

```
In [1]: import math
from matplotlib.lines import Line2D
import matplotlib.pyplot as plt
from matplotlib.ticker import PercentFormatter
import numpy as np
import pandas as pd
import re
import seaborn as sns
import statsmodels.api as sm

pd.set_option("display.max_colwidth", None) # Set maximum column width
pd.set_option("display.max_columns", None) # Set maximum number of columns displayed
pd.set_option("display.width", 300) # Set width for entire output
from IPython.display import display, HTML
display(HTML("<style>.container { width:100% !important; }</style>"))

In [2]: ##### Notes
# We process the following datasets:
# * ZIP_Locale_Detail.tsv
# * CAASPP entities
# * CAASPP
# * CA_Schl_Table_EthPct.tsv provides LCFF+ indicator
# * UC Admissions "FR Eth by Yr" https://www.universityofcalifornia.edu/about-us/information-center/admissions-source-school
# * UC Admissions "FR GPA by Yr" https://www.universityofcalifornia.edu/about-us/information-center/admissions-source-school
# We only utilize data about California public schools, including charter schools.

# UC Admissions data is indexed by high school name. Each school has a numeric ID, but the IDs differ from the CAASPP IDs.
# Several high schools in the state have the same name. There are even a few duplicate school names in the same county.
# UC Admissions data also includes City and County but not district.
# CAASPP data is indexed by high school name. CAASPP data also includes County, District, and ZIP Code, but not City. We look up City from ZIP using a USPS table.

campuses = ["Berkeley", "UC Davis", "UCI", "UCLA", "UC Merced", "UCR", "UCSB", "UCSC", "UCSD"]

testID2Name = {
    "1": "English (ELA)",
    "2": "Math"
}

In [3]: ##### Read the ZIP code data to be able to map ZIP code to City.
# The PHYSICAL CITY column indicates the city name. It is blank for a handful of records: Pandas converts these to NaN. PHYSICAL STATE is blank for the same records.
# The LOCALE NAME often indicates the neighborhood name within the city such as "CARMEL MOUNTAIN", "LA JOLLA". It contains other text such as "ANNEX" that we clean.
# Sometimes the LOCALIE indicates a street location rather than a neighborhood.
filename = "./ZIP_Locale_Detail.tsv"
keepColumns = ["DELIVERY ZIPCODE", "LOCALE NAME", "PHYSICAL CITY", "PHYSICAL STATE"]
dfZIPS = pd.read_csv(filename, delimiter="\t", dtype=str, usecols=keepColumns)
zip2CityDict = {}
zip2City = {}
zip2State = {}
for index, row in dfZIPS.iterrows():
    city = row["PHYSICAL CITY"]
    locale = row["LOCALE NAME"]
    zip = row["DELIVERY ZIPCODE"]
    if city != city: # NaN
        print(zip, locale, city, row["PHYSICAL CITY"])
        continue

    # Clean the Locale
    locale = re.sub(r" +", " ", locale)
    locale = re.sub(r"\DHGT\$|", "HEIGHTS", locale)
    locale = re.sub(r"\^(:?AIRPORT|CPU|DOWNTOWN|FEDERAL(?: BUILDING)?|MAIN OFFICE|VPO) ", "", locale)
    locale = re.sub(r"\^([A-Z]{2})(?: .*)? S&PDC$", "", locale) # E.g. "SANTA ANA CA NORTH GRAND SDC"
    locale = re.sub(r"\^(:? \d+)?", "", locale) # E.g. "NORTHRIDGE CARRIERS 24", "CARMEL VALLEY 21 CARRIERS"
