Predication of user's subscription based on user activity data

Supervision: Mayya Lihovodov

Idan Zorer **Amit Yahav**

Matan Mizrahi **Yigor**

Shira Lugasi

Introduction:

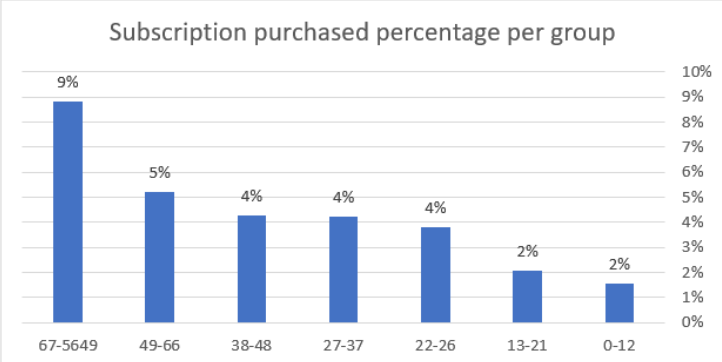
Youkid is an app that offers a wide range of interactive games, captivating movies, and delightful characters.

The main objective of the project is to create a predictive model that determines whether a user is likely to subscribe to the Youkid application or not.

Data description:

Our data included information about users' actions in the application, such as activity times, games played, gender, platform, and more.

Using the various parameters and subscription purchase data, we were able to build a predictive purchase model



Machine Learning step:

we used logistic Regression model to visualize the subscription purchased.
In our case, the binary outcome is subscription purchased or not.
The model is a classification model that fits the problem of buying a subscription.

Stages of work:

- Data merging – In this project, we merged a variety of data files obtained by Mixpanel (a product analytics tool) with data from YouKid, we merged the files in the best way to allow us to work and analyze them.
- Data cleaning- we cleaned the data from features that are unlearnable and contain a lot of missing values, as well as outliers.
- EDA – In order to find out if there is a correlation between using the application and purchasing a subscription, we binned our data (binning is a technique for grouping continuous or numeric data).

Confusion Matrix:

```
In [6]: confusion_matrix(smote_df.Actual, smote_df.Predicted)
Out[6]: array([[1303,  269],
               [  17,   34]], dtype=int64)

In [10]: smote_df.Actual.count()
Out[10]: 1623
```

Final Results:

```
Accuracy: 0.8274799753542822
Precision: 0.11186440677966102
Recall: 0.6470588235294118
F1 Score: 0.19075144508670522
```

In this case, the model has a relatively high number of false positives and struggles to correctly identify positive samples.