Recipe Reviews and User Feedback Analysis

1. Brief Overview of Business/Organization Challenges

Our company, as a recipe sharing platform, experiences fluctuations in user interaction and satisfaction. We've noticed differences in user behaviors, such as varying levels of interaction with recipes and different emotions expressed in comments. Understanding these trends is crucial for maintaining user satisfaction and optimizing platform performance. Our company possesses a dataset containing recipe data and user feedback, including ratings, comments, likes, and user reputation. By analyzing this data, we aim to understand user preferences and behaviors and improve the user experience of the platform.

2. Problem Summary/Define

The primary challenge is to understand past trends and dynamics regarding user interaction and sentiment on our recipe sharing platform. We need to identify patterns in user behaviors such as recipe popularity, effectiveness of user comments, and overall satisfaction with recipes. By understanding these factors, we can make informed decisions to enhance user experience and optimize platform performance.

The main challenges faced by the company are as follows:

- Identifying which recipes users like the most and why.
- Examining the overall distribution of ratings given by users and how it changes over time.
- Investigating the relationship between user interaction with comments and user reputation.
- Determining who the active users of the platform are and how frequently they interact.
- Analyzing comments on recipes and performing sentiment analysis based on interaction statuses to identify high-quality recipes and make improvements through advertising campaigns.

3. Method (Exploratory Data Analysis - EDA)

- **Descriptive Statistics of the Dataset:**

In this stage, the general characteristics of the dataset are examined, and basic statistical information is obtained for each feature. In particular, statistics such as mean, standard deviation, minimum, and maximum values are calculated for numerical data. Additionally, the presence of missing or null values for each feature is checked, and these values are cleaned if necessary.

- **Correlation Analysis:**

A correlation matrix is used to understand the relationships between features. The correlation matrix contains Pearson correlation coefficients between each pair of features. These coefficients indicate the strength and direction of the linear relationship between features. Additionally, pair plot graphs can be used to visualize relationships between features.

- **Popular Recipes Analysis:**

The popularity of recipes is evaluated based on a combination of factors. These factors include the number of comments, average rating, user interaction criteria (likes/dislikes, reply count), etc. Popular recipes are determined and ranked based on these factors.

- **User Interaction Analysis:**

User interaction metrics are examined to understand users' interaction with recipes and comments. These metrics include comment reply count, likes/dislikes, user reputation, etc. This analysis is important for understanding the level of interaction and behaviors of users on the platform.

4.Analysis

Data Preparation and Preprocessing:

The dataset was included in the project and converted into a dataframe. Missing values were cleaned, and unnecessary columns were removed. Key features were identified, and the dataset was filtered according to these features. Null or missing values were detected and cleaned. Variables in the dataset were made suitable for the DataFrame.

