

Expérimentation pédagogique sur le thème

Big Data et Intelligence Artificielle

Des activités en collège et lycée ayant pour thème les notions de Big Data et d'IA

« Sondage »

Cycle 4 – 4e

Testée dans une classe de 4e au collège Pierre et Marie Curie.

1. Travail préparatoire

Les élèves ont déjà travaillé en classe les notions statistiques d'effectif, de moyenne, d'étendue et de regroupement par classe.

Ils ont également utilisé l'outil tableur pour se réapproprier les notions de cellule, de plage de données. Ils ont retravaillé sur la mise en place de formule et création de graphiques sur des exemples simples n'utilisant que quelques valeurs.

Ils ont également travaillé l'algorithmique avec la notion de condition SI ... ALORS ..., SI ... ALORS ... SINON ... Cela les aidera à aborder certaines fonctions du tableur pour faire des regroupements par classe.

2. Première phase : élaboration d'un sondage

Temps : 1 séance.

Présentation du projet

En classe entière, les élèves découvrent le projet proposé qui consiste à élaborer un sondage commun, puis à le proposer à 10 personnes (de tout âge) de notre entourage pour finalement regrouper toutes les valeurs ensemble afin de réaliser une étude plus proche de la réalité que les exercices d'entraînement que l'on peut trouver dans les livres.

Choix du sondage

Par îlot, ils doivent trouver des idées de questions pour un sondage.

Premier problème : trouver un fil directeur pour le sondage. Quelques idées commencent à émerger.

- La garde-robe (le nombre de pulls, de chaussures...) : les élèves conviennent rapidement que cela va être difficile à mettre en place et que les gens ne connaissent pas forcément l'inventaire chiffré de leur garde-robe.
- La musique : trop vaste !
- ...
- Le nombre de temps passé à faire quelque chose : c'est finalement ce qui semble faire l'unanimité.

Élaboration du sondage

Trouver 6 questions que l'on posera à nos sondés : pas si simple que cela pour toucher toutes les tranches d'âge. Il a fallu supprimer et retravailler plusieurs questions, adapter certaines questions sur un temps par jour ou par semaine pour aboutir finalement à 6 propositions qui plaisent à la grande majorité de la classe.

Je propose d'ajouter une question préalable au sondage, car on a constaté lors des questions que nous étions souvent confrontés à un problème : l'âge.

Un élève trouve cela judicieux, car on pourra comparer les résultats en fonction de l'âge. Super ! J'ai au moins un élève (mais j'espère sûrement d'autres, plus timides) qui a compris ce que nous allions faire.

Les élèves ont, le jour même, une fiche (voir ci-dessous) qu'ils devront remplir et utiliser pour compléter un fichier que je leur préparerai et mettrai à disposition sur leur espace numérique de travail (elyco).

Sondage réalisé par _____	classe de _____
---------------------------	-----------------

- Q1 Combien de temps dormez-vous par jour (nuit + sieste) ?
- Q2 Combien de temps passez-vous pour vous préparer (du lever du lit au début de votre journée) ?
- Q3 Combien de temps passez-vous dans les déplacements par semaine (trajets "travail", "courses") ?
- Q4 Combien de temps consacrez-vous au repas par jour (à manger uniquement) ?
- Q5 Combien de temps regardez-vous la télévision (ou films, séries sur écran) par semaine ?
- Q6 Combien de temps consacrez-vous à vos loisirs par semaines (hors question précédente) ?

	Âge	Q1	Q2	Q3	Q4	Q5	Q6
Personne 1							
Personne 2							
Personne 3							
Personne 4							
Personne 5							
Personne 6							
Personne 7							
Personne 8							
Personne 9							
Personne 10							

3. Seconde phase : le sondage en lui-même

Temps : 15 jours laissés au rythme des élèves

Les élèves doivent, avant de poser la question à une personne, s'assurer que celle-ci n'a pas déjà répondu au sondage.

Mes collègues étaient prévenus, car, comme je m'en doutais, certains ont attaqué la série de questions dès la séance d'élaboration. Pourquoi ai-je accepté leurs questions ? J'ai été forcément le premier sondé !

