

L'objectif de ce tutoriel est la prise en main de SAS Viya sur de petits cas simples avec SAS Viya for Learners.

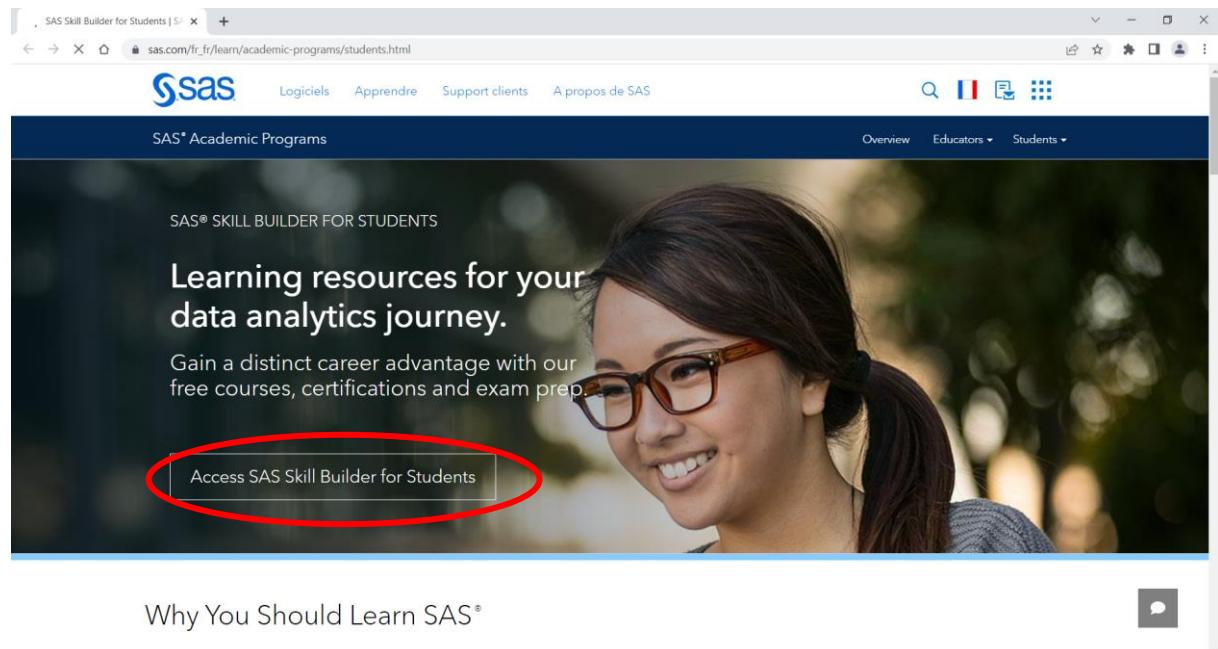
Nous commencerons par utiliser le cas Insight Toy pour découvrir SAS Visual Analytics. Nous utiliserons ensuite Organics pour conceptualiser le Machine Learning.

Table des matières

Bienvenue dans le programme SAS® Skill Builder :.....	2
SAS Visual Analytics	6
Présentation du cas Insight Toy.....	6
Création de rapport avec SAS Visual Analytics.....	13
Création de hiérarchie et de tableau-croisé	22
Création de hiérarchie géographique	26
Création de tableau de bord	32
Enregistrer votre travail.....	37
Exercices.....	39
Machine Learning avec SAS Visual Analytics.....	40
Le cas Organics.....	40
Exploration	44
Arbre.....	48
Régression	51
Forêt.....	53
Gradient Boosting.....	54
Réseau neuronal	55
Comparaison de modèle.....	56
Scoring avec la régression.....	63
Exercices.....	69
Qualité des données	71
Machine Learning en Pipeline	79
Changer la langue du navigateur.....	85
Changer la langue du navigateur Firefox	85
Changer la langue du navigateur Chrome	87
Vidéos complémentaires :.....	89
Préparer de la certification de programmation de base.....	89
Bibliographie	106
Site Web.....	106

Bienvenue dans le programme SAS® Skill Builder :

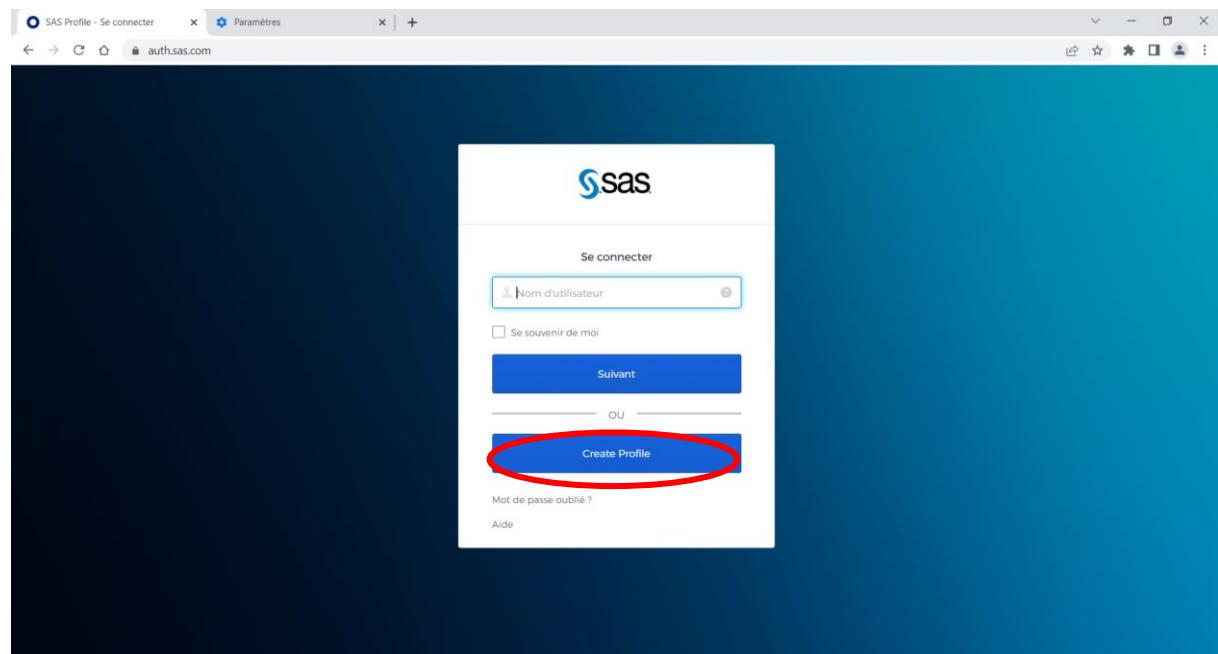
Aller à la page https://www.sas.com/fr_fr/learn/academic-programs/students.html



Why You Should Learn SAS®



Cliquer sur "Access SAS Skill Builder for students"



Si vous n'avez pas de profile SAS avec un e-mail d'une institution académique, une université, une grande école, cliquez sur « Create Profile » ?

Confirmation | SAS

SAS Profile

https://www.sas.com/profile/ui/#/create?fromURI=https%3F%2Fauth.sas.com%2Fapp%2Fsa...

Rechercher

Profil SAS

Etape 1 sur 2 : vos informations

Choisissez votre langue

Prénom *

Nom *

Email *

Pays/Région *

Relation avec SAS *

Niveau d'expertise SAS *

Raison sociale *

Entrer vos coordonnées, votre langue, vos prénom et nom, votre adresse e-mail d'une institution académique, votre pays, votre relation avec SAS, ...

Confirmation | SAS

SAS Profile

https://www.sas.com/profile/ui/#/create?fromURI=https%3F%2Fauth.sas.com%2Fapp%2Fsa...

Rechercher

Email *

Pays/Région *

Relation avec SAS *

Niveau d'expertise SAS *

Raison sociale *

*Requis

Oui, je souhaite recevoir occasionnellement des informations par email à propos des logiciels et solutions SAS : événements, livres blancs, offres spéciales, formations, publications. Je comprends que l'ensemble de mes données personnelles sera traité conformément aux dispositions de la politique SAS de confidentialité des données personnelles. [En savoir plus.](#)

J'accepte les conditions générales *

Créer votre profil

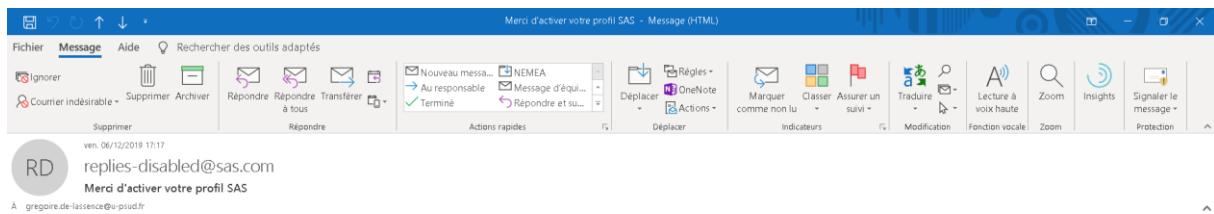
Après avoir cliqué sur "Créer votre profil", vous recevrez un email de vérification contenant les instructions pour créer votre mot de passe et activer votre profil.

Votre niveau en SAS est simplement à titre d'information,

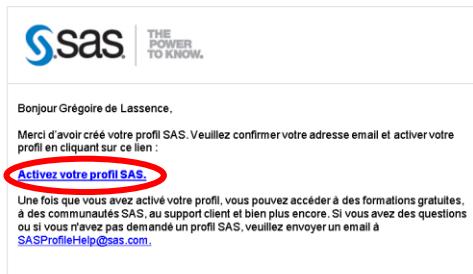
La raison sociale est le nom de votre établissement d'enseignement

Accepter les conditions et cliquer sur créer un profil.

Vous recevrez un e-mail comme celui-ci :

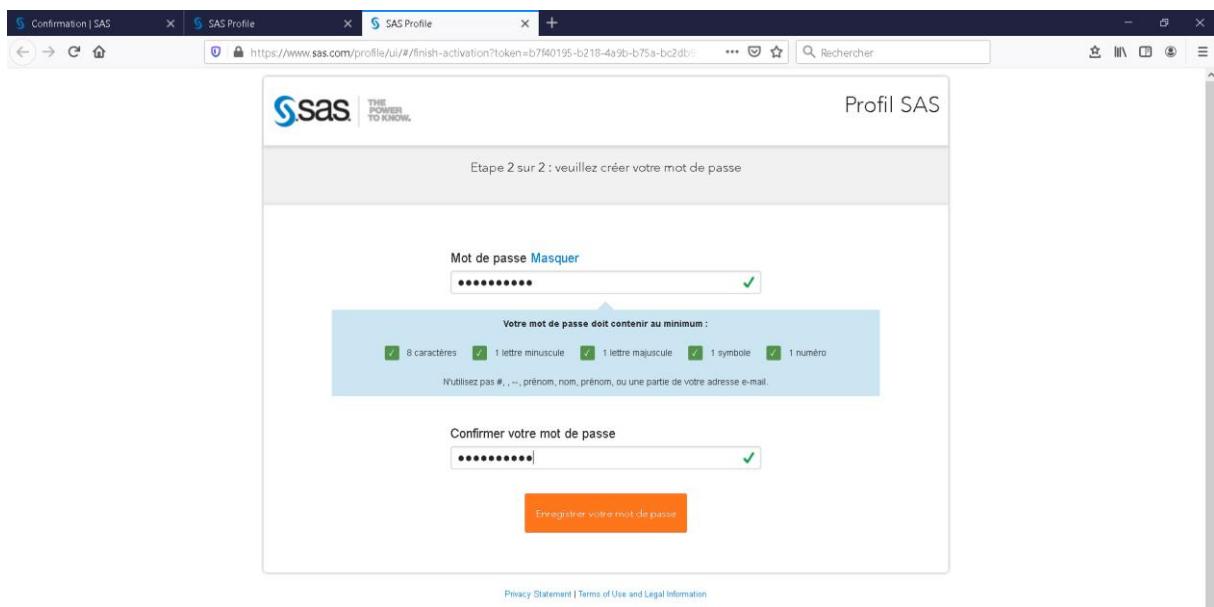


EXTERNAL



SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. © SAS Institute Inc. All rights reserved.

Cliquer sur le lien pour activer le profil.



Créer un mot de passe et l'enregistrer.

Si besoin, aller à la page https://www.sas.com/fr_fr/learn/academic-programs/students.html

SAS Training in -- My Training

We're here to help. Alors que nous faisons face au COVID-19 ensemble, notre engagement envers vous reste fort. Si vous souhaitez développer des compétences essentielles à votre carrière, nous vous invitons à suivre nos formations en ligne gratuites ou à vous inscrire à des cours en classe connectée. Suivez le cours avec un formateur SAS et des machines virtuelles pour pratiquer en temps réel, comme dans une salle de formation.

Formation Sélectionnez votre formation Rechercher | My Training Panier +33 1 60 62 11 00 Contactez-nous

Training

- Accueil
- Trouvez votre formation
- Certification
- Offres de formation
- Discounts
- Centre de formation
- Conditions générales
- Tutorials gratuits

SAS Books

My Training

Activate Your Product(s)

Listed below you will find the product(s) you can activate. Once you accept the license agreement and click submit, you will be able to access your products at any time using the "My Training" link as long as you are logged in using your SAS profile. If you want to use a different profile, save the activation code and come back to this page when you are ready to activate.

SAS Skill Builder for Students

License Agreement

Important: Merci de lire attentivement les termes et conditions du contrat de licence ("Agreement"). En cochant la case "J'accepte", vous vous engagez à respecter les conditions.

Yes, I accept the License Agreement

I must accept the License Agreement in order to complete your activation of any training items.

Submit

Accepter les conditions.

Course: SAS Skill Builder for Studi

Welcome Grégoire to SAS Skill Builder for Students!

THE CURIOSITY CUP: A GLOBAL SAS STUDENT COMPETITION

Looking for a way to inspire your students? The Curiosity Cup is the perfect group project!

Learn More

Learn SAS

Get SAS Certified

Career Resources

Help me get started

I want to build analytics skills

SAS Visual Analytics

Présentation du cas Insight Toy

Nous allons commencer avec une table modeste d'environ 1,6 millions de lignes et 57 colonnes (2 Go) sur la société fictive « Insight Toy Company », implantée dans différentes entités (Facility = centre de production = centre d'assemblage = point de vente) dans le monde. Cette compagnie produit et commercialise des jouets high-techs. Nous avons des données sur 8 ans, de 1998 à fin 2005, sur les aspects financiers, productions, distributions, ventes et marketing.

Liste des principales variables	
Variable	Description
Customer	Il y a plus de 80 000 clients uniques
CustomerLat	Latitude de l'adresse du client
CustomerLon	Longitude de l'adresse du client
CustomerSatisfaction	Satisfaction du client en pourcentage
Facility	Facility = entité = usine et point de vente
FacilityAge	Age de l'usine
FacilityCity	Ville de l'usine
FacilityCityLat	Latitude de la ville de l'usine
FacilityCityLon	Longitude de la ville de l'usine
FacilityContinent	Continent de l'usine
FacilityContinentLat	Latitude du continent
FacilityContinentLon	Longitude du continent
FacilityCountry	Pays de l'usine
FacilityCountryLat	Latitude du pays
FacilityCountryLon	Longitude du pays
FacilityEfficiency	Efficacité de l'usine : une évaluation de l'efficacité des opérations de l'usine, en fonction de multiples facteurs de gestion. Les scores varient de 30% à 100%
FacilityLat	Latitude de l'usine
FacilityLon	Longitude de l'usine
FacilityOpeningDate	Date d'ouverture : utilisée pour calculer l'âge de l'usine
FacilityRegion	Région de l'usine
FacilityRegionLat	Latitude de la région
FacilityRegionLon	Longitude de la région
Order	Numéro d'identification de la commande pour une vente
Product	Numéro d'identification du produit. Un produit appartient à un style de produit (voir ci-dessous)

Liste des principales variables	
Variable	Description
ProductBrand	2 marques de produits : "Novelty" and "Toy"
ProductCostOfSale	Coût de vente du produit. Le coût d'achat et de production du produit vendu
ProductLine	8 lignes de produits. Une ligne appartient à une marque de produit (voir ci-dessus)
ProductMake	71 sous-marques. Une sous-marque appartient à une seule ligne de produits (voir ci-dessus).
ProductSale	Les revenus réels de la vente de ce produit = le chiffre d'affaires
ProductMaterialCost	La composante matière première du coût de vente
ProductPriceActual	Le chiffre d'affaires réel de la vente de ce produit
ProductPriceTarget	Le prix standard idéal du produit. Si le produit était considéré comme à 100% de qualité et que les conditions du marché étaient idéales, le prix de vente du produit serait celui-ci.
ProductQuality	Une évaluation de la qualité de fabrication du produit. Les scores varient de 60 à 100%
ProductStyle	335 styles de produits. Un style appartient à une sous-marque de produit (voir ci-dessus)
SalesRep	ID du vendeur qui a effectué la vente
SalesRepCustomerBase	Montant des revenus potentiels pour tous les clients possibles dans la région d'un vendeur donné
SalesRepCustomers	Nombre de clients auxquels un vendeur est responsable à une date donnée
SalesRepID	ID du vendeur qui a fait cette vente
SalesRepRating	Evaluation interne de la performance d'un vendeur
TransactionDate	Date de la transaction
TransactionMonth	Mois de la transaction
TransactionWeekday	Jour de la semaine de la transaction
TransactionYear	Année de la transaction
Unit	Unité de fabrication
UnitActual	Nombre de produits réalisés par une unité de fabrication donnée à ce moment-là
UnitAge	Age de l'unité de fabrication
UnitCapacity	Capacité de l'unité de fabrication
UnitDiscardRate	Pourcentage représentant le nombre de produits rejetés pour des raisons de qualité, divisé par l'objectif de l'unité
UnitReliability	Un ratio représentant la fiabilité d'une unité de fabrication. Il est composé de produits rejetés pour des raisons de qualité, divisés par la quantité totale de produits assemblés par l'unité, pour une période de temps donnée

Liste des principales variables	
Variable	Description
UnitTarget	Les produits qui devraient être fabriqués à partir d'une unité de fabrication donnée à ce moment-là
UnitYieldRate	Un ratio des produits fabriqué ('Unit Actual') par rapport à l'objectif ('Unit Target')

Aller à la page https://www.sas.com/fr_fr/learn/academic-programs/students.html

SAS Skill Builder for Students | SAS

sas.com/fr_fr/learn/academic-programs/students.html

SAS® Academic Programs

Overview Educators Students

SAS® SKILL BUILDER FOR STUDENTS

Learning resources for your data analytics journey.

Gain a distinct career advantage with our free courses, certifications and exam prep.

Access SAS Skill Builder for Students

Why You Should Learn SAS®

Cliquer sur “Access SAS Skill Builder for students”

SAS Profile - Se connecter

auth.sas.com

Se connecter

b00198424@essec.edu

Se souvenir de moi

Suivant

OU

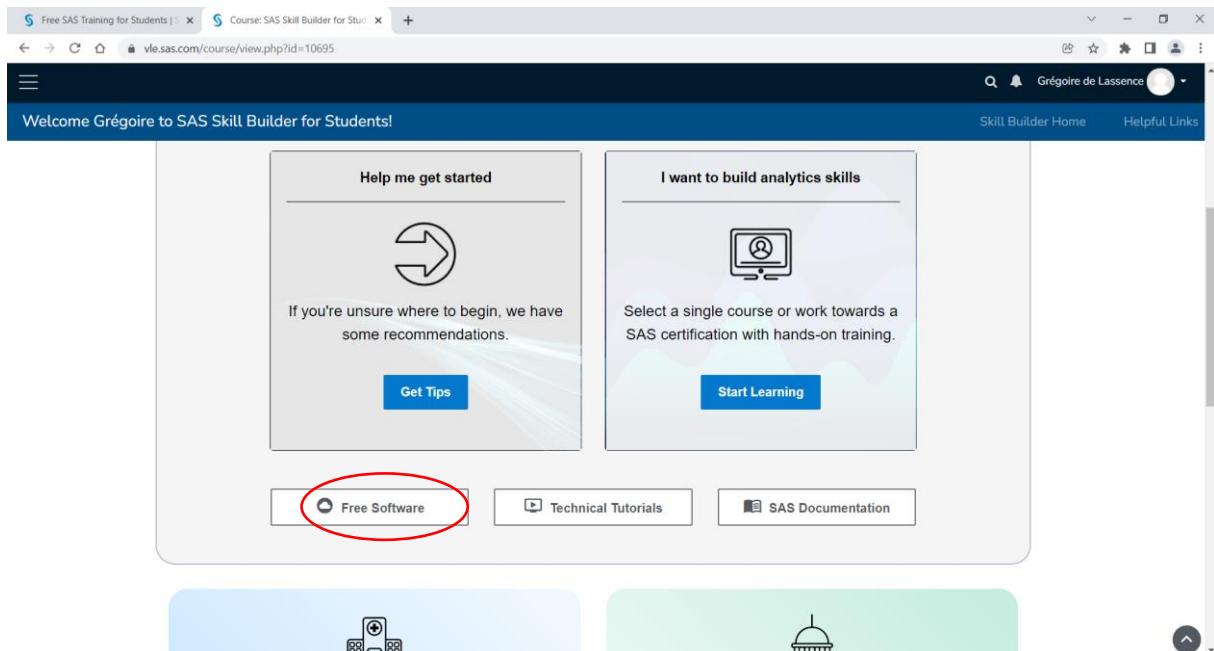
Create Profile

Mot de passe oublié?

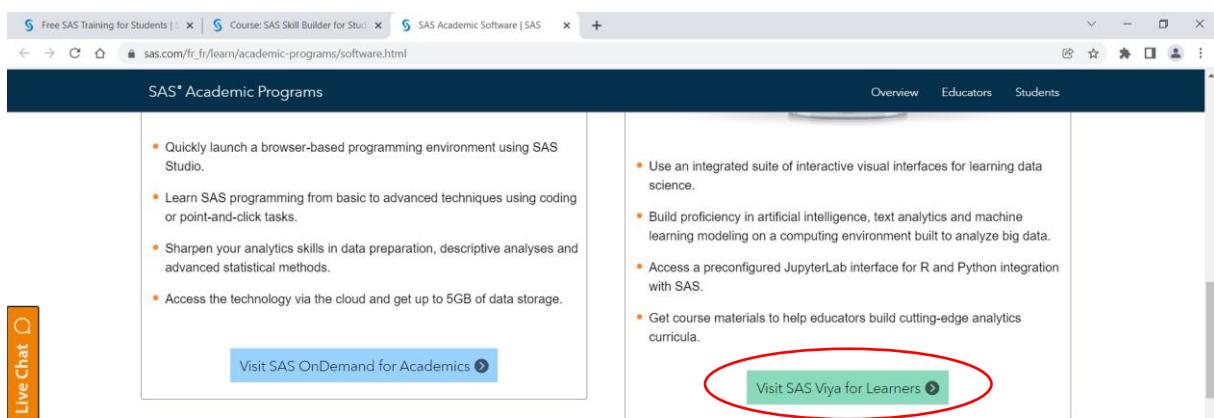
Aide

https://auth.sas.com/#

Suivant, Se connecter



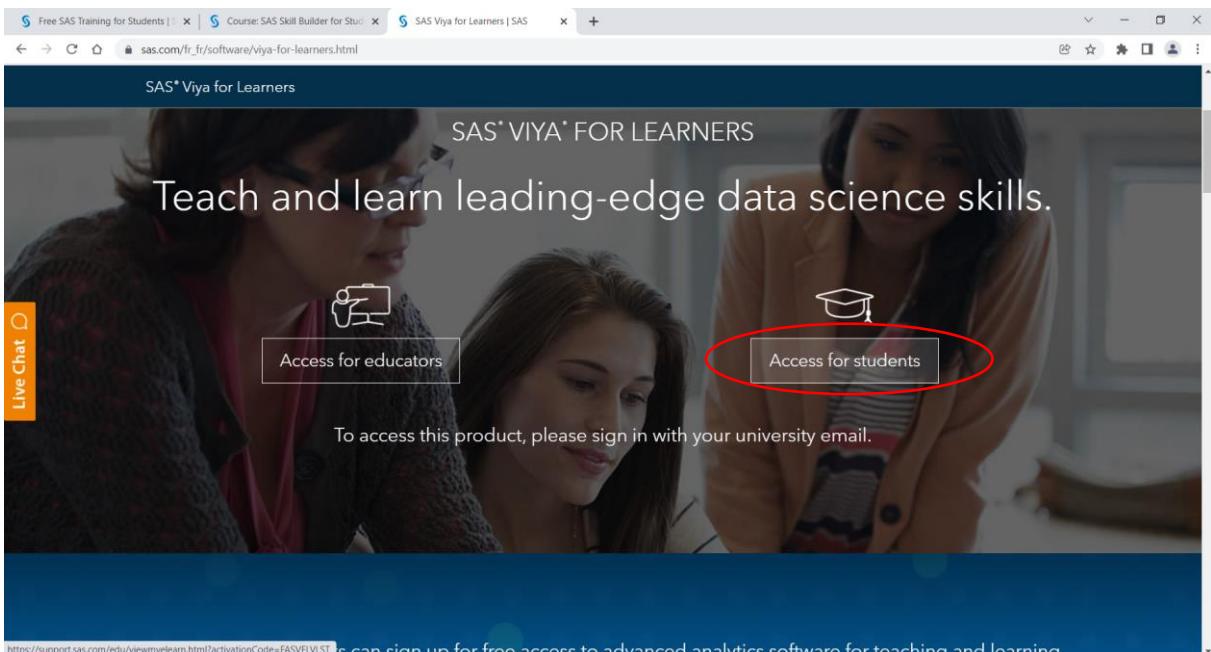
Cliquer sur “Free Software”



SAS keeps me current in a field that changes every day.
And it gives my students a leg up when they go into the
workforce because they can make decisions from a data



Cliquer sur “Visit SAS Viya for Learners”



Cliquer sur “Access for Students”

SAS Viya for Learners

Dashboard / Courses / SAS Viya for Learners

Your progress

Launch SAS Viya for Learners using the button below. Bookmark this page and return here each time you launch SAS Viya for Learners.

SAS Viya for Learners 3.5

Launch

More Info User Data Upload Info

SAS Viya for Learners Data Repository

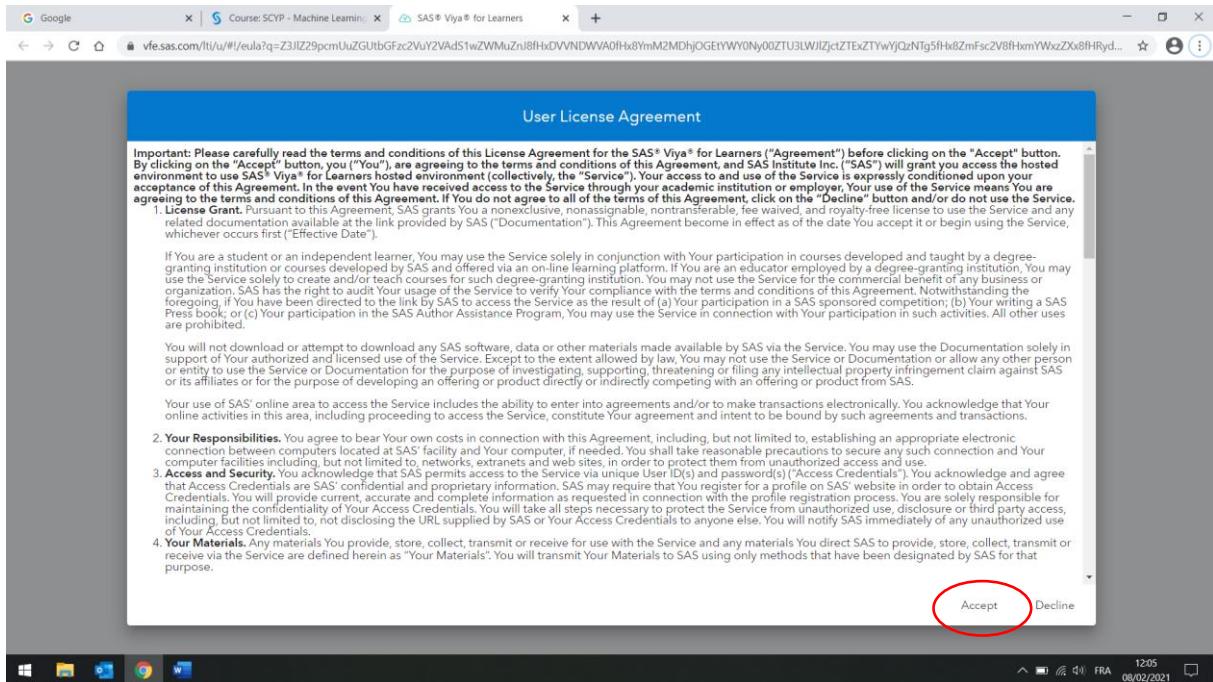
Access Now

Important Information

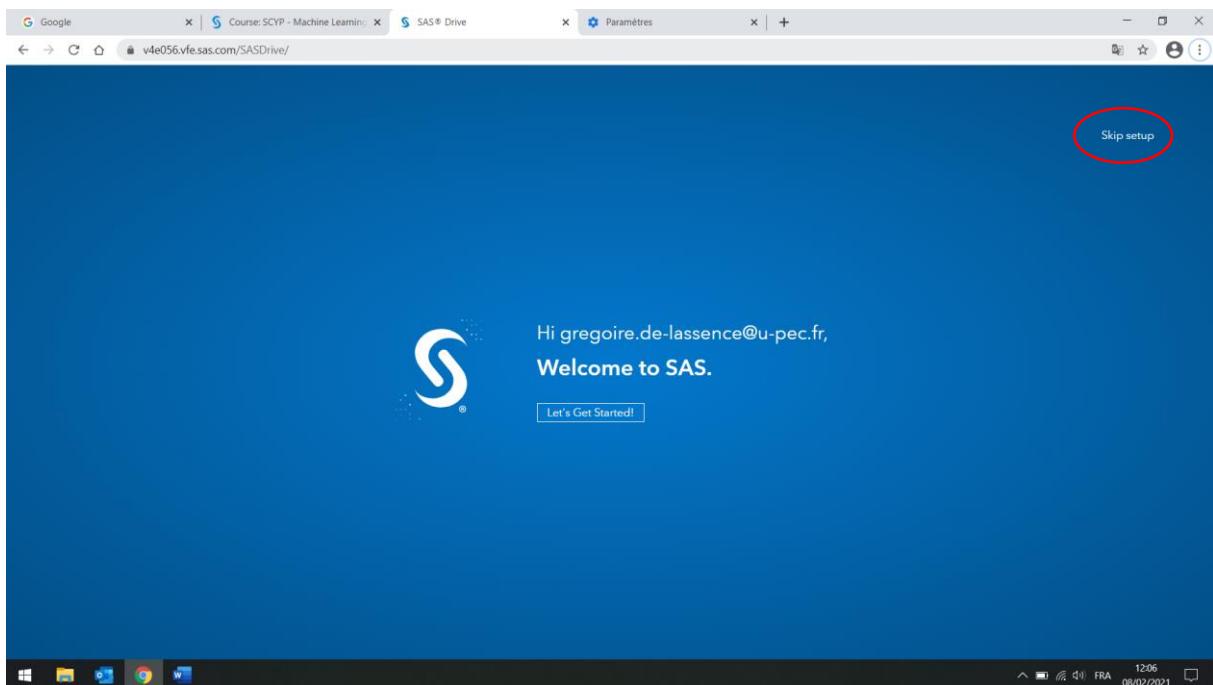
- Autotuning is disabled in SAS Viya For Learners. The software will return an error when a user requests autotuning.

https://vle.sas.com/mod/ri/view.php?id=176320

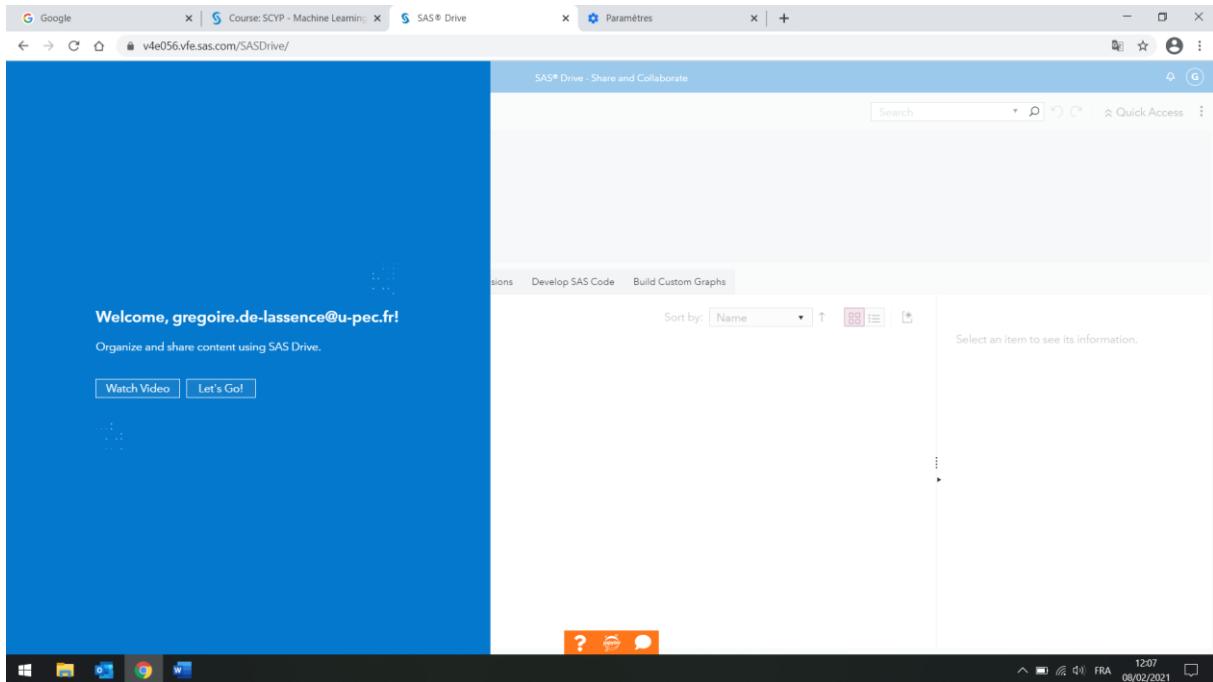
Cliquer sur “Launch SAS Viya for Learners 3.5”



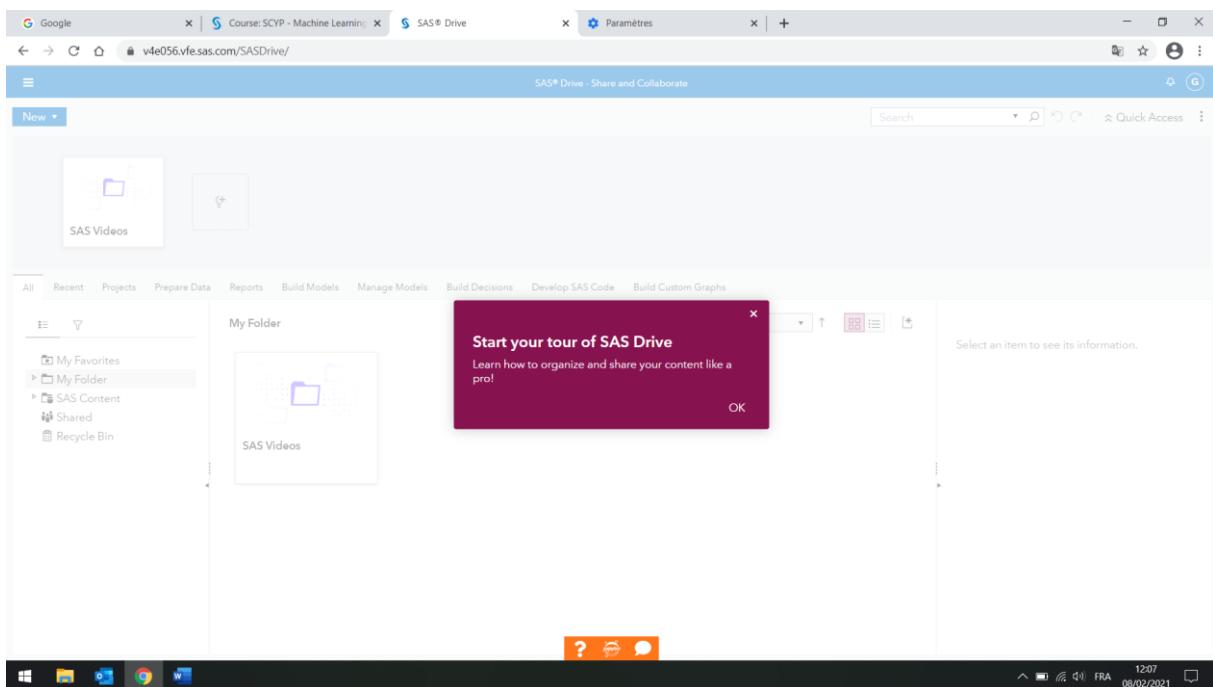
Cliquer sur "Accept"



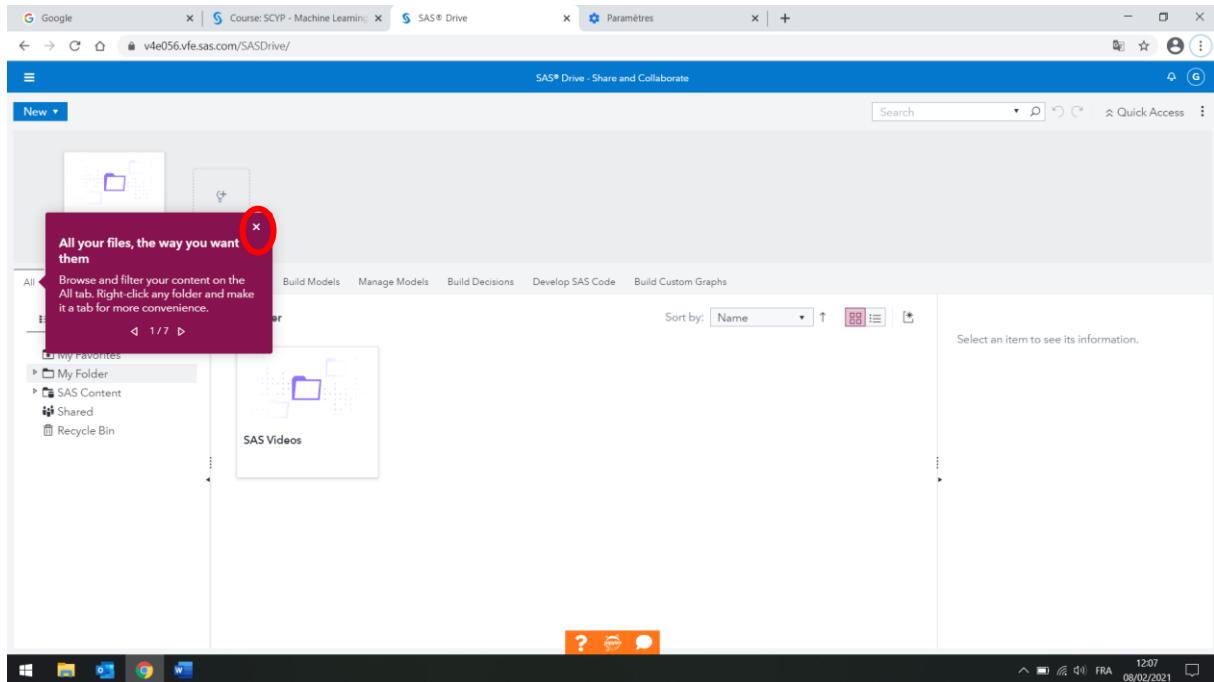
Vous pouvez ignorer la configuration (Skip setup)



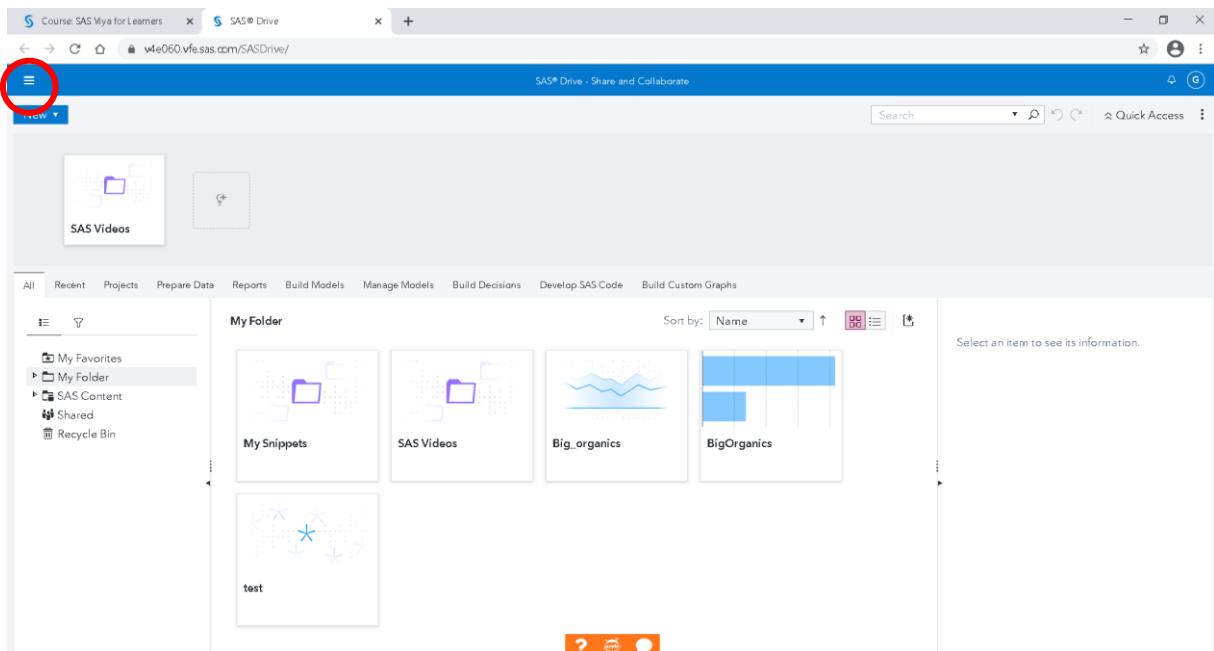
Let's go!



