

Data Analytics Research Individual Final Project Report Draft

Amy Enyenihi

Fall 2023

Contents

DAR Project and Group Members	1
Instructions (DELETE BEFORE SUBMISSION)	2
More thoughts on your notebook submission	2
NOTEBOOK SUBMISSION CHECKLIST: Things to check before you submit (DELETE BEFORE SUBMITTING)	3
Instructions: Submitting a Notebook via github (DELETE BEFORE SUBMISSION)	4
Notebook User Guide (DELETE BEFORE SUBMISSION)	4
Abstract	4
Introduction and Background	5
Problems Tackled	5
Data Description	5
Data Analytics Methods	6
Experimental Results	6
Discussion of Results and Key Findings	6
Conclusions	6
Directions for Future Investigation	7
Bibliography	7
Files and Github Commits	7
Contribution	7
Appendix	8

DAR Project and Group Members

- Project name: *Hockey Analytics*
- Project team members: *Amy Enyenihi, Jeff Jung, Caleb Smith, Ashley Woodson, Lieben Zhang*

Instructions (DELETE BEFORE SUBMISSION)

- Consider the document your are creating to be a *research paper* with embedded code or, if this is not possible, links to where the embedded code can be found.
- This document should “read” like a standalone scientific report written in complete sentences (i.e. not bullets) using good rules of grammar. It should be readable as a paper even if all the code is not shown, and if only the results of running your code are shown.
- The document should be written for someone who is not already familiar with the project. As such you will need to describe the basics of the problem. You need to describe the basics of the data. You should have sufficient details for scientific reproducibility including documentation of the code. You will need to describe the analysis methods that can be used together with the code to reproduce your work. This is especially important if you use several R files. *Ideally someone should be able to go back years later and understand exactly what you did and reproduce your results.*
- A suggested report structure is given below, but you can customize this to meet the needs of your project. You must make sure the elements below – data description, problems tackled, methods, results, discussion, and future work – are addressed even if you change the organization. If you tackled multiple separate projects feel free to create a section for each problem and then separately discuss the data, problems tackled, methods, results and discussion. Make sure to provide a final summary and future work at the end.
- **ProTip: Create dataframes once:** If you have code chunks with long run times that create data frames, you can execute them once, and save them to an Rds. Comment out that code for subsequent runs or indicate in the code chunk header that the code should not execute; simply read in the Rds when knitting your notebook.
- If you have other documents that are part of your submission then put them on github and link to them from this document. You can provide excerpts or screen shots of the external files and reference the longer version.

More thoughts on your notebook submission

- Every student’s final project notebook should be written individually or, in some cases, as a small group. You must discuss with the instructor if you think you should do a small group notebook; otherwise assume you’re creating an individual until approved.
- As noted above, your final notebook serves as a written presentation of your work this semester so it must be written like a research paper. You should use proper R Markdown code chunk syntax to hide code chunks that don’t need to be shown. **Most of your code does not need to be visible in your final knitted report!**
- The R code that executes the results should be embedded in this notebook if possible.
 - It’s also okay to “source” external scripts from within your notebook, including those provided by your instructors.
 - PLEASE make sure all source code is in your team’s github repository!
- Fall 2023 students may have work that is not appropriate to be embedded on your final notebook
 - You should describe the work in the notebook and provide figures generated elsewhere (e.g. screen shots, graphs).
 - Indicate if that work has been committed to github. If necessary put details in Appendix including the names of the committed files.
- Your writing style should be suitable for sharing with external partners/mentors and useful to future contributors. Do not assume that your reader is familiar with the problem. Again, write as if this is a research paper.
- Focus on results; please don’t summarize everything you did this semester!

- Discuss only the *most important* aspects of your work.
- Ask yourself *what really matters?*
- **IMPORTANT:** Discuss any insights you found regarding your research.
- If there are limitations to your work, discuss, in detail.
- Include any **background** or **supporting evidence** for your work.
 - For example, mention any relevant research articles you found – and be sure to include references!
- Include your recommendations for continuing or extending your project
 - Data utilization...
 - Analytics...
 - Visualizations
 - User interface design...
 - Apps...
 - Open questions that could be investigated...

