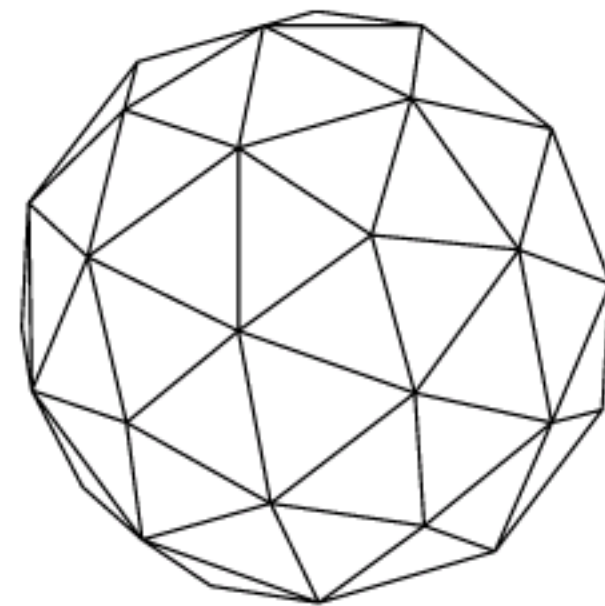


# Programa Atlântico Academy Future



**DISCOVER EVERYONE**  
DATA SCIENCE

Professor:

**Madson Luiz Dantas Dias**

Integrantes:

