EXECUTIVE SUMMARY

EdTech Synergy Innovators (Team 53)

Kay Quiballo, Rohan Tandon, Lauren Vaught, Ge Li, Brandon Eubank

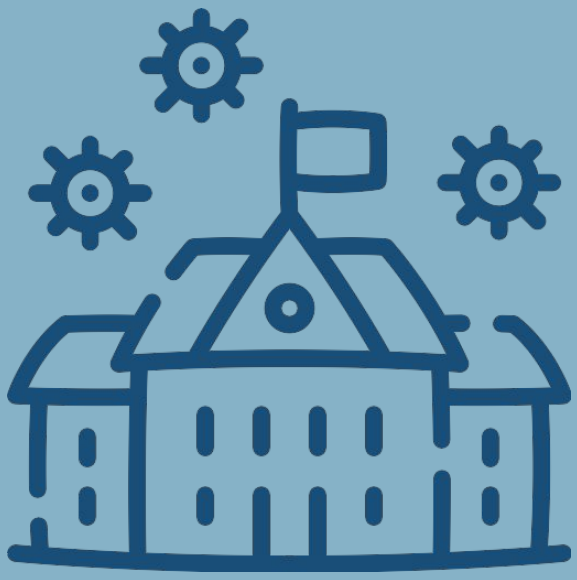
CONTENTS

01 Problem Statement

Background

In 2020, the COVID-19 pandemic affected the education ecosystem, impacting students, teachers, and school districts. In order to continue melding young minds, school districts deployed online learning tools to facilitate classes during quarantine.

In light of this historic moment, we as a data science team have information at our fingertips to discover how school districts handled the pandemic and if their efforts to deploy online tools were truly successful



Objectives

1. **To explore the data available to us. This includes but is not limited to: identifying frequently used learning platforms, identifying demographics that may have had less access to online learning, and identifying geographic regions that may have had less access to online learning**

2. **To quantify differences in tech engagement across different populations. This will be accomplished using various statistical models.**

3. **To provide educators with resources to inform them of tech engagement level in their district and top ed-tech products. Educators can leverage the dashboard and chat bot that our team builds to advocate for more resources.**

For future directions, school districts may use this report and the deliverables we create to advocate for better funding in EdTech online solutions. Ensuring that every student has the tools they need to succeed is our objective.

01 Project Goals

Goal 1
Exploratory Data Analysis

Through EDA, our team hopes to implement visualizations to better understand the data and interactions between variables.

Goal 2
Model Building

Through Modeling, our team hopes to quantify the relative digital engagement across different demographics.

Goal 3
Deliverables

Through Deliverables, our team hopes to advise educators with a dashboard and chatbot.

To address our problem statement and complete our project goals, we must first take a deeper look at the data we wish to utilize.

*Cody Bakley, Julia Elliott, Mary Styers, Scott McQuiggan, Zarifa Zakaria. (2021). LearnPlatform COVID-19 Impact on Digital Learning. Kaggle. https://kaggle.com/competitions/learnplatform-covid19-impact-on-digital-learning

