CAPSTONE 2

Ashrae Energy Predictor

Problem statement

Q: How much does it cost to cool a skyscraper in the summer?
A: A lot! And not just in dollars, but in environmental impact.

Input Data

- Training Data

building_id - Foreign key for the building metadata.

meter - The meter id code. Read as {0: electricity, 1: chilledwater, 2: steam, 3: hotwater}. Not every building has all meter types.

timestamp - When the measurement was taken

meter_reading - The target variable. Energy consumption in kWh (or equivalent). Note that this is real data with measurement error, which we expect will impose a baseline level of modeling error. UPDATE: as discussed here, the site 0 electric meter readings are in kBTU.

	building_id	meter	timestamp	meter_reading
0	0	0	2016-01-01 00:00:00	0.0
1	1	0	2016-01-01 00:00:00	0.0
2	2	0	2016-01-01 00:00:00	0.0
3	3	0	2016-01-01 00:00:00	0.0
4	4	0	2016-01-01 00:00:00	0.0

- Building Metadata

site_id - Foreign key for the weather files.

building_id - Foreign key for training.csv

primary_use - Indicator of the primary category of activities for the building based on EnergyStar property type definitions

square_feet - Gross floor area of the building

year_built - Year building was opened

floor_count - Number of floors of the building

	site_id	timestamp	air_temperature	cloud_coverage	dew_temperature	precip_depth_1_hr	sea_level_pressure	wind_direction	wind_speed
0	0	2016-01-01 00:00:00	25.000000	6.0	20.00000	NaN	1019.5	0.0	0.000000
1	0	2016-01-01 01:00:00	24.406250	NaN	21.09375	-1.0	1020.0	70.0	1.500000
2	0	2016-01-01 02:00:00	22.796875	2.0	21.09375	0.0	1020.0	0.0	0.000000
3	0	2016-01-01 03:00:00	21.093750	2.0	20.59375	0.0	1020.0	0.0	0.000000
4	0	2016-01-01 04:00:00	20.000000	2.0	20.00000	-1.0	1020.0	250.0	2.599609

- Weather Data

site_id

air_temperature - Degrees Celsius

cloud_coverage - Portion of the sky covered in clouds, in oktas

dew_temperature - Degrees Celsius

precip_depth_1_hr - Millimeters

sea_level_pressure - Millibar/hectopascals

wind_direction - Compass direction (0-360)

wind_speed - Meters per second

	site_id	building_id	primary_use	square_feet	year_built	floor_count
0	0	0	Education	7432	2008.0	NaN
1	0	1	Education	2720	2004.0	NaN
2	0	2	Education	5376	1991.0	NaN
3	0	3	Education	23685	2002.0	NaN
4	0	4	Education	116607	1975.0	NaN

Data Cleaning

```
inspect_missing(train_data)
```

	Total	Percent
building_id	0	0.0
meter	0	0.0
timestamp	0	0.0
meter_reading	0	0.0
meter_type	0	0.0

```
inspect_missing(weather_train)
```

	Total	Percent
site_id	0	0.000000
timestamp	0	0.000000
air_temperature	55	0.039350
dew_temperature	113	0.080845
wind_speed	304	0.217496
wind_direction	6268	4.484414
sea_level_pressure	10618	7.596603
precip_depth_1_hr	50289	35.979052
cloud_coverage	69173	49.489529

```
inspect_missing(building_metadata)
```

	Total	Percent
site_id	0	0.000000
building_id	0	0.000000
primary_use	0	0.000000
square_feet	0	0.000000
year_built	774	53.416149
floor_count	1094	75.500345

