

xpto Analytics



```
...mirror_mod = modifier_ob.  
...mirror object to mirror  
...mirror_mod.mirror_object =  
...operation == "MIRROR_X":  
...mirror_mod.use_x = True  
...mirror_mod.use_y = False  
...mirror_mod.use_z = False  
...operation == "MIRROR_Y":  
...mirror_mod.use_x = False  
...mirror_mod.use_y = True  
...mirror_mod.use_z = False  
...operation == "MIRROR_Z":  
...mirror_mod.use_x = False  
...mirror_mod.use_y = False  
...mirror_mod.use_z = True  
...selection at the end -add  
...mirror_ob.select= 1  
...modifier_ob.select=1  
...context.scene.objects.active  
...("Selected" + str(modifier_ob.  
...mirror_ob.select = 0  
...= bpy.context.selected_object  
...data.objects[one.name].select  
...print("please select exactly  
...  
...-- OPERATOR CLASSES --  
...  
...types.Operator):  
...X mirror to the selected  
...object.mirror_mirror_x"  
...mirror X"  
...  
...context):  
...context.active_object is not
```



“ Você verá aqui, **trechos** de uma análise técnica onde aplica-se ciência de dados para diagnosticar um problema empresarial. Somente trechos permitidos pelo cliente e com os devidos cortes estão sendo publicados, com intuito de possibilitar avaliação curricular do cientista de dados.



“ Explorar e desenvolver a melhor versão que sua empresa pode alcançar desafiando o senso comum , responde ao porque do nosso trabalho. Data Science é a maneira como fazemos isso.

Luiz Almeida Júnior

Cientista de dados | Black Belt Lean & Six-Sigma

(95) 98114-6319

luiz.almeida.jr@hotmail.com

<https://www.linkedin.com/in/luiz-almeida-datascience>

Índice

Diagnóstico em camadas: afinal, qual o problema?	<u>04</u>
Olhando para frente: Vendas no futuro	<u>08</u>
Aconselhamento com uma I.A.	<u>18</u>
Menos é mais	<u>21</u>
Resumindo	<u>25</u>
Próximos passos.....	<u>27</u>





**Diagnóstico em camadas:
afinal, qual o problema?**

```
..._mod = modifier_ob.  
... mirror object to mirror  
... mirror_mod.mirror_object =  
... operation == "MIRROR_X":  
... mirror_mod.use_x = True  
... mirror_mod.use_y = False  
... mirror_mod.use_z = False  
... operation == "MIRROR_Y":  
... mirror_mod.use_x = False  
... mirror_mod.use_y = True  
... mirror_mod.use_z = False  
... operation == "MIRROR_Z":  
... mirror_mod.use_x = False  
... mirror_mod.use_y = False  
... mirror_mod.use_z = True  
... selection at the end -add  
... mirror_ob.select= 1  
... mirror_ob.select=1  
... context.scene.objects.active  
... ("Selected" + str(modifier_ob.  
... mirror_ob.select = 0  
... = bpy.context.selected_object  
... data.objects[one.name].select  
... print("please select exactly  
...  
... OPERATOR CLASSES ...  
... types.Operator):  
... X mirror to the selected  
... object.mirror_mirror_x"  
... mirror X"  
... context):  
... context.active_object is not
```

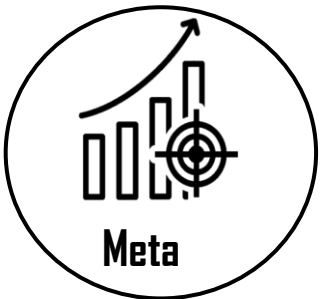
Diagnóstico em camadas: afinal, qual o problema?



“ Uma das fases mais delicadas do projeto foi identificar o problema a ser enfrentado. Uma vez errado o diagnóstico, todo projeto fica comprometido.



O pedido inicial era para ajudar na projeção das receitas para o ano 20xz. Os valores não faziam sentido.



Cavando um pouco mais descobri que o orçamento estava atrelado a projeção da receita. Portanto estava comprometido.

Naturalmente a realização orçamentária estava atrelada a conquista de uma meta para a receita, na ponta final dessa cadeia, o resultado final da empresa, seu lucro.



**Olhando para trás:
Comportamento das vendas**



“ Explorar a base de dados é fundamental para responder vários questionamentos, inclusive se há presença de **outliers**.

