**A/B Testing Final Project Rubric**

|  |  |  |
| --- | --- | --- |
| **Criteria** | **Meets Specifications** | **Exceeds Specifications**  (Completely Udacious) |
| **Metric Choice** | | |
| Has the student chosen good invariant and evaluation metrics for the experiment? | The student chose a good set of metrics for the experiment, and did not miss any necessary or valuable metrics. | N/A |
| Has the student given a well-reasoned justification of their choice of metrics? | All metrics had clear and well-reasoned explanations of why they were or were not chosen. | N/A |
| Has the student stated for which results they would launch the experiment? | The student clearly stated what results they were looking for to launch the experiment, and the stated results were aligned with the experiment goals. | N/A |
| **Variability** | | |
| Is the standard deviation for all evaluation metrics correctly calculated? | The standard deviation for all evaluation metrics is correctly calculated. | N/A |
| Has the student correctly reasoned about whether each analytic standard deviation is likely to be accurate? | Each evaluation metric has a clear and correct explanation of whether the analytic variability is likely to match the empirical variability. | N/A |
| **Sizing** | | |
| Does the number of pageviews correctly take into account the planned analysis? | The number of pageviews given is correct given the students choice of whether to use the Bonferroni correction. | N/A |
| Has an appropriate level of exposure for the experiment been chosen based on the risk? | The student has made a well-reasoned argument about how risky the experiment will be and chosen a fraction of traffic to divert accordingly. | N/A |
| Does the duration of the experiment correctly take the exposure chosen into account? | The duration of the experiment is correctly calculated given the fraction of traffic the student chose to divert. | N/A |
| **Sanity Checks** | | |
| Has the student correctly performed sanity checks? | The student has correctly calculated sanity checks for all chosen invariant metrics. | N/A |
| Has the student taken the sanity checks into account? | All sanity checks passed or the student did not proceed to the rest of the experiment and analyzed why the sanity checks may have failed. | N/A |
| **Effect Size Tests** | | |
| Has the student calculated confidence intervals around the difference of all evaluation metrics? | Correctly calculated confidence intervals have been reported for the difference in all evaluation metrics. | N/A |
| Has the student correctly evaluated statistical and practical significance? | Statistical and practical significance have been correctly reported for all evaluation metrics. | N/A |
| **Sign Tests** | | |
| Has the student correctly reported a sign test p-value for each evaluation metric and indicated whether the sign test is statistically significant? | P-value and statistical significance have been correctly reported for all evaluation metrics. | N/A |
| **Results Summary** | | |
| Has the student correctly chosen whether to use the Bonferroni correction? | The student has given good justification for their choice of whether to use the Bonferroni correction. | N/A |
| Has the student correctly analyzed all discrepancies between the effect size tests and the sign tests? | The student has given well-reasoned and plausible explanations for each discrepancy between the effect size tests and the sign tests. | N/A |
| **Recommendation** | | |
| Has the student made a well-reasoned recommendation based on the results of the experiment? | The student has made a recommendation that is well reasoned and supported by the data. | N/A |
| **Follow-Up Experiment** | | |
| Has the student chosen a plausible experiment for the purpose given with a clearly stated hypothesis? | The student has described a plausible experiment that would be worth testing and the hypothesis is clearly stated. | The student has described a creative or innovative change that Udacity would be happy to test. |
| Has the student chosen good metrics to evaluate the proposed experiment with good reasoning to support them? | The metrics the student has chosen will be sufficient to evaluate the hypothesis of the experiment, would be possible to measure under most infrastructures, and are well-supported by the students reasoning. | N/A |
| Has the student chosen a well-reasoned unit of diversion for the experiment? | The student has chosen a reasonable unit of diversion and given good support for their choice. | N/A |