I**ntroduce the presentation topic**

I'm going to talk about my home assignment and interesting facts that I was able to discover. I’ve described whole data exploration process and my assumption in the report that I’ve send yesterday. So here I would like to focus on things that are the most interesting and crucial in my opinion. If you have any questions, feel free to interrupt.

So let's go ahead and begin.

**Outline**

My talk will be in five parts.

In the first part I will describe task goals.

Then in the second part I will try to explain what data are available.

Next I’ll focus on the facts that could be used to identify risky transactions and some interesting facts about customers.

Then I’ll present predictive models that could be build based on available data.

Finally, I’ll go on to talk about tools and framework I was using during the process.

So let's go ahead and begin.

**Goals**

So, first let's talk about what I was asked to do.

*<points from presentation>*

**Data**

There are two tables available: customer and transaction. One customer has one or more transaction. The last transaction of the client I treat as my credit application, and the previous as a credit history.

**Findings x 3**

Now I would like to tell about the facts that I was able to discover from the data.

*<points from presentation>*

**Modeling x 4**

I’d now like to change direction and talk about models that I’ve build based on the data.

To reduce the complexity I’ve focused only on customers that has at least one previous transaction, and only od predicting the default event.

*<points from presentation>*

Next slide (features)

I spent some time to do feature engineering and then I use Information Value to measure their predictive power. This measure helped me to classify features into one of the category: strong, medium, weak.

*<points from presentation>*

Next slide (LR)

*“The sensitivity, or true positive rate of the model, is shown on the y-axis. And the false positive rate, or 1 minus the specificity, is given on the x-axis.*

*The ROC curve always starts at (0, 0) corresponding to a threshold of 1. This means we have 0 sensitivity and we won’t catch a good care cases. But since our false positive rate is 0 as well, that means that we correctly label all the poor care cases.*

*The ROC curve always ends at (1, 1) which corresponds to a threshold of 0. So, the threshold decreases as we move from (0, 0) to (1, 1).*

*Let’s take an approximate point (0.4, 0.65) on the curve. This point signifies that we correctly label 65% of the cases with a false positive rate of 40%” –* [source](https://medium.com/@shivangisareen/roc-curves-80ad099915b9)

Next slide (XGBoost)

**Modeling x 4**

At the end I would like to describe the tool I’ve used.

**End**

That concludes my talk for today.

Thank you for your attention.