OK



Fermer



Création de rapport avec SAS Visual Analytics

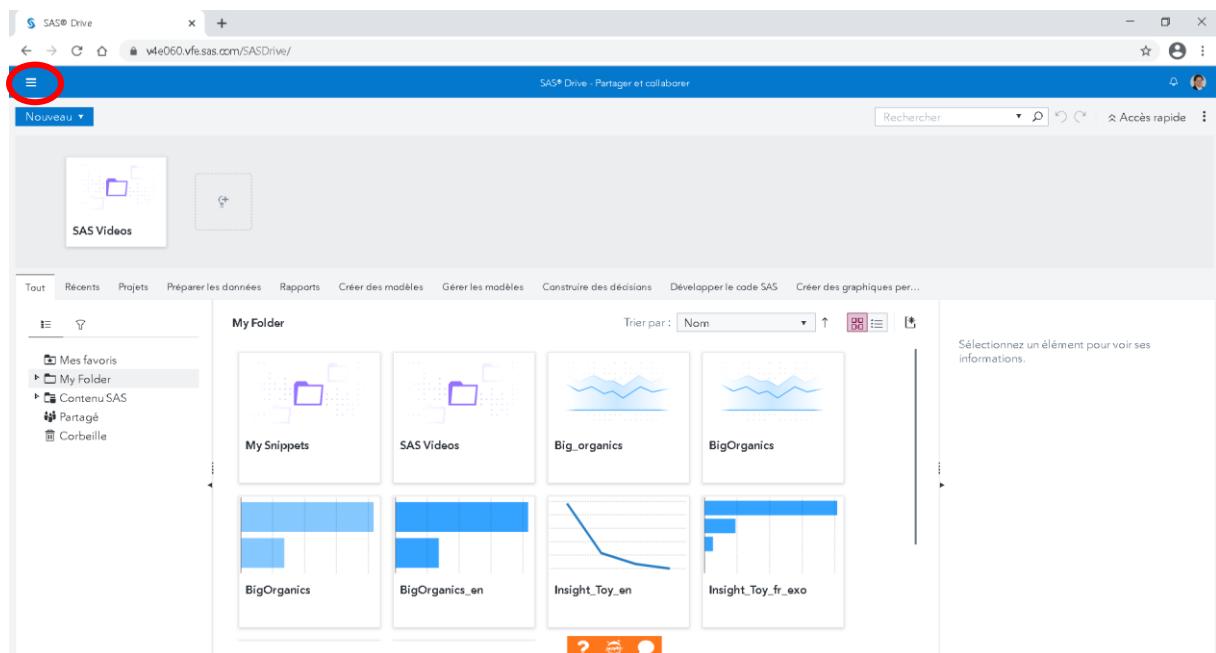
SAS VA (Visual Analytics) est une interface Web sur SAS Viya, basée sur des serveurs analytiques dans le nuage : CAS (Cloud Analytics Servers). SAS VA permet de visualiser très rapidement et simplement des tables gigantesques afin de découvrir simplement quelques patterns¹.

SAS Visual Analytics en version 8 ou supérieur est sur SAS Viya alors que les versions précédentes, jusqu'à 7, étaient sur SAS 9.4.

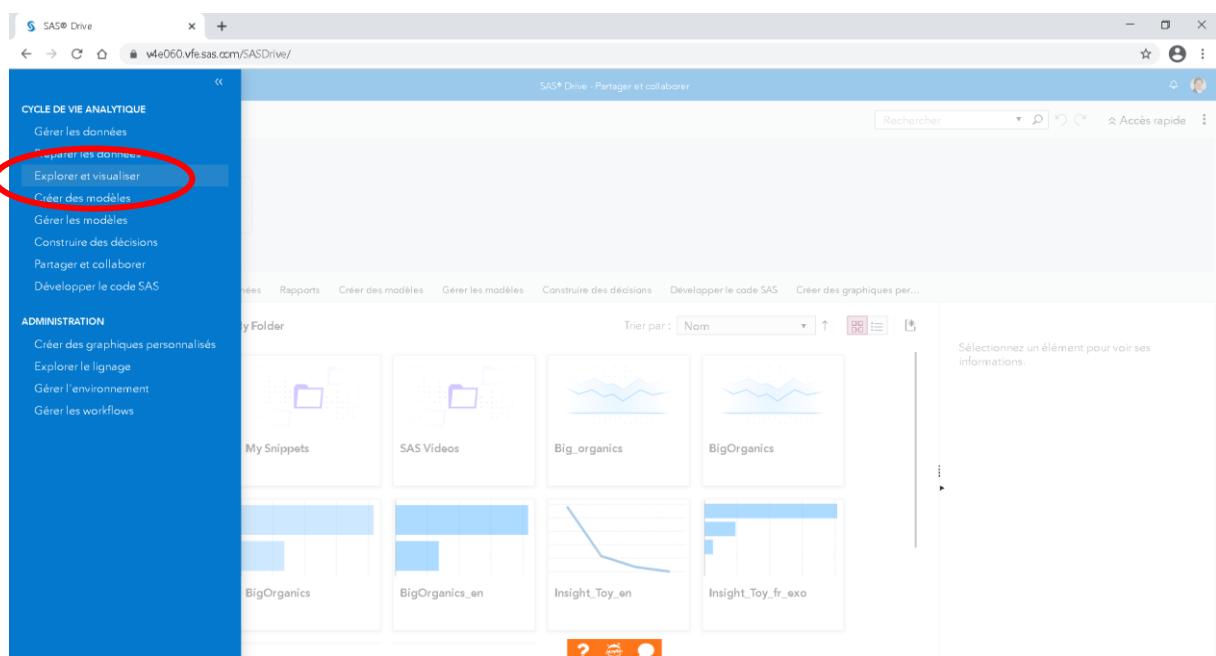
¹ Pattern : mot anglais utilisé pour désigner une structure, une forme, une règle, un motif ou bien encore un type, notamment dans les bases de données.

SAS VA est basé sur des serveurs CAS qui permettent de monter les données en mémoire vive, la RAM, de façon distribuée, afin d'offrir une fluidité maximale pour visualiser et analyser de grandes bases.

L'interface VA permet de masquer la complexité algorithmique afin de permettre, même à des personnes ayant peu de compétences informatiques, de visualiser simplement les données. Cela doit permettre une démocratisation de la culture analytique au sein des organisations. Cela n'enlève en rien le besoin de Data Scientists mais bien au contraire montre qu'ils sont incontournables pour résoudre des problèmes plus complexes.



Si vous n'arrivez pas dans SAS Visual Analytics, mais par exemple sur le Drive, dans le menu en haut à gauche, il faut sélectionner **Explorer et Visualiser**.



Dans le cadre de ce tutoriel, l'interface est en français.

Par défaut, ce sont les paramètres régionaux du navigateur qui sont pris en compte.

Une solution simple pour changer de langue est d'aller dans les paramètres du navigateur pour sélectionner son choix et relancer l'application en se reconnectant. C'est expliqué dans le détail à la fin de ce document.

Pour les TP de ce document, les données sont déjà chargées dans CAS (Cloud Analytics Server).

The screenshot shows the SAS Visual Analytics interface. At the top, there's a header bar with the title "Rapport 1". Below it is a toolbar with various icons. A red circle highlights the "Modification..." button. The main area is a blank canvas with a placeholder message: "Faites glisser les éléments de données ou les objets ici." Below the canvas is a row of five data visualization cards:

- 2018 Company Revenue: A bar chart showing revenue in millions of dollars: \$24M, \$528M, \$924M, \$17M.
- Measure by Time Axis: A line chart with time on the x-axis and a measure on the y-axis.
- A scatter plot with a circular legend.
- A bar chart with two series: 1.7K and 5.4K.
- Regional Revenue: A horizontal bar chart.

Cliquer sur Données

The screenshot shows the "Choisir des données" (Select Data) dialog box. On the left, there's a list of available data sources under "Disponible". A red circle highlights the table "INSIGHT_TOY_DEMO". On the right, there's a detailed view of the selected table "INSIGHT_TOY_DEMO". The table has 57 columns and 1,6 Md rows. The "OK" button at the bottom right is also highlighted with a red circle.

Sélectionner la table Insight_Toy_Demo (la table dont le nom est exactement Insight_Toy_Demo, pas celles « company » ou « sales »)

OK

The screenshot shows the SAS Visual Analytics interface. On the left, the 'Données' (Data) panel is open, displaying a list of data sources and categories. A red box labeled [1] highlights the 'Données' icon, and another red box labeled [2] highlights the pin icon used to keep the panel open. The main workspace on the right shows a placeholder message 'Faites glisser les éléments de données ou les objets ici.' (Drag data elements or objects here.) and a section titled 'Démarrer à partir d'un modèle de page' (Start from a page template) with several visualization preview cards.

Voici l'interface de création de rapport.

A gauche se trouvent les éléments de données, les objets et la gestion du contenu du rapport. Si l'on clique sur les l'icône des données [1], puis sur la punaise [2], ce menu reste figé.

A droite, se trouvent les paramètres de l'élément sélectionné.

Si l'on ne dépose pas d'objet, mais simplement un élément de donnée, le graphique le plus adapté par défaut sera créé.

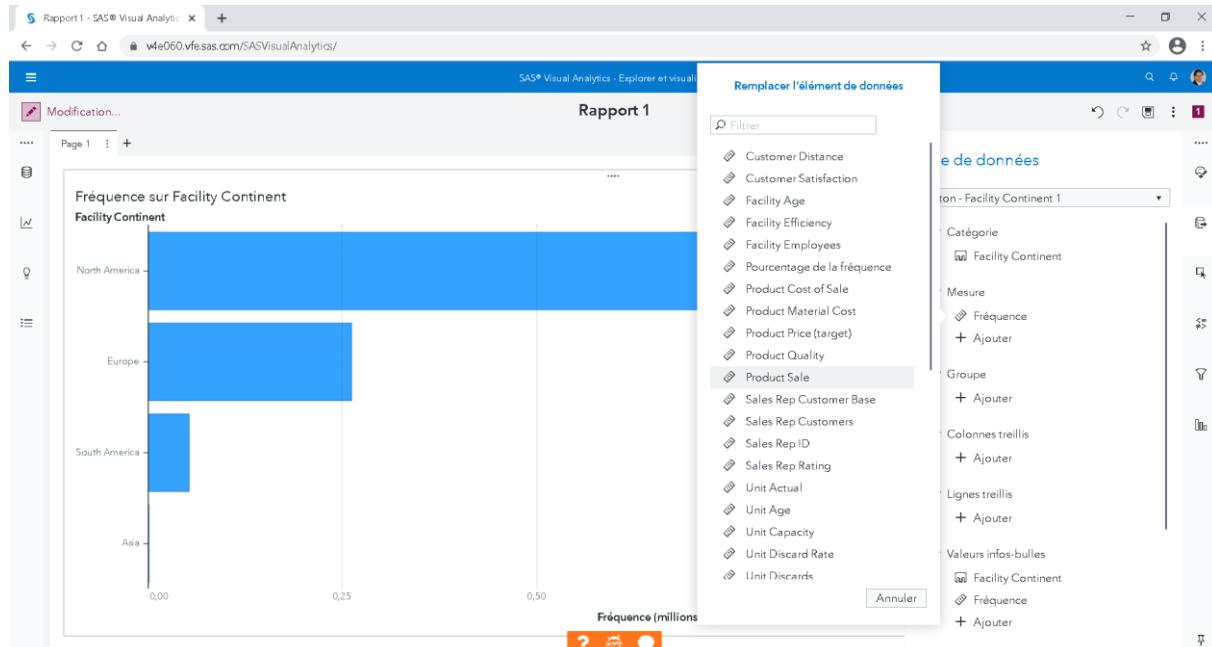
This screenshot is similar to the previous one but includes a red arrow pointing to the 'Facility Continent' item in the 'Données' (Data) panel. A tooltip box is overlaid on the panel, showing details about the selected item: 'Nom : FacilityContinent', 'Valeurs distinctes : 4', 'Nom dans les données : FacilityContinent', and 'Format : \$'. The main workspace remains the same, showing the drag-and-drop placeholder and the page template section.

Cliquer sur Facility Continent, glisser au centre de l'interface et lâcher ; ou bien, double-cliquer sur Facility Continent

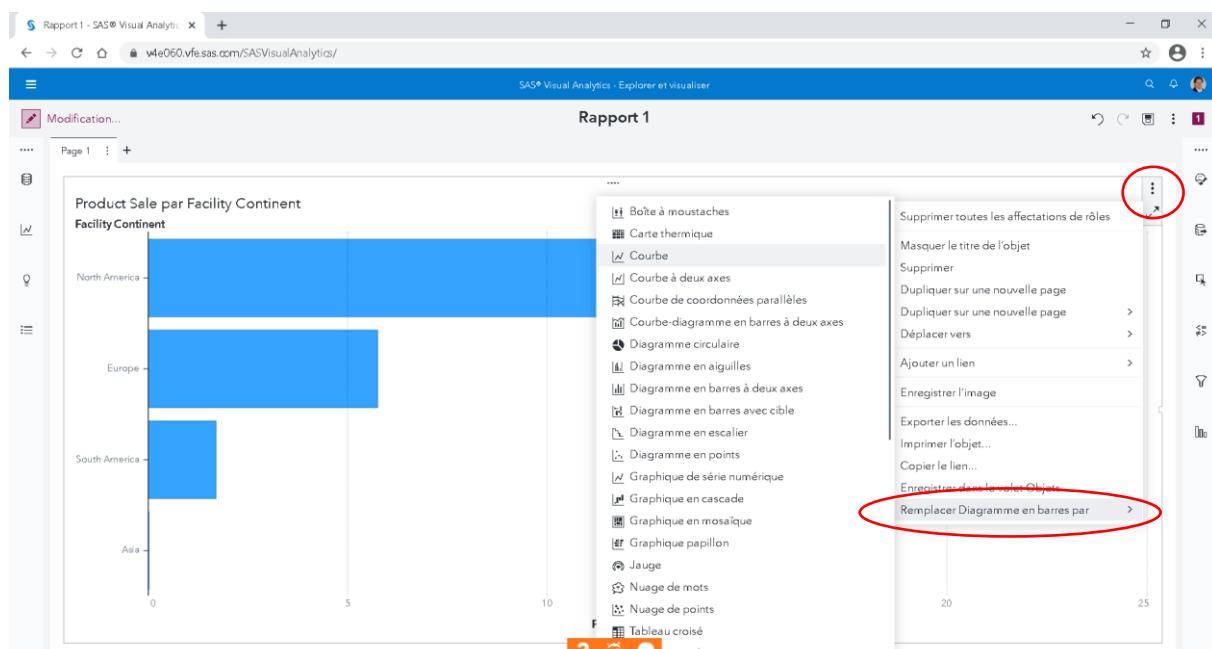
Si aucun objet n'est défini, SAS Visual Analytics utilisera un graphique par défaut.

S'il n'y a pas de mesure sélectionnée, par défaut, la fréquence² sera utilisée ; c'est-à-dire dans notre cas, le nombre de transactions par continent.

Donc, avec simplement le continent qui est une catégorie, on obtient un diagramme en barres.

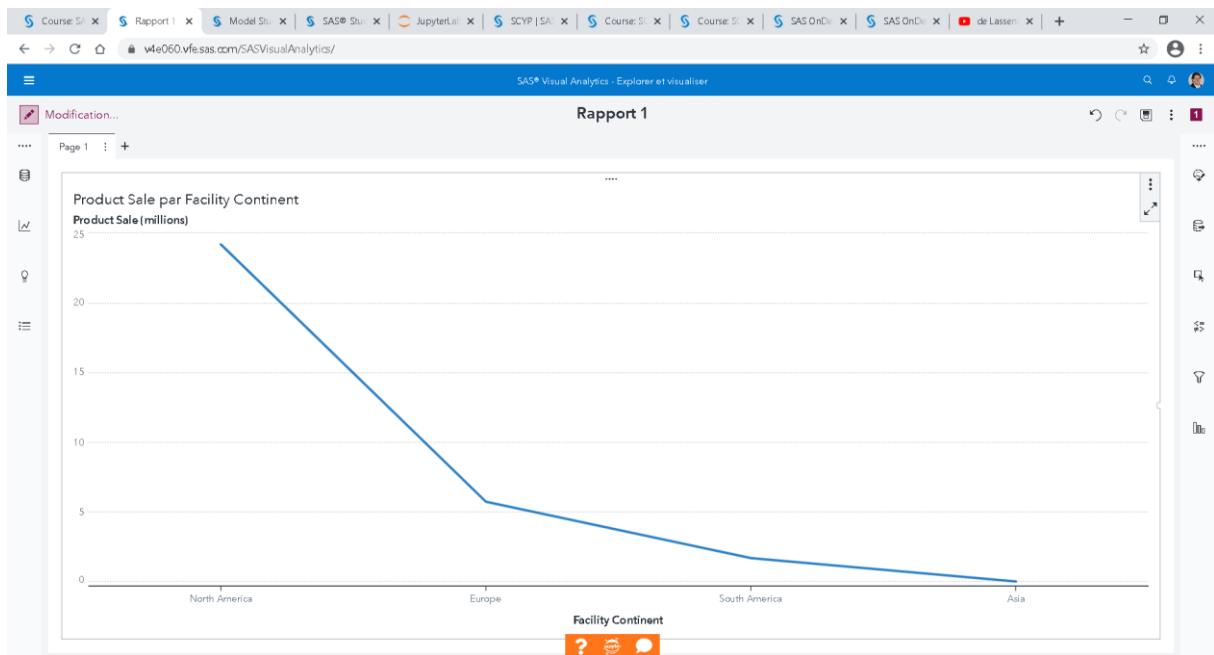


Si l'on double-clique sur « Product Sale », la fréquence sera remplacée par le chiffre d'affaires.

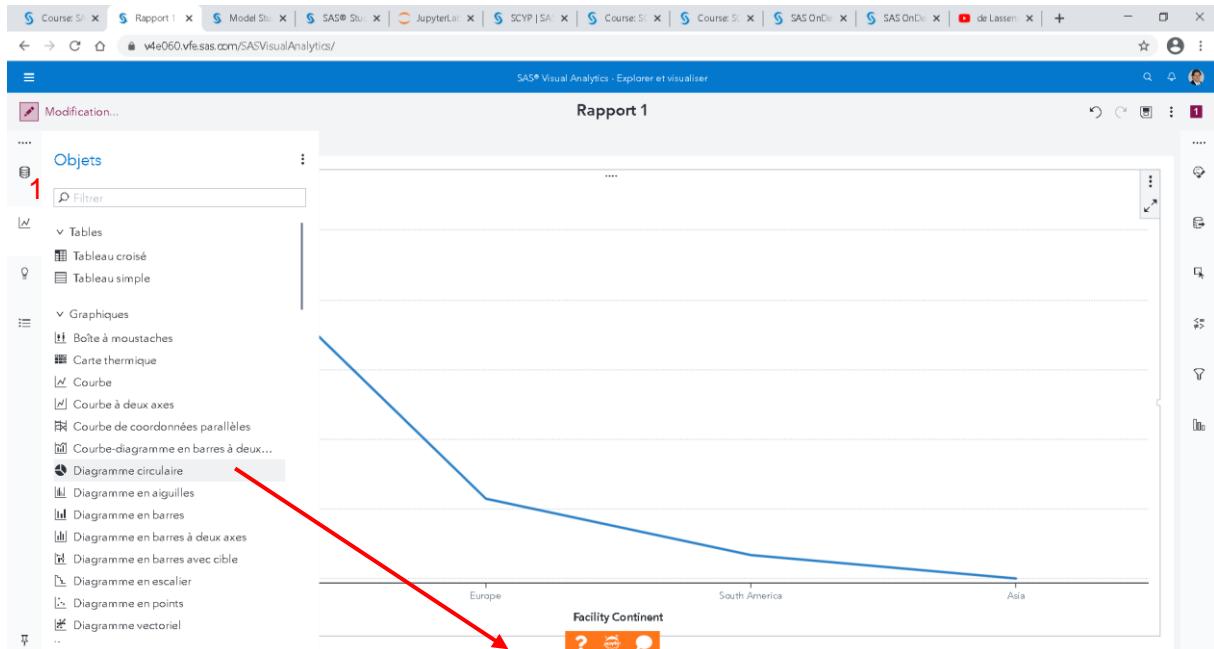


En cliquant sur les trois petits points en haut à droite du graphique, on peut changer de représentation et sélectionner la courbe.

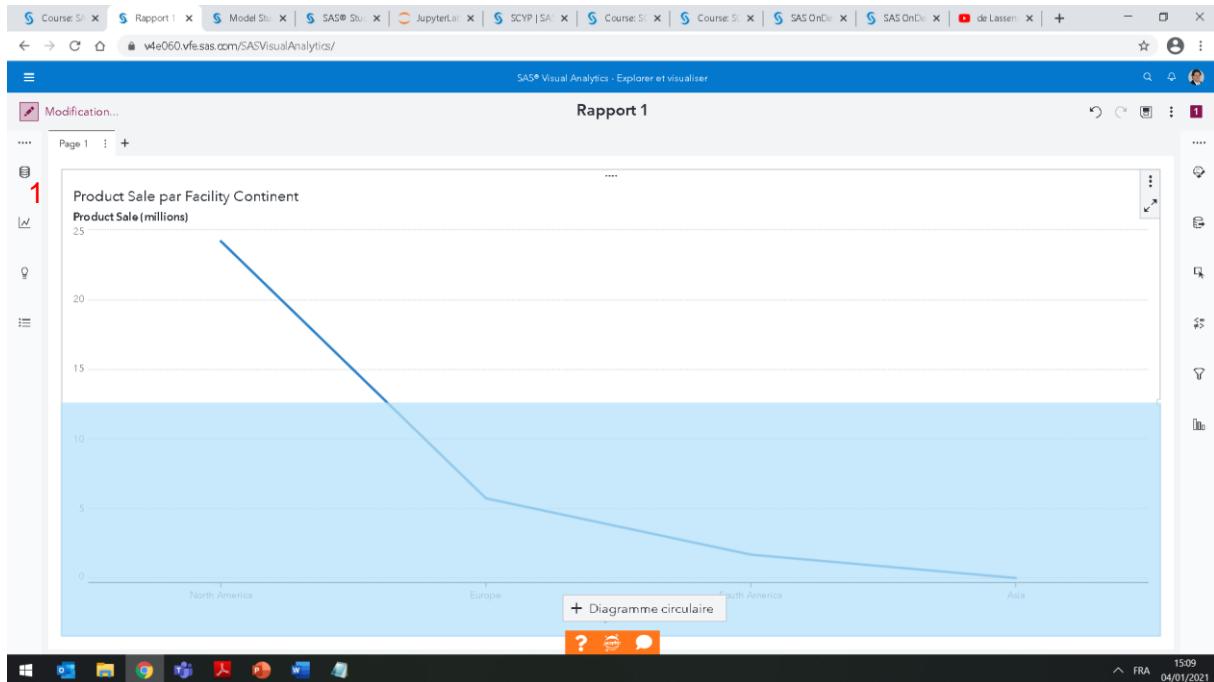
² « fréquence » est utilisé dans son acceptation anglaise : dénombrement.



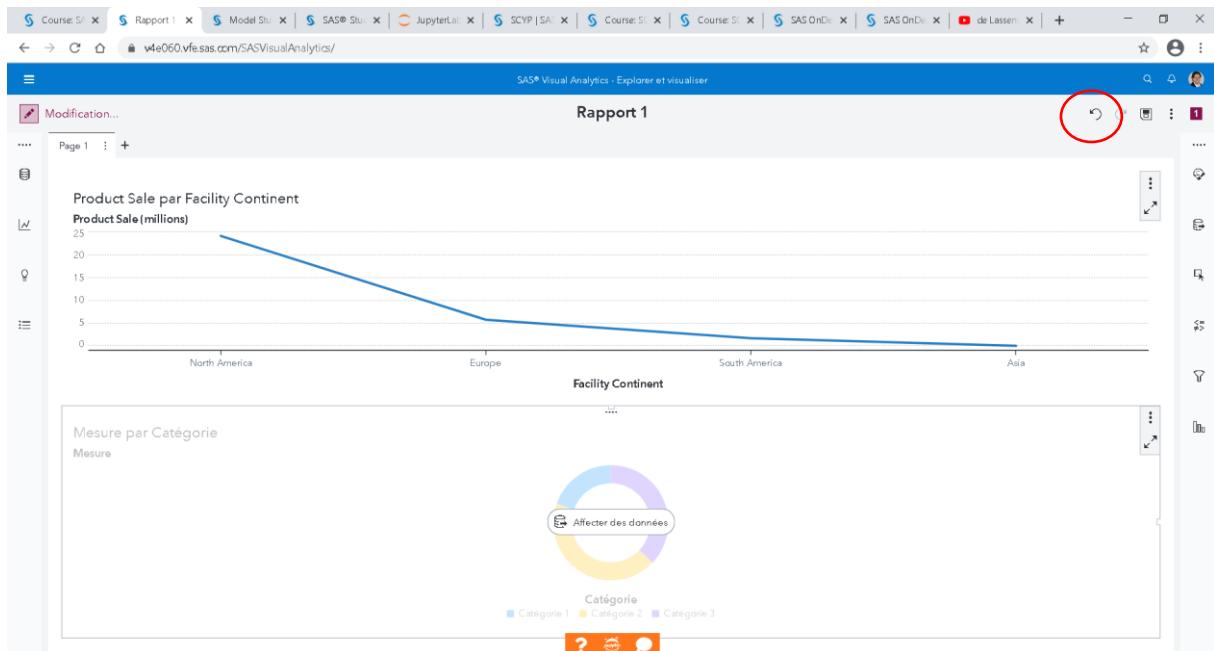
Cela représente les mêmes données, mais de façon différente.



Si l'on veut un diagramme circulaire en-dessous de la courbe, il suffit de cliquer sur le bouton des objets [1], puis de sélectionner cet élément « diagramme circulaire » et le glisser-lâcher.



Donc on clique sur diagramme circulaire, dans la partie gauche, on ne lâche pas le clic, on glisse dans la partie centrale, une barre bleu clair apparaît en dessous du premier graphique lorsque la souris est en bas. A ce moment, on lâche. En dessous de votre e-mail, en haut à droite, il y a la possibilité de revenir en arrière.



Dans la partie droite, sur le rôle des Données, on sélectionne le « continent » comme catégorie et on remplace la fréquence par « Product Sale » comme mesure. On obtient ce graphique qui permet de mieux visualiser les proportions.

Rapport 1

Product Sale par Facility Continent

Product Sale(millions)

Facility Continent	Product Sale (millions)
North America	25
Europe	5

Fréquence sur Facility Continent

Fréquence

1,6 M

Facility Continent

North America Europe Autre

Facility Continent

Product Sale

Customer Distance
Customer Satisfaction
Facility Age
Facility Efficiency
Facility Employees
Pourcentage de la fréquence
Product Cost of Sale
Product Material Cost
Product Price(target)
Product Quality
Product Sale
Sales Rep Customer Base
Sales Rep Customers
Sales Rep ID
Sales Rep Rating
Unit Actual
Unit Age
Unit Capacity
Unit Discard Rate
Unit Discards

Annuler

Rapport 1

Product Sale par Facility Continent

Product Sale(millions)

Facility Continent	Product Sale (millions)
North America	25
Europe	5
South America	1

Product Sale par Facility Continent

Product Sale

32 M

Facility Continent

North America Europe South America Asia

Options

Diagramme circulaire - Facility Continent 1

Direction : En sens inverse des aiguilles d'heure

Section "Autre" : Pourcentage minimum pour la section "Autre" : 4

Style de groupement : En pile

Libellé du diagramme circulaire

LIBELLÉS DE DONNÉES

Libellés de catégorie

Valeurs réelles

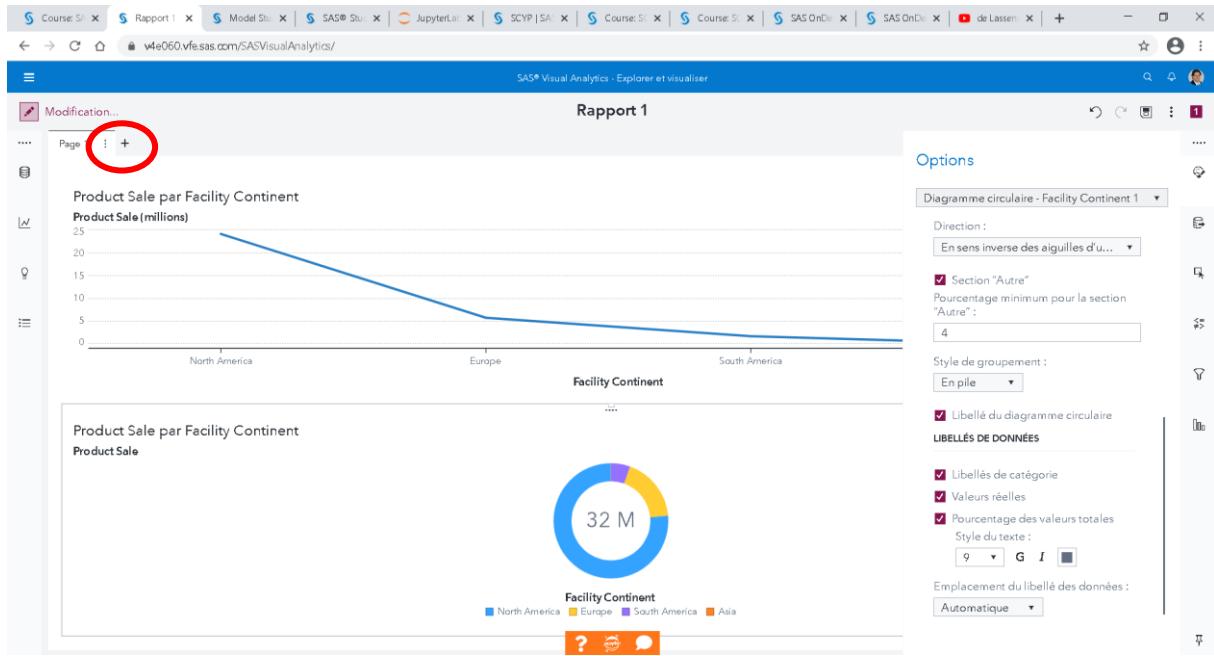
Pourcentage des valeurs totales

Police du texte : Arial

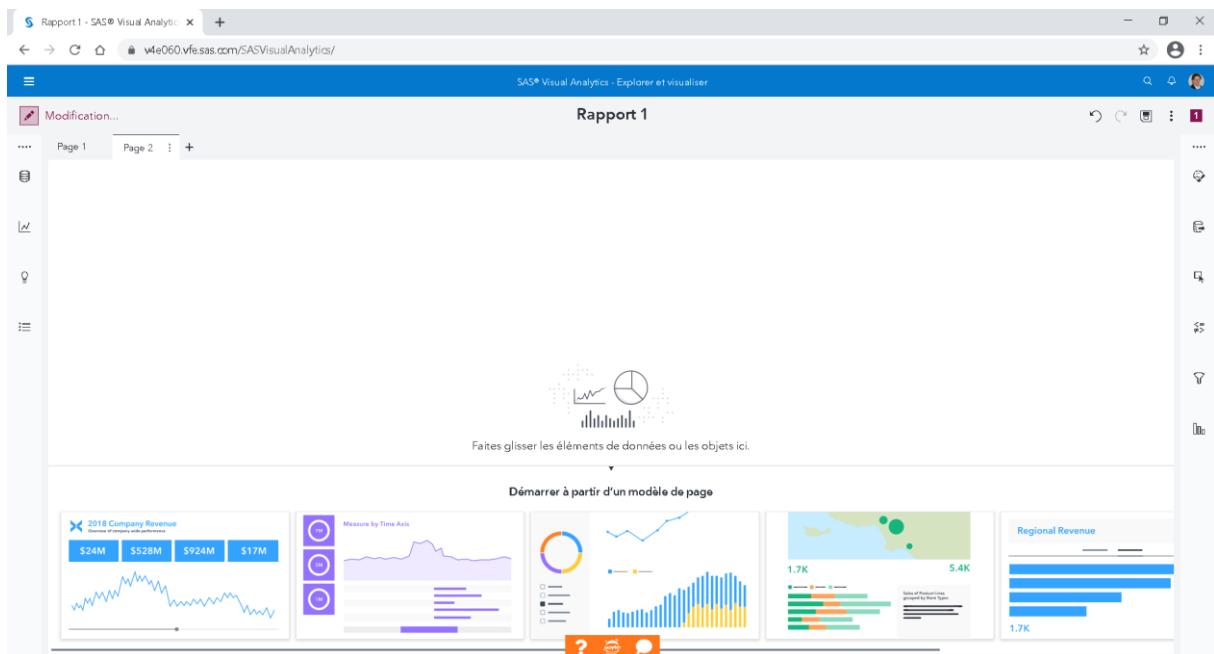
Emplacement du libellé des données : Automatique

Dans la partie droite, si le diagramme circulaire est sélectionné (encadré), on peut sélectionner le libellé et le pourcentage des valeurs.

Si la résolution n'est pas suffisante, il peut être nécessaire de décocher le libellé ou d'agrandir la taille de la fenêtre. On peut aussi réduire le zoom de la page : les légendes sont plus petites, mais on voit plus de choses.



Pour ajouter une page à votre rapport, cliquer sur le + à côté de la « Page 1 ».



Nous avons une seconde page vierge.

Création de hiérarchie et de tableau-croisé

The screenshot shows the SAS Visual Analytics interface. On the left, the 'Données' (Data) panel is open, displaying a list of data sources and their counts: INSIGHT_TOY_DEMO (1.6M), Product Brand (2), Product Line (8), Product Make (71), Product Style (335), Sales Rep (874), and Transaction Date (2K). A red circle highlights the 'Nouvel élément de données' (New data element) section, specifically the 'Hiérarchie' (Hierarchy) option. The main workspace shows a placeholder message 'Faites glisser les éléments de données ou les objets ici.' (Drag data elements or objects here.) and a 'Démarrez à partir d'un modèle de page' (Start from a page template) section with four preview cards.

Aller dans le menu des données pour sélectionner la création d'un nouvel élément de données → une hiérarchie

The screenshot shows the 'Nouvelle hiérarchie' (New hierarchy) dialog box. In the 'Eléments disponibles (15):' (Available elements (15)) list, several items are listed, including Facility, Facility City, Facility Continent, Facility Country, Facility Opening Date, Facility Region, Order, Sales Rep, Transaction Date, Transaction Month, Transaction Weekday, Transaction Year, Unit, and Unit Status. In the 'Eléments sélectionnés (5):' (Selected elements (5)) list, five elements are selected: Product Brand (2), Product Line (8), Product Make (71), Product Style (335), and Product (1.6M). The 'OK' button is visible at the bottom right of the dialog.

Vous pouvez la nommer : « produit ».

Sélectionner dans l'ordre :

Product Brand

Product Line

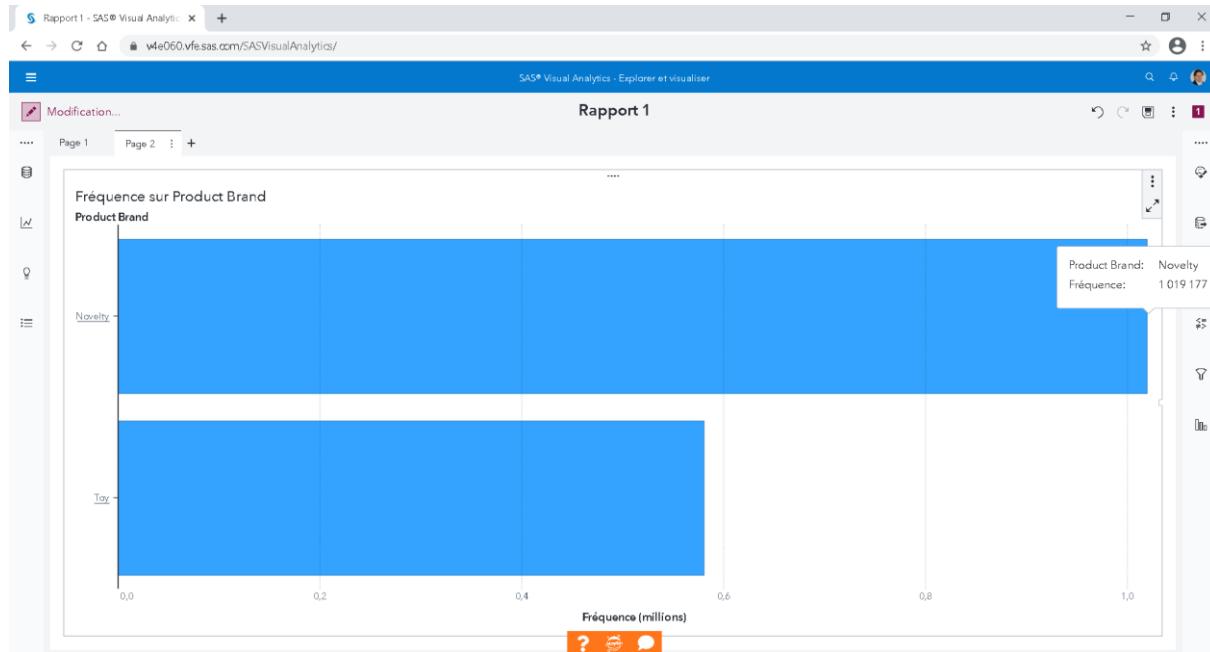
Product Make

Product Style

Product

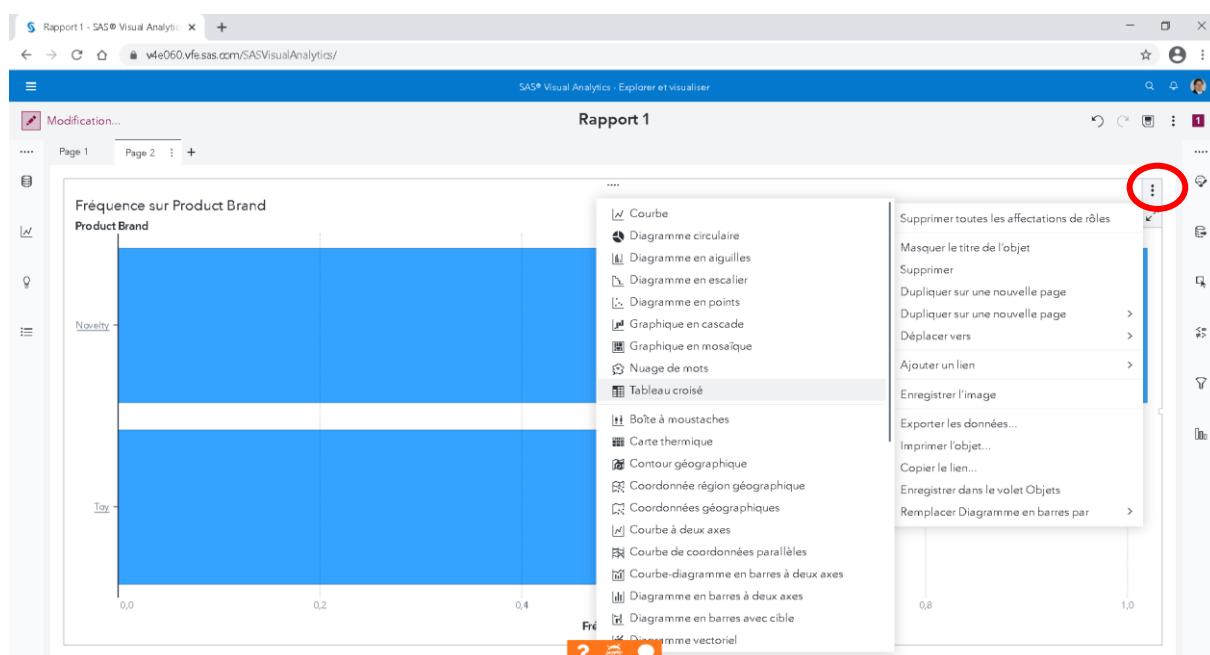
Si vous n'avez pas sélectionné ces éléments dans le bon ordre, les flèches sur la droite permettent de corriger cela.

OK



En ajoutant la hiérarchie produit, le graphique par défaut est un graphique en barres.

Si l'on clique sur « Novelty », on obtient le détail.



On peut remplacer cette représentation par un tableau croisé.

Product Line	Product Make	Fréquence
Figurine	Bear (l)	255 845
Figurine	Bear (m)	140 903
Figurine	Bear (s)	8 723
Figurine	Big Cats (l)	9 060
Figurine	Big Cats (m)	8 966
Figurine	Big Cats (s)	9 134
Game	Cat (l)	8 791
Game	Cat (m)	8 923
Game	Cat (s)	8 675
Game	Dog (l)	8 663
Game	Dog (m)	9 183
Game	Dog (s)	8 423
Plush	Elephant (l)	8 675
Plush	Elephant (m)	8 565
Plush	Elephant (s)	8 948
Plush	Horse (l)	8 853
Plush	Horse (m)	8 748
Plush	Horse (s)	8 581

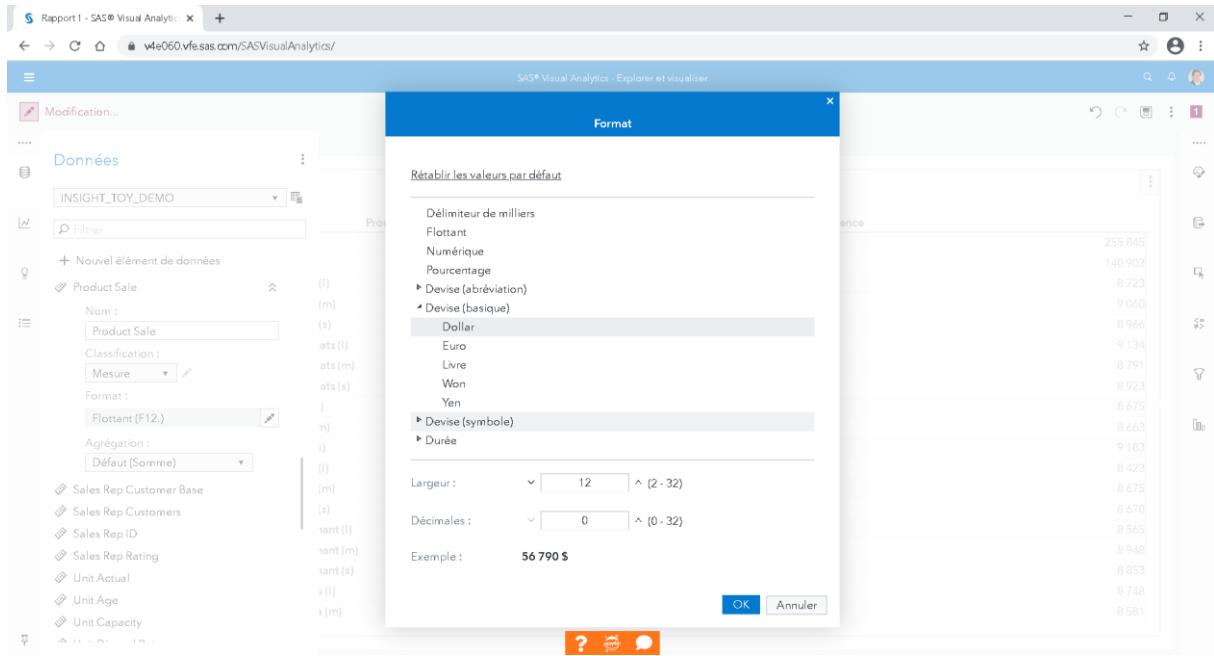
On peut rajouter « Product Sale » pour remplacer la Fréquence.

