Table Review

There are three tables, one captures home water consumption, another captures home electricity consumption and solar panel electricity production and another one has weather data. They all are aggregated to hourly readings during calendar years 2017 - 2020 at 109 Austin homes. There are 99 single family homes and 10 townhomes.

Homes have varying amounts of electric and water data over the four year period. Gaps in the data are caused by sensor connectivity issues or erroneous sensor readings being removed. The logic for data inclusion was if a house had water data for more than 3 months it was included and paired with all available electricity consumption and solar production data for any of the years it had water data. Thus, some homes have 4 years of electricity of water data, while others just have a few months of water data.

Water Data Table captures water data at one hour time intervals.

hourly_time - (Timestamp with timezone) Hourly Interval of data.

home_id - (String) Each home_id represents one housing unit.

hourly_gal - (Decimal) The amount of water in gallons that was used since the last meter read. Data is reported at an hourly interval. If there is a gap in the data there are two possible explanations for the reading immediately following a gap in data.

1. **Case #1** If our sensor only lost communication this will display the sum of water consumed during the entire period the sensor was not reporting.

Case #1 - Meter Records Data But Temporarily Loses Communication

HYPOTHETICAL CONSUMPTION			DATASET REPRESENTATION		
hourly_time	home_id	gal_water_ consumed	hourly_time	home_id	gal_water_ consumed
4/26/18 0:00	668	0.00	4/26/18 0:00	668	0.00
4/26/18 1:00	668	0.00	4/26/18 1:00	668	0.00
4/26/18 2:00	668	1.87	4/26/18 2:00	668	1.87
4/26/18 3:00	668	7.86	4/26/18 3:00	668	7.86
4/26/18 4:00	668	0.33			
4/26/18 5:00	668	0.16			
4/26/18 6:00	668	0.07	Data for hours not in the dataset are added to the next entry seen below		
4/26/18 7:00	668	6.24			
4/26/18 8:00	668	1.52			
4/26/18 9:00	668	0.00			
4/26/18 10:00	668	0.96	4/26/18 10:00	668	9.29
4/26/18 11:00	668	0.03	4/26/18 11:00	668	0.03
4/26/18 12:00	668	4.84	4/26/18 12:00	668	4.84
4/26/18 13:00	668	0.04	4/26/18 13:00	668	0.04
4/26/18 14:00	668	0.00	4/26/18 14:00	668	0.00
4/26/18 15:00	668	0.00	4/26/18 15:00	668	0.00
4/26/18 16:00	668	0.00	4/26/18 16:00	668	0.00
4/26/18 17:00	668	0.02	4/26/18 17:00	668	0.02
4/26/18 18:00	668	25.57	4/26/18 18:00	668	25.57
4/26/18 19:00	668	7.56	4/26/18 19:00	668	7.56

2. **Case #2** If our sensor was offline or bad data was deleted the reading immediately following a gap will only show water consumption for the corresponding hour and NOT the period with a gap in data.

Case #2 - Meter Goes Offline, Data is Lost

HYPOTHETICAL CONSUMPTION			DATASET REPRESENTATION		
hourly_time	home_id	gal_water_ consumed	hourly_time	home_id	gal_water_ consumed
4/26/18 0:00	668	0.00	4/26/18 0:00	668	0.00
4/26/18 1:00	668	0.00	4/26/18 1:00	668	0.00
4/26/18 6:00	668	1.87	4/26/18 6:00	668	1.87
4/26/18 7:00	668	7.86	4/26/18 7:00	668	7.86
4/26/18 10:00	668	0.33			
4/26/18 12:00	668	0.16			
4/26/18 13:00	668	0.07	Data for hours not in the dataset are not recorded		
4/26/18 16:00	668	6.24			
4/26/18 17:00	668	1.52			
4/26/18 18:00	668	0.00			
4/26/18 20:00	668	0.96	4/26/18 20:00	668	0.96
4/26/18 21:00	668	0.03	4/26/18 21:00	668	0.03
4/26/18 22:00	668	4.84	4/26/18 22:00	668	4.84
4/27/18 0:00	668	0.04	4/27/18 0:00	668	0.04
4/27/18 3:00	668	0.00	4/27/18 3:00	668	0.00
4/27/18 4:00	668	0.00	4/27/18 4:00	668	0.00
4/27/18 5:00	668	0.00	4/27/18 5:00	668	0.00
4/27/18 8:00	668	0.02	4/27/18 8:00	668	0.02
4/27/18 12:00	668	25.57	4/27/18 12:00	668	25.57
4/27/18 14:00	668	7.56	4/27/18 14:00	668	7.56

Electric Data Table captures all electricity consumption at the home and any solar power produced at the home.

hourly_time - (Timestamp with timezone) Hourly interval of data.

home_id - (String) Each home_id represents one housing unit.

hourly_kwh- (decimal) This is the kWh consumed at the home during the given hour of the day. Unlike the water data, electricity reads are only for the hour shown. If there are any gaps in the hourly data that data is lost (Case #2 above).

hourly_solar_kWh - (decimal) This is the hourly solar electricity production in kWh at the homes. Over 70% of the homes in the sample have solar panels on them.

Weather Data All weather data is collected at Pecan Street's Lab in the Mueller neighborhood of Austin.

Localhour - (Timestamp without timezone) Hourly interval of data in UTC time. To join with water or electric time you need to add the TZ Offset to this time so the weather time matches correctly. For example Austin in the summer is Localhour - 5hours. If you join one of the hourly_time fields to this field you need to incorporate the TZ Offset. One way to check if this is correct is to look at solar data vs. your localhour timestamp and solar production should begin and end at sunrise and sunset.

TZ Offset – (Integer) This is the number of hours offset from UTC for the weather data in a given line. For example Austin in the winter time is UTC-6 so a tz_offset of -6 means the weather data is for Austin's timezone.

Temperature - (Decimal) Temperature in degrees Fahrenheit.

Precip Intensity - (Decimal) The amount of precipitation that has fallen during the hour

Irradiance - (Decimal) Global Horizontal Irradiance in Watts per meter squared. We only have irradiance data for 2020 but another good source is here:

https://dashboard.solarforecastarbiter.org/

Cloud Cover - (Decimal) Percentage of clouds covering the sky (0 -1)

Humidity - (Decimal) Percent humidity of the air (0 - 100)