Capgemini Interview Case Study

Center Parcs Europe

Author: Mathieu Tardy

Center Parks Europe is a network of “holiday villages”.

ENTERTAINMENT & ACTIVITIES

VILLAS

SHOPPING

RESTAURANTS

22 resorts over Europe.

We will focus on **Sherwood** and **Elveden** forest in the UK.

Abundance of precious business insights in customer reviews that can be extracted with new data science techniques

25,5k reviews were extracted from TripAdvisor.

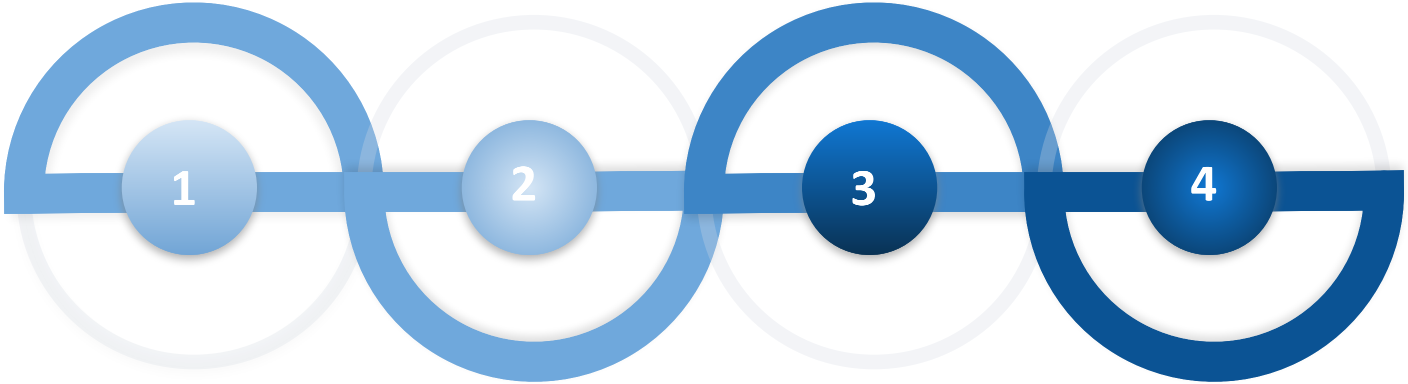
On average, they have a length of 170 words.

DATA SCIENCE NLP
Techniques

Business Insights
Recommendations

Structure

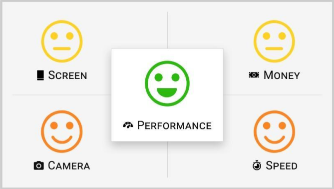
Table of Contents



The main task is to extract concrete and relevant information.

NLP: Aspect Based Sentiment Analysis

*Aspect Based Sentiment Analysis is the task of extracting **opinion terms** and **aspect terms** (opinion targets) and the **relations** between them in a given corpus (text).*



1 How to find **Aspect Words?**

- Aspect words should be domain specific (e.g: Staff)
- Comments and sentences might contain multiple aspects/topics

2 How to determine **Sentiment?**

- How to match a sentiment to the right Aspect Word ?
- How to best quantify positive and negative Sentiment ?

3 How to best **aggregate data**?

- Only focus on relevant insights (high frequency)
- Different wordings and spellings might contain the same message

Implemented from scratch Part of Speech tagging Algorithm.

NLP: Part of Speech Algorithm

*The PoS algorithm first **loops through the pre-processed review** to find **opinion words** defined by a list of positive and negative words from the NLTK open-source tool. Once the algorithm finds an opinion word, it looks for the **immediate syntactic dependence** of the opinion word and **assigns the sentiment** of the opinion word to the aspect found.*

1 Opinion words

- A list of positive and negative words are used from the "opinion_lexicon" of nltk corpus.
- The list of positive words includes opinion words such as: good, comfortable, etc.
- The list of negative words includes opinion words such as: bad, expensive, etc

2 Adjective modifier and negation

- If an adjective modifier (i.e pretty, very, etc.) is detected, more weights is added to the sentiment.
- If a negation is detected, the sign of the sentiment is then flipped