NOTEBOOK SUBMISSION CHECKLIST: Things to check before you submit (DELETE BEFORE SUBMITTING)

- Have you given your file(s) a clear, sensible name?
 - `my_notebook.Rmd` or `joey.Rmd` are ****not*** acceptable!
- Is your document readable as a research paper even if all the code is suppressed?
 - Try suppressing all the code using hint below and see if this is true.
- Did you proofread your document? Does it use complete sentences and good grammar?
- Is every figure clearly labeled and titled?
- Does every figure serve a purpose?
 - Does the figure have a useful title? **Hint:** What *question* does the figure answer?
 - You can put extra (non-essential) figures in your **Appendix**.
 - Is the figure captioned?
 - Are the figure and its associated findings discussed in the text?
 - Is it clear which figure is being discussed? **Hint:** use captions!
- Are your tables clearly labeled and legible?
 - We recommend using `kable()` (built into `knitr`) or the `gt` (“Grammar of Tables”) package
 - `xtable()` might also work for you (not as easy as `kable()` and not as elegant as `gt`)
 - Is the table and what its associated findings discussed in the text?
 - Is the table titled and captioned?
 - Are the table and its associated findings discussed in the text?
 - Is it clear which figure is being discussed (hint use the captions)?
- **CRITICAL:** Have you given enough information for someone to reproduce, understand and extend your results?
 - Where can they *find* the data that you used?
 - Have you *described* the data that used?
 - Have you *documented* your code?
 - Are your figures *clearly labeled*?
 - Did you *discuss your findings*?
 - Did you use good grammar and *proofread* your results?
 - Finally, have you *committed* your work to github and made a *pull request*?
 - Did you indicate the github *issues* that you addressed?
- Use this notebook as a template for your FINAL project notebook.
- Use the sections starting with **Work Summary** as your outline for your submitted notebook.
- Summarize ALL of your work that is worthy of being preserved in this notebook; Feel free to include work in the appendix at end. It will not be judged as being part of the research paper but rather as additional information to be preserved. **if you don’t show and/or link to your work here, it doesn’t exist for us!**

Instructions: Submitting a Notebook via github (DELETE BEFORE SUBMISSION)

1. Create a new copy of this notebook in the appropriate assignment sub-directory of your team's github repository using the following naming convention
 - `dar_final_rcsid_ddmmmyyyy.Rmd` and `dar_final_rcsid_ddmmmyyyy.html`
 - For example, `dar_final_erickj4_1nov2021.Rmd`
 - Look for `FinalProject` under `StudentNotebooks`
2. Delete the "INSTRUCTIONS" section!!
3. You **MUST** include figures and/or tables to illustrate your work. *Screen shots are okay*; exported PNGs are preferred.
4. You **MUST** include links to other important resources (knitted HTML files, Shiny apps). See the guide below for help.
5. Commit the source (`.Rmd`), pdf (`.pdf`) and knitted (`.html`) versions of your notebook and push to github. Turn in the pdf version to lms.
6. **Submit a pull request.** Please notify Dr. Erickson if you don't see your notebook merged within one day.
7. **DO NOT MERGE YOUR PULL REQUESTS YOURSELF!!**

See LMS for guidance on how the contents of this notebook will be graded.

DELETE THE SECTIONS ABOVE!

Notebook User Guide (DELETE BEFORE SUBMISSION)

R Notebooks are meant to be dynamic documents. Provide any relevant technical guidance for users of your notebook. Also take care of any preliminaries, such as required packages. Sample text:

This report is generated from an R Markdown file that includes all the R code necessary to produce the results described and embedded in the report. Code blocks can be suppressed from output for readability using the command code `{R, echo=show}` in the code block header. If `show <- FALSE` the code block will be suppressed; if `show <- TRUE` then the code will be show.

```
# Set to TRUE to expand R code blocks; set to FALSE to collapse R code blocks
show <- TRUE
```

Executing this R notebook requires some subset of the following packages:

- `ggplot2`
- *List other packages...*

These will be installed and loaded as necessary (code suppressed).

Abstract

Briefly summarize your project and findings. A good rule of thumb is to write your abstract *after* you've prepared most of the notebook.

-will add after the rest of the notebook is complete.

Introduction and Background

Frame your project, including any relevant background that argues for the significance of the work, previous approaches taken by others, and motivation for the approach taken by you and your team. Note the goal of this document is to create a research paper that documents your primary research and either includes the code.

The goal of this project was to manipulate data from the RPI Women's Hockey team to create a meaningful product for hockey players. RPI's IDEA tracking tool tracks players' position on the rink frame by frame. The data at hand was generated from 130+ shots on goal plays from the 2022-2023 season.

