**Final Project Dataset**

1. Names of each group member.

Lin Meng, Matt Sarmiento, Qingyi Sun

1. Link to dataset.

https://www.kaggle.com/c/ashrae-energy-prediction/overview

1. Number of observations.

1048576

1. Number of raw features.

15

1. Sample row of dataframe with column names.

train.csv

|  |  |  |  |
| --- | --- | --- | --- |
| building\_id | meter | timestamp | meter\_reading |
| 0 | 0 | 2016/1/1 0:00 | 0 |
| 1 | 0 | 2016/1/1 0:00 | 0 |
| 2 | 0 | 2016/1/1 0:00 | 0 |

building\_metadata

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| site\_id | building\_id | primary\_use | square\_feet | year\_built | floor\_count |
| 0 | 0 | Education | 7432 | 2008 |  |
| 0 | 1 | Education | 2720 | 2004 |  |
| 0 | 2 | Education | 5376 | 1991 |  |

weather\_train.csv

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| site\_id | timestamp | air\_temperature | cloud\_coverage | dew\_temperature | precip\_depth\_1\_hr | sea\_level\_pressure | wind\_direction | wind\_speed |
| 0 | 2016/1/1 0:00 | 25 | 6 | 20 |  | 1019.7 | 0 | 0 |
| 0 | 2016/1/1 1:00 | 24.4 |  | 21.1 | -1 | 1020.2 | 70 | 1.5 |
| 0 | 2016/1/1 2:00 | 22.8 | 2 | 21.1 | 0 | 1020.2 | 0 | 0 |