

# Projeto: Capstone de análise preditiva

## Tarefa 1: Determine formatos de loja para as lojas existentes

### 1. Qual é o número ideal de formatos de loja? Como você chegou a esse número?

R: Com base no relatório K-Means Diagnostic, e índices Adjusted Rand e Calinski-Harabasz, o número ideal de formatos de lojas é 3. Uma vez que ambos os índices registram o maior valor mediano.

#### K-Means Cluster Assessment Report

##### Summary Statistics

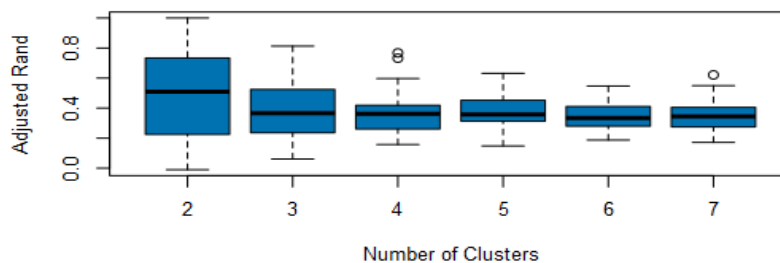
Adjusted Rand Indices:

|              | 2         | 3        | 4        | 5        | 6        | 7        |
|--------------|-----------|----------|----------|----------|----------|----------|
| Minimum      | -0.009746 | 0.061122 | 0.157263 | 0.147999 | 0.187466 | 0.171662 |
| 1st Quartile | 0.242753  | 0.239522 | 0.262007 | 0.313963 | 0.281025 | 0.275832 |
| Median       | 0.509299  | 0.366248 | 0.362164 | 0.358178 | 0.333944 | 0.344909 |
| Mean         | 0.475104  | 0.387225 | 0.368014 | 0.379542 | 0.346238 | 0.351392 |
| 3rd Quartile | 0.724028  | 0.511592 | 0.417343 | 0.451396 | 0.40961  | 0.403465 |
| Maximum      | 1         | 0.813119 | 0.76761  | 0.631759 | 0.547099 | 0.621149 |

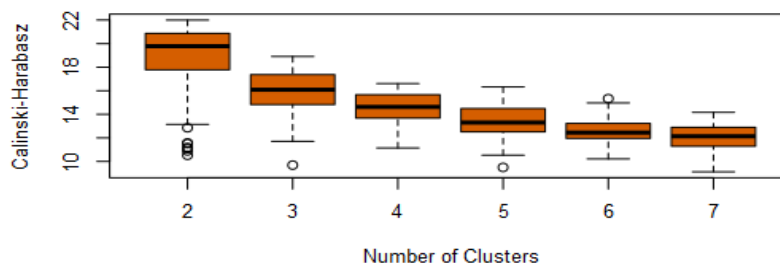
Calinski-Harabasz Indices:

|              | 2        | 3         | 4        | 5        | 6        | 7         |
|--------------|----------|-----------|----------|----------|----------|-----------|
| Minimum      | 10.518   | 9.683921  | 11.14097 | 9.49051  | 10.21393 | 9.120745  |
| 1st Quartile | 17.8029  | 14.837656 | 13.67875 | 12.50456 | 11.95352 | 11.288284 |
| Median       | 19.77924 | 16.086514 | 14.62825 | 13.29649 | 12.44466 | 12.129699 |
| Mean         | 18.82134 | 15.966673 | 14.57619 | 13.3812  | 12.58452 | 12.005648 |
| 3rd Quartile | 20.869   | 17.364245 | 15.64198 | 14.48288 | 13.22522 | 12.887625 |
| Maximum      | 21.99265 | 18.908142 | 16.60523 | 16.32568 | 15.32989 | 14.165491 |

#### Adjusted Rand Indices



#### Calinski-Harabasz Indices



### 2. Quantas lojas enquadram-se em cada formato?

R: O número de loja por formato é representado na imagem abaixo:

| Cluster | Count |
|---------|-------|
| 1       | 23    |
| 2       | 29    |
| 3       | 33    |

