

HACKATOW

First Challenge: Prediction of churn for an online retail store

February 2025

A business of Marsh McLennan

CONFIDENTIALITY

Our clients' industries are extremely competitive, and the maintenance of confidentiality with respect to our clients' plans and data is critical. Oliver Wyman rigorously applies internal confidentiality practices to protect the confidentiality of all client information.

Similarly, our industry is very competitive. We view our approaches and insights as proprietary and therefore look to our clients to protect our interests in our proposals, presentations, methodologies, and analytical techniques. Under no circumstances should this material be shared with any third party without the prior written consent of Oliver Wyman.

© Oliver Wyman

PREDICTING CHURN FOR AN ONLINE RETAIL STORE



CONTEXT

- Oliver Wyman has been engaged by an online retailer which has lately been having trouble with customer retention to develop a model to detect which customers are more likely to churn
- Your work will directly feed their future strategy, where the client hopes to target customers with personalized marketing strategies, driving customer loyalty and retaining them for longer
- To support the Oliver Wyman team, the client has shared multiple datasets for us to understand the current behaviour of the clients. The information also includes a flag that indicates whether the customer churned after the observation date. **The churn flag should be used as the target variable in the model**



YOUR TASK

- Your manager, an expert in the field of online commerce, provides you with the following guidelines:
 - 1 Consider a **logistic regression** to start with as being a widely used industry standard. In particular, make sure to consider the following:
 - In order to test the performance of your model, please **use data from 2016 to 2018 as Out-Of-Time (OOT)** sample
 - Please **select appropriate variables**
 - Please **assess the performance of your model**
 - 2 Your supervisor wants to leverage **Machine Learning techniques**
 - Can you think of other models that use a more sophisticated approach?
 - What are the advantages and disadvantages of such an approach in this setting?
- Details on the datasets are in the following slides

DATASET 1. CLIENT INFORMATION



client_info.csv

Column	Description
User ID	Unique internal ID of the user
Client location	Location where the user resides
Activity	User's activity profile on website
Marital	User's marital status
Tenure	User's account tenure
Gender	User's gender
Customer satisfaction score	Metric 1: calculated to describe the user's satisfaction
Reduced customer satisfaction score	Metric 2: calculated to describe the user's satisfaction
Spending to lifetime ratio	Metric encapsulating how much a user spends
Number of accounts	Number of unique accounts a user has (e.g., one for parents, one for children)
Observation date	Date of observation
Churn	Flag indicating whether the user has churned

DATASETS 2 & 3. PURCHASE AND LOCATION INFORMATION



purchase_info.csv

Column	Description
User ID	Unique internal ID of the user
Account ID	ID of accounts a user has
Total spending	Total spent over user's lifetime
Average order value	Average order/shopping cart value



city_info.csv

Column	Description
City	City where customer resides
Population	Population of city
GDP per Capita (2020)	GDP per capita of city in January 2020

OTHER IMPORTANT INFORMATION



ADDITIONAL COMMENTS

- You can work in any programming language you prefer, although we would strongly recommend working in **Python and Jupyter Notebooks**
- The goal is not to assess how well you can code – you should rather familiarize yourself with the **concept of modelling and** be able to explain the **steps needed** to produce a good model
- Think about whether **all of the variables are appropriate** to be used in a churn prediction model **or** whether you can **think of any reason not to use them**, even if they are predictive
- **For Task 2 you do not need to implement the model** – a purely theoretical approach is sufficient (however, if you want to implement it, please do so – you will probably notice a few interesting things!)
- **You can leverage any resources that you can find** on the web – the main thing is that you can show how well you understood the concepts of:
 - Building a good model in general
 - What to consider when choosing a model depending on its use



SUBMISSION DETAILS

- Deadline to submit the deliverables is:
4th of April before 23:59h CEST
- Deliverables should be submitted by email to:
recruiting.iberia@oliverwyman.com
- The subject of the email should be:
HackatOW Challenge Iberia 2025 - *Name of the Team* - Results Submission
- Candidates must share codes in a **zip file** named **Hackatow2025_*Team Name*_Team Member(s) Name(s)** and code will be expected to run directly from raw datasets. This is a requirement for challenge submission

QUALIFICATIONS, ASSUMPTIONS, AND LIMITING CONDITIONS

This report is for the exclusive use of the Oliver Wyman client named herein. This report is not intended for general circulation or publication, nor is it to be reproduced, quoted, or distributed for any purpose without the prior written permission of Oliver Wyman. There are no third-party beneficiaries with respect to this report, and Oliver Wyman does not accept any liability to any third party.

Information furnished by others, upon which all or portions of this report are based, is believed to be reliable but has not been independently verified, unless otherwise expressly indicated. Public information and industry and statistical data are from sources we deem to be reliable; however, we make no representation as to the accuracy or completeness of such information. The findings contained in this report may contain predictions based on current data and historical trends. Any such predictions are subject to inherent risks and uncertainties. Oliver Wyman accepts no responsibility for actual results or future events.

The opinions expressed in this report are valid only for the purpose stated herein and as of the date of this report. No obligation is assumed to revise this report to reflect changes, events, or conditions, which occur subsequent to the date hereof.

All decisions in connection with the implementation or use of advice or recommendations contained in this report are the sole responsibility of the client. This report does not represent investment advice nor does it provide an opinion regarding the fairness of any transaction to any and all parties. In addition, this report does not represent legal, medical, accounting, safety, or other specialized advice. For any such advice, Oliver Wyman recommends seeking and obtaining advice from a qualified professional.