

EXPLORER TWITTER AVEC SAS® VISUAL ANALYTICS



Dans notre monde interconnecté, il est plus important que jamais de comprendre non seulement des détails sur vos données, mais aussi la façon dont ses différentes parties sont reliées les unes aux autres. Les réseaux sociaux révèlent des détails souvent surprenants sur ce que les gens pensent de vos produits ou services, comment ils sont liés à d'autres communautés sociales qui pourraient influencer sur votre entreprise, et même comment cela peut aider à la prise de décisions, comme par exemple connaître les personnes à cibler dans une prochaine campagne de marketing.

Cet article s'inscrit dans cette optique et présente, en quelques étapes, comment créer votre premier rapport **SAS Visual Analytics**, solution pour des décisions plus rapides et plus intelligentes, à partir de données extraites du **réseau social Twitter**.



Caractéristiques :

Catégories : Visual Analytics
OS : Windows, Unix
Version : SAS® 9.4M2 / SAS® Visual Analytics 6.4, 7.1
Vérifié en Décembre 2014



Sommaire

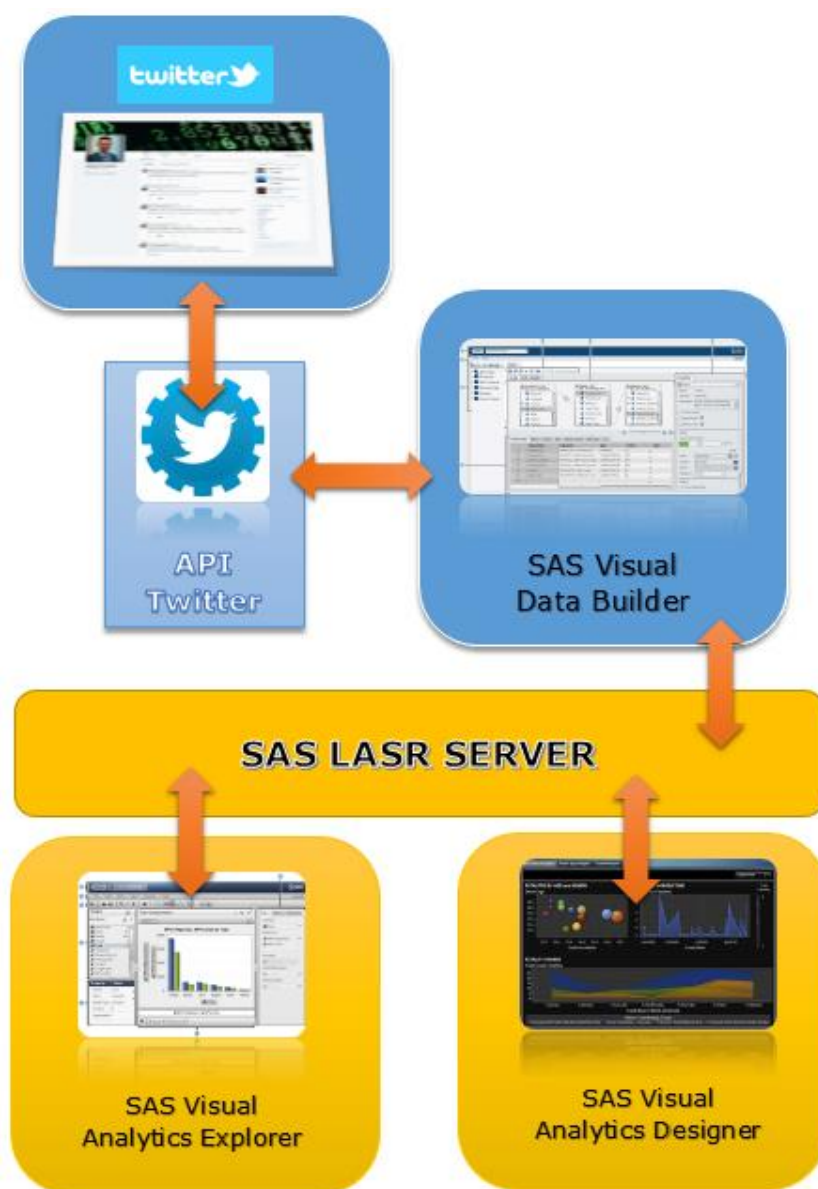
1.	Explorer Twitter avec SAS® Visual Analytics	1
2.	Fonctionnement	2
3.	Pré-requis et composants SAS	3
4.	Scénario	3
5.	Préparation des données.....	3
6.	Exploration dans les données.....	11
7.	Liens utiles.....	17
8.	Conclusion	17

1. FONCTIONNEMENT

Le schéma ci-dessous présente les interactions entre Twitter et SAS Visual Analytics.

Les données sont importées depuis Twitter par l'intermédiaire de l'API Twitter (Twitter's v1.1 REST API: <https://dev.twitter.com/docs/using-search>) En effet, cette API permet à n'importe quelle application de se connecter au réseau social et d'effectuer des recherches.

Les données en provenance de Twitter sont ensuite stockées dans une table LASR pour être explorées et analysées dans SAS Visual Analytics.



2. PRÉREQUIS ET COMPOSANTS SAS

Pour suivre les étapes de cet article, il est nécessaire d'avoir les prérequis suivants :

- SAS 9.4m2
- SAS Visual Analytics 6.4 ou 7.1
- Vous devez posséder un compte Twitter afin de permettre à SAS d'utiliser l'API.

3. SCENARIO

Comprendre comment un hashtag voyage dans la tweetsphère est un élément essentiel pour mener une campagne hashtag efficace. SAS Visual Analytics est un outil essentiel pour étudier, surveiller, analyser et archiver des hashtags. SAS Visual Analytics **permet** de mettre en valeur l'information.

Aussi, pour illustrer l'intégration de données Twitter et leur utilisation dans SAS Visual Analytics, nous allons nous connecter à Twitter via l'API pour collecter les Tweets relatifs au hashtag #Porsche.

Une fois les Tweets rapatriés dans le serveur LASR, nous créons une exploration de ces données. Cette exploration visuelle et interactive, ainsi que la représentation graphique des données, permet aux utilisateurs métier de détecter des phénomènes.

Après la lecture de cet article, vous serez capable de :

- Lier SAS Visual Analytics à Twitter
- Importer des données
- Explorer les résultats.

