ENSF 692 Final Project Summary

Group 4

Ibrahim Khan
Danish Shahid

Project Overview

Our project analyzes the performance and relationships between three major asset classes: Oil, Gold, and the S&P 500. The notebook allows users to select a custom time range and then explores trends, returns, and correlations using Pandas and Matplotlib.

Rubric Requirements & How They Are Met

Stage 1: Dataset Selection

- We use three separate datasets (Oil, Gold, S&P 500), each from a different Excel/CSV file.
- The merged DataFrame has more than 200 rows.
- Our merged dataset would have had more than 10 columns if we had not optimized our joins and dropped unnecessary columns before merging.
- No data is hard-coded except for column names and original files are not modified.

Stage 2: DataFrame Creation

- Data is imported into Pandas DataFrames.
- At least two merge/join operations are performed to combine the datasets.
- Duplicate columns are removed after merging.
- A hierarchical index (Year, Month, Day) is created for the combined DataFrame.
- Data is sorted and checked for nulls.

Stage 3: User Entry

- The user is prompted for a start date, end date, and rolling period.
- Input is validated with try/except blocks and clear instructions.

- No hard-coded data values are used.
- Output is clearly labeled and presented as tables or sentences.

Stage 4: Analysis and Calculations

- The describe() method is used for aggregate statistics.
- At least two new columns are added (e.g., percent return, cumulative return).
- Aggregation, masking, and groupby operations are used (e.g., monthly averages, trend classification, overall_trend).
- A pivot table is created for market direction trends (e.g., overall_pivot_df).
- At least two user-defined functions are included (e.g., get_user_input, alpha_beta_plot).

Stage 5: Export and Matplotlib

- The final DataFrame can be exported to Excel with index and headers.
- Multiple plots are generated using Matplotlib and Seaborn (e.g., SMA, returns, heatmaps, regression).
- Plots can be saved as .png files.

Data Sources
Gold: Kaggle - Daily Gold Price Historical Dataset
Oil: Kaggle - Fuels Futures Data
S&P 500: Kaggle - S&P 500 Dataset

Citations:

Daily Gold Price Historical Dataset
F. Janjua, Daily Gold Price Historical Dataset, Kaggle, [Online]. Available:
https://www.kaggle.com/datasets/faisaljanjua0555/daily-gold-price-historical-dataset

Fuels Futures Data
G. Servera, Fuels Futures Data, Kaggle, [Online]. Available:
https://www.kaggle.com/datasets/guillemservera/fuels-futures-data

S&P 500 Dataset
YouCantTouchThis, S&P 500 Dataset, Kaggle, [Online]. Available:
https://www.kaggle.com/datasets/youcanttouchthis/s-p-500-dataset