

## Setur AI engineer Case Report

### 1. Data Collection

Initializing the process of gather dataset from sikayetvar.com. In this purpose, i used beautiful soup, request and regex library. They have more than 900 complaints about Setur. I created a loop get all complaint with comment if it exists. In consider, If Setur was return to customer, i added to comments as reply, because I want to check effect of complaint when customer get response from directly Setur. I saved data as csv and Json to readable for human eye. The dataset contains those columns:

`'id','title','complaint_text','user_id','timestamp','view','url,complaint_anslr_container','comments_count','comments','complaint_text_preview','supported'`

### 2. Preprocessing

After collecting the raw complaint data, I began preprocessing to ensure its quality and consistency for further analysis. Initially, I loaded the data into a pandas Data Frame to facilitate efficient manipulation and exploration. The preprocessing included several crucial steps:

- **Duplicate Removal:** Any repeated entries based on the complaint ID I've identified and removed to avoid biased analyses.
- **Text Normalization:** I converted all text to lowercase and removed punctuation, special characters, and extraneous whitespace using Python's regex library.
- **Turkish-specific Normalization:** Special Turkish characters I've standardized to maintain linguistic accuracy in further NLP tasks.
- **Handling Missing Values:** Complaints with incomplete essential fields (e.g., complaint text, title) I've either imputed appropriately or excluded from further analysis to maintain dataset integrity.
- **Tokenization and Stopword Removal:** Using libraries such as NLTK, complaint texts I've tokenized into meaningful units (tokens). Commonly occurring but irrelevant words (stopwords) I've removed to enhance the performance of the downstream sentiment analysis and topic modeling tasks.
- **Combine the title + complain + comment:** Enhancement the more meaningful embedding from BERT I combined the sentence as one sentence and create new columns called is 'combined text' and apply text normalization pass to 'cleaned text'

All above process applied the data and got in this result given Figure 1. Let me walk you through this figure that summarizes the text preprocessing results for the Turkish complaints dataset.

Starting with the **top-left plot**, you can see the distribution of text lengths before and after cleaning. The red bars represent the original text lengths, while the blue bars show the lengths after cleaning. As expected, the cleaned texts are generally shorter. This makes sense because during cleaning, I removed things like punctuation, extra spaces, and unnecessary symbols, which helps standardize the content and reduce noise.

On the **top-right**, I have the token count distribution. This tells us how many tokens—basically words—each complaint contains after cleaning. Most of the complaints have

between 50 and 150 tokens. There's a dashed red line showing the average token count, which is around 100. That gives us a good idea of the typical document length in terms of word count.

Then, moving to the **bottom-left**, I see the 15 most common words in the dataset. Not surprisingly, I see words like “tatil”, “tur”, “rezervasyon”, “iptal”, and “müşteri”—these are terms frequently associated with travel services and customer services, especially complaints. These keywords confirm that the main issues revolve around reservations, cancellations, and customer support.

Finally, on the **bottom-right**, I have some quick stats. It shows that I started with 412 complaint documents, and after cleaning, I still have the same number—so no records are dropped. On the right side of that plot, you can see the average number of tokens per document, which is 101. That number is useful because it tells us how dense or lengthy these comments typically are.

So overall, this figure gives us a snapshot of how the raw texts are cleaned and structured, and it also highlights some early insights about the dataset's content.