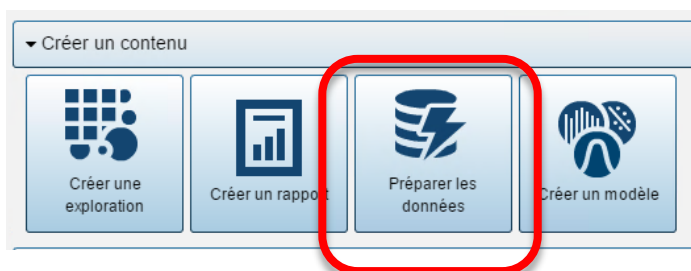
4. PRÉPARATION DES DONNÉES

La première étape concerne la préparation des données, c'est-à-dire l'import des données Twitter dans une table SAS.

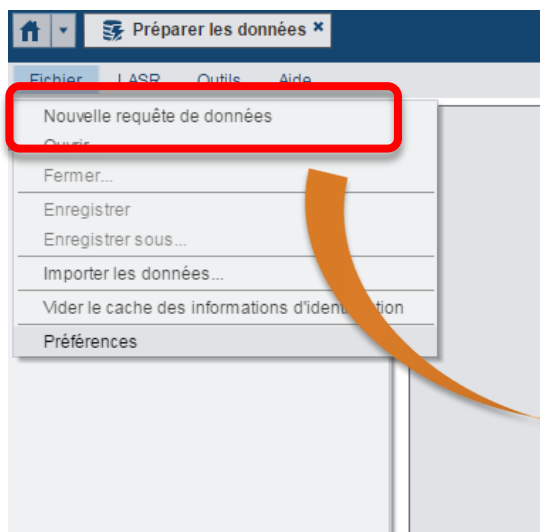
Connectons-nous à SAS Visual Analytics Hub :

<http://<votre serveur SAS>/SASVisualAnalyticsHub>

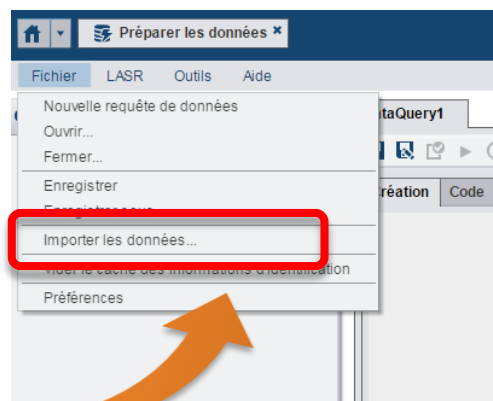
Une fois connecté, sélectionnons « Créer une requête de données », soit depuis le hub, soit depuis le menu Fichier :



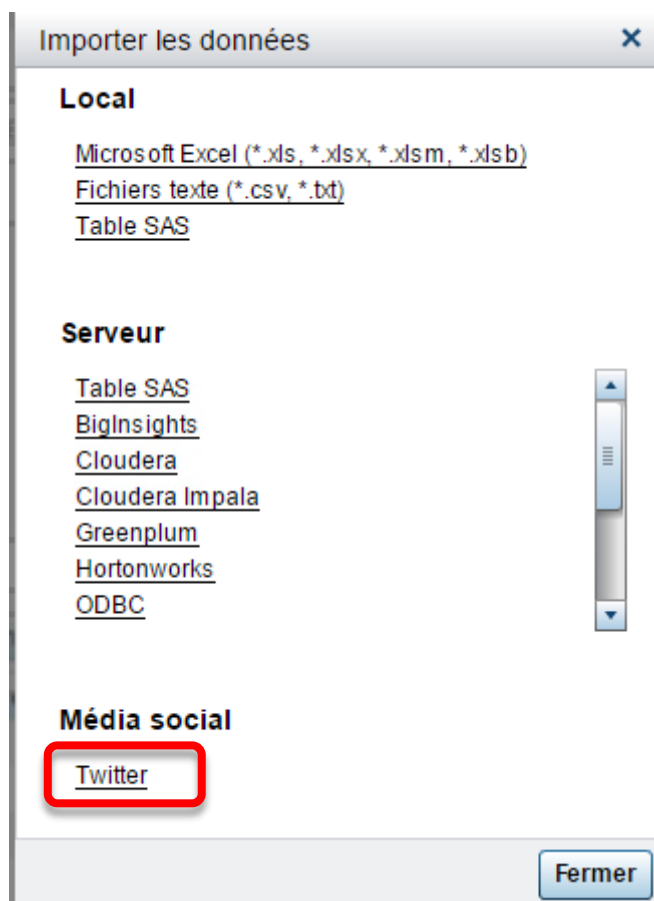
Nouvelle requête de données



Importer des données :



Dans la fenêtre « Importer les données », sélectionnons Twitter :





Vous devez alors autoriser SAS Visual Analytics à utiliser un compte Twitter. L'idéal est donc de créer un compte Twitter utilisé pour accéder aux données à travers l'API. Dans les exemples de cet article, j'utilise mon compte Twitter personnel.

Cette étape sera à réaliser une seule fois. L'application est autorisée à :

- Lire les Tweets de votre fil,
- Regarder vos abonnements.



Concernant la sécurité sur Twitter, l'application SAS Visual Analytics n'est pas capable de :

- Suivre de nouvelles personnes,
- Mettre à jour votre profil,
- Poster des Tweets pour vous,
- Accéder à vos messages privés,
- Voir votre mot de passe Twitter.

Une fois que vous avez autorisé l'application SAS, vous pouvez retourner dans SAS Visual Analytics :

The dialog box titled "Importer les données Twitter" contains the following fields and options:

- Terme de recherche : [Empty text box]
- Maximum de tweets à retourner : 2000
- Table LASR : [Empty text box]
- Nom : Twitter_Output
- Description : Importé le dimanche 31 août 2014 18:23:23 GMT+0200 vers "Twitter_Output" par "sasdemo"
- Options avancées (expanded)
- Proxy Server (expanded)
- Buttons at the bottom: "Supprimer les jetons d'autorisation", "OK", and "Annuler".

Note : A tout moment vous pouvez supprimer les jetons d'autorisation à votre compte Twitter en cliquant sur le bouton **Supprimer les jetons d'autorisation**

Vous pouvez également interdire l'accès à **SAS Visual Analytics** depuis votre compte Twitter :



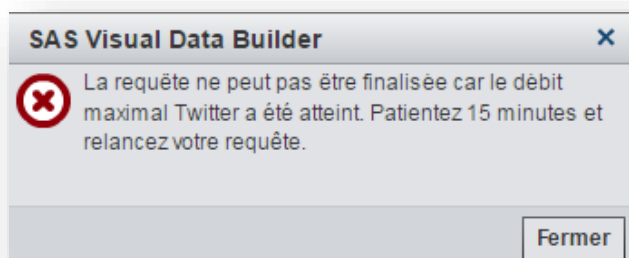
Maintenant que nous avons autorisé le lien entre SAS Visual Analytics et Twitter, nous pouvons importer les données :

- Nous allons importer les Tweets contenant le terme : #PORSCHE
- La requête va s'effectuer sur 20 000 Tweets (la valeur par défaut est 2 000)
- Les données seront stockées dans une table LASR appelée « Twitter_Porsche » et stockée dans /Shared Data/LASR (emplacement par défaut)

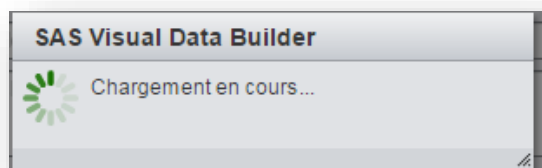


- Aujourd'hui, l'import Twitter ne fonctionne qu'en anglais.
- Les résultats importés dans SAS Visual Analytics et ceux retournés par une recherche Twitter ne sont pas rigoureusement identiques. Les deux moteurs de recherche utilisent des mécanismes différents pour rapatrier les Tweets. Pour plus d'informations sur ce sujet, vous pouvez consulter la documentation Twitter <https://dev.twitter.com/rest/reference/get/search/tweets>

Le nombre de requêtes que vous pouvez faire dans Twitter est limité. Si vous dépassez la limite, le message d'erreur suivant s'affiche :



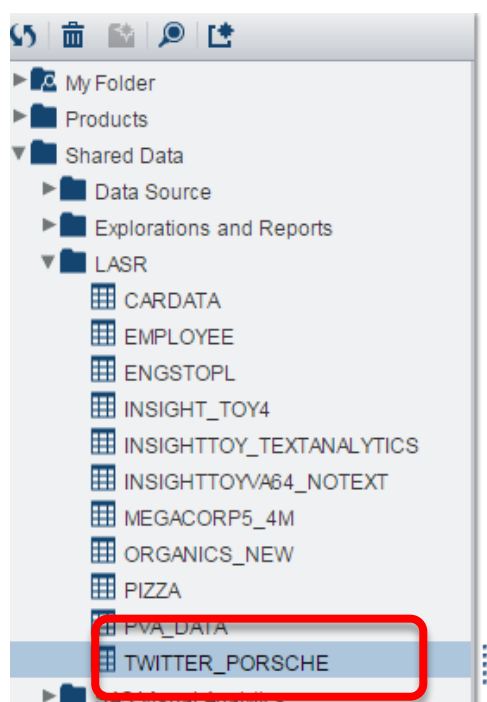
Si vous souhaitez en savoir plus sur cette limitation, vous pouvez consulter la documentation Twitter à ce sujet disponible à l'adresse suivante : <https://dev.twitter.com/rest/public/rate-limiting>
Si tout va bien, nous devons avoir ce message à l'écran :



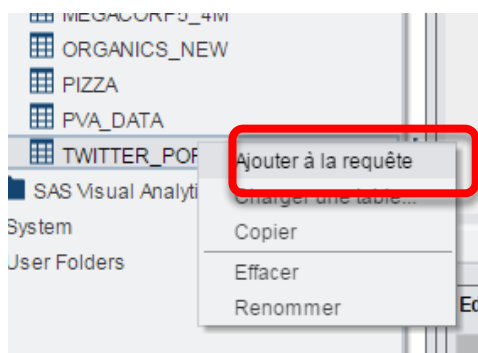
La durée du chargement varie en fonction du nombre de Tweets à retourner.

Maintenant que nos données ont été rapatriées, nous allons les « préparer » en créant une nouvelle requête dans SAS Visual Data Builder :

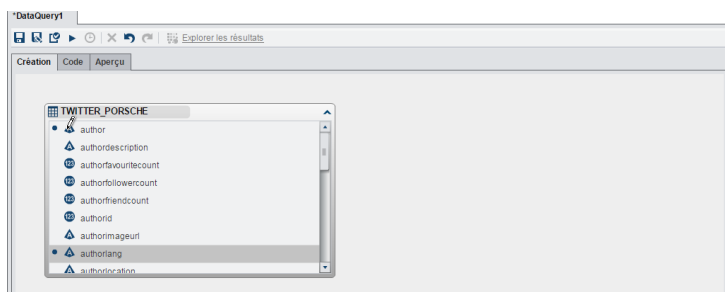
Localiser votre nouvelle table



Puis sélectionner « Ajouter à la requête » :



En cliquant sur les noms de colonnes, nous sélectionnons les données qui nous intéressent :



- Author
- Authorlang
- Body
- Isretweet
- Publisheddatetime
- Tags
- Retweetcount

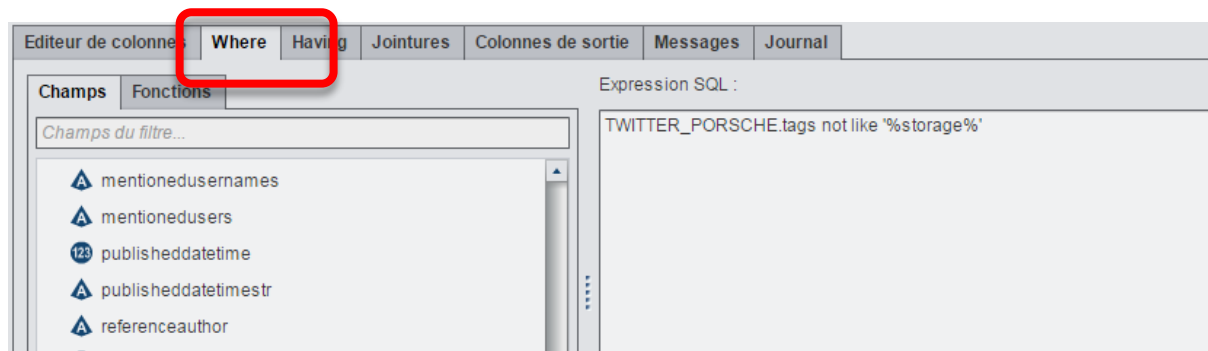
Editeur de colonnes		Where	Having	Jointures	Colonnes de sortie	Messages	Journal		
		Nom de colonne	Expression	Type	Format	Libellé	Agrégations		
1		author	TWITTER_PORSCHE.author	CHARACTER(128)	\$128.				
2		authorlang	TWITTER_PORSCHE.authorlang	CHARACTER(2)	\$2.				
3		body	TWITTER_PORSCHE.body	CHARACTER(1024)	\$1024.				
4		isretweet	TWITTER_PORSCHE.isretweet	NUMERIC	BEST12.				
5		deviceinfo	TWITTER_PORSCHE.deviceinfo	CHARACTER(1024)	\$1024.				
6		publisheddatetime	TWITTER_PORSCHE.publisheddatetime	NUMERIC	BEST12.				
7		tags	TWITTER_PORSCHE.tags	CHARACTER(128)	\$128.				
8		retweetcount	TWITTER_PORSCHE.retweetcount	NUMERIC	BEST12.				
+									

Après avoir sélectionné ces colonnes, nous devons modifier le format de la colonne *Publishdatetime*.