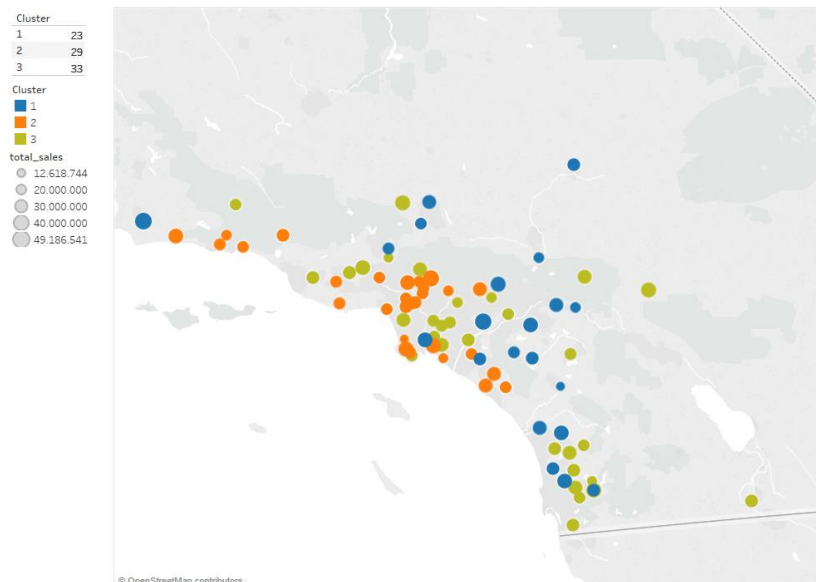
3. Com base nos resultados do modelo de agrupamento, de que forma os *clusters* diferem um do outro?

R: Conforme imagem abaixo, os clusters diferem em percentual total de vendas (Floral\_pct\_of\_total). As lojas o Cluster 1 tem as maiores médias de vendas totais (Avg\_store\_sales) quando comparada as outras 2.

| Cluster | Floral_pct_of_total | Avg_store_sales | Avg_floral_sales |
|---------|---------------------|-----------------|------------------|
| 1       | 0.007573            | 32253841.90     | 244272.97        |
| 2       | 0.010486            | 27472964.45     | 288072.19        |
| 3       | 0.006941            | 28356954.96     | 196828.43        |

4. Envie um dashboard do Tableau (salvo como um arquivo público do Tableau) que mostre a localização das lojas e utilize cores para mostrar os *clusters* e tamanhos para mostrar as vendas totais.

Lojas por formato



[https://public.tableau.com/profile/ortega.diego#!/vizhome/Tarefa1\\_15604780197720/Painel](https://public.tableau.com/profile/ortega.diego#!/vizhome/Tarefa1_15604780197720/Painel)

## Tarefa 2: Formato das lojas novas

1. Qual metodologia você usou para prever o melhor formato para as lojas novas?

R: Eu usei o modelo Boosted. Enquanto a acurácia dos três modelos é idêntica, o valor F1 do modelo Boosted é superior. Veja imagem abaixo:

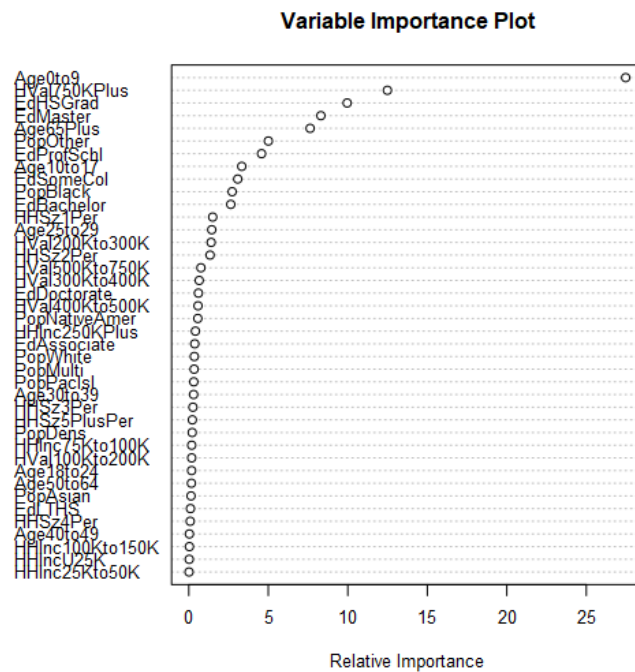
### Model Comparison Report

| Fit and error measures |          |        |            |            |            |
|------------------------|----------|--------|------------|------------|------------|
| Model                  | Accuracy | F1     | Accuracy_1 | Accuracy_2 | Accuracy_3 |
| TREE                   | 0.8235   | 0.8251 | 0.7500     | 0.8000     | 0.8750     |
| FOREST                 | 0.8235   | 0.8251 | 0.7500     | 0.8000     | 0.8750     |
| BOOST                  | 0.8235   | 0.8543 | 0.8000     | 0.6667     | 1.0000     |

