Joshua Vong

Reflections of Peer Review

Reading the peer reviews really helped me on improving my project over time. The first

peer review, my group seemed concerned about my graphs, since at the time I only had a few,

and they weren't that interesting. My graphs needed to tell a story about my data, so after some

cleaning up and restructuring my graphs, I managed to remake my graphs a lot more visually

appealing and help make it tell a better story. The second set of peer reviews for my second

project involved some criticisms for my modeling. At the time I didn't have any models to show,

but instead I asked questions about how to go about modeling my data. My group were able to

explain to me clearly on how to model my data. Overall both sets of peer reviews were extremely

helpful in creating my final version of my project

Peer Reviews

Joseph's Project:

Cynthia's Comment: Look's awesome! Really like the images of the stadiums. Just curious why you picked that specific player, is there something that makes him stand out from others?

Josh's Comment: When doing model planning don't forget the fail rate for number of hits for the player.

Darlyn's Comment: The pictures really help to visualize the stadiums. Think about what "success" and "failure" look like with respect to these questions.

Zak's Comment: Maybe show what success and failure looks like for the 3 data science questions.

Darlyn's Project:

Cynthia's Comment: Great! Maybe make an age group column and count the "yes/no" from the categorical variables from each age group to make a numerical variable out of that.

Joseph's comment: Include limitations to answering the data science question with just your data. You could even say, "More data is needed about this…"

Josh's Comment: Could possibly build off of the correlations between columns and make a new one with a predictive model in mind?

Zak's Comment: Maybe include some more visualizations for the data. Adding new manipulated variables to get more numerical variables will be helpful in model predicting.

Cynthia's Project:

Zak's Comment: Possibly state what success and failure looks like for each data science question

Darlyn's Comment: The graphs are very good and help visualize the data. Think about which question you want to use and how to define success and failure for this question.

Joseph's Comment: If you choose to predict a categorical variable, look into "logistic regression". It is used for classification predicting if there are only 2 categories.

Josh's Comment: For the predict model could maybe predict some of the variables of what they would look like in the future.

Josh's Project:

Darlyn's Comment: Really interesting story about why you chose this data and how it relates to you personally. Make sure to include limitations this data set might have when modeling your data.

Cynthia's Comment: Great data. Predicting why some areas caught on fire while others did not is a great idea. Try thinking of what other numerical variables can be predicted using the rest of the variables. Also, think about what success or failure would look for each.

Joseph's comment: Good data science question! Be sure to mention that fires are also dependent on other factors than weather.

Zak's Comment: Possibly add some visualizations for the data set to get an idea of what the data looks like. You could possibly predict the size of the fire and how much land it will burn.

Zak's Project:

Cynthia's Comment: Great! For the visualizations where you're showing the defensive rating, maybe show teams instead of each individual player. So show total defensive rating per team.

Darlyn's Comment: Great job so far. Include more detailed explanations of exactly what the graphs are showing.

Joseph's Comment: Work on the problem of focusing on teams instead of players, if you choose to predict a team metric

Josh's Comment: For modeling you could focus on the teams rather than individual players or you could do just a single player's defensive stats or any other stats.

# Zoe K.

## What parts of the presentation were most effective?

Zoe

Reilly
Informative, easy to read plots. Good at showing the general spread of the data with the EDA and model plots.

Joshua - I would say that the graphs are the most effective, they seemed neat and easy to understand.

Jordan

## Did you find any aspects of the presentation particularly confusing or unclear? If so, what might help to clarify those parts?

Zoe

Reilly

Joshua - Splitting the data into training and tests seemed a little messy.

Jordan

## Are the prediction tasks being pursued easy to understand? Do they address the relevant data science questions?

Zoe

Reilly
Looks good, very prediction-oriented which is nice!

Joshua - Seems easy to understand.

Jordan - They are definitely easy to understand and very clear. They relate to the data science questions enough, that these data science questions will be answered with your models.

## Are the definitions of "success" and "failure" clear?

Zoe

Reilly

Joshua - Not really any definitions of success and failure

Jordan

## Are the models well justified?

Zoe

Reilly

Great thinking on the exponential model, check out the in-class assignment from 11/4 (Simple Models) or somewhere around chapter 24.2.3 in R for Data Science to hopefully figure out that issue you're having with the log-scale lm().

