



# California Public School Enrollment vs Census Unemployment Rate



*Rookie Corgis*

Collins  
Ellis  
Natali  
Seidy

# Outline

---



1. Background
2. Data Retrieval
3. Data Exploration
4. Data Cleaning
5. Analysis Process
6. Conclusions
7. Implementations
8. Next Steps



# Background



## The Spark

- Our team has first hand knowledge of current enrollment trends in Los Angeles County.
- Enrollment drops were widely recorded and published at the beginning of the COVID-19 pandemic across the country.



## The Questions

- Has public school enrollment changed significantly in California in recent years?
- Is public school enrollment in California correlated with unemployment?



Los Angeles Times

LOG IN



CALIFORNIA

LAUSD enrollment decline continues during online learning, with unexpected drop of 6,000 kindergarteners



Analysis Background



Data Retrieval



Data Exploration



Cleaning Data



Analysis Process



Conclusions



Implementations



Limitations

# Hypotheses



#1

*Public school enrollment has been trending downward across the state of California even before the beginning of the COVID-19 pandemic.*

#2

*Unemployment rates are heavily correlated with public school enrollment across the state of California.*



Analysis Background



Data Retrieval



Data Exploration



Cleaning Data



Analysis Process



Conclusions



Implementations



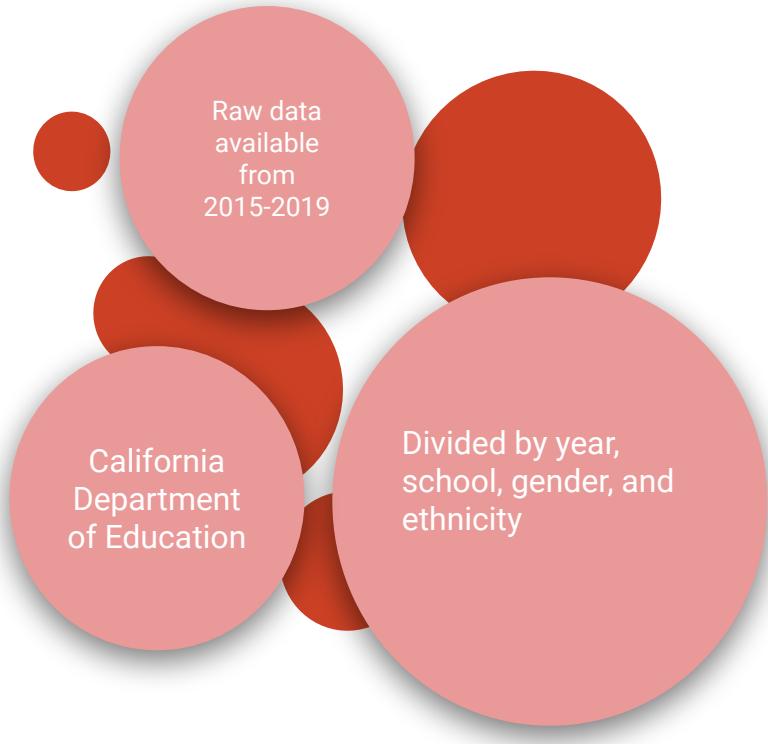
Limitations



## 01 California School District CA Department of Education

## 02 Unemployment Rate U.S. Census Bureau *American Community Survey* Subject Tables & Variables API





California Department of Education

Raw data available from 2015-2019

Divided by year, school, gender, and ethnicity

## 01 California School District

The screenshot shows the California Department of Education's website with a navigation bar including 'Teaching & Learning', 'Testing & Accountability', 'Finance & Grants', 'Data & Statistics', 'Specialized Programs', and 'Learning'. The main content area is titled 'Enrollment by School' and describes downloadable data files for school-level enrollment. It notes that the data includes primary enrollments only. Below this is a table listing data files from 2012-13 to 2019-20, each with a link to its file structure.

Year of Data	File Name	File Structure
2019–20	<a href="#">enr19</a> (TXT; 14MB; Posted 13-Apr-2020)	<a href="#">File Structure: School Enrollment (2007–19)</a>
2018–19	<a href="#">enr18</a> (TXT; 14MB; Posted 28-Mar-2019)	<a href="#">File Structure: School Enrollment (2007–19)</a>
2017–18	<a href="#">enr17</a> (TXT; 14MB; Posted 11-Apr-2018)	<a href="#">File Structure: School Enrollment (2007–19)</a>
2016–17	<a href="#">enr16</a> (TXT; 13MB; Posted 11-Apr-2017)	<a href="#">File Structure: School Enrollment (2007–19)</a>
2015–16	<a href="#">enr15</a> (TXT; 13MB; Posted 17-May-2016)	<a href="#">File Structure: School Enrollment (2007–19)</a>
2014–15	<a href="#">enr14</a> (TXT; 13MB; Posted 24-Apr-2015)	<a href="#">File Structure: School Enrollment (2007–19)</a>
2013–14	<a href="#">enr13</a> (TXT; 13MB; Posted 28-Apr-2014)	<a href="#">File Structure: School Enrollment (2007–19)</a>
2012–13	<a href="#">enr12</a> (TXT; 13MB; Posted 05-Apr-2013)	<a href="#">File Structure: School Enrollment (2007–19)</a>



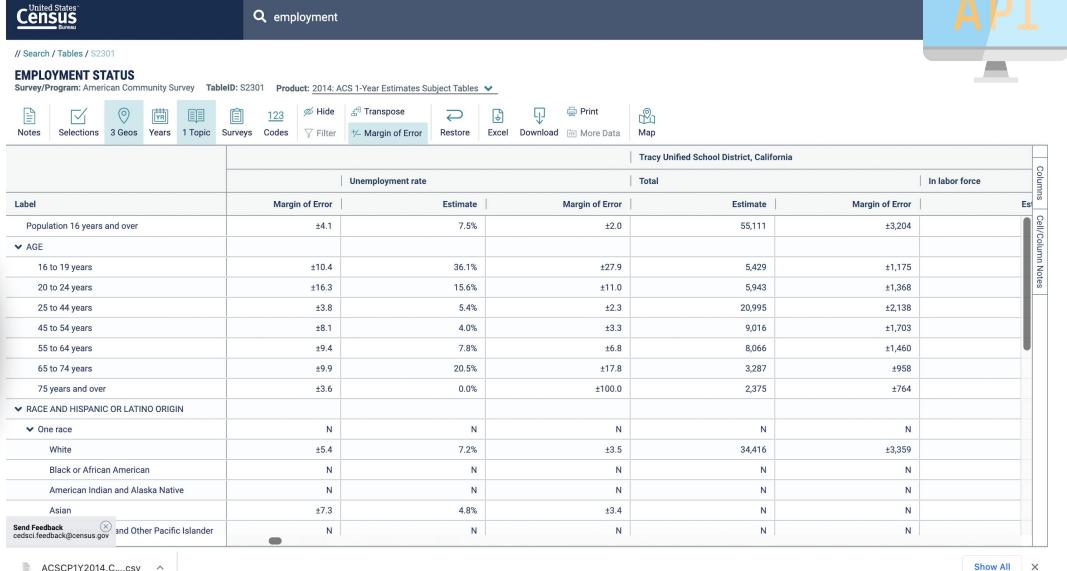


API Data available from 2015-2019

U.S. Census Bureau

Divided by year, school districts, and school district types (Unified, Elementary, Secondary)