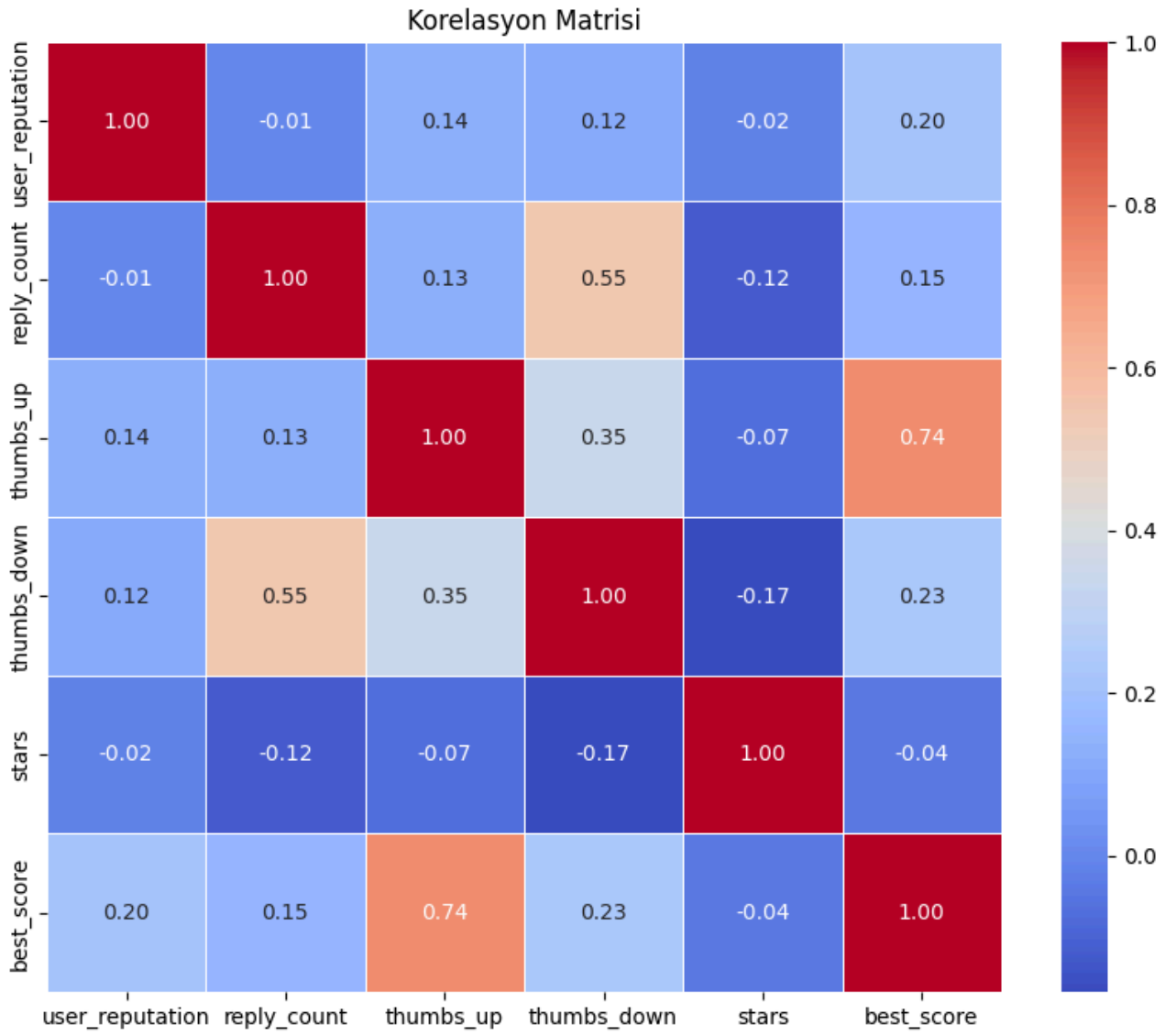
Descriptive Statistics of Features along with Information about the Dataset:

	index	cipe_number	ecipe_code	r_reputation	reated_at	eply_count	humbs_up	humbs_down	stars
count	.000000	060.000000	060.000000	059.000000	059000e+03	3059.000000	059.000000	3059.000000	59.000000
mean	80392	3.349020	2735.070915	1.832298	23740e+09	0.014384	0.940830	0.442955	.529585
std	54446	1.922429	384.127476	4.828822	72997e+06	0.159059	3.973107	2.781608	.270235
min	000	1.000000	832.000000	0.000000	13035e+09	0.000000	0.000000	0.000000	.000000
25%	00000	2.000000	309.000000	1.000000	22717e+09	0.000000	0.000000	0.000000	.000000
50%	00000	3.000000	4299.000000	1.000000	22718e+09	0.000000	0.000000	0.000000	.000000
75%	00000	5.000000	7826.000000	1.000000	22718e+09	0.000000	0.000000	0.000000	.000000
max	00000	000	6.000000	00000	627e+09	000	0000	0000	000

Results and Recommendations:

There are a total of 18182 entries and 15 columns. Important columns include recipe name, username, stars, and comments. Each entry has 15 features. Features include recipe number, recipe name, comment ID, user ID, username, user reputation, creation date, reply count, likes and dislikes counts, star rating, and best score.

Correlation Analysis between Features:



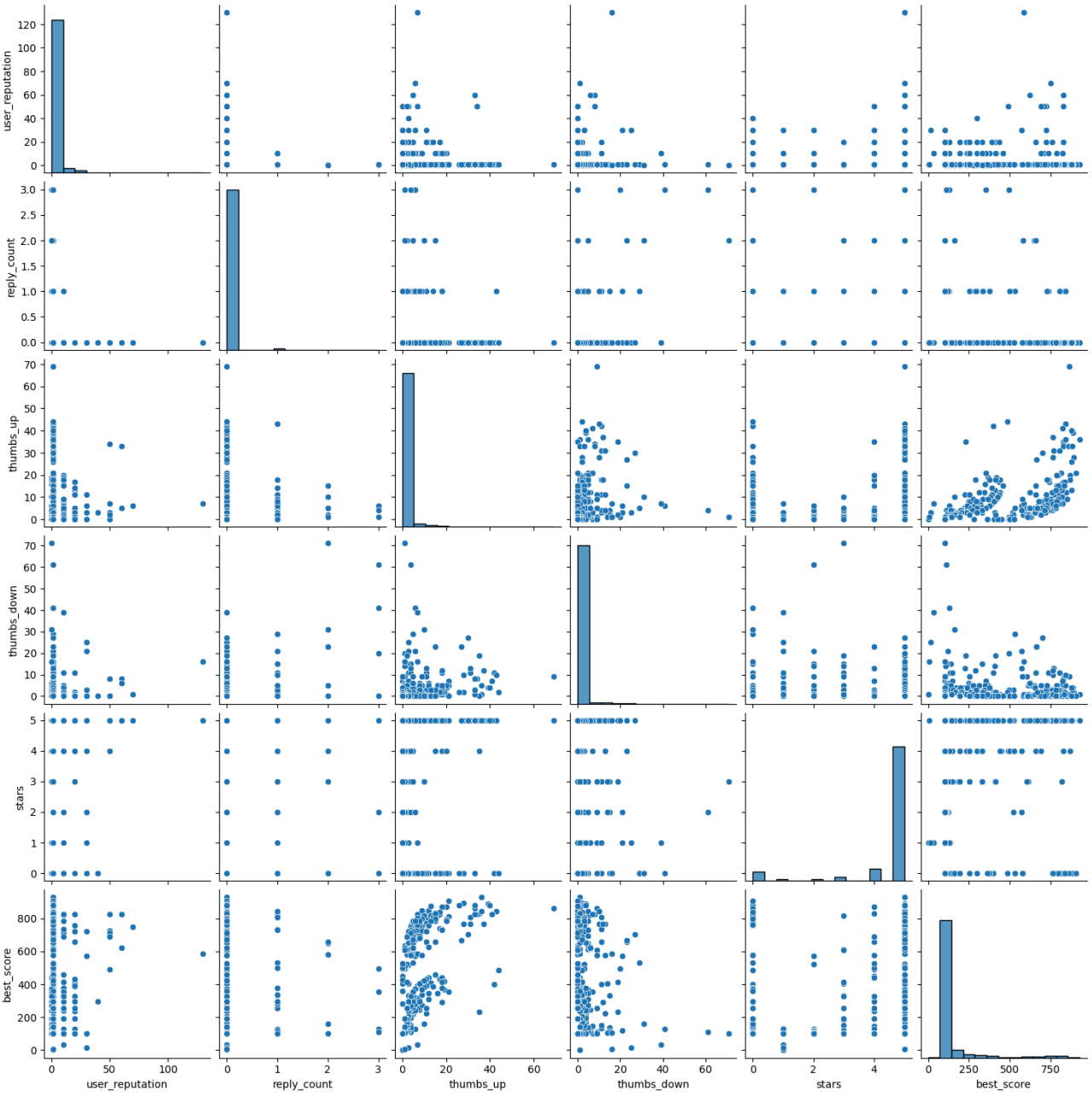
- **Relationship between user_reputation and reply_count (0.010609):** This correlation coefficient is close to zero, indicating a very weak relationship between user_reputation and reply_count. This suggests that user reputation has very little effect on the number of responses to recipes.
- **Relationship between thumbs_up and reply_count (0.208786):** This positive correlation coefficient indicates a moderate relationship between thumbs_up (likes) and reply_count (number of

responses). Thus, there is a certain connection between the likes a recipe receives and the number of responses it gets.

- **Relationship between thumbs_up and user_reputation (0.057860):** This correlation coefficient indicates a slight relationship between user_reputation and likes. Thus, users with higher reputations tend to give more likes to recipes.
- **Relationship between stars and thumbs_down (-0.142549):** This negative correlation coefficient indicates a moderate inverse relationship between star ratings and thumbs_down (dislikes). Thus, as the number of dislikes a recipe receives increases, its star rating generally decreases.
- **Relationship between best_score and thumbs_up (0.684834):** This positive correlation coefficient indicates a strong relationship between the best score of a recipe and the number of likes it receives. Thus, if a recipe has a high best score, it generally receives a high number of likes.
- **Relationship between best_score and reply_count (0.201303):** This positive correlation coefficient indicates a moderate relationship between the best score of a recipe and the number of responses it receives. Thus, as the number of responses to a recipe increases, its best score generally increases.

