Table of Contents

Executive Summary

In this paper, it thoroughly examines the work light products of Horizon's competitor brands, with emphasis on the KODA brand. The key goals are to forecast sales for the remainder of the year and to do brand perceptual mapping with Natural Language Processing (NLP).

Time series analysis is used to forecast future sales trends, which helps find possible high-performing products. The study identifies key patterns and trends that can help guide strategic decisions.

NLP approaches are used to understand customer feelings and product attributes through brand perceptual mapping, which uses customer feedback and reviews.Based on the findings of this analysis, it can help KODA's go-to-market strategy, assuring targeted efforts to increase sales and improve brand perception.

Company overview

Horizon Brands is committed to developing creative, high-quality solutions that value its client's time, money, and resources. The company is founded on basic values of honesty, quality, innovation, cooperation, accountability, and respect, and it focuses on making long-lasting goods that improve people's daily lives. Horizon Brands strives for excellence via continual improvement and open communication, aiming to create higher value where existing goods fall short.

Horizon Brands is deeply involved in philanthropy, organizing events to benefit global communities, providing internships to enhance youth talents, and implementing environmentally friendly methods. The organization promotes a diverse and inclusive workplace, valuing different cultures and perspectives in order to boost innovation and productivity. Horizon Brands is an equal opportunity company that ensures non-discrimination and reasonable accommodations for those with disabilities, encouraging all employees to thrive.

Introduction

Horizon Brands is a company which pushes the limits to add more to your everyday life. The result is a thriving and ever-expanding portfolio of uniquely useful Automotive Accessories, Consumer Electronics, and Home & Lifestyle Products for people who care about quality and value.This project aims to leverage advanced analytical techniques to accurately predict sales and understand the brand perception of Horizon's requirements, with a specific focus on the KODA brand. By analyzing customer feedback and performing sentiment analysis, we seek to identify key factors influencing customer preferences and pricing. The insights gained from this analysis will be instrumental in formulating a robust go-to-market strategy for KODA, ensuring the brand can effectively compete and thrive in the market.

The objective of this analysis is two-fold: First, to predict sales for the remainder of the year using time series analysis, which helps in understanding the sales dynamics and identifying potential high-performing products. Second, to perform brand perceptual mapping using Natural Language Processing (NLP) on customer reviews and feedback. This will provide insights into customer sentiments and the perception of KODA relative to its competitors. By leveraging these advanced analytical techniques, Horizon Brands aims to optimize product offerings, enhance customer satisfaction, and strengthen KODA's market position. The findings from this analysis will enable Horizon Brands to devise strategies that not only boost sales but also differentiate KODA’s position in a competitive market landscape.

Data Preprocessing

Loading Datasets

- Loaded the primary dataset containing the top 100 work light products, along with the master sheet that includes detailed specifications for all work light products.

Handling Missing Values

- Replaced 'N/A' values with pandas' pd.NA and dropped rows containing missing values to clean the dataset.

Merging Datasets

- Ensured that the 'ASIN' columns in both datasets are of the same type (str).
- Merged the primary dataset with another dataset containing customer reviews on the 'ASIN' column to integrate product attributes with customer sentiment and ratings.

Grouping and Aggregating Data

- Grouped the dataset by 'Brand' and calculated the sum of 'Revenue' for each brand to identify the top 10 brands by revenue.
- Aggregated data by 'Brand' to compute the average sentiment score and average ratings.

Standardizing Column Names

- Ensured consistent naming conventions across datasets for streamlined analysis and merging processes.

Analysis

Exploratory Data Analysis (EDA)

Descriptive Analysis has been performed to summarize and understand the main features of the dataset. It provides insights into the distribution, central tendency, and variability of key variables, which helps in identifying patterns and anomalies. Also provides a clear and concise overview of the dataset's key metrics. This foundational understanding helps in identifying trends, guiding further analysis, and making informed business decisions.(Table 1)