    locale = re.sub(r"\^(:? CARRIERS)? (:?ANNEX|ANX|CPLX)(?: .*)? ", "", locale) # E.g. "CARMEL VALLEY CARRIER ANNEX", "MOUNTAIN VIEW CARRIERS ANNEX", "NAPOLEON STREET CARRIER CPLX", "LA JOLLA ANNEX UC"
    locale = re.sub(r"\^(:?POSTAL|RETAIL|STAMP) STORES", "", locale) # E.g. "CARMEL MOUNTAIN POSTAL STORE", "MACYS RETAIL STORE"
    locale = re.sub(r"\^(:?FIN|?ANCE)? (:?STA|STATION|UNIT))?", "", locale) # E.g. "PARK MANOR FINANCE"
    locale = re.sub(r"\^(:?DPOBU|PRS)$", "", locale) # E.g. "BIG BEAR LAKE DPOBU"

    if locale == "" or locale == city:
        pass
    elif re.search(r"\b(?:AIRPORT|AVENUE|BANK|BRANCH|CENTER|CTR|DRUGS|HOSPITAL|HWY|MALL|MARKET|MARKETPLACE|MKT|PLAZA|SERVICE|STA|STATION|STORE|STREET|THE|UNIV|[A-Z]|\\d+|[\\#[0-9A-Z]+]\\b", locale): # E.g. "NAPOLEON STREET CARRIER CPLX"
        pass
    elif re.search(city, locale): # E.g. "WASHINGTON LOS ANGELES"
        pass
    else:
        if zip not in zip2CityDict:
            zip2CityDict[zip] = {}
        zip2CityDict[zip][locale] = 1

    zip2City[zip] = city
    zip2State[zip] = row["PHYSICAL STATE"]

for zip in zip2CityDict:
```

```

sCity = "/".join(zip2CityDict[zip].keys())
zip2City[zip] = "/".join([sCity, zip2City[zip]])
if len(zip2CityDict[zip]) > 1 and zip2State[zip] == "CA":
    print(zip2City[zip])

# Add ZIP codes not available in the USPS file
zip2City["90010"] = "LOS ANGELES"
zip2City["90227"] = "unknown"
zip2City["90747"] = "CARSON"
zip2City["91719"] = "unknown"
zip2City["92093"] = "LA JOLLA"
zip2City["92415"] = "SAN BERNARDINO"
zip2City["93740"] = "FRESNO"
zip2City["93741"] = "FRESNO"
zip2City["95057"] = "unknown"
zip2City["95331"] = "unknown"
zip2City["96311"] = "Fleet Post Office AP"
print(len(zip2City))

```

```

32304 WESTSIDE nan nan
32305 WESTSIDE nan nan
32310 WESTSIDE nan nan
32316 WESTSIDE nan nan
ALAMEDA/ARCADE/LOS ANGELES
BARRINGTON/VILLAGE/LOS ANGELES
LINCOLN/LINCOLN HEIGHTS/LOS ANGELES
EL SERENO/LINCOLN HEIGHTS/LOS ANGELES
BICENTENNIAL/MIRACLE MILE/LOS ANGELES
EAGLE ROCK/HIGHLAND PARK/LOS ANGELES
LA TIJERA/WESTVERN/LOS ANGELES
ARCO/BUNKER HILL/FOY/LOS ANGELES
HUB CITY/WILLOWBROOK/COMPTON
FIESTONE/HOLLYDALE/SOUTH GATE
BELMONT SHORE/LOMA/LONG BEACH
BIXBY/NORTH/LONG BEACH
SPRING/VIKING/LONG BEACH
BRYANT/SPRING/LONG BEACH
NORTHRIDGE/RANCH/PORTER RANCH
ROMOLAND/SUN CITY/MENIFEE
QUAIL VALLEY/SUN CITY/MENIFEE
PLAYA/SOUTH LAGUNA/LAGUNA BEACH
BOLSA/GOLDENWEST/WESTMINSTER
BRISTOL/SOUTH MAIN/SANTA ANA
DOYLE COLONY/POPLAR/PORTERVILLE
PINEDALE/WOODWARD PARK/FRESNO
BARTON/CEDAR/FRESNO
CHESTNUT/LINDEN/SOUTH SAN FRANCISCO
GEARY/MENDELL/SAN FRANCISCO
BAYVIEW/STANTON/BURLINGAME
MARINA/PRESIDIO/SAN FRANCISCO
TOWNSEND/ISLAND COVE/SAN FRANCISCO
MCCLAREN/VISITACION/SAN FRANCISCO
MARINA/MARINA GREEN/SAN FRANCISCO
CAMBRIDGE/PALO ALTO/EAST PALO ALTO
MISSION SAN JOSE/WARM SPRINGS/FREMONT
DIMOND/WEST GRAND/OAKLAND
GRAND LAKE/WEST GRAND/OAKLAND
HILLTOP/MIRA VISTA/RICHMOND
BEAR VALLEY/CAMP CONNELL/ARNOLD
BROADWAY/FORT SUTTER/SACRAMENTO
CAMELLIA/FORT SUTTER/SACRAMENTO
37995

```

```

