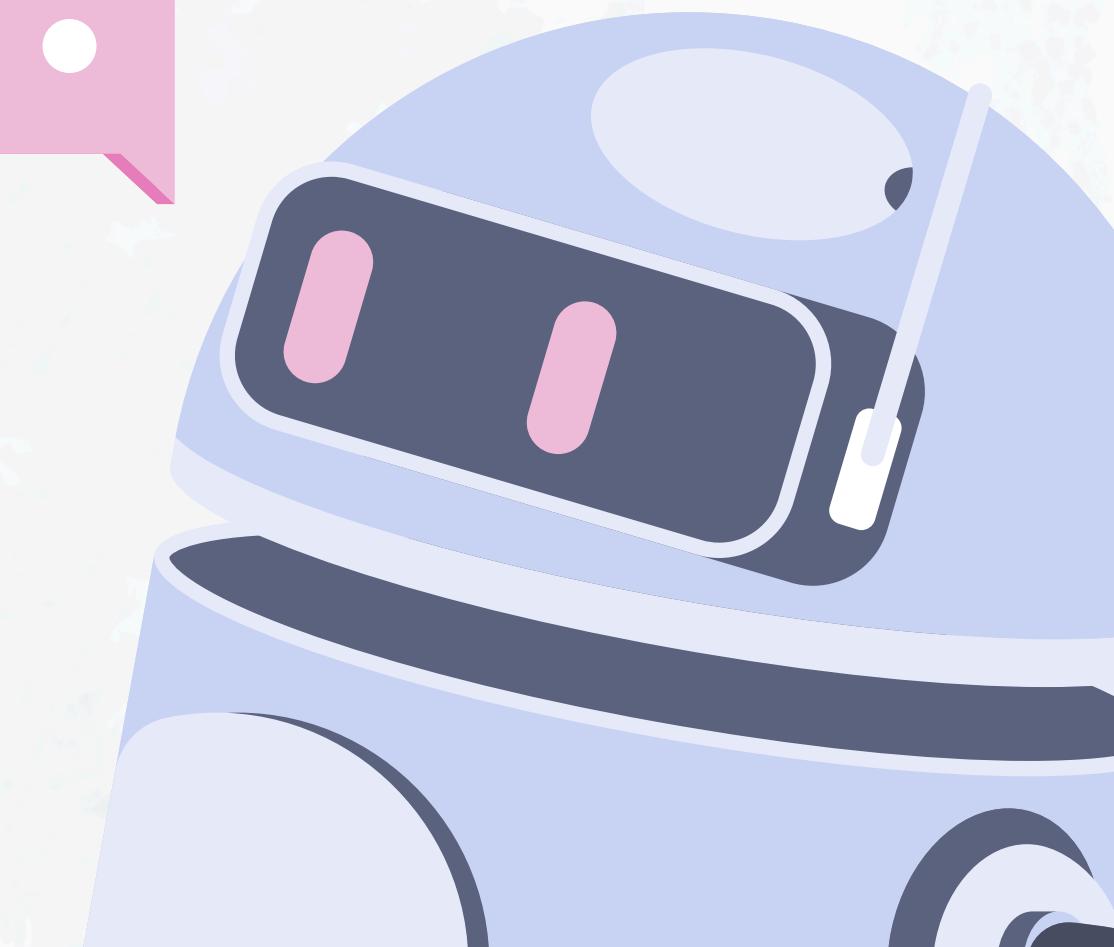


Sales Analysis

Forecasting discounts on goods.

Analysis of comment tone.

Building a recommendation system



Prepared by Popazov Petro AI-221

Content

- 01 → Tasks to Be Completed
- 02 → Dataset
- 03 → Used Technologies
- 04 → Task Implementation

01 Tasks to Be Completed

(a) Preprocessing and Data Preparation

Data cleaning from missing values, anomalies, and duplicates. Normalization and standardization of variables to ensure the correctness of further analysis.

(b) Data visualization

Creation of charts such as histograms, box plots, scatter plots, and heatmaps to identify patterns, correlations, and trends. Visualization of dependencies between product categories, prices, ratings, and sales volumes. Determination of the number of items by categories.

(c) Analysis of comments

Determination of mood (positive, negative, neutral) by utilizing text processing tools.

01 Tasks to Be Completed

(d) Forecasting reduced prices for goods.

Using linear regression and a neural network – build a model for forecasting reduced prices.
Conduct model evaluation (MAE, MSE, RMSE, R2).

(e) Building a personalized offer model.

Function returns the 5 best products with the highest similarity scores that the user has not yet purchased.

02 Dataset

Amazon Sales Dataset

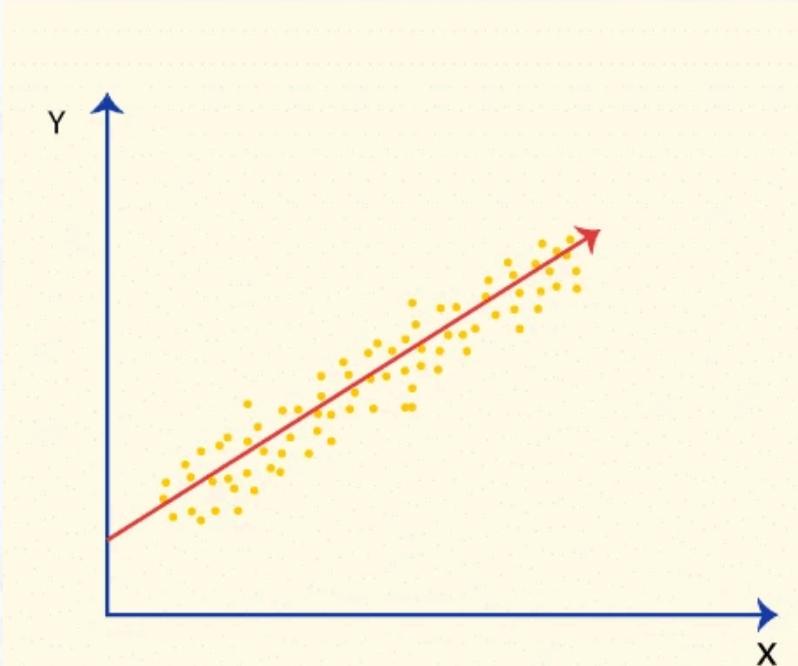
Contains detailed information about various products sold through Amazon, including such attributes as product pricing data, ratings, number of reviews, sales volume, and product descriptions



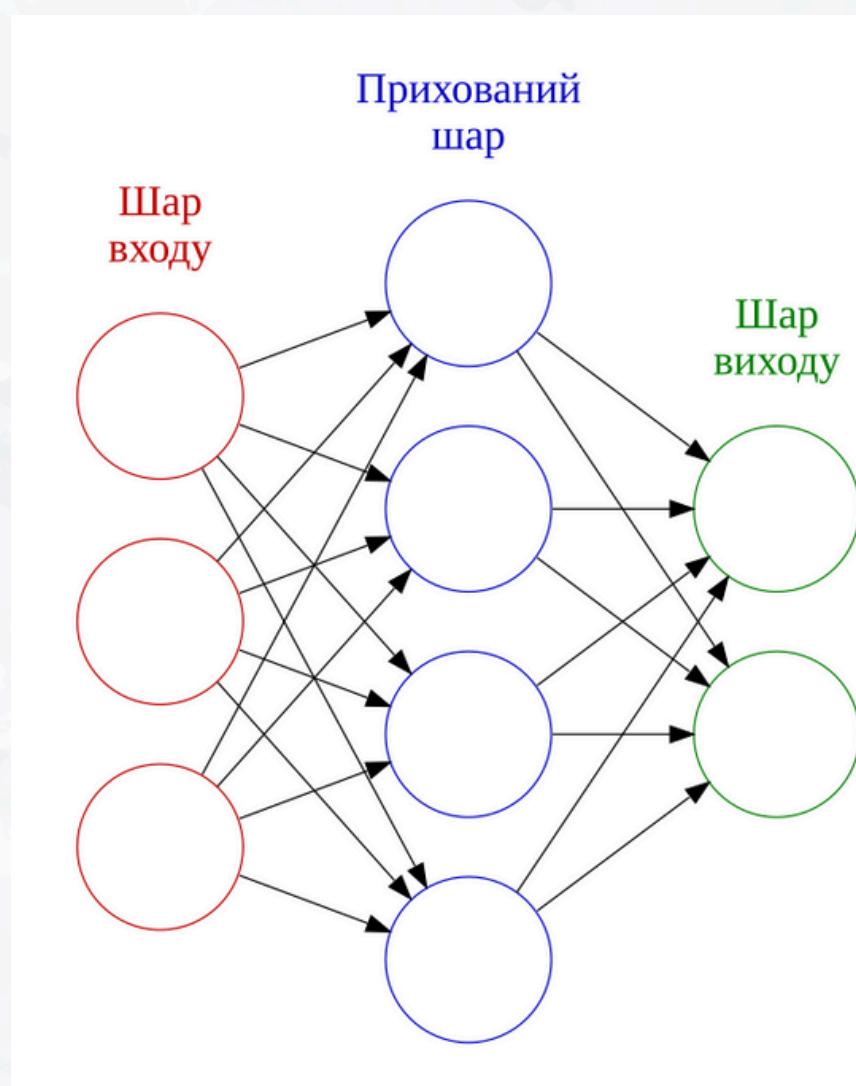
- product_id
- product_name
- category
- discounted_price
- actual_price
- discount_percentage
- rating
- about_product
- user_id
- review_title
- review_content

