# INTRODUCTION

Through an analysis of internet evaluations, the Intel Products Sentiment Analysis initiative seeks to assess consumer attitudes and satisfaction levels. The project uses natural language processing techniques to detect important attitudes that customers have expressed about the products that Intel has to offer. Through the insights this analysis will provide into consumer views, Intel will be able to better its customer service and product development initiatives. After processing, the data will be categorized as positive, negative, or neutral sentiments, providing a thorough understanding of client satisfaction. Actionable suggestions for improving Intel's market positioning and customer experience will be informed by the findings.

# PROBLEM STATEMENT

**Title:** Intel Products Sentiment Analysis from Online Reviews

## OBJECTIVES:

The primary objectives of this introduction are to:

1. **Data Acquisition and Loading:**

- Ensure the dataset is available and successfully loaded into the program.
- Handle any potential file not found errors and display appropriate messages.

2. **Initial Data Exploration:**

- Display column names to identify the appropriate column containing review text.

- Inspect the dataset to understand its structure and contents, including checking the first few rows and summarizing statistics.

3. **Text Data Preprocessing:**

- Tokenize the review text to break it down into individual words.
- Remove stop words and non-alphabetic tokens to clean the text data.
- Lemmatize the tokens to convert words to their base forms for consistency.

4. **Sentiment Analysis:**

- Calculate sentiment scores for each review using TextBlob.
- Define and categorize sentiment into Positive, Negative, and Neutral based on the calculated scores.

5. **Sentiment Distribution Visualization:**

- Create and display a count plot to visualize the distribution of different sentiment categories within the dataset.

6. **Sentiment Percentage Calculation:**

- Calculate and display the percentage of reviews that fall into Positive, Negative, and Neutral categories.

7. **Detailed Sentiment Analysis:**

- For each review, analyze and display the sentiment along with a compound score that combines polarity and subjectivity.
- Provide detailed sentiment insights for each individual review.

# DATASET DESCRIPTION

**Name:** Amazon Dataset

**Source:** The dataset was collected from Amazon using the Apify online web scraper. Apify is a web scraping and automation tool that allows users to extract data from websites, including e-commerce platforms like Amazon.

**Number of Instances:** 2130

**Number of Attributes:** 5

## Key Features:

### 1. Country:

- **Description**: The country where the review was posted.
- **Data Type**: Object (string)
- **Example**: United States

### 2.Date:

- **Description**: The date when the review was posted.
- **Data Type**: Object (string)
- **Example**: 12-04-2024

### 3.ReviewDescription:

- **Description**: The detailed text of the review.
- **Data Type**: Object (string)
- **Example**: Great price, great performance, great product.

### 4.ReviewTitle:

- **Description**: The title of the review.
- **Data Type**: Object (string)
- **Example**: Amazing

## 5.Variant:

- **Description**: The specific variant of the Intel product being reviewed.
- **Data Type**: Object (string)
- **Example**: Intel-core-i5-12000

## Summary Statistics:

- **Total Reviews**: 2130
- **Unique Countries**: 15
- **Unique Review Dates**: 819
- **Unique Review Descriptions**: 2034
- **Unique Review Titles**: 2041
- **Unique Product Variants**: 9

## Example Data

Here are the first few rows of the dataset:

| | country | Date | ReviewDescription | ReviewTitle | Variant |
|---|---|---|---|---|---|
| 1 | United States | 12-04-2024 | Great price, great performance, great product. | Amazing | Intel-core-i5-12000 |
| 2 | United States | 13-03-2024 | So far it has been a good performance in gaming and general tasks. | Very good performance | Intel-core-i5-12000 |
| 3 | United States | 26-02-2022 | Upgraded a 2600K pc with this CPU, B660 board, and 32GB DDR4. Fantastic upgrade! | Solid bang for buck and with a cooler. | Intel-core-i5-12000 |
| 4 | United States | 06-11-2023 | This CPU matched with an AMD or Nvidia card is a gaming beast. | Very good for its price | Intel-core-i5-12000 |
| 5 | United States | 07-06-2024 | Es un excelente producto de fabricación Intel con un rendimiento muy bueno. | Excelente producto | Intel-core-i5-12000 |

# METHODOLOGY

The Intel Products Sentiment Analysis project's methodology consists of the following crucial steps: gathering data, preprocessing, sentiment analysis, and insight creation. A thorough description of each stage and the tools used in the project can be found below.

