Merging and Joining datasets

The advent of relational database management is essentially based the concept of being able to combine datasets. Having separate datasets containing different data but somehow being 'linked' together allows for many advantages in management of the datasets. For example, having a dataset containing the longitude and latitude of a house. Once converted to zip code, it yields one or more houses with the same zip code but different longitude and latitude. In this case, for one house, we can find information about all the other houses in the same zip code. This is a one-to-many relationship. The relationship between two datasets can be many-to-one, many-to-many or even one-to-one. In all cases, we can combine them and query the results (i.e. for house1 in longitude = 12345678, latitude = 98765431, list the price, number of bedrooms, number of bathrooms, square footage of all neighboring homes listed.

References

Combining Datasets: Merge and Join -

https://jakevdp.github.io/PythonDataScienceHandbook/03.07-merge-and-join.html