## 02 Unemployment Rate



The screenshot shows a data retrieval interface for the United States Census. The search bar at the top contains the query "employment". Below the search bar, the title "EMPLOYMENT STATUS" is displayed, along with the survey program ("American Community Survey"), table ID ("S2301"), and product ("2014: ACS 1-Year Estimates Subject Tables"). The interface includes various filters and download options (Excel, PDF, Map). The main content area displays a table of unemployment rates for different demographic groups in Tracy Unified School District, California. The table includes columns for Label, Margin of Error, Estimate, Margin of Error, Total Estimate, Margin of Error, and In labor force. The data is categorized by AGE and RACE AND HISPANIC OR LATINO ORIGIN.

Label	Unemployment rate		Total		In labor force	
	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error	Estimate
Population 16 years and over	±4.1	7.5%	±2.0	55,111	±3,204	
▼ AGE						
16 to 19 years	±10.4	36.1%	±27.9	5,429	±1,175	
20 to 24 years	±16.3	15.6%	±11.0	5,943	±1,368	
25 to 44 years	±3.8	5.4%	±2.3	20,995	±2,138	
45 to 54 years	±8.1	4.0%	±3.3	9,016	±1,703	
55 to 64 years	±9.4	7.8%	±6.8	8,066	±1,460	
65 to 74 years	±9.9	20.5%	±17.8	3,287	±958	
75 years and over	±3.6	0.0%	±100.0	2,375	±764	
▼ RACE AND HISPANIC OR LATINO ORIGIN						
▼ One race	N	N	N	N	N	N
White	±5.4	7.2%	±3.5	34,416	±3,359	
Black or African American	N	N	N	N	N	N
American Indian and Alaska Native	N	N	N	N	N	N
Asian	±7.3	4.8%	±3.4	N	N	N
Send Feedback <a href="mailto:cdecsl.feedback@census.gov">cdecsl.feedback@census.gov</a> and Other Pacific Islander	N	N	N	N	N	N



# Data Exploration



## School Enrollment Data

```
list_of_files = ["Resources/2014-15-enrollment.csv",
                 "Resources/2015-16-enrollment.csv",
                 "Resources/2016-17-enrollment.csv",
                 "Resources/2017-18-enrollment.csv",
                 "Resources/2018-19-enrollment.csv",
                 "Resources/2019-20-enrollment.csv"]

enroll_dfs = []
for i,filename in enumerate(list_of_files):
    X1 = pd.read_csv(filename)
    X1["year"] = 2014 + i
    enroll_dfs.append(X1)

# Combine the data into a single dataset

combined_data = pd.concat(enroll_dfs)
enrollment_df = pd.DataFrame(combined_data)
```



## Employment Data

```
import pandas as pd
import requests

# Census Subject Table API for Employment Status data within Elementary School Districts in California for 2015
url = "https://api.census.gov/data/2015/acs/acs1/subject?get=group(S2301)&for=school%20district%20(elementary)&in-state:06"

#Request for HTTP Data from Census API, which is working <Response [200]>
response = requests.get(url)

#Resetting data from API Data for future formatting, lists data in one column
response_json = response.json()

#Places data in a dataframe and drops index column 0 (with headers), which has 51 schools and 702 columns of variables
elementaryschool_df = pd.DataFrame(response_json,columns=response_json[0]).drop(0)

#Census Subject Table Variables for Employment Status data
variableurl = "https://api.census.gov/data/2015/acs/acs1/subject/variables.json"

#Request for HTTP Data from Census API and reset data
variables_json = requests.get(variableurl).json()

#View variable for column title
variables_json['variables'][0]['label']

#Label: 'Total!Estimate!Population 16 years and over',
#concept: 'EMPLOYMENT STATUS',
#predicateType: 'int',
#group: 'S2301',
#limit: 0,
#attributed: 'S2301_C01_001M,S2301_C01_001EA'

#Find and replace all columns with variable titles
new_labels = []
for col in elementaryschool_df.columns:
    label = variables_json['variables'].get(col)
    if label is not None:
        label = label['label']
    else:
        if col[-2] == 'E':
            label = variables_json['variables'].get(col[:-1])
            label = label['label'] + '|Annotation'
        elif col[-1] == 'M':
            label = variables_json['variables'].get(col[:-1] + 'E')
            label = label['label'] + '|MarginOfError'
        elif col[-2] == 'MA':
            label = variables_json['variables'].get(col[:-2] + 'E')
            label = label['label'] + '|MarginOfErrorAnnotation'
    new_labels.append(label)
```



Analysis Background



Data Retrieval



Data Exploration



Cleaning Data



Analysis Process



Conclusions



Implementations



Limitations

# Cleaning Data - School Enrollment by district (feat Jupyter Notes)



## Criteria for cleaning:

- Group enrollment data by District and Year, aggregate enrollment numbers
- Add new columns to the dataframe, calculate year-on-year percentage changes
- Format results and replace NaN with zero

```
enr_sub_summary_df = enrollment_df.groupby(["DISTRICT", "YEAR"])\n    .agg({"ENR_TOTAL":"sum"}).reset_index()
```

```
enr_sub_summary_df.head(10)
```