In [4]: ##### dfEntities
# Read the CAASPP entities data. This mostly gives us the mapping of county code to county name.
# Read the file. "latin1" (ISO-8859-1) is needed due to Spanish characters in the names of 3 schools.
filename = "CAASPP_sb_c2024entities_csv.txt"
allEntitiesColumns = ["County Code", "District Code", "School Code", "Type ID", "Filler", "Test Year",
                      "County Name", "District Name", "School Name", "Zip Code"]
keepEntitiesColumns = ["County Code", "District Code", "School Code", "Type ID",
                      "County Name", "District Name", "School Name", "Zip Code"]
dfEntities = pd.read_csv(filename, delimiter="^", dtype=str, encoding="latin1", usecols=keepEntitiesColumns)
dfEntities["School Name"] = dfEntities["School Name"].str.strip() # Trim whitespace

entitiesCounties = dfEntities["County Name"].value_counts() # Count of county names in dataset

code2County = {}
for index, row in dfEntities.iterrows():
    code = row["County Code"]
    name = row["County Name"]
    if code in code2County:
        if code2County[code] != name:
            print("Error")
    else:
        code2County[code] = name

```

```

code2City = {}
code2ZIP = {}
for index, row in dfEntities.iterrows():
    code = row["School Code"]
    zip = row["Zip Code"]
    if code == "0000000":
        continue
    if code in code2ZIP:
        if code2ZIP[code] != zip:
            print(code, code2ZIP[code], zip)
    else:
        code2ZIP[code] = zip
        if zip in zip2City:
            code2City[code] = zip2City[zip]
        else:
            print("Unknown ZIP Code:", zip)

```

```

In [5]: ##### dfTests 2024
# Read the CAASPP test score data. This includes number of students at the school and test results.
# Fields:
# "Total Students Enrolled" may be "", in which case "Total Students Tested" = 0, "Total Students Tested with Scores" = 0, and Percentage values are NaN.
# Dataframes:
# df: Whole file. Temporary.
# dfTests: Filter out records for grade != 11, etc.
# dfRight = dfTests[keepColumns], then add additional columns.
filename = "CAASPP_sb_ca2024_all_csv_v1.txt"
keepColumns = ["County Code", "District Code", "District Name", "School Code", "School Name", "Type ID", "Test Year", "Test Type",
              "Test ID", "Student Group ID", "Grade",
              "Total Students Enrolled", "Total Students Tested", "Total Students Tested with Scores",
              "Mean Scale Score",
              "Percentage Standard Exceeded", "Count Standard Exceeded",
              "Percentage Standard Met", "Count Standard Met",
              "Percentage Standard Met and Above", "Count Standard Met and Above",
              "Percentage Standard Nearly Met", "Count Standard Nearly Met",
              "Percentage Standard Not Met", "Count Standard Not Met",
              "Overall Total"]

# Read the file. "latin1" (ISO-8859-1) is needed due to Spanish characters in the names of 3 schools.
df = pd.read_csv(filename, delimiter="\t", dtype=str, encoding="latin1", usecols=keepColumns)
df["School Name"] = df["School Name"].str.strip() # Trim whitespace. School Name in the CAASPP and CAASPP Entities data is truncated to 50 characters.

dfTests = df[(df["School Code"] != "0000000") & (df["Grade"] == "11") & (df["Student Group ID"] == "1")] # Keep only high schools. Use "&" operator for Series of booleans.
