Which topic did you choose to apply the data science methodology to? **(2 marks)**

I have selected the application of data science in Credit Card fraud as for this task . I have chose this topic because I am interest in the credit and banking markets .

Next, you will play the role of the client and the data scientist.

Using the topic that you selected, complete the Business Understanding stage by coming up with a problem that you would like to solve and phrasing it in the form of a question that you will use data to answer. **(3 marks)**

You are required to:

1. Describe the problem, related to the topic you selected.
2. Phrase the problem as a question to be answered using data.

For example, using the food recipes use case discussed in the labs, the question that we defined was, "Can we automatically determine the cuisine of a given dish based on its ingredients?".

A major challenge in the credit card industry is the risk associated with fraud ( "28.65 billions dollars were lost just in 2019 do to credit card fraud"¹) and the difficulty of detecting it. It is unsustainable investment to have thousands of employees manually verifying millions of translations that occur every single day. However utilizing a model could offer a more reliable, faster, and cost-effective model for determine whether a transaction is suspicious or not.

Therefore, our main question becomes; "Can we develop a model capable of identifying whether a specific transaction is suspicious ?

¹ - Source: Medium

Aviso

Briefly explain how you would complete each of the following stages for the problem that you described in the Business Understanding stage, so that you are ultimately able to answer the question that you came up with. **(5 marks)**:

1. Analytic Approach
2. Data Requirements
3. Data Collection
4. Data Understanding and Preparation
5. Modeling and Evaluation

You can always refer to the labs as a reference with describing how you would complete each stage for your problem.

1. Analytic Approach: Determine the feasibility of our model based and available data and resources. 2. Data Requirements: To develop an analytical model capable of accurately identifying suspicious transactions, we need information related to bank clients,this information should include personal data, transaction timestamps, transaction geolocation, and more. 3. Data Collection: Daily data on individual credit card transactions and customer behavior will be acquired and stored in a database, containing unique identifiers and personal information. 4. Data Understanding and Preparation: Statistical analysis (descriptive, diagnostic) will be employed to assess the suitability of data for model creation. This phase involves evaluating different variables to gain a better understanding using statistical parameters such as median, pearson correlation coefficient, and the presence of outliers (which can serve as indicators of potential fraud). 5. Modeling and Evaluating: Subsequently, a model will be developed using clustering techniques to identify outliers in geolocations based on individual consumption habits. The model's trained outcomes will be evaluated, and necessary adjustments will be made. If the model successfully passes the train, it will pass for the test phase, then evaluation phase, then the implementation phase can start, followed by feedback collection.