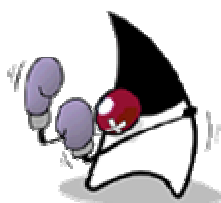


Laboratoire de Technologies de e-commerce (Java, mobiles & R) Partie "Technologies Java et Data mining" : projet " **InpresEnjoyYourHolidays**"- suite

3^{ème} Informatique de gestion
2022-2023



Projet *InPrESEnjoyYourHolidays* - la suite



Claude Vilvens et Christophe Charlet



1. Préambule

L'Unité d'Enseignement "**Programmation réseaux, web et mobiles**" ou "**Programmation réseaux et web**" (10 ECTS - 135h, 9 ECTS - 120h et 7 ECTS - 90h selon les orientations) comporte des activités d'apprentissage :

- ◆ AA: Réseaux et technologies Internet (dans toutes les options);
- ◆ AA: Programmation.Net (dans toutes les options);
- ◆ AA: Technologie de l'e-commerce et mobiles (informatique de gestion seulement);
- ◆ AA: Compléments de programmation réseaux (informatique réseaux-télécoms seulement).

Le contexte de ce laboratoire de "Technologie de l'e-commerce et mobiles" est le même que celui du laboratoire de Réseaux et technologies Internet, à savoir celui de "InpresEnjoyYourHolidays" qui vise à la gestion d'une infrastructure touristique de villages de vacances.

2. Règles d'évaluation

Comme on sait, la note finale pour l'UE considérée se calcule par une moyenne des notes des AA constitutives, sachant que le seul cas de réussite automatique d'une UE est une note de 10/20 minimum dans chacune des AAs.

Pour ce qui concerne l'évaluation de l'AA "Technologie de l'e-commerce et mobiles", voici les règles de cotation utilisées par les enseignants de l'équipe responsable de cette AA.

1) L'évaluation établissant la note de l'AA "Technologie de l'E-commerce et mobiles" est réalisée de la manière suivante :

- ◆ examen de théorie: un examen **oral** en janvier 2023 (sur base d'une liste de points de théorie à développer fournis au fur et à mesure de l'évolution du cours théorique) et coté sur 20;
- ◆ laboratoire en évaluation continue: une évaluation ("évaluation 1" ci-dessous) cotée sur 20 qui constitue la note d'évaluation continue (non remédiable en 2^{ème} session);
- ◆ examen de laboratoire: un examen oral en janvier 2023 consistant en la présentation de la 2^{ème} partie du laboratoire ("évaluation 2" ci-dessous) et coté sur 20;
- ◆ **note finale : moyenne géométrique de la note de l'examen de théorie (poids de 50%), de la note d'évaluation continue (poids de 20%) et de la note de l'examen de laboratoire (poids de 30%).**

Dans ces conditions, *il est clair qu'une note beaucoup trop basse parmi les trois ne peut que conduire à l'échec de l'AA considérée.*

2) Dans le cas où les travaux sont présentés par une équipe de deux étudiants, chacun d'entre eux doit être capable d'expliquer et de justifier l'intégralité du travail sans de longues recherches dans le code de l'application proposée (pas seulement les parties du travail sur lesquelles il aurait plus particulièrement travaillé).

3) Dans tous les cas, tout étudiant doit être capable d'expliquer de manière générale (donc, sans entrer dans les détails) les notions et concepts théoriques qu'il manipule dans ses travaux (par exemple: keystore SSL, régression multiple et tests, etc).

4) En 2^{ème} session, un **report de note** est possible pour **des notes supérieures ou égales à 10/20** en ce qui concerne :

- ◆ la note de théorie;
- ◆ la note de laboratoire de l'évaluation 2 (examen de laboratoire).

Les évaluations de théorie et du laboratoire d'examen ayant des **notes inférieures à 10/20** sont donc **à représenter dans leur intégralité** (le refus de représenter une évaluation complète de laboratoire entraîne automatiquement la cote de 0).

Note finale : **moenne géométrique de la note de l'examen de théorie (poids de 50%) et de la note de l'examen de laboratoire (poids de 50%)**.

La note de laboratoire de l'évaluation 1 n'est pas remédiable et n'intervient plus dans le calcul de la note finale (principe général de l'évaluation continue).

Le laboratoire de "Technologie de l'e-commerce et mobiles" comportera donc deux évaluations. La première (data mining avancé et Big data, Android) sera **évaluée** par l'un des professeurs du laboratoire **à partir du 7 novembre 2022** (avec rentrée d'un dossier papier tel que décrit dans l'énoncé). La deuxième (SSL, messagerie électronique, Big data et exploration des données, ...) sera évaluée lors de l'examen de laboratoire en **janvier 2023** (un dossier papier ne sera plus nécessaire).

3. Agenda des évaluations

Pour chaque évaluation, le délai est à respecter impérativement.

Evaluation	Evaluation continue : semaine d'évaluation	Remédiable en 2^{ème} session
Evaluation 1 :	7/11/2022-11/11/2022	non
Examen de laboratoire:	Date de l'examen de laboratoire de janvier 2023	oui

Remarque importante : Pour rappel, lors de chaque évaluation, chaque étudiant est sensé connaître les bases théoriques qui lui ont permis de réaliser les développements proposés. Dans le cas contraire, on sera amené à considérer qu'il a développé sans comprendre ce qu'il faisait ...

Les travaux de l'évaluation 1 : ACP, messagerie électronique, graphiques statistiques, régression multiple

Compétences développées :

- ◆ Maîtriser les concepts et l'utilisation des techniques d'exploration des données quantitatives (ACP) et de leur modélisation par régression multiple ou ANOVA 2;
- ◆ Maîtriser les concepts et l'utilisation des techniques de graphiques (JFreeChart);
- ◆ Savoir utiliser les commandes des protocoles SMTP et POP3.

Dossier attendu :

1. Principales conclusions des explorations de données traitées par les ACP;
2. Traitement des régressions multiples et ANOVA.

1. Exploration des données: ACP

1.1 Le maïs

Le ministère de l'Agriculture du Barbareland a commandité une étude sur les plants de maïs afin d'optimiser les techniques de culture. Un échantillon de 100 pieds de maïs a été constitué (sur 50000 pieds possibles) et les résultats ont été compilés dans le fichier **etude-agro-mais.csv**. On y trouve comme variables:

Hauteur ; Masse ; Nb.grains ; Masse.grains ; Couleur ; Germination.epi ; Enracinement ; Verse ; Attaque ; Parcelle ; Hauteur.J7 ; Verse.Traitement ; Nb.jours.attaque ; Censure.droite

Certaines variables s'interprètent par leur nom et pour les autres :

- ◆ Masse: masse totale du plant
- ◆ Masse.grains: masse de l'ensemble des grains du plant
- ◆ Germination.epi: le grain est-il germé sur épi ?
- ◆ Verse: le pied est-il penché ou tombé ?
- ◆ Attaque: attaqué par des insectes ?
- ◆ Hauteur.J7: hauteur 7 jours après la récolte
- ◆ Verse.Traitement: verse après traitement ?
- ◆ Nb.jours.attaque: nombre de jours entre la pousse jusqu'à l'attaque
- ◆ Censure.droite: non utilisée dans la suite

Il s'agit d'effectuer une ACP sur ces données afin de tenter d'y voir un peu plus clair.

Il n'aura pas échappé au lecteur attentif que certains individus (donc, certains plants) ne sont pas complètement documentés: autrement dit, il y a des NAs ! Pour les variables quantitatives, on peut remplacer les valeurs manquantes par la moyenne de la colonne sur les valeurs non NA : ceci n'affecte pas l'ACP avec des variables centrées réduites. Pour réaliser cela, il suffit d'utiliser une fonction

```
remplaceNAparMOY<-function(x)
{
  return ( ifelse(is.na(x), mean(x,na.rm = TRUE), x) )
}
```

dans une fonction `apply()` appliquée aux colonnes considérées. La fonction `ifelse()` est documentée comme suit :

ifelse(test_expression, x, y)

Here, test_expression must be a logical vector (or an object that can be coerced to logical). The return value is a vector with the same length as test_expression.

This returned vector has element from x if the corresponding value of test_expression is TRUE or from y if the corresponding value of test_expression is FALSE.
This is to say, the i-th element of result will be x[i] if test_expression[i] is TRUE else it will take the value of y[i].
The vectors x and y are recycled whenever necessary.

1.2 Les postes budgétaires d'un aéroport

Le tableau suivant fournit le bilan, c'est-à-dire la ventilation en % selon les postes, d'un aéroport de 1998 à 2013 :

Année	CAP	INT	SUBS	DLMT	DCT	INFR	CHAFF	CREAN	t°m	pol
1998	17.93	3.96	0.88	7.38	19.86	25.45	5.34	19.21	16.1	DR
1999	16.21	3.93	0.94	9.82	19.11	26.58	5.01	18.40	15.6	DR
2000	19.01	3.56	1.91	9.43	17.87	25.94	5.40	16.88	15.1	CD
2001	18.05	3.33	1.73	9.72	18.83	26.05	5.08	17.21	16.1	CD
2002	16.56	3.10	2.14	9.39	20.36	23.95	6.19	18.31	18.6	CD
2003	13.09	2.64	2.44	8.10	25.05	19.48	11.61	17.59	21.3	CD
2004	13.43	2.42	2.45	10.83	22.07	22.13	11.17	15.49	20.6	CG
2005	9.83	2.46	1.79	11.81	24.10	22.39	11.31	16.30	17.3	CG
2006	9.46	2.33	2.30	11.46	24.45	23.07	11.16	15.77	16.5	CG
2007	10.93	2.95	2.25	10.72	23.16	24.17	9.64	16.20	15.2	G
2008	13.02	3.74	2.21	7.99	23.04	19.53	12.60	17.87	16.6	G
2009	13.43	3.60	2.29	7.09	23.59	17.61	16.67	15.72	14.3	G
2010	13.37	3.35	2.58	6.76	23.94	18.04	15.42	16.54	14.0	CG
2011	11.75	2.74	3.11	7.37	25.04	18.11	14.71	17.18	13.6	CG
2012	12.59	3.05	3.85	7.12	23.40	19.17	11.86	18.97	15.6	CG
2013	13.00	3.00	4.00	7.00	24.00	20.00	12.00	17.00	16.1	DR

Les postes du bilan sont les suivants :

CAP : l'ensemble des capitaux propres de l'entreprise.

INT : Intérêts = l'ensemble des frais financiers supportés par l'entreprise.

SUBS : Subventions accordées par l'Etat et les entités fédérées.

DLMT : Dettes à long et moyen terme.

DCT : Dettes à court terme.

INFR : Immobilisations ; représente l'ensemble des infrastructures de l'entreprise (matériel, terrains).

CHAFF : Chiffre d'affaire, valeurs d'exploitation.

CREAN : Valeurs réalisables et disponibles ; ensemble des créances à court terme de l'entreprise.

Les deux dernières variables sont :

- ♦ la température moyenne annuelle
- ♦ la tendance politique du gouvernement du pays.

Quelles informations, quels modèles peut-on dégager de ces données ?

2. Les graphiques statistiques

Sur base du dataset des plants de maïs évoqué ci-dessus, représenter graphiquement, en utilisant **R** d'une part et la librairie Java **JFreechart** d'autre part,

1. la relation entre Masse et Enracinement;
2. la relation entre Nb.grains et Masse.grains;
3. la répartition de l'orientation de Parcelle pour l'échantillon considéré;
4. la relation entre Couleur et Enracinement.

3. SMTP/POP3 en Java

Il s'agit de développer une petite application **Application_Mail_Java** qui est un mail user agent, c'est-à-dire permettant de traiter le courrier électronique relatif aux informations échangées par les employés de la société. Plus précisément, elle comporte une fonctionnalité GUI **Java** de gestion classique d'e-mail (type "**mini-outlook**"), permettant d'envoyer et de recevoir un mail

- soit simple de type texte;
- soit composite avec des pièces attachées qui sont des **images** gif ou jpg, ou un **digest** de contrôle d'intégrité ou encore un objet sérialisé (instance de la petite classe **PieceAttachee** qui se limite à encapsuler le type d'information avec les éléments de son contenu.

L'utilisateur devrait idéalement être prévenu dans un délai de 5 minutes de l'arrivée d'un nouveau message. A cet effet, l'application utilisera un thread de polling qui "interrogera" périodiquement la boîte aux lettres.

De plus, une fonctionnalité supplémentaire permettra de "tracer" un message reçu, c'est-à-dire de lister les agents mails (les MTAs) par qui ce message est passé.

4. Régression multiple : les accidents sur les routes du Minnesota

Le fichier **accidents2.csv** contient des données sur le taux d'accidents de voiture sur les autoroutes du Minnesota (1973). Ces données, collectées sur 39 grands segments d'autoroute, ont évidemment été collectées pour essayer de déterminer les raisons de ces accidents.

- ◆ rate : taux d'accidents par million de véhicules
- ◆ len : longueur du segment de collecte
- ◆ adt: densité moyenne du trafic (en milliers)
- ◆ trks: pourcentage de camions
- ◆ sigs1: nombre de signaux routiers par mile
- ◆ slim: vitesse maximale autorisée
- ◆ shld: largeur (en pieds) de la bande d'urgence latérale
- ◆ lane: nombre de bandes
- ◆ acpt: nombre d'entrées par mile
- ◆ itg: nombre d'échangeurs par mile
- ◆ lwid: largeur d'une bande (en pieds)
- ◆ htype: type de chaussée, soit MC ("major collector"), FAI ("Federal interstate highways"), PA ("principal arterial highway") or MA ("major arterial highways").
- ◆ rate: nombre d'accidents par million de véhicules en 1973

On demande d'étudier l'éventuelle relation entre ce taux d'accidents et nombre de signaux routiers par mile associé à la largeur de la bande d'urgence latérale. Dans un second temps, on demande d'ajouter d'autres variables explicatives dans le modèle testé, comme par exemple le nombre d'entrées par mile d'autoroute.

