



FRENCH BAKERY



BRAND ANALYSIS

Fumagalli, Durante,
Giannone, Rutigliano

CONTENTS

THE PROJECT ¹

THE DATASET ²

SCRAPING ³

REVIEW ANALYSIS ⁴

CROSS VALIDATION ⁵



CONTENTS

ALGORITHMS TRAINING ⁶

EMOTIONS ANALYSIS ⁷

DRIVER ANALYSIS ⁸

COMPLETE ANALYSIS TABLE ⁹

BEST BAKERY ¹⁰



Scenario

An **investment** firm is evaluating the purchase of one or more brands within a specific industry.

To guide this **strategic decision**, the firm wants to conduct an in-depth analysis of consumer perceptions of different brands, using data from online reviews and social media conversations.

Target

The **aim** of this project is to **analyze** consumer perceptions through a series of quantitative and qualitative parameters, to provide data-based recommendations on which brands represent the most promising investment.

the PROJECT





The Dataset

The dataset contained:

- Social
- Reviews (text)
- Date (data_created_at)
- Language (lang_value)
- Number of Twitter likes (likes)
- Rating (score_rating)
- Brand name (Players)



Data

Check the following columns of the dataset to correct any errors:

- likes
- score rating

Language

We tried to use deep learning packages to fill the empty values in the **lang_value** column, however we were unable to do so due to the inaccuracies of the tool and therefore we only analyzed reviews in French.

Data Scraping



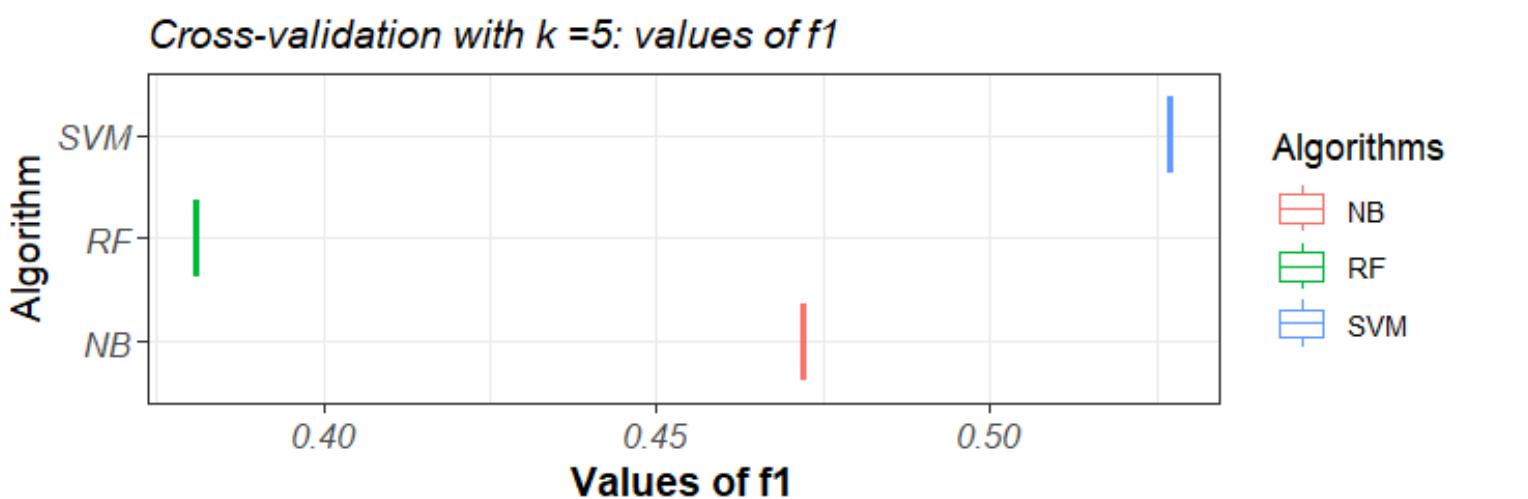
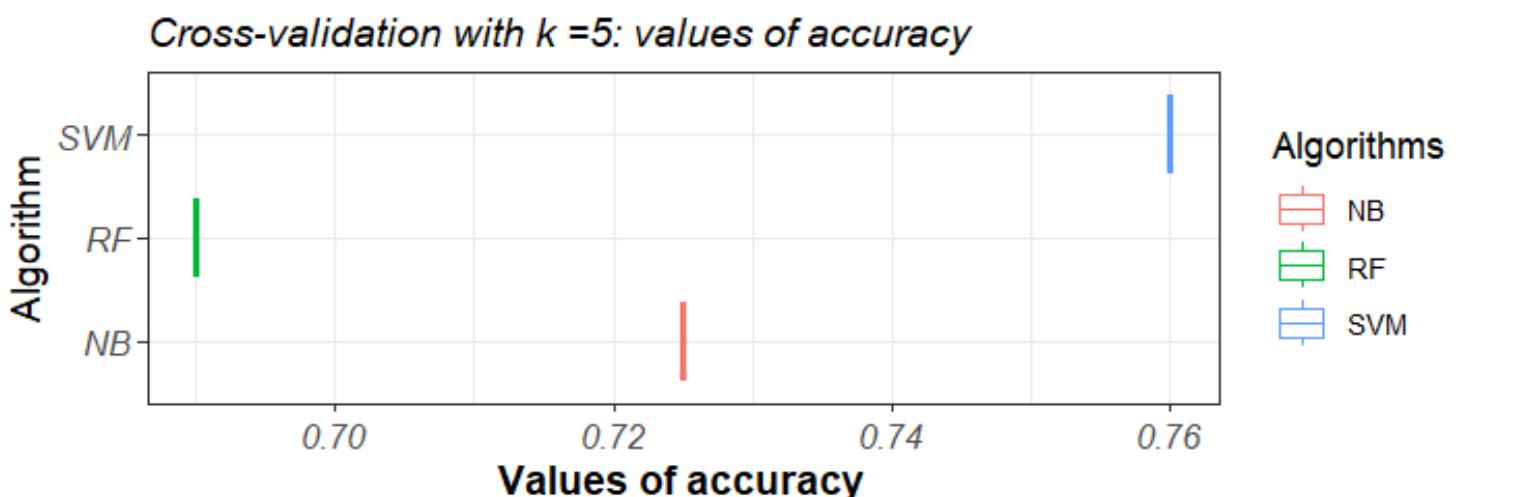
Review analysis