Outliers são valores muito diferente dos demais, gerando desconfiança sobre sua origem.

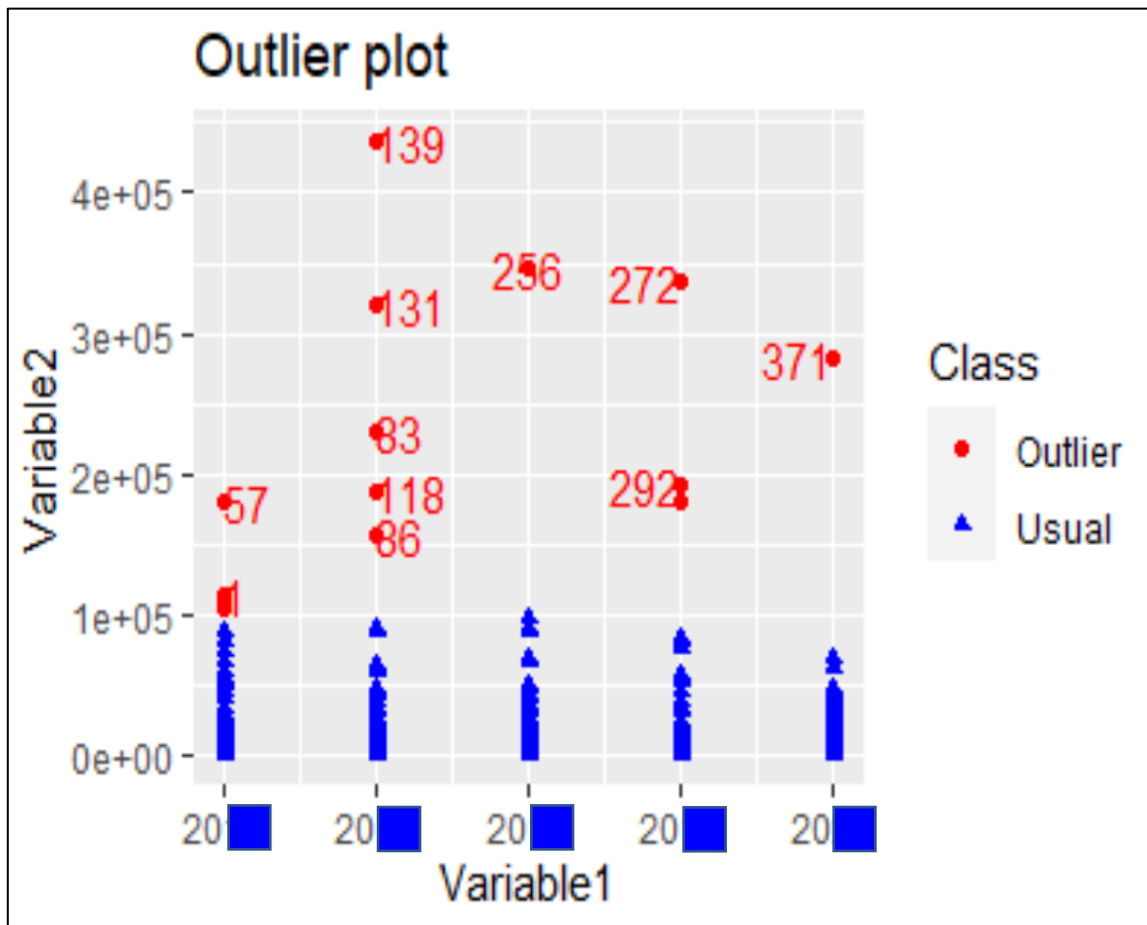


Gráfico 1 – Outliers das vendas distribuídas por ano

0%	25%	50%	75%	100%
145,92	2.560,00	5.655,00	15.077,40	436.040,00

Quadro 1 – Distribuição das vendas por quartis

Analizados os eventos de venda entre os anos de 20xy e 20xz, percebemos que:

1. As vendas consideradas normais ficam em torno de 100 mil reais.
2. Outliers **não** são exceções como se espera.

Constatamos pelo Quadro 1 uma alta quantidade de outliers, pois 75% das vendas tem valor até R\$ 15.077,40.

A distribuição dos dados se mostra assimétrica a esquerda, com uma longa calda a direita, semelhante a uma distribuição logarítmica.

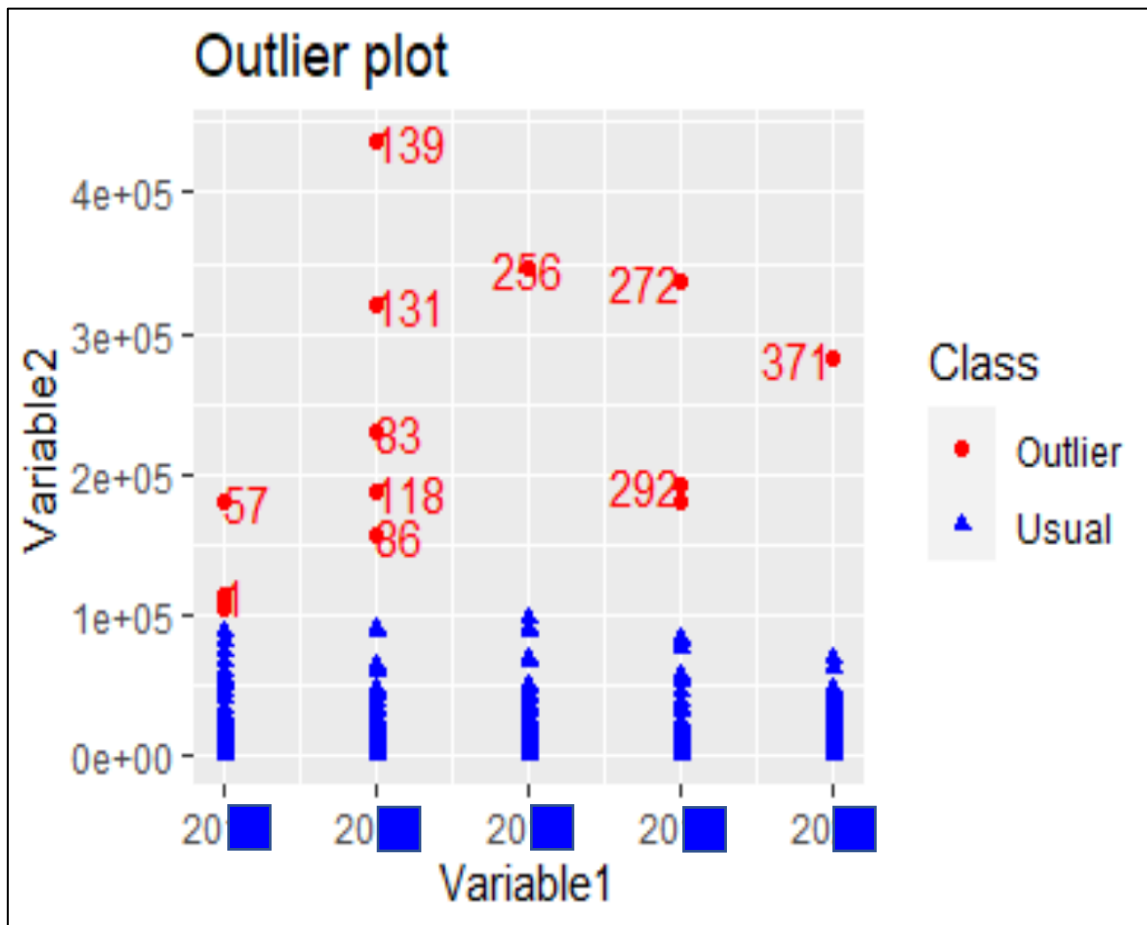


Gráfico 1 – Outliers das vendas distribuídas por ano

0%	25%	50%	75%	100%
145,92	2.560,00	5.655,00	15.077,40	436.040,00

Quadro 1 – Distribuição das vendas por quartis

Seguindo a lógica de Ishikawa, a falta de homogeneidade nas vendas pode ocorrer em função de:

- falta de padronização no processo de vendas,
- problemas com representatividade das métricas utilizadas para medir o processo,
- fonte de dados não confiáveis e/ou
- falta de treinamento da equipe de vendas.



**Olhando para frente:
Vendas no futuro**



“ Para o ano de 20xz foi proposto pela alta direção um incremento de 12% sobre as vendas de 20xy.