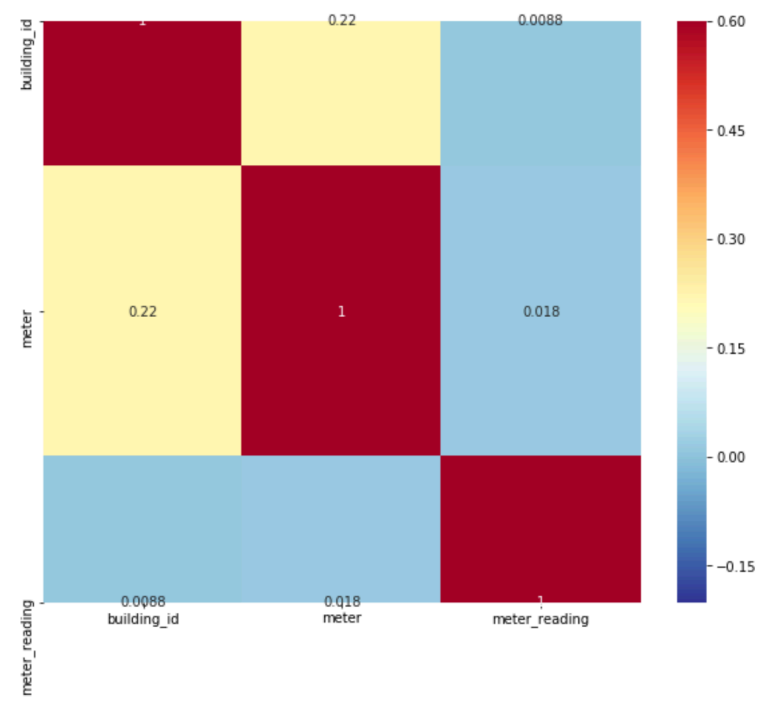
Data Interpolation

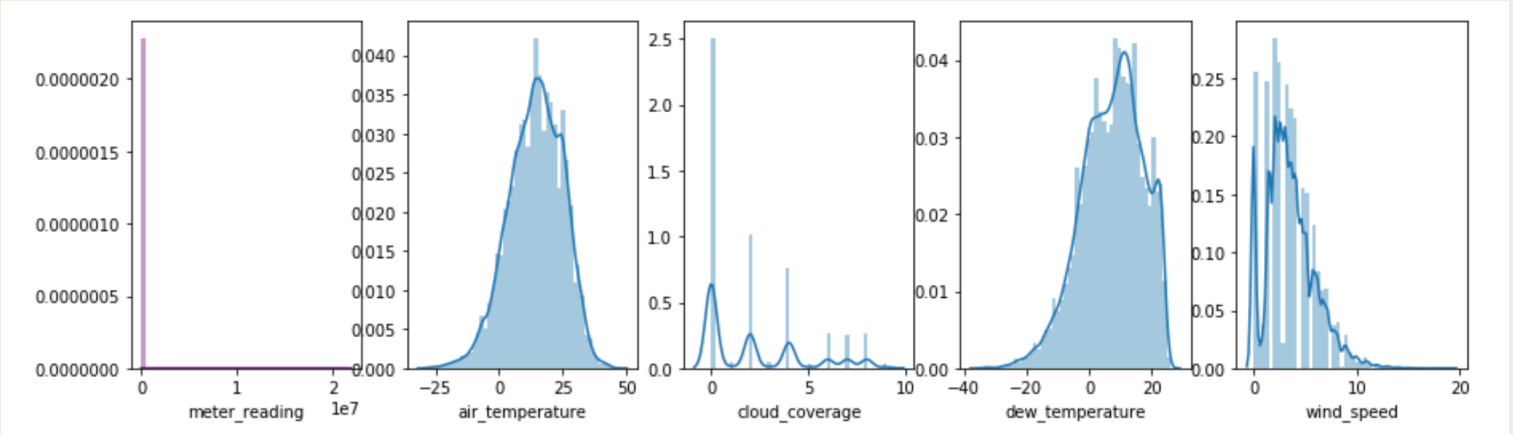
Interpolating Weather Data