Si l'on clique sur le nom du niveau de la hiérarchie, on le sélectionne.

Si l'on clique sur la flèche après le nom, on le détaille.

Product Line	Product Make	Fréquence
Figurine	(l)	255 845
Figurine	(m)	140 903
Figurine	(s)	8 723
Figurine	ats (l)	9 060
Figurine	ats (m)	8 966
Figurine	ats (s)	9 134
Game)	8 791
Game	m)	8 923
Game	i)	8 675
Game	[l]	8 663
Game	[m]	9 183
Game	[s]	8 423
Plush	tant (l)	8 675
Plush	tant (m)	8 565
Plush	tant (s)	8 948
Plush	x (l)	8 853
Plush	x (m)	8 748
Plush	x (s)	8 581

Dans le menu des données, sur la double-flèche de Product Sale, il est possible de changer le format



Afin de le mettre avec le symbole \$ par exemple.

OK

The screenshot shows the 'Remplacer l'élément de données' dialog box. It lists various data elements such as Customer Distance, Customer Satisfaction, Facility Age, Facility Efficiency, Facility Employees, Pourcentage de la fréquence, Product Cost of Sale, Product Material Cost, Product Price (target), Product Quality, Product Sale, Sales Rep Customer Base, Sales Rep Customers, Sales Rep ID, Sales Rep Rating, Unit Actual, Unit Age, Unit Capacity, Unit Discard Rate, and Unit Discards. 'Product Sale' is highlighted. At the bottom right are 'Annuler' and 'OK' buttons.

Ajouter une troisième page à votre rapport.

Création de hiérarchie géographique

Pour représenter les données sur une carte du monde, nous allons ici créer une hiérarchie géographique « Continent → Pays → Région → Ville → Entité » en utilisant les coordonnées « World Geodetic System (WGS84) ».

Dans les données :

Cliquer sur la double flèche permettant de modifier les propriétés de « Facility Continent »

Au lieu de catégorie, cliquer sur Géographie

Sélectionner Latitude et longitude des données au lieu de la recherche de noms ou de codes géographiques.

Sélectionner la latitude et la longitude du continent

Sélectionner « xyFacility Continent Lat » pour la Latitude

Et sélectionner « xyFacility Continent Lon » pour la Longitude

OK

Refaire les mêmes étapes pour Country, Région, City et Facility

On obtient 5 éléments géographiques.

On peut alors créer une hiérarchie : Nouvel élément de données → pour ajouter une nouvelle hiérarchie.

Sélectionner les éléments dans l'ordre : Continent, Country, Région, City et Facility.

Renommer cette hiérarchie Geo.

OK

Rapport 1 - SAS® Visual Analytics

SAS® Visual Analytics - Explorer et visualiser

Rapport 1

Données

- INSIGHT_TOY_DEMO
- Filtrer
- + Nouvel élément de données
- Facility - 99
- Facility City - 99
- Facility Continent - 4
- Facility Country - 18
- Facility Region - 68
- Hiérarchie**
- Geo
- Produit
- Mesure
- Customer Distance
- Customer Satisfaction
- Facility Age
- Facility Efficiency
- Facility Employees

Annuler la sélection

Nom : Facility
Contenu : Product
Utilisé par : Tableau croisé - Produit 1 rôle

Démarrez à partir d'un modèle de page

Faites glisser les éléments de données ou les objets ici.

Regional Revenue

1.7K 5.4K

Ajouter cette hiérarchie à la troisième page de votre Rapport.

En double-cliquant sur le point de l'Europe, on obtient :

Rapport 1 - SAS® Visual Analytics

SAS® Visual Analytics - Explorer et visualiser

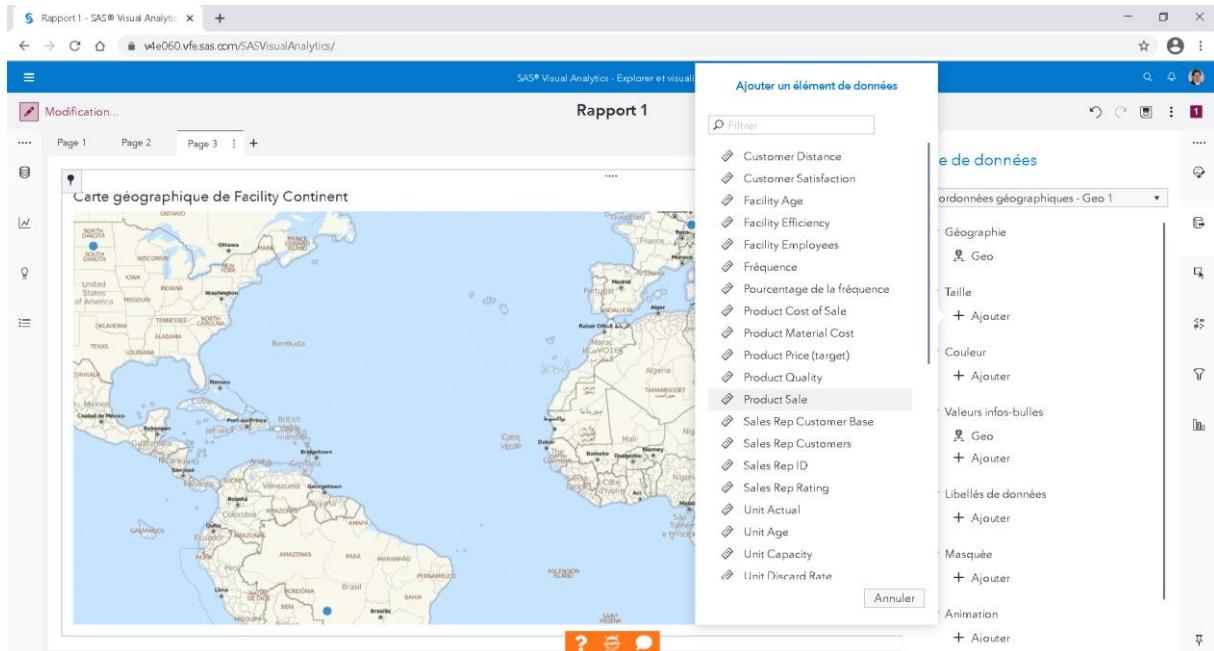
Rapport 1

Page 1 Page 2 Page 3

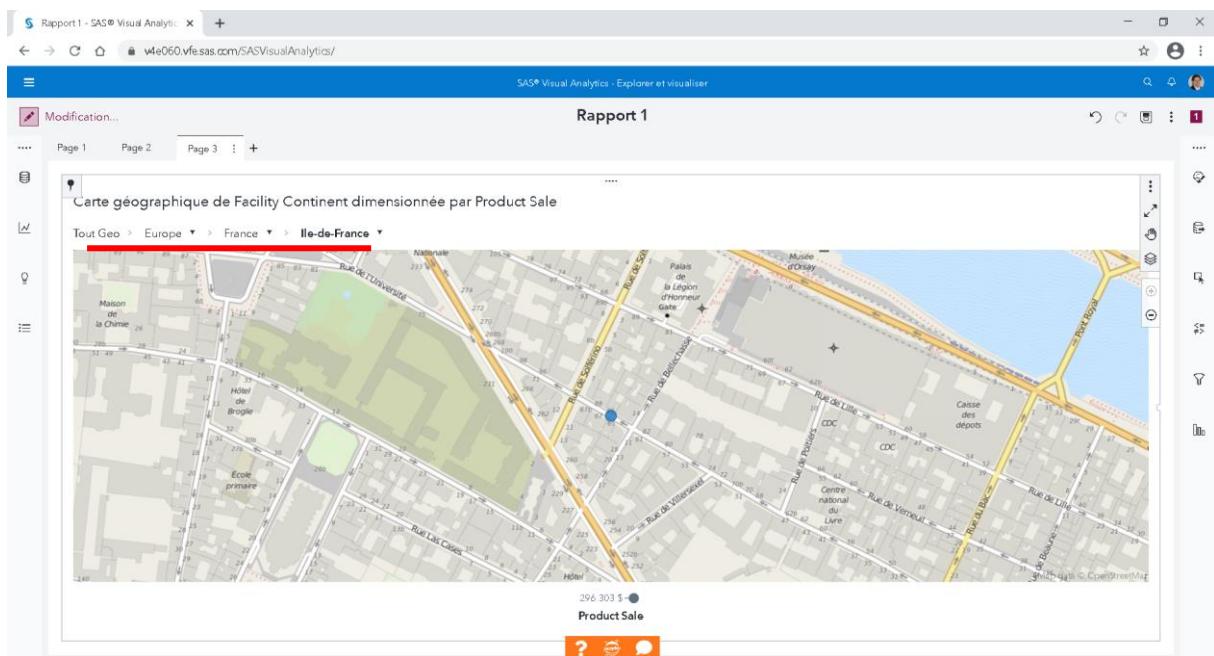
Carte géographique de Facility Continent

Map data © OpenStreetMap

Si l'on veut le chiffre d'affaires, il faut ajouter « Product sale ».

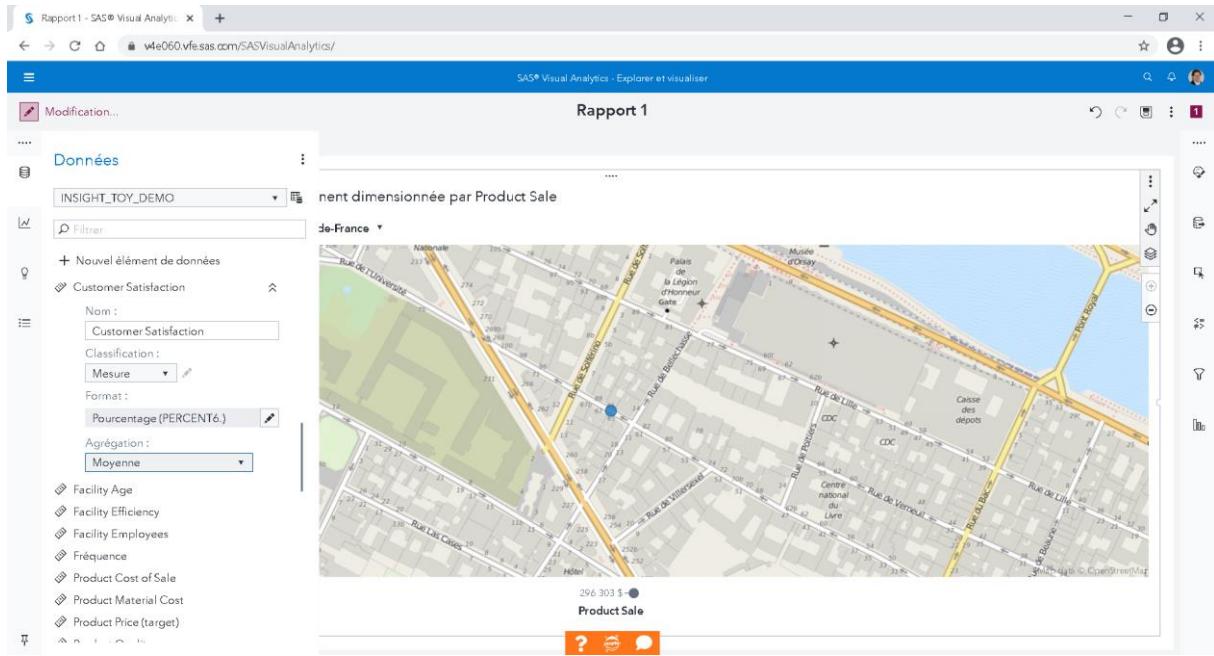


En haut à gauche du graphique, on peut remonter dans la hiérarchie.



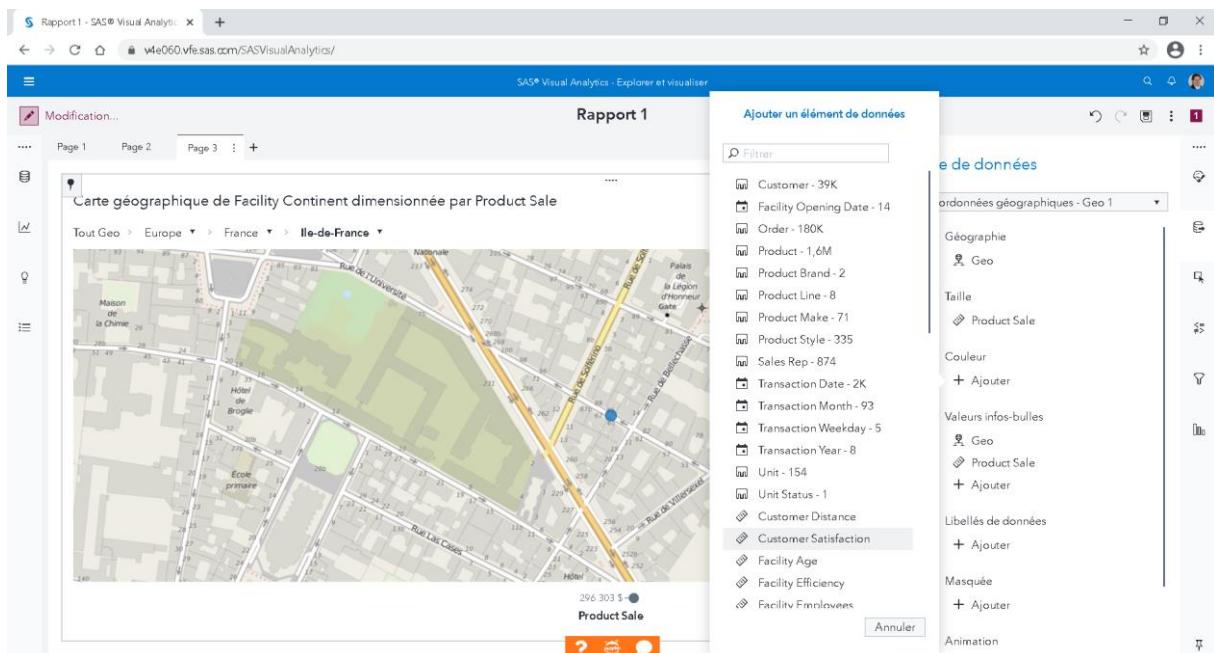
Si l'on rajoute aussi la satisfaction client, on obtient le graphique suivant avec le dégradé de couleurs en fonction de ce dernier élément.

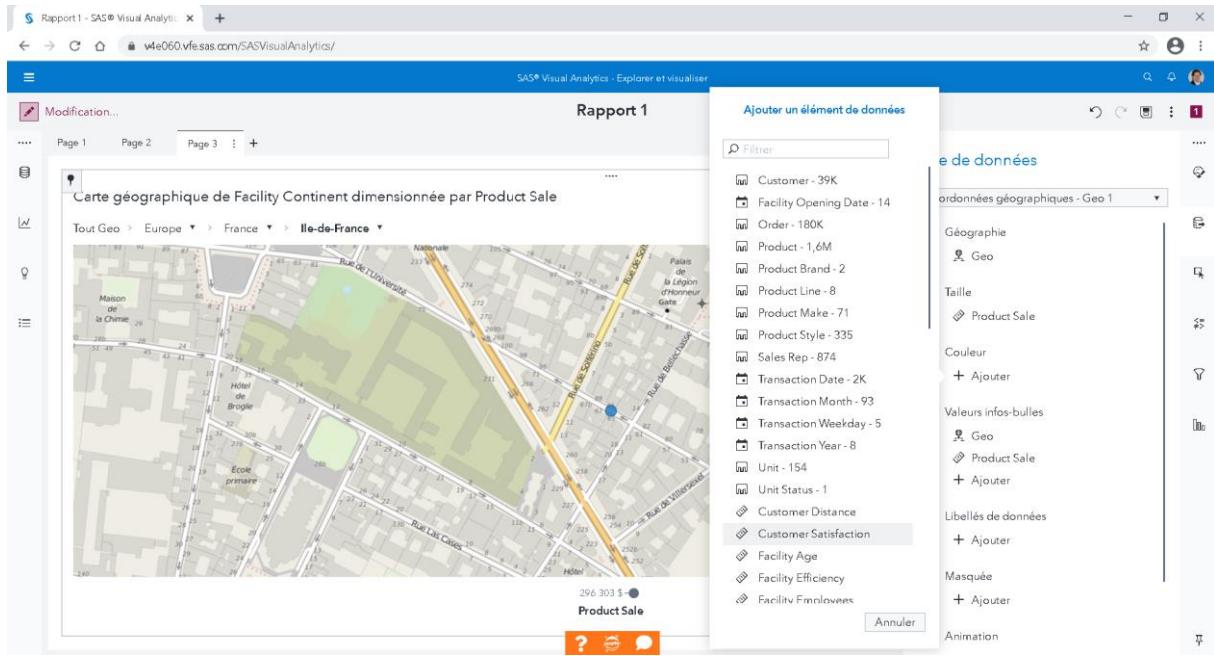
La satisfaction client est un pourcentage. La somme d'un pourcentage n'a pas de sens mais la moyenne oui.



Donc pour la satisfaction client, sélectionner l'agrégation par la moyenne.

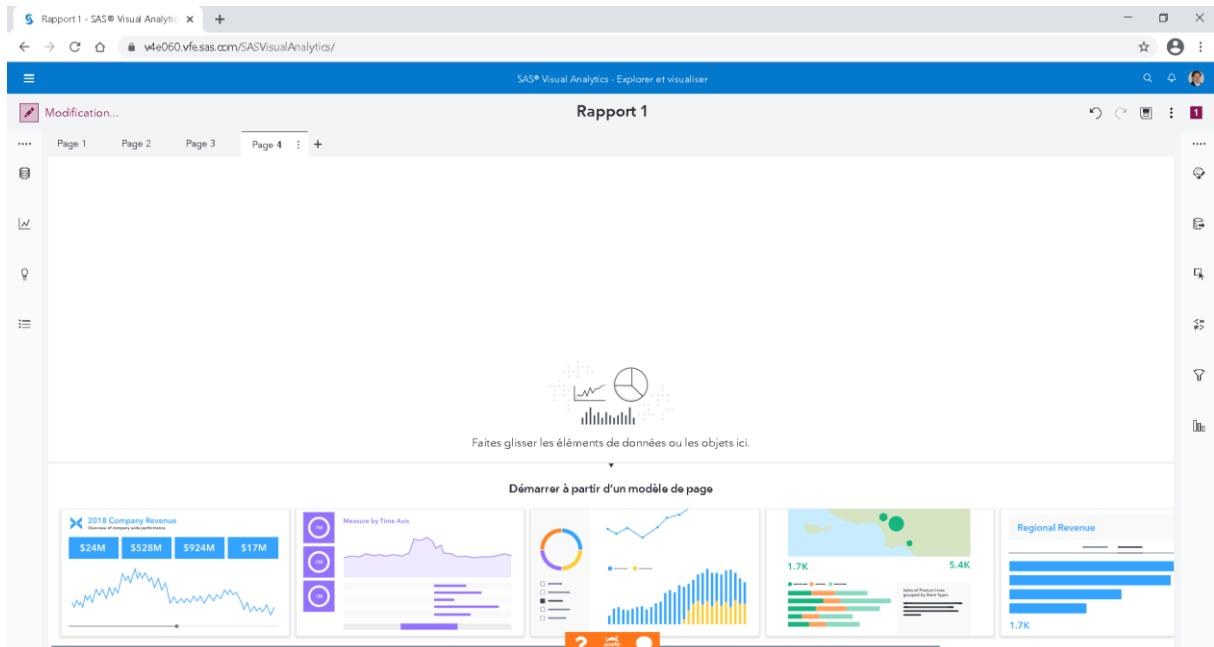
Ajouter la satisfaction client comme couleur des points sur votre carte.





Ajouter une quatrième page à votre rapport.

Création de tableau de bord

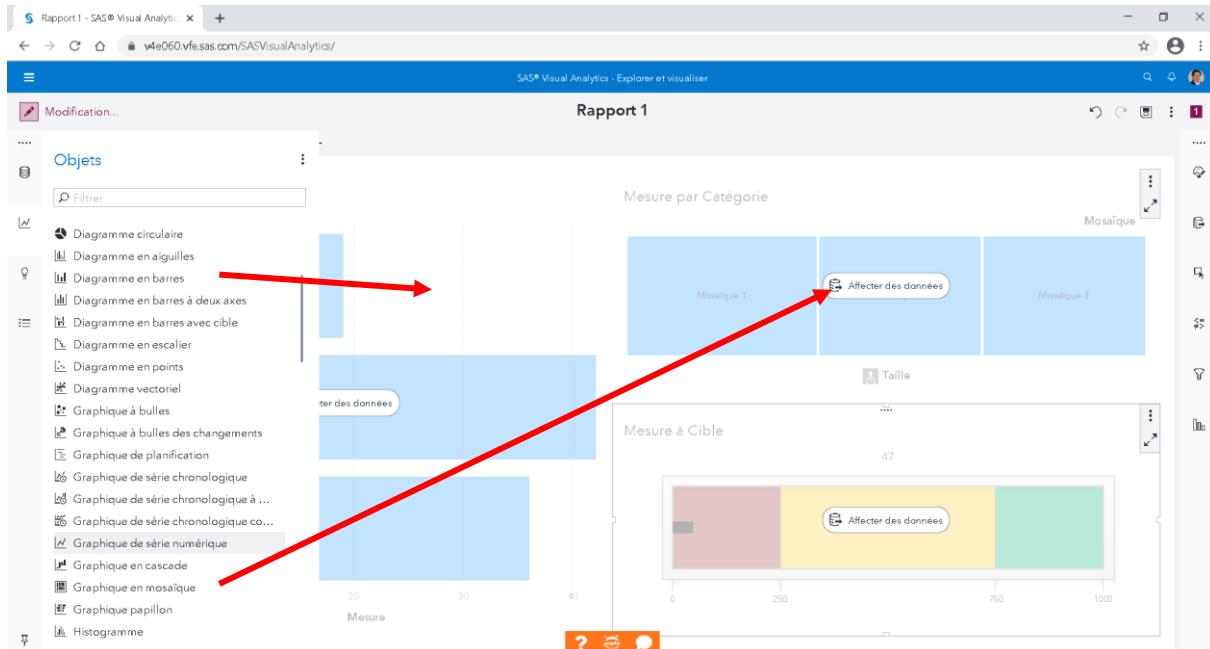


Dans votre 4^{ème} page, depuis les Objets, sélectionner puis glisser-lâcher

Le diagramme en barres au milieu,

Le graphique en mosaïque à droite,

Et la jauge en bas à droite.



La satisfaction client est un pourcentage. Une somme de pourcentages ne veut rien dire.

Dans les données, aller sur la mesure « Customer Satisfaction » afin de modifier dans les propriétés, l'agrégation de Somme à Moyenne.

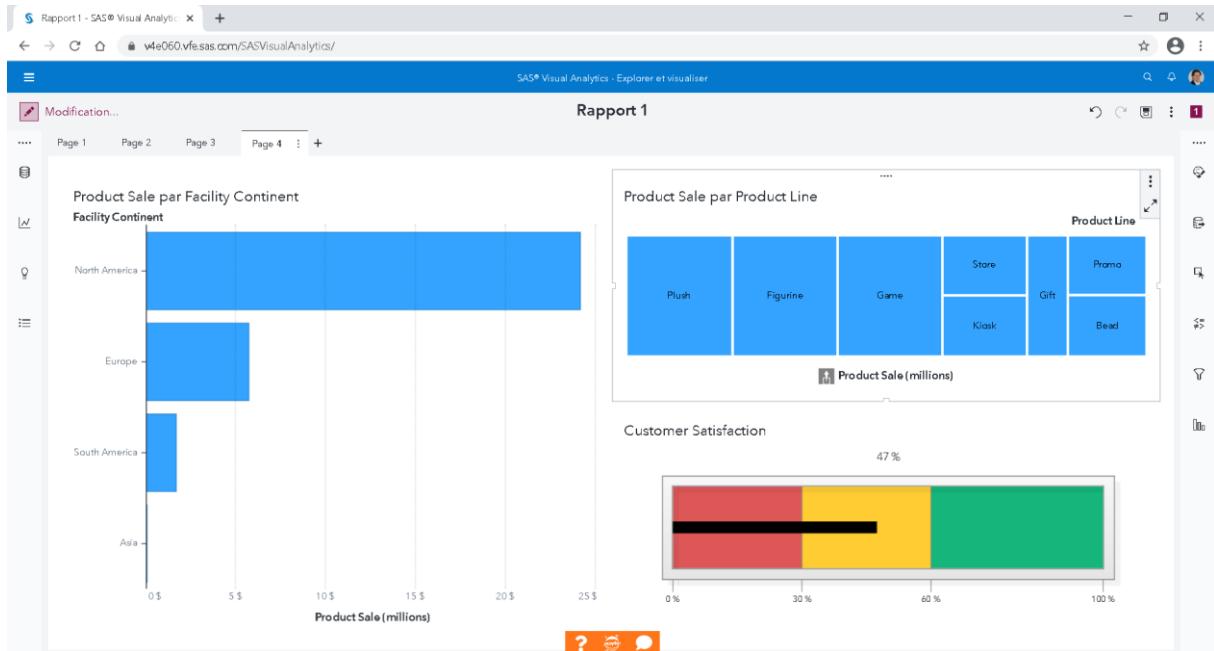
Sélectionner cette Customer Satisfaction pour la glisser-lâcher sur la jauge.

Sélectionner le continent (de Géographie) pour le glisser-lâcher sur le diagramme en barres.

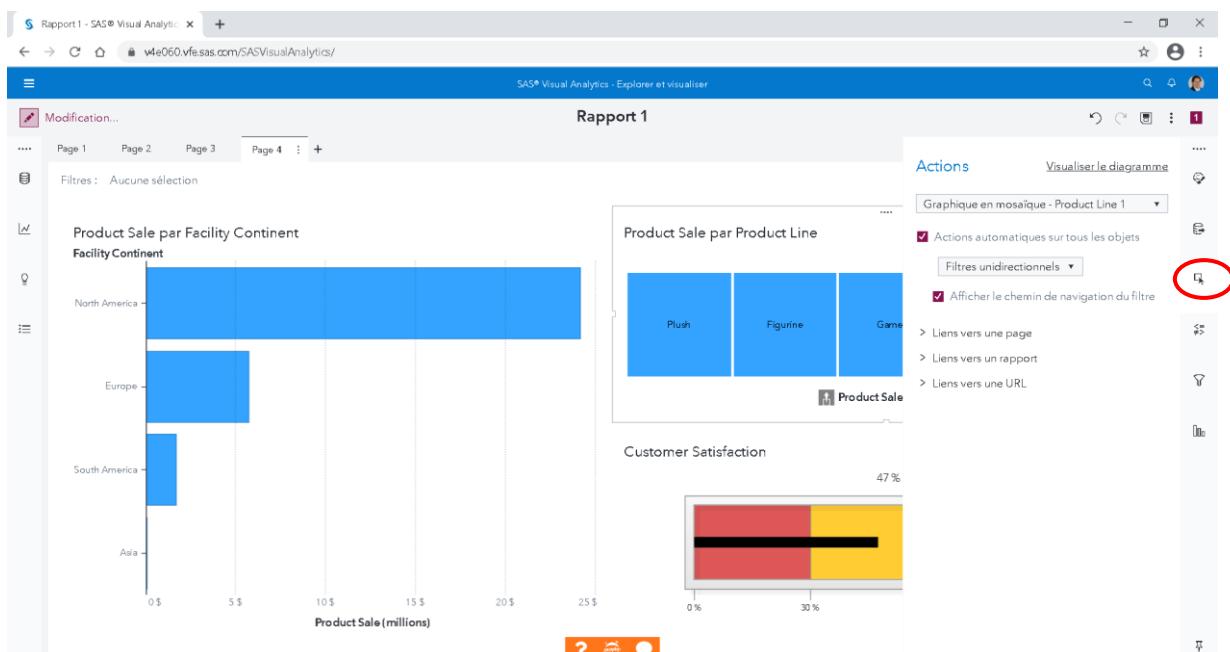
Sélectionner « Product Line » pour le glisser-lâcher sur Graphique en mosaïque.

Sélectionner « Product Sale » pour le glisser-lâcher sur le diagramme en barres et sur Graphique en mosaïque.

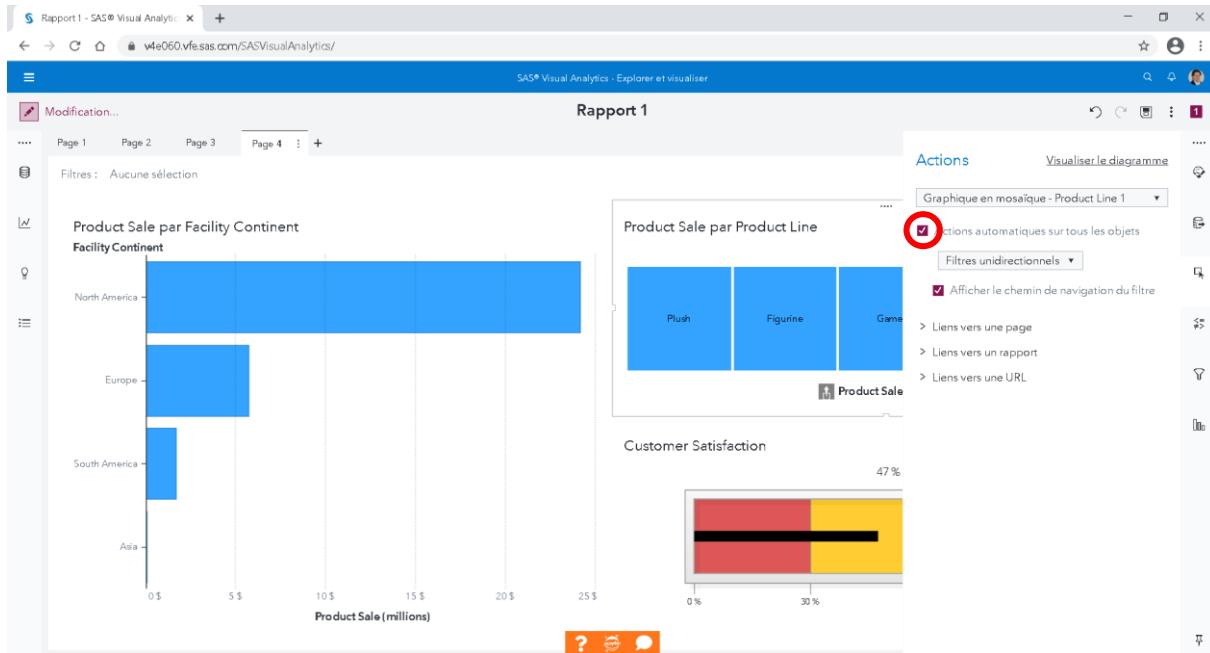
Vous pouvez aussi sélectionner le diagramme en barres ou le graphique en mosaïque et remplacer la fréquence depuis le rôle des données par Product Sale.



Sélectionner le diagramme en barres, puis sur la droite l>Action.



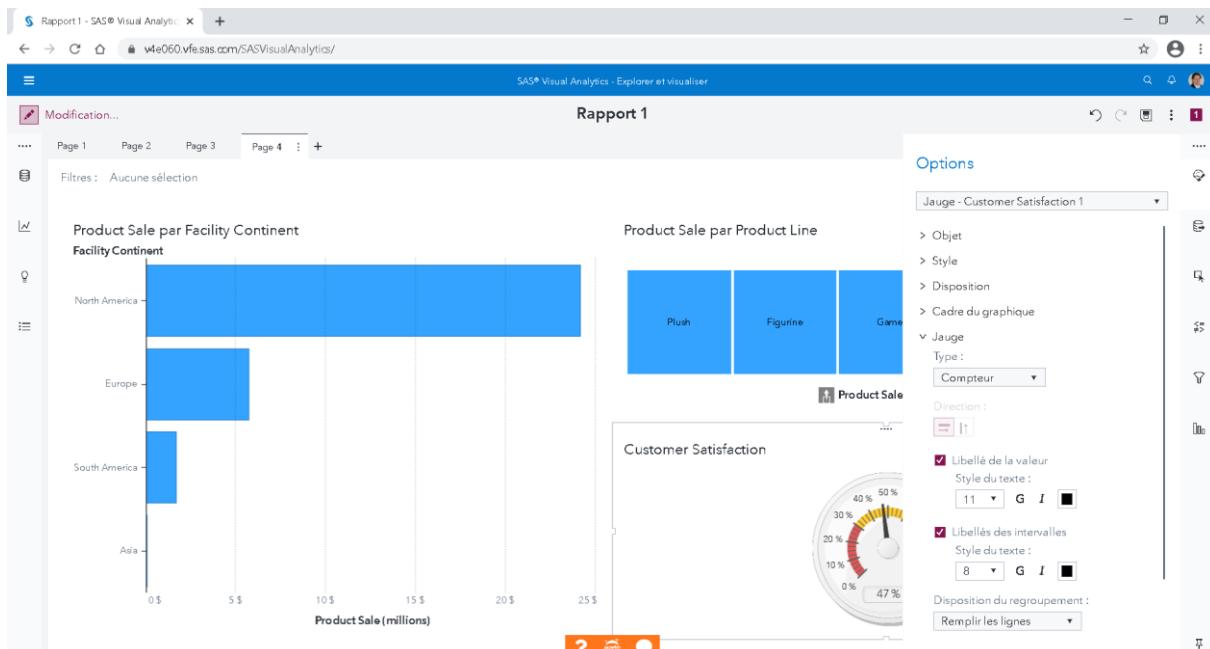
Cocher l'action automatique sur tous les objets

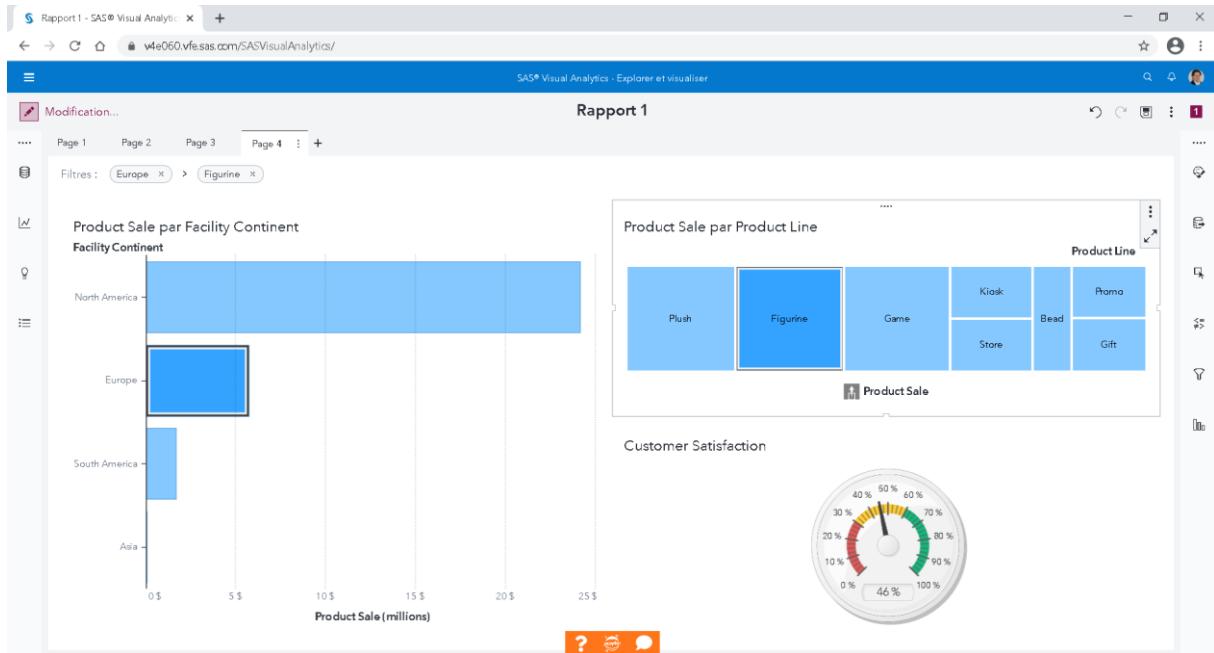


Sélectionner le filtre bidirectionnel

Si l'on clique sur un bâton du diagramme en barres, cela filtre les éléments sur les deux autres objets.

Sélectionner la jauge et dans les options, sélectionner le compteur comme type de jauge.

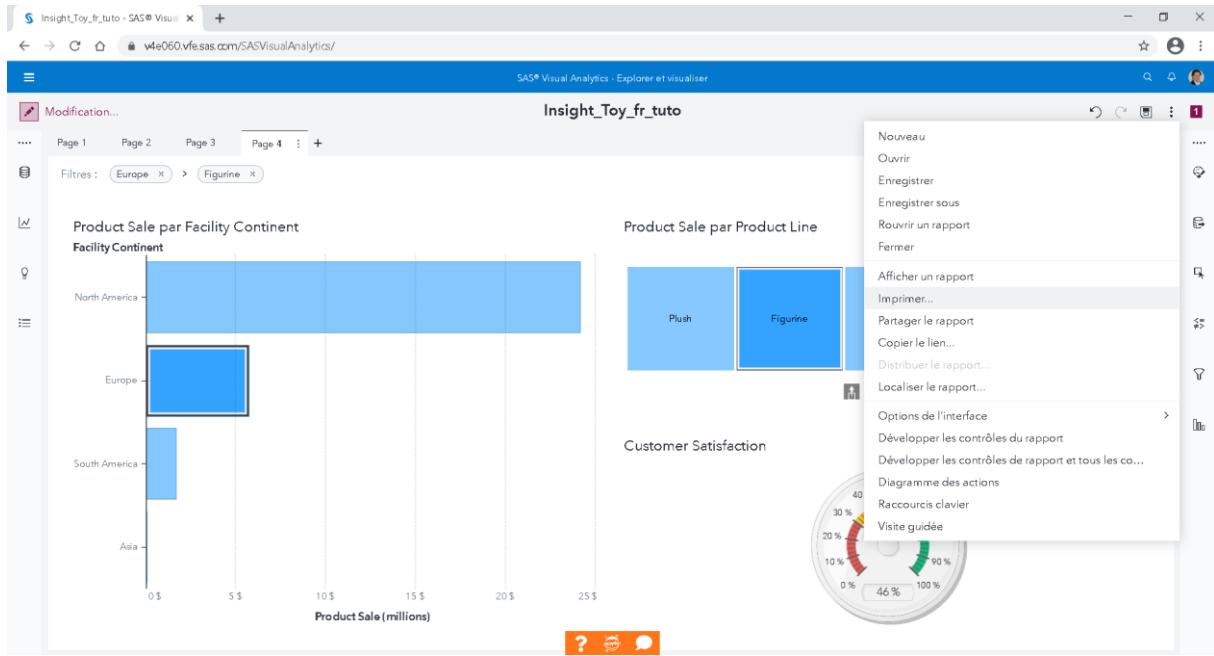




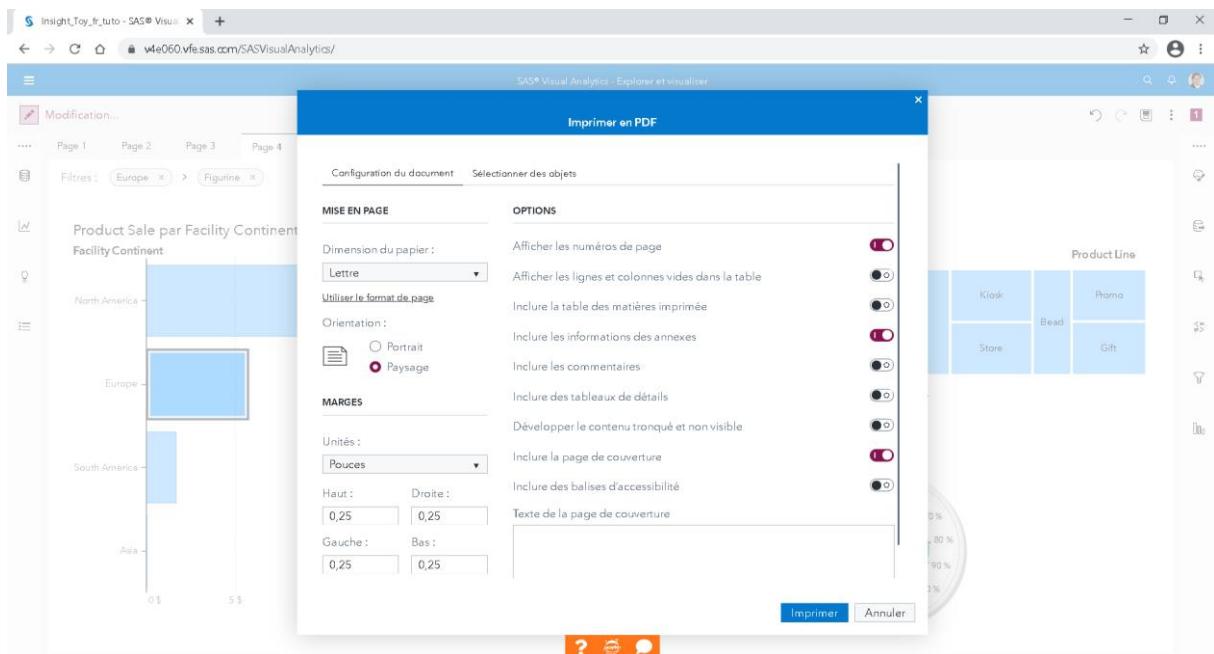
Enregistrer votre travail

The screenshot shows two windows of the SAS Visual Analytics interface. The top window is titled 'Rapport 1 - SAS® Visual Analytics' and displays a dashboard with three visualizations: 'Product Sale par Facility Continent' (a bar chart), 'Product Sale par Product Line' (a bar chart), and 'Customer Satisfaction' (a gauge chart). A context menu is open on the right side of the dashboard, with the 'Enregistrer sous' (Save As) option highlighted and circled in red. The bottom window is titled 'Insight_Toy_fr_tuto - SAS® Visual Analytics' and shows the 'Enregistrer sous' (Save As) dialog box. The dialog box has a search bar and a tree view of 'Dossiers'. Under 'My Folder', there is a folder named 'My Folder' containing sub-folders like 'My Snippets' and 'SAS Videos'. The user is prompted to enter a name ('Nom') and select a type ('Type'). The 'Nom' field contains 'Insight_Toy_fr_tuto' and the 'Type' dropdown is set to 'Report'. There are 'Enregistrer' (Save) and 'Annuler' (Cancel) buttons at the bottom.

Vous pouvez l'enregistrer dans votre dossier : « My Folder »



Vous pouvez l'imprimer pour générer un pdf.



Exercices

Créer un nouveau rapport.

Créer un nouveau rapport sur la même table « Insight Toy Demo ».

