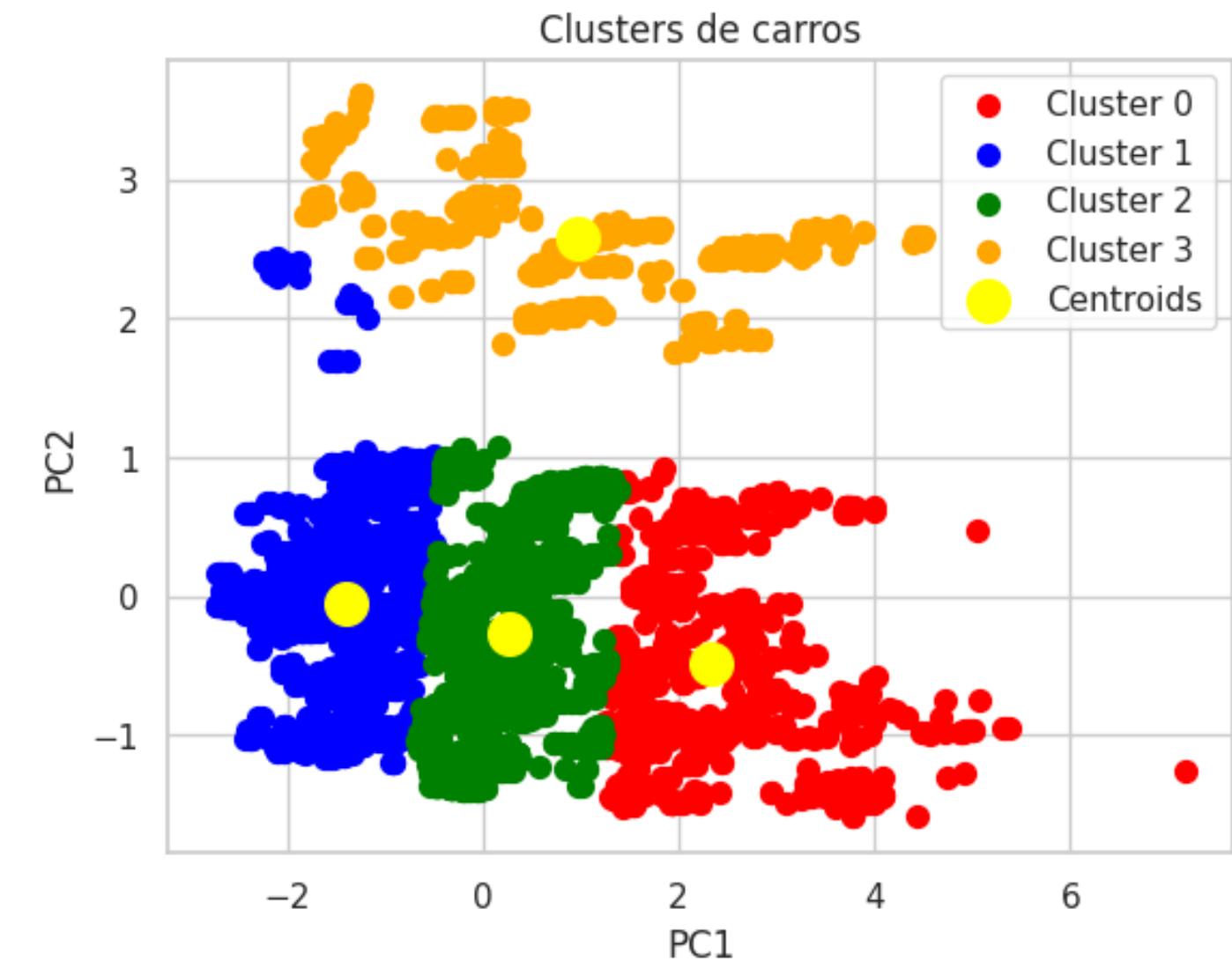
Comparação da emissão de CO<sub>2</sub> dos carros entre os 4 clusters



