

Case Hotmart

Principais Descobertas – Série temporal de vendas

1. Visão Geral

```
Qtidade de afiliados únicos: 22947
Qtidade de produtores únicos: 8020
Qtidade de compradores únicos: 1100649
Qtidade de compras únicas: 1599828
Qtidade de produtos únicos: 17883
Receita Total: 15998280.019973695
Qtidade de origens únicas: 9603
Qtidade de devices únicos: 5
Qtidade de nichos únicos: 25
Qtidade de categorias únicas: 10
```

OBS.: Para cada valor contido na variável 'purchase_value' foi adicionado o valor de 10 (int), para eliminar o efeito dos valores negativos em agregações de soma.

A base de dados continha apenas 1 amostra com dados faltantes, que foi excluído pois todos os valores estavam vazios.

2. Clientes

As vendas se mostram bastante diluídas entre os clientes. Mesmo os clientes com maior histórico de compras (tanto em quantidade quanto em receita), individualmente representam uma porção muito pequena (menos de 1%) em relação ao total de vendas.

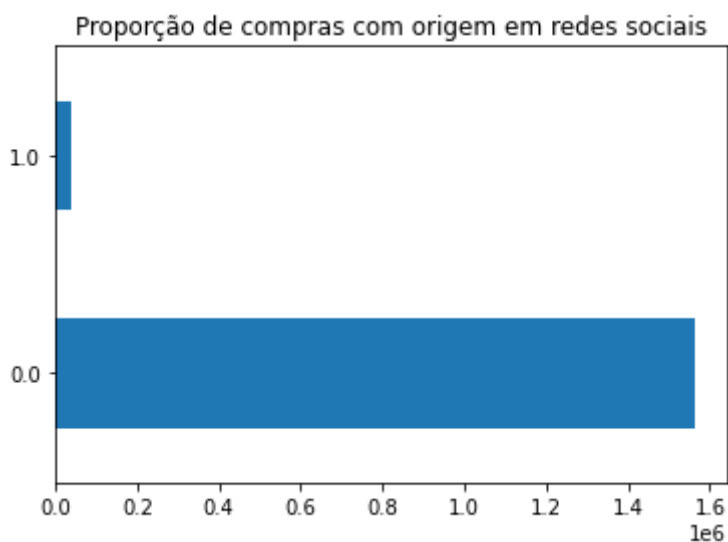
Os top 10 compradores possuem entre 41 e 66 compras, e gasto total entre 405 e 656.

buyer_id	compras	receita_total	%_compras	%_receita	ticket_medio
4763605	66	656.25	0.0	0.0	9.94
701523	62	614.26	0.0	0.0	9.91
158226	62	604.00	0.0	0.0	9.74
336781	59	579.07	0.0	0.0	9.81
1128526	51	569.00	0.0	0.0	11.16
137316	44	422.23	0.0	0.0	9.60
2196492	43	417.49	0.0	0.0	9.71
37734	43	413.28	0.0	0.0	9.61
32277	42	405.10	0.0	0.0	9.65
314082	41	404.99	0.0	0.0	9.88

No geral, os clientes analisados realizam em média 1,45 compras, sendo que grande parte deles (75%) realizou apenas 1 única compra.

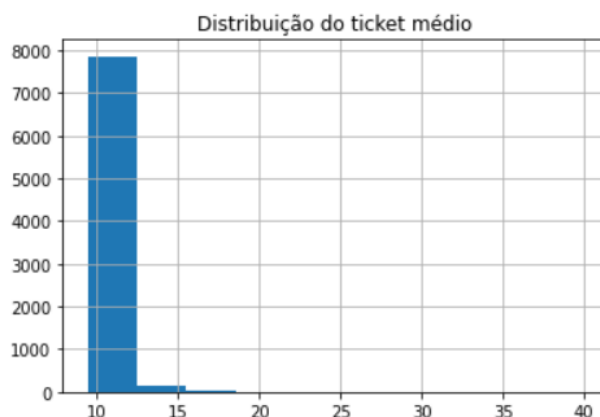
Também por isso, existe pouca variabilidade entre quantidade de produtos, categorias e nichos distintos que são adquiridos por esses clientes em média. Os que compram mais de 1 tipo de produto/categoria/nicho são outliers, representando menos de 25% da base total de clientes.

