**Day 9&10 - Group ML Challenge**

**Challenge 1**

**Customer segmentation**

You are given a dataset of [credit card transactions](https://drive.google.com/uc?export=download&id=1TU-HZ_FdvGge8AfSuaT5LtmSJSbAfKhc). The dataset contains 8500 customers and their activity over a 6-month period. Each transaction is characterized by 18 features described below.

Your task is to find the most useful customer segmentation to improve the marketing campaigns of the company.

**The features:**

* customer\_id : id of the credit card holder
* balance\_account : balance amount left in the account to make purchases at the end of the 6-month period
* purchases\_amount : amount of purchases made from account
* paid\_advance : number of transactions made with “Cash in Advance” method
* full\_purchases : maximum purchase amount done in full payment
* balance\_frequency\_update : how frequently the balance has been updated, score between 0 and 1 (1 = frequently - updated, 0 = not frequently updated)
* installments : amount of purchases done in installments
* purchases\_frequency : how frequently the purchases are being made, score between 0 and 1 (1 = frequently - - purchased, 0 = not frequently purchased)
* full\_purchases\_frequency : how frequently purchases are happening in full payment (1 = frequently purchased, - 0 = not frequently purchased)
* purchases\_installments\_frequency : how frequently purchases in installments are being done (1 = frequently - done, 0 = not frequently done)
* cash\_advance\_frequency : how frequently the cash in advance being paid
* nr\_cash\_advances : cash in advance given by the user
* nr\_purchases : number of purchase transactions made
* credit\_limit : limit of credit card for user
* fixed\_rate\_period : duration of fixed interest rate of credit card service for user (in months)
* payments : amount of payment done by user
* min\_payments : minimum amount of payments made by user
* full\_payment\_prc : percent of full payment paid by user

**Challenge 2**

**Telcom Churn Prediction**

Here’s a [dataset of telecom customers](https://drive.google.com/uc?export=download&id=1A3MUldrs0z09DlYR6Y1utfySwKNO9Qsz), which you can also find on [kaggle](https://www.kaggle.com/datasets/blastchar/telco-customer-churn). There is data available on 5’986 customers.

Which customers are likely to churn? What are the attributes that make you think so?

**The features:**

* customerID - customer id
* gender - client gender (male / female)
* SeniorCitizen - is the client retired (1, 0)
* Partner - is the client married (Yes, No)
* tenure - how many months a person has been a client of the company
* PhoneService - is the telephone service connected (Yes, No)
* MultipleLines - are multiple phone lines connected (Yes, No, No phone service)
* InternetService - client’s Internet service provider (DSL, Fiber optic, No)
* OnlineSecurity - is the online security service connected (Yes, No, No internet service)
* OnlineBackup - is the online backup service activated (Yes, No, No internet service)
* DeviceProtection - does the client have equipment insurance (Yes, No, No internet service)
* TechSupport - is the technical support service connected (Yes, No, No internet service)
* StreamingTV - is the streaming TV service connected (Yes, No, No internet service)
* StreamingMovies - is the streaming cinema service activated (Yes, No, No internet service)
* Contract - type of customer contract (Month-to-month, One year, Two year)
* PaperlessBilling - whether the client uses paperless billing (Yes, No)
* PaymentMethod - payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic))
* MonthlyCharges - current monthly payment
* TotalCharges - the total amount that the client paid for the services for the entire time
* Churn - whether there was a churn (Yes or No)

**Challenge 3**

**Olympic Athletes**

You will be working with a [historical dataset on the modern Olympic Games](https://drive.google.com/uc?export=download&id=17IPb-3hwhv9wRreC_NWd8p5YBB2T3n6M), which includes all the games from Athens 1896 to Rio 2016. The data was scraped from www.sports-reference.com in May 2018.

The dataset athlete\_events.csv contains 271’116 rows and 15 columns. Each row corresponds to an individual athlete competing in an individual Olympic event. Can you predict who will win?

**The features:**

* ID - Unique number for each athlete
* Name - Athlete’s name
* Sex - M or F
* Age - Integer
* Height - In centimeters
* Weight - In kilograms
* Team - Team name
* NOC - National Olympic Committee 3-letter code
* Games - Year and season
* Year - Integer
* Season - Summer or Winter
* City - Host city
* Sport - Sport
* Event - Event
* Medal - Gold, Silver, Bronze, or NA