Objectifs du CAS

Evaluations des ratings et review d'un magasin de vêtements online

Analyse des sentiments

Techniques de traitement du langage.

Technologie utilisée

Simuler grossièrement le processus d'apprentissage automatique avec l'approche du " bag of words " et "lexicon"

DATA PREP

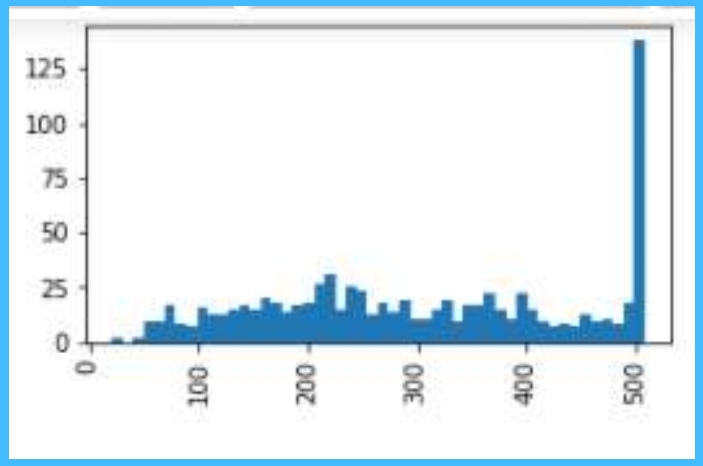
Quality check

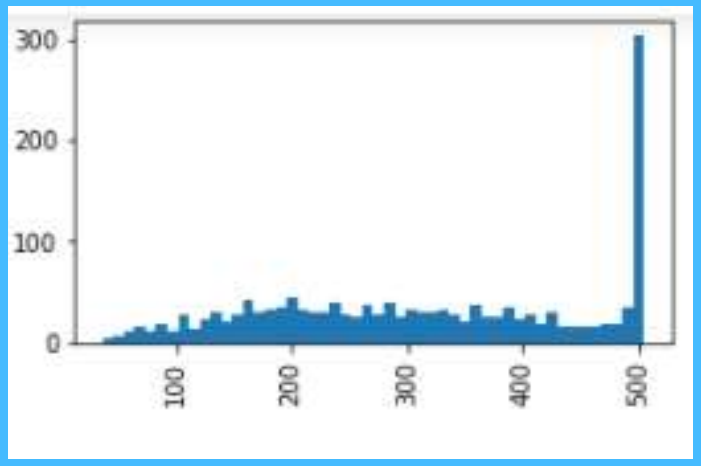
Dtype, Info,
Keep only "Review text" et "Ratings"

