**ETL Process Report**

***Extraction:***

-Extracted data from two different csv files (sources below) using pandas read\_csv

<https://www.kaggle.com/mdabbert/ultimate-ufc-dataset?select=ufc-master.csv>

<https://www.kaggle.com/mdabbert/ufc-fights-2010-2020-with-betting-odds>

***Transformation:***

- created a new column “New\_ID” for both datasets

-Copied columns from ufc betting data that were relevant and renamed the column as well – Red\_Corner\_Fighter, Blue\_Corner\_Fighter, Red\_Fighter\_odds, Blue\_Fighter\_odds, Winner, Red\_profit\_on\_a\_100\_credit\_winnin\_bet, "Blue\_profit\_on\_a\_100\_credit\_winnin\_bet"

-combined the two datasets into one dataset

-dropped the rows with NaN values

***Loading***

-Created database connection to postgres using SQLAlchemy ‘create\_engine’

-Confirmed table names from ETLProject\_db database created in postgres (see separate sql documentation for scheme set up)

-Loaded final pandas dataframes using ‘to\_sql’ method and connection created by SQLAlchemy dependency

**Fight ID:**

**Primary Dataset:**

Yes