4. Troisième phase : récupérer les résultats

Temps : 15 minutes sur une de mes séances (après une semaine du début du sondage).

Afin de récupérer les valeurs de tous les élèves (sans devoir les taper), mais également pour commencer à les faire manipuler sur leurs données, je leur explique la marche à suivre qui sera à faire, soit chez soi, soit au cdi, permanence...

Ils devront, à partir d'un fichier mis à leur disposition (sous la forme d'un objet « **devoir** » d'elyco) :

- saisir les résultats en respectant le format **hh:mm** (j'ai formaté les cases pour cela),
- calculer la moyenne et l'étendue pour chaque question en utilisant une formule.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R			
1		Age	Q1	Q2	Q3	Q4	Q5	Q6													
2	Personne 1								Q1	Combien de temps dormez-vous par jour ?											
3	Personne 2								Q2	Combien de temps passez-vous pour vous préparer ?											
4	Personne 3								Q3	Combien de temps passez-vous dans les déplacements par semaine (trajets "travail", "courses") ?											
5	Personne 4								Q4	Combien de temps consacrez-vous au repas par jour (à manger uniquement) ?											
6	Personne 5								Q5	Combien de temps regardez-vous la télévision (ou films, séries sur écran) par semaine ?											
7	Personne 6								Q6	Combien de temps consacrez-vous à vos loisirs par semaines (hors question précédente) ?											
8	Personne 7																				
9	Personne 8																				
10	Personne 9																				
11	Personne 10																				
12	Personne 11																				
13	Personne 12																				
14	Personne 13																				
15	Personne 14																				
16	Personne 15																				
17																					
18																					

	Age	Q1	Q2	Q3	Q4	Q5	Q6
Moyenne							
Etendue							

Facultatif

Format à saisir pour les durées : 0:30 pour 30 minutes
2:00 pour 2 heures
2:30 pour 2 heures et 30 minutes

Attention si vous étiez une formule de l'age vers les questions, vous risquez de perdre le format des cellules !

Un retour sur cette phase

Certains élèves n'ont pas joué le jeu, soit en n'ayant posé aucune question (donc aucune valeur à proposer), soit en ayant fait tous les calculs sur une feuille sans utiliser le tableur, soit encore en saisissant des valeurs fantaisistes tant dans les données que dans les calculs de moyenne (réalisés sans l'aide des formules, mais à la main).

Les élèves n'auront donc que 180 données à traiter sur les 250 attendues.

5. Quatrième phase : traitement des données

Temps : 2 séances en salle multimédia.

Les élèves ont accès à 2 fichiers :

- un fichier tableur avec 2 onglets (un pour les données brutes, un autre pour la synthèse en partie commencée),
- un document leur donnant quelques outils pour les aider à gérer un plus grand nombre de données.

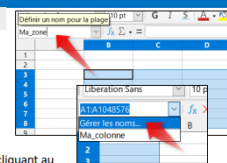
	A	B	C	D	E	F	G	H
1	Personne	Age	Q1	Q2	Q3	Q4	Q5	Q6
172	171	13	7:30	1:00	6:00	1:00	2:00	3:30
173	172	33	8:00	0:15	7:30	1:30	6:00	10:00
174	173	15	7:30	0:20	2:15	1:30	1:30	6:00
175	174	15	7:00	0:40	5:00	1:00	1:00	4:30
176	175	42	9:00	0:15	3:00	1:00	8:00	10:00
177	176	33	8:00	0:40	7:00	1:30	4:00	7:00
178	177	69	7:00	0:20	0:30	1:30	13:00	23:00
179	178	65	8:00	0:30	0:30	2:00	12:00	20:00
180	179	41	9:00	0:15	12:00	0:45	8:00	8:00
181	180	40	8:00	1:30	3:00	2:00	2:00	10:00