1. Faire un graphique automatique sur la première page qui représente « Product Cost of Sale » par « Facility Continent ».
2. Ajouter une seconde page à votre rapport afin de créer une boite à moustache représentant « Product Cost of Sale » par « Product Line ».
3. Ajouter une troisième page qui représente « Unit Capacity » par le mois de la transaction. Ajouter sur cette troisième page un graphique en mosaïque qui représente le chiffre d'affaires par pays.
4. Ajouter une 4^{ième} page à votre rapport. Combien de chiffre d'affaires est généré pour les villes d'Europe ? Créer un diagramme en barres. Dans la partie droite, ajouter un filtre sur les continents afin de n'afficher que l'Europe.
5. Ajouter une 5^{ième} page afin d'avoir le chiffre d'affaires en Angleterre à Manchester et à Londres entre 2003 et 2004. Faire au moins 2 graphiques différents sur cette page pour représenter la même chose. Attention, il y a deux Manchester : un US et un UK.
6. Créer au moins deux autres sections (section=page) avec au moins 6 objets différents : graphiques, jauge, contrôles, tableaux, etc... Pour cela, relire la description du cas au début de ce document. Cette dernière question est libre. Testez des graphiques, cherchez ... La présentation compte beaucoup.

Enregistrer ce rapport dans votre dossier.

L'imprimer pour l'envoyer par e-mail à votre enseignant.

Machine Learning avec SAS Visual Analytics

Nous allons faire ici du Machine Learning simplement en quelques clics. C'est très simple et surtout cela permet présenter les modèles de Machine Learning aux métiers, au business, à votre manager ou selon, à votre client.

Dans les organisations actuelles, il est primordial de pouvoir communiquer sur ce que vous faites.

Plus que de se concentrer sur la performance du modèle, nous allons nous intéresser principalement au retour sur investissement : le Rol !

L'idée n'est pas de faire des modèles de décisions qui par exemple reconnaissent l'écriture manuscrite, auquel cas on mesure la performance du modèle notamment par le taux de mal classés.

Le but n'est pas non plus de faire de l'estimation, par exemple du revenu potentiel, où l'on regarderait la différence entre la valeur prédite par le modèle et celle de la réalité.

On s'intéresse ici au ranking. C'est-à-dire à ordonner ou classer les clients depuis celui ayant la probabilité la plus forte jusqu'à celui ayant celle la plus faible d'acheter. Le ranking permet simplement de prioriser.

On ne prédit pas l'achat ou non mais simplement cherche à se concentrer sur les clients ayant la probabilité la plus forte d'acheter ou inversement.

Si le Rol est important, cela peut être très intéressant.

Le cas Organics

Un supermarché britannique souhaite lancer une nouvelle ligne de produits bio. Ils aimeraient cibler parmi tous les clients ceux ayant la probabilité la plus forte d'achat. Ce supermarché a déjà mis en place une carte de fidélité, ce qui permet d'avoir sur les clients quelques informations sociodémographiques et de les compléter avec des informations comportementales.

Une première campagne a été faite sur quelques clients sélectionnés au hasard afin d'obtenir une base permettant de modéliser leur comportement.

La table Organics est composée de 12 colonnes et 22 000 observations.

Liste des variables		
Variable	Type	Long.
DemAffl	Num.	Nombre de visite moyenne par mois (entre 1 et 32)
DemAge	Num.	Age du client
DemGender	Texte	Sexe
DemReg	Texte	Région
DemTVReg	Texte	Zone de télévision
ID	Texte	Identifiant du client
PromClass	Texte	Niveau de fidélité

Liste des variables		
Variable	Type	Long.
PromSpend	Num.	CA généré
TargetAmt	Num.	Si le client a acheté : combien

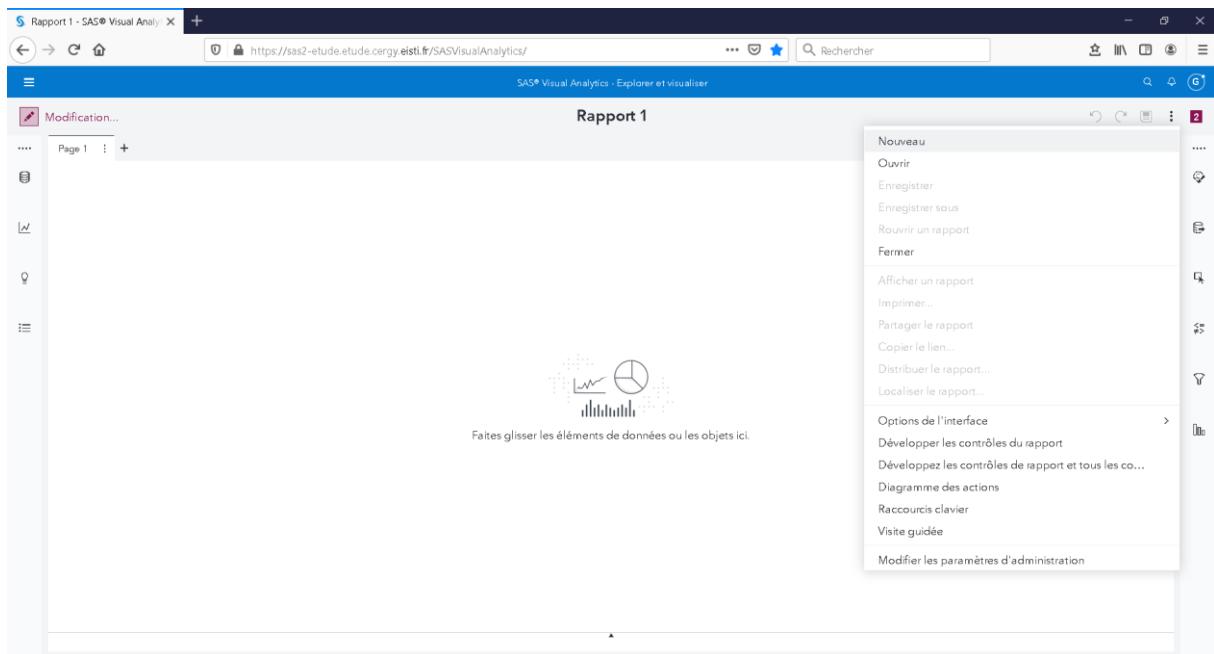
Aller dans SAS Visual Analytics : Explorer et Visualiser

The screenshot shows the SAS Drive interface. On the left, a sidebar titled 'CYCLE DE VIE ANALYTIQUE' has 'Explorer et visualiser' selected. The main area displays a folder structure under 'Contenu SAS > Courses > YVA285'. A right-hand panel shows details for a folder named 'VA285' with tabs for 'Details', 'Commentaires', and 'Miniatures'. The bottom status bar shows the date as 29/12/2020.

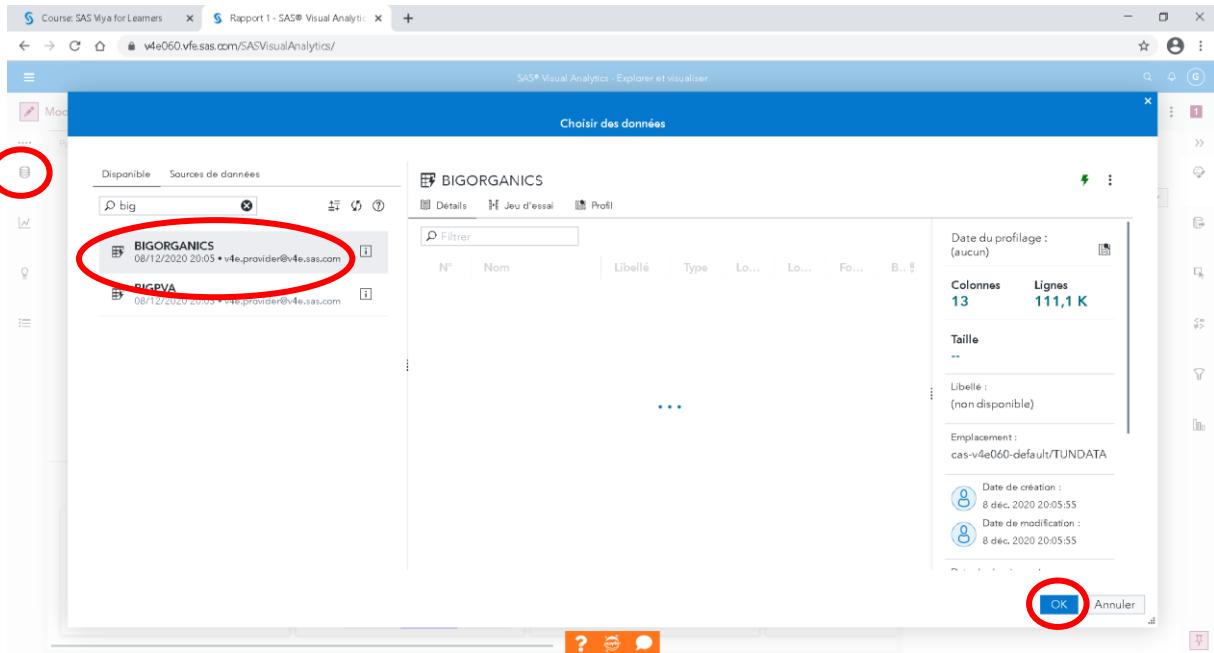
Depuis le Drive : « Explorer et visualiser »

The screenshot shows the SAS Visual Analytics interface. The top navigation bar includes 'Course: SAS Mya for Learners', 'SAS Drive', and a search bar. The main area features a 'Explorer et visualiser' section with a 'Nouveau rapport' button highlighted by a red circle. To the right, there's an 'Options' panel with tabs for 'Page 1', 'Général', 'Style', 'Disposition', and 'Contrôles de la page'. A small preview of a report is visible on the right.

Nouveau rapport



Pour ceux étant déjà dans l'application, en haut, à droite, sur les trois petits points, créer une nouvelle visualisation



Dans données, sélectionner une nouvelle source.

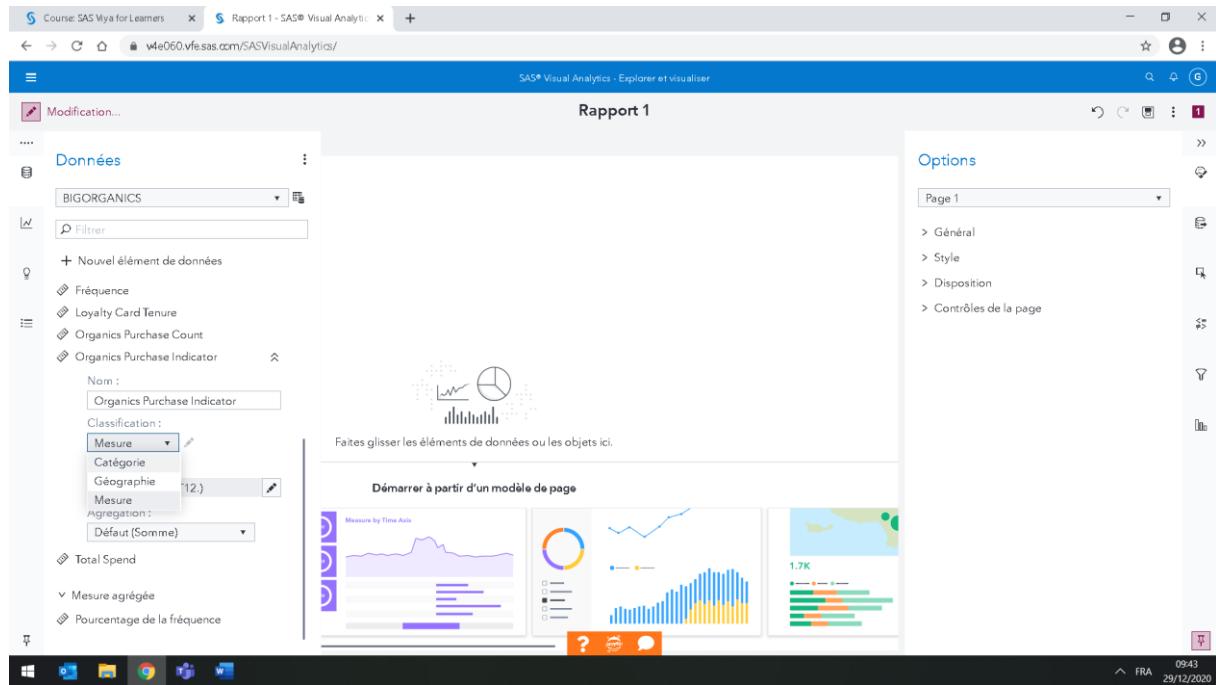
Sélectionner la table **Bigorganics**

OK

Exploration

Dans un premier temps, il peut être intéressant de découvrir la table. Dans les objets on peut prendre 4 diagrammes en barres.

Modifier les propriétés de la variable « Organics Purchase Indicator » afin de remplacer la classification en mesure par celle de catégorie.

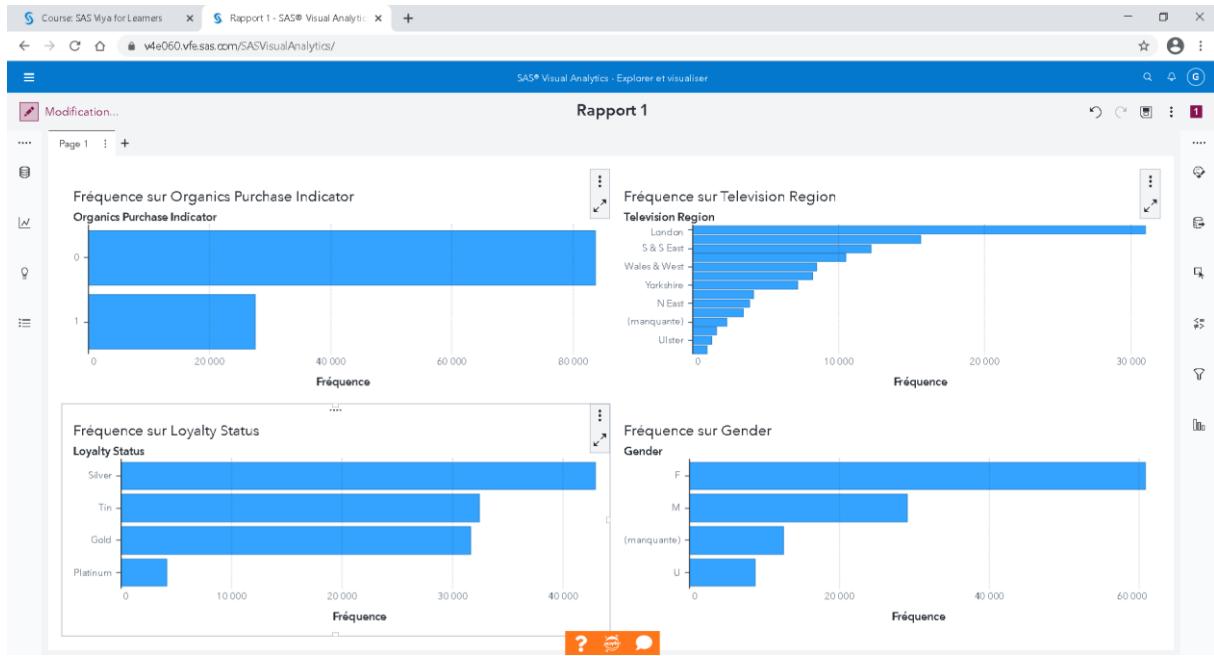


Ajouter « Organics Purchase Indicator » au premier diagramme en barres

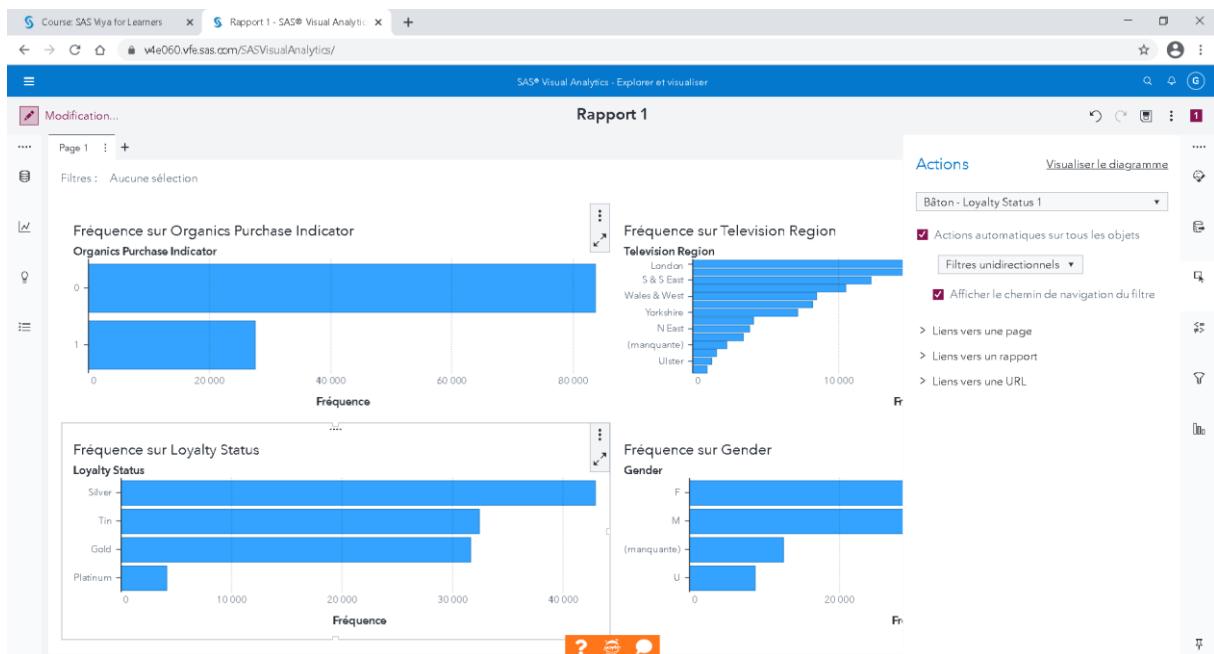
Ajouter « Television Region » au second diagramme en barres

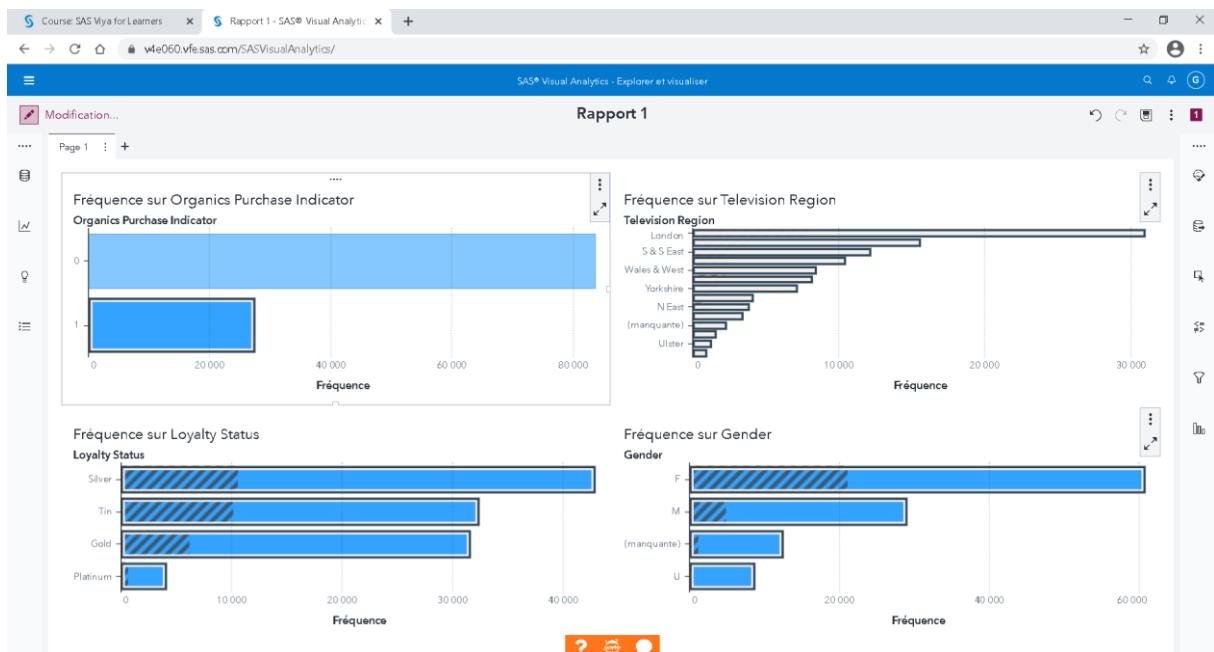
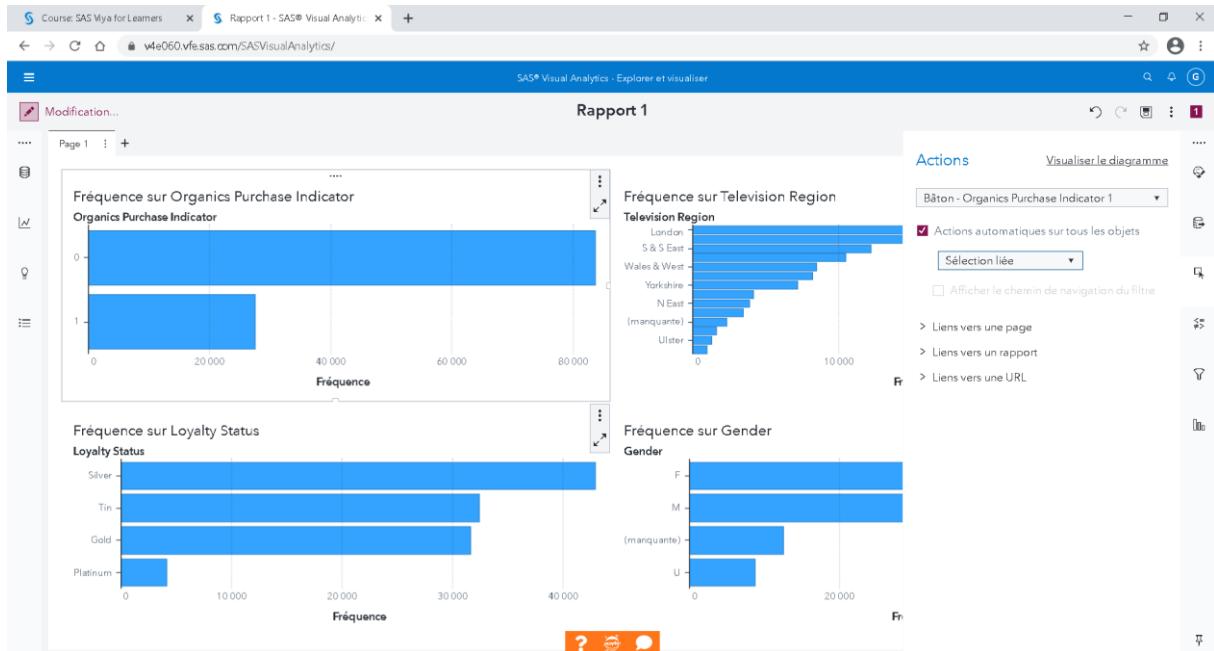
Ajouter « Loyalty Status » au troisième diagramme en barres

Ajouter « Gender » au quatrième diagramme en barres



Sélectionner le diagramme en barres « Organics Purchase Indicator » puis sur la droite, dans les Actions. Sélectionner les actions automatiques sur tous les objets et une sélection liée.





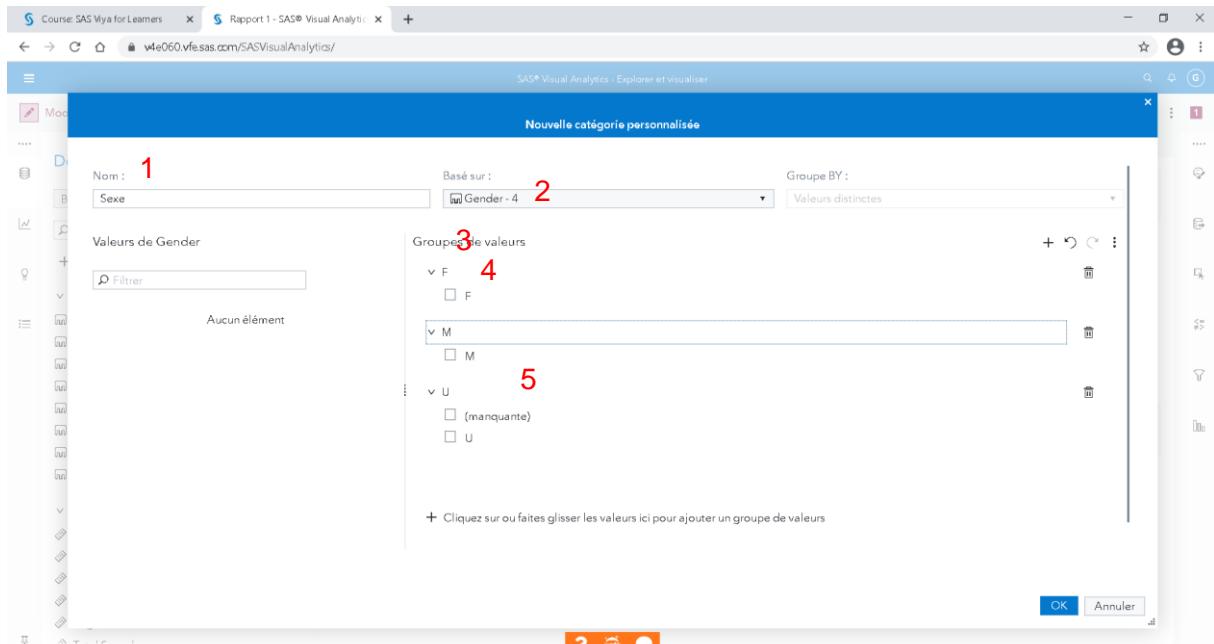
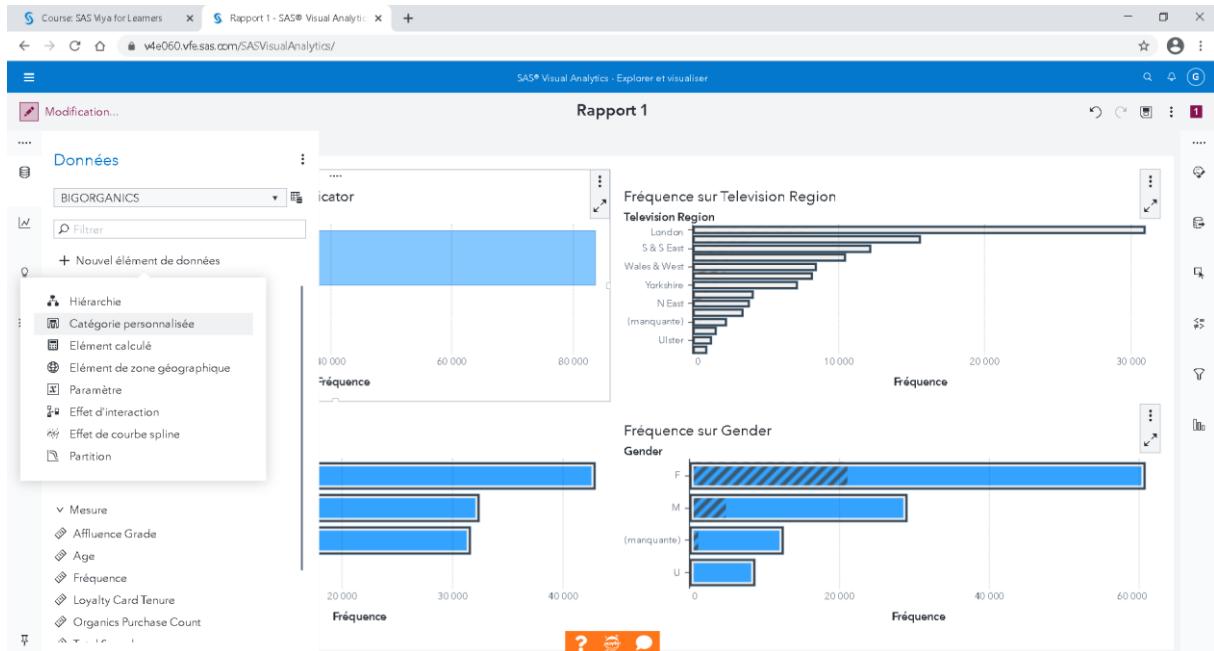
Il est alors possible en cliquant sur une barre, de voir où les individus sélectionnés se retrouvent sur les autres diagrammes en barres.

On remarque notamment :

que la proportion de femmes achetant est plus importante.

pour le sexe, il y a 4 possibilités. Les femmes (F= Female), hommes (M = Male), U = Unknow, et manquant. Il serait intéressant de créer un groupe rassemblant « U » et manquant.

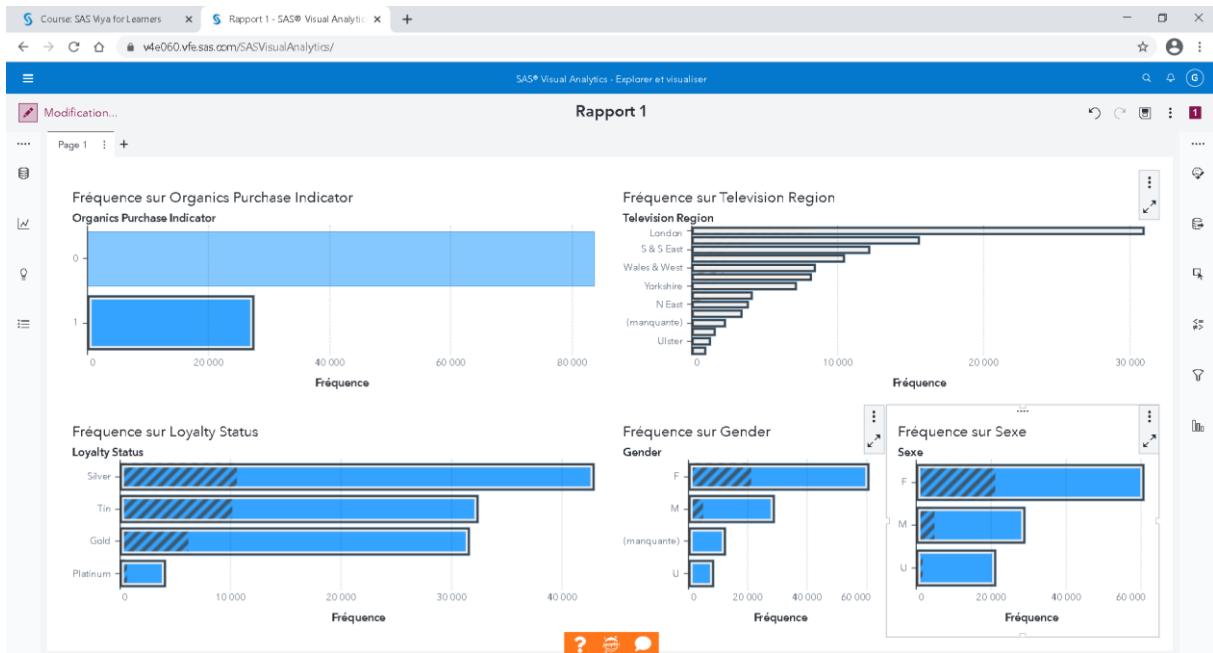
Nous allons donc ajouter une catégorie personnalisée (A gauche, données → Ajouter)



On peut lui donner un nom [1], sélectionner la variable Gender [2], renommer le premier groupe en « Femme » en faisant un clic-droit pour modifier le nom du groupe [3], glisser-lâcher la colonne « F » [4], ajouter un second groupe (avec le + en haut à droite) et refaire la même chose pour les hommes. Enfin créer un groupe contenant « U » et « manquant »[5].

OK

On peut rajouter un diagramme en barres de cette nouvelle variable.



Arbre

Ajouter une page au rapport.

Dans cette seconde page, ajouter l'objet de l'arbre de décision.

The figure shows the SAS Visual Analytics interface with the 'Objets' (Objects) panel open on the left side. A red arrow points to the 'Arbre de décision' (Decision Tree) option under the 'SAS Visual Statistics' section. The main workspace is currently empty, displaying a placeholder message: "Faites glisser les éléments de données ou les objets ici." (Drag data elements or objects here.).

Dans la partie droite, dans le rôle des données,

Ajouter « Organics Purchase Indicator » comme variable de réponse.

Ajouter le « Geographic Region », « Loyalty Status », « sexe » (mais pas Gender), « Television Region », « Affluence Grade », « Age » et « Total Spend » comme Prédicteurs.

Ok

Il est possible de maximiser l'affichage en enlevant les punaises à droite et à gauche.

Il y a quatre parties sur ce rapport.

En haut à droite, parmi les 6 variables explicatives, 3 sont importantes : l'âge, l'affluence (nombre de visite par mois), et le sexe.

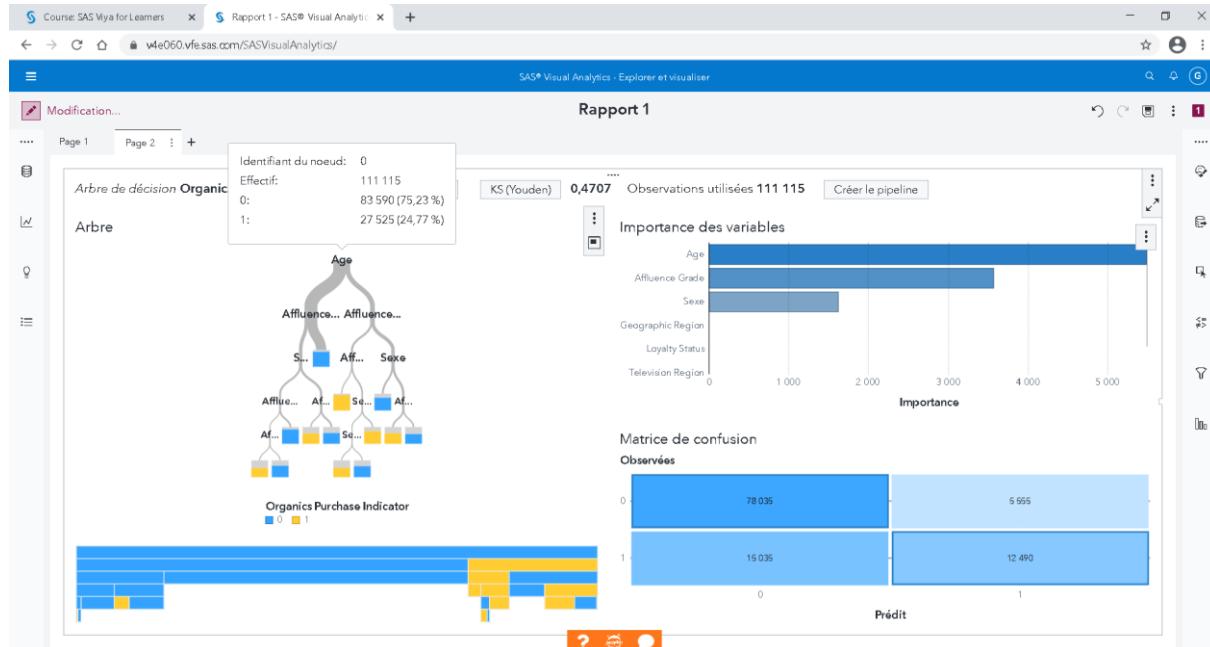
Il est possible de zoomer sur l'arbre. La première variable discriminante est l'âge.

Il y a dans la base 75% de clients qui n'achètent pas de produits bio et donc 25% qui en consomment.

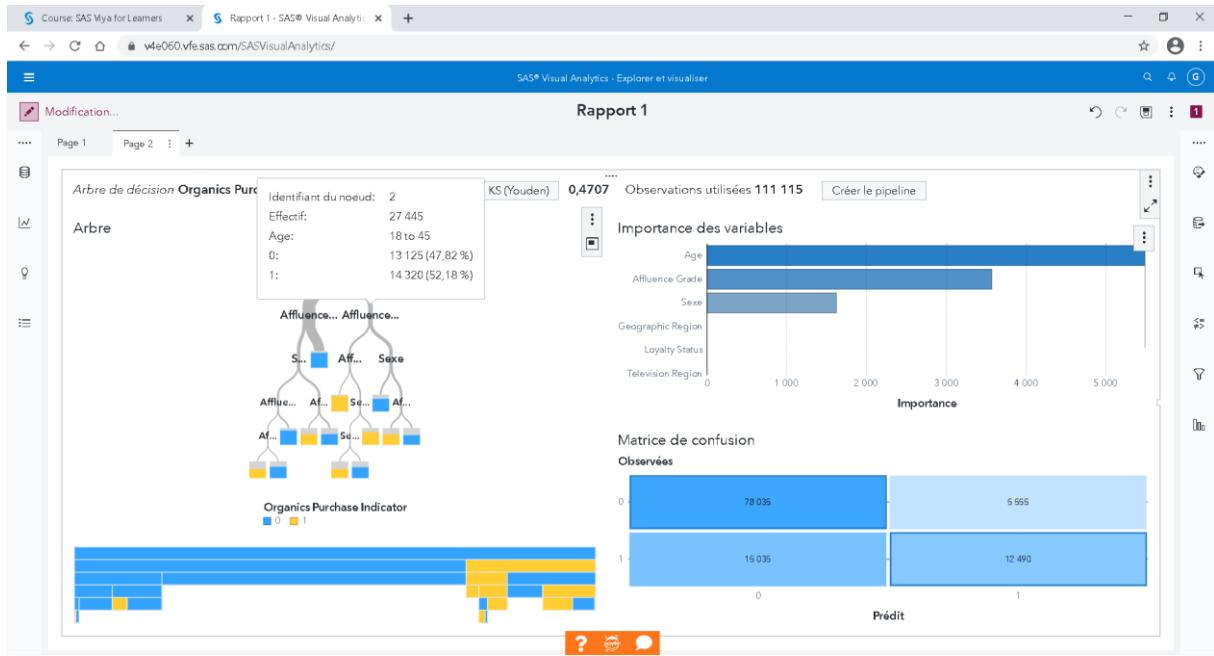
Plus la taille du trait est importante, plus la population est grande.

La racine de l'arbre se trouve en haut. Il est découpé en branche. Les terminaisons sont les feuilles. Plus une feuille est orange, plus la proportion de clients qui achètent des produits bio est importante.

Pour interpréter l'arbre, on peut glisser la souris dessus. Lorsque l'on s'arrête, le détail s'affiche.

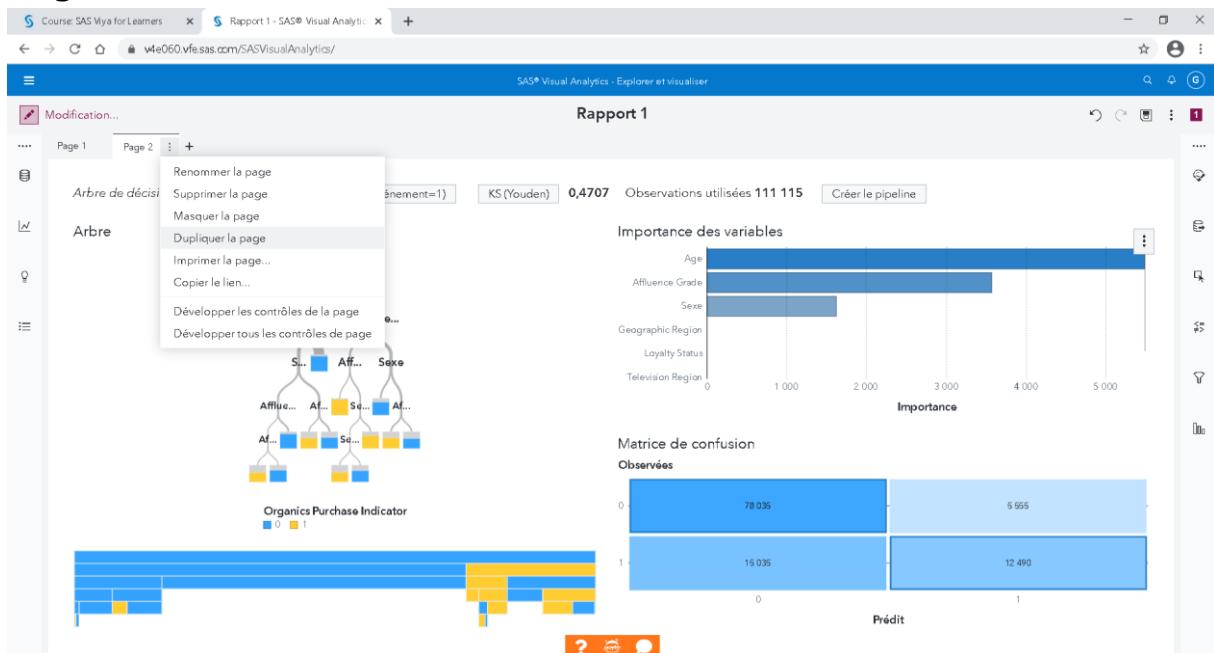


Lorsque la souris est en haut de l'arbre, l'identifiant de la racine est le zéro. Il y a 111 115 personnes dans la base. 75,23% de clients n'achètent pas. 24,77% ont acheté du bio soit 27 525 personnes. (Vous pouvez avoir des chiffres légèrement différents).