# df = df.astype({"Total Students Enrolled": "int", "Total Students Tested": "int", "Total Students Tested with Scores": "int"}) # This generates an error due to "Total Students Enrolled" = ""
df = pd.DataFrame() # Free the memory of the big dataframe
print(dfTests.shape[0])
print(dfTests)

```

County Code District Code with Scores Mean Scale Score \			District Name School Code			School Name Type ID Test Year Test Type Test ID Student Group ID Grade Total Students Enrolled Total Students Tested Total Students Tested											
2206	01	10017	Envision Academy for Arts & Technology	0112607	Envision Academy for Arts & Technology	9	2024	B	1	1	11		50		48		
47	2517.4																
2373	01	10017	Envision Academy for Arts & Technology	0112607	Envision Academy for Arts & Technology	9	2024	B	2	1	11		50		48		
48	2472.5																
4099	01	10017	Alameda County Office of Education	0130401	Alameda County Juvenile Hall/Court	7	2024	B	1	1	11		13		5		
5	*																
4212	01	10017	Alameda County Office of Education	0130401	Alameda County Juvenile Hall/Court	7	2024	B	2	1	11		13		6		
6	*																
4329	01	10017	Alameda County Office of Education	0130419	Alameda County Community	7	2024	B	1	1	11		15		10		
9	*																
...
4050261	58	72769	Wheatland Union High	0123570	Wheatland Community Day High	7	2024	B	2	1	11		*		*		
*	*																
4050294	58	72769	Wheatland Union High	0133751	Edward P. Duplex	7	2024	B	1	1	11		22		21		
21	2469.6																
4050358	58	72769	Wheatland Union High	0133751	Edward P. Duplex	7	2024	B	2	1	11		22		21		
21	2435.6																
4050422	58	72769	Wheatland Union High	5838305	Wheatland Union High	7	2024	B	1	1	11		240		237		
237	2582.1																
4050524	58	72769	Wheatland Union High	5838305	Wheatland Union High	7	2024	B	2	1	11		240		237		
237	2545.9																
Percentage Standard Exceeded Count Standard Met Above Count Standard Met and Above Count Standard Met and Above Percentage Standard Nearly Met Count Standard Nearly Met Percentage Standard Not Met			Percentage Standard Exceeded Count Standard Met Above Count Standard Met and Above Count Standard Met and Above Percentage Standard Nearly Met Count Standard Nearly Met Percentage Standard Not Met			Percentage Standard Exceeded Count Standard Met Above Count Standard Met and Above Count Standard Met and Above Percentage Standard Nearly Met Count Standard Nearly Met Percentage Standard Not Met											
Count	Standard	Not Met	Overall Total	6.38	3	21.28	10	27.66	13	31.91	15		40.43				
2206																	
19	47			0.00	0	12.50	6	12.50	6	14.58	7		72.92				
2373																	
35	48																
4099	*	*		*	*	*	*	*	*	*	*		*		*	*	*
*	*			*	*	*	*	*	*	*	*		*		*	*	*
4212	*	*		*	*	*	*	*	*	*	*		*		*	*	*
*	*			*	*	*	*	*	*	*	*		*		*	*	*
4329	*	*		*	*	*	*	*	*	*	*		*		*	*	*
*	*																
...
4050261	*	*		*	*	*	*	*	*	*	*		*		*	*	*
*	*			*	*	*	*	*	*	*	*		*		*	*	*
4050294	0.00	0		9.52	2	9.52	2	9.52	2	28.57	6		61.90				
13	21			0.00	0	0.00	0	0.00	0	0.00	0		0.00		0	100.00	
4050358																	
21	21																
4050422	20.25			48	33.76	80	54.01	128	24.89	59		21.10					
50	237			5.91	14	17.30	41	23.21	55	26.16	62		50.63				
120	237																

[5428 rows x 26 columns]

```
In [6]: #### dfTests2025
filename = "CAASPP_sb_ca2025_all_csv_v1.txt"
keepColumns = ["County Code", "District Code", "District Name", "School Code", "School Name", "Type ID", "Test Year", "Test Type",
              "Test ID", "Student Group ID", "Grade",
              "Total Students Enrolled", "Total Students Tested", "Total Students Tested with Scores",
              "Mean Scale Score",
              "Percentage Standard Exceeded", "Count Standard Exceeded",
              "Percentage Standard Met", "Count Standard Met",
              "Percentage Standard Met and Above", "Count Standard Met and Above",
              "Percentage Standard Nearly Met", "Count Standard Nearly Met",
              "Percentage Standard Not Met", "Count Standard Not Met",
              "Overall Total"]
```