Examples Output

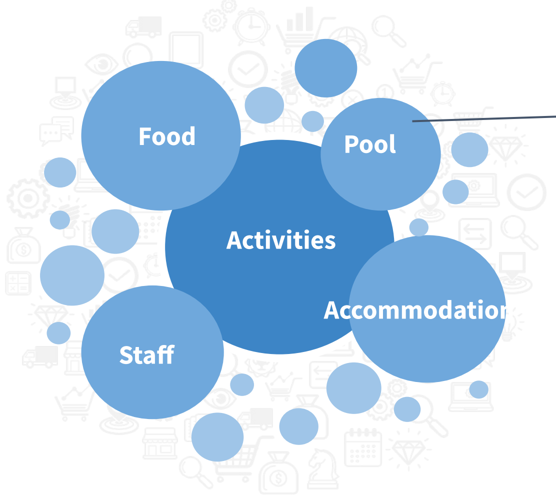
Bed : (-1.5, 'very uncomfortable')
Pool : (1, 'not small')

3 Output

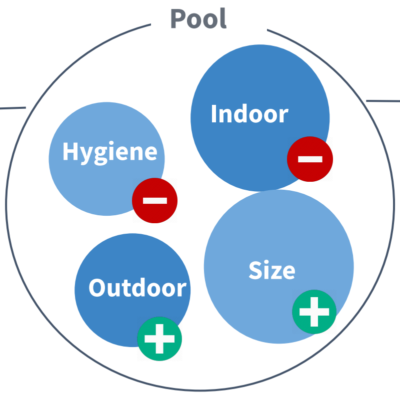
- For each pre-processed review, the output of the Part of Speech algorithm will be the sentiment of each aspect detected, as well as the opinion word describing the aspect.
- The advantage of this algorithm is that we can not only find the sentiment towards each aspect but also the opinion word describing the aspect so that we can do **deeper analysis** on how people say about each aspect.

Once we processed the Reviews, we aggregate the Data on a Topic and Aspect Level

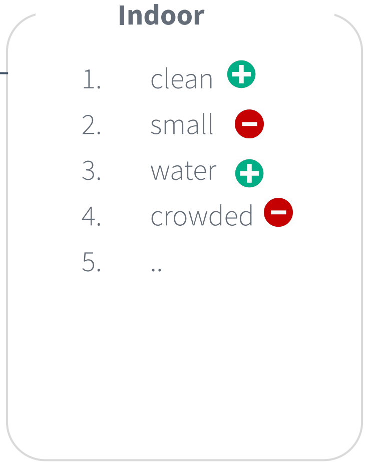
Topic Analysis



Aspect Analysis



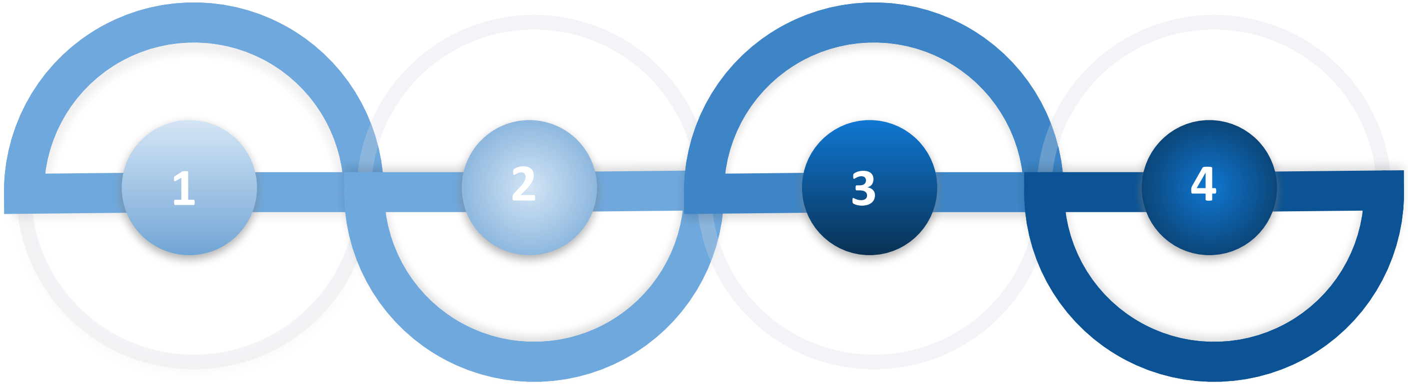
Opinion Analysis



- Using Domain Specific Topics
- Identify which topics are trending
- Looking at Aspect Words for each Topic
- Focus on frequent and sentiment Aspects
- Find out Reason for Sentiment
- Essential for Recommendation

Structure

Table of Contents



Key pain points are similar in both parcs.

