In [5]:
```
import os
import pandas as pd
import requests
from bs4 import BeautifulSoup
import numpy as np
import re
import math
from fake_useragent import UserAgent
os.environ["OPENAI_API_KEY"] = "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"
```

In [6]:
```
headers={'User-Agent':'Mozilla/5.0 (Windows NT 6.3; Win 64 ; x64) Apple WeKit /537.36(KH
webpage = requests.get('https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BR
```

In [7]:
```
soup=BeautifulSoup(webpage,'lxml')
#soup.prettify()
```

In [8]:
```
title=[]
price=[]
ratings=[]
about_this_item=[]
#filling the lists
title.append(soup.find_all('h1', class_ = 'a-size-large')[0].text.strip())
price.append(soup.find_all(class_='a-offscreen')[0].text.strip())
ratings.append(soup.find_all(id='acrCustomerReviewText')[0].text.strip())
about_this_item.append(soup.find_all(id='feature-bullets')[0].text.strip())
```

In [9]:
```
#dictionary
specification = {}
```

In [10]:
```
#soup.find('table', id='productDetails_techSpec_section_1').find_all('th', class_='a-col
length = len(soup.find('table', id='productDetails_techSpec_section_1').find_all('th', c

for i in range(length):
    specification[soup.find('table', id='productDetails_techSpec_section_1').find_all('t
```

In [11]:
```
#creating a dataframe
df1=pd.DataFrame({'title':title,
   'price':price,
   'ratings':ratings,
   'about_this_item':about_this_item,
   })
df1
#converting dict to dataframe
new_dataframe = pd.DataFrame(specification, index=[0])
new_dataframe
#concatenate
updated_df = pd.concat([df1, new_dataframe], axis=1)
updated_df
```

Out[11]:

	title	price	ratings	about_this_item	OS	RAM	Product Dimensions	Batteries	Item model number	Wireles communicatio technologie
0	Samsung Galaxy S23 Ultra 5G (Cream, 12GB, 256G...	₹1,24,999.00	1,025 ratings	About this item More innovation, less footp...	Android 12.0	12 GB	0.8 x 7.1 x 14.6 cm; 233 Grams	1 Lithium Ion batteries required. (included)	SM-S918B	Cellul

1 rows × 23 columns

In [12]:
```
# Now training model on the Phone Dataframe
from langchain.agents import create_pandas_dataframe_agent
from langchain.llms import OpenAI
agent = create_pandas_dataframe_agent(OpenAI(temperature=0), updated_df, verbose=False)
agent
agent.run("how many rows are there?")
# agent.run("what is the dataframe related to?")
# agent.run("what is a mango")
# agent.run("can you give me a brief description about samsung s23 ultra 5g phone?")
# agent.run("how much ram in samsung s23 ultra?")
```

Out[12]:
```
'There are 1 rows in the dataframe.'
```

In [13]:
```
#Review Analysis
product_url = 'https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5'
```

In [14]:
```
def scrape_amazon_reviews(product_url, max_pages=None):
    reviews = []
    page_num = 1

    while True:
        url = f"{product_url}?ie=UTF8&reviewerType=all_reviews&pageNumber={page_num}"
        print(url)
        ua = UserAgent()
        headers = {"User-Agent": ua.random}
        response = requests.get(url, headers=headers)
        soup = BeautifulSoup(response.text, "html.parser")

        reviews_divs = soup.find_all("div", {"data-hook": "review"})

        if not reviews_divs:
            break

        for review in reviews_divs:
            author_elem = review.find("span", {"class": "a-profile-name"})
            rating_elem = review.find("i", {"data-hook": "review-star-rating"})
            title_elem = review.find("a", {"data-hook": "review-title"})
            date_elem = review.find("span", {"data-hook": "review-date"})
            body_elem = review.find("span", {"data-hook": "review-body"})

            data = {
                "author": author_elem.text if author_elem else None,
                "rating": rating_elem.find("span").text if rating_elem and rating_elem.f
                "title": title_elem.find("span").text if title_elem and title_elem.find(
                "date": date_elem.text if date_elem else None,
                "body": body_elem.find("span").text if body_elem and body_elem.find("spa
            }
            reviews.append(data)

        if max_pages and page_num >= max_pages:
            break

        page_num += 1

    return pd.DataFrame(reviews)
```

In [15]:
```
%%time
df = scrape_amazon_reviews(product_url)
```

```
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerT
ype=all_reviews&pageNumber=1
```

https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=2
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=3
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=4
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=5
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=6
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=7
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=8
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=9
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=10
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=11
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=12
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=13
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=14
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=15
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=16
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=17
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=18
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=19
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=20
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=21
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=22
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=23
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=24
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=25
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=26
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=27
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=28
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=29
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=30
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=31
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=32
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=33
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=34

```
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=35
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=36
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=37
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=38
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=39
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=40
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=41
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=42
https://www.amazon.in/Samsung-Galaxy-Ultra-Cream-Storage/dp/B0BRSLH4B5?ie=UTF8&reviewerType=all_reviews&pageNumber=43
CPU times: total: 23.8 s
Wall time: 3min 12s
```

In [16]:
```
df.shape
```

Out[16]:
```
(320, 5)
```

In [17]:
```
from langchain.document_loaders import DataFrameLoader
from langchain.text_splitter import CharacterTextSplitter
from langchain.chat_models import ChatOpenAI
from langchain import OpenAI
from langchain.embeddings import OpenAIEmbeddings
from langchain.vectorstores import Chroma
loader = DataFrameLoader(df, page_content_column="body")
doc = loader.load()
```

In [18]:
```
text_splitter = CharacterTextSplitter(chunk_size=163, chunk_overlap=100)
texts = text_splitter.split_documents(doc)
```

In [19]:
```
embeddings = OpenAIEmbeddings(openai_api_key="sk-fwnAWQxJ7kyHKPOLZHCOT3BlbkFJAvzYITwfQkf
docsearch = Chroma.from_documents(texts, embeddings)
```

In [20]:
```
from langchain.chains import RetrievalQA
from langchain.chat_models import ChatOpenAI
```

In [21]:
```
qa = RetrievalQA.from_chain_type(llm=OpenAI(), chain_type="map_reduce", retriever = docs
```

In [22]:
```
#query about the reviews
query = "What negative reviews have people left about the phone?"
qa.run(query)
```

Out[22]:
```
' Negative reviews people have left about the phone include: "Battery life is poor", "sl
ight screen leakage", "hangs/screen freeze issue", "heats like hell while charging", "sc
reen refresh rate is not stable while multitasking", "Bixby issues", and "battery backup
does not last whole day".'
```

In [23]:
```
#query about the phone
agent.run("can you give me a brief description about samsung s23 ultra 5g phone?")
```

Out[23]:
```
'The Samsung Galaxy S23 Ultra 5G is a phone with a striking symmetrical design made from
recycled and eco-conscious materials. It has a built-in S Pen, a Pro-grade Camera, a 200
MP resolution Wide-angle Camera, and a Snapdragon 8 Gen 2 Mobile Platform for optimized
and streamlined gaming. It also has Fast Charging Support, Wireless Charging, and a 5000
Battery Power Rating.'
```

In [25]:
```
#query about the reviews
query = "What is the overall impression of these reviews, give most prevalent examples i
qa.run(query)
```

Out[25]:
```
' Overall impression: The reviews of the product are generally positive. Examples of pos
itive reviews include: good display, performance, battery life, portrait shots, super fa
st, no lag, overall great experience, improved battery life, overall performance, proces
sor, battery life, camera, display, and UI.'
```

In [27]:
```
#query about the reviews
query = "What do you suggest we should focus on improving based on these reviews?"
qa.run(query)
```

Out[27]:
```
' Focus on improving camera shutter lag, indoor shots, and human face shots.'
```

In [29]:
```
#query about the reviews
query = "Create a copy for a Facebook ad based on these reviews: {text} as far as text g
qa.run(query)
```

Out[29]:
```
' Get a Compact Phone with Awesome Camera, Great Processing & Fast Charging! Primary tex
t (125 characters): Get a compact phone with great battery backup and fast charging opti
ons that come with great camera and processing speed! Enjoy an awesome photography exper
ience and faster processing speed with this phone. Description (30 characters): Get a co
mpact phone with awesome features!'
```

In [30]:
```
Reviews_S23_Series = df
```

In [36]:
```
Reviews_S23_Series.to_csv('Reviews_S23_Series.csv', header=True, index=False)
```

In [ ]:
```

```