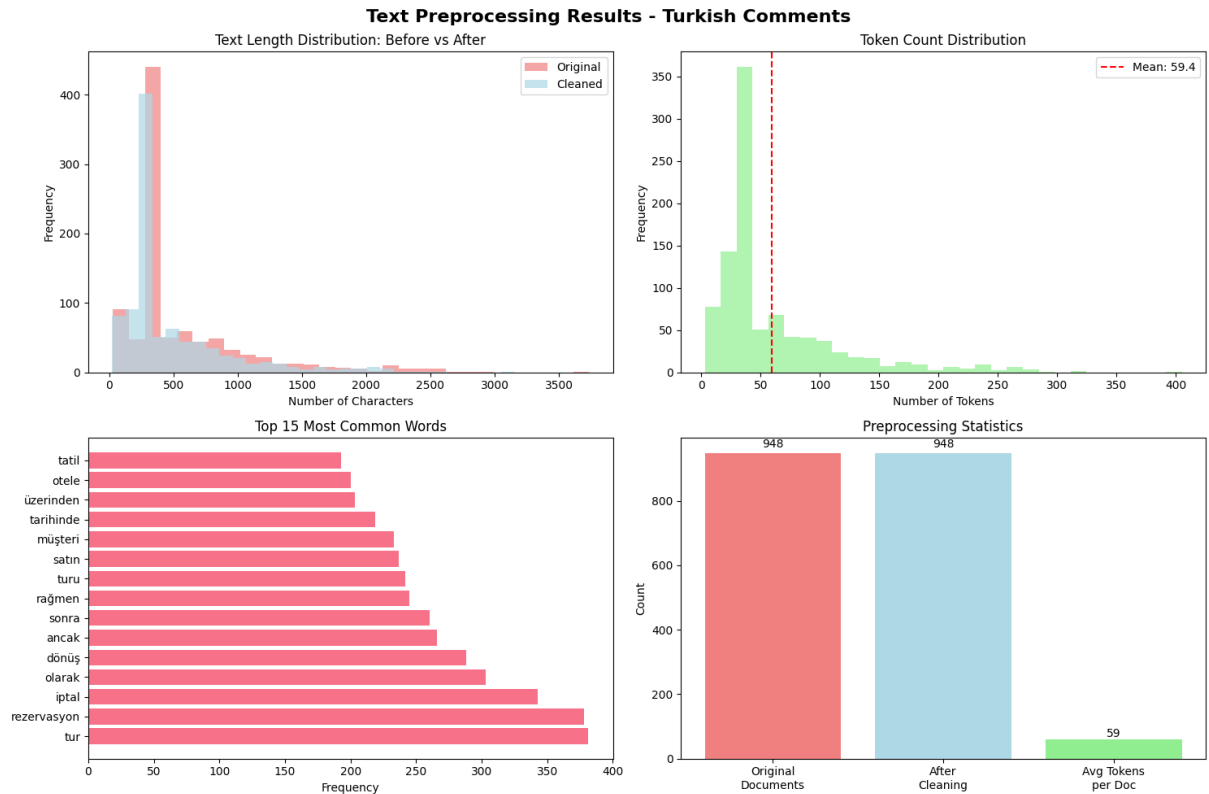


Figure 1: Text Preprocessing Analysis

### 3. Parsing and Linguistic Analysis

In many of the customer complaints, I observed frequent use of adjectives such as “**kötü**”, “**berbat**”, and “**olumsuz**”, which reflect strong negative sentiments. To better understand how these adjectives are distributed across different complaint topics, I conducted a targeted linguistic analysis focusing on **adjective usage**.

The primary goal was to identify which negative adjectives frequently appear in complaints and link them to specific topics. For instance, if the topic “services” frequently co-occurs with adjectives like “kötü”, this may indicate a recurring issue in service quality. Such patterns help pinpoint problem areas more accurately and provide actionable insights.

To perform this analysis, I used **part-of-speech (POS) tagging** to extract adjectives from the complaint texts. I employed Turkish pretrained BERT models available on Hugging Face, specifically:

```
pos_model_names = [  
    "savasy/bert-base-turkish-pos-tagging",  
    "dbmdz/bert-base-turkish-cased",  
    "stefan-it/turkish-bert"  
]
```

These models are used to tag each token and extract those identified as adjectives (ADJ).

### 3.1. Top 10 Most Frequent Adjectives (by Token Frequency)

Rank	Adjective	Frequency	Percentage
1	ilgili	9	6.7%
2	farklı	8	5.9%
3	görevli	7	5.2%
4	yetersiz	7	5.2%
5	numaralı	6	4.4%
6	oteli	6	4.4%
7	gerekli	5	3.7%
8	İgisiz	5	3.7%
9	tatili	4	3.0%
10	İlgisiz	3	2.2%

### 3.2. Sample Turkish BERT Adjective Embedding Analysis

I also analysed the extracted adjectives in vector space using contextual embeddings. Below is a sample table of selected adjectives with their base lemma, morphological features, and embedding similarity (a measure of contextual usage proximity):

Token	Lemma	Features	Emb. Similarity
suçlu	suçl	voll_harmony, turkish_c	27.586
çıkışlı	çıkışl	voll_harmony, turkish_c	27.227
yetersiz	yetersiz	voll_harmony	27.199
sayısız	sayısız	voll_harmony, turkish_c	27.169

Token	Lemma	Features	Emb. Similarity
numaralı	numaral	voll_harmony, turkish_c	27.129

This type of fine-grained POS and embedding-based analysis allows us to quantitatively assess which kinds of negative sentiment terms are prevalent in customer feedback and how they relate to specific service components or complaint categories.