	DISTRICT	YEAR	ENR_TOTAL
0	ABC Unified	2015	20863
1	ABC Unified	2016	20768
2	ABC Unified	2017	20550
3	ABC Unified	2018	20465
4	ABC Unified	2019	20269
5	Acalanes Union High	2015	5546
6	Acalanes Union High	2016	5530
7	Acalanes Union High	2017	5662
8	Acalanes Union High	2018	5683
9	Acalanes Union High	2019	5635

```
# create a new column that refers to the enrollment total from previous year\nenr_sub_summary_df["PREV YR TOT"] = enr_sub_summary_df.groupby(by="DISTRICT")\\n    ["ENR_TOTAL"].shift((1))
```

```
# create a new column that shows percentage change
```

```
enr_sub_summary_df["PERC CHANGE"] = (((enr_sub_summary_df["ENR_TOTAL"]\\n        - enr_sub_summary_df["PREV YR TOT"]) \\n        / enr_sub_summary_df["PREV YR TOT"]) * 100).round(2)
```

```
enr_sub_summary_df
```

	DISTRICT	YEAR	ENR_TOTAL	PREV YR TOT	PERC CHANGE
0	ABC Unified	2015	20863	NaN	NaN
1	ABC Unified	2016	20768	20863.0	-0.46
2	ABC Unified	2017	20550	20768.0	-1.05
3	ABC Unified	2018	20465	20550.0	-0.41
4	ABC Unified	2019	20269	20465.0	-0.96
...	...	...	...	...	...
5077	Yucaipa-Calimesa Joint Unified	2015	9889	NaN	NaN
5078	Yucaipa-Calimesa Joint Unified	2016	9969	9889.0	0.81
5079	Yucaipa-Calimesa Joint Unified	2017	10063	9969.0	0.94
5080	Yucaipa-Calimesa Joint Unified	2018	9982	10063.0	-0.80
5081	Yucaipa-Calimesa Joint Unified	2019	9831	9982.0	-1.51



# Cleaning Data - Unemployment Rate by district (feat Jupyter Notes)



## Preparing Census Employment Status Data:

- American Community Survey
  - Subject Tables & Variables for API Format
- 15 in Total, one per School District Type and year
  - Elementary, Secondary, Unified
  - 2015, 2016, 2017, 2018, 2019
- Consolidated to 9 data point columns including the year, NCES District Code, and District Type
- Combined 15 into one & formatted for merge

```
#Connect python and data
census = pathlib.Path('/Users/nataligracia/git/ca-school-enrollment-trend/all_employment_data_merged.csv')

#read research data to analyze the study results
data = pd.read_csv(census)

cleanone = data.drop(columns=['Unnamed: 0','Unnamed: 0.1','Unnamed: 0.1.1'])

cleantwo = cleanone.loc[cleanone.iloc[:,0].isnull()]
    .dropna(axis=1)
    .rename(columns=(col:col.replace('Estimate!!','') for col in cleanone.columns))

cleanthree = cleanone.loc[cleanone.iloc[:,0].notnull()]
    .dropna(axis=1)
    .rename(columns=(col:col.replace('Estimate!!','') for col in cleanone.columns))

cleanfour = pd.concat([cleantwo,cleanthree],ignore_index=True)

cleanfive = cleanfour.replace(-999999999.0,np.nan)

final = cleanfive.astype({'year':str})

final.to_csv("/Users/nataligracia/git/ca-school-enrollment-trend/CensusDataFinal.csv")
```

```
#Change column titles for columns labeled "None"
assert len(new_labels) == len(elementaryschool_df.columns)

#Confirm the number of columns without titles
sum([1 for x in new_labels if x is None])

2

#Setup new Labels of columns labeled "None"
new_labels[-2] = 'NAME'
new_labels[-1] = 'STATE'
new_labels[1] = 'SCHOOL DISTRICT'

#Create new labels of columns labeled "None"
elementaryschool_df.columns = new_labels

#Create a new dataframe for data without the columns that have "Annotation" in the title, which is 702 columns
without_annotation = elementaryschool_df[[col for col in elementaryschool_df.columns if 'Annotation' not in col]].copy()

#Create a new dataframe for data without the columns that have "MarginOfError" in the title, which is 142 columns
withoutmarginerror = without_annotation[[col for col in without_annotation.columns if 'MarginOfError' not in col]].copy()

#Create a new dataframe for data without the columns that have "Labor Force Participation Rate" in the title, which is
withoutlaborforce = withoutmarginerror[[col for col in withoutmarginerror.columns if 'Labor Force Participation Rate' not in col]].copy()

#Create a new dataframe for data without the columns that have "Sex" in the title, which is 89 columns
withoutsex = withoutlaborforce[[col for col in withoutlaborforce.columns if 'SEX' not in col]].copy()

#Create a new dataframe for data without the columns that have "Poverty Status" in the title, which is 83 columns
withoutpoverty = withoutsex[[col for col in withoutsex.columns if 'POVERTY STATUS' not in col]].copy()

#Create a new dataframe for data without the columns that have "Disability Status" in the title, which is 80 columns
withoutdisabilitystatus = withoutpoverty[[col for col in withoutpoverty.columns if 'DISABILITY STATUS' not in col]].copy()

#Create a new dataframe for data without the columns that have "Educational Attainment" in the title, which is 65 columns
withouteducation = withoutdisabilitystatus[[col for col in withoutdisabilitystatus.columns if 'EDUCATIONAL ATTAINMENT' not in col]].copy()

#Create a new dataframe for data without the columns that have "Age" in the title, which is 35 columns
withoutage = withouteducation[[col for col in withouteducation.columns if 'AGE' not in col]].copy()

#Create a new dataframe for data without the columns that have "Latino" in the title, which is 31 columns
withoutlatino = withoutage[[col for col in withoutage.columns if 'Latino' not in col]].copy()

#Create a new dataframe for data without the columns that have "Race" in the title, which is 10 columns
withoutrace = withoutlatino[[col for col in withoutlatino.columns if 'RACE' not in col]].copy()

#Create new column with combined state and district ID
withoutrace['NCERDist'] = withoutrace['STATE'].astype(str) + withoutrace['SCHOOL DISTRICT']

#Add Year column
withoutrace['year'] = '2015'

withoutrace['School District Type'] = 'Elementary'

elementary2015 = withoutrace.drop(columns=['STATE','SCHOOL DISTRICT'])

elementary2015.to_csv("/Users/nataligracia/git/ca-school-enrollment-trend/Elementary2015.csv")
```



Analysis Background



Data Retrieval



Data Exploration



Cleaning Data



Analysis Process



Conclusions



Implementations



Limitations

# Cleaning Data - Merging Both (feat Jupyter Notes)



## Criteria for data:

- Merging on NCES District code AND year
- Ensuring that NCES District code matches data types
- Cleaning up column names
- Resetting -9999999 results to NaN

	District	Year	Total Enrollment	Percent Change	NCESDist	School District Type	Population, 16+	Employment Ratio, 16+	Unemployment Rate, 16+	Population, 20-64	Employment Ratio, 20-64	Unemployment Rate, 20-64
0	Rocklin Unified	2018	14070	3.814654	600013	Unified	51542.0	59.9	5.6	38279.0	73.4	5.6
1	Rocklin Unified	2019	14125	0.390903	600013	Unified	51882.0	58.6	4.4	36253.0	75.1	4.4
2	Hesperia Unified	2015	23970	NaN	600014	Unified	74447.0	50.4	14.7	58032.0	61.2	14.7
3	Hesperia Unified	2016	23829	-0.588235	600014	Unified	82334.0	48.6	8.7	62228.0	60.4	8.7
4	Hesperia Unified	2017	24138	1.296739	600014	Unified	80751.0	49.3	11.8	60671.0	62.7	11.8
...	...	...	...	...	...	...	...	...	...	...	...	...
969	Val Verde Unified	2015	19833	NaN	691135	Unified	68697.0	55.6	10.7	53982.0	65.5	10.7
970	Val Verde Unified	2016	19924	0.458831	691135	Unified	61066.0	54.0	5.9	46370.0	66.1	5.9
971	Val Verde Unified	2017	20215	1.460550	691135	Unified	72651.0	55.7	10.8	57404.0	67.8	10.8
972	Val Verde Unified	2018	20109	-0.524363	691135	Unified	NaN	NaN	NaN	NaN	NaN	NaN

```
[54]: # merging economic and district enrollment data based on district code and year
econ_district_df = pd.merge(district_df, econ_df, on=["NCESDist", "year"])

[56]: econ_district_df.columns
Index(['Unnamed: 0_x', 'District', 'year', 'ENR_TOTAL', 'Percent Change',
       'Total!!Population 16 years and over',
       'Employment/Population Ratio!!Population 16 years and over',
       'Unemployment rate!!Population 16 years and over',
       'Total!!Population 20 to 64 years',
       'Employment/Population Ratio!!Population 20 to 64 years',
       'Unemployment rate!!Population 20 to 64 years'],
      dtype='object')

[64]: # dropping old index columns
econ_district_df = econ_district_df.drop(columns=["Unnamed: 0_x", "Unnamed: 0_y"])
***

[65]: #verifying data types again
econ_district_df.dtypes
District          object
year            int64
ENR_TOTAL        int64
Percent Change   float64
NCESDist         int64
School District Type    object
Total!!Population 16 years and over   float64
Employment/Population Ratio!!Population 16 years and over float64
Unemployment rate!!Population 16 years and over float64
Total!!Population 20 to 64 years     float64
Employment/Population Ratio!!Population 20 to 64 years     float64
Unemployment rate!!Population 20 to 64 years     float64
dtype: object

[65]: District          object
year            int64
ENR_TOTAL        int64
Percent Change   float64
NCESDist         int64
School District Type    object
Total!!Population 16 years and over   float64
Employment/Population Ratio!!Population 16 years and over float64
Unemployment rate!!Population 16 years and over float64
Total!!Population 20 to 64 years     float64
Employment/Population Ratio!!Population 20 to 64 years     float64
Unemployment rate!!Population 20 to 64 years     float64
dtype: object

[86]: # cleaning up column names to make it more readable
econ_district_df = econ_district_df.rename(columns={
    "year": "Year",
    "ENR_TOTAL": "Total Enrollment",
    "Total Population, 16+": "Population, 16+",
    "Employment/Population Ratio!!Population 16 years and over": "Employment Ratio, 16+",
    "Unemployment Rate, 16+": "Unemployment Rate, 16+",
    "Total!!Population 20 to 64 years": "Population, 20-64",
    "Employment/Population Ratio!!Population 20 to 64 years": "Employment Ratio, 20-64",
    "Unemployment rate!!Population 20 to 64 years": "Unemployment Rate, 20-64"})
[91]: econ_district_df = econ_district_df.replace(-99999999.0,np.nan)
econ_district_df
```





## Definitions

- **Unemployment rate**
  - Percentage of population either 16 years and older or between 20-64 years old who are unemployed and looking to be employed
- **Total enrollment**
  - The aggregate number of students enrolled within a district or across the state
  - Public school enrollment is calculated every year in October on a statewide day called “Census Day”
- **Average enrollment**
  - The average number of students enrolled per school within a district or across the state
- **Enrollment percent change**
  - The percent increase or decrease of enrollment within a district year over year

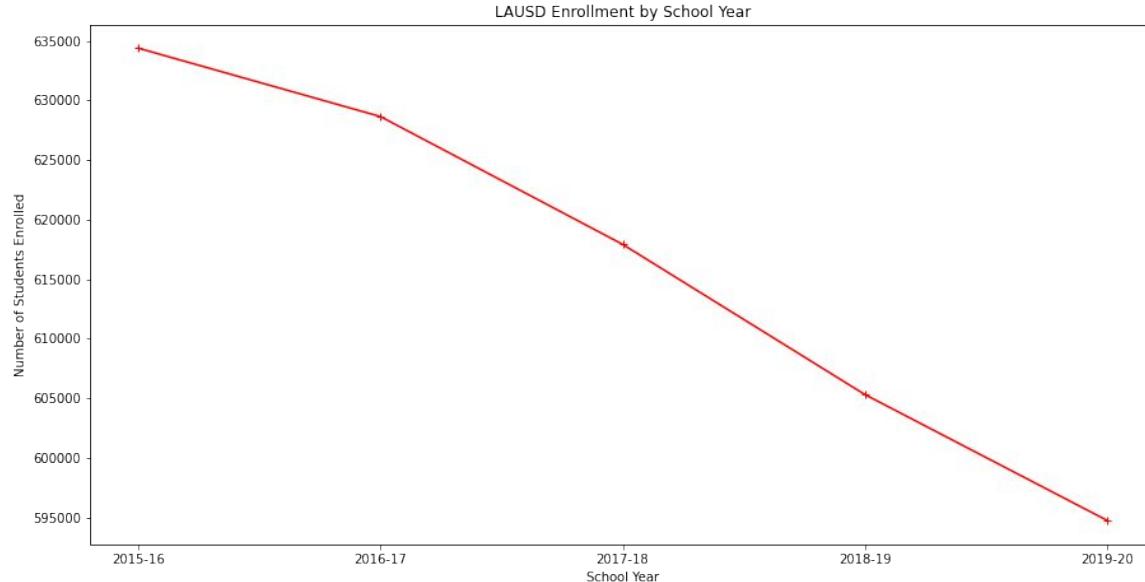




# Analysis Process



## Los Angeles Unified Enrollment Trend



*Enrollment across Los Angeles Unified has decreased every year since the 2015-16 school year.*

