Task 2 Response:

To use this software an account needed to be made on the databricks website. A community edition account was made since it did not require any subscription to be made to use. This allowed for the use of one database along with 15GB of memory available. After completing the sign up process of linking an email account, specifying a name along with a company name, the tutorial "Get started as a Databricks Workspace user" which was previously named "Explore the Quickstart Tuturial" was followed to set up a big data environment. This tutorial can be found under the documentation section on the databricks website.

The first task on the tutorial was to orient oneself with the layout of the platform which was pretty simple due to the layout of the platform. Through this step I was able to learn the layout of the platform which includes a navigation bar on the left hand side of the screen with accessible buttons to all the functions of the system including the data, cluster, and workspace sections. The second thing that was learned from this section was the ability to create a cluster which is the first step needed to run a big data program on the platform. To do this you go to the cluster section of the platform which can be found on the initial page under the common tasks heading called "create cluster" or through the respective button on the navigation bar. Once on this page the button "create cluster" was clicked and the option to create a new cluster opened. On here the option was given to chose the version of runtime. My instructor suggested to use the version 7.2 ML (Scala 2.12, Spark 3.0.0) however this option was no longer available so the version 7.4 LTS (Scala 2.12, Spark 3.0.1) was chosen as suggested by the tutorial. After naming the cluster "csc410" and waiting for the cluster to activate, the feature spaces were then added to the system. The tutorial taught us to do this by going to the data page which can be found the same was as the cluster page. on this page the "create new table" button was clicked to add the respective spreadsheets. The appropriate feature space was uploaded then the table was created using the create table with UI option which allowed the selection of a cluster which in this case was csc410 the active cluster. For each table the options of using the first row as headers was selected along with the option to infer the datatype. Then one by one each feature space was added to the system each receiving the name of table1, table2, table3... respectively. Lastly the tutorial taught how to create a python notebook which to access this option was similar to finding the create cluster option. Once on the respective site one goes to users then selects the current user and by clicking on the dropdown arrow next to the username an option to create a notebook is given. After creating a python notebook and giving it a name the process of creating a machine learning program was then able to begin.