External Factors Affecting Average California SAT Scores

Introduction

It is well known that different environments can have an impact on students. Our aim this time is to find the relationship between each external factor and SAT scores of California students through exploration and visualization in the given data.(SAT is the American College Entrance Examination)

Meaning of Factors

Grade: grade level. county: counties in California. calwork:Percent qualifying for CalWorks (income assistance). lunch:Percent qualifying for reduced-price lunch. computer:Number of computers. expenditure: Expenditure per student. income: District average income (in USD 1,000). english: Percent of English learners. read: Average reading score. math: Average math score. average_score: mean of read and math score

Question Explored by Visualization

1. Can you display the distribution of scores of different counties in different grades?
2. For variables calworks,lunch, computer, expenditure, income and english, which has higher effects on the scores?

Bubble Map of Grades

It is the Bubble Map of Grades of Different Counties in California.
It can be seen that KK-06 is mostly in the coastal area.
The distribution of KK-08 covers the whole of California, and there are more in the central and upper part of California.

Checkbox of Grade

Allow users to change the Grades

Checkbox of County

Allow users to change the Counties

Student ID:27181944, Name:Weisheng Jiang, Tutors: Mohit Gupta, Vaibhavi Bhardwaj

Selected Counties

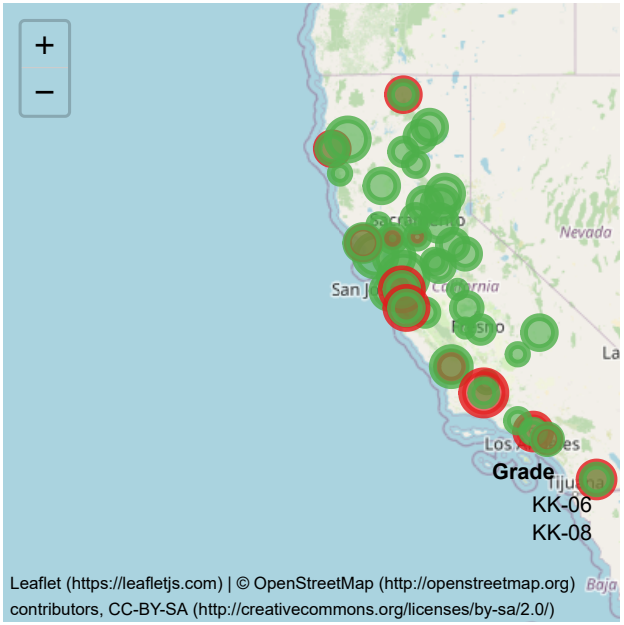
All

Grade

Select grades:

- [x] KK-06
- [x] KK-08

Bubble Map of Grades of Different Counties in California



Select the Counties You Wish To See

County:

- [x] All
- [ ] Alameda
- [ ] Butte
- [ ] Calaveras
- [ ] Contra Costa
- [ ] El Dorado
- [ ] Fresno
- [ ] Glenn
- [ ] Humboldt
- [ ] Imperial
- [ ] Inyo
- [ ] Kern
- [ ] Kings
- [ ] Lake
- [ ] Lassen
- [ ] Los Angeles
- [ ] Madera
- [ ] Marin
- [ ] Mendocino
- [ ] Merced
- [ ] Monterey
- [ ] Nevada
- [ ] Orange
- [ ] Placer
- [ ] Riverside
- [ ] Sacramento
- [ ] San Benito
- [ ] San Bernardino
- [ ] San Diego
- [ ] San Joaquin
- [ ] San Luis Obispo
- [ ] San Mateo
- [ ] Santa Barbara
- [ ] Santa Clara
- [ ] Santa Cruz
- [ ] Shasta
- [ ] Siskiyou
- [ ] Sonoma
- [ ] Stanislaus
- [ ] Sutter
- [ ] Tehama
- [ ] Trinity
- [ ] Tulare
- [ ] Tuolumne
- [ ] Ventura
- [ ] Yuba

Select bar of X-axis

Allow users to change the X-axis

Select bar of Y-axis

Allow users to change the Y-axis

Bubble Chart of Y-Axis on X-Axis

The bubble chart shows the relationship between two variables.
When Y-axis is average score, it can be seen that calworks, lunch, english have a strong negative relationship with average score
Income has a strong positve relationship with average score.
Computer has no relationship with average score.
Expenditure have a weak positive relationship with average score.