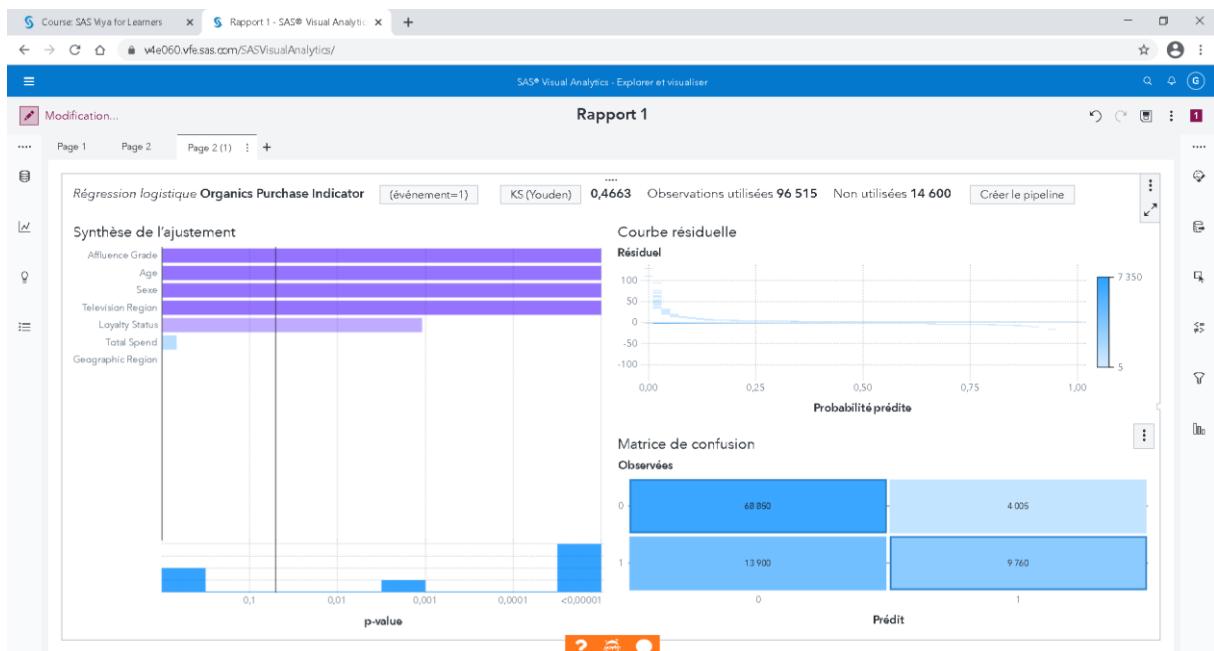
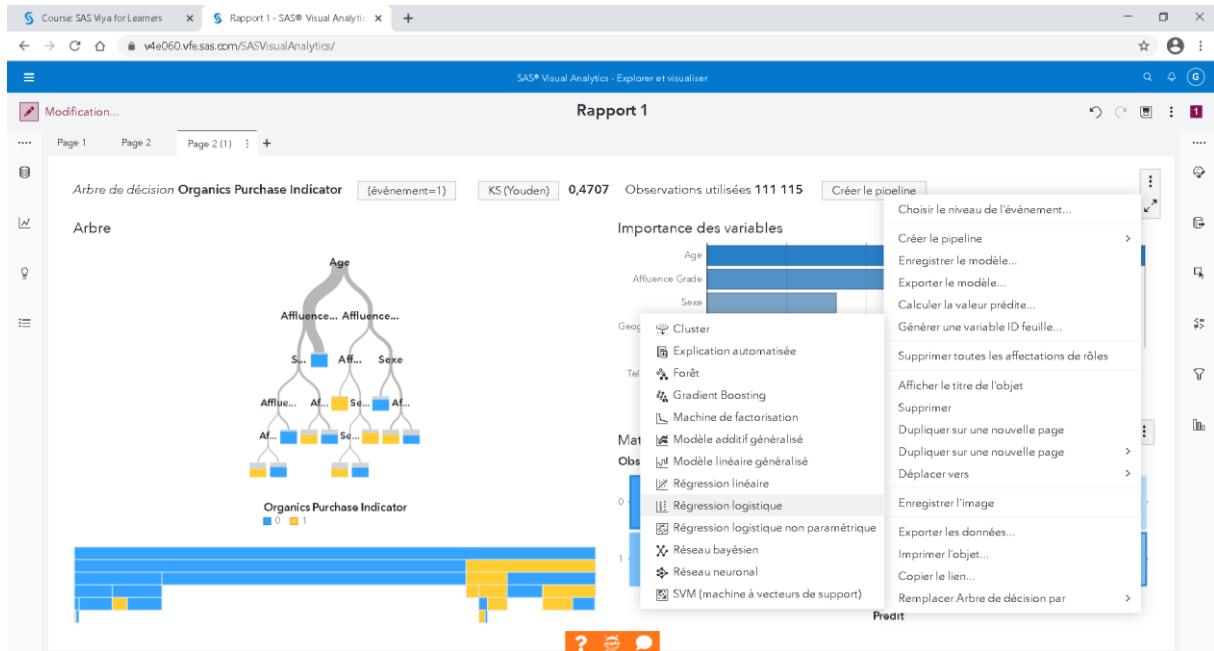


Si l'on descend légèrement la souris sur la branche de droite, sur le second nœud ; il y a 27 445 clients de moins de 45 ans et parmi eux 52,18% de clients qui achètent du bio.

Régression

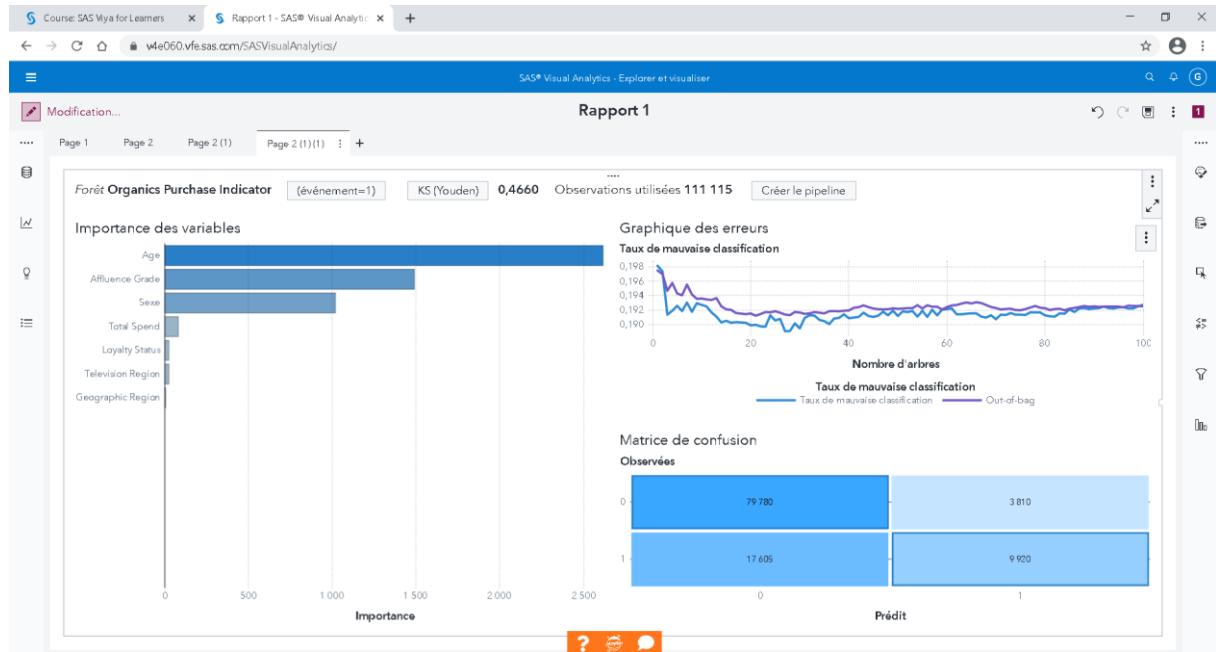


On peut dupliquer la page et remplacer l'arbre par une régression logistique.



Forêt

On peut dupliquer la page et remplacer la régression logistique par une forêt.

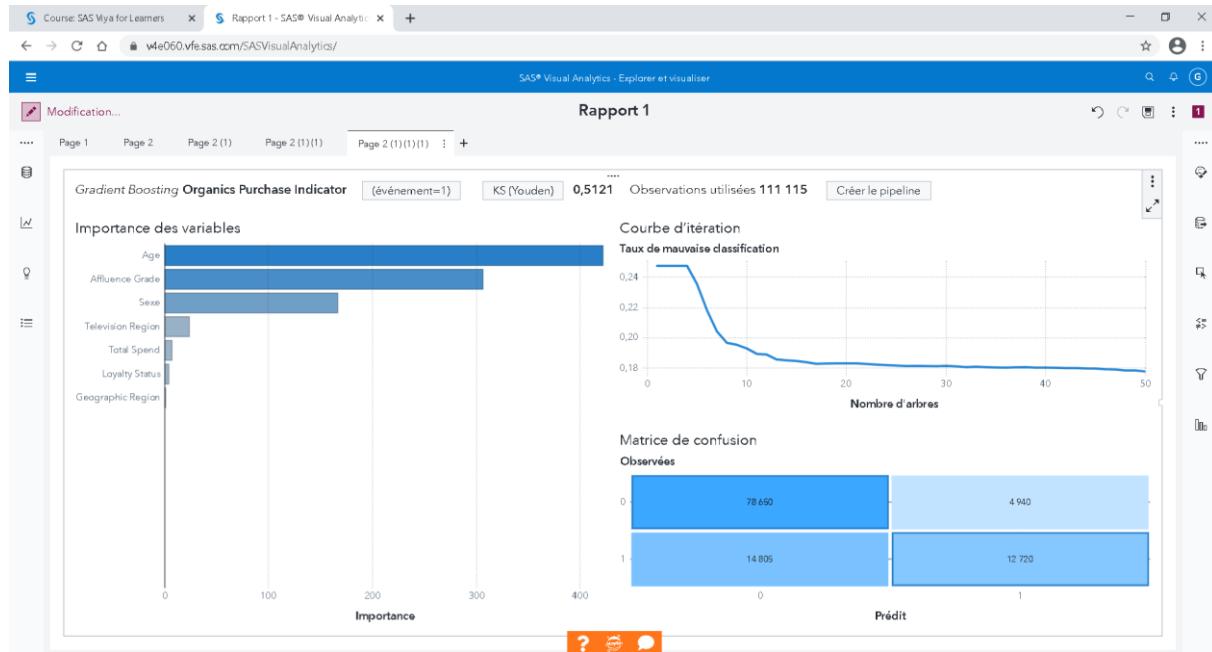


Comme son nom l'indique, une forêt décisionnelle est un ensemble d'arbres qui forment une forêt. Le principe est de faire un tirage aléatoire avec remise sur les individus afin de créer N tables semblables et de créer N arbres sur ces différentes tables en réduisant le nombre de variables aléatoirement aussi. N arbres apprennent donc sur N tables différentes et on en fait généralement tout simplement la moyenne.

Ce n'est donc pas un arbre mais plusieurs et souvent cela fonctionne mieux.

Gradient Boosting

On peut dupliquer la page et remplacer la forêt par un Gradient Boosting.



On obtient les résultats du Gradient Boosting pour les mêmes paramètres.

Le principe du Gradient Boosting est assez proche de la forêt.

On affecte le même poids à tous les individus. Ils ont donc tous la même probabilité d'être tirés.

On commence par faire un tirage aléatoire avec remise sur les individus et on construit un premier arbre.

La performance de ce premier modèle est mesurée pour chaque individu. On compare la prédiction de l'arbre que l'on vient de construire avec les vraies valeurs.

On corrige les poids de tous les individus en augmentant relativement ceux pour lesquels le modèle était mauvais et l'on diminue le poids des individus relativement bien prédit.

On fait un nouveau tirage aléatoire avec remise pour l'arbre suivant en ayant une plus grande chance de tirer les exemples difficiles et moins de chance de sélectionner des exemples faciles.

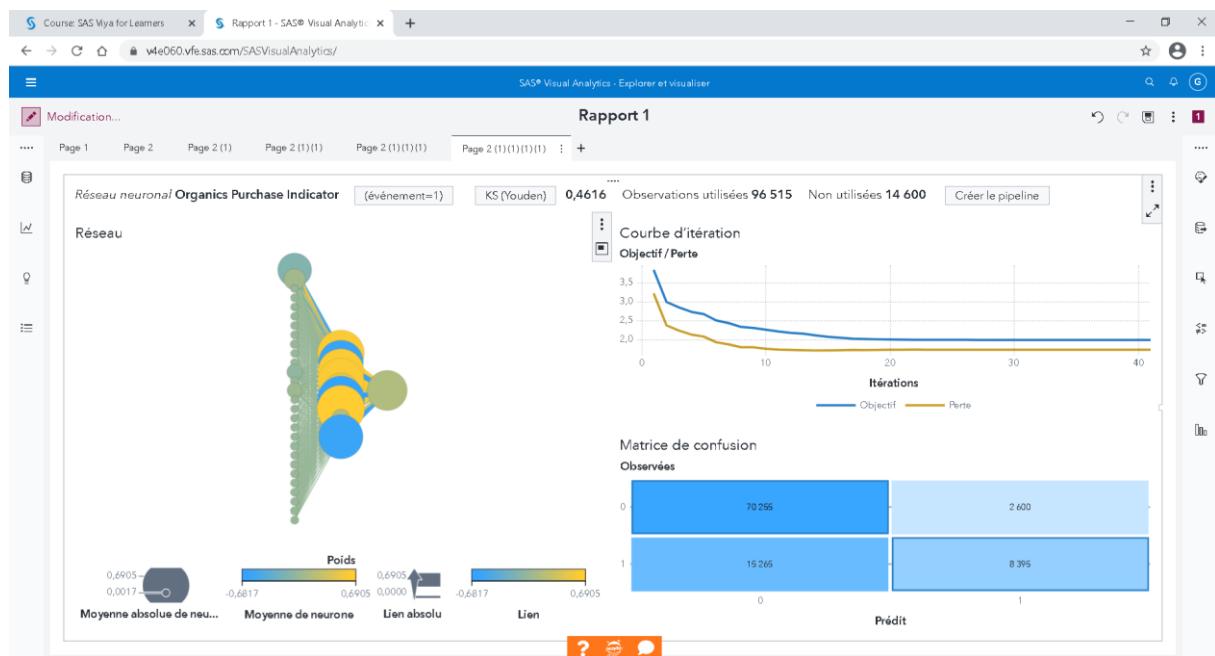
Ce nouvel arbre apprendra donc plus les exemples difficiles et moins les faciles, en fonction de la performance de son prédecesseur.

On recommence de processus plusieurs fois.

Donc, comme pour la forêt, au lieu d'un arbre, on en a plein et l'on fait la moyenne.

Ce processus où chaque arbre apprend en fonction des erreurs de son prédecesseur est souvent très efficace.

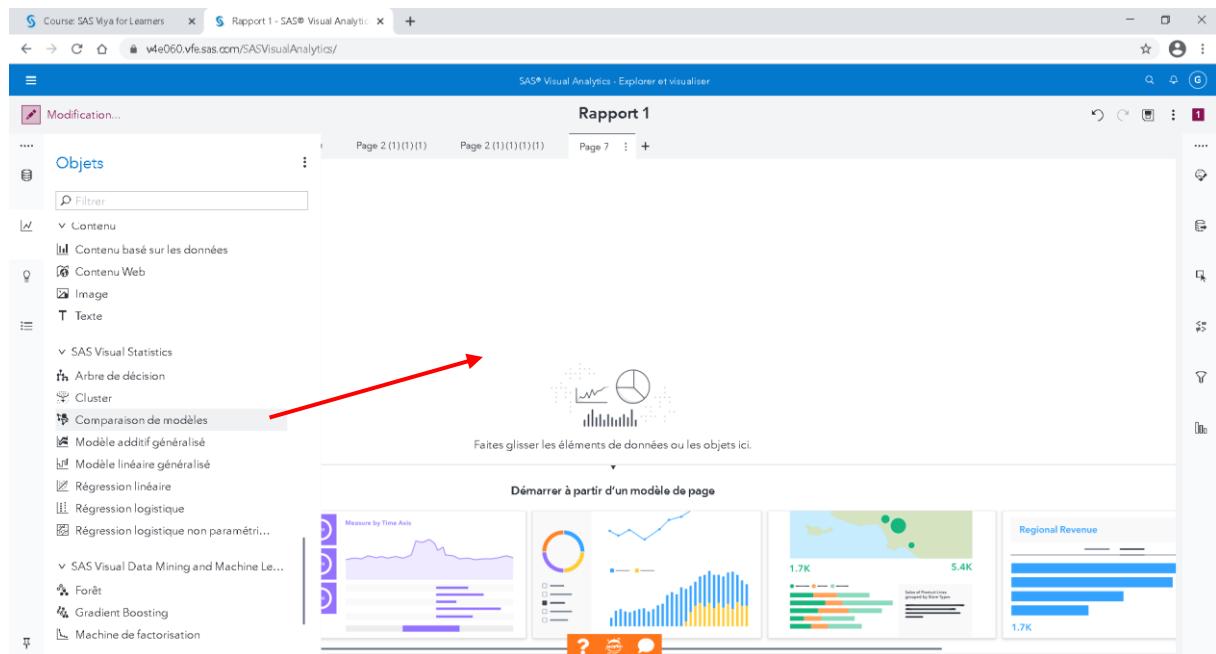
Réseau neuronal



De même, on peut dupliquer notre page et remplacer le Gradient Boosting par un réseaux de neurone.

Comparaison de modèle

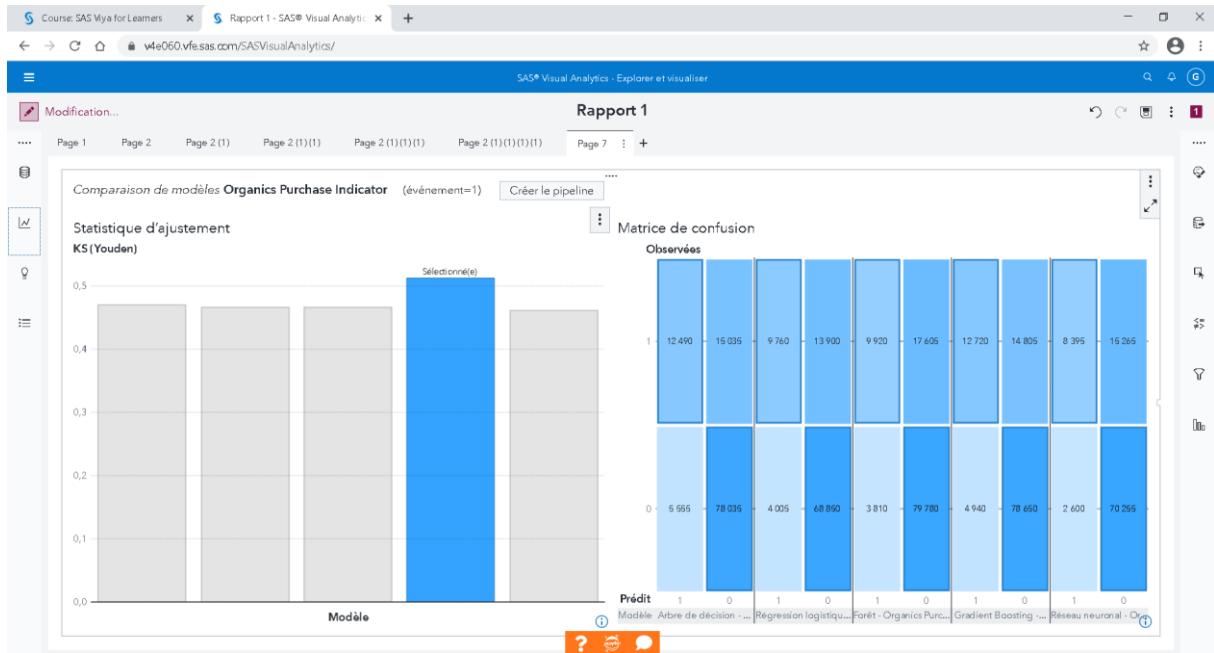
Ajouter une autre page pour faire une comparaison de modèle.



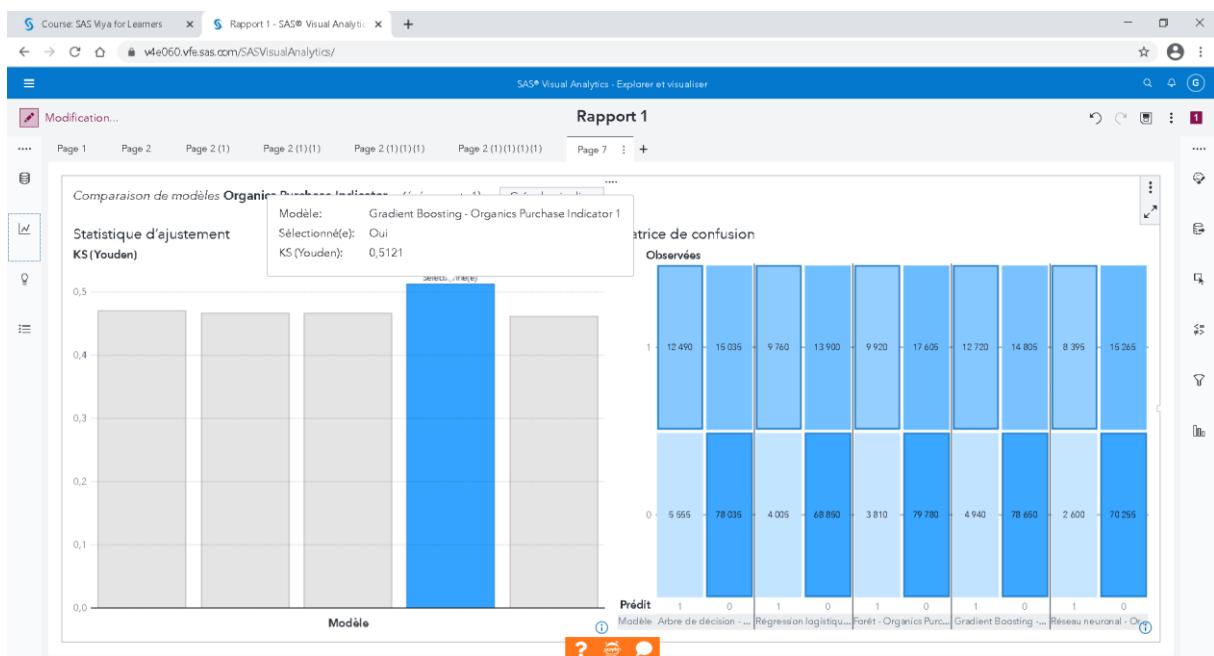
The screenshot shows the SAS Visual Analytics interface with the title "Rapport 1". On the left, there is a sidebar titled "Objets" (Objects) with a list of available objects. A red arrow points to the "Comparaison de modèles" (Model Comparison) item, which is highlighted with a blue background. To the right of the sidebar, there is a preview area showing various data visualizations like charts and maps. Below the sidebar, there is a modal dialog titled "Ajouter une comparaison de modèles" (Add a model comparison). This dialog contains several configuration fields: "Données:" set to "BIGORGANICS", "Partition:" set to "(aucune partition)", "Réponse:" set to "Organics Purchase Indicator", "Niveau de l'événement:" set to "1", and "Groupe BY:" set to "(aucun)". Below these fields is a section titled "Modèles disponibles:" (Available models) which lists several predictive models with checkboxes: "Sélectionner tout" (Select all) is checked, followed by five other models: "Arbre de décision - Organics Purchase Indicator 1", "Régression logistique - Organics Purchase Indicator 1", "Forêt - Organics Purchase Indicator 1", "Gradient Boosting - Organics Purchase Indicator 1", and "Réseau neuronal - Organics Purchase Indicator 1". At the bottom of the dialog are "OK" and "Annuler" buttons.

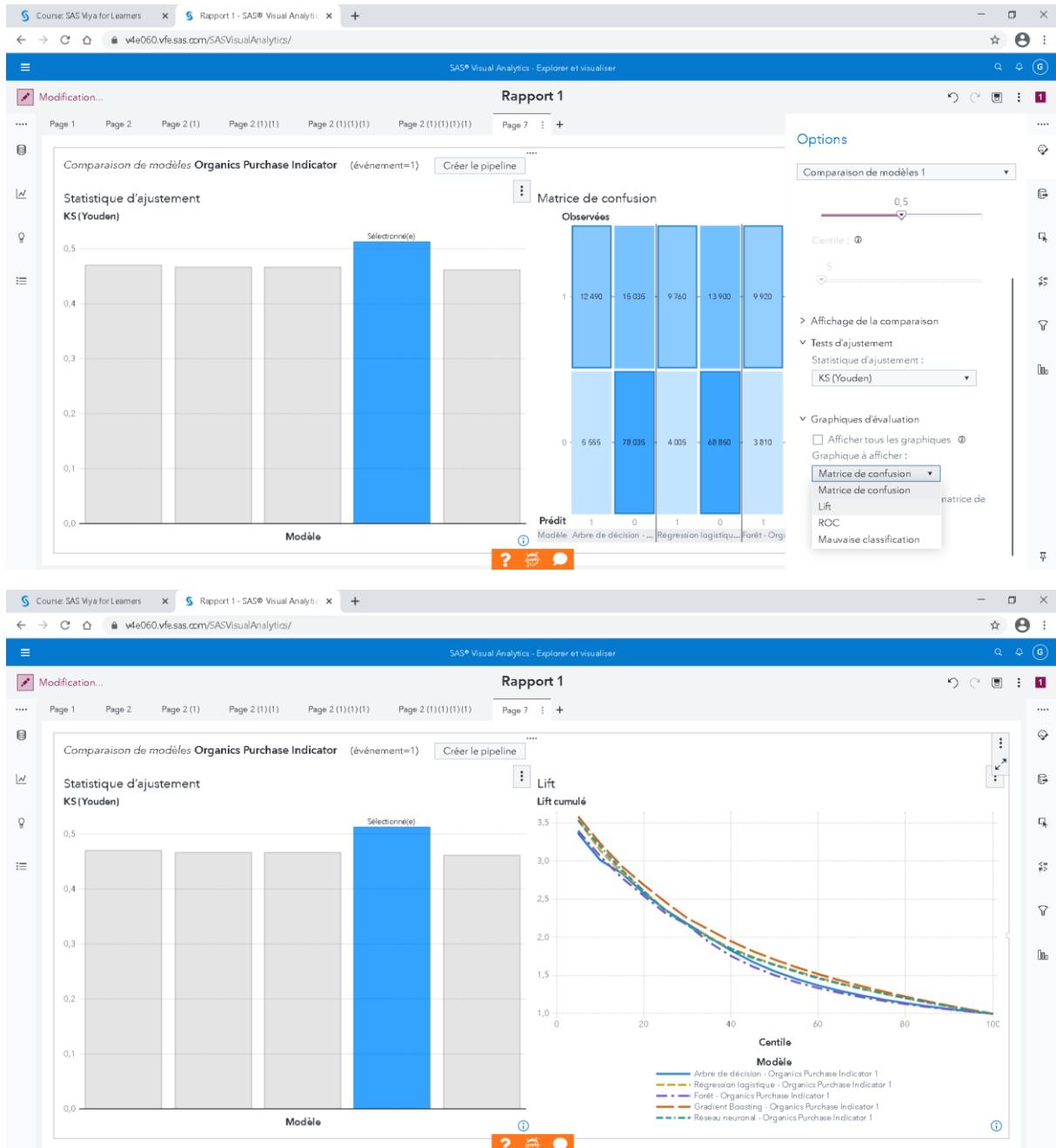
Sélectionner tous les modèles.

OK



D'après la statistique du KS (Youden), le meilleur modèle dans notre cas est le Gradient Boosting.





Sur la courbe du levier (lift), c'est aussi le Gradient Boosting.

L'axe horizontal de la courbe de lift est l'ordonnancement des individus depuis celui ayant la probabilité la plus forte d'acheter un produit bio à 0% jusqu'à 100% où se trouve le client ayant la probabilité la plus faible d'achat.

L'axe vertical, le lift, représente le coefficient multiplicateur du taux de retour.

En positionnant la souris sur le point le plus haut, on peut lire que le Gradient Boosting à un lift de 3,6397 à 5%. C'est-à-dire que si l'on sélectionne les 5% ayant la probabilité la plus forte d'acheter d'après notre modèle, on peut espérer multiplier le taux de retour de base par 3,6397.

Il y a 25% des clients qui achètent des produits bio dans notre base. Pour rappel, nous avons vu cela dans l'exploration dans la première page de notre rapport. De même, ce pourcentage se trouvait à la racine de l'arbre.

Donc, si l'on sélectionne les 5% des clients ayant la probabilité la plus forte d'acheter, $3,6377 \times 25 = 90,94\%$ des clients devraient acheter.

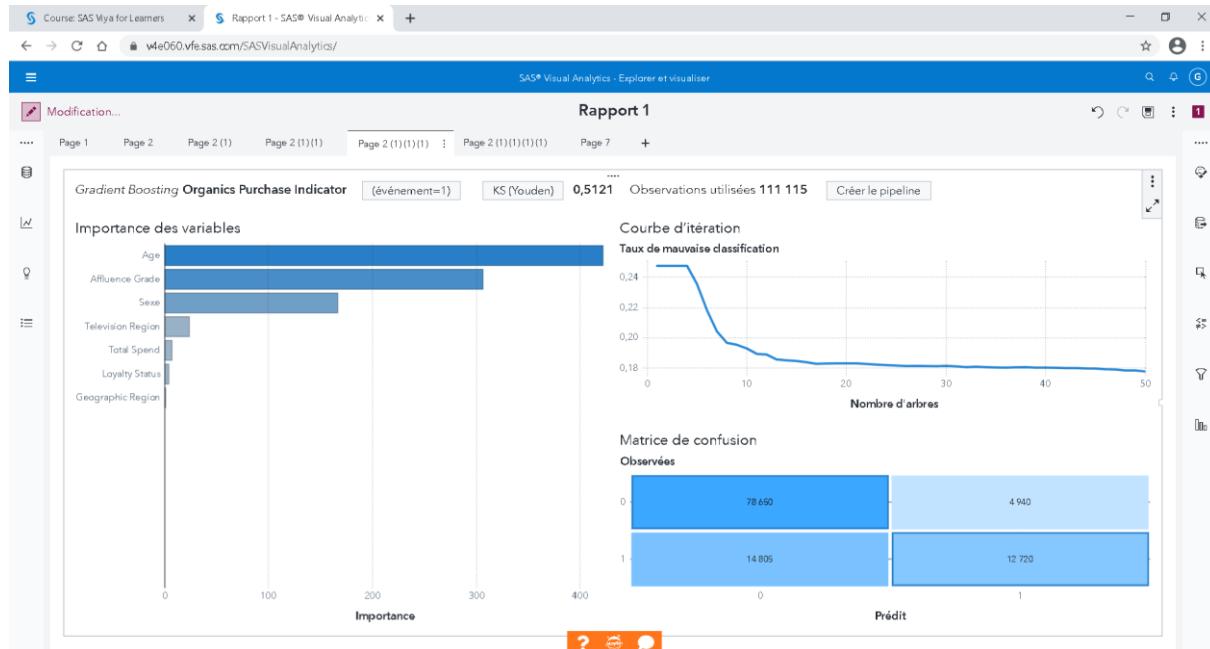
Dans notre cas, le coût moyen d'une communication est de 2€ et la marge nette moyenne est de 5€.

Si nous avons 10 000 personnes à qui nous envoyons une communication à 2€ et que seulement 25% achètent pour une marge net moyenne de 5€, nous perdons de l'argent.

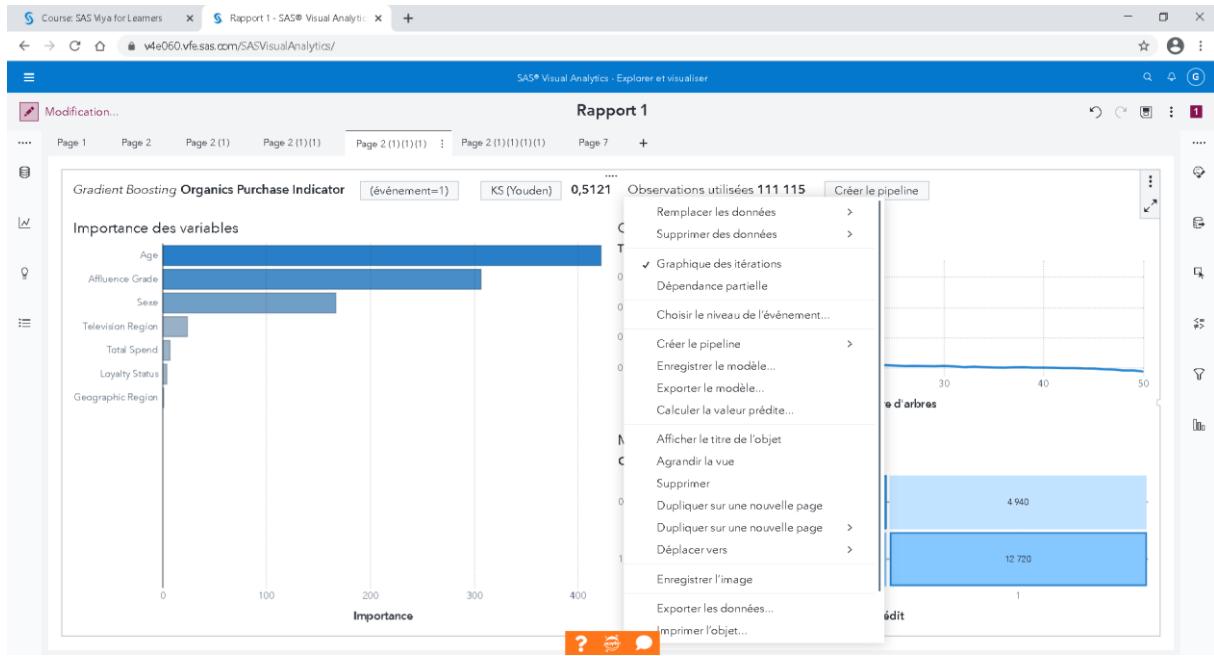
$$10\ 000 (-2€ + 5€ \times 25\%) =$$

Ayant un Gradient Boosting qui a appris, sur les 5% de notre nouvelle base ayant la proba la plus forte, le taux de retour ne sera plus de 25% mais de 90%.

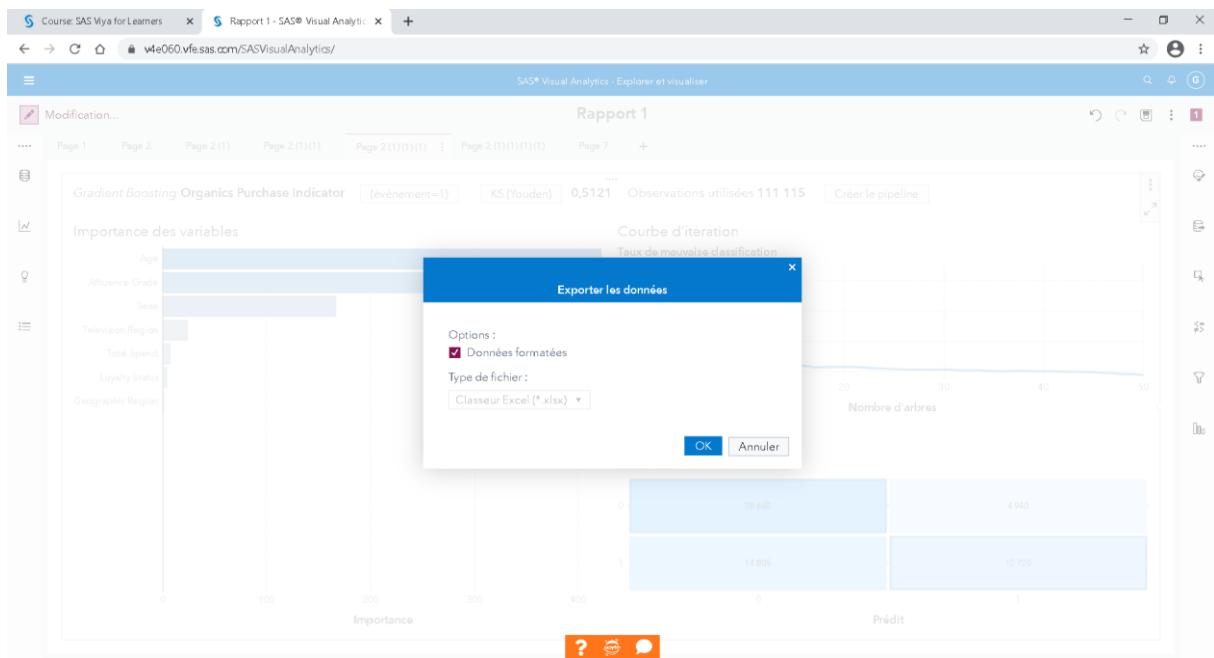
$$10\ 000 \times 5\% (-2€ + 5€ \times 25\% \times 3,6377) =$$



Revenons sur la page 2 (1) (1) du Gradient Boosting.



Clic-droit → Exporter les données



OK

Ouvrir le fichier téléchargé avec Excel

OK

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	Centile	Observations	Evénements	Modèle	Best	Modèle cumulé	Best cumulé														
2	5,00	5556	4931	3,5829 ####		3,5829	4,0371														
3	10,00	5556	3971	2,8854 ####		3,2342	4,0371														
4	15,00	5556	3202	2,3266 ####		2,9316	4,0371														
5	20,00	5556	2695	1,9582 ####		2,6883	4,0371														
6	25,00	5556	2151	1,5629 ####		2,4632	4,0000														
7	30,00	5556	1640	1,1914 ####		2,2512	3,3333														
8	35,00	5556	1650	1,1992 ####		2,1009	2,8571														
9	40,00	5556	1234	0,8969 ####		1,9504	2,5000														
10	45,00	5556	1066	0,7743 ####		1,8198	2,2222														
11	50,00	5556	985	0,7157 ####		1,7094	2,0000														
12	55,00	5556	835	0,6067 ####		1,6091	1,8182														
13	60,00	5556	735	0,5341 ####		1,5195	1,6667														
14	65,00	5556	624	0,4531 ####		1,4375	1,5385														
15	70,00	5556	500	0,3636 ####		1,3608	1,4286														
16	75,00	5556	426	0,3095 ####		1,2907	1,3333														
17	80,00	5556	325	0,2361 ####		1,2248	1,2500														
18	85,00	5556	231	0,1676 ####		1,1626	1,1765														
19	90,00	5556	194	0,1412 ####		1,1059	1,1111														
20	95,00	5556	85	0,0618 ####		1,0509	1,0526														
21	100,00	5556	45	0,0527 ####		1,0000	1,0000														
22																					
23																					
24																					
25																					
26																					
27																					
28																					
29																					

Aller dans la feuille du Lift

Activer la modification

Créer une colonne du ROI

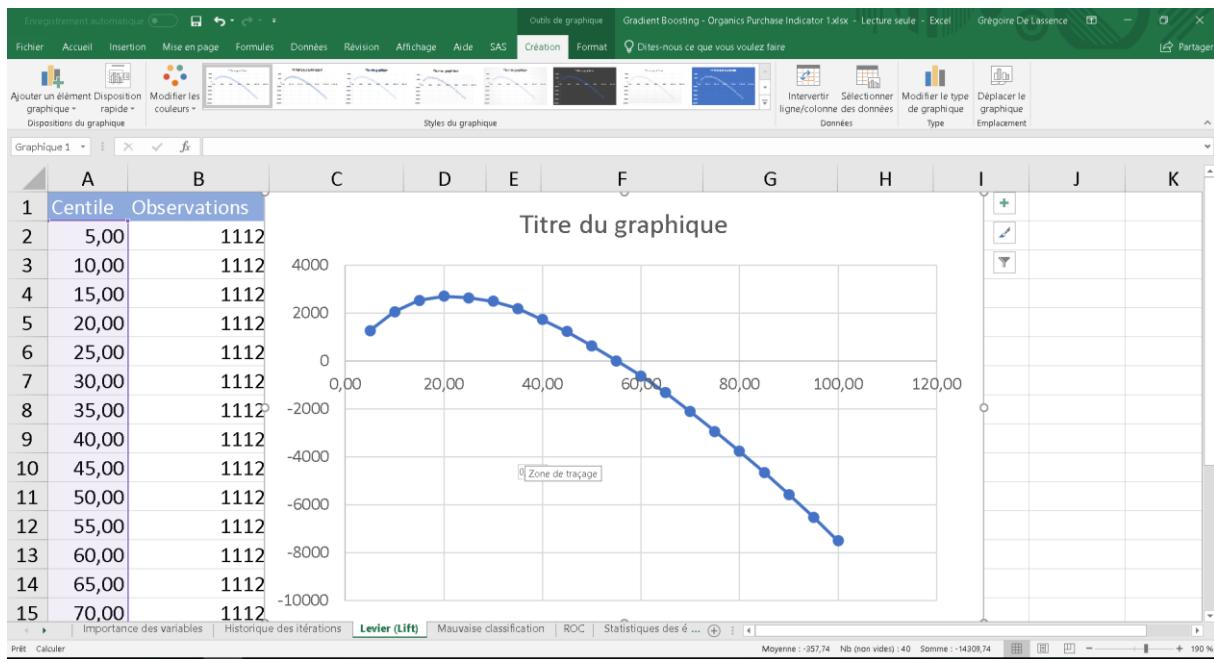
La formule est

$$=10000*A2/100*(-2+5*F2*25/100)$$

A	B	C	D	E	F	G	H	I	J	K
1	Centile	Observations	Evénements	Modèle	Best	Modèle cumulé	Best cumulé	ROI		
2	5,00	1112		1002	3,6397 ####		3,6397	4,0400	1274,83	
3	10,00	1112		792	2,8767 ####		3,2582	4,0400	2072,74	
4	15,00	1112		645	2,3446 ####		2,9537	4,0400	2538,15	
5	20,00	1112		519	1,8856 ####		2,6866	4,0400	2716,62	
6	25,00	1112		410	1,4896 ####		2,4472	4,0000	2647,59	
7	30,00	1112		376	1,3645 ####		2,2668	3,3333	2500,39	
8	35,00	1112		311	1,1299 ####		2,1044	2,8571	2206,57	
9	40,00	1112		237	0,8626 ####		1,9491	2,5000	1745,69	
10	45,00	1112		217	0,7893 ####		1,8203	2,2222	1238,99	
11	50,00	1112		180	0,6530 ####		1,7035	2,0000	647,139	
12	55,00	1112		163	0,5922 ####		1,6025	1,8182	17,257	
13	60,00	1112		162	0,5886 ####		1,5180	1,6667	-614,896	
14	65,00	1112		133	0,4832 ####		1,4384	1,5385	-1312,9	
15	70,00	1112		90	0,3259 ####		1,3589	1,4286	-2109,21	

On peut glisser cette formule sur la colonne du ROI

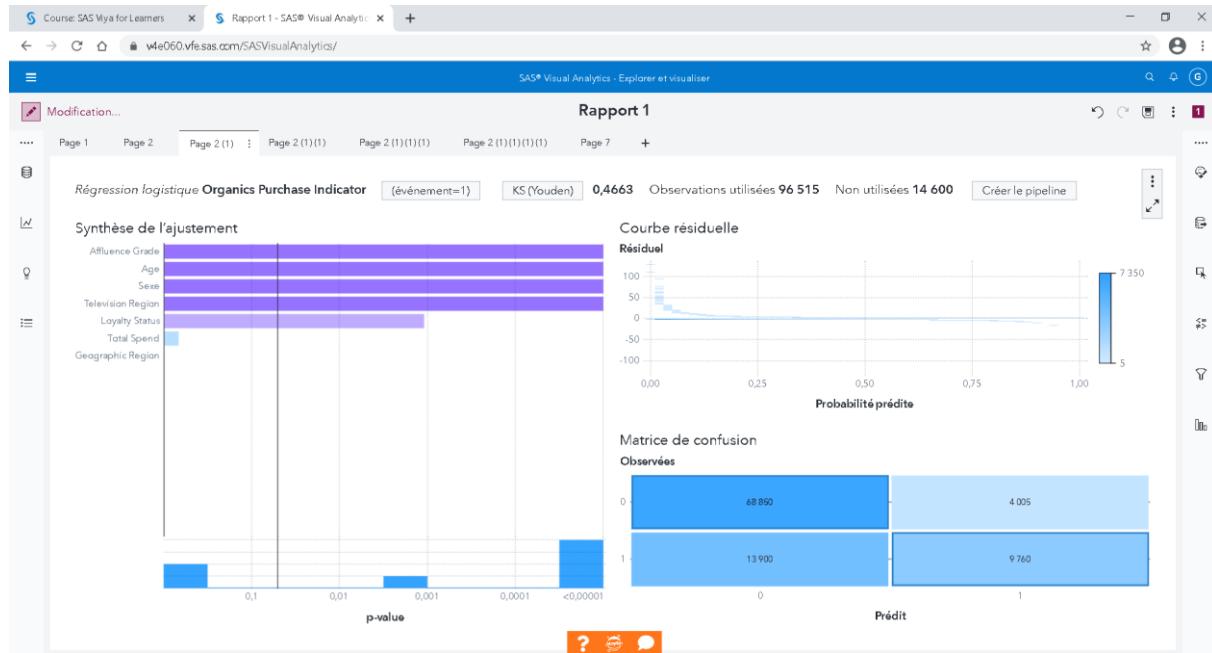
Et éventuellement faire un graphique



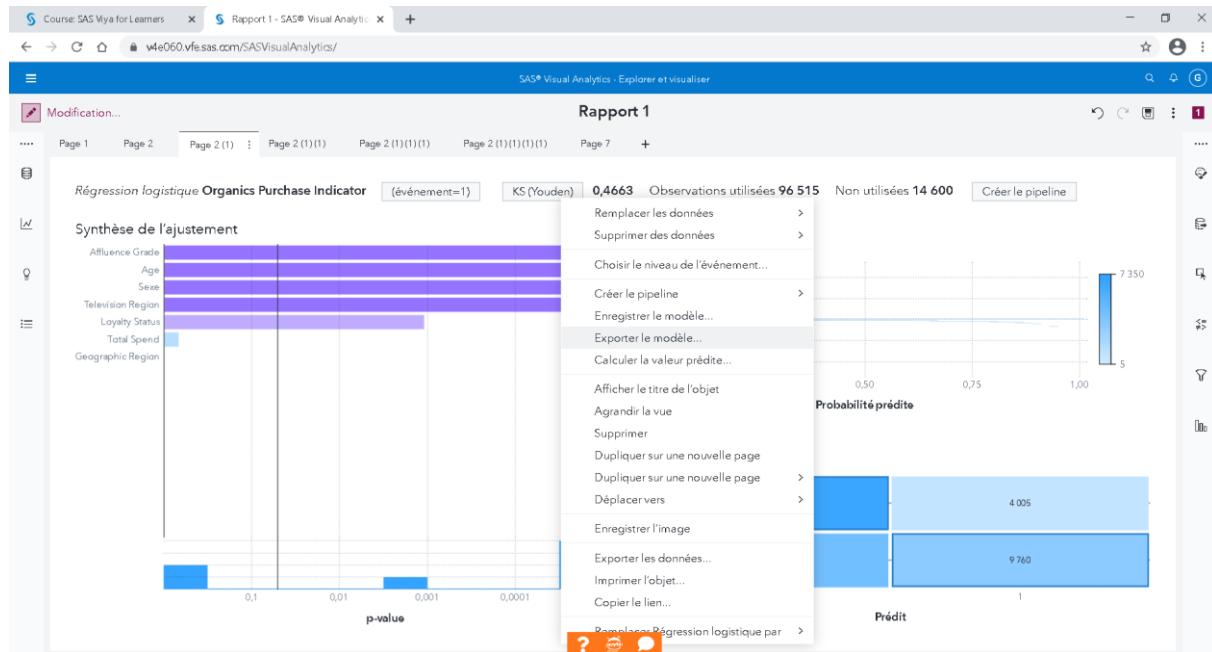
Le maximum du Rol est à 20%.