Visualizations:

Using the Seaborn library, relationships between numerical variables in the dataset were visualized with pairplot graphs. These visualizations were used to better understand relationships between features.



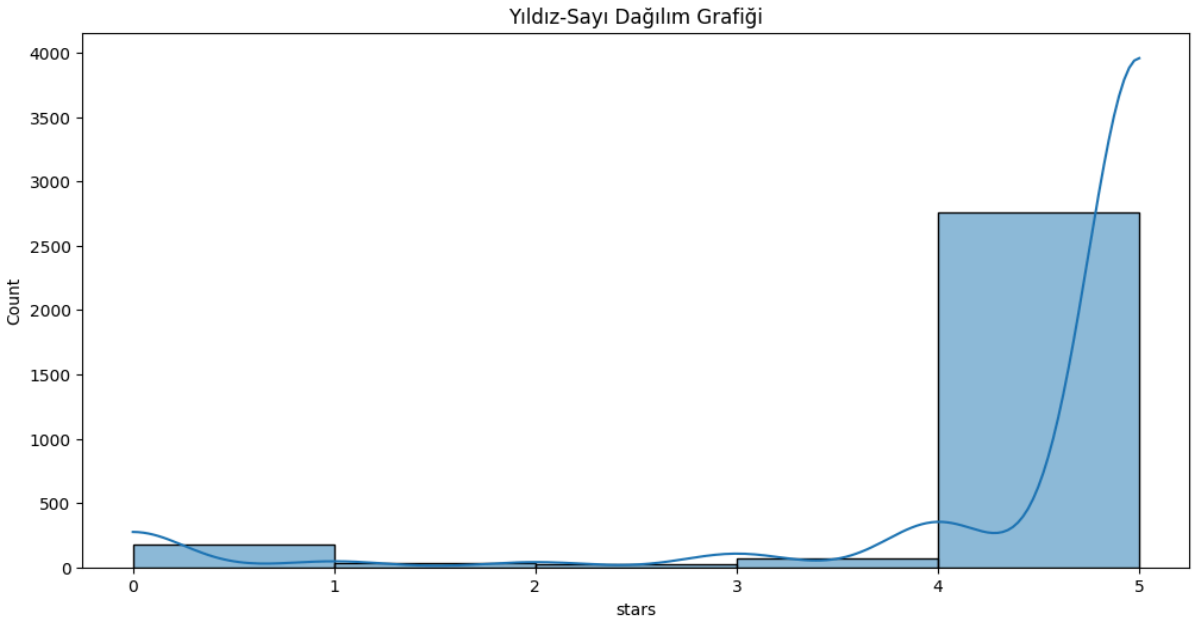
- **humbs_up and best_score:** There appears to be a positive relationship between likes and best score. Thus, recipes that receive more likes generally have higher best scores. This indicates that users generally like and give high scores to recipes they find satisfying.

- **reply_count and user_reputation:** There seems to be a relationship between reply count and user reputation. Users with higher reputations tend to receive more responses. This suggests that active and contributing users on the platform receive more interaction within the community.