*Is this unique to Los Angeles Unified or has this been the trend across all major districts?*





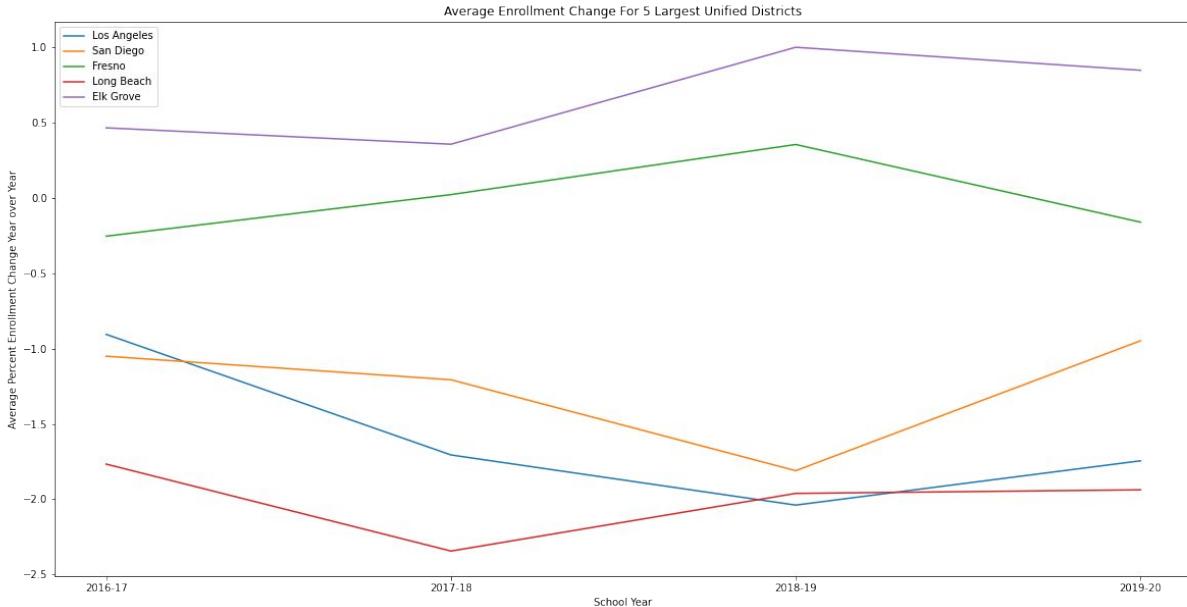
# Analysis Process



## Largest Districts Average Enrollment Changes

*The four largest unified districts had seen a negative trend in their total enrollment since 2015.*

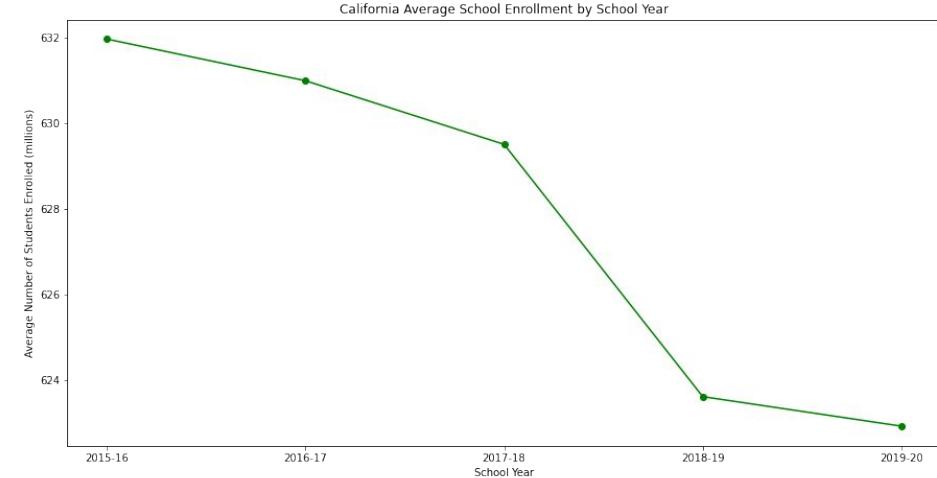
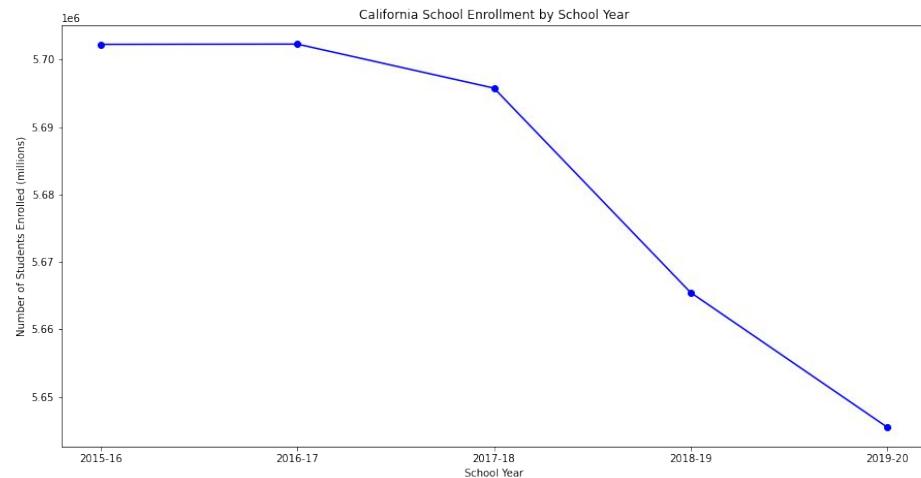
*Is this the case across California?*



# Analysis Process



## California Total and Average Enrollment by School Year



*Total and average public school enrollment across California has been decreasing.*

*What is the relation with unemployment rates? Are economic factors forcing families out?*



