UDACITY

Nanodegree Project Evaluation

STUDENT

Teguh Wijaya

COURSE

Intro to Data Science

Hi Teguh,

Thank you for your project submission. This is a well written project, and 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 dataanalyst-project@udacity.com if you have any questions. I look forward to your resubmission!

Stephen and the Udacity Team

Click here to tell us whether this feedback was helpful.

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 — Does Not Meet 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.
    Comment: In the plot of "Entries hourly by day of week", it isn't clear what is being plotted here. Are these total hourly entries for all weekdays? Are these averages? If you would like to plot aggregate data (such as averages or totals), you should aggregate the data before creating the plot by using the .groupby() function. If you are using a geom_bar layer with ggplot, then you should probably be using stat = "bar" as an option. I think that you will find plot very different than it appears now (there are some strong trends that will appear in the plot when this is done correctly).

Quality of Analysis **Does Not Meet Specifications**

- When using statistical tests and linear regression models, the choice of test type and features are always well justified based on the characteristics of the data.
- Statistical tests or linear regression models are not described thoroughly, or the reasons for choosing them are not clearly articulated.

  Comment: In Section 2, it is not quite correct to say that dummy variables "are stored in the UNIT field of given data." The dummy variables are created from the UNIT feature, but not stored there. There is one dummy variable for each value of UNIT.

  For question 2.6, could you elaborate on why the model will not be able to predict future values well? What is the range for R^2? What is considered a good R^2, and what is not? Why? For reference, here is a helpful blog post: http://blog.minitab.com/blog/adventures-in-statistics/how-high-should-r-squared-be-in-regression-analysis
- Mistakes are made in use or interpretation of statistical techniques.

  Comment: In Section 1, the default p-value given by the Mann-Whitney U test in Python is a one-tailed p-value. You will need to double this in order to get the two tailed p-value.

  Optional: The statement of the null hypothesis for the Mann-Whitney U test captures the general idea of the test. An exact statement of the null hypothesis can be found in the downloadables from Lesson 3. The downloadable notes about the Mann-Whitney U test can be accessed by clicking on the appropriate link below the video window of any of the Lesson 3 videos.
- 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.

PROJECT EVALUATION

Project Does Not Meet Specifications