03 Used Technologies

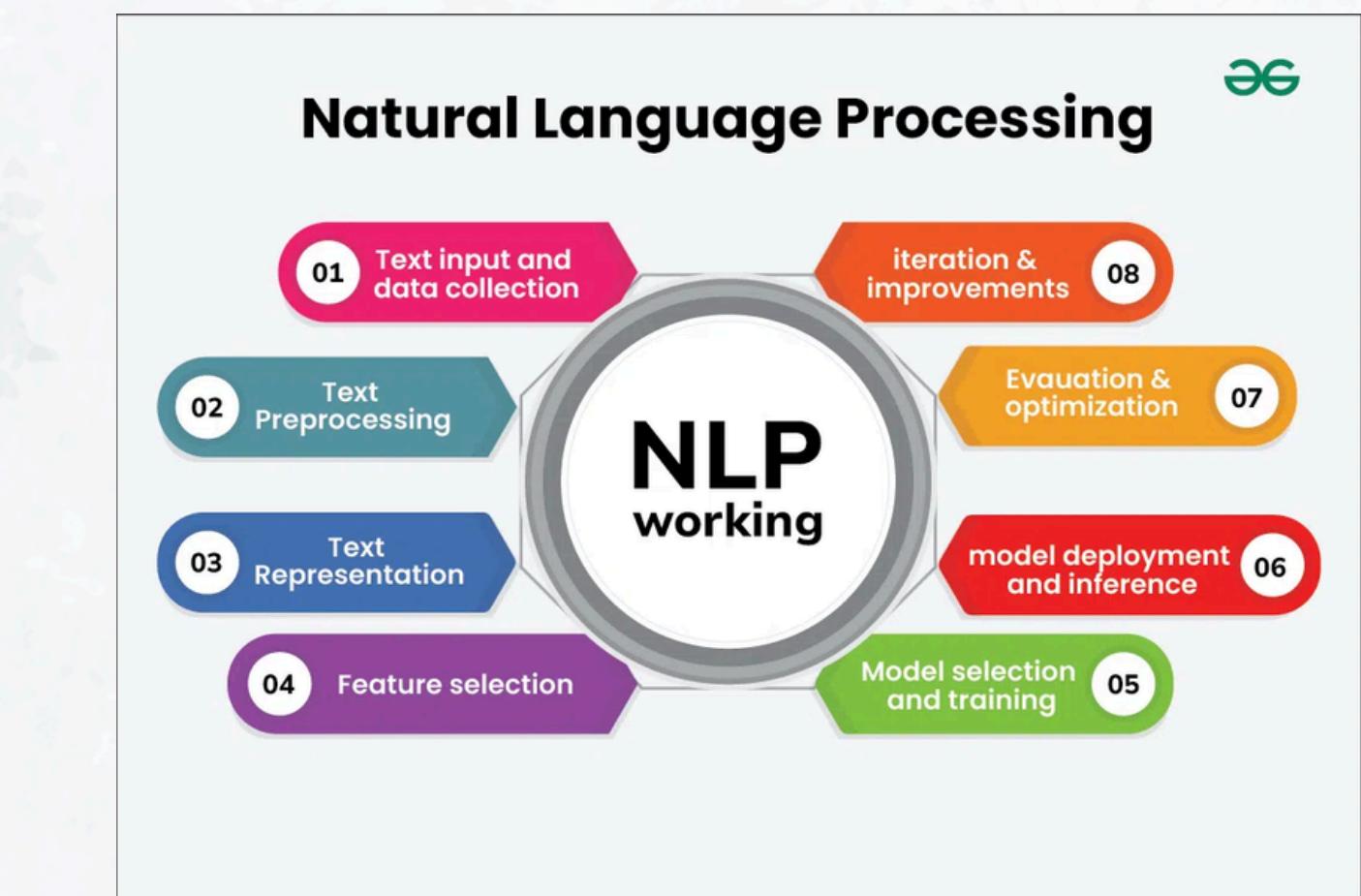
Linear regression



Neural network



NLP



TfidfVectorizer

cosine_similarity

04 Task Implementation

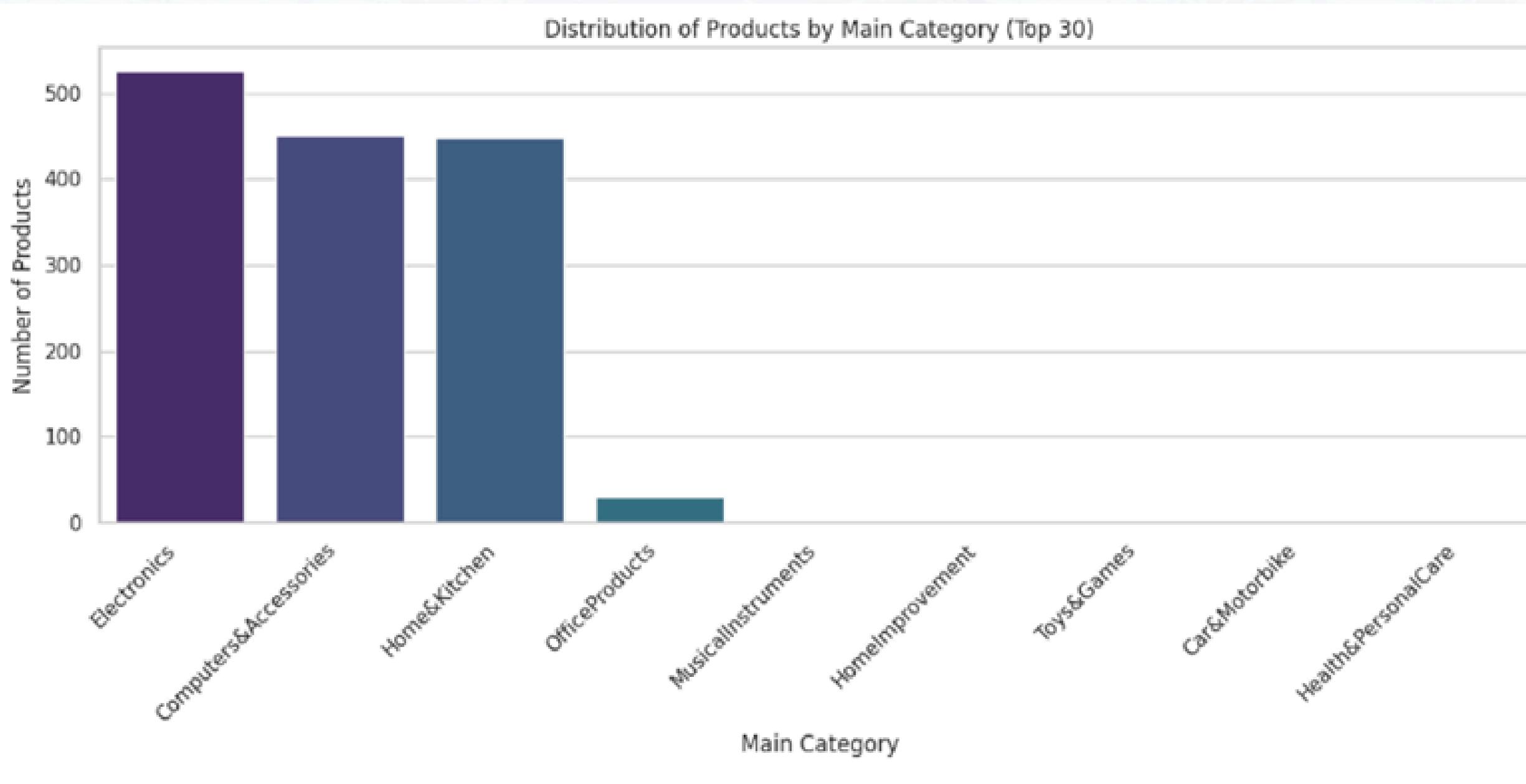
01 Previous processing

| product_name | category | discounted_price | actual_price | discount_percentage | rating | rating_count |
|---|---|------------------|--------------|---------------------|--------|--------------|
| Wayona Nylon Braided USB to Lightning Fast Cha... | Computers&Accessories Accessories&Peripherals ... | ₹399 | ₹1,099 | 64% | 4.2 | 24,269 |
| Ambrane Unbreakable 60W / 3A Fast Charging 1.5... | Computers&Accessories Accessories&Peripherals ... | ₹199 | ₹349 | 43% | 4.0 | 43,994 |
| Source Fast Phone Charging Cable & Data Sync U... | Computers&Accessories Accessories&Peripherals ... | ₹199 | ₹1,899 | 90% | 3.9 | 7,928 |

