End-to-End Machine Learning Project:
Customer Churn Prediction Machine Learning Model

Qiuhong Wei

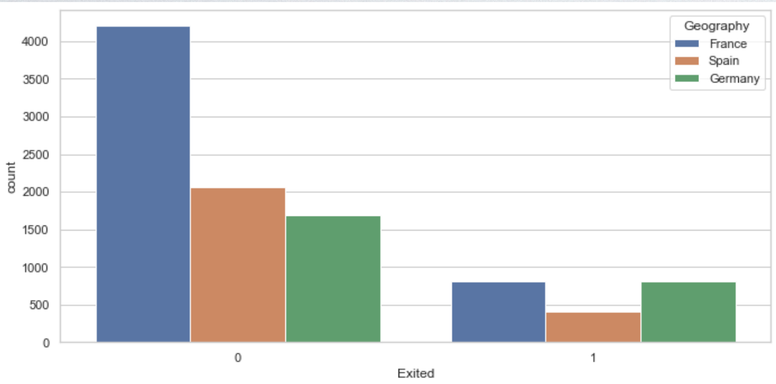
BACKGROUND INFORMATION

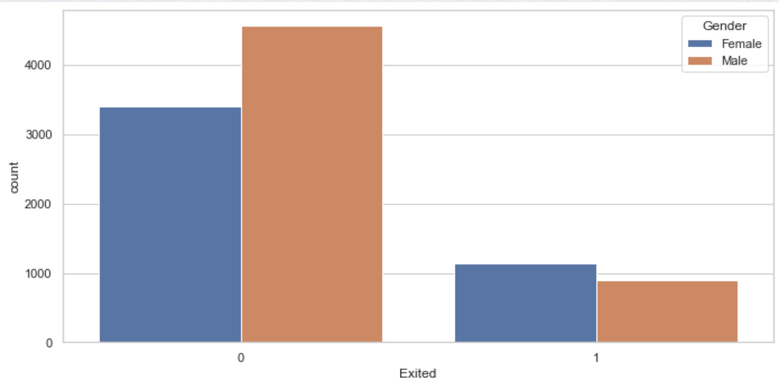
- Customer churn prediction means knowing which customers are likely to leave from bank.
- **Methods:** The churn rate is predicted by using the LightGBM, RandomForest and SVM.
- **Significance:** Acquiring new customers often costs more than retaining existing ones.
  - If a customers may churn, firm should make with each customer to maximize their likelihood of staying.
- **Impact of customer churn on businesses:** A company with a high churn rate loses many customers, resulting in lower growth rates and a greater impact on sales and profits. Companies with low churn rates can retain customers.

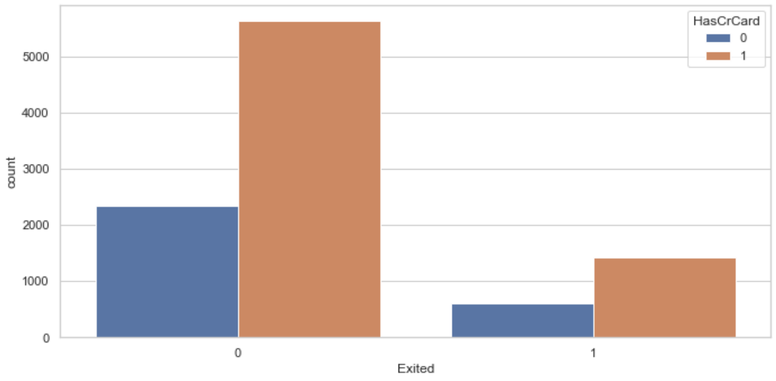
BUSINESS QUESTION

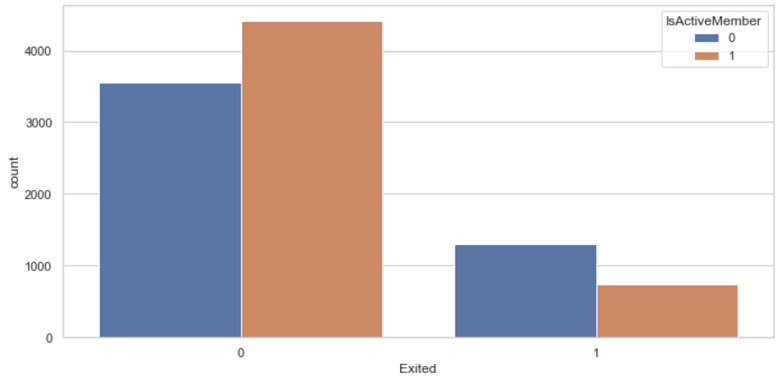
- **Challenge**: Everyone has different behaviours, preferences, and reasons for churning.
  - Firms find marketing activities that are most effective for each individual customer.
- **Business Questions:**
  - What is the significant feature that relates to the churning rate?
  - What is the potential strategy to keep customer staying?