# K-means

## Resumo cluster 0

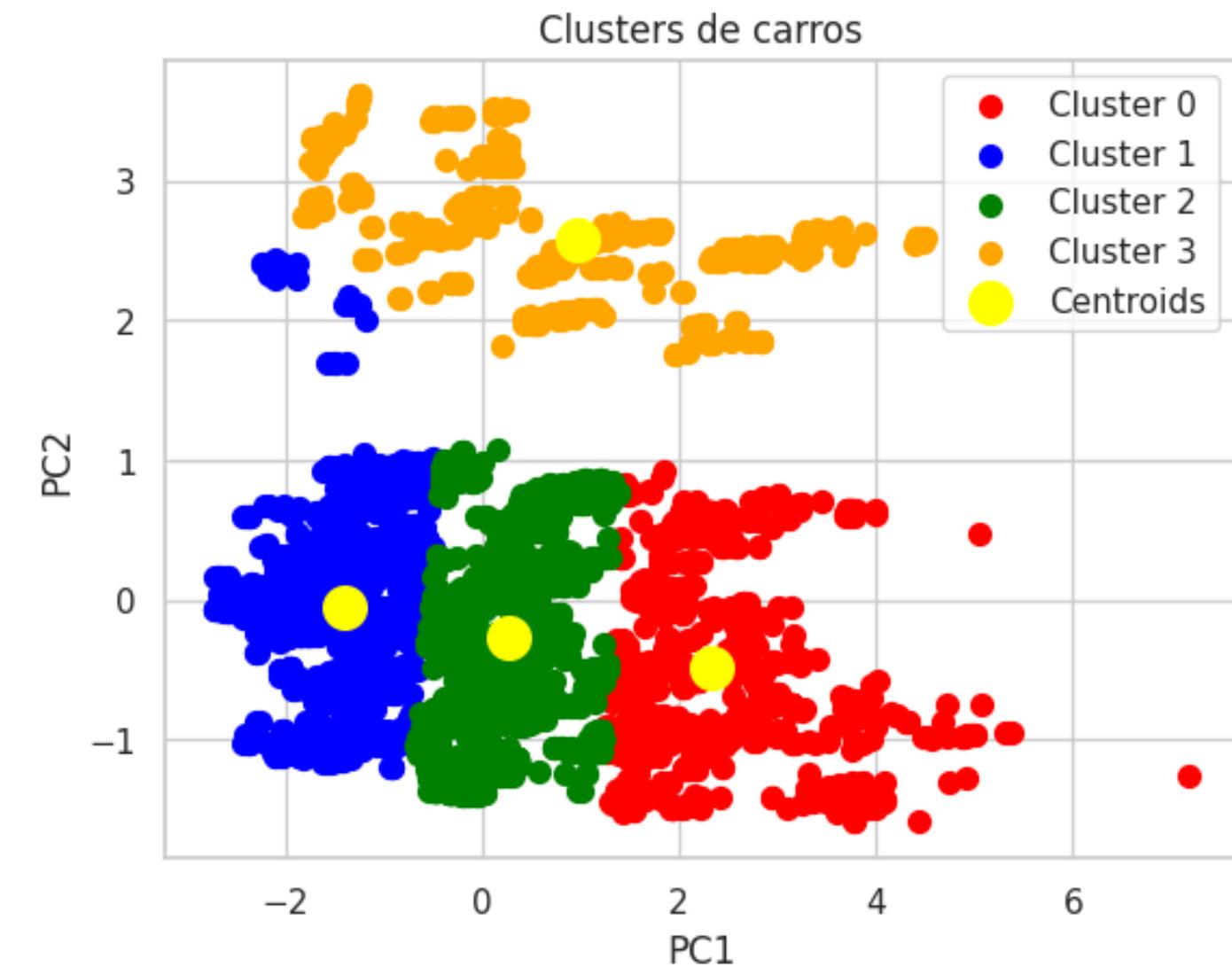
- Tamanho do motor: 5.33
- Cilindros: 8.5
- Consumo de combustível (L/100km): 14.41
- Emissão de CO<sub>2</sub> (g/km): 335.55
- Tipo de combustível: Z (68.2%), X (31.8%)
- Tipo de câmbio: AS8 (21.7%), A6 (15.7%), A8 (14.2%)
- Fabricantes: Chevrolet (13.6%), Mercedes-Benz (13.1%), GMC (10.4%)
- Modelo de carro:
  - **SUV\_Standard** (21.7%),
  - **Pickup\_Truck** (17.2%),
  - **Two-Seater** (14%)



