Project: ETL for Breast Cancer Dataset

Dataset: <https://www.kaggle.com/iabhishekbhardwaj/breast-cancer-prediction>

Instructions:  
1. I need to be able to separate out patients based on the mean of tumor radius   
for a research project.  
2. I’d like the final datasets to be in CSVs, and I’d like 5 different CSVs.

3. Create five json files from the created CSV files

4. Load json files into MongoDB

Steps and Time Resources:

1. Reviewed dataset.
   1. Found parent dataset/analysis in Kaggle and its documentation – 1   
      hours.
   2. Reviewed documentation and obtained original medical journal article   
      – 0.5 hours.
   3. Wrote up summary – 0.25 hours.

2. Communicated findings with supervisor – 0.25 hours.

3. Split and exported dataset per supervisor’s instructions – 1.35 hours.

Total man hours: 3.55 hours.