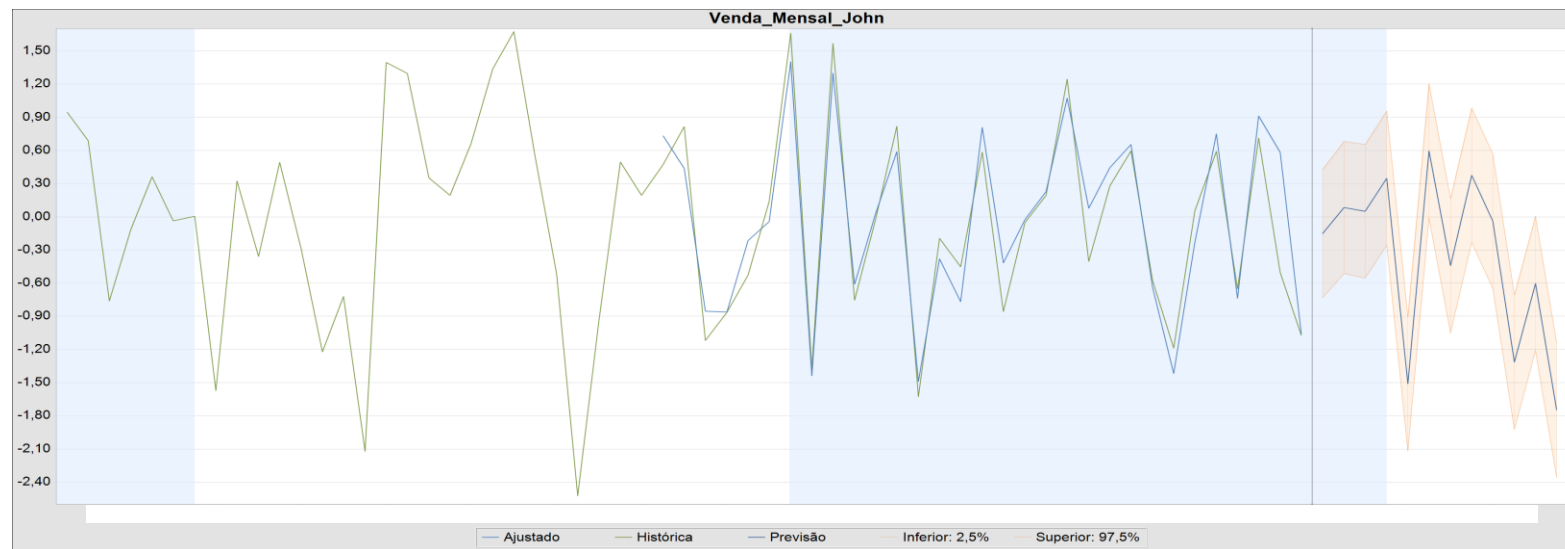
O que investigamos e queremos responder, é se essa meta é realista frente ao histórico das vendas da empresa.

Embora haja vários modelos que possam projetar valores, optamos por um baseado em séries temporais, pelo tipo de base de dados que foi disponibilizado. O que obteve a melhor pontuação foi o $SARIMA(1,0,1)(1,1,1)$.

Olhando para frente: Vendas no futuro

Período	Inferior: 2,5%	Previsão	Superior: 97,5%
jan/21	82.231,35	119.901,05	185.150,85
fev/21	94.423,58	142.292,19	228.954,96
mar/21	91.919,29	138.588,45	223.330,69
abr/21	111.671,64	174.155,77	290.426,46
mai/21	25.550,86	49.851,36	74.087,92
jun/21	132.489,27	212.645,64	363.542,14
jul/21	67.729,32	98.818,05	151.148,04
ago/21	113.538,14	177.741,25	297.532,74
set/21	86.706,60	129.910,48	207.890,54
out/21	34.048,71	57.275,98	83.670,33
nov/21	61.272,96	89.248,03	134.304,18
dez/21	13.864,58	40.931,34	64.052,13

Quadro 1 – Tabela de prisão das vendas.



Observando o quadro e o gráfico ao lado, deduzimos que meta perseguida para 20x2 é provável se ocorrer o melhor cenário, portanto acima do valor que se espera, destacado na coluna “Previsão”.

Apesar de estatisticamente provável, medir o quão provável permitirá uma visão realista de quanto trabalho diferenciado será exigido da equipe.

