# Amazon Review Sentiment Analysis System

# WHAT

1. **It is a Natural Language Processing Application** that analyzes sentiment from **Amazon product reviews**.

2. The application predicts sentiment into **three categories**: **Positive**, **Negative**, and **Neutral**.

3. It **visualizes** the sentiment results based on **different factors** like **product category**, **rating**, **verified purchase status**, etc.

4. It **fetches** data from **Amazon reviews dataset** (scraped or publicly available datasets).

---

# WHY

1. This project has many use cases, such as **product feedback monitoring**, **brand sentiment tracking**, **competitor analysis**, and **customer satisfaction analysis**.

2. It demonstrates your skills in **programming**, **Machine Learning**, and **Natural Language Processing (NLP)** with **real-world data**.

3. It is a **strong portfolio project** — recruiters love projects with real-world e-commerce data.

---

# HOW

- **Backend**:

  - **Data Collection**: Use publicly available **Amazon Reviews datasets** (ex: Kaggle) or build a simple scraper (optional).

  - **Data Organization**: **Pandas** for data cleaning and processing.

- ○ **Sentiment Analysis**: Use **NLTK**, **VADER Sentiment**, or even fine-tune **HuggingFace models** for more advanced analysis.

- ○ **Data Visualization**: **Plotly** for interactive visual charts (pie charts, bar graphs, word clouds, etc.).

- ● **Frontend**:

  - ○ **User Interface**: **Streamlit** web app where users can upload Amazon review CSV files or enter review text manually.

  - ○ **Output**: Display sentiment analysis and visual insights .

## GOOGLE FORM AND SHEET

Steps Google acc, Google project, Enable google Api, Creating consent app Download credential

```python
from google_auth_oauthlib.flow import InstalledAppFlow
from googleapiclient.discovery import build
f=InstalledAppFlow.from_client_secrets_file("key.json",["https://www.googleapis.com/auth/spreadsheets"])
cred=f.run_local_server(port=0)
service=build("Sheets", "v4", credentials=cred).spreadsheets().values()
d=service.get(spreadsheetId="1pqEGewb_LuMY-f999yooJan5Qz6zmjPh8za5oKYIJfk", range="A:E").execute()
data=d['values']
print(data)
```

(AmazonTextAnalysis) F:\finalprojects\AmazonTextAnalysis\Scripts>python backend.py
Please visit this URL to authorize this application: https://accounts.google.com/o/oauth2/auth?response_type=code&cl
_id=801028058486-tett78ei2gdukde4vrg3qg50p24968dd.apps.googleusercontent.com&redirect_uri=http%3A%2F%2Flocalhost%3A5
%2F&scope=https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fspreadsheets&state=ELOPpEqxfe7yAF9UcnHy8YdPLazn3h&access_type=of
e
[['name', 'description', 'rating star', 'country', 'size'], ['Kristal Kincaid', '. Perfect to wear to work and styli
s well.', '5', 'USA', '15'], ['Johannes ', 'Received wrong shoe size 2 times in a row, ordered size 42 but got 40.5
 times', '1', 'SWEDEN', '7.5'], ['Human', "My kid says it's alright but annoys the top of his heels aka Achilles ten
area .", '2', 'UAE', '7'], ['RBDiego', 'Very comfortable shoe. Does not feel heavy either.', '5', 'Turkey', '7'], ['
lter Collins', 'True to size and light weight. Perfect walking shoes', '5', 'USA', '8'], ['Amazon Customer', ' Great
e, great value', '5', 'USA', '13'], ['Sarah Johnson', 'Terrible quality, ripped after two weeks. Waste of money.', '
'USA', '9']]

Frontend part

```python
import streamlit as st
from google_auth_oauthlib.flow import InstalledAppFlow
from googleapiclient.discovery import build
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
import pandas as pd
import plotly.express as px


st.title("AMAZON ANALYSIS SYSTEM")

choice=st.sidebar.selectbox("My Menu", ("HOME", "ANALYSIS","RESULTS"))
if (choice == "HOME"):
    st.image("https://miro.medium.com/v2/1*_JW1JaMpK_fVGld8pd1_JQ.gif")
    st.write("1.It is a Natural Language Processing Application which can analyze the sentiment on a text data.")
    st.write("2. This application predicts the sentiment into 3 categories Positive, Negative and Neutral.")
    st.write("3.This Application then visualizes the results based on different factors such as name,description,rating star and manymore.")
```

