

## Análise do Amazon Fine Food Reviews dataset

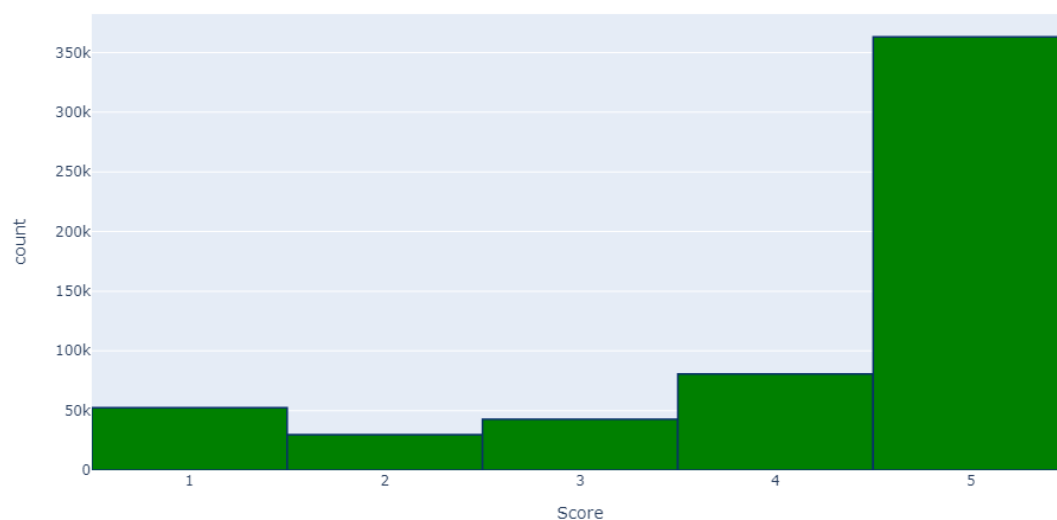
A análise a seguir se refere ao Amazon Fine Food Reviews, um *dataset* público que contém reviews de *fine foods* (comidas que não são do dia-a-dia, por exemplo, caviar, vinhos, ostras etc) e das outras categorias da Amazon. Os dados foram coletados ao longo de 10 anos até outubro de 2012 e inclui mais de 500.000 reviews. Os reviews são principalmente compostos por: id do produto, id do usuário, *score* (nota de 1-5), resumo do review (Summary) e o review completo (*Text*). Download feito no site <https://www.kaggle.com/snap/amazon-fine-food-reviews?select=Reviews.csv>. Utilizou-se as técnicas de *natural language processing (NLP)* – *TextBlob* e *Latent Dirichlet Allocation (LDA)* feito na linguagem de programação *python* para a análise do *dataset*.

O objetivo do estudo é verificar se as reviews são positivas ou negativas e, no caso das negativas, avaliar quais são os prováveis tópicos dessas avaliações e recomendar o que poderia ser feito diminuir o número de avaliações negativas.

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Text
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	1303862400	Good Quality Dog Food	I have bought several of the Vitality canned d...
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	1346976000	Not as Advertised	Product arrived labeled as Jumbo Salted Peanut...
2	3	B000LQOCH0	ABXLMWJDXAIN	Natalia Corres "Natalia Corres"	1	1	4	1219017600	"Delight" says it all	This is a confection that has been around a fe...
3	4	B000UA0QIQ	A395BORC6FGVXV	Karl	3	3	2	1307923200	Cough Medicine	If you are looking for the secret ingredient i...
4	5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham "M. Wassir"	0	0	5	1350777600	Great taffy	Great taffy at a great price. There was a wid...
...	...	...	...	...	...	...	...	...	...	...
568449	568450	B001EO7N10	A28KG5XORO54AY	Lettie D. Carter	0	0	5	1299628800	Will not do without	Great for sesame chicken. this is a good if no...
568450	568451	B003S1WTCU	A3I8AFVPEE8K15	R. Sawyer	0	0	2	1331251200	disappointed	I'm disappointed with the flavor. The chocolat...
568451	568452	B004I613EE	A121AA1GQV751Z	pkds "pk_007"	2	2	5	1329782400	Perfect for our maltipoo	These stars are small, so you can give 10-15 o...
568452	568453	B004I613EE	A3IBEVCTXKNOH	Kathy A. Welch "katwei"	1	1	5	1331596800	Favorite Training and reward treat	These are the BEST treats for training and rew...
568453	568454	B001LR2CU2	A3LGQPJCZVL9UC	srfe117	0	0	5	1338422400	Great Honey	I am very satisfied ,product is as advertised,...