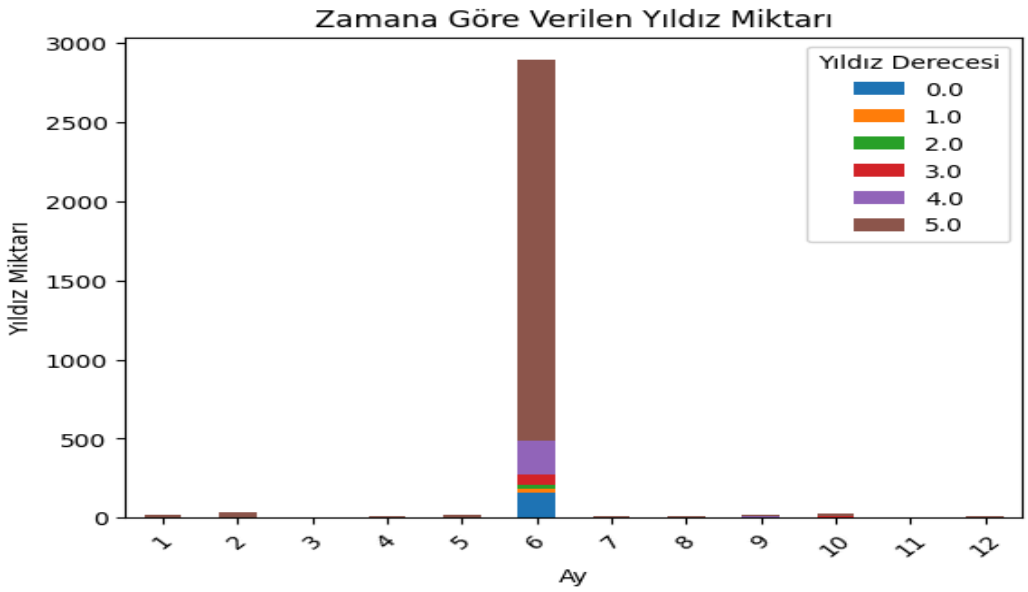
- **stars and best_score:** A relationship between star ratings and best score is observed. This indicates that recipes with high star ratings generally have higher best scores. This relationship between these two features, which reflect overall user satisfaction, is naturally expected.

Analysis Results:

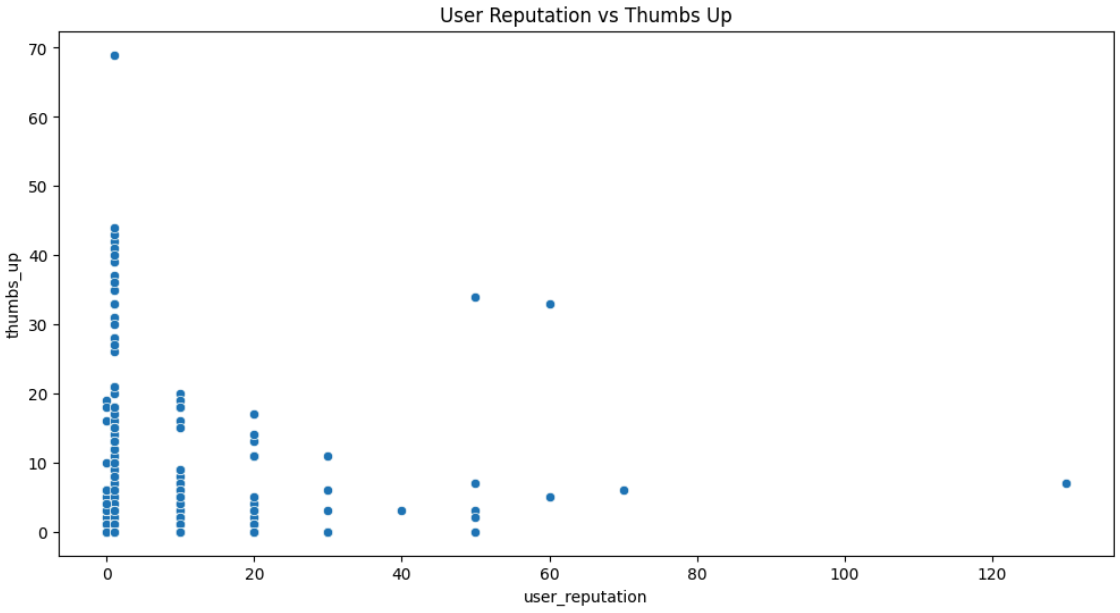
It was observed that the ratings given to recipes are generally high.