5. ANOVA à deux facteurs

5.1 Bière et petits maux

L'Administration de la Santé Publique de Bidendumie a recensé le nombre de patients atteints de l'une des 4 maladies bénignes les plus fréquentes et ayant consommé l'une des 3 bières locales les plus répandues. Elle a mesuré un coefficient biochimique représentatif sur 6 patients (si possible) choisis aléatoirement et a obtenu (chaque ligne correspond à une maladie) :

	bière : La tueuse						bière : La vaisselle						bière : Le nectar royal					
gastro entérite	42	44	36	13	19	22	33	26	33	21	29	31	31	-3	25	25	24	17
mal de tête	28	23	34	42	13	27	34	33	31	36	34	38	3	26	28	32	4	16
nausées	1	29	19	29	18	22	11	9	7	1	-6	5	21	1	9	3	12	17
sinusite	24	9	22	-2	15	12	27	12	12	-5	16	15	22	7	25	5	12	19

Est-il possible d'interpréter de tels résultats ? Ils se trouvent dans le fichier
bieres_petits_maux.csv.

5.2 Le retour du maïs

Sur base du dataset du maïs évoqué ci-dessus, étudier l'éventuelle dépendance de la masse des grains par rapport à la couleur et l'enracinement.

Les travaux de l'évaluation 2 : Android, RServe, ACM, SSL, classifications

Compétences développées :

- ◆ Maîtriser les techniques de base du développement Android.
- ◆ Savoir utiliser les commandes des protocoles SMTP et POP3;
- ◆ Maîtriser la sécurisation d'une communication réseau par TSL-SSL;
- ◆ Maîtriser les concepts et l'utilisation des techniques d'exploration des données ACP, ACM et les classifications.

Dossier attendu :

1. principales conclusions des explorations de données traitées par les ACP et ACM;
2. copie écran de la visualisation des handshakes SSL entre les deux serveurs considérés.

6. L'application mobile Android pour les réservations

Dans le contexte d'Inpres-EnjoyYourHolidays, l'idée est d'élargir le plus possible les parts de marché d'IEYH, et donc de permettre au grand public d'effectuer des réservations pour les motels et les villages au moyen de son smartphone.

Le serveur **Serveur_Reservations** (voir l'énoncé de "Réseaux et technologies Internet") doit donc aussi satisfaire ces requêtes en considérant que l'application **Mobile_Applic_Reservations** tournant sur le smartphone est capable d'utiliser le protocole ROMP (mais sans LROOMS).

Cependant, par mesure de sécurité, cette possibilité n'est offerte qu'aux clients qui sont déjà connus de la société, autrement dit à des voyageurs qui ont déjà séjournés dans un des villages d'IEYH (autrement dit, des voyageurs qui ont effectué une réservation et qui l'ont payée) et qui ont reçu à cette occasion un mot de passe spécial qui leur est propre. Ils devront donc produire ce mot de passe pour pouvoir effectuer une réservation par smartphone. Techniquement parlant, cette application utilise une architecture basée sur les activity et les intents.

Il s'agit d'une version

- ◆ adaptée : les résultats principaux sont sauvegardés dans une "base décisionnelle" SQLite locale au téléphone; on pourra consulter cette "base" à la demande;
- ◆ améliorée : l'utilisateur peut choisir sa langue d'utilisation.

6.2 Le retour des mails

Les applications **Web_Reservations** et **Mobile_Applic_Reservations** envoient un mail de confirmation. L'application Web ajoute une facture en pièce attachée lorsqu'une réservation est terminée.

7. Data mining avec le serveur RServe

7.1 Un problème de Big data

Il s'agit de réaliser ici une application Java qui permet de réaliser quelques tests statistiques classiques (ANOVA à 1 ou 2 facteurs, REGRESSION-CORRELATION simple ou multiple, test du chi-carré) en utilisant le serveur RServe, ceci sur base de l'énoncé proposé ci-dessous. En pratique :

- ◆ on commencera par résoudre les problèmes en **R interactif**;
- ◆ on passera ensuite à la programmation au moyen d'une application Java avec interface graphique et cliente de RServe; les résultats statistiques proprement dits seront donc réalisés

en utilisant RServe tandis que les graphiques statistiques correspondants seront réalisés avec la librairie JFreeChart.

