Data Cleaning Report

- Checked Errors
    - Empty price, size, or timestamp
    - Missing or invalid data entry
    - Negative price or size
    - Duplicate entries
    - Outliers

- Methodology
    - Given that the dataset was extremely large (1.76 million values) I simply removed any empty, or invalid fields as this would scarcely effect the results. Retrospectively, creating an algorithm to fill in these values via prediction may have been better practice.
    - As far as duplicates, I simply did not add any non initial instance of a given row to my dataset of cleaned values. In other words, before a row was added to my cleaned set, it was vetted to ensure that it was not already present.
    - Outliers were simply removed from the data, and to achieve this I used a simple IQR calculation. If any value was below Q1-1.5(IQR) or above Q3+1.5(IQR) I deemed it an outlier and removed it from the data.