# K-means

## Resumo cluster 1

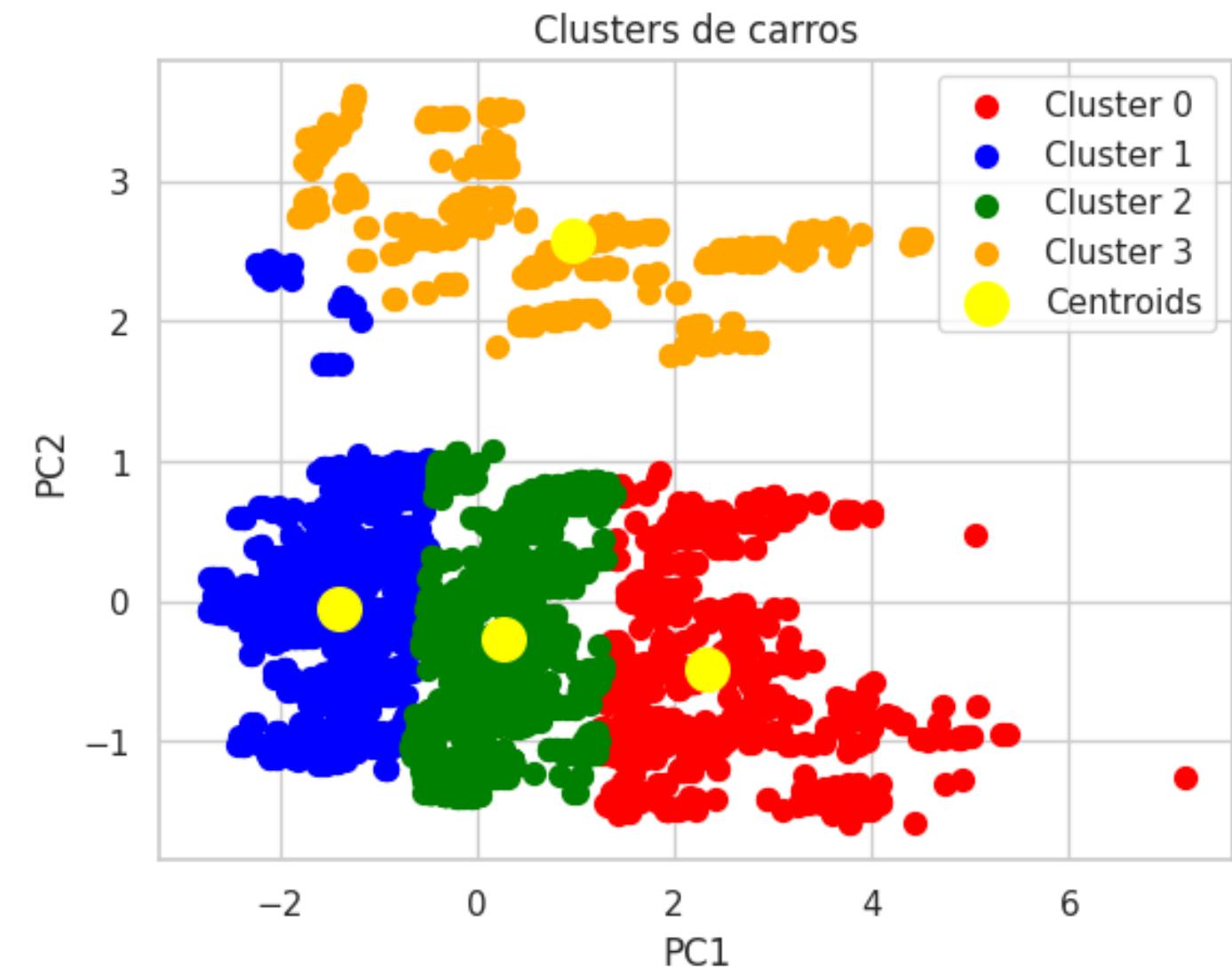
- Tamanho do motor: 1.98
- Cilindros: 3.99
- Consumo de combustível (L/100km): 8.68
- Emissão de CO<sub>2</sub> (g/km): 201.98
- Tipo de combustível: X (65.6%), Z (33.6%)
- Tipo de câmbio: AS6 (24.4%), M6 (19.1%), AS8 (12.9%)
- Fabricantes: Ford (8.2%), Mini (7.4%), Chevrolet (6.8%)
- Modelo de carro:
  - **SUV\_Small (23.6%)**,
  - **Compact (23.5%)**,
  - **Mid-Size (20.4%)**



# K-means

## Resumo cluster 2

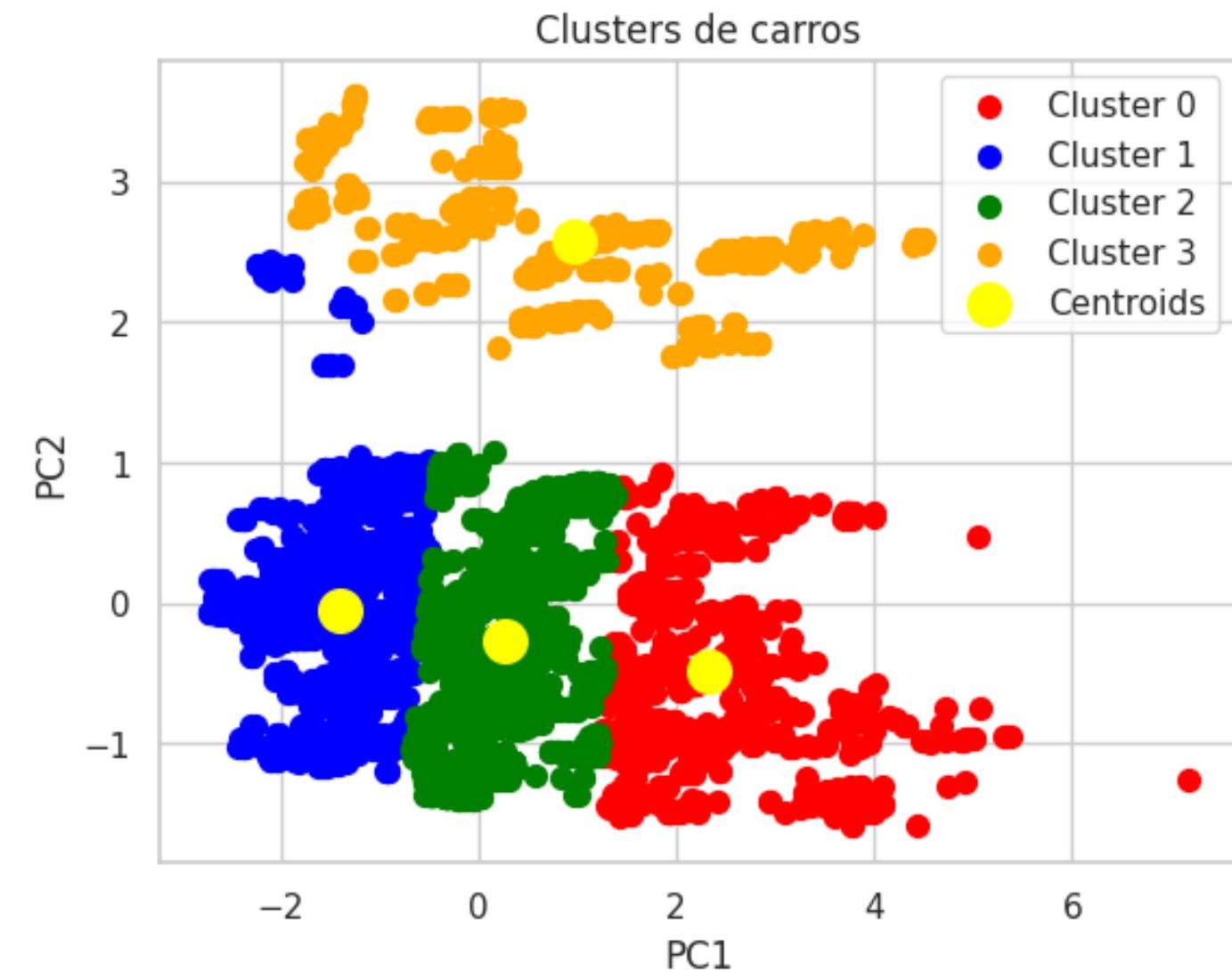
- Tamanho do motor: 3.35
- Cilindros: 6
- Consumo de combustível (L/100km): 11.35
- Emissão de CO<sub>2</sub> (g/km): 264.36
- Tipo de combustível: Z (54.3%), X (45.6%)
- Tipo de câmbio: AS8 (21%), AS6 (17.6%), AM7 (8.8%)
- Fabricantes: BMW (12.8%), Porsche (11.7%), Ford (8.9%)
- Modelo de carro:
  - SUV\_Small (16.4%),
  - Mid-Size (13.7%),
  - SUV Standard (11%)



