Amazon Product Review Extractor & Sentiment Analysis Tool (Python + Streamlit + BeautifulSoup + VADERSentiment)

Project Links

- Live WebApp (Deployed on Cloud): https://amazon-review-extractor-and-sentiment-analyser.streamlit.app/
- GitHub Repository: https://github.com/Ashish-Bonde/Amazon-Product-Review-Sentiment-Analysis-webscrap-extractor-/tree/main

Overview

The Amazon Product Review Extractor and Sentiment Analysis Tool is a web-based application designed to help users extract, analyze, and visualize product reviews from Amazon. This tool allows businesses, marketers, and researchers to understand customer feedback by extracting reviews and performing sentiment analysis using the VADER model.

This project provides the following core functionalities:

- 1. Product Review Extraction
 - Extract customer names, review dates, star ratings, short and long reviews.
- 2. Data Aggregation
 - Store extracted data in a structured format (Pandas DataFrame).
- 3. Sentiment Analysis
 - Analyze short or long reviews for sentiment: Positive, Neutral, or Negative.
- 4. Visualization
 - Visualize sentiment results using interactive charts (Pie Chart, Histogram, Scatter Plot).

Navigation Menu

The app has a sidebar with the following options:

- 1. A Home Introduction and how it works.
- 2. Parazon Product Review Extractor Extract reviews from an Amazon product link.
- 3. II Review Analysis Perform sentiment analysis on extracted or uploaded data.
- 4. Result Visualization View sentiment results in charts (Pie Chart, Histogram, Scatter Plot).

X How to Use Each Section

1. Amazon Product Review Extractor

☑ Steps:

- 1. Go to the sidebar and select " Amazon Product Review Extractor".
- 2. Paste the Amazon product URL into the text box.
- 3. Click the "Extract" button.
 - Reviews will be fetched and displayed below.
 - You can also extract multiple pages by entering next page URLs.
- 4. Click "Download as CSV" to save the extracted data for later use.

Note: Amazon frequently changes its website structure. If extraction fails, try again or check if the product page uses JavaScript rendering (not supported by this tool).

2. II Review Analysis

You can analyze sentiment from:

- Extracted Amazon reviews
- Uploaded CSV file
- Google Sheets via public link

✓ Steps:

1. Select "II Review Analysis" from the sidebar.

- 2. Choose input method:
 - Extracted Amazon Reviews
 - Upload CSV File
 - Google Sheet Link
- 3. Upload/select your file or paste the link.
- 4. Choose the column containing the review text.
- 5. Click "Analyze".
 - The system will classify each review as Positive , Negative , or Neutral using the VADER model .
- 6. Download the analyzed results as CSV.

3. Result Visualization

After performing sentiment analysis, you can visualize the results.

☑ Steps:

- 1. Select "✓ Result Visualization" from the sidebar.
- 2. Choose visualization type:
 - Pie Chart: Overall sentiment distribution.
 - Histogram : Sentiment across selected columns.
 - Scatter Plot : Sentiment vs other features.

Advantages of This Tool

Feature	Description
⊕Cloud-Based	No installation needed; just open link in any browser.
File Upload Support	Analyze sentiment from local CSV files.
Google Sheets Integration	Paste a shareable Google Sheet link for live analysis.
NLP-Powered Sentiment Analysis	Uses VADER for quick and accurate sentiment classification.
Interactive Visualizations	Understand customer opinions visually with Plotly charts.
⋑ Data Export	Save both raw and analyzed data as CSV for further use.

! Limitations

- Amazon Website Changes: Amazon frequently changes its HTML structure, which may affect scraping accuracy.
- No Login Required Detection: Some pages require login and are not accessible through this tool.
- Limited Sentiment Model: VADER is best suited for social media texts; may not capture nuanced product sentiment accurately.

Example Use Cases

☑ Business Use

- Analyze customer feedback on new products.
- Track customer satisfaction over time.
- Identify common complaints or praises in reviews.

Academic / Research Use

- Study consumer behavior.
- Compare sentiment between different products.

Build datasets for machine learning projects.

Rersonal Use

- Decide whether to buy a product based on aggregated sentiment.
- Compare products before making a purchase.

