

CS3 – Airline Scorecard Rubric

DS 4002 – David Nu Nu Marip

Why are you doing this? Which major US airline is most likely to make passengers arrive late? You will practice working with a large, messy dataset, defining a clear performance metric, and turning your analysis into a recommendation that a non-technical reader can understand. This work reinforces skills in data preparation, exploratory analysis, visualization, and written communication.

What you are going to do:

1. Start from the provided Airline Scorecard GitHub repository and follow the data preparation instructions to obtain or load airline on time performance data for domestic US flights
2. Create a clean analysis dataset that focuses on nine major US airlines and includes an indicator for whether each flight arrived at least 15 minutes late.
3. Compute and compare delay rates across airlines and decide how to handle cancelled flights.
4. Build an airline scorecard that shows which airlines are more or less likely to make passengers late.
5. Compare your results to public rankings and summaries of airline performance provided in the supplemental materials.
6. Write a short report that explains your methods, presents your scorecard, and answers the main question for a typical traveler.

How will I know I have Succeeded? You will meet expectations on this case study when you meet the criteria in the rubric below.

Spec Category	Spec Details
Formatting	<ul style="list-style-type: none">• You work in a single GitHub repository for this case study.• The top level of the repository is organized and contains at least:<ul style="list-style-type: none">○ README.md○ DATA folder○ SCRIPTS folder○ SUPPLEMENTAL_MATERIALS folder○ Your written report• File and folder names are descriptive so that another student can quickly see where to start and where key pieces of the project live.
Data and Data Preparation	<ul style="list-style-type: none">• You use the airline on time performance data described in the case study materials.

	<ul style="list-style-type: none"> ● Your analysis focuses on domestic US flights for the specified years and nine major US airlines. ● You apply the filters described in the preparation guide, such as excluding diverted flights. ● You ensure that key variables exist and are usable, including: <ul style="list-style-type: none"> ○ Airline codes and names ○ Flight date ○ Arrival delay in minutes ○ A binary indicator for flights arriving 15 or more minutes late ○ A cancelled flag ● You run basic checks such as counts by year and airline and simple summaries of delay rates to confirm that the dataset looks reasonable. ● Any important choices or deviations from the preparation guide are briefly documented in comments or in your report.
Analysis and Methods	<ul style="list-style-type: none"> ● You define a clear metric for delay risk, such as the proportion of flights with arrival delay of 15 or more minutes for each airline. ● You state explicitly how you handle cancelled flights in your calculations and give a short justification. ● You compute delay metrics correctly for each airline and, if used, for any subgroups such as year or distance group. ● Any extra analyses you perform are connected to the main question and help you understand which airlines are more likely to make passengers late. ● Your analysis is transparent. A reader can see how you went from raw data to final metrics.
Scorecard and Visualizations	<ul style="list-style-type: none"> ● You create at least one main figure or table that functions as your airline scorecard. ● The scorecard makes it easy to compare airlines and see which ones have higher or lower delay rates. ● Visuals have clear titles, axis labels, and legends where needed. ● The design of the scorecard is simple and readable. A non-technical reader could look at it and understand the main message without reading your code. ● Additional plots, if included, support your argument and are not redundant.

Interpretation and Comparison to Benchmarks	<ul style="list-style-type: none"> • You clearly answer the core question: which airline in your dataset is most likely to make passengers late. • You interpret your delay metrics in plain language. You explain what the numbers mean for someone choosing an airline. • You compare your findings to at least two external references from the supplemental materials, such as official rankings or public articles about airline performance. • You note where your results agree with those references and where they differ. • You offer reasonable explanations for differences, such as different time periods, different sets of airlines, or different ways of treating cancellations.
Spec Category: Written Report	<ul style="list-style-type: none"> • Your report is concise and organized with a clear structure, for example: <ul style="list-style-type: none"> ○ Introduction and question ○ Data and methods ○ Results and airline scorecard ○ Comparison to benchmarks ○ Limitations and recommendation • The writing uses clear, direct sentences and smooth transitions between sections. • Technical terms are explained the first time they appear so that a second year student or interested non technical reader can follow. • You discuss at least two important limitations of your analysis, such as coverage of the data, possible biases, or modeling choices. • You end with a concrete recommendation for travelers and a short statement about how confident you are in that recommendation.
Code Quality and Reproducibility	<ul style="list-style-type: none"> • Your notebook or scripts in the SCRIPTS folder run from top to bottom without errors when the required data files are in place. • Code is grouped into logical sections with short headings and comments. • You avoid unnecessary repetition and keep intermediate outputs manageable. • Another student could clone your repository, follow the instructions in your README, and recreate your main tables and figures.

References and Supplemental Materials	<ul style="list-style-type: none"> • You use the articles and documentation in the SUPPLEMENTAL_MATERIALS folder to understand airline performance and definitions of on time. • You refer to these materials when comparing your results to public rankings or explanations. • You include a simple reference list in your report that cites any external data sources or articles you rely on. • If you add any new articles or notes, you place them in the SUPPLEMENTAL_MATERIALS folder and label them clearly.
---------------------------------------	--

Note: I just copied Prof. Alonzi original CS3Rubric.docx and made edit to them.