Olhando para frente: Vendas no futuro

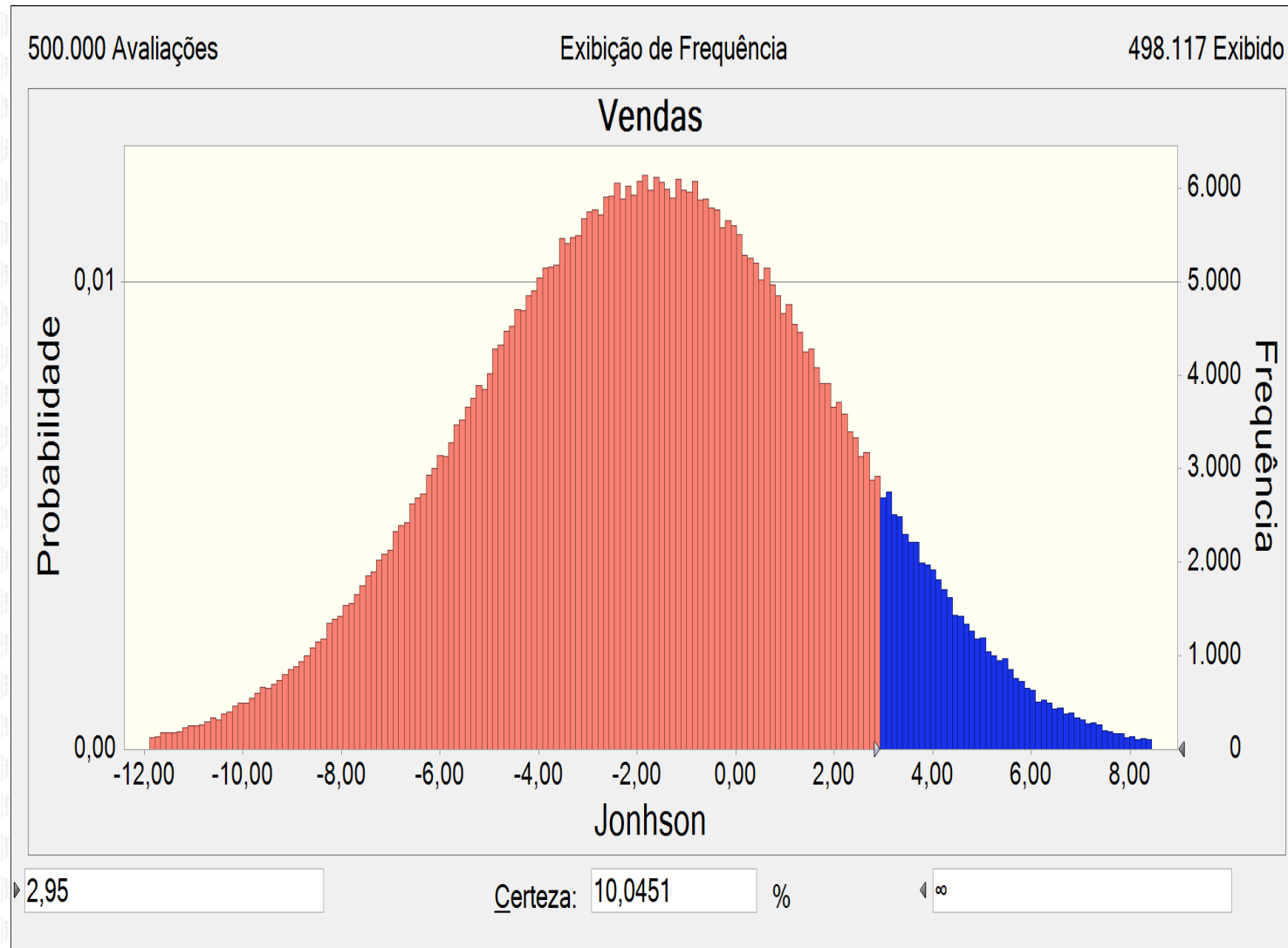


Gráfico 2 – Simulação de Monte Carlo com 500.000 simulações

Para responder ao tamanho do esforço foi necessário superar a dificuldade da granularidade de dados anuais. A solução foi utilizar simulação de Monte Carlo.

Realizada 500 mil simulações, ficou consolidado que a **probabilidade** de ocorrer um valor igual ou superior a meta **é de 10,04%**.

