

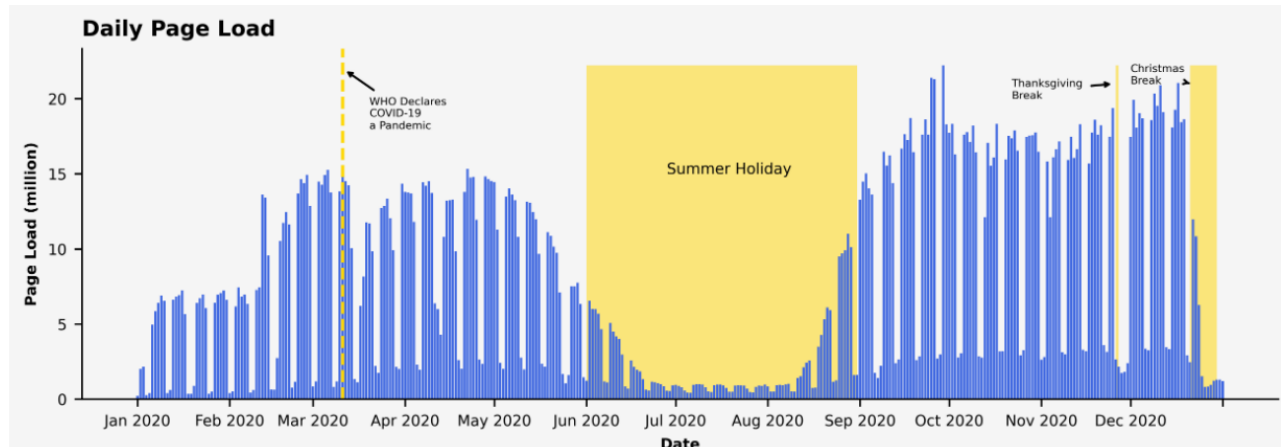
Exploratory Data Analysis

Distribution

Distribution:

- What is the picture of digital connectivity and engagement in 2020?
- How many products were being accessed daily?

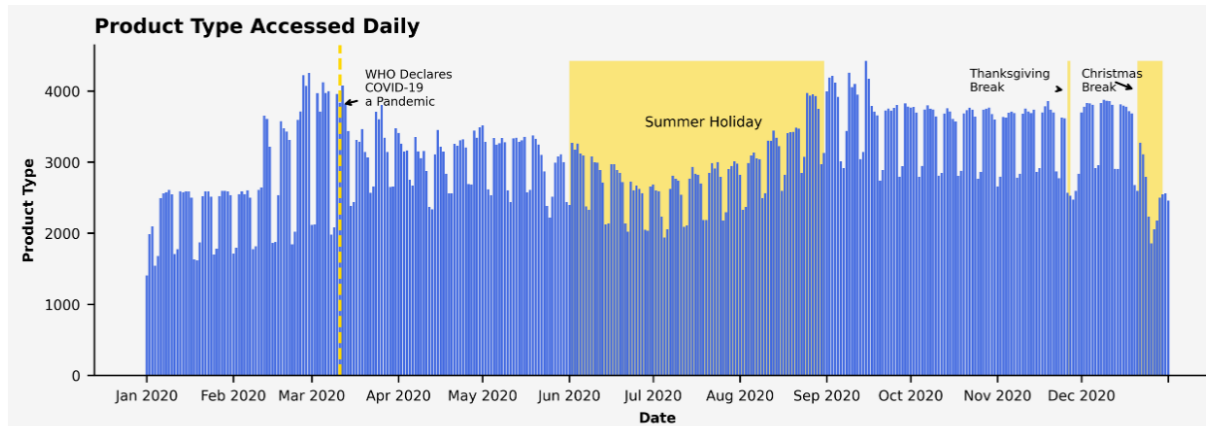
a) Page-Load



Observations:

- Overall, throughout the year, the page-load record is in its highest from September to December whereas at its lowest during Summer holiday (June-August).
- In early March, after WHO declared COVID-19 a pandemic, there is a slight increase in page-load but the difference is not significant. Then, it declines gradually until summer holiday where it reaches its lowest record. As the fall semester started in August, there is an increase in page-load.
- We can also see that page load is consistently much higher in weekdays than in weekends. Similarly, page load is also much lower during holiday seasons. In other words, students are not using the products intensively during weekends and holidays.

b) Product Accessed



Observations:

- There were total of 8646 products in 2000 and approximately 2000-4000 products being used daily.
- From Jan to mid-Feb 2020, there were approximately more than 2500 products used daily and it increased to more than 3000 and even close to 4000 products by end of February. The frequency trend is similar with the page load distribution.

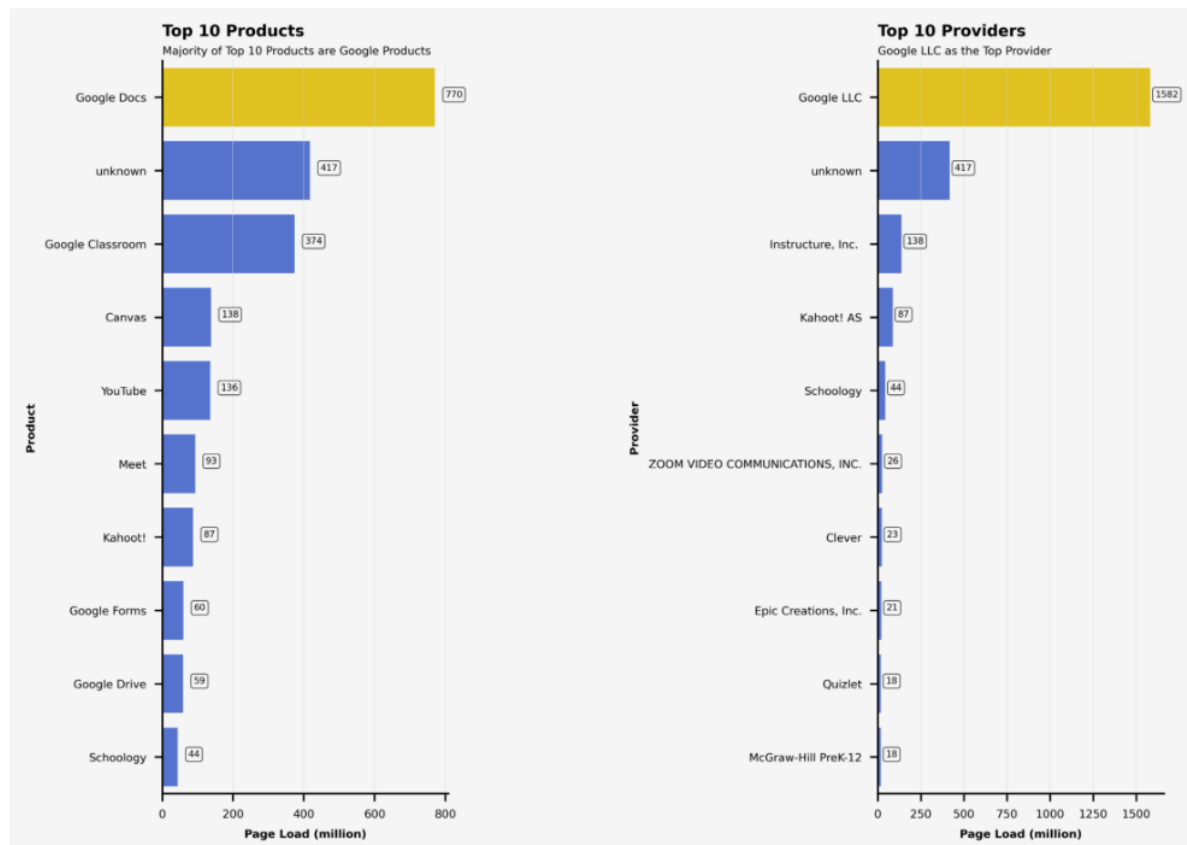
2.2 Product

What are the top 20 most frequent used products, in what sectors were these products used in and what functionality do these products provide?