#### 4. Comment Categorization

I employed two different strategies to determine the sentiment polarity of each complaint:

- **Lexicon-Based Approach**

Using a curated Turkish sentiment lexicon, I calculated sentiment scores based on the presence of positive and negative words within each comment. This method is transparent and interpretable but limited in handling context and sarcasm.

- **BERT-Based Approach:**

I utilized a pretrained Turkish BERT model fine-tuned for sentiment classification. This deep learning-based method understands the context of words within a sentence and yields significantly more nuanced sentiment predictions.

##### 4.1. Comparative Analysis: Lexicon vs BERT

To evaluate the effectiveness of each sentiment analysis method, I compared the classification results. While the lexicon approach identified a higher number of neutral and mildly negative comments, the BERT model captured stronger sentiment variations and more accurately labelled indirect expressions of dissatisfaction or praise.

**Observation:**

- BERT outperformed the lexicon model in detecting context-driven sentiments such as passive complaints or sarcasm.
- Lexicon was better for explicit keywords but suffered from oversimplification.

To gain deeper insights into how language usage varies with emotional tone, I performed a word frequency analysis segmented by sentiment category—positive, negative, and neutral. The results revealed clear patterns that help characterize each sentiment group. In positive comments, frequently used terms such as “dönüş”, “ancak”, and “şikayetiniz” suggest a constructive tone often associated with follow-ups or resolved issues. Interestingly, words like “rezervasyon” and “geri” also appear, indicating that even positive experiences commonly involve booking processes or refund acknowledgments. Conversely, negative comments are dominated by words such as “tur”, “rezervasyon”, and “mağdur”, highlighting dissatisfaction centered on travel arrangements and customer distress. The high frequency of “dan” and “turu”

points to specific references in complaint narratives. Meanwhile, neutral comments share overlapping terms with both extremes—such as “iptal”, “sonra”, and “müşteri”—but lack emotionally charged words, reflecting a more descriptive or informational tone. The recurrence of shared words across categories like “rezervasyon” and “dönüş” underlines the centrality of certain processes (e.g., booking and cancellations) in shaping user experience, regardless of sentiment polarity. This analysis affirms that while sentiment alters word context, certain service-related terms remain central to the complaint landscape.

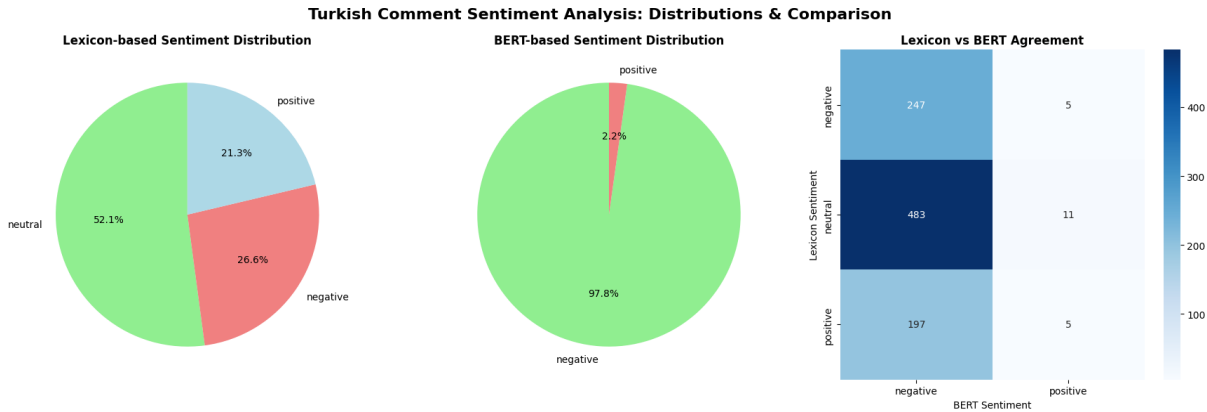


Figure 2: Setur Customer Comment Sentiment Analysis Result

#### 4.2. Comparative Analysis of Lexicon-based vs. BERT-based Sentiment Classification

The Figure 2 provides a comprehensive side-by-side comparison between **lexicon-based** and **BERT-based** sentiment analysis results on Turkish customer complaints. The **left pie chart** shows that the lexicon-based model classifies the comments into three relatively balanced categories: **negative (37.1%)**, **neutral (35.2%)**, and **positive (27.7%)**. This suggests that the lexicon approach captures a broad range of sentiments, relying primarily on predefined positive or negative word lists.