X Technology Stack

Layer	Technology Used	Description
Frontend	Streamlit	For building an intuitive and responsive web interface
Backend	Python	Handles logic, scraping, and analysis
Web Scraping	BeautifulSoup + Requests	To parse HTML content and extract relevant review data
NLP Engine	VADERSentiment	For performing sentiment analysis on text
Data Visualization	Plotly Express	Interactive visualizations for sentiment breakdown
Data Handling	Pandas	Structured data manipulation and storage

Coding:

```
import streamlit as st
import requests
from bs4 import BeautifulSoup
import pandas as pd
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
import plotly.express as px
import nltk
import ssl
st.set_page_config(
   page_title="Amazon Product Review Analysis Tool",
   layout="wide",
   initial_sidebar_state="expanded"
if "global_df" not in st.session_state:
   st.session_state.global_df = pd.DataFrame(columns=["Name", "Date", "Star_Rating",
"Short_Review", "Long_Review"])
def extract_reviews(url):
   headers = {
        'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) '
                    'AppleWebKit/537.36 (KHTML, like Gecko) '
                   'Chrome/120.0.0.0 Safari/537.36',
        'Accept-Language': 'en-US,en;q=0.9',
        'Referer': 'https://www.amazon.com/', # Helps mimic natural browsing behavior
```

```
'Accept-Encoding': 'gzip, deflate, br', # Supports compressed responses for faster
        'Connection': 'keep-alive', # Keeps the session open for multiple requests
        'Cache-Control': 'max-age=0', # Prevents caching issues that may affect dynamic
        'Upgrade-Insecure-Requests': '1', # Ensures secure browsing for better stability
   web = requests.get(url, headers=headers)
    if web.status code != 200:
        st.error("Failed to connect. Please check the link.")
        return pd.DataFrame()
    soup = BeautifulSoup(web.content, 'html.parser')
    names = [i.get_text(strip=True) if i else "N/A" for i in soup.find_all('span', {'class':
'a-profile-name'})]
   dates = [i.get_text(strip=True).replace("Reviewed in India on ", "") if i else "N/A" for i
in soup.find_all('span', {'class': 'review-date'})]
    star_ratings = [i.get_text(strip=True) if i and "out of 5 stars" in i.text else "N/A" for
i in soup.find_all('span', {'class': 'a-icon-alt'})]
    short reviews = []
    for link in soup.find_all('a', {'class': 'review-title'}):
        review_text = link.get_text(strip=True)
        cleaned_text = review_text.split("out of 5 stars")[-1].strip()
       if cleaned text:
            short_reviews.append(cleaned_text)
    long_reviews = []
    review_text_containers = soup.find_all('span', class_=lambda c: c and 'a-size-base' in c
and 'review-text' in c)
    for container in review_text_containers:
        review_body_span = container.find('span')
        review_text = review_body_span.get_text(strip=True) if review_body_span else
container.get_text(strip=True)
        read_more_link = container.find_next_sibling('div', class_=lambda c: c and 'a-
expander-header' in c)
        if read_more_link:
           text_parts = [c.strip() for c in container.contents if isinstance(c, str) or
c.name == "span"]
            review_text = " ".join(text_parts)
       if review_text:
            long_reviews.append(review_text)
    max_length = max(len(names), len(dates), len(star_ratings), len(short_reviews),
len(long_reviews))
   names += ["N/A"] * (max_length - len(names))
```

```
dates += ["N/A"] * (max_length - len(dates))
    star_ratings += ["N/A"] * (max_length - len(star_ratings))
    short_reviews += ["N/A"] * (max_length - len(short_reviews))
    long_reviews += ["N/A"] * (max_length - len(long_reviews))
    df = pd.DataFrame({
       "Name": names,
        "Date": dates,
        "Star_Rating": star_ratings,
        "Short_Review": short_reviews,
       "Long_Review": long_reviews,
    })
    df.replace("N/A", pd.NA, inplace=True)
    df.dropna(how='all', inplace=True)
    return df
st.markdown(
   <style>
        body { background-color: #121212; color: #ffffff; }
        .reportview-container { background-color: #121212; }
        .sidebar .sidebar-content { background-color: #1f1f1f; }
        h1, h2, h3, h4, h5, h6 { color: #ffffff; }
    </style>
   unsafe_allow_html=True
choice = st.sidebar.selectbox(
    "# Select an Option",
    ("♠ Home", "♥ 🛒 Amazon Product Review Extractor", "📊 Review Analysis", "⊿ Result
Visualization")
if choice == "â Home":
    st.title("♯ Amazon Product Review Analysis Tool ♯")
    # **Stylized Welcome Message**
    st.markdown(
        <div style="text-align: center; font-size:22px; font-weight: bold;</pre>
                    color: #FFA500; background-color: #1f1f1f; padding: 10px;
                    border-radius: 8px:">