EXPLORATORY DATA ANALYSIS









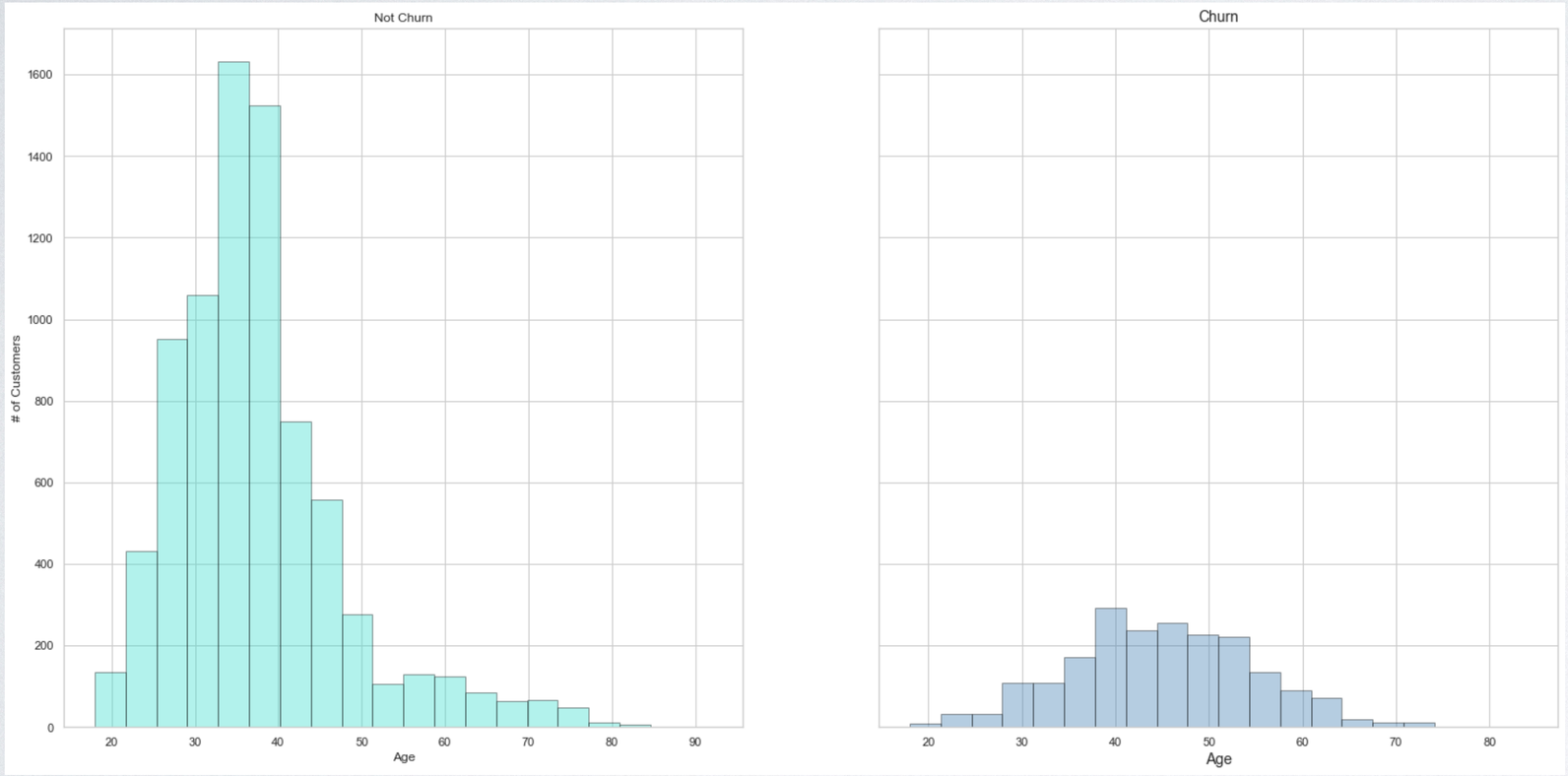
From **Geography**, Spanish are more likely to stay than German and French.