Si l'on sélectionne les 20% ayant la probabilité la plus forte d'acheter d'après notre modèle, on pourrait gagner 2 716€.

Scoring avec la régression



Revenir sur la page de la régression logistique, sur les 3 petits points en haut à droite ou clic droit au milieu, exporter le modèle.



SAS® Visual Analytics - Explorer et visualiser

Exporter le modèle

```

169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202

```

Créer le pipeline

7 350

4 005

9 760

?

Annuler

Sélectionner tout le code généré et le copier.

Vous pouvez l'enregistrer dans un fichier texte sur votre poste en local.

« Annuler » pour revenir à SAS Visual Analytics.

SAS® Visual Analytics - Explorer et visualiser

Rapport 1

Analytics Purchase Indicator (événement=1) KS (Youden) 0,4663 Observations utilisées 96 515 Non utilisées 14 600 Créer le pipeline

Course résiduelle

Résiduel

Matrice de confusion

Observées

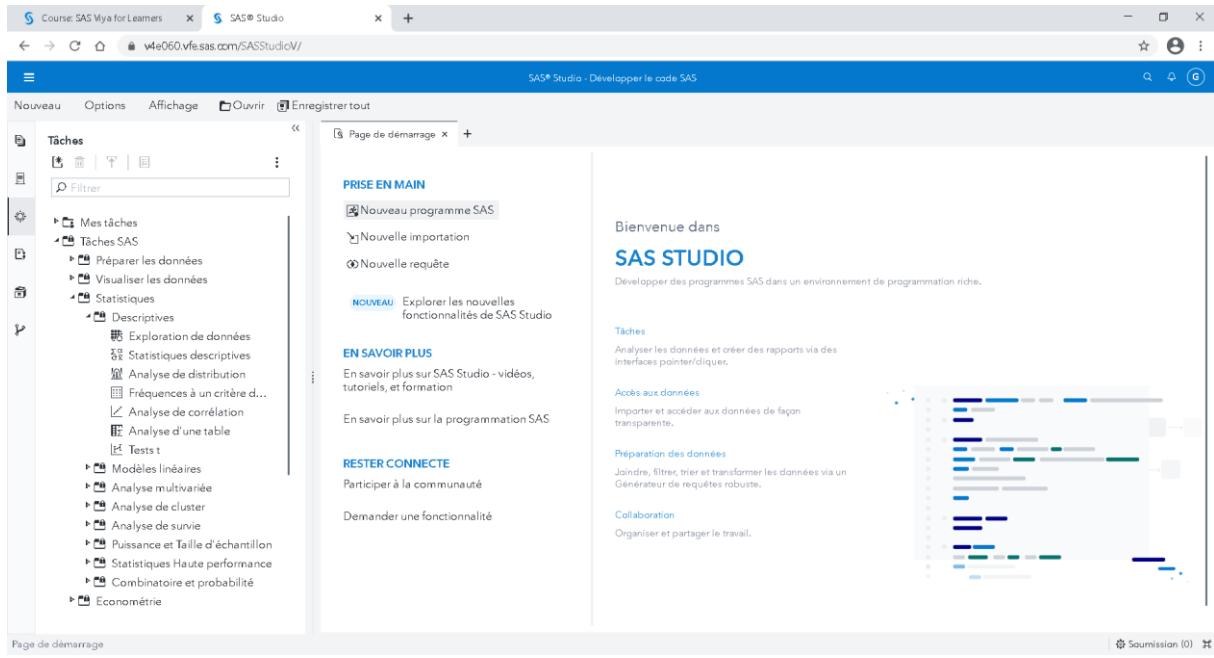
0	68 890	4 005
1	13 900	9 760

Prédit

?

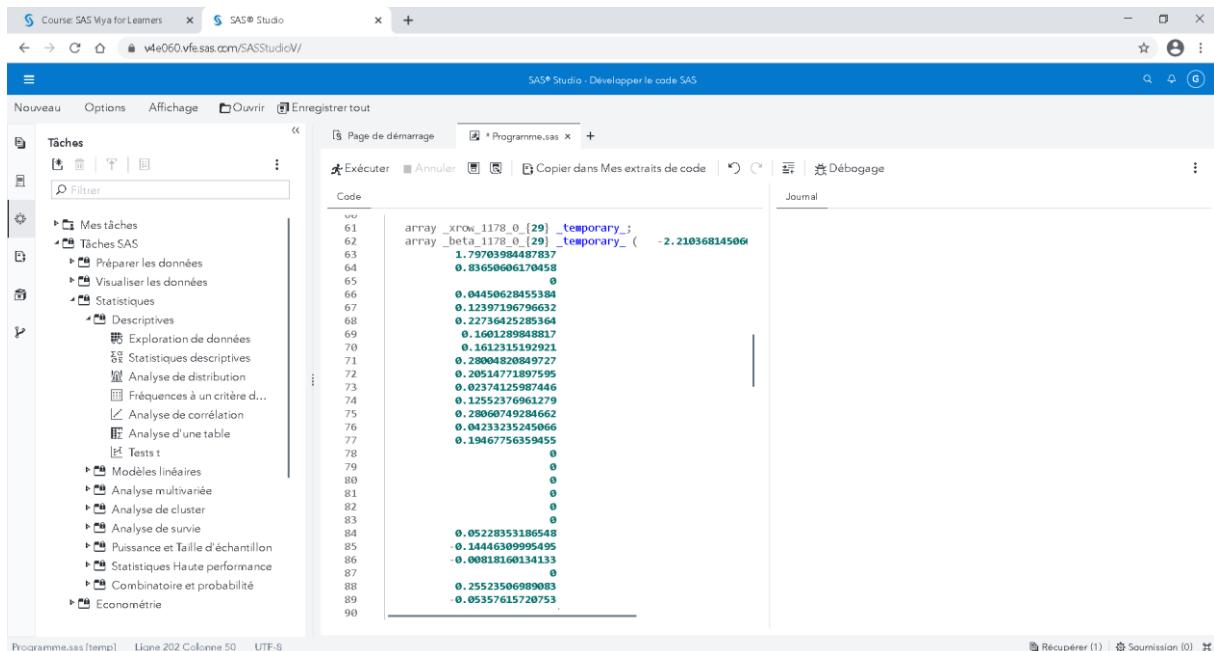
Annuler

Aller dans « Développer le code SAS »



Cliquer sur « Nouveau programme SAS »

Coller le code que vous avez copié



Tout à la fin, retour à la ligne, taper « run; »

```

1/*  
2label P_TargetBuy = 'Prédit : TargetBuy=1';  
3if (_badval_ eq 0) and not missing(_linp_) then do;  
4  if (_linp_ > 0) then do;  
5    P_TargetBuy1 = 1 / (1+exp(-_linp_));  
6  end; else do;  
7    P_TargetBuy1 = exp(_linp_) / (1+exp(_linp_));  
8  end;  
9  P_TargetBuy0 = 1 - P_TargetBuy1;  
10 if P_TargetBuy1 >= 0.5 then do;  
11   _TargetBuy = _levels_1178_[1];  
12 end; else do;  
13   _TargetBuy = _levels_1178_[2];  
14 end;  
15 end; else do;  
16   _linp_ = .;  
17   P_TargetBuy1 = .;  
18   P_TargetBuy0 = .;  
19 end;  
20 /* VA DROP*/ drop 'va_c_Sexe'n '_TargetBuy'n 'P_Targ  
length '_TargetBuy_1178'n $32;  
'P_TargetBuy1_1178'n '_TargetBuy'n;  
'P_TargetBuy0_1178'n '_TargetBuy0'n;  
/*  
run;
```

Tout au début,

```

1 caslib _all_ assign;  
2  
3 data work.MaTableBigOrganicsScore;  
4 set tundata.bigorganics;  
5 /*  
6 L'instruction options ci-dessous doit être placée  
7 avant l'étape DATA lors de la soumission du code.  
8 -----  
9 options VALIDMEMNAME=EXTEND VALIDVARNAME=ANY;  
10 */  
11 /* Code de scoring SAS générée  
12 Date : 29Dec2020:03:50:05  
13 Paramètres régionaux : fr_FR  
14 Type du modèle : Régression logistique  
15 Variable quantitative : DemAff1(Affluence Grade)  
16 Variable quantitative : DemAge(Age)  
17 Variable quantitative : PromSpend(Total Spend)  
18 Variable quantitative : va_c_Sexe(Sexe)  
19 Variable qualitative : DemRegion(Television Region)  
20 Variable qualitative : DemAge(Geographic Region)  
21 Variable qualitative : PromCusts(Loyalty Status)  
22 Variable qualitative : TargetBuy(Organics Purchase Intent)  
23 Variable de réponse : TargetBuy(Organics Purchase Intent)  
24 Distribution : Binaire  
25 Fonction de liaison : Logit  
26 -----  
27 /* Colonne calculée temporaire */  
28 Tmonth * va_c_Sexe as E;
```

Ajouter le code :

```

«
cas;
caslib _all_ assign;
```

```

data work.MaTableBigOrganicsScore;
set tundata.bigorganics;
```

»

Cliquer sur le bouton Exécuter

```

1 cas;
2 caslib _all_ assign;
3
4 data work.MATABLEBIGORGANICSSCORE;
5 set tundata.biorgantics;
6
7 /*-----
8 L'instruction Options ci-dessous doit être placée
9 avant l'étape DATA lors de la soumission du code.
10 -----*/
11 options VALIDMEMNAME=EXTEND VALIDVARNAME=ANY;
12 /*
13 Code de scoring SAS généré
14 Date : 29Dec2020:03:50:05
15 Paramètres régionaux : fr_FR
16 Type du modèle : Régression logistique
17 Variable quantitative: DemAffl(Affluence Grade)
18 Variable quantitative: DemAge(Age)
19 Variable quantitative: PromSpend(Total Spend)
20 Variable qualitative : _va_c_Sexe(Sexe)
21 Variable qualitative : DemReg(Geographic Region)
22 Variable qualitative : PromClass(Loyalty Status)
23 Variable qualitative : TargetBuy/Organics Purchase
24 Variable de réponse : TargetBuy/Organics Purchase
25 Distribution : Binnaire
26 Fonction de liaison : Logit
27
28 /* colonne calculée temporaire */
29 *month * va_c_Couleur *
30 */

```

NOTE: The session CASAUTO connected successfully to Cloud Analytic Ser... |
 NOTE: The SAS option SESSREF was updated with the value CASAUTO.
 NOTE: The SAS macro _SESSREF was updated with the value CASAUTO.

	P_TargetBuy1_1178	P_TargetBuy1
1	0.0293707586	0.9704
2	0.027295427	0.974
3	0.0629150809	0.937
4	0.1050125252	0.8945
5	0.2520535703	0.7475
6	0.019626544	0.980
7	0.2070148572	0.7925
8	0.2502124163	0.7497
9	0.1271967955	0.8728
10	.	.
11	0.1795642928	0.8204
12	.	.
13	0.5997737305	0.2277

Dans les données de sortie, on retrouve une colonne en plus P_TargetBuy1***** qui est par abus de langage la probabilité que le client achète un produit bio.

Dans données de sortie, On peut trier de manière décroissante selon la colonne de la probabilité la plus forte d'acheter P_TargetBuy1

SAS Studio - Développer le code SAS

Exécuter Annuler Copier dans Mes extraits de code Débogage 29 déc. 2020 à 10:22:14

Nouveau Options Affichage Ouvrir Enregistrer tout

Filtrer

Code

```

1 cas;
2 caslib _all_ assign;
3
4 data work.MatTablebigorganicsscore;
5 set tundata.bigorganics;
6
7 /*----- L'instruction Options ci-dessous doit être placée avant l'étape DATA lors de la soumission du code. -----
11 options VALIDMEMNAME=EXTEND VALIDVARNAME=ANY;
12 */
13
14 Code de scoring SAS généré
15 Date : 29dec2020:03:50:05
16 Paramètres régionaux : fr_FR
17 Type du modèle : Régression logistique
18 Variable quantitative: DemAffl(Affluence Grade)
19 variable quantitative: DemAge(Age)
20 variable quantitative: PromSpend(Total Spend)
21 variable qualitative : _va_c_Sexe(Sexe)
22 variable qualitative : DemVReg(Television Region)
23 variable qualitative : PromClass(Loyalty Status)
24 variable qualitative : TargetBuy(Organics Purchase)
25 variable de réponse : TargetBuy(Organics Purchase)
26 Distribution : Binaire
27 Fonction de liaison : Logit
28
29 /* colonne calculée temporaire */
30

```

Journal Données de sortie

WORKMATABLEBIGORGANICSSCORE Colonnes : 16 sur 16

Saisir l'expression

	Tri	getBuy
5	0.2	0.970
6	0.1	0.97
7	0.2	0.970
8	0.2502124163	0.749
9	0.1271967955	0.872
10	.	.
11	0.1795642928	0.820
12	.	.
13	0.2227727209	0.447

Récupérer (1) Soumission (0)

SAS Studio - Développer le code SAS

Exécuter Annuler Copier dans Mes extraits de code Débogage 29 déc. 2020 à 10:22:14

Nouveau Options Affichage Ouvrir Enregistrer tout

Filtrer

Code

```

1 cas;
2 caslib _all_ assign;
3
4 data work.MatTablebigorganicsscore;
5 set tundata.bigorganics;
6
7 /*----- L'instruction Options ci-dessous doit être placée avant l'étape DATA lors de la soumission du code. -----
11 options VALIDMEMNAME=EXTEND VALIDVARNAME=ANY;
12 */
13
14 Code de scoring SAS généré
15 Date : 29dec2020:03:50:05
16 Paramètres régionaux : fr_FR
17 Type du modèle : Régression logistique
18 Variable quantitative: DemAffl(Affluence Grade)
19 variable quantitative: DemAge(Age)
20 variable quantitative: PromSpend(Total Spend)
21 variable qualitative : _va_c_Sexe(Sexe)
22 variable qualitative : DemVReg(Television Region)
23 variable qualitative : PromClass(Loyalty Status)
24 variable qualitative : TargetBuy(Organics Purchase)
25 variable de réponse : TargetBuy(Organics Purchase)
26 Distribution : Binaire
27 Fonction de liaison : Logit
28
29 /* colonne calculée temporaire */
30

```

Journal Données de sortie

WORKMATABLEBIGORGANICSSCORE Colonnes : 16 sur 16

Saisir l'expression

	P_TargetBuy0_1178
1	0.0019383997
2	0.0019383997
3	0.0019383997
4	0.0019383997
5	0.0019383997
6	0.0035599181
7	0.0035599181
8	0.0035599181
9	0.0035599181
10	0.0035599181
11	0.0036305675
12	0.0036305675

Récupérer (1) Soumission (0)

Exercices

Cet exercice consiste à refaire sur la table PVA (ce nom exactement) comme ce cas Organics.

La partie « code » avec SAS Studio n'est pas à faire.

Target_B (Target Gift Flag) est la variable à expliquer. Il faut bien la mettre en catégorie.

Il faut utiliser toutes les variables sauf Target_D* (Target Gift Amount et Target Gift Amount with zero) et l'ID (Control Number).

Pour le calcul du ROI sur la table PVA : Coût moyen du gadget : 2€ - Marge net moyenne : 3€.

A national veterans' organization seeks to better target its solicitations for donation. By soliciting only the most likely donors, less money is spent on solicitation efforts and more money is available for charitable concerns. Solicitations involve sending a small gift to an individual and include a request for a donation. Gifts to donors include mailing labels and greeting cards.

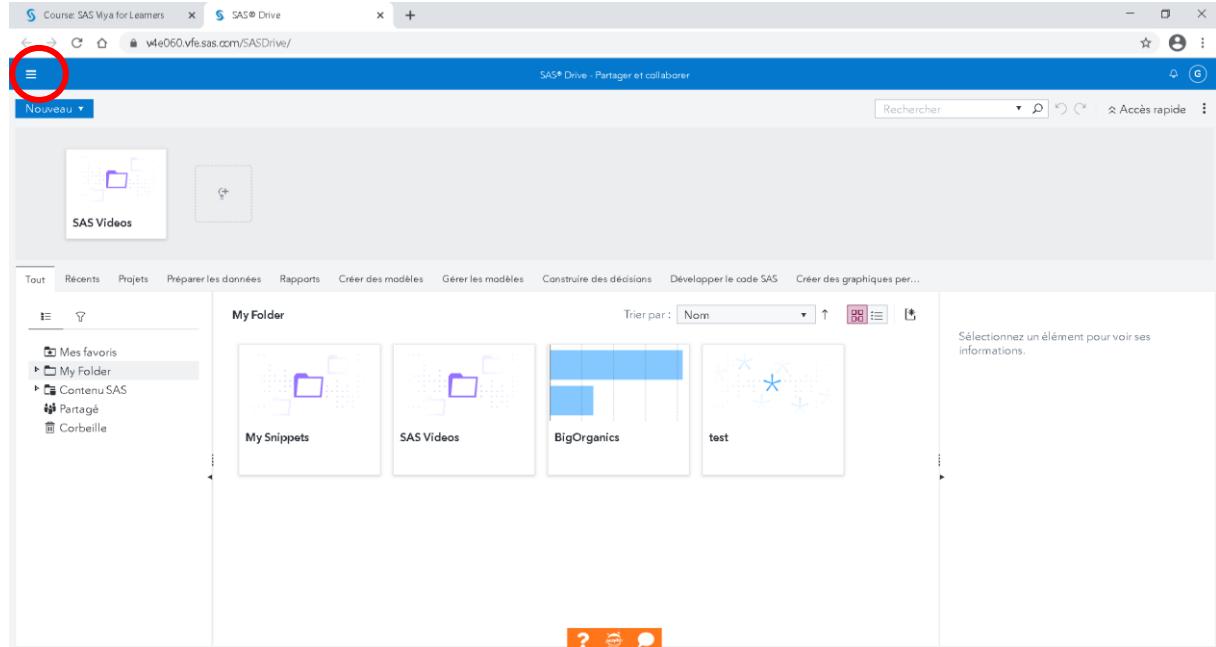
Name	Description
DemAge	Age
DemCluster	Demographic Cluster
DemGender	Gender
DemHomeOwner	Home Owner
DemMedHomeValue	Median Home Value Region
DemMedIncome	Median Income Region
DemPctVeterans	Percent Veterans Region
GiftAvg36	Gift Amount Average 36 Months
GiftAvgAll	Gift Amount Average All Months
GiftAvgCard36	Gift Amount Average Card 36 Months
GiftAvgLast	Gift Amount Last
GiftCnt36	Gift Count 36 Months
GiftCntAll	Gift Count All Months
GiftCntCard36	Gift Count Card 36 Months
GiftCntCardAll	Gift Count Card All Months
GiftTimeFirst	Time Since First Gift
GiftTimeLast	Time Since Last Gift
ID	Control Number
PromCnt12	Promotion Count 12 Months
PromCnt36	Promotion Count 36 Months
PromCntAll	Promotion Count All Months
PromCntCard12	Promotion Count Card 12 Months
PromCntCard36	Promotion Count Card 36 Months
PromCntCardAll	Promotion Count Card All Months

StatusCat96NK	Status Category 96NK
StatusCatStarAll	Status Category Star All Months
TARGET_B	Target Gift Flag
TARGET_D	Target Gift Amount

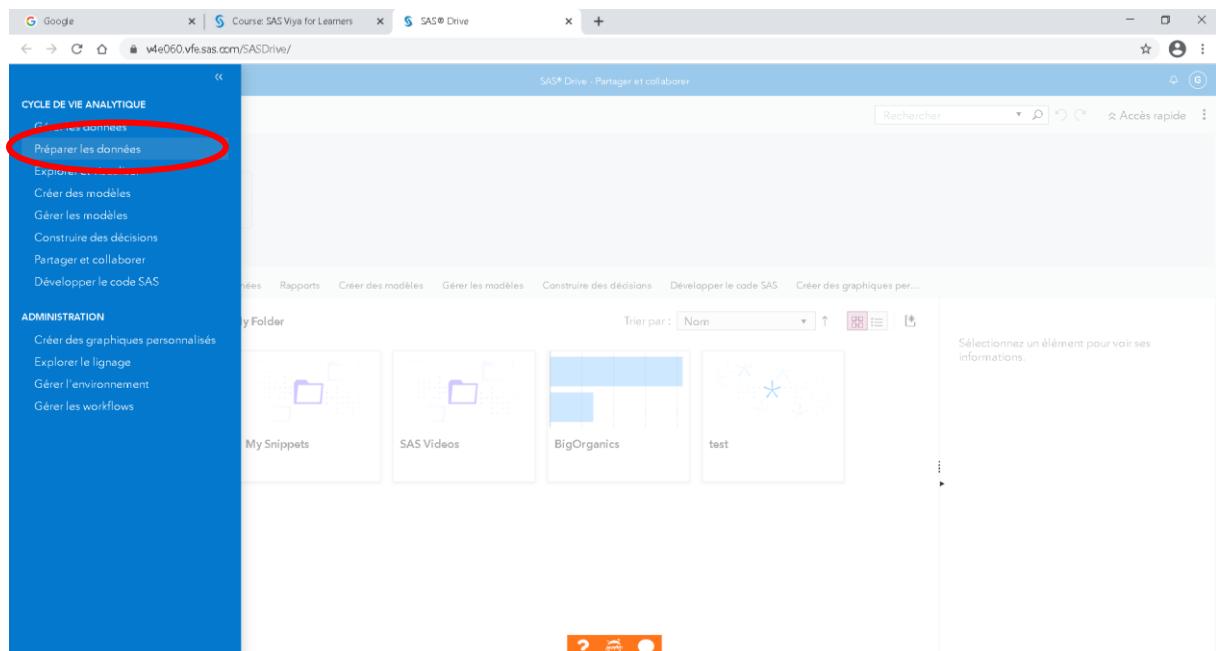
Qualité des données

Voici un petit exercice de qualité de données.

Depuis le Drive

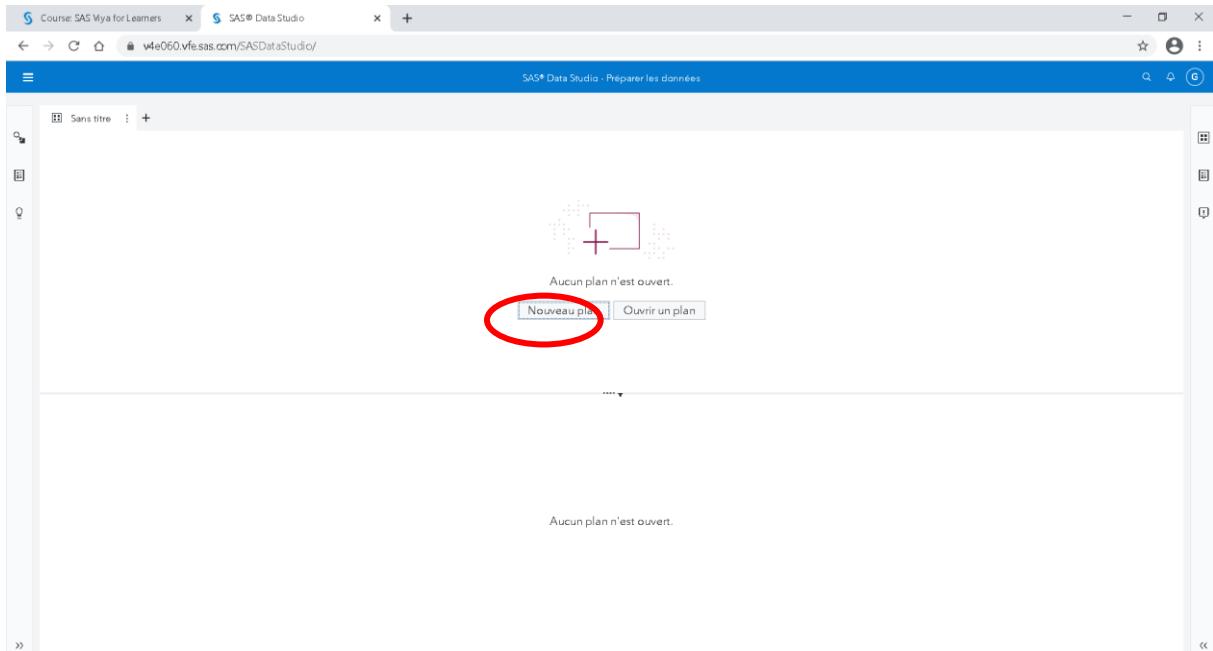


The screenshot shows the SAS Drive interface with a blue header bar. In the top-left corner, there is a menu icon (three horizontal lines) which is circled in red. Below the header, there is a search bar and an 'Accès rapide' (Quick Access) button. The main area displays a 'My Folder' view with four items: 'My Snippets', 'SAS Videos', 'BigOrganics', and 'test'. On the left side, there is a sidebar with navigation links: 'Mes favoris', 'My Folder', 'Contenu SAS', 'Partagé', and 'Corbeille'. At the bottom of the interface, there are three orange buttons: a question mark, a circular arrow, and a speech bubble.

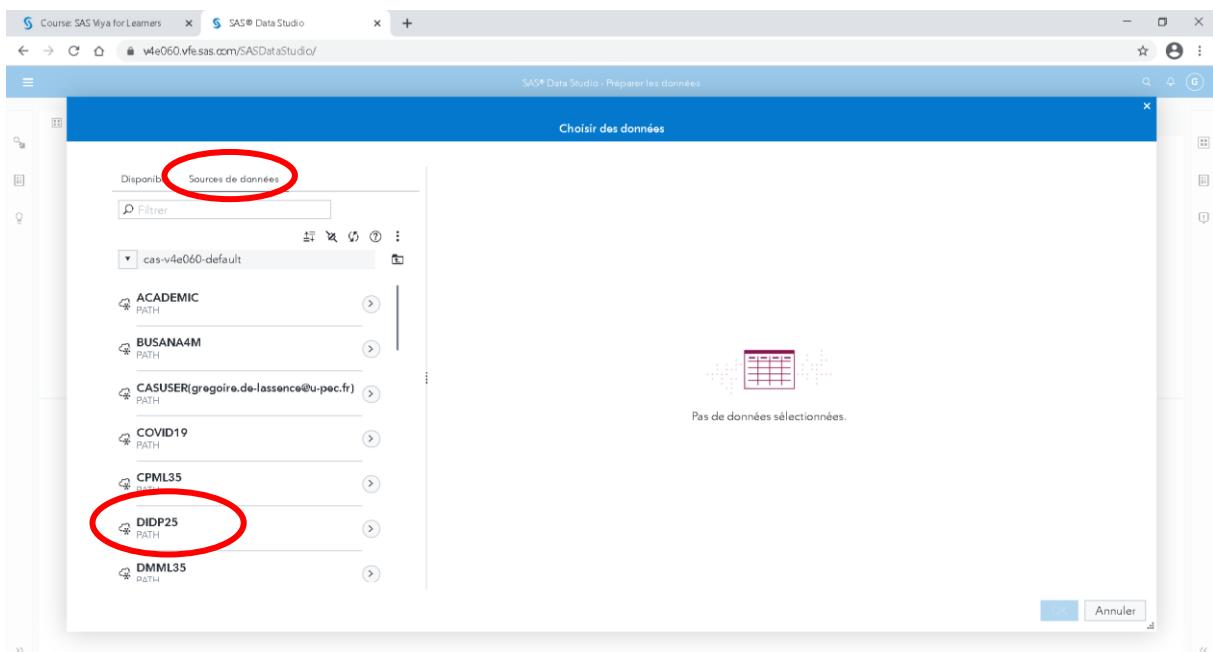


This screenshot is similar to the one above, but it includes a vertical sidebar on the left. The sidebar has two main sections: 'CYCLE DE VIE ANALYTIQUE' and 'ADMINISTRATION'. Under 'CYCLE DE VIE ANALYTIQUE', the 'Préparer les données' link is highlighted with a red circle. Other links in this section include 'Explorer les données', 'Créer des modèles', 'Gérer les modèles', 'Construire des décisions', 'Partager et collaborer', and 'Développer le code SAS'. Under 'ADMINISTRATION', the links are 'Créer des graphiques personnalisés', 'Explorer le lignage', 'Gérer l'environnement', and 'Gérer les workflows'. The rest of the interface is identical to the first screenshot, showing the 'My Folder' view with four items and the same set of orange buttons at the bottom.

Préparer les données



Nouveau Plan



Dans Source de données, dans la bibliothèque DIDP25

SAS® Data Studio - Préparer les données

Choisir des données

Disponible Sources de données

CLIENT_INFO

Détails Jeu d'essai Profil

N°	Nom	Libellé	Type	Lo...	Lo...	Fo...
1	ID	d...	8	12		
2	NAME	c...	100	100		
3	ADDRESS	c...	100	100		
4	CITY	c...	100	100		
5	STATE	c...	100	100		
6	ZIP	c...	100	100		
7	PHONE	c...	100	100		
8	PRODUCT	c...	100	100		
9	PURCHASE_DATE	d...	8	19	D...	

Date du profilage : (aucun)

Colonnes 9 Lignes 30

Taille ..

Libellé : (non disponible)

Emplacement : cas-v4e060-default/DIDP25

Date de création : 8 déc. 2020 20:04:31

Date de modification : 8 déc. 2020 20:04:31

Date du dernier accès : 21 déc. 2020 01:02:38

OK Annuler

Sur la table Client_Info, Jeu d'essai permet d'avoir un aperçu de la table

SAS® Data Studio - Préparer les données

Choisir des données

Disponible Sources de données

CLIENT_INFO

Détails Jeu d'essai Profil

ID	NAME	ADD...	CITY	STATE	ZIP	PHO...	PRO...
1	Sharon ...	3540 Wi...	Redwo...	CA	94065	902-86...	BMV
2	Abigail ...	4211 S ...	Palo Alto	california	94303	860-95...	Base + ...
3	Lou Voss	777 S. H...	Bellevue	WA	98006-...	357-98...	dfPower...
4	Phillip ...	PO Box ...	Skokie	IL	60076-...	360-68...	dfPower...
5	Rob Ru...	3939 Ru...	St Louis	MO	63021	297-07...	dfPower...
6	Margaret...	3001 W ...	Dallas	texas	75247	883-27...	dfPower...
7	Jose Ro...	2800 W ...	Ballwin	MO	63011	721-14...	dfPower...
8	K.Sekeres	5541 C...	Marlow	NH	03456	991-31...	dfPower...
9	Andrie ...	5024 Fa...	Sunnyvale	ca	94089	618-12...	Studio ...
10	Ashley ...	PO. Box...	N. Ridg...	OH	44039	271-47...	BMV
11	Gary Str...	5153 C...	Sunnyvale	calif	94086	350-96...	dfPower...
12	Michelle...	4000 Ea...	Chester...	missouri	63005	426-75...	dfPower...

Lignes d'exemples : 100

OK Annuler

OK

SAS Data Studio - Préparer les données

Plan 1

Table Profil Métagdonnées

CLIENT_INFO

ID	NAME	ADDRESS	CITY	STATE
1	Sharon Mandelb...	3540 Wilshire Blvd	Redwood Shores	CA
2	Abigail Sargent	4211 S Rushford St	Palo Alto	california
3	Lou Voss	777 S. Harbor Bl...	Bellevue	WA
4	Philip Gerstle	PO Box 13507	Skokie	IL

Ajouter une transformation au plan.

Plan

Nom : Plan 1
Modifié : 29/12/2020 12:54

Nous allons commencer par une analyse de genre. Trouver le genre en fonction du prénom.

Double-clic sur « Analyse de genre ».

SAS Data Studio - Préparer les données

Plan 1 *

CLIENT_INFO

La table de sessions n'est pas à jour pour le plan. Exécutez le plan pour me... Lignes de résultats : 100

ID	NAME	ADDRESS	CITY	STATE
1	Sharon Mandelb...	3540 Wilshire Blvd	Redwood Shores	CA
2	Abigail Sargent	4211 S Rushford St	Palo Alto	california
3	Lou Voss	777 S. Harbor Bl...	Bellevue	WA
4	Philip Gerstle	PO Box 13507	Skokie	IL

1. Analyse de genre

Exécuter

Colonne source : Nom de la nouvelle colonne :
 NAME_GNDR Paramètres régionaux : English (United States)

Définition : Longueur de caractères :
 256

Sélectionner la colonne source NAME.

Une colonne NAME_GNDR sera créée.

La langue et le pays, ici, l'anglais – Etats-Unis, sont très importants.

Exécuter

CLIENT_INFO (session)

NAME	ID	NAME	AD...	CITY	STATE	ZIP	F
F	1	Sharon...	3540 W...	Redwo...	CA	94065	902-1
F	2	Abigail ...	4211 S ...	Palo Alto	california	94303	860-1
M	3	Lou Voss	777 S. H...	Bellevue	WA	98006...	357-1
M	4	Phillip ...	PO Box ...	Skokie	IL	60076...	360-1
..

1. Analyse de genre

Colonne source : NAME Nom de la nouvelle colonne : NAME_GNDR Paramètres régionaux : English (United States)

Définition : Name Longueur de caractères : 256

La colonne NAME_GNDR a été créée. Sharon est une femme et Phillip un homme.

CLIENT_INFO (session)

NAME_GNDR	ID	NAME	AD...	CITY	STATE	ZIP
U	10	Ashley ...	P.O. Box...	N. Ridg...	OH	44039
M	11	Gary Str...	5153 C...	Sunnyvale	calif	94086
F	12	Michelle...	4000 Ea...	Chester...	missouri	63005
F	13	Cindy P...	515 E. B...	St Louis	MO	63146
..

1. Analyse de genre

Colonne source : NAME Nom de la nouvelle colonne : NAME_GNDR Paramètres régionaux : English (United States)

Définition : Name Longueur de caractères : 256

Pour certain, le genre n'a pas été trouvé.

L'état (= STATE) est parfois avec les initiales, parfois en minuscule, etc.
Nous allons le standardiser.

SAS® Data Studio - Préparer les données

CLIENT_INFO (session)

NAME_GNDR	ID	NAME	AD...	CITY	STATE	ZIP
M	9	Andre ...	5024 Fa...	Sunnyvale	ca	94089
U	10	Ashley ...	P.O. Box...	N. Ridg...	OH	44039
M	11	Gary Str...	5153 C...	Sunnyvale	calif	94086
F	12	Michell...	4000 Ea...	Chester...	missouri	63005

2. Standardiser

Colonne source : Paramètres régionaux : Définition :

Remplacer la colonne source Créer une nouvelle colonne

Options pour les nouvelles colonnes

Double-clic sur Standardiser

SAS® Data Studio - Préparer les données

CLIENT_INFO (session)

NAME_GNDR	ID	NAME	AD...	CITY	STATE	ZIP
M	9	Andre ...	5024 Fa...	Sunnyvale	ca	94089
U	10	Ashley ...	P.O. Box...	N. Ridg...	OH	44039
M	11	Gary Str...	5153 C...	Sunnyvale	calif	94086
F	12	Michell...	4000 Ea...	Chester...	missouri	63005

2. Standardiser

Colonne source : Paramètres régionaux : Définition :

Remplacer la colonne source Créer une nouvelle colonne Options pour les nouvelles colonnes

Sélectionner la colonne source STATE.

Sélectionner la création d'une nouvelle colonne.

La langue et le pays, ici, l'anglais – Etats-Unis, sont très importants.

Sélectionner la définition : State/province (abréviation)

Exécuter

Course: SAS Mya for Learners SAS® Data Studio

SAS® Data Studio - Préparer les données

Plan 1 * : +

Table Profil Méta données

CLIENT_INFO (session)

La table de sessions est à jour pour le plan.

Lignes de résultats : 100

STA...	NA...	ID	NAME	AD...	CITY	STATE	ZIP
CA	M	9	Andre...	5024 Fa...	Sunnyvale	ca	9408
OH	U	10	Ashley ...	P.O. Box...	N. Ridg...	OH	4403
CA	M	11	Gary Str...	5153 C...	Sunnyvale	calif	9408
MO	F	12	Michelle...	4000 Ea...	Chester...	missouri	6300
MO	r	13	Christ...	445 E...	East ...	MO	6314

2. Standardiser

Exécuter

Colonne source : Paramètres régionaux : Définition :

STATE English(United States) State/Province(Abbreviation) +

Remplacer la colonne source Créer une nouvelle colonne Options pour les nouvelles colonnes

Plan

Nom : Plan 1
Modifié : 29/12/2020 12:59

1. Analyse de genre
2. Standardiser

La colonne STATE_STND semble correctement standardisée.

Nous allons maintenant extraire du numéro de téléphone la première partie grâce à une analyse de parsing.

Course: SAS Mya for Learners SAS® Data Studio

SAS® Data Studio - Préparer les données

Plan 1 * : +

Table Profil Méta données

CLIENT_INFO (session)

La table de sessions est à jour pour le plan.

Lignes de résultats : 100

STA...	NA...	ID	NAME	AD...	CITY	STATE	ZIP
CA	M	9	Andre...	5024 Fa...	Sunnyvale	ca	9408
OH	U	10	Ashley ...	P.O. Box...	N. Ridg...	OH	4403
CA	M	11	Gary Str...	5153 C...	Sunnyvale	calif	9408
MO	F	12	Michelle...	4000 Ea...	Chester...	missouri	6300
MO	r	13	Christ...	445 E...	East ...	MO	6314

2. Standardiser

Exécuter

Colonne source : Paramètres régionaux : Définition :

STATE English(United States) State/Province(Abbreviation) +

Remplacer la colonne source Créer une nouvelle colonne Options pour les nouvelles colonnes

Plan

Nom : Plan 1
Modifié : 29/12/2020 12:59

1. Analyse de genre
2. Standardiser

Double-clic sur Analyse (parsing).

Sélectionner la colonne source PHONE.

La langue et le pays, ici, l'anglais – Etats-Unis, sont très importants.