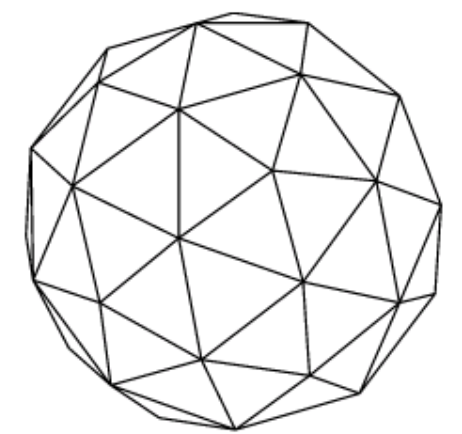
**Arildo Alves**

**Germano Fenner**

**Maurício Moura**

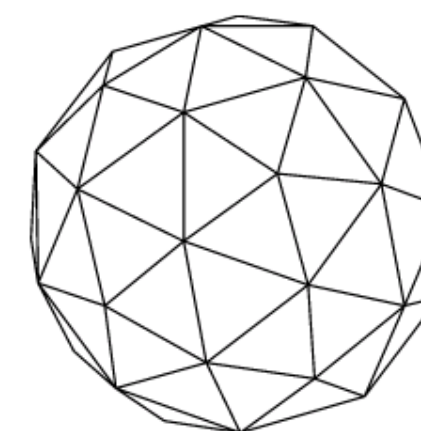
**Nator Júnior**

**Rodolfo Ricardo**



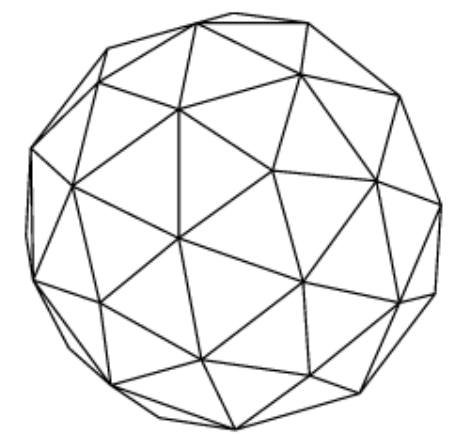
**DISCOVEREYONE**  
DATA SCIENCE

# A Organização

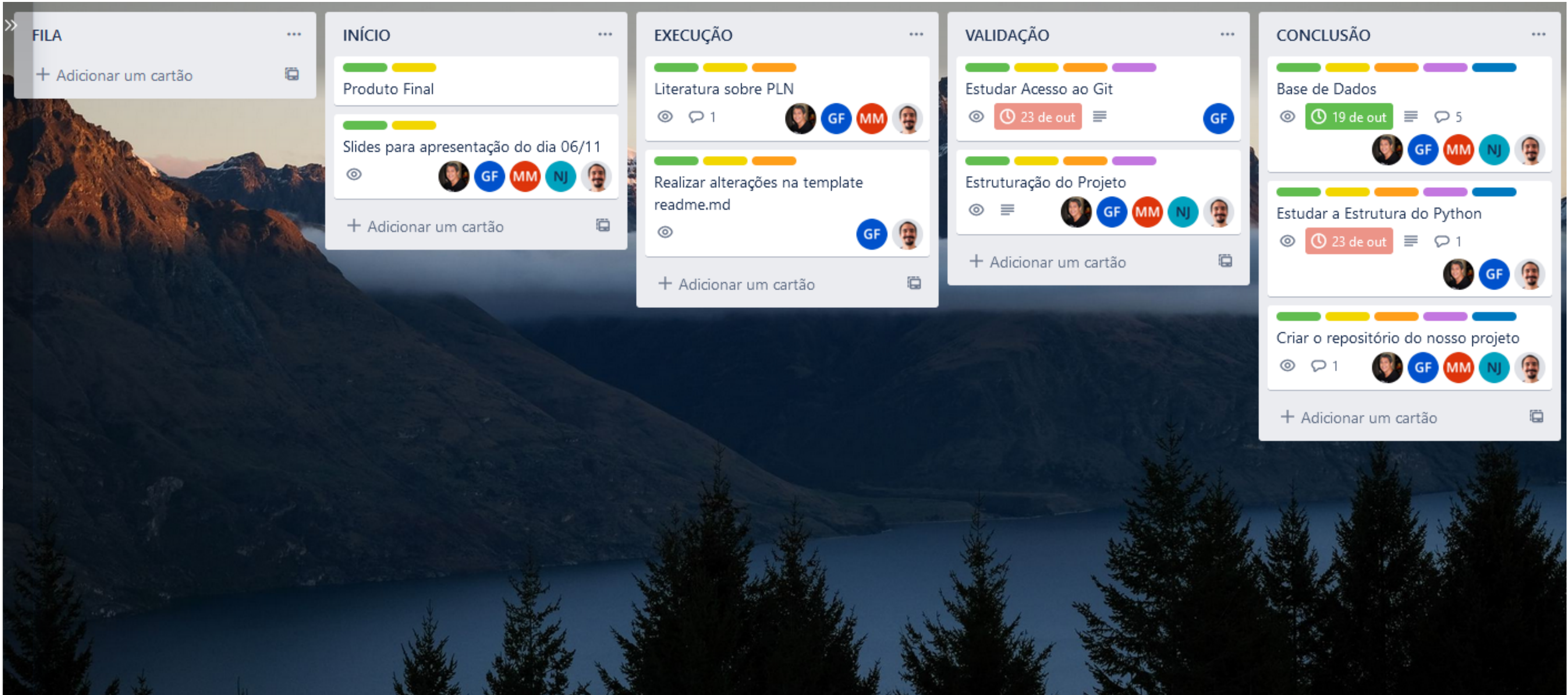


**DISCOVER EVERYONE**  
DATA SCIENCE

- *Trello*
- Dataset
- *Cookiecutter*
- *GitHub*







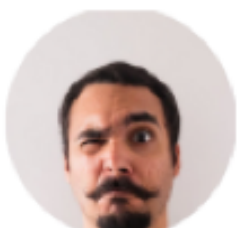
**Arildo Alves**  
Celular: (61) 9 8189-5600  
E-mail: arildoalves@gmail.com