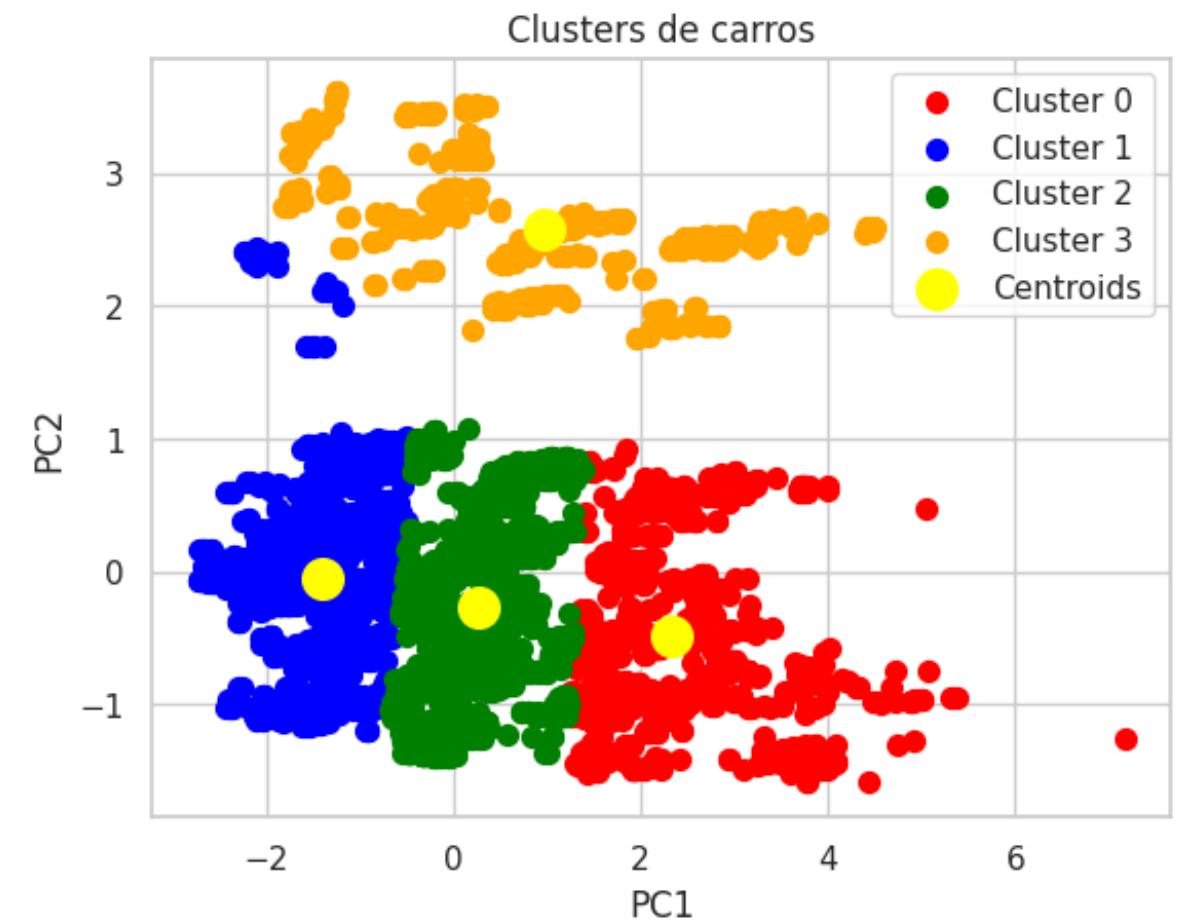
# K-means

## Resumo cluster 3

- Tamanho do motor: 3.73
- Cilindros: 6.16
- Consumo de combustível (L/100km): 14.67
- Emissão de CO<sub>2</sub> (g/km): 267.67
- Tipo de combustível: E (70%), D (30%)
- Tipo de câmbio: A6 (40.2%), AS6 (16.2%), AS8 (11.8%)
- Fabricantes: Ford (24%), Chevrolet (20.1%), GMC (17.8%)
- Modelo de carro:
  - Pickup-Truck Standard (29.8%),
  - SUV Standard (21.8%),
  - SUV Small (10.7%)