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1465 entries, 0 to 1464
Data columns (total 16 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   product_id       1465 non-null   object 
 1   product_name     1465 non-null   object 
 2   category         1465 non-null   object 
 3   discounted_price 1465 non-null   object 
 4   actual_price     1465 non-null   object 
 5   discount_percentage 1465 non-null   object 
 6   rating           1465 non-null   object 
 7   rating_count     1463 non-null   object 
 8   about_product    1465 non-null   object 
 9   user_id          1465 non-null   object 
 10  user_name        1465 non-null   object 
 11  review_id        1465 non-null   object 
 12  review_title     1465 non-null   object 
 13  review_content   1465 non-null   object 
 14  img_link         1465 non-null   object 
 15  product_link     1465 non-null   object 
dtypes: object(16)
```

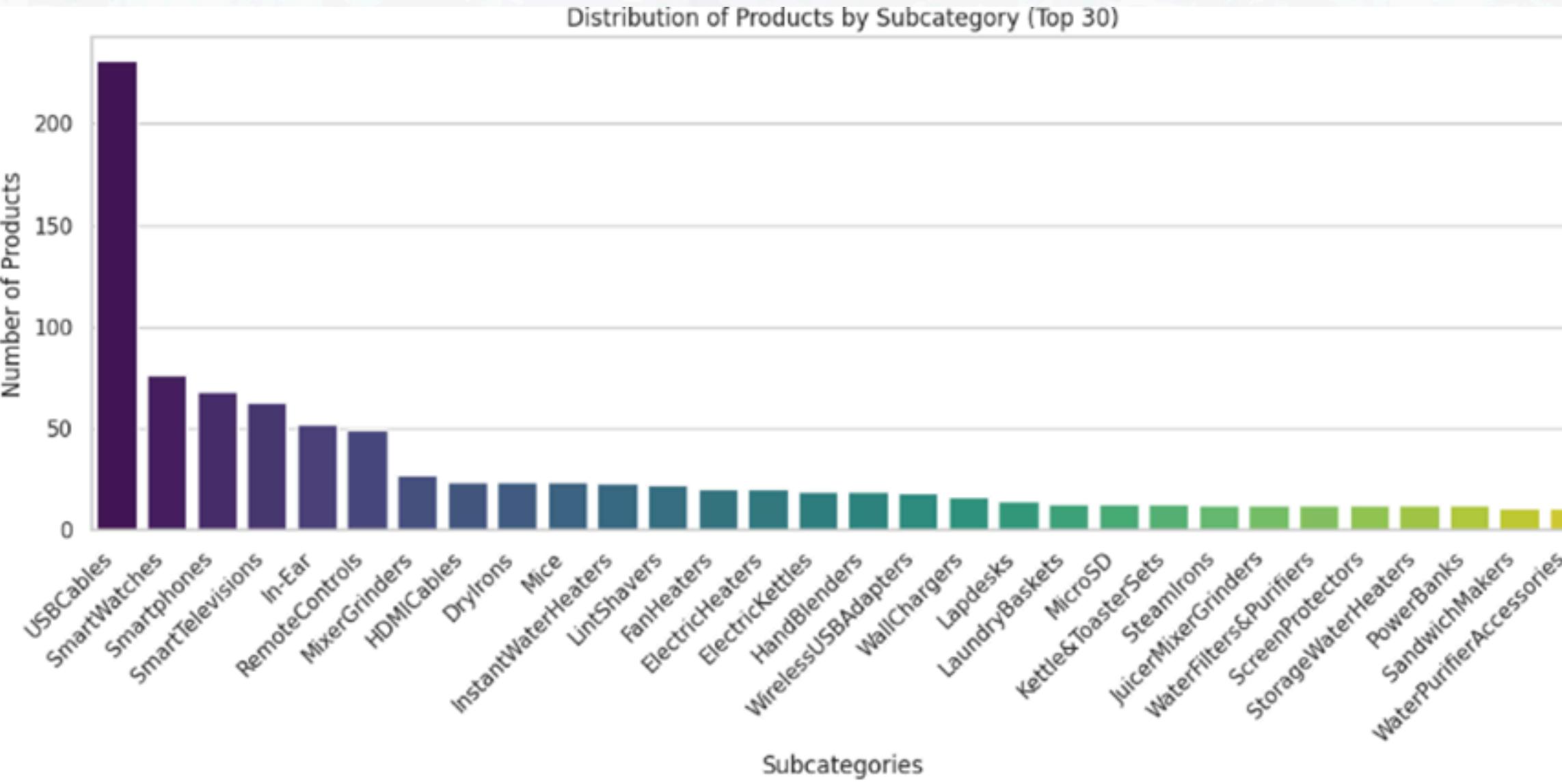
04 Task Implementation

02 Data visualization



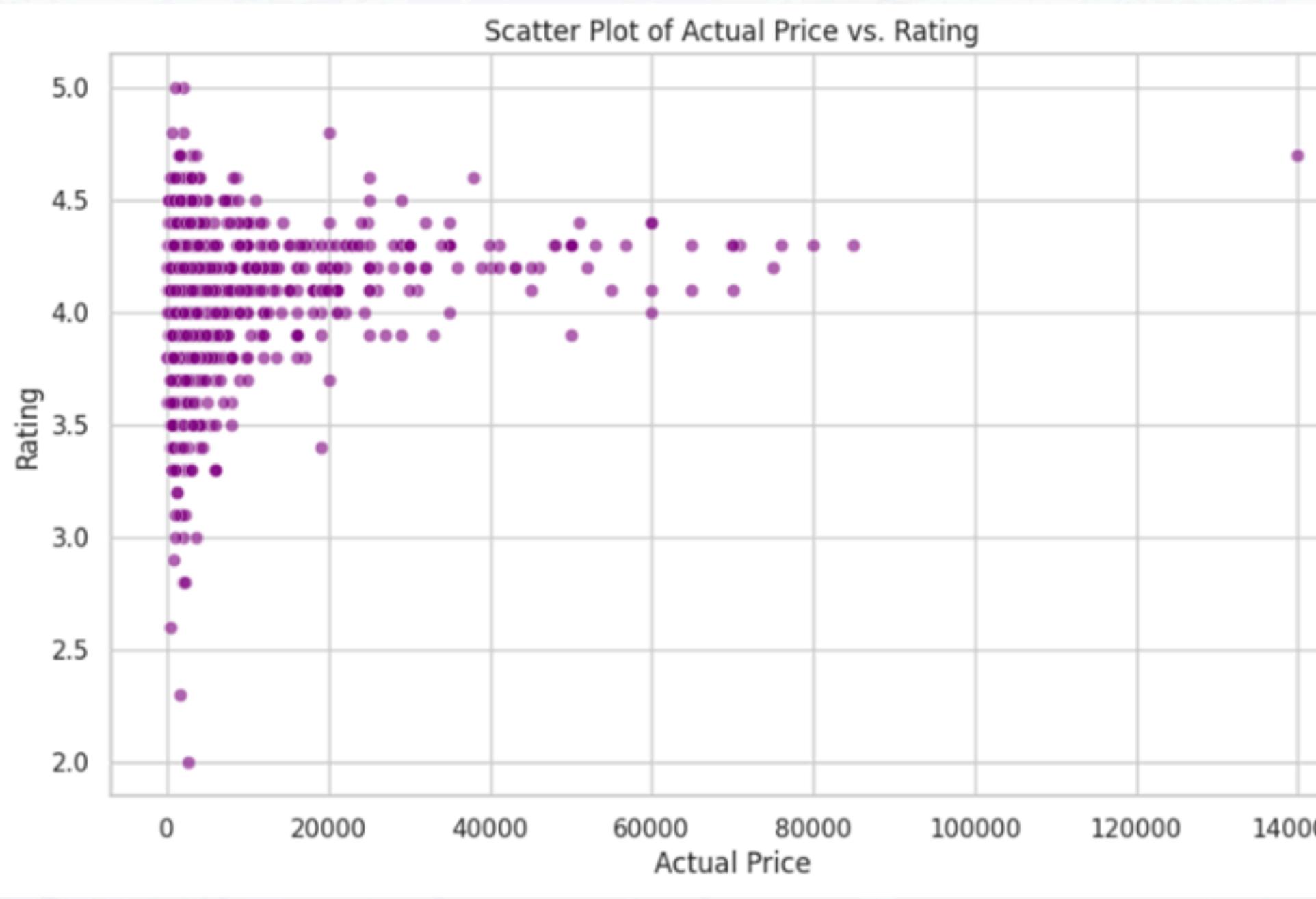
04 Task Implementation

02 Data visualization



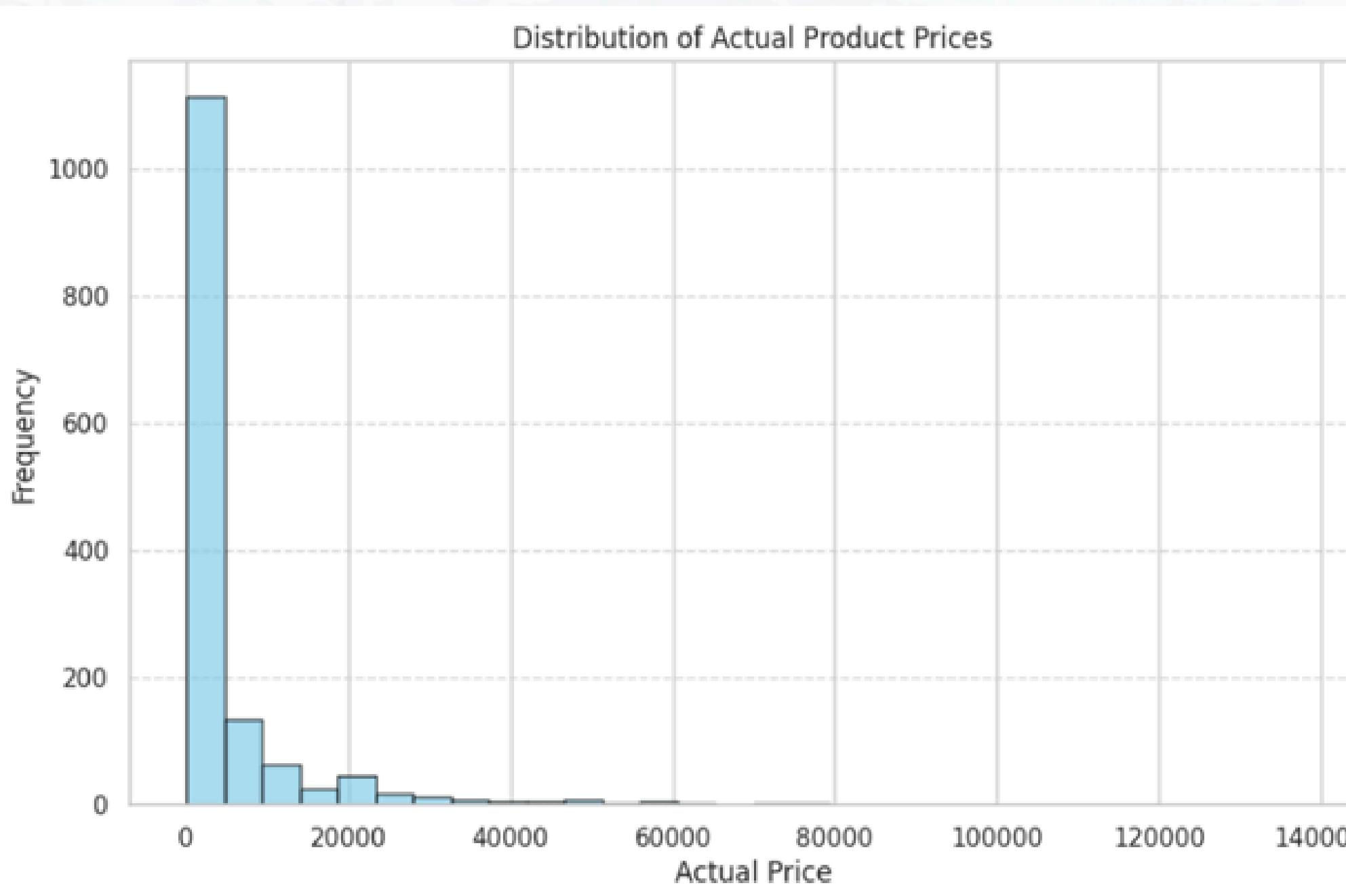
04 Task Implementation

02 Data visualization



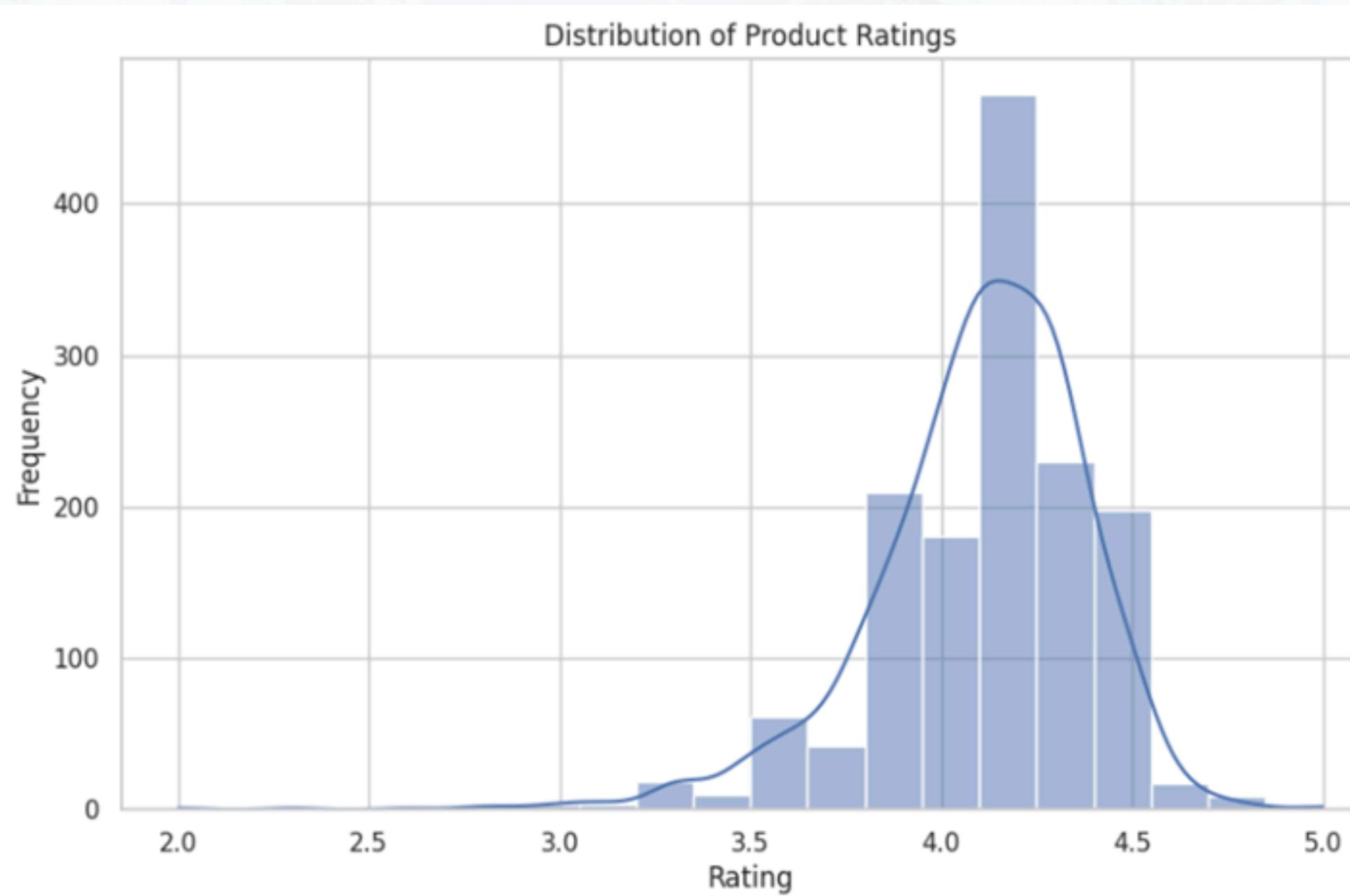
