**Day 2**

## Project: Predict customer churn in a bank

### **Data Content**

* RowNumber—corresponds to the record (row) number and has no effect on the output.
* CustomerId—contains random values and has no effect on customer leaving the bank.
* Surname—the surname of a customer has no impact on their decision to leave the bank.
* CreditScore—can have an effect on customer churn, since a customer with a higher credit score is less likely to leave the bank.
* Geography—a customer’s location can affect their decision to leave the bank.
* Gender—it’s interesting to explore whether gender plays a role in a customer leaving the bank.
* Age—this is certainly relevant, since older customers are less likely to leave their bank than younger ones.
* Tenure—refers to the number of years that the customer has been a client of the bank. Normally, older clients are more loyal and less likely to leave a bank.
	+ Balance—also a very good indicator of customer churn, as people with a higher balance in their accounts are less likely to leave the bank compared to those with lower balances.
	+ NumOfProducts—refers to the number of products that a customer has purchased through the bank.
	+ HasCrCard—denotes whether or not a customer has a credit card. This column is also relevant, since people with a credit card are less likely to leave the bank.
	+ IsActiveMember—active customers are less likely to leave the bank.
	+ EstimatedSalary—as with balance, people with lower salaries are more likely to leave the bank compared to those with higher salaries.
	+ Exited—whether or not the customer left the bank.

You are requested to answer the following:

* + 1. How many exited customers from bank (n and %);
		2. How many exited customers from bank per location (n and %);
		3. How many exited customers from bank per location divided by gender (n and %);
		4. Using Binomial distribution approach, we want to predict the probability out of 1000 customers, how many will exit the bank.
		5. Using Binomial distribution approach, we want to predict the probability out of 1000 customers, how many will exit the bank for each country.
		6. If the management want to implement an advertisement, using the results from (4) and (5) whom are the specific target customers.
		7. Using a statistical test, is there is a significant difference among the countries based on customers who left the bank.
		8. Using a statistical test, is there is a significant difference among the gender based on customers who left the bank.
		9. Using a statistical test, is there is a significant difference among gender levels based on customers who left the bank.
		10. Calculate the descriptive statistics for customer “balance” in each country.
		11. Calculate the descriptive statistics for customer “balance” in each country divided by “Exited”.
		12. What is the probability of a customer with balance more that 80,000 from exited records?
		13. What is the probability of a customer with balance more that 80,000 from non-exited records?
		14. Giving that, the average salary of Bank Accounts Manager is 1,750 KWD with standard deviation equal 400 KWD, assume that the salaries followed the normal distribution, what is the probability that employee salary will be more that 2,300 KWD.

Useful calculator:

1. Binomial distribution:

<https://homepage.divms.uiowa.edu/~mbognar/applets/bin.html>

1. Normal Distribution

<https://onlinestatbook.com/2/calculators/normal_dist.html>