- 1.Extracting 200 Random Reviews in Excel (Training Set)
 2. **Sentiment Analysis** and Definition
 3. **Reinserting Sentiment Column in R**
 4. Test Set
 - a.Dataframe creation
 - b.Corpus creation
 - c.Dfm creation
 5. **Feature Match** between training set and test set
-

Cross Validation

- Training for Random Forest algorithm to get the best number of trees
- Cross Validation for all three algorithms
- After an analysis we can see that the **Support Vector Machine algorithm** has higher precision

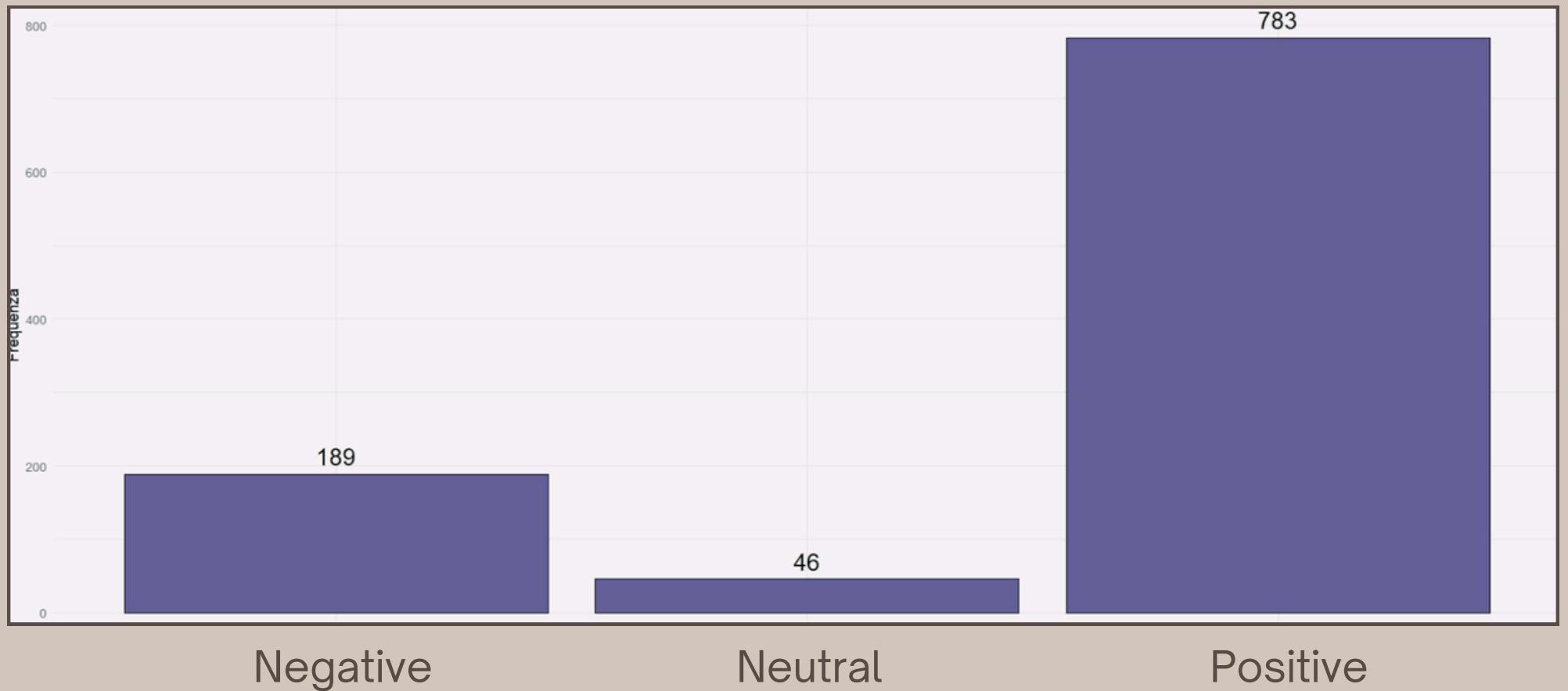


Algorithms

Training

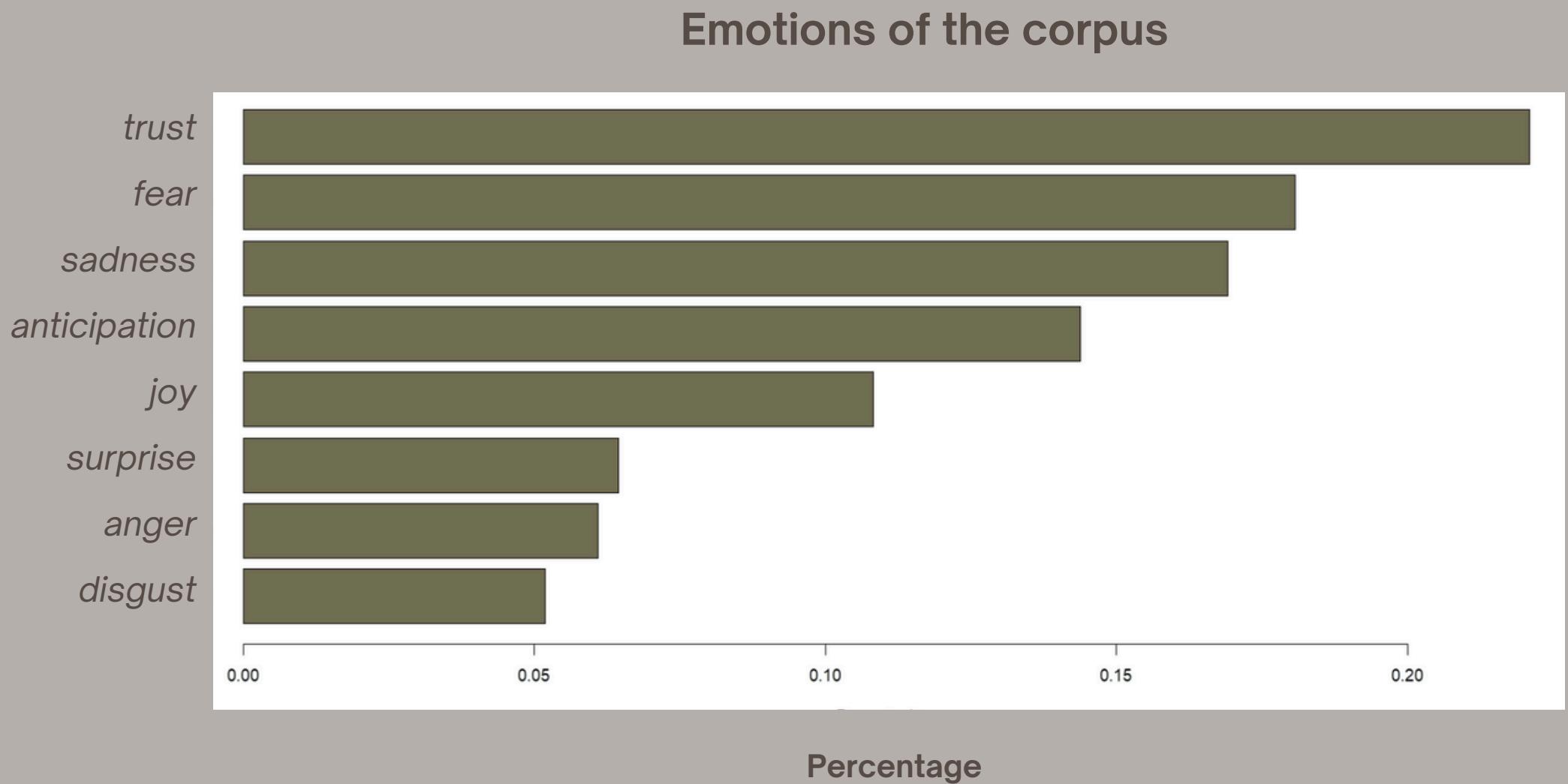
After applying the best **algorithm** to the training set we obtained the following results:

Post-algorithm sentiment distribution



Emotions Analysis

We did the emotion analysis to be able to analyze the **emotions** that filter from the reviews, the graph indicates the percentage of the specific emotion on the total reviews.



Driver Analysis



- Trimming dfm: remove all words that were not repeated at least twice in the reviews
- Recovered the 50 most present words in the reviews of the dataset
- Created a **dictionary** with 5 categories:
 - Price
 - Quality
 - Location
 - Staff (team)
 - Products

A word cloud visualization where words are colored based on their category. The categories are represented by specific colors: green for Price, orange for Quality, blue for Location, red for Staff (team), and purple for Products. The size of each word indicates its frequency in the dataset.

endroit autr
j'ai quandbeaucoup
goût demand correct pizz chocolat
avantsalad trop viennoiser temp
tous excellent boulanger rapport faut
matin riencherça sandwich mang part
lieu symp tout produit vrai aussformul
dommag client pain rapid aimabl
prendr serv plus tabl tres choixpâtisser
vu car souri c'est bon prix fait pet aucun
terr toujour bien a si accueil caf just être
non vendueux personnel agréabl pass délici
comm sup qualit servic j'ai fairalor
déjeun recommand plac top vendeur
jam prop grand peu sympath toilet
équip sourirfois bref
donc command serveux person n'est
croissanter

We created the dictionary so that we could get a sense of sentiment based on the dominant category in the review.

Final analysis table

BRAND	N.	Avg	Avg	PRODUCT				TEAM				LOCATION				QUALITY				PRICE	
	REVIEWS	RATING	SENTIMENT	N.	RATING	SENT.	N.	RATING	SENT.	N.	RATING	SENT.	N.	RATING	SENT.	N.	RATING	MEDIA REVIEWS	MEDIA REVIEWS		
Average	1018	3,47	82,45	239	3,25	75,97	492	4,10	93,11	148	2,54	70,88	62	2,76	78,08	77	3,04	72,80			
Ange Boulangerie	397	3,92	87,61	66	3,45	76,36	245	4,30	93,31	52	3,02	75,38	19	3,53	89,47	15	3,33	84,00			
Brioche Doree	172	3,17	81,40	39	3,28	82,56	64	3,62	90,00	27	2,15	71,85	15	2,80	78,67	27	3,15	70,37			
Le mie Caline	217	3,47	82,86	58	3,28	76,55	93	4,00	93,12	42	2,74	73,33	12	2,83	76,67	12	3,50	73,33			
la croissanterie	232	3,33	77,93	76	3,00	68,42	90	4,49	96,00	27	2,26	62,96	16	1,88	67,50	23	2,17	63,48			

On the graph we can see the complete analysis of the 4 bakeries, with the respective average values of the reviews for the sentiment and for the score rating. This comparison is made to have two different measures to better analyze the brands. For the **score rating** we based ourselves on the values of the dataset, while for the **sentiment** we gave a numerical value based on the emotional evaluation.