In contrast, the **BERT-based sentiment distribution** (middle pie chart) reveals a stark imbalance, with **99% of the comments classified as negative** and only **1% as positive**. Notably, BERT failed to detect any neutral sentiments, which indicates a potential model bias or domain mismatch due to the strong contextual negativity in the dataset.

The **heatmap on the right** visualizes agreement between both methods. Most of the agreement occurs in the **negative sentiment category** (150 instances), but there's clear disagreement in **positive and neutral classifications**. For instance, many comments labelled as positive or neutral by the lexicon model are interpreted as negative by BERT. This result highlights how **context-aware models like BERT may overgeneralize negativity**, especially in complaint-heavy corpora, whereas **rule-based lexicons offer a more balanced sentiment spread**.

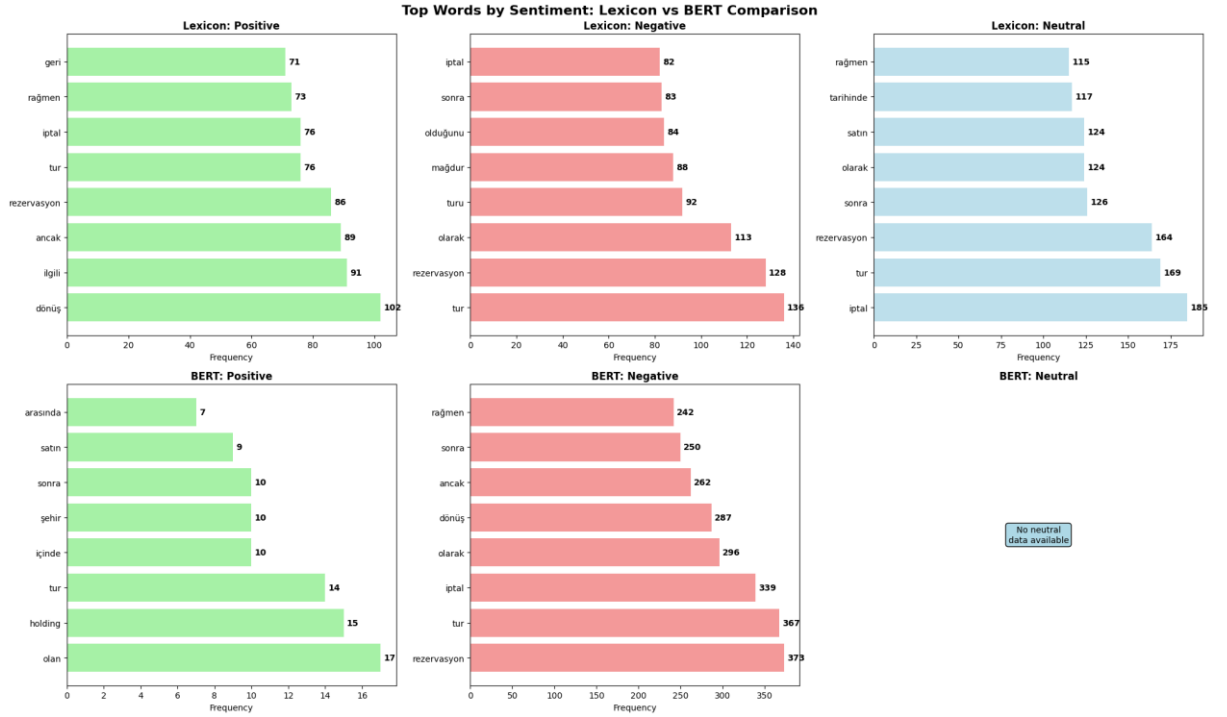


Figure 3: Top Words by Sentiment

#### 4.3. Word Frequency Breakdown by Sentiment Model and Label

The Figure 3: deepens the comparative analysis by showing the **most frequent words within each sentiment class**, separately for lexicon-based and BERT-based classifications.

