HORIZON
BRANDS

Price Prediction Analysis

Akshaya Ramprasad
Nivedya Manikandanpillai
Ardha Anil

CONTENTS

HORIZON
BRANDS

Introduction

Horizon Brands is an innovative consumer goods company offering high-quality products across Home Appliances, Personal Care, Food and Beverages, and Electronics.This project aims to predict sales for the competitor brands of KODA for the rest of the year using advanced analysis and understand customer views through feedback and sentiment analysis. By doing so, Horizon Brands seeks to improve product offerings, enhance customer satisfaction, and strengthen KODA's market position. The insights gained will help create a robust strategy for KODA to effectively compete in the market.

Objective 1-Sales trend analysis on Koda’s competitiors

Methodology

- Data Collection: Gather 6 months of competitors' work light sales data.
- Data Preprocessing: Clean and prepare data, handle missing values, and transform date columns for time series analysis.
- Prophet Library: Use Prophet tool for accurate time series forecasting.
- Model Building: Train the Prophet model on historical sales data to capture trends and patterns.
- Forecasting: Generate future sales predictions with the trained model.

Objective 2: To identify top competitive brands using Ratings and performing sentiment analysis on customer reviews

Methodology:

- Data Collection: Gather customer reviews and feedback from Amazon.
- Data Preprocessing: Clean and standardize the text data.
- Sentiment Analysis: Use TextBlob to calculate sentiment polarity scores for each review, ranging from -1 (negative) to 1 (positive).
- Aggregating Sentiment Scores: Calculate the average sentiment score for KODA’s competitors to gauge overall customer satisfaction.
- Perceptual Mapping: Create a visual representation of competitors' brand positioning based on customer sentiment and review ratings to identify top brands.

Data Collection and Cleaning

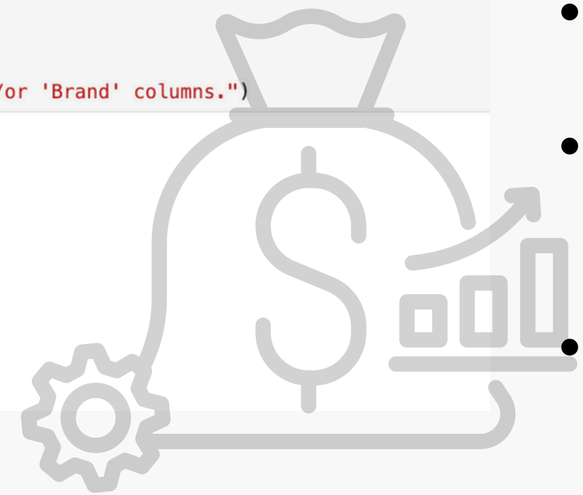
```python
df = pd.read_excel('Master Sheet updated- Top 100 Work Light.xlsx')

df = df[['Date','ASIN','Top 100 ASIN','Sales','Price','Ratings','Reviews']]

# Remove rows where 'Top 100 ASIN' is either missing or marked as '#N/A'
df = df[df['Top 100 ASIN'].notnull() & (df['Top 100 ASIN'] != '#N/A')]

# Ensure that 'Sales' has a value
df = df[df['Sales'].notnull()]


df['Date'] = pd.to_datetime(df['Date'], format='%m/%d/%y')
df = df.rename(columns={'Date': 'ds', 'Sales': 'y'})
```

Removes rows with missing or invalid 'Top 100 ASIN' values and ensures the 'Sales' column has values. The 'Date' column is converted to a datetime format, and columns are renamed ('Date' to 'ds' and 'Sales' to 'y') for further analysis or modeling.

Top 10 brands by revenue

```python
import pandas as pd

# Load the dataset
file_path = 'Master Sheet updated- Top 100 Work Light.xlsx'
df = pd.read_excel(file_path)

# Ensure that the columns 'Revenue' and 'Brand' exist in the DataFrame
if 'Revenue' in df.columns and 'Brand' in df.columns:
    # Group by 'Brand' and sum the 'Revenue'
    brand_revenue = df.groupby('Brand')['Revenue'].sum().reset_index()

    # Sort the grouped DataFrame by 'Revenue' in descending order
    top_10_brands_by_revenue = brand_revenue.sort_values(by='Revenue', ascending=False).head(10)

    # Display the top 10 brands by revenue
    print(top_10_brands_by_revenue)
else:
    print("The DataFrame does not contain 'Revenue' and/or 'Brand' columns.")
```

	Brand	Revenue
19	DEWALT	5551500.22
16	Coquimbo	922766.42
54	Milwaukee	886578.53
48	LIVOWALNY	785993.69
36	GUWELL	719618.55
31	Ecarke	520797.67
112	ropelux	486500.07
40	HOTLIGH	410057.46
81	Streamlight	368546.46
58	NEIKO	334989.35



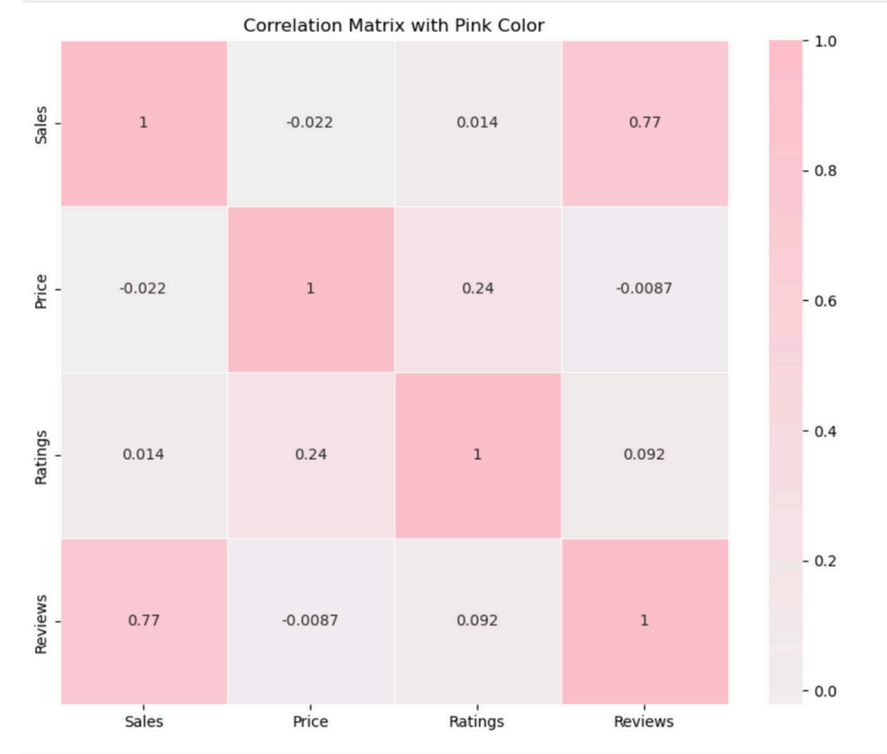
- Loaded the dataset into a pandas DataFrame and verified the 'Revenue' and 'Brand' columns.
- Grouped the data by brand and calculated total revenue for each.
- Sorted the total revenues in descending order.
- Identified the top 10 revenue-generating brands, with DEWALT leading at $5,551,500.22.
- Highlights top-performing brands in the work light market.

Descriptive Statistics

	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

The dataset shows a mean revenue of $32,741.12, with a wide range from $543.91 to $747,150.00. Average sales are 791.52 units, with prices averaging $42.99 and a maximum of $229.00. Reviews average 1,674.69, highlighting significant diversity in revenue, sales, pricing, and customer feedback.