Tabela 1 – Parte do *dataset*

Na tabela 1, podemos observar todas as features descritas na introdução, como *Score* e *Text*, e o número total de observações que foram registradas.



O primeiro passo foi analisar a feature *Score* e como ela está distribuída no dataset. Seu intervalo é de 1-5 e, quanto maior o número, melhor foi avaliado o produto. Na figura 1, vemos que a maioria das reviews foram avaliadas com 5, portanto podemos esperar que a maioria dos reviews escritos (Text) foram positivos também.

Para confirmar que a maioria dos reviews foram positivos, temos na figura 2, palavras em destaque, como: *love, amazon, taste, delicious* etc. Isso confirma que a maioria dos reviews estão, em geral, falando bem dos produtos comprados no site e é positivo para a empresa.

Em seguida, para separar as reviews entre positivas e negativas, foi feito um Sentiment Analysis com TextBlob de cada um dos *Text* para verificar sua polaridade, significando -1 muito negativa e +1 muito positiva.







Figura 7 - Wordcloud com sentiment < -0.1

Realizando a mesma análise, só que retirando as áreas (1) e (2), portanto, as reviews neutras, obtemos a figura 6, e logo percebe-se uma mudança na palavra em destaque, *love*, demonstrando o sentimento positivo. Na figura 7, vemos o contrário, as palavras *disappointed* e *bad* (circuladas em preto), as mesmas da figura 5, ganharam mais destaque, revelando um maior sentimento negativo.

Para obter alguma informação sobre as reviews negativas (área a esquerda da linha amarela na figura 3), utilizou-se da técnica de LDA para *Topic Modeling*, assim será possível identificar os possíveis principais tópicos dessas avaliações negativas. Configurado o parâmetro para obter 3 tópicos dentre os reviews, foi obtido:

```
[(0,
  '0.025*"like" + 0.025*"taste" + 0.016*"coffee" + 0.015*"flavor" + 0.013*"tea" + 0.011*"just"
+ 0.011*"bad" + 0.008*"tastes" + 0.008*"br" + 0.008*"product"'),
(1,
  '0.017*"product" + 0.013*"amazon" + 0.011*"disappointed" + 0.011*"box" + 0.008*"just" + 0
.008*"order" + 0.008*"buy" + 0.008*"ordered" + 0.007*"bag" + 0.007*"received"'),
(2,
  '0.018*"chicken" + 0.017*"food" + 0.016*"dog" + 0.011*"eat" + 0.011*"like" + 0.011*"br" + 0
.009*"treats" + 0.008*"dogs" + 0.007*"loves" + 0.007*"just"')]
```

Analisando os tópicos, temos:

Tópico 0: bebidas, gosto, sabor e *bad* (com sentido de ser ruim)

Tópico 1: desapontado, pedido, entrega

Tópico 2: animais de estimação, comidas para animais de estimação.

Começando pelo tópico 2, ele não é de muita ajuda, pois está falando de animais de estimação e não há nenhum sentimento negativo presente. O tópico 0 aborda bebidas e seu gosto e apresenta também a palavra *bad*, que pode significar que quem comprou alguma bebida não gostou do sabor. Já o tópico 1 aborda pedido e entrega e um sentimento de desapontamento,

portando o cliente não ficou satisfeito com o que foi entregue, seja o produto ou alimento não atendeu a expectativa, porque no anúncio estava diferente ou o sabor não estava tão bom, caso seja uma comida, ou foi entregue o pedido errado.

Por fim, com os resultados obtidos, seria interessante a empresa verificar se o que está sendo entregue é o que foi anunciado, a qualidade das comidas e bebidas que estão sendo vendidas, pois, como a maioria do *dataset* são observações de *fine food*, o preço desses itens são elevados, gerando alta expectativa, que caso não foi atendida, gera reviews ruins, prejudicando as vendas da empresa e verificar se os pedidos estão sendo entregues corretamente pela transportadora/correios.