a) Product and Provider

Observations:

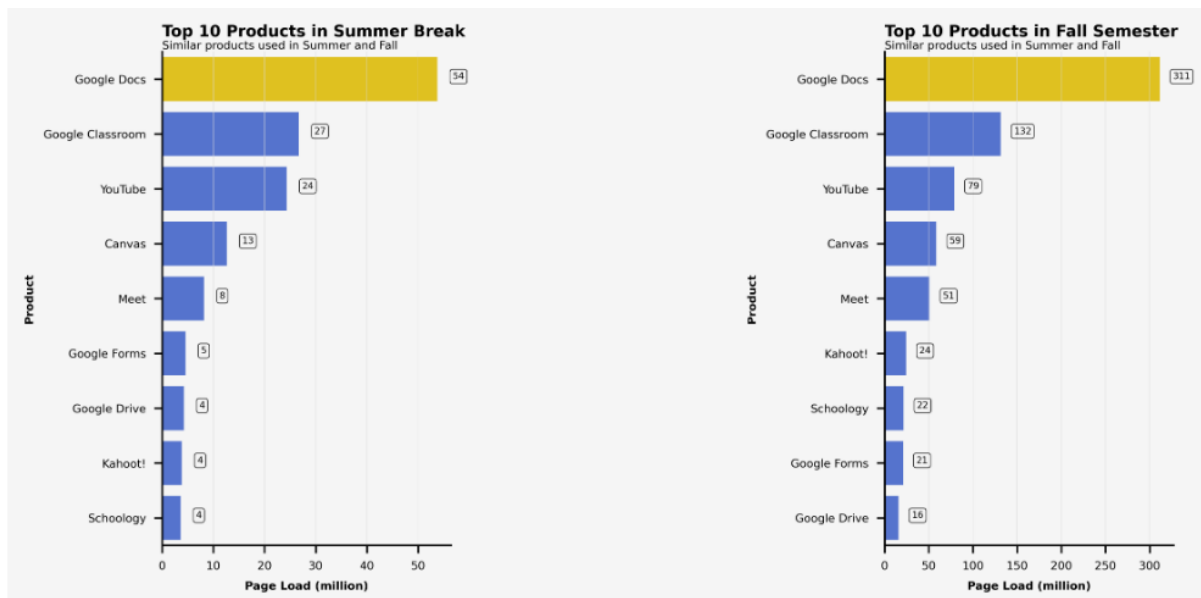
- Majority of the top 10 products used are Google products and the top provider is Google which contributed around 1600-million-page load in year 2020.
- On the other hand, there are also many missing data, which makes the 'unknown' as the second highest in both Top 10 Products and Top 10 Providers.
- Besides, YouTube also ranks as the top 5 product which suggest that besides education purpose, students may also be using the product for entertainment purpose.



b) Product Preference in Summer Break and Fall Semester

Observations:

