ETL Project

Used major marathon data joined from multiple datasets collected from the Abbot World Major Marathon men's winners. Most of the 6 separate csv files were consistent with the same information. However, there were some standardization needed for key names before concatenating the lists. Such as country vs. national or country vs. state/province. Time keys were needed to all be changed from variations and note columns that need to be dropped. The data was then loaded into non-relational Mongo DB using pymongo and Jupyter notebook. Struggled with converting the list to JSON. This was an action we did not need to do as the data was in the correct type already. This created a string which which could not be inserted into MongoDB.