Customer Sentiment Analysis

Objective:

As a Data Analyst at Flipkart, you have been tasked with gauging customer sentiment towards the iPhone 15 128GB model. The primary goal of this project is to analyze public perception and evaluate customer reactions by performing sentiment analysis on product reviews posted by users. By extracting and processing customer reviews, you will derive insights about the overall sentiment (positive or negative) surrounding the product, which can be useful for decision-making, improving customer experience, and identifying key areas for product improvement.

Tasks:

1. Data Collection (Web Scraping):

Tool: Selenium and BeautifulSoup

Task: Scrape at least 300 customer reviews from Flipkart's product page for the iPhone 15 128GB model. Each review should include**:**

Username: The name of the reviewer.

Rating: The rating provided by the user (1 to 5 stars).

Review Text: The content of the customer's review, which may contain valuable information regarding their experience with the product.

Steps:

- Set up Selenium to automate browser interactions, navigate to Flipkart's product page for iPhone 15 128GB, and extract the reviews.
- Use BeautifulSoup to parse the HTML of the reviews and extract the relevant details (username, rating, and review text).
- Ensure that the scraper handles pagination to retrieve reviews from multiple pages if necessary.

2. Data Cleaning and Preprocessing:

Tool: Pandas

Task: Clean and preprocess the scraped data for analysis.

Steps:

- **Remove duplicates:** Eliminate any duplicate reviews to ensure data quality.
- Handle missing values: Address missing or incomplete data, such as missing review text or rating, by either removing rows or filling in missing values if applicable.

Text preprocessing:

- **Convert the review text to lowercase.**
- **Remove irrelevant characters (e.g., special characters, punctuation, and extra spaces).**
- **Tokenize the text into individual words.**
- **Remove stop words (commonly used words that do not add significant meaning to sentiment analysis).**
- **Perform lemmatization to convert words into their base form (e.g., "running" → "run").**

3. Sentiment Analysis:

Tool: TextBlob

Task: Analyze the sentiment of each review to classify them as either positive or negative.

Steps:

- **Use TextBlob to perform sentiment analysis on the review text.**

TextBlob will provide a polarity score between -1 (negative) and +1 (positive), as well as a subjectivity score.

- **Define a threshold to classify the sentiment:**

Positive sentiment: Polarity score ≥ 0.1

Negative sentiment: Polarity score < 0.1

- **Store the sentiment classification for each review in the dataset.**

4. Data Analysis and Insights:

Tool: Pandas and Matplotlib/Seaborn for visualization

Task: Perform an analysis on the sentiment of reviews and extract actionable insights.

Steps:

- **Sentiment Distribution:** Calculate the overall distribution of positive and negative sentiments for the 300 reviews.
- **Average Rating vs Sentiment:** Analyze if there is any correlation between the numeric ratings (1-5 stars) and sentiment polarity. Do higher ratings correspond with more positive sentiments?
- **Word Cloud:** Create a word cloud to identify the most frequently mentioned words in the positive and negative reviews.
- **Review Length Analysis:** Investigate if longer reviews are associated with more detailed sentiments, either positive or negative.

5. Reporting:

Summarize the findings in a clear, concise report with the following sections:

- **Overview** of the data collection and cleaning process.
- **Sentiment Analysis Results:** Distribution of positive/negative reviews, average sentiment per rating, etc.
- **Insights:** Key trends from the sentiment analysis, such as common issues with the product or positive highlights.
- **Recommendations:** Based on customer sentiment, suggest improvements for the iPhone 15 128GB model or potential areas Flipkart can focus on for marketing.

Libraries and Tools:

- **Selenium:** For automating the web scraping process.
- **BeautifulSoup:** For parsing HTML and extracting review details.
- **Pandas:** For data cleaning, processing, and analysis.
- **TextBlob:** For performing sentiment analysis on the review text.
- **Matplotlib/Seaborn:** For visualizations like sentiment distribution and word clouds.