02 Overview of Data

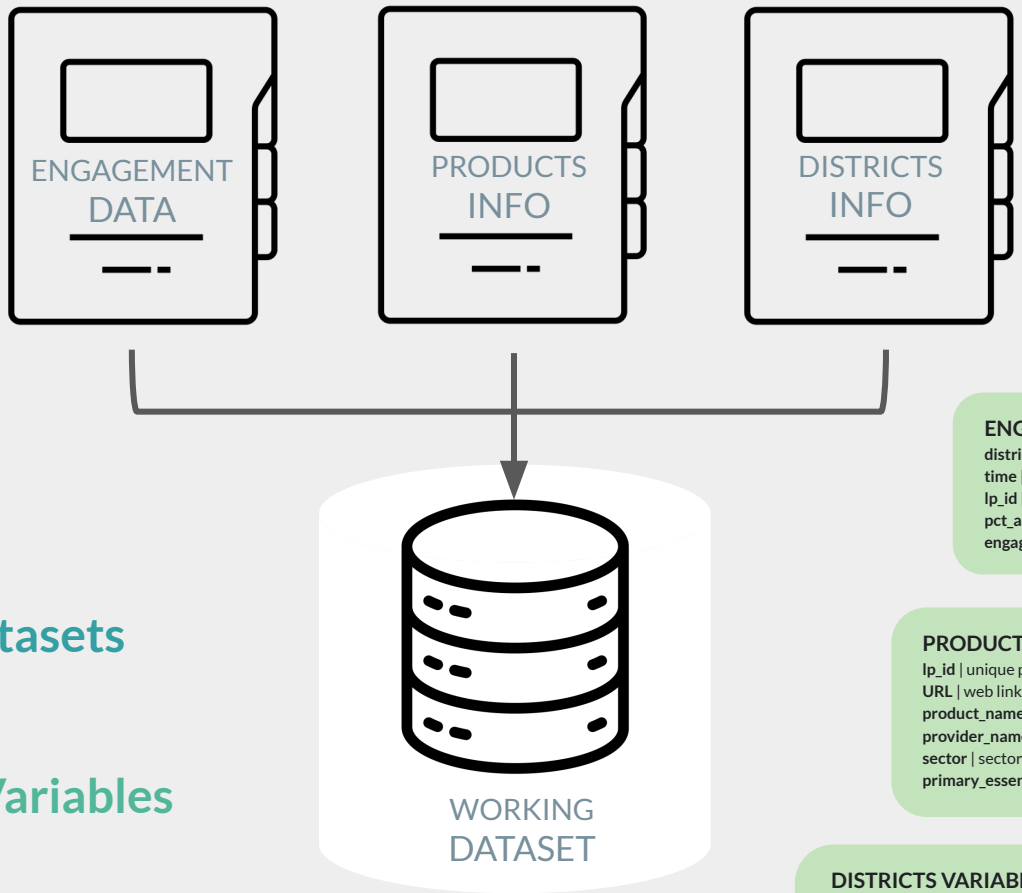
BACKGROUND

The datasets that are provided contain information from edtech engagement during 2020. Over 200 school districts provided daily engagement data.

Engagements: This file provides information on the percentage of students who logged into digital platforms such as Google Docs, YouTube, Canvas, etc

Products: This file encompasses a list of all the digital products utilized by schools to deliver online education

Districts: This file contains a list of district IDs across various US states



ENGAGEMENT VARIABLES
district_id | district ID
time | date
lp_id | unique product ID
pct_access | % students 1-pg load/day
engagement_index | total pg load/1k students/day

PRODUCTS VARIABLES
lp_id | unique product ID
URL | web link to product
product_name | name of product
provider_name | name of provider
sector | sector of education
primary_essential_function | function of product

DISTRICTS VARIABLES
district_id | district ID
state | US state of district
locale | NCES locale classification of US territory
pct_black.hispanic | % students black/hispanic
pct_free.reduced | % students free/reduced lunch
county_connections_ratio | % high speed connection
pp_total_raw | per pupil total expenditure

