1. **Problem Introduction**

   This data contains key insights in maintaining user happy service. Our task in this project is to develop a model that predict which users have potential to cancel premium services, etc...
   My work will give business more chances to provide suitable service to their customers.

2. **Strategy to solve the problem**

   - I have prepared several insighs research to clarify throughout the project. You can find out at step `Feature Engineering`
   1. What is the total number of thumbs down by users for each churn?
   2. And the opposite, How many thumbs up of the users?
   3. Find out the number of errors that users encountered

3. **Metrics**

   - The label which needs to be predicted is churn which is defined by a user who has commited a "Submit Downgrade" event or "Cancellation Confirmation" event.

4. **EDA**

   - Raw dataset may have many problems about data quality. I have set up entire process to clean and reformat to a clean dataset that can be used for ML model
   - I have prepared several insigh problems that need to find out. In the future, we can give more insigh to be better understand the situation
   1. What is the total number of thumbs down by users for each churn?
   2. And the opposite, How many thumbs up of the users?
   3. Find out the number of errors that users encountered

5. **Modelling**

   - I choose to use Linear Regression with several features to build up model. The data is compressed in a DataFrame for easy to querying, In the last part, I also calculate F1-score and accuracy of the model

6. **Hyperparameter tuning**

   - At this time, we do not apply hyperparmeter tuning, this will have in an improved version of this work

7. **Results**

   - F1-score is 0.63 and accuracy is 0.6 of the model

8. **Conclusion/Reflection**

   - We can apply more algorithms in the next version for better improvements. With this work, business can decide which services should be given to potential customers

9. **Improvements**

   - We can research more features and more algorithms to this work, makes it a better version of model