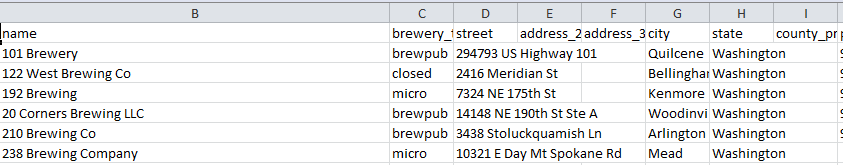
**Mile Stone 1**

**Data Set:**



**API:**

<https://www.openbrewerydb.org/documentation>

<https://api.openbrewerydb.org/breweries/>

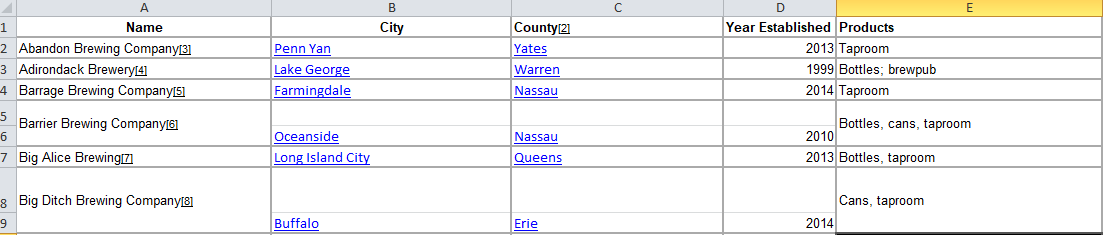


**Website:**

<https://en.wikipedia.org/wiki/List_of_breweries_in_the_United_States>

<https://en.wikipedia.org/wiki/List_of_breweries_in_Maryland>

<https://en.wikipedia.org/wiki/List_of_breweries_in_New_York>



1. Data is about the list of breweries located in United States of America with all the details like address, what type of brewery and online website link.
2. All the data sets have a common column **name** (name of the brewery) which links them together.
3. Milestones needs data be cleansed extracted from all the sources and then data has to be inserted into a database and also can be used to visualize.

* Dataset from flat file is having all the information and needs to be checked for duplicates and some bad data.
* Data from API source needs to extracted and converted into a readable csv format for the states that we need
* Data from Wikipedia source is not in a proper format, we have to extract that as a tabular format and also needs some additional columns to be added like the address details, those were provided as a hyperlink which I think has to be handled manually.

1. Once all the above steps are done, I feel data sets will be ready to be inserted into Database