



Analysis & Prediction of Covid-19 Impact on Digital Learning

- Bhavya Patlolla

Agenda



Introduction

Data Overview

Exploratory Data Analysis

Packages & Algorithm

Results

Challenges

Summary

Introduction

- 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 about the exacerbating digital divide and long-term learning loss among America's most vulnerable learners continue to grow.
- This analysis is done to explore and show how the data can show us the picture of digital connectivity and engagement in 2020 and the effect of the COVID-19 pandemic on online distance learning.

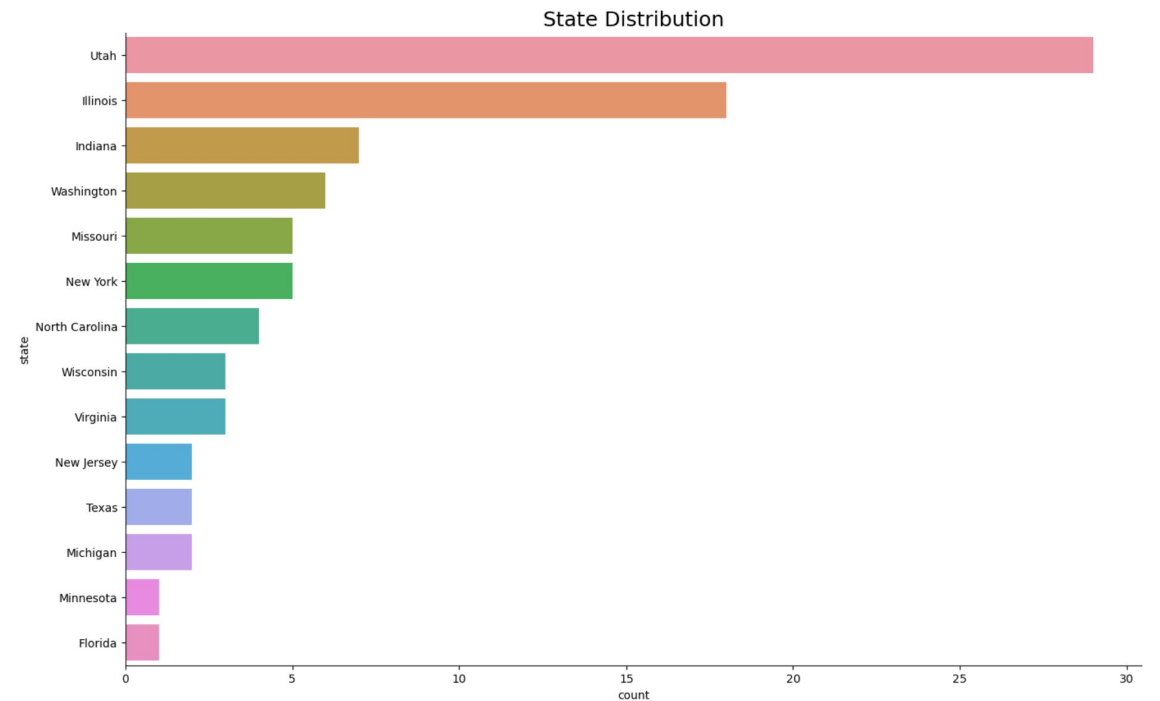
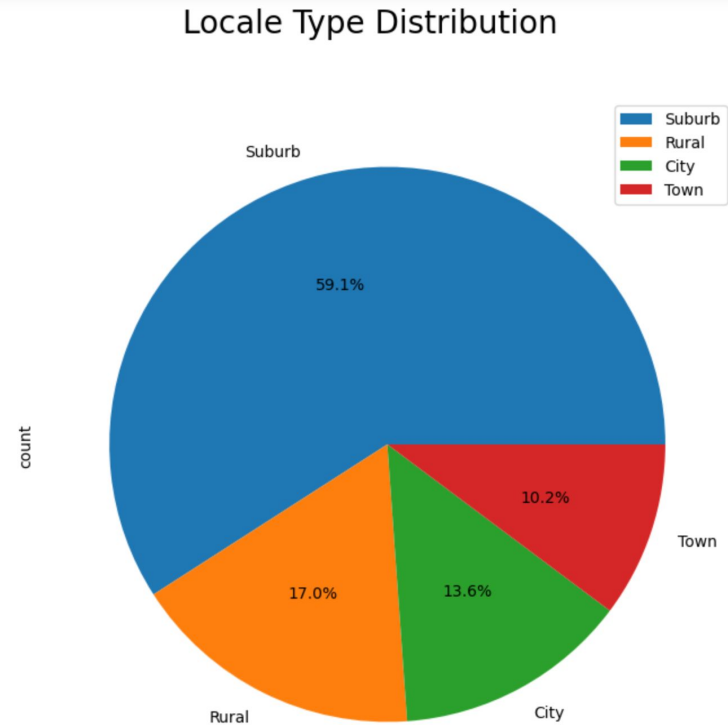
Data Overview