Code

Problems Tackled

Discuss the goals of your work and the questions you addressed. Describe how the rest of your final project report is organized.

The initial and overall ask was to determine what variable or combination of variables has the greatest impact on not scoring? Because all data was generated with RPI on defense, we were scouting for defensive success.

Code

Data Description

Provide a detailed discussion of the dataset, including:

- IDEA Tracking tool gathered positions of players on the ice. The data was saved into CSV files.
- The original data set contained 12 variables, but by the end of the semester it grew to at least 14.
- Whether the features are categorical, continuous, or a mixture of the two
- The following variables are all numerical:
 - Puck distance from the center of the goal crease at the point of the shot
 - Puck speed at the start of its trajectory
 - Puck trajectory angle (in rink coordinate system) w.r.t. the center of the goal crease
 - Shooter puck possession time prior to taking the shot
 - Shooter skating speed at the point of the shot
 - Goalie distance from the center of the goal crease at the time of the shot
 - Angle between goalie location and point of the shot
 - Angle between a given player and the puck at a given point along the puck trajectory
- The following variables are categorical:
 - Handedness of the shooter
 - Goalie screened/not screened by player(s) from either team
 - Number of offensive team players participating in the play
 - Number of defensive team players participating in the play
- The CSVs were processed by Mohammed prior to the beginning of the fall semester. The resulting data frame was used and expanded on by my group.

```
# Code
```

Include an infographic overview of dataset — a visualization of its structure.

```
# Code
```

```
#view header of data set
```

Data Analytics Methods

Describe the data analytics methods you used and why you chose them.

- What machine learning techniques did you use?
- Were you able to use pre-existing implementations?
- If the techniques required user-specified parameters, how did you choose what parameter values to use?
- Discuss your data analytics “pipeline” from *data preparation* and *experimental design*, to *methods*, to *results*.

I did not use machine learning for my portion of the project.

(explain how visualization work lead to the app idea, include images of the earlier rink plots)

The app I created uses the data that has been through Jeff’s preparation methods. (mention said methods + how I altered his resulting data frame)

```
# Code
```

Experimental Results

Include a summary of the results obtained, including (if relevant):

- Training and testing errors
- Relevant visual illustrations of findings

(will go over puck distance example and shooter speed case example)

```
# Code
```

Make sure all figures are clearly labelled; always use meaningful titles (please)! *Hint: In Data INCITE we highly recommend that you include a short caption with each figure to explain the contents of the figure. This not only helps the reader, but also helps dramatically with re-use of the notebook code.*

Discussion of Results and Key Findings

Discuss the significance of your experimental results

- What are your key findings?
- Why are those findings important?

(The app can reveal the obvious [puck distance] as well as the unexpected [shooter speed])

```
# Code
```

Conclusions

What conclusions and recommendations can you make as a result of this research?

-What features or combination of features increases defensive success? (No goal)

- far shots do not score
- show relationship about number of players
- shooting speed case example

Directions for Future Investigation

Discuss proposed future work and possible extensions of your project.

- Have the hockey team use the app to do analysis
- Use feedback to further develop the dashboard
- Use additional data from the added cameras
- Increase sample size to improve analysis accuracy
- Compare data from RPI's defense against the away team's defense
- Add a camera to the other side of the rink
- Draw conclusions to improve RPI's offense in addition to defense

NOTE: Feel free to interleave problems, methods, results, and discussion sections for different parts of your project.

Bibliography

Provide a listing of references and other sources.

- Citations from literature. Give each reference a unique name combining first author last name, year, and additional letter if required. e.g.[Bennett22a]. If there is known author, make something reasonable up.
- Significant R packages used

- Reference ColorBrewer (not used above but will be in additional work in appendix)
- list packages used

Files and Github Commits

Commit your final files and any supporting files to github. List your files and their locations here and explain purpose of each one. Make sure to include all files necessary to reproduce your results.

- add link to AnalysisCodeFunc.R (functions used for graphing)
- add link to raster jpeg (used for graphing)
- add link to UpdatedFactoredDiscretizedData.Rds (data frame used for app)
- add link to my app

Contribution

If this is joint work, then discuss the contribution of each teammate.

- Jeff: did the data cleaning for the variable categorization

Appendix

Include here whatever you think is relevant to support the main content of your notebook. For example, you may have only include example figures above in your main text but include additional ones here. Or you may have done a more extensive investigation, and want to put more results here.