3 Datasets

23 Variables

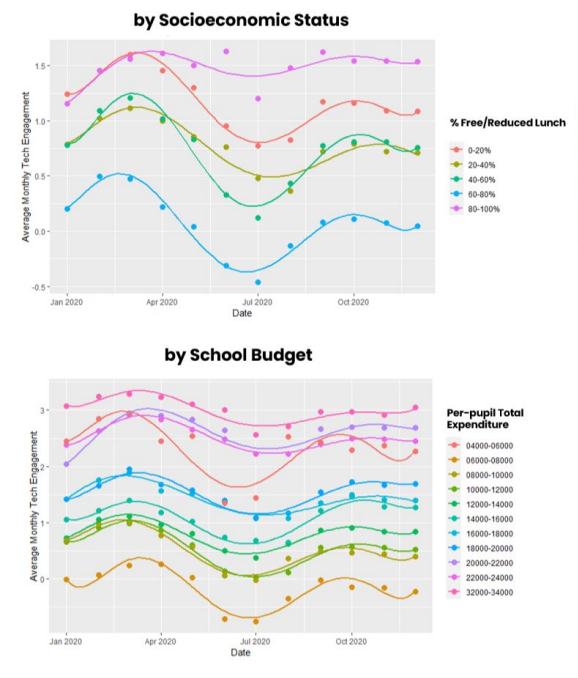
22+ Million Data Points

*Cody Bakley, Julia Elliott, Mary Styers, Scott McQuiggan, Zarifa Zakaria. (2021). LearnPlatform COVID-19 Impact on Digital Learning. Kaggle. https://kaggle.com/competitions/learnplatform-covid19-impact-on-digital-learning

What's Next? | Using Our Data to Learn Something New

Our team carefully examined the datasets, cleaning and transforming necessary metrics. We performed an Exploratory Data Analysis (EDA) consisting of visualizations (which can be found in the full report).

STUDENT TECH ENGAGEMENT DURING COVID-19



PURPOSE

We are interested in examining student engagement with digital platforms over time because it is an indicator of access to resources that are essential for student success.

TAKEAWAYS

From socioeconomic status, race, and school budget, we also see trends in engagement based off demographics that may cue us to their importance and influence on different populations. Including these variables when analyzing tech engagement will give us more accurate predictions.

INTERPRET

School districts with the 60%-80% of free/reduced lunch students have the lowest rates of average monthly tech engagement indicated by the blue line in the Socioeconomic Status chart.

Schools with higher per-pupil expenditures also have higher rates of average monthly tech engagement from their students as indicated by the pink and purple lines in the School Budget chart.

Additional Exploratory Data Analysis (EDA) key findings below:

LINE CHARTS

We identified time trends in our data that allowed us to create a semesters time reference which will enhance interpretability of our upcoming statistical models

Socioeconomic status, Race, and school budget are likely influencing the studied populations and need to be thought about with predictive.

RADAR CHARTS

We identified profiles that behave similarly and suggest a correlation between lower tech engagement and limited internet access, as well as lower diversity and affluence.

Many profiles focus on predominantly white and affluent communities. These profiles have better internet access which suggests they are not at risk of low tech-engagement.

TREE MAPS

We learned that selecting highly represented groups as reference points enhances our understanding of geographic areas' behavior in terms of student tech engagement, boosting statistical power for more accurate predictions and identifying populations in need of improved resource access. When creating our models we can use these as reference groups and understand the bias incorporated within our dataset.

03 Statistical Models

The preliminary findings of the EDA will help us bridge the understanding between our dataset and our problem statement and allow us to form hypotheses which will drive our statistical models.

Through Modeling, our team quantified the association of several key factors that impact student tech engagement rates.

STATISTICAL MODELS

Our goal is to identify key factors that impact student success in the form of tech engagement.

To quantify the impact from our dataset, we constructed 3 models that predict tech engagement (table on right). The tuned random forest performed best with the lowest error rate

METHOD

`Glm, dtree, and randomForest in RStudio`

PREDICTORS

`locale, regions, pef2, sector2, Semester, pct_black.hispanic, pct_free.reduced, pp_total_raw`

RESPONSE

`Log_engagement_index` | *"Total page-load events per one thousand students of a given product and on a given day."*

RANDOM FOREST

~190 trees, and 6 branch nodes yielded the best tuning for appropriate fitting and highest accuracy

MODEL	ROOT MEAN SQUARE ERROR
generalized linear model (AIC- stepwise)	2.735586
decision tree (pruned)	2.848849
random forest (tuned)	**2.718192** (lowest error, highest accuracy)

MODEL TAKEAWAYS

Overall, we produced 3 separate statistical models to predict tech engagement from school district data in an attempt to discover what variables play a key role in the success of digital learning.