04 Task Implementation

02 Data visualization



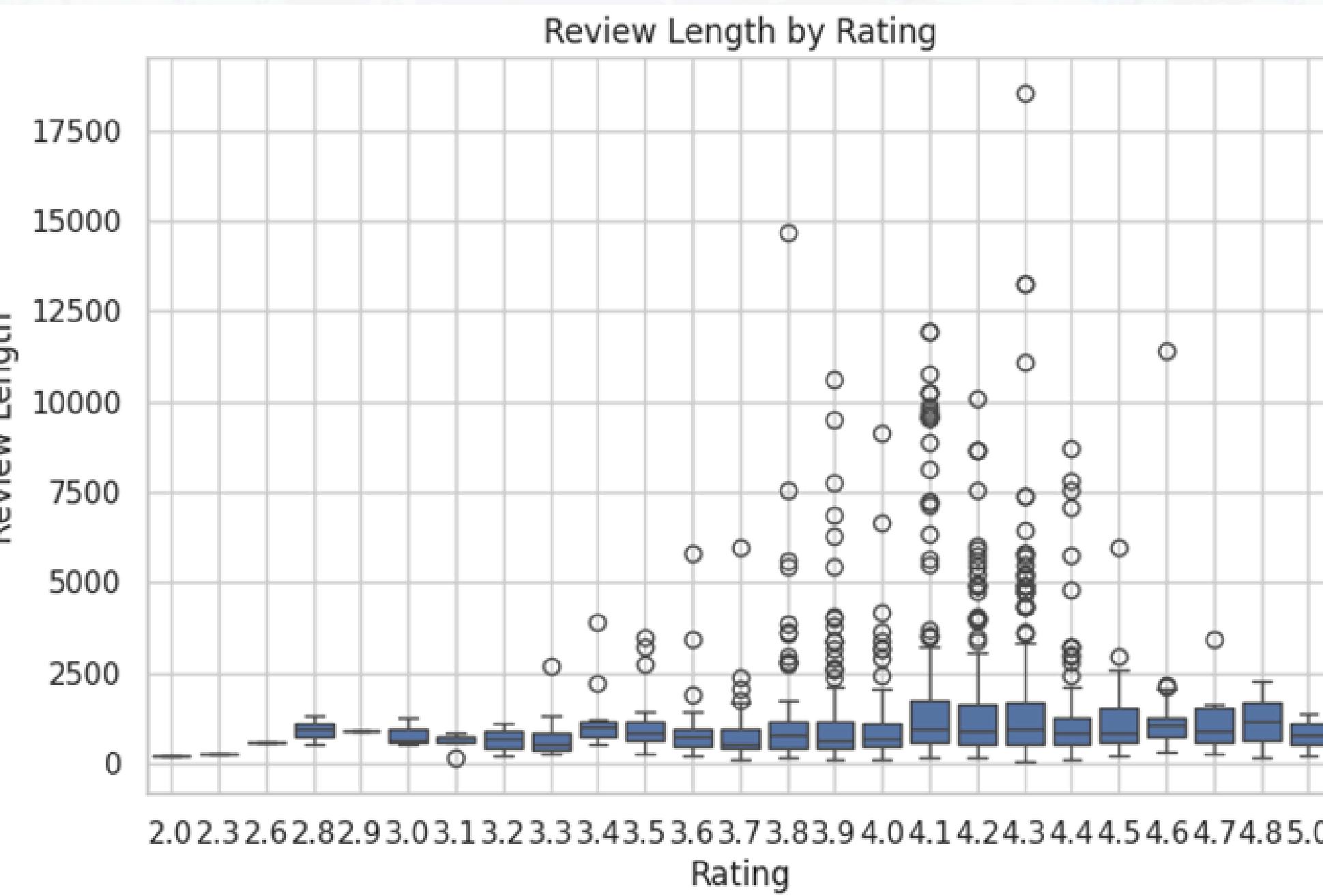
04 Task Implementation

02 Data visualization



04 Task Implementation

02 Data visualization



04 Task Implementation

03 Analysis of reviews

Work Steps:

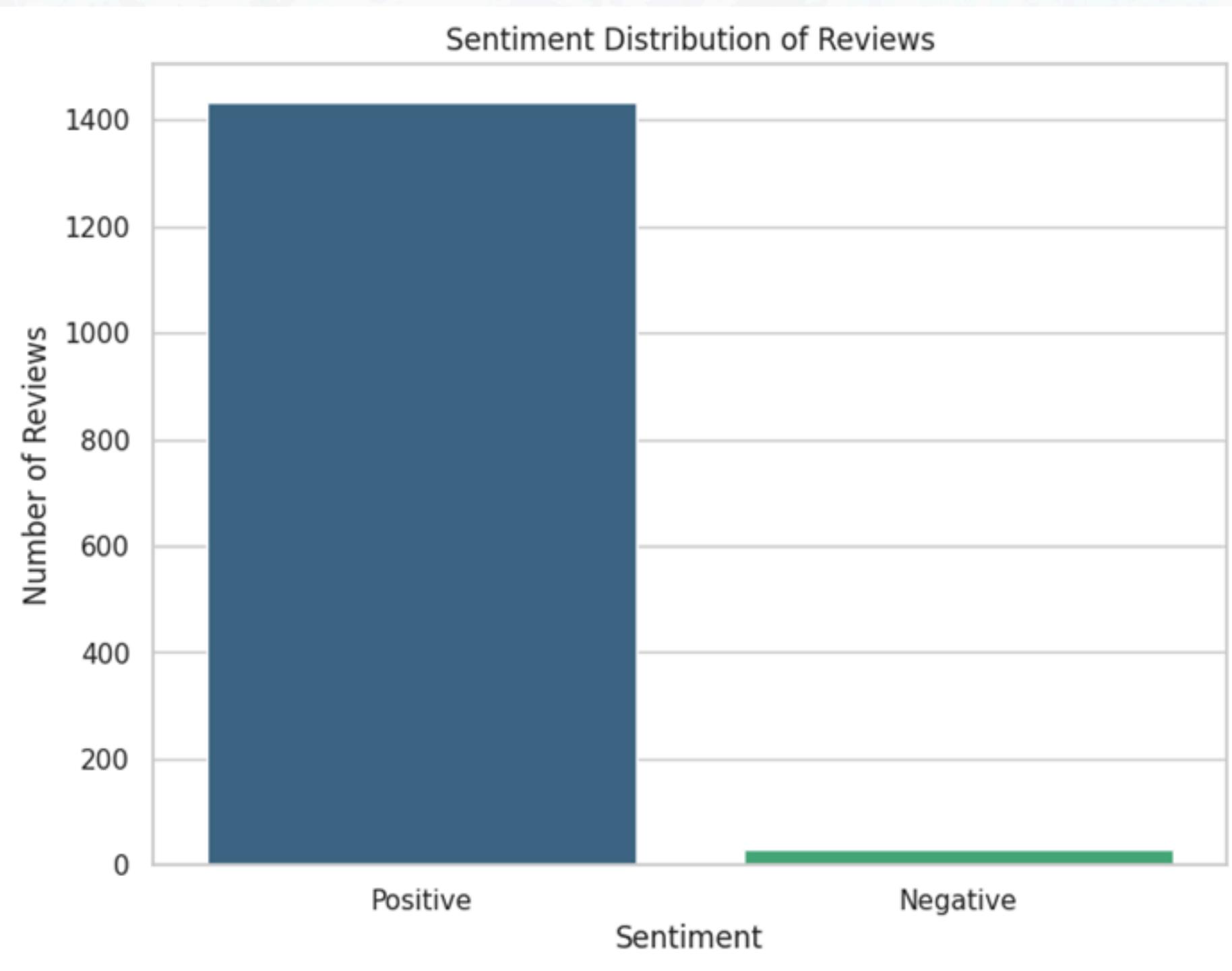
1 Pre -processing Text2 Analysis of
tonality using Vader3 Classification of

Tonality

4 Tone Distribution Visualization

04 Task Implementation

03 Analysis of reviews



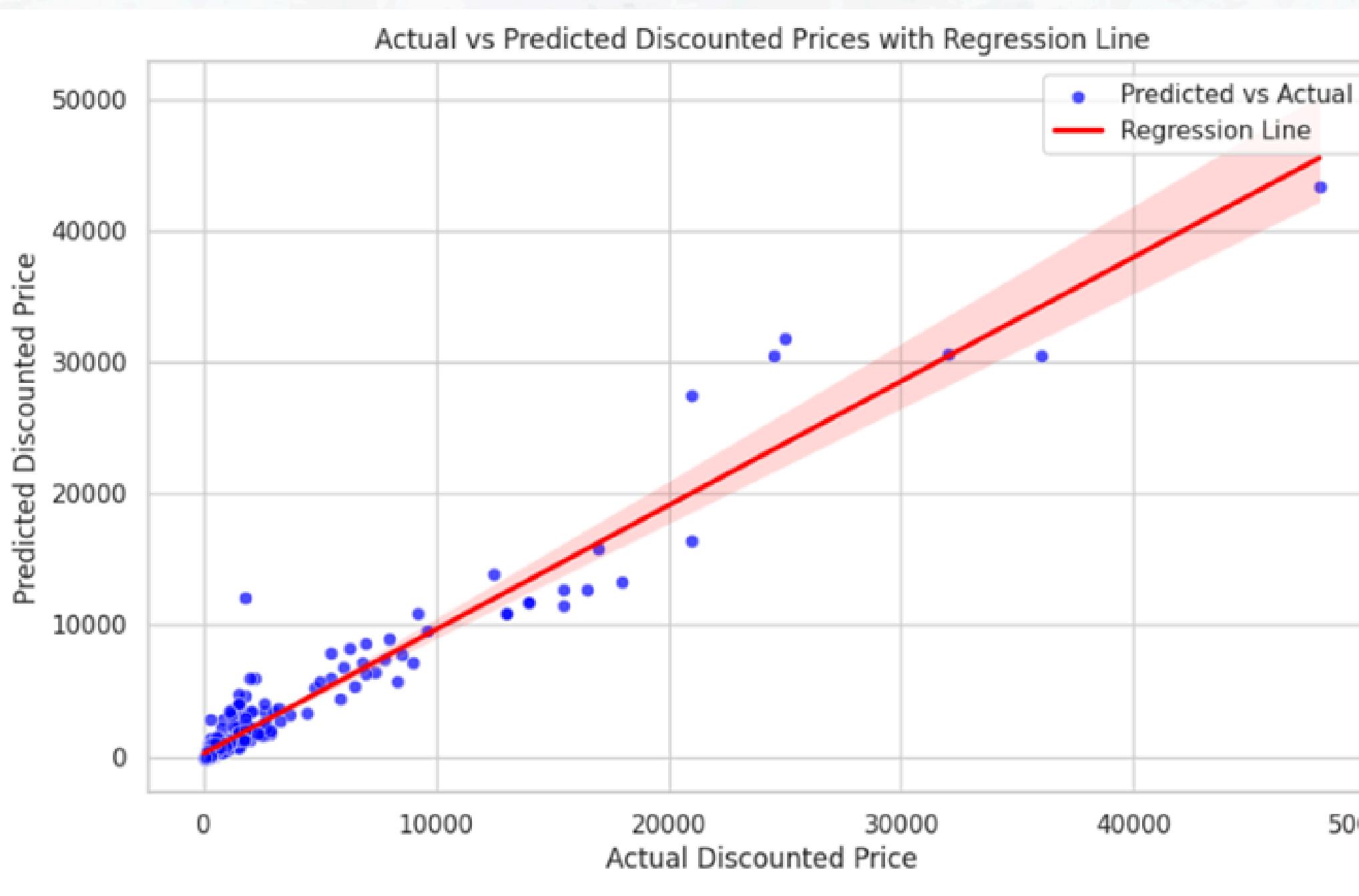
04 Task Implementation

03 Analysis of reviews

| 152 | Samsung 80 cm (32 inches) Wondertainment Serie... | Good, Sound is very low another brand comparing... | -0.5574 | Negative | |
|-----|--|--|---------|----------|--|
| 155 | 7SEVEN® Compatible for Tata Sky Remote Original... | do not buy | -0.7425 | Negative | |

04 Task Implementation

04 Price Reduction Forecast



04 Task Implementation

04 Price Reduction Forecast

Linear Regression

Mean Absolute Error (MAE): 684.12
Mean Squared Error (MSE): 1913777.78
Root Mean Squared Error (RMSE): 1383.39
R-squared (R^2): 0.93

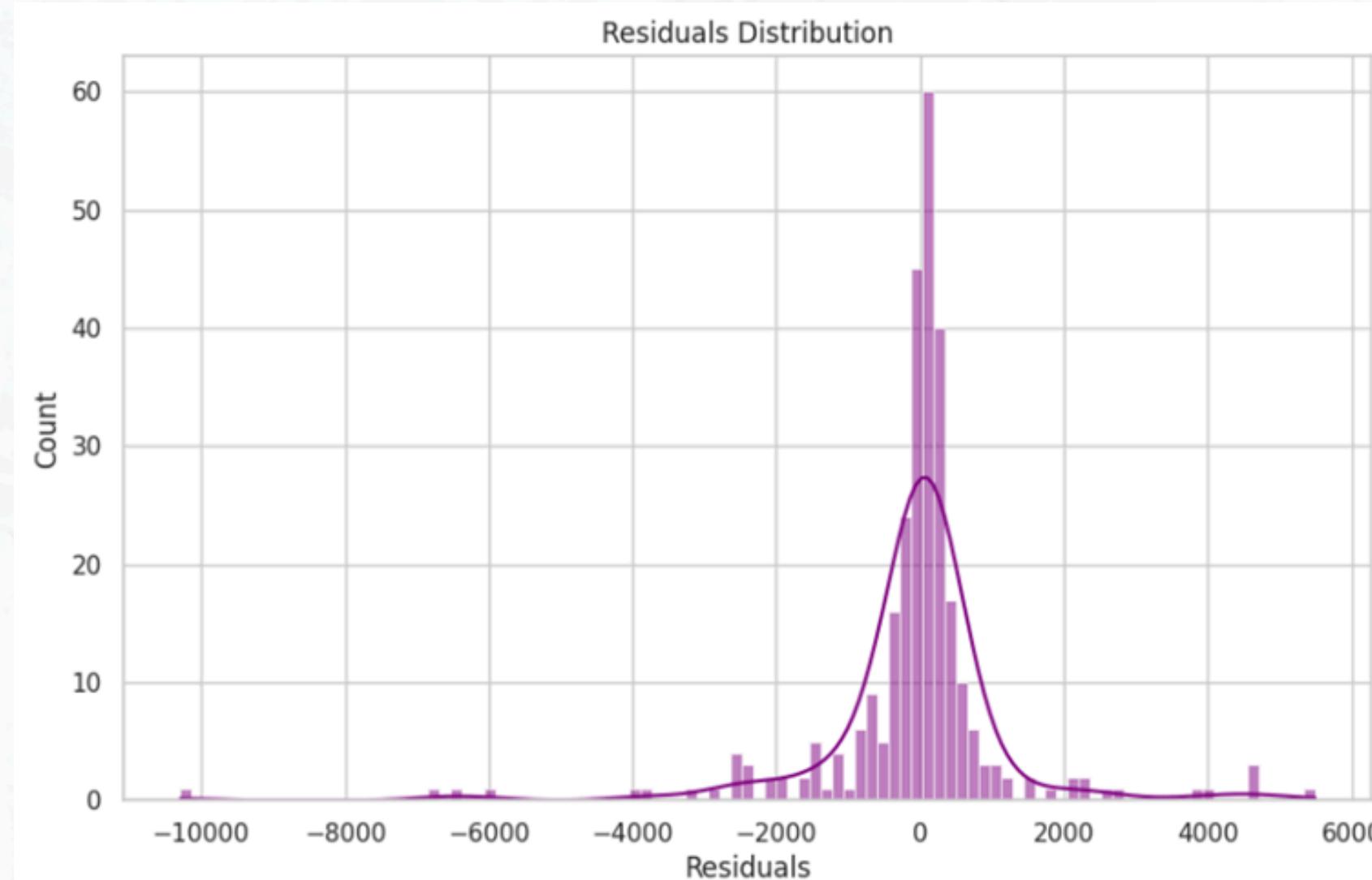
Neural network

Test Loss (MSE): 1870265.50
Test Mean Absolute Error (MAE): 653.66