Percentage stacked Bar Chart

The percentage is divided into 5 levels, each present 20%.
When mouse on the bar, it shows the amount and percentage for this county.
When Y-axis is average score,it shows that Alameda have the highest percent of very high average score with 1 unit. Mateo have 5 units of very high average score, but only have 29% inside San Mateo.

Correlation Matrix

This matrix shows the correlations between each variable.
Based on our question, we will focus on the last row and last column.
It shows the correlationship of age score and othe external features.
From the graph and number it can be seem that, average score have a strong negative correlationship with calwork, lunch,

Factor Affecting (Please Ensure X and Y different Selection)

X-axis:

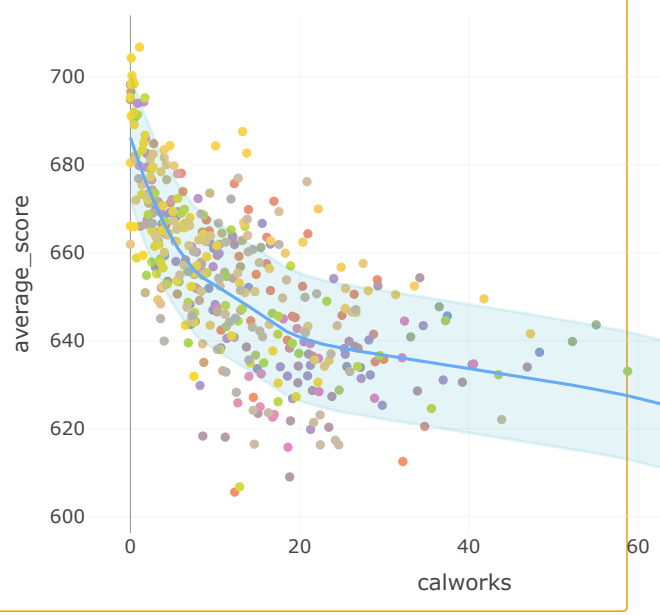
calworks

Factor Affected (Please Ensure X and Y different Selection)

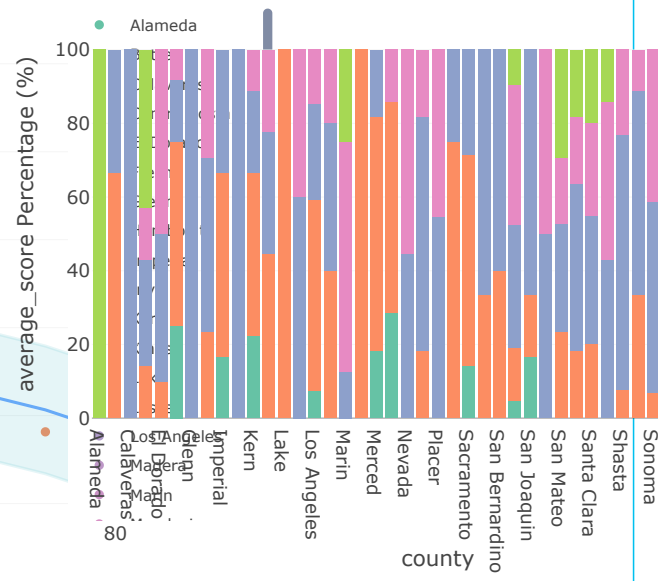
Y-axis:

average_score

Bubble Chart of Y-Axis on X-Axis



Percentage Stacked Bar Chart of Y-Axis in Different Counties



Correlation Matrix of Different features for All/Specific Counties and Grades

Conclusion

For question 1, the distribution is shown in the bubble map format.KK-06 is concentrated on the coast. KK-08 covers a wide area in California.
For question 2, the lunch has the highest negative effects on average score. The income has the highest positive effect on average score.
Positive factor: strong: income, weak: Expenditure
Negative factor: strong: calwork, lunch, english
Minimal or no effect: computer

english.
And average score have a positive correlationship with income.
Average score have no unclear correlationship with computer in weak negative side, and weak positive correlationship with expenditure.