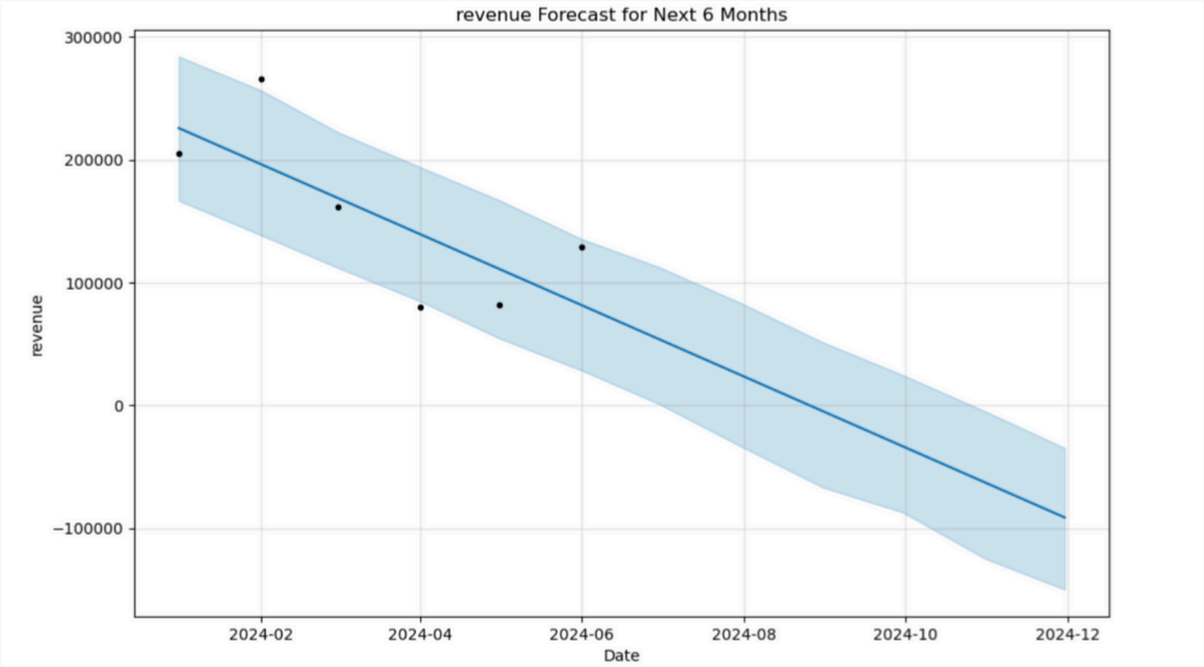
Correlation matrix



- The correlation matrix ranges from -1 to 1, showing relationships between dataset variable.
- The strongest positive correlation is between Sales and Reviews (0.77), indicating more reviews lead to higher sales.
- A moderate positive correlation between Price and Ratings (0.24) suggests higher-priced products often receive better ratings.
- Weak correlations close to zero indicate little to no linear relationship between other variable pairs.

Time Series model using Prophet model with one ASIN based on Revenue

```python
# Creating the DataFrame from the provided data
data = {
    'date': ['1/1/24', '2/1/24', '3/1/24', '4/1/24', '5/1/24', '6/1/24'],
    'sales': [204780.47, 266046.41, 161307.39, 80427.38, 81605.56, 128599.21]
}

df = pd.DataFrame(data)
df['date'] = pd.to_datetime(df['date'], format='%m/%d/%y')
df = df.rename(columns={'date': 'ds', 'sales': 'y'})

# Initialize the Prophet model
model = Prophet()

# Fit the model
model.fit(df)

# Create a DataFrame for future predictions (next 6 months)
future = model.make_future_dataframe(periods=6, freq='M')

# Predict future sales
forecast = model.predict(future)

# Plot the forecast
fig = model.plot(forecast)
plt.title('Sales Forecast for Next 6 Months')
plt.xlabel('Date')
plt.ylabel('Sales')
plt.show()

# Display forecasted values
forecast[['ds', 'yhat', 'yhat_lower', 'yhat_upper']].tail(6)
```

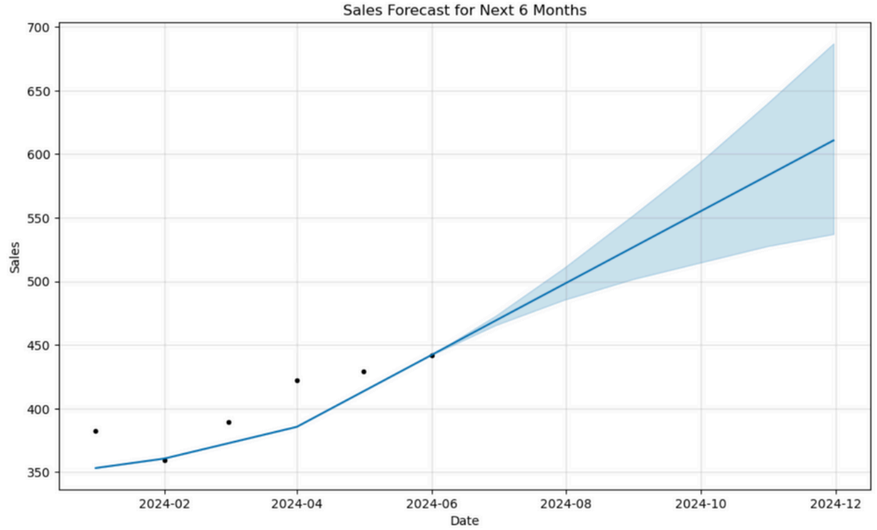


We used the Prophet model to forecast the next six months of revenue. First, we created a DataFrame with historical revenue data and formatted it for the Prophet model. we fit the model for one ASIN based on the revenue for the model to predict the future sales