2. Quais são as três variáveis mais importantes que ajudam a explicar a relação entre os indicadores demográficos e o formato das lojas?

**R:** As variáveis mais importantes são:

- Ave0to9
- HVal750KPlus
- EdHSGrad



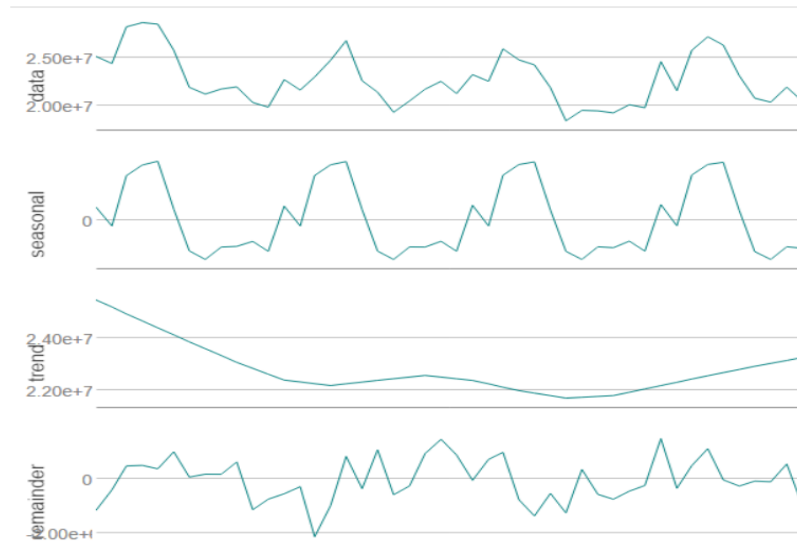
3. Em que formato cada uma das 10 lojas novas se enquadra? Preencha a tabela abaixo:

| Número da loja | Segmento |
|----------------|----------|
| S0086          | 1        |
| S0087          | 2        |
| S0088          | 3        |
| S0089          | 2        |
| S0090          | 2        |
| S0091          | 1        |
| S0092          | 2        |
| S0093          | 1        |
| S0094          | 2        |
| S0095          | 2        |

## Tarefa 3: Prevendo a vendas de produtos

1. Qual tipo de modelo, ETS ou ARIMA, você usou para cada previsão? Use a notação ETS (a, m, n) ou ARIMA (ar, i, ma). Como você chegou a essa decisão?

**R:** Escolhi o modelo ETS (M,N,M). Os gráficos a seguir mostram que o erro é multiplicativo, não há tendência e a sazonalidade é multiplicativa, e é por isso que escolhi o modelo ETS (M,N,M).



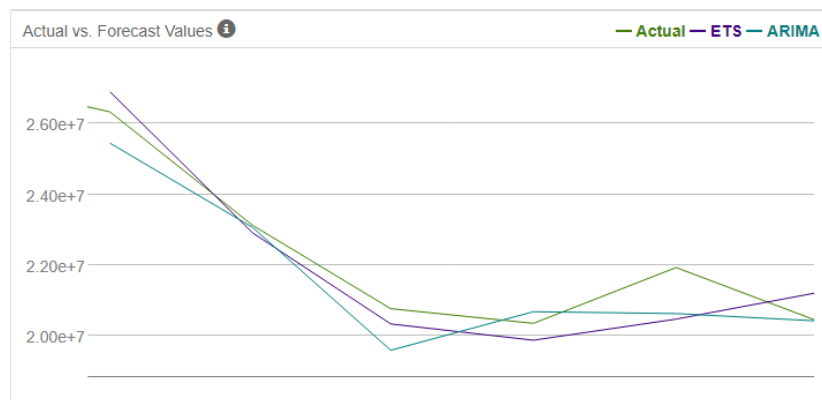
Os seguintes valores de previsão e modelos de precisão são o que eu usei para determinar que o modelo ETS foi o melhor modelo para prover as vendas de produtos.