A maior parte das compras registradas veio de clientes que não tiveram as redes sociais como origem, de acordo com a variável 'is_origin_page_social_network'. Os clientes que vieram de redes sociais e efetuaram a compra corresponde a apenas 2,3% do total.



3. Produtores

Os produtores que mais vendem são os maiores responsáveis pelo faturamento da Hotmart. Isso é um comportamento natural considerando que os valores de ticket médio possuem pouca variação entre si, e se apresentam bastante homogêneos. Esse fator contribui para que a participação em vendas fique muito próxima da participação em receita.

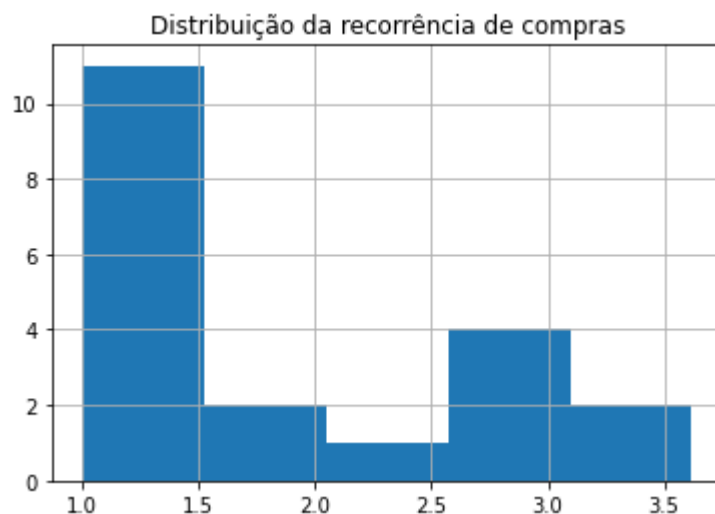


Os 10 produtores com maior quantidade de vendas são responsáveis por 17,6% das vendas, e 17% da receita total.

	produtor	vendas	receita_total	%_vendas	%_receita	ticket_medio
7500	6697083	41220	395158.75	2.576527	2.470008	9.59
5002	3992235	39331	379743.50	2.458452	2.373652	9.66
1100	464846	35470	342262.50	2.217113	2.139371	9.65
877	349701	34568	342514.70	2.160732	2.140947	9.91
556	166090	28895	279678.68	1.806132	1.748180	9.68
1054	442241	27798	272007.86	1.737562	1.700232	9.79
3408	2307584	21720	210073.17	1.357646	1.313098	9.67
4415	3382787	20199	192690.83	1.262573	1.204447	9.54
5570	4580574	16386	155324.88	1.024235	0.970885	9.48
1544	671256	16096	158555.78	1.006108	0.991080	9.85

Para medir a recorrência de compras, foi criado um indicador que leva em conta a quantidade de produtos criados e a quantidade de clientes distintos que compraram produtos desse produtor. Esse índice foi nomeado como taxa de recorrência. De certa forma é um índice que pode ajudar a mensurar a fidelização dos clientes com os produtores, e entender quais se destacam.

Foi observado que a grande maioria dos produtores tem uma taxa de recorrência baixa, o que significa que os clientes fazem apenas uma única compra. Porém, existem alguns produtores que apresentam altas taxas de recorrência, com média entre 2 e 3,62 produtos vendidos por cliente.

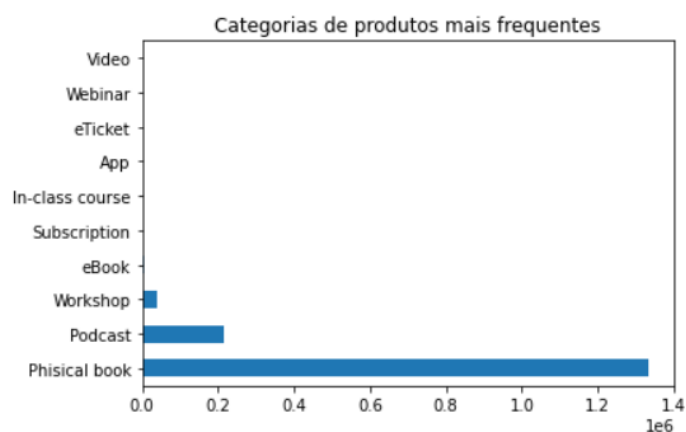


compradores_unicos	produtos_unicos	vendas_totais	tx_recorrencia
producer_id			
442241.0	7682	6	27798
4580574.0	4563	1	16386
2026525.0	3309	2	10094
906263.0	2995	1	9044
166090.0	10156	8	28895
3512.0	4875	4	12931
3992235.0	18393	5	39331

