

Taxi Predictive Model Case Study Rubric

DS 4002 – Siobhan Flood

Submission format: Upload link to GitHub repository on UVA Canvas

Individual Assignment

Why am I doing this?

Following this case study challenges your skills as a data scientist and prepares you for future coursework and career opportunities. This project specifically tests your coding abilities and your experience working with time series models. You will practice building predictive models, analyzing data, and becoming more comfortable with different modeling approaches.

What am I going to do?

The GitHub repository for this case study can be found [here](#).

You will obtain the data from the Chicago Data Portal, available [here](#). The GitHub repository also contains the original CSV and a cleaned version of the data.

After obtaining and cleaning the data (or using the provided cleaned data), you will create a graph to explore taxi ridership during the month of October using R. Once you understand the trends, you will select two predictive time series models, specifically SARIMA and BSTS, to compare.

You will:

- Train both models
- Test their predictive accuracy on October ridership
- Evaluate and compare their performance.

Your final deliverables should include:

- The original and cleaned datasets you used,
- A data dictionary,
- Graphs showing ridership over time,
- Well-documented and commented source code,
- A GitHub repository containing all project materials.

Tips for success:

- Research SARIMA and BSTS models to understand how they work, so you can build them effectively and interpret your evaluation results.
- Organize your GitHub repository using folders — avoid putting too many files in one place to keep it readable and easy to navigate.
- Get comfortable using R, as it will be the main language for this project.
(Also, remember to install the **BSTS** package in R.)

How will I know I have succeeded? You will meet expectations on this case study when you successfully follow and complete the criteria in the rubric below:

Spec Category	Spec Details
Formatting	<p>One GitHub repository (submitted via link on Canvas)</p> <ul style="list-style-type: none">▪ Create a new GitHub repository for this assignment titled ‘CS2_Taxi Predictive Modeling’ that contains▪ README.md▪ LICENSE.md▪ Source Code File▪ Your data (original and cleaned)▪ REFERENCES.md
README.md	<ul style="list-style-type: none">● Goal: This file serves as an orientation to everyone who comes to your repository, it should enable them to get their bearings.● Use markdown headers to divide content.● Make an H2 (##) section explaining the contents of the repository● Section 1: Software and platform section<ul style="list-style-type: none">○ The type(s) of software you used for the project.○ The names of any add-on packages that need to be installed with the software.○ The platform (e.g., Windows, Mac, or Linux) you used.● Section 2: A Map of your documentation. <p>In this section, you should provide an outline or tree illustrating the hierarchy of folders and subfolders contained in your Project Folder, and listing the files stored in each folder or subfolder.</p>

Source Code File	<ul style="list-style-type: none"> ● Goal: This folder contains all the source code for your project. Include all the scripts you used. <p>In this project the source code must include:</p> <ul style="list-style-type: none"> ● Code for cleaning the Data ● Code for building the BSTS model ● Code for building the SARIMA model ● Code for building the BSTS model ● Code for evaluating the BSTS and SARIMA model ● Graph detecting patterns in ridership data <p>● All script files should include header comments at the beginning of a script to provide information that anyone working with or executing the script should be aware of. Throughout all your scripts, you should include copious comments explaining what each command or sequence of commands accomplishes and what the purpose is</p>
REFERENCES.md	<ul style="list-style-type: none"> ● All references should be listed in the README.md ● Use IEEE Documentation style

Acknowledgements: Thank you to Professor Alonzi for providing the rubric structure.