

novobanco

DATA SCIENTIST

Technical assessment



RECRUITMENT CHALLENGE FOR DATA SCIENTIST

Applicants will be scheduled for an 1h individual interview that includes the presentation of the following challenge

Context: A bank is facing the problem of customer attrition in their credit card services and they want to reduce customer churn.

Objective: Identify the customers that will churn in the near future.

- Perform a complete analysis of the dataset.
- Apply advanced analytics and machine learning methods to train a model that identifies the customers who are most likely to churn.

Presentation and Evaluation:

- You'll deliver a 15 to 20 minutes (max) oral presentation.
- The presentation should clearly convey the project business problem, motivation of work, methods used and results obtained, alongside with relevant business and technical analysis, contributions and/or discussion questions.
- You are expected to complete the assessment within approximately 1 week.
- The presentation document (e.g. PDF, PPT) with the main outputs of your work must be delivered until 2 days before the scheduled presentation slot.
- It is also mandatory to deliver your code (only Python will be considered) - feel free to deliver it as a Jupyter Notebook or as .py file(s).

Resources & Tips:

- Consider CRISP-DM or other project methodology to organize your work.
- The dataset is available on [Kaggle](#) (along with an overview of this challenge) - yes, we added some minor twists for your assessment!
- All the details of the dataset are available on the next slide.

ABOUT THE DATASET

The dataset consists of approximately 10,000 customers mentioning their age, salary, credit card limit and others - there are 22 columns in total

Variables description:

- **CUSTOMER_ID**: Customer ID number. Unique identifier for the customer holding the account.
- **YEAR_MONTH**: Year and month of reference for the data.
- **CHURN_FLAG**: Binary variable that indicates if the customer has churned or is an existing customer.
- **CUSTOMER_AGE**: Customer's age in years.
- **GENDER**: M=Male, F=Female.
- **DEPENDENT_COUNT**: Number of dependents.
- **EDUCATION_LEVEL**: Educational qualification of the account holder (e.g. High School, College).
- **MARITAL_STATUS**: Civil status (e.g. Married, Single).
- **INCOME_CATEGORY**: Annual income category of the account holder (< 40K, 40K - 60K, 60K - 80K, 80K-120K, > 120K).
- **CARD_CATEGORY**: Type of Credit Card (Blue, Silver, Gold, Platinum).
- **MONTHS_ON_BOOK**: Period of relationship with the bank.
- **TOTAL_PRODUCT_COUNT**: Total number of products held by the customer.
- **MONTHS_INACTIVE_12_MON**: Number of months inactive in the last 12 months.
- **CONTACTS_COUNT_12_MON**: Number of contacts in the last 12 months.
- **CREDIT_LIMIT**: Credit Limit assigned to the Credit Card.
- **TOTAL_REVOLVING_BAL**: Total revolving balance on the Credit Card.
- **AVG_OPEN_TO_BUY**: Open-to-Buy Credit line – amount of remaining credit available at a given time on the Card holder's account (Average of last 12 months).
- **TOTAL_TRANS_AMT**: Total transaction amount (Last 12 months).
- **TOTAL_TRANS_CT**: Total transaction count (Last 12 months).
- **TOTAL_AMT_CHNG_LAST_3M_6M**: Change in transaction amount between the last 3 months and the 3 months before (4th to 6th months)*.
- **TOTAL_CT_CHNG_LAST_3M_6M**: Change in transaction count between the last 3 months and the 3 months before (4th to 6th months)**.
- **AVG_UTILIZATION_RATIO**: Average card utilization ratio.

$$* \left(\frac{\text{transaction amount in the last 3 months}}{\text{transaction amount in the 3 months before}} \right)$$

$$** \left(\frac{\text{transaction count in the last 3 months}}{\text{transaction count in the 3 months before}} \right)$$

novobanco

MANY THANKS!