Correlation analysis was performed to examine the relationships between different numerical attributes. A correlation matrix was generated and visualized using a heatmap to identify significant correlations. We have used Price, Ratings, Reviews, Sales. The correlation matrix was computed using the .corr() method, which calculates the pairwise correlation of the columns. The correlation values range from -1 to 1, where 1 implies a perfect positive correlation and -1 implies a perfect negative correlation and 0 implies no correlation. The result of this process is a heatmap that visually represents the correlation coefficients between the selected variables. Each cell in the heatmap shows the correlation value between a pair of variables. This visualization helps quickly identify strong correlations (either positive or negative) and can guide further analysis or feature selection for modeling.(Table 2)

Time Series Analysis

In this analysis, we utilized the Prophet model to forecast the revenue for the next six months based on historical revenue data from January to June 2024. The historical data was first converted to a suitable format and then fed into the Prophet model, which was trained to understand the underlying trends. We then generated future dates for the next six months and used the model to predict future revenues. The resulting forecast indicated a declining trend in revenue, which was visualized through a plot showing both the forecasted values and their confidence intervals. The detailed forecast included the predicted revenue (yhat), along with lower and upper bounds (yhat_lower and yhat_upper), providing a comprehensive outlook on expected future performance.(Table 3 & Graph 1)

After renaming and formatting the data, we fit the Prophet model, incorporating price, ratings, and reviews as additional regressors. We then generated future predictions for the next six months, assuming constant values for these regressors and making logical adjustments to the reviews. The model provided a detailed forecast with upper and lower confidence intervals, indicating a steady increase in sales, which was visualized in a plot along with the forecasted values table. This analysis helps in understanding potential sales trends and preparing for future demand.The table (Table 4) provides specific forecast values for each month, including the predicted sales (yhat), and the lower (yhat_lower) and upper (yhat_upper) bounds of the confidence interval. The model predicts a steady increase in sales over the forecasted period, indicating positive growth trends. This forecast helps in strategic planning and resource allocation to meet the anticipated demand.(Graph 2)

We performed a time series analysis to forecast sales for the next six months. Initially, we prepared a dataset containing monthly sales data from January to June 2024. The resulting plot, combining historical sales and predicted values, illustrates a steady upward trend in sales, aiding in strategic planning and resource allocation. This analysis helps in understanding potential future sales trends and making informed business decisions.(Graph 3)

Testing and Training for the model

Model review: We used the Prophet framework for sales forecasting, incorporating additional factors like price and ratings to improve prediction accuracy.(Table 6)

Data preparation: Historical sales data was cleaned and prepared, ensuring completeness and accuracy. Data was split into training and testing sets for each unique ASIN.

Model training: The Prophet model was trained on historical sales data. Additional regressors (price and ratings) were included to enhance the model's predictions.

We trained and tested a sales forecasting model using the Prophet framework, incorporating price and ratings as additional factors. The model was evaluated using several metrics:

- Mean Absolute Error (MAE): 304.07
- Mean Squared Error (MSE): 307,887.46
- Mean Absolute Percentage Error (MAPE): 0.4576

The MAE value of 304.07 indicates the average absolute difference between the predicted and actual sales.(Table 7)

Web scraping

We utilized Outscraper to gather customer reviews for each product under various ASINs. This process yielded approximately 8,000 customer reviews. These reviews were then analyzed using Natural Language Processing (NLP) techniques to gain insights into customer sentiments and feedback. This analysis helps us understand customer perceptions and improve product offerings.

NLP Analysis

Natural Language Processing was performed for the customer reviews(Table 8). It was used to analyze customer reviews and feedback for Horizon Brands' competitor work light products.

Specifically, TextBlob was utilized to perform sentiment analysis on the reviews, enabling us to gauge customer satisfaction and identify common themes. Sentiment analysis(Table 9) involved calculating the sentiment polarity of each review, which ranges from -1 (negative) to 1 (positive). This was done by applying the TextBlob library to the 'Reviews' column of the dataset, resulting in a new column that contained the sentiment scores for each review.It was performed for top 100 ASIN data which is shown in Table 10. These scores were then aggregated by brand to determine the average sentiment for each brand, providing insights into overall (Table 11). Additionally, the analysis included creating a perceptual map that visually represents KODA's brand and its competitors based on customer sentiments and product features. By plotting the average sentiment scores against the average ratings for each brand, we were able to visualize the positioning of each brand in the market. This perceptual map helps in identifying top brands. Specifically highlighting areas for potential improvement and differentiation. The visual representation of sentiment and ratings provides actionable insights for strategic decision-making and enhancing the brand's market positioning.

