

Ce que j'ai appris des défis hebdomadaires de Friends of SAS

Conférence du Club des utilisateurs SAS de Québec,
automne 2017

Un seul objectif

Éveiller votre curiosité pour les
défis de Friends of SAS

Friends of SAS

- <https://sas.influitive.com/>
- Site géré l'équipe par SAS Canada
- Apport de la communauté des utilisateurs
- Volets promotion, information et éducation
 - Tip of the Day!
 - Tips and Tricks
 - Trivia Tuesday
 - Série thématique (ex. Semaine contre la fraude)
 - Et Weekly: Posed Programming Problem

Trivia Tuesday

- Quelques questions sur un même thème
- Soumis par des utilisateurs
- Avec choix de réponses

Weekly: Posed Programming Problems

- Tous les mercredis depuis au moins octobre 2015
- Défi soumis par un utilisateur
 - Idéalement avec une solution
- Niveau de difficulté variable
- Solutions diffusées sur SAS Canada Community
 - <http://sascanada.ning.com>, groupe Friends of SAS Programming Problems Discussions

Quelques exemples

Problème du 15 février 2017

Spécial St-Valentin



1. IF 'TED' < 'ALICE' < 'BOB' < 'CAROL'
THEN PUT 'TRUE LOVE';
2. IF 'TED' < ('ALICE' < 'BOB') < 'CAROL'
THEN PUT 'TRUE LOVE';
3. IF ('TED' < 'ALICE') < ('BOB' < 'CAROL')
THEN PUT 'TRUE LOVE';

- Quelles lignes affichent « TRUE LOVE »?

Pas à pas, ligne 1

'TED' < 'ALICE' < 'BOB' < 'CAROL'

La série de < est évaluée d'un coup et donne faux

Donc pas de TRUE LOVE

Pas à pas, ligne 2

'TED' < ('ALICE' < 'BOB') < 'CAROL'

L'opération entre parenthèses donne vrai.

Pour SAS, un booléen est une variable numérique,
où Faux = 0 et Vrai = 1

'TED' < 1 < 'CAROL' ≠ vrai

NOTE : Données numériques incorrectes, 'TED'

NOTE : Données numériques incorrectes, 'CAROL'

Donc pas de TRUE LOVE

Pas à pas, ligne 2

('TED' < 'ALICE') < ('BOB' < 'CAROL')

La parenthèse de gauche donne faux, donc 0

La paranthèse de droite donne vrai, donc 1

-> 0 < 1 = vrai

Donc TRUE LOVE

Ce que j'ai appris

- Importance de savoir que pour SAS, il n'y a que des chiffres et des lettres

Problème du 19 octobre 2016

Quand 6,4 n'égal pas 6,4!



- Questionnaire pour peindre une pièce
- Chaque couleur de 1 à 10
- 5 personnes donnent leur avis
- Formule du score final :
 - 10 % de Francis + 10 % de Leslie + 50 % de Kelly + 30 % de Dave – 10 % Chester
 - Kelly a un droit de veto pour les couleurs qu'elle n'aime pas (score de 1)
 - En cas d'égalité, Dave décide
- Quel est le top 3 des couleurs?

Opinion de chacun et résultat

Couleur	Francis	Leslie	Chester	Kelly	Dave	Score	
Blue	8	5	7	8	4	5,8	
Grey	7	3	6	6	8	5,8	
Brown	8	1	8	4	5	3,6	
Green	6	10	2	7	5	6,4	
Red	1	5	4	9	5	6,2	
Pink	1	8	1	5	10	6,3	
Yellow	6	5	10	4	2	2,7	
Black	10	7	3	8	4	6,6	
Orange	5	6	6	10	1	5,8	
Purple	6	3	4	7	8	6,4	
Bronze	8	1	2	7	6	6,0	
Silver	7	4	7	1	9	0	Veto de Kelly
Gold	9	3	5	9	4	6,4	
Teel	3	9	6	1	7	0	Veto de Kelly
Beige	5	5	9	7	9	6,3	

Identification du top 3

```
...  
    score = 0,1*F+0,1*L+0,5*K+0,3*D-0,1*C;  
...  
proc sort data=couleur;  
    by descending score descending dave;  
run;  
data top3;  
    set couleur;  
    by descending score descending dave;  
    if _n_ le 3;  
run;  
proc print data=top3;
```

Top 3

Couleur	Francis	Leslie	Chester	Kelly	Dave	Score	Rang
Black	10	7	3	8	4	6,6	1
Green	6	10	2	7	5	6,4	2
Gold	9	3	5	9	4	6,4	3
Purple	6	3	4	7	8	6,4	

- Triple égalité en deuxième place, donc le choix revient à Dave.
- Pourquoi « Purple » est en 4^e place?
- Pourquoi Proc sort n'aime pas Dave?

Top 3 dans Excel

Couleur	Francis	Leslie	Chester	Kelly	Dave	Score	Rang
Black	10	7	3	8	4	6,6	1
Purple	6	3	4	7	8	6,4	2
Gold	9	3	5	9	4	6,4	3
Green	6	10	2	7	5	6,4	

- Pourquoi « Green » est en 4^e place?
- Pourquoi le tri n'aime pas Dave?