Il s'agit d'une valeur numérique à transformer en champs de type date :

5	publisheddatetime	TWITTER_PORSCHE.publisheddatetime	DATETIME	DATETIME.
---	-------------------	-----------------------------------	----------	-----------

Il est également possible de conditionner les résultats à retourner en utilisant l'onglet Where :

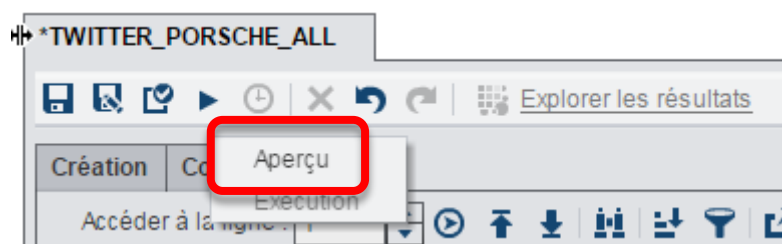


Pour filtrer les résultats, il suffit d'écrire directement l'expression SQL à appliquer :

Dans notre exemple, nous voulons exclure les Tweets contenant le tag « storage ». L'expression SQL à utiliser est la suivante

```
TWITTER_PORSCHE.tags not like '%storage%'
```

Une fois les modifications apportées, nous pouvons lancer l'aperçu afin de vérifier si nos données et le format de date choisis nous conviennent :



Création Code Aperçu							
Accéder à la ligne : 1							
	author	authorlang	body	isretweet	publishedat...	tags	retweetcount
1	ajrf993GT2	en	RT @GT3_Wheel_House: 1986 #Porsche 911 Coup...	1	01SEP2014	Porsche	2
2	ajrf993GT2	en	RT @GT3_Wheel_House: 2010 #Porsche 911 997 G...	1	01SEP2014	Porsche	6
3	mossadkqb	en	RT @i_s_a_a_c: Next week!!! #porsche #911 #t...	1	01SEP2014	porsche...	1
4	PFFDE_EN	en	#Porsche, #911, #964 :Chassis 964 http://t.co/cBvLrv...	0	01SEP2014	Porsche	0
5	WheelDynamics_	en	RT @RabyPorsche: Here is a previous #Porsche blo...	1	01SEP2014	Porsche	2
6	MattJimmyTaylor	en	Fantastic day driving this beauty #porsche911 #porsc...	0	01SEP2014	porsche...	0
7	mattcollins907	en	RT @GT3_Wheel_House: 2010 #Porsche 911 997 G...	1	01SEP2014	Porsche	6
8	panamera_tweets	en	RT @Porsche_Ohio: Fall is upon us! Where are you a...	1	01SEP2014	Porsche...	2
9	WheelDynamics_	en	RT @Porsche_Ohio: Fall is upon us! Where are you a...	1	01SEP2014	Porsche...	2
10	panamera_tweets	en	RT @RabyPorsche: Here is a previous #Porsche blo...	1	01SEP2014	Porsche	2
11	panamera_tweets	en	RT @porschens: Fall is upon us! Where are you and ...	1	01SEP2014	Porsche...	1
12	95octanecom	en	#Porsche 911 Turbo S: Exclusive GB Edition http://t...	0	01SEP2014	Porsche	0
13	CoastalSpeed	es	Amazing CARRERA 4 #porsche #porsche911 #porsc...	0	01SEP2014	porsche...	0
14	MitchChambliss	en	RT @GT3_Wheel_House: 2010 #Porsche 911 997 G...	1	01SEP2014	Porsche	6
15	RabyPorsche	en	Here is a previous #Porsche blog post Porsche 996 -...	0	01SEP2014	Porsche	2
16	vierenzestig	nl	Niets te doen dit weekend? RS Fest! http://t.co/3LH...	0	01SEP2014	porsche	0
17	Porsche_Ohio	en	Fall is upon us! Where are you and your #Porsche tra...	0	01SEP2014	Porsche...	2

Maintenant que les données sont prêtes, il faut les enregistrer dans une table LASR afin qu'elles puissent être exploitées :

Entrées **Sorties** Propriétés

Vider

Sortie

Vider

Table : DATA_TWITTER_PORS

Emplacement : /Shared Data/LASR

Bibliothèque : /Products/SAS Visual Analytic

Partition par : (néant)

Puis validons la requête :

TWITTER_PORSCHE_ALL

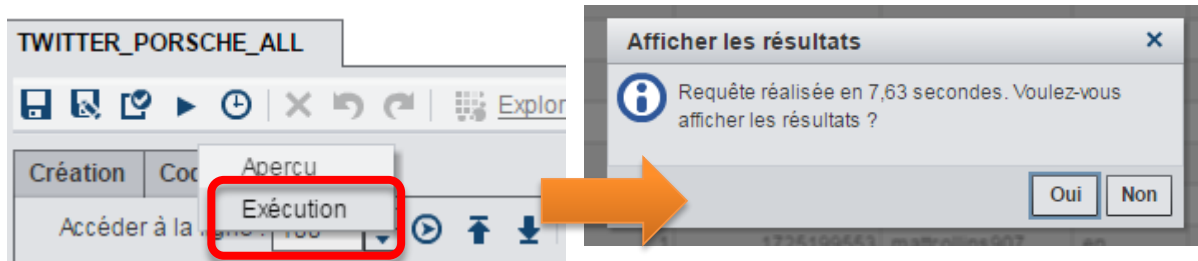
Création Code Aperçu

Valider

La requête est valide.