Root Mean Squared Error (RMSE): 1367.58
 R^2 Score: 0.94

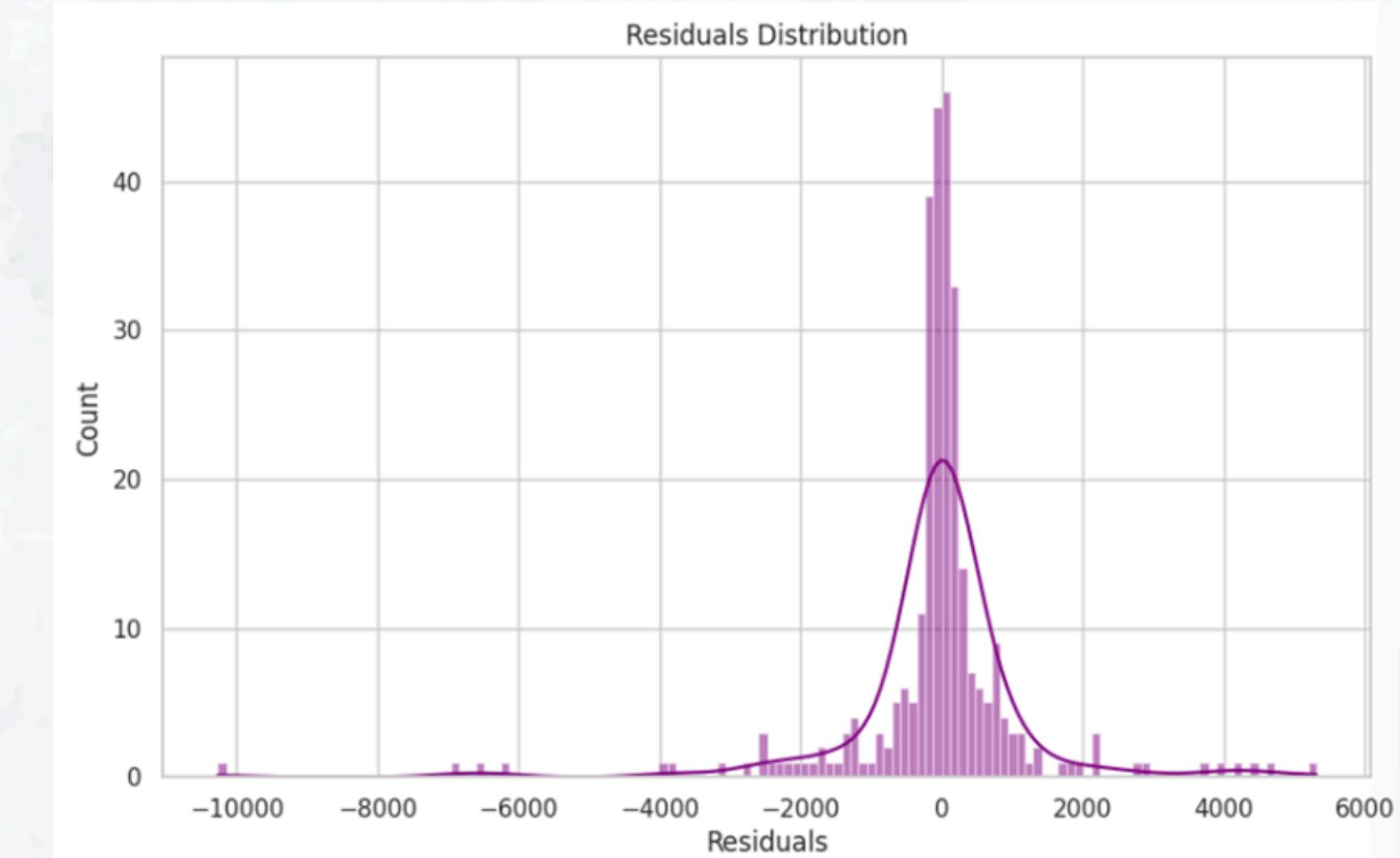
04 Task Implementation

04 Price Reduction Forecast

Linear Regression



Neural network



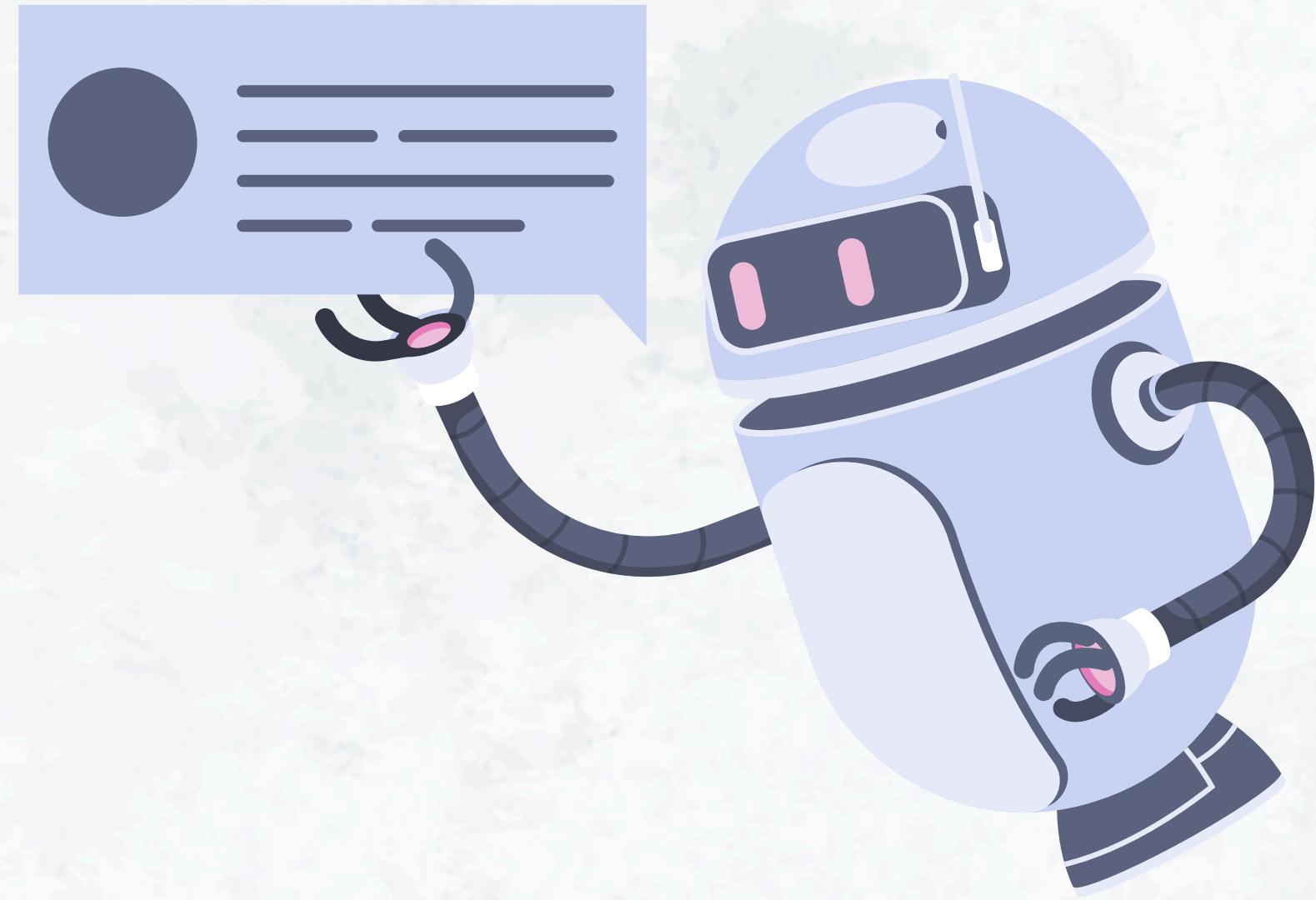
04 Task Implementation

04 Price Reduction Forecast

Conclusion

Both models showed similar results.

Linear regression is the best choice for simple, linear tasks with lower computational requirements, while neural networks are suitable for more complex, nonlinear problems with large datasets.



04 Task Implementation

05 Recommendation Model

a Encoding of tags for user
identifiers
b Frequency Table

| User ID | Frequency |
|---------|-----------|
| 0 | 1045 |
| 1 | 622 |
| 2 | 673 |
| 3 | 253 |
| 4 | 87 |

TF-IDF Matrix

TfidfVectorizer is used to convert product descriptions(column about_product) into numerical vectors. These vectors represent the importance of each word in the context of the entire product description corpus, with common words having less weight.

04 Task Implementation

05 Recommendation Model

Calculation of Cosine Similarity