**Nator Júnior**  
Celular: (88) 9 9422-0929  
E-mail: natorjuniorccc@gmail.com



**Maurício Moura**  
Celular: (85) 9 9936-0320  
E-mail: mouramauricio99@gmail.com

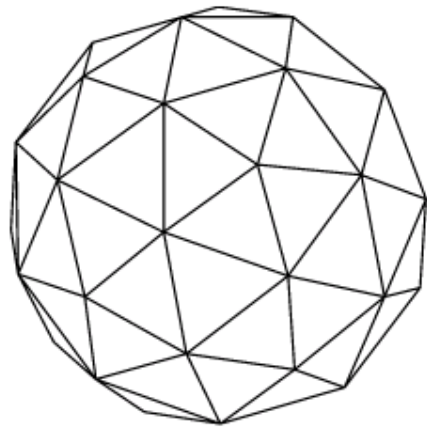


**Rodolfo Ricardo**  
Celular: (85) 9 9612-6313  
E-mail: rodolforicardotech@gmail.com



**Germano Fenner**  
Celular: (85) 9 9959-5900  
E-mail: germanofenner@gmail.com

- Fenner - Projeto
- Maurício - Dataset
- Arildo - Análise de sentimentos
- Rodolfo - Tratamentos iniciais
- Nator - Próximos passos

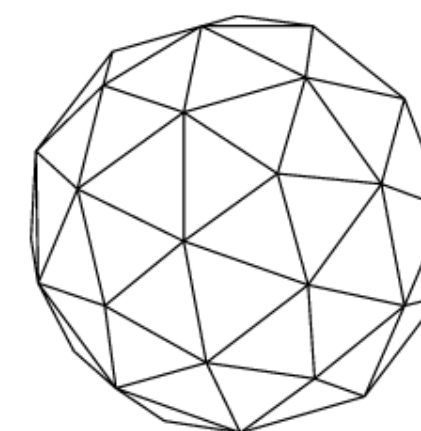


**DISCOVEREVERYONE**  
DATA SCIENCE

# O Dataset

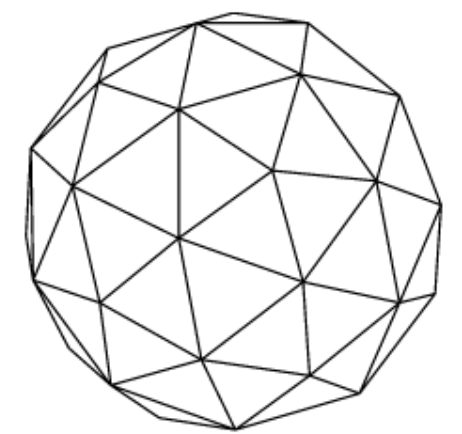


**Maurício Moura**



**DISCOVER EVERYONE**  
DATA SCIENCE

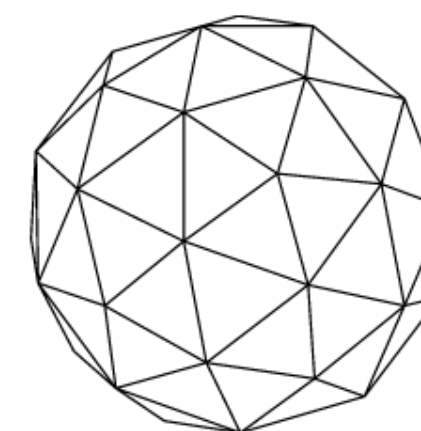
# Explicações sobre o *dataset* da Olist e suas possibilidades



