Table of Contents

Abstract: the report contains a detailed analysis of the energy supply market, in order to unlock the potential of the newly founded energy company Octopus Energy, a new player in the Italian sector. The data-driven approach ranges across different marketing techniques, in order to offer useful insights and perspectives for the development of efficient and effective marketing campaigns.

Introduction

Octopus Energy was established in the United Kingdom in 2016 with the objective of driving a global green revolution and providing an energy supply that is both sustainable and customer-centric.
The company seeks to innovate the energy sector by offering competitive pricing in both the short and long term, while also enhancing the customer experience through value-added services, such as exceptional customer support. Octopus Energy is committed to facilitating the transition to renewable energy by promoting the adoption of heat pumps, photovoltaic systems, and other smart home technologies. Additionally, it leverages artificial intelligence and machine learning algorithms to monitor market trends in real time, helping consumers optimize their energy usage and maximize savings.
As a recent entrant into the Italian energy market, how does Octopus Energy position itself against its more established competitors? What is the demographic composition of energy service users? And which customer segment should be prioritized for expansion?

Methodology

The analysis of this extensive energy consumer market employs a diverse range of advanced marketing analytics techniques, including need-based segmentation, comparative positioning, and conjoint analysis.
The data utilized in this study was sourced partly from the company and partly gathered through a rigorous, multi-phase data collection process. This process comprised in-depth interviews, sentiment analysis of customer reviews, and the administration of a survey.
Following data collection, the information was systematically processed and structured into datasets for analysis using both supervised and unsupervised learning methods. These included clustering, principal component analysis, linear regression, classification models, and other advanced analytical techniques.

Data Collection

Interviews

The data collection process began with interviews carried out on accurately chosen individuals, that represent demographically and behaviourally different groups of consumers to which the energy market is directed.
A total of 12 interviews have been conducted by a group of 5 people (*reports for references*); the questions ranged from the discovery of personal features (such as age, profession, domicile) and how they connected with their energy consumption, to the creation of an ideal energy plan to infer possible factors that may be essential for the sale/good outcome of a tariff.
Then, the interviewees expressed themselves about their feelings and perspectives on a selection of 3 companies operating in their country, allowing us to predict the aspects that influence a company's brand identity.

The interview process was the one that lasted the longest: an initial slot of two interviews (*on the professor Massimo Aliberti's advice during the meeting that took place on Teams on the 23rd of December*) have been carried out: analyzing the results, the questions have been adapted and the process proceeded with another slot of four interviews.
Halfway through the interviews, the difficulty of inferring positioning and conjoint attributes emerged, therefore, the structure of positioning and conjoint has changed again (*on the professor Massimo Aliberti's advice during the meeting that took place on Teams on the 8th of January*), and with the new slots of interviews, the factors have been extracted directly from behavioural features and segmentation-focused questions.
This process - combined with sentiment analysis - allowed us to establish five attributes for our conjoint analysis, which became four later (*under professor's suggestion of focusing more on positioning during the meeting that took place on Teams on the 21st of January*), and three attributes for positioning.
In an effort to refine our survey design and capture more nuanced customer decision-making criteria, two additional interviews were conducted using varied questions and approaches. The interviewees were asked what energy companies they were familiar with in the Italian market, mentioning Eni, Enel and A2A. Personal experience was reported only with Enel and A2A, and the highlights exhibited a preference for Enel (subsequently chosen for electricity) thanks to its user-friendly and complete app, and its straightforward approach for account transfer.
However, the non-efficiency of technical assistance led to a switch to A2A for gas provision. Although the price differences were noted to be marginal, the decision was primarily driven by speed rather than cost.

Sentiment Analysis

Sentiment analysis has been carried out at the same time as the interviews, representing another fundamental step in identifying attributes related to conjoint and positioning, having allowed us to observe what consumers consider to be a priority.
Furthermore, it allows an external observation of the company, in order to improve the targeting approach and increase business growth and profitability.

1) **Trustpilot web scraping**

In today's digital landscape, customer feedback plays a pivotal role in shaping business strategies. Online review platforms, such as Trustpilot, offer a wealth of publicly available customer opinions that can provide valuable insights into consumer sentiment, service quality, and recurring concerns.
To harness this information, a web scraping project was conducted to extract customer reviews for three energy companies—Sorgenia, Enel, and Octopus Energy. The objective was to collect structured data that would facilitate sentiment analysis, trend identification, and survey design by pinpointing the most relevant attributes influencing customer satisfaction.
The primary aim of this project was to develop a structured dataset from Trustpilot reviews, capturing key elements such as Customer usernames, Locations, Review dates, Ratings, Review content, Review titles, Experience dates. By transforming unstructured customer feedback into a structured format, the extracted data could be leveraged for sentiment analysis, topic modeling, and trend analysis, offering a deeper understanding of customer experiences.

Methodology

The web scraping process was implemented using **Selenium**, a powerful automation tool for web interactions. The methodology involved several critical steps to ensure efficient and accurate data collection.

To execute the scraping task, the **Python libraries and tools** were employed:
- **Selenium:** Automates browser interactions to navigate web pages and extract data.
- **Chrome WebDriver:** Facilitates Selenium's interaction with Google Chrome for seamless automation.

Data extraction process

1. **Setting Up Selenium:** A headless Chrome WebDriver instance was initialized to ensure efficient execution without rendering the browser interface.
2. **Navigating Through Pages:** The script iterated through multiple pages (up to **574 pages per company**) by dynamically modifying the URL.
3. **Locating and Extracting Elements:**
    - CSS selectors were used to extract key components such as usernames, review content, ratings, and review dates.
    - The page was scrolled dynamically to load all customer reviews, as Trustpilot employs lazy loading for content rendering.
4. **Data Cleaning and Processing:**
    - If the **experience date** appeared in the review content, it was removed to avoid redundancy.
    - Exception handling mechanisms were incorporated to manage missing data and prevent script failures.
5. **Data Storage:** The extracted reviews were compiled into a **Pandas DataFrame** and subsequently exported to a **CSV file** for further analysis.

Challenges

While executing the web scraping process, **several challenges** were encountered. Effective solutions were implemented to mitigate these issues, ensuring a smooth and efficient extraction process.

1. **Network interruption:**
- **Issue:** Temporary network failures could disrupt the scraping process, causing data loss.
- **Solution:** A **retry mechanism** was implemented to resume scraping from the last successful page, minimizing data loss.

2. **Dynamic Content Loading**
- **Issue:** Trustpilot loads reviews dynamically, making it challenging to extract content efficiently.
- **Solution:** The **send_keys(Keys.END)** function was used to scroll to the bottom of the page, ensuring all content was fully loaded before extraction.

3. **Avoiding IP Blocks**

- **Issue:** Frequent and automated requests could trigger Trustpilot's security measures, potentially blocking access.
- **Solution:** The script introduced **random delays** between requests to simulate human-like browsing behavior, reducing the risk of detection.

Results

The web scraping process was successfully executed, yielding a **large dataset of customer reviews** from Trustpilot. The extracted data included:

- **1274 pages processed** across the three energy companies
- **Thousands of customer reviews collected**
- **Structured CSV datasets** ready for further analysis

This dataset serves as a valuable resource for deriving meaningful insights into customer experience and satisfaction levels.

2) Improve Customer Experience

To begin, a threshold for ratings and sentiment scores within the dataset was set to categorize customer comments into positive and negative reviews. This allowed for a clear distinction between customer complaints and praises, which could then be further analyzed to understand the core issues and strengths customers identified.
Once the reviews were categorized into complaints and praises, key terms that appeared frequently within each category were extracted. This helped in identifying the specific areas where customers expressed dissatisfaction or satisfaction.

- **Frequent Keywords in Complaints**:
  - The most common terms in the negative reviews included: *'change', 'company', 'contract', 'cost', 'customer', 'email', 'energy', 'month', 'octopus', 'operator', 'price', 'rate', 'service', 'supplier', 'time'*.
  - These words pointed to issues such as dissatisfaction with billing, contract terms, energy prices, and delays in service delivery or customer support.
- **Frequent Keywords in Praises**:
  - In the positive reviews, the recurring terms were: *'company', 'customer', 'energy', 'excellent', 'fast', 'great', 'kind', 'octopus', 'operator', 'problem', 'quick', 'rate', 'service', 'thank', 'thanks'*.
  - These terms highlighted aspects such as quick and helpful service, positive interactions with customer support, and satisfaction with energy quality and pricing.

After identifying the key themes from both positive and negative reviews, clustering analysis was performed to group the reviews based on common themes. This helped in understanding the most common issues in complaints (e.g., billing or service delays) and the key strengths that customers praised (e.g., customer support efficiency or great pricing). Clustering allows for more detailed insights that could inform targeted improvements in service and customer communication.

The following clusters were identified in the complaints:

- **Cluster 1**:
  - **Keywords**: *awesome, advantage, condition, consultant, advice, courteous*.
  - **Theme**: Mixed feedback, possibly related to initial interactions or consultation issues. Customers might have expressed both positive and negative sentiments in the early stages of their engagement.
- **Cluster 2**:
  - **Keywords**: *better, assistant, deal, compliment, answer*.
  - **Theme**: This cluster represents dissatisfaction with customer support or assistant handling.
- **Cluster 3**:
  - **Keywords**: *convenience, dropped, appreciated, commercial*.
  - **Theme**: Problems with service convenience, dropped communication, or commercial terms. Customers likely expressed frustration with service interruptions, lack of follow-up, or unclear commercial offerings.
- **Cluster 4**:
  - **Keywords**: *couple, competition, active, deal*.
  - **Theme**: Complaints about competitive pricing, active deals, or promotions. This cluster indicates that customers are concerned with pricing strategies, competition, and the perceived value of current offers.
- **Cluster 5**:
  - **Keywords**: *bureaucratic, contacted, confirmation, document, debit*.
  - **Theme**: Issues with billing, documentation, or bureaucratic processes. Customers likely expressed frustration with the complexity and inefficiency of internal procedures, including issues with account management, billing confirmation, or transaction processing.

Similarly, the praises were clustered based on common themes:

- **Cluster 1**:
  - **Keywords**: *customer, operator, excellent, energy, thanks*.
  - **Theme**: Praise for overall customer service and operator efficiency. Customers appreciated the general quality of service and the professionalism of the operators handling their requests.
- **Cluster 2**:
  - **Keywords**: *quick, friendly, helpful, kind, efficient*.
  - **Theme**: Positive feedback about the friendly and quick service. This cluster reflects satisfaction with the speed and attitude of customer support staff, who were perceived as helpful, courteous, and efficient.
- **Cluster 3**:
  - **Keywords**: *service, clear, efficient, response, fast*.
  - **Theme**: Recognition of clear and efficient customer support. Customers in this cluster emphasized the clarity and effectiveness of the responses received, as well as the overall efficiency in handling inquiries or issues.
- **Cluster 4**:
  - **Keywords**: *efficient, rate, impeccable, excellent*.
  - **Theme**: Highlights high efficiency and competitive rates. Customers expressed their appreciation for both the service efficiency and the competitive pricing offered by the company.
- **Cluster 5**:
  - **Keywords**: *kind, support, great, customer*.