### Cluster 3



### Cluster 1



### Cluster 2

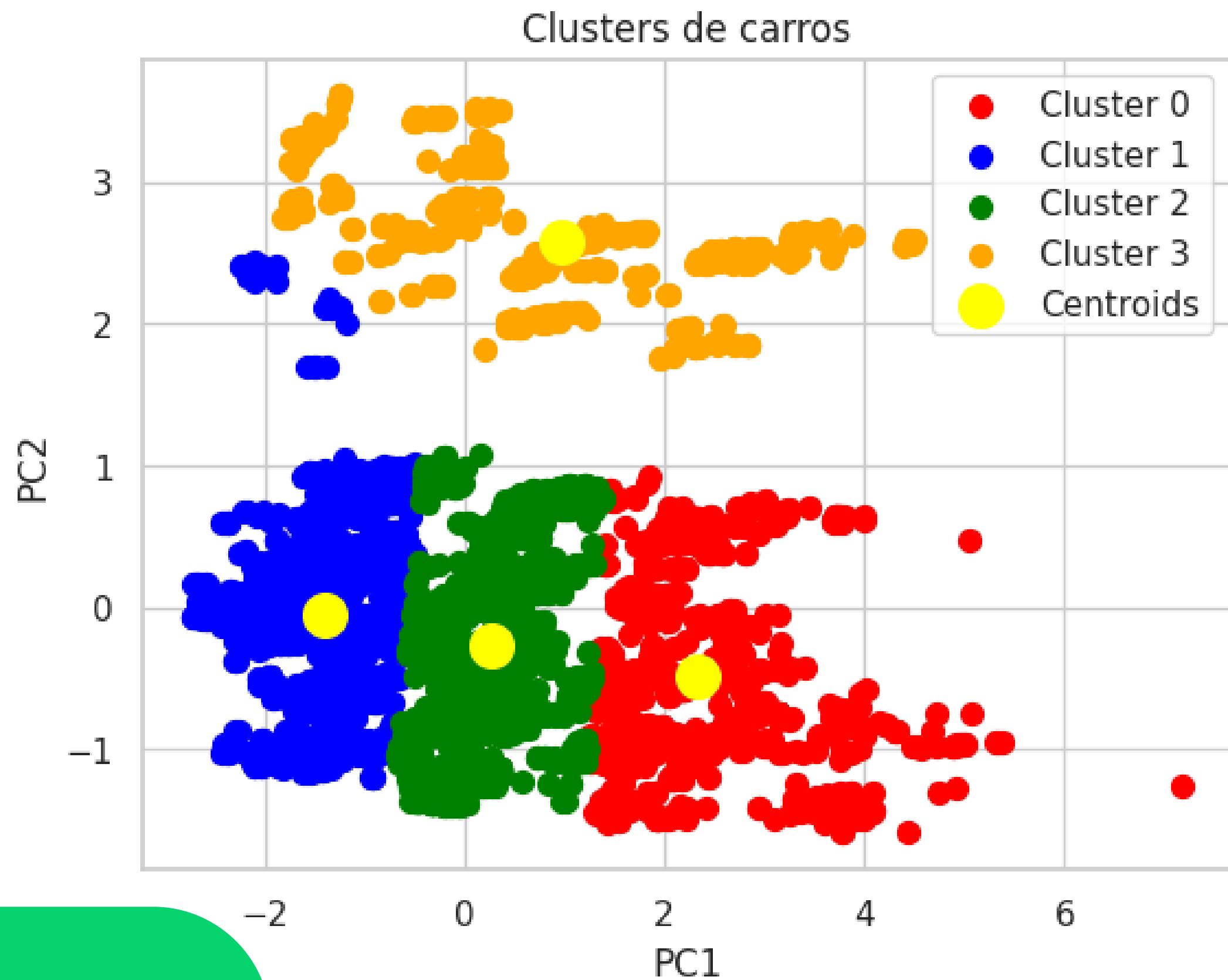