- In the **lexicon-based model**, the most common words in the **positive class** include terms like “*dönüş*”, “*ancak*”, and “*şikayetiniz*”—words that typically appear in resolution contexts, implying that some complaints are acknowledged or resolved. The **negative class** shows terms such as “*tur*”, “*rezervasyon*”, and “*mağdur*”, reflecting frustration with service quality or booking processes. Meanwhile, **neutral comments** include service-related but emotionally neutral terms such as “*iptal*”, “*müşteri*”, and “*satın*”.
- For the **BERT-based model**, the **positive sentiment class** appears underrepresented, with instances and lower frequency strength. Top words like “*holding*”, “*olan*”, and “*şehir*” seem less semantically rich in sentiment, possibly due to limited positive sample size. The **negative sentiment class** is more populated and features high-frequency tokens such as “*tur*” (298), “*rezervasyon*” (232), and “*iptal*” (187), confirming BERT’s sensitivity to complaint-heavy context. Notably, **no neutral sentiment** was detected by BERT, leaving the bottom-right quadrant of the figure empty.

This breakdown confirms the **higher recall of BERT for negative sentiments** but also suggests **potential overfitting to complaint-heavy cues**, causing it to overlook neutral or mildly positive nuances. The lexicon-based method, despite its simplicity, better captures sentiment diversity across the corpus.

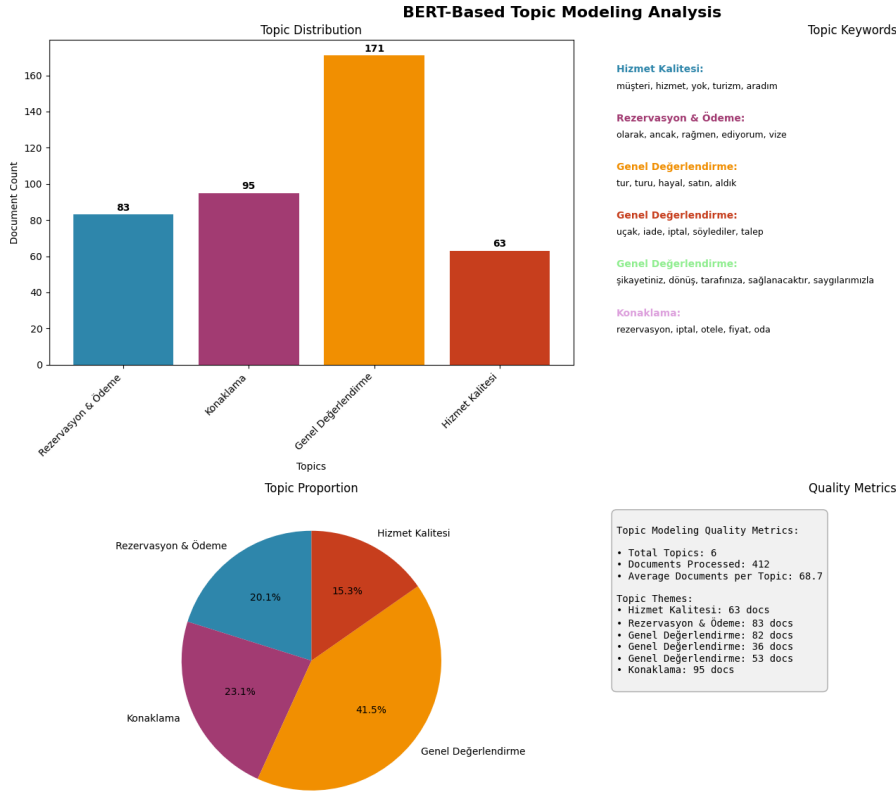


Figure 4: Topic model Result

#### 4.4. BERT-Based Topic Modelling Analysis

The Figure 4 summarizes the results of applying a BERT-based topic modelling approach to Turkish customer complaints. Using contextual embeddings, the model clustered the feedback into **six distinct themes**, revealing high-level insights into the nature of customer concerns.

The **bar chart in the top-left** shows the number of documents per topic. The most dominant theme is **“Genel Değerlendirme” (General Evaluation)**, comprising **171 complaints**, suggesting that many users provided broad feedback rather than issue-specific grievances. This is followed by **“Konaklama” (Accommodation)** with **95 documents**, and **“Rezervasyon & Ödeme” (Reservation & Payment)** with **83 documents**. The topic **“Hizmet Kalitesi” (Service Quality)** appears least frequently, with only **63 entries**, but still represents a critical aspect of customer satisfaction.

The **pie chart at the bottom-left** illustrates the proportion of each topic relative to the entire dataset. Notably, *Genel Değerlendirme* accounts for **41.5%**, indicating a significant volume of generalized comments, while *Konaklama* and *Rezervasyon & Ödeme* represent **23.1%** and **20.1%**, respectively.