Sélectionner la définition : Phone

Sélectionner Base Number

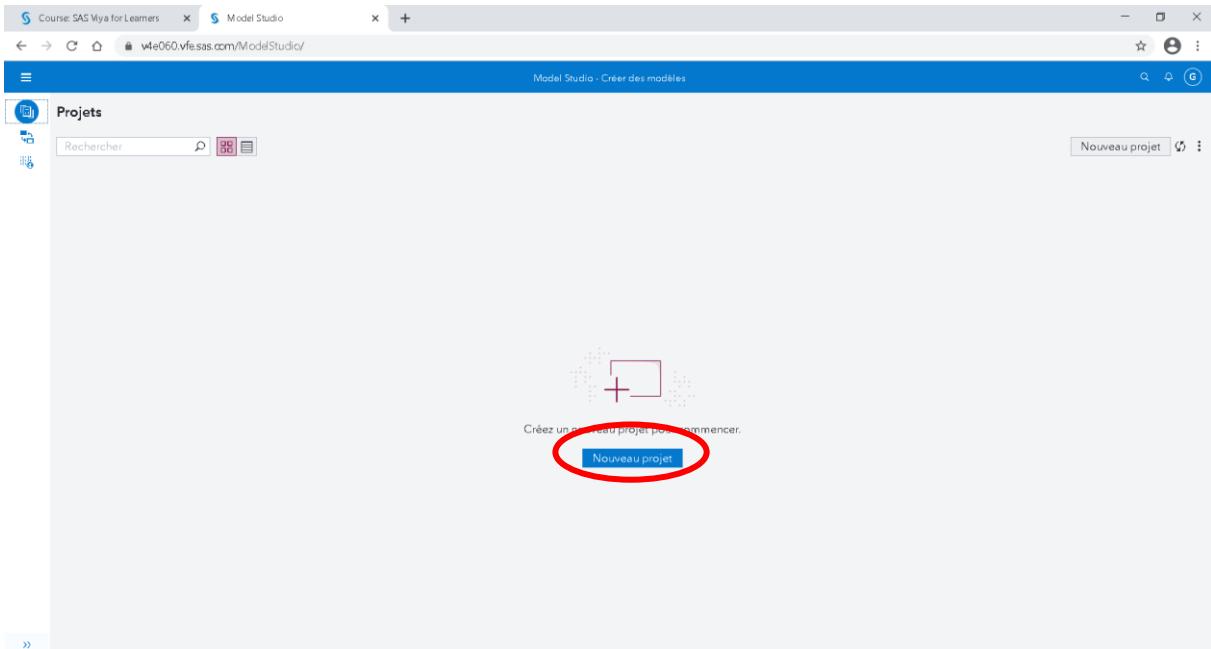
Exécuter

Machine Learning en Pipeline

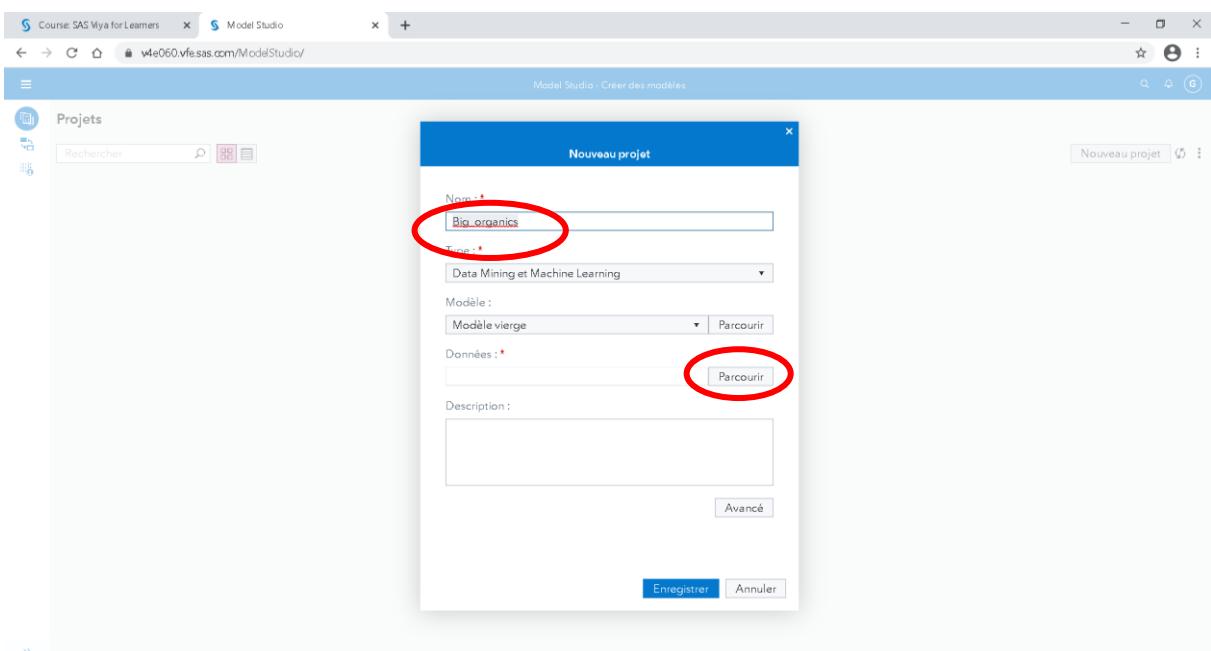
Le machine Learning en « pipeline » dans SAS Viya ressemble beaucoup aux « flux de processus » de SAS Enterprise Miner. « Pipeline » et « flux de processus » sont des expressions synonymes.

The screenshot shows the SAS Viya interface with two main panels. On the left, a sidebar titled 'CYCLE DE VIE ANALYTIQUE' (Analytics Life Cycle) lists several categories: Gérer les données, Préparer les données, Explorer et visualiser, **Create models** (highlighted with a red circle), Construire des décisions, Partager et collaborer, Développer le code SAS, and ADMINISTRATION. On the right, the main workspace is titled 'SAS® Drive - Partager et collaborer' and shows a list of items under 'My Folder': 'My Snippets', 'SAS Videos', 'BigOrganics', and 'test'. A message at the bottom right says 'Sélectionnez un élément pour voir ses informations.'

Créer des modèles



Nouveau Projet



Lui donner un nom et sélectionner la table

The screenshot shows the 'Choisir des données' (Select Data) dialog in SAS Model Studio. On the left, there's a list of available datasets under 'Disponible'. The 'BIGORGANICS' dataset is selected and highlighted with a red circle. On the right, there's a detailed view of the 'BIGORGANICS' dataset, including its columns (id, DemAffl, DemAge, DemClusterGroup, DemGender, DemReg, DemTVReg), types (categorical or numerical), and statistics (13 columns, 111,1 K lines). At the bottom right of the dialog, there's an 'OK' button highlighted with a red circle.

Sélectionner BigOrganics

OK

The screenshot shows the 'Nouveau projet' (New Project) dialog in SAS Model Studio. It's a form for creating a new project named 'Big_organics' with a type of 'Data Mining et Machine Learning'. The 'Modèle' is set to 'Modèle vierge' and the 'Données' are set to 'TUNDATA.BIGORGANICS'. There's an 'Avancé' (Advanced) button and an 'Enregistrer' (Save) button at the bottom right, which is highlighted with a red circle.

Enregistrer

Nom de la variable	Libellé	Type	Rôle	Niveau	Ordre
DemAffi	Affluence Grade	Numérique	Entrée	Intervalle	Par défaut
DemAge	Age	Numérique	Entrée	Intervalle	Par défaut
DemCluster	Neighborhood Cluster-55 Level	Caractère	Rejeté	Nominal	Par défaut
DemClusterGroup	Neighborhood Cluster-7 Level	Caractère	Entrée	Nominal	Par défaut
DemGender	Gender	Caractère	Entrée	Nominal	Par défaut
DemReg	Geographic Region	Caractère	Entrée	Nominal	Par défaut
DemTVReg	Television Region	Caractère	Entrée	Nominal	Par défaut
id	Customer Loyalty ID	Caractère	ID	Nominal	Par défaut
PromClass	Loyalty Status	Caractère	Entrée	Nominal	Par défaut
PromSpend	Total Spend	Numérique	Entrée	Intervalle	Par défaut
PromT	Loyalty Card Tenure	Numérique	Entrée	Intervalle	Par défaut
<input checked="" type="checkbox"/> TargetAmt	Organics Purchase Count	Numérique	Cible	Nominal	Par défaut
TargetBuy	Organic Purchase Type	Caractère	Entrée	Binaire	Par défaut

Rôle :

- Cible
- Cible
- Classification
- Cle
- Décalage
- Entrée
- Evaluation
- Filtre
- ID
- ID temps
- Partition
- Rejeté
- Résiduel
- Segment

Sélectionner la colonne TargetAMT et changer son rôle à rejeté.

Données Pipelines Comparaison de pipelines Tableaux de bord

Pipeline 1

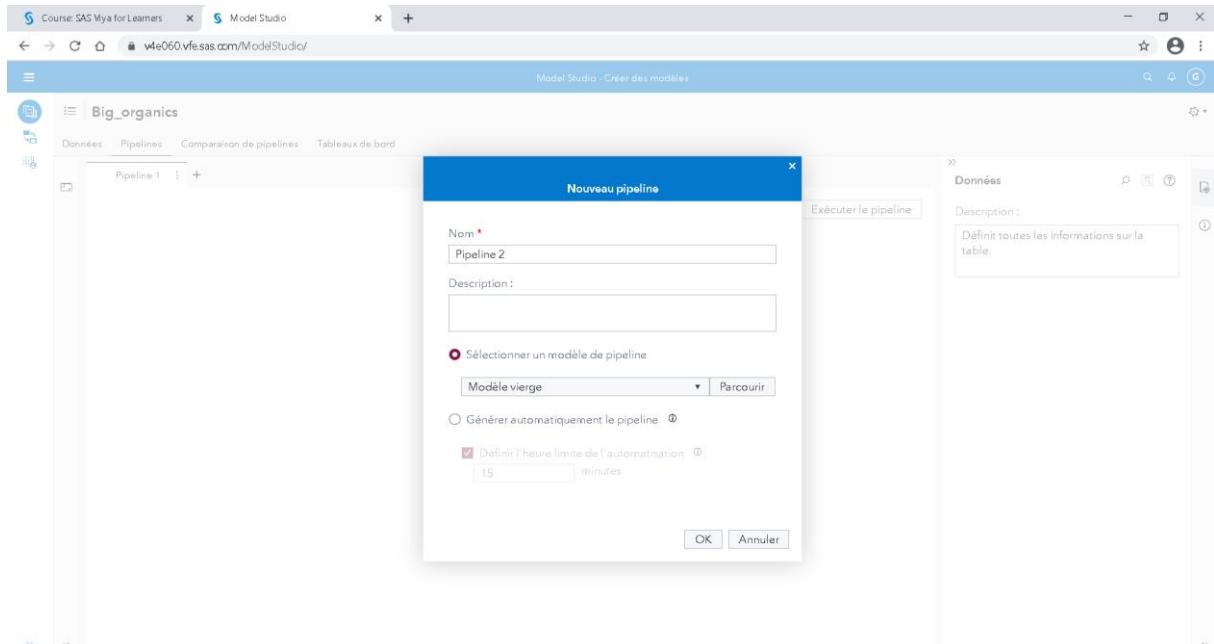
Exécuter le pipeline

Données

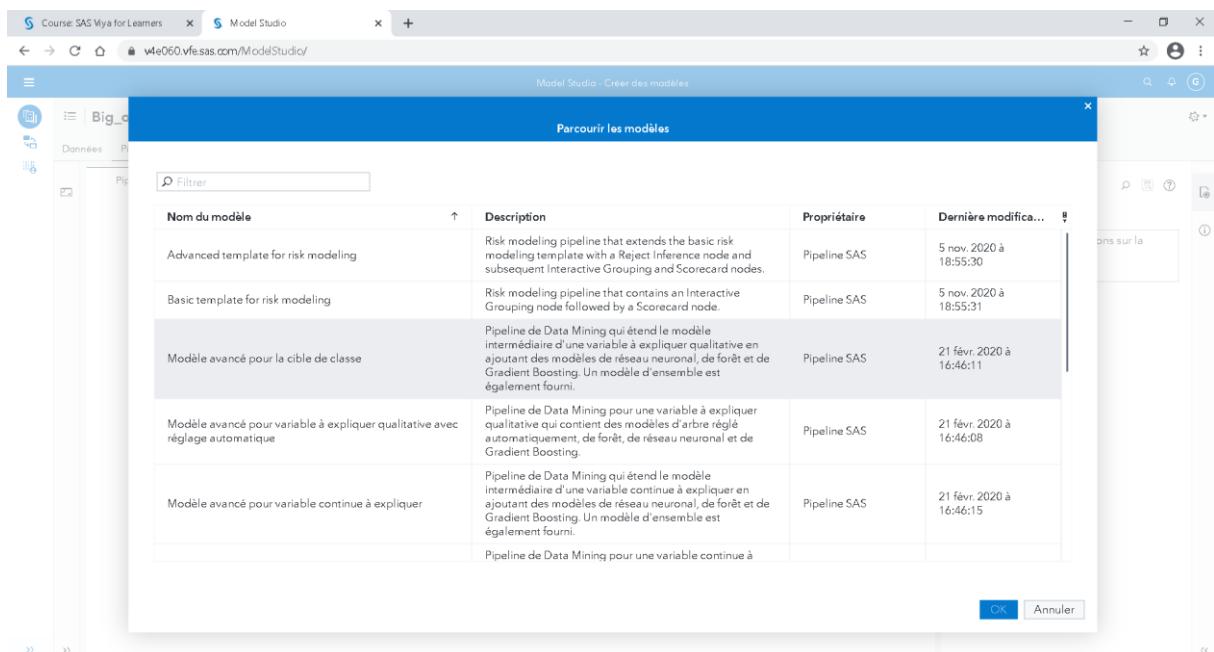
Description : Définit toutes les informations sur la table.

Pipeline

+ ➔ nouveau pipeline

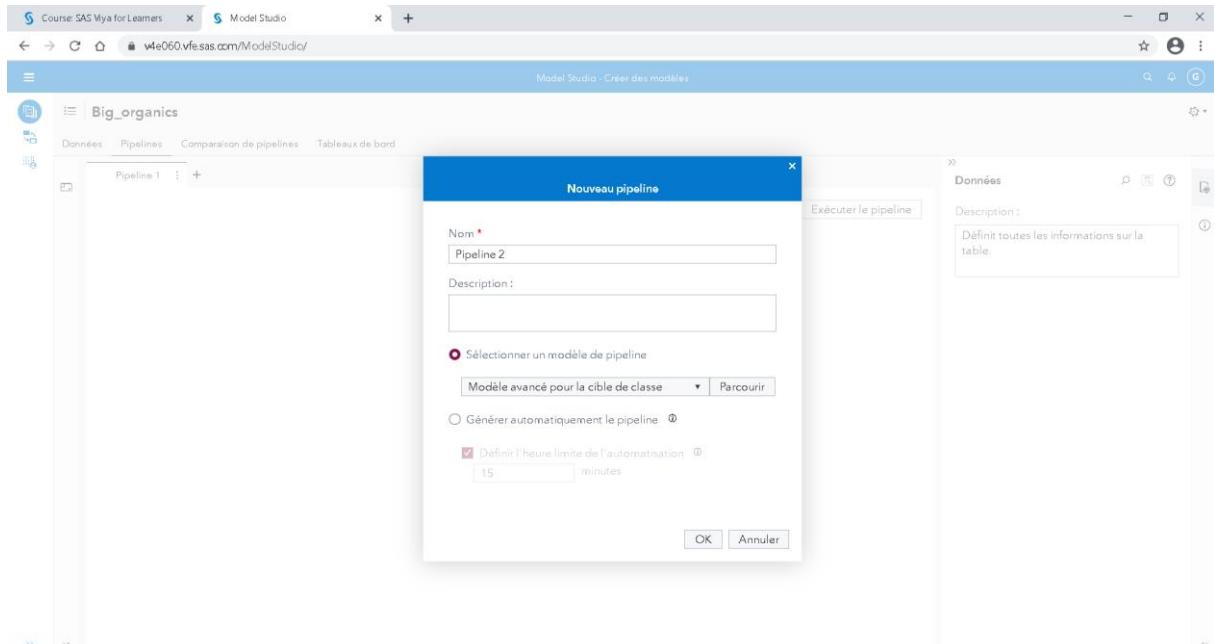


Parcourir

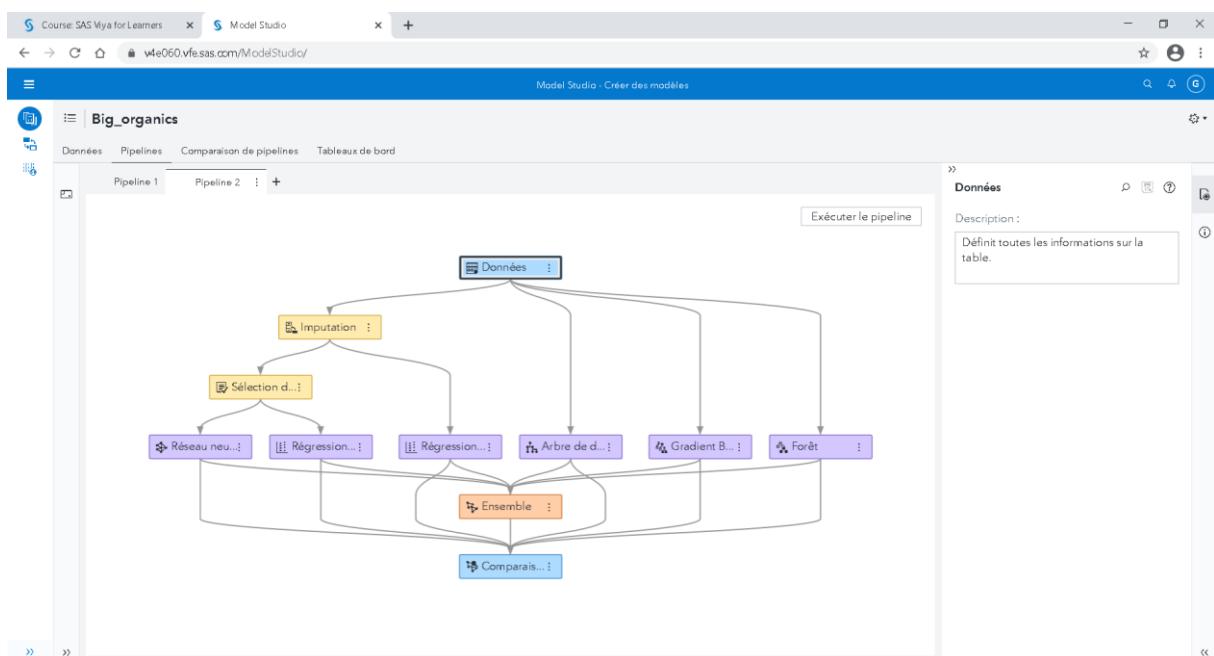


Modèle avancé pour une variable cible de classe

OK



OK



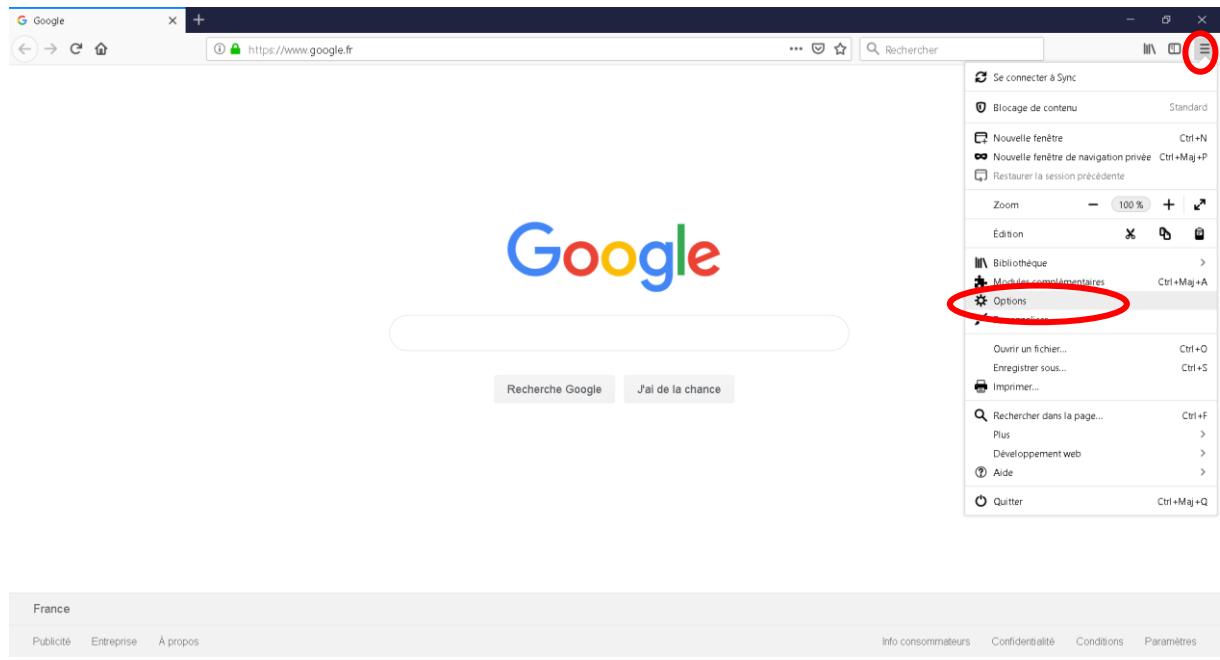
Ceci est une introduction très rapide.

Pour aller plus loin, reprendre le cours de machine Learning dans SAS Academic Hub. Il est nécessaire et suffisant pour préparer la certification de Machine Learning sur SAS Viya.

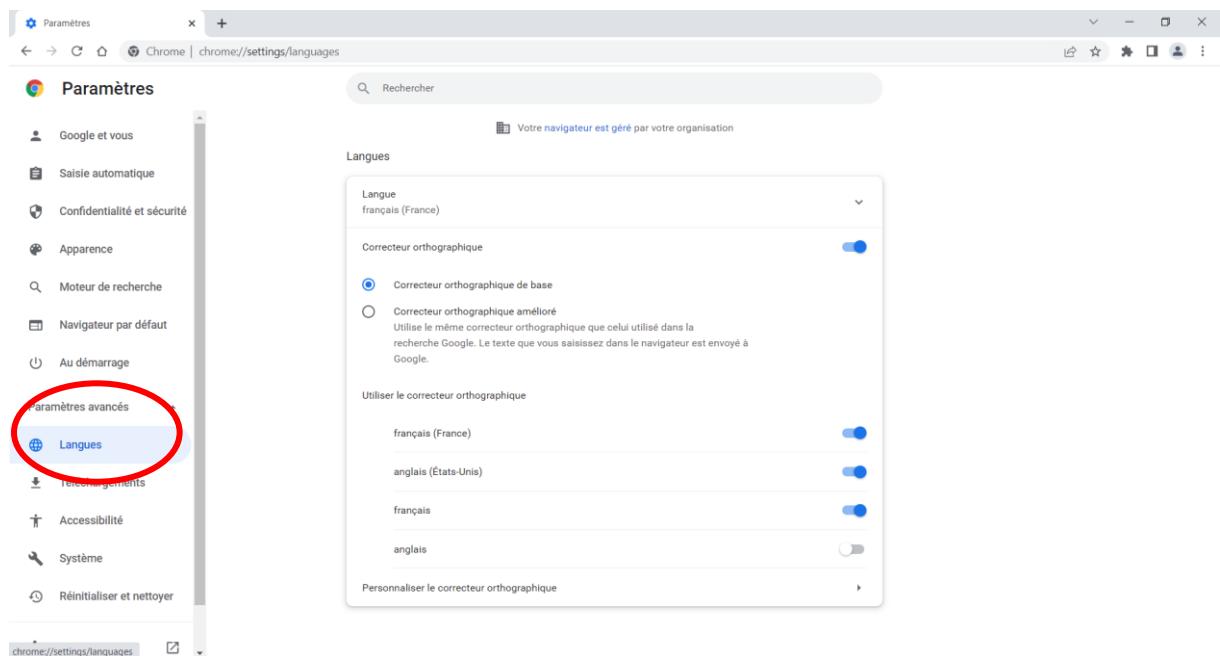
Changer la langue du navigateur

Les interfaces SAS utilisent notamment la langue d'affichage du navigateur pour choisir la langue de l'interface SAS. Donc pour avoir SAS en français par exemple, il faut changer la langue de d'affichage du navigateur.

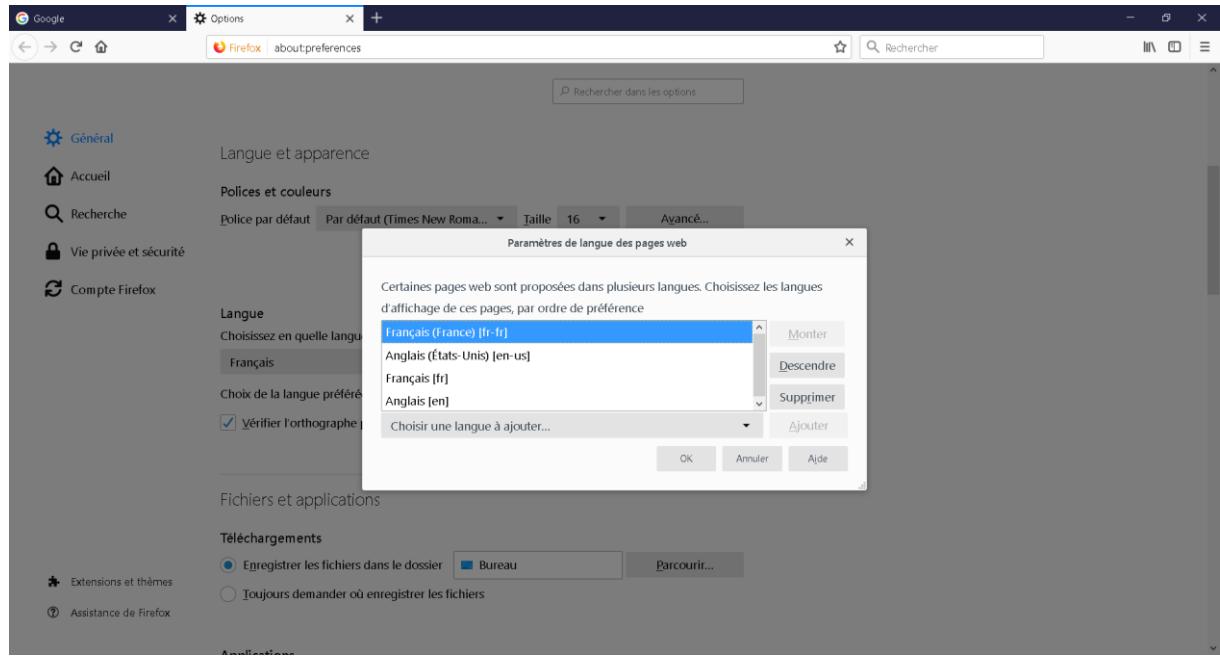
Changer la langue du navigateur Firefox



Dans le menu de Firefox en haut à droite → Options



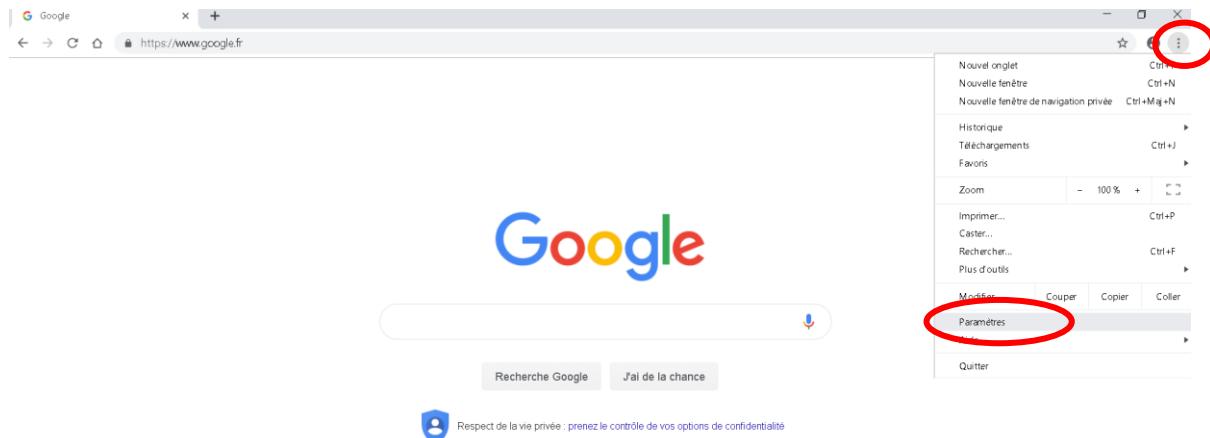
Descendre jusqu'à langue et apparence → langue et **choisir** la langue préférée pour l'affichage des pages



Il faut que la première ligne soit **Français (France) [fr-fr]**

Utiliser les boutons **Ajouter**, **Monter** et **Descendre** si besoin.

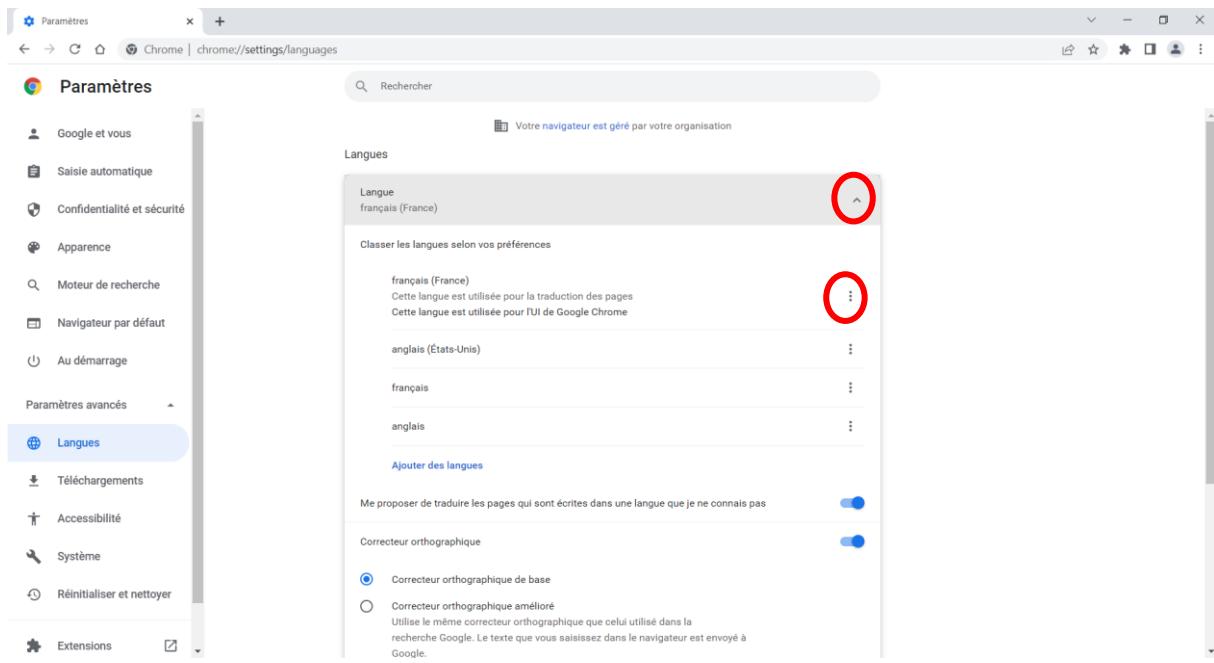
Changer la langue du navigateur Chrome



Aller dans le menu de Chrome en haut à droite → **Paramètres**

A screenshot of the Chrome settings page, specifically the "Langues" section. On the left, the sidebar shows "Paramètres avancés" with "Langues" selected and circled in red. The main area displays language settings for the spell checker and browser interface. The "Langue" dropdown is set to "français (France)". Under "Utiliser le correcteur orthographique", "français (France)" is selected. A second screenshot of the context menu from the previous image is shown on the right, with "Paramètres" again highlighted and circled in red.

A gauche, sélectionner les **Paramètres avancés** → **Langue**



Déplier **Langues**, il faut que **France (France)** soit en premier. Il est possible d'ajouter une langue et de les monter ou descendre.

Vidéos complémentaires :

Stratégie d'optimisation de la recherche d'hyperparamètres de modèles de Machine Learning dans Model Studio sur SAS Viya

<https://www.youtube.com/watch?v=IDhoSQPi-HI>

Présentation de régressions logistiques avec différentes procédure SAS

<https://www.youtube.com/watch?v=W9qpShm0HaQ>

Amélioration de la performance d'un modèle de Machine Learning avec le Text Mining

<https://www.youtube.com/watch?v=LKOdQMbd0yl>

Définition de la différence entre une prévision et une prédiction

<https://www.youtube.com/watch?v=2t9cznGK0HU>

Interprétation de modèle « boîte noire » (RN – GB – SVM – etc.) dépendance partielle

<https://www.youtube.com/watch?v=1xtiZFhBYcs>

Lancer un code SAS créer via un Job SAS Studio V, depuis SAS Visual Analytics

<https://www.youtube.com/watch?v=mNJkalYPToQ>

Différentes interfaces - Machine Learning - SAS Viya

<https://www.youtube.com/watch?v=5c5cGCIUeTq>

Introduction aux réseaux neuronaux d'apprentissage profond (Deep Learning) sous SAS® Viya®

<https://www.youtube.com/watch?v=6iMeaCya-FY>

Préparer de la certification de programmation de base

Se connecter à SAS Skill Builder

<https://www.youtube.com/watch?v=uUVxpfT68oE>

SAS Skill Builder - Certification programmation SAS Base

https://www.youtube.com/watch?v=_TSCSrPUAV8

<https://www.youtube.com/watch?v=U-yk3wO1aeM>

Passage de l'examen

<https://www.youtube.com/watch?v=tKAno46PhfM>

Aller à la page https://www.sas.com/fr_fr/learn/academic-programs/students.html
Se connecter à SAS Skill Builder

The screenshot shows the SAS Skill Builder for Students landing page. At the top, there are two main sections: one with a "Get Tips" button and another with a "Start Learning" button. Below these are three buttons: "Free Software" (circled in red), "Technical Tutorials", and "SAS Documentation". The "Free Software" button has a circular icon with a play symbol. Below these buttons are two large callout boxes. The left box, titled "PREPARE FOR A CAREER IN HEALTHCARE AND LIFE SCIENCES", features a hospital icon and describes fields like epidemiology, health care operations, and biostatistics. The right box, titled "PREPARE FOR A CAREER IN GOVERNMENT", features a government building icon and describes fields like defence, national security, and transportation.

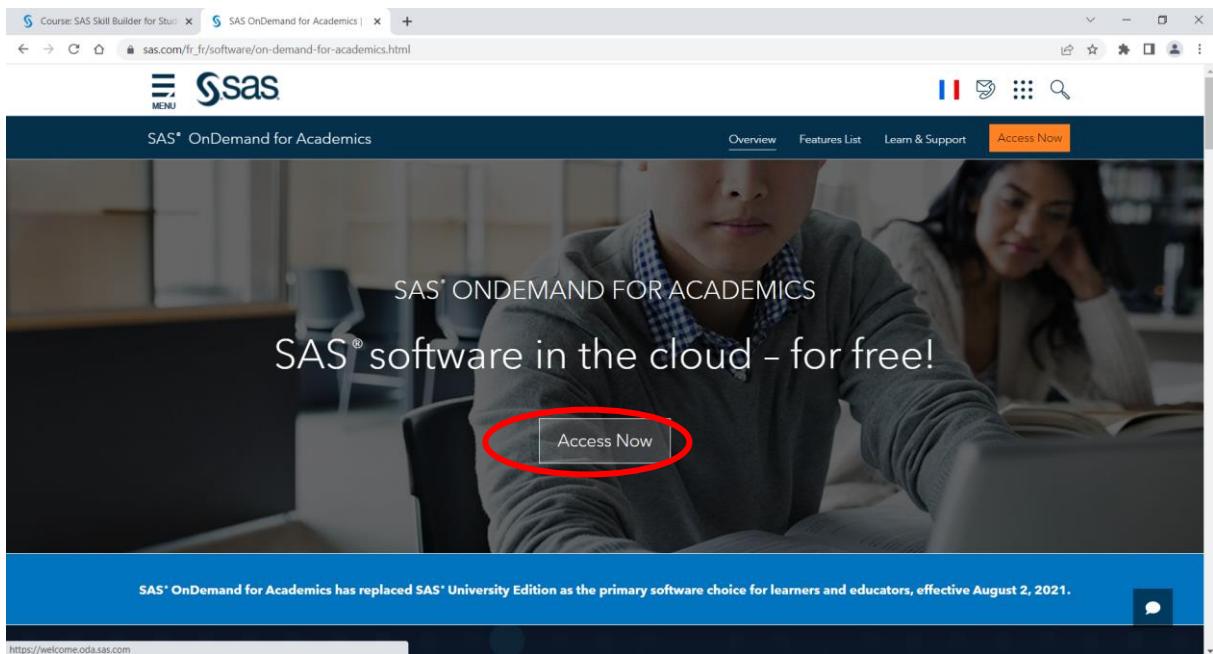
Dans Free software

The screenshot shows the SAS Academic Programs page. It features a central monitor displaying SAS Studio and SAS Viya interfaces. To the left, a sidebar lists various academic programs. Below the monitor, a list of benefits includes:

- Quickly launch a browser-based programming environment using SAS Studio.
- Learn SAS programming from basic to advanced techniques using coding or point-and-click tasks.
- Sharpen your analytics skills in data preparation, descriptive analyses and advanced statistical methods.
- Access the technology via the cloud and get up to 5GB of data storage.

A blue button at the bottom left of the monitor area is circled in red and labeled "Visit SAS OnDemand for Academics". To the right, another monitor displays a dashboard with various data visualizations. A green button below it is labeled "Visit SAS Viya for Learners".

Lancer SAS OnDemand for Academics



Access Now

Cliquer sur **SAS OnDemand for Academics Control Center**.

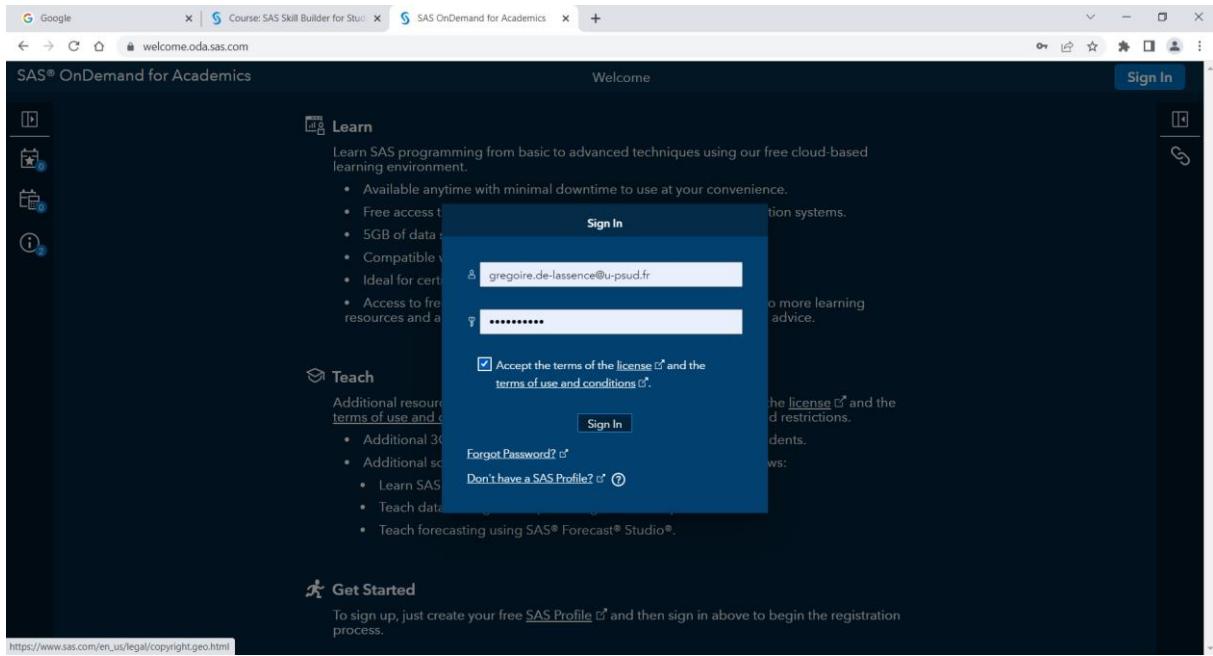
Selectionner l'**Europe**. Cela permet de réduire la latence de l'application.

Si vous n'avez pas encore de compte sur SAS OnDemand for Academics, il faut en créer un.

Se connecter avec son profil SAS.

A screenshot of the SAS OnDemand for Academics "Welcome" page. The page has a dark blue header with the "Welcome" text. Below the header, there are three main sections: "Learn", "Teach", and "Get Started". The "Learn" section contains a list of bullet points about the learning environment. The "Teach" section contains a list of bullet points about teaching resources. The "Get Started" section contains text about creating a SAS Profile and signing in. A red circle highlights the "Sign In" button located in the top right corner of the page.

« Sign In »

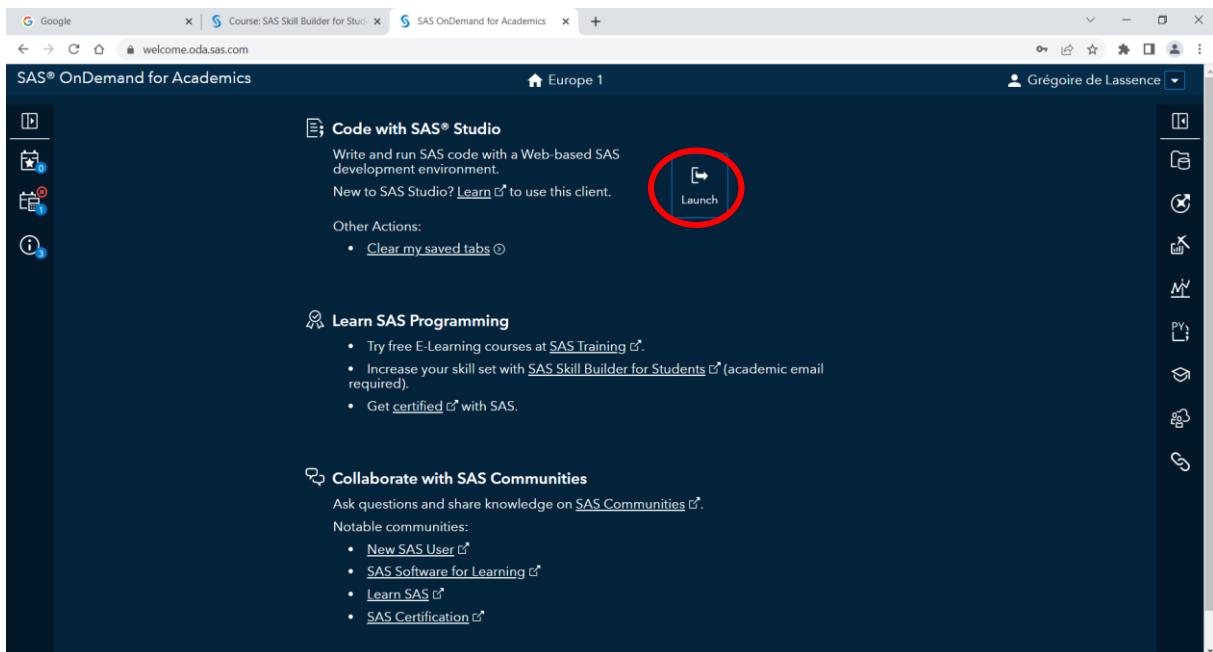


Utiliser votre identifiant ou votre adresse e-mail.