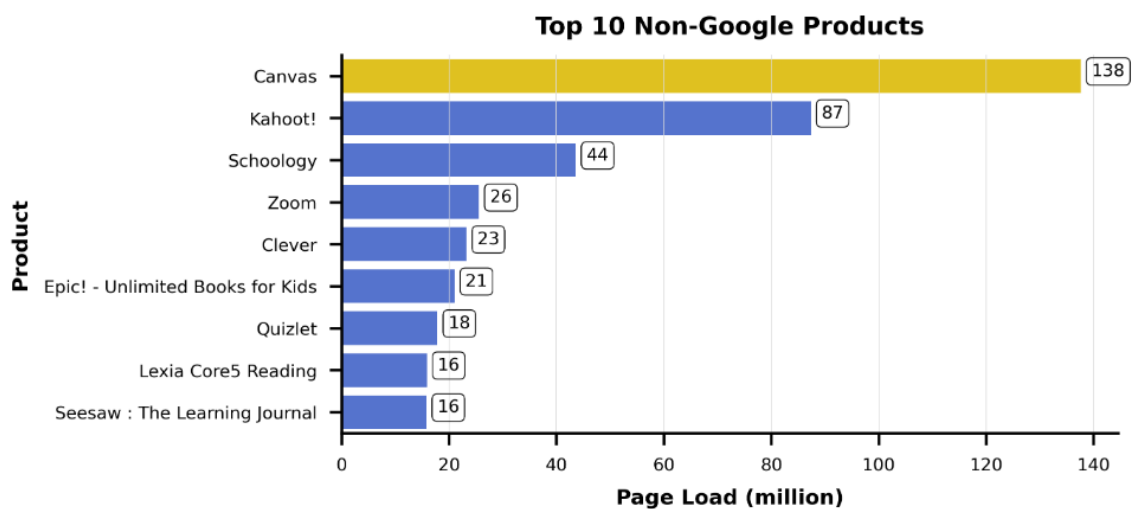
There is not much difference in the top 10 product preference in Summer Break and Fall Semester. Majority of the top 10 products in both periods are from Google.



c) Non-Google Product

Observations:

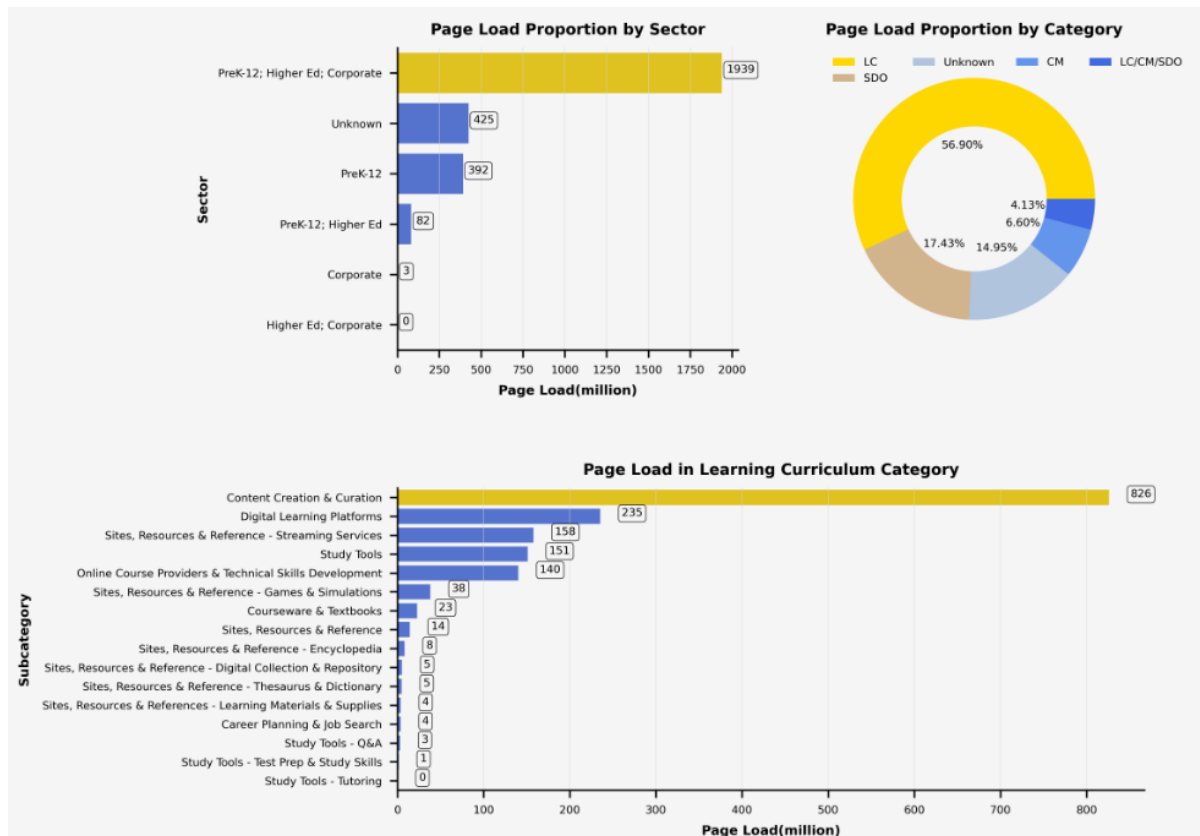
Canvas, Kahoot! and Schoology are ranked as the top 3 products when we exclude Google products.



