# Homework 1. Customer segmentation

Should be done in mini groups of 3-4 students. The deadline **is 22nd of October, 23:59.**

**The task:**

Imagine that you are the marketing analysts of a marketplace. The company wants to drive sales and customer activity using a data-driven approach.

You are provided with a dataset, describing the following data:

1. Customer features (id, age, city,country, registration date on the website or in ecom app, communication channels available)

2. Transactional features:  
- OrderId - unique order number of the customer  
- Order Date - date when the order was placed  
- Ship Date - date of order shipment  
- Ship Mode - delivery method  
- Customer ID - unique customer number  
- Country/Region - delivery region  
- City - delivery city  
- State - delivery state  
- Postal Code - postal code  
- Region - region of product delivery  
- Product ID - product identifier  
- Category - product category  
- Sub-Category - product sub-category  
- Product Name - product name  
- Sales - sales volume   
- Quantity - number of ordered units  
- Discount - discount in % (0.8 means an 80% discount)  
- Profit - margin profit

You should segment the sample of customer base in order to find the key consumer segments and describe marketing strategy for each of segments to achieve the company’s goals.

1. Analyze the data. What features are there in the dataset? Which of them may be useful for the segmentation? Clean the dataset, if needed (1 points).

2. Use the cohorts to analyze the customers activity in dynamics. What can be said about the customers behavior? (1.5 points)

3. Perform the cluster analysis and RFM analysis. Choose the optimal number of clusters (should be shown) and features (try various combinations to find the most appropriate one). Generate additional features, if they are needed. Compare the results of 2 segmentations, choose the best one. (1.5 points)

4. Describe the customers in each group. Which significant characteristics can be underlined? (1 point)

5. Choose the KPI (“metrics of success”) for each segment (what can be grown/ decreased/ increased), explain your choice. (1 point)

6. Think of possible marketing initiatives that could potentially drive the chosen metrics for each segment. Generate 3-5 hypothesis and test them using statistical criteria (remember that we can use parametric criteria like t-test only if the data distribution is normal). (2 points)

7. Write the marketing strategy based on the results of hypothesis testing. Don’t forget about the communication channels! Suggest the formats and place of communication message delivery (emails, banners in mobile app, etc.). Evaluate the potential effect of the suggested strategy on the chosen metrics (2 points)

As the result of the homework your team should provide:

1. The colab file \ Jupiter notebook with the code, compilation results, the findings and all the necessary

Explanation \ the group repository link

2. The presentation (ppt\pdf) with the key learnings and the marketing strategy description

Some of you will show your presentations on 23rd of October during the class. Presentation timing – max 15 min, 5-10 min for questions. Others – online at another time – timetable will be announced some time after.

Useful materials:

The example of cohort analysis and RFM analysis:

<https://colab.research.google.com/drive/1FnI3P5c0OS726-2hmwRhPU8mjQmnVJ2f?usp=sharing>

The dataset from the example:

<https://drive.google.com/file/d/1kZP16STWCKtpLcE7RGBUH6-at311hcGt/view>