Fermer

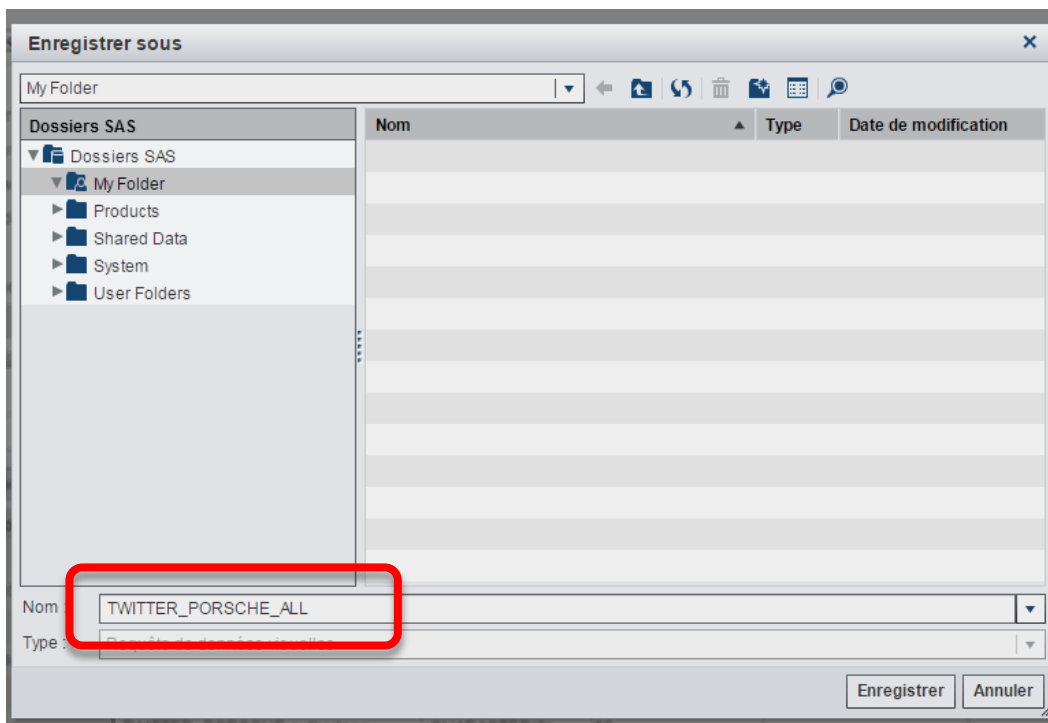
Enfin, dernière étape de la préparation des données, sauvegardons puis exécutons la requête :



Notre table contient 8977 lignes.



Enregistrons maintenant notre requête :



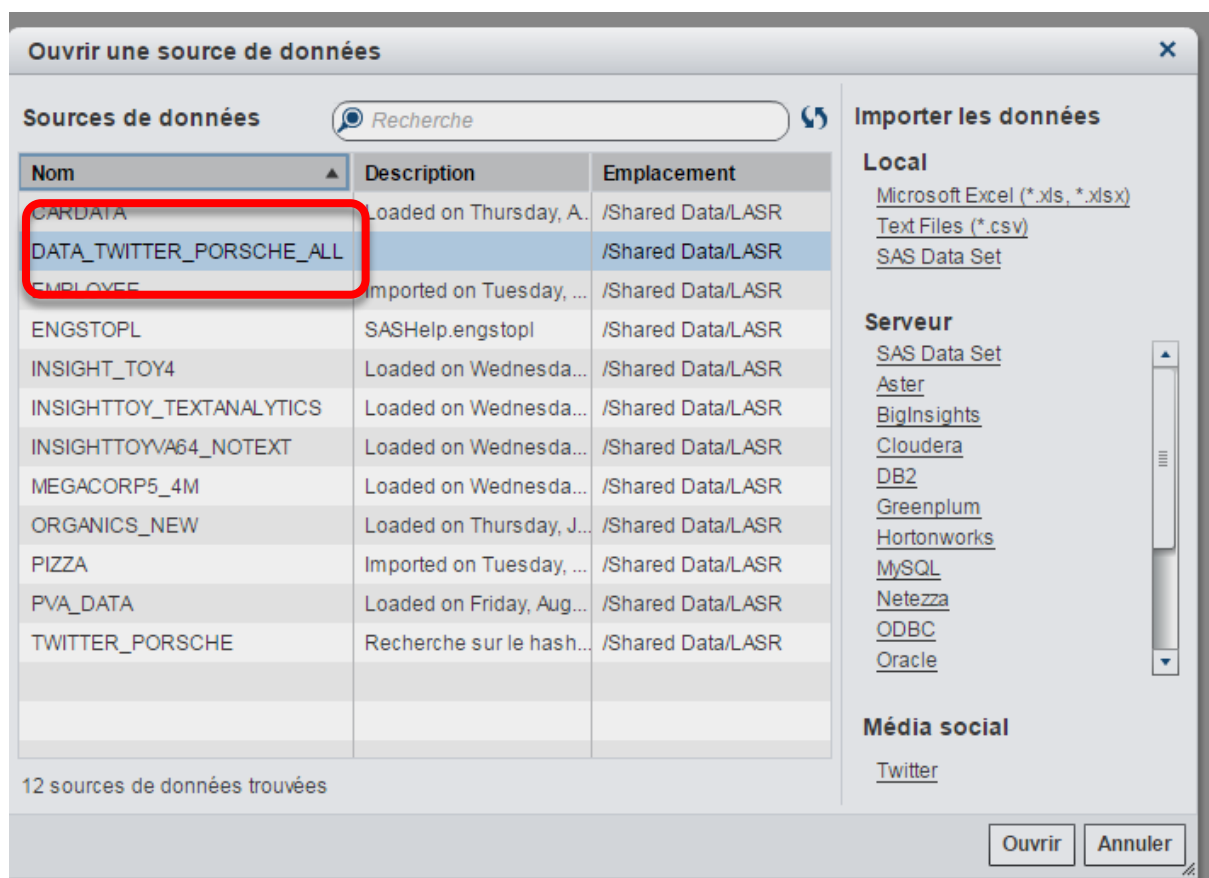
Voilà, l'étape de préparation des données est terminée.
Passons maintenant à l'exploration.

5. EXPLORATION DANS LES DONNÉES

Toujours depuis le Visual Analytics HUB, nous allons créer une nouvelle exploration :

