Exploratory data analysis has shown that the name doesn’t uniquely identify the site since there are a few site with multiple addresses (e.g. post codes) and each address has it’s own set of menu items.

For instance, the site name Bertucci's appears with 4 different addresses and although three of them appear to be of the same franchise having similar menu items with the same description and price, one of the three is missing the Margherita Pizza. Moreover the 4th one, the one in Illinois seems to be a completely different restaurant since it only has one menu item that is different from the menu items of the other 3 Bertucci’s.

The above observations have highlighted the need to combine the name with at least one more column before we can create a new instance of a site otherwise the data would incorrectly assume that because the name is the same all 4 Bertucci's sites are actually the same site.

Also it has rendered the assumption that all sites sharing the same name are part of the same franchise/chain incorrect.

Currency: even though the currency logically belongs with the price, we are making a modelling assumption that it actually is a property of the site, since a site cannot be selling items at different currencies. An even more generic assumption would be to make the currency a property of the country the site is in but since we only have data from the US and all currencies where specified are USD we don’t have more evidence that that assumption would be correct.