From **Gender**, Females have higher churn rate than Male.

With a credit card, customers are more likely to churn.

From **IsActiveMember**, non-active customers are more likely to churn.

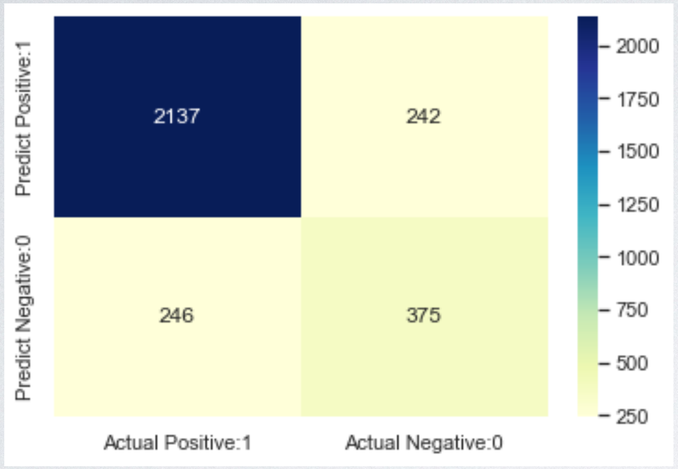
EXPLORATORY DATA ANALYSIS



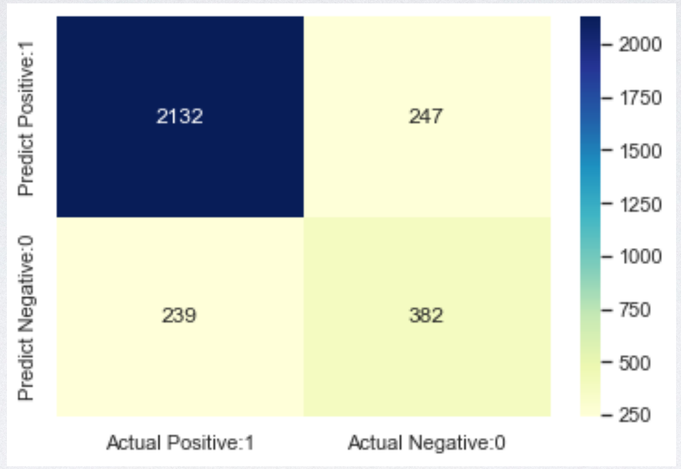
For **churned customers**, the age of majority of customers are from 30 to 40 years old.

For **not churned customers**, the age of majority of customers are from 40 to 50 years old.