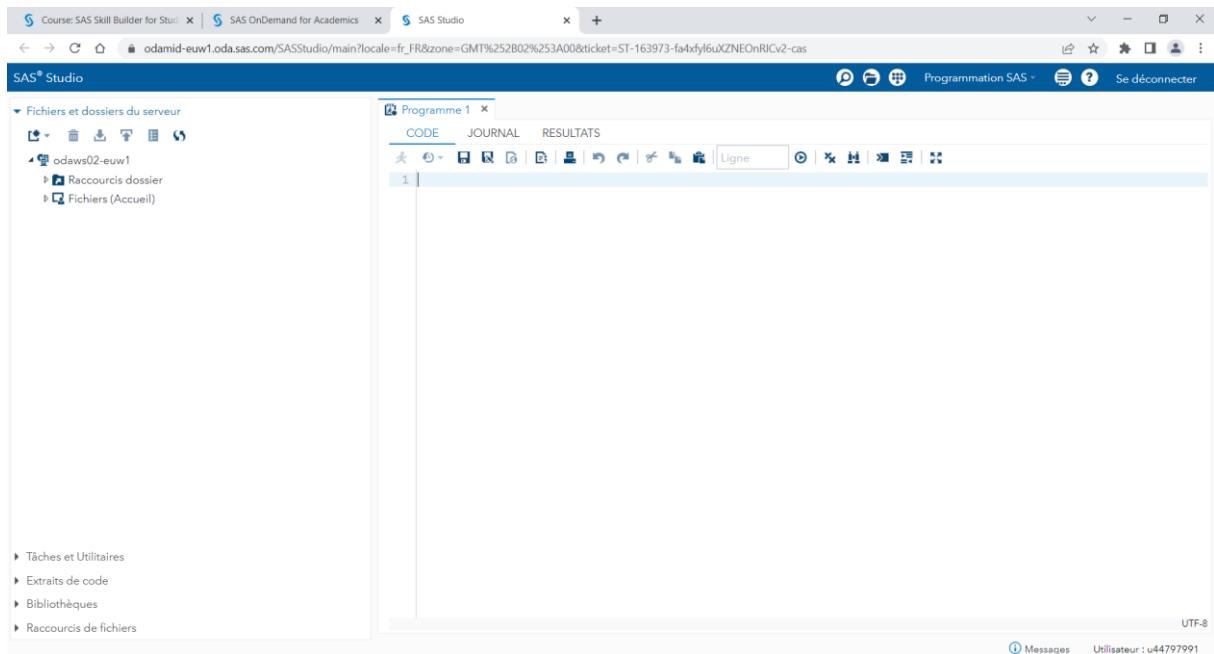
Si vous commencer sur SAS On Demand for Academics, vous n'avez accès qu'à SAS Studio

Si vous êtes étudiant, vous ne pouvez pas créer un cours.

Si vous avez créé un profil SAS avec le statut d'enseignant, il faut attendre quelques jours pour qu'une personne de SAS valide ce statut en vérifiant avec vos prénom et nom, issu de votre adresse e-mail académique, que tel est bien le cas. Avec ce profil enseignant, il y a deux éléments qui diffèrent : vous pouvez créer un cours et y inviter vos étudiants, et en plus des 5 Go d'espace disponibles comme les étudiants, vous avez 3 Go d'espace pour partager avec les étudiants enrôlés à vos cours, des données ou des codes par exemple.



Aller dans SAS Studio.



Pour préparer la certification de programmation SAS,

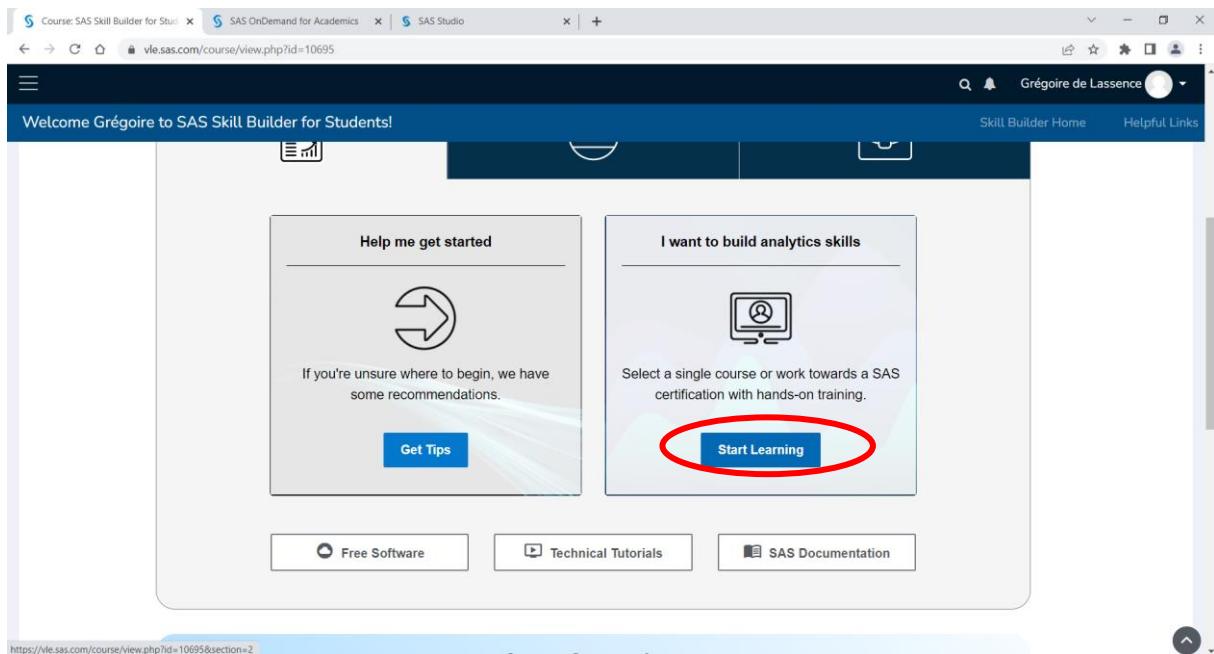
Aller dans SAS Skill Builder, l'onglet est resté ouvert (si non, https://www.sas.com/en_us/learn/academic-programs/students.html)

Welcome Grégoire to SAS Skill Builder for Students!

SAS EXPLORE
A Free Virtual Event for Technologists
Sept. 27-29
[REGISTER NOW](#)

[Learn SAS](#) [Get SAS Certified](#) [Career Resources](#)

[Help me get started](#) [I want to build analytics skills](#)



Start Learning

The screenshot shows the SAS Virtual Learning Environment. It displays the 'E-Learning Courses and Certification Pathways' section. A 'Back to Learn SAS' button is visible. The 'Programming' category is expanded, showing 'SAS Programming 1: Essentials' and 'SAS Programming 2: Data Manipulation Techniques', both of which are circled in red. An 'Expand All' button is located in the top right corner. The URL in the browser bar is <https://vle.sas.com/course/view.php?id=10695§ion=2>.

Déplier « Programming »

Aller dans « SAS Programming 1: Essentials »

The screenshot shows the SAS Global Academic Student enrollment page. A 'SAS Global Academic Student' dropdown is open. A red circle highlights the 'Enroll me' button.

S'enrôler dans le cours

The screenshot shows a web browser window with four tabs open: "Course: SAS Skill Builder for Stud...", "Course: SAS Programming 1: Es...", "SAS OnDemand for Academics", and "SAS Studio". The main content area displays the "Programming 101" course page. A red circle highlights the first lesson, "Course Overview and Data Setup (REQUIRED)". Below it, a note says: "Important: If you are not using the SAS Virtual Lab, you must perform the data setup at the end of the Course Overview to create the files for this course." A sidebar on the right includes a "Rate This Course" section with a 5-star rating and a feedback form.

Lessons

- Course Overview and Data Setup (REQUIRED)**
- Lesson 1: Essentials
- Lesson 2: Accessing Data
- Lesson 3: Exploring and Validating Data
- Lesson 4: Preparing Data
- Lesson 5: Analyzing and Reporting on Data
- Lesson 6: Exporting Results
- Lesson 7: Using SQL in SAS®

Resources

- Extended Learning: SAS® Programming 1: Essentials

Course Overview and Data Setup (REQUIRED)

The screenshot shows a web browser window with four tabs open: "Course: SAS Skill Builder for Stud...", "EPG1V2: Course Overview and D...", "SAS OnDemand for Academics", and "SAS Studio". The main content area displays the "Course Overview and Data Setup (REQUIRED)" page. A sidebar on the left lists course modules: "Course Overview and Data Setup", "Welcome to the Course 1:50", "Demo: The SAS Programming Process 7:17", "Practicing in This Course 1:02", and "Setting Up Your Practice Files (REQUIRED)". The main content area features a large blue background with the text "SAS® Programming 1: Essentials" and a "Resume" button.

En regardant les vidéos, on peut mettre les sous-titres en anglais.

Dans les « Setting Up Your Practice Files (REQUIRED) » cliquer sur le bouton Open pour «Setup instructions for SAS OnDemand for Academics».

The screenshot shows a web browser window with the title "Setup for SAS onDemand for Academics: SAS Studio - Google Chrome". The URL in the address bar is "vle.sas.com/pluginfile.php/800779/mod_scorm/content/96/setup_soda_new.htm". The page content is titled "Setup Instructions for SAS OnDemand for Academics: SAS Studio". It contains two main sections: "Task 1 of 2: Access SAS OnDemand for Academics" and "Task 2 of 2: Create your course files.". The first section instructs users to sign in with their SAS profile email and password, and to register for a SAS OnDemand for Academics account. It includes a "Register or Sign in" button. The second section provides steps to set up data sets and practice files for SAS Studio, assuming an active account. Step 1 is circled in red.

**Setup Instructions for SAS OnDemand for Academics:
SAS Studio**

Task 1 of 2: Access SAS OnDemand for Academics

Sign in with your SAS profile email address and password (the same one you use to access this training). Then, register for a SAS OnDemand for Academics account. You can use this same link to sign in to SAS OnDemand for Academics after you register.

[Register or Sign in](#)

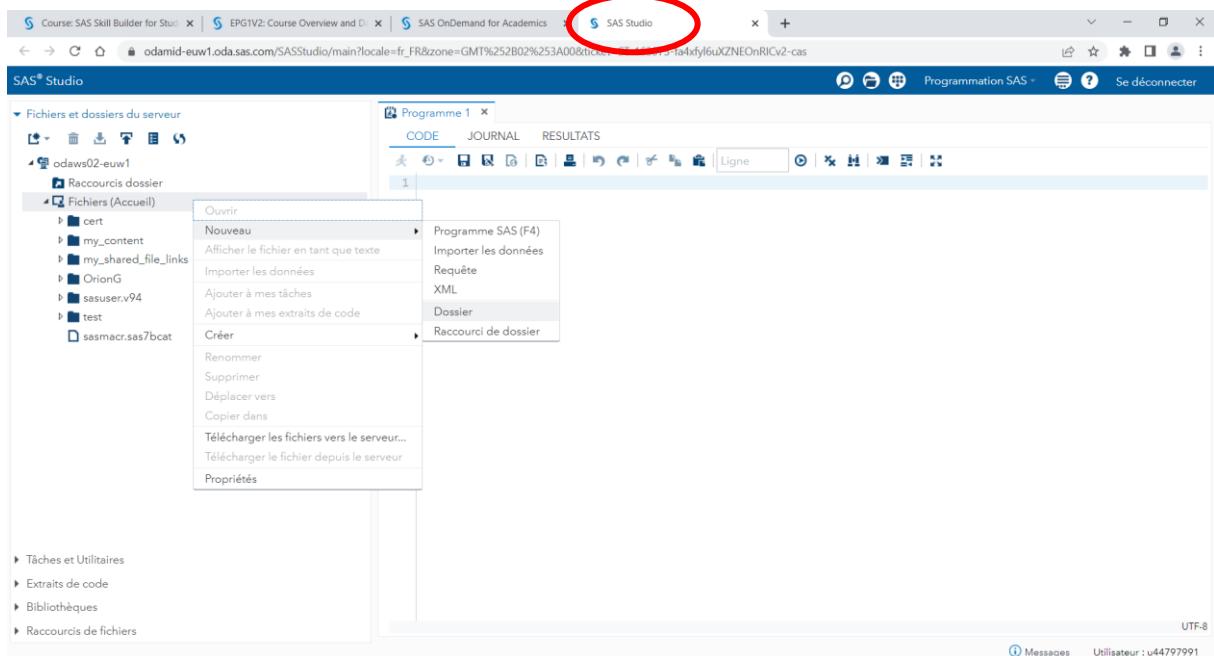
Task 2 of 2: Create your course files.

To complete the practices in this course, you must **complete the steps below** to set up the data sets and practice files for SAS Studio. These instructions assume that you have an active SAS OnDemand for Academics account.

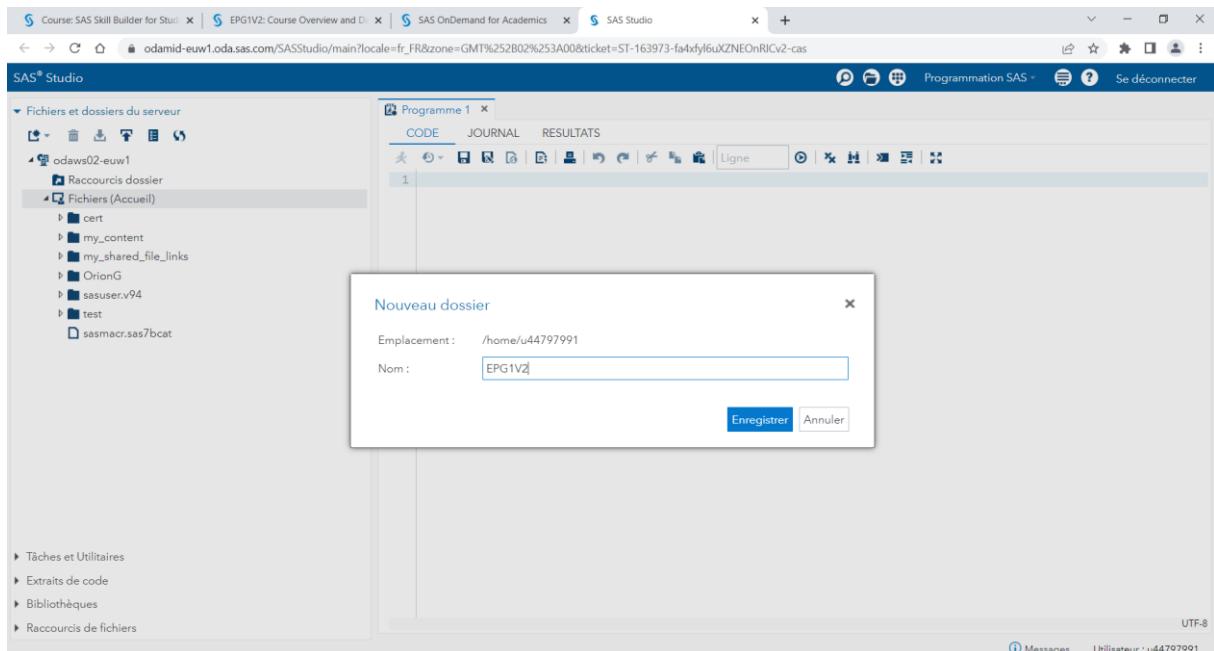
1. Open [this zip file](#) and extract it to any location on your computer. The unzip process creates a folder named **EPG1V2** that includes one SAS program: **createdataPG1v2_oda.sas**.
2. Sign in to your [SAS OnDemand for Academics Dashboard](#) and start SAS Studio.
3. At the top of the **Server Files and Folders** panel, click **New** and select **Folder**.
4. In the **Name** box, type **EPG1V2** in uppercase exactly as shown. Click **Save**.
5. If necessary, expand **Files(Home)** and verify that the **EPG1V2** folder was created.

Télécharger le ZIP

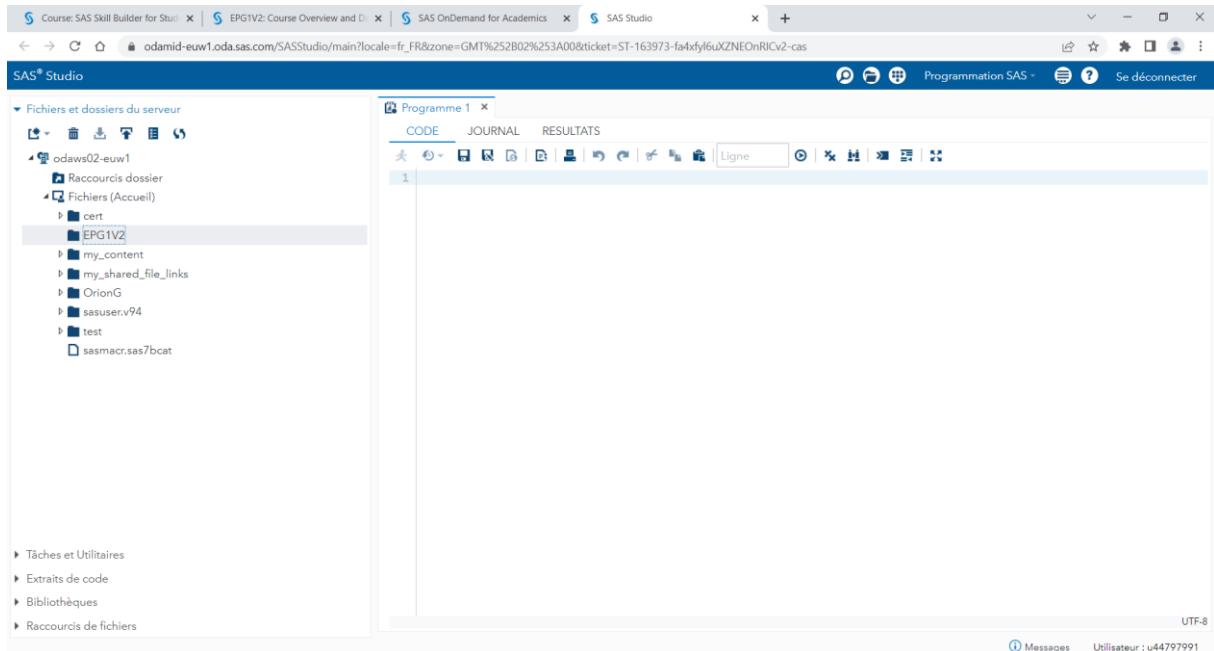
Dans SAS Studio, l'onglet doit être resté ouvert



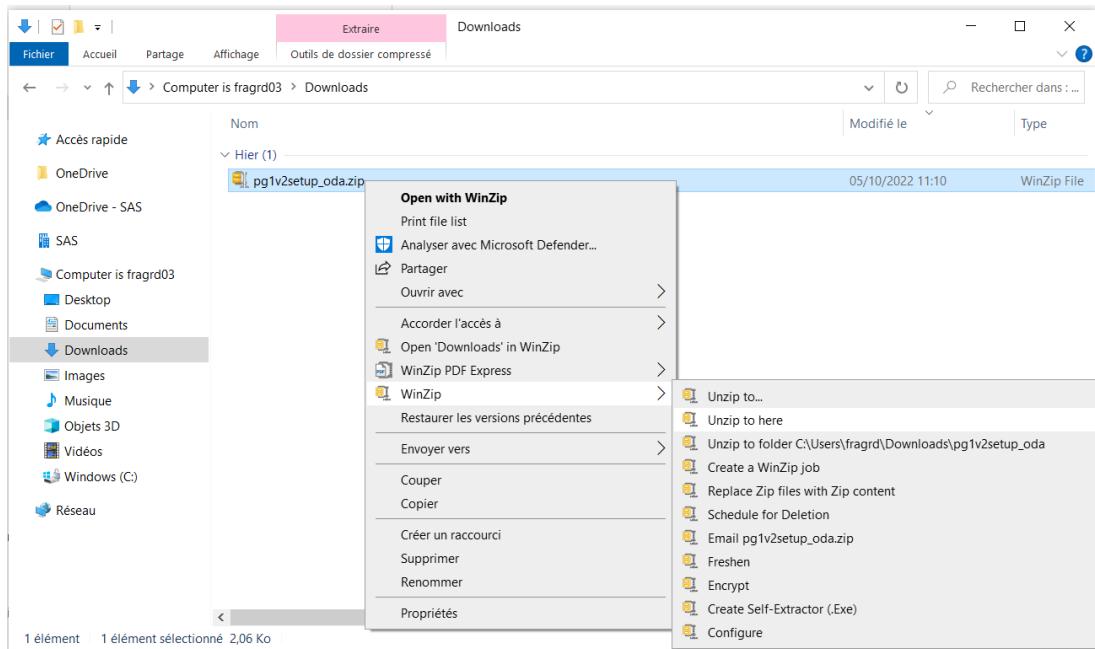
Déplier odaws0X-euw1, clique-droit sur « Fichier (Accueil) » → Nouveau → Dossier

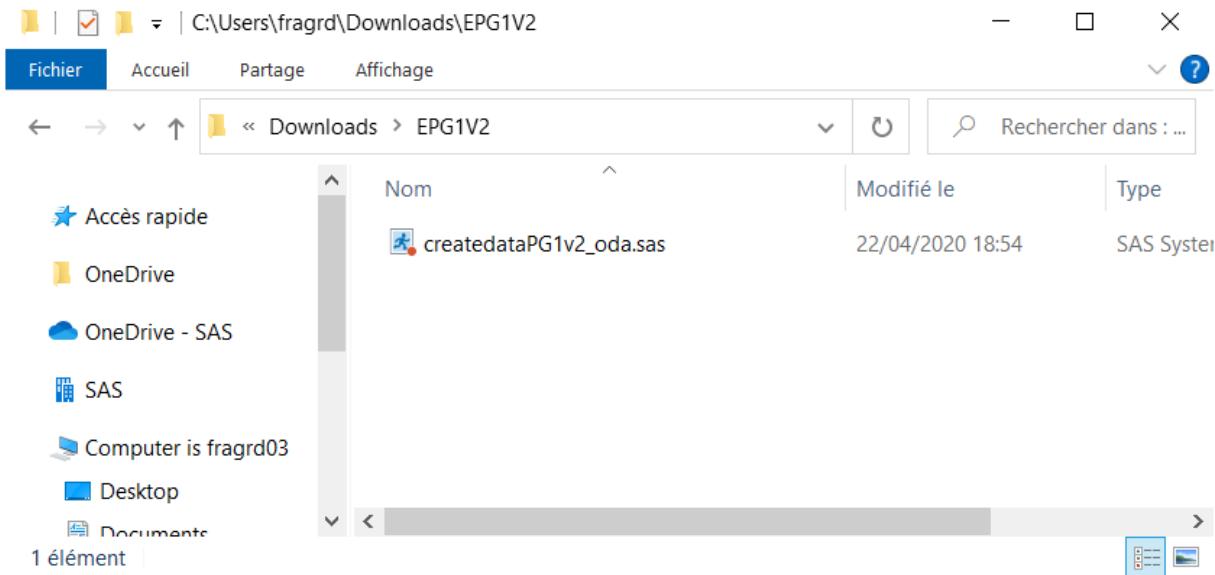


Entrer le nom exactement **EPG1V2**
Enregistrer



Dans l'explorateur de fichier, dans votre dossier de téléchargement, décompresser le Zip téléchargé.

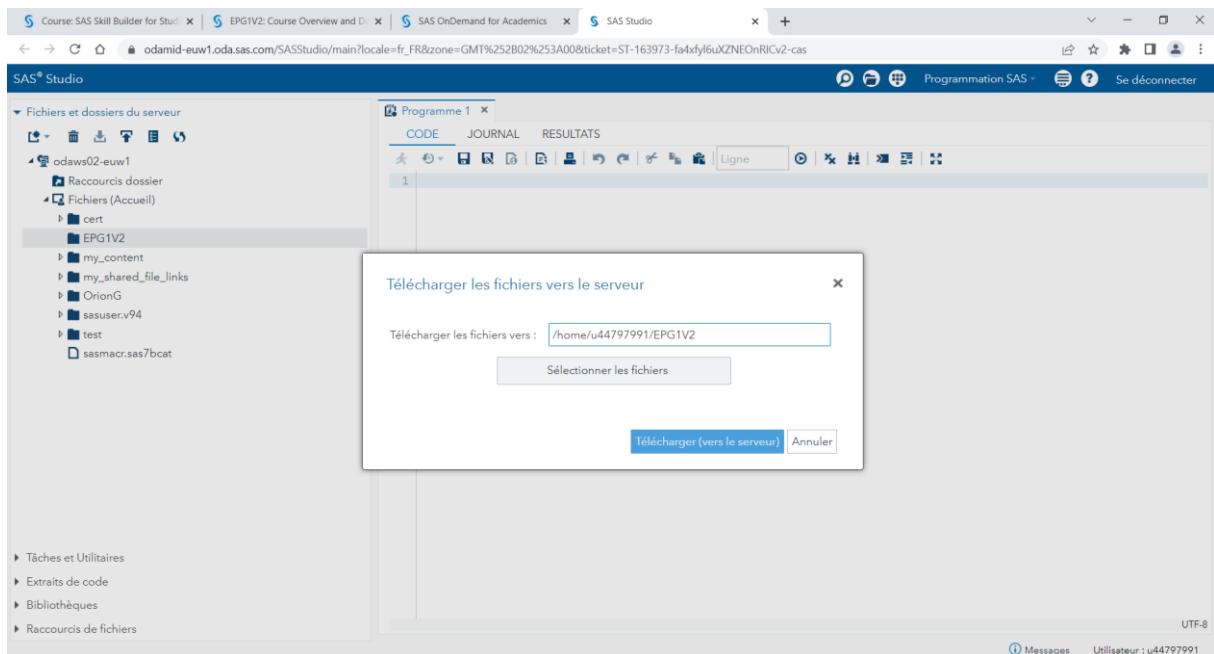




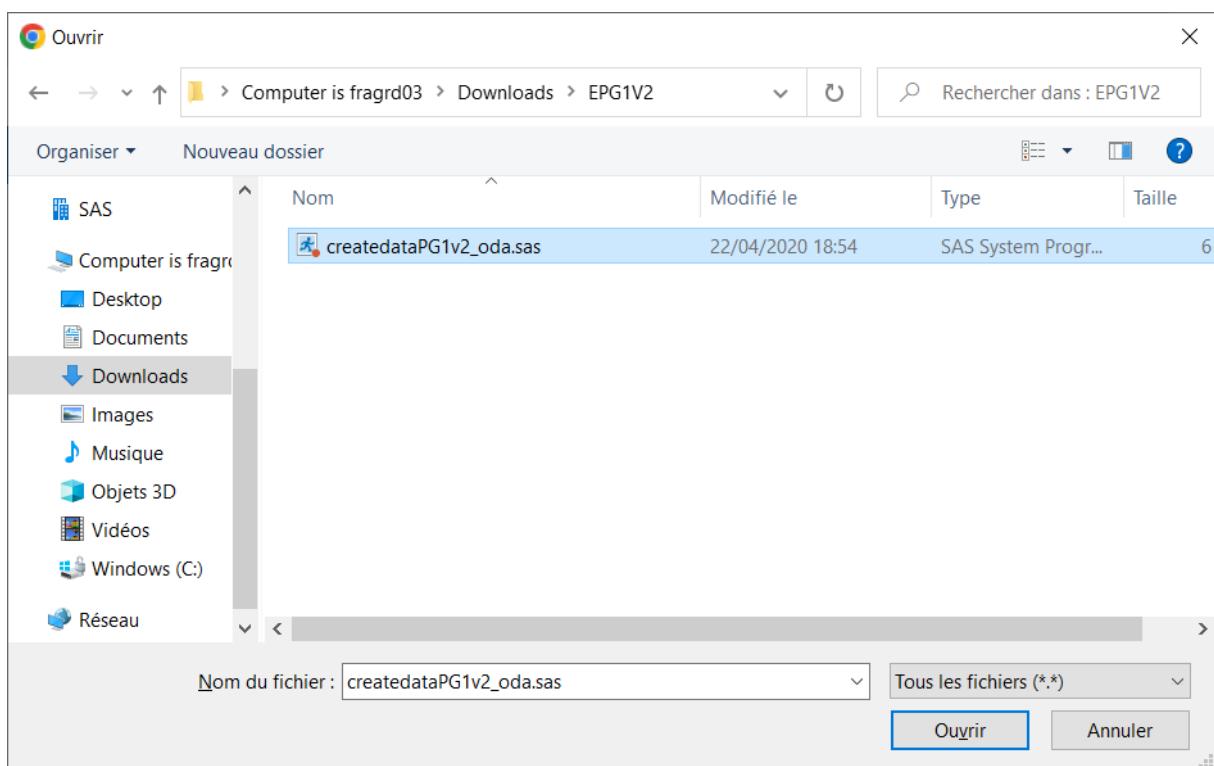
Dans SAS Studio,

The screenshot shows the SAS Studio interface. The top navigation bar includes tabs for 'Course: SAS Skill Builder for Stud...' and 'SAS OnDemand for Academics'. The main workspace shows a 'Programme 1' editor with three tabs: CODE, JOURNAL, and RESULTATS. The left sidebar, titled 'SAS Studio', contains a tree view of files and folders. A context menu is open over the 'EPG1V2' folder in the tree view, listing options such as 'Ouvrir', 'Nouveau', 'Afficher le fichier en tant que texte', 'Importer les données', 'Ajouter à mes tâches', 'Ajouter à mes extraits de code', 'Créer', 'Renommer', 'Supprimer', 'Déplacer vers', 'Copier dans', 'Télécharger les fichiers vers le serveur...', 'Télécharger le fichier depuis le serveur...', and 'Propriétés'. The status bar at the bottom right shows 'UTF-8' and 'Utilisateur : u44797991'.

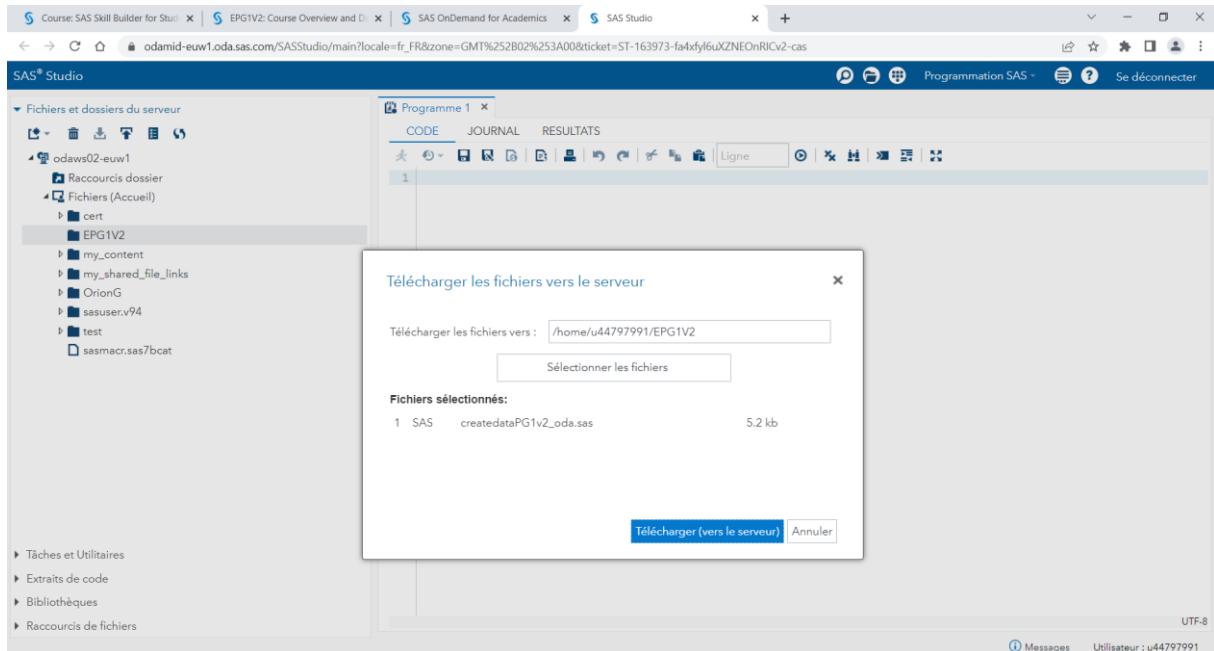
Clique droit sur le dossier EPG1V2 → Télécharger les fichiers vers le serveur ...



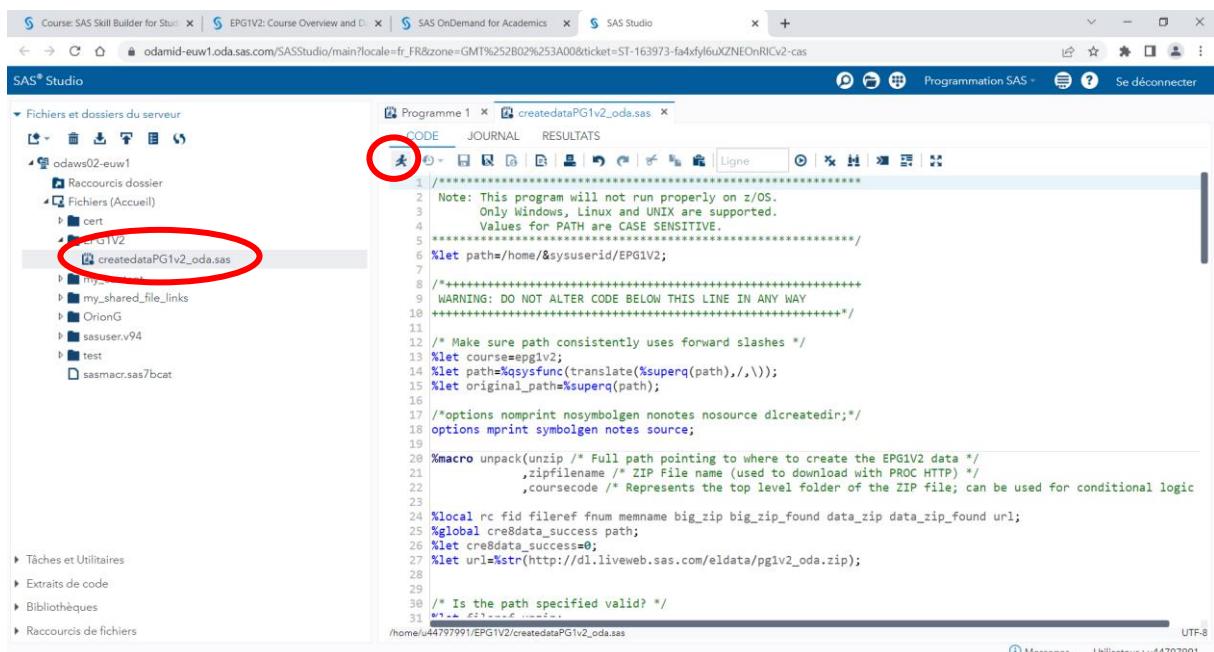
Sélectionner le fichier de programme createdataPG1v2_oda.sas



Ouvrir



Télécharger (vers le serveur)



Ouvrir le programme et l'exécuter

SAS Studio

Fichiers et dossiers du serveur

- odaws02-euw1
 - Raccourcis dossier
 - Fichiers (Accueil)
 - cert
 - EPG1V2
 - activities
 - data
 - demos
 - output
 - practices
 - createdataPG1v2_oda.sas
 - epg1v2.zip
 - setup.sas
 - my_content
 - my_shared_file_links
 - OrionG
 - sasuser.v94
 - test
 - sasmacro.sas7bcat
- Tâches et Utilitaires
- Extraits de code
- Bibliothèques
- Raccourcis de fichiers

Programme 1 createdataPG1v2_oda.sas

CODE JOURNAL RESULTATS DONNEES EN SORTIE

Table des matières

La procédure CONTENTS

	Répertoire
Libref	PG1
Moteur	V9
Nom physique	/home/u44797991/EPG1V2/data
Nom du fichier	/home/u44797991/EPG1V2/data
Numéro inode	131997723
Autorisation d'accès	non-rwx
Nom du propriétaire	u44797991
Taille du fichier	4KB
Taille de fichier (octets)	4096

#	Nom	Type de membre	Taille du fichier	Modifié(e) le
1	CLASS_BIRTHDATE	DATA	256KB	06/10/2022 09:00:44
2	CLASS_TEACHERS	DATA	256KB	06/10/2022 09:00:44
3	CLASS_TEST2	DATA	256KB	06/10/2022 09:00:44
4	CLASS_TEST3	DATA	256KB	06/10/2022 09:00:44
5	CLASS_UPDATE	DATA	256KB	06/10/2022 09:00:44
6	EU_OCC	DATA	512KB	06/10/2022 09:00:45
7	NP_CODELOOKUP	DATA	384KB	06/10/2022 09:00:45
8	NP_FINAL	DATA	256KB	06/10/2022 09:00:45
9	NP_LARGE PARKS	DATA	256KB	06/10/2022 09:00:45
10	NP_MULTIYR	DATA	7MB	06/10/2022 09:00:45
11	NP_SPECIES	DATA	8MB	06/10/2022 09:00:45
12	NP_SUMMARY	DATA	256KB	06/10/2022 09:00:45
13	NP_TRAFFIC	DATA	384KB	06/10/2022 09:00:45
14	NP_WESTWEATHER	DATA	1MB	06/10/2022 09:00:45
15	STORM 2017	DATA	256KB	06/10/2022 09:00:45

Messages : 1 Utilisateur : u44797991

Revenir dans l'onglet EPG1V2

Course Overview and Data Setup

- Welcome to the Course 1:00
- Demo: The SAS Programming Process 7:17
- Practicing in This Course 1:02
- Setting Up Your Practice Files (REQUIRED)

Using the SAS Virtual Lab

The SAS Virtual Lab already has the course files loaded for you. All you have to do is launch the lab and start using SAS Studio.

[Open](#) Instructions for using the SAS Virtual Lab

Using SAS OnDemand for Academics

[Open](#) Setup instructions for SAS OnDemand for Academics

Using your own SAS software

Start your SAS software. Click [Open](#) next to the SAS software interface you have, and follow the instructions to set up your data.

[Open](#) Setup Instructions for SAS Studio

[Open](#) Setup Instructions for SAS Enterprise Guide

[Open](#) Setup Instructions for SAS Windowing Environment

Text Version ▾

This is not a video. See the main content area.

Jump to...

Lesson 1: Essentials ▶

You are logged in as Grégoire de Lassence (Log out)

Get the mobile app

Lesson 1 : Essentials

The screenshot shows a web browser window with the following tabs open:

- Course: SAS Skill Builder for Stats
- EPGIV2: Lesson 1: Essentials
- SAS OnDemand for Academics
- SAS Studio

The main content area displays the SAS Virtual Learning Environment. The URL in the address bar is vle.sas.com/mod/scorm/player.php?a=8497¤torg=ORG-EPGIV2&scoid=19204.

The page title is "Virtual Learning Environment". The top right corner shows "English (United States) (en_us)".

The navigation path is: Dashboard / Courses / SAS Programming 1: Essentials / Lessons / Lesson 1: Essentials.

A button labeled "Exit activity" is located in the top right corner of the main content area.

The main content area displays "Lesson 1: Essentials".

The left sidebar contains a "Contents" tab and a "Search" tab. The "Contents" tab is active, showing the following structure:

- Lesson 1: Essentials**
 - Lesson Overview** 0:17
 - Using SAS Programming Tools**
 - SAS Programming Interfaces** 1:35
 - Demo: Programming with SAS Studio 4:25
 - Demo: Programming with SAS Enterprise Guide 4:38
 - Practice Using SAS Software**
 - Understanding SAS Syntax**
 - SAS Program Structure** 2:40
 - Activity
 - SAS Program Syntax** 0:56
 - Demo: Understanding SAS Program Syntax 2:12
 - Demo: Finding and Resolving Syntax Errors 2:12

Dans SAS Studio

The screenshot shows the SAS Studio interface. On the left, there's a sidebar with a tree view of files and folders. The main area has a code editor titled "Programme 1" with the file "setup.sas" open. The code in the editor is as follows:

```
1 ****  
2 /* DO NOT EDIT THE CODE BELOW */  
3 ****  
4 %let path=~/EPGIV2/data;  
5 libname PG1 "&path";  
6  
9 %include "&path/pg1v2.sas";  
10  
11 proc contents data=pg1._all_ nobs;  
12 run;
```

Ouvrir le programme setup qui sera à lancer à chaque début de session pour ces leçons.

The screenshot shows the SAS Studio interface with the 'RESULTS' tab active. The left sidebar shows a library tree with a node 'PG1' expanded, containing various datasets like CLASS_BIRTHDATE, CLASS_TEACHERS, etc. The main pane displays the 'La procédure CONTENTS' table and a list of files with their details.

#	Nom	Type de membre	Taille du fichier	Modifié(e) le
1	CLASS_BIRTHDATE	DATA	256KB	06/10/2022 11:33:38
2	CLASS_TEACHERS	DATA	256KB	06/10/2022 11:33:38
3	CLASS_TEST2	DATA	256KB	06/10/2022 11:33:38
4	CLASS_TEST3	DATA	256KB	06/10/2022 11:33:38
5	CLASS_UPDATE	DATA	256KB	06/10/2022 11:33:38
6	EU_OCC	DATA	512KB	06/10/2022 11:33:38
7	NP_CODELOOKUP	DATA	384KB	06/10/2022 11:33:38
8	NP_FINAL	DATA	256KB	06/10/2022 11:33:38
9	NP_LARGE PARKS	DATA	256KB	06/10/2022 11:33:38
10	NP_MULTIYR	DATA	7MB	06/10/2022 11:33:38
11	NP_SPECIES	DATA	8MB	06/10/2022 11:33:39
12	NP_SUMMARY	DATA	256KB	06/10/2022 11:33:38
13	NP_TRAFFIC	DATA	384KB	06/10/2022 11:33:39
14	NP_WESTWEATHER	DATA	1MB	06/10/2022 11:33:39
15	STORM_2017	DATA	256KB	06/10/2022 11:33:39

On obtient donc une bibliothèque PG1 qui correspond à un raccourci vers les données se trouvant dans le dossier ~/EPG1V2/data (dans certains code ou support, c'est s:/workshop/data)

The screenshot shows the SAS Studio interface with the 'RESULTS' tab active. The left sidebar shows a library tree with a node 'odaws01-euw1' expanded, containing a 'data' folder with various CSV and XLSX files. The main pane displays the 'La procédure CONTENTS' table and a list of files with their details.

#	Nom	Type de membre	Taille du fichier	Modifié(e) le
1	CLASS_BIRTHDATE	DATA	256KB	06/10/2022 11:33:38
2	CLASS_TEACHERS	DATA	256KB	06/10/2022 11:33:38
3	CLASS_TEST2	DATA	256KB	06/10/2022 11:33:38
4	CLASS_TEST3	DATA	256KB	06/10/2022 11:33:38
5	CLASS_UPDATE	DATA	256KB	06/10/2022 11:33:38
6	EU_OCC	DATA	512KB	06/10/2022 11:33:38
7	NP_CODELOOKUP	DATA	384KB	06/10/2022 11:33:38
8	NP_FINAL	DATA	256KB	06/10/2022 11:33:38
9	NP_LARGE PARKS	DATA	256KB	06/10/2022 11:33:38
10	NP_MULTIYR	DATA	7MB	06/10/2022 11:33:38
11	NP_SPECIES	DATA	8MB	06/10/2022 11:33:39
12	NP_SUMMARY	DATA	256KB	06/10/2022 11:33:38
13	NP_TRAFFIC	DATA	384KB	06/10/2022 11:33:39
14	NP_WESTWEATHER	DATA	1MB	06/10/2022 11:33:39
15	STORM_2017	DATA	256KB	06/10/2022 11:33:39

Bibliographie

Pour préparer la certification de programmation de base et celle de programmation avancée, le livre : *SAS, 4^{ième} édition* - Sébastien Ringuédé aux Edition Eyrolles est fortement recommandé.

Site Web

<http://www.sas.com> (site commercial de SAS)

<http://support.sas.com> (site des utilisateurs de SAS)

<https://communities.sas.com/welcome>