Conclusion and Recommendation

This analysis of Horizon Brands' work light products involved thorough data preprocessing, exploratory data analysis, time series forecasting, and NLP-based sentiment analysis. By cleaning and merging datasets, we ensured a cohesive dataset for analysis. The EDA phase highlighted key relationships between sales, price, ratings, and reviews, which informed our understanding of product performance.

Time series analysis using the Prophet model provided forecasts of future sales trends, aiding strategic planning. NLP techniques, specifically sentiment analysis with TextBlob, revealed customer satisfaction levels and common themes in reviews. The resulting perceptual map highlighted KODA's brand positioning relative to competitors. Overall, the insights from this analysis offer a clear framework for strategic decision-making, optimizing product offerings, and enhancing KODA's market position.

We recommend Horizon Brands to focus on the top 20 brands identified through the brand perceptual map and NLP analysis. By analyzing customer sentiments and sales trends, these brands show strong potential for driving revenue and enhancing market presence. We suggest incorporating sales trend analysis to understand their performance over time, which will guide strategic decisions and optimize marketing efforts. This approach aligns with our objective to leverage advanced analytics for boosting sales and improving brand perception, ensuring that KODA effectively competes and thrives in the market.

Appendices

Column Header	Description	Type	Values
Date	The date of the record	Continuous	Date values
Product_Category	Category of the product	Categorical	Work Light
Product Details	Details of the product	Categorical	Text description
ASIN	Amazon Standard Identification Number	Categorical	Unique identifier
URL	URL to the product page	Categorical	URL string
Image URL	URL to the product image	Categorical	URL string
Brand	Brand name of the product	Categorical	Brand names
Price $	Price of the product	Continuous	Monetary values
Sales	Number of units sold	Continuous	Numerical values
Revenue	Total revenue generated	Continuous	Monetary values

Variables used for analysis

```
In [236]: df.describe()
```

	Date	Lumens	Price	Sales	Revenue	BSR	Fees	Active Sellers	Ratings	Reviews	Images	Review velocity	Weight
count	600	401.000000	598.000000	559.000000	557.000000	600.000000	594.000000	600.000000	600.000000	600.000000	600.000000	587.000000	596.000000
mean	2024-03-16 20:00:00	3987.581047	42.991722	791.522361	32741.117307	20811.841667	12.587475	4.895000	4.526500	1674.685000	7.138333	26.689949	2.294950
min	2024-01-01 00:00:00	140.000000	4.990000	21.000000	543.910000	559.000000	0.000000	1.000000	0.000000	0.000000	1.000000	0.000000	0.020000
25%	2024-02-01 00:00:00	1000.000000	23.040000	359.000000	10557.360000	13475.750000	8.930000	1.000000	4.400000	188.000000	6.000000	8.000000	0.729413
50%	2024-03-16 12:00:00	2100.000000	32.990000	505.000000	17936.550000	21369.000000	10.350000	1.000000	4.500000	708.500000	7.000000	15.000000	1.190599
75%	2024-05-01 00:00:00	4800.000000	45.990000	739.000000	29473.080000	28747.500000	13.427500	3.000000	4.700000	2075.750000	8.000000	26.000000	1.832500
max	2024-06-01 00:00:00	21000.000000	229.000000	15659.000000	747150.000000	41210.000000	55.780000	57.000000	5.000000	20252.000000	15.000000	1415.000000	20.954115
std	NaN	5241.632893	32.063496	1260.043891	63967.383460	9412.059434	7.200212	8.394726	0.254734	3022.350013	1.849748	65.808642	3.309542

Table 1- Descriptive statistics

Table 2- Correlation Matrix

	ds	yhat	yhat_lower	yhat_upper
6	2024-06-30	53926.176612	-591.164955	112104.679433
7	2024-07-31	24499.206671	-32109.643594	80715.072503
8	2024-08-31	-4927.763269	-64067.615579	52877.153518
9	2024-09-30	-33405.476115	-83951.124768	26665.417796
10	2024-10-31	-62832.446056	-120898.361479	-9832.331984
11	2024-11-30	-91310.158901	-144771.004833	-31954.453465