Data Viz

Analyse de la len() des commentaires en fonction de leurs ratings

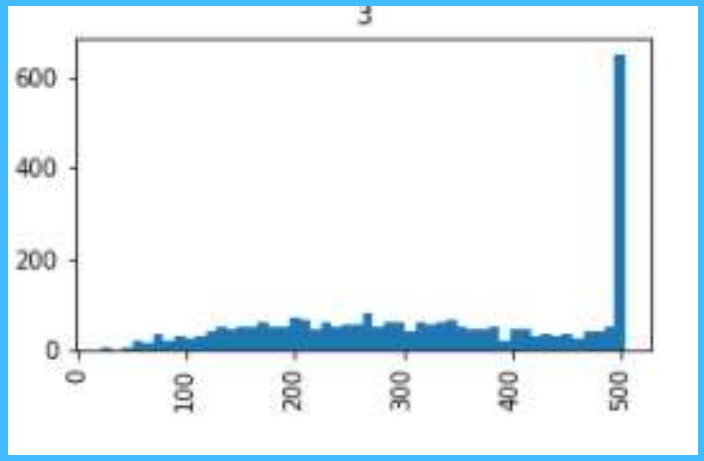
Nos étapes

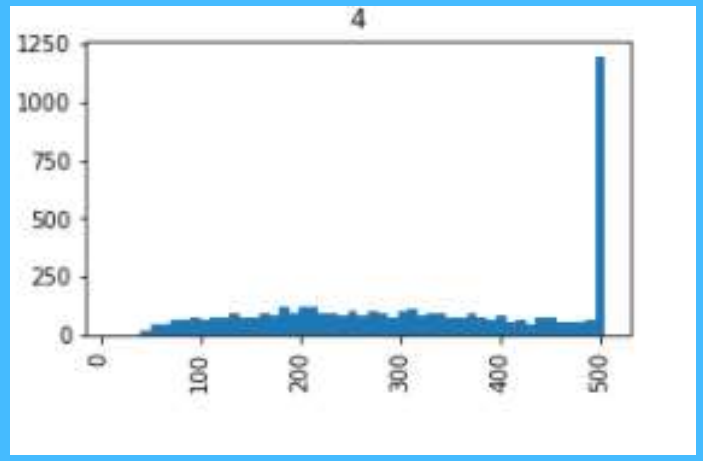


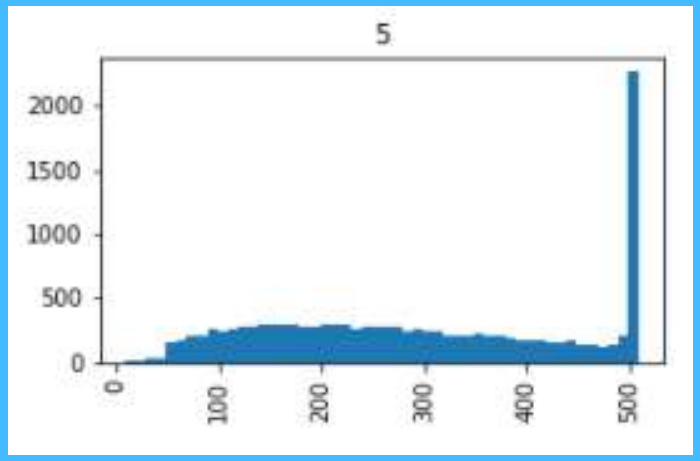


Distribution de la longueur des commentaires par rating

- Répartition similaire entre les ratings, engagement fort
- On peut imaginer que la limite des caractères pour les avis est de 500.







DATA PREP

Quality check

Dtype, Info,
Keep only "Review text" et "Ratings"

Data Viz

Analyse de la len() des commentaires en fonction de leurs ratings

Nos étapes

Seulement les valeurs non N/A

```
df[df['Review Text'].notna()]
```

Supprimer la ponctuation

```
test_punctuation=str.maketrans(",",string.punctuation)
test=test.translate(test_punctuation)
```

