

Expérimentation pédagogique sur le thème

Vers le Big Data

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

« Donner au train des Idées d'Avance »

Utiliser le tableur pour exploiter un grand nombre de données de la SNCF.

Cycle 4 et Lycée

Testée dans deux classes de 3e au collège Emmanuel De Martonne et une classe de 2nde au lycée Aimé Césaire

1. Présentation de l'activité

Il s'agit d'étudier à l'aide d'un tableur les différents objets trouvés/restitués dans deux gares en partant d'un fichier de données téléchargé sur le site de la SNCF : <https://data.sncf.com/pages/accueil/>

Cette étude est proposée pour des élèves de 3^{ème} et des élèves de 2nd, avec deux approches adaptées.

2. Contexte

Version collège	Version lycée
Les élèves n'ont pas ou très peu utilisé le tableur depuis la 6e. Il s'agit donc, dans un premier temps, de leur proposer de petits exercices pour s'appropriier les différentes spécificités et quelques fonctionnalités de tri du tableur. Ensuite, ils pourront réinvestir ces apprentissages pour traiter des fichiers de données plus volumineux. Les séances auront lieu en salle multimédia en groupe ce qui permet d'avoir 1 ordinateur par élève.	L'activité intervient après le chapitre « Information chiffrée » où les notions de fréquences, proportions, pourcentages, coefficient de proportionnalité, taux d'évolution, coefficient multiplicateur ont été abordées. Les élèves n'ont pas beaucoup utilisé le tableur donc au moins 3 séances seront nécessaires pour cette activité. Les séances auront lieu en salle classique en groupe (2 ou 3 élèves par groupe) chaque élève apporte son ordinateur (suite à la distribution des ordinateurs par la Région).

3. Déroulé

Version collège	Version Lycée
<p>Etape 1 :</p> <p>1^{er} temps : je leur présente le site de la SNCF et le type de fichier que l'on peut y trouver puis leur montre le fichier de la tâche finale. Il n'y reste que les gares de Laval et Quimper et les informations ont déjà été allégées. La 1^{ère} consigne est "Que peut-on faire de toutes ces données ?"</p> <p>Assez rapidement, l'idée de savoir combien d'objets de chaque type ont été trouvés et restitués arrive, l'idée de comparer les résultats des 2 gares aussi. Viennent alors les remarques/questions du type "On ne va pas tout faire à la main, il y en a beaucoup trop ", "Comment on fait pour tout compter ?" "Il faudrait tout classer".</p> <p>Avant de se lancer dans l'exploitation il va donc falloir (ré)apprendre à utiliser le tableur en particulier saisir une formule, trier des données, compter des cellules.</p> <p>2^{ème} temps (re)découvrir le tableur. Il s'agit d'explorer une fonctionnalité du tableur par feuille. Les élèves sont à leur disposition une feuille papier avec les consignes et un espace " à retenir" pour y noter les fonctions et leur utilisation ainsi que le fichier tableur "pour prendre un bon départ".</p> <p>Feuille 1 "calculatrice VS tableur" → saisir une formule</p> <p>Feuille 2 "ticket de caisse" → étirer une formule</p> <p>Feuille 3 "stock" → tester les différents tris, extraire sans doublon les valeurs d'une liste</p> <p>Feuille 4 "bâtiment" → dénombrer avec les fonctions NBVAL et NBSI</p> <p>Les 1^{ères} difficultés apparaissent pour les calculs de prix du ticket de caisse : ce n'est pas le tableur qui bloque, mais la compréhension même du calcul d'un prix et d'un pourcentage. Pour le dernier groupe, cette feuille a été changée pour éviter les difficultés extérieures au tableur.</p> <p>J'avais prévu 1h pour cette partie il en a fallu deux.</p> <p>Etape 2 : étudier les objets trouvés dans les gares</p> <p>La gare de Laval :</p> <p>Il s'agit de dénombrer chaque type d'objet. Les données sont assez nombreuses : sans tableur c'est très fastidieux.</p> <p>La manipulation du tableur étant assez fastidieuse sur les 1^{ères} séances, j'ai pris le parti de ne pas en rajouter en nommant les plages de cellules ou en utilisant les filtres dans les premiers groupes, mais j'ai par la suite modifié la feuille pour permettre l'utilisation des filtres.</p> <p>Difficulté : il faut compter les valeurs dans la colonne A (ou B) en fonction de la valeur dans la colonne C ; le choix a été fait de trier la colonne C et ensuite de</p>	<p>Etape 1 :</p> <p>La première séance les élèves découvrent le document « brut » et la manière de le télécharger à partir du site de la SNCF. Dans un second temps ils accèdent au fichier sur lequel ils vont travailler disponible sur leur espace de travail elyco. Un document papier expliquant les consignes leur est distribué en plus d'un document rappelant quelques fonctionnalités du tableur dont ils pourraient avoir besoin.</p> <p>Au bout de cette séance, tous les groupes avançaient sur la gare de Laval. Le groupe le plus avancé était sur la question concernant le nombre d'objets trouvés par année,</p> <p>Observation/difficulté: certains élèves ont utilisé la fonction « tri » sans étendre la sélection,</p> <p>Etape 2 :</p> <p>La deuxième séance, les élèves ont continué à avancer sur la gare de Laval. Un seul groupe a entamé la gare de Quimper à la fin de cette séance.</p> <p>Observation : la fonction NBSI a posé problème sur certains ordinateurs (fenêtre qui se fermait) sans comprendre pourquoi. Problème résolu en redémarrant les ordinateurs</p> <p>Etape 3 :</p> <p>Lors de cette séance les élèves ont pu pour la grande majorité arriver au bout de la gare de Laval et entamer la gare de Quimper, certains (4 groupes) ont fini Quimper.</p>

compter dans la A (ou la B) avec NBVAL .
Problème rencontré et non résolu : la fonction "tri filtre spécial sans doublon" redonne systématiquement la 1^{ère} ligne.

La gare de Quimper :

Les données sont encore beaucoup plus nombreuses, le fichier a été retravaillé pour ne plus travailler sur des données littérales, mais sur des données chiffrées :

- 1) La restitution a été codée par 1 ou 0
- 2) Le type d'objet est codé par un nombre

On voit avec les élèves l'intérêt d'un tel codage : si on additionne les deux codes 10 donne un objet bagagerie non restitué et 11 un objet bagagerie restitué. On dénombre alors des valeurs chiffrées sans avoir besoin de tenir compte des autres colonnes, le NBSI suffit.

Dans les premiers groupes rares ont été les élèves à arriver jusqu'ici c'est pourquoi, dans les groupes suivants, une fois la non-praticité éprouvée de la feuille Laval, on est passé à la feuille Quimper.

Les élèves terminent beaucoup plus facilement le travail pour la gare de Quimper.

Aucun groupe n'a eu le temps d'analyser les résultats et de comparer les deux gares.

4. Prolongements possibles

Par la suite, il pourrait être intéressant de faire modifier la feuille Laval par les élèves pour faciliter le comptage comme pour la feuille Quimper.

Dans le cadre d'une comparaison entre les 2 gares, on peut aussi changer les critères et trier par saisons pour voir les différences et éventuellement infirmer ou confirmer l'hypothèse d'une différence due au tourisme par exemple.

On peut aussi bien entendu faire tracer des graphiques (pour le collège) et déterminer ceux qui sont les plus pertinents.