Ecco i valori assegnati:

- **Positive** - 100
- **Neutral** - 60
- **Negative** - 20

Score Rating

Bakery



Staff



Location



Products



Quality



Price

Sentiment



Staff



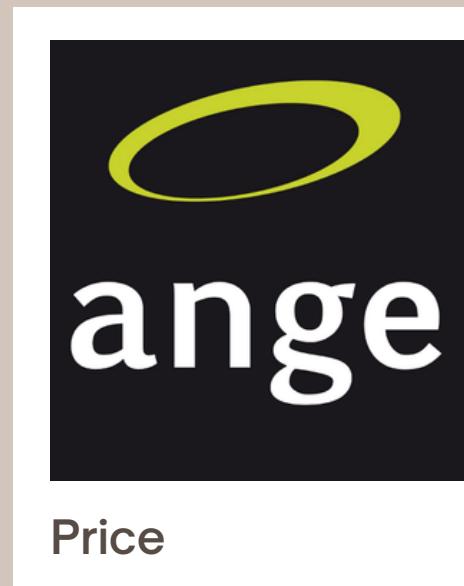
Location



Products



Quality



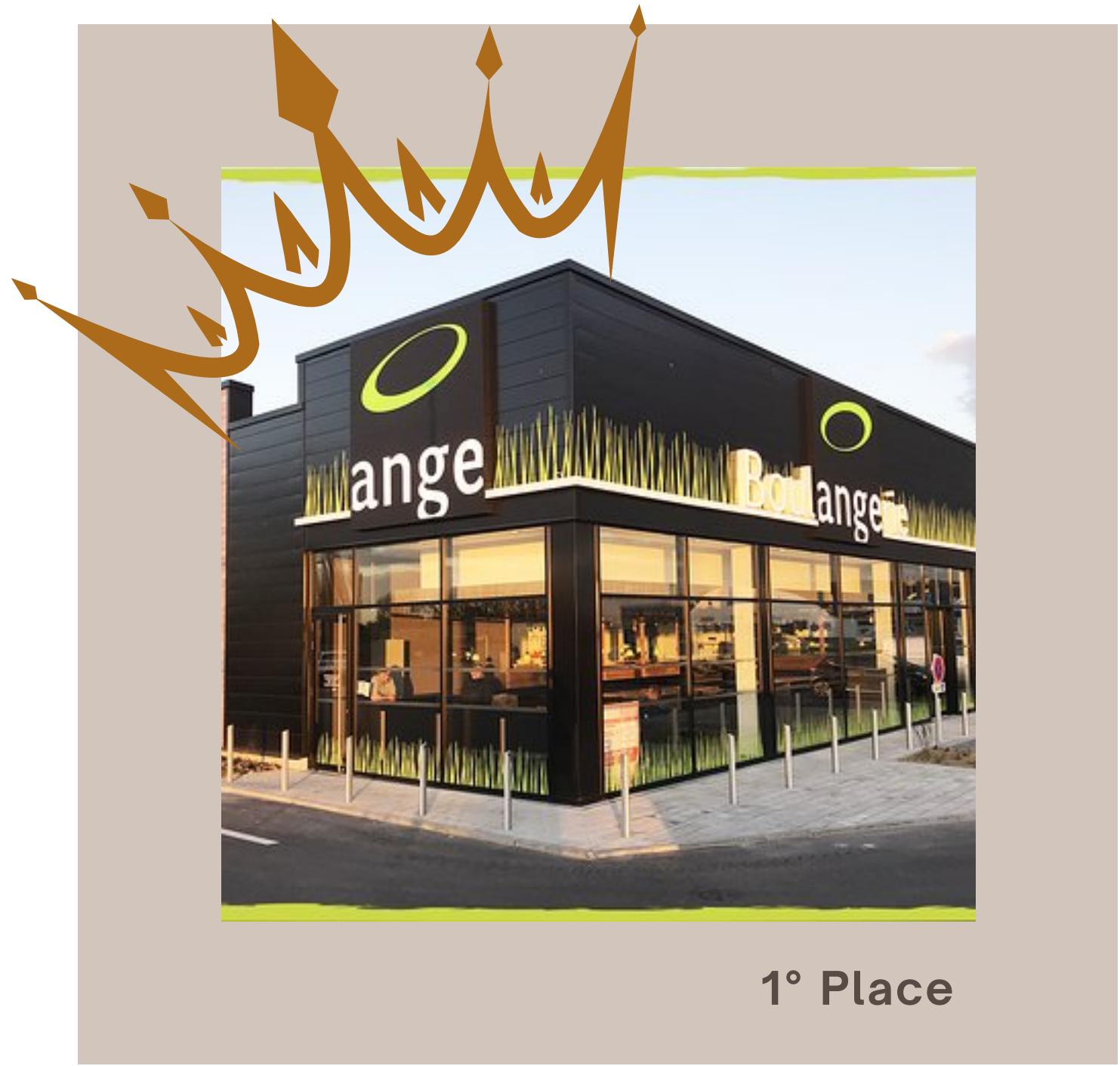
Price

Best Bakery
for category

Best Bakery Brand



2° Place



1° Place



3° Place

Merci!

