Analyzing the NYC Subway Dataset

Student Notes | Project Review | Project History

Does Not Meet Specifications

Communication

SPECIFICATION

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.

MEETS SPECIFICATION

Reviewer Comments

Well Done!!

SPECIFICATION

The answers are a well-formed summary of the analyses and do not leave out important information (i.e. fully answering the question).

MEETS SPECIFICATION

Reviewer Comments

Since the project is a report it might make sense to include the code and the very long tables in an appendix at the end of the report. Another option that you might want to consider is to hide the code in the HTML.

Quality of Visualizations

SPECIFICATION

Plots depict relationships between two or more variables.

MEETS SPECIFICATION

Reviewer Comments

Well Done for including charts throughout the report that depict different aspects of the data set and the analysis.

SPECIFICATION

All plots and data are of the appropriate type.

MEETS SPECIFICATION

Reviewer Comments

Well done!!

SPECIFICATION

All plots are appropriately labeled and titled. Plot is given an appropriate title. X-axis and y-axis are appropriately labeled. Visual cues (colors, size, etc) are easy to distinguish. It is clear what data are represented.

MEETS SPECIFICATION

Reviewer Comments

Just few comments here,
For the chart that depict the "Relation between average ENTRIESn_hourly for each hour and day_week" please consider to use the actual day instead of the indexes.er to use larger fonts, in some charts the text is barely readable.

Quality of Analysis

How satisfied are you with this feedback?

Resubmit Project

atures are

MEETS SPECIFICATION

Reviewer Comments

Well Done for the implementation, the clear interpretation and explanation about the statistical test. I do have few comments here.

1. Please consider to use more descriptive null hypothesis, for example describe what distinguish the two groups.
2. Since you use a two tailed p-value and since the statistical test is non parametric, it is appropriate to include the averages that you calculated for each condition as part of the interpretation for the statistical test results.

SPECIFICATION

Statistical tests and linear regression models are described thoroughly, and the reasons for choosing them are articulated clearly.

MEETS SPECIFICATION

Reviewer Comments

Well Done for the clear explanation about the reasons for choosing features to the regression model.

Currently, the coefficient for rain in your regression model is negative. One interpretation for this is that, holding all other variables constant, rain will have a negative effect on ridership. Since there are many variables included in the regression, it might be the case that some of the variables are highly correlated, which can cause your coefficient estimates to be unstable. It might be a good idea to build the regression model gradually, to ensure that highly correlated variables are not included together. For more information about multicollinearity, including information about the condition number, see the following Wikipedia article: http://en.wikipedia.org/wiki/Multicollinearity.

Please note that percipi is a continues feature and not discrete.

SPECIFICATION

The use and interpretation of statistical techniques are correct.

MEETS SPECIFICATION

Reviewer Comments

Well Done for the clear interpretation of R squared and the explanation about the result from the regression model.

SPECIFICATION

All conclusions are correctly justified with data.

DOES NOT MEET SPECIFICATION

Reviewer Comments

Well Done for using the results from the statistical test and the regression model in the conclusion section.
However in section 4.2 it is stated "But the p-value from linear regression summary (0.210) is greater than the common alpha level of 0.05, which indicates that:Rain is not statistically significant effect on ENTRIESn_hourly", this is not accurate. The p-value of the model coefficient, only indicate whether the contribution of the feature (rain) to the model is significant. Keep in mind that the model coefficient and the corresponding p-value are strongly depend on the exact set of features that are included in the model and will change when you changing the features the included in the model. Therefore, you can use the coefficient and the corresponding p-value as an indicator, but the average that you calculate for each group is more robust in this case.

SPECIFICATION

No incorrect conclusions are drawn from the data.

MEETS SPECIFICATION

Reviewer Comments

Please see my comment above.

SPECIFICATION

Some shortcomings of the dataset and statistical tests or regression techniques used are appropriately acknowledged.

MEETS SPECIFICATION

Reviewer Comments

Well Done for the comprehensive shortcoming section.

Download project

Learn the best practices for revising and resubmitting your project.

Have a question about your review? Email us at review-support@udacity.com and include the link to this review.