- It includes three basic sets of files that help us to perform the analysis
- The **districts_info.csv** file includes information about the characteristics of school districts, including data from NCES and FCC.
- The **products_info.csv** file includes information about the characteristics of the top 372 products with the most users in 2020.
- The **engagement_data.csv** is based on LearnPlatform's Student Chrome Extension. The engagement data have been aggregated at the school district level.

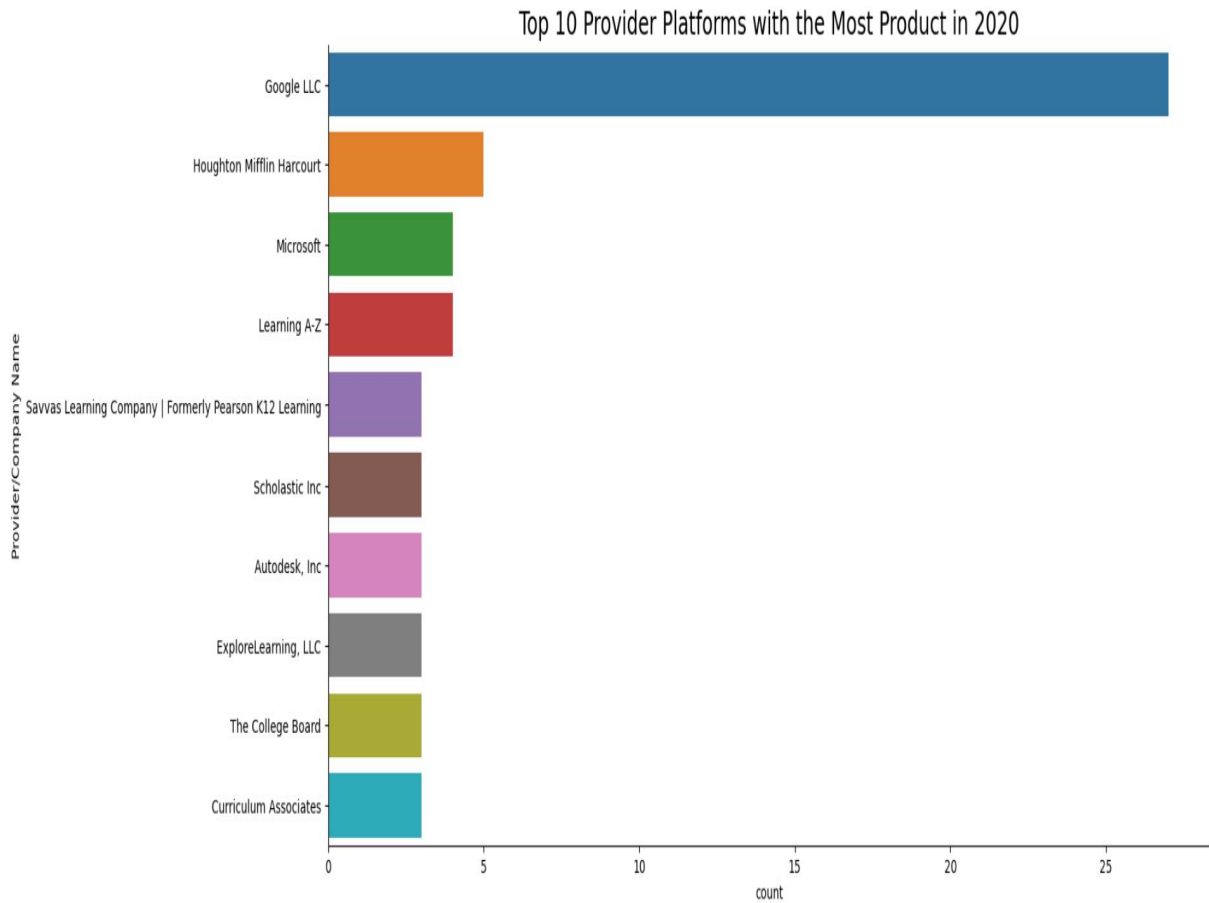
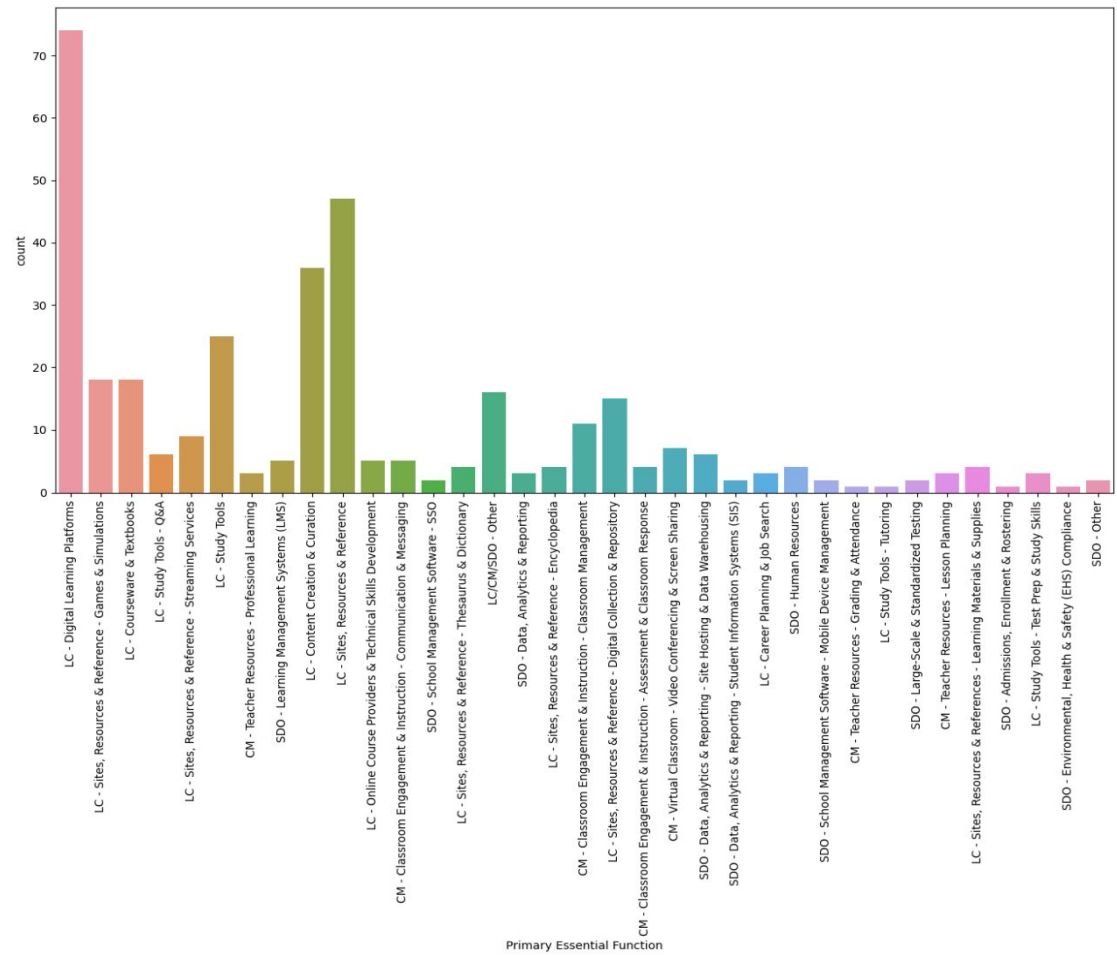