Table 3-Display Revenue forecasted values

7]:

	ds	yhat	yhat_lower	yhat_upper
6	2024-06-30	469.016148	465.430897	472.822330
7	2024-07-31	497.727660	485.355988	510.507675
8	2024-08-31	526.439171	501.537768	551.579284
9	2024-09-30	554.219094	514.350373	592.655590
10	2024-10-31	582.930605	527.540964	639.768116
11	2024-11-30	610.710529	537.071746	686.986763

Table 4-Display Sales forecasted values

```
Error processing ASIN B09G9RPR4G: Dataframe has less than 2 non-NaN rows.
Error processing ASIN B0CCKT3HJH: Dataframe has less than 2 non-NaN rows.
             ds        ASIN         yhat   yhat_lower   yhat_upper
0    2024-05-31  B088XWTWPM   253.740907   195.383210   306.435508
1    2024-06-30  B088XWTWPM    32.856671   -21.218577    82.208012
2    2024-07-31  B088XWTWPM  -195.390372  -246.770358  -141.856723
3    2024-08-31  B088XWTWPM  -423.637416  -475.858448  -369.209413
4    2024-09-30  B088XWTWPM  -644.521651  -695.411470  -593.170664
..          ...         ...          ...          ...          ...
604 2024-07-31  B001BM5YZW -1694.731609 -2034.245999 -1333.578327
605 2024-08-31  B001BM5YZW -2403.403345 -2738.267367 -2050.578902
606 2024-09-30  B001BM5YZW -3089.214703 -3451.209259 -2750.973814
607 2024-10-31  B001BM5YZW -3797.886439 -4156.593846 -3450.279676
608 2024-11-30  B001BM5YZW -4483.697796 -4814.884021 -4128.422517

[609 rows x 5 columns]
```

Table 5 -Display All ASIN Sales forecasted values

```python
# Calculate accuracy metrics
if 'y' in all_predictions.columns and 'yhat' in all_predictions.columns:
    y_actual = all_predictions['y'].dropna()
    y_pred = all_predictions['yhat'].dropna()

    print(f"Number of actual values: {len(y_actual)}")
    print(f"Number of predicted values: {len(y_pred)}")

    if len(y_actual) > 0 and len(y_pred) > 0:
        mae = mean_absolute_error(y_actual, y_pred)
        mse = mean_squared_error(y_actual, y_pred)
        mape = mean_absolute_percentage_error(y_actual, y_pred)

        print(f"MAE: {mae}")
        print(f"MSE: {mse}")
        print(f"MAPE: {mape}")
    else:
        print("Error: Mismatch in number of actual and predicted values")
else:
    print("Error: 'y' or 'yhat' column not found in all_predictions")

# Save all predictions to a CSV file
all_predictions.to_csv('all_asin_predictions.csv', index=False)

# Print all predictions
print(all_predictions)
```

Table 6- Training and testing the accuracy

```
MAE: 304.07122560418253
MSE: 307887.46219569637
MAPE: 0.45761107405569107
            ds         yhat       y   yhat_lower   yhat_upper
0   2024-05-01   328.244501   515.0   -29.668554   714.304388
1   2024-06-01  4341.783121  5698.0  2527.720559  6023.637009
2   2024-06-01   268.356057   411.0   254.165270   283.260370
3   2024-06-01   331.401669   576.0   331.401633   331.401714
4   2024-05-01  1704.319290  2045.0   720.831478  2743.251591
..         ...          ...     ...          ...          ...
64  2024-06-01   663.487645  1100.0   552.485270   776.067899
65  2024-06-01   245.265568   497.0   110.547621   349.750785
66  2024-05-01   440.839240   291.0   440.839231   440.839251
67  2024-06-01  5741.066951  8579.0  2063.163871  9560.408447
68  2024-05-01   486.286236   600.0  -889.675784  1837.212732

[69 rows x 5 columns]
```

Table 7 - Model Accuracy results

	product_asin	body
0	B07G9X19G1	I received one of these as a gift and liked th...
1	B07G9X19G1	I have one of these in each car, around the ho...
2	B07G9X19G1	I'm updating my review with more information a...
3	B07G9X19G1	There are tons of work lights of this type out...
4	B07G9X19G1	I was surprised and how versatile these lights...