HOME ⌄
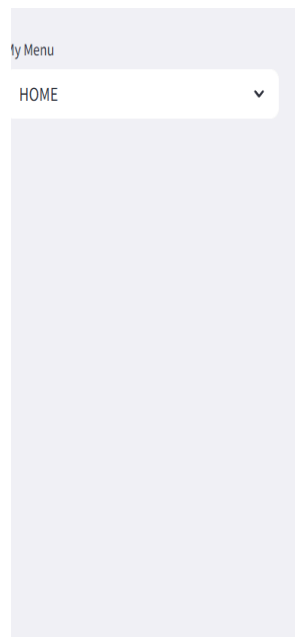
# AMAZON ANALYSIS SYSTEM

I would go agai

😧 😕 😐 🙂 😃

1.It is a Natural Language Processing Application which can analyze the sentiment on a text data.

2. This application predicts the sentiment into 3 categories Positive, Negative and Neutral.

3.This Application then visualizes the results based on different factors such as name,description,rating star and manymore.

```python
    st.write("3.This Application then visualizes the results based on different factors such as name,description,r
elif (choice == "ANALYSIS"):
    sid=st.text_input("Enter your Google Sheet ID")
    r=st.text_input("Enter Range between first column and last columns")
    c=st.text_input("Enter column name that is to be analyzed")
    btn=st.button("Analyze")
    if btn:
        if 'cred' not in st.session_state:
            f=InstalledAppFlow.from_client_secrets_file("key.json", ["https://www.googleapis.com/auth/spreadsheets
            st.session_state['cred']=f.run_local_server (port=0)
        mymodel=SentimentIntensityAnalyzer()
        service=build("Sheets", "v4", credentials=st.session_state['cred']).spreadsheets().values()
        k=service.get(spreadsheetId=sid, range=r).execute()
        d=k['values']
        df=pd.DataFrame(data=d[1:], columns=d[0])
        l=[]
        for i in range(0,len(df)):
            t=df._get_value(i,c)
            pred=mymodel.polarity_scores(t)
            if(pred['compound' ]>0.5):
                l.append("Positive")
            elif (pred['compound' ]<-0.5):
                l.append("Negative")
            else:
                l.append("Neutral")
        df['Sentiment']=l
        df.to_csv("results.csv",index=False)
        st.subheader("The Analysis results are saved by the name of a results.csv file")
```

# AMAZON ANALYSIS SYSTEM

Enter your Google Sheet ID

1pqEGewb_LuMY-f999yooJan5Qz6zmjPh8za5oKYIJfk

Enter Range between first column and last columns

A:E

Enter column name that is to be analyzed

description

Analyze

## The Analysis results are saved by the name of a results.csv file

| | | | |
|---|---|---|---|
| main | 4/19/2025 7:21 PM | Python File | 3 KB |
| normalizer | 4/18/2025 10:42 AM | Application | 106 KB |
| numpy-config | 4/18/2025 10:44 AM | Application | 106 KB |
| pip | 4/17/2025 2:46 PM | Application | 106 KB |
| pip3.13 | 4/17/2025 2:46 PM | Application | 106 KB |
| pip3 | 4/17/2025 2:46 PM | Application | 106 KB |
| pyrsa-decrypt | 4/18/2025 10:42 AM | Application | 106 KB |
| pyrsa-encrypt | 4/18/2025 10:42 AM | Application | 106 KB |
| pyrsa-keygen | 4/18/2025 10:42 AM | Application | 106 KB |
| pyrsa-priv2pub | 4/18/2025 10:42 AM | Application | 106 KB |
| pyrsa-sign | 4/18/2025 10:42 AM | Application | 106 KB |
| pyrsa-verify | 4/18/2025 10:42 AM | Application | 106 KB |
| python | 10/7/2024 10:27 AM | Application | 245 KB |
| pythonw | 10/7/2024 10:27 AM | Application | 242 KB |
| results | 4/24/2025 4:42 PM | Microsoft Excel Co... | 1 KB |
| streamlit | 4/18/2025 10:45 PM | Windows Comma... | 1 KB |
| streamlit | 4/18/2025 10:46 PM | Application | 106 KB |
| watchmedo | 4/18/2025 10:43 PM | Application | 106 KB |

