**Kedir Nasir Omer Date: Nov. 28/2022**

**UC SanDiego- Big Data Specialization**

**2.2. Data Partitioning and Modeling**

* **Classifying in KNIME to identify big spenders in Catch the Pink Flamingo**
* The training data set was used to build the decision tree model, and the data was divided into test and training datasets.
* The test dataset was then fed into the trained model.
* This is significant because it enables us to check the precision of the trained model by dividing the data set into training and test data.
* Setting the random seed is crucial when using sampling to partition the data because it enables you to get consistent results every time you run the partition.

Below is a screenshot of the resulting decision tree:

