FLATIRON SCHOOL

Student Name: Michael Flores
Instructor: Abhineet Kulkarni
Date: 08/03/20
Attempt Number: 1

Data Science Online Data Science Bootcamp
Module 2 Final Project Review

Technical Notebook

Project Specifications	Metric for success	Developing		Accomplished		Exemplary (X-Factor)		Notes
README.md	Student has a clear README, highlighting important aspects of the project.	Student does not have a readme, or has a readme that is just a copy of the notebook.	☐	Student has a readme with a clear and well organized outline, conclusion and recommendation section. Visualizations are present.	☐	Student has a readme with a clear and well organized outline, conclusion and recommendation section. Visualizations are present. Language and markdowns lend themselves to succinctness.	☑	
Pick a novel interesting problem at the appropriate challenge level.	For the House prices dataset, constructed the business case around it well.	Business case not clearly articulated. Answered an obvious business question.	☐	Business case constructed clearly. And answered an obvious question, like clearly articulated the business stakeholder requirements that the project aims to accomplish. Combine questions in unique ways like how does zipcode impact house prices?	☑	Created original and meaningful work - Created a unique business case for the House prices dataset. For example, analyzed the dataset from the point of view of sellers, buyers, a startup that would use this business model, etc.	☐	
Preprocess data	Import the data and preprocess the data that includes cleaning, scrubbing, handling missing values, etc.	Data not fully ready for later analysis. 100% correctly structured data. Handled missing values.	☐	Explored different methods.	☐	Handled especially tricky issues. Explored different methods with benchmarking.	☑	
Describe data	Use data exploration, visualizations (at least 3 different kinds), distributions, and discussed the distributions as pertaining to the given dataset. Calculated few statistics, models, and analyses.	Inadequate visualizations (less than 3 different ones), fit inappropriate distributions (or no discussion of distributions), inadequate descriptive statistics	☐	Visualizations (at least 3) enough to understand data. Explored enough to understand data. Fit the most appropriate distributions.	☐	Created novel distributions. Compared multiple distributions. Fit distributions outside of Module./class. Found insights in data.	☑	

Fit models/Hypothesis testing	Fit at least one model. Summarize model impact and meaning.	Attempted basic model fitting (or forgot to model fit). Incorrect application. Misinterpreted results.	☐	Correctly fit a single model. Correctly interpreted model results. Summarized model meaning & impact.	☐	Compared multiple models. Fit models outside of class materials. Detailed numerical and visual analysis of models.	☑	
Present to technical audience	Present work done to a technical audience with code, insights, summary, future work, and even a live demo (for extra credit).	Unintelligible, hard to follow. Unclear. Incomplete.	☐	Engaging talk with insights & lessons. Explained code examples.	☑	Live demo! Ran code and changed parameter values.	☐	
Write quality code	Code does what the analysis says it does. It is clear, concise, easy to read and understand.	Code is incomplete. Code NOT in GitHub. Code does NOT work. Code is hard to read. Code does not have README. Commit messages are not helpful.	☐	Repeated some analyses covered in sections/class. Showed some creativity.	☐	Code has comments and tests. Professional level/ pep 8. GitHub repo is public (if appropriate).	☑	
Conclusion	Notebook contains a conclusion with business recommendations that are driven by analysis.	No conclusion present.	☐	Conclusion present but only states findings and contains 1 or 2 relevant business recommendations.	☐	Conclusion is present and contains at least 3 recommendations that are business relevant.	☑	Make sure you calculate the rmse whenever you have to predict the value of a continous variable
X - factor: Did something out of the box	Went above and beyond to research some additional topic, concept, Python package(s).	Routine project. Repeated analysis covered in class/sections of the module.	☐	Showed creativity.	☑	Ground breaking.	☐	

Non-Technical Presentation

Project Specifications	Metric for success	Developing		Accomplished		Exemplary (X-Factor)		Notes
Present to non-technical audience	Present work done to a non-technical (business focused) audience with problem statement, business value, methodology explained simply, business recommendations, summary, and future work.	Unintelligible, hard to follow. Unclear. Incomplete. Slides are too verbose, slide notes non existent.	☐	Engaging talk with insights & lessons. Explained methodology. Slides have images, less text, slide notes present on slide that mirror the script of the presenter. One slide for each of the following - Problem statement, business value, methodology, business recommendations (each recommendation on a separate slide), future work/next steps.	☐	Additional slides like findings, or use of engaging images, graphics, material showing expertise in communicating to business stakeholders.	☑	
Slide Quality	Slides are light on text, engaging and tell a story.	Slides are very text heavy or highly unorganized and all over the place.	☐	Slides are organized and tell a story, but contain too much text at times, especially when a visualization will suffice.	☐	Slides are organized, contain visualizations that relay information and slides tell a story.	☑	
Duration	Your presentation should be between 5 and 8 minutes.	Presentation is over 10 minutes or under 3 minutes.	☐	Presentation is over 8 minutes or under 5 minutes.	☐	Presentation is between 5 and 8 minutes.	☑	

Non Technical	Presentation contains great data science that is delivered using non technical language.	Presentation uses technical terms without succinct explanations more than 3 times.	☐	Presentation uses technical terms without succinct explanations once or twice.	☐	Presentation does not use technical terms or provides succinct explanations when using them.	☑	
Test Results	Hypothesis test results are shown and made relevant to the business, driving the recommendations from the project.	No tests are shown or tests shown do not relate to business.	☐	Test results are shown and made clear to business case.	☐	Test results are shown, made relevant to business case and also highlight deeper insights into the business.	☑	
Visualizations	Slides contain visualizations that take the place of text and give the viewer insight.	Slides do not contain visualizations or the visualizations present are not relevant to the story.	☐	Slides contain visualizations that are relevant to the story but hard to interpret.	☐	Slides contain visualizations that are relevant and easy to understand.	☑	
Recommendations	A great presentation contains business recommendations and steps moving forward.	No recommendations are made	☐	At least 3 recommendations are made, but are not driven by data analysis or model.	☐	At least 3 recommendations are made and are driven by analysis and model.	☑	
Future Work	A data scientist will never have enough time to explore all aspects of dataset. If you had more time, what other aspects of the dataset would you explore?	No slide on Future work.	☐	Future work slide content not well defined and/or articulated.	☐	Future work clearly articulated, explored, and its potential business impact (s) described.	☑	
Thank You Slide	Thank your audience for their time, it's a great practice.	Thank You Slide is not present.	☐	Thank You Slide is present.	☐	Thank You Slide is present. Appendix includes additional work.	☑	

Qualitative Assessment

1. Problem Statement - how well was it defined for this project the readme was detailed and provides some good details

2. Things you did well: the modelling and the presentation were great

3. Things to work on/ consider : making sure you calculate all the required metrics

4. Action items: none!