**DISCOVER EVERYONE**  
DATA SCIENCE



# O objetivo



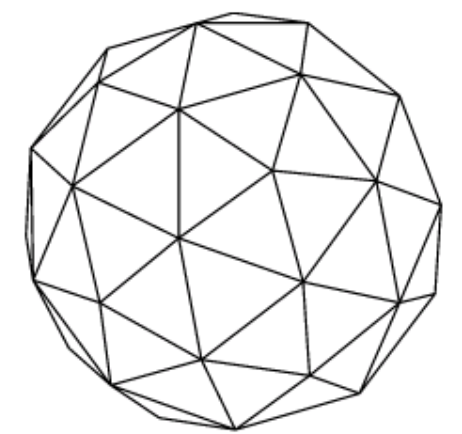
**DISCOVER EVERYONE**  
DATA SCIENCE



Nosso objetivo é a **análise** de  
**sentimentos** no *dataset*  
escolhido

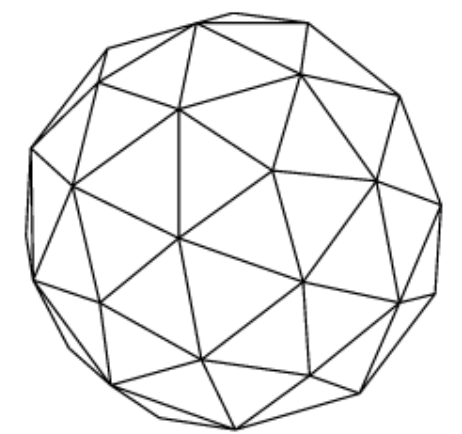


**Arildo Alves**



**DISCOVER EVERYONE**  
DATA SCIENCE

# E.D.A. ou Análise Exploratória de Dados

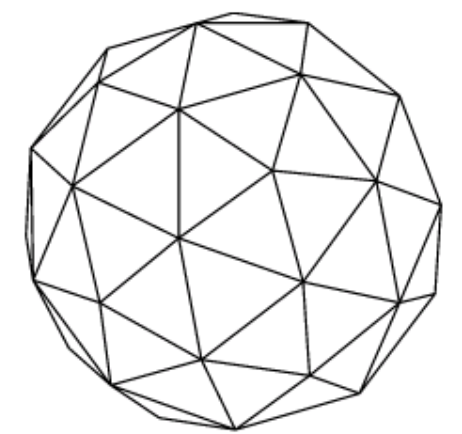


**DISCOVEREYONE**  
DATA SCIENCE

# O que encontramos até agora e **como** encontramos



Rodolfo Ricardo



**DISCOVER EVERYONE**  
DATA SCIENCE

```
In [4]: df = pd.read_csv('../data/raw/olist_order_reviews_dataset.csv')
```

```
In [5]: df.head()
```

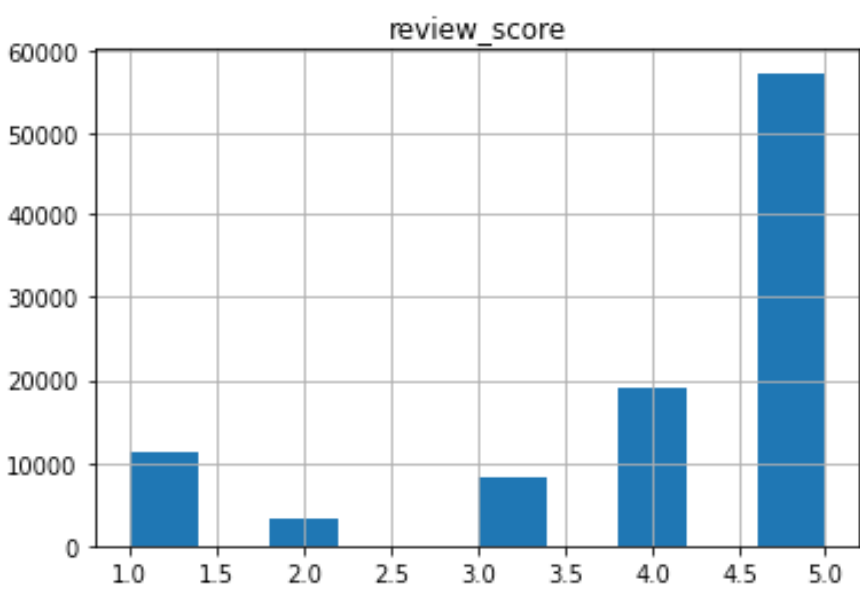
Out[5]:

	review_id	order_id	review_score	review_comment_title	review_comment_message	review_creation_date	review_answer_timestamp
0	7bc2406110b926393aa56f80a40eba40	73fc7af87114b39712e6da79b0a377eb	4	NaN	NaN	2018-01-18 00:00:00	2018-01-18 21:46:59
1	80e641a11e56f04c1ad469d5645fdfde	a548910a1c6147796b98fdf73dbeba33	5	NaN	NaN	2018-03-10 00:00:00	2018-03-11 03:05:13
2	228ce5500dc1d8e020d8d1322874b6f0	f9e4b658b201a9f2ecdecbb34bed034b	5	NaN	NaN	2018-02-17 00:00:00	2018-02-18 14:36:24
3	e64fb393e7b32834bb789ff8bb30750e	658677c97b385a9be170737859d3511b	5	NaN	Recebi bem antes do prazo estipulado.	2017-04-21 00:00:00	2017-04-21 22:02:06
4	f7c4243c7fe1938f181bec41a392bdeb	8e6bfb81e283fa7e4f11123a3fb894f1	5	NaN	Parabéns lojas lannister adorei comprar pela l...	2018-03-01 00:00:00	2018-03-02 10:26:53

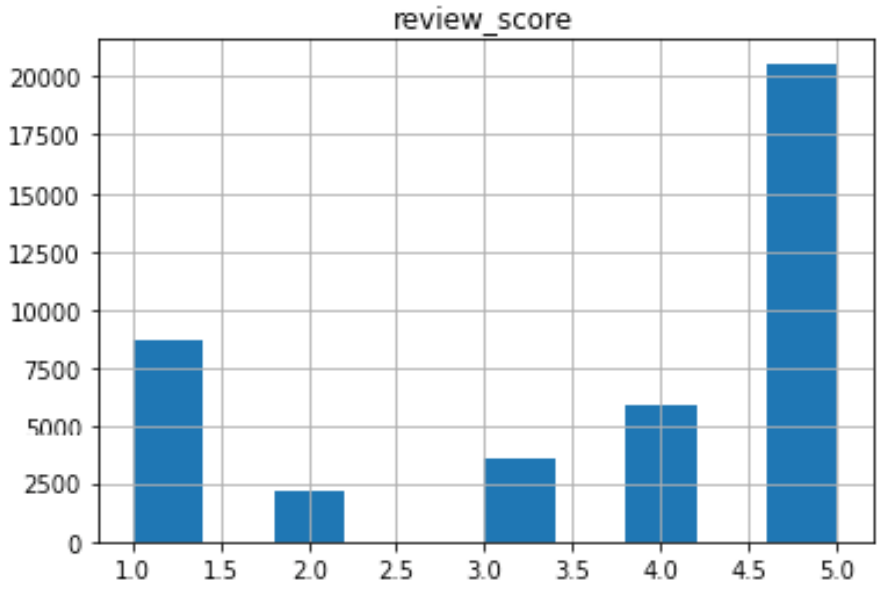
Quantificando os reviews

```
In [6]: df['review_score'].value_counts()

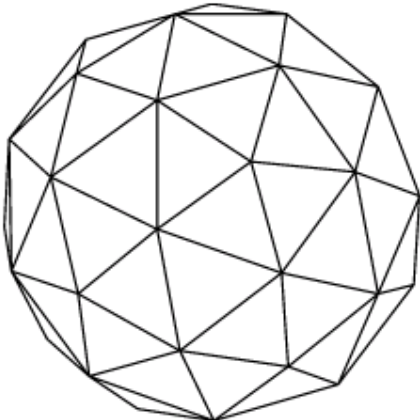
Out[6]: 5    57328
        4    19142
        1    11424
        3     8179
        2     3151
        Name: review_score, dtype: int64
```