Venons-en au problème considéré. L'IRMSB, Institut de Recherche du Ministère de la Santé de Batracie, petit état au niveau de vie élevé, étudie le comportement de sa population par rapport aux consultations médicales ainsi qu'aux hospitalisations. Il a réuni des données comportant 3874 observations pour 17 variables dans un fichier **sante_batracie_2.csv**. Les variables sont :

1. docvis: nombre de consultations chez le médecin (0-121)
2. hospvis: nombre de jours d'hospitalisation (0-51)
3. edlevel: niveau éducation (1: niveau primaire, 2-niveau secondaire inférieur, 3-niveau secondaire supérieur, 4-niveau supérieur)
4. age: entre 25 et 64 ans
5. outwork: actif (0) ou pas (1)
6. female: femme (1) ou homme (0)
7. married: marié (1) ou célibataire (0)
8. kids: a des enfants (1) ou pas (0)
9. hhninc: rentrées d'argent du ménage par an (en Marks Batraciens !)
10. educ: nombre d'années d'enseignement reçu (7-18)
11. self: indépendant (1) ou salarié (0)
12. edlevel1: pas de diplôme du secondaire (1) ou au moins diplôme du secondaire (0)
13. edlevel2: diplômé du secondaire sans plus (1) ou mieux (0)
14. edlevel3: études supérieures (1) ou pas (0)
15. edlevel4: diplômé du supérieur (1) ou pas (0)

L'institut souhaite obtenir des réponses à diverses questions, comme :

- 1) Le nombre de consultations varie-t-il en fonction du niveau d'études ?
- 2) Le nombre de consultations des femmes varie-t-il en fonction des rentrées d'argent ?
- 3) En Batracie, le fait d'être un homme ou une femme influence-t-il le fait d'être marié ou célibataire ?

Note: Avant d'attaquer les deux questions suivantes, il est recommandé de d'abord résoudre les deux exercices proposés au point 2. "Régression multiple et ANOVA à deux facteurs")

- 4) Le nombre de jours d'hospitalisation dépend-il significativement du nombre de consultations et de l'âge ? Même question en ajoutant le nombre d'années d'enseignement reçu.
- 5) Le nombre de jours d'hospitalisation dépend-il significativement du niveau éducation et/ou du fait d'être marié ou pas ?

7.2 Le serveur Serveur DataMining

Dans le contexte des villages de vacances, IEYH a décidé de se doter d'un serveur **Serveur DataMining** dans le but d'améliorer ses offres et de mieux répondre aux souhaits des voyageurs ("un voyageur satisfait est un voyageur qui revient" - comme disait on ne sait plus qui). Il utilise en back-end un serveur **RServe** pour effectuer toutes les études statistiques.

Ce serveur, multithread Java/Windows-Unix (en modèle pool de threads), attend sur le PORT_DM des requêtes formulées par les analystes d'IEYH. Ceux-ci manipulent une application **Application DataMining**, dont le rôle est de demander le traitement (notamment statistique) des informations disponibles dans BD_HOLIDAYS. Ceci sous-entend que cette base de données contient les informations nécessaires aux questions qui seront posées. De

plus, le serveur écrit aussi dans une autre base BD_DECISIONS qui est la base décisionnelle : elle mémorise les résultats stratégiques importants obtenus lors des requêtes.

Le **protocole applicatif** (basé TCP) est **TUSTADO** (**TU** **STA**tistiqueras sans la **DO**uleur) permet, après un login classique, d'effectuer les études suivantes :

1. P_RES_NAT: relation entre le prix payé par réservation et la nationalité les voyageurs; l'application client génère ensuite (avec JFreeChart), sur base des résultats retournés par le serveur, un graphique "en camembert" et un "diagramme à moustache";
2. P_RES_NB_ACT: relation entre le prix payé par réservation et le nombre d'activités pratiquées; l'application client génère ensuite (avec JFreeChart), sur base des résultats retournés par le serveur, un graphique "nuage de points";
3. P_RES_NB_ACT_NB_ENF: relation du prix payé par réservation avec le nombre d'activités pratiquées et le nombre d'enfants
4. P_RES_NAT_ACC: relation du prix payé par réservation avec la nationalité les voyageurs et le fait d'être seul ou accompagné
5. RES_SEX: relation entre le nombre de réservations et le sexe du voyageur référent.

Pour que ces graphiques ne soient pas trop simplistes, on générera avec R des fichiers de données simulées avec suffisamment de données pour obtenir des graphiques plus réalistes.

