**Project 2**

Group 9 Team Members:

Chad Dubiel, David Martinez, Katy Fuentes

Scope of Research:

Correlation between cryptocurrency pricing and Covid case counts.

Github Repo:

<https://github.com/cdubiel08/ETL-Project-Group-9>

Data Sources:

* API Key- SFOX <https://www.sfox.com/developers/?python#market-data> or Gemini <https://docs.gemini.com/websocket-api/#market-data>
* Source
  + Covid- <https://www.kaggle.com/imdevskp/corona-virus-report>
    - Full\_grouped.csv (date, country, confirmed)
  + Cryptocurrency Historical Chart- <https://www.kaggle.com/mczielinski/bitcoin-historical-data>
    - (Timestamp, close?, or average open/close or average high/low)

Other:

What useful investigation could be done with the final database?

Use the output and compare to markets, commodities, or US dollar.

Whether final database will be relational or non-relational. Why?

Relational because the information will be interconnected based on a timeframe.

Considerations: Dates not a good join method, need a unique ID for primary key

Data Analysis

* Pandas for data formatting- date cleaning, reduce columns,
* Database- Mongo better for skipping null values which would skip data column, any covid/crypto overlaps captured.

Steps:

Data Sources:

- At least 2 (or more) sources

- If possible, try to incorporate a web API as one of your data sources.

ETL Process:

- Within Jupyter, build out the ETL process to extract your data from their sources, apply some level of transformation, and

load the resulting data to a database (relational or non-relational)

Flask API:

- Build a Flask application that has a route that will execute a query to your database and return the results in JSON format.

Final Report:

- Write up a short report that details your 3 ETL steps.

- More details on a later slide.

Github Repo:

- Store all of your project files in a well-organized project repository

- Each member of your team will submit a link to your project repo to BCS by the end of class Tuesday

Write Up Process Summary:

● What data sources you chose, and why?

● Detailing the process of the extraction, transformation, and loading steps

● Explain why you have performed the types of transformation you did

● Why you chose the type of final database

● Schema of the tables/collections in the final database

● Hypothetical use case(s) for your database