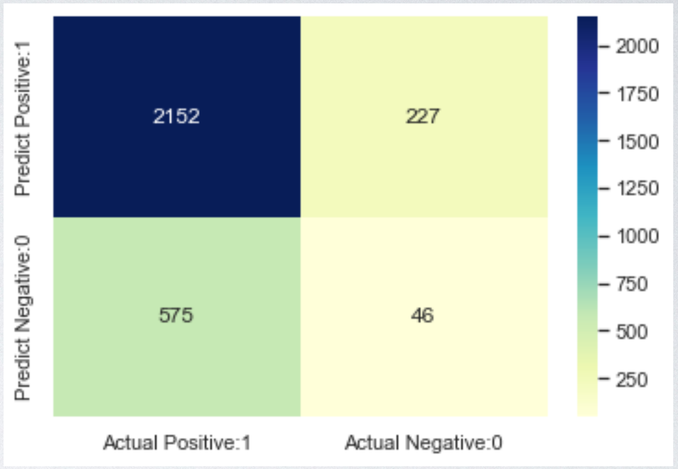
INSIGHTS



LightGBM
Accuracy: 0.84
False Positive: 242
False Negative: 246



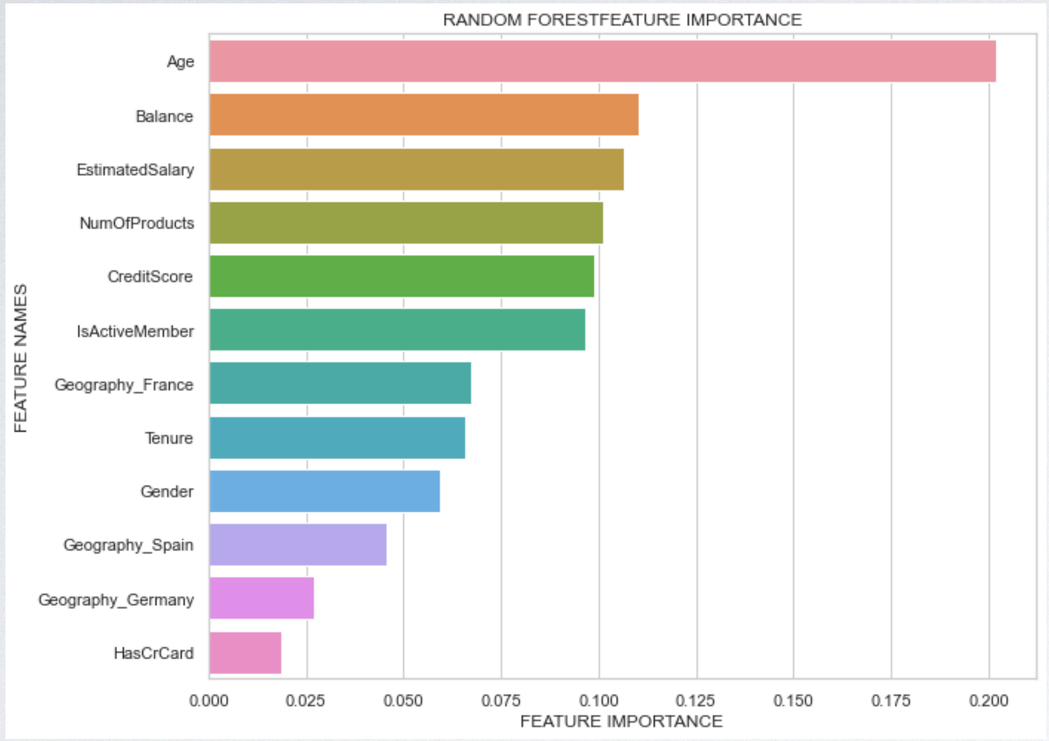
Random Forest
Accuracy: 0.84
False Positive: 247
False Negative: 239