- **Theme**: Praise for the kind and supportive attitude of the staff. This cluster indicates that customers valued the empathy, kindness, and personalized support they received from the staff.

3) Time-based Trend Analysis

Subsequently, a time based trend analysis was carried out with the the goal of identifying and understanding patterns and changes in customer behavior, service usage, or key business metrics over time. By analyzing data across different time periods, the aim is to uncover seasonal trends, peak demand periods, and any significant fluctuations.
The first stage of the analysis involved plotting two key metrics to understand the trends over time: **Review Volume** and **Average Rating**.

1. **Review Volume Over Time:** The graph shows the number of reviews received over the given time period. The data indicates significant peaks and dips at various times. For example, there is a large spike around late 2023 and early 2024, suggesting that a notable event or change in the service may have prompted a large number of reviews. Following the peak, there's a notable decline in the number of reviews by mid-2024, followed by another smaller peak later in 2024. The volume drops sharply in January 2025.
2. **Average Rating Over Time:** The second graph tracks the average customer rating over the same time period. It shows fluctuations in ratings, with notable dips observed in specific periods such as late 2022 and early 2023 where ratings drop below 4.7. These dips might reflect a temporary decline in service satisfaction or issues during those periods. However, the overall trend appears to show an improvement over time, with ratings gradually increasing towards 2024 and beyond, indicating an overall positive shift in customer satisfaction.

Analysis of Sentiment During Peaks (January 2024 and October 2024)

The sentiment analysis performed on customer reviews during the peak periods of January 2024 and October 2024 provides valuable insights into how customer satisfaction evolved during these high-engagement months. With the sentiment scale ranging from -1 (strongly negative) to 1 (strongly positive), and 0 representing neutral sentiment, the analysis helps gauge whether the increased volume of reviews in these months is a reflection of heightened satisfaction or frustration.
The overall sentiment across the entire dataset yielded an average sentiment score of **0.328**. This suggests that the general sentiment of customer reviews is slightly positive, albeit closer to neutral than strongly positive. This value serves as a baseline, providing context for evaluating sentiment trends during the two specific peak periods —January 2024 and October 2024.
The sentiment analysis revealed different trends for the two peak months:

- **January 2024 (Peak 1)**: During this period, the average sentiment was **0.311**, which is slightly below the overall average of **0.328**. This drop suggests a mild decrease in customer sentiment during this time. While the sentiment score remains in the slightly positive range, it indicates that some dissatisfaction may have contributed to the overall feedback during January 2024. This could be due to specific issues or events that prompted reviews, reflecting mixed customer experiences with some negative feedback, but not overwhelmingly so.

- **October 2024 (Peak 2)**: In contrast, October 2024 saw a slight improvement in sentiment, with an average sentiment score of **0.364**. This score is higher than both the overall average and the sentiment observed in January 2024, indicating that customers generally felt more positive during this period. The increase in sentiment suggests that improvements in service, communication, or other customer-facing processes may have contributed to better customer perceptions during this time, leading to more favourable reviews.

Two important patterns emerge from the sentiment analysis of both peak periods:

1. **Sentiment Levels**: Sentiment scores for both January and October remain within the slightly positive range, indicating that there were no extreme fluctuations in customer emotions during these periods.
2. **Review Volume vs Sentiment**: Although review volumes were high in both peaks, there was no clear correlation between increased reviews and significantly improved sentiment. This suggests that while customer engagement increased, it was likely driven by specific issues, promotions, or events that prompted customers to leave feedback, rather than a direct reflection of improved satisfaction. Consequently, while the company may have seen an increase in customer engagement, the sentiment in both peaks remained fairly consistent, highlighting that volume alone is not a definitive indicator of customer satisfaction.

Analysis of Ratings During Peaks

In addition, the ratings analysis of customer feedback during the peak periods offers important insights into customer satisfaction during times of heightened review volume.
The **overall average rating** across the entire dataset, excluding peak periods, is an exceptionally high **4.91**, which reflects strong customer satisfaction overall. This value provides a benchmark for comparing the ratings during the peak periods.
During **January 2024**, the **average rating** was **4.89**, which is slightly lower than the overall average of **4.91**, but still indicates very strong customer satisfaction. The analysis reveals that there were **872 positive ratings** and only **15 negative ratings**. This significant disparity between positive and negative ratings highlights that the majority of customers were satisfied, with only a minimal number expressing dissatisfaction.
In **October 2024**, the **average rating** increased slightly to **4.94**, marginally higher than the overall average and first peak. This further improvement in ratings indicates that customer satisfaction continued to improve during this period. The total number of **positive ratings** surged to **1,084**, with only **9 negative ratings**, reinforcing the positive trend seen during this peak. This suggests that the company's efforts to improve service quality or address customer feedback between January and October were successful, as reflected in the stronger customer sentiment.

As a result:

1. Both Peaks Show High Satisfaction

2. The difference in average ratings between the two peaks is minimal (4.89 in January vs. 4.94 in October). However, the higher ratings in October suggest that the company may

have made improvements based on feedback received during the earlier peak period, which led to better customer satisfaction.

3. Negative ratings in both peaks were exceptionally low

Analysis of Sentiment Dips (2022-11 and 2023-02)

In the next step, sentiment analysis was conducted to understand two specific dips in average rating over time during the months of **November 2022 (First Dip)** and **February 2023 (Second Dip)**. By comparing the sentiment during these periods with the overall dataset, valuable insights into customer dissatisfaction and areas for improvement were gained.

During the **first dip in November 2022**, the **average sentiment** was calculated at **0.327**, which is slightly below the overall average sentiment of **0.331.** This represents a marginal negative impact of **-0.003** on customer satisfaction during this period. The relatively small difference suggests that while there was some decline in sentiment, it was not severe. This dip is likely due to transient issues that caused temporary dissatisfaction among customers, though the impact was minimal. This period likely reflects specific complaints from a smaller group of customers rather than systemic issues affecting the entire customer base.

In contrast, the second dip, occurring in **February 2023**, saw a more significant decline in sentiment. The **average sentiment** for this period was **0.238**, which is **-0.093** lower than the overall average sentiment. This indicates a **significant drop** in customer satisfaction compared to the first dip. The sentiment during this period suggests that customer dissatisfaction was more severe, possibly due to recurring or critical issues that required immediate attention.

To further understand the causes of these dips, a word frequency analysis was performed. The top words identified during each dip revealed recurring concerns that contributed to the dissatisfaction:

- **First Dip (November 2022)**:
  - The top words during this dip included *"whatsapp," "email," "time," "request," "bill," "activation,"* and *"contract."* These words indicate that communication delays (especially through email or WhatsApp) and concerns related to billing, contract terms, and activation processes were significant contributors to the dip in sentiment. Issues such as slow response times and service delays were likely the primary drivers of dissatisfaction during this period.
- **Second Dip (February 2023)**:
  - The key words during this dip were *"octopus," "company," "also," "bill," "activation,"* and *"contract."* These terms highlight that while communication issues persisted, there was an additional layer of dissatisfaction with the company itself, its operators, and its service quality. The word "octopus" suggests that the company's reputation or specific aspects of its service were under scrutiny, and concerns regarding activation and billing continued to contribute to the overall negative sentiment.

A comparison between the two dips revealed several common themes:

- **Billing, Activation, and Contract Issues**: These concerns were prevalent in both dips, indicating that recurring problems with these processes need to be addressed.
- **Service and Operator Performance**: The second dip revealed broader dissatisfaction.
- Additionally, **positive themes** such as fast service, kind staff, and excellent service were notably absent during both dips. This highlights that even though the company generally performs well, these qualities were not sufficiently evident during these periods of dissatisfaction.

4) Customer Feedback Classification

In this section, the goal is to classify and understand customer sentiments and extract meaningful insights using Natural Language Processing (NLP) techniques. The analysis includes text preprocessing, feature extraction, topic modeling, and classification techniques to derive insights from customer reviews. We have assumed three different clusters for customer reviews: (we obtained the best result with 3 classes, The more classes, the less satisfactory the classification result.)

- *Pricing*: Majority of reviews highlight concerns related to cost and billing.
- *Customer Service*: Positive and negative feedback on service quality.
- *Product Quality*: Comments on energy services and product features.

Main steps for this task includes:

1. Feature extraction: To convert text into numerical features, *CountVectorizer* was used. This method transforms text into a matrix where each column represents a word, and each row corresponds to a document. (At first, we used *TF-IDF* method, but the accuracy of classification result was not satisfying)
2. Topic Modeling by LDA: LDA is an unsupervised machine learning technique used to discover topics within a collection of texts. The model was trained on the vectorized text data.
3. Extracting Top Words for Each Topic: then we have extracted 12 top words for each topic which are as follows:
    - Topic1: *kind, problem, operator, quick, thanks, thank, helpful, response, kindness, speed, professional, fast*
    - Topic 2: *service, customer, customer service, whatsapp, via, excellent, fast, efficient, rate, change, always, request*
    - Topic 3: *octopus, energy, bill, rate, company, price, offer, octopus energy, supplier, first, month, also*
4. Assigning Topic Labels: Using cosine similarity between LDA topics and predefined category descriptions to assign topics to categories automatically. As a result: Topic 1 is equivalent to Product Quality, Topic 2 is equivalent to Customer Service, and Topic 3 is equivalent to Pricing.
5. Classification: this step was performed on the extracted features using **K-Means Clustering** to group customer reviews into predefined categories. The number of clusters was set to **3**, corresponding to the number of predefined categories. For each cluster, the most frequent words were identified to help interpret the meaning of the cluster which are as follows:
    - Cluster 1: *kind, operator, thanks, fast, excellent, problem, service, company, thank, quick, clear, response, great, octopus, rate*

- Cluster 2: *customer, service, customer service, excellent, octopus, rate, kind, fast, whatsapp, thanks, operator, great, always, excellent customer, company*
- Cluster 3: octopus, energy, customer, rate, bill, operator, offer, contract, month, service, company, change, supplier, thanks, energy

As you can see, based on the topic modeling we've done before Cluster 1 is related to *Product Quality*, Cluster 2 is related to *Customer Service* and Cluster 3 is related to *Pricing*.
Finally, we have assigned one category to each customer review

6. Result Evaluation: In this step, the goal was to analyze overlaps or ambiguities between clusters and predefined categories. This step helps evaluate whether clusters are distinct or have significant thematic overlap. The result is as follow:
    - Cluster 1 & 2: *excellent, thanks, fast, service, kind*
    - Cluster 1 & 3: *operator*
    - Cluster 2 & 3: *customer, octopus, rate*

    Cluster Distributions: Cluster 1 contains 6097 reviews, cluster 2 contains 1698 reviews and cluster 3 contains 1050 reviews. So as conclusion we can say:

    - **Distinctiveness of Clusters:** While clusters are mostly distinct, significant overlap exists between Cluster 1 and Cluster 2, which suggests shared themes related to service and customer satisfaction.
    - **Cluster Sizes:** Cluster 1 contains the majority of the data, indicating that its topics are more dominant.