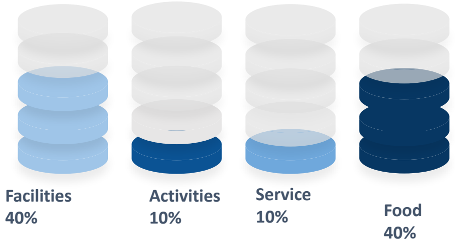
Facilities: cramped
Beach: crowded
Park: nightmare
Lodge: expensive
Park: small
Place: little

Restaurant: pricey, expensive
Meal: Cold, limited
Drinks: Smell

Pool floor: dirty
Water: cold
Activities: lack
Pool: small

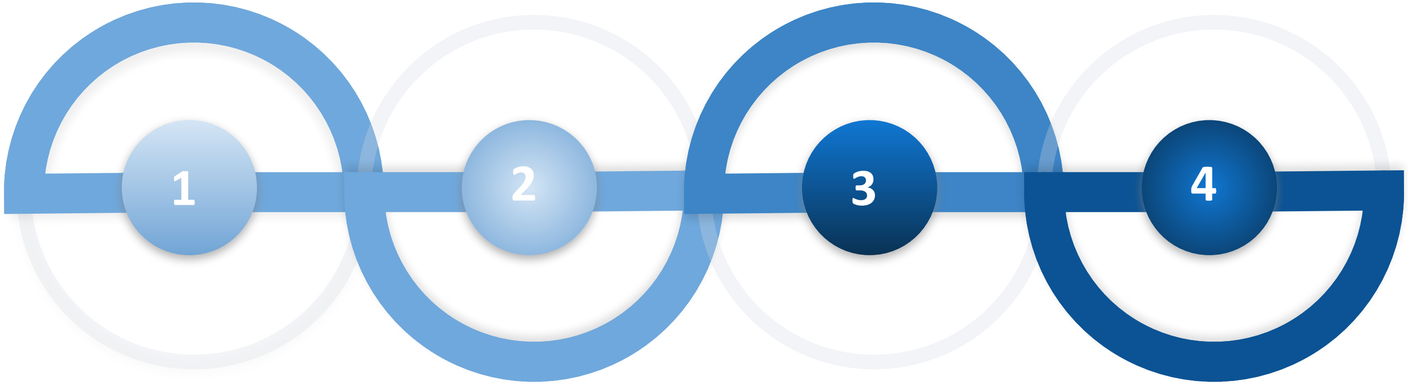
Staff: little, tired, hard
Guard: grumpy
Support: bad

Negative Reviews



Structure

Table of Contents



Based on the previous Business Insights we crafted the following specific recommendations for Center Parks Europe

Facilities

- The resorts are cramped.
- Limit the number of maximum possible customers at a specific time in the resort.
- Increase amount of free space.

Activities

- Limited activities.
- Create more sessions of popular activities.
- Increase number of activities.