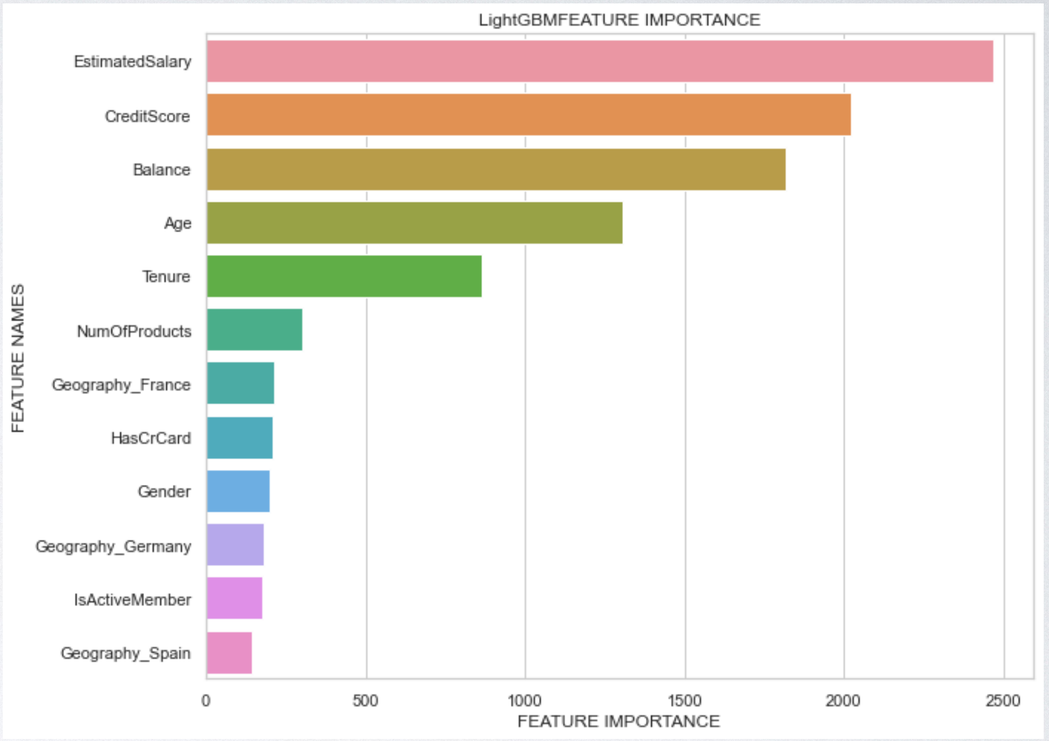


SVM
Accuracy: 0.78
False Positive: 227
False Negative: 575

- Based on the 3 models, the **random forest** has the best performance in accuracy.
- Random forest has the lowest false negative, and it results the small mistake in the churn prediction among 3 models. The bank can find the majority customers who may churn.

BUSINESS ACTION





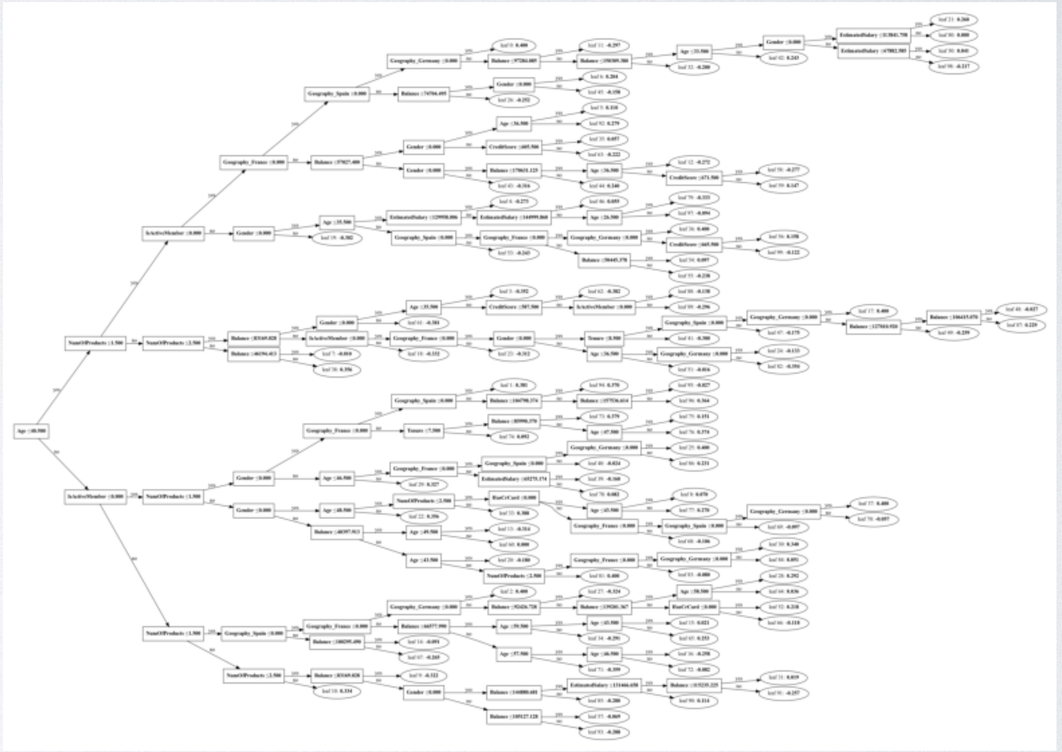
- For the **estimated salary, balance, age** and **credit score**, they have significant roles in the prediction of the churn rate in the random forest and LightGBM models, which suggests bank to focus in these customers' conditions.

BUSINESS OUTCOME

- Due to the high importance scores, bank should focus on below features.
- For **estimated salary**, bank should improve its model to predict accurate salary. It will benefit the customer churn prediction model.
- For **balance, age, and credit score**, bank can collect the data from the internal database, and the database should be update frequently.

TREE PLOT (OUTRO SLIDE)

- This is the tree plot of the LightGBM, and the **advantages is the interpretation of decision tree**.
- Random forest has many decision tree, and it is difficult to plot the model.
- SVM is a hyperplane due to the number of feature is over



THANK YOU.