5) Improve Marketing Strategies

The aim of this analysis is to identify actionable insights from customer feedback to refine and enhance marketing strategies. The approach focuses on extracting recurring themes from *positive* and *negative* reviews to guide targeted messaging and service improvement.

1. To extract *positive* feedback, we filter out ratings of greater or equal to 4 with Sentiment scores above 0.27 (indicating strongly positive sentiment). After extracting the 20 most frequent words in this category and visualizing them, the following key themes identified:
    - **Customer Service Excellence**:
        - Words like "customer," "service", "kind," and "thanks" indicate that customers highly value support.
        - Specific mentions of "operator" and "WhatsApp" highlight ease of interaction.
    - **Efficiency**: Words like "fast" and "always" showcase customer appreciation for quick responses and reliable service.
    - **Clarity and Transparency**: Words like "clear", "bill" and "rate" suggest that customers value transparent communication and billing.

2. To extract *Negative* feedback, we filter out ratings of less than or equal to 1 with Sentiment scores below 0 (indicating dissatisfaction). By extracting the 25 most frequent words and visualizing the following Key Issues Identified**:**

- **Service Delays and Customer service issues**:
    - Words like "waiting", "response," and "time" highlight dissatisfaction with delays and unresponsiveness.
- **Billing Problems**:
    - Words like "bill", "charges" and "overcharged" reveal common frustrations with pricing and transparency.
- **Regional or Demographic Issues**:
    - Specific mentions of "area", "location" or "availability" suggest location-based complaints.

*Check file in the **Sentiment Analysis** folder for the sentiment analysis codes (customer comment analysis, web scraping of Trustpilot, improve customer experience, improve marketing strategies, time based trend analysis, customer feedback classification)*

Survey process

After a detailed and meticulous analysis of customers' feelings and behaviour, the survey processed followed, summarizing the most interesting and potentially useful variables that emerged during the previous procedures.

The aim of survey is to gather insights and opinions on energy tariff and brands, helping to understand key trends, preferences and challenges faced by the target audience. By collecting and analyzing the responses, this survey seeks to form the foundation of our dataset and the starting point of the identification of patterns.

The survey has been developed with a structure in four sections:

1) Section 1: demographics information.
    - Age group (demographic variable, categorical)
    - Gender (demographic variable, categorical)
    - Profession (demographic variable, categorical)
    - Family members number (demographic variable, categorical)
    - Type of house (demographic variable, categorical)
    - Nationality (demographic variable, categorical)
    - Country and region of residence (demographic variable, categorical)

    More references: Need-based Segmentation

2) Section 2: Italian residents have been provided with three companies (Octopus Energy, Enel and A2A) and they have been asked to rate from 1 to 5 a set of six attributes.
    - Popularity
    - Sustainability
    - Pricing
    - Easy and speed of initial activation
    - Clarity of the contract
    - User-friendly app and online platform

    More references: Positioning

3) Section 3: information related to consumption habits.
    - Owning of the energy contract (demographic variable, binary)

- Average cost of the bill (demographic variable, categorical)
- Rank a set of house devices from 1 to 5 based on consumption (demographic variable, categorical)
- Time of the day with highest consumption (demographic variable, categorical)
- Season with highest consumption (demographic variable, categorical)
- Loyalty towards energy provider (demographic variable, categorical)

More references: Need-based Segmentation

4) Section 4: respondents have been given a set of bundles, containing each four attributes with different levels, and they have been asked to rate them from 1 to 5.

More references: Conjoint analysis

*Check file in the **Interview-survey** folder for the interviews' file and the survey.*

Conjoint Analysis

Conjoint analysis is a widely used statistical technique in market research, aimed at understanding how consumers make choices and what factors influence their decision-making processes. In our case in particular, it is needed to identify the relative importance of various features composing an energy tariff.
In order to comprehensively understand the nuanced preferences of potential Octopus Energy costumers, the research identified and selected the set of key attributes for inclusion in the conjoint analysis. These features aim to encapsulate a diverse array of factors that influence costumers' decisions and contribute to their experience, as well, as considering the complex nature of the energy sector, where diplomatic and government choices largely influence the final tariff proposal.

- The first inferred attribute is the chance of providing **digitalized tools to monitor consumption,** with the possibility of a *daily summary on the website*, a standard *bill* monitoring, and a *real-time track through app*.
  This factor has been widely mentioned during the interview process, suggesting that consumption control services are conceived as strong tools for unwanted overconsumption, allowing customers to maximize savings.

- In this sense, instead of considering the factor 'pricing', '**flexibility of the tariff**' has been selected, working perfectly as a pricing reference. The levels proposed in the bundles were a *fixed rate* (same rate no matter the time of consumption), a *two-hourly rate* (time slots are divided in two bands: F1 -daytime-, F2 - evening), a *multi-hourly rate* (a third band is added, introducing an F3 related to nighttime).

- In today's fast-paced world, consumers increasingly demand quicker and more efficient service. For this reason, the second attribute chosen is **'speed of contracts'**, with the chance of picking among a contract to be signed *on the website directly*, a *setup with quick verification* and an *activation after one week*.
  It is important to state that sentiment analysis spotted "contract" as one of the words with the highest reference linked to negative emotions, highlighting how frustration towards lack of clarity and long processes may affect business growth.

- Reviews and sentiment analysis have also emphasized the critical importance of **fast and efficient technical support**, as well as a responsive and effective customer service.

Several interviewees expressed dissatisfaction with how previous energy providers handled power shortages and overloads, leading some to consider switching to a different supplier. The options contained in the bundles were a *24h available* assistance, a *standard business hours* one and an *email support* only chance.

Analysis of ratings

In the survey, respondents were asked to rate nine different energy plan bundles, which were randomly generated using the DOE (Design of Experiments) function through a partial factorial design of experiment.

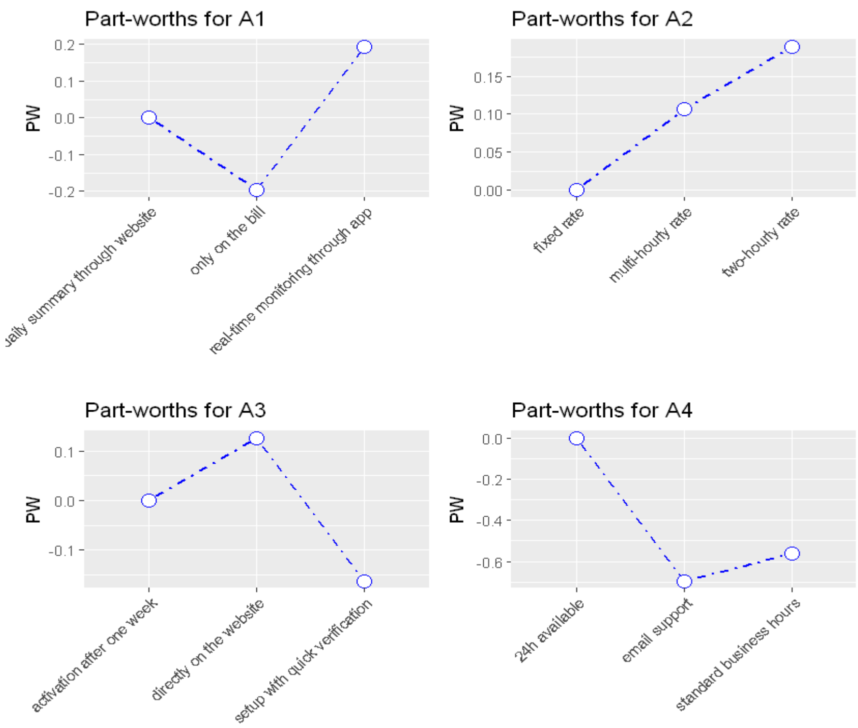
- **Bundle 1:** fixed rate, activation after one week, standard business hours, only on the bill.
- **Bundle 2:** two-hourly rate, directly on the website, 24h available, only on the bill.
- **Bundle 3:** multi-hourly rate, directly on the website, 24h available, only on the bill.
- **Bundle 4**: fixed rate, activation after one week, email support, daily summary through website.
- **Bundle 5:** two-hourly rate, setup with quick verification, standard business hours, daily summary through website.
- **Bundle 6:** multi-hourly rate, activation after one week, 24h available, daily summary through website.
- **Bundle 7:** fixed rate, setup with quick verification, 24h available, real-time monitoring through app.
- **Bundle 8:** two-hourly rate, activation after one week, email support, real-time monitoring through app.
- **Bundle 9:** multi-hourly rate, directly on the website, standard business hours, real-time monitoring through app.

The dataset has been organized with an ID for each response, an ID for each respondent, each attribute to each column and a rating from 1 to 5 to be provided for each combination of level.

Subsequently, the CONJOINT function has been used to infer the part-worth utilities, which quantify the contribution of each attribute level to the overall preference for a product.

Attributes	**Levels**	**PW**
A1	daily summary through website	0 reference level
A1	only on the bill	-0.196
A1	real-time monitoring through app	0.193
A2	fixed rate	0 reference level
A2	two-hourly rate	0.107
A2	multi-hourly rate	0.189
A3	activation after one week	0 reference level
A3	directly on the website	0.126
A3	setup with quick verification	-0.163
A4	24h available	0 reference level

A4	email support	-0.696
A4	standard business hours	-0.563



The fourth attribute (reliability of technical assistance) is the one yielding the highest weight (0.445 against 0.249 - A1 -, 0.121 - A2-, 0.185 -A3).

The bundle2 (two-hourly rate, directly on the website, 24h available, only on the bill) has the highest utility (3.412), indicating it would be the most preferred option.
The bundle 6(multi-hourly rate, activation after one week, 24h available, daily summary through website) also performs well with a total utility of 3.400.
Bundles such as bundle 1 (fixed rate, activation after one week, standard business hours, only on the bill) and bundle 4 (fixed rate, activation after one week, email support, daily summary through website) have lower utilities, making them less preferred.
Overall, respondents **prioritize real-time access to consumption data, dynamic pricing models, fast and easy activation, and round-the-clock customer support**.
Features like **infrequent updates, limited support, and slow activation** processes are viewed **negatively.**

It is important to emphasize to focus attention on **technical assistance services**, highly valued by all segments through:

1. **Customer Service Improvements:** Invest in faster response times and better training for support teams.support.
2. **Clear and Efficient Support:** Reinforcing the clarity and efficiency of customer support communication will further solidify the company's image as a reliable and clear communicator.
3. **Streamlined Communication**: Ensure customer queries and emails are resolved promptly and effectively.