The generalized linear model performed well, and informed us that **Suburbs in the South and West regions had lower rates of Engagement.**

The random forest model performed best, and informed us that **school districts with lower per pupil total expenditures and high rates of free/reduced lunch had lower rates of Engagement.**

When consulting stakeholders, we must consider their regional makeup, the funding their school has available for students, and the affluence of the neighborhood. These variables play the biggest role in students' access to digital learning.

What's Next? | From Models to Solutions

Statistical and predictive models help us identify region and population with low edtech engagement. We then come up with practical solutions with cutting edge deliverables to address the gaps.

To inform stakeholders on **Who** needs the resources to improve edtech engagement, and **What** digital learning solutions can align with their needs, our team has created a **dashboard, a mobile dashboard, and a chatbot.**

Our goal is student success via digital learning, and these deliverables utilize the data and findings to help stakeholders make informed decisions.

DASHBOARD

Our Tableau Dashboard provides educators and parents with an immersive experience of interactive data analytics.

Users navigate effortlessly between student engagement analysis and digital learning platform insights, placing analytical power directly in the hands of users.

Complemented by a sleek mobile application, this dashboard is easily accessible on the go, ensuring a seamless and convenient experience.

CHATBOT

Our EdTech Product Chatbot deliverable will fulfill the mission of LearnPlatform by Instructure by addressing inquiries from educators about online learning solutions, enabling them to make well-informed choices.

This chatbot is an artificial intelligence (AI) bot that employs the OpenAI's ChatGPT algorithms to generate new responses based on the ed-tech products information we gathered through their websites

The chatbot is capable of answering common educational platform inquiries and needs. In essence, our chatbot closes the loop, embodying a seamless transition from artificial intelligence to a user-friendly, practical solution.

04 Visualization Dashboards - Tableau

Our team initiated a visualization dashboard development using Tableau to gain deeper insights into student profiles and their pandemic usage of digital products. Tableau, including Tableau Mobile, provides convenient access to data on the go.