Joshua - I believe that the model is justified.

Jordan - The models are well justified

## Are any unjustifiable conclusions drawn?

Zoe

Reilly

Joshua - It doesn't seem like there are unjustifiable conclusions, since the project isn't finished yet.

Jordan

## Are data science questions clear and forward looking? Could they form the basis of more formal falsifiable hypotheses?

Zoe

Reilly

Definitely forward-looking, though I wonder if comparing to 2020 would cause issues with the test data, as that's not a particularly great year for baseball records. Is the plan to use 2019 data as test data to predict what 2020 *should* have been like?

Joshua

Jordan - The questions are clear on what you are trying to achieve and forward looking.

Can you think of any other models or techniques that could be useful in addressing the data science questions?

Zoe

Reilly

Joshua -  Could potentially use a naive bayes classifier?

Jordan - definitely look into the simple models assignment we did and also in the book they use good examples of different models that can be used for your data.

Can you think of any potential social or ethical implications of the project?

Zoe

Reilly

Joshua- I believe it would be helpful for those interested in baseball statistics, I don't really see much ethical implications.
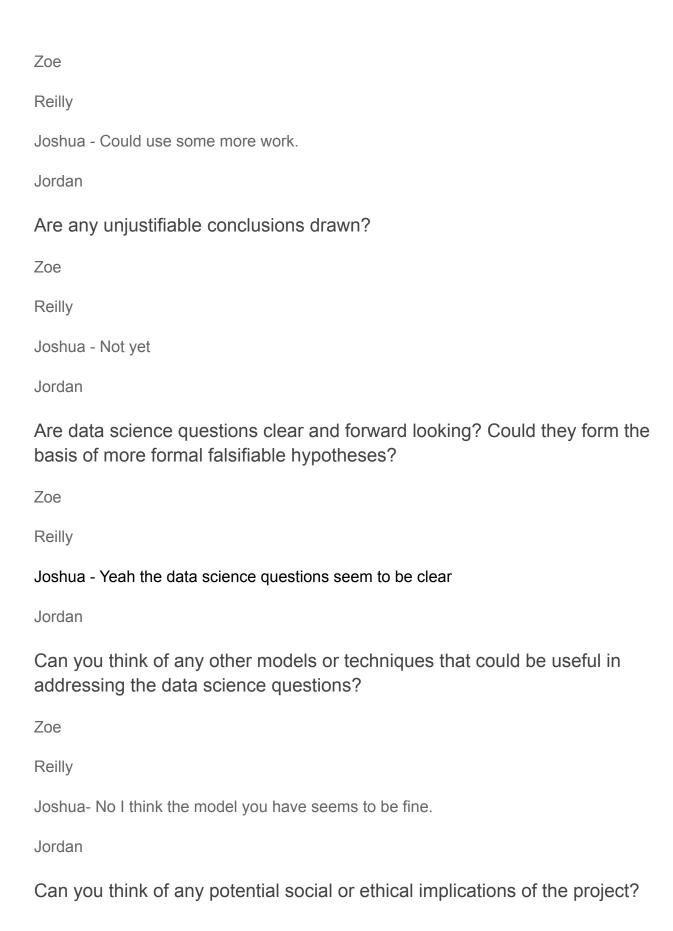
Jordan

# Reilly

What parts of the presentation were most effective?

Zoe

Reilly

Joshua - The topic is interesting, since chess is competitive I feel like good strategies can come from this dataset.

Jordan

Did you find any aspects of the presentation particularly confusing or unclear? If so, what might help to clarify those parts?

Zoe

Reilly

Joshua - No not really, the presentation seems clear of what your trying to accomplish.

Jordan

Are the prediction tasks being pursued easy to understand? Do they address the relevant data science questions?

Zoe

Reilly

Joshua - Yeah they seem easy to understand.

Jordan

Are the definitions of "success" and "failure" clear?

Zoe

Reilly

Joshua - Not really since the project is incomplete yet, there's still more work to be done.

Jordan

Are the models well justified?

Zoe

Reilly

Joshua - Could use some more work.

Jordan

Are any unjustifiable conclusions drawn?

Zoe

Reilly

Joshua - Not yet

Jordan

Are data science questions clear and forward looking? Could they form the basis of more formal falsifiable hypotheses?