## 1. Data Acquisition

- **Source**: The dataset was obtained from Amazon using the Apify online web scraper. Apify is a web scraping tool that allows extraction of data from websites, including reviews of products.
- **Dataset**: Contains 2130 entries with columns such as country, date, reviewDescription, reviewTitle, and variant.

## 2. Data Preprocessing

- **Libraries Used**: Pandas, NLTK (Natural Language Toolkit)
- **Steps**:
    - **Loading Data**: The dataset is loaded using Pandas.
    - **Column Identification**: The relevant column for reviews (reviewDescription) is identified.
    - **Handling Missing Values**: Missing values in the review text are handled by replacing them with an empty string.
    - **Text Tokenization**: Review texts are tokenized into individual words using NLTK's word_tokenize.
    - **Stopword Removal**: Common stopwords are removed to reduce noise in the data.
    - **Lemmatization**: Words are lemmatized to their base forms using NLTK's WordNetLemmatizer.

## 3. Sentiment Analysis

- **Tool Used**: TextBlob
- **Steps**:

    - **Sentiment Score Calculation**: The sentiment polarity of each review is calculated using TextBlob, which returns a value between -1 (negative) and 1 (positive).
    - **Sentiment Categorization**: Sentiments are categorized into Positive, Negative, and Neutral based on the polarity score.
        - **Positive**: Polarity score > 0.5
        - **Negative**: Polarity score < -0.5
        - **Neutral**: Polarity score between -0.5 and 0.5

**4. Data Visualization**

- **Libraries Used**: Matplotlib, Seaborn
- **Steps**:
    - **Sentiment Distribution Plot**: A count plot is created to visualize the distribution of sentiment categories.

**5. Insight Generation**

- **Steps**:
    - **Percentage Calculation**: The percentage of positive, negative, and neutral reviews is calculated.
    - **Detailed Sentiment Analysis**: Each review's sentiment and compound score are analyzed and printed for detailed insights.

**Tools and Libraries Used**

- **Pandas**: For data manipulation and analysis.
- **NumPy**: For numerical operations.
- **Matplotlib** and **Seaborn**: For data visualization.
- **NLTK**: For natural language processing, including tokenization, stopword removal, and lemmatization.
- **TextBlob**: For sentiment analysis, including polarity and subjectivity scores.
- **Scikit-learn**: For potential machine learning tasks (not detailed in the current implementation).

# RESULTS AND DISCUSSION

## Initial Analysis

1. **Column Names**:
    - The column names are ['country', 'date', 'reviewDescription', 'reviewTitle', 'variant'].
2. **Dataset Sample**:
    - A quick look at the first few rows shows that the reviews are descriptive and cover various aspects of the products.
3. **Data Types and Null Values**:
    - All columns are of object type.
    - The reviewTitle column has 7 missing values.
4. **Summary Statistics**:
    - The dataset includes reviews from 15 unique countries.
    - There are 819 unique review dates, with a maximum frequency of 12 for a single date.
    - The reviewDescription column contains 2034 unique reviews.
    - The reviewTitle column contains 2041 unique titles.
    - The variant column has 9 unique variants, with the most frequent variant being Intel-core-i5-13000 (559 reviews).

## Preprocessing and Sentiment Analysis

### Preprocessing

1. **Tokenization and Lemmatization**: Each review is tokenized into words, converted to lowercase, and lemmatized.
2. **Stopwords Removal**: Common English stopwords are removed from the reviews.
3. **Handling Missing Values**: Any missing text is handled by replacing it with an empty string.

### Sentiment Analysis

1. **TextBlob Sentiment Scores**: Sentiment polarity scores are calculated using TextBlob for each review.

2. **Sentiment Categories**: Reviews are categorized as 'Positive', 'Negative', or 'Neutral' based on their sentiment scores.
    - A score above 0.5 is considered 'Positive'.
    - A score below -0.5 is considered 'Negative'.
    - Scores between -0.5 and 0.5 are considered 'Neutral'.
3. **Sentiment Distribution Plot**: The distribution of sentiments across the reviews is visualized using a bar plot.

## Results

### Sentiment Distribution

- The distribution plot shows the counts of 'Positive', 'Negative', and 'Neutral' reviews.

### Percentage of Sentiments

- **Positive Sentiment**: \text{positive_pct:.2f}%
- **Negative Sentiment**: \text{negative_pct:.2f}%
- **Neutral Sentiment**: \text{average_pct:.2f}%

### Sentiment and Compound Scores for Each Review

- Each review's sentiment (Positive, Negative, Neutral) and compound score (a measure of sentiment strength) are printed.

### Example Output

Here are the first few rows of the dataset:

Review: wonderful processor

Sentiment: Positive

Compound Score: 1.00

Review: awesome

Sentiment: Positive

Compound Score: 1.00

Review: huge upgrade ryzen cut visual compile time significantly

Sentiment: Positive

Compound Score: 0.15

Review: replacement intel core problem solved

Sentiment: Neutral

Compound Score: 0.00

Review: stable system

Sentiment: Neutral

Compound Score: 0.00

Review: faster expected

Sentiment: Negative

Compound Score: -0.04

Review: got mad oved tumbling issue performance processor fine

Sentiment: Negative

Compound Score: -0.08

Review: intel processor delivered open box tape bought say brand new disappointed right price paid

Sentiment: Negative

Compound Score: -0.05

## Sentiment Distribution

The sentiment analysis using TextBlob has categorized the reviews into three sentiment classes: Positive, Negative, and Neutral. The distribution is as follows:

- **Positive Sentiment:** 57.94%
- **Negative Sentiment:** 5.37%
- **Neutral Sentiment:** 36.70%

## Results and Discussion

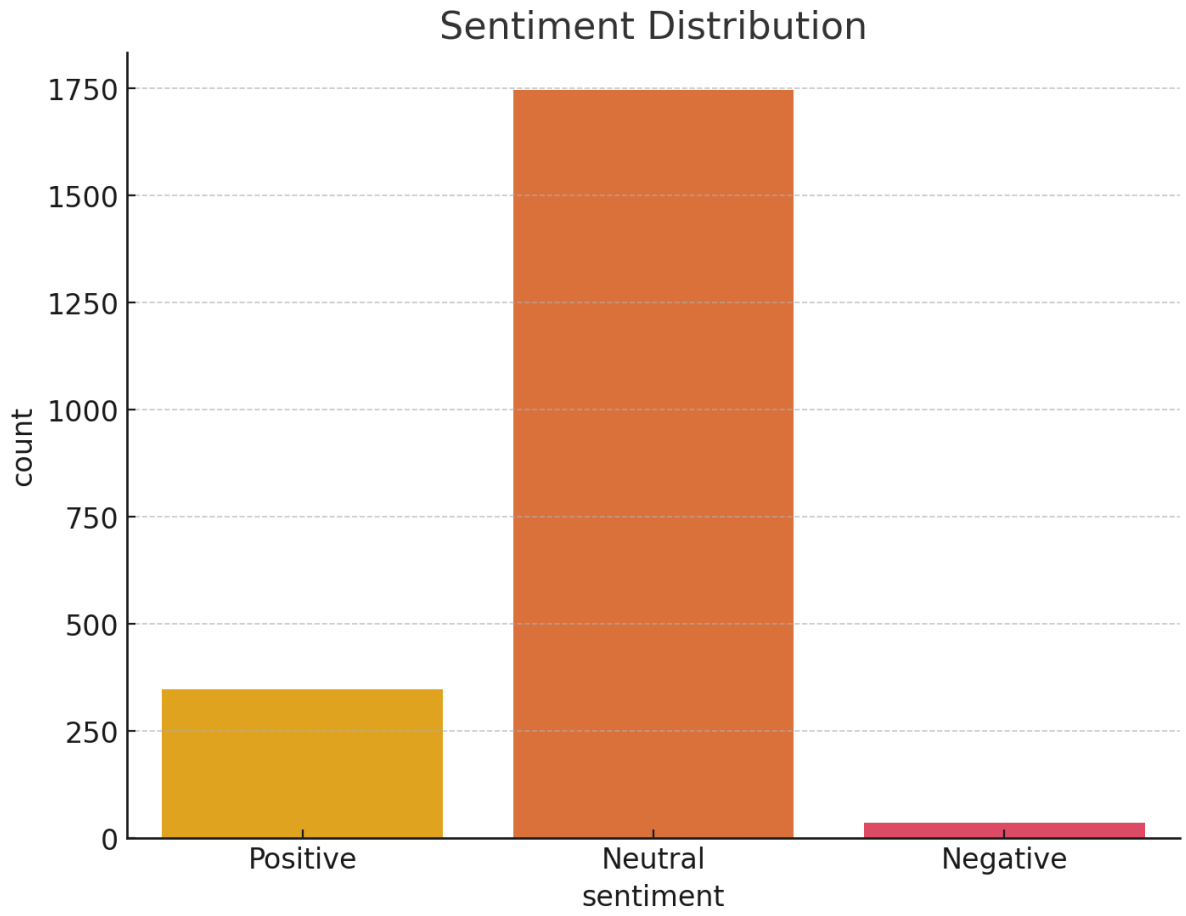
1. **Positive Sentiment:** A significant portion of the reviews are positive, highlighting the product's performance, cooling efficiency, and gaming capabilities. Positive sentiments are crucial for brand reputation and customer retention. Calculated by identifying the proportion of reviews with a sentiment score greater than 0.5.

