

Netflix Advanced Data Science Boot Camp – Midterm Project Grading Rubric

Instructions for the instructional team: Evaluate the student's submitted midterm project assignment and presentation against the outlined criteria in the rubric table in this document, and assign a rating to each criterion. But, assign an Incomplete if any of the following occur:

- No submission was received.
- The submission is empty or blank.
- The submission contains evidence of academic dishonesty.

Note to the instructional team: We encourage the students to collaborate and share ideas during the project weeks. So, you might notice shared code, documentation, and written explanations across the student submissions. This is acceptable and should not be a consideration when assigning a rating to the student's performance.

Midterm Project Rubric

Criterion	Demonstrating proficiency (30 points)	Approaching proficiency (26 points)	Developing proficiency (22 points)	Emerging (18 points)
Phase I: Data Modeling and ETL	<ul style="list-style-type: none">• There is a final Jupyter notebook that thoroughly describes the data extraction and transformation process, and the CSV files don't contain errors when they are exported.• A test suite with four passing tests is created that tests the functionality of the data extraction or transformation process.• A database is designed, the schema creates the tables with the appropriate relationships, and data is loaded into the tables without error.	<ul style="list-style-type: none">• There is a final Jupyter notebook that adequately describes the data extraction and transformation process, and the CSV files don't contain errors when they are exported.• A test suite with three passing tests is created that tests the functionality of the data extraction or transformation process.• A database is designed, the schema creates the tables but some relationships are missing, but the data is loaded into the tables without errors.• Advanced SQL queries are written that answer three	<ul style="list-style-type: none">• There is a final Jupyter notebook that partially describes the data extraction and transformation process, but the CSV files contain errors when they are exported.• A test suite with at least two passing tests is created that tests the functionality of the data extraction or transformation process.• The database is designed, the schema creates the tables, there are errors in the relationships, and some CSV files can't be loaded into the tables.• Advanced SQL queries are written that answer one to two	<ul style="list-style-type: none">• The final Jupyter notebook is incomplete and doesn't describe all data extraction and transformation processes, and the CSV files contain errors when they are exported.• A test suite with at least one passing test is created that tests the functionality of the data extraction or transformation process.• The database is poorly designed, the schema creates some tables, there are no relationships, and the CSV files can't be loaded into the tables.• There are no advanced SQL queries written, or the SQL

Netflix Advanced Data Science Boot Camp – Midterm Project Grading Rubric

	<ul style="list-style-type: none"> Advanced SQL queries are written that answer all four questions of the data. The final repository presents with professional quality. 	<p>questions of the data.</p> <ul style="list-style-type: none"> The final repository is acceptable but needs more work toward professional quality. 	<p>questions of the data.</p> <ul style="list-style-type: none"> The final repository needs ample work toward professional quality. 	<p>queries are very basic.</p> <ul style="list-style-type: none"> The README file is empty.
Criterion	Demonstrating proficiency (30 points)	Approaching proficiency (26 points)	Developing proficiency (22 points)	Emerging (18 points)
Phase II: Migrate ETL Pipeline to the Cloud and Test for Quality Assurance	<ul style="list-style-type: none"> The final Google Colab Jupyter Notebook runs without errors to extract and transform the data using PySpark or SparkSQL. Code is written to store data in parquet format, partitioned, or cached and query execution times improve substantially. Four successful assertions are written to test for quality assurance during the extraction or transformation process. 	<ul style="list-style-type: none"> The final Google Colab Jupyter Notebook has a few minor errors but is able to extract and transform the data using PySpark or SparkSQL. Code is written to store data in parquet format, partitioned, or cached but query execution times slightly improve. At least three successful assertions are written to test for quality assurance during the extraction or transformation process. 	<ul style="list-style-type: none"> The final Google Colab Jupyter Notebook has errors and the extraction and transformation process using PySpark or SparkSQL isn't complete. There is an attempt to write code to store the data in parquet format, partitioned, or cached. At least two successful assertions are written to test for quality assurance during the extraction or transformation process. 	<ul style="list-style-type: none"> The final Google Colab Jupyter Notebook has errors and the extraction and transformation process using PySpark or SparkSQL isn't complete. There is no attempt to write code to store the data in parquet format, partitioned, or cached. At least one successful assertion is written to test for quality assurance during the extraction or transformation process.
Criterion	Demonstrating proficiency (10 points)	Approaching proficiency (8 points)	Developing proficiency (6 points)	Emerging (4 points)
Analysis and Conclusion	<ul style="list-style-type: none"> The write-up summarizes the major findings and implications at a professional level. Each question in the project 	<ul style="list-style-type: none"> The write-up sufficiently summarizes the findings and implications. Each question in the project proposal is answered but could 	<ul style="list-style-type: none"> The write-up lacks a sufficient summary, and the data and findings need more coherence. The question responses are 	<ul style="list-style-type: none"> The write-up doesn't summarize the findings and implications. No analysis exists. The question responses are

