**Product Analyst - Assignment**

Dear candidate,

Thanks in advance for doing this exercise – as well as helping us get to know you, we think it also provides a great way for you to understand and explore the type of topics you will be working on at OLX Group if you join us.

We have prepared for you an assignment that we would like you to focus on.

The objective of this task is to calculate retention rate - which is defined as percentage of users who visited the platform on a given day and returned within the next 3 days. This ratio helps us to understand if users are satisfied with services by platform and are willing to revisit it. Based on given data we ask you to analize retention.

**Retention rate**

Retention is defined as the percentage of users who visited the platform on a given day and returned within the next 3 days.

**Example**

Users Alice, Bob, Eve and Charlie were browsing real estate offers on our platform on 1st of June. Alice revisited the platform the next day and on 3rd of June. Bob returned on the 4th of June. Eva revisited on 5th of June. Charlie did not return to our platform. Retention rate for 1st of June is 2/4 = 50%.

**Technical part**

* Prepare SQL query, which calculates the retention rate per day (can be in separate file .sql or a part of script in Python/R).
* Prepare a notebook in Python/R with codes to calculate retention rate per day. Prepare daily data with calculated retention rate for date range from 1st of June to 7th of June. Additionality check what is the average rate for this timeframe.

**Analytical part**

Prepare full data analysis, including answers to following questions:

* Analyse retention rate based on provided data. What is influencing retention rate? In which area users are less likely to return? What may be the cause of that?
* Are users only visiting one category? What is the retention rate for users, who visit only one category? Compare differences between categories.
* Verify hypothesis that users using ios have higher retention rate than users using mobile web version.

**Form**

1. Prefered Jupyter Notebook or R markdown for analysis.
2. Scripts may be in separate files or as a part of a notebook.
3. Final results and most important findings please include in a presentation (eg. Google Slides).

**There are 3 data files available**

Data has been randomly selected for date range 01-10.06.2021.

* **data\_visits.csv** - Users’ activities (events). Any activity means that a user visited the platform. Possible activities:
  + listing - displaying results of a search in a listing form
  + ad\_page - displaying a single offer

Columns: user\_id,session\_id, time\_stamp, event\_name, ad\_id,

* **data\_users.csv** - information about device(s) used by a user

Columns: user\_id, session\_id, channel

* **data\_ads.csv** - mapping to category tree

Columns: ad\_id, category\_name\_1, category\_name\_2