[Video Transcript]

Good morning,

Today, I will provide an overview of the initial steps of our project, which utilizes data science to study home loans. This presentation will cover the "Data Science Lifecycle," outlining the process that any data science solution follows. We will discuss the project overview, process overview, data description, analysis, modeling and model evaluation process, and conclude with our recommendations based on our findings.

We employed the CRISP-DM methodology, which stands for Cross Industry Standard Procedure for Data Mining, to tackle this challenge. This process entails understanding the company and the data, preparing the data for modeling, modeling according to the business objectives, and evaluating the model's performance to ensure it aligns with those objectives. If necessary, we revisit our understanding of the problem and data to improve data preparation for better modeling outcomes.

After gaining a deeper understanding of the industry, we have identified the main issue the company seeks to address: the lengthy and manual home loan application process, which can take applicants up to three days to receive a response on their application. The company believes that by automating this process and leveraging tools like artificial intelligence and machine learning, it can be completed within seconds.

Our recommendation is to develop a system where potential borrowers can submit their home loan applications online using either a phone or a computer. Applicants will complete the application by providing their information as they would on any other application. Subsequently, our machine learning model will predict the application status within seconds.

To train this model, we utilized historical data comprising 614 records of past home loan applications. Among these, 422 were accepted, while 192 were denied. The dataset consists of 13 features: 8 categorical and 5 numerical.

Some key insights derived from our data analysis include the observation that males tend to submit more applications and have a higher acceptance rate compared to females. Additionally, we discovered an interesting relationship between loan amounts and individual earnings.

Having understood the business goal and performed thorough analysis, we proceeded to the modeling phase. Given that this is a binary classification problem with two classes (loan acceptance or declination), we needed to appropriately prepare the data to maximize the performance of our chosen model. Data preparation involves handling missing data, scaling features to a fixed range to facilitate learning for certain models, and converting categorical attributes into numerical values (e.g., "Female" = 0 and "Male" = 1), as most models only work with numerical inputs. It is worth noting that Automated Machine Learning (AutoML) typically requires minimal data preparation. In our solution, we combined AutoML with traditional machine learning to achieve the best results.

To ensure fair assessment, we divided our data into two independent sets: the training set, used to fit and train the models, and the test set, used for model evaluation. We allocated 80% of the data for training and 20% for testing.

After training both the chosen model and AutoML, we achieved accuracy levels of 77% for the chosen model and 78% for AutoML. Accuracy is calculated as the total number of correctly predicted events divided by the total number of forecasts.

Based on our work, we believe that the bespoke model outperforms AutoML, despite AutoML scoring 2% lower in our tests. This is because we have a better understanding of the bespoke model's construction and believe that with further iterations, we can improve its performance. Additionally, the bespoke model requires fewer resources and trains more quickly.

We are confident that with a few more iterations following the CRISP-DM framework, we can achieve satisfactory results that meet the business objectives. We also anticipate integrating our solution with mobile phone and web applications, making the application process quick and user-friendly for prospective clients.

Thank you.