Table 8-The top 5 head of the reviews

	product_asin	body	sentiment
0	B07G9X19G1	I received one of these as a gift and liked th...	0.182130
1	B07G9X19G1	I have one of these in each car, around the ho...	0.233333
2	B07G9X19G1	I'm updating my review with more information a...	0.260123
3	B07G9X19G1	There are tons of work lights of this type out...	0.364773
4	B07G9X19G1	I was surprised and how versatile these lights...	0.080357
...	...	...	...
8604	B0CRRKWJ1Q	Dewalt Batteries work perfect. They weight th...	0.361111
8605	B0CRRKWJ1Q	The media could not be loaded. This LED light ...	0.276296
8606	B0CRRKWJ1Q	The media could not be loaded. This is a nice ...	0.276440
8607	B0CRRKWJ1Q	It works great, Very bright and handy to be ab...	0.702500
8608	B0CRRKWJ1Q	Nice and bright. High brightness & low brightn...	0.502000

ꞁ609 rows × 3 columns

Table 9- Sentiment Analysis of the Reviews

	ASIN	Top 100 ASIN	Brand	Ratings
0	B0CD7B5SQ6	B0CD7B5SQ6	Tresda	4.7
1	B0BZY4J3PV	B0BZY4J3PV	TECHMUR	4.3
3	B088XWTWPM	B088XWTWPM	DEWALT	4.8
4	B00SKOCRCW	B00SKOCRCW	DEWALT	4.8
6	B07G9FWBCL	B07G9FWBCL	Milwaukee	4.8
...	...	...	...	...
583	B0CHDQCKK4	B0CHDQCKK4	wokelux	4.3
584	B0CHDQCKK4	B0CHDQCKK4	wokelux	4.3
585	B001BM5YZW	B001BM5YZW	Woods	4.4
591	B001BM5YZW	B001BM5YZW	Woods	4.4
592	B001BM5YZW	B001BM5YZW	Woods	4.4

