So guys, till now whatever videos we have seen in the previous session was something related to supervised machine learning problems. All the algorithms like linear regression, logistic regression, Decision Tree, Random Forest, XGBoost, Gradient Boost, AdaBoost. All these algorithms were examples of supervised machine learning algorithm. Now, whenever I say supervised machine learning algorithm, that basically means let's say you have a data set in this specific data set in supervised machine learning algorithm, the data set that is basically given you used to have some features, let's say f1, f2, F3, f4, and you would always have a dependent feature, one dependent feature. And this specific is nothing. But this is my dependent or dependent or output feature right. And all the other features used to be my independent features or input features. Right. Now this was the kind of problem statement we used to be given, uh, in supervised machine learning and two types of problem statement we were solving either regress or classification. Regression or classification. Now, if I consider unsupervised machine learning. In unsupervised machine learning and will try to also understand the difference. So in unsupervised machine learning you will be seeing that you will be solving a problem statement which is called as clustering. Clustering. That basically means you will be grouping your data. Grouping your data into similar clusters. Into similar clusters. Now, let's say that I have been given a data set. In this data set. Uh, you know, I may have features. Let's say the feature is age, years of experience. And your salary. Here. Our aim is not to find out anything as such. That basically means I don't have a specific output feature in this particular data set. Okay, in this entire data set, I don't have a specific output, no output okay. That basically means this is not a supervised machine learning problem statement. I'm not saying that, okay, your model should be able to predict salary based on your age and experience. No, I'm not saying that. Let's say if you have some data which has all this particular data points with respect to age, experience and salary. Now, if I try to apply any unsupervised machine learning algorithm, and I'll also talk about which all unsupervised machine learning algorithm are there. If I try to apply an unsupervised machine learning algorithm on to this specific data set. Okay then here we will be able to group this data together. Group or cluster this data. Or cluster this data. Okay. Now what does this basically mean? So suppose I have this entire data set. So when I'm grouping the information that I may be able to get right from this particular data set is that people who are having similar age, similar years of experience and similar salary. So when I apply unsupervised machine learning algorithm, you will be able to create clusters that may have the similar types of values within those clusters. Okay. Let's say this particular, uh, group or cluster, I will mention it as cluster one. This may have similar age, experience and salary. So all the people who are inside this particular cluster with respect to this particular data points may have similar age, experience and salary. So I may be having some data points which will be falling in this particular data in this cluster. That basically means let's say some of the records like this may be falling in this cluster. Some of the record may be falling in this cluster. Similarly, I may be having more data points that may be falling in other clusters. Right. So that is what an unsupervised machine learning algorithm will do. At the end of the day, we are clustering this data points into a specific group okay. But why do we do this okay. One basic example. Let me talk about a real world scenario. There is a use case which is called as customer segmentation. Then customer segmentation. What do we do? Let's say I own a specific, uh you know, product okay. And I have a list of data. Of which all people have bought this specific product. Okay. And I may also be having some information like, uh, the salary they're spending score. Let's say I have all this information from those people who have bought this product. Then what may happen is that when I bring a new product into the market. You know, by just seeing this data points, I will be able to understand that who are the people will definitely buy this particular product for the first time in the first day itself. And if they are probably buying this specific product on the first

day, I may give them some kind of discount, right? So obviously, I can definitely apply an unsupervised machine learning algorithm. And I can say that, okay, there is a group of customers considering this salary spending score. There is another group of customers. I know that in this group of customers, people like to buy the product. As soon as a new feature comes, let's say Apple. Okay. So I may be saying that okay. Let's go ahead and probably, you know, give this people who who buys my product regularly around 15% of discount. Let's say these are the other cluster of people who buys this product. Not on a regular basis, but yes, sometimes. So as soon as this product launch get launches, you know, I may probably increase the discount rate to 20% so that they will be interested to buy the product as soon as a new feature will come up. So this example is called as customer segmentation. And based on segmenting this people based on some features that basically means clustering, right? I am providing some features over here. Right. So this can be a perfect example with respect to customer segmentation. Now in unsupervised machine learning what all algorithms will be learning. So the first algorithm uh that we are going to learn in unsupervised machine learning. Okay. The first algorithm is nothing but. K-means clustering. K-means algorithm. The second algorithm that we will be learning is something called as hierarchical. Clustering all algorithm. Okay. The third algorithm that we'll be learning is something called as DBscan. Pull a string. Okay. And finally, the fourth thing that we are going to see is something called as sill height. Scoring. And this mechanism will be basically used to see the model that I have probably got. In short, I will be using this particular mechanism to validate the specific models. Okay super super important slide scoring. So we'll be covering this algorithm one by one will be understanding the maths in depth intuition behind it the geometric intuition behind it. Then we'll go to hierarchical clustering DBscan clustering. And finally we'll be seeing silhouette scoring okay. So yes let's go to the next video. And let's start our first unsupervised machine learning algorithm that is K-means clustering.