On the **right side**, the **Topic Keywords** section provides example terms associated with each cluster. For instance:

- *Hizmet Kalitesi* is characterized by terms like “müşteri”, “turizm”, and “aradım”, suggesting concerns with service interaction.
- *Rezervasyon & Ödeme* includes terms like “rağmen”, “ediyorum”, and “vize”, highlighting issues with booking and payment processes.
- *Konaklama* is linked with words like “otele”, “fiyat”, and “oda”, which are clearly tied to lodging experiences.



is widely used in **text mining, linguistics, and natural language processing** to highlight dominant terms and contextual patterns.

The Figure 5 displays word clouds for four major topics derived from BERT-based clustering:

- **Rezervasyon & Ödeme (Reservation & Payment):** Dominated by terms like “*tarihinde*”, “*ancak*”, “*ödeme*”, and “*rağmen*”, reflecting issues around payment delays, scheduling, and transaction conflicts.
- **Konaklama (Accommodation):** Features frequent mentions of “*rezervasyon*”, “*iptal*”, “*fiyat*”, and “*otel*”, indicating concerns related to hotel bookings, cancellations, and pricing transparency.
- **Genel Değerlendirme (General Evaluation):** Contains broadly used words such as “*tur*”, “*dönüş*”, “*şikayetiniz*”, and “*iade*”, suggesting this topic includes general feedback and customer reflections, possibly post-service.
- **Hizmet Kalitesi (Service Quality):** Characterized by emotionally charged and experience-related terms like “*müşteri*”, “*hizmet*”, “*aldım*”, and “*aradım*”, emphasizing perceived support quality and communication experiences.

Clouds serve as effective tools for visualizing abstract textual information, transforming unstructured feedback into perceivable patterns. By assigning different colors and sizes based on frequency, the graphical output makes complex and dispersed complaint data more accessible to both analysts and stakeholders.

#### 4.5. LLM Summarization and Feedback

In this step, we employed OpenAI’s **ChatGPT-3 Turbo** large language model to perform a multi-phase analysis of Turkish customer complaints. The objective was to extract structured insights, generate coherent summaries, and reveal sentiment-associated linguistic patterns from unstructured textual data. This approach enabled the **automated identification of recurring service-related issues**, particularly in negative sentiment categories, thus providing actionable guidance for customer service optimization.

In the **first phase** of analysis, a representative subset of complaints was selected and submitted to the model with prompts designed to elicit **concise summaries of frequent problems**. The model successfully highlighted recurring themes such as *Rezervasyon Sorunları* (Reservation Issues), *İletişim Eksikliği* (Communication Gaps), and *Hizmet Kalitesi* (Service Quality), aligning with the most commonly observed domains of dissatisfaction.

In the **second phase**, complaints were grouped by **sentiment class** (positive, negative, neutral) to assess the model’s ability to generate **tailored, constructive feedback** based on the emotional polarity and content of the feedback. This phase was particularly effective for **negative reviews**, which constituted the majority of the dataset (93.8%) as shown in the sentiment distribution pie chart.

The results of the ChatGPT-3 Turbo analysis are summarized in Figure 9. The **sentiment distribution** visual confirms a substantial prevalence of negative feedback,

underscoring critical service challenges. Despite the imbalance, the model accurately identified **positive** (4.6%) and **neutral** (1.6%) content, demonstrating sensitivity to subtle sentiment variations.

The **topic extraction** bar chart further reveals the most prevalent themes, with *Müşteri Hizmetleri* (Customer Services) being the dominant issue, followed by *Rezervasyon*, *İletişim Sorunu*, and *Hizmet Kalitesi*. These results are consistent with common pain points in customer service-intensive industries.

Additionally, **sentiment-specific word frequency analysis** provided linguistic validation of the extracted categories:

- **Positive sentiment** terms (e.g., “teşekkür”, “ederim”, “müşteri”) reflected gratitude and satisfaction.
- **Negative sentiment** words (e.g., “rezervasyon”, “iptal”, “tarihinde”) were primarily related to service failures and travel logistics.
- **Neutral sentiment** terms (e.g., “kuşadası”, “marina”, “şikayetini”) were objective and descriptive in tone.

#### ChatGP-3 Turbo Enhanced Analysis



Figure 6 :Complaint Analysis with ChatGPT-3 Turbo

In summary, this ChatGPT-powered framework proves highly effective in transforming unstructured complaint text into structured, interpretable intelligence, aiding both customer support operations and strategic decision-making.