d) Product Function and Sector

Observations:

- About 2 billion (68%) of page load was contributed by the product which can be used for PreK-12, Higher Ed and Corporate, followed by the unknown sector which contributed 425 million (15%) of page load, whereas the third place was from product for preK-12 sector (close to 4 million of page load).
- By category, 57% of page load was for Learning & Curriculum category, followed by 17.43% for the School & District Operations category and 14.95% from the unknown.
- Under the Learning & Curriculum category, 51% of page load was from product for Content Creation & Curation subcategory, followed by Digital Learning Platforms (15%) and Sites, Resources & Reference- Streaming Services (9.8%).

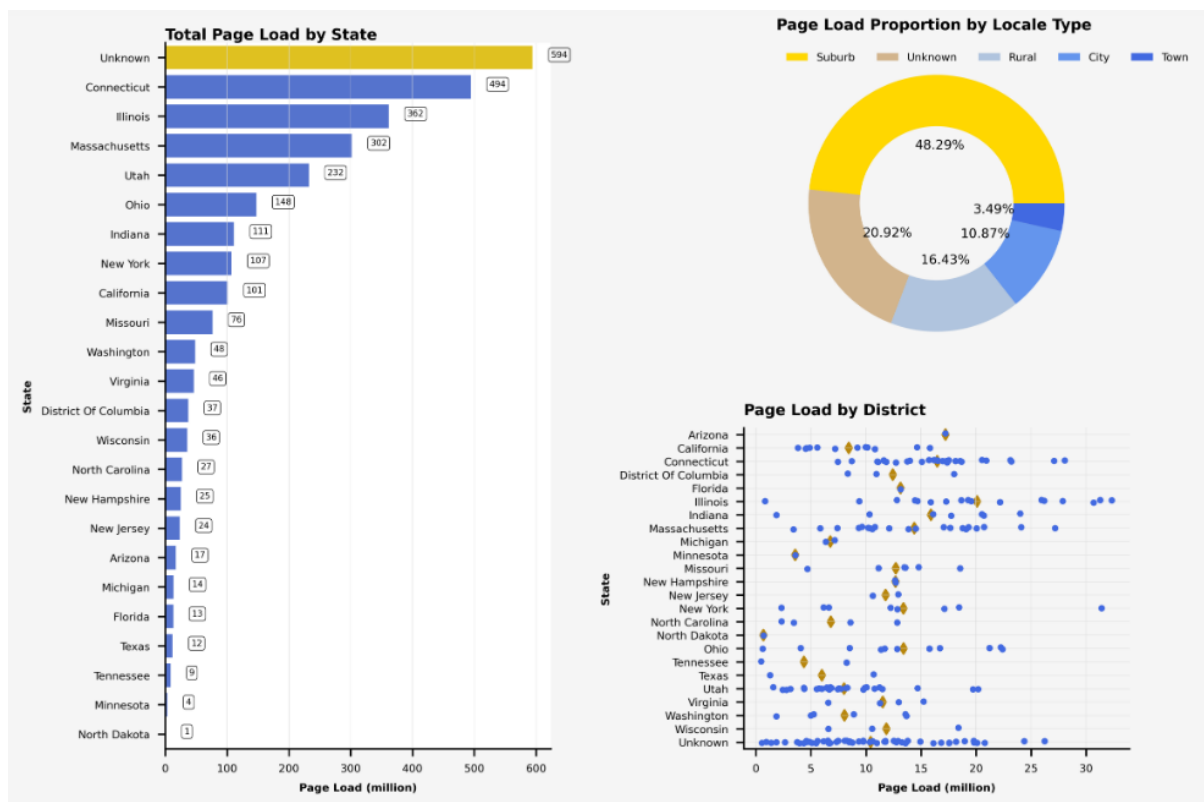


2.3 DEMOGRAPHIC

Demographic: How does page load differ across different location? What are the effects from demographic context (e.g., race/ethnicity, ESL, learning disability) and socioeconomic status?

