

# Data Science challenge

Nordeus - Job Fair 2023

## About the challenge

The usual phases tackled while doing ML (Machine Learning) projects are detecting business opportunities, formalizing the problems, getting and cleaning the data, developing and training the model, and integrating all into the production pipeline. For this assignment, we defined the problem, prepared the data for you, and would like to see you develop and train an ML model!

Before making an ideal matchmaking algorithm we want to be able to predict league rank (position in the league table) at the end of the season in order to provide the most fair and challenging experience to our users.

We will provide Top Eleven data describing users at the last day of the previous season and id of the league in the current season. The goal of this challenge is to predict **league rank** at the end of the current season for each user within the league.

## Data description

There are two datasets:

\* jobfair\_train.csv

\* jobfair\_test.csv (same as training dataset but without target variable)

You can find the dataset description below:

Variable	Definition
season	Season identifier. Within one season, various competitive activities happen, including League Matches, Cups, the Champions League and other matches. Each season lasts for 28 days.
club_id	Unique club number.
league_id	Unique league number.  Each league consists of 14 clubs that compete against each other.
dynamic_payment_segment	Segmentation of users: <ul style="list-style-type: none"><li>• '0) NonPayer' - a user who does not spend money in the game.</li><li>• '1) ExPayer' - a user who used to spend money in a game but no longer does.</li><li>• '2) Minnow'</li><li>• '3) Dolphin'</li><li>• '4) Whale'</li></ul>

cohort_season	Number of seasons since the user's registration date.
avg_age_top_11_players	Average number of years of the top 11 players.
avg_stars_top_11_players	Average number of stars of the top 11 players. Stars represent the quality of players.
avg_stars_top_14_players	Average number of stars of the top 14 players. Stars represent the quality of players.
avg_training_factor_top_11_players	Average training factor of the top 12 players.
days_active_last_28_days	Number of days the user was active in the last 28 days.
league_match_watched_count_last_28_days	Number of league matches the user watched in the last 28 days.
session_count_last_28_days	Number of sessions the user had in the last 28 days.
playtime_last_28_days	Total playing time in minutes in the last 28 days.
registration_country	Country of origin of the first session (registration).
registration_platform_specific	Platform through which the user had their first session (registration).
league_match_won_count_last_28_days	Number of wins the user has achieved in league matches in the last 28 days.
training_count_last_28_days	Number of training the user has done in the last 28 days.
global_competition_level	Users can take part in global competition and can progress through this competition system. The higher the competition level the greater the progress the user has made. Users can also opt out from this competition thus having no global_competition_level set.
tokens_spent_last_28_days	Amount of tokens spent in the last 28 days.
tokens_stash	Amount of tokens in the stash.
rests_stash	Amount of rests in the stash.
morale_boosters_stash	Amount of morale boosters in the stash.
league_rank	Target variable. Position in the league table at the end of the season. 1 - means club finished first in the league 14 - means club finished last in the league This field is available only in the training and not in the test dataset.

## Target variable

Target variable is **league\_rank**. It is an integer describing the position of the club (user) at the end of the current season within the league identified by .

## EVALUATION

### What is your goal? 🎯

It is your job to predict the league rank within each league (identified by league\_id) for each club (identified by club\_id) in the test dataset.

### Metric 📊

Submissions are evaluated on [Mean-Absolute-Error \(MAE\)](#) between predicted league rank and the observed league rank.

### Submission format 📄

You should save your results in the “league\_rank\_predictions.csv” file.

The submission file should contain a row for each club in the test dataset and 2 columns: club\_id and predicted\_league\_rank.

```
club_id,league_rank
6048216,2
14503204,5
etc.
```

We expect the submission to be sent via email to [jobfair@nordeus.com](mailto:jobfair@nordeus.com) with a link to your GitHub repository (email subject: Data Science challenge). Besides the file with predictions, the repository should contain scripts/notebooks with code that shows how predictions were made. For this challenge you can use the language of your choice preferably Python or R.

**Challenge is open until November 20, 2023** end of day. Good luck!