First and foremost, I would make sure that they get a solid understanding of the fundamentals of machine learning. It is common for people to use the terms artificial intelligence and machine learning interchangeably; this isn't technically correct and may lead to some misconceptions about what machine learning actually is and its practical capabilities. Artificial intelligence is a border term that refers to the autonomous solving of problems by mimicking human intelligence. AI algorithms will take in input and react to it to produce a certain output. Within the umbrella of artificial intelligence lies machine learning. Machine learning revolves around the idea of pattern recognition in data and is more specific technically speaking as it refers to certain kinds of algorithms and methods, whereas artificial intelligence is far broader. At this point, my friends will likely understand that running a machine learning algorithm will not result in a robot coming to life with a mind of its own. It may still be a little difficult for them to fully understand what these ideas mean in a more concrete sense, so some examples would help solidify their understanding.

There are two main forms of machine learning: supervised and unsupervised machine learning. Simply put, supervised learning deals with a question and comes up with an answer, which can often be tested using existing data. Unsupervised learning, on the other hand, doesn’t have a direct answer.

Two excellent examples of practical applications for supervised learning are knn regression and classification. Every student in Vancouver strives to find affordable housing, but it’s often quite difficult to come by. However, it is quite beneficial to compare rent at prospective apartments and decide whether or not they are actually good deals. This would be a great time to use knn regression! By taking in a large amount of housing data from around the Vancouver area, one could train a model to predict how much a certain apartment would charge for rent. The student can then see if the place they are considering is actually worth it. When training the model, the student could research several different factors that may influence how expensive the apartment will be. Examples of different predictors they could use would be the number of bedrooms, the number of bathrooms, the total area of the apartment, how close it is to downtown or campus, the year it was built, and other numerical factors. By separating the data into training and testing data, the student could then create a model using the training data and test its accuracy using the testing data. If the accuracy proves to be high, then the student can trust that it may potentially be a valuable resource. A challenge that the student may face would be attaining a large amount of relevant data to use. Additionally, the algorithm can’t take into account factors that aren’t numerical such as how nice the apartment looks, or the color of the exterior. Another example of supervised learning is classification. Most people don’t realize it, but they are benefitting from machine learning algorithms every day by opening up their emails! Companies like Google take in great amounts of data from different unwanted emails and create an algorithm that will automatically pick out different emails that it classifies as spam and put them in a spam folder. A difficulty that this example poses is that there is a possibility that important emails may get tagged as spam based on the contents (such as the initial invitation for an interview with UBC Launch Pad!).

On the other hand, unsupervised learning isn’t based on answering a question with a definitive answer, but rather the clustering of data to discover patterns. Once again, this is heavily used all around us! When someone goes on a website like Amazon, there are often items advertised that seem to fit the users' interests. This is because Amazon takes data from

different users based on what they search and purchase, and group people together. So if someone spends a lot of time looking at hiking boots and ends up purchasing them, the next day they may find that Amazon has suggested they purchase a camping tent. The computer doesn't intuitively know that a tent and hiking boots would likely be purchased by the same person. Instead, it takes all of this data and goes through a process known as clustering to group different users and their preferences together. In a simple clustering problem with two variables and a set number of groups, one can go through and use a clustering algorithm to assign the different data points to different groups. This works by randomly assigning points to different groups, and then reassigning the points based on the location of the centroid of each group after each iteration. With each reassignment, the centroid may alter slightly as different points change to different groups. When no cluster assignments change after an iteration, the assignment of points to groups has concluded. This same method can be used on a much greater scale to accomplish a task like clustering people's online shopping habits. A challenge of this method is that depending on how accurate the first assignment is, it can potentially affect the rest of the iterations and result in a seemingly inaccurate clustering of the data.

After people understand these everyday examples of how machine learning affects us, they will likely understand that it is a useful tool that is growing with potential and possibilities. Although it likely won't single-handedly solve world hunger, it is by no means a fad; instead, it's a useful tool with growing applications.