

STUDENT

Richard Haughton

COURSE

Intro to Data Science

Well done on a good project! It is not far from meeting specifications. I have just a few comments for you in the rubric below. The bullet points **highlighted in blue** need to be addressed for your project to meet specifications. Feel free to email nd002-project@udacity.com if you have any questions. I look forward to your resubmission!

Charlotte and the Udacity Team

Communication**Meets Specifications**

- Analysis done using methods learned in the course is explained in a way that would be understandable to a student who has completed the class.
- The answers are a well-formed summary of the analyses.

Quality of Visualizations**Meets Specifications**

- Plots depict relationships between two or more variables.
- All plots are of the appropriate type.
- **Some plots are not appropriately labeled and titled or visual cues are not always easy to distinguish. It is not clear what data are represented.**
The plot of Average NYC Turnstile Entries by Hour of the Day has no legend. Please include a legend or explanation of the visual encodings used. What do the grey area and darker blue line represent? This is a good idea for a plot, and you make some good points in your analysis of it.
Your plot of "Entries Per Hour By Day And Temperature" is a really good idea, very well done for your work on implementing jitter!
It seems to suffer from overplotting which obscures the density of the points and, as you mention, the colours could perhaps be misleading. In my view, these issues are overcome by your clarifying comments. It is difficult to make a very clear plot that encodes so many variables at once, you have made a good attempt, well done!

Quality of Analysis**Does Not Meet Specifications**

- **The choice of statistical test type, features, and linear regression models are sometimes not**

appropriate based on the characteristics of the data.

You should state whether you use a one- or two-tailed test. In this case it might be a good idea to use a two-tailed statistical test. Picking a one-tailed test means that we assume in advance (before we collect the data) that rain will not be associated with lower ridership, which is a very strong assumption. For more information, see the following:

http://graphpad.com/guides/prism/6/statistics/index.htm?one-tail_vs_two-tail_p_values.htm

- Statistical tests or linear regression models are not described thoroughly, or the reasons for choosing them are not clearly articulated.

Did you use a one or two tailed test, and what was your significance level? These should be chosen before you start the test.

You mention that you use 'ones' as a feature. This is not strictly a feature, it allows you to have an intercept to your linear model.

- The use and interpretation of statistical techniques are correct.
- All conclusions are correctly justified with data.
- No incorrect conclusions are drawn from the data.
- Some shortcomings of the statistical tests or regression techniques used are appropriately acknowledged.

You have some very interesting ideas on ways to improve the dataset and your analysis. I hope you work on this more when you have learned more data analysis techniques!

PROJECT EVALUATION

Project Does Not Meet Specifications