```

```
Welcome to the cutting-edge sentiment analysis platform! 🖋
        </div>
        unsafe_allow_html=True,
    st.markdown(
        <div style="text-align: center; font-size:20px; font-weight: bold;</pre>
                    color: #00FF00; background-color: #333333; padding: 10px;
                    border-radius: 8px;">
        Powered by VADER Sentiment Model, this tool decodes and understands Amazon product
reviews with precision.
        </div>
       unsafe_allow_html=True,
    # **Visually Engaging Flow Section**
    st.subheader(" How It Works:")
    st.success("1\subseteq **Extract Reviews** - Enter an Amazon product link to retrieve customer
    st.info("2□ **Analyze Sentiment** - Process extracted reviews to determine sentiment
(Positive, Neutral, Negative).")
    st.warning("3□ **Visualize Results** — View interactive charts and insights based on
analyzed sentiments.")
   # **Call-to-Action**
    st.markdown(
        "<h3 style='text-align: center;'><sup>†</sup> Get started by selecting an option from the
sidebar! 'ᡮ</h3>",
       unsafe allow html=True,
elif choice == "" 🛒 Amazon Product Review Extractor":
    st.subheader("Amazon Product Review Extractor")
    st.write("Enter an Amazon product link to extract and analyze customer reviews
efficiently.")
    url = st.text input("Enter Link and then press enter:")
    st.info("Amazon's website structure changes frequently, which may impact review
extraction. You may need to click 'Extract' button multiple times.")
    st.warning(' Before proceeding with Sentiment Analysis on the downloaded CSV file, ensure
the data is cleaned, properly formatted, and free of empty values for accurate results.')
    if url:
        st.write("Press 'Extract' to fetch reviews.")
        if st.button("Extract"):
            extracted_df = extract_reviews(url)
            if not extracted_df.empty:
```

```
separator_df = pd.DataFrame([["-", "-", "-", "-", "-"]], columns=["Name",
"Date", "Star_Rating", "Short_Review", "Long_Review"])
                st.session_state.global_df = pd.concat([st.session_state.global_df,
separator_df, extracted_df], ignore_index=True)
                st.success("Reviews extracted successfully!")
                st.write(st.session_state.global_df)
       next_url = st.text_input("Enter the next page URL and press Enter")
        if next_url and st.button("Add Extra Data"):
            next_df = extract_reviews(next_url)
            if not next df.empty:
                separator_df = pd.DataFrame([["-", "-", "-", "-", "-"]], columns=["Name",
"Date", "Star_Rating", "Short_Review", "Long_Review"])
                st.session_state.global_df = pd.concat([st.session_state.global_df,
separator_df, next_df], ignore_index=True)
                st.success("Next page reviews added successfully!")
                st.write(st.session state.global df)
        if st.sidebar.button("Reset Data");
           st.session_state.global_df = pd.DataFrame(columns=["Name", "Date", "Star_Rating",
"Short Review", "Long Review"])
           st.success("Data reset successfully!")
    if st.button("Download as CSV"):
        if st.session state.global df.empty:
            st.warning("No reviews extracted yet!")
       else:
            st.session state.global df.to csv("Amazon Extracted Data.csv", index=False)
            st.success("Amazon extracted data saved as 'Amazon Extracted Data.csv'.")
            st.success("To proceed with sentiment analysis, please upload
'Amazon_Extracted_Data.csv' in the 'Analysis with CSV file' section.")
elif choice == " Review Analysis":
    st.subheader("Amazon Product Review Analysis")
    st.write("Upload a CSV file, enter a Google Sheet link, or use extracted Amazon reviews
for analysis.")
    input_choice = st.selectbox("Select Input Method", ("Extracted Amazon Reviews", "Upload
CSV File", "Google Sheet Link"))
   df = None
    if input choice == "Extracted Amazon Reviews":
       if "global_df" in st.session_state:
            df = st.session_state.global_df.copy()
```

```
else:
            st.warning("No reviews extracted yet! Please extract reviews first.")
    elif input choice == "Upload CSV File":
        uploaded_file = st.file_uploader("Upload a CSV file", type=["csv"])
        if uploaded_file:
            encodings = ["utf-8", "ISO-8859-1", "latin1"]
            for encoding in encodings:
                try:
                    df = pd.read_csv(uploaded_file, encoding=encoding, on_bad_lines="skip")
                    st.write(f"CSV file successfully read with encoding: {encoding}")
                    break
                except Exception:
                    st.warning(f"Failed to read file with encoding {encoding}. Trying
           if df is None:
                st.error("Failed to read the CSV file.")
    elif input choice == "Google Sheet Link":
        st.success('''Confirm the sheet link is proper and sharing settings allow public
access. 🖋 This helps prevent data parsing issues!''')
        google sheet link = st.text input("Enter Google Sheet link and press Enter")
        if google_sheet_link:
            if "/edit?usp=sharing" in google_sheet_link:
                google_sheet_link = google_sheet_link.replace("/edit?usp=sharing",
"/export?format=csv")
                st.success("✓ Google Sheets link converted for CSV export!")
           df = None
            encodings = ["utf-8", "ISO-8859-1", "latin1"]
            for encoding in encodings:
                try:
                    ssl. create default https context = ssl. create unverified context
                    df = pd.read_csv(google_sheet_link, encoding=encoding,
on_bad_lines="skip")
                    st.write(f"Google Sheet successfully read with encoding: {encoding}")
                    break
                except Exception:
                    st.warning(f"▲ Failed to read file with encoding {encoding}. Trying
next...")
    if df is not None and not df.empty:
       st.write("Successfully loaded data:")
       st.dataframe(df)
        if "global_df" not in st.session_state:
```

```
st.session_state.global_df = df.copy()
        c = st.selectbox("Choose Column", df.columns)
        if c:
            st.write(f"Column to analyze: {c}")
            if st.button("Analyze"):
                st.success("Performing sentiment analysis...")
                mymodel = SentimentIntensityAnalyzer()
                sentiments = []
                for i in range(len(df)):
                    text = df._get_value(i, c)
                    if pd.isna(text):
                        text = ""
                    else:
                        text = str(text)
                    pred = mymodel.polarity_scores(text)
                    if pred["compound"] > 0.5:
                        sentiments.append("Positive")
                    elif pred["compound"] < -0.5:</pre>
                        sentiments.append("Negative")
                        sentiments.append("Neutral")
                df["Sentiment"] = sentiments
                st.session_state.processed_df = df.copy()
                st.write("Sentiment Analysis Results:")
                st.dataframe(st.session_state.processed_df)
                st.session_state.processed_df.to_csv("Sentiment_Analysis_Results.csv",
index=False)
                st.success("Sentiment analysis completed. Results saved to
'Sentiment_Analysis_Results.csv'.")
   else:
        st.warning("No data available. Please upload a valid file or provide a valid Google
Sheet link.")
elif choice == "✓ Result Visualization":
    st.subheader("Result Visualization")
    st.write("Visualize the sentiment analysis results.")
    df = pd.read_csv("Sentiment_Analysis_Results.csv")
    if df.empty:
        st.warning("No sentiment analysis results available. Please perform sentiment analysis
first.")
   else:
        st.write("Sentiment Analysis Results:")
        st.dataframe(df)
```

```
v_choice = st.selectbox("Select Visualization Type", ("None","Pie Chart", "Scatter Plot",
"Histogram"))
    if v_choice == "Pie Chart":
        pos_per = (len(df[df["Sentiment"] == "Positive"]) / len(df)) * 100
        neg_per = (len(df[df["Sentiment"] == "Negative"]) / len(df)) * 100
        neu_per = (len(df[df["Sentiment"] == "Neutral"]) / len(df)) * 100
        fig = px.pie(values=[pos_per, neg_per, neu_per], names=["Positive", "Negative",
"Neutral"], title="Sentiment Analysis")
       st.plotly_chart(fig)
    elif v choice == "Histogram":
        k=st.selectbox("Choose Column", df.columns)
       if k:
            fig = px.histogram(x=df[k], color=df["Sentiment"], title="Sentiment Analysis")
           st.plotly chart(fig)
    elif v_choice == "Scatter Plot":
        k=st.selectbox("Choose Column", df.columns)
       if k:
            fig = px.scatter(x=df[k], y=df["Sentiment"], title="Sentiment Analysis")
            st.plotly_chart(fig)
st.write("---")
st.markdown(f"""*** Developed by Ashish Bonde** <br>
> **Interested in the Amazon Review Extractor and Sentiment Analysis WebApp?** <br
Connect with me on :<br/>LinkedIn](https://www.linkedin.com/in/ashish-bonde/)<br/>CitHub
Profile](https://github.com/Ashish-Bonde)<br>
[WhatsApp](https://api.whatsapp.com/send?phone=918484864084&text=Hi%20Ashish!%20I%20came%20acr
oss%20your%20Amazon%20Review%20Extractor%20and%20Sentiment%20Analysis%20WebApp%20and%20would%2
0love%20to%20connect%20to%20learn%20more%20about%20it.%20Let's%20connect!)
""", unsafe_allow_html=True)
```

Screenshot:





