TOP DIGITAL PROVIDER / STUDENT ENGAGEMENT TAKEAWAYS

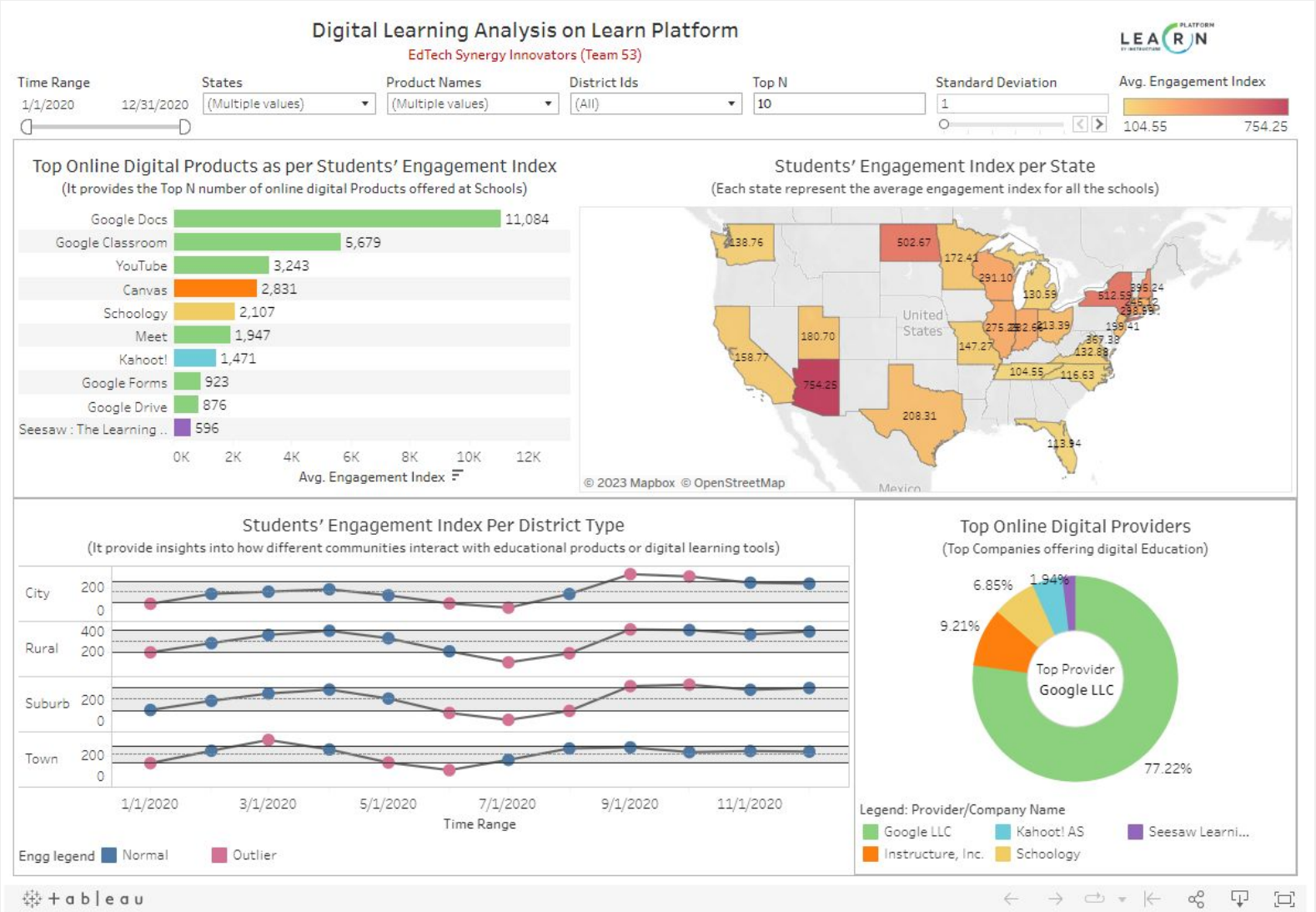
- **Google LLC was the top online digital provider.**
- **Student engagement peaked in April and September. Usage decreased during summer and winter breaks.**

We used standard deviation to find outliers, indicating high and low engagement indexes.

DASHBOARD OBJECTIVES:

To identify top online products used by school districts to deliver education to students

