**Introduction**

As is known, Kiwi.com sells flight tickets (not only flight, also bus, train, etc. but that’s not important here). Together with these tickets a person can buy additional products called ancillaries. One of these ancillaries is Kiwi.com’s Virtual Fare Type which, for some price, gives the person e.g. an option to use e-mail communication with our customer support, use free change of reservation or even getting back 90 % of the price if you decide to cancel the reservation (there’s more, feel free to browse through our website, fill some data about “passenger” and check it yourself).

**Task**

Predict probability that a customer will buy Kiwi.com’s Virtual Fare Type, try to find some relationship between predictors and target.

**Terminology**

In the next section there is used our internal terminology which we briefly explain here, in particular terms sector and segment.

Each itinerary consists of sectors - separate logical parts of the itinerary. Oneway consists of one sector (way there), roundtrip from two (way there and way back), multicity and nomad itineraries can consist of one or multiple sectors.

Each sector can consist of one or more segments, i.e. particular flights a passenger is taking.

**Data**

You can find data about whether or not a customer bought this ancillary for a given itinerary here (link na drive). It’s a sample data from our sales. There are two datasets, first (itineraries.csv) giving overall information about an itinerary together with target and second (segments.csv) giving information about each segment of each sector of an itinerary.

Feel free to use any data which you consider useful.

Brief explanation of some columns (others should be quite self-explanatory):

* *fare\_type* stands for target, if it’s saverit means that the ancillary wasn’t bought, standard and flexi mean that virtual fare type was bought (standard and flexi are two of the paid options)
* *content\_type* - specifies whether the itinerary is virtually interlined or not, i.e. whether we connected some flights/airlines that don’t cooperate together or not
* *is\_hold\_bags\_disabled* - specifies whether checked baggage can be bought or not for this itinerary
* *base\_price* - price for which we buy the ticket from our provider (i.e. price without our markup)
* *fare\_category* - specifies class of the flight ticket, M stands for economy class, W for premium economy, C for business class and F for first class

For joining of the two datasets use column *bid* (booking id).

**Output**

Some shareable and reproducible format which we can check, ideally a (private) github repository with some code, whether it’s jupyter notebooks or whatever, together with a brief description of why you did what you did. Share the results with us, we will go through it and during the next round of the interview you will present your findings, evaluate the performance of your model and we will have a discussion about your solution. We are not expecting the best solution, we are more interested in your methodological approach to the task at hand.