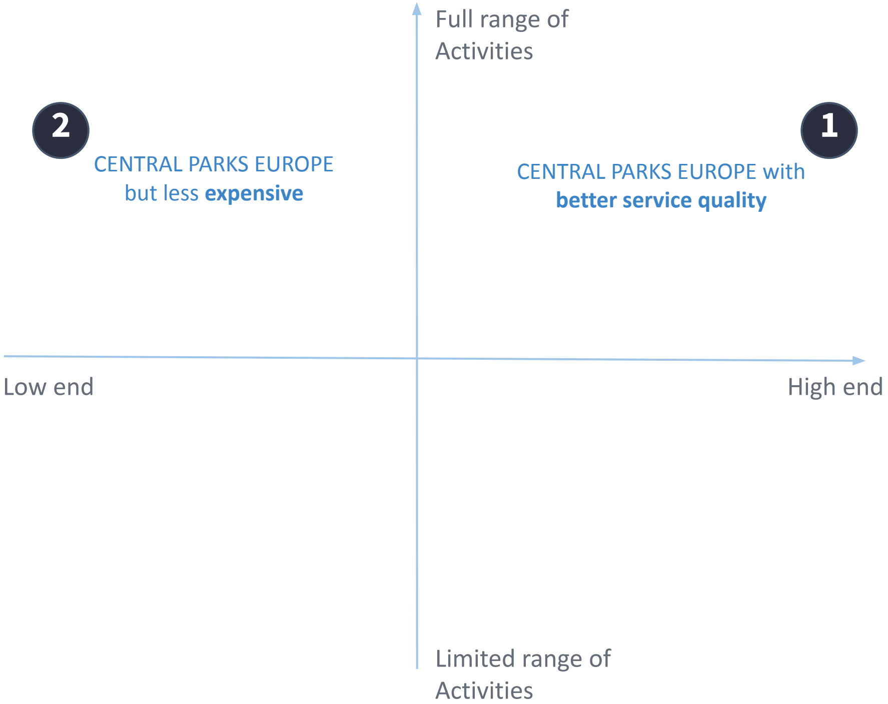
Service

- Staff is pressured by the high demand of customers.
- Think of hiring more.
- More part-time staff to compensate for higher demand.

Food

- Expensive food.
- Lack of service does not justify the price of meals.

Raising service level will improve positioning of CPE as a high end player and will justify higher prices of accommodation and food



In Conclusion, this proposed method is able to analyze customer reviews on a highly granular level.

Project Overview

Built Scalable & Reusable NLP Pipeline

Derived Domain-Specific Recommendations

Advantages of our Method

- Collected Segment specific Insight that can help to advise respective actors.
- Data Pipeline can be used to collect analyze any dataset and any topics.
- PoS NLP technique allows for low level analysis which helps to find reasons for sentiment.

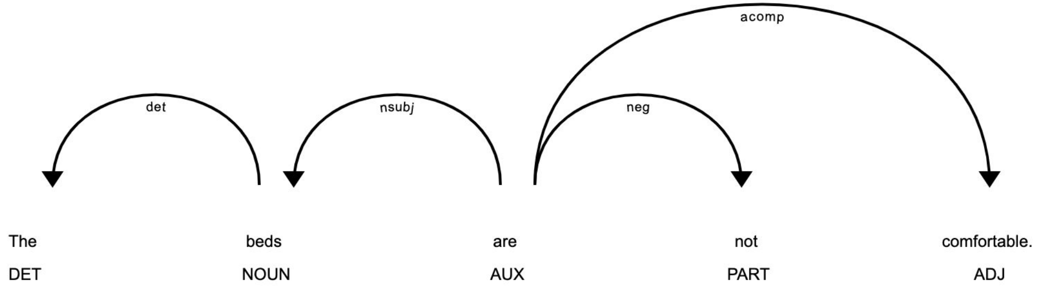
Limitations and possible Extensions

- Presented model is not able to account for the use of irony.
- Manual PoS tagging algorithms needs to be improved to accommodate complex reviews.
- Should investigate Unsupervised clustering techniques and advanced sentiment analysis method.

APPENDIX

PoS Algorithm - Part 1

1. Split reviews into sentences.
2. Determine PoS tag. Example using a sentence:



3. Output for token “beds” is (-1,”comfortable”).

PoS Algorithm - Part 2

1. For every topic, use Word2Vec to find similar words in the corpus (reviews). Similar words are the aspect words.
2. When you print result of a specific topic, also get results of those similar words.

Example of output for topic “pool”

```
---------- NEW WORD: pool --------
Mean Value of all Sentiments towards this topic: 0.46
Aggregate Number of Sentiments on that topic 2346
Positive and Negative count towards that topic (1715, 631)
Positive Opinions of that topic {'great': 222, 'lovely': 108, 'good': 96, 'clean': 81, 'nice':
Negative Opinions of that topic {'dirty': 46, 'expensive': 42, 'cold': 39, 'little': 33, 'small'
Aspects from that topic ['swim', 'garden', 'small', 'water', 'grass', 'set', 'floor', 'pool', '
Mean Value of Sentiments towards each Aspect in one topic {'swim': 0.4368131868131868, 'garden':
Count of each Aspect in one topic {'swim': 91, 'garden': 8, 'small': 2, 'water': 105, 'grass':
Positive and Negative Count of Aspects {'swim': (65, 26), 'garden': (5, 3), 'small': (1, 1), 'wa
Positive Opinions of each Aspect in one topic {'swim': {'free': 8, 'paradise': 7, 'great': 5, '
Negative Opnions of each Aspect in one topic {'swim': {'expensive': 4, 'cold': 1, 'freezing': 1,
```