Low Caps

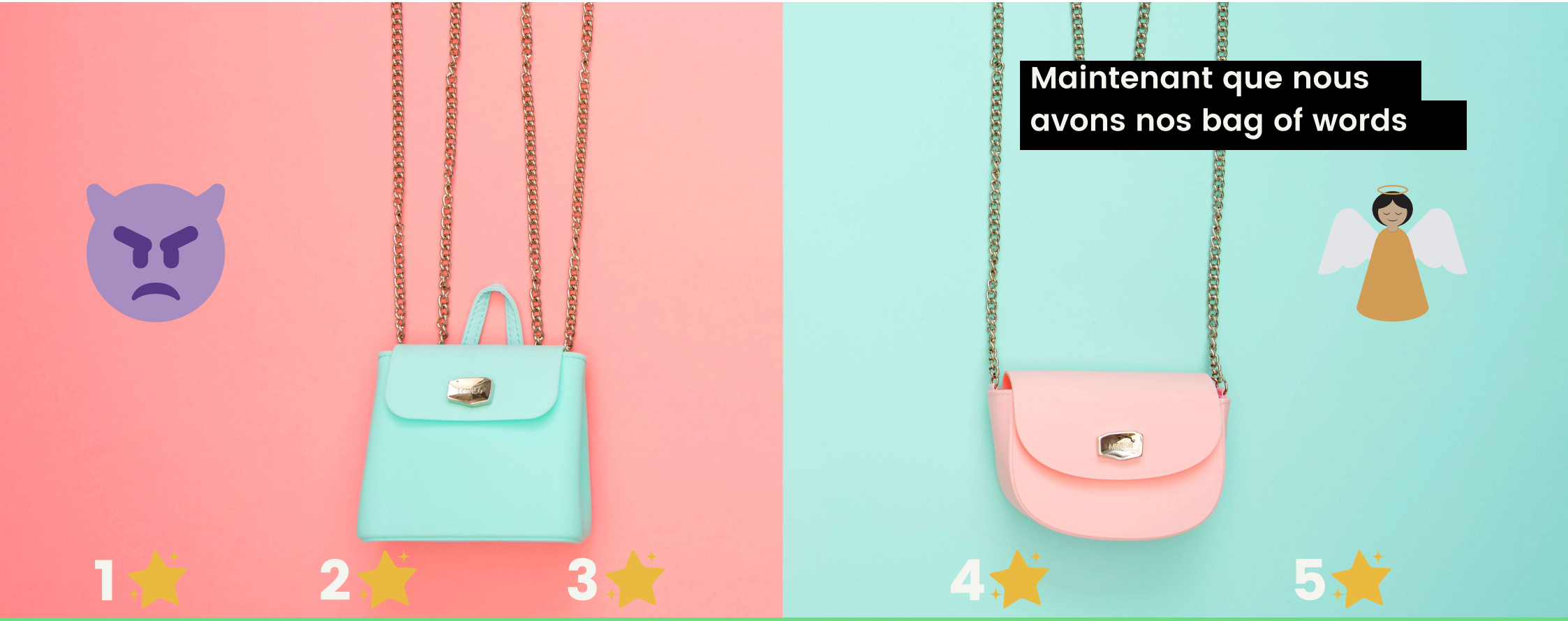
```
test=test.lower()
```

Séparer les mots dans une liste

```
test.split()
```

Suppr les mots courts et stop words

```
dfgood=df[df['Rating'] > 3]
```



Créer son lexicon Good et Bad

Choisir une pondération du Lexicon Good et Bad

Définir la frontière Good et Bad

GOOD Lexicon

Choix des mots

**['love','great','perfect','flattering','nice',
,beautiful','comfortable','cute','best','pretty']**

Pondération

+3

BAD Lexicon

Choix des mots

['small','large','would','too','even',"didn't",'than','looked',"but","not"]

Pondération

–]

"i love this dress !

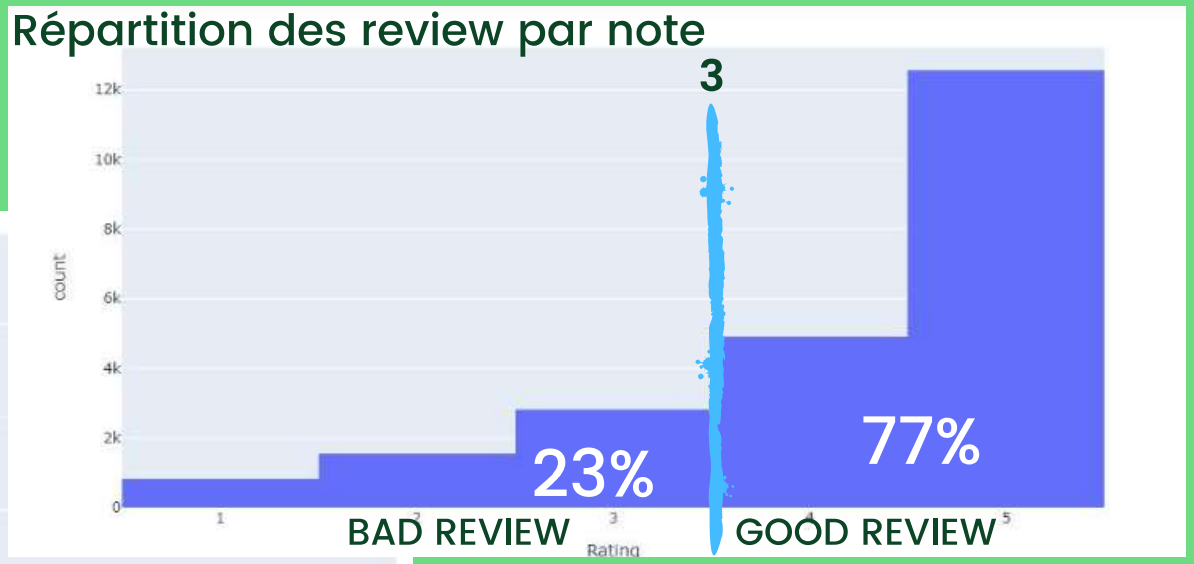
"love" = +3 points
TOTAL : +3 POINTS

"Looked nice on the model but too small for me "