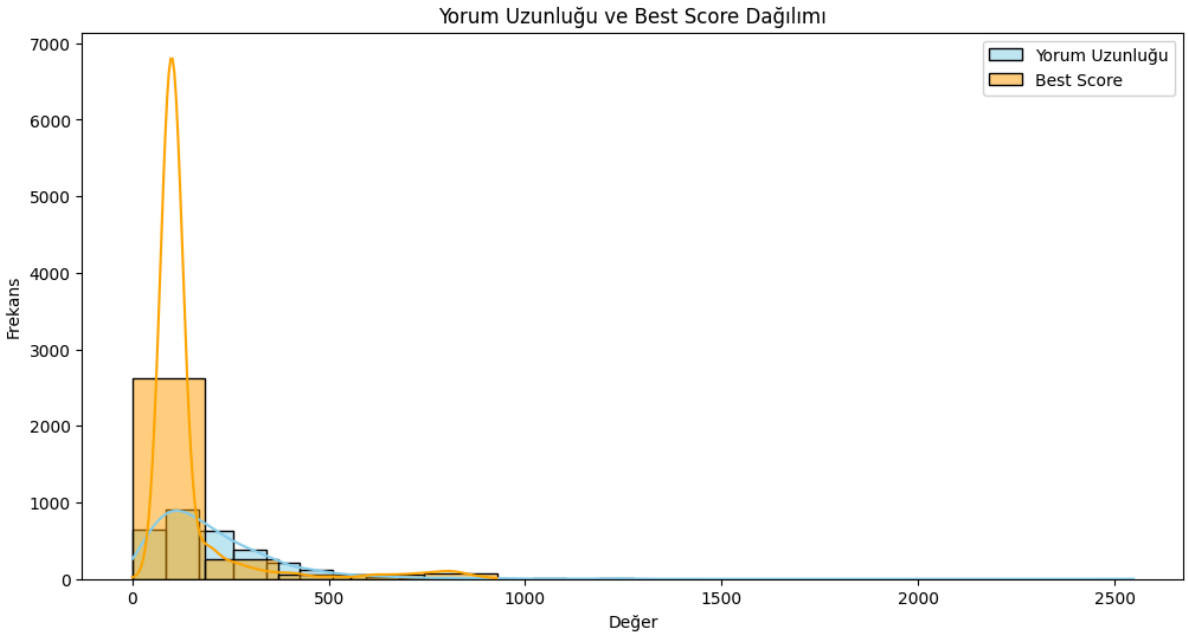
The 6th month of the year 2022 was determined to be the period with the highest interaction and ratings given to recipes.



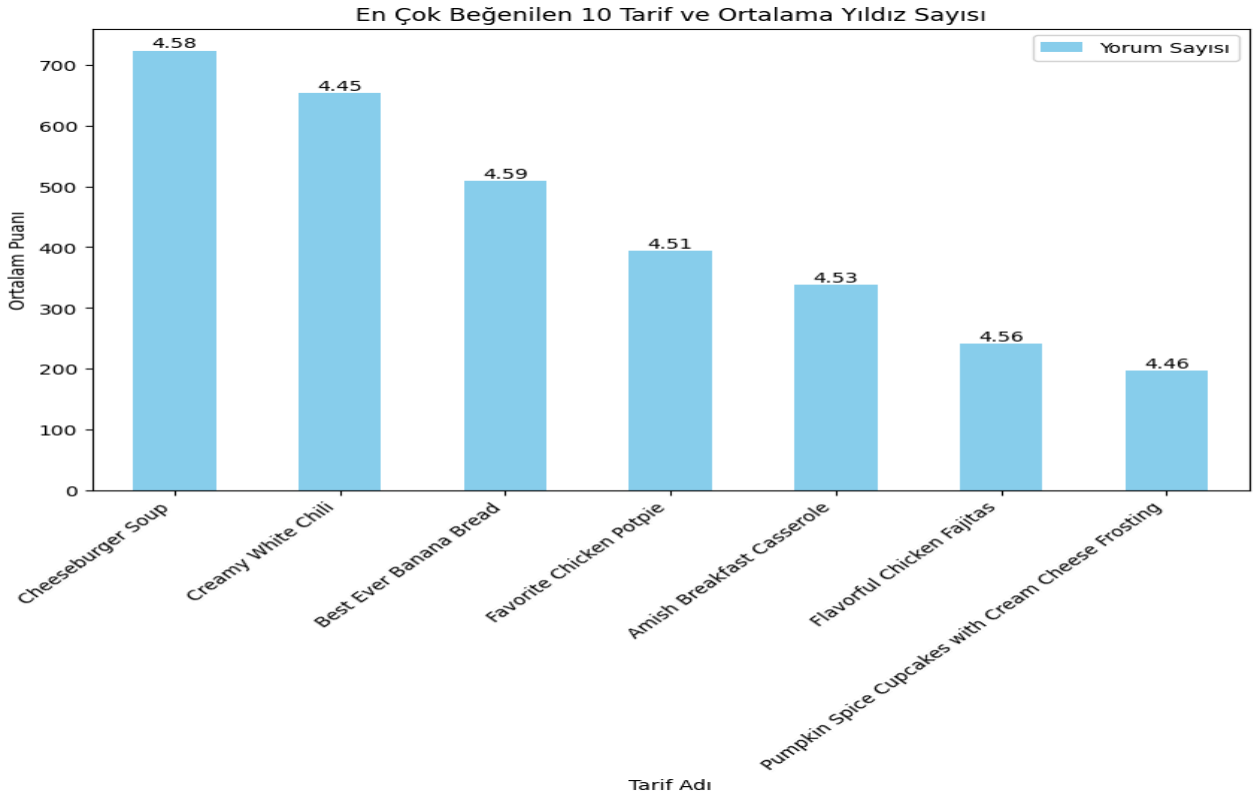
It was found that comments made by users with high user reputation receive more interaction.