```
# Read the file. "Latin1" (ISO-8859-1) is needed due to Spanish characters in the names of 3 schools.
df = pd.read_csv(filename, delimiter="\t", dtype=str, encoding="latin1", usecols=keepColumns)
df["School Name"] = df["School Name"].str.strip() # Trim whitespace. School Name in the CAASPP and CAASPP Entities data is truncated to 50 characters.
```

```
dfTests2025 = df[(df["School Code"] != "0000000") & (df["Grade"] == "11") & (df["Student Group ID"] == "1")] # Keep only high schools
df = pd.DataFrame() # Free the memory of the big dataframe
print(dfTests2025.shape[0])
```

5422

```
In [7]: # CAASPP school names that require special treatment
# * Parentheses are automatically treated as a first name unless added as an exception here.
# * "ALTERNATIVE" (65 schools) and "CONTINUATION" (241 schools) are automatically removed unless added as an exception here.
testParents = [
    # ALTERNATIVE that occurs in UC Schools
    # "not in CAASPP": "SANTA CRUZ ALTERNATIVE", # UCID = 50446
    "VISTA ALTERNATIVE": "VISTA ALTERNATIVE",

    # ALTERNATIVE and CONTINUATION that should not be stripped to avoid duplicates within the same county.
```

```

# (Duplicates are rare and may be automatically handled via Lookup.psv, so not critical to avoid them here.)
"ENTERPRISE ALTERNATIVE": "ENTERPRISE ALTERNATIVE",
"Live Oak Alternative": "Live Oak Alternative",
"mENDOCINO ALTERNATIVE": "mENDOCINO ALTERNATIVE",
"pOMONA ALTERNATIVE": "pOMONA ALTERNATIVE",
"ROUND VALLEY CONTINUATION": "ROUND VALLEY CONTINUATION",
"SIERRA VISTA HIGH (ALTERNATIVE)": "SIERRA VISTA ALTERNATIVE HIGH",
"TEMPLE CITY ALTERNATIVE": "TEMPLE CITY ALTERNATIVE",

# Parentheses
"ATASCADERO CHOICES IN EDUCATION ACADEMY (ACE)": "ATASCADERO CHOICES IN EDUCATION ACADEMY (ACE)",
"BURBANK UNIFIED INDEPENDENT LEARNING ACADEMY (BUIL)": "BURBANK UNIFIED INDEPENDENT LEARNING ACADEMY (BUILA)",
"EPIC CHARTER (EXCELLENCE PERFORMANCE INNOVATION CI)": "EXCELLENCE PERFORMANCE INNOVATION CITIZENSHIP (EPIC) CHARTER",
"Five Keys Charter (SF Sheriff's)": "Five Keys Charter (SF Sheriff's)",
"Five Keys Independence HS (SF Sheriff's)": "Five Keys Independence HS (SF Sheriff's)",
"Harold McAlister High (Opportunity)": "Harold McAlister High (Opportunity)",
"Humanities and Arts (Harts) Academy of Los Angeles": "Humanities and Arts (Harts) Academy of Los Angeles",
"Juvenile Hall (Endeavor/Voyager Secondary)": "Endeavor/Voyager Secondary (Juvenile Hall)",
"Keith B. Bright High (Juvenile Hall)": "Keith B Bright High (Juvenile Hall)",
"Lakeport Alternative (Home School)": "Lakeport Alternative (Home School)",
"Pacific Valley (K-12)": "Pacific Valley K-12",
"S.F. County Opportunity (Hilltop)": "San Francisco County Opportunity - Hilltop",
"SOAR High (Students on Academic Rise)": "Students on Academic Rise (SOAR) High",
"Walnutwood High (Independent Study)": "Walnutwood High (Independent Study)",
"William J. (Pete) Knight High": "William J (Pete) Knight",
"Yosemite School Community Day (7-12)": "Yosemite School Community Day 7-12",
}

```

```
In [8]: # Abbreviations are common in UC names, rare in CAASPP names.