Zoe

Reilly

Joshua - Yeah the data science questions seem to be clear

Jordan

Can you think of any other models or techniques that could be useful in addressing the data science questions?

Zoe

Reilly

Joshua- No I think the model you have seems to be fine.

Jordan

Can you think of any potential social or ethical implications of the project?

Zoe

Reilly

Joshua - Could have an impact on the chess world, but not really any social or ethical implications.

Jordan

# Jordan Archuleta-

What parts of the presentation were most effective?

Zoe

Reilly

Joshua - Definitely the graphs, the topic itself is interesting, but the graphs makes it tell a story

Jordan

Did you find any aspects of the presentation particularly confusing or unclear? If so, what might help to clarify those parts?

Zoe

Reilly
"Gun laws" are not clearly defined, and seem to require more research or elaboration to support claims made about the data.

Joshua - maybe go into more details about some of the columns? Not that big of an issue really.

Jordan

Are the prediction tasks being pursued easy to understand? Do they address the relevant data science questions?

Zoe

Reilly
The second data science question (about predicting via gun laws) doesn't seem to be supported by the dataset; how could this be addressed without outside data regarding gun laws?

Joshua - Yes I believe the questions are easy to understand.

Jordan

## Are the definitions of "success" and "failure" clear?

Zoe

Reilly
Lacking extensive model work as of yet, so not currently.

Joshua - model isn't finished yet so no definitions of success and failure

Jordan

## Are the models well justified?

Zoe

Reilly
Good idea on the first model prediction plan, that may be useful. The second one seems even more useful, and likely easier to correlate. The third one about gun laws I'm not sure on, mostly due to the amorphous definition of that variable to begin with.

Joshua - has a good start, but unknown to how it will play out

Jordan

## Are any unjustifiable conclusions drawn?

Zoe

Reilly
No real conclusions yet, though the murder vs. prisoners plot is spread out a bit too much to draw too many useful conclusions from I think; I find the interesting part of it to be the year colors on the dots, but the points themselves only form a fairly slight pattern.

Joshua - not yet

Jordan

## Are data science questions clear and forward looking? Could they form the basis of more formal falsifiable hypotheses?

Zoe

Reilly

Joshua

Jordan

## Can you think of any other models or techniques that could be useful in addressing the data science questions?

Zoe

Reilly

Some more EDA would be helpful, exploring demographic and economic factors in the database in particular. Work done here can be exceptionally useful in narrowing down model parameters.

Joshua - No not really

Jordan

## Can you think of any potential social or ethical implications of the project?

Zoe

Reilly

The results of this data may bring some information to the table that makes people uncomfortable, such as the idea that more prisoners means higher murder rates. American society doesn't take kindly to the idea of not treating prisoners like abominable fifth-class citizens.

Joshua - Definitely has a social implications involving some laws as to see which play has the highest crime.

Jordan

# Joshua Vong

## What parts of the presentation were most effective?

Zoe: Good Plots with a lot of colors, pretty easy to understand.

Reilly
Great plots! Very colorful and varied. Using things like xlim() or ylim() might help make the data more readable. This may be especially useful on the "solar_radiation_langley" plot, to truncate the outliers.

Joshua

Jordan - Plots are nice and visually easy to understand.

## Did you find any aspects of the presentation particularly confusing or unclear? If so, what might help to clarify those parts?

Zoe-some of the graphs could use more variation in color to make them easier to read for the reader.

Reilly
Being unfamiliar with the domain variables and with no data science questions, it's a bit tricky to quickly get in the headspace of the domain; with solid data science questions this should be easy to remedy.

Joshua

Jordan

## Are the prediction tasks being pursued easy to understand? Do they address the relevant data science questions?

Zoe- there are opportunities to be more forward looking, such as using the amount of toxicity in the fire, and predicting other factors based off that.

Reilly
Predicting the target may be a potential area of interest; the process for that would likely look similar to the Naive Bayes' Theorem model we did on A5, unless you could find a way to get a linear or exponential model to fit the data. Hopefully you do, since Naive Bayes' is a bit of work…

Joshua

Jordan

## Are the definitions of "success" and "failure" clear?

Zoe

Reilly

Joshua

Jordan - Not clear yet.

## Are the models well justified?

Zoe-there aren't any models yet, but he wants to predict some of the targets with some of the variables.

Reilly
No models, but intriguing plans in the works regarding the target variable.

Joshua

Jordan

## Are any unjustifiable conclusions drawn?

Zoe-not yet

Reilly
No conclusions yet!

Joshua

Jordan - Not yet

## Are data science questions clear and forward looking? Could they form the basis of more formal falsifiable hypotheses?

Zoe-there aren't any written data science questions yet, but he has a good idea of some that he could use.

Reilly

Joshua

Jordan - No data science questions yet. But you could totally predict target based upon the soil temps or really most of the variables that you have in there.

## Can you think of any other models or techniques that could be useful in addressing the data science questions?

Zoe-I think it would be helpful to focus on the area of the fires, and you can help predict which areas are more susceptible.

Reilly
Agree with Zoe; looking at specific focus areas and doing EDA on them might show some key patterns in the data.

Joshua

Jordan - Definitely look into the assignment we did and at the R for data science book that covers simple models. They have a ton of good examples to use, that could be very useful for your data.

## Can you think of any potential social or ethical implications of the project?

Zoe-This work could potentially help firefighters know where to make new fire stations, or help predict how high the temperature of a fire will go based on other variables.

Reilly

This could be an ethically fantastic project, as being able to predict the toxin content of smoke given the location or temperature, or being able to even predict likely fire areas, could save many lives.

Joshua

Jordan

-Jordan A.

## What parts of the presentation were most effective?

Zoe-I thought the graphs were very neat and effective

Reilly

Joshua

Jordan

## Did you find any aspects of the presentation particularly confusing or unclear? If so, what might help to clarify those parts?

Zoe-Reilly brought up what is classified as a gun law? Does this mean that there are laws against anyone owning a gun or just simply that there are no concealed carry licenses etc.

Reilly

Joshua

Jordan

## Are the prediction tasks being pursued easy to understand? Do they address the relevant data science questions?

Zoe-there are not any prediction tasks yet, I think an interestding one to use might be the prisoners incarcerated, and the rate of murder in the area. It is interesting to me that the number of murders has gone down over the years.

Reilly

Joshua

Jordan

## Are the definitions of "success" and "failure" clear?

Zoe-there is not really definitions of successes and failures yet.

Reilly

Joshua

Jordan

## Are the models well justified?

Zoe-not models yet

Reilly

Joshua

Jordan

## Are any unjustifiable conclusions drawn?

Zoe-not yet.

Reilly

Joshua

Jordan

## Are data science questions clear and forward looking? Could they form the basis of more formal falsifiable hypotheses?

Zoe-I do think that they are clear and forward looking.

Reilly

Joshua

Jordan

## Can you think of any other models or techniques that could be useful in addressing the data science questions?

Zoe

Reilly

Joshua

Jordan

## Can you think of any potential social or ethical implications of the project?

Zoe-there could be many social implications of this, lawmakers and politicians could use this data to argue policy for sure.

Reilly

Joshua

Jordan

-Reilly.

What parts of the presentation were most effective?

Zoe-the actual set up of the project in HTML is soooo organized. I almost want to change mine to be that way. I thought the graphs were well organized, as well as his explanation of them.

Reilly

Joshua

Jordan

Did you find any aspects of the presentation particularly confusing or unclear? If so, what might help to clarify those parts?

Zoe-

Reilly

Joshua

Jordan

Are the prediction tasks being pursued easy to understand? Do they address the relevant data science questions?

Zoe-

Reilly

Joshua

Jordan

Are the definitions of "success" and "failure" clear?

Zoe-.

Reilly

Joshua

Jordan

## Are the models well justified?

Zoe-no models yet

Reilly

Joshua

Jordan

## Are any unjustifiable conclusions drawn?

Zoe-not yet

Reilly

Joshua

Jordan

## Are data science questions clear and forward looking? Could they form the basis of more formal falsifiable hypotheses?

Zoe- I think they could be forward looking, they just need to make sure which variables they are going to use are clear.

Reilly

Joshua

Jordan

## Can you think of any other models or techniques that could be useful in addressing the data science questions?

Zoe- I like the idea of trying to predict who wins the game

Reilly

Joshua

Jordan

## Can you think of any potential social or ethical implications of the project?

Zoe- The implications could be more withdrawals while playing if it is easier to predict who will win off the first few moves.

Reilly

Joshua

Jordan