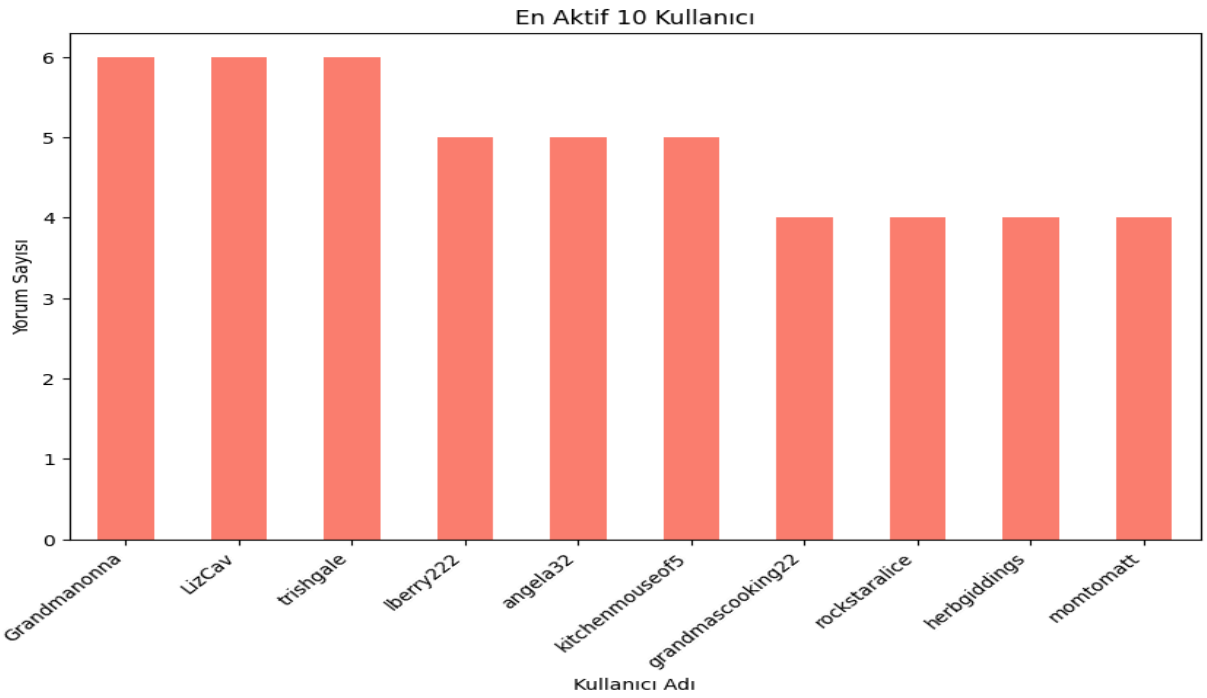
There is a strong relationship between the word count in comments and the best score.