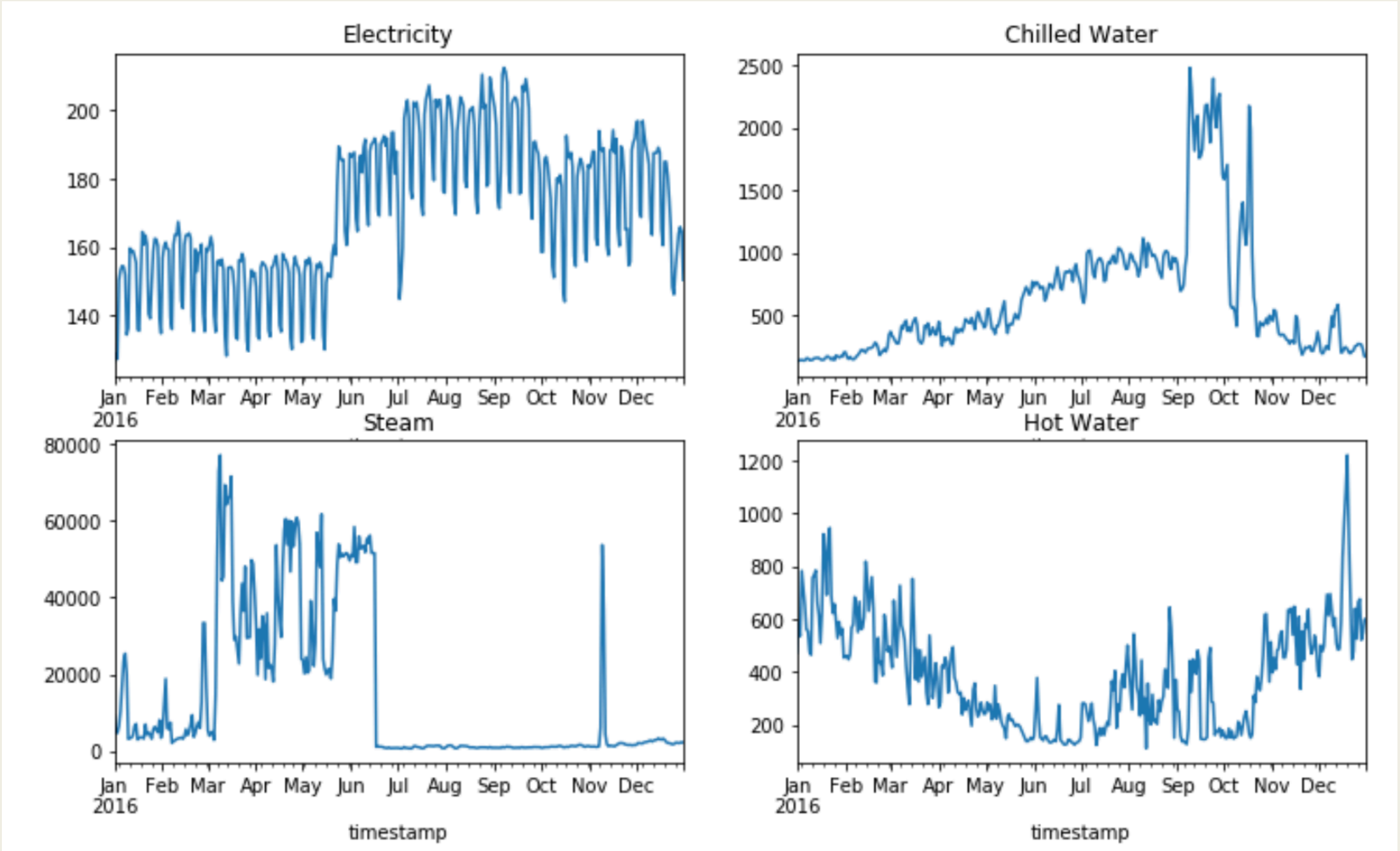
```
weather_train = weather_train.groupby('site_id'
weather_train.groupby('site_id').apply(lambda g
```