Analysis Background



Data Retrieval



Data Exploration



Cleaning Data



Analysis Process



Conclusions



Implementations



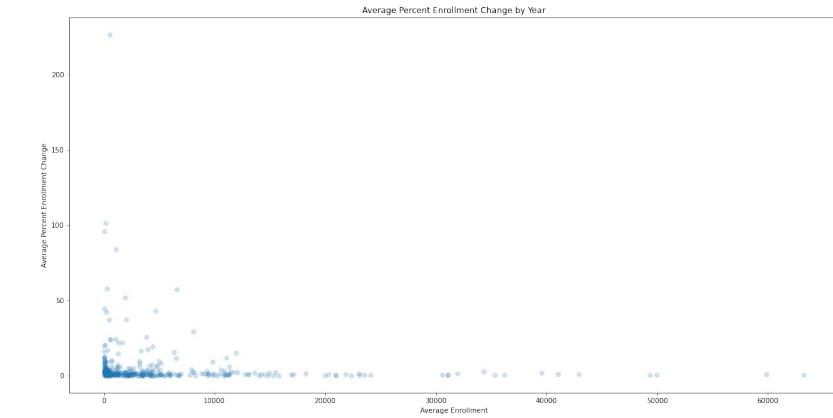
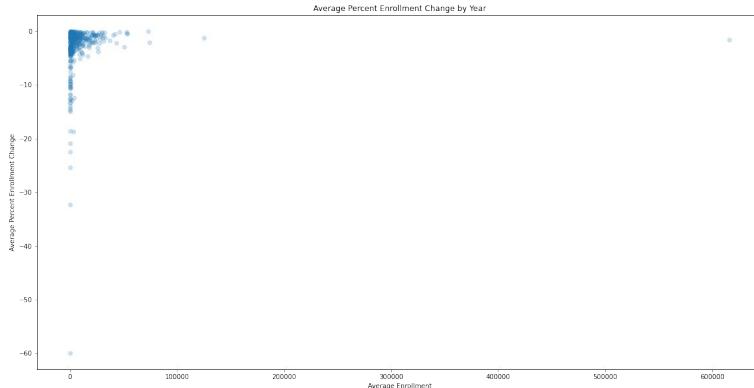
Limitations

# Analysis Process



## Average Enrollment Change by District Size

*Negative Average Enrollment Change*

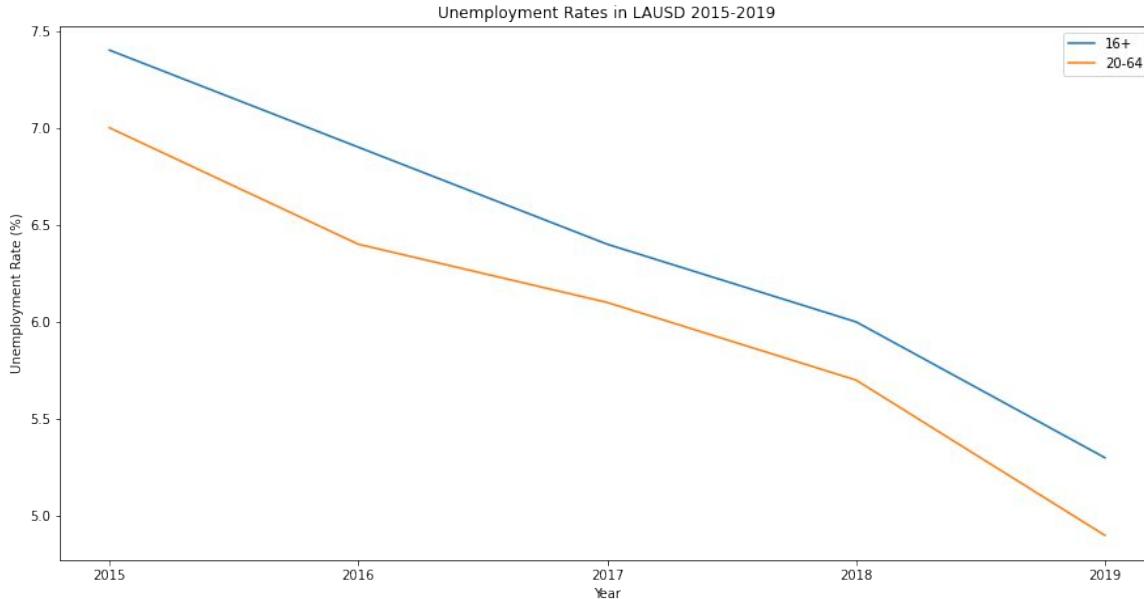




# Analysis Process



## Los Angeles Unified Area Unemployment Rates



*Unemployment for people over 16 and between 20-64 years old has been decreasing in the area bounded by LAUSD since 2015.*

*What does this look like compared to enrollment data?*

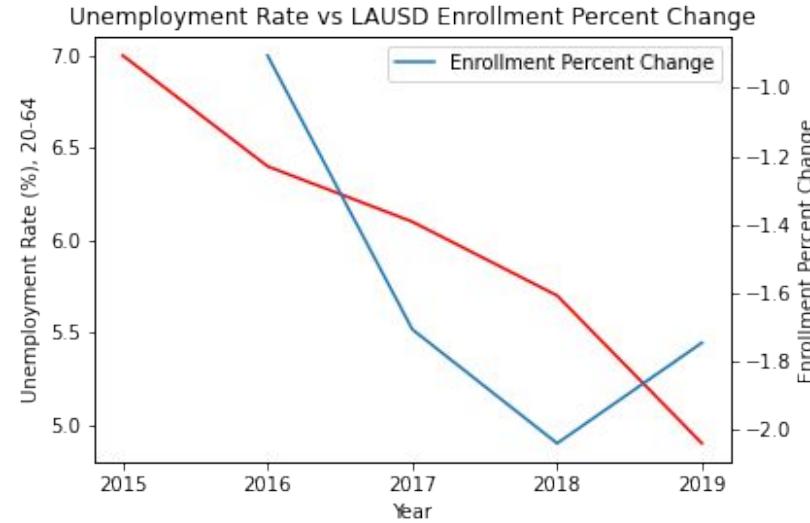
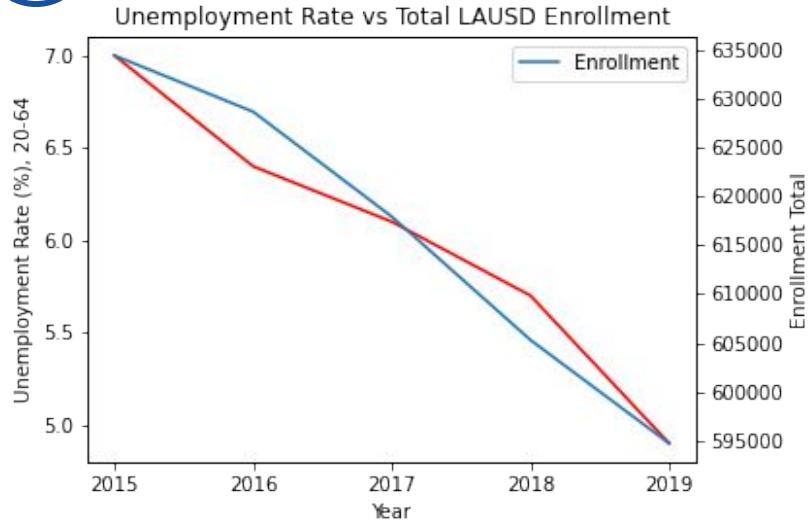