### Cluster 0

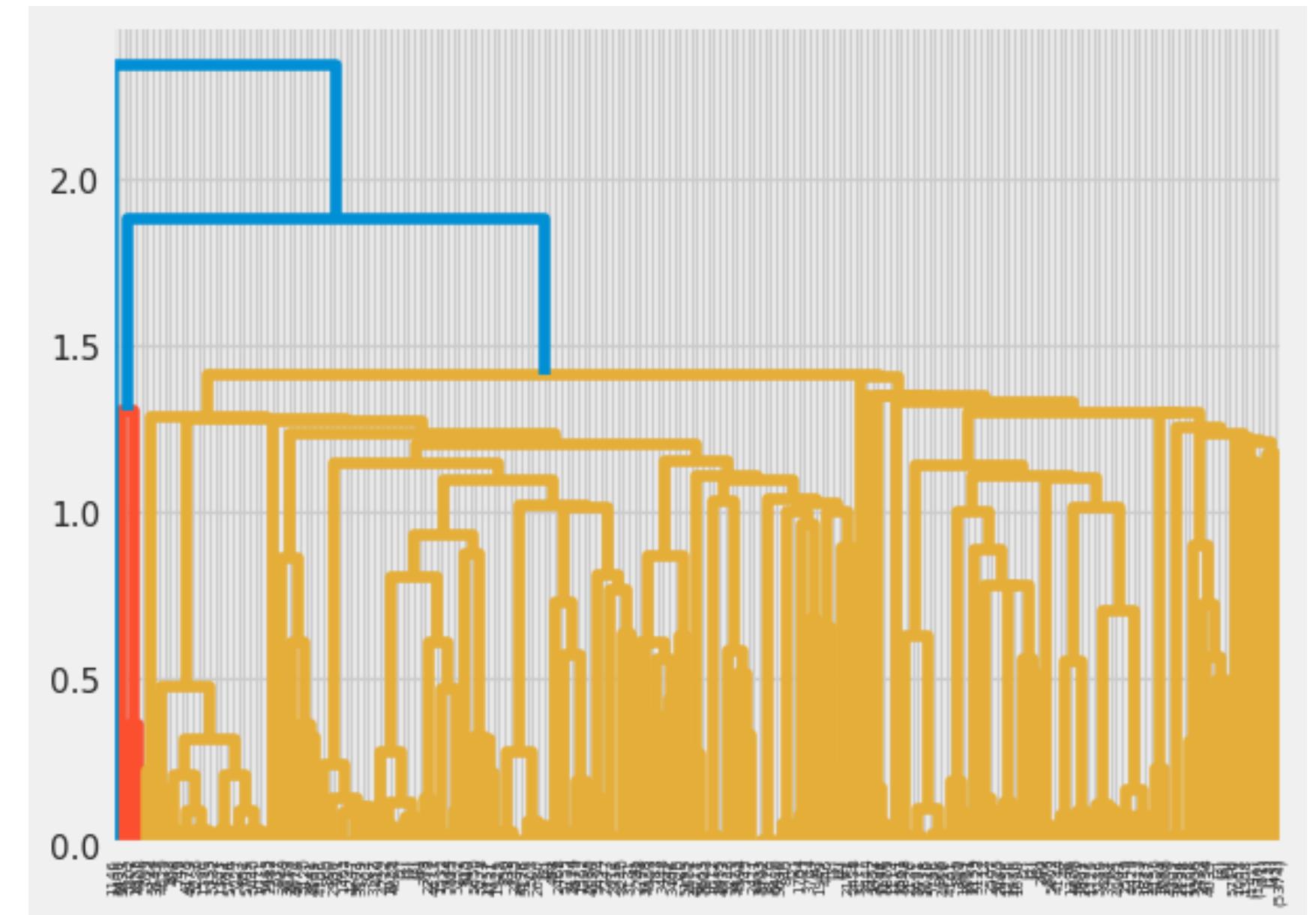
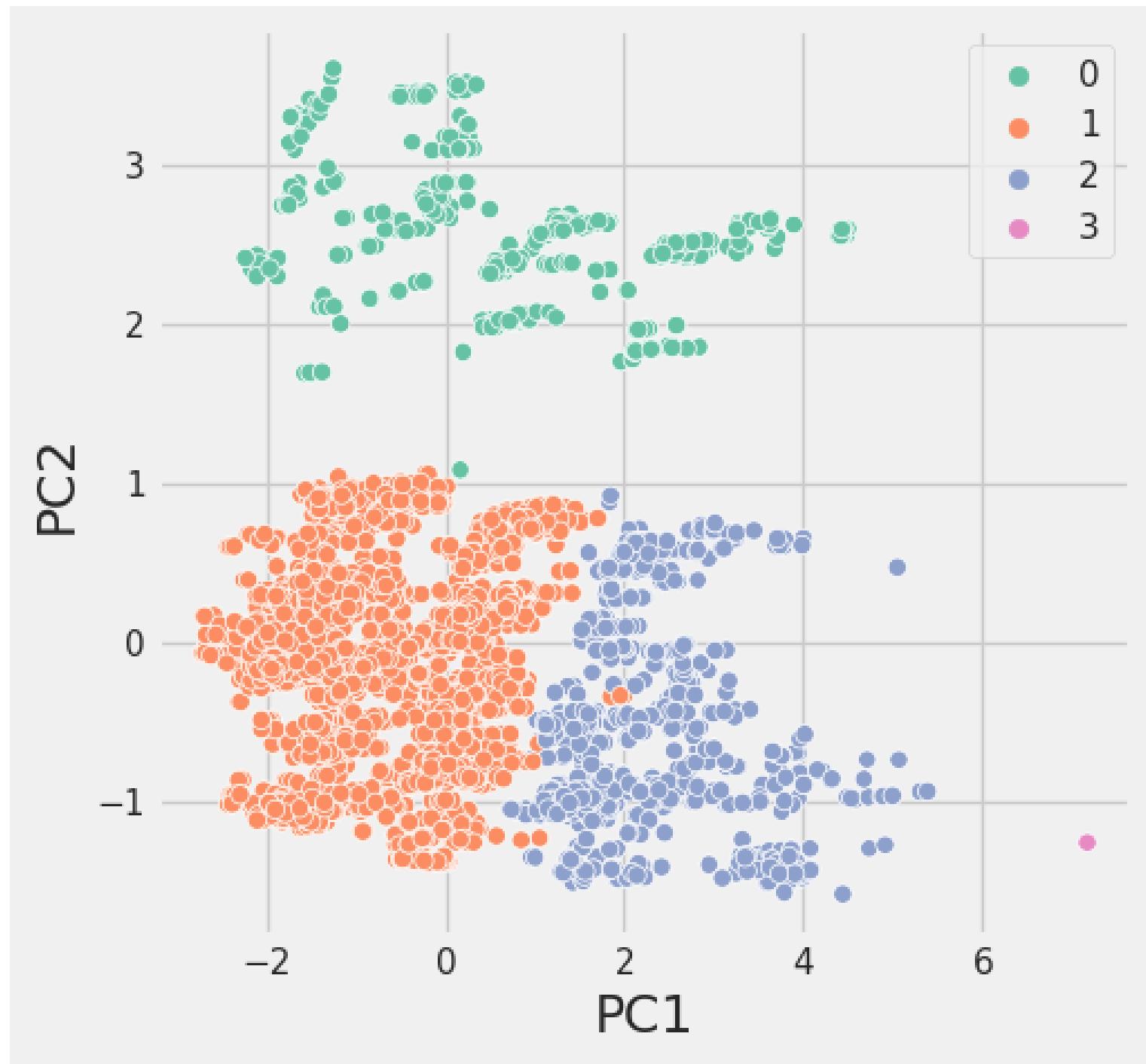