Considerando que nada de diferente seja feito em 20xz.



“ **Overfitting** ocorre quando desenvolvemos um modelo superespecializado no conjunto de dados treinados, tornando-se incapaz de realizar uma previsão acurada com uma nova base de dados.

Analisar a qualidade das projeções geradas pelo modelo é fundamental para afastar esse risco.

Esse modelo aproveitou os valores outliers utilizando uma transformação conhecida como normalização a partir do algoritmo $-1,50724 + 0,938140 \times \text{Asinh}((X - 49998,5) / 34967,8)$.



**Aconselhamento
com uma I.A.**

```
..._mod = modifier_ob.  
... mirror object to mirror  
..._mod.mirror_object =  
... operation == "MIRROR_X":  
... mirror_mod.use_x = True  
... mirror_mod.use_y = False  
... mirror_mod.use_z = False  
... operation == "MIRROR_Y":  
... mirror_mod.use_x = False  
... mirror_mod.use_y = True  
... mirror_mod.use_z = False  
... operation == "MIRROR_Z":  
... mirror_mod.use_x = False  
... mirror_mod.use_y = False  
... mirror_mod.use_z = True  
... selection at the end -add  
..._ob.select= 1  
..._ob.select=1  
... context.scene.objects.active  
... ("Selected" + str(modifier_...  
... mirror_ob.select = 0  
... = bpy.context.selected_object  
... data.objects[one.name].select  
... print("please select exactly  
...  
... OPERATOR CLASSES ...  
... types.Operator):  
... X mirror to the selected  
... object.mirror_mirror_x"  
... mirror X"  
... context):  
... text.active_object is not
```




“ **O modelo GMB** foi o modelo de machine learning que corroborou e aperfeiçoou a análise de agrupamentos, apontando com um bom nível de precisão as variáveis que mais impactam as vendas.

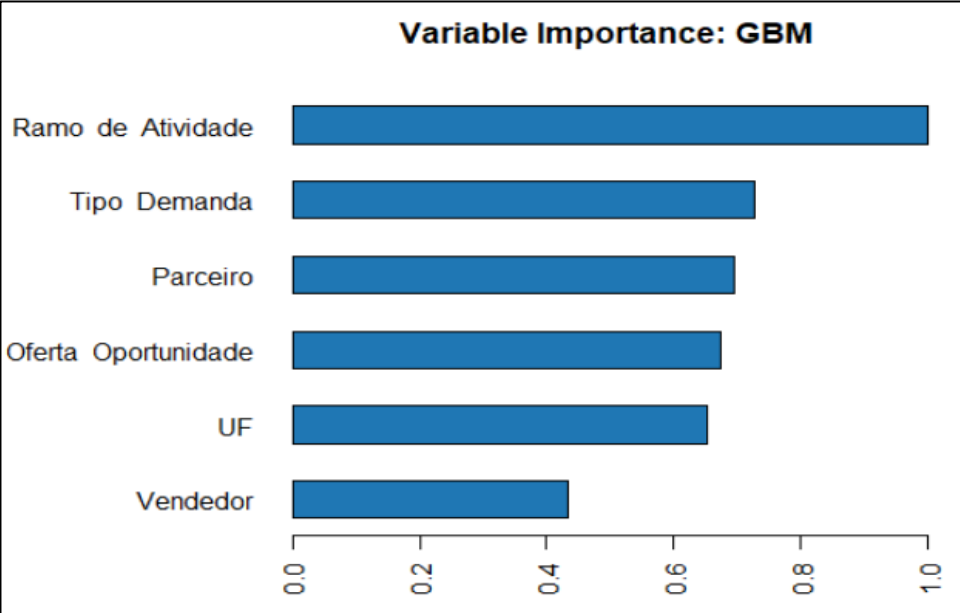


Gráfico 6 – Previsão de vendas usando SARIMA(1,0,1)(1,1,1)

Machine learning: O uso do modelo pretendia encontrar uma equação que explicasse explicitamente quais e quanto cada variável impacta nas vendas. Entretanto o melhor modelo encontrado, Gradient Boosting Machine (GBM), consegue apontar quais variáveis são mais relevantes para o resultado final das vendas, como demonstrado no gráfico a esquerda.

Performance do modelo GBM.