# Analysis Process



## Los Angeles Unified Enrollment vs Unemployment



*It looks like total and percent change of enrollment and unemployment trend in the same direction.*

*What does this trend look like statewide?*

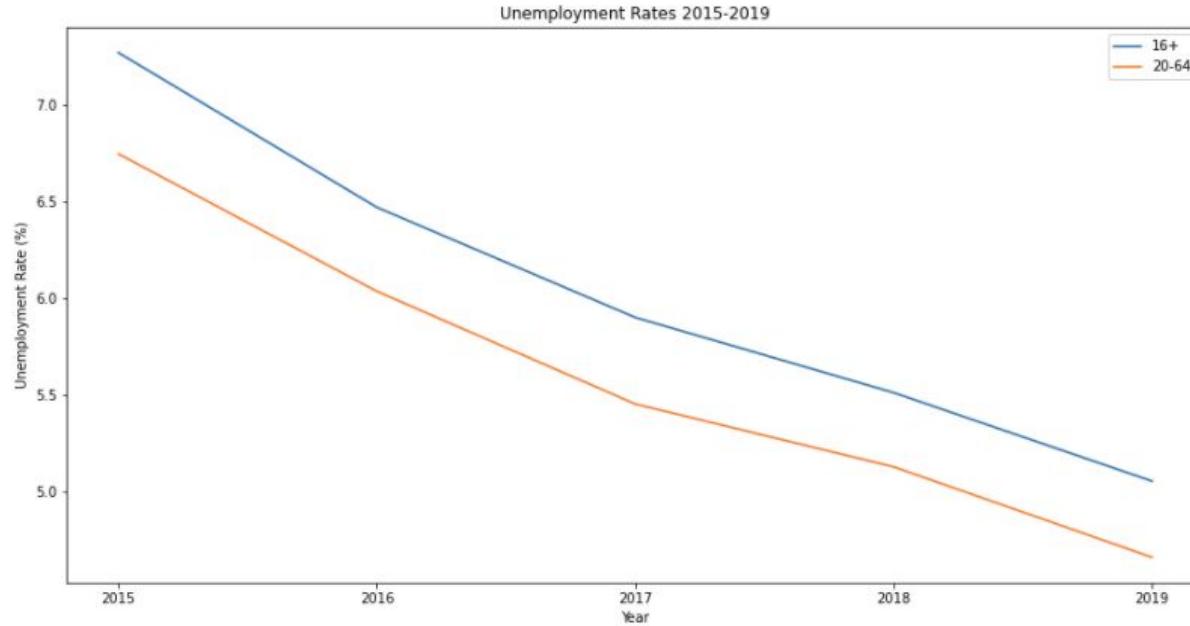




# Analysis Process



## California Unemployment Rates



*Across the state of California, unemployment for both age groups has been falling consistently since 2015.*

*Do enrollment changes correlate with unemployment rate?*

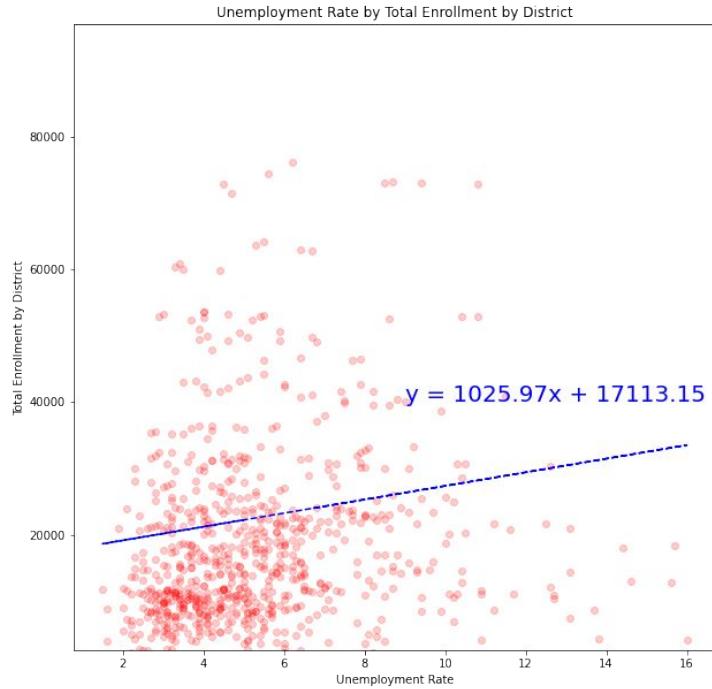




# Analysis Process



## Unemployment Rates by Total Enrollment



*Using linear regression, we found the correlation coefficient to be 0.05. There is a minimal positive correlation between unemployment rate and total enrollment by district.*

*In fact, this suggests that as unemployment rate increases, enrollment does as well.*