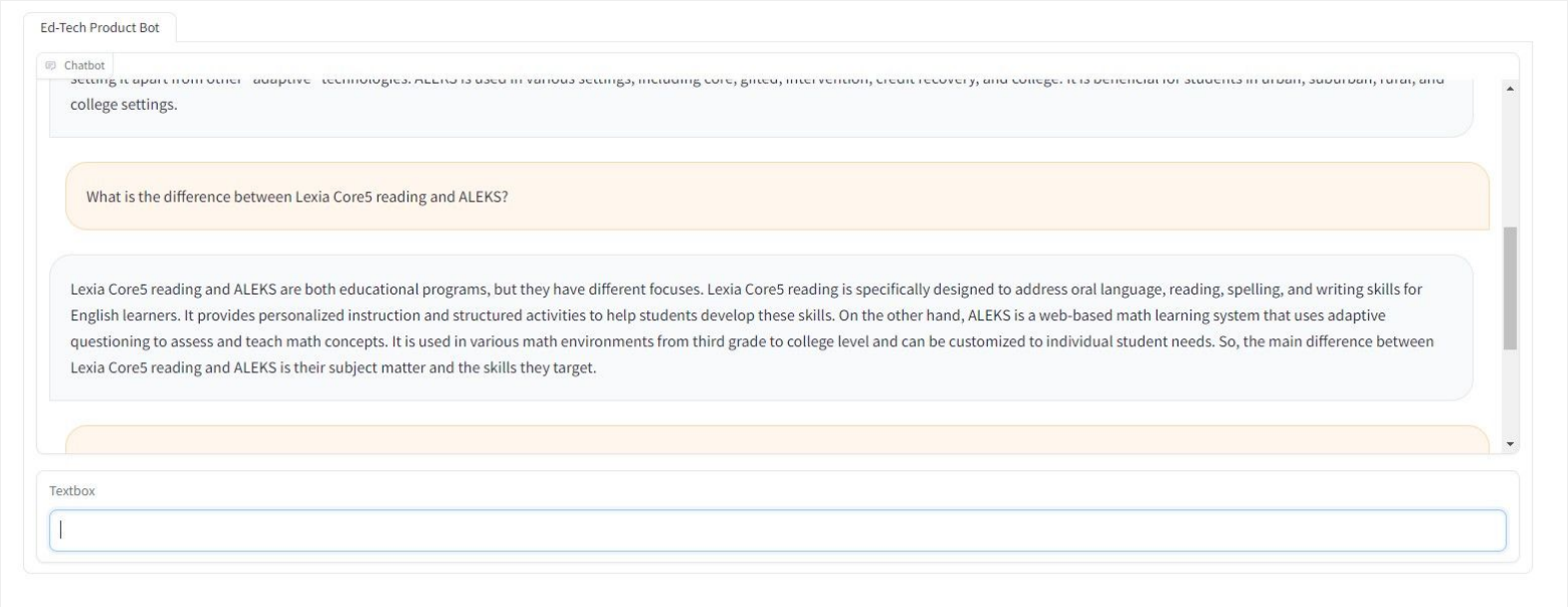
To visualize a breakdown of tech engagement at the district level

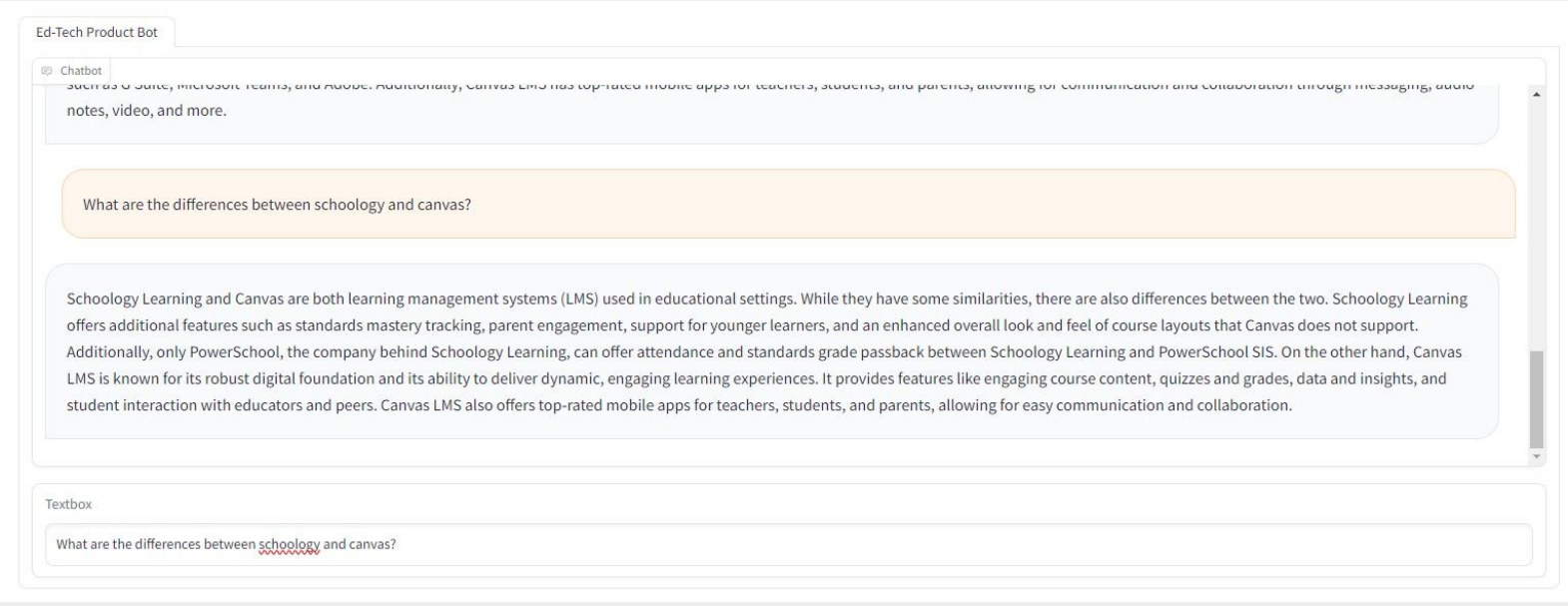


05 Chatbot

An Ed-Tech Product Chatbot, powered by OpenAI's ChatGPT, has been developed to respond to inquiries about various online learning solutions. Integrated with a user-friendly web interface through Gradio, this chatbot can compare online learning solutions and assist educators in identifying products that best suit their requirements, especially in a landscape where there is an abundance of online offerings.

- **Customer Support and Engagement** answer common questions about prominent online learning solutions, offering assistance that enhances user satisfaction and engagement.
- **Data Analytics** provide valuable insights into user behavior, preferences, and pain points through Chatbot data collection.
- **Customization and Integration** tailor to the specific needs of LearnPlatform and integrate with existing systems and processes.





06 Conclusions & Recommendations

RECAP OF CONCLUSIONS

EXPLORATORY DATA ANALYSIS

The graphs and charts we produced in the EDA indicated patterns of an association between per-pupil expenditures and student tech engagement.

This observation and others informed us of hypotheses to test in modeling.

MODELING

From these models, we were able to quantify the impact of the following different demographics on student tech engagement.

School districts with lower per pupil total expenditures and high rates of free/reduced lunch had lower rates of Engagement.

DELIVERABLES

The Tableau Dashboard is a user-friendly tool for exploring student engagement and learning data. With a mobile app, it's easy to use on the go.

Our EdTech Chatbot uses artificial intelligence to help with edtech product questions, offering an efficient way for educators to pinpoint products that meet their requirements

RECOMMENDATIONS FROM OUR TEAM

Our goal from the beginning was to ensure that all students have access to digital learning resources to ensure their success in school no matter what demographics they come from. Our objective was to make information accessible to educators via our deliverables.

- We recommend the continued research and collaboration of school districts with data scientists to identify at-risk populations of low tech engagement rates. Ensuring they have access to the resources is key to students' success. These populations may include school districts of Suburbian locale, low per-pupil expenditures, high populations of free/reduced and Black/Hispanic student populations.
- We recommend the continued improvement of the deliverables. In predictive modeling, we covered 3 different models, but would recommend testing different models like XGBoost which could provide higher predictive accuracy, and would have been built out if time allowed.
- In the dashboard, we recommend the incorporation of data filters when consulting different geographical regions and locales. Seeing the importance of geolocational variables from our EDA, tuning the dashboards for personalized consulting sessions with school districts will make the data you present representative of the communities you interact with.
- In the chat bot, we recommend making it accessible through different online portals to school administrators who may have questions about digital learning resources. Having that type of assistance at their fingertips will help them make informed decisions about providing the resources and funding for their students' success.

Overall, there is still more work that can be done on the front of research, but we hope that our recommendations provide a solid launching point to advise school districts on the next steps they can follow to ensure the success of each and every student via digital learning resources.