Group 2 Project Report

Valense Acquah-Louis, Sam Caldwell, Logan Miles

EXTRACT:

Both datasets used in this project were obtained from kaggle.com.
Data sources:

- Complete Historical Cryptocurrency Financial Data
  https://www.kaggle.com/philmohun/cryptocurrency-financial-data

- Stock Exchange Data
  https://www.kaggle.com/mattiuzc/stock-exchange-data

The dataset was pulled and saved into a folder named Resources in a CSV format. Python/pandas were used to consume the data and present the outcome. This was done by importing the necessary dependencies. These dependencies were:

Import os (to import the operating system), import csv (to import the csv files) and import pandas as pd (to import pandas in the form pd).

After which the two csv files were converted to a pandas data frame and read.

TRANSFORM:

Summary statistics were performed for each table.

The two data sources had different date ranges, the Complete Historical Cryptocurrency Financial Data was from the year 2013 to 2019 and the Stock Exchange Data from 1965 to 2021. The dates prior and after the crypto currency financial dataset were dropped from the stock exchange dataset to align the date ranges. To do this the earliest and latest dates in the crypto currency dataset were determined and used for the exclusions. Next the date formats were harmonized.

Renamed the currency column of the complete historical cryptocurrency financial dataset to index to match the stock exchange dataset. Additionally, tags of 's' and 'c' were added to separate currency indices from stock indices.

Our two disparate tables were concatenated into one, and the redundant timestamp column was removed.

A data frame of primary keys was created, and the ID's from the three tables were removed, as that data was now stored elsewhere.

LOAD:

From SQLAlchemy 'create_engine' was imported, as well as psycopg2 and the password from the config file.

Next an attempt to connect to postgres is made, with a conditional failure print. After the connection is made and a cursor created, a check is run to see if 'stock_crypto_db' exists, and the user is informed whether the database already exists, or was created.

Another connection is made, this time with the intent of adding our tables to the database. First we run a check to see if any tables are already in the database, and then one by one add our tables to the database, printing after each step to ensure proper loading.

POTENTIAL USES

1)   Ease of access: Cryptocurrency is a peer-to-peer contract with less paperwork as opposed to regular bank red tape, this makes it transparent and dependent on the availability instead of solely on customer sentiment.

2)   Spot trends: Helps the end users determine areas where customers are spending, what they are buying and also what they are investing in. This can help investors potentially alleviate some of the inherent risks associated with investing in the notoriously volatile cryptocurrency market.