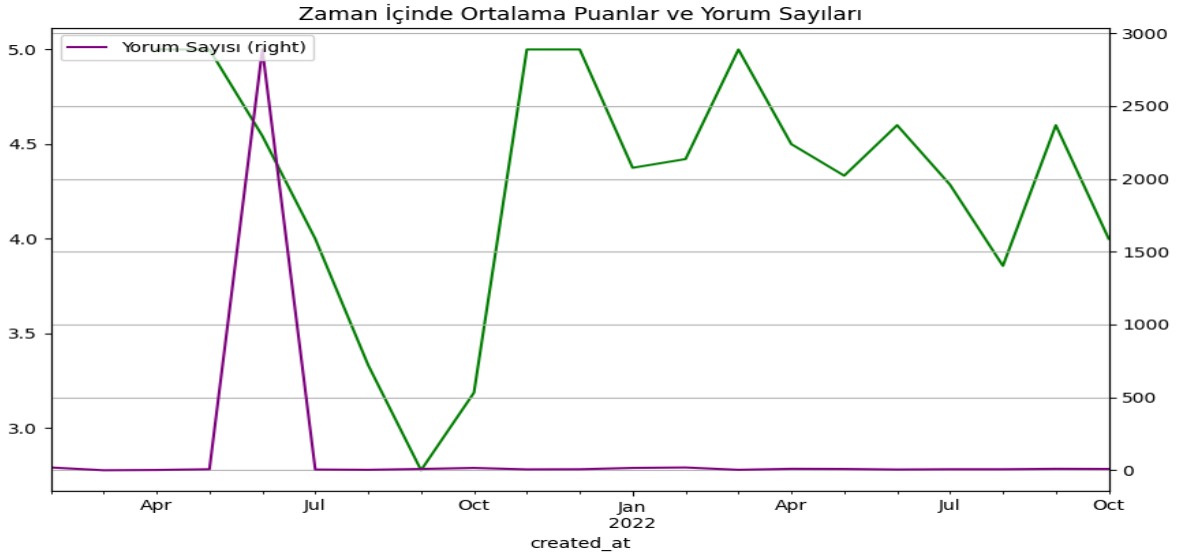
The "Cheeseburger Soup" recipe was identified as the most liked recipe.



Active users who comment most frequently have been identified and the top 5 most interacted user comments, user information and recipe names are presented.



Using time series graphs, the number of comments made on recipes and the change in average star values over time are shown.



Top 5 User Comments with the Highest Interaction, User Information, and Recipe Names

Comments that are generally positive or provide details about the recipe's content and express user satisfaction have received the most interaction. These comments have attracted users' attention and have had a significant impact on the platform. Here are the top 5 user comments with the highest interaction, along with user information and recipe names:

1. **Recipe Name: Cheeseburger Soup**
   - **User Name:** CJR58
   - **Comment:** "This looks very delicious. What would happen if I use ground turkey instead of beef? Also, can I use cauliflower instead of potatoes?"
2. **Recipe Name: Amish Breakfast Casserole**
   - **User Name:** Gail Lee
   - **Comment:** "This is just for me and my husband, so I made half the recipe. However, would you recommend making the full recipe instead of half?"
3. **Recipe Name: Amish Breakfast Casserole**
   - **User Name:** lpetty
   - **Comment:** "This is my favorite breakfast casserole recipe ever! I always make it for my guests, and they all love it."
4. **Recipe Name: Cheeseburger Soup**
   - **User Name:** tanjan4evr
   - **Comment:** "CAN I USE CAULIFLOWER INSTEAD OF POTATOES?"
5. **Recipe Name: Amish Breakfast Casserole**
   - **User Name:** krysfielder
   - **Comment:** "This is an excellent recipe! Yes, you can make a healthier option using lean sausages and low-fat cheese. Also, I love adding green peppers."

Monitoring and Evaluation Plan

- Implementation of Recommendations: It is essential to implement the recommendations based on the insights obtained from the analysis. This includes optimizing popular recipes, improving user interaction features, and addressing concerns related to sentiment.

- Monitoring and Feedback: Continuous monitoring of user interaction metrics, sentiment trends, and platform performance indicators is necessary. User feedback should be collected, and adjustments based on feedback should be made to continually improve platform performance.

- Regular Reporting: Regular reports should be presented to the management team, highlighting key findings, implemented changes, and their impact on user satisfaction and platform performance.

Final Recommendation to the Board of Directors

Based on the analysis, I recommend focusing on implementing the recommended improvements, with a focus on increasing user interaction in popular recipes and comment features. Additionally,

prioritizing sentiment analysis and addressing user feedback can lead to a more satisfying user experience and improved platform performance. By implementing these recommendations and continuously monitoring platform performance, user satisfaction can be maintained, leading to long-term growth and success.