Business Problem: Improving Point Calculation and Review Ranking in E-commerce

Introduction: In the world of e-commerce, two crucial challenges arise: accurately calculating points for aftermarket products and ensuring the correct ordering of product reviews. Overcoming these challenges holds the key to enhancing customer satisfaction on e-commerce platforms, boosting product visibility for sellers, and providing a seamless shopping experience for buyers. The aim is to address these issues, ultimately driving increased sales for e-commerce sites and sellers while facilitating a hassle-free purchasing journey for customers. This project focuses on developing solutions to improve point calculation and review ranking, thereby benefiting all stakeholders in the e-commerce ecosystem.

Point Calculation: Accurately determining the points awarded to aftermarket products is a critical concern. Points play a significant role in customer decision-making, reflecting product quality, reliability, and value. By ensuring precise point calculation, e-commerce sites can enhance customer satisfaction. This, in turn, leads to improved trust in the platform, increased repeat purchases, and positive word-of-mouth recommendations. Sellers also benefit from higher product prominence, increased visibility, and greater sales opportunities. Therefore, solving this problem holds substantial benefits for all stakeholders involved.

Review Ranking: Another challenge in the e-commerce landscape is the presence of misleading or deceptive comments among product reviews. Such comments can lead to financial losses and the loss of customers' trust. To mitigate this issue, the Amazon company, for example, aims to rank reviews based on their sentiment (positive/negative). By implementing a robust review ranking system, e-commerce platforms and sellers can ensure that genuine and helpful reviews receive higher visibility. This empowers buyers to make well-informed purchasing decisions, driving customer satisfaction and fostering a positive shopping experience.

Benefits and Conclusion: By addressing these two fundamental problems, e-commerce platforms and sellers stand to reap numerous benefits. Accurate point calculation enhances customer satisfaction and loyalty, leading to increased sales and positive business growth. Transparent review ranking based on sentiment analysis allows buyers to navigate through reviews more effectively, making informed decisions while minimizing the impact of misleading comments. Ultimately, these solutions contribute to a seamless and hassle-free purchasing journey for customers, positively impacting the overall e-commerce ecosystem.

The project focuses on analyzing and visualizing data from customer reviews using Python. It involves extracting insights from a dataset containing customer reviews and performing various data analysis and visualization techniques to gain valuable information. The code snippets provided demonstrate different aspects of the project, including sentiment analysis, missing values analysis, data summary, and categorical variable visualization.

Description: The project utilizes popular Python libraries such as Pandas, TextBlob, VADER Sentiment Analyzer, and Plotly to process, analyze, and visualize the customer review data. Here's an overview of the key functionalities:

Sentiment Analysis: The code includes sentiment analysis techniques using TextBlob and VADER Sentiment Analyzer. TextBlob is used to calculate polarity and subjectivity scores for each review text, while VADER Sentiment Analyzer measures the sentiment polarity using a predefined lexicon.

Missing Values Analysis: A function is implemented to identify columns with missing values in the dataset. It calculates the total count and ratio of missing values for each column, providing insights into data completeness and potential data cleaning requirements.

Data Summary: Another function provides an overview of the dataset, including its shape (number of rows and columns) and the data types of each column. It also incorporates the analysis of the missing values to give a comprehensive summary of the dataset.

Quantiles of Numerical Data: The code calculates the quantiles (0%, 0.05%, 0.50%, 0.95%, 0.99%, and 100%) for the numerical data in the dataset. This information helps in understanding the distribution and range of the numerical variables.

Categorical Variable Visualization: The project includes a function that visualizes categorical variables using two types of plots: a countplot and a pie chart. The countplot displays the count of each category in a bar chart, while the pie chart represents the percentage distribution of each category.

Conclusion: In conclusion, this project provides a comprehensive analysis and visualization of customer reviews. By leveraging various Python libraries, it offers insights into sentiment analysis, missing values, data summary, numerical data distribution, and categorical variable visualization. These analyses contribute to a better understanding of customer feedback, enabling businesses to make informed decisions and improve their products or services based on customer insights.