CUSTOMER SENTIMENT ANALYSIS

Objective of the project:

As a Data Analyst at Flipkart, the goal here is to analyze customer sentiment regarding 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, I then 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.

1. Data Collection (Web Scraping & Automation):

Tools used: Selenium and BeautifulSoup

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

Usernames: The name of the customer.
Ratings: The rating provided by the user (1 to 5 stars).
Reviews: The content of the customer's review, which may contain valuable information regarding their experience with the product.
Region: The place of the customer/user.

Install and import all the Python libraries that are required for the task. Such as,

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from selenium import webdriver
from bs4 import BeautifulSoup
from selenium.webdriver.common.by import By
from selenium.webdriver.common.keys import Keys
import time
from textblob import TextBlob
from nltk import sent_tokenize
import re
import nltk
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
```

After importing, start extracting the data from the Flipkart's website using BeautifulSoup for Web Scraping and Selenium for Automation.

```
driver = webdriver.Chrome() #Initialize Selenium WebDriver (opens Chrome browser)
#create empty lists to store the user data such as names,ratings, reviews and region
name = []
rating = []
review = []
region = []
url = """https://www.flipkart.com/apple-iphone-15-pink-128-gb/product-reviews/itm7579ed94ca647?pid=MOBGTAGPNMZA5PU5&lid=LSTMOBGTAGPNMZA5PU5E1UCRJ&
marketplace=FLIPKART&page="""

for page_num in range(1,20):
    new_url = url+str(page_num)
    driver.get(new_url)     #Open the page in Chrome
    time.sleep(3)   #Wait for JavaScript content to fully load
    soup = BeautifulSoup(driver.page_source,"html.parser")  #Parse the page content

# requests.get(url) is used to fetch a webpage, but it only retrieves static HTML but Flipkart is a dynamic page
    #extract customer names
    Names = soup.find_all("p",{"class":"_2NsDsF AwS1CA"})
    for n in Names:
        name.append(n.text)

    #extract customer ratings
    Ratings = soup.find_all("div",{"class":"XQDdHH Ga3i8K"})
    for ra in Ratings:
        rating.append(ra.text)

    #extract customer reviews
    Reviews = soup.find_all("div",{"class":"ZmyHeo"})
    for re in Reviews:
        review.append(re.text)

    #extract customer region
    Region = soup.find_all("p",{"class":"MztJPv"})
    for reg in Region:
        region.append(reg.text)

driver.quit()    #Close the Selenium WebDriver (closes Chrome browser)
print("Names: ",name)
print("Ratings: ",rating)
print("Reviews: ",review)
print("Region: ",region)
```

```
mer', 'MONIRUL MOLLA', 'Rupesh jha', 'Mohammad  Zishan ', 'Akash  Deep', 'neeraj gour', 'K Deepak', 'KaryD', 'Dr Babu  Mandal', 'Venkat Ramana', 'De
epti Rankhambe', 'Mohammad  Mubarak', 'anujyadav anuj', 'Mohamed Suheb', 'Sudipto Sarkar', 'Dev Sarda', 'Naresh Soni', 'ABDUR RAHMAN ', 'Mayank Vars
hney', 'Flipkart Customer', 'Athrav Ankur', 'Vanshu  Customer', 'SOURAV NANDI', 'Nitin Patil', 'Tekchand Biswal', 'vinay mohan som', 'Durga prasad',
'Deepsayan Dam', 'Yash', 'Amit Biswas', 'Abhishek Kumar', 'Syed Intekhab  Hussain', 'R K', 'PROSIT DAS', 'Dhruv M Golwala', 'Flipkart Customer', 'Pr
atyay Sarkar', 'Ravi Gangrale', 'Flipkart Customer', 'santosh  parida', 'Prasanth Rajendran', 'Inderjit Singh', 'VENKATESH REDDY', 'Sunny  Patel ',
'Satyam Kumar', 'Vamshi Reddy', 'mrityunjay kanwar', 'Naseer Ahmed', 'Shrikant Panigrahi', 'Ashis Karamakar', 'Satendra  Singh', 'jithu reddy', 'Sah
il Badkul', 'Aditya  Arya', 'Munish  Choudhary', 'Abnish  Sen ', 'Ajay G', 'Ritesh Kumar', 'Arman Ali', 'Motam Jamatia']
Ratings:  ['4', '5', '5', '5', '5', '5', '5', '4', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '4', '5', '5', '4', '5', '5', '5',
'5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '4', '5', '5', '5', '5', '4', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5',
'5', '5', '5', '5', '5', '5', '5', '5', '4', '5', '5', '5', '5', '4', '5', '5', '5', '5', '5', '5', '5', '5', '5', '4', '4', '4', '5', '4', '5', '5',
'5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '4', '5', '5', '5', '3', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5',
'5', '5', '5', '5', '5', '4', '4', '5', '5', '5', '5', '5', '5', '5', '5', '5', '4', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5',
'5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '4']
Reviews:  ['Very niceREAD MORE', 'Just go for it.Amazing one.Beautiful camera with super fast processorREAD MORE', 'High quality camera😍READ MORE',
'Switch from OnePlus to iPhone I am stunned with camera performance. Everything is perfect on  iPhone 15.READ MORE', 'Camera Quality Is Improved Lov
ing ItREAD MORE', 'So beautiful, so elegant, just a vowww😍❤️READ MORE', 'Awesome photography experience. Battery backup is good . Display is much
better than 14 version.READ MORE', 'Best mobile phoneCamera quality is very nice Battery backup is very good Sound quality is amazing.READ MORE', 'T
otally happy!Camera 5Battery 5 Display 5Design 5READ MORE', 'Simply premium.READ MORE', 'Just loved the product , colour , design is wow and camera
```

Put the extracted data into a table using DataFrame from Pandas.

```
#create a dataframe for the extracted data
df = pd.DataFrame({"Customer Name":name, "Ratings":rating, "Reviews":review, "Region":region})
df
```

	Customer Name	Ratings	Reviews	Region
0	Mousam Guha Roy	4	Very niceREAD MORE	Certified Buyer, Matialihat
1	bijaya mohanty	5	Just go for it.Amazing one.Beautiful camera wi...	Certified Buyer, Baleshwar
2	Ajin V	5	High quality camera 😍 READ MORE	Certified Buyer, Balaghat
3	Nikhil Kumar	5	Switch from OnePlus to iPhone I am stunned wit...	Certified Buyer, Meerut Division
4	Prithivi Boruah	5	Camera Quality Is Improved Loving ItREAD MORE	Certified Buyer, Bokajan
...	...	...	...	...
164	Abnish Sen	5	SuperREAD MORE	Certified Buyer, Kanpur
165	Ajay G	5	Back to iPhone family from android, best upgra...	Certified Buyer, Hyderabad
166	Ritesh Kumar	5	Best phone nice camera and performanceREAD MORE	Certified Buyer, Buxar District
167	Arman Ali	5	Pros48MP Camera With 2x Zoom. Dynamic Island. ...	Certified Buyer, Bettiah
168	Motam Jamatia	4	I phone is phone nothing but it's a brand that...	Certified Buyer, Teliamura

169 rows × 4 columns

2. Data Cleaning and Preprocessing:

Tool used: Pandas

Task: Clean and preprocess the scraped data for analysis.
Remove duplicates, handle missing values, 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 like “Read More”).

```
#Captialize the first letter of each word in Customer Name(Proper Case)
df["Customer Name"] = df["Customer Name"].str.title().str.strip()
df
```

	Customer Name	Ratings	Reviews	Region
0	Mousam Guha Roy	4	Very niceREAD MORE	Certified Buyer, Matialihat
1	Bijaya Mohanty	5	Just go for it.Amazing one.Beautiful camera wi...	Certified Buyer, Baleshwar
2	Ajin V	5	High quality camera 😍 READ MORE	Certified Buyer, Balaghat
3	Nikhil Kumar	5	Switch from OnePlus to iPhone I am stunned wit...	Certified Buyer, Meerut Division
4	Prithivi Boruah	5	Camera Quality Is Improved Loving ItREAD MORE	Certified Buyer, Bokajan
...	...	...	...	...
164	Abnish Sen	5	SuperREAD MORE	Certified Buyer, Kanpur
165	Ajay G	5	Back to iPhone family from android, best upgra...	Certified Buyer, Hyderabad
166	Ritesh Kumar	5	Best phone nice camera and performanceREAD MORE	Certified Buyer, Buxar District
167	Arman Ali	5	Pros48MP Camera With 2x Zoom. Dynamic Island. ...	Certified Buyer, Bettiah
168	Motam Jamatia	4	I phone is phone nothing but it's a brand that...	Certified Buyer, Teliamura

169 rows × 4 columns

```
#replacing "read more" in the reviews column & convert to lower case
df["Reviews"] = df["Reviews"].str.replace("READ MORE","").str.lower()
df
```

	Customer Name	Ratings	Reviews	Region
0	Mousam Guha Roy	4	very nice	Certified Buyer, Matialihat
1	Bijaya Mohanty	5	just go for it.amazing one.beautiful camera wi...	Certified Buyer, Baleshwar
2	Ajin V	5	high quality camera😍	Certified Buyer, Balaghat
3	Nikhil Kumar	5	switch from oneplus to iphone i am stunned wit...	Certified Buyer, Meerut Division
4	Prithivi Boruah	5	camera quality is improved loving it	Certified Buyer, Bokajan
...	...	...	...	...
164	Abnish Sen	5	super	Certified Buyer, Kanpur
165	Ajay G	5	back to iphone family from android, best upgra...	Certified Buyer, Hyderabad
166	Ritesh Kumar	5	best phone nice camera and performance	Certified Buyer, Buxar District
167	Arman Ali	5	pros48mp camera with 2x zoom. dynamic island. ...	Certified Buyer, Bettiah
168	Motam Jamatia	4	i phone is phone nothing but it's a brand that...	Certified Buyer, Teliamura

169 rows × 4 columns

```
#removing "Certified Buyer" and keeping the place name in Region column
df["Region"] = df["Region"].str.replace("Certified Buyer, ","").str.strip()
df
```

	Customer Name	Ratings	Reviews	Region
0	Mousam Guha Roy	4	very nice	Matialihat
1	Bijaya Mohanty	5	just go for it.amazing one.beautiful camera wi...	Baleshwar
2	Ajin V	5	high quality camera😍	Balaghat
3	Nikhil Kumar	5	switch from oneplus to iphone i am stunned wit...	Meerut Division
4	Prithivi Boruah	5	camera quality is improved loving it	Bokajan
...	...	...	...	...
164	Abnish Sen	5	super	Kanpur
165	Ajay G	5	back to iphone family from android, best upgra...	Hyderabad
166	Ritesh Kumar	5	best phone nice camera and performance	Buxar District
167	Arman Ali	5	pros48mp camera with 2x zoom. dynamic island. ...	Bettiah
168	Motam Jamatia	4	i phone is phone nothing but it's a brand that...	Teliamura

169 rows × 4 columns

3. Sentiment Analysis:

Tool: TextBlob

Task: Perform sentiment analysis on the review text. Analyze the sentiment of each review to classify them as either positive or negative.
Define a threshold to classify the sentiment as **Positive sentiment and Negative sentiment** based on *polarity score*. Store the sentiment classification for each review in the dataset.

```
#Tokenizing the Reviews column as Reviews_1:
df["Reviews_1"] = df["Reviews"].apply(sent_tokenize)
df
```

	Customer Name	Ratings	Reviews	Region	Reviews_1
0	Mousam Guha Roy	4	very nice	Matialihat	[very nice]
1	Bijaya Mohanty	5	just go for it.amazing one.beautiful camera wi...	Baleshwar	[just go for it.amazing one.beautiful camera w...
2	Ajin V	5	high quality camera😍	Balaghat	[high quality camera😍]
3	Nikhil Kumar	5	switch from oneplus to iphone i am stunned wit...	Meerut Division	[switch from oneplus to iphone i am stunned wi...
4	Prithivi Boruah	5	camera quality is improved loving it	Bokajan	[camera quality is improved loving it]
...	...	...	...	...	...
164	Abnish Sen	5	super	Kanpur	[super]
165	Ajay G	5	back to iphone family from android, best upgra...	Hyderabad	[back to iphone family from android, best upgr...
166	Ritesh Kumar	5	best phone nice camera and performance	Buxar District	[best phone nice camera and performance]
167	Arman Ali	5	pros48mp camera with 2x zoom. dynamic island. ...	Bettiah	[pros48mp camera with 2x zoom., dynamic island...
168	Motam Jamatia	4	i phone is phone nothing but it's a brand that...	Teliamura	[i phone is phone nothing but it's a brand tha...

169 rows × 5 columns

```
#to get the polarity of the sentences
def get_polarity(sentences):
    return [TextBlob(sentence).sentiment.polarity for sentence in sentences]
df['Polarity'] = df['Reviews_1'].apply(get_polarity)

from statistics import mean   #to get the avg of the polaritites
def calculate_average_polarity(polarities):
    return mean(polarities) if polarities else 0

df['Average_Polarity'] = df['Polarity'].apply(calculate_average_polarity)
df.head(10)
```

	Customer Name	Ratings	Reviews	Region	Reviews_1	Polarity	Average_Polarity
0	Mousam Guha Roy	4	very nice	Matialihat	[very nice]	[0.78]	0.780000
1	Bijaya Mohanty	5	just go for it.amazing one.beautiful camera wi...	Baleshwar	[just go for it.amazing one.beautiful camera w...	[0.26666666666666666]	0.266667
2	Ajin V	5	high quality camera😍	Balaghat	[high quality camera😍]	[0.16]	0.160000
3	Nikhil Kumar	5	switch from oneplus to iphone i am stunned wit...	Meerut Division	[switch from oneplus to iphone i am stunned wi...	[0.0, 1.0]	0.500000
4	Prithivi Boruah	5	camera quality is improved loving it	Bokajan	[camera quality is improved loving it]	[0.6]	0.600000
5	Akshay Meena	5	so beautiful, so elegant, just a vowww😍❤️	Jaipur	[so beautiful, so elegant, just a vowww😍❤️]	[0.675]	0.675000
6	Flipkart Customer	5	awesome photography experience. battery backup...	Aizawl	[awesome photography experience., battery back...	[1.0, 0.7, 0.5]	0.733333
7	Sheetla Prasad Maurya	4	best mobile phonecamera quality is very nice b...	Sultanpur	[best mobile phonecamera quality is very nice ...	[0.738]	0.738000
8	Rahul Shedge	5	totally happy!camera 5battery 5 display 5design 5	Satara	[totally happy!camera 5battery 5 display 5desi...	[0.0]	0.000000
9	Saurabh Gothwad	5	simply premium.	Paradip	[simply premium.]	[0.0]	0.000000

```python
#creating the function to know where the polarity lies
def sentiment(pol):
    if pol >= 0.75:
        return "Extremely Positive"
    elif pol > 0:
        return "Postive"
    elif pol == 0:
        return "Neutral"
    elif pol <= -0.75:
        return "Negative"
    else:
        return "Extremely Negative"
df["sentiment"] = df["Average_Polarity"].apply(sentiment)
df.head(10)
```

	Customer Name	Ratings	Reviews	Region	Reviews_1	Polarity	Average_Polarity	sentiment	Review Length
0	Mousam Guha Roy	4	very nice	Matialihat	[very nice]	[0.78]	0.780000	Extremely Positive	2
1	Bijaya Mohanty	5	just go for it.amazing one.beautiful camera wi...	Baleshwar	[just go for it.amazing one.beautiful camera w...	[0.26666666666666666]	0.266667	Postive	10
2	Ajin V	5	high quality camera😍	Balaghat	[high quality camera😍]	[0.16]	0.160000	Postive	3
3	Nikhil Kumar	5	switch from oneplus to iphone i am stunned wit...	Meerut Division	[switch from oneplus to iphone i am stunned wi...	[0.0, 1.0]	0.500000	Postive	17
4	Prithivi Boruah	5	camera quality is improved loving it	Bokajan	[camera quality is improved loving it]	[0.6]	0.600000	Postive	6
5	Akshay Meena	5	so beautiful, so elegant, just a vowww😍❤️	Jaipur	[so beautiful, so elegant, just a vowww😍❤️]	[0.675]	0.675000	Postive	7
6	Flipkart Customer	5	awesome photography experience. battery backup...	Aizawl	[awesome photography experience., battery back...	[1.0, 0.7, 0.5]	0.733333	Postive	15

4. Data Analysis and Insights:

Tool: Pandas and Matplotlib/Seaborn for visualization**.**

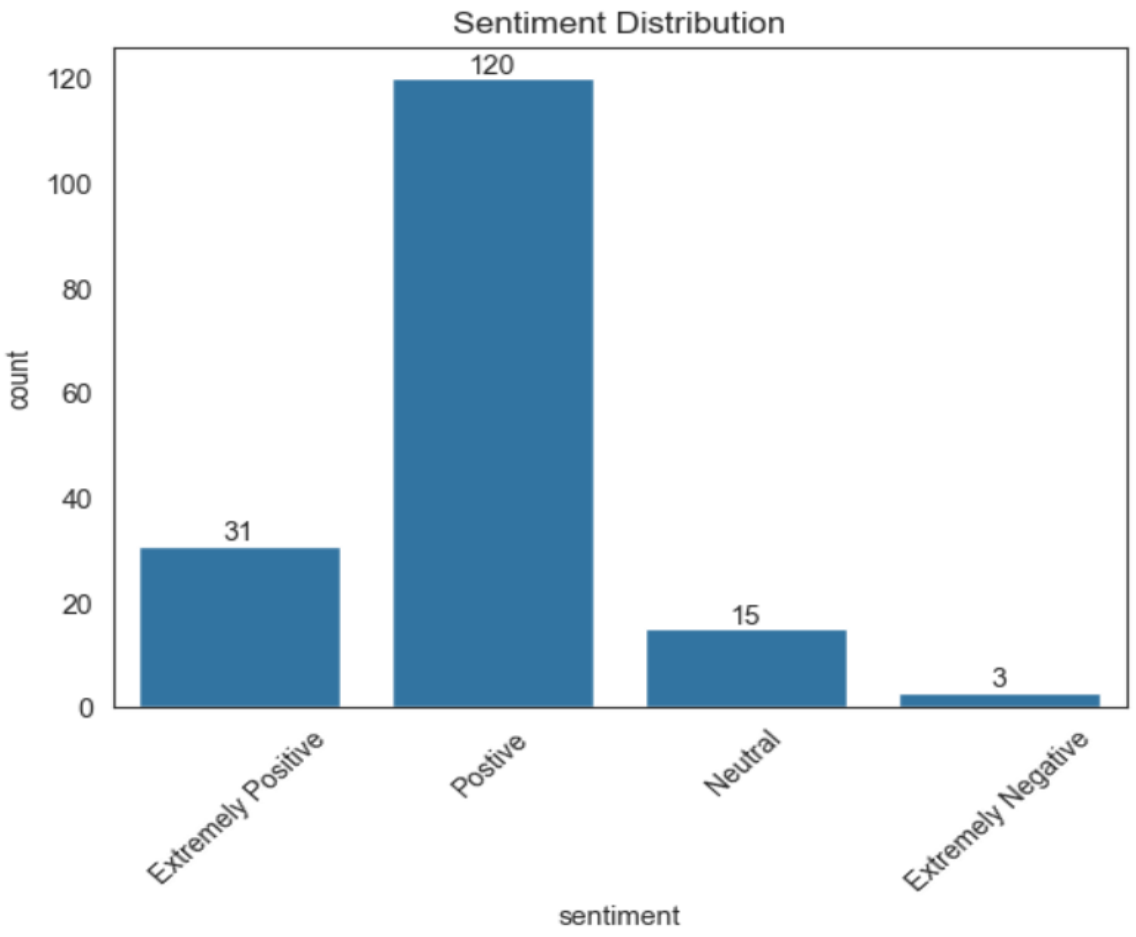
Task: Perform an analysis on the sentiment of reviews and extract actionable insights.

- **Sentiment Distribution:** Calculate the overall distribution of positive and negative sentiments for all the reviews.
- **Review Length Analysis:** Investigate if longer reviews are associated with more detailed sentiments, either positive or negative.

```python
#4. Data Analysis and Insights:
#This will calculate and also print the mean polarity value
df["Average_Polarity"].mean()
```

```
np.float64(0.4741037885428911)
```

```
sns.set_style("white")  # Clean, no gridlines
ax = sns.countplot(x = df["sentiment"]) #creating a Seaborn chart on the sentiments
ax.bar_label(ax.containers[0])
plt.xticks(rotation = 45)
plt.title("Sentiment Distribution")
plt.show()
```



Sentiment Distribution:
This bar graph represents the distribution of customer sentiments. The x-axis shows different sentiment categories: **Extremely Positive, Positive, Neutral, Negative and Extremely Negative**, while the y-axis indicates the count of each sentiment.

Positive sentiments are the most frequent (120 counts),
followed by Extremely Positive (31),
Neutral (15),
and Extremely Negative sentiment (3).

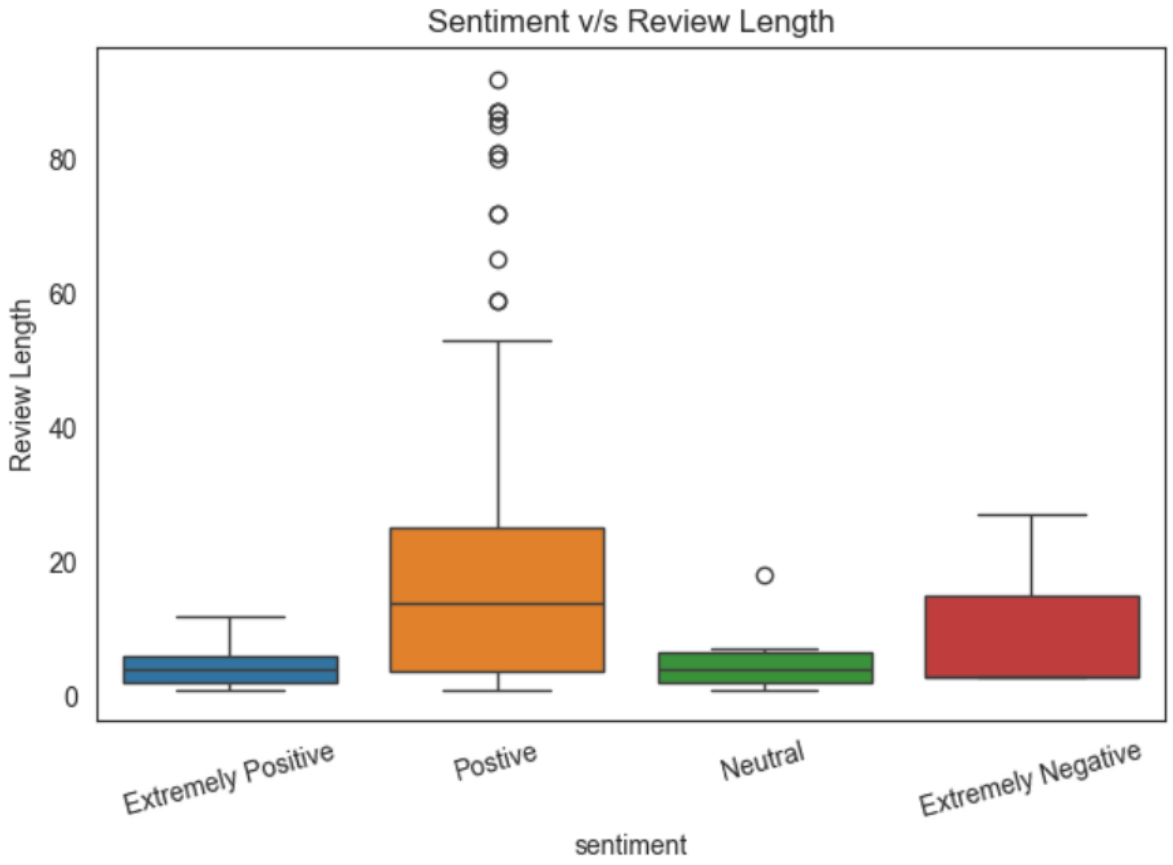
This suggests that the majority of customers have expressed positive or favourable feedback for iPhone 15 128 GB mobile phone.

```
#create a new column to know the number of words in reviews
df["Review Length"] = df["Reviews"].apply(lambda x : len(x.split()))
df
```

	Customer Name	Ratings	Reviews	Region	Reviews_1	Polarity	Average_Polarity	sentiment	Review Length
0	Mousam Guha Roy	4	very nice	Matialihat	[very nice]	[0.78]	0.780000	Extremely Positive	2
1	Bijaya Mohanty	5	just go for it.amazing one.beautiful camera wi...	Baleshwar	[just go for it.amazing one.beautiful camera w...	[0.26666666666666666]	0.266667	Postive	10
2	Ajin V	5	high quality camera😍	Balaghat	[high quality camera😍]	[0.16]	0.160000	Postive	3
3	Nikhil Kumar	5	switch from oneplus to iphone i am stunned wit...	Meerut Division	[switch from oneplus to iphone i am stunned wi...	[0.0, 1.0]	0.500000	Postive	17
4	Prithivi Boruah	5	camera quality is improved loving it	Bokajan	[camera quality is improved loving it]	[0.6]	0.600000	Postive	6
...	...	...	...	...	...	...	...	...	...
164	Abnish Sen	5	super	Kanpur	[super]	[0.3333333333333333]	0.333333	Postive	1
165	Ajay G	5	back to iphone family from android, best upgra...	Hyderabad	[back to iphone family from android, best upgr...	[0.5, 0.5, -0.03571428571428571, 0.21333333333...	0.294405	Postive	43
166	Ritesh Kumar	5	best phone nice camera and performance	Buxar District	[best phone nice camera and performance]	[0.8]	0.800000	Extremely Positive	6
167	Arman Ali	5	pros48mp camera with 2x zoom. dynamic island. ...	Bettiah	[pros48mp camera with 2x zoom., dynamic island...	[0.0, 0.0, 0.0, 1.0, 0.0, -0.30000000000000004...	0.094444	Postive	72
168	Motam Jamatia	4	i phone is phone nothing but it's a brand that...	Teliamura	[i phone is phone nothing but it's a brand tha...	[0.0, 0.5, 0.0]	0.166667	Postive	22

169 rows × 9 columns

```
sns.boxplot(x= "sentiment", y= "Review Length", data = df, hue ="sentiment")
plt.xlabel("sentiment")
plt.ylabel("Review Length")
plt.xticks(rotation = 15)
plt.title("Sentiment v/s Review Length")
plt.tight_layout()
plt.show()
```



Observation:
Reviews with moderately positive sentiment tend to be longer and more detailed, often including a wide range of feedback — even outliers. In contrast, neutral and extreme sentiments (like “Amazing!” or “Terrible!”) are usually short and to the point.

5. Summary of Sentiment Analysis:

🔹 **Objective:**
To analyze public sentiment about the iPhone 15 (128GB) using customer reviews from Flipkart, aiming to extract insights for improving customer experience and understanding product perception.

🔹 **Method:**
- Web scraping using **Selenium** and **BeautifulSoup**
- Data cleaning with **Pandas**
- Sentiment analysis using **TextBlob**
- Data visualization using **Seaborn**

🔹 **Sentiment Insights:**
- Positive sentiments were most frequent (120), followed by extremely positive (31), neutral (15), and extremely negative (3).
- Indicates a strong overall **positive reception** of the product.

🔹 **Review Length Observation:**
- **Moderately positive reviews** were longer and more detailed, often including diverse user experiences and feedback.
- **Neutral and extreme sentiments** (e.g., “Amazing!”, “Terrible!”) were typically **short and brief**, lacking elaboration.

🔹 **Conclusion:**
Customers generally feel good about the product. The most useful, constructive feedback comes from **moderately satisfied users**, while short emotional reactions are common at both extremes.