

1 问题描述

原始数据为一台热水器一天内的信息记录。此数据分析目的涉及以下几方面：

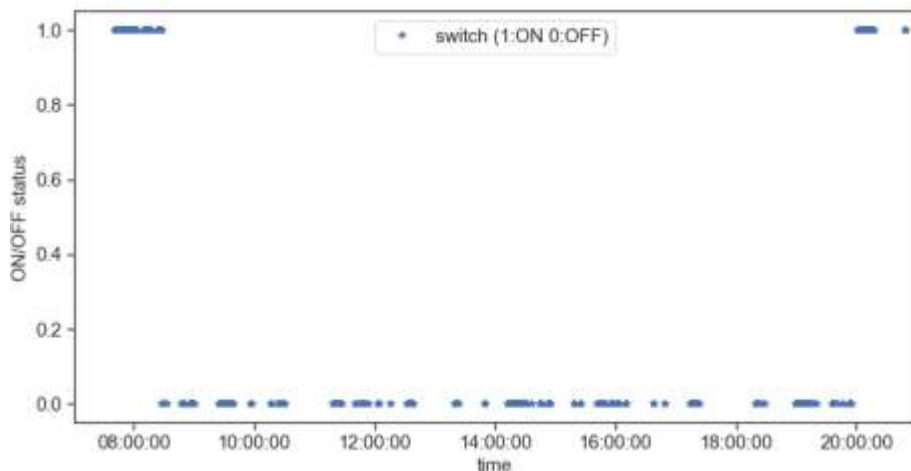
- 用户
 - 该用户这一天的用水行为如何？包括用水次数、用水量多少、耗电多少等等。
 - 该用户的潜在需求有哪些？此热水器是否能满足其需求？
 - 该用户属于我们目标用户的哪一类细分人群？其跟其它类的用户行为有何异同？
- 热水器
 - 此热水器目前的性能如何？比如功能多少、加热速度快慢等等。
 - 此热水器跟行业其它热水器性能比较有哪些优缺点？
- 其它
 - 可改进方向
 - 数据疑问

2 数据分析

原始文件为一台热水器 2020 年 3 月 18 日这天从早上 7 点到晚上 21 点之间的数据，包含 1137 条行记录和 94 列记录。整个数据集里有缺失数据，比如“场景”模式这一列缺失，但并不影响直接使用。用于分析的数据集需要精简，因为 1) 大部分列记录里的值只有一个（比如“夏季模式”均为“关”；“播放模式”均为“关”——此热水器功能很多，但用户并未全部使用），对分析无用；2) 原始数据集含有多个时间序列，保留一个即可。精简之后数据维度为 1137*11。留下的 11 列特征为：时间（time）、开关（switch）、加热状态（heatingstatus）、保温状态（keepwarmstatus）、实际温度（currenttemp）、剩余水量（waterremaining）、水流量（waterflow）、加热剩余时间（heatremainmins）、设定温度（settemp）、热水器顶部温度（heatertoptemp）、进水温度（inwatertemp）。

2.1 开关状态

该用户早上 7:40 开始用水，此时热水器处于开启状态，45 分钟之后关闭热水器一整天，直至晚上 20:00 再次开启热水器。数据截止至晚上 20:48，如下图所示。



值得注意的是，早上 7:40 的数据显示，此时水的实际温度为 41°C，远低于热水器的设定温度 55°C。这说明热水器在此之前处于关闭状态。而晚上 20:48 时，热水器处于开启状态，剩余水量为 60L。假设用户每天用水行为一致，基于早晨第一条记录显示的剩余水量 30L 可推测，晚 20:48 之后，用户依然在用热水器，但这一部分数据缺失。因此，我们的数据只是 3 月 18 日这天该用户的部分用水量，而非全部。

综上，该用户在 24 小时内，一共开启两次热水器，即早上洗漱（45 分钟左右）和晚饭后的洗漱（数据不完全，无法推测用时长短）。其余时间，即白天和晚上睡觉时，热水器处于关闭状态。由此推测，耗电量应为该用户在乎的因素，因此需要计算耗电量。

2.2 剩余水量

整个数据集内，最小剩余水量为 30L，最大剩余水量为 60L，而约 4/5 的时间，剩余水量为 40L 或者 50L。由此推断，1) 此热水器容量为 60L；2) 此热水器对该用户来说可能有些偏大。如果空间不宽裕的话，小一些的热水器也能满足该用户需求。

2.3 加热剩余时间

此热水器从关闭状态转为开启状态时，需要 20 分钟才能达到设定温度 55°C。这段时间对用户来说是否过长？市场上其它类似热水器的加热速度跟此相比如何？

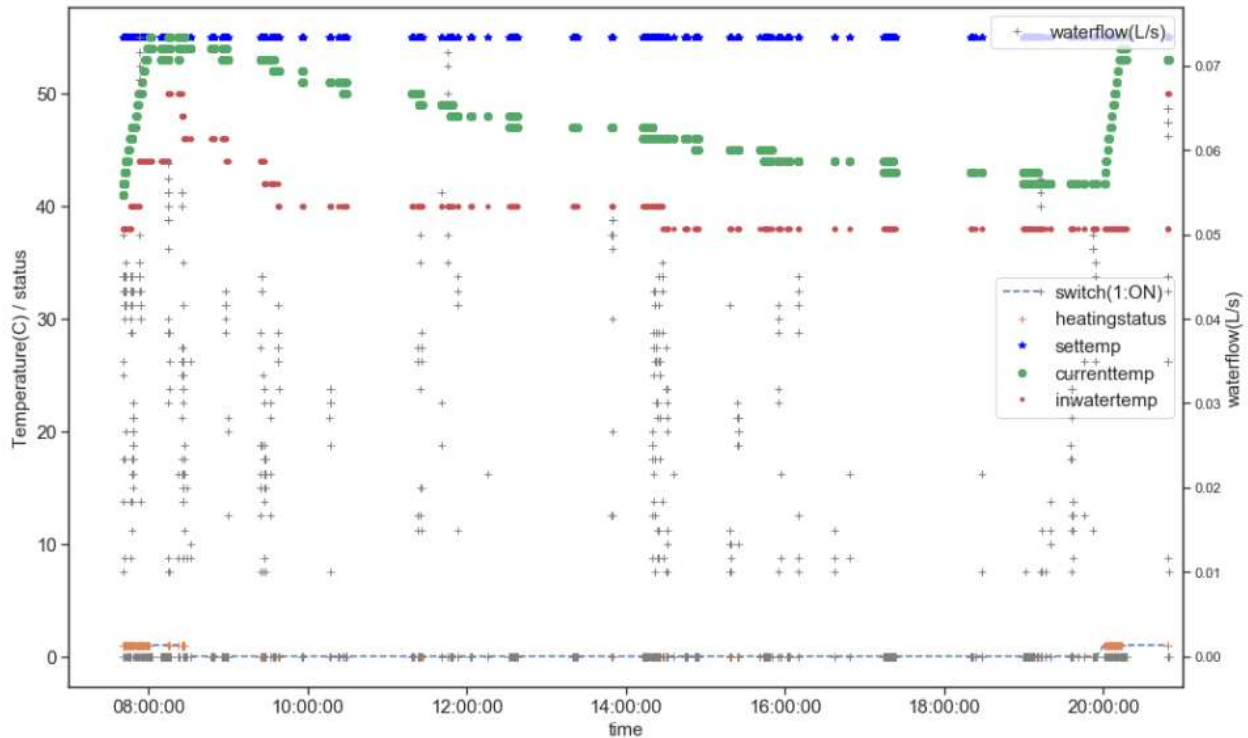
2.4 热水器顶部温度

热水器顶部温度和实际温度变化趋势相同且值相差不大，两者反应的是同一信息，所以只保留“实际温度”这一特征。

2.5 变量综合分析

下图表示了六个变量在 14 个小时内的变化，包括热水器开关状态、加热状态、设定温度、实际温度、进水温度以及水流量。整体来看，热水器只在早晨 8 点左右和晚上 20 点左右开启，实际水温随着加热稳步上升，表示该热水器性能正常。基于此图有以下几点分析。

- 从早上 7 点后至晚上 9 点前，不管热水器开关与否，一直有水流量显示。而 2020/03/18 为周三工作日，说明此用户要么并非上班族，要么是上班族但因为疫情呆在家里。
- 热水器的设置温度一直为 55°C。这个值是用户自己设置的还是默认标准？基于节能的考虑，或许可以考虑针对不同使用场景设置低一些的温度，比如 50°C。
- 在白天热水器关闭的 12 小时之内，实际水温先保持设定温度 55°C 左右不变，随后稳步下降至 40°C。此段时间内，用户一直在用水。如果用户是为了节能才关闭了热水器，那么低于 55°C 的水温能否满足用户的需求？如果不能，我们是否可以设计拥有更好物理保温性能的水热水器，使温度下降得更慢一些？
- 进水温度在 38°C 至 50°C 之间变化。这个范围的进水温度是否合理？
- 水流量一直处于变化状态，且数据记录并非每秒连续。如果计算总用水量的话，可行方法之一是先计算用户每一次的用水量（平均水流量乘以用水时长），然后再求和。



2.6 耗电量计算

在热水器开启时，因为保温状态耗电很低，所以只需计算加热状态时的耗电量。根据公式

$$\text{耗电量(KWh)} = m \times C_p \times (\text{Temperature Difference}) / (3412 \text{ KWh/BTU} \times \text{UEF})$$

这里，

C_p ：水的热容量 (1 BTU/lb/F)

m ：加热水量 (gallons)；假设 1 gallon = 8.33 lbs 水；3,412 BTU = 1 kWh

UEF: Energy Uniform Factor，电热水器的变化范围为 0.75 - 0.95，这里选取值 0.85 进行计算。

温度差异是设置温度和进水温度之差，加水量可根据水流量 (L/min) 乘以用水时长来计算。

3 初步结论及未来方向

3.1 用户行为归纳

1. 在早 7 点至晚 21 点这段时间内，该用户用水时长达 55 分钟，一共用水 103L；在热水器开启的状态下，该用户用水时长达 20 分钟，用水量 41L，一共用电 1.6KWh，主要集中在早 7:40 至 8:30 之间。由该用户晚 20:48 热水器开启及用水量推断，晚间应该也有一个用水高峰期，但这部分数据缺失。
2. 以 4 分钟间隔和用水量 100 毫升作为阈值，可以得出早 7 点至晚 21 点这段时间内，该用户一共有 16 个用水行为，最长的一次是下午 14:19 开始，用水时间为 11 分 35 秒，用水量约 20L；次长的是早上 8:22，7:40 及 9:24 开始，用时分别为 9 分 1 秒、7 分 51 秒、7 分 38 秒，用水量分别为 18L，16L，11L。由此推断，此用户可能为三口之家。根据中国居民生活用电人均 700 度/年计算，三口之家用电大约 6 度/天。而此三口之家仅热水器一

项就超过 1.6 度电/14 小时，如果按 2 度电/天来算，仅热水器一项就占总用电量的 30%，说明该三口之家的热水器用电有值得节省的地方，或许这也是其热水器在白天关闭的原因之一。

3. 此热水器容量为 60L，剩余水量一直在 30L 或者以上。如果用户家里空间不大，是否小尺寸、小容量的热水器会更受欢迎？这个可以通过做用户调查来得出结论。
4. 数据显示该用户倾向于省电，但为什么“节能模式”并未开启？“节能模式”是通过什么来达到节能的？是降低设置温度吗？这里需要比较“节能模式”和“正常模式”的耗电、功用等等，以便更好满足用户需求。

3.2 热水器性能归纳

1. 该热水器工作性能正常，开启之后温度稳步上升，从开启到加热完全需要 20 分钟。这个时长对于用户来说是否太长？该热水器有“云预约”功能，且用户开启使用，这个是否有效帮助用户削减等待时间？
2. 该热水器功能很多，比如分人洗模式、冬季模式、节能模式等等，但用户并未使用。如果用户不在意这些附加功能的话，是否可以针对这类用户来设计功能简单、价格低廉的热水器？
3. 用户白天的热水器处于关闭状态。或许可以设计关闭状态下保温性能更好的热水器，满足用户对热水和省电的双重需要。

3.3 未来方向

1. 此用户属于我们目标用户的哪一个细分市场？其用水模式是否能代表整个群体？或许多个热水器的数据能帮我们做进一步分析。
2. 此类数据采集时间为 2020/3/18，采集地区为哪里？进水温度是否受外在温度的影响？这个时间，该用户是否因为疫情呆在家里？如果是，此段时间的采集数据是否能代表 2020 全年或者用于预测以后？
3. 用户方面，可以帮其追踪每月热水器的用电量，并跟周围同类型的用户（比如三口之家）对比，帮助其制定省电策略。

4 一些疑问

1. 进水温度是自来水温度吗？为什么会在 38-55℃ 之间变化？
2. 为什么当日晚 20:48 之后的数据缺失？