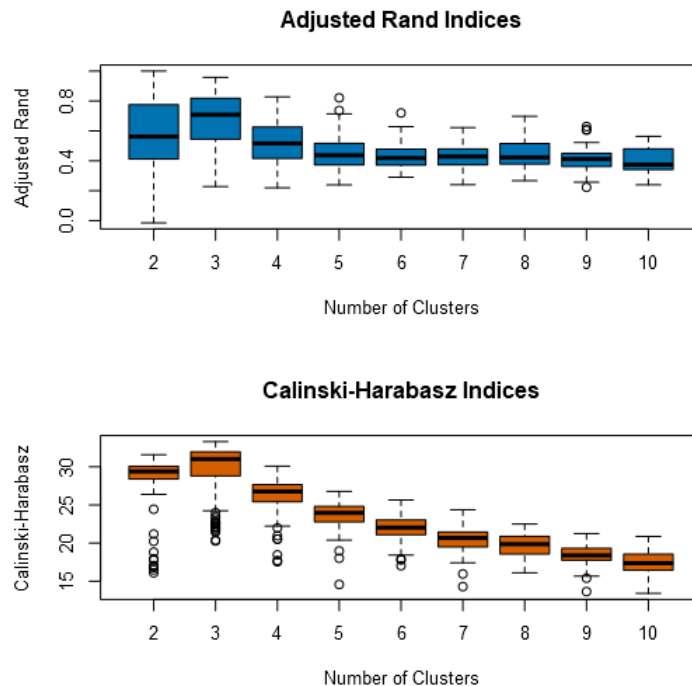


Projeto: Capstone de análise preditiva

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?

Inicialmente foi realizado o cálculo da importância das categorias para cada loja. Depois do cálculo, foi rodado um *K-Centroids Diagnostics* para determinar o número ideal de grupos. O resultado pode ser visto abaixo.



Pelo que observamos acima, a opção com 3 grupos aparenta ser a mais indicada, apresentando maior estabilidade (maior mediana Rand) e compacidade / separação (maior mediana CH).

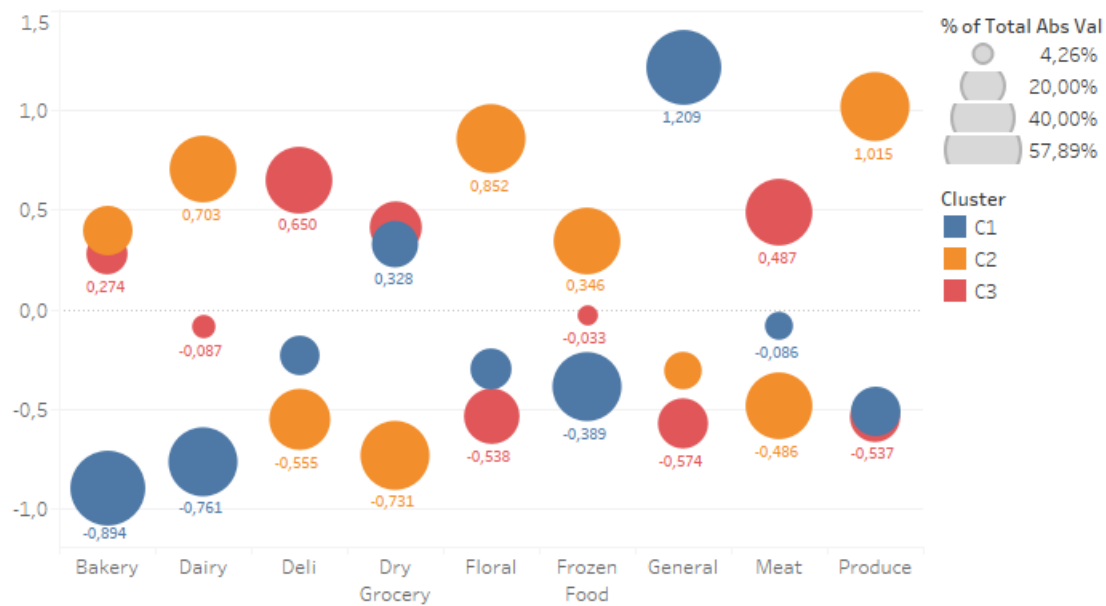
2. Quantas lojas enquadram-se em cada formato?

