Introduction

Current research shows educational outcomes are far from equitable. The imbalance was exacerbated by the COVID-19 pandemic. There's an urgent need to better understand and measure the scope and impact of the pandemic on these inequities.

In this challenge, you'll work to uncover trends in digital learning. Accomplish this with data analysis about how engagement with digital learning relates to factors like district demographics, broadband access, and state/national level policies and events.

Problem Statement

The COVID-19 Pandemic has disrupted learning for more than 56 million students in the United States. In the Spring of 2020, most states and local governments across the U.S. closed educational institutions to stop the spread of the virus. In response, schools and teachers have attempted to reach students remotely through distance learning tools and digital platforms. Until today, concerns of the exacerbating digital divide and long-term learning loss among America's most vulnerable learners continue to grow.

Challenge

Your analysis should be motivated by the answering of these two questions :

- **What is the state of digital learning in 2020?**
- **And how does the engagement of digital learning relate to factors such as district demographics, broadband access, and state/national level policies and events?**

We encourage you to guide the analysis with questions that are related to the themes that are described above (in bold font). Below are some examples of questions that relate to the

problem statement:

- What is the picture of digital connectivity and engagement in 2020?
- What is the effect of the COVID-19 pandemic on online and distance learning, and how might this also evolve in the future?
- How does student engagement with different types of education technology change over the course of the pandemic?
- How does student engagement with online learning platforms relate to different geography? Demographic context (e.g., race/ethnicity, ESL, learning disability)? Learning context? Socioeconomic status?
- Do certain state interventions, practices or policies (e.g., stimulus, reopening, eviction moratorium) correlate with the increase or decrease online engagement?

Dataset Description

You are provided a set of daily edtech engagement data from over 200 school districts in 2020. They include three basic sets of files :

- The `engagement_data` folder is based on LearnPlatform's Student Chrome Extension. The extension collects page load events of over 10K education technology products in our product library, including websites, apps, web apps, software programs, extensions, ebooks, hardwares, and services used in educational institutions. The engagement data have been aggregated at school district level, and each file represents data from one school district.
- The `products_info.csv` file includes information about the characteristics of the top 372 products with most users in 2020.
- The `districts_info.csv` file includes information about the characteristics of school districts, including data from NCES and FCC.

File Structure

The organization of data sets is described below:

```
Root/
  -engagement_data/
```

```
    -1000.csv
    -1039.csv
    -...
  -districts_info.csv
  -products_info.csv
```

Data Definition

Engagement data

The engagement data are aggregated at school district level, and each file in the folder `engagement_data` represents data from one school district. The 4-digit file name represents `district_id` which can be used to link to district information in `district_info.csv`. The `lp_id` can be used to link to product information in `product_info.csv`.

Name	Description
time	date in "YYYY-MM-DD"
lp_id	The unique identifier of the product
pct_access	Percentage of students in the district have at least one page-load event of a given product and on a given day
engagement_index	Total page-load events per one thousand students of a given product and on a given day

District information data

The district file `districts_info.csv` includes information about the characteristics of school districts, including data from NCES (2018-19), FCC (Dec 2018), and Edunomics Lab. In this data set, the identifiable information about the school districts was removed. An open source tool ARX (Prasser et al. 2020) was used to transform several data fields and reduce the risks of re-identification. For data generalization purposes some data points are released with a range where the actual value falls under. Additionally, there are many missing data marked as 'NaN' indicating that the data was suppressed to maximize anonymization of the dataset.

Name	Description
district_id	The unique identifier of the school district
state	The state where the district resides in
locale	NCES locale classification that categorizes U.S. territory into four types of areas: City, Suburban, Town, and Rural. See Locale Boundaries User's Manual for more information.
pct_black/hispanic	Percentage of students in the districts identified as Black or Hispanic based on 2018-19 NCES data
pct_free/reduced	Percentage of students in the districts eligible for free or reduced-price lunch based on 2018-19 NCES data
county_connections_ratio	`ratio` (residential fixed high-speed connections over 200 kbps in at least one direction/households) based on the county level data from FCC From 477 (December 2018 version). See FCC data for more information.
pp_total_raw	Per-pupil total expenditure (sum of local and federal expenditure) from Edunomics Lab's National Education Resource Database on Schools (NERD$) project. The expenditure data are school-by-school, and we use the median value to represent the expenditure of a given school district.

Product information data

The product file `products_info.csv` includes information about the characteristics of the top 372 products with most users in 2020. The categories listed in this file are part of LearnPlatform's product taxonomy. Data were labeled by our team. Some products may not have labels due to being duplicate, lack of accurate url or other reasons.

Name	Description
LP ID	The unique identifier of the product
URL	Web Link to the specific product

Name	Description
Product Name	Name of the specific product
Provider/Company Name	Name of the product provider
Sector(s)	Sector of education where the product is used
Primary Essential Function	The basic function of the product. There are two layers of labels here. Products are first labeled as one of these three categories: LC = Learning & Curriculum, CM = Classroom Management, and SDO = School & District Operations. Each of these categories have multiple sub-categories with which the products were labeled