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

I have chosen as topic for this task the application of data science in the field of credit cards. The reason behind this choice is that it is related with my finance education.

### **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?".

So the main problem for banks regarding credit cards is that they have to create a model to know to who they can provide them. Certain clients will not be feasible as they do not have the economic strenght to back up this service.

So our question would be " Can we automatically determine if a client is suitable to obtain a credit card?

### **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: As the problem requires a yes/no answer we will use a classification model
2. Data Requirements: To create the classification model we will require information regarding the bank clients. This info should include personal data of the client and should include the ones that defaulted and the one that paid.

3: Data Collection: We would use techiques like descriptive statistics and data evalution should be implemented in this phase to make sure that we have useful data for our model.

4: Data Undestanding and Preparation: In this step we need to evaluate the different variables of our data in order to undestant it better. For example we would calculate univariate statistics, such as mean or median and the correlation between variables. So we need to evaluate the quality of the data. In the data preparation phase we have to prepare the data in an specific way depending on the model.

5: Modeling and Evaluation: Lastly we create a classification model, evaluate the outcome and perform the corresponding changes untill we have a suitable model.

### **Titre**

data science methodology appliqué au problème de la réadmission en néonatologie

Data science methodology applied to the problem of readmission in neonatology

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

I chose to apply data science methodology to the topic of Hospitals because I work in this field as a researcher for a health agency.

### **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?".

So the main problem for banks regarding credit cards is that they have to create a model to know to who they can provide them. Certain clients will not be feasible as they do not have the economic strenght to back up this service.

So our question would be " Can we automatically determine if a client is suitable to obtain a credit card?

En tant que responsable du département de néonatalogie dans un hôpital (le client), je dois m’assurer de minimiser le taux de réhospitalisation dans les 30 jours. À cette fin, j’aimerais pouvoir évaluer la probabilité qu'une patiente revienne à l’hôpital en me basant sur l’évaluation de son bilan de sortie

Comme data scientist, je pourrais traduire ce problème en formulant la question suivante: “Quel est la probabilité de réhospitalisation d’une patiente quittant le département de néonatalogie en se basant sur l’historique des données disponibles?”

sur l'évolution médicale durant l'hospitalisation, les traitements administrés et les diagnostics établis?”

1. As the head of the neonatology department in a hospital (the client), I must ensure that the 30-day readmission rate is minimized. To this end, I would like to assess the probability of a patient returning to the hospital based on the evaluation of their discharge summary.
2. As a data scientist, I could translate this problem by formulating the following question: “What is the probability of readmission for a patient leaving the neonatology department based on the available historical data?”

### **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: The question we are trying to answer is, “What is likely to happen?” This represents a predictive analytics approach.
2. Data Requirements: To create the classification model we will require information regarding the bank clients. This info should include personal data of the client and should include the ones that defaulted and the one that paid.

Data Requirements: Pour créer un modèle prédictif de la réhospitalisation dans les 30 jours, nous pourrions recourir aux données historiques de l'évolution médicale durant l'hospitalisation, des traitements administrés et des diagnostics établis.

Data Requirements: To create a predictive model for 30-day readmissions, we could use historical data on medical progress during hospitalization, administered treatments, and established diagnoses.

1. Data Collection: We would use techiques like descriptive statistics and data evalution should be implemented in this phase to make sure that we have useful data for our model.

Data collection: Data scientists apply descriptive statistics and visualization techniques to thoroughly assess the content, quality, and initial insights gained from the collected data, identify gaps, and determine if new data is needed or to substitute existing data

Data collection: À l’aide de statistiques descriptives et de méthodes de visualisation, nous pourrions évaluer le contenu, la qualité et l’utilité des données disponibles parmi celles jugées pertinentes.

Data collection: Using descriptive statistics and visualization methods, we could evaluate the content, quality, and usefulness of the available data among those considered relevant in the “Data Requirements” stage.

1. Data Undestanding and Preparation: In this step we need to evaluate the different variables of our data in order to undestant it better. For example we would calculate univariate statistics, such as mean or median and the correlation between variables. So we need to evaluate the quality of the data. In the data preparation phase we have to prepare the data in an specific way depending on the model.

Data Undestanding and Preparation: Dans un premier temps, nous devons nous efforcer de comprendre de quelle manière les données recueillies sont en mesure d’expliquer la réhospitalisation des patientes. Pour ce faire, nous pouvons utiliser des statistiques descriptives comme la moyenne ou l’écart-type ou encore des statistiques prédictives comme la corrélation. Pour la préparation des données, nous devons nous attarder à certains problèmes comme les données manquantes, les duplicates et s’assurer qu’elles soient présentées dans le bon format pour s’assurer de l’efficacité du modèle.

Data Undestanding and Preparation: To begin, we would like to understand how the collected data can explain patient readmissions. This can be achieved using descriptive statistics such as the mean or standard deviation, as well as predictive statistics like correlation. For data preparation, attention must be given to issues such as missing data, duplicates, and ensuring that the data is presented in the correct format to maximize the model's effectiveness.

1. Modeling and Evaluation: Lastly we create a classification model, evaluate the outcome and perform the corresponding changes untill we have a suitable model.

Modeling and Evaluation: The data modeling process uses a training data set. Data scientists test multiple algorithms on the training set data to determine whether the variables are required and whether the data supports answering the business question. The outcome of those models are either descriptive or predictive. During the Evaluation stage, data scientists and others assess the quality of the model and determine if the model answers the initial Business Understanding question or if the data model needs adjustment.

Modeling and Evaluation: Dans la phase de modélisation, nous pouvons entraîner différents algorithmes à l’aide d’un training data set. Cela doit nous permettre d’évaluer quel modèle le mieux la question d’affaires et quelles données sont les plus pertinentes. Dans la phase d’évaluation, nous devons évaluer le meilleur modèle et déterminer si celui-ci requiert des ajustements.

Modeling and Evaluation: In the modeling phase, we can train various algorithms using a training dataset. This should help us determine which model best addresses the business question and which data is most relevant. In the evaluation phase, we need to assess the best-performing model and identify whether it requires any adjustments.

Final 3

1. Analytic Approach: The question we are trying to answer is, “What is likely to happen?” This represents a predictive analytics approach.
2. Data Requirements: To create a predictive model for 30-day readmissions, we could use historical data on medical progress during hospitalization, administered treatments, and established diagnoses.
3. Data collection: Using descriptive statistics and visualization methods, we could evaluate the content, quality, and usefulness of the available data among those considered relevant in the “Data Requirements” stage.
4. Data Undestanding and Preparation: To begin, we would like to understand how the collected data can explain patient readmissions. This can be achieved using descriptive statistics such as the mean or standard deviation, as well as predictive statistics like correlation. For data preparation, attention must be given to issues such as missing data, duplicates, and ensuring that the data is presented in the correct format to maximize the model's effectiveness.
5. Modeling and Evaluation: In the modeling phase, we train various algorithms using a training dataset. This should help us determine which model best addresses the Business question and which data is most relevant. In the evaluation phase, we need to assess the best-performing model and identify whether it requires any adjustments.