adityamhaske / **SP23_MGMT_Project** Public

SP23 MGMT Course - Capstone Project

0 stars 0 forks

Star Watch

Code Issues Pull requests Actions Projects Wiki Security Insights Se

main

adityamhaske Update README.md ... now 8

View code

README.md

Financial Data Analysis using MongoDB and Pandas

This project involves querying financial data stored in a MongoDB database, and using the Pandas library to perform various analysis tasks, such as calculating stock returns, volatility, and correlation.

Google Colab Link for interactive Charts :
https://colab.research.google.com/drive/1OdqbH67mdPbOkEKC8y31wl6x0QlETwGU?usp=sharing

Github Repository Link: https://github.com/adityamhaske/SP23_MGMT_Project

Requirements

To run this project, you'll need:

- Python 3.x
- The following Python packages: pandas, pymongo, matplotlib, seaborn

Getting started

- Clone the repository to your local machine.
- Install the required Python packages using pip install -r requirements.txt.
- Make sure that you have a MongoDB instance running on your local machine. You can download MongoDB here.
- Download the Apple stock dataset in CSV format from here and place it in the data directory.

Project structure

The project is structured as follows:

- 'data': Contains the CSV file with the Apple stock data.
- 'notebooks': Contains the Jupyter notebooks used for data analysis.
- 'scripts': Contains Python scripts used for cleaning and preprocessing the data.

Data cleaning

Before performing any analysis, the financial data is cleaned and preprocessed using Pandas. This involves:

- Dropping unnecessary columns
- Renaming columns
- Converting date strings to datetime objects
- Filling missing values

Data exploration

Various visualizations are created to explore the data, including:

- Line plots showing stock prices over time
- Histograms showing daily stock returns
- Box plots showing stock returns by year
- Heat maps showing the correlation matrix of the data

Data analysis

Using Pandas and MongoDB, various analysis tasks are performed, including:

- Calculating daily stock returns and volatility
- Calculating moving averages
- Calculating the correlation matrix of the data
- Storing the data in a MongoDB database

Usage

1. Start by importing the necessary libraries and connecting to the MongoDB database.
2. Clean and preprocess the financial data by removing missing values and calculating additional columns.
3. Explore the data using descriptive statistics and visualizations.
4. Perform various analysis tasks such as calculating stock returns, volatility, and correlation.
5. Use interactive visualizations to better understand the data.
6. Perform CRUD operations and data pipeline operations using MongoDB.
7. Store the results of the analysis in a MongoDB collection for future use.

Conclusion

This project demonstrates how to use Pandas and MongoDB to analyze financial data, including cleaning and preprocessing the data, exploring the data with visualizations, and performing various analysis tasks. By storing the data in a MongoDB database, it is easy to perform further analysis and integrate with other applications.

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

- Jupyter Notebook 100.0%