Olá ! segue minha análise .

1. Informações sobre o Conjunto de Dados

- Qual o conjunto de dados escolhido? Onde podemos encontrar ele?
  Avaliação de dispositivos mobiles provenientes da Amazon => Amazon-reviews-unlocked-mobile-phones (https://www.kaggle.com/PromptCloudHQ/amazon-reviews-unlocked-mobile-phones)

- Este conjunto de dados foi criado com o intuito de ser útil para que cenário?

O principal cenário para utilização desses dados, aplicando fundamentos que norteiam a extração e preparação dos dados que resultam em uma mineração de dados eficiente, seria a revisão das 'Reviews' através do processamento e análise de texto ,ou seja, o perfilamento dos produtos ao monitorar suas avaliações pelos próprios clientes , a fim de monitorar o comportamento e a satisfação do cliente ao receber e utilizar o produto, bem como acompanhar as possíveis falhas ou durabilidade através de previsão de vínculo entre as possíveis avaliações negativas ou erro de fabricação.

- Por que você achou interessante?

  Não somente pelo fato desses dados emanarem de uma empresa destaque no meio tecnológico e vendas, é uma base de dados que constitui atributos que se relacionam e podem ter análises consistentes. Também pelo fato de ser algo atual, o fato de o cliente buscar avaliações anteriores do produto para assegurar sua segurança de compra.

2. Explorando o Conjunto de Dados.

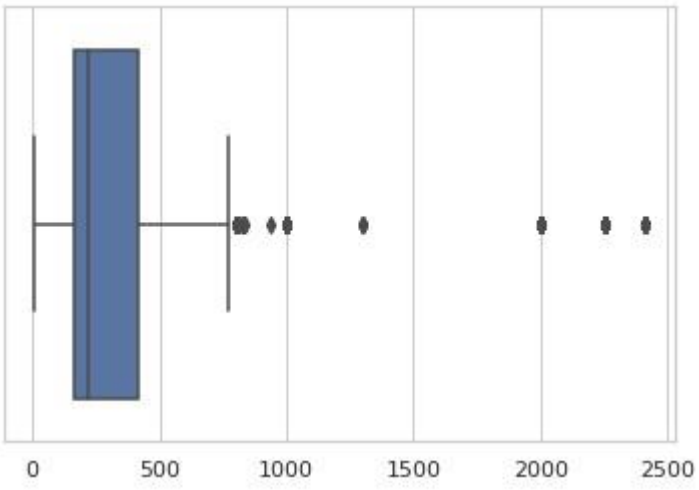
- Quais os atributos disponíveis no conjunto de dados? Qual o tipo desses atributos?

  O atributo:'Produt Name' corresponde ao nome do produto.
  'Brand Name' corresponde a marca do produto.
  'Price' corresponde ao preço do produto.
  'Rating' corresponde a uma nota de 1-5 de satisfação.
  'Reviews' corresponde a uma descrição da experiência do usuário.
  'Review Votes' corresponde ao numero de pessoas que avaliaram o comentário como útil.

	Product Name	Brand Name	Price	Rating	Reviews	Review Votes
0	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	samsung	199.99	5	feel so lucky to have found this used phone t...	1.0
1	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	samsung	199.99	4	nice phone nice up grade from my pantach revue...	0.0
2	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	samsung	199.99	5	very pleased	0.0
3	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	samsung	199.99	4	it works good but it goes slow sometimes but i...	0.0
4	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	samsung	199.99	4	great phone to replace my lost phone the only ...	0.0
5	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	samsung	199.99	1	already had a phone with problems know it st...	1.0
6	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	samsung	199.99	2	the charging port was loose got that soldered...	0.0
7	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	samsung	199.99	2	phone looks good but wouldnt stay charged had ...	0.0
8	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	samsung	199.99	5	originally was using the samsung s galaxy for...	0.0
9	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	samsung	199.99	3	its battery life is great its very responsive ...	0.0
10	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	samsung	199.99	3	my fiance had this phone previously but caused...	0.0
11	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	samsung	199.99	5	this is a great product it came after two days...	0.0
12	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	samsung	199.99	5	these guys are the best had a little situatio...	2.0
13	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	samsung	199.99	1	really disappointed about my phone and servic...	1.0
14	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	samsung	199.99	5	ordered this phone as a replacement for the sa...	1.0

- Quantos exemplos há no conjunto de dados? Existem dados faltantes? Outliers?

A quantidade inicial de amostras é 413.840 mil , porém , após remover linhas com dados faltantes, resta 334.335 mil, ou seja , uma diferença de 79.505 mil amostras.
Foram detectados Outliers nos preços dos produtos da Samsung, confira abaixo.



- Qual a média, mediana e desvio padrão dos atributos?