Regarding the stipulation of contracts, in addition to emphasizing speed, it is advisable to enhance the aspect of transparency by providing clear and comprehensive billing statements to mitigate any potential confusion.

Product configuration and market-share simulation

Following the analysis of customer ratings, a simulation of three distinct energy tariff plans was conducted—Standard, Premium, and Low-Cost—each designed to assess consumer preferences and estimate potential market shares.
Each tariff was structured by combining different levels of key attributes identified in the conjoint analysis:
-Standard Plan: a baseline configuration with a **daily summary through the website, a two-hourly rate, contract with setup with quick verification and a standard business hours technical assistance and customer support.**
-Premium Plan: a higher-end option that includes **real-time monitoring through an app, a fixed rate, a contract to be signed on the website directly and 24h technical assistance and customer support.**
-Low-cost Plan: a budget-friendly option offering energy consumption **details only on the bill, a multi-hourly rate, an activation of the contract after one week and customer support limited to communication through emails.**
Using the part-worth utilities derived from the conjoint model, the relative preference share has been estimated for each tariff (plan):

Tariff type	**Estimated market share (%)**
Standard	31.05175%
Premium	40.69282%
Low-cost	28.25543%

All three tariff plans demonstrate competitive positioning, with the Premium Plan achieving the highest estimated market share (40.69%). Despite the higher cost associated with a fixed rate, the added benefits of real-time monitoring and 24/7 support appear to outweigh this drawback, making it the most attractive option.
The Standard Plan, with a market share of 31.05%, indicates a strong preference for a straightforward, predictable offering with reliable customer support.
Meanwhile, the Low-Cost Plan, while still appealing to a notable segment (28.25%), has a comparatively lower market share due to its limitations in service availability and monitoring capabilities, despite its affordability and flexible pricing structure.

Introduction of a new option and strategic recommendations

Building on our previous market simulation, a fourth tariff option—the Balanced Plan—was introduced to evaluate how an alternative configuration might influence consumer preferences. This new plan was designed to offer a strategic balance of flexibility, pricing, and support features, appealing to a broader consumer base. The Balanced Plan includes a **daily summary accessible via the website, a two-hourly pricing structure, quick contract verification upon setup, and 24/7 technical assistance**.

The inclusion of the Balanced Plan shifts the estimated market share distribution as follows:

Tariff Type	**Estimated market share (%)**
Standard	22.60024%

Premium	29.61725%
Low-cost	20.56501%
Balanced	27.21750%

While the **Premium Plan** continues to hold the largest market share (29.61%), it has experienced a notable decrease compared to the previous simulation (40.69%). This suggests that while premium features such as real-time monitoring and 24/7 support remain attractive, a segment of consumers still finds value consider **dynamic pricing** as an appealing option. The **Balanced Plan** has captured a substantial share (27.22%), largely drawing consumers away from the **Standard Plan** (previously 31.05%, now 22.60%). This shift is a predictable result: many users prioritize **enhanced technical assistance**, demonstrating a growing demand for reliable customer support, even in non-premium plans. The **Low-Cost Plan** (20.56%) continues to appeal to a niche market of highly price-sensitive consumers but remains the least popular choice. The limited availability of **customer support and digital tools** appears to be a key deterrent, reinforcing the importance of service accessibility in consumer decision-making.

Despite the varying responses to the revised energy plans, the four tariff options still demonstrate a notable degree of homogeneity in their market share, indicating the presence of distinct consumer segments. This supports the notion of a targeted approach that focuses on developing tariffs tailored to the specific characteristics of each segment.
Consequently, it was deemed appropriate to conduct a segment-based conjoint analysis, informed by the results of the segmentation process and the k-means clustering technique.

Segmentation

Customer segmentation is essential for understanding and addressing the diverse needs of a customer base. In this task, segmentation was done based on survey responses to identify distinct groups with common characteristics and behaviors. This enables the company to create targeted marketing strategies, improve customer experiences, and optimize service offerings to meet the specific needs of each segment. Ultimately, segmentation helps enhance customer satisfaction and drive business growth.
For the customer segmentation task, irrelevant or non-key columns were removed from the dataset to reduce noise and improve the precision of the segmentation results. Non-numerical data was then labeled and transformed into numerical values for easier analysis. Key columns used for segmentation included *age_group, income, household_size, house_type, average_bill_cost, season, loyalty,* and *times_energy_provider_changed,* as they provided meaningful insights into customer behaviors and characteristics.
Based on the analysis of the clustering results using the elbow method and other evaluation metrics (inertia, distortion score, silhouette score, and Calinski Harabasz score), the optimal number of clusters for segmenting the customers was determined to be 4. The elbow method indicated a clear inflection point at **k = 4**, where the inertia reduction starts to level off, suggesting that adding more clusters does not significantly improve the model. Additionally, the silhouette score and Calinski Harabasz score both supported this conclusion, with the highest scores occurring at **k = 4**, indicating the best balance between cluster cohesion and separation. Therefore, 4 clusters were chosen for the segmentation, providing a meaningful division of the customer base.

Customer Segmentations by K-mean Clustering Method

After applying the K-Means clustering method, four distinct customer segments were identified, each with specific characteristics:

- **Group 1** (Older, Large-House Consumers): This segment is characterized by older individuals with larger households and higher-than-average energy needs. They exhibit moderate loyalty to their energy provider. With higher energy consumption, these customers might benefit from tailored plans that address their energy requirements.
- **Group 2** (High-Income, Large-House, Seasonal Users): These customers have significantly higher incomes, larger households, and show seasonal variations in energy consumption. Their moderate loyalty suggests that personalized plans or loyalty incentives may increase retention.
- **Group 3** (Low-Income, Moderate Energy Consumers): Customers in this group have lower incomes but higher-than-average energy bills. Their energy consumption is impacted by seasons, and they exhibit low loyalty. These price-sensitive customers may benefit from energy-saving options or subsidies.
- **Group 4** (Small Household, High Loyalty, Low Seasonal Sensitivity): This segment consists of small households, often in smaller homes or apartments, with high loyalty to their energy provider. Their energy consumption does not vary much with the seasons, making them ideal candidates for low-consumption or retention strategies.

Customer Segmentations by Hierarchical Clustering Method

In addition to the K-Means clustering, hierarchical clustering was also applied as an alternative method for customer segmentation. The resulting segments are described as below:

- **Group 1** (Small, Loyal Households): This cluster represents small households or apartments with relatively low energy needs. Customers in this segment exhibit high loyalty to their energy providers. They are stable consumers with lower energy consumption due to their small household size and home type. Tailored plans focusing on stability and energy efficiency would be beneficial for these customers.
- **Group 2** (Older, Large-House, Low-Income Consumers): This segment consists of older individuals who live in larger homes but have below-average income. Despite having larger homes, their energy consumption might be less predictable. These customers are more likely to benefit from energy efficiency programs or cost-saving plans designed to reduce their energy bills.
- **Group 3** (High-Income, Seasonal, Large-House Consumers): Customers in this cluster are the wealthiest among all segments, with larger households. Their energy usage significantly varies with the seasons, making them ideal candidates for dynamic pricing plans or premium green energy solutions that align with their energy usage patterns.
- **Group 4** (Small, Seasonal, Low-Loyalty Users): This group consists of smaller households with a high influence of seasonal energy consumption. They exhibit low loyalty to their energy providers, making them ideal targets for strategies that focus on flexible or short-term energy plans to improve their satisfaction and retention.

Targeting Strategies Based on Hierarchical Clustering

The following targeting strategies have been identified for each customer segment to better meet their needs and improve engagement:

- **Segment 1** (Small, Stable Households): To reinforce loyalty among small, stable households, **retention rewards** should be provided. This can help strengthen long-term customer relationships by rewarding their consistent usage and loyalty to the company.
- **Segment 2** (Older Consumers with Larger Homes): For this segment, **energy efficiency programs** should be offered, targeting older consumers living in larger homes. These programs can help reduce their energy consumption and costs, while also promoting sustainability in their energy use.
- **Segment 3** (High-Income, Large Households): Market **premium plans or green energy options** to high-income, large households. This segment is well-suited for dynamic pricing models and environmentally conscious solutions, which could align with their energy needs and values.
- **Segment 4** (Small, Seasonal, Low-Loyalty Users): Focus on **loyalty-building strategies** and offer **flexibility** for their seasonal energy needs. This group is less loyal, so implementing short-term energy plans or flexible pricing options can increase their satisfaction and improve retention.

In the process of visualizing the clusters, the **PCA (Principal Component Analysis)** diagram did not yield satisfactory results in terms of explaining the variance. As a result, **t-SNE (t-distributed Stochastic Neighbor Embedding)** was also applied for dimensionality reduction and visualization. The t-SNE method provided much clearer separation between the clusters, allowing for a more evident and distinct visualization of the different customer segments. This approach ensured that the customer groups were effectively represented and easily distinguishable.

*Check the file in the **Segmentation** folder for the Segmentation code.*

K-Means Segment-based conjoint analysis

Segment-based conjoint analysis recognizes that different groups of consumers may have distinct preferences and behaviors. This approach segments the market into groups with similar characteristics or preferences and estimates how each segment values the various product features.

The segments considered were the result of the application of a k-means clustering, which identified four categories that can be summarised as: **Group 1** (Older, Large-House Consumers), **Group 2** (High-Income, Large-House, Seasonal Users)**, Group 3** (Low-Income, Moderate Energy Consumers)**, Group 4** (Small Household, High Loyalty, Low Seasonal Sensitivity).

Segment	Most weighted attribute	PW
Segment 1	A4	0.441
Segment 2	A4	0.434
Segment 3	A4	0.476
Segment 4	A1	0.413

- The trend in utility concerning the first attribute (**digital tools for consumption monitoring**) is largely consistent across segments, with Segment4 exhibiting the highest utility. In contrast, Segment1 stands out as an anomaly, displaying notable indifference (importance weight 0.054) across nearly all levels of the attribute. This segment is predominantly composed of older users with consumption levels above the average. While the moderate energy usage might imply a need for monitoring, the age factor could present a barrier to the adoption of digital services.
- In line with previous observations, **Segment1** derives the greatest benefit from the establishment of **price ranges**, owing to its consumption habits that exceed the average. **Segment2** and **Segment3**, representing respectively the highest and lowest income users, exhibit distinct characteristics; however, they follow a strikingly similar pattern, likely due to different underlying factors (with a slightly higher utility for **Segment3**). A high-consumption, low-income group clearly benefits from the presence of various price bands. Conversely, **Segment2** is influenced by a strong seasonal consumption pattern, making a flexible tariff particularly suitable for addressing this specific need. **Segment4** consists of small households residing in modest dwellings. Given their stable consumption levels, their sensitivity to flexibility is minimal.
- All the segments agree that the closing of a **contract** should be **quick and painless,** even though **Segment1** is the cluster yielding the lowest utility from a quick contract. This behaviour is consistent with the previous result and with the disinterest that this segments manifests towards digital services.
- **Segment4** is the only segment that does not assign the highest weight to this attribute; in contrast, all other segments demonstrate significant sensitivity to **technical assistance,** instead exhibiting greater enthusiasm for digital applications. The size of the dwelling may also contribute to this outcome, positioning this segment as ideal candidates for low consumption with a low likelihood of energy overloads or shortages. Furthermore, the predominance of apartment-style housing suggests that technical assistance is typically managed and organized by the building administration, which diminishes the priority of this factor for this particular cluster.