	A	B	C	D	E	F	G	H	I	J	
1	Q1	Combien de temps dormez-vous par jour ?									
2	Q2	Combien de temps passez-vous pour vous préparer ?									
3	Q3	Combien de temps passez-vous dans les déplacements par semaine (trajets "travail", "courses") ?									
4	Q4	Combien de temps consacrez-vous au repas par jour (à manger uniquement) ?									
5	Q5	Combien de temps regardez-vous la télévision (ou films, séries sur écran) par semaine ?									
6	Q6	Combien de temps consacrez-vous à vos loisirs par semaines (hors question précédente) ?									
7											
8				Effectif total							
9	Calculs des temps moyens par question										
10		Age	Q1	Q2	Q3	Q4	Q5	Q6			
11	Moyenne										
12	Etendue										
13											
14											
15	Calculs des temps moyens par question et par classe d'âge										
16		Age		Effectif	Q1	Q2	Q3	Q4	Q5	Q6	
17	min (exclus)	max (inclus)									
18	0	15									
19	15	30									
20	30	45									
21	45	60									
22	60	75									
23	75	90									
24	Total										
25											
26											
27											
28											

Quelques outils pour vous aider à compléter correctement le tableau

Donner un nom à toute une zone :

- Sélectionner la zone, puis lui attribuer un nom dans le cadre en haut à gauche (voir figure ci-contre).
- On peut également utiliser le menu (Données -> Définir la plage : LibreOffice ; Formule -> définir un nom : Excel)
- La zone peut être une colonne entière, une zone rectangulaire.
- Conseil : donner un nom à chacune des colonnes de votre tableau de données.
- On peut gérer les noms créés (modification, suppression...) en cliquant au même endroit (voir l'image ci-contre).



Des fonctions :

- NB ; NB.SI ; NB.SI.ENS (NB, NBSI, NBSI.ENS : sur les vieilles versions de LibreOffice)
Cette fonction permet de compter combien il y a de nombre (NB), combien il y a de nombre si une condition est remplie (NOMBRE.SI), combien il y a de nombre si un ensemble de condition est requis (NOMBRE.SI.ENS).
- SOMME ; SOMME.SI ; SOMME.SI.ENS (SOMME, SOMME SI, SOMMESIENS : sur les vieilles versions de LibreOffice)
Cette fonction permet de calculer des sommes...
- MOYENNE ; MOYENNE.SI ; MOYENNE.SI.ENS (MOYENNE, MOYENNESI, MOYENNESIENS : sur les vieilles versions de LibreOffice)
Cette fonction permet de calculer des moyennes...

Exemples d'utilisation :

- =NB(colonne) permet de compter tous les valeurs de la zone nommée colonne.
- =NB.SI(colonne ; >7) permet de compter dans la zone nommée colonne tous les nombres plus grand que 7.
- =NB.SI.ENS(colonne ; >7 ; colonne ; <=15) permet de compter dans la zone nommée colonne tous les nombres plus grand que 7 mais plus petit ou égal à 15.
- Si les valeur 7 et 15 sont le contenu des cellules A4 et A5 on peut écrire : =NB.SI.ENS(colonne ; >=A4 ; colonne ; <=A5)
- Pour les fonctions SOMME et MOYENNE, il y a un premier argument qui correspond à la plage de données pour laquelle on souhaite effectuer la somme ou la moyenne.

Utilisation du signe \$ pour bloquer la ligne ou la colonne d'une cellule.

Si dans une formule, on tape F5 alors si :

Il va de soi que lorsqu'on manipule un plus grand nombre de données, il devient nécessaire de commencer par nommer des plages de données pour faciliter l'exploitation par la suite. Sélectionner une zone à la souris est souvent source d'erreurs.

Un premier tableau leur suggère de refaire les mêmes calculs de moyenne et d'étendue, mais non plus sur leurs 10 valeurs, mais sur les 180 valeurs mises à leur disposition.

Dans un second temps, ils doivent réaliser la première tâche, plus complexe, de regrouper les données par classe en utilisant les fonctions NB.SI.ENS, et MOYENNE.SI.ENS qui permet de réaliser des calculs pour un ensemble de conditions.