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | name | description | rating star | country | size | Sentiment | |
| 2 | Kristal Kinc | . Perfect to | 5 | USA | 15 | Positive | |
| 3 | Johannes | Received v | 1 | SWEDEN | 7.5 | Neutral | |
| 4 | Human | My kid say | 2 | UAE | 7 | Neutral | |
| 5 | RBDiego | Very comf | 5 | Turkey | 7 | Positive | |
| 6 | Walter Col | True to siz | 5 | USA | 8 | Positive | |
| 7 | Amazon C | Great sho | 5 | USA | 13 | Positive | |
| 8 | Sarah Johr | Terrible qu | 1 | USA | 9 | Negative | |
| 9 | | | | | | | |

```
    st.subheader("The Analysis results are saved by the name of a results.csv file")
elif(choice=="RESULTS"):
    df=pd.read_csv("results.csv")
    choice2=st.selectbox("Choose Visualization", ("NONE", "PIE CHART", "HISTOGRAM", "SCATTER PLOT"))
    st.dataframe (df)
    if (choice2=="PIE CHART"):
        posper=(len (df [df[ 'Sentiment'] == 'Positive'])/len (df))*100
        negper=(len(df [df[ 'Sentiment']=='Negative'])/len (df))*100
        neuper=(len(df [df['Sentiment']=='Neutral'])/len (df))*100
        fig=px.pie(values = [posper, negper, neuper], names=['Positive', 'Negative', 'Neutral'])
        st.plotly_chart(fig)
    elif (choice2=="HISTOGRAM"):
        k=st.selectbox("Choose column", df.columns)
        if k:
            fig=px.histogram (x=df[k], color=df[ 'Sentiment'])
            st.plotly_chart(fig)
    elif (choice2=="SCATTER PLOT"):
        k=st.text_input("Enter the continous column name")
        if k:
            fig=px.scatter (x=df[k],y=df[ 'Sentiment'])
            st.plotly_chart(fig)
```
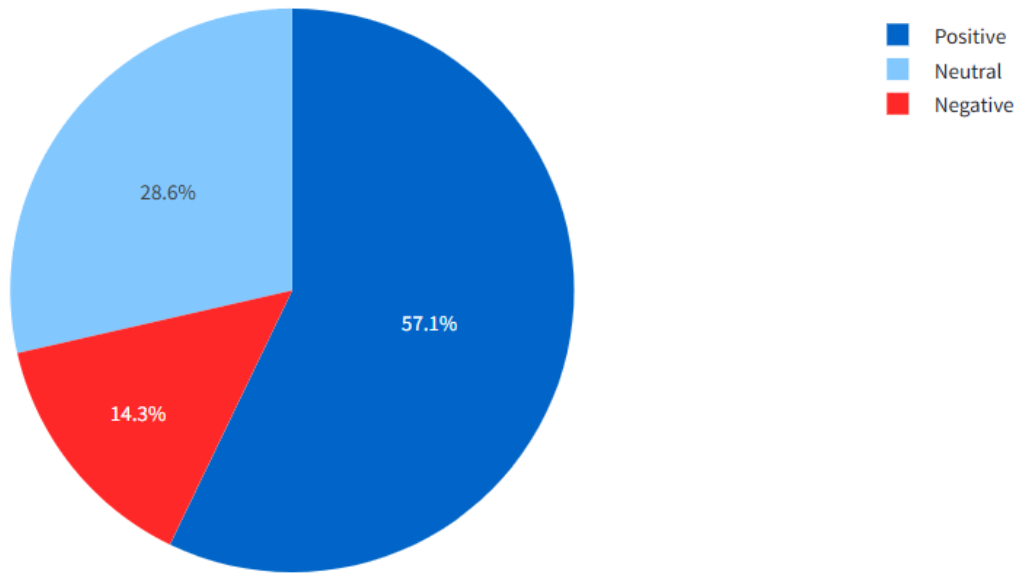
# AMAZON ANALYSIS SYSTEM

Choose Visualization

| NONE | ⌄ |

| | name | description | rating |
|---|---|---|---|
| 0 | Kristal Kincaid | . Perfect to wear to work and stylish as well. | |
| 1 | Johannes | Received wrong shoe size 2 times in a row, ordered size 42 but got 40.5 both times | |
| 2 | Human | My kid says it's alright but annoys the top of his heels aka Achilles tendon area . | |
| 3 | RBDiego | Very comfortable shoe. Does not feel heavy either. | |
| 4 | Walter Collins | True to size and light weight. Perfect walking shoes | |
| 5 | Amazon Customer | Great shoe, great value | |
| 6 | Sarah Johnson | Terrible quality, ripped after two weeks. Waste of money. | |

Choose column

country ⌄

Negative

Neutral

Positive