Observations:

- Out of 2840 million page load, 21% (600 million) of page load comes from an unknown state which stands the first position of total page load ranking by state, followed by 18% from Connecticut (500 million page-load), 13% from Illinois (362 million) and 11% Massachusetts (300 million).
- By locale type, suburb area accounts for 48% of the total page load, followed by unknown (21%) and rural area (16%); whereas town has the lowest contribution (3.5%). As the number of school districts recorded varies in all states, we can check on their mean page load for a better picture. Illinois has the highest mean page load (20 million). Interestingly, there is only one school from Arizona been included. Also, the graphs suggest that there are still many school districts have not been mapped in the dataset, which may shift the current observation patterns.

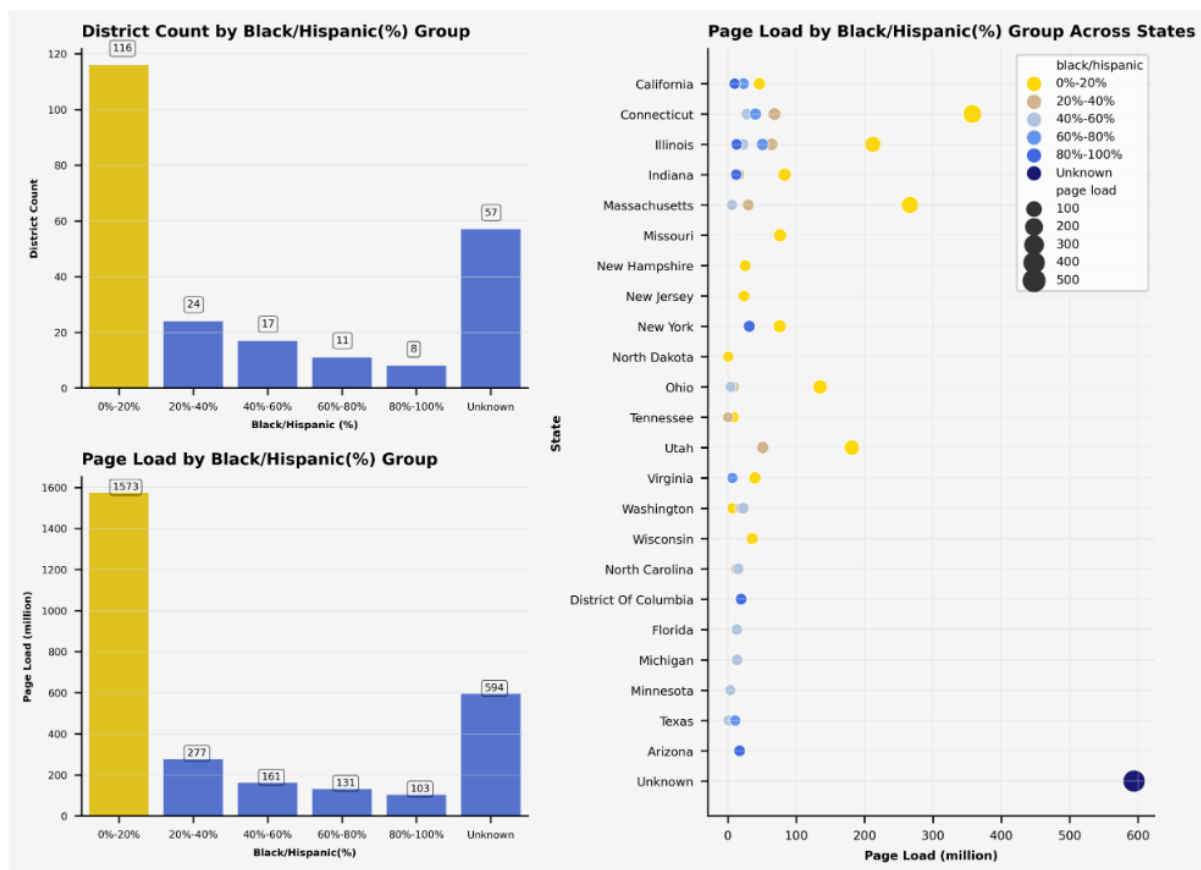


2.4 RACE/ETHNICITY:

What is the distribution of black/hispanic percentage in states and does this make a difference in total page load?

Observations:

- Out of 233 districts, 50% have 0-20% black/hispanic, followed by 25% with unknown information and 10% with 20-40% black/hispanic. 55% of page-load comes from school districts with 0-20% black/hispanic, followed by 21% unknown and 10% from districts with 20%-40% black/hispanic.
- Of all the states included, more than 100-million-page load were from districts with 0-20% black/hispanic. On the other hand, there are close to 600-million-page load do not have state and black/hispanic percentage information, which also accounts for the biggest proportion in the data.

