EDA notes – beer\_reiews.csv

Nulls in:

brewery\_name 1586599 non-null object

review\_profilename

beer\_abv

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Column** | **Description** | **API Expected Parameter** | **Nulls?** |  |  |
| **brewery\_id** | Identifier of brewery | No | No |  |  |
| **brewery\_name** | Name of brewery | **Yes** | **Yes** |  |  |
| **review\_time** | Timestamp of review | No | No |  |  |
| **review\_overall** | Overall score given by reviewer | No | No |  |  |
| **review\_aroma** | Score given by reviewer regarding beer aroma | **Yes** | **No** |  |  |
| **review\_appearance** | Score given by reviewer regarding beer appearance | **Yes** | **No** |  |  |
| **review\_profilename** | Profile name of reviewer | No | Yes |  |  |
| **review\_palate** | Score given by reviewer regarding beer palate | **Yes** | **No** |  |  |
| **review\_taste** | Score given by reviewer regarding beer taste | **Yes** | **No** |  |  |
| **beer\_style (target)** | Type of beer | No | No |  |  |
| **beer\_name** | Name of beer | No | No |  |  |
| **beer\_abv** | Alcohol by volume measure | **Yes** | **Yes - ~4.3%** |  |  |
| **beer\_beerid** | Identifier of beer | No | No |  |  |

There are 104 different beer\_styles (the modelling target) and 5742 different breweries (brewery\_name is required as input to the model).

Many beer styles and beer\_style – brewery\_name combinations are very rare.

**Decisions:**

Brewery\_name is an api expected parameter but it does not need to be used in making predictions. There are 5742 breweries.