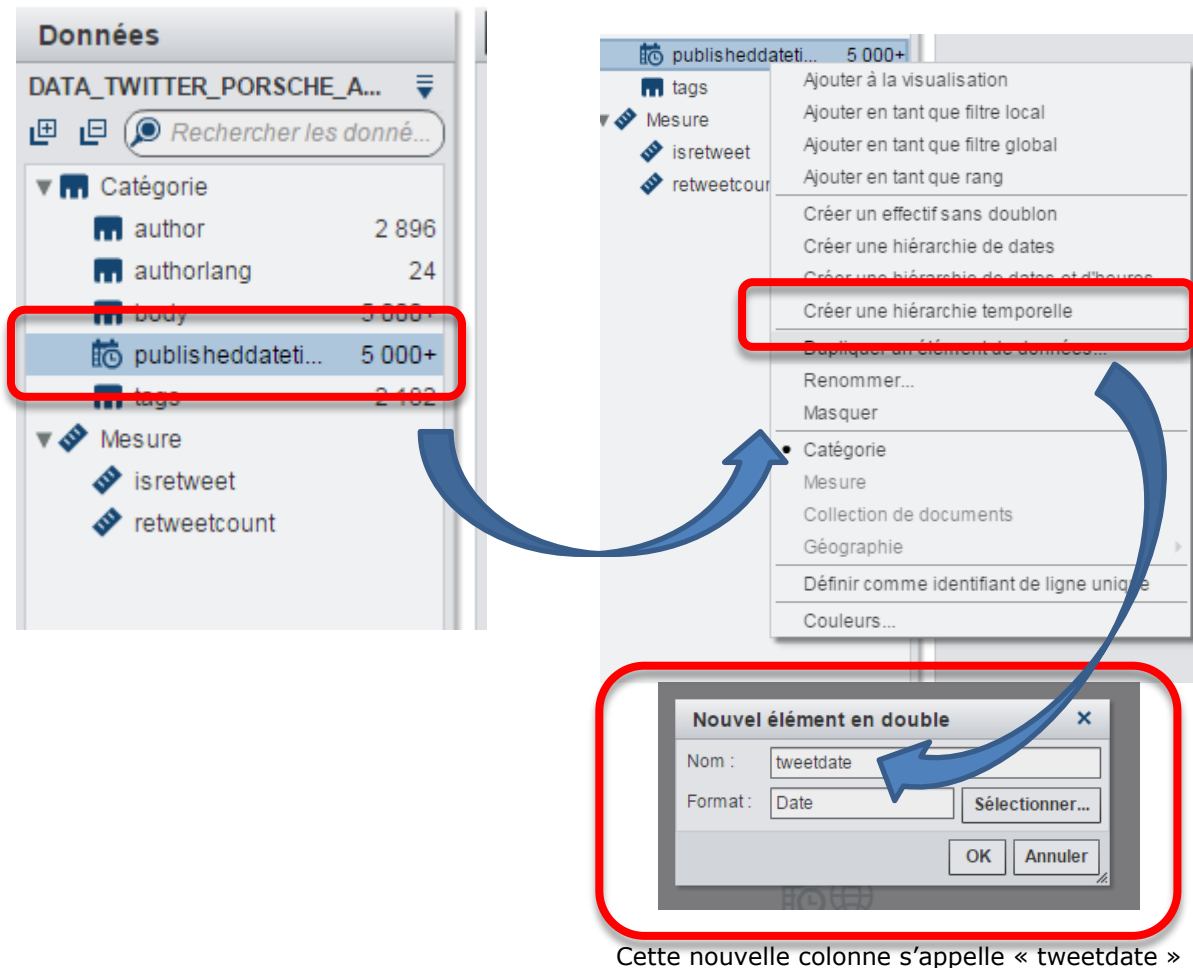


La première étape consiste à choisir une source de données.
Sélectionnons le résultat de notre préparation, la table DATA_TWITTER_PORSCHE_ALL :

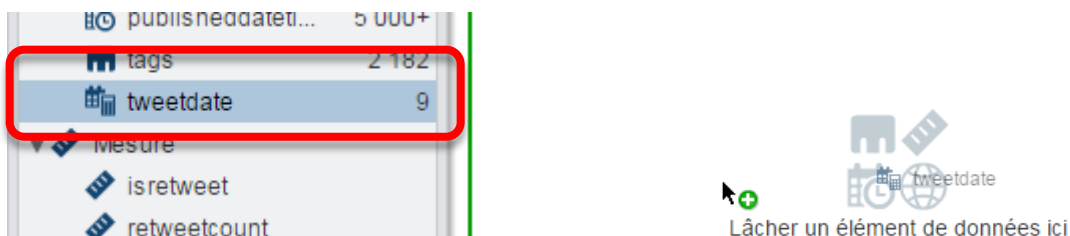


Comme nous l'avons vu lors de la préparation des données, la colonne *publisheddatetime* est un champ de type DATETIME contenant la DATE du Tweet au format JOUR MOIS ANNEE HEURE MINUTES SECONDE.

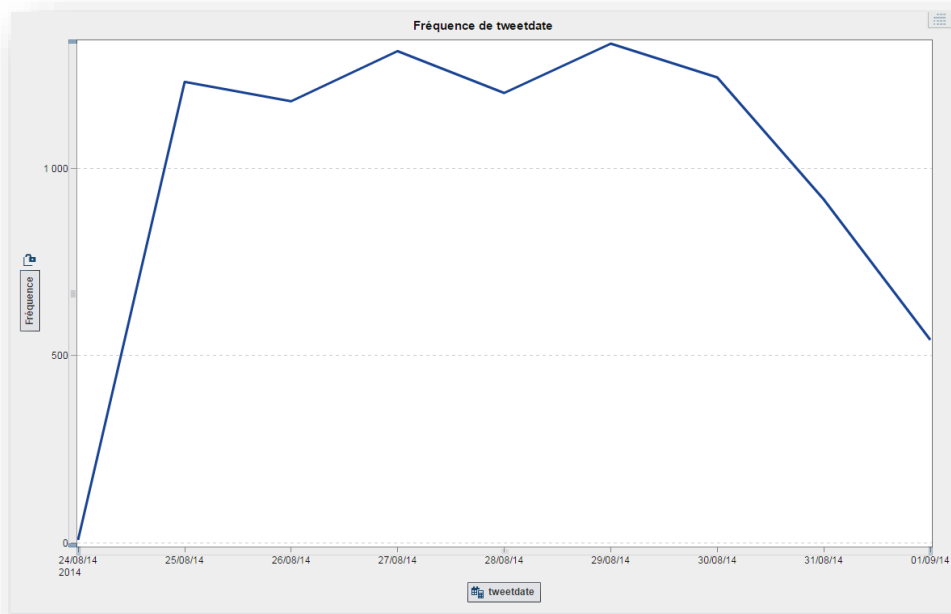
Afin de créer un graphique, l'idée est de dupliquer cette colonne pour créer une colonne DATE au format JOUR MOIS ANNEE :



Maintenant que nos données sont prêtes pour l'exploration, nous allons glisser la variable de catégorie « tweetdate » vers la fenêtre de visualisation :