Product configuration and market-share simulation

Following the analysis of the ratings carried out by the CONJOINT function on each segment, the reaction to the three initial configurations have been analyzed:

Segment	**Standard**	**Premium**	**Low Cost**
Segment 1	32.51534%	35.17382%	32.31084%
Segment 2	28.46034%	44.01244%	27.52722%
Segment 3	31.39073%	42.11921%	26.49007%
Segment 4	32.41650%	39.68566%	27.89784%

- **Segment1** derives a relatively consistent utility from all tariffs, with a slightly higher preference for the Premium option, primarily due to the enhanced technical assistance. This outcome suggests that price remains a significant factor for this cluster, and that the additional services offered by the Premium plan may be considered interchangeable if more competitive pricing options are made available.

- As the segment with the highest income, **Segment 2** exhibits low sensitivity to basic options without additional services. They value the extra features (like real-time monitoring and 24h technical assistance), that align with their willing to pay for these added services.
- **Segment 3** displays a similar pattern, with a slightly higher market share for the Standard and Low-Cost configurations. This result is particularly interesting: despite their distinct features, Segment 3 and Segment 2 exhibit similar behavior toward the tariffs,suggesting that Segment3 could be additionally segmented to to highlight any differences that lead one segment to tend towards a premium offering, and another segment to tend towards a more standard one. Both clusters consist of households with lower-than-average ages, which may contribute to a shared enthusiasm for digital tools. On the other hand, the need for efficient technical assistance likely stems from different circumstances: for the lower-income cluster, assistance may be needed since new device may be unaffordable, while for the higher-income cluster, the larger household requires more attention and support.
- **Segment 4** shows the strongest preference for the standard tariff. Given their low seasonal sensitivity, the simple structure of the Standard plan suits their stable energy consumption patterns. At the same time, the Premium Tariff also has a notable preference, consistent with the lack of interest this cluster has in dynamic pricing.

Introduction of a new option and strategic recommendations

Segments	Standard	Premium	Low Cost	Balanced
Segment1	23.41679%	25.33137%	23.26951%	27.98233%
Segment2	20.93822%	32.37986%	20.25172%	26.43021%
Segment3	22.67943	30.43062%	19.13876%	27.75120%
Segment4	23.77522%	29.10663%	20.46110%	26.65706%

The balanced option primarily absorbs market share from the Standard and Premium configurations, though it fails to attract Low-Cost users, who prefer to maintain lower prices and a reduced range of additional services. The decrease in market share is approximately consistent across all segments, indicating that the needs related to the tariff plan are quite similar despite different demographic characteristics, and that an overly fragmented set of offers may not yield the desired results.
Despite the varying characteristics of consumers, energy consumption needs are relatively similar, and a limited set of tariff options could effectively meet the demand.

Based on the results of the customer segmentation, the following targeting strategies can be implemented to effectively reach and address the specific needs of each identified customer group.

- **"Segment 1 – The Boomers":** As previously mentioned, this group demonstrates a modest increase in utility and a general disinterest in digital tools or online contract signing, likely due to their higher-than-average age and limited digital literacy. However,

their larger-than-average household size results in a positive response to dynamic pricing and the availability of technical assistance.
This segment can be effectively targeted with a **Balanced tariff,** which aligns well with their desire for both affordability and support.: providing a relatively low level of digital tools does not diminish the utility for these consumers. Instead, their utility is maximized through cost-effective tariffs and efficient customer support.
Further segmentation could reveal more clearly whether the **Low-Cost option** could still be relevant for highly price-sensitive individuals.

- Segment 2 and Segment 3 - **"The Opposite Poles"**: These two segments exhibit the most contrasting characteristics, yet their patterns are strikingly similar. This reveals important insights: income may not be the most significant demographic factor, whereas age could play a more crucial role, and different needs can be addressed with the same solution, but they must be marketed in distinct ways.
Premium and **Balanced** tariffs should be the focus for these segments, with a slight emphasis on the **Balanced** option for Segment 2 to cater to those who desire more services but at a lower price point than the **Premium** plan. Highlighting the value of comprehensive support and fixed pricing will be key in retaining this group. On the other side, **Low-Cost** tariffs are not likely to resonate with this group.
- Segment 4 - “**The loyalists”**: the loyal component of this cluster suggests a choice mostly determined by the popularity of the company. This segments showed a little reaction to dynamic pricing - due to the low consumption of a small household - and to an efficient technical assistance - due to the apartment-based component of the segment. Better to prioritize the Premium and Balanced offers, focusing on the value of technical assistance and high-quality service. Given their high loyalty, Premium can be marketed as a premium option for long-term customers, while Balanced as a more affordable yet feature-rich alternative.

Positioning

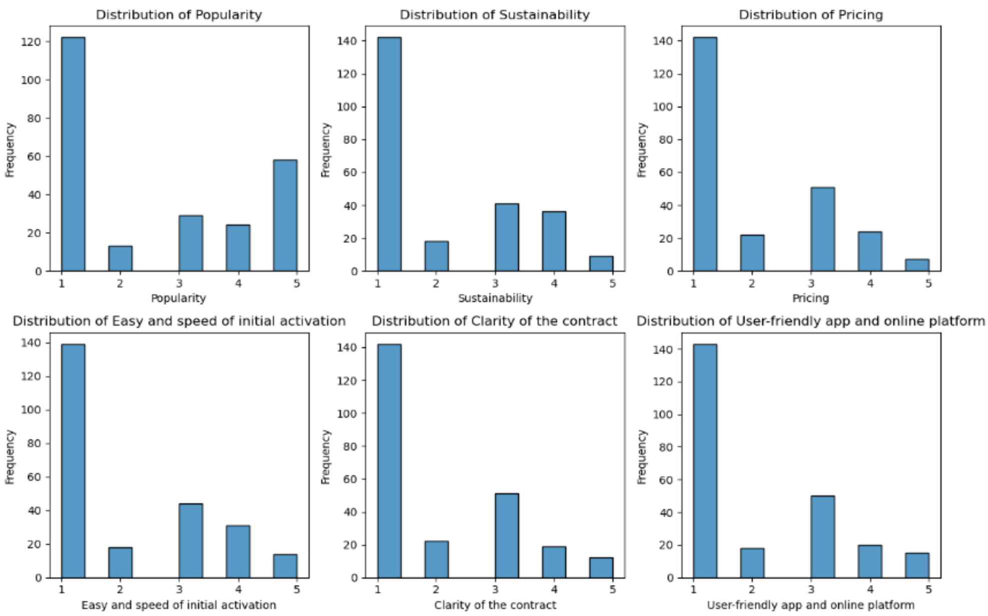
As a new company, understanding how Octopus Energy image is defined in the minds of consumers is essential to developing effective marketing campaigns.
In this regard, a set of questions containing six attributes referring to three energy companies (Enel, A2A, Octopus Energy) was inserted into the survey, and Italian consumers were asked to rate them from 1 to 5. The goal was to understand how Octopus Energy is positioned alongside larger companies with well-established brand images in the minds of consumers across several key dimensions that influence the decision-making process for selecting an energy provider.
The choice of attributes was based on the same process as the conjoint.

- Interviews have shown **that brand popularity** is a significant factor in choosing an energy company, to the point that additional services or comparisons between tariffs may not even be taken into consideration. Brands that maintain a strong public presence through strategic advertising, influencer partnerships, and consistent messaging often experience higher customer engagement and loyalty. Popularity serves as a foundational attribute, as brands that are widely recognized tend to enjoy greater consumer confidence and market dominance.
- **Sustainability** measures a company's commitment to environmentally friendly practices. Modern consumers, especially younger demographics, prioritize brands that adopt eco-friendly operations, ethical sourcing, and carbon-neutral initiatives. Businesses that emphasize sustainability not only contribute to global conservation efforts but also strengthen brand loyalty by aligning with socially conscious values.

- **Pricing** is a crucial determinant of consumer preference, as it directly impacts the perceived value of a product or service. Transparent pricing models that eliminate hidden costs can enhance consumer trust and reinforce positive brand perceptions, in addition to attracting the segment of consumers that the company targets.
- The **ease and speed of initial activation** refer to how simple and efficient it is for customers to begin using a brand's product or service. A streamlined onboarding process minimizes friction and accelerates customer adoption rates.
- In sentiment analysis, contract issues and the lack of transparency of the latter were among the hottest topics of negative reviews, making the **clarity of contract** a suitable factor for analysing brand image, as it represents the company's first impression.
- In the digital age, a brand's online presence significantly impacts customer engagement and satisfaction. Interviews highlighted how a **user-friendly app and online platform** that consumers can easily navigate for services, making purchases, and accessing support

Initial Data Exploration

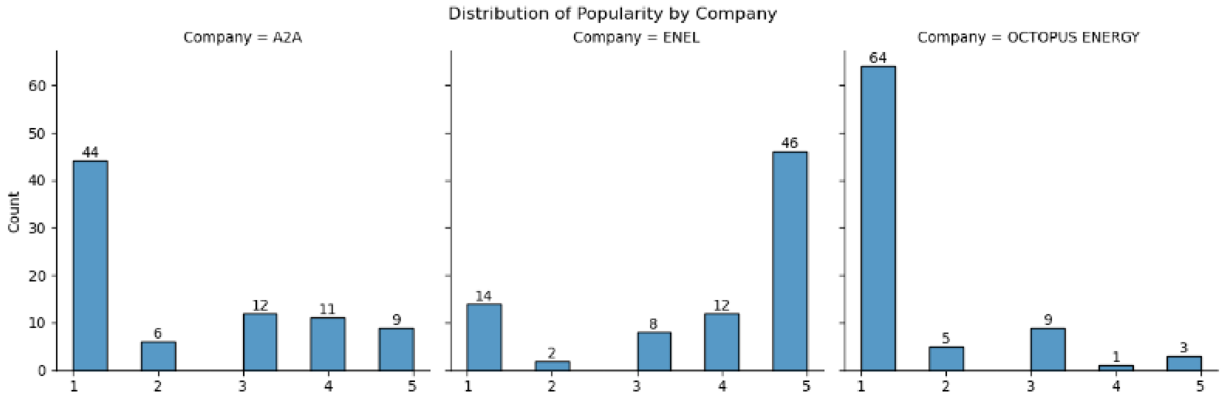
First of all, records with missing values were removed to ensure data consistency. This resulted in a cleansed and final dataset suitable for in-depth examination.
A close examination of the overall rating distributions across all respondents reveals a pronounced tendency toward lower scores (1–2) in several critical service dimensions. Notably, **Clarity of the Contract** and **Ease of Activation** are heavily skewed, with a sizable number of respondents rating them at the lowest end of the scale. This pattern signals that many customers may find the contract terms confusing or overly complicated, while also encountering difficulties or frustrations in initiating service.
By contrast, in categories like **Popularity**, there is a cluster of higher ratings (4–5) for certain providers—particularly ENEL—indicating a robust brand image and recognition among a segment of consumers. This highlights a stark contrast: while some companies enjoy strong customer endorsement in terms of brand awareness, the same companies or others may fall behind when it comes to clarity, ease of onboarding, and possibly other operational aspects.