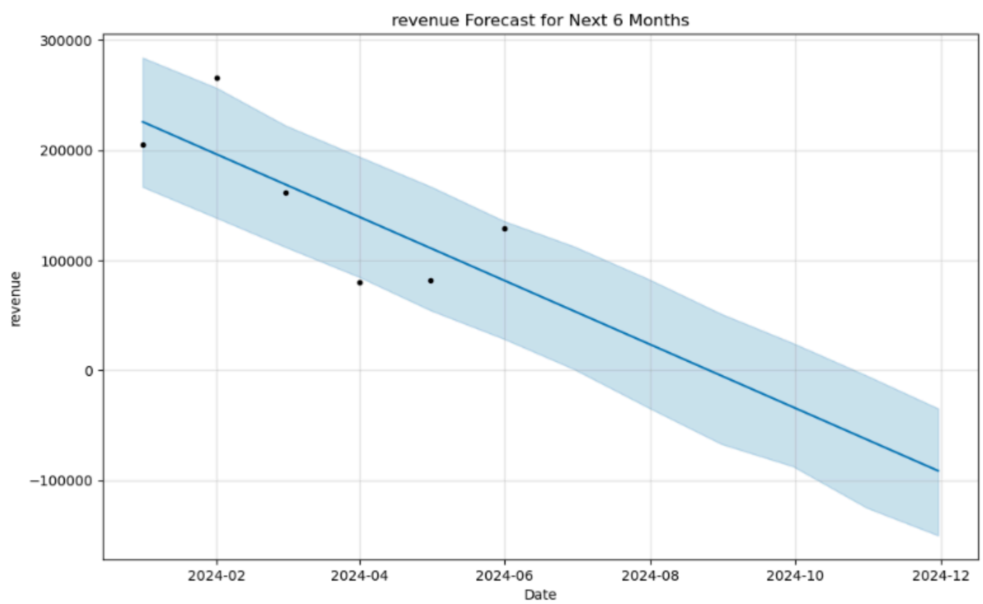
418 rows × 4 columns

Table 10-Top 100 ASIN Ratings

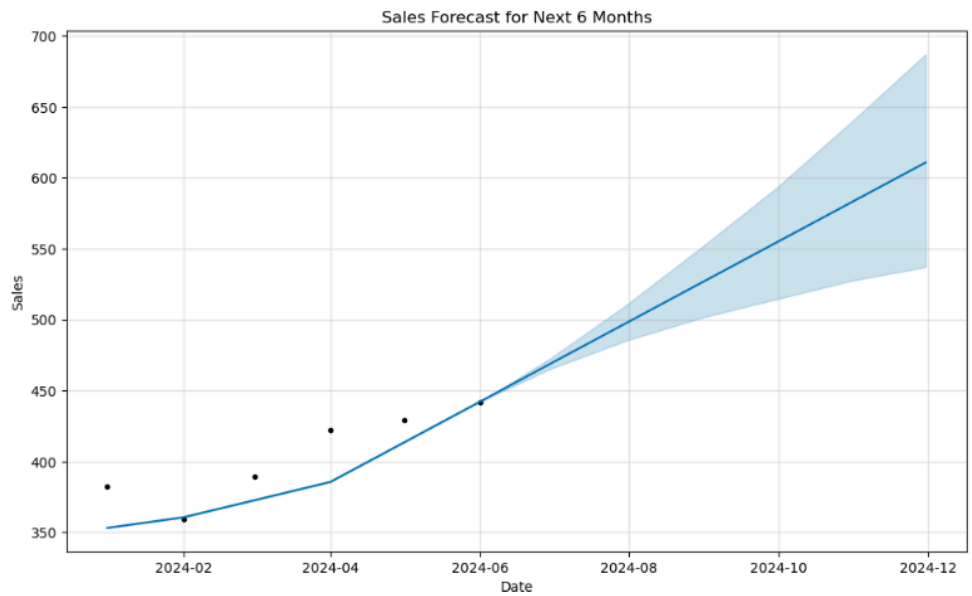
	Brand	sentiment	Ratings
0	Aiptertet	0.275072	4.275000
1	Anylight	0.332744	4.700000
2	Astro Pneumatic Tool	0.257464	4.540000
3	Azocek	0.367173	4.550000
4	Bell+Howell	0.264357	4.200000
..	...	...	...
70	lexall	0.302238	4.183333
71	ropelux	0.328035	4.350000
72	sunzone	0.335079	4.500000
73	tekstap	0.348013	4.280000
74	wokelux	0.305653	4.441667

[75 rows x 3 columns]

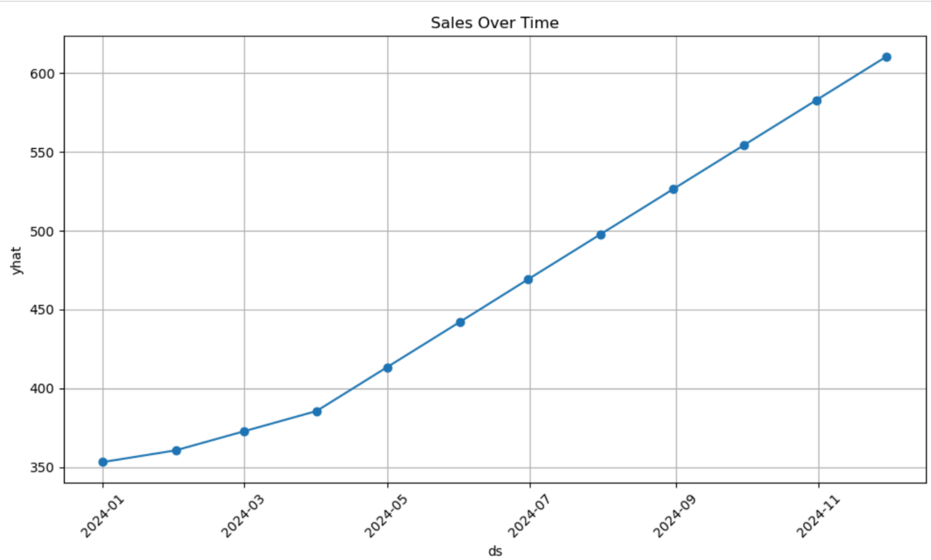
Table 11-Average Sentiment score and ratings



