Assignment

- Create a Github Actions Workflow to run the application everyday at 5pm & store its run time log in the log file
- You can consider this as application.

```
import streamlit as st
import pandas as pd
import numpy as np

# Function to load data
@st.cache
def load_data():
    data = pd.read_csv('sales_data.csv')
    return data

# Function to transform data
def transform_data(df):
    df['Date'] = pd.to_datetime(df['Date'])
    df['Year'] = df['Date'].dt.year
    df['Month'] = df['Date'].dt.month
```

```
    df['Sales_to_Profit_Ratio'] = df['Sales'] / df['Profit']
    df['Cumulative_Sales'] = df['Sales'].cumsum()
    return df

# Function to extract summary statistics
def get_summary_statistics(df):
    summary = df.describe()
    return summary

# Main function to run the Streamlit app
def main():
    st.title("ETL Process for Sales Data")

    # Step 1: Extract
    st.header("Step 1: Extract")
    data = load_data()
    st.write("Raw Data")
    st.write(data)

    # Step 2: Transform
    st.header("Step
```

- Sales_data.csv File

```
Date,Product,Sales,Profit
2023-01-01,Product A,100,20
2023-01-02,Product B,150,30
2023-01-03,Product A,200,40
2023-01-04,Product C,250,50
2023-01-05,Product B,300,60
```