**Team members:**

**Tony Gardella**

**Emily Keymon**

**Assignment:**

We are Data Engineers at SpaceX. We were tasked to provide data to ensure the media coverage for our space launches was effective. Therefore, bringing in more investors.

**Extract:**

**Data Sources:**

**Next Space Flight: https://nextspaceflight.com**

\* This dataset contained flight information

\* Started with 9 columns and 4,324 rows of data

**HuffPost: https://www.huffpost.com/**

\* This dataset contained news stories

\* Started with 6 columns and 200,853 rows of data

**Transform:**

**Data source one was a csv file. Pandas library was used for transformation.**

\* pd.read\_csv was used to read the file into Jupyter Notebook

\* df.drop was used to drop extra columns

\* df.value\_counts was used to review the amount of data collected

\* df.loc was used to filter out SpaceX flights

\* df.to\_csv was used to save the completed data frame as a csv file

**Data source two was a json file. Pandas library was used for transformation.**

\* pd.read\_json was used to read the file into Jupyter Notebook

\* df.str was used to search for SpaceX news

\* df.value\_counts was used to review the amount of data collected

\* df.Index was used to index the data

\* df.to\_csv was used to save the completed data frame as a csv file

\* pd.concat was used to concatenate the two data frames together

**Load**

The last step was to transfer our final output into a database. We created a database and respective table to match the columns from the final Panda's Data Frame using MongoDB. We chose this non-relational database because it proves to be ideal for storing data that may be changed frequently.

\* PyMongo was used to work with MongoDB

\* PyMongo.MongoClient was used to establish the connection

\* client.db was used to declare the database

\* db.collection.insert\_many was used to add the data frame to the database

\* df.to\_dict was used to convert the records to dictionary for storage in the database

**Final Tables**

\* Company Name

\* Location

\* Date

\* Detail

\* Status Rocket

\* Rocket

\* Status Mission

\* Category

\* Headline

\* Authors

\* Link

\* Short Description