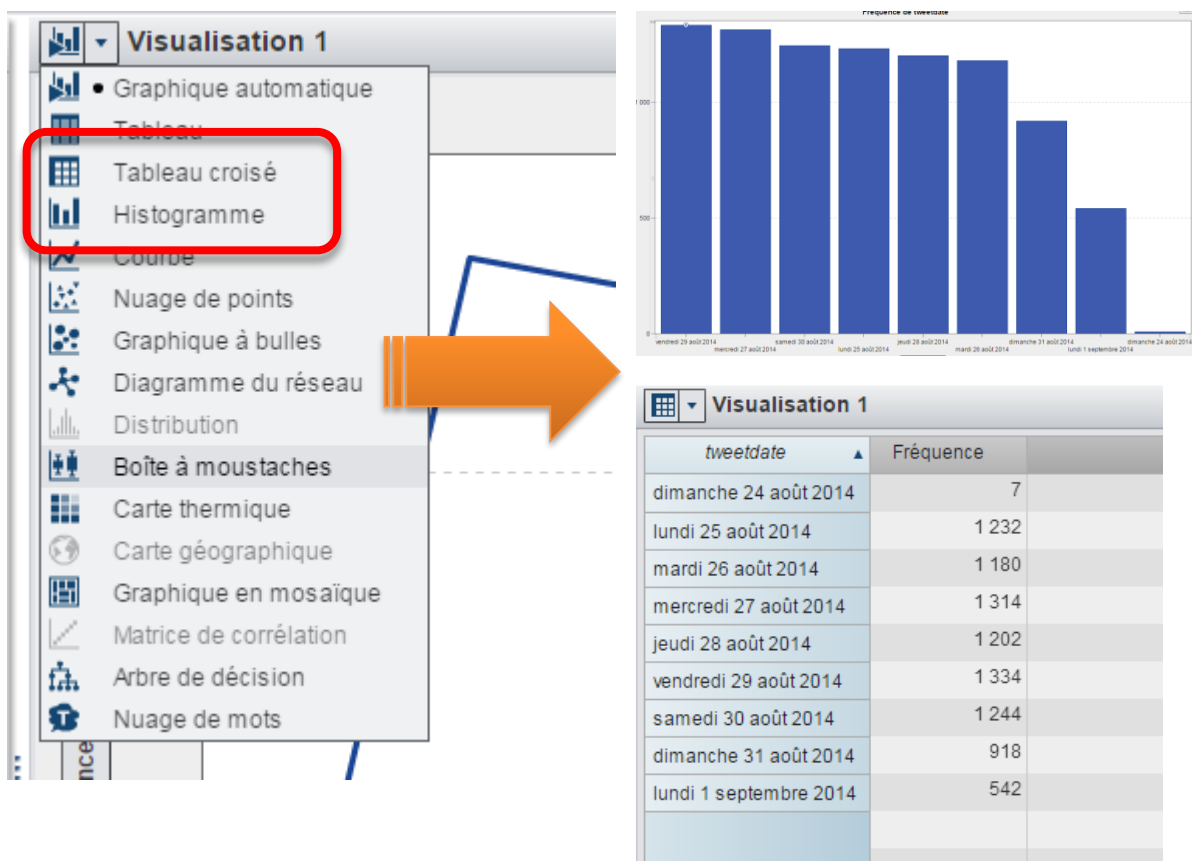


Un premier graphique, simple, est ainsi généré :



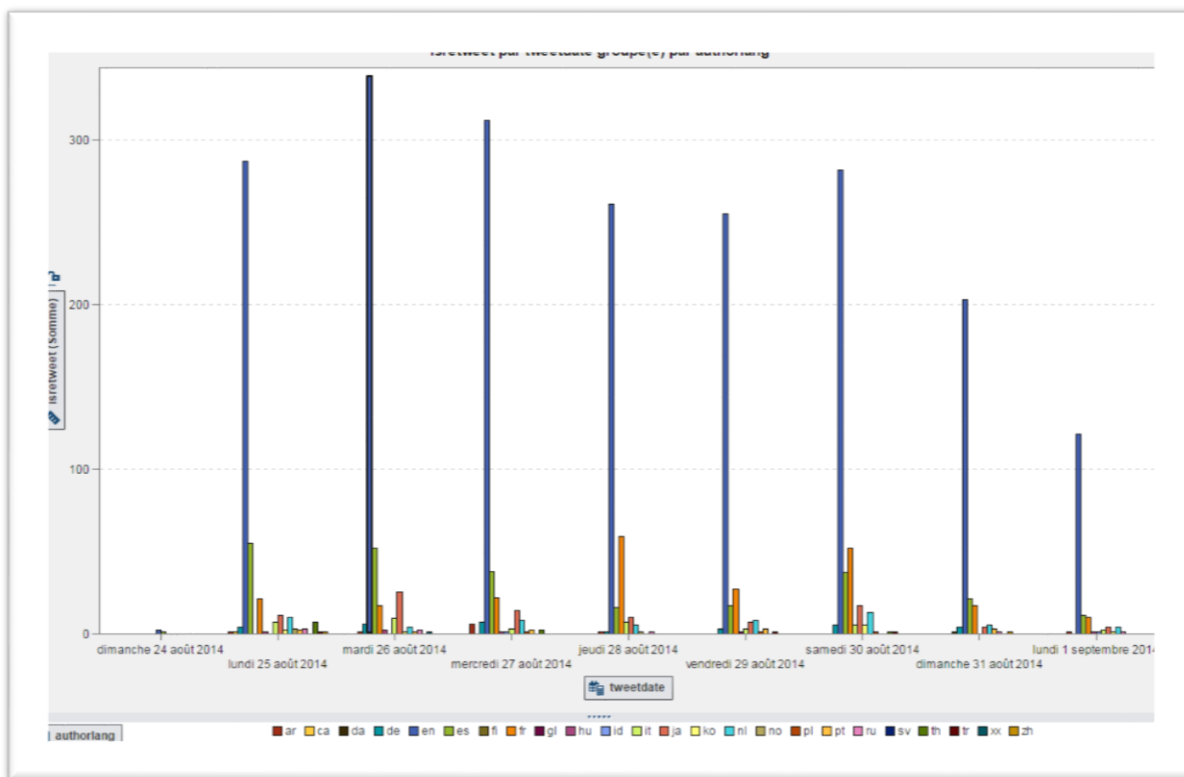
Ce graphique présente le nombre de « Tweets » contenant le hashtag « #porsche » entre le 24 août et le 1^{er} septembre. Sa lecture est simple. Nous constatons, par exemple, une chute du nombre de Tweets à partir du 1^{er} septembre.