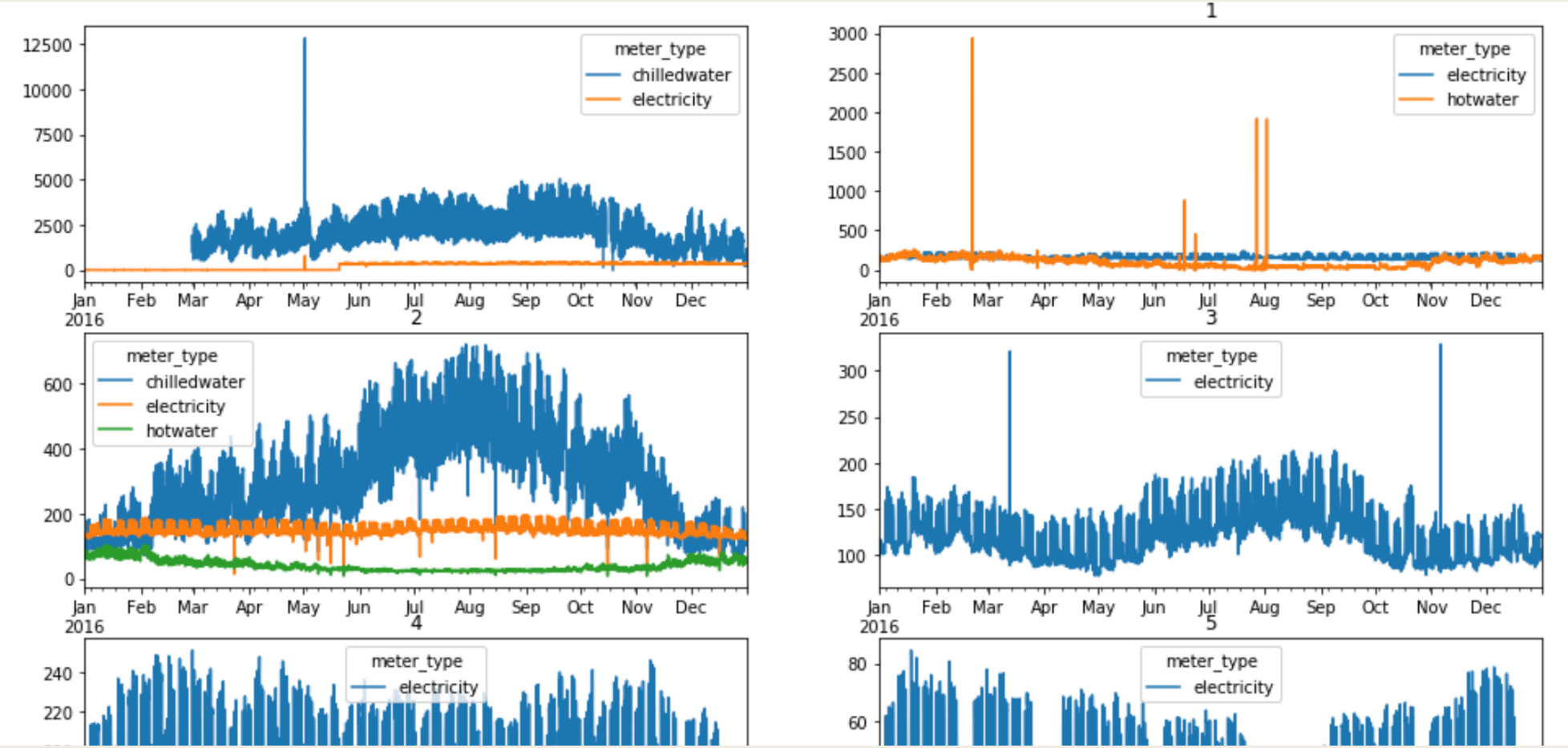
site_id	site_id	timestamp	air_temperature	cloud_coverage	de
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	

EDA


building_id
meter
meter_reading
0.22
0.0088
0.22
1
0.018
0.0088
0.018
1
building_id
meter
meter_reading
0.60
0.45
0.30
0.15
0.00
−0.15


meter_reading
1e7
air_temperature
cloud_coverage
dew_temperature
wind_speed


Electricity
200
180
160
140
Jan
2016
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Chilled Water
2500
2000
1500
1000
500
Steam
80000
60000
40000
20000
0
timestamp
Hot Water
1200
1000
800
600
400
200


meter_type
chilledwater
electricity
12500
10000
7500
5000
2500
0
Jan
2016
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
1
meter_type
electricity
hotwater
3000
2500
2000
1500
1000
500
0
2
meter_type
chilledwater
electricity
hotwater
600
400
200
0
3
meter_type
electricity
300
250
200
150
100
4
meter_type
electricity
240
220
5
meter_type
electricity
80
60

Outlier Detection