Cluster Information:

Cluster	Size	Ave Distance	Max Distance	Separation
1	23	2.320539	3.55145	1.874243
2	29	2.540086	4.475132	2.118708
3	33	2.115045	4.9262	1.702843

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

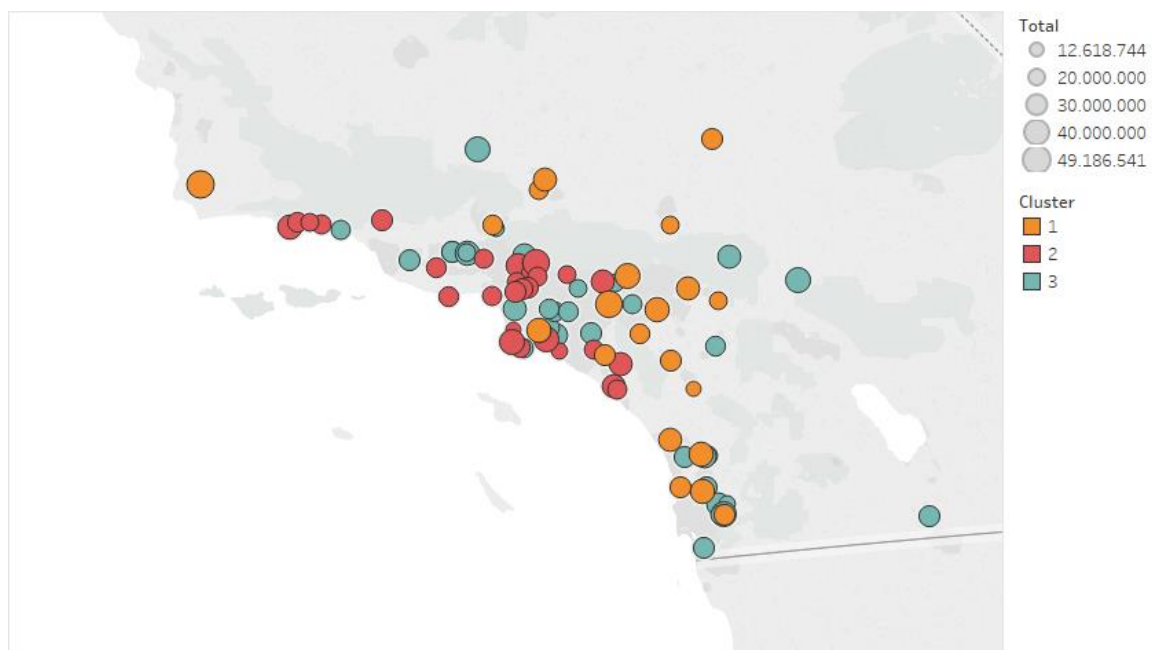
	Sum_Dry_Grocery	Sum_Dairy	Sum_Frozen_Food	Sum_Meat	Sum_Produce	Sum_Floral	Sum_Deli
1	0.327833	-0.761016	-0.389209	-0.086176	-0.509185	-0.301524	-0.23259
2	-0.730732	0.702609	0.345898	-0.485804	1.014507	0.851718	-0.554641
3	0.413669	-0.087039	-0.032704	0.48698	-0.53665	-0.538327	0.64952
	Sum_Bakery	Sum_General_Merchandise					
1	-0.894261	1.208516					
2	0.396923	-0.304862					
3	0.274462	-0.574389					



Tamanho representa importância percentual de um grupo para a soma dos absolutos de uma categoria.

- Grande distância entre o Cluster 2 e os outros na categoria *Produce*
- Grande distância entre o Cluster 1 e os outros na categoria *General Merchandise*
- Grande distância entre o Cluster 3 e os outros na categoria *Bakery*

- 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.



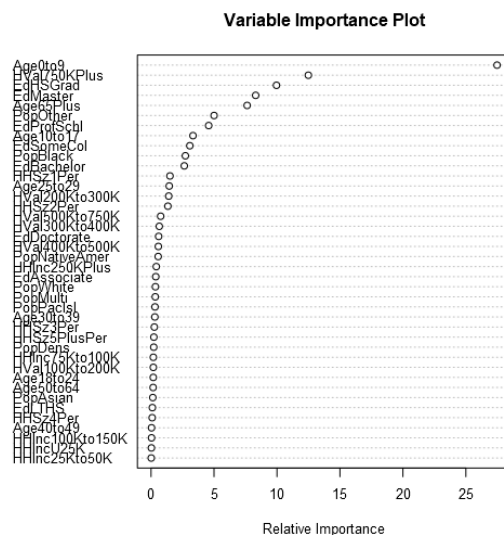
Tarefa 2: Formato das lojas novas

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

Apesar de o Modelo de Floresta e o Modelo Boosted apresentarem a mesma acurácia geral, o Modelo Boosted acaba se mostrando como melhor por apresentar um escore F1 maior e possuir maior acurácia nos grupos com maior rendimento médio.

