**Project Instructions - Investment Portfolio**

**Project Overview**

You have been tasked by Timberline Financial to prepare a pipeline to analyze investment portfolios for their clients. Timberline Financials offers a variety of investment options including stocks, index funds, ETFs, and Forex. Market data will primarily be sourced from polygon.io but additional market data may be included from reliable external sources. Using the Polygon API, you must construct an ETL pipeline using Pandas to load historical and daily data into a SQL Server database. Your analysis should summarize each investment’s performance as well as its relative performance relative to the market. Utilize Matplotlib or Seaborn visualizations alongside Report Ready Tables in SQL to present your analysis.

**Core Technologies**

* Jupyter Notebooks
* Python – Pandas, Matplotlib, Seaborn, pyodbc
* Microsoft SQL Server

**Investment Portfolio**

Your portfolio analysis pipeline should be able to handle any level of investment, but as proof of concept your sample portfolio begins with an initial investment of $100,000 on January 1, 2023. The portfolio is diversified to include a minimum of five carefully selected stocks and one Exchange-Traded Fund (ETF) representing a broad market index. Additionally, to further enhance the portfolio's performance and compound returns, you may optionally include an additional investment of $1,000 monthly into some subset of the portfolio’s components. The portfolio's performance is tracked through the present date, with continuous analysis and adjustments made to optimize returns and manage risks effectively.

**Data Sources and Requirements**

* Primary Data Source: Polygon.io
* At least 5 total Stocks must be included in the portfolio
* At least one Exchange Traded Fund (ETF)
* Additional data may be sourced from any established reliable vendor to be included in the portfolio analysis as necessary. For example:
* Company earnings
* Economic indicators

**Data Analysis and Visualization**

**Data Quality and Validation**

* Data Cleaning: Handle missing values, outliers, and ensure consistency.
* Data Transformation: Normalize data formats, convert time zones, and aggregate data.
* Validation: Check data against known benchmarks and verify through summary statistics.

**Investment Metrics**

At a minimum the following metrics should be included in your analysis. Additional metrics may be included as you see fit.

* Portfolio
  + Total Return
  + Cumulative Return
  + Annualized Return
  + Volatility
  + Sharpe Ratio
* Stocks and ETFs
  + Total Return
  + Cumulative Return
  + Volatility
  + 10-Day and 100-Day Moving Averages Relative to Market Indexes
  + Sharpe Ratio
  + Beta

**Visualization**

* Data visualization should be done through the matplotlib and seaborn libraries
* Visualizations should include charts of relevant metrics and KPIs
* Visualizations should be included in the final PowerPoint presentation

**Deliverables**

The demo portfolio serves as a proof of concept, demonstrating the effectiveness and versatility of the ETL pipeline and analysis framework. The deliverables for this project are designed to showcase the capabilities of the pipeline and provide a comprehensive analysis of the investment portfolio. The pipeline should be adaptable and capable of processing any selection of stocks or index funds. The deliverables include:

**PowerPoint Presentation:**

* A detailed presentation summarizing the entire project.
* Overview of Portfolio
* Introduction to the selected stocks and ETF included in the portfolio.
* Explanation of the initial investment strategy and (optionally) regular monthly investments.
* Overview of the ETL Pipeline
* Description of the ETL (Extract, Transform, Load) process, including data sources, cleaning, transformation, and loading into the SQL database.
* Steps taken to ensure data quality, consistency, and accuracy.
* Mock Dashboards
* Visual representations of the performance for each individual investment and the entire portfolio.

**MSSQL .bak File:**

* A backup file of the Microsoft SQL Server database containing all the cleaned and transformed data.
* Ready for querying and further analysis, allowing for easy replication and validation of results.

**Jupyter Notebooks:**

* Comprehensive notebooks documenting the entire data processing and analysis workflow.
* Data Extraction and Cleaning
* Code for extracting data from the various sources and cleaning it to ensure it is in a usable format.
* Data Transformation and Loading
* Scripts for transforming the data into the required format and loading it into the SQL database.
* Data Analysis and Visualization
* Detailed analysis using Python libraries such as Pandas, NumPy, Matplotlib, and Seaborn.
* Creation of visualizations to illustrate key insights and performance metrics.