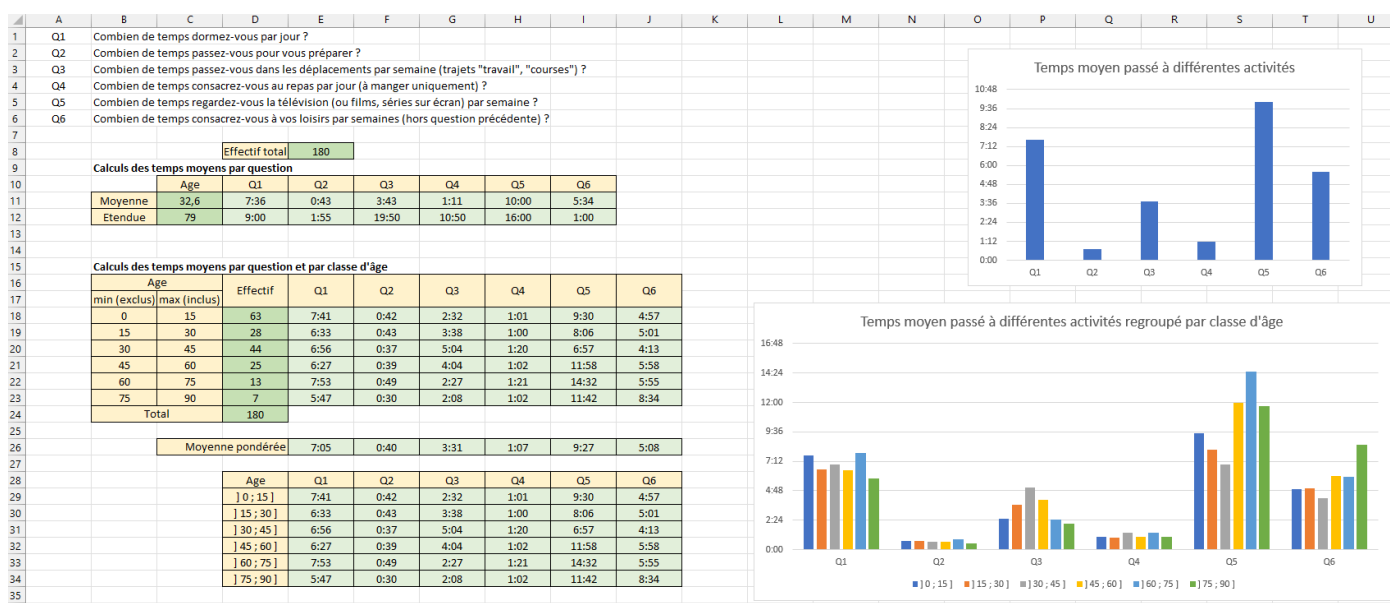
Certains élèves parviennent à finir cette tâche en 1 séance.

Dans un second temps, ils doivent travailler sur la réalisation de graphiques illustrant les résultats des 2 tableaux. Des essais ont été faits sur des diagrammes circulaires, mais rapidement, ceux qui ont abordé cette question reviennent sur des graphiques en barres.

Pour les plus rapides, il leur est proposé de mettre le dernier tableau en forme pour qu'il ressemble à ce que nous avons l'habitude de voir en classe pour des regroupements par classe.

Ils ont pu tester l'utilisation du \$ dans les formules ainsi que les concaténations.

Ci-dessous, un exemple de réalisation complète par un élève assez rapide (même s'il a fallu le motiver à mettre des titres à ses graphiques et à soigner un peu la présentation de ses résultats).



Un retour de ces 2 séances

- Pour une petite minorité de la classe, ils n'ont entamé ni le tableau final, ni les graphiques.
- Beaucoup d'élèves ont passé énormément de temps à taper chaque formule sans penser à l'étirer et se sont arrêtés aux graphiques
- L'utilisation des \$ n'a été testée que pour 3 à 4 élèves.

6. Le retour en classe

Temps : une demi-séance.

Les élèves devaient revenir avec leurs 10 résultats afin que l'on puisse les comparer avec ceux obtenus sur les 180 valeurs.

Le retour n'a pas été aussi complet que je l'espérais, les élèves ont fait des tâches sans une réelle analyse. Néanmoins certains élèves ont fait remarquer que leurs résultats avaient pu être faussés :

- ils avaient des réponses que pour une certaine catégorie de personnes,
- ils avaient dans leurs données des valeurs loin de la moyenne, importantes pour eux, mais noyées dans les 180 valeurs.