Une question de décimales

- $\text{Score} = 0,1 * F + 0,1 * L + 0,5 * K + 0,3 * D - 0,1 * C$
- Une fraction est conservée avec une certaine précision (voir *Numerical Accuracy in SAS Software*)
- $\text{Score modifié} = F + L + 5 * K + 3 * D - 1 * C$

Couleur	Francis	Leslie	Chester	Kelly	Dave	Score	Rang
Black	10	7	3	8	4	66	1
Purple	6	3	4	7	8	64	2
Green	6	10	2	7	5	64	3

Ce que j'ai appris

- Utilisation d'entiers, lorsque possible, dans les opérations
- Utilisation de la fonction `round()` pour « effacer » les décimales excédentaires

Problème du 8 novembre 2017

Données manquantes et boucle



- Id du vendeur et vente dans différents trimestres

2014 1,554.34 4,571.54 1758.32 1511.70

2010 4211.54 7845.12

2016 1,413.25 1,735.24 4,256.12

2111 1,784.96 5,648.89

1132 4,578.10 7,656.00 1,211.32

3421 4578.25

- Obtenir un data avec id, vente et trimestre
 - Attention : trimestres manquants et formats différents pour les ventes

Ma solution (définitivement pas optimale)

```
data t;
  format id_num $4. q1-q4 $20.;
  infile DATALINES dlm=" " missover;
  input id_num $ q1 $ q2 $ q3 $ q4 $;

  q1=compress(q1, ",");
  q2=compress(q2, ",");
  q3=compress(q3, ",");
  q4=compress(q4, ",");

  datalines;
2014 1,554.34 4,571.54 1758.32 1511.70
2010 4211.54 7845.12
2016 1,413.25 1,735.24 4,256.12
2111 1,784.96 5,648.89
1132 4,578.10 7,656.00 1,211.32
3421 4578.25
;
run;
```

```
data t2;
  set t;
  format sales 7.2;

  if q1 ne "" then do;
    sales=put(q1, 7.2);
    quarter=1;
    output;
  end;

  ...

  if q4 ne "" then do;
    sales=put(q4, 7.2);
    quarter=4;
    output;
  end;

  drop q1 q2 q3 q4;
run;
```

Solution d'Horst Wolter

```
data multi;
  infile datalines missover ;
  input ID num @;
  do quarter=1 to 4;
    input Sales : comma8.2 @ ;
    if Sales ne . then output;
  end;
  datalines;
2014 1,554.34 4,571.54 1758.32 1511.70
2010 4211.54 7845.12
2016 1,413.25 1,735.24 4,256.12
2111 1,784.96 5,648.89
1132 4,578.10 7,656.00 1,211.32
3421 4578.25
;
run;
```

Bonification d'Aroop Ghosh

```
data multi;
  infile datalines missover ;
  input id_num $ sale_amt $ @;
  do Quarter=1 by 1 until (sale_amt= ' ');
    output;
    input sale_amt $ @;
  end;
cards;
2014 1,554.34 4,571.54 1758.32 1511.70
2010 4211.54 7845.12
2016 1,413.25 1,735.24 4,256.12
2111 1,784.96 5,648.89
1132 4,578.10 7,656.00 1,211.32
3421 4578.25
;
run;
```

Ce que j'ai appris

- Utilisation d'`input` avec
 - : -> d'un espace à l'autre
 - @ -> arrêt de la lecture
- Utilisation d'`infile` avec l'option `data lines`
- Faire confiance aux formats

Problème du 1^{er} mars 2017

Avalanche d'espaces



- Soit le jeu de données suivant :

SAN FRANCISCO		GIANTS		FORTYNINERS		
LOS ANGELES	RAMS	RAIDERS		LAKERS		DODGERS
BOSTON	CELTICS		BRUINS		RED SOX	
CHICAGO	CUBS	BULLS	WHITE SOX		BEARS	BLACKHAWKS
CHEYENNE						
NEW YORK	YANKEES			METS	RANGERS	KNICKS
TORONTO			MAPLE LEAFS		BLUE JAYS	

- Combien d'équipes par ville?
 - Attention : des noms contiennent un espace

Ma solution

```
data m1;
  infile "C:\SAS\FOS\raw20170301.txt" DLM='09'x;
  format rawData $255. city $16.;
  input rawData $;
  rawData = trim(rawData);
  rawData = tranwrd(rawData, " ", "%");
  nbWordT = countw(rawData, "%");
  nbTeams = 0;
  city = "";
  do i = 1 to nbWordT;
    if i = 1 then city = scan(rawData, 1, "%");
    else if scan(rawData, i, "%") ne " " then
nbTeams +1;
  end;
  keep city nbTeams;
run;
```

Solution d'Art Tabachneck

```
data have;  
  infile cards truncover;  
  informat city team1-team6 $16.;  
  input city & (team1-team6) (&);  
  array teams(*) team1-team6;  
  number_of_teams=dim(teams)-cmiss(of teams(*));  
  cards;  
SAN FRANCISCO  GIANTS  FORTYNINERS  
LOS ANGELES  RAMS  RAIDERS  LAKERS  DODGERS  
BOSTON  CELTICS  BRUINS  RED SOX  
CHICAGO  CUBS  BULLS  WHITE SOX  BEARS  BLACKHAWKS  
CHEYENNE  
NEW YORK  YANKEES  METS  RANGERS  KNICKS  
TORONTO  MAPLE LEAFS  BLUE JAYS  
;  
run;
```

Ce que j'ai appris

- Utilisation d'`input` avec `&` pour des valeurs contenant un espace et l'espace comme séparateur
- Utilisation d'`input` avec une liste de variables et de formats entre parenthèses