Hello everyone, I'm Jiapeng Wang, and today I'll introduce my work on feature engineering.

Here are some examples about my data procession. For binary variables like Attrition flag and Gender, a simple 0-1 encoding was applied. Meanwhile, for categorical features like Card Category, I employed one-hot encoding. Besides, as we saw before, we have some numerical features, like Customer\_Age. Thus here I use Standardization to rescale. And for the Income\_Category feature, I used different numbers to illustrate different income levels .

After that, here is a heatmap for all the features after my procession. Now we could find that some features are highly correlated, like these two. So I drop some of them for better accuracy.

I also tried some less successful approaches in my work. Here’s one example about imbalanced data. As we saw in data analysis, there are only around 15% of attrited customers in our data. So I used resampling methods. The results were unexpected—resampling negatively influenced accuracy. However, a significant increase can be seen in recall score, which means the resampled model is better at finding customers who’s gonna leave. This may be helpful in our real life.

Thank you very much.Next please introduce GELI about the models we use.