

Item	Description
Title	Include the name of your project, the team members who contributed to the project, and your assigned project group number (example: group7).
Table of Contents	table of contents for the entire report which includes the section numbers.
Abstract	<ul style="list-style-type: none"> <li>• Provide a high-level overview of your project</li> <li>• Brief data set description</li> <li>• Link to data set</li> <li>• Provide a bulleted or numbered of interesting / surprising data exploration insights you found</li> <li>• Provide a bulleted or numbered list of the specific predictions you made</li> <li>• Provide a bulleted or numbered list of the specific inferences you made.</li> <li>• Provide a brief conclusion summary outlining the success / failure of your predictions and inference</li> <li>• this should be 1 block and have no supporting code or graphs</li> </ul>
Data Collection/ Cleaning	<ul style="list-style-type: none"> <li>• Provide brief details about what you had to do to clean the data. For example, did the data have a lot of NA values. Did you drop whole columns of data or whole rows of data. How did you handle missing values?</li> </ul>
Data Exploration Insights Using Standard Statistical Techniques – Excluding Machine Learning	<ul style="list-style-type: none"> <li>• Describe the results of your data exploration.</li> <li>• Provide some data exploration visualizations.</li> <li>• Tell me something interesting you learned about the data.</li> <li>• Maybe provide some statistical summaries</li> <li>• Maybe look at feature correlation</li> <li>• Example data exploration insight for a group using an airline flight delay dataset: We found that there was a higher frequency of delayed flights which originated at Chicago O'Hare airport as compared to other airports. We found evidence that these delays were caused mainly by ...</li> <li>• Do not include model inference in this section. Machine learning model inference goes in it's own section described below.</li> </ul>
Methodology	
Model Prediction	<ul style="list-style-type: none"> <li>• A good technique for this section is to put each model in its own sub section with a descriptive title like Linear Regression Prediction of Flight Delay. A lot of the requested information can be efficiently conveyed with a table (the same general table can be used for each model). Or provide a combined summary table and text that describes the combined summary table.</li> <li>• For each summary table, include the following: <ul style="list-style-type: none"> <li>o Value being predicted.</li> <li>o Model type: Random Forest, Linear Regression, neural network</li> <li>o Scoring metric: F1, AUC, MSE, RMSE, etc.</li> <li>o If cross validation was used. Number of folds if used</li> <li>o Resulting Model Score</li> </ul> </li> <li>• Provide brief text explanations for each table. Describe things like transformations or feature engineering that was not common to all models and not described in the methodology section. Describe things particular to this model that needed to be performed in order to meet your goals.</li> <li>• There is no need to overly complicate things in this section. Don't make things more complicated than they need to be.</li> </ul>

Item	Description
Model Inference	<ul style="list-style-type: none"> <li>• Same rules as for prediction above: A good technique for this section is to put each model in its own sub section with a descriptive title like Linear Regression Inference of Flight Delay. A lot of the requested information can be efficiently conveyed with a table (the same general table can be used for each model). Or provide a combined summary table and text that describes the combined summary table.</li> <li>• For each summary table, include the following in some sorted order: <ul style="list-style-type: none"> <li>o Feature Name</li> <li>o Feature metric being compared: Example: Regression coefficient, gini index, entropy, etc.</li> <li>o Scoring metric: F1, AUC, MSE, RMSE, etc.</li> <li>o If cross validation was used indicate that fact and the number of folds used to score your model.</li> </ul> </li> <li>• For each table, provide a brief text explanation which highlights the following. Provide a bullet or numbered list for each of the following: <ul style="list-style-type: none"> <li>o The inference goal</li> <li>o What you learned through your analysis</li> <li>o Any special transformations / feature engineering you had to perform in order to correctly perform the inference. For example, transformed the data using a standard scaler so regression coefficients representing real data could be compared between each other.</li> <li>o It is encouraged to stick with inference techniques covered in class.</li> <li>o There is no need to overly complicate things in this section. Don't make things more complicated than they need to be.</li> </ul> </li> <li>• 2 to 3 pages are advised for this section</li> </ul>
Conclusion	<ul style="list-style-type: none"> <li>• Describe the results of the project by expanding upon the brief conclusion results outlined in the abstract. Summarize model comparison results in a prediction summary table. Summarize inference results in an inference summary table. Sort the summary tables. Use table column names which are similar to columns in previous sections. Maybe add sub sections for prediction and inference summaries. Provide results for every prediction and inference item presented in the abstract.</li> </ul>
Appendix	<ul style="list-style-type: none"> <li>• Include references you used. For example, if there are specific web sites or books you used, include them here and link to your references from the document.</li> <li>• You can include any additional information you consider important that might not be appropriate or too large to include in the main document.</li> <li>• The appendix is not the place to put lots of tables and figures that didn't fit elsewhere</li> </ul>