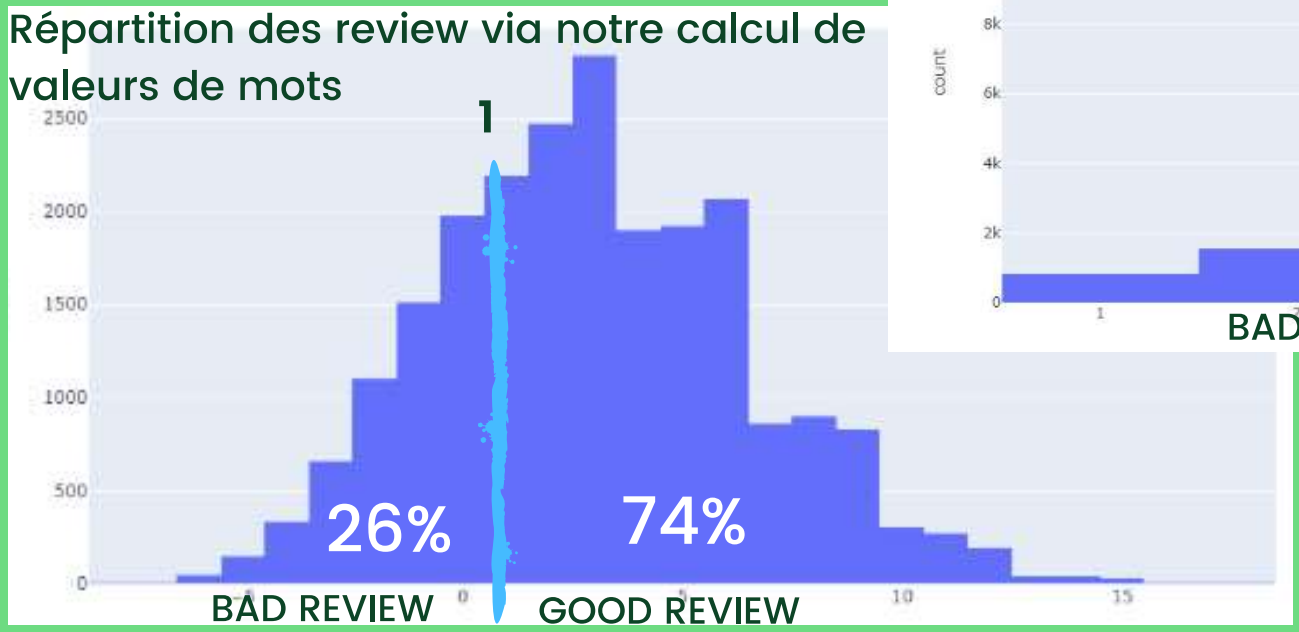
"looked" = -1
"Nice" = +3
"but" = -1
"small" = -1
TOTAL : 0 POINTS

Où situer la frontière GOOD et BAD?

Répartition des review par note



Répartition des review via notre calcul de valeurs de mots



Notre limite = 1

"i love this dress !