*How do enrollment percent changes correlate with unemployment rate?*





# Analysis Process

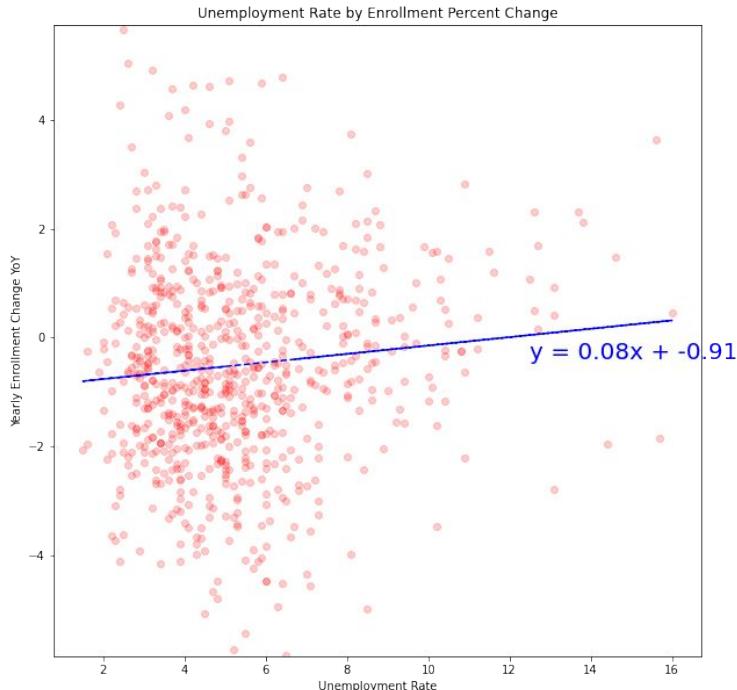


## Unemployment Rates by Enrollment Percent Change

*Using linear regression, we found the correlation coefficient to be 0.07. The correlation is slightly higher than the previous analysis.*

*Once again, suggesting that as unemployment rate increases, the enrollment percent change becomes more positive.*

*What conclusions can we draw from this data?*



# Hypotheses



#1

*Public school enrollment has been trending downward across the state of California even before the beginning of the COVID-19 pandemic.*

#2

*Unemployment rates are heavily correlated with public school enrollment across the state of California.*



Analysis Background



Data Retrieval



Data Exploration



Cleaning Data



Analysis Process



Conclusions



Implementations



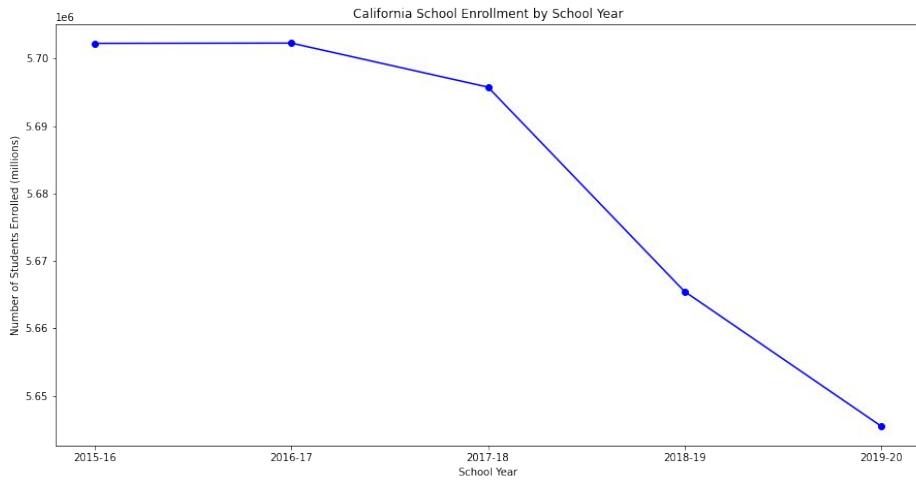
Limitations



# Conclusions

#1

*Public school enrollment has been trending downward across the state of California even before the beginning of the COVID-19 pandemic.*



Based on our analysis, public school enrollment has **decreased** by about 100,000 students across the state of California between 2015 and 2019.

Not all districts are experiencing decreasing enrollment, but four out of the five largest unified districts have had decreasing enrollment since 2015.



Analysis Background



Data Retrieval



Data Exploration



Cleaning Data



Analysis Process



Conclusions



Implementations



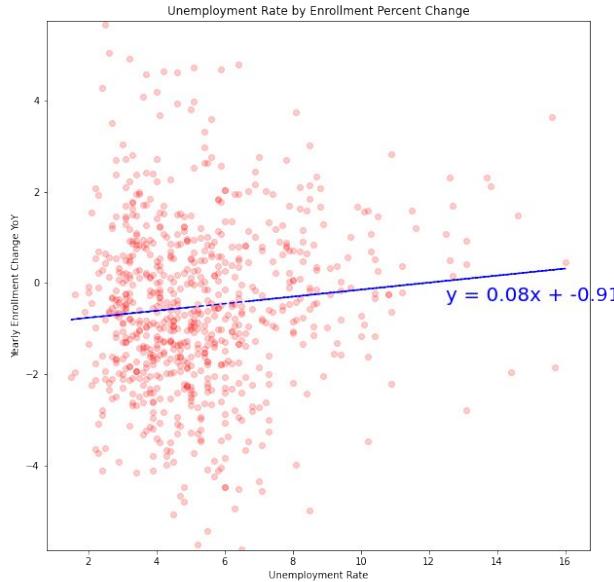
Limitations



# Conclusions

#2

*Unemployment rates are heavily correlated with public school enrollment across the state of California.*



Our statistical analysis revealed that there is a **minimal** correlation of 0.07 between unemployment rate and enrollment percent change year over year.

The correlation is slightly positive, so this means that with an increase in unemployment, enrollment should be increasing or become more positive.



Analysis Background



Data Retrieval



Data Exploration



Cleaning Data



Analysis Process



Conclusions



Implementations



Limitations

# Implementations



## Enrollment

- Decreased in LAUSD
- Decreased statewide
- Not all districts are experiencing decreasing enrollment

## Unemployment

- Unemployment had been decreasing across California before COVID-19

## Both

- There is minimal positive correlation between enrollment percent change and unemployment rates





## Limitations

- *Consumer Price Index data from the U.S. Bureau of Labor Statistics is not available by county, city, or school district, only state, region, and nation*
- *Time to pursue other economic indicators like Housing, Cost of Living, Income, etc...*
- *Time to analyze changes in enrollment and unemployment within ethnic/racial groups*



## Next Steps

- *Analyzing Housing data and statewide Consumer Price Index with enrollment data*
- *Pursuing private versus public school data in California*
- *Considering district birth rate and immigration data*



# Q&A

---



[Analysis Background](#)



[Data Retrieval](#)



[Data Exploration](#)



[Cleaning Data](#)



[Analysis Process](#)



[Conclusions](#)



[Implementations](#)



[Next Steps](#)

# Thank You!

---



[Analysis Background](#)



[Data Retrieval](#)



[Data Exploration](#)



[Cleaning Data](#)



[Analysis Process](#)



[Conclusions](#)



[Implementations](#)



[Next Steps](#)