0.1 Engineering a “usage” feature

The dublin bike dataset used for this analysis consists of snapshots of station occupancy at five-minute intervals. Because of this it is not possible to determin from the dataset the exact number of times bikes have been returned to or taken from a station. We can look at the difference in bike occupancy at the beginning and end of a five minute interval to establish a *lower-bound* on the number of interactions with the station, but it is possible for there to have been more interactions (returns or borrowings) than this difference. For instance, it may be that a bike station has 10 bikes at time t and 10 bikes at time $t + 5$m, but that in fact 5 bikes were borrowed and 5 bikes were deposited.

But, we can detect such cases, because the dataset also includes a LAST UPDATED field. If the occupancy is not changed, but the time of last update is between the start and end of the interval, then we know that there were at least 2 interactions with the bike station.