Contagem Total  
(99.224)



Contagem de reviews  
com texto  
(37420)



Out[ 24 ] :

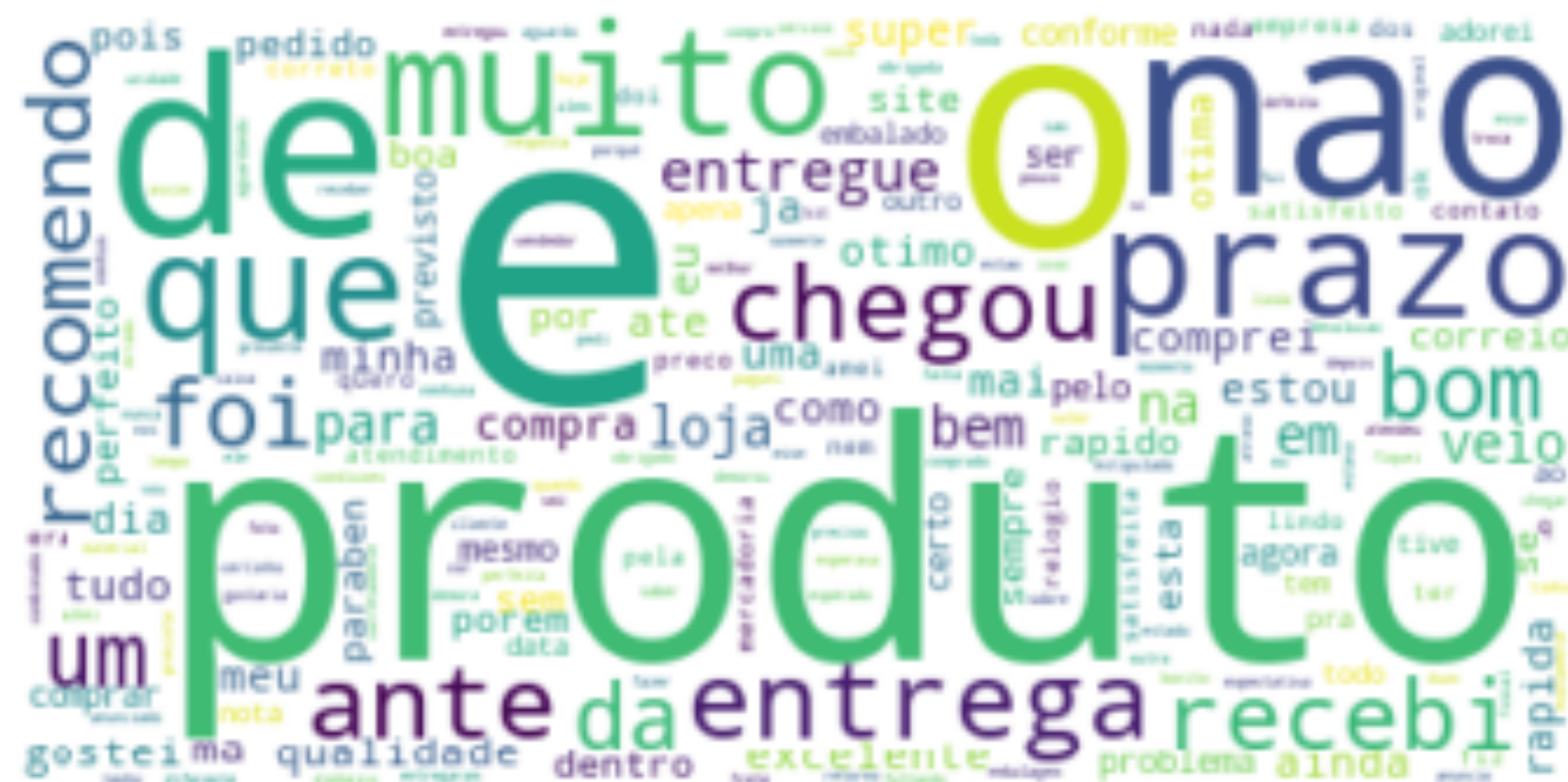
	review_score	review_comment_message	target
3	5	recebi bem antes do prazo estipulado	1
4	5	parabens lojas adorei comprar pela internet ...	1
9	4	aparelho eficiente no site a marca do aparelh...	1
12	4	mas um pouco travando pelo valor ta boa \r\n	1
15	5	vendedor confiavel produto ok e entrega antes...	1
...	...	...	...
99203	2	tive um problema na entrega em que o correio c...	0
99205	4	para este produto recebi de acordo com a compr...	1
99208	5	entregou dentro do prazo o produto chegou em ...	1
99221	5	excelente mochila entrega super rapida super...	1
99223	1	meu produto chegou e ja tenho que devolver po...	0

