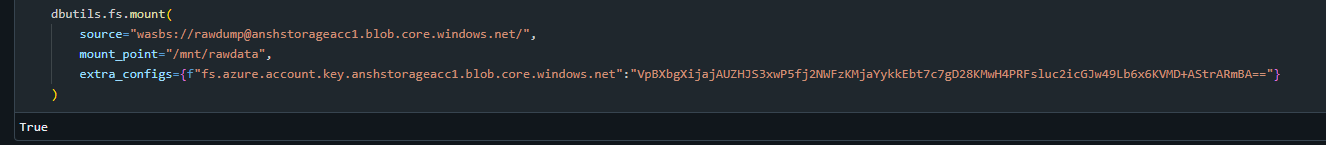
**Ansh Ranjan  
Databricks Case Study  
Exercise 2 – Data Ingestion**

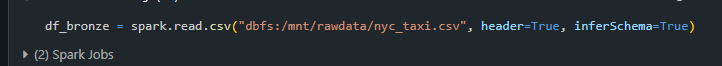
**TASK 1: Load a dataset into your Databricks using Spark**

1. Our data resides in a ‘rawdump’ named container in our storage account.   
   A screenshot of a computer

   AI-generated content may be incorrect.
2. Creating a spark session  
   A screen shot of a computer

   AI-generated content may be incorrect.
3. Now we have to mount our data point to our workspace.  
     
   You can get your access key by going to storage account > Security + networking > Access Keys  
   A screenshot of a computer

   AI-generated content may be incorrect.
4. Reading data into a dataframe  
   A screenshot of a computer

   AI-generated content may be incorrect.  
   

**TASK 2: Inspect the data**

1. Displaying the data  
   A screenshot of a computer

   AI-generated content may be incorrect.
2. Dataframe schema  
   A screen shot of a computer program

   AI-generated content may be incorrect.
3. Number of rows  
   A screenshot of a computer

   AI-generated content may be incorrect.

**TASK 3: Perform data cleaning**

1. Dropping null values from the df  
   A screen shot of a computer screen

   AI-generated content may be incorrect.  
   Now there should not be any null values  
   A computer screen shot of a computer code

   AI-generated content may be incorrect.
2. As we saw earlier, our pickup and dropoff columns are of double data type. We need to convert them into standard datetime format  
   A screen shot of a computer

   AI-generated content may be incorrect.
3. Checking for any rows where dropoff time might be before pickup time for data quality  
   A screen shot of a computer code

   AI-generated content may be incorrect.