Time Series model using Prophet model with one ASIN based on Sales

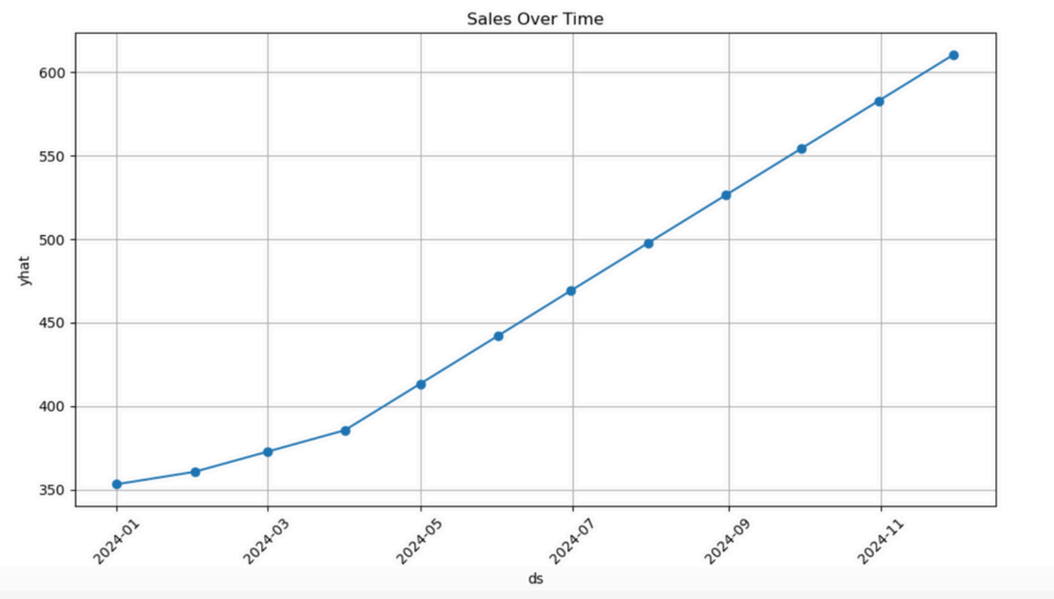
```python
# Creating the DataFrame from the provided data
data = {
    'date': ['1/1/24', '2/1/24', '3/1/24', '4/1/24', '5/1/24', '6/1/24'],
    'sales': [382, 359, 389, 422, 429, 442],
    'price': [112.79, 118.99, 118.99, 112.16, 118.99, 118.99],
    'ratings': [4.7, 4.8, 4.7, 4.7, 4.7, 4.8],
    'reviews': [30166, 26464, 22164, 22308, 23667, 25131]
}
```

	ds	yhat	yhat_lower	yhat_upper
6	2024-06-30	469.016148	464.571100	472.017507
7	2024-07-31	497.727660	484.063163	507.771349
8	2024-08-31	526.439171	500.322157	547.411148
9	2024-09-30	554.219094	514.890921	589.148889
10	2024-10-31	582.930605	528.860892	633.766236
11	2024-11-30	610.710529	539.074742	680.109219



Time Series graph

- Created a time series graph to forecast sales over the next six months.
- X-axis: dates; Y-axis: predicted sales ('yhat')
- Graph features: connected data points, markers, grid, and rotated x-axis labels.
- Visualization highlights sales trends and patterns.
- Combines historical sales (January to June 2024) with predictions.
- Shows a steady upward sales trend.



Time Series model using Prophet model with all ASIN's based on Sales

	ds	ASIN	yhat	yhat_lower	yhat_upper
0	2024-05-31	B088XWTWPM	392.833707	203.728771	582.431277
1	2024-06-30	B088XWTWPM	270.661024	-316.480108	884.251123
2	2024-07-31	B088XWTWPM	144.415917	-1013.717626	1343.340136
3	2024-08-31	B088XWTWPM	18.170811	-1779.934075	1888.798695
4	2024-09-30	B088XWTWPM	-104.001873	-2589.868105	2552.339863
..	...	...	...	...	...
604	2024-07-31	B001BM5YZW	-719.551529	-3834.559208	2433.932056
605	2024-08-31	B001BM5YZW	-1169.059646	-6224.647005	3738.179627
606	2024-09-30	B001BM5YZW	-1604.067501	-8734.250251	5414.652546
607	2024-10-31	B001BM5YZW	-2053.575618	-11534.539435	7208.426595
608	2024-11-30	B001BM5YZW	-2488.583474	-14262.147616	9686.505830

[609 rows x 5 columns]

We built a model using prophet to run the ASINs in a loop and predict the sales data for each ASIN over the next 6 months. As shown in the table below , the yhat , y lower and y upper values are given for each ASIN representing the mean , upper bound and lower bound values for forecasting.

