**(Video Transcript)**

**Hello**

**Today, I'll walk you through the initial steps of the project, which uses data science to study home loans.**

**This presentation will go over the "Data Science Lifecycle," which is the process that any data science solution follows, as well as the project overview, the process overview, a description of the data we used, the analysis we performed, the modelling and model evaluation process, and finally our recommendation(s) based on our findings.**

**We used a methodology known as CRISP-DM, or the Cross Industry Standard Procedure for Data Mining, to try to solve this challenge. This process requires that we comprehend the company and the data, prepare the data for modelling, model in accordance with the business's objectives, and then assess if the model's performance is in line with those objectives before moving the model's production toward listing. If not, we go back to work to improve our comprehension, which will help us better grasp and, as a consequence, better prepare the data to be modelled for good outcomes.**

**As we have learned more about the industry, we believe that the issue the firm is trying to solve is the lengthy and manual home loan application process, which might take applicants up to three days to learn whether or not their application was successful. The company thinks that by automating this process and utilising tools like artificial intelligence and machine learning, it can be done in a matter of seconds.**

**We suggest a system where a potential borrower may submit an application for a home loan online using either a phone or a computer. The applicant will fill out the application by giving us their information like they would on any other application, and after doing so, our machine learning model will predict the applicant's application status in a matter of seconds.**

**The historical data that will be used to train this model consists of 614 records, or historical home applications, of which 422 were accepted and 192 were denied. There are 13 characteristics in this data: 8 categorises and 5 are numerical.**

**Some conclusions drawn from our examination of the data indicate that males submit more applications and are accepted proportionally more often than their female counterparts. The relationship between the applicants' loan amounts and individual earnings is another intriguing conclusion.**

**Understanding the business goal and doing an analysis have given us the confidence to go on to the modelling step. As we are expecting two classes in this problem—loan acceptance or declination—it is a binary classification issue. We must prepare the data appropriately in order to get the most performance out of the model we ultimately choose. In order to prevent the model from favouring attributes with higher values without taking units into account, this preparation includes how we handle missing data, scaling the features to a fixed range that makes the learning process "easier" for some models, and converting categorical attributes to numbers, such as "Female" = 0 and "Male" = 1, since most models only understand numbers. It is crucial to remember that Automated Machine Learning, or AutoML, seldom ever needs any data preparation. We combined AutoML with bespoke/traditional ML in our solution.**

**In order to guarantee that we can assess our models fairly, we also divide our data into two independent sets: the training set, which will serve as the section on which the models are fitted and trained, and the test set, which will serve as the portion on which the models are tested. In comparison to 20% for testing, 80% was used for training.**

**We acquired the following accuracy levels after training the chosen model and AutoML: 78% for AutoML and 77% for the chosen model. It's vital to remember that accuracy is calculated as the total number of correctly predicted events divided by the total number of forecasts.**

**Based on the work completed, we feel that bespoke is superior to AutoML, despite the fact that the latter fared 2% worse in our tests. This is because we are more familiar with the model's construction and believe that with more iterations, we can improve the results. It also requires fewer resources and trains more quickly.**

**We are confident that with a few more CRISP-DM iterations, we will be able to achieve results that are satisfactory and meet the business's objectives. We also anticipate that our solution will eventually be integrated with mobile phone and web applications and placed in the hands of our prospective clients, making the application process quick and simple for them.**

**Thank You**