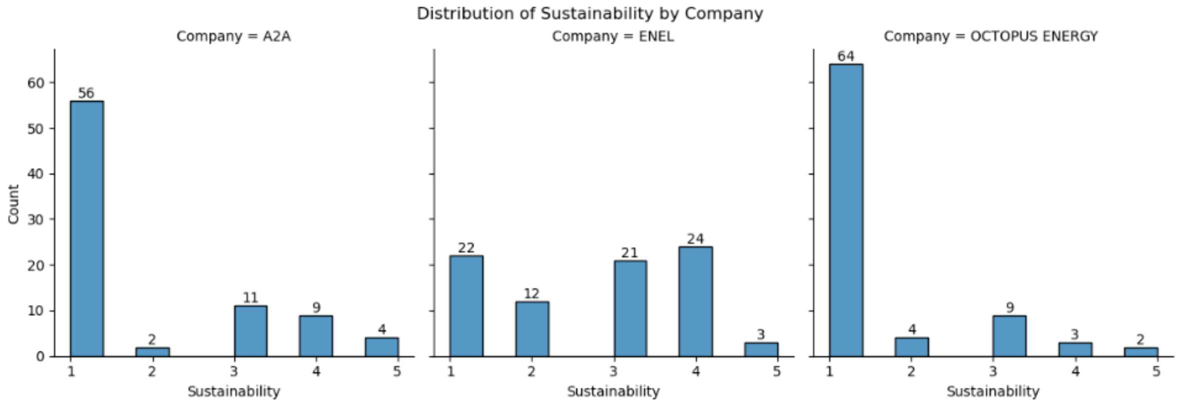


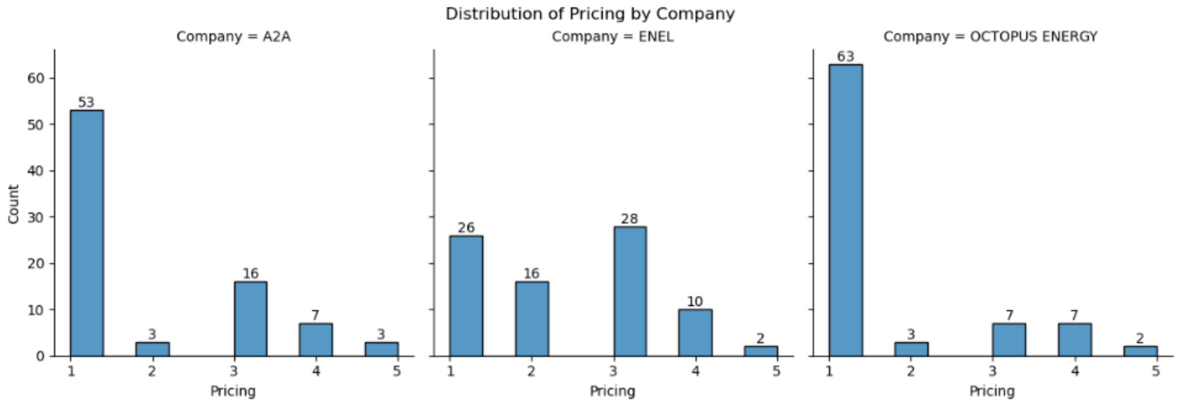
From these initial distributions, it becomes clear that addressing contract transparency and improving the onboarding process could be immediate priorities for companies receiving the majority of these lower scores. At the same time, providers with stronger Popularity ratings could leverage their existing reputation by also focusing on clarifying service details, ensuring that high brand equity does not get undermined by poor customer experiences.

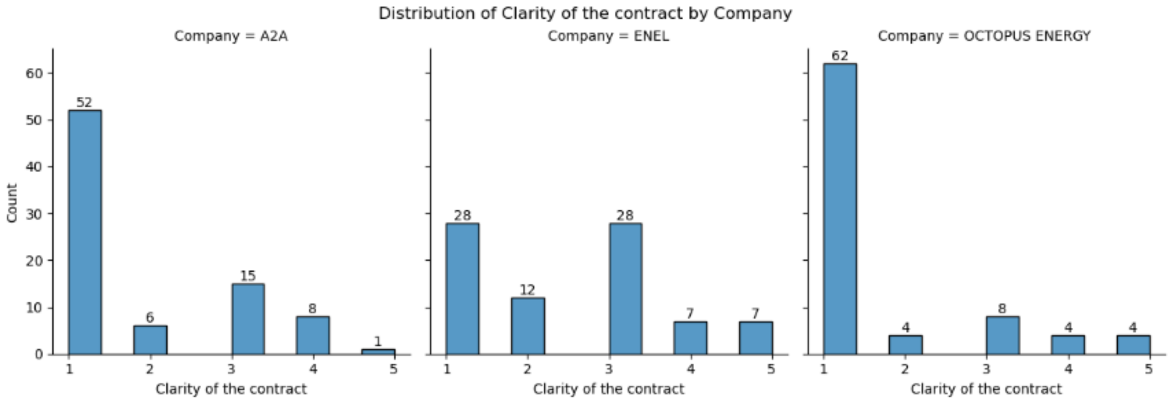
Comparing Distributions by Company

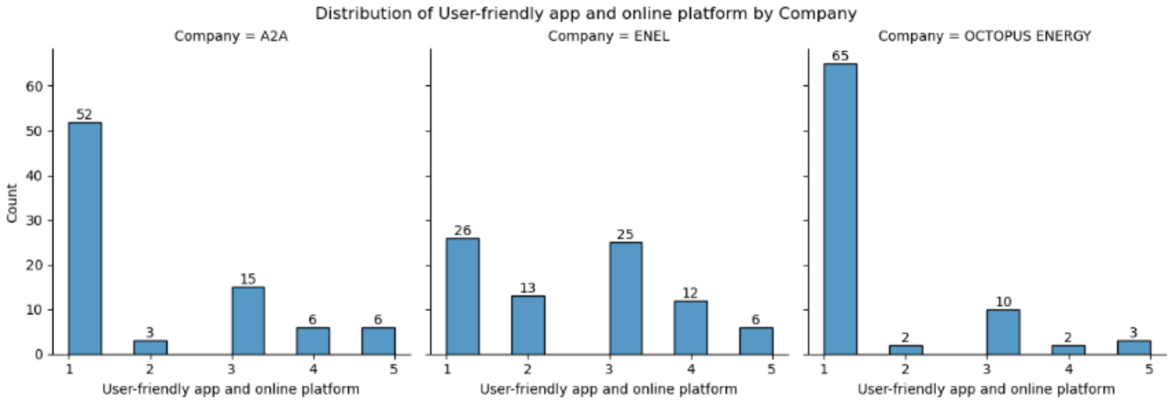
To explore variations among the three companies, individual histograms were generated for each dimension grouped by company. These side-by-side plots revealed distinct rating patterns:

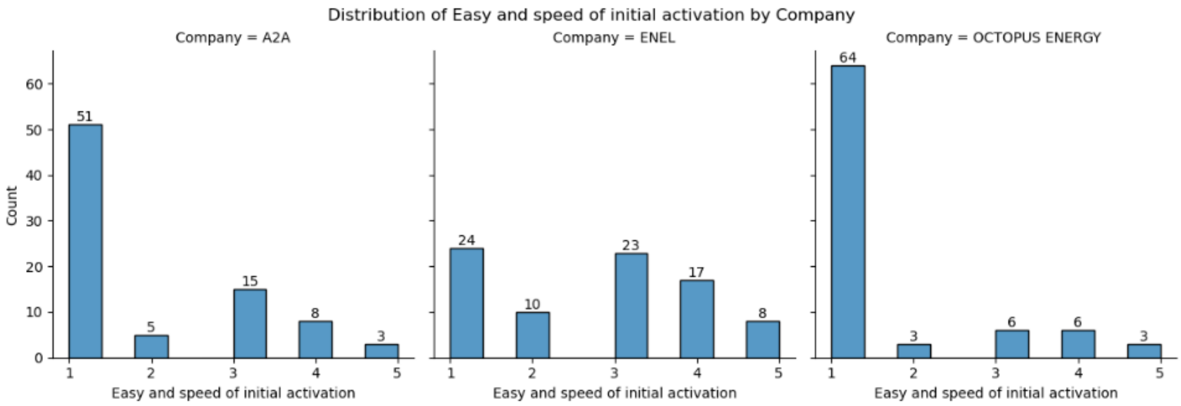












- **A2A** often shows a moderate distribution of ratings, with a notable number of low scores (1–2) in areas like Popularity and Sustainability.

- **ENEL** tends to have a healthier spread, often exhibiting higher ratings (3–5) in Popularity, Sustainability, and Ease of Activation.
- **OCTOPUS ENERGY** frequently appears at the lower end of the rating scale across multiple categories (e.g., many 1s in Popularity, Clarity of the Contract, and Ease of Activation).