Web Scraping for customer reviews

Outscraper

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.

Customer reviews sentiment analysis using Textblob

```
reviews_main.head()
```

	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...

The respesentation for the top 5 head of the customer sentiment analysis reviews

This table shows customer reviews with sentiment scores for different products. Each row includes the product ASIN, review text, and a sentiment score from -1 (negative) to 1 (positive). Positive reviews, like those for ASIN B07G9X19G1, indicate high customer satisfaction. This helps understand customer feedback and overall sentiment towards products.

	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

8609 rows × 3 columns

Training and Testing the model

```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)
```

```
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]
```

We trained and tested a sales forecasting model using the Prophet framework, incorporating price and ratings as additional factors. The model was evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE) with results of 304.07, 307,887.46, and 0.4576 respectively.The output shows the predicted sales (yhat), actual sales (y), and the prediction intervals (yhat_lower and yhat_upper) for future dates.

Average Brand Sentiment Ratings

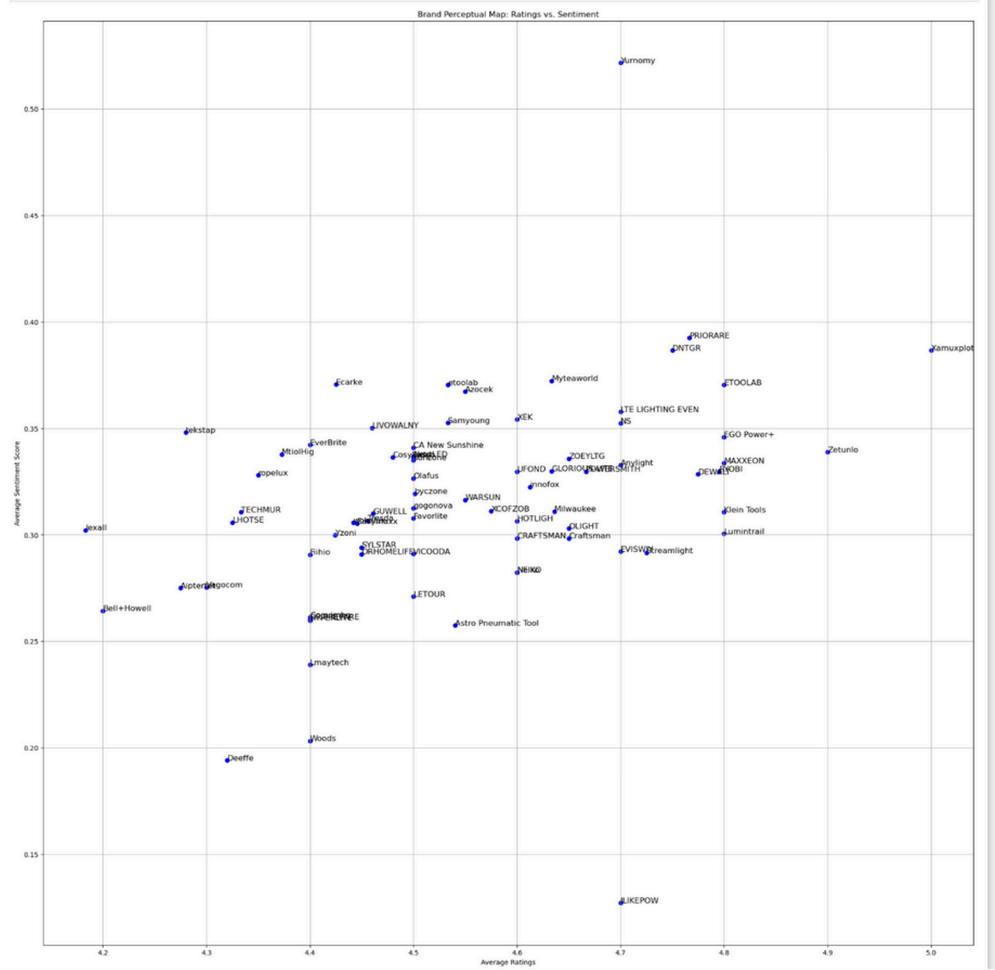
Dataframe showing the average sentiment score and ratings for each brand. Brands such as Anylight and Azocek have high ratings of 4.7 and 4.55, respectively, with corresponding sentiment scores of 0.332744 and 0.367173. Other brands like Bell+Howell and Astro Pneumatic Tool have lower sentiment scores and 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]

