Analytic Results for Sprint 3

Treatment Group – houses who were subjected to dynamic time of use energy prices

Mean maximum daily energy consumption for the treatment group (over a half hour period): 0.79 kWh

Standard deviation of maximum daily energy consumption for the treatment group: 0.60 kWh

Mean of total daily energy consumption for the houses in the treatment group: 9.50 kWh

Standard deviation of total daily energy consumption for the treatment group: 8.08 kWh

Treatment group table size: 706,031 rows.

Control Group – houses on a flat rate tariff of 14.288 pence/kWh

Mean maximum daily energy consumption for the houses in the control group: 0.85 kWh

Standard deviation of maximum daily energy consumption for the control group: 0.68 kWh

Mean of total daily energy consumption for the houses in the control group: 10.28 kWh

Standard deviation of total daily energy consumption for the control group: 9.37 kWh

Control group table size: 2,804,402 rows.

A two-sample t-test for a difference in the means between the treatment and control groups with respect to both max daily energy use and sum of daily energy use result in a rejection of the null hypothesis of no decrease in energy usage as a result of dynamic time of use pricing being implemented at any conceivable statistical significance threshold (the t-statistic for the former was -70.11 and -73.04 for the latter).

Text

Description automatically generatedBut because the daily\_dataset table has plenty of cartoonishly large outliers in it compared to what you would expect it to have if it followed a Gaussian distribution, I also performed a two-sample Wilcoxon test, the non-parametric equivalent test for a difference in means, on the join of the daily\_dataset and the informations\_households tables for good measure. The results of that hypothesis test for the difference in the means of the energy\_sum column are shown below:

Text

Description automatically generatedI also performed a Wilcoxon rank sum test, also known as the Mann-Whitney test, on the difference in the means of the maximum daily usage column, the results of this test are shown below: