

impact de la  
pandémie COVID-19 sur l’Apprentissage aux USA

4BIGF – Big Data Fundamentals

BENYOUCEF Lynda

2022 HULLOUX Guillaume

Campus de Lille LEPLAE Louan

Table des matières

[Introduction 2](#_Toc97582903)

[Objectifs 2](#_Toc97582904)

[Données disponibles 3](#_Toc97582905)

[Engagement étudiant 3](#_Toc97582906)

[Technologies éducatives 3](#_Toc97582907)

[Districts scolaires 4](#_Toc97582908)

[Préparation des données 5](#_Toc97582909)

[Connectivité et engagement numérique 5](#_Toc97582910)

[Informations générales 5](#_Toc97582911)

[Etat de la connectivité au niveau des comtés 6](#_Toc97582912)

[COVID-19 et apprentissage à distance 6](#_Toc97582913)

[Engagement en fonction du temps en 2020 6](#_Toc97582914)

[Engagement étudiant et technologies éducatives 7](#_Toc97582915)

[Catégories et secteurs des produits d’apprentissage 7](#_Toc97582916)

[Produits et fournisseurs prédominants 7](#_Toc97582917)

[Engagement étudiant et facteurs sociogéographiques 8](#_Toc97582918)

[Répartition géographique des engagements étudiants 8](#_Toc97582919)

[Engagements des communautés noire et hispanique 8](#_Toc97582920)

[Engagement étudiant et politiques d’Etat 8](#_Toc97582921)

[Dépenses totales locales et fédérales par élève 8](#_Toc97582922)

[Engagement des élèves de foyers modestes 8](#_Toc97582923)

[Conclusion 8](#_Toc97582924)

# Introduction

La COVID-19, une pandémie apparue en novembre 2019 à Wuhan en Chine, a entraîné, par sa virulence, une multitude de défis sans précédent pour l'apprentissage et l'éducation des élèves à travers le monde. La flambée mondiale de cas de COVID-19 a contribué à la fermeture de plusieurs écoles, collèges et universités en 2020 dans presque toutes les régions du monde et le passage à l'apprentissage en ligne, ce qui a eu un impact sur l'apprentissage des élèves de différentes manières.

Cela a conduit les éducateurs et les étudiants à passer plus que jamais du temps sur Internet, et par conséquent s'adapter à l'apprentissage à distance en apprenant et se familiarisant avec les informations, les outils, les applications mis à leur disposition.

L'étude de ces comportements web, sous la forme d'exploration et d'analyse de Big Data, provenant de différents pays du monde offre la possibilité d'identifier, d'enquêter et de quantifier les besoins, les intérêts et les défis liés à l'apprentissage en ligne dans différents pays du monde imposés par la COVID-19.

A travers ce projet, nous utiliserons des données concernant les étudiants américains ainsi que des outils d'analyse de données pour déterminer les tendances de l'apprentissage numérique et son efficacité envers les communautés étudiées. Nous comparerons les districts et les États sur des facteurs tels que la démographie, l'accès à Internet, l'accès aux produits d'apprentissage et les finances. Enfin, nous résumerons notre rapport et indiquerons les domaines qui nécessitent notre plus grande attention pour rendre l'éducation accessible à tous les étudiants des États-Unis.

# Objectifs

L’objectif de ce projet est d’explorer l'état de l'apprentissage numérique en 2020 et comment l'engagement de l'apprentissage numérique est lié à des facteurs tels que la démographie du district ; mais aussi, d’analyser les ensembles de données, trouver des informations significatives et créer des visualisations simples et pertinentes.

Nous étudierons l’engament des étudiants à l’apprentissage à distance selon cinq axes :

* L’état de la connectivité et engagement numérique en 2020
* L’effet de la COVID-19 sur l’apprentissage à distance
* La relation entre l’engagement étudiant et les technologies éducatives
* Le potentiel lien entre l’engagement étudiant et les facteurs sociogéographiques
* Le rôle des politiques d’Etat dans l’engagement étudiant

# Données disponibles

Le jeu de données fourni est un ensemble de données sur l'engagement « edtech » de plus de 200 districts scolaires des Etats-Unis en 2020. Celui-ci inclut trois ensembles de fichiers de base.

## Engagement étudiant

Il s’agit d’un dossier comportant un ensemble de fichiers .csv basé sur l'extension Student Chrome de la plateforme LearnPlatform. L'extension collecte les événements de chargement de page de plus de 10 000 produits de technologie éducative dans la bibliothèque de produits, y compris les sites Web, les applications, les applications Web, les programmes logiciels, les extensions, les livres électroniques, les matériels et les services utilisés dans les établissements d'enseignement. Les données sur l'engagement ont été agrégées au niveau du district scolaire, et chaque fichier représente les données d'un district scolaire.

Les données d'engagement sont agrégées au niveau du district scolaire, et chaque fichier du dossier **engagement\_data** représente les données d'un district scolaire. Le nom de fichier à 4 chiffres représente **district\_id** qui peut être utilisé pour établir un lien vers les informations du district, **lp\_id** peut être utilisé pour créer un lien vers des informations sur le produit de technologie éducative.

|  |  |
| --- | --- |
| ***Nom*** | ***Description*** |
| temps | Date en "AAAA-MM-JJ" |
| lp\_id | Identifiant unique du produit |
| pct\_access | Pourcentage d'élèves du district ayant au moins un événement de chargement de page d'un produit donné et un jour donné |
| engagement\_index | Nombre total d'événements de chargement de page pour mille étudiants d'un produit donné et un jour donné |

## Technologies éducatives

Ces données se présente sous la forme d’un fichier .csv comprenant des informations sur les caractéristiques des 372 meilleurs produits avec le plus d'utilisateurs en 2020. Les catégories répertoriées dans le fichier **products\_info.csv** font partie de la taxonomie des produits de LearnPlatform. Certains produits peuvent ne pas avoir d'étiquettes en raison d'un doublon, d'un manque d'URL précise ou d'autres raisons.

|  |  |
| --- | --- |
| ***Nom*** | ***Description*** |
| LP ID | Identifiant unique du produit |
| URL | Lien Web vers le produit spécifique |
| Product Name | Nom du produit spécifique |
| Provider/Company Name | Nom du fournisseur du produit |
| Sector(s) | Secteur de l'éducation où le produit est utilisé |
| ***Nom*** | ***Description*** |
| Primary Essential Function | Fonction de base du produit.  Il y a deux couches d'étiquettes ici. Les produits sont d'abord étiquetés dans l'une de ces trois catégories : LC = Apprentissage et programme d'études, CM = Gestion de classe et SDO = Opérations de l'école et du district. Chacune de ces catégories comporte plusieurs sous-catégories avec lesquelles les produits ont été étiquetés |

## Districts scolaires

Il s’agit d’un fichier .csv comprenant des informations sur les caractéristiques des districts scolaires, y compris des données du NCES (pour la connexion IP) et de la FCC (pour les impayés).

Dans cet ensemble de données, **district districts\_info.csv**, les informations identifiables sur les districts scolaires ont été anonymisés. À des fins de généralisation des données, certains points de données sont publiés avec une plage dans laquelle se situe la valeur réelle. De plus, il existe de nombreuses données manquantes marquées comme "NaN", indiquant que les données ont été supprimées pour maximiser l'anonymisation de l'ensemble de données.

|  |  |
| --- | --- |
| ***Nom*** | ***Description*** |
| distict\_id | Identifiant unique du district scolaire |
| state | Etat où le district se trouve |
| local | Classification locale NCES qui classe le territoire américain en quatre types de zones : ville, banlieue, ville et campagne |
| pct\_black/hispanic | Pourcentage d'élèves dans les districts identifiés comme noirs ou hispaniques sur la base des données du NCES 2018-19 |
| pct\_free/reduced | Pourcentage d'élèves dans les districts éligibles à un déjeuner gratuit ou à prix réduit sur la base des données du NCES 2018-19 |
| county\_connections\_ratio | Ratio (connexions haut débit fixes résidentielles supérieures à 200 kbps dans au moins une direction/foyers) sur la base des données au niveau du comté de FCC From 477 (version de décembre 2018) |
| pp\_total\_raw | Dépenses totales par élève (somme des dépenses locales et fédérales) d'après le projet National Education Resource Database on Schools (NERD$) d'Edunomics Lab |

# Préparation des données

Dans le but de faciliter les analyses, nous commençons par regrouper les trois sources de données, concernant l’engagement étudiant, les technologies éducatives, et les districts scolaires, en une unique source.

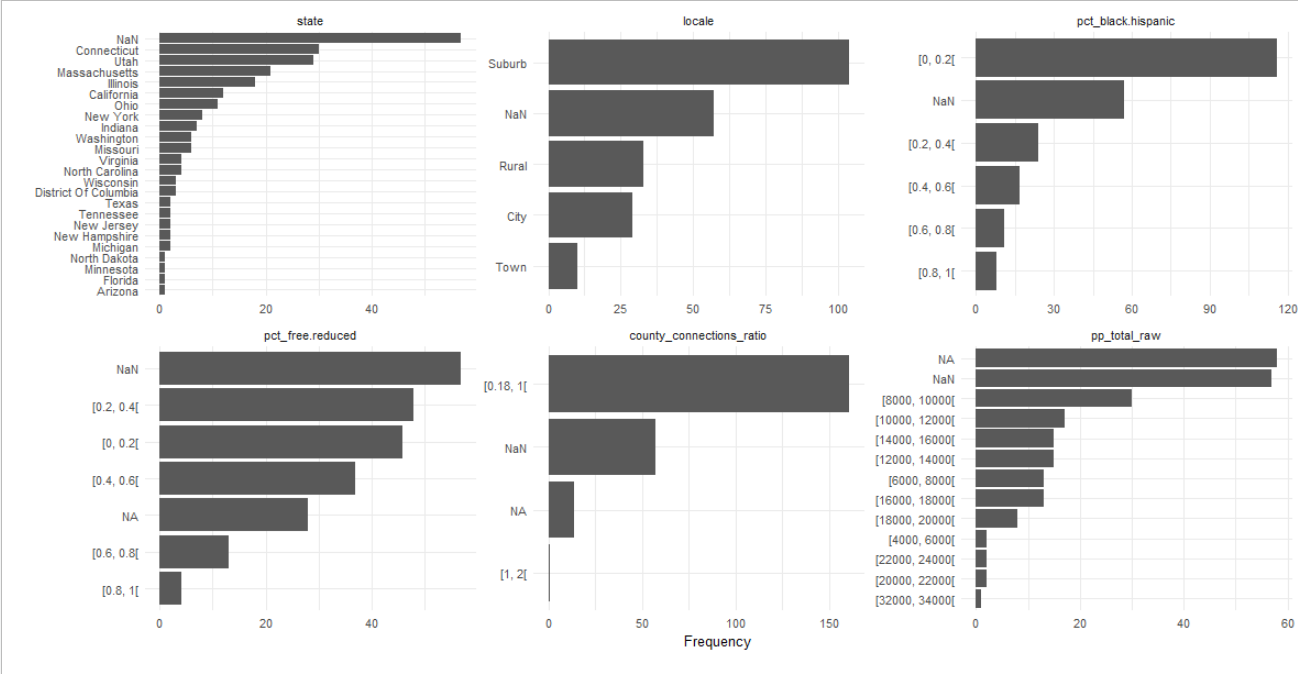
Pour constituer la base commune, on utilise les noms à 4 chiffres des fichiers présents dans le dossier **egagement\_data**, qui représente la valeur **district\_id** pour établir un lien vers les informations du district du fichier **district\_info.csv**; et on utilise le champ **lp\_id** de ces fichiers pour créer un lien vers des informations sur le produit de technologie éducative du fichier **product\_info.csv**.

Une fois la source de données définie, nous allons mettre en lumière les proportions de valeurs inconnues ou non définies pour avoir un aperçu sur la précision de nos analyses.

# Connectivité et engagement numérique

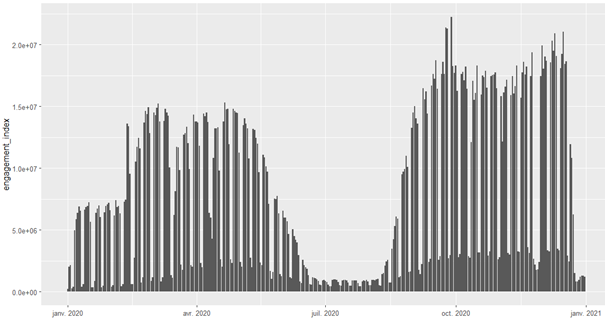
## Informations générales

## Etat de la connectivité au niveau des comtés



# COVID-19 et apprentissage à distance

## Engagement en fonction du temps en 2020



Ce graphique représente l’engagement des étudiant, par le nombre total d'événements de chargement de page d'une solution éducative, sur la période du 1er janvier au 31 décembre 2020.

On remarque que :

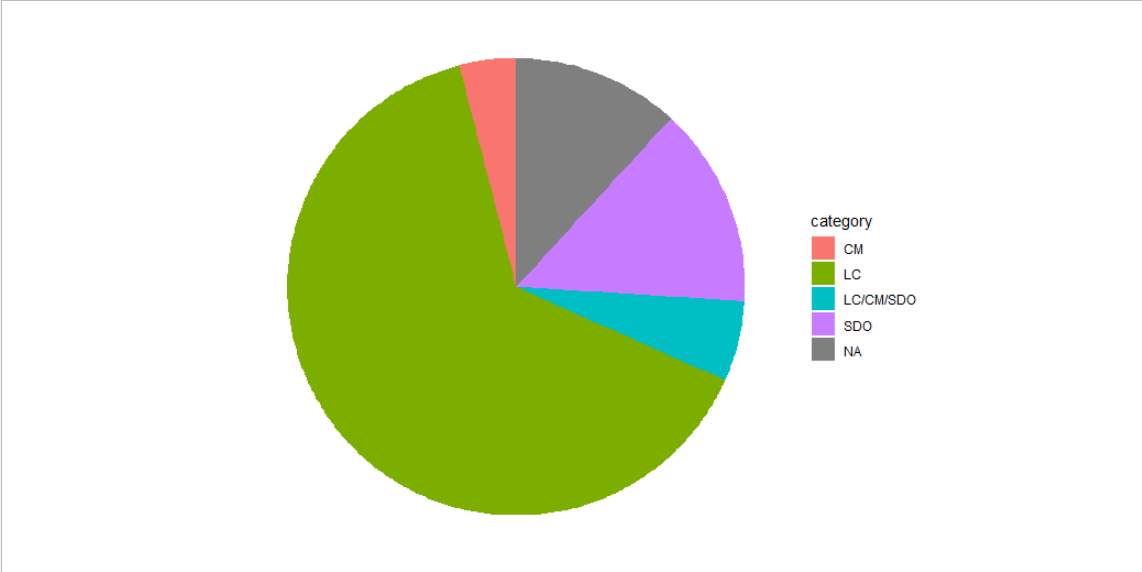
* L’épidémie a provoqué une fermeture temporaire des écoles à partir de mi-février avant l’annonce par l’OMS officielle du 11 mars 2020 qui définit le COVID-19 comme pandémie. Et c’est précisément à partir de cette période que le nombre de pages consultées augmente, ce qui signifie que l'apprentissage numérique est utilisé plus fréquemment qu'auparavant.
* Le très faible nombre de pages consultées chaque week-end transparaît distinctement sur le graphique, ce qui dépeint la plage d’apprentissage privilégiée des étudiants est du lundi au vendredi.
* Tout comme les activités les week-ends, le nombre de pages consultées est également très faible pendant les vacances d'été (fin juin à début septembre), ce qui signifie également que les étudiants utilisent peu les produits en dehors des plages réservées à l’apprentissage.

\*\*\*\*R\*\*\*\*

# Engagement étudiant et technologies éducatives

## Catégories et secteurs des produits d’apprentissage

## Produits et fournisseurs prédominants



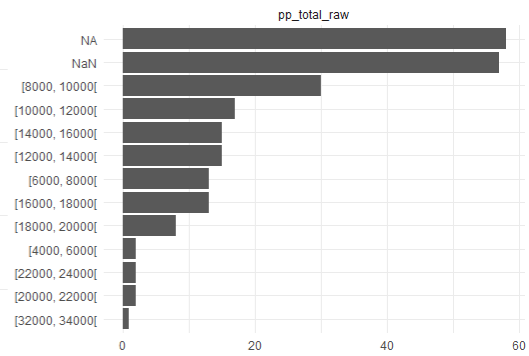
# Engagement étudiant et facteurs sociogéographiques

## Répartition géographique des engagements étudiants

## Engagements des communautés noire et hispanique

# Engagement étudiant et politiques d’Etat

## Dépenses totales locales et fédérales par élève

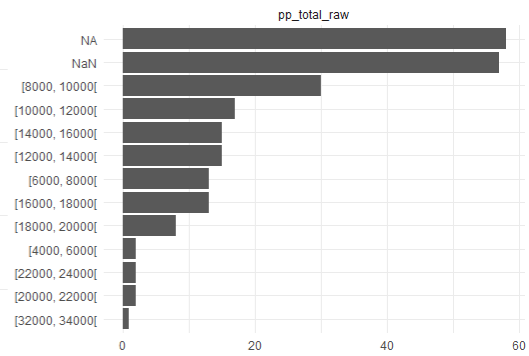


Ce graphique montre le nombre de pages chargées en fonction du montant des budgets alloués aux écoles provenant du projet National Education Resource Database on Schools (NERD$) du Edunomics Lab.

On observe que :

* Près de la moitié des informations sur les dépenses totales par élève sont marquées comme inconnues.
* À partir des données restantes, nous constatons que la plupart des pages chargées proviennent de districts scolaires dont les dépenses sont comprises entre 8 000 et 10 000.
* Nous pouvons remarquer également que les districts scolaires dont les dépenses sont comprises entre 10 000 et 18 000 ont plus de 200 millions de pages chargées en 2020.
* Les deux résultats les plus faibles sont celles des districts dont les dépenses sont comprises entre 4 000 et 6 000 et entre 32 000 et 34 000, soient les dépenses les plus faibles et les plus élevées de l'ensemble de données. Cela signifie que le montant des dépenses n’est pas forcément lié à l’engagement des étudiants

\*\*\*\*R\*\*\*\*



Ce graphique montre le nombre de districts scolaires par rapport au montant des budgets alloués aux écoles provenant du projet National Education Resource Database on Schools (NERD$) du Edunomics Lab.

On observe que :

* Ce graphique avec le nombre de districts scolaires dans une catégorie de budget semble être en accord avec le graphe concernant le nombre de pages chargées pour un même budget ; et préciser les observations qui en découlent.
* Une fois encore, la catégorie ayant le plus grand nombre de districts scolaires est celle regroupant les valeurs non renseignées.
* Il y a 30 districts scolaires dont les dépenses totales sont comprises entre 8 000 et 10 000, ce qui explique également le nombre de pages chargés plus élevé dans le graphique précédent.

\*\*\*\*R\*\*\*\*

## Engagement des élèves de foyers modestes

# Conclusion