4. Produtos

4.1.Categorias

Com relação às categorias principais dos produtos vendidos, o destaque fica com a categoria 'physical book' que está presente e, 83,4% das vendas e 83,8% da receita, seguido da categoria 'podcast' com 13,5% das vendas e 12,9% da receita.



categoria	vendas	receita_total	%_vendas	%_receita	ticket_medio
Physical book	1334610	13413697.23	83.42	83.84	10.05
Podcast	216720	2076434.04	13.55	12.98	9.58
Workshop	38246	408120.43	2.39	2.55	10.67
eBook	3965	38259.31	0.25	0.24	9.65
Subscription	1970	19314.80	0.12	0.12	9.80
App	1557	15048.82	0.10	0.09	9.67
In-class course	1561	15920.87	0.10	0.10	10.20
eTicket	847	8097.57	0.05	0.05	9.56
Webinar	295	2809.75	0.02	0.02	9.52
Video	57	577.20	0.00	0.00	10.13

4.2.Nichos

Os nichos de produtos com maior relevância são 'Negotiation', 'Anxiety management' e 'Personal finance'. Juntos, as vendas de produtos nesses nichos representam 44,9% das vendas e da receita total.

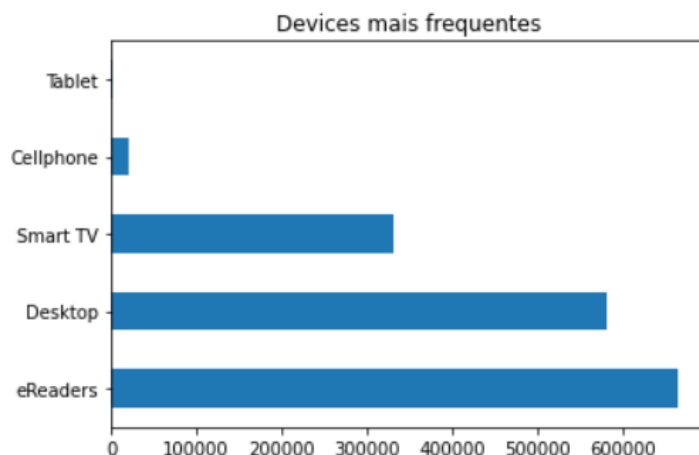


nicho	vendas	receita_total	%_vendas	%_receita	ticket_medio
Negotiation	279921	2785509.95	17.50	17.41	9.95
Anxiety management	251675	2505463.04	15.73	15.66	9.96
Personal finance	186849	1903377.88	11.68	11.90	10.19
Presentation skills	140460	1440358.50	8.78	9.00	10.25
Immigration	95660	924308.12	5.98	5.78	9.66
Government	93365	905453.58	5.84	5.66	9.70
YouTube video creation	80206	781778.16	5.01	4.89	9.75
Online course creation	60431	597396.45	3.78	3.73	9.89
Careers	52929	555969.90	3.31	3.48	10.50
Organization	48632	480817.78	3.04	3.01	9.89

4.3.Devices

O device mais utilizado no momento da compra são os 'eReaders', o que parece ser meio contraditório considerando que a principal categoria vendida é a de livros físicos ('physical book'). Em seguida, vêm os devices 'Desktop' e 'Smart TV'.

Somados, esses 3 devices (ou 60% do total de devices) correspondem à mais de 98% das vendas e da receita total.



device	vendas	receita_total	%_vendas	%_receita	ticket_medio
eReaders	664197	6656705.59	41.52	41.61	10.02
Desktop	581900	5923618.16	36.37	37.03	10.18
Smart TV	330073	3176822.56	20.63	19.86	9.62
Cellphone	20708	211128.86	1.29	1.32	10.20
Tablet	2950	30004.86	0.18	0.19	10.17

4.4.Origin

Existe uma grande variedade de níveis distintos nessa variável (são 9.603 níveis distintos). As origens que ocorrem com maior frequência são: 'Origin ef2b', 'Origin 5187' e 'Origin adf0', que, juntas, correspondem à cerca de 35% de todas as vendas e do total de receita.

origem	vendas	receita_total	%_vendas	%_receita	ticket_medio
Origin ef2b	330077	3176861.32	20.63	19.86	9.62
Origin 5187	167028	1716864.93	10.44	10.73	10.28
Origin adf0	77857	753694.06	4.87	4.71	9.68
Origin 18eb	28693	276151.53	1.79	1.73	9.62
Origin 3ade	13069	126356.04	0.82	0.79	9.67
Origin cf02	12066	126335.52	0.75	0.79	10.47
Origin d8b2	11033	111064.25	0.69	0.69	10.07
Origin cd46	10723	105034.36	0.67	0.66	9.80
Origin a144	10049	98993.21	0.63	0.62	9.85
Origin 3745	9404	91052.36	0.59	0.57	9.68