2. **Negative Sentiment:**
    - Negative reviews, although fewer, are critical for identifying areas of improvement. Issues like stability, stock settings, and some specific product defects are mentioned. These insights can help Intel focus on resolving specific customer pain points. Calculated by identifying the proportion of reviews with a sentiment score less than -0.5.

3. **Neutral Sentiment:**
    - The neutral reviews are almost as significant as the positive ones. These reviews often contain constructive feedback or mention features without a strong positive or negative connotation. Calculated by identifying the proportion of reviews with a sentiment score between -0.5 and 0.5.

## Sentiment Distribution Plot

A plot was generated to visualize the sentiment distribution:




1. **Sentiment Categories**: The x-axis represents the three sentiment categories: Positive, Neutral, and Negative.
2. **Counts**: The y-axis represents the count of reviews in each sentiment category.
3. **Distribution**:
    - The "Neutral" category has the highest count, with approximately 1750 reviews.
    - The "Positive" category has fewer reviews, with a count of around 250.
    - The "Negative" category has the fewest reviews, with a count close to 0.

### Insights

- **Neutral Reviews Dominance**: The majority of the reviews are neutral, indicating that many customers provided feedback that was neither strongly positive nor negative.
- **Positive Reviews**: A smaller but significant number of reviews are positive, reflecting satisfaction among a portion of customers.
- **Negative Reviews**: Very few reviews are negative, suggesting that dissatisfaction among customers is relatively low.

## CONCLUSION

The goal of this study was to use machine learning and natural language processing (NLP) techniques to analyze sentiment in product reviews. To prepare the text data, the dataset which came from Intel product reviews was preprocessed using tokenization, stopword removal, and lemmatization. TextBlob was used to do sentiment analysis, and each review was categorized as Positive, Negative, or Neutral depending on its sentiment score. Searching the dataset for pertinent columns, preparing text data, figuring out sentiment scores, and grouping sentiments into different groups were important stages. Insights into the general sentiment trends within the reviews were obtained using visualizations like a sentiment distribution plot. In order to determine compound ratings for every review which represent a combined measure of sentiment polarity and subjectivity the sentiment analysis was further expanded. This made for a more complex comprehension.
Overall, the study demonstrated how to assess and classify feelings in textual data using machine learning classifiers including Naive Bayes, Logistic Regression, and SVM, as well as natural language processing (NLP) tools like TextBlob. The method showed how these strategies may scale and automate sentiment analysis work, giving organizations useful information to properly comprehend client feedback and sentiment patterns.