Chenhui Zheng

Lingyu Li

**Bitcoin Comparisons Project**

**Project Background**

The purpose of this project is to compare bitcoin prices from 2014 to 2019. In addition, our team has added a third dataset to show the correlation of bitcoin prices between Euro as well as Gold prices. In order to set a consistent standard, all prices are denoted in U.S dollars.

**Project Steps:**

**Process 1: Extract Data**

To achieve our goal, we have gathered three different datasets (Bitcoin dataset, Coinbase dataset, and Gold dataset) from Kaggle. The sources of these datasets are listed below:

Dataset Sources:

<https://www.kaggle.com/altruistdelhite04/gold-price-data>

<https://www.kaggle.com/mczielinski/bitcoin-historical-data>

**Process 2: Transform Data**

**Step 1: Remove NaN**

Though both the Bitcoin and Coinbase dataset are over 25 MB, much of the information inside are NaN, so the first step we did is to remove all the NaN cells with the code #pd.dropna() command.

**Step 2: Scoping**

To further clean up the bitcoin data, we have removed some information that are not crucial to our comparison projects, which include: open, high, low, volume and weighted price. Instead we narrow the dataset down to date and closing price of that day.

On the other hand, we have also narrowed down the Goldprice.csv data to only include Gold price and Euro currency exchange rate.

**Step 3: Date Modification and Drop Duplicated Columns**

Step 3 is the hardest part of this project because not only different dataset has different date range but also different data types and data format. Therefore, we have created #translate\_function and #translate\_to function to clean up the date time to "%Y-%m-%d” format. Moreover, we’ve only kept the first closing price of each day to make sure we don’t receive duplicated values.

**Step 4: Rename Columns**

Due to the ambiguity of the naming we have named all columns in a consistent way.

**Process 3: Load Data**

In the final process, we have created a SQL database bitcoin\_comparison\_db as well as the table bitcoin\_comparisons. The final table demonstrated the fluctuation of bitcoin, gold, euro, in relation to U.S Dollars from 2014 to 2018.