Netflix Advanced Data Science Boot Camp – Midterm Project Grading Rubric

	<p>proposal is answered with precise descriptions and findings.</p> <ul style="list-style-type: none"> The findings are strongly supported with tables. Each question response is supported with a well-discerned analysis from the lessons—for example, an aggregation, correlation, and comparison. 	<p>use more description and analysis.</p> <ul style="list-style-type: none"> The findings are sufficiently supported with numbers and visualizations. Each question response is supported with a statistical analysis but lacks accuracy in application. 	<p>minimal.</p> <ul style="list-style-type: none"> The visualizations don't relate to the questions. The question responses don't relate to the questions, and they demonstrate minimal analysis. 	<p>either missing or irrelevant.</p> <ul style="list-style-type: none"> Very few or no visualizations exist.
Criterion	Demonstrating proficiency (20 points)	Approaching proficiency (16 points)	Developing proficiency (12 points)	Emerging (8 points)
Slide Presentation	<ul style="list-style-type: none"> The slides appear clean and professional. The slides relate to the project. The slides effectively demonstrate the scope of the project. 	<ul style="list-style-type: none"> The slides look good but have some minor problems. The slides almost entirely relate to the project. The slides are mostly effective at demonstrating the scope of the project. 	<ul style="list-style-type: none"> The slides look okay, but the design could be improved. The slides somewhat relate to the project. The slides are somewhat effective at demonstrating the scope of the project. 	<ul style="list-style-type: none"> The slide designs have major issues. The slides don't relate to the project. The slides don't effectively demonstrate the scope of the project.
Criterion	Demonstrating proficiency (10 points)	Approaching proficiency (8 points)	Developing proficiency (6 points)	Emerging (4 points)
Individual Video Presentation	<ul style="list-style-type: none"> The video summary is about two minutes in length. The video concisely summarizes the purpose, results, and conclusions of the project. The sound is clear and 	<ul style="list-style-type: none"> The video summary is either two to three minutes or about one minute in length. The video mostly summarizes the purpose, results and conclusions of the project. The sound is mostly clear and 	<ul style="list-style-type: none"> The video summary is either three to five minutes or significantly shorter than one minute in length. The video somewhat summarizes the purpose, results and conclusions of the 	<ul style="list-style-type: none"> The video summary is more than five minutes long. The video doesn't summarize the purpose, results and conclusions of the project very well. The sound isn't clear or

Netflix Advanced Data Science Boot Camp – Midterm Project Grading Rubric

	<p>intelligible.</p> <ul style="list-style-type: none">• Titles and transitions are used effectively.• The presenter seems well prepared.• The presentation relates to the project.• The presentation maintains the interest of the audience.	<p>intelligible</p> <ul style="list-style-type: none">• Titles and transitions are mostly effective.• The presenter seems mostly prepared and encounters only minor difficulties.• The presentation almost entirely relates to the project.• The presentation mostly maintains the interest of the audience.	<p>project.</p> <ul style="list-style-type: none">• The sound is barely clear and intelligible.• Titles and transitions are used sparingly.• The presenter seems fairly well prepared but encounters some major difficulties.• The presentation mostly relates to the project.• The presentation barely maintains the interest of the audience.	<p>intelligible.</p> <ul style="list-style-type: none">• Titles and transitions aren't used• The presenter seems unprepared, and the presentation is scattered or confusing.• The presentation doesn't relate to the project.• The presentation doesn't maintain the interest of the audience.
--	--	---	---	---