



AI&D Technical Assessment

Deloitte Spain

Artificial Intelligence & Data

Assessment overview

The time available for this assessment will be **24 hours**, although the **estimated time** required for the **technical resolution** is approximately **3 hours**.

In this activity you will be asked to **build a model to predict the churn rate** of a Telecom Company.

The churn rate is a measure of the number of contractual customers who leave the Telecom Company during a given timeframe.

To solve this task, you are expected to perform **data preparation, modeling, evaluation** and **summarize a set of conclusions** using **Python or R (Jupyter Notebook or RMarkdown)**. When you finish the assessment, submit the code to replicate your results. Include some **concise comments in your code** to justify your decisions. For questions that you solve only theoretically, you can include the answers as code comments as well.

Also, a **PowerPoint presentation of 4-5 slides** with the results and main conclusions is expected.

The exercise is demanding from the point of view of time available, so the use of **efficient and pragmatic** solutions is highly recommended. They should let you tackle all questions with the shortest development time. If you are running out of time, we recommend that you set out the remaining sections with a **brief description of how you would approach the problem**. In particular for the last two sections (next steps and conclusions), a theoretical explanation is sufficient.

Please note that all comments, explanations and answers should be **written in English**.

Data sources

This assessment uses the **churn_all.csv** dataset, which you should have received.

The column names of should be self-explanatory. However, if more detail were to be needed, you can find a data table with the variable names and descriptions.

Assessment steps

To solve the proposed business problem, complete the following tasks:

- Explore and decide **which variables could be significant** for the model and generate at least one **data visualization resource** that supports your claim.
- Create a **new categorical variable** to reasonably **group customers into two consumption levels**.
- Build a **classification model** by selecting a set of variables to include while justifying your decision.
- **Evaluate the accuracy** of the model by leveraging:
 - Which **metric or metrics** to use

- Which **validation strategy** ensures generalizable results
- **Discuss and analyze how each variable contributes** to the churn rate according to the model. Which variables are the most impactful? How are they affecting the churn? Does it make sense?
- Imagine that you have now been informed that the Telcom Company is planning to carry out **commercial actions** on a sample of **500 customers** to avoid potential outflows. How **would you select** the 500 customers that **maximize the success** and **profitability** of these actions?
- Summarize in **8-10 lines the most important messages, learnings and conclusions of the analysis and the model you have developed** (write it as if you were to present it to the Telcom Company).

Delivery format

After completing the tasks described above, please submit the code and the presentation you developed for the business case via email. If you are using R, submit it as an **R-Markdown** file; if you are using Python, submit it as a **Jupyter Notebook**. Additionally, it would be helpful if you also export the document to **HTML** format and include it in the email.

Finally, as it has been mentioned, make sure all your comments and annotations are written in **English**.

Variable dictionary

Variable Name	Description
state	Two-letter code from the given USA state
area_code	Telephone area code (prefix)
phone_number	Telephone number
international_plan	The customer has a contracted telephone international plan
voice_mail_plan	The customer has a contracted voice mail service
number_vmail_messages	Number of messages in voice mail
total_day_minutes	Morning – Consumption in minutes
total_day_calls	Morning – Number of calls
total_day_charge	Morning – Amount to pay
total_eve_minutes	Evening – Consumption in minutes
total_eve_calls	Evening – Number of calls
total_eve_charge	Evening – Amount to pay
total_night_minutes	Night – Consumption in minutes
total_night_calls	Night – Number of calls
total_night_charge	Night – Amount to pay
total_intl_minutes	International - Consumption in minutes
total_intl_calls	International - Number of calls
total_intl_charge	International – Amount to pay
number_customer_service_calls	Number of customer service calls