# Validação K-Means

- Silhueta: 0.533
- Davies\_Bouldin: 0.678

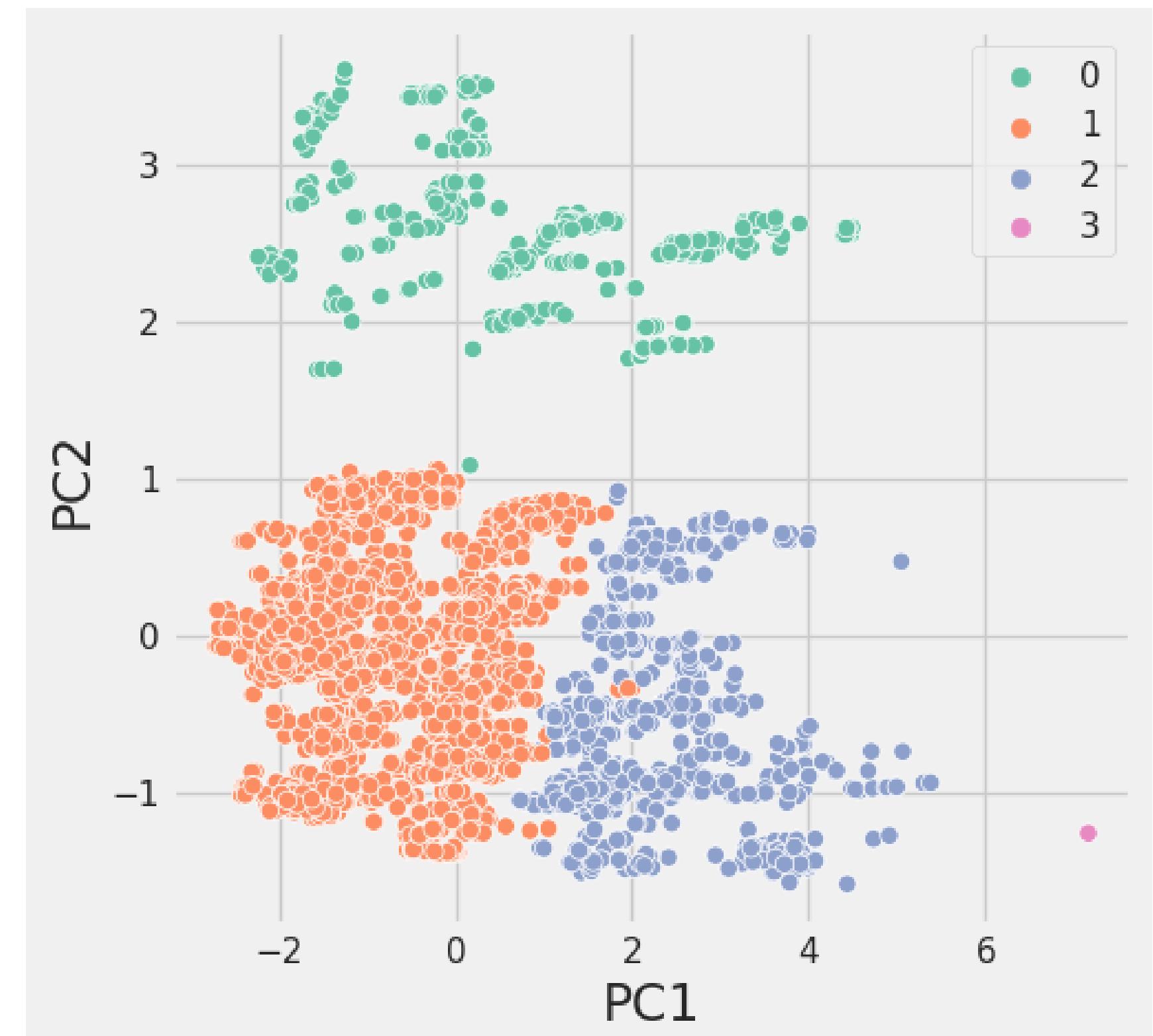


# Hierárquico

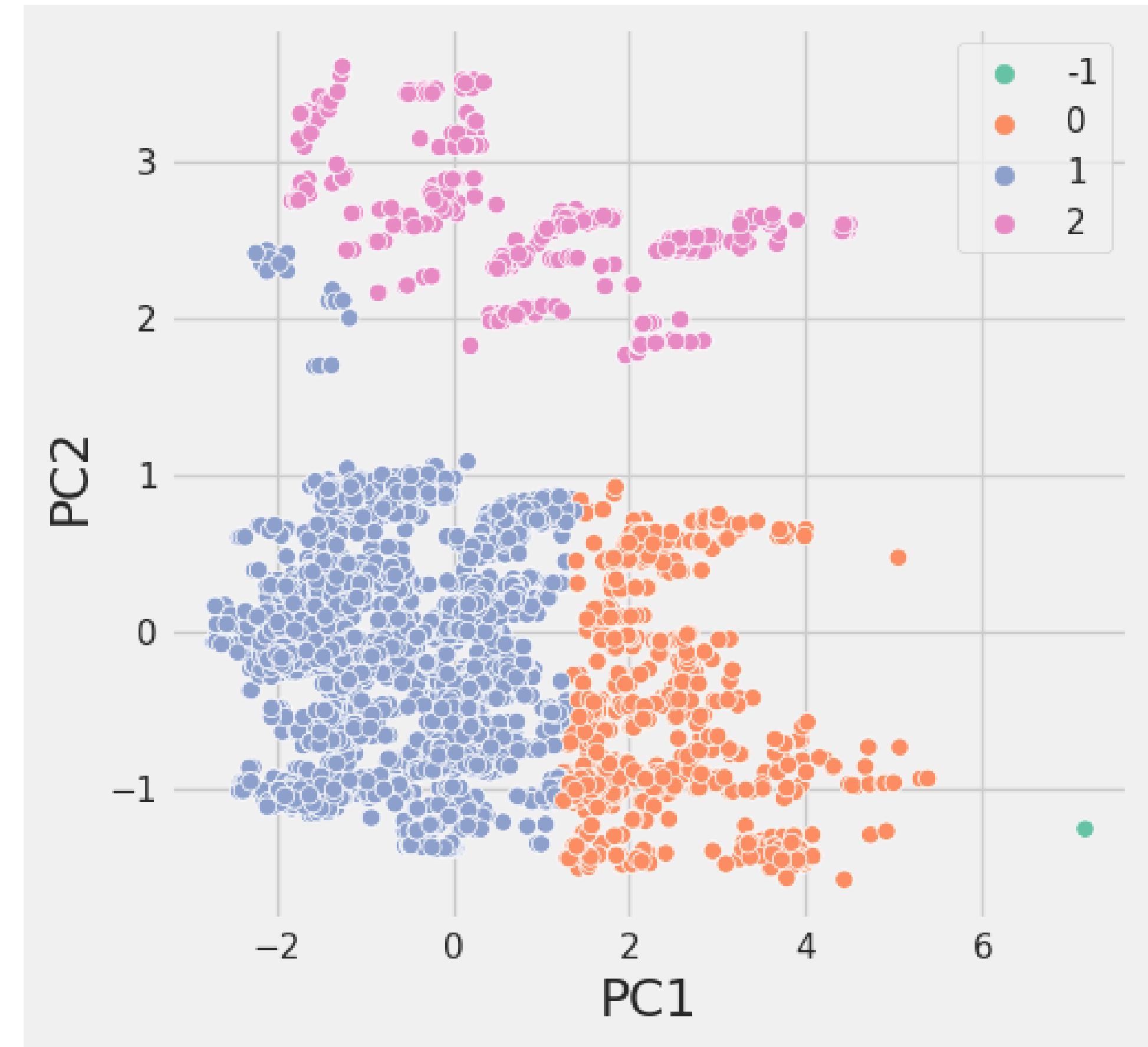


# Validação Hierárquico

- Silhueta: 0.510
- Davies\_Bouldin: 0.589



# Densidade



# Validação Densidade

- Silhueta: 0.526
- Davies\_Bouldin: 0.664