using `cosine_similarity` from `sklearn.metrics.pairwise`, the code calculates similarity indicators between products. This is done by comparing vectors of TF-IDF products with which the user interacted (bought) with the entire product catalog

e Recommended products

Output - results are stored in a DataFrame

04 Task Implementation

05 Recommendation Model

Call of the function returns recommend_products(df, 253, matrix)

| | user_id_encoded | recommended_product | score |
|---|-----------------|---|-------|
| 0 | 253 | boAt Deuce USB 300 2 in 1 Type-C & Micro USB S... | 1.0 |
| 1 | 253 | boAt Deuce USB 300 2 in 1 Type-C & Micro USB S... | 1.0 |
| 2 | 253 | boAt Rugged v3 Extra Tough Unbreakable Braided... | 1.0 |
| 3 | 253 | boAt Deuce USB 300 2 in 1 Type-C & Micro USB S... | 1.0 |
| 4 | 253 | boAt Deuce USB 300 2 in 1 Type-C & Micro USB S... | 1.0 |

04 Task Implementation

05 Recommendation Model

Checking the purchase history of the same user

| | user_id_encoded | product_name | sub_category |
|-----|-----------------|---|--------------|
| 3 | 253 | boAt Deuce USB 300 2 in 1 Type-C & Micro USB S... | USBCables |
| 11 | 253 | boAt Rugged v3 Extra Tough Unbreakable Braided... | USBCables |
| 92 | 253 | boAt Deuce USB 300 2 in 1 Type-C & Micro USB S... | USBCables |
| 258 | 253 | boAt Rugged V3 Braided Micro USB Cable (Pearl ... | USBCables |
| 392 | 253 | boAt Deuce USB 300 2 in 1 Type-C & Micro USB S... | USBCables |
| 443 | 253 | boAt Rugged v3 Extra Tough Unbreakable Braided... | USBCables |
| 628 | 253 | boAt Deuce USB 300 2 in 1 Type-C & Micro USB S... | USBCables |

Conclusion

The work covers the most important methods in machinery, from pre -processing of text data of revaluation of forecast models and generation personalized product recommendations.

The analysis of moods makes it clear customer feedback, aggression models and neural networks provide value prices for products.

The recommendation system guarantees that user holds individual suggestions based on content. Methodologies involved are the basis of data processing processes for plathorme electronic commerce, where understanding the behavior of users, creating accurate forecasts and providing personalized recommendations are resolute to improve the interaction with clients of business results.



QUESTION?

CREDITS: This presentation template was created by Slidesgo and includes icons by Flaticon, infographics & images by Freepik and content by Eliana Delacour

Please, keep this slide as attribution