8. Exploration des données: ACP et ACM

8.1 Le retour de l'ACP

Le fichier "**HeartDiseasesBorsolavia.csv**" contient des données provenant d'une étude réalisée par le Ministère de la santé de Boursoulavie. Cette étude concerne l'influence de certains paramètres sur les maladies cardiaques. Les variables utilisées sont:

1. Age : Age of the patient
2. Sexe : Sex of the patient
3. DouleurPoitrine : Chest Pain type chest pain type
 - a. Value 1: typical angina
 - b. Value 2: atypical angina
 - c. Value 3: non-anginal pain
 - d. Value 4: asymptomatic
4. TensionAuRepos : resting blood pressure (in mm Hg)
5. TauxCholesterol : cholestoral in mg/dl fetched via BMI sensor
6. TauxSucre : (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
7. ResultatElectroCardigram : resting electrocardiographic results
 - a. Value 0: normal
 - b. Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
 - c. Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
8. MaxPulsations: maximum heart rate achieved
9. RisqueDouleurPoitrine: exercise induced angina (1 = yes; 0 = no)
10. PicExercice: ST depression induced by exercise relative to rest
11. TauxPicExercice: the slope of the peak exercise ST segment
12. Vaisseaux: number of major vessels (0-3) colored by fluoroscopy
13. ThalassémiePresente (maladie génétique des globules rouges): 3 = normal; 6 = fixed defect; 7 = reversable defect
14. RisqueAttaque : 0= less chance of heart attack 1= more chance of heart attack

On demande d'explorer ces données et de voir ensuite si il est possible d'en extraire un (ou des) modèle linéaire de données expliquant certaines variables en fonction d'autres.

8.2 Une enquête sur les chiens

Le Ministère de la Protection animale de Boursoulavie a réalisé une enquête sur les caractéristiques de divers chiens auprès d'éleveurs spécialisés dans l'élevage d'une seule race de chien. Le résultat de cette enquête se trouve dans le fichier **chiens.csv**.

Quelles conclusions tirer de cette enquête ?

Remarque : La boucle suivante pourrait vous être utile :

```
for (i in colnames(x))  
{  
  x[,i]<-factor(x[,i])  
}
```

8.3 Une enquête sur les OGMs

Les étudiants en informatique de l'université de ClaudeLeGrand de Christocharvie ont réalisé une enquête sur les OGMs en interrogeant 135 personnes sur leur campus. Les données se trouvent dans le fichier **gmo.csv**, avec des noms de variables anglicisés à des fins d'utilisation internationale. Ces variables sont :

- ◆ **Implicated**: Vous sentez-vous concernés par la problématique des OGMs ?
- ◆ **Position.Culture**: Quelle est votre position par rapport à la culture des OGMs en France ?
- ◆ **Position.Al.H**: Quelle est votre position quant à l'incorporation d'OGMs dans l'alimentation humaine ?
- ◆ **Position.Al.A**: Idem pour l'alimentation animale ?
- ◆ **Protest**: Avez-vous participé à une manifestation anti-OGMs ?
- ◆ **Media.Passive**: Pensez-vous que les médias communiquent assez sur le sujet ?
- ◆ **Info.Active**: Effectuez-vous la démarche de vous informer sur les OGMs ?
- ◆ **Phytosanitary.products**: Pensez-vous que l'usage des OGMs puisse réduire l'utilisation des fongicides ?
- ◆ **Hunger**: Pensez-vous que l'usage des OGMs puisse réduire la faim dans le monde ?
- ◆ **Animal.feed**: Pensez-vous que l'usage des OGMs puisse faciliter la vie des agriculteurs ?
- ◆ **Future.Progress**: Pensez-vous que l'usage des OGMs puisse permettre de futurs progrès scientifiques ?
- ◆ **Danger**: Pensez-vous que les OGMs représentent un danger pour notre santé ?
- ◆ **Threat**: Pensez-vous que les OGMs représentent un danger pour l'environnement ?
- ◆ **Finan.risk**: Pensez-vous que les OGMs représentent un risque pour les agriculteurs ?
- ◆ **Useless.practice**: Pensez-vous que les OGMs représentent un procédé scientifique inutile ?
- ◆ **Grandparents**: Pensez-vous que nos grands-parents avaient une alimentation plus saine ?
- ◆ **Sex, Age, Profession, Political.Party**: signification évidente ...
- ◆ **Relation**: Exercez-vous un métier ou suivez-vous des études en rapport avec l'agriculture, l'agro-alimentaire ou l'industrie pharmaceutique ?

Quels résultats peut-on tirer d'une telle enquête ?

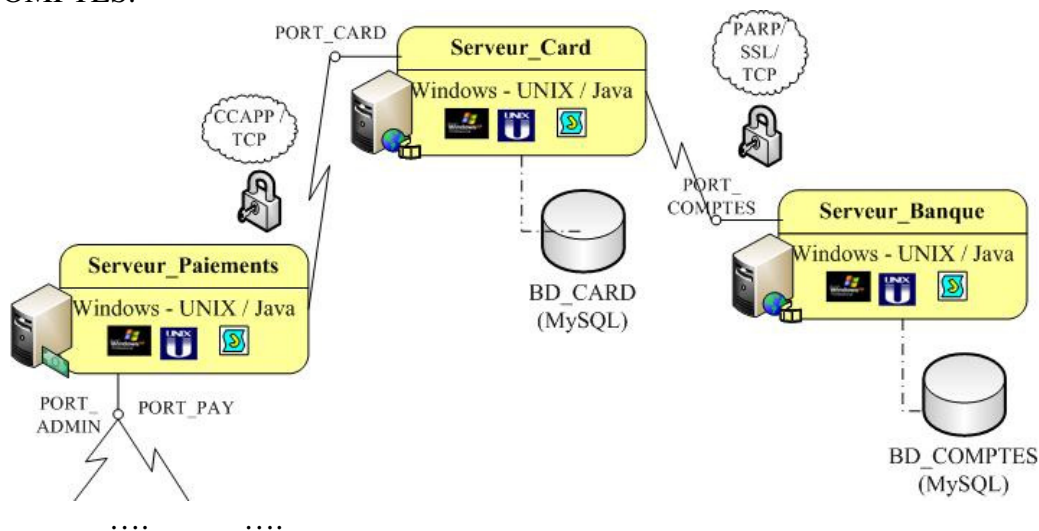
9. SSL

On se souviendra (peut-être) que, dans la structure de IEHY, il existe un serveur **Serveur_Card**, utilisé par **Serveur_Paiements** et répondant au seul type de requête d'une vérification/débit (protocole **CCAPP** [Check **C**ArD and **P**ayment **P**rotocol]). Ce

Serveur_Card utilise sa propre base de données BD_CARD, qui contient essentiellement les comptes et les numéros de cartes (un compte peut avoir plusieurs cartes de crédit associées).

Dans la vraie vie, le compte est en réalité hébergé sur serveur propre à chaque banque : c'est en fait ce serveur qui règle les débits/crédits. Dans ce cas, Serveur_Card ne dispose plus que des numéros de cartes et contacte le serveur de la banque concernée pour finaliser (ou refuser) le paiement des réservations. Cette communication, réglée par un protocole **PARP** [Payment Agreement or Refusal Protocol] qu'il vous appartient de définir et d'implémenter, **est intégralement chiffrée et authentifiée en utilisant le sous-protocole SSL**.

Pour faciliter les choses en fonction du temps de développement disponible, nous considérerons qu'il n'existe qu'un seul **Serveur_Banque** qui utilise sa propre base de données BD_COMPTEs.



10. Les classifications

10.1 Ces quelques fleurs ...

Un classique : explorer le dataset "**Iris**" intégré à R et chargé par

```
>data(iris).
```

Il concerne 150 fleurs relevant, à priori, de trois espèces différentes (à vérifier). Quelles sont les variables les plus importantes pour séparer les trois espèces ? Ces trois espèces semblent-elles valides ? Etc

10.2 Le retour des températures

Etudier la question du regroupement hiérarchisé en classes de villes européennes par rapport à leurs températures mensuelles. Les données sont (fichier **temperatures_europennes.csv** fourni !):

