

✓ Data Loading from previous sentiment analysis

```
import pandas as pd

# Load sentiment stats and sample reviews
stats = pd.read_csv("product_sentiment_stats.csv")
top_reviews = pd.read_csv("top_reviews_per_product.csv")

PROMPT_TEMPLATE = """
Product: {product_name}
Total Reviews: {review_count}
Positive: {pos_pct:.1%}, Negative: {neg_pct:.1%}

Top positive review excerpt:
"{text_pos}"

Top negative review excerpt:
"{text_neg}"

Task: Write a 2-3 sentence summary highlighting:
1. Main strengths
2. Common issues
3. Overall recommendation or verdict.
"""
```

✓ Third Party LLM Setup (Mistral)

```
from huggingface_hub import login

login(token="*") # paste your Hugging Face token (removed for security)

import torch
from transformers import AutoTokenizer, AutoModelForCausalLM, pipeline

model_id = "mistralai/Mistral-7B-Instruct-v0.1" # gated model with access

tokenizer = AutoTokenizer.from_pretrained(model_id, use_auth_token=True)
model = AutoModelForCausalLM.from_pretrained(model_id, torch_dtype=torch.float16, device_map="auto", use_auth_token=True)

generator = pipeline(
    "text-generation",
    model=model,
    tokenizer=tokenizer,
    max_new_tokens=150,
    temperature=0.7,
    device_map="auto"
)
```

```

/usr/local/lib/python3.11/dist-packages/transformers/models/auto/tokenization_auto.py:902: FutureWarning: The `use_auth_token` argument is deprecated in favor of `use_auth_token` and will be removed in a future version. Please use `use_auth_token` instead.
warnings.warn(
/usr/local/lib/python3.11/dist-packages/huggingface_hub/utils/_auth.py:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://huggingface.co/settings/tokens), set it as the secret `HF_TOKEN` in your Colab secrets, and restart this notebook.
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public models or datasets.
warnings.warn(
tokenizer_config.json: 100% 2.10k/2.10k [00:00<00:00, 173kB/s]
tokenizer.model: 100% 493k/493k [00:00<00:00, 22.7MB/s]
tokenizer.json: 100% 1.80M/1.80M [00:00<00:00, 2.80MB/s]
special_tokens_map.json: 100% 414/414 [00:00<00:00, 45.9kB/s]
/usr/local/lib/python3.11/dist-packages/transformers/models/auto/auto_factory.py:476: FutureWarning: The `use_auth_token` argument is deprecated in favor of `use_auth_token` and will be removed in a future version. Please use `use_auth_token` instead.
warnings.warn(
config.json: 100% 571/571 [00:00<00:00, 60.3kB/s]
model.safetensors.index.json: 100% 25.1k/25.1k [00:00<00:00, 3.17MB/s]
Fetching 2 files: 100% 2/2 [00:36<00:00, 36.21s/it]
model-00002-of-00002.safetensors: 100% 4.54G/4.54G [00:21<00:00, 217MB/s]
model-00001-of-00002.safetensors: 100% 9.94G/9.94G [00:35<00:00, 461MB/s]
Loading checkpoint shards: 100% 2/2 [00:04<00:00, 1.94s/it]
generation_config.json: 100% 116/116 [00:00<00:00, 15.3kB/s]
Device set to use cuda:0

```

✓ Using LLM to generate summaries of Product

We see missing reviews for products, we will omit missing product reviews to generate a summary for products with reviews

```

missing = set(stats["name"]) - set(reviews["name"])
print("Products missing review entries:", missing)
stats = stats[stats["name"].isin(reviews["name"])]

```

```

Products missing review entries: {'Certified Refurbished Amazon Fire TV Stick (Previous Generation - 1st),,\nCertified Re

```

We make a good prompt which will help maintain consistency and help us generate summaries for the product based on the reviews

```

import pandas as pd

stats = pd.read_csv("product_sentiment_stats.csv")
reviews = pd.read_csv("top_reviews_per_product.csv")

PROMPT = """
Product: {name}
Reviews: {review_count}, 👍 {pos_pct:.0%}, 👎 {neg_pct:.0%}
Top Positive: "{text_pos}"
Top Negative: "{text_neg}"

Write a concise 1-2 sentence summary: key strength, main issue, and overall recommendation.
"""

summaries = []
for _, r in stats.iterrows():
    revs = reviews[reviews["name"] == r["name"]]
    if revs.empty:
        print(f"⚠️ Skipped '{r['name']}' - no review entry found.")
        continue

    rev = revs.iloc[0]
    prompt = PROMPT.format(
        name=r["name"],
        review_count=r["review_count"],
        pos_pct=r["pos_pct"],
        neg_pct=r["neg_pct"],
        text_pos=rev["reviews.text_pos"],

```

```

    text_neg=rev["reviews.text_neg"],
)
out = generator(prompt)[0]["generated_text"].strip()
summaries.append({"name": r["name"], "summary": out})

pd.DataFrame(summaries).to_csv("product_summaries.csv", index=False)
print("✅ Summaries generated and saved.")

🔗 Certified Refurbished Amazon Fire TV (Previous Generation - 1st),,, - no review entry found.
⚠️ Skipped 'Certified Refurbished Amazon Fire TV Stick (Previous Generation - 1st),,,
Certified Refurbished Amazon Fire TV Stick (Previous Generation - 1st),,, - no review entry found.
⚠️ Skipped 'Certified Refurbished Amazon Fire TV Stick (Previous Generation - 1st),,,
Kindle Paperwhite,,, - no review entry found.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
⚠️ Skipped 'Certified Refurbished Amazon Fire TV with Alexa Voice Remote,,,
Certified Refurbished Amazon Fire TV with Alexa Voice Remote,,, - no review entry found.
⚠️ Skipped 'Coconut Water Red Tea 16.5 Oz (pack of 12),,,
Amazon Fire Tv,,, - no review entry found.
⚠️ Skipped 'Echo (Black),,,
Amazon 9W PowerFast Official OEM USB Charger and Power Adapter for Fire Tablets and Kindle eReaders,,, - no review entry fo
⚠️ Skipped 'Echo (Black),,,
Echo (Black),,, - no review entry found.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
⚠️ Skipped 'Echo (White),,,
Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes Special Offers, Tangerine" - no review entry found.
⚠️ Skipped 'Echo Dot (Previous generation)' - no review entry found.
⚠️ Skipped 'Echo Spot Pair Kit (Black)' - no review entry found.
⚠️ Skipped 'Expanding Accordion File Folder Plastic Portable Document Organizer Letter Size' - no review entry found.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
⚠️ Skipped 'Fire HD 8 Tablet with Alexa, 8 HD Display, 16 GB, Tangerine - with Special Offers,' - no review entry found.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
⚠️ Skipped 'Fire HD 8 Tablet with Alexa, 8 HD Display, 32 GB, Tangerine - with Special Offers,' - no review entry found.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
⚠️ Skipped 'Fire TV Stick Streaming Media Player Pair Kit' - no review entry found.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
⚠️ Skipped 'Kindle Dx Leather Cover, Black (fits 9.7 Display, Latest and 2nd Generation Kindle Dxs),, - no review entry fou
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
⚠️ Skipped 'Kindle Keyboard,,,
Kindle Keyboard,,, - no review entry found.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
⚠️ Skipped 'Kindle Paperwhite,,,
Kindle Paperwhite,,, - no review entry found.
⚠️ Skipped 'Kindle PowerFast International Charging Kit (for accelerated charging in over 200 countries)' - no review entry
⚠️ Skipped 'Kindle Voyage E-reader, 6 High-Resolution Display (300 ppi) with Adaptive Built-in Light, PagePress Sensors, Fre
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
⚠️ Skipped 'New Amazon Kindle Fire Hd 9w Powerfast Adapter Charger + Micro Usb Angle Cable,,,
' - no review entry found.
Setting `pad_token_id` to `eos_token_id`:2 for open-end generation.
⚠️ Skipped 'Two Door Top Load Pet Kennel Travel Crate Dog Cat Pet Cage Carrier Box Tray 23" - no review entry found.
✅ Summaries generated and saved.

```

We now will save these summaries into a csv

```

pd.DataFrame(summaries).to_csv("product_summaries.csv", index=False)
print("✅ Completed - final summaries saved.")

```

🔗 ✅ Completed - final summaries saved.