37420 rows x 3 columns

Depois de todo o tratamento realizado restaram **37420 linhas** para darmos continuidade ao projeto.



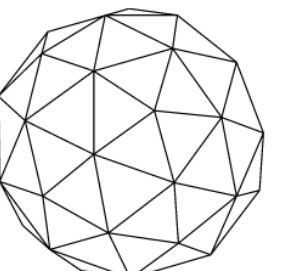




## Nuvem de palavras COM Stopwords



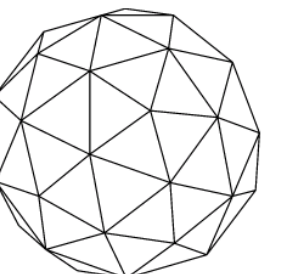
## Nuvem de palavras SEM Stopwords

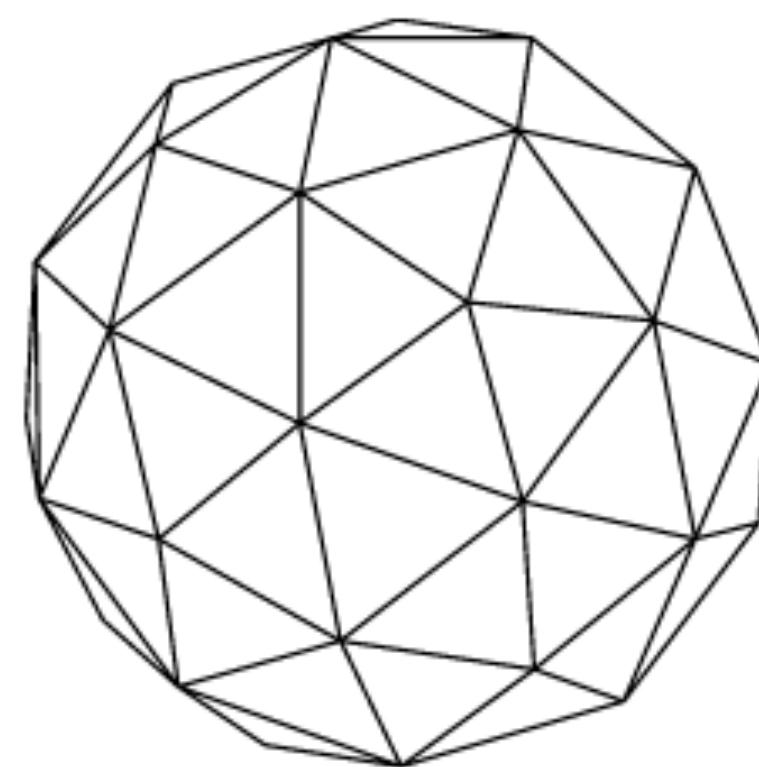


# Próximos Passos e técnicas a serem utilizadas



**Nator Júnior**





**DISCOVER EVERYONE**  
DATA SCIENCE

Obrigado!