



Data Boot Camp Grading Rubric

Unit 22 - Big Data Homework - "Alexa, can you handle big data?"

Instructions:

Evaluate the homework against the outlined criteria in the below rubric, assigning a rating to each criterion. Add points earned across all criteria and convert the total points to a letter grade, assigning a "+" or "-" letter grade designation at your discretion.

A (+/-)	35+	C (+/-)	15-24	F (+/-)	<5
B (+/-)	25-34	D (+/-)	5-14		

Notes:

The deployed assignment utilizes **AWS** and the **PySpark** library to complete 1 of 2 levels of challenges. The source code should also be deployed to **Github** or **Gitlab**. **The homework grade should be graded on Level 1 only.** Level 2 is **optional**, and the rubric is included only as reference.

Rubric for Big Data - Level 1:

	Mastery 20 points	Approaching Mastery 15 points	Progressing 10 points	Emerging 5-0 points	Incomplete
Extract	In both notebooks the student did all of the following: <ul style="list-style-type: none">✓ Connects to and loads in datasets from AWS to dataframes using pyspark✓ Correctly handles the header and has column names as the first row✓ Ensures the data is retrieved by outputting the head of the dataframe✓ Discovers the size of the dataframe by outputting the number of rows in it	In both notebooks the student did 3 of the following: <ul style="list-style-type: none">✓ Connects to and loads in datasets from AWS to dataframes using pyspark✓ Correctly handles the header and has column names as the first row✓ Ensures the data is retrieved by outputting the head of the dataframe✓ Discovers the size of the dataframe by outputting the number of rows in it	In both notebooks the student did 2 of the following: <ul style="list-style-type: none">✓ Connects to and loads in datasets from AWS to dataframes using pyspark✓ Correctly handles the header and has column names as the first row✓ Ensures the data is retrieved by outputting the head of the dataframe✓ Discovers the size of the dataframe by outputting the number of rows in it	In both notebooks the student did 0-1 of the following: <ul style="list-style-type: none">✓ Connects to and loads in datasets from AWS to dataframes using pyspark✓ Correctly handles the header and has column names as the first row✓ Ensures the data is retrieved by outputting the head of the dataframe✓ Discovers the size of the dataframe by outputting the number of rows in it	No submission was received -OR- Submission was empty or blank -OR- Submission contains evidence of



Data Boot Camp Grading Rubric

Unit 22 - Big Data Homework - "Alexa, can you handle big data?"

					academic dishonesty
Transform & Load	<p>Student does all of the following with the dataframes:</p> <p>Transform</p> <ul style="list-style-type: none"> ✓ Removed duplicate rows ✓ Kept and renamed only necessary columns to match the current database table schema ✓ Matched dataframe column types with the database column types <p>Load</p> <ul style="list-style-type: none"> ✓ Successfully pushed dataframes to AWS 	<p>Student does 3 of the following with the dataframes:</p> <p>Transform</p> <ul style="list-style-type: none"> ✓ Removed duplicate rows ✓ Kept and renamed only necessary columns to match the current database table schema ✓ Matched dataframe column types with the database column types <p>Load</p> <ul style="list-style-type: none"> ✓ Successfully pushed dataframes to AWS 	<p>Student does 2 of the following with the dataframes:</p> <p>Transform</p> <ul style="list-style-type: none"> ✓ Removed duplicate rows ✓ Kept and renamed only necessary columns to match the current database table schema ✓ Matched dataframe column types with the database column types <p>Load</p> <ul style="list-style-type: none"> ✓ Successfully pushed dataframes to AWS 	<p>Student does 0-1 of the following with the dataframes:</p> <p>Transform</p> <ul style="list-style-type: none"> ✓ Removed duplicate rows ✓ Kept and renamed only necessary columns to match the current database table schema ✓ Matched dataframe column types with the database column types <p>Load</p> <ul style="list-style-type: none"> ✓ Successfully pushed dataframes to AWS 	

Rubric for Big Data - Level 2 (Activity is optional):

	Mastery 20 points	Approaching Mastery 15 points	Progressing 10 points	Emerging 5-0 points	Incomplete
Extract & Transform	<p>The student did all of the following to extract and clean the data:</p> <ul style="list-style-type: none"> ✓ Connected to and loaded in datasets from AWS to dataframes using Spark ✓ Removed any unnecessary columns ✓ Dropped rows with null values ✓ Dropped duplicated rows 	<p>The student did 3 of the following to extract and clean the data:</p> <ul style="list-style-type: none"> ✓ Connected to and loaded in datasets from AWS to dataframes using Spark ✓ Removed any unnecessary columns ✓ Dropped rows with null values ✓ Dropped duplicated rows 	<p>The student did 2 of the following to extract and clean the data:</p> <ul style="list-style-type: none"> ✓ Connected to and loaded in datasets from AWS to dataframes using Spark ✓ Removed any unnecessary columns ✓ Dropped rows with null values ✓ Dropped duplicated rows 	<p>The student did 0-1 of the following to extract and clean the data:</p> <ul style="list-style-type: none"> ✓ Connected to and loaded in datasets from AWS to dataframes using Spark ✓ Removed any unnecessary columns ✓ Dropped rows with null values ✓ Dropped duplicated rows 	<p>No submission was received</p> <p>-OR-</p> <p>Submission was empty or blank</p>
Analysis	<p>Student does all of the following to analyze if "vine" reviews are trustworthy:</p> <ul style="list-style-type: none"> ✓ Splits the reviews between vine (paid) and non-vine (unpaid) ✓ Compares metrics between vine and non-vine reviews such 	<p>Student does all of the following to analyze if "vine" reviews are trustworthy:</p> <ul style="list-style-type: none"> ✓ Splits the reviews between vine (paid) and non-vine (unpaid) ✓ Compares metrics between vine and non-vine reviews such as, but 	<p>Student does 2 of the following to analyze if "vine" reviews are trustworthy:</p> <ul style="list-style-type: none"> ✓ Splits the reviews between vine (paid) and non-vine (unpaid) ✓ Compares metrics between vine and non-vine reviews such as, but 	<p>Student does 0-1 of the following to analyze if "vine" reviews are trustworthy:</p> <ul style="list-style-type: none"> ✓ Splits the reviews between vine (paid) and non-vine (unpaid) ✓ Compares metrics between vine and non-vine reviews such as, but 	<p>-OR-</p> <p>Submission contains evidence of academic dishonesty</p>



Data Boot Camp Grading Rubric

Unit 22 - Big Data Homework - "Alexa, can you handle big data?"

	<p>as, but not limited to:</p> <ul style="list-style-type: none">Number of reviewsNumber of 5-star reviewsAverage RatingNumber of helpful votes <p>✓ Comes up with a conclusion on the trustworthiness of vine reviews with data to back up their claim</p>	<p>not limited to:</p> <ul style="list-style-type: none">Number of reviewsNumber of 5-star reviewsAverage RatingNumber of helpful votes <p>✓ Comes up with an invalid conclusion on the trustworthiness of vine reviews or does not provide data to back up their claim</p>	<p>not limited to:</p> <ul style="list-style-type: none">Number of reviewsNumber of 5-star reviewsAverage RatingNumber of helpful votes <p>✓ Comes up with a conclusion on the trustworthiness of vine reviews with data to back up their claim</p>	<p>not limited to:</p> <ul style="list-style-type: none">Number of reviewsNumber of 5-star reviewsAverage RatingNumber of helpful votes <p>✓ Comes up with a conclusion on the trustworthiness of vine reviews with data to back up their claim</p>	
--	--	--	--	--	--