

Mise en contexte

Le programme « [Stop-and-Frisk](#) » mis en œuvre depuis 2002 dans la région de New York autorise la police (NYPD) à arrêter les individus dans la rue ou ailleurs et à procéder à des fouilles ou arrestations si nécessaire. L'objectif principal de cette politique, jugée inconstitutionnelle par les défenseurs des droits de l'homme, s'inscrit dans le cadre d'un programme de dissuasion du crime.

Vous êtes engagés à titre d'analyste dans un nouvel organisme de défense des droits de l'homme installé à New York et on vous confie le mandat de manipuler les données des arrestations totales effectuées par la police (NYPD) en **2021**.

Les tables de données qui accompagnent ce cas sont disponibles dans le classeur EXCEL « **Stop_Frisk_21.xlsx** » fournit sur Zone cours.

Le classeur Excel contient les 2 onglets suivants :

- **Data_2021**
- **Full_Adress_2021**

Les données de l'onglet (**Data_2021**) contiennent des observations correspondantes chacune à une **arrestation** dans les rues de la ville de New York. Plus de 80 variables sont répertoriées dans cette table. Ces différentes variables procurent de l'information sociodémographique en plus de fournir une description physique de l'individu, la raison de l'arrestation, la raison de la fouille, l'usage de la force, s'il y a lieu, le type de produit trouvé lors de la fouille ou l'arrestation, etc.

Les données de l'onglet (**Full_Adress_2021**) contiennent l'adresse exacte de chaque **arrestation**, y compris son arrondissement (**District**).

Remarque : Deux tables format **sas7bdat** sont fournies également pour contourner le problème d'importation des tables le cas échéant.

Mandat

Vous êtes appelés à effectuer les manipulations indiquées dans les questions suivantes :

- 1- **(2 points)** À partir de la table (**Data_2021**), créer une table temporaire, nommée **Temp** qui ne contient que les 10 variables suivantes :

*STOP_ID, STOP_FRISK_DATE, STOP_FRISK_TIME, STOP_DURATION_MINUTES,
FRISKED_FLAG, SEARCHED_FLAG, WEAPON_FOUND_FLAG, SUSPECT_REPORTED_AGE,
SUSPECT_SEX, SUSPECT_RACE_DESCRIPTION*

- a. Donner le nombre d'observations totales ainsi que le type de chacune des variables contenues dans la table **Temp**.
- 2- **(8 points)** À partir de la table **Temp**, créer une nouvelle table **Temp2** qui contient ce qui suit :
- a. Le **Mois** de l'arrestation, le **Jour** de l'arrestation et la **Date de l'arrestation** (format de date appliqué **yymmdd10**).
- b. L'**heure** de l'arrestation.
- c. Une variable catégorielle **Quart_jour** qui prend la valeur AM si l'arrestation est effectuée avant midi, PM sinon.
- d. Trois variables indicatrices **Frisked**, **Searched**, **FS** et **Arme où**.
- i. **Frisked** prend 1 si **FRISKED_FLAG** = 'Y', 0 sinon.
- ii. **Searched** prend 1 si **SEARCHED_FLAG** = 'Y', 0 sinon.
- iii. **FS** prend 1 si **FRISKED_FLAG** = 'Y' et **SEARCHED_FLAG** = 'Y', 0 sinon.
- iv. **Arme** prend 1 si **WEAPON_FOUND_FLAG** = 'Y', 0 sinon.
- e. Une variable catégorielle indiquant la catégorie de l'âge de l'individu arrêté. Assurez-vous que la variable **Age_cat** est alphanumérique.

Table : Définition de la catégorie d'âge

SUSPECT_REPORTED_AGE	< 15	15-24	25-34	35-44	45-54	55-64	>= 65
Age_cat	1	2	3	4	5	6	7

- f. La nouvelle table **Temp2**, ne doit contenir que les informations se rapportant aux individus **dont la race est clairement identifiée**.

- 3- (2 points) Calculer les statistiques suivantes :
- Le nombre de femmes arrêtées en 2021
 - Le nombre d'arrestations par race identifiée
 - Le nombre total d'arrestations par mois
 - La durée moyenne d'une arrestation (en minutes) par quart de jour

Veillez à assigner un titre à chacune des sorties.

- 4- (2 points) On veut obtenir la table en sortie suivante :

	SUSPECT_RACE_DESCRIPTION	Aucune_ame	Arme_Found
1	AMERICAN INDIAN/ALASKAN NATIVE	7	1
2	ASIAN / PACIFIC ISLANDER	180	12
3	BLACK	4506	898
4	BLACK HISPANIC	631	113
5	MIDDLE EASTERN/SOUTHWEST ASIAN	60	11
6	WHITE	666	66
7	WHITE HISPANIC	1482	231

Pour ce faire, adapter le code suivant à vos données :

- Calculer les fréquences croisées (**Arme* Race**) et créer une table en sortie **TableB** (à même la procédure proc freq).
- Trier la **TableB**
- Transposer la **TableB** pour obtenir la table désirée (**TableC**).

```
*****;
Proc freq data=TableA;
    table Race*Arme/norow nocol nopercent
        Out=TableB (options tables de données);
run;

proc sort data=TableB; by race; run;

Proc transpose data = TableB out =TableC (options de tables de données)
;
    by ...;
    id ...;
    var ...;
run;

*****;
```

- 5- (1 point) À partir de la table **Full_Adress_2021**, créer une nouvelle table **Temp3** qui comporte les deux variables suivantes : **STOP_ID** et **District**. District, étant, l'arrondissement de l'arrestation.

Veillez prendre note que vous êtes appelés à extraire le District de la variable Full_Adress.

- 6- (2 points) À partir de **Temp3**, créer une nouvelle table **Freq_district** tout en respectant l’affichage suivant:

	District	Nombre d'Arrestations	Pourcentage du Total
1	BRONX	1898	21.21%
2	BROOKLYN	2962	33.11%
3	MANHATTAN	2368	26.47%
4	QUEENS	1399	15.64%
5	STATEN ISLAND	320	3.58%

- 7- (1 point) Fusionner les deux tables **Temp2** et **Temp3** et nommer la nouvelle table **All_2021**.

Astuce: La table **All_2021** doit contenir tous les éléments en commun de **Temp2** et **Temp3**.

- 8- (2 points) Agréger les valeurs de la variable **FS** (Frisked and searched) par mois et par District pour pouvoir obtenir les deux tables suivantes :

i. Table **AG_FS**

	District	JANVIER	FEVRIER	MARS	AVRIL	MAI	JUIN	JUILLET	AOUT
1	BRONX	53	43	54	20	46	27	50	50
2	BROOKLYN	57	49	48	31	61	65	72	61
3	MANHATTAN	37	32	64	40	42	48	43	43
4	QUEENS	29	29	33	34	40	41	41	35
5	STATEN ISLAND	9	5	6	5	5	9	5	4

ii. Table **AG_FS2**

	mois	BRONX	BROOKLYN	MANHATTAN	QUEENS	STATEN_ISLAND
1	1	53	57	37	29	9
2	2	43	49	32	29	5
3	3	54	48	64	33	6
4	4	20	31	40	34	5
5	5	46	61	42	40	5
6	6	27	65	48	41	9
7	7	50	72	43	41	5
8	8	50	61	43	35	4
9	9	43	67	39	42	7
10	10	74	118	61	37	9
11	11	100	68	47	37	9
12	12	104	54	49	36	8