1. the URLs which is used in both dataset, both belong to same website