Extract

The original data source for both the GPU and CPU prices were CSV files from Kaggle.

Transform

CPU: The CPU data was imported into a Pandas dataframe. The recommended price column was refined to drop rows with no data and with price ranges since we were only looking for data with one starting price. This column was then modified to become a float and to drop the dollar sign and commas within the numbers. Finally the data frame was narrowed down to only include types of data that we found most useful for the future analysis. These were Collection Name, Vertical Segment, Processor Number, Launch Date, Recommended Price, Number of Cores, Number of Threads, Max Memory Size, and max memory bandwidth.

FINAL: In order to create the final table we would be uploading, a second dataframe was created from each dataframe (Cryptocurrency, CPU, and GPU) that grouped the average price by release quarter. These will then be merged to create the final table.

Load