Brand Perceptual Map: Ratings vs. Sentiment

Brand Perceptual Map visualizes the relationship between average ratings and sentiment scores for various brands. It plots brands on a scale with average ratings from 4.3 to 5.0 (x-axis) and sentiment scores from 0.35 to 0.525 (y-axis).



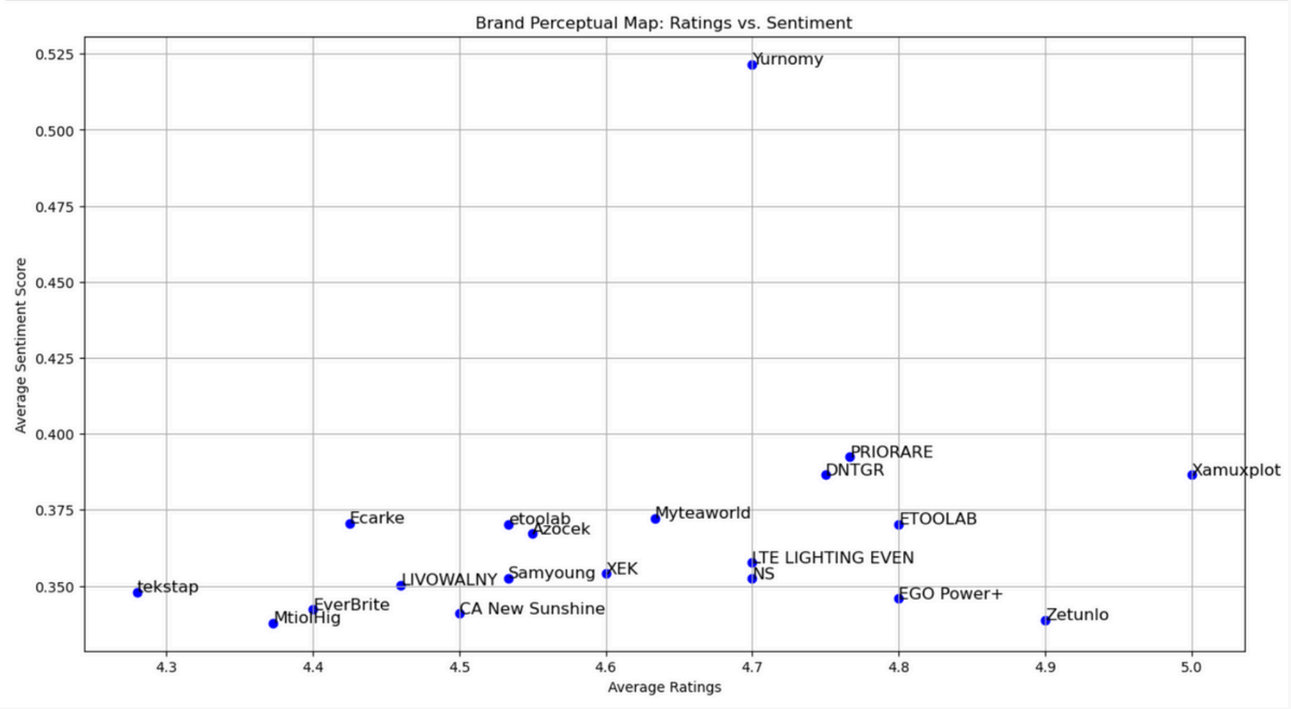
Brand Perceptual Map: Ratings vs. Sentiment for Top 20

The Brand Perceptual Map shows the relationship between average ratings and sentiment scores for top 20 brands:

- X-axis: Average ratings (4.3 to 5.0)
- Y-axis: Average sentiment scores (0.35 to 0.525)

Key insights:

- Xamuxplot: High rating (5.0) and high sentiment (0.5)
- Yurnomy: High sentiment (0.525) but lower rating (4.6)
- Zetunlo and EGO Power+: High ratings (above 4.8) but moderate sentiment (around 0.425)



Recommendation

- 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.

Thank you!