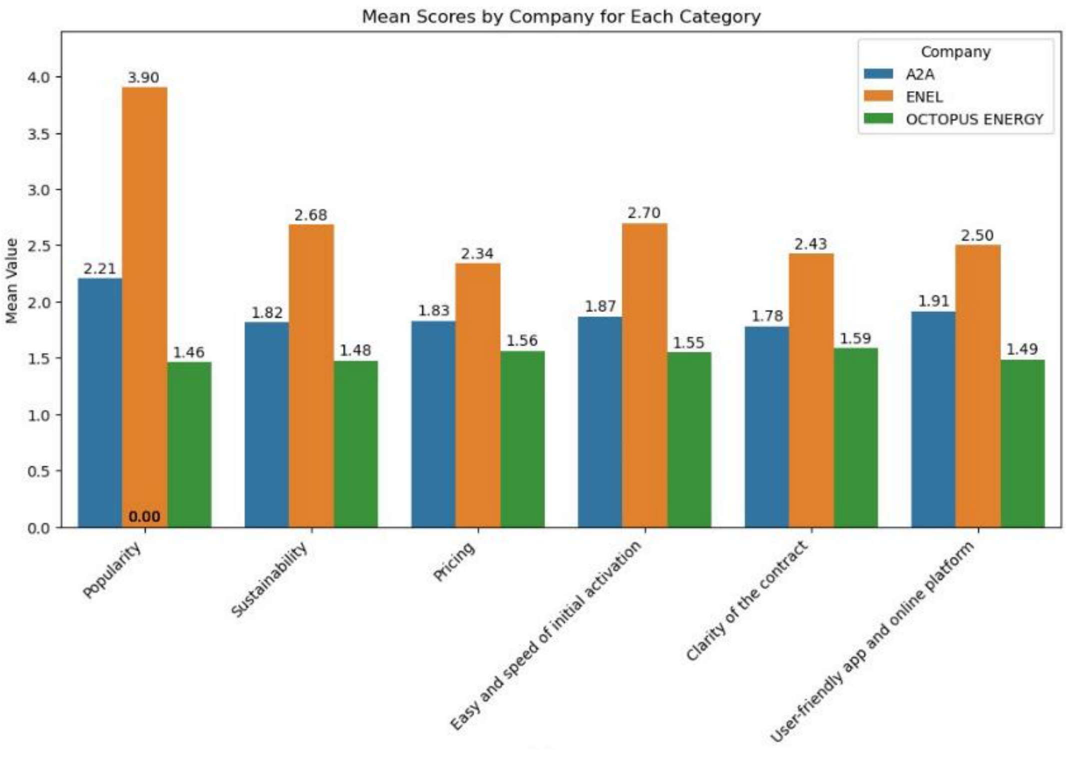
These detailed histograms underscore ENEL's relative dominance in most categories, whereas OCTOPUS ENERGY struggles with a high proportion of very low ratings. A2A occupies a middle ground, surpassing OCTOPUS in certain respects but lagging significantly behind ENEL.

Mean Ratings by Category

To gain a concise overview of each company's performance, we computed the average (mean) score for all six dimensions—Popularity, Sustainability, Pricing, Ease and Speed of Initial Activation, Clarity of the Contract, and User-Friendly App/Online Platform. The results are summarized in **Table 1** (below), and a visual representation is provided in **Figure 2** (a grouped bar chart).

Company	Popularity	Sustainability	Pricing	Ease & Speed of Activation	Clarity of the Contract	User-Friendly App
A2A	2.21	1.82	1.83	1.87	1.78	1.91
ENEL	3.90	2.68	2.34	2.70	2.43	2.50
OCTOPUS ENERGY	1.46	1.48	1.56	1.55	1.59	1.49

one sees that **ENEL** achieves the highest average ratings across all six dimensions, with particularly high Popularity (3.90) and a strong digital platform (2.50). **A2A** maintains mid-range performance, while **OCTOPUS ENERGY** receives lower average scores overall.

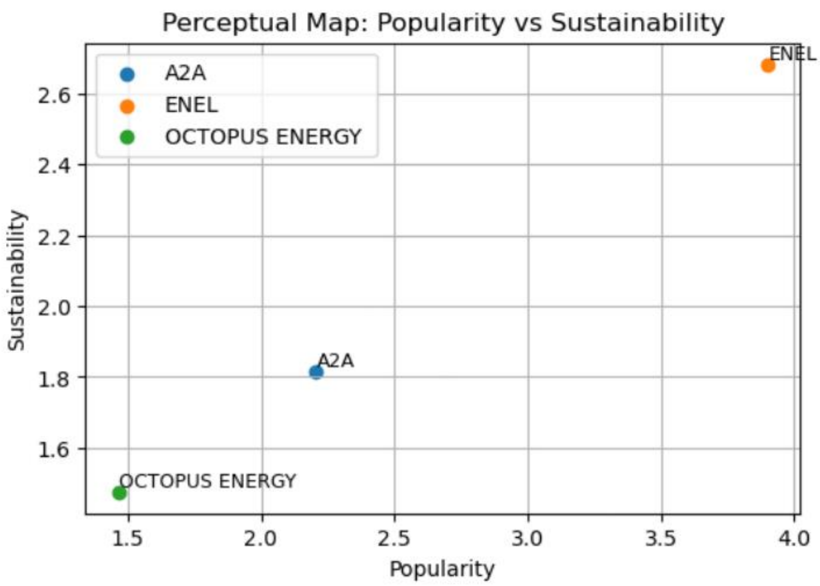


From this figure, it is apparent that **ENEL** achieves top marks in every category, distinguishing itself in areas like **Popularity** (3.90) and **User-Friendly App** (2.50). In contrast, **A2A** generally hovers around more moderate scores, while **OCTOPUS ENERGY** underperforms across the board, especially in categories such as **Popularity** (1.46) and **Sustainability** (1.48). This pattern suggests that ENEL holds a strong competitive edge in terms of both brand perception and service quality. A2A, despite not matching ENEL's peaks, outperforms OCTOPUS ENERGY in most areas, positioning itself somewhere between a leading premium brand and a lower-performing competitor.
For a clearer visual comparison of these average scores. This grouped bar chart places each category along the x-axis, with the mean score (ranging from 1 to 5) on the y-axis, while color-coding each company's bar. Such a graphical representation underscores ENEL's leadership and highlights the gap in performance faced by OCTOPUS ENERGY. It also quickly reveals the relative strengths and weaknesses of A2A, which may help stakeholders prioritize specific areas such as raising sustainability initiatives or improving contract clarity in order to challenge ENEL's market dominance.

Perceptual Mapping for Market Positioning

To delve deeper into how each company is positioned relative to its competitors, **perceptual maps** were created. These are two-dimensional plots that pair important attributes and plot the average rating of each company on the X-Y plane.

- Popularity vs. Sustainability

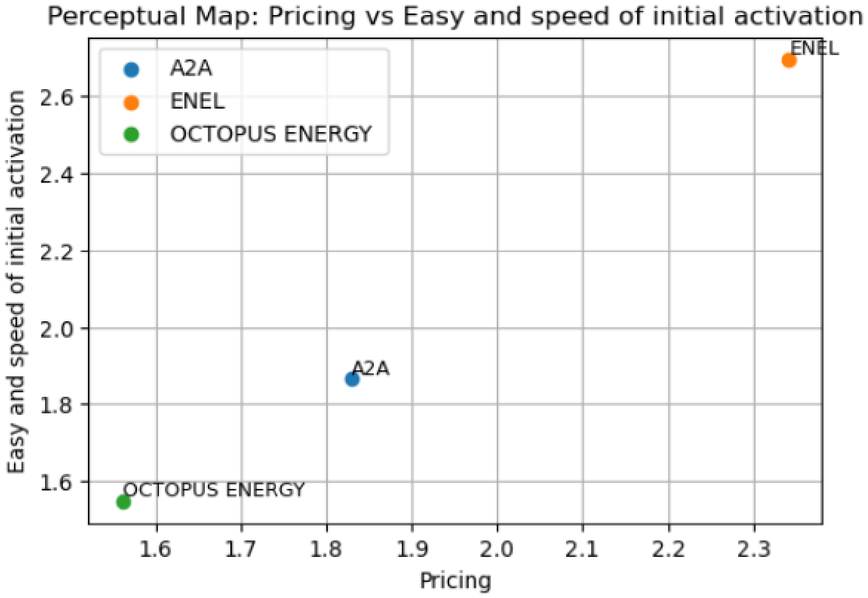


-ENEL sits at the top-right quadrant, with both high Popularity (~4.0) and higher Sustainability (~2.7).

-A2A appears more moderate, around 2.2 in Popularity and 1.8 in Sustainability.

-OCTOPUS ENERGY is at the lower-left corner, indicating poor brand awareness (1.46) and the weakest sustainability perception (1.48).

- Pricing vs. Ease of Activation

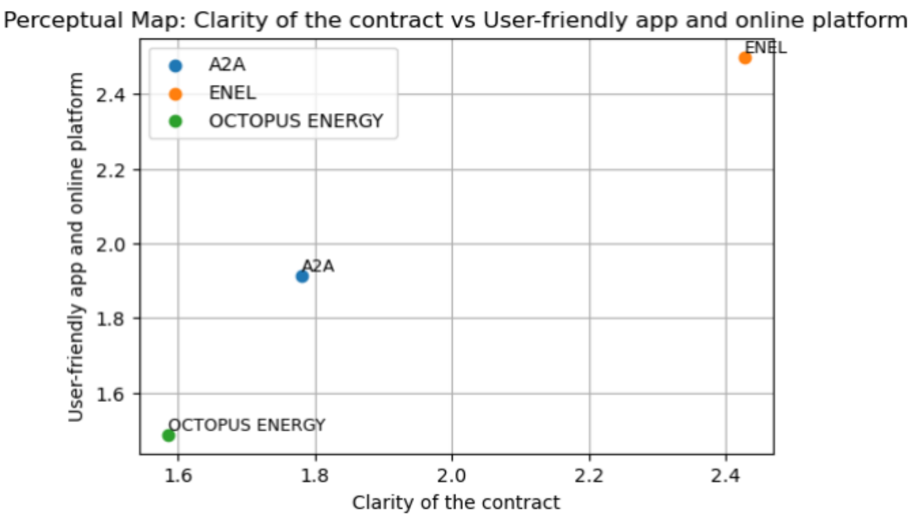


-ENEL is associated with higher pricing (~2.34) but also the easiest activation (~2.70), positioning it as a somewhat premium yet convenient option.

-A2A offers moderate pricing (~1.83) and activation (~1.87).

-OCTOPUS ENERGY is perceived as the least expensive (~1.56) but struggles with ease of signup (~1.55).

- Clarity of the Contract vs. User-Friendliness



-ENEL once again leads, with the most transparent contract terms (2.43) and a strong online platform (2.50).

-A2A lands in a mid-low zone (1.78 for contract clarity and 1.91 for user-friendliness).

-OCTOPUS ENERGY shows the lowest clarity (1.59) and digital platform rating (1.49).

These perceptual maps highlight clear competitive distinctions. **ENEL** appears well-rounded, whereas **A2A** remains middling with some room to improve both contract clarity and sustainability. **OCTOPUS ENERGY** lags in multiple dimensions, indicating a pressing need for brand-building, environmental positioning, and a better user experience.