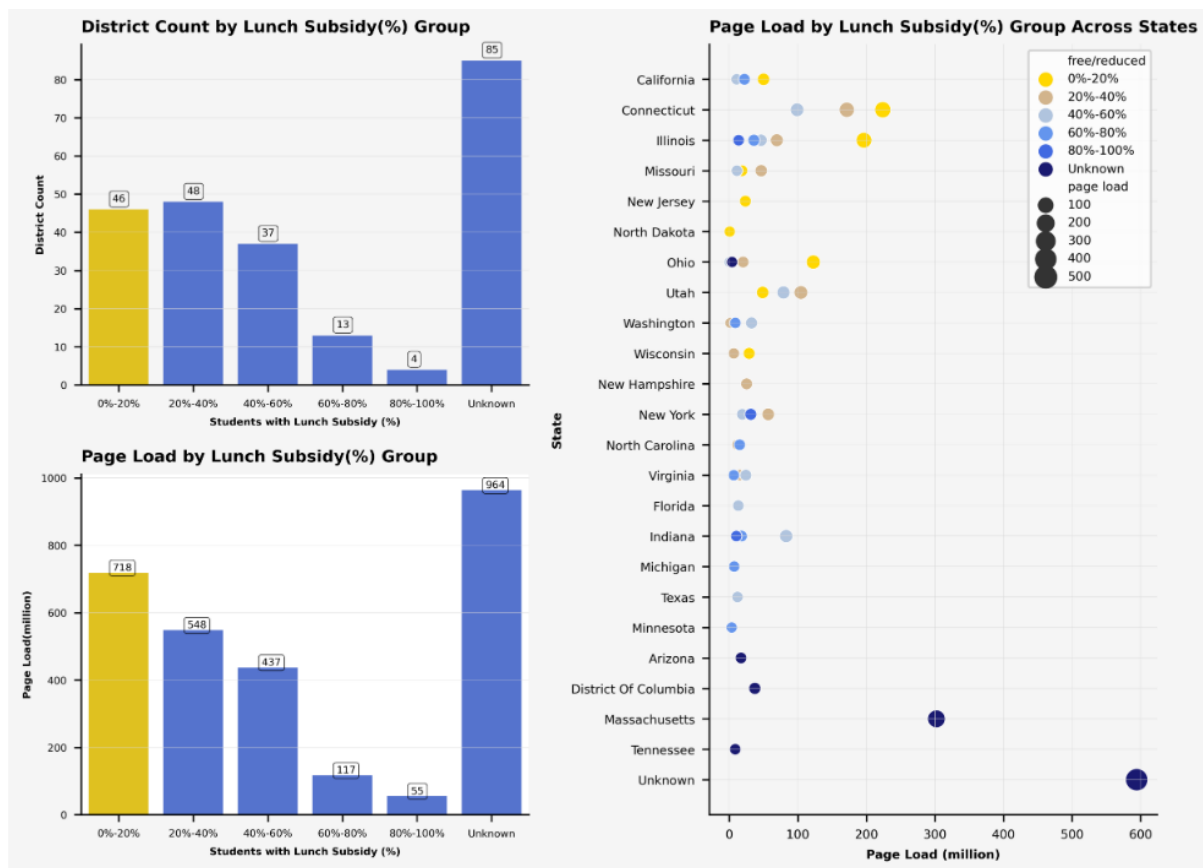


2.5 Financial Status

Students Percentage with Lunch Subsidy

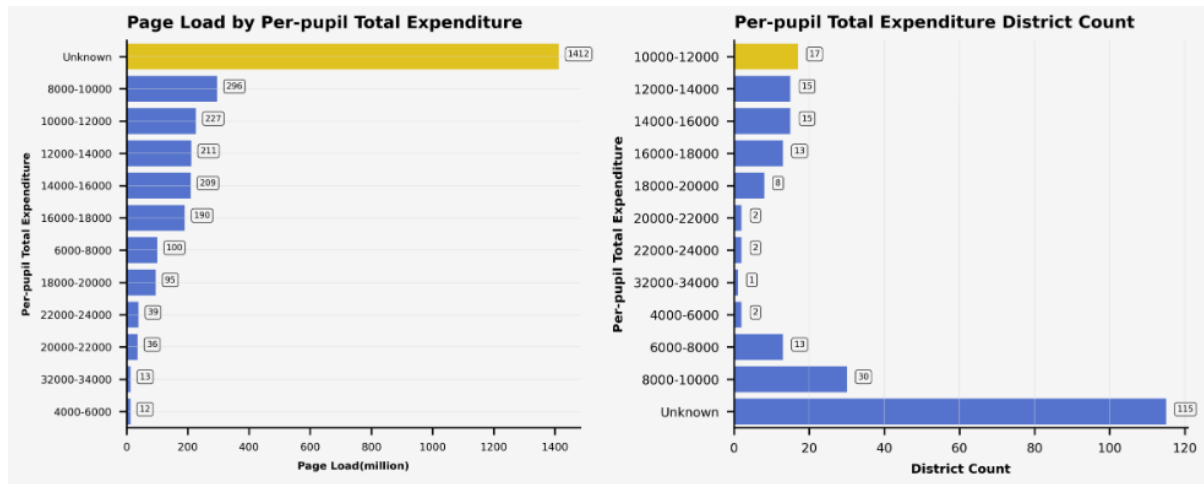
Observations:

- Unfortunately, majority of the districts and page load counts do not have information on the student percentage with eligibility for free or reduced-price lunch.
- About 40% of districts have 0-40% of students eligible for free or reduced-price lunch, whereas only 4 districts have 80-100% of students eligible for free or reduced price lunch.
- 25% of page load was contributed by school district with 0-20% of students eligible for lunch subsidy, followed by 19% page load from district with 20-40% and 15% from districts with 40-60%.
- Excluding the missing values, states with more than 100-million-page load comes from district with 0-20% and 20-40% of students eligible for free or reduced price lunch.



Per-pupil Total Expenditure

- The expenditure data are school-by-school and the median value was used to represent the expenditure of a given school district.
- About 50% of the page load and district does not have per-pupil total expenditure information.
- About 40% of page-load comes from group with USD 8000-18000 per-pupil total expenditure, whereas only 3% of page load comes from group with more than USD20000 per-pupil total expenditure.
- 44% of district have USD 6000-18000 per-pupil total expenditure, while only 5% of districts have more than USD18000 per-pupil total expenditure.



2.6 CONNECTIONS

County	Connections Ratio	District Count
0	100%-200%	1
1	18%-100%	161
2	Unknown	71

Observations:

- 30% of the districts have missing information on the county connections ratio.
- 69% of the districts have 18-100% of county connections ratio where only 1 district with 100-200% county connections ratio.