Model Comparison Report					
Fit and error measures					
Model	Accuracy	F1	Accuracy_1	Accuracy_2	Accuracy_3
FM	0.8235	0.8251	0.7500	0.8000	0.8750
BM	0.8235	0.8543	0.8000	0.6667	1.0000
DT	0.7059	0.7327	0.6000	0.6667	0.8333

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



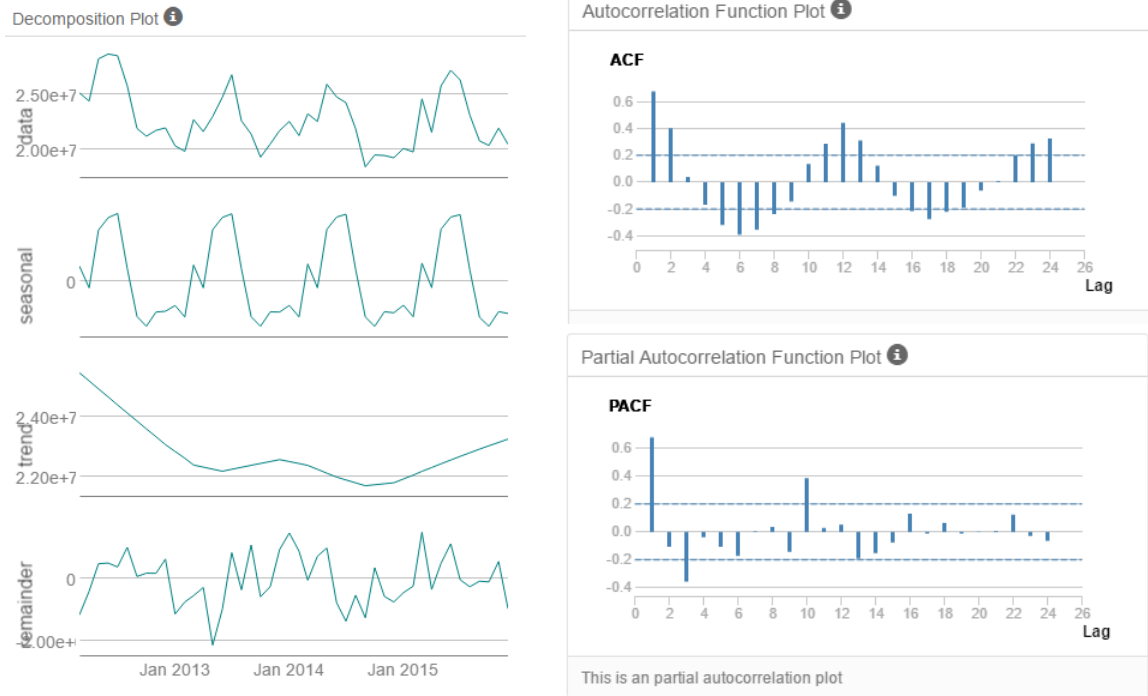
1. Age0to9
2. HVal750kPlus
3. EdHSGrad

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

Nº 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

- 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?



Observando os gráficos de decomposição e de autocorrelação acima, percebemos que a série apresenta sazonalidade com ligeiro decremento, não possui tendência aparente e erro inconstante. Com isso em mente, foram testados diversos modelos ETS e ARIMA, ficando entre os 3 abaixo.

Modelo	RMSE	MASE	AIC
ETS (M,N,M)	760,267	0.38	1,283
ARIMA (1,1,0)(1,1,0)[12]	1,935,636	1.06	848
ARIMA (1,0,0)(1,1,0)[12]	1,050,239	0.55	880

Com base no resultado da validação contra a amostra de retenção, podemos concluir que o modelo **ETS (M,N,M)** se mostra como a melhor opção para o caso por apresentar o maior ganho em relação ao método *naive* e possuir menor desvio padrão empírico.

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.

Mês	Novas	Existentes
jan/16	2,587,450.85	21,539,936.01
fev/16	2,477,352.89	20,413,770.60
mar/16	2,913,185.24	24,325,953.10
abr/16	2,775,745.61	22,993,466.35
mai/16	3,150,866.84	26,691,951.42
jun/16	3,188,922.34	26,989,964.01
jul/16	3,214,745.65	26,948,630.76
ago/16	2,866,348.66	24,091,579.35
set/16	2,538,726.85	20,523,492.41
out/16	2,488,148.29	20,011,748.67
nov/16	2,595,270.39	21,177,435.49
dez/16	2,573,396.63	20,855,799.11