Confira abaixo as estatísticas das 20 marcas mais frequentes na base de dados.
*Observação: a variável 'Sum votes' representa a somatória da quantidade de pessoas que classificaram os comentários das respectivas marcas como útil

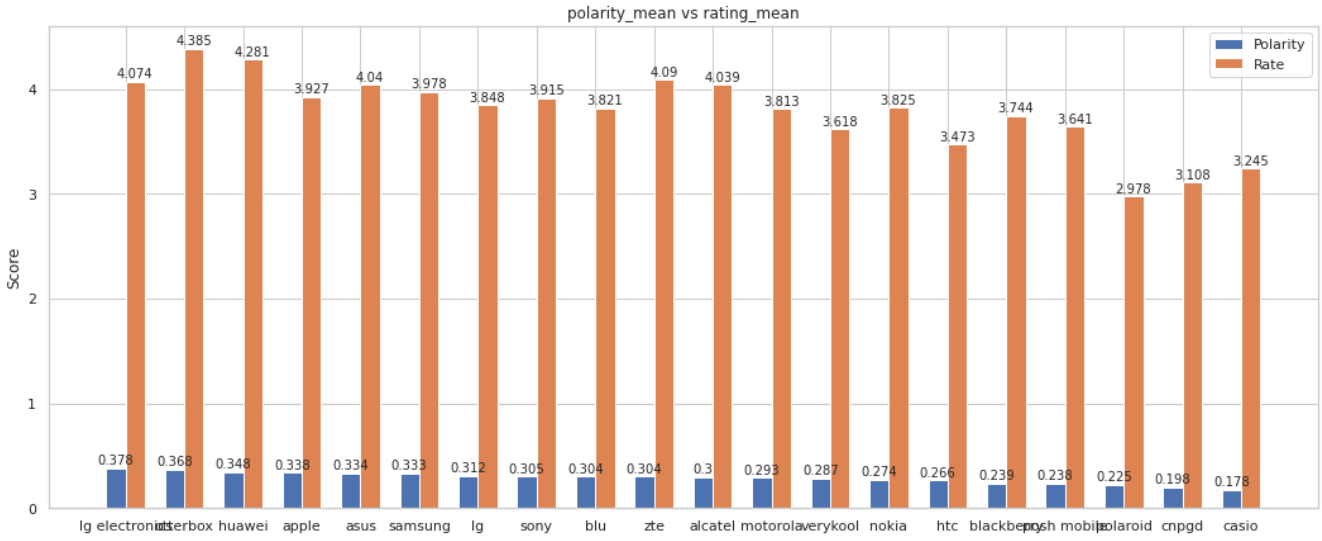
	Brand Name	Mean Rate	Median Rate	Mode Rate	Frequency	Sum Votes
0	samsung	3.978	5.0	[5]	65917	262235
1	blu	3.821	5.0	[5]	59176	226118
2	apple	3.927	5.0	[5]	56102	220290
3	lg	3.848	5.0	[5]	21642	83277
4	blackberry	3.744	5.0	[5]	17525	65618
5	nokia	3.825	5.0	[5]	16199	61956
6	motorola	3.813	5.0	[5]	13033	49693
7	htc	3.473	4.0	[5]	12539	43548
8	cnpgd	3.108	3.0	[5]	12302	38233
9	otterbox	4.385	5.0	[5]	7880	34556
10	sony	3.915	5.0	[5]	7652	29955
11	posh mobile	3.641	4.0	[5]	6516	23724
12	huawei	4.281	5.0	[5]	3682	15763
13	lg electronics	4.074	5.0	[5]	3043	12397
14	asus	4.040	5.0	[5]	1862	7523
15	zte	4.090	5.0	[5]	1619	6622
16	polaroid	2.978	3.0	[1]	1609	4792
17	alcatel	4.039	5.0	[5]	1380	5574
18	verykool	3.618	4.0	[5]	1111	4020
19	casio	3.245	4.0	[5]	1076	3492

3.Visualizando o Conjunto de Dados.

- Como é distribuição desses dados? Se houver "outliers", como podemos representá-los graficamente?

  Analisando o rank das 20 marcas mais frequentes, percebe-se que ela corresponde mais de 60% dos dados, porém ao se relacionarem entre si há algumas anuências comuns na variação de preço e frequência. Uma das principais formas de representar esses outliers é por BoxPlot, possibilitando visualizar a media, mediana, moda e até mesmo os quartis, porém é possível visualizar a distribuição em um scatter também em certas ocasiões (gráfico de pontos).

- Há uma relação pequena entre uma avaliação ('polarity') positiva (tendo o limite 1 como totalmente positivo) do cliente por marca, e a média de avaliações do produto(Rate), apesar das frequências serem diferentes



4.Quais as tecnologias que você utilizou para sua análise?

A fim de manipular os dados e analisá-los, foram utilizadas estas ferramentas

A fim de gerar processamento de texto e avaliar o sentimento e a subjetividade das avaliações: