Create a graphical user interface (ui)

Subtask:

Develop a simple web application using Streamlit to display the clustered data and the cluster visualization.

Reasoning: The subtask is to create a Streamlit application to display the clustered data and the visualization. I need to write the code for the Streamlit app to a file named `app.py`.

```
%%writefile app.py
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import streamlit as st

st.title("Store Clustering Dashboard")

df = pd.read_csv("store_clusters.csv")

st.write("### Clustered Store Data")
st.dataframe(df)

st.write("### Cluster Plot")
fig, ax = plt.subplots(figsize=(10, 6))
sns.scatterplot(data=df, x='pca_x', y='pca_y', hue='cluster', s=100, palette='Set2', ax=ax)
ax.set_title("Store Clusters (PCA Projection)")
st.pyplot(fig)
```

```
Overwriting app.py
```

Deploy the ui

Subtask:

Deploy the Streamlit application using ngrok for local testing in Colab.

Reasoning: Install ngrok and pyngrok, then authenticate ngrok and run the streamlit app to expose it via a public URL.

```
!pip install ngrok pyngrok
from pyngrok import ngrok
import os

NGROK_AUTHTOKEN = os.environ.get("2za7WNjHkTi0CaOsPQaVxHTlklY_6FnQkcNF6KHKZoKTL65CE")
if NGROK_AUTHTOKEN:
  ngrok.set_auth_token(NGROK_AUTHTOKEN)
  # Run the Streamlit app and expose it via ngrok
  !streamlit run app.py &>/dev/null&
  public_url = ngrok.connect(8501) # Streamlit's default port
  print(f"Streamlit app is running at: {public_url}")
else:
  print("i have added the secret key but it is failing to recognize")
```

```
Requirement already satisfied: ngrok in /usr/local/lib/python3.11/dist-packages (1.4.0)
Requirement already satisfied: pyngrok in /usr/local/lib/python3.11/dist-packages (7.2.11)
Requirement already satisfied: PyYAML>=5.1 in /usr/local/lib/python3.11/dist-packages (from pyngrok) (6.0.2)
i have added the secret key but it is failing to recognize
```

Summary:

Data Analysis Key Findings

- The dataset contains 4313 store entries with features including sales, cost, forecast, and PCA components (`pca_x`, `pca_y`), along with a pre-assigned `cluster` label.
- The `unit_cost_amount` column is entirely empty and does not contain any usable data.
- Store clusters have been visualized using PCA components, showing distinct groupings of stores based on their features.
- The clustered data and the KMeans model used for clustering were successfully saved to CSV and pickle files, respectively, for potential future use or deployment.
- A basic Streamlit application (`app.py`) was successfully created to display the clustered data and the cluster visualization.

Insights or Next Steps

- The lack of `unit_cost_amount` data might impact analyses requiring cost information; consider if this data can be sourced or imputed.
- To fully deploy the Streamlit UI, the `NGROK_AUTHTOKEN` must be added to the Colab secrets manager as instructed in the final steps of the process.