

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 (+/-)	70+	C (+/-)	30-49	F (+/-)	<10
B (+/-)	50-69	D (+/-)	10-29		

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**.

Rubric for Big Data - Level 1:

	Mastery 20 points	Approaching Mastery 15 points	Progressing 10 points	Emerging 5-0 points	Incomplete
Extract	<p>In both notebooks the student did all of the following:</p> <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 	<p>In both notebooks the student did 3 of the following:</p> <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 	<p>In both notebooks the student did 2 of the following:</p> <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 	<p>In both notebooks the student did 0-1 of the following:</p> <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 	<p>No submission was received</p> <p>-OR-</p> <p>Submission was empty or blank</p> <p>-OR-</p> <p>Submission contains evidence of academic dishonesty</p>
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 	<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 	<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 	<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 Load ✓ Successfully pushed dataframes to AWS	necessary columns to match the current database table schema ✓ Matched dataframe column types with the database column types Load ✓ Successfully pushed dataframes to AWS	necessary columns to match the current database table schema ✓ Matched dataframe column types with the database column types Load ✓ Successfully pushed dataframes to AWS	necessary columns to match the current database table schema ✓ Matched dataframe column types with the database column types Load ✓ Successfully pushed dataframes to AWS	
--	---	---	---	---	--

Rubric for Big Data - Level 2:

	Mastery 20 points	Approaching Mastery 15 points	Progressing 10 points	Emerging 5-0 points	Incomplete
Extract & Transform	The student did all of the following to extract and clean the data: ✓ Connected to and loaded in datasets from AWS to dataframes using Spark ✓ Removed any unnecessary columns ✓ Dropped rows with null values ✓ Dropped duplicated rows	The student did 3 of the following to extract and clean the data: ✓ Connected to and loaded in datasets from AWS to dataframes using Spark ✓ Removed any unnecessary columns ✓ Dropped rows with null values ✓ Dropped duplicated rows	The student did 2 of the following to extract and clean the data: ✓ Connected to and loaded in datasets from AWS to dataframes using Spark ✓ Removed any unnecessary columns ✓ Dropped rows with null values ✓ Dropped duplicated rows	The student did 0-1 of the following to extract and clean the data: ✓ Connected to and loaded in datasets from AWS to dataframes using Spark ✓ Removed any unnecessary columns ✓ Dropped rows with null values ✓ Dropped duplicated rows	No submission was received -OR- Submission was empty or blank -OR- Submission contains evidence of academic dishonesty
Analysis	Student does all of the following to analyze if "vine" reviews are trustworthy: ✓ Splits the reviews between vine (paid) and non-vine (unpaid) ✓ Compares metrics between vine and non-vine reviews such as, but not limited to: Number of reviews Number of 5-star reviews Average Rating Number of helpful votes ✓ Comes up with a conclusion on the trustworthiness of vine reviews with data to back up their claim	Student does all of the following to analyze if "vine" reviews are trustworthy: ✓ Splits the reviews between vine (paid) and non-vine (unpaid) ✓ Compares metrics between vine and non-vine reviews such as, but not limited to: Number of reviews Number of 5-star reviews Average Rating Number of helpful votes ✓ Comes up with an invalid conclusion on the trustworthiness of vine reviews or does not provide data to back up their claim	Student does 2 of the following to analyze if "vine" reviews are trustworthy: ✓ Splits the reviews between vine (paid) and non-vine (unpaid) ✓ Compares metrics between vine and non-vine reviews such as, but not limited to: Number of reviews Number of 5-star reviews Average Rating Number of helpful votes ✓ Comes up with a conclusion on the trustworthiness of vine reviews with data to back up their claim	Student does 0-1 of the following to analyze if "vine" reviews are trustworthy: ✓ Splits the reviews between vine (paid) and non-vine (unpaid) ✓ Compares metrics between vine and non-vine reviews such as, but not limited to: Number of reviews Number of 5-star reviews Average Rating Number of helpful votes ✓ Comes up with a conclusion on the trustworthiness of vine reviews with data to back up their claim	