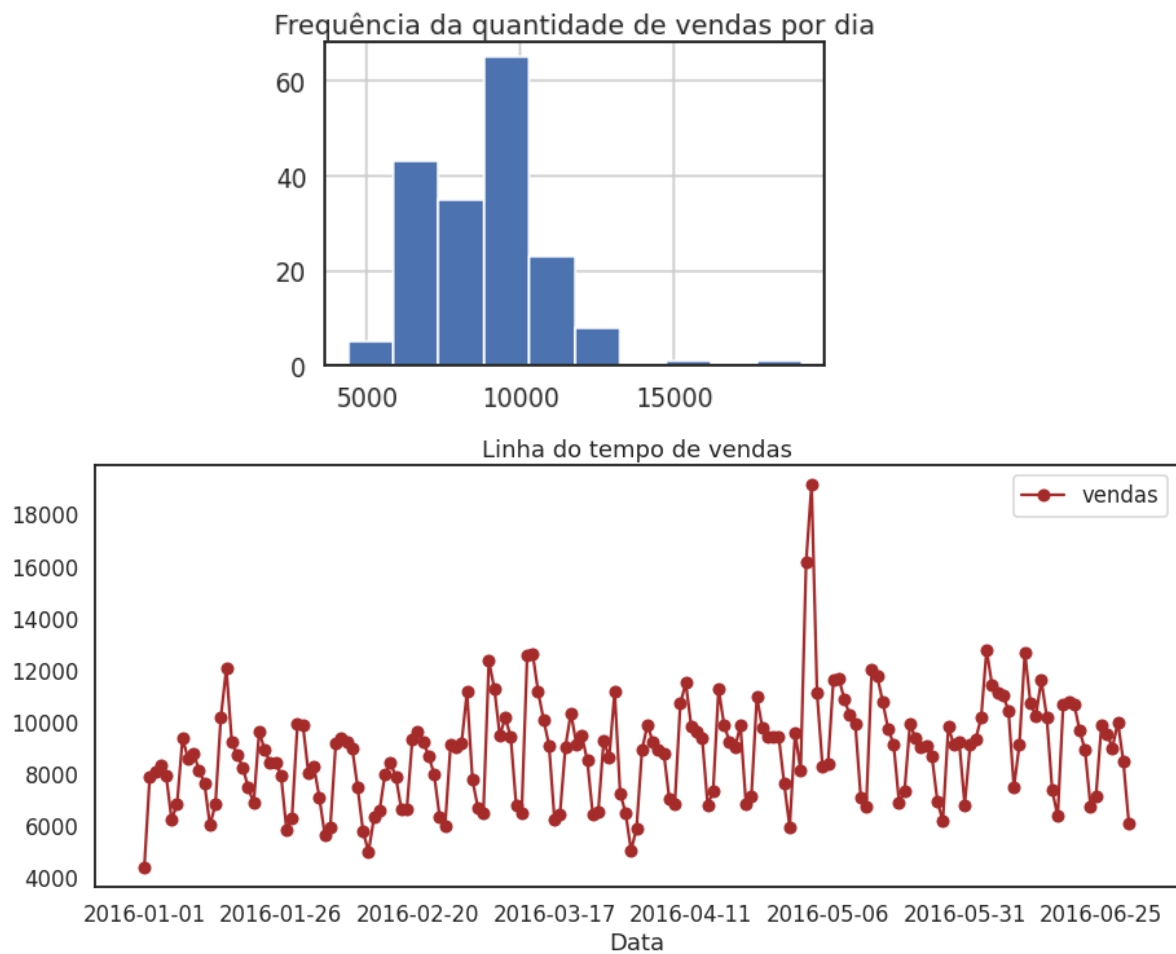
5. Vendas

5.1. Linha do tempo de vendas

É provável que nem todos os valores da variável 'creation_date' estejam corretos, pois existem datas de criação de produtos posteriores à primeira data de venda registrada. Outro comportamento observado é que existem cursos que apresentam uma data de criação muito antiga, apresentando uma diferença muito grande, de anos inclusive, em relação à data da primeira venda registrada.

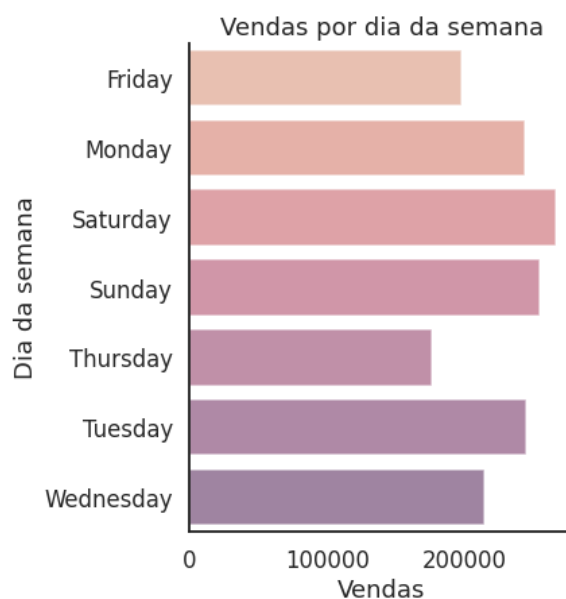
```
Data mínima de criação: 2008-10-27
Data máxima de criação: 2016-12-31
Data mínima de venda: 2016-01-01
Data máxima de venda: 2016-06-30
```

A média diária é de 8.838 vendas por dia, sendo que o dia 01/01/2016 é a data com a menor quantidade de vendas (4.378) e o dia 03/05/2016 é a data com maior quantidade de vendas (19.167).



5.2.Dia da semana

Sábado e Domingo são os dias em que ocorrem a maior parte das vendas. As vendas realizadas aos fins de semana representam 32,6% de todas as vendas.

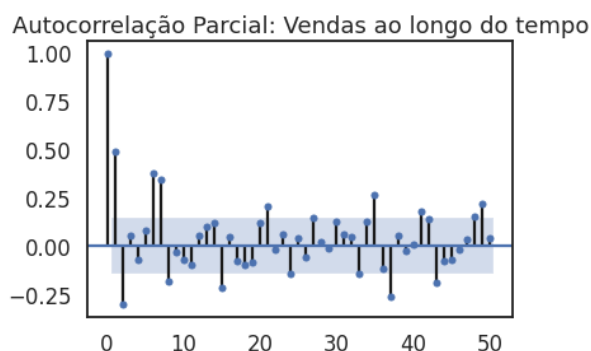
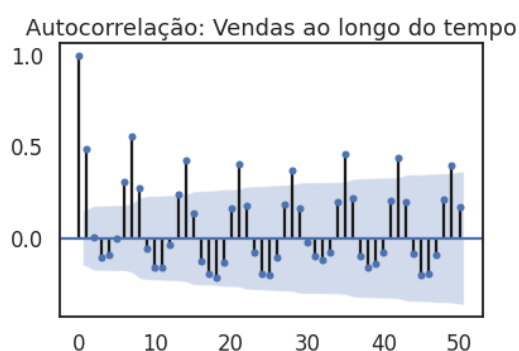
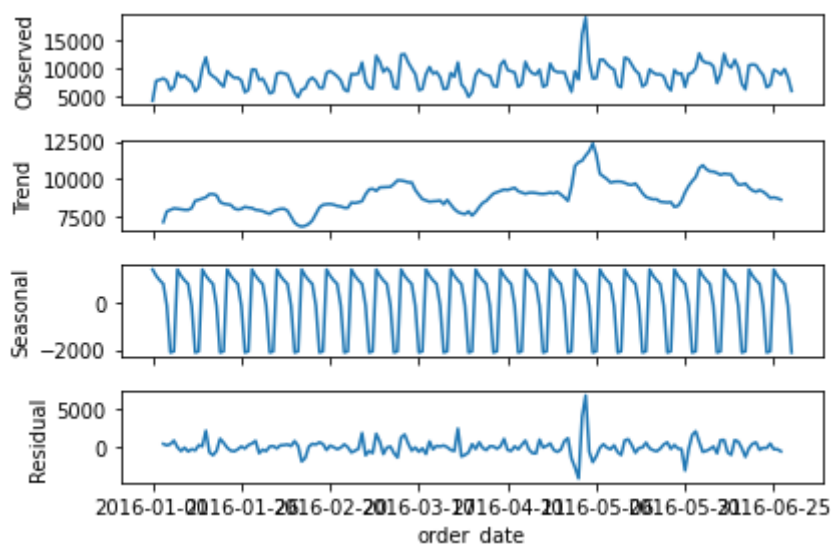