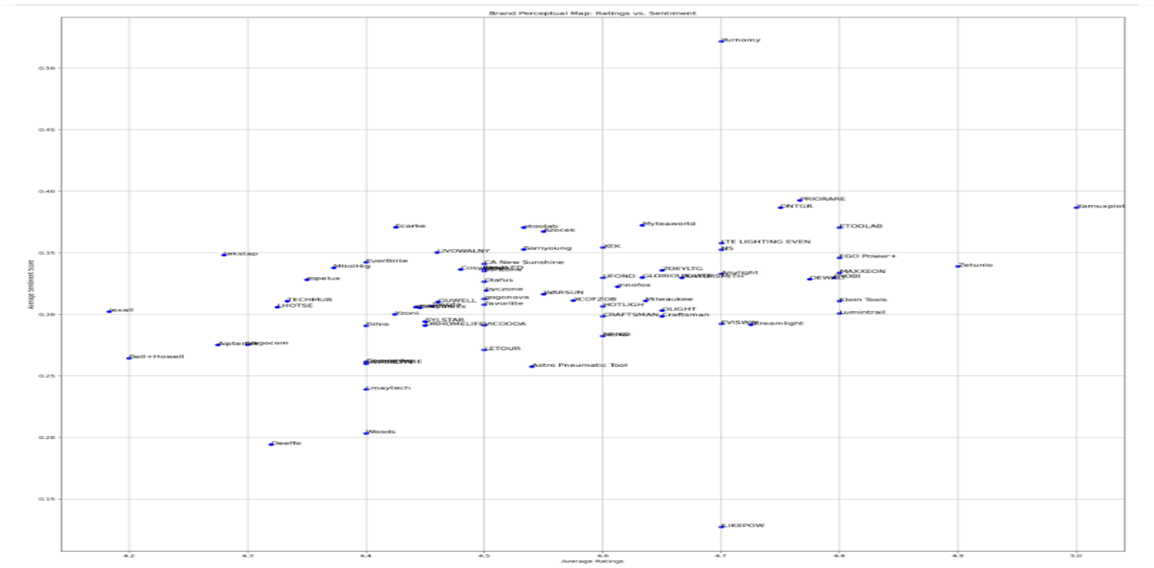
Graph 1-Revenue forecast for next 6 months



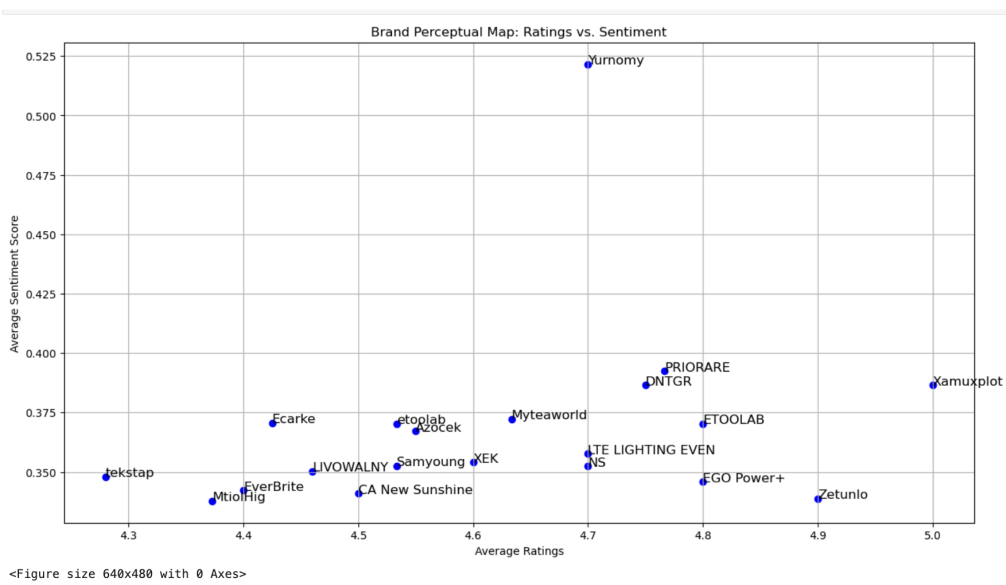
Graph 2-Sales forecast for next 6 months



Graph 3-Time series graph



Graph 4- Brand Perceptual Map: Ratings vs. Sentiment



Graph 5-Brand Perceptual Map: Ratings vs. Sentiment(FOR TOP 20)