Data project task

As we learned during the lectures, platforms play an important role in the current digital business environment. Marketplaces (including marketplaces for the financial services) is one of the key platform types. To create the network effect, marketplaces need to attract both buyers and sellers.

You are provided with the data from Vinted marketplace:

column	type	description
portal	character	portal
id	numeric	item id
created_at	POSIXct	time when item was uploaded
user_id	numeric	user id
listing_price_eur_fixed	numeric	item price
status	character	condition of the item
sale_time	POSIXct	time when listing was sold
gmv_eur_fixed	numeric	price for how much the item was sold
brand	character	item's brand
brand_is_verified	logical	condition whether brand was verified
declined_at	POSIXct	time when item was restricted
color_primary	character	primary color of the item
listing_quality_string	character	string explaining item's quality
suggested_price_maximum	numeric	suggested maximum price for the item by our system
catalog_code_1	character	top level item category
catalog_code_2	character	2nd level item category if available
catalog_code_3	character	3rd level item category if available
catalog_code_4	character	4th level item category if available
catalog_code_5	character	5th level item category if available
gender	character	user gender
country_code	character	user country
lister_nth_listing	numeric	number of listings made by the same user
first_listing_local_date	Date	time when first listing was made by the user
listing_platform	character	device platform in which item was uploaded
registration_platform	character	device platform that was used during registration
registration_local_date	Date	time when user registered
total_positive_feedback_count	numeric	number of positive feedbacks received
total_negative_feedback_count	numeric	number of negative feedbacks received
window_items_listed	numeric	number of items listed by user in 30 days
window_items_bought	numeric	number of items bought by user in 30 days
window_items_sold	numeric	number of items sold by user in 30 days
listings_in_first_7days_detailed	character	number of items listed by user in the first 7 days after registration

Analyze the training dataset <u>from seller's perspective</u> and create the model to classify the listed item to one of two categories:

- item will be sold,
- item will not be sold.

Use the model to classify the listings in the test dataset.

<u>Prepare the following:</u>

- presentation summarizing the work done, models used, key assumptions, validation performed and insights from the data on the following topics:
    - How would you describe liquidity/efficiency/health of the Vinted marketplace?
    - What are the main factors having an impact on item's success to be sold?
    - What actions could be taken to have as many listings sold and as quickly sold as possible?
- csv file with the classification result for each test dataset item id (id field).

The competition winner will be announced based on the test dataset F_2-score:

$$F_2 = 5 \times (Precission \times Recall)/(4 \times Precission + Recall)$$