	weekday	vendas
2	Saturday	266326
3	Sunday	254922
5	Tuesday	244935
1	Monday	244063
6	Wednesday	215013
0	Friday	198255
4	Thursday	176314

6. Predição de vendas

Para prever as vendas dos próximos 90 dias realizei uma análise da correlação entre a série de dados, além de uma decomposição sazonal para identificar indícios de estacionariedade, sazonalidade e tendência da série. De forma resumida, essas análises são importantes para escolher qual modelo de previsão de séries temporais é o mais adequado para o conjunto de dados em questão.

Essas análises demonstraram a existência de estacionariedade fraca (a média e a variância dos dados não se mantêm estáveis ao longo do tempo), tendência e sazonalidade.

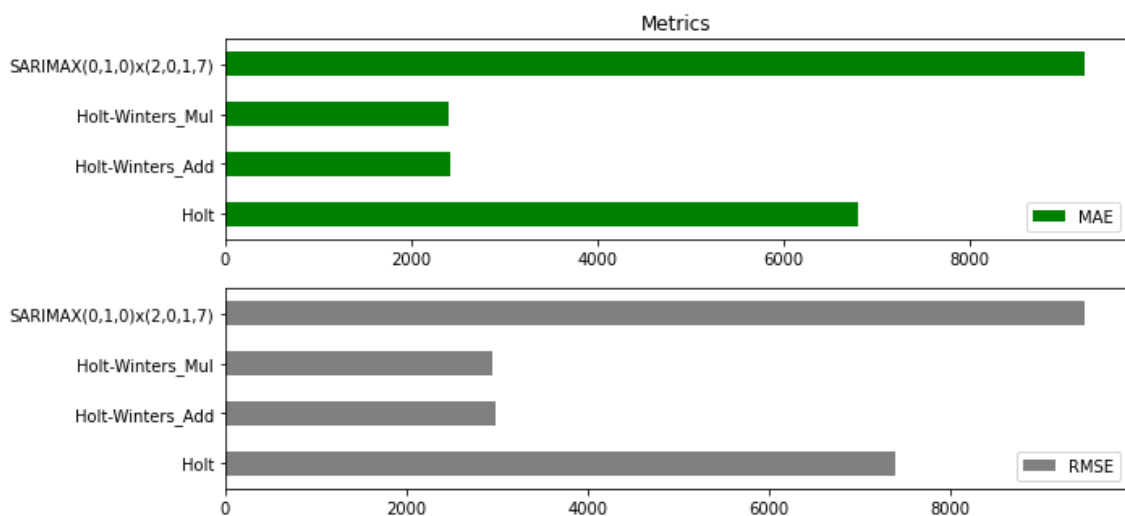


Para prever a quantidade de vendas, foram utilizados 5 modelos diferentes, com base em algoritmos bastante conhecidos para previsões de séries temporais. São eles:

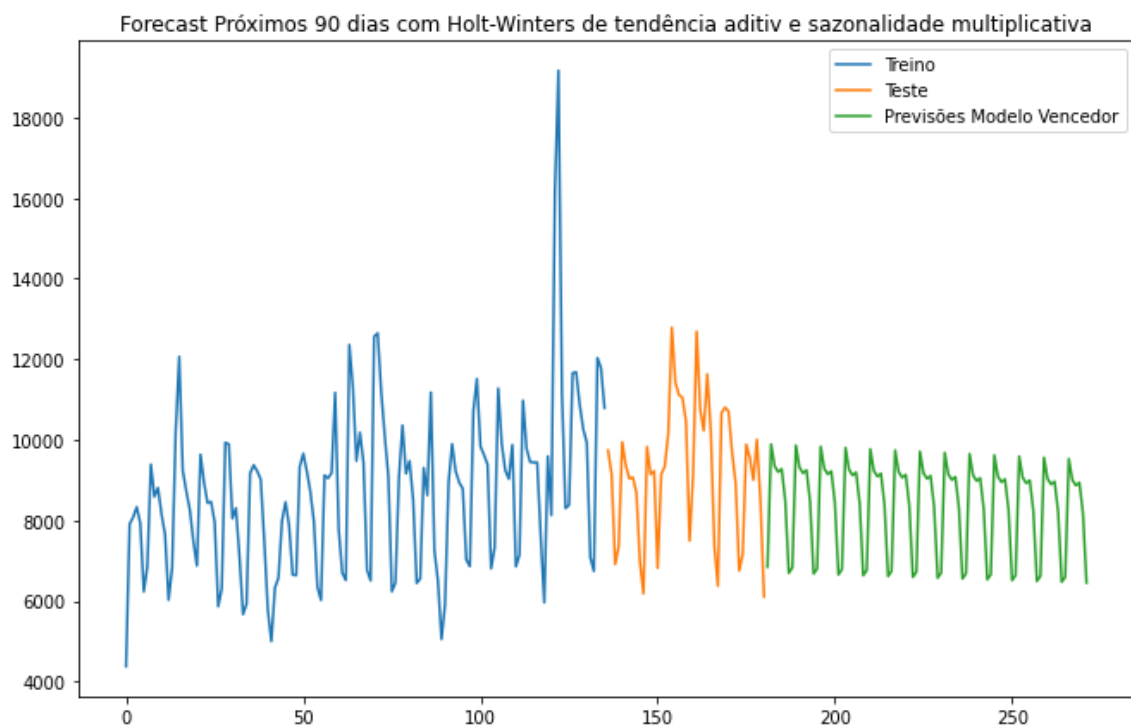
- 1) Modelos de suavização exponencial tripla (Holt-Winters com tendência aditiva; Holt-Winters com tendência e sazonalidade aditiva e Holt-Winters com tendência e sazonalidade multiplicativa)
- 2) Modelos auto-regressivo integrado de médias móveis com componente de sazonalidade (SARIMA)
- 3) Modelo de arquitetura de rede neural recorrente (LSTM)

A base de dados foi dividida em uma base de treino (75% da base) e uma base de teste (25% da base). Ou seja, os dados da base de treino contemplam as vendas realizadas entre 01/01 e 16/05/2016, e a base de teste contempla as vendas de 17/05 a 30/06/2016.

As métricas utilizadas para avaliação dos modelos foram o erro absoluto médio (MAE), média do erro quadrático (MSE) e raiz do erro quadrático médio (RMSE). Com base nos resultados alcançados, o modelo de Holt-Winters com tendência aditiva foi o que trouxe os melhores resultados.



Com o modelo vencedor (Holt- Winters com tendência aditiva e sazonalidade multiplicativa), o RMSE ficou em 2.950, o que equivale a dizer que em média esse modelo está errando 2.950 vendas para mais ou para menos. Com base na média de vendas por dia (8.838) o erro apresentado equivale a 33% dessa média, indicando que ainda existe bastante espaço para melhorar o desempenho dessas previsões.



Referências

- Ehlers, R.S. (2009) Análise de Séries Temporais, disponível em:
<http://www.icmc.usp.br/~ehlers/stemp/stemp.pdf>
- StatsmodelsSeasonal-Trend decomposition using LOESS (STL), disponível em
https://www.statsmodels.org/dev/examples/notebooks/generated/stl_decomposition.html

Links externos

[Notebook](#) utilizado nas análises e na criação dos modelos de predição temporal de vendas.

[Repositório](#) do projeto.