### Actual and Forecast Values:

| Actual      | ETS            | ARIMA          |
|-------------|----------------|----------------|
| 26338477.15 | 26907095.61191 | 25454225.03787 |
| 23130626.6  | 22916903.07434 | 23071096.30787 |
| 20774415.93 | 20342618.32222 | 19598371.02787 |
| 20359980.58 | 19883092.31778 | 20688679.39787 |
| 21936906.81 | 20479210.4317  | 20635860.61787 |
| 20462899.3  | 21211420.14022 | 20431492.19787 |

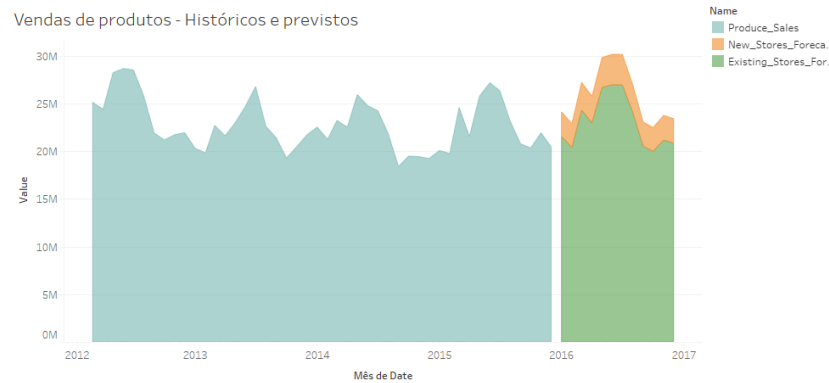
### Accuracy Measures:

| Model | ME       | RMSE     | MAE      | MPE    | MAPE   | MASE   |
|-------|----------|----------|----------|--------|--------|--------|
| ETS   | 210494.4 | 760267.3 | 649540.8 | 1.0288 | 2.9678 | 0.3822 |
| ARIMA | 520597   | 813457.4 | 630163.2 | 2.2909 | 2.8291 | 0.3708 |



2. Envie um dashboard do Tableau (salvo como um arquivo público do Tableau) que inclua uma tabela e um gráfico das três previsões mensais; um para as existentes, um para as novas e um para todas as lojas. Nomeie a aba no arquivo "Tarefa 3" do Tableau.

**R:** Na tabela abaixo é apresentado a previsão de vendas para as novas lojas e lojas existentes e o total de cada mês. Também é apresentado um gráfico de áreas empilhadas que representa as vendas de produtos históricos com previsões de vendas de produtos para as novas lojas e lojas existentes.



Tabela

| Year_Month | New_Stores_Forecast | Existing_Stores_Forecast | Total_Produce_Forecast |
|------------|---------------------|--------------------------|------------------------|
| 2016-01    | 2.587.451           | 21.539.936               | 24.127.387             |
| 2016-02    | 2.477.353           | 20.413.771               | 22.891.123             |
| 2016-03    | 2.913.185           | 24.325.953               | 27.239.138             |
| 2016-04    | 2.775.746           | 22.993.466               | 25.769.212             |
| 2016-05    | 3.150.867           | 26.691.951               | 29.842.818             |
| 2016-06    | 3.188.922           | 26.989.964               | 30.178.886             |
| 2016-07    | 3.214.746           | 26.948.631               | 30.163.376             |
| 2016-08    | 2.866.349           | 24.091.579               | 26.957.928             |
| 2016-09    | 2.538.727           | 20.523.492               | 23.062.219             |
| 2016-10    | 2.488.148           | 20.011.749               | 22.499.897             |
| 2016-11    | 2.595.270           | 21.177.435               | 23.772.706             |
| 2016-12    | 2.573.397           | 20.855.799               | 23.429.196             |

[https://public.tableau.com/profile/ortega.diego#!/vizhome/Tarefa3\\_15602229735920/Tarefa3?publish=yes](https://public.tableau.com/profile/ortega.diego#!/vizhome/Tarefa3_15602229735920/Tarefa3?publish=yes)