"love" = +3 points
TOTAL : +3 POINTS

3>1

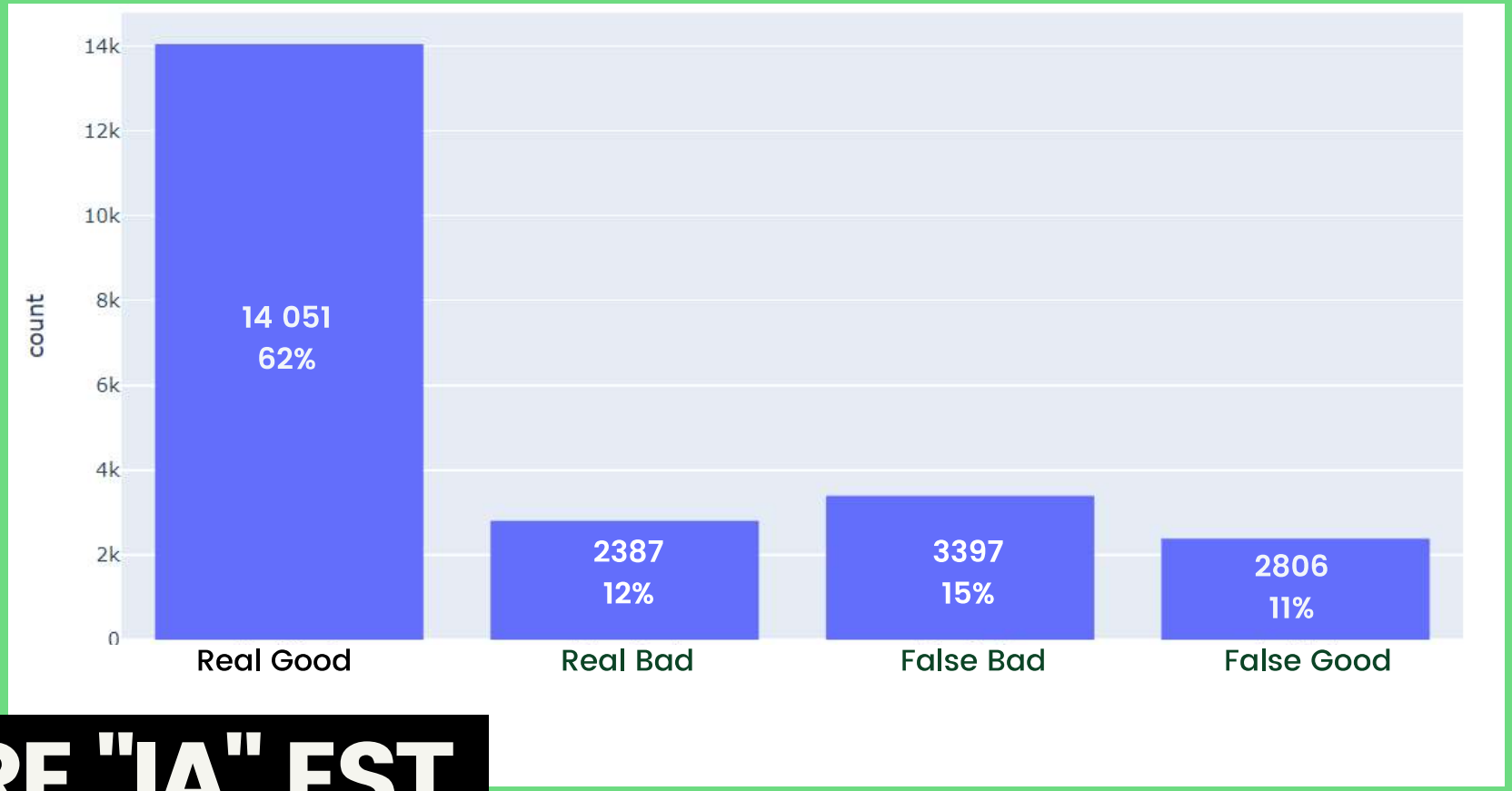
Good Review

"Looked nice on the model but too small for me "

"looked" = -1
"Nice" = +3
"but" = -1
"small" = -1
TOTAL : 0 POINTS

0<1

Bad Review



NOTRE "IA" EST EFFICACE A 74%

Commentaires positifs : 16% d'erreurs dans les prédictions

Commentaires négatifs : 58% d'erreurs dans les prédictions

Conclusion

Axes d'amélioration :

- Amélioration du Lexique BAD - GOOD
  1. Savoir prendre en compte des groupes de mots dans les lexiques.
  2. Travailler du Regex

- Amélioration du scoring BAD-GOOD
  1. Prendre en compte l'occurence d'un mot dans un même commentaire
  2. Utiliser Vader et text.blob

Questions & Answers

Merci de votre attention !