# The below abbreviations are unambiguous, so we can expand and then match vs trying to match raw and expanded names.
```

```

abbreviations = {
    "ACD": "ACADEMY", # Not used in CAASPP
    "ACAD": "ACADEMY", # Could also mean "ACADEMIES". Not used in CAASPP.
    "ACDM": "ACADEMY", # Not used in CAASPP
    "CHAR": "CHARTER", # Not used in CAASPP
    "CHARTR": "CHARTER",
    "CHTR": "CHARTER", # Not used in CAASPP
    "COL": "COLLEG", # Could be "COLLEGE" or "COLLEGIATE". Not used in CAASPP
    "CLOG": "COLLEG", # Could be "COLLEGE" or "COLLEGIATE". Not used in CAASPP
    "COLL": "COLLEG", # Could be "COLLEGE" or "COLLEGIATE".
    # "CTR": "?", # Could be "CENTER" or "CHARTER"
    "CTRL": "CENTRAL",
    "HS": "HIGH SCHOOL", # Used 3x in CAASPP
    "INST": "INSTITUTE", # Not used in CAASPP
    "INTERNATIO": "INTERNATIONAL",
    "JR": "JUNIOR",
    "JR.": "JUNIOR", # Only CAASPP
    "JUNIOR-SENIOR": "JUNIOR SENIOR",
    "JUNIOR/SENIOR": "JUNIOR SENIOR",
    "JR/SR": "JUNIOR SENIOR",
    "JR./SR.": "JUNIOR SENIOR",
    "LDRSHIP": "LEADERSHIP",
    "LRN": "LEARNING",
    "LRNG": "LEARNING", # Not used in CAASPP
    "L.A.": "LOS ANGELES", # Only CAASPP
    "MATHEMA": "MATH",
    "MATHEMATICS": "MATH",
    "MT": "MOUNT",
    "MT.": "MOUNT",
    "ONE.CHARTER": "ONE CHARTER", # Only CAASPP
    "PREP": "PREPARATORY", # Common use vs strictly abbreviation, in CAASPP.
    "SC": "SCHOOL", # Not used in CAASPP
    "SCH": "SCHOOL", # Used once in CAASPP
    "SCHL": "SCHOOL", # Not used in CAASPP
    "SCI": "SCIENCE", # Not used in CAASPP
    "SF": "SAN FRANCISCO",
    "SIERRA@HOME": "SIERRA AT HOME", # Only CAASPP
    "SR": "SENIOR",
    "SR.": "SENIOR", # Only CAASPP
    "S.F.": "SAN FRANCISCO", # Only CAASPP
    # "TECH": "TECHNOLOGY", # Could also mean "TECHNICAL", in CAASPP
    "TECHLGY": "TECHNOLOGY",
    "W.E.B.": "W E B", # Only CAASPP
    "-SENIOR": "SENIOR", # Only CAASPP
    "&": "AND",
    "&": "-", # Should be "AT"
    "@MCATEER": "- MCATEER",
}

```

```
In [9]: ##### Formatting
# Hyphens -- Most often, a hyphen surrounded by spaces indicates that a location follows, whereas
# a hyphen not surrounded by whitespace is a hyphenated name, two locations, K-12, etc. This is true in CAASPP and UC data.
# But the use of hyphens and whitespace around hyphens are not consistent. E.g., in CAASPP data:
```

```

# ALLIANCE GERTZ-RESSLER RICHARD MERKIN 6-12 COMPLEX/Los Angeles
# ALLIANCE LