Proposal

we are going to extract data sources from kaggle (sunspot data, bitcoin data, tesla stock data and gas price data) we will transform the data to remove unnecessary columns and put all dates in the same format we will then load the data to a sql database

Technical Document

Data Sources

​ Sunspots Data: <https://www.kaggle.com/abhinand05/daily-sun-spot-data-1818-to-2019>

​ Gas Price Data: <https://www.kaggle.com/mruanova/us-gasoline-and-diesel-retail-prices-19952021>

​ Tesla Stock Data: <https://www.kaggle.com/cosmosankur/tesla-stock-price-complete-data>

​ Bitcoin Data: <https://data.world/crispydata/bitcoin-historical-data>

Extract

For our project we selected 4 sources of data from the data sources above, while we were hoping to use a wider range of data formats, the data we found interesting only came in csv.

Transform

* Tesla data came in a csv and originally had 7 columns of which we used the Date, Open price, and Closing price. We had to reformat the Date column as pandas was reading as an object and we were planning on joining the data sets on the date column and had to make sure they would all match. We also wanted to see the daily change and where able to accomplish that easily in pandas.
* Gas Prices data only needed little to no cleaning.
* Sunspot data, however, proved to be a lot more finicky, as the dates were split up into 3 separate columns and we spent the longest time in this project trying to figure out how to convert those three columns into the proper datetime format which was finally solved with an impressive display of ‘google-fu’ befitting of a blue belt. We used the pd.to\_datetime function.
* Bitcoin data was again not too much of an issue, we dropped the column we didn’t need, along with null values within the price column. We had to change the date column from an object type to a datetime type. We also renamed columns to sync them up with our database table in postgress.

Load  
We created a database called ETL\_project to store our tables using pgAdmin. We created 4 tables for our database and mapped the data from our various sources to the tables. The tables are listed belowSunspots

* Date - primary key
* Total\_sunspots - INT
* Observations - INT

Gas Prices

* Date - primary key
* Gas Price - Float

Tesla Stock

* Date - primary key
* Open - float
* Close - float
* Daily\_change - float
* Volume - INT

Bitcoin

* Date - primary key
* Price - float
* Exchange volume - INT
* Generate - INT

We chose this data to see what the relation is between Tesla’s rise in market value and the undying increase in gas prices. We also thought it would be interesting to see if there was any correlation between Bitcoin and Tesla, and since Mr. Musk has a hard time not sharing his love of crypto currency on Twitter, perhaps he knows something we do not. The sunspots was chosen for a laugh, and because we though it would be interesting.