	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre	Moyenne	Amplitude	Latitude	Longitude	Région
Amsterdam	2.9	2.5	5.7	8.2	12.5	14.8	17.1	17.1	14.5	11.4	7	4.4	9.9	14.6	52.2	4.5	Ouest
Athènes	9.1	9.7	11.7	15.4	20.1	24.5	27.4	27.2	23.8	19.2	14.6	11	17.8	18.3	37.6	23.5	Sud
Berlin	-0.2	0.1	4.4	8.2	13.8	16	18.3	18	14.4	10	4.2	1.2	9.1	18.5	52.3	13.2	Ouest
Bruxelles	3.3	3.3	6.7	8.9	12.8	15.6	17.8	17.8	15	11.1	6.7	4.4	10.3	14.4	50.5	4.2	Ouest
Budapest	-1.1	0.8	5.5	11.6	17	20.2	22	21.3	16.9	11.3	5.1	0.7	10.9	23.1	47.3	19	Est
Copenhague	-0.4	-0.4	1.3	5.8	11.1	15.4	17.1	16.6	13.3	8.8	4.1	1.3	7.8	17.5	55.4	12.3	Nord
Dublin	4.8	5	5.9	7.8	10.4	13.3	15	14.6	12.7	9.7	6.7	5.4	9.3	10.2	53.2	6.1	Nord
Helsinki	-5.8	-6.2	-2.7	3.1	10.2	14	17.2	14.9	9.7	5.2	0.1	-2.3	4.8	23.4	60.1	25	Nord
Kiev	-5.9	-5	-0.3	7.4	14.3	17.8	19.4	18.5	13.7	7.5	1.2	-3.6	7.1	25.3	50.3	30.3	Est
Cracovie	-3.7	-2	1.9	7.9	13.2	16.9	18.4	17.6	13.7	8.6	2.6	-1.7	7.7	22.1	50	19.6	Est
Lisbonne	10.5	11.3	12.8	14.5	16.7	19.4	21.5	21.9	20.4	17.4	13.7	11.1	15.9	11.4	38.4	9.1	Sud
Londres	3.4	4.2	5.5	8.3	11.9	15.1	16.9	16.5	14	10.2	6.3	4.4	9.7	13.5	51.4	0	Nord
Madrid	5	6.6	9.4	12.2	16	20.8	24.7	24.3	19.8	13.9	8.7	5.4	13.9	19.7	40.2	3.4	Sud
Minsk	-6.9	-6.2	-1.9	5.4	12.4	15.9	17.4	16.3	11.6	5.8	0.1	-4.2	5.5	24.3	53.5	27.3	Est
Moscou	-9.3	-7.6	-2	6	13	16.6	18.3	16.7	11.2	5.1	-1.1	-6	5.1	27.6	46.2	1.5	Est
Oslo	-4.3	-3.8	-0.6	4.4	10.3	14.9	16.9	15.4	11.1	5.7	0.5	-2.9	5.6	21.2	59.5	10.5	Nord
Paris	3.7	3.7	7.3	9.7	13.7	16.5	19	18.7	16.1	12.5	7.3	5.2	11.2	15.3	48.5	2.2	Ouest
Prague	-1.3	0.2	3.6	8.8	14.3	17.6	19.3	18.7	14.9	9.4	3.8	0.3	9.2	20.6	50	14.2	Est
Reykjavik	-0.3	0.1	0.8	2.9	6.5	9.3	11.1	10.6	7.9	4.5	1.7	0.2	4.6	11.4	64.1	21.6	Nord
Rome	7.1	8.2	10.5	13.7	17.8	21.7	24.4	24.1	20.9	16.5	11.7	8.3	15.4	17.3	41.5	12.3	Sud
Sarajevo	-1.4	0.8	4.9	9.3	13.8	17	18.9	18.7	15.2	10.5	5.1	0.8	9.4	20.3	43.5	18.3	Sud
Sofia	-1.7	0.2	4.3	9.7	14.3	17.7	20	19.5	15.8	10.7	5	0.6	9.6	21.7	42.4	23.2	Est
Stockholm	-3.5	-3.5	-1.3	3.5	9.2	14.6	17.2	16	11.7	6.5	1.7	-1.6	5.8	20.7	59.2	18	Nord
Anvers	3.1	2.9	6.2	8.9	12.9	15.5	17.9	17.6	14.7	11.5	6.8	4.7	10.3	15	51.1	4.2	Ouest
Barcelone	9.1	10.3	11.8	14.1	17.4	21.2	24.2	24.1	21.7	17.5	13.1	10	16.2	15.1	41.2	2.2	Sud
Bordeaux	5.6	6.7	9	11.9	15	18.3	20.4	20	17.6	13.5	8.5	6.1	12.7	14.8	44.5	0.3	Ouest
Edimbourg	2.9	3.6	4.7	7.1	9.9	13	14.7	14.3	12.1	8.7	5.3	3.7	8.3	11.8	55	3	Nord
Francfort	0.2	1.8	5.4	9.7	14.3	17.5	19	18.3	14.8	9.8	4.9	1.7	9.8	18.8	50.1	8.4	Ouest
Genève	0.1	1.9	5.1	9.4	13.8	17.3	19.4	18.5	15	9.8	4.9	1.4	9.7	19.3	46.1	6.1	Ouest
Gênes	8.7	8.7	11.4	13.8	17.5	21	24.5	24.6	21.8	17.8	12.2	10	16.1	15.9	44.3	9.4	Sud
Milan	1.1	3.6	8	12.6	17.3	21.3	23.8	22.8	18.9	13.1	6.9	2.6	12.6	22.7	45.3	9.2	Sud
Palerme	10.5	11.5	13.3	16.9	20.9	23.8	24.5	22.3	22.3	18.4	14.9	12	16.6	14	38.1	13.1	Sud
Séville	10.7	11.8	14.1	16.1	19.7	23.4	26.7	26.7	24.3	19.4	14.5	11.2	18.2	16	37.2	5.6	Sud
St. Pétersbourg	-8.2	-7.9	-3.7	3.2	10	15.4	18.4	16.9	11.5	5.2	-0.4	-5.3	4.5	26.6	59.6	30.2	Est

Zurich	-0.7	0.7	4.3	8.5	12.9	16.2	18	17.2	14.1	8.9	3.9	0.3	8.7	18.7	47.2	8.3	Ouest
--------	------	-----	-----	-----	------	------	----	------	------	-----	-----	-----	-----	------	------	-----	-------

Bon travail !

s: CV & C.C.