Il est possible de modifier le type de graphique via le sélecteur présent au-dessus du graphique :

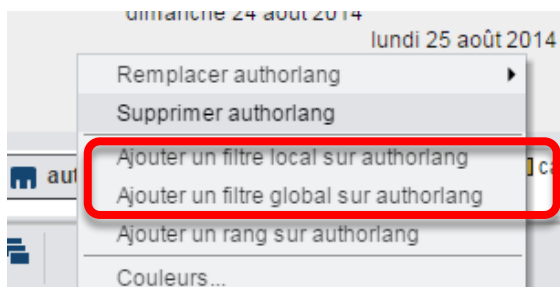
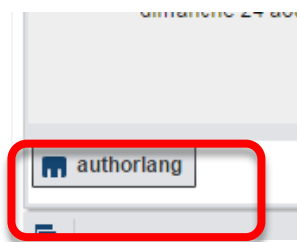


Il est également possible d'ajouter des données aux graphiques afin d'affiner les informations recherchées :

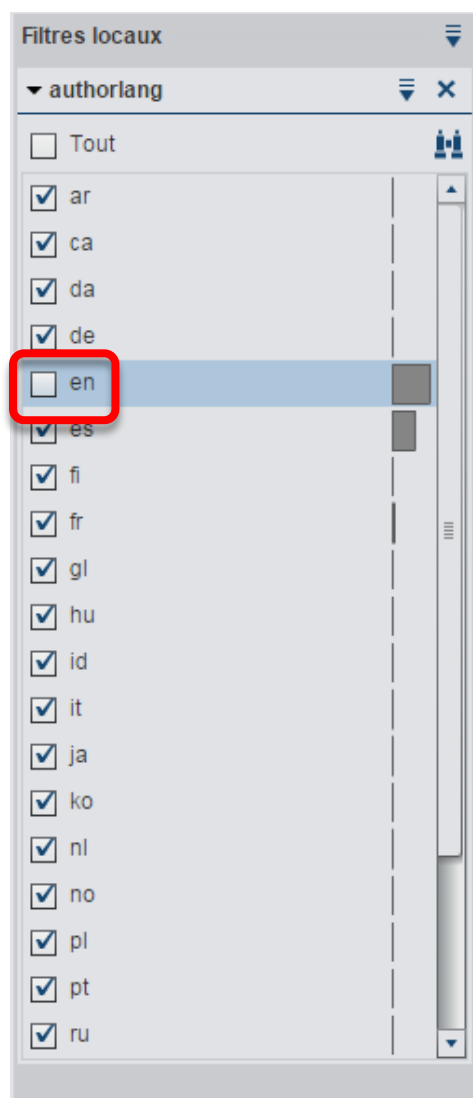
Par exemple, en ajoutant la colonne mesure « isretweet » et la catégorie « authorlang », nous créons ainsi un graphique indiquant le nombre de Tweets retweetés par langue :



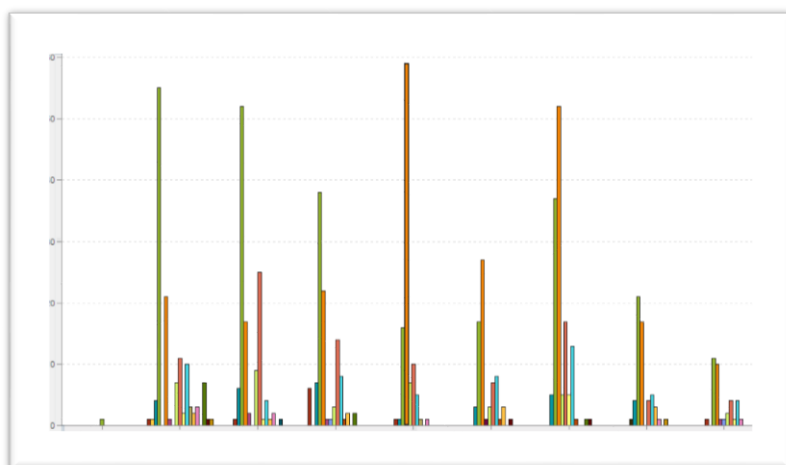
Il est également possible de filtrer ces résultats :



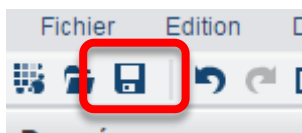
Nous allons exclure les Tweets anglais afin d'affiner notre graphique.



La mise à jour de la visualisation est automatique :



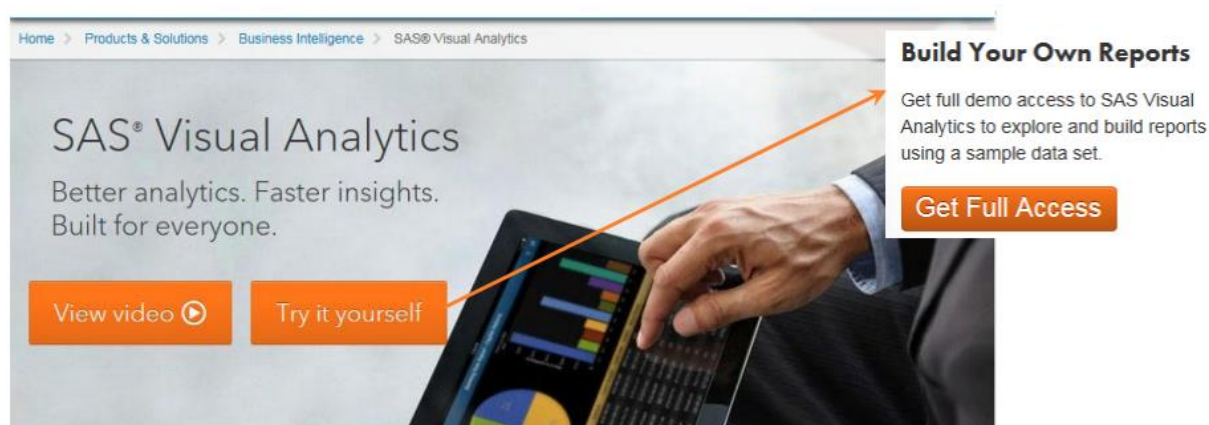
Maintenant que notre exploration est terminée, nous pouvons la sauvegarder.



6. LIENS UTILES

Quelques liens utiles pour aller plus loin :

Le démonstrateur SAS Visual Analytics :



http://www.sas.com/en_us/software/business-intelligence/visual-analytics.html

What's New in SAS(R) 9.4

<http://support.sas.com/documentation/cdl/en/whatsnew/64788/HTML/default/viewer.htm#n11x3oquj3in1sn10mctyzmugzvtv.htm>

SAS VISUAL ANALYTICS 6.4 USER'S GUIDE

<HTTP://SUPPORT.SAS.COM/DOCUMENTATION/CDL/EN/VAUG/67270/PDF/DEFAULT/VAUG.PDF>

Enfin, si vous souhaitez en savoir plus sur l'API Twitter :

<https://dev.twitter.com/overview/documentation>

7. CONCLUSION

Cet article est une introduction à ce qu'il est possible de faire avec SAS Visual Analytics et Twitter. SAS Visual Analytics propose des techniques d'exploration et d'analyse des données à la portée de chaque utilisateur afin d'améliorer les prises de décision et apporte des réponses rapides à des questions complexes.

Les utilisateurs métiers peuvent aller plus loin, établir des corrélations, repérer des tendances, identifier des opportunités et facilement diffuser les résultats. Tout cela avec une moindre sollicitation du service informatique. En effet, les utilisateurs bénéficient d'un nouveau mode d'accès à l'information.

Nicolas HOUSSET

Consultant Support Clients SAS France