Conclusion and Strategic Recommendations

Drawing together the insights from the distribution histograms, mean scores, and perceptual

maps, we observe three distinct competitive positions among ENEL, A2A, and OCTOPUS ENERGY.
ENEL certainly enjoys a **very strong positioning** in the market, given by various factors ranging from the historical-political context (for several years, this company had a monopoly on the sector) to its important dimensions and long years of activity.
This allows it to drive its image with a popularity well rooted in the imagination of consumers, which manifests, however, a certain laziness in the implementation of customer-based tariffs and **shows some critical points** that emerged in the reviews and in the setting analysis in relation **to the transparency of contracts**. Despite these negative sides, the company still offers a pricing that, despite being perceived as higher, remains sufficiently competitive.
A2A, meanwhile, establishes a solid presence in the middle tier, offering moderate pricing and decent ease of activation. Although the company's focus on an image built around the circular economy (as also highlighted by the LOGO), in which the company name is located in a triangle of arrows symbolizing recycling, it is positioned behind ENEL for sustainability. It therefore fails to make **its image as a company that focuses on clean source of energy strong enough.**
It is in this context that the brand image opportunities of OCTOPUS ENERGY develop, which focuses on building an image linked to sustainability.
OCTOPUS ENERGY sets itself apart by offering **the lowest pricing,** a factor that could appeal to price-sensitive customers. However, this advantage is overshadowed by very low popularity, weak sustainability perceptions, a poor user experience, and unclear contract terms.
The sentiment analysis shows extremely discordant results with those that emerged from the survey: the negative reviews represent a small number of 8 out of 1110 positive ones, highlighting how OCTOPUS ENERGY has good market retention and a mainly positive image among its consumers.
This contrast brings out two fundamental insights:
1) **popularity is a driving factor** of positioning in the market, almost absorbing the impact that other attributes have on the image.
The main recommendation is to focus primarily on **customer base growth**, possibly aspiring to occupy the middle-tier position currently held by A2A.
2) A common point between the survey results and the sentiment analysis is the **poor reference to sustainability**, a key element of OCTOPUS ENERGY's image. The site structure and logo do not allow consumers to automatically identify the company's image with this value. Therefore, a strong rebranding campaign, perhaps supported by coordinated initiatives with non-profits dedicated to the topic, would benefit OCTOPUS ENERGY's brand image.

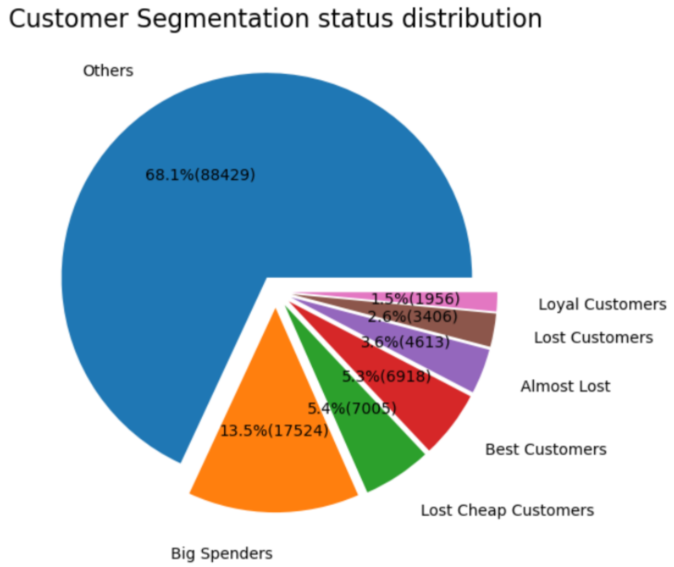
*Check file in the **Positioning** folder for the Positioning code and the company's ratings.*

Complementary

RFM analysis

The RFM (Recency, Frequency, Monetary) analysis was conducted on transactional data from January 1, 2022, to March 18, 2022, to segment customers based on their purchasing behavior. Recency indicates how recently a customer made a purchase, Frequency represents how often they buy, and Monetary measures their total spending. We assigned quartile-based scores to each metric and combined them into an RFM Score to classify customers into meaningful segments such as Best Customers, Loyal Customers, Big Spenders, Almost Lost, Lost Customers, and Lost Cheap Customers. These segments provide insights into customer engagement, retention, and spending patterns.

The analysis revealed that most customers have low purchase frequency and monetary value, with a few high-value customers contributing significantly to revenue. Best Customers (6,918) and Loyal Customers (1,956) should be nurtured with exclusive rewards, while Almost Lost (4,613) and Lost Customers (3,406) require re-engagement strategies.

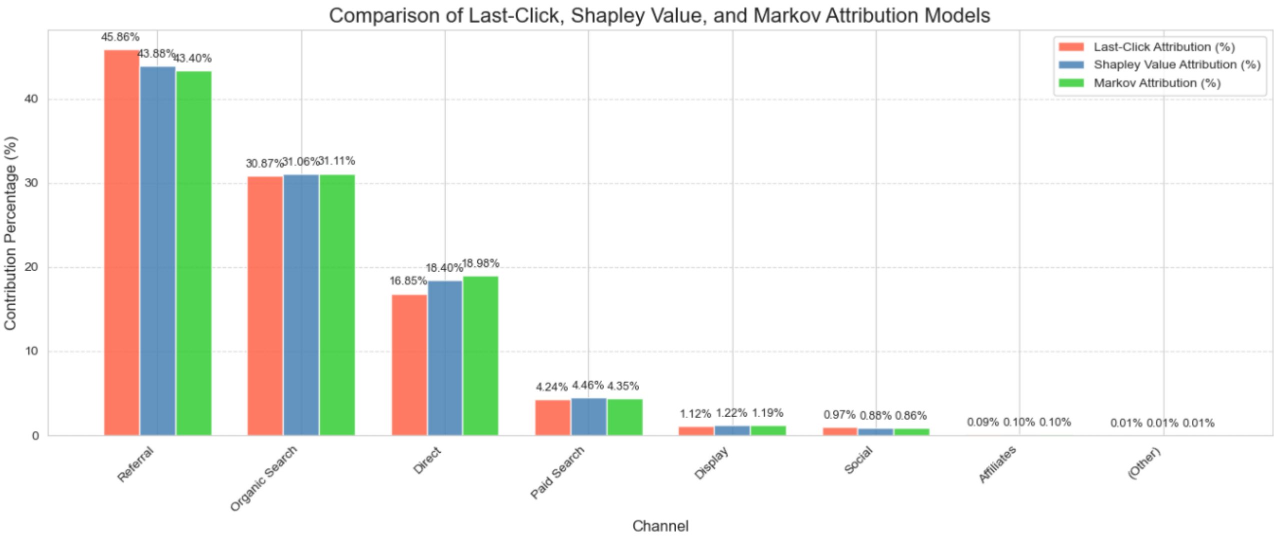


Understanding these patterns within the given timeframe helps in designing targeted marketing campaigns to enhance customer retention and maximize revenue. Future work could involve tracking changes in RFM scores over time to identify trends, predict customer behavior, and develop proactive strategies for retention and engagement.

Marketing Attribution

This analysis applies a data-driven attribution approach to assess the effectiveness of marketing channels, using a 30-day attribution window and comparing Last-Click, Shapley, and Markov attribution models. By analyzing visitor session data, timestamps, marketing sources, and conversion outcomes, the aim was to better understand how users interact with different marketing touchpoints before converting. This approach provides a more accurate way to allocate conversion credits across channels, revealing key insights into channel performance and user behavior. Findings indicate that Referral (~46%) and Organic Search (~30%) are the dominant drivers of conversions, while Direct (~17%) and Paid Search (~4%) generate substantial traffic but lower conversion rates. Display, Social, and Affiliates contribute minimally, signaling opportunities for strategic improvements. When comparing attribution models, Shapley and Markov methods distribute conversion credit more evenly across multiple touchpoints, acknowledging the role of indirect influences. Direct traffic's attribution increases under Markov (18.98%), highlighting its importance in multi-touch journeys, while Paid Search, Display, and Social receive slightly more credit in Shapley, showing their early-stage influence on conversions.

These insights suggest that optimizing Referral and Organic Search while refining Paid Search and Direct can enhance overall marketing performance. Additionally, Affiliates and Social channels may require strategy adjustments or budget reallocation to improve their impact. By leveraging a multi-attribution model approach, marketers can make more informed decisions on budget distribution, channel optimization, and multi-touch engagement strategies, leading to higher efficiency in conversion-driven marketing efforts.



Churn prediction

Churn prediction is a critical component in customer retention strategies, helping businesses identify at-risk customers before they leave. In this section Octopus customer subscription details dataset are used to predict churn. Each row represents a unique customer and includes identifiers such as anonymised_supplypoint_id, anonymised_customer_id, and signup_id. It also records customer acquisition details, including sales_channel (e.g., Direct, Price Comparison), anonymised_sales_subchannel, and whether the customer used a has_promo_code or has_referral_code during signup.

Customer demographics include supply_province, age_of_customer, and sex_of_customer. The dataset tracks service details such as first_product_category, and last_product_category, indicating product changes. The date_of_signup, supply_start_date, and supply_end_date fields help calculate customer tenure, while has_churned indicates whether the customer has left the service. After the data cleaning process, several new features were created to enhance the churn analysis. Time-based features such as supply_start_year_month and supply_end_year_month were derived to track customer subscription trends over time. The start_to_end_days feature was introduced to measure the duration of a customer's service, ensuring that ongoing subscriptions were correctly accounted for by replacing missing end dates with the current date. Similarly, signup_to_start_days captures the delay between signup and service activation, while total_service_days reflects the entire customer lifecycle from signup to churn or the present date.

To better understand customer engagement, recency-related features were created. days_since_last_activity calculates the time since the customer's last recorded activity, providing insight into their likelihood of churn. Additionally, categorical variables such as sales_channel, supply_province, sex_of_customer, and product_change were one-hot encoded

to prepare the data for machine learning. Irrelevant or redundant columns were dropped to optimize the dataset, ensuring that only meaningful features were retained for modeling.
Three models were implemented for customer churn prediction: SVM, Logistic Regression, and Logistic Regression with 5-Fold Cross-Validation. The dataset was preprocessed, scaled using StandardScaler, and split into 80% training and 20% test data with stratification to maintain class balance.

Model	**Precision**	**Recall**
SVM	%100	%93.49
Logistic regression	%100	%98.37
Logistic regression with 5 fold cross validation	%100	%98.37

All models performed exceptionally well. The high scores suggest strong predictive power, but further validation is recommended to ensure real-world applicability.

Check file in the ***Complementary*** *folder for the churn prediction code, the RFM analysis code and Marketing attribution code.*

Conclusions

In concluding our marketing analytics report on **Octopus Energy**, we have conducted a comprehensive assessment of the company's potential. By integrating advanced analytical tools with a deep understanding of consumer behavior, our objective has been to provide actionable insights that could contribute to business growth and enhance the company’s market position. The survey and sentiment analysis were essential components of our study, offering a detailed examination of the factors influencing consumer choices and decisions regarding energy providers. These analyses enabled a deeper understanding of the strengths and weaknesses of Octopus Energy’s tariff offerings. Furthermore, through conjoint analysis and segment-based conjoint analysis, we obtained valuable insights into consumer preferences across various attributes, underscoring the significance of personalized marketing approaches in expanding the customer base. Additionally, the increasing demand for digital tools and efficiency among consumers highlights the importance of promoting these aspects in the company’s offerings.
Our positioning analysis, which compared Octopus Energy with other Italian energy providers, revealed a lack of strong brand recognition in the Italian market, primarily due to limited awareness. This suggests the need for strategic investments in marketing campaigns to attract new consumers, as well as a redesign of the company’s website and logo to reinforce its image as an environmentally conscious brand.
With the energy sector undergoing a shift towards sustainability and innovation, Octopus Energy has the potential to emerge as a leading force in this transformation.