A tabela ao lado mostra que se um conjunto de dados novos for usado no modelo, acertará em média 73,41% das vezes se a venda estará no intervalo de muito baixa a muito alta.

	Alto	Baixo	Medio	Muito Alto	Muito Baixo	Error	Rate	Precision
Alto	98	9	7	1	3	0.1695	20 / 118	0.75
Baixo	9	98	12	0	15	0.2687	36 / 134	0.66
Medio	16	17	91	1	8	0.3158	42 / 133	0.75
Muito Alto	4	0	0	11	1	0.3125	mai/16	0.85
Muito Baixo	3	24	12	0	94	0.2932	39 / 133	0.78
Total	130	148	122	13	121	0.2659	142 / 534	
Recall	0.83	0.73	0.68	0.69	0.71			

Tabela 1 – Matriz de confusão do modelo GBM.



Menos é mais



“ **Reduzir a dimensão** e o encadeamento entre as variáveis que explicam o resultado, significa apontar com clareza qual o caminho a ser trilhado.

Menos é mais

Ramo Atividade	Valores	Ticket Médio	Participação Acumulada	Gráficos
Call center	2.989.515,02	14.303,90	45,16%	
Financeiro	781.152,51	10.278,32	56,96%	
Saúde	464.229,91	10.796,04	63,97%	
Comunicação/Editoras/Publici	374.080,97	15.586,71	69,62%	
Serviços	342.903,52	10.085,40	74,80%	
Editoras e Jornais	281.373,28	31.263,70	79,05%	
TI / Telecom	190.849,72	11.928,11	81,94%	
TIC	176.436,32	9.802,02	84,60%	
Outros	121.862,70	24.372,54	86,44%	
Imóveis	121.210,72	7.575,67	88,27%	
Varejista	102.115,40	8.509,62	89,82%	
Setor Público	89.896,00	14.982,67	91,18%	
Comunicação/Publicidade	86.242,24	9.582,47	92,48%	
Comércio - Atacado	79.370,24	26.456,75	93,68%	
Educação	76.052,76	12.675,45	94,83%	
Atacadista	75.733,88	25.244,63	95,97%	
Comércio - Varejo	73.778,92	5.675,30	97,08%	
Indústria	61.045,34	8.720,76	98,01%	
Alimentos	42.781,56	8.556,31	98,65%	
Indústria - Outros	37.846,00	6.307,67	99,22%	
Indústria - Alimentos/Bebidas	28.664,44	4.094,92	99,66%	
Turismo	12.040,00	12.040,00	99,84%	
Agropecuário	10.609,08	10.609,08	100,00%	

Tabela 2 – Classificação de Pareto para variável Ramo de Atividade

Observar as relações da tabela 2 mostra para onde os esforços de marketing e comercialização devem ser focados.

Algumas questões que podem ajudar no enfoque das decisões são:

Quais ramos devem ser esquecidos e quais contemplados e por que?

As variações observadas nos gráficos, é compreendida pela equipe?



O modelo esquemático mostra a dinâmica onde um ramo de atividade localizado em uma UF, manifesta uma necessidade por uma solução, que é atendida por um vendedor de um parceiro e classificada em um tipo de demanda para ser solucionada por uma oferta de oportunidade.

Ramo de Atividade é a origem do processo, portanto a variável focal para um estudo de melhoria dos resultados.



Concluindo



- Previsão de valores de vendas por médias, nesse caso não são adequadas pela distribuição irregular dos valores e a forte presença de outliers.
- O crescimento de 12% sobre o ano de 20xy é possível, entretanto a probabilidade de ocorrer é de 10,04%.

- Distribuição irregular das vendas e forte presença de outliers é um clássico problema em processos. Verificar a padronização, métricas e treinamento é fundamental para iniciar a correção.
- Estratégias e táticas mercadológicas devem focar primeiro a variável “Ramo de atividade” para definir suas ações.
- 26% dos ramos de atividades respondem por 79% do volume financeiro das vendas.
- Editoras e jornais que apresenta o maior ticket médio parou de gerar negócios em 20xk.
- 9 dos 23 ramos de atividades apresentam descontinuidades significativas em relação as vendas, no período de 5 anos.



Obrigado

Luiz Almeida Júnior

Cientista de dados | Black Belt Lean & Six-Sigma

luiz.almeida.jr@hotmail.com

<https://www.linkedin.com/in/luiz-almeida-datascience>