Exploratory Data Analysis

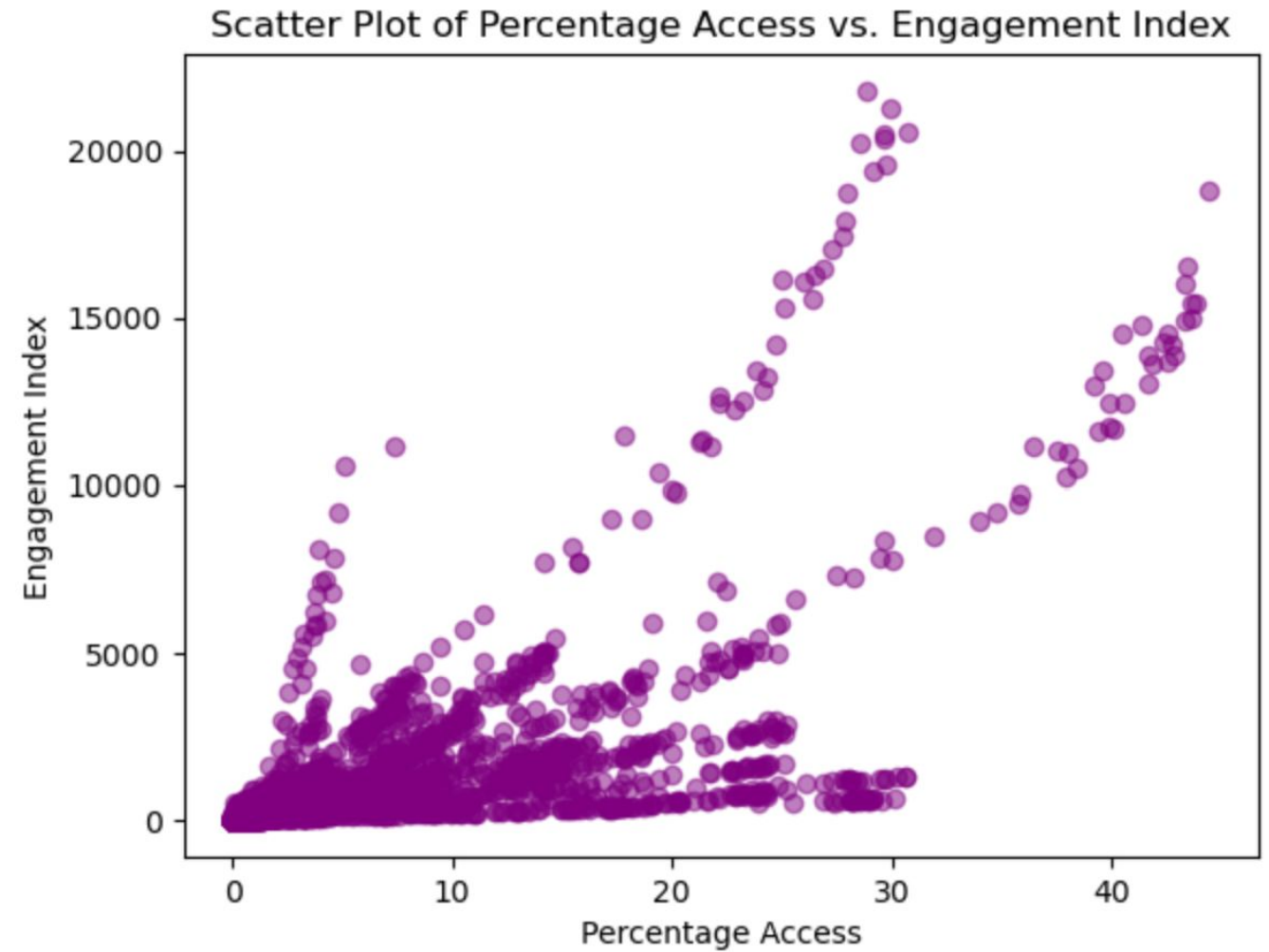
Districts Data



Products Data



Engagement Data



Packages & Algorithm

- Pandas
 - Numpy
 - Matplotlib
 - Seaborn
 - sklearn
- **Random Forest Classifier**- Random Forest Classifier's ensemble approach, ability to handle overfitting, feature importance analysis, versatility, robustness to missing data, and parallelization make it a powerful and widely used choice in the machine learning community.

Results

- Accuracy for model prediction -1.0
- Classification Report

Classification Report:

	precision	recall	f1-score	support
High	1.00	1.00	1.00	8
Low	1.00	1.00	1.00	17656
Medium	1.00	1.00	1.00	59
accuracy			1.00	17723
macro avg	1.00	1.00	1.00	17723
weighted avg	1.00	1.00	1.00	17723

Challenges



Data Cleaning- Removing Null values, Removing Duplicates, Handling Missing Data



Data Preprocessing- Preparing Train & Test Data



Model Selection- Selecting the appropriate model for the data set.

Summary



The COVID-19 pandemic has boosted the usage of digital learning platforms and studying from home, forcing every student to use it to help them get an education.



The suburb area had the highest distribution of overall students.



Among all the states, Utah had more students using digital platforms.



Google LLC was the top provider of all digital learning platforms.



Those with greater than 20% access gained the highest engagement index with digital platforms.



The overall analysis is restricted to the US. Participation in the Kaggle competition is a great learning.



Thank You