I’ve not been able to find the answers to question 4.2 and 5.1, please provide a detailed answer specifically to each question following the order propose by the question set..

I corrected that in this version.

From your work it is not clear whether you effectively doubled the P value to compare it to your chosen P critical value, please make this explicit in your next submission. Although you have correctly chosen, in 1.1, to use a two-tailed test, the p-value reported in 1.3 is a one-tailed value. Please take a look at the documentation for the SciPy implementation of the Mann-Whitney U-test to learn how to convert this to a two-tailed value.

I explained this in Section 4.2

Did you use a different dataset from the one used in the course? If so please specify that because the mean values and the P value reported don’t match the ones of the base dataset. (Please refer to problem set 3.3 for the actual values or specify the reason for that difference in the beginning of your submission)

There was a mistake in the previous version with one of the pandaSQL GROUP BY queries that led to incorrect mean values and P values, and completely changed the conclusion of the Mann-Whitney analysis. My numbers may differ slightly from that in the Lesson, because I removed 6% of the data, explained in Section 1.3, and 2.3. Thanks to the grader for pointing this out

In order to meet specification you would need to discuss more thoroughly the meaning of the R squared and the result you obtained in your model in 2.6, a generic statement is not sufficient. What is the meaning of the R squared?

I improved the explanation in Section 2.6.

You mention using a two tailed test but then it is not obvious whether you compare a one tailed P value to the P critical value. The P critical refers to a value (often .05) established beforehand that the p value reported from the statistical test will be compared against. In order to be deemed significant the p value should be <= p critical.

I explained this in Section 4.2

The statistical test performed as in Problem Set 3.3 actually confirmed that there is a difference: the two tailed P value is less than the chosen P critical value. We reject the null hypothesis at the 95% confidence level that the distributions are the same and accept the alternate hypothesis that they are different.

There was a mistake in the previous version with one of the pandaSQL GROUP BY queries that led to incorrect mean values and P values, and completely changed the conclusion of the Mann-Whitney analysis. My numbers may differ slightly from that in the Lesson, because I removed 6% of the data, explained in Section 1.3, and 2.3. Thanks to the grader for pointing this out

In plot 3.1 the bins abruptly stop at 1000, this is not requested, please provide a rationale for your choice in your next submission. In 3.2 rescaling the X axis and increasing the binning would help the overall readability. In 3.3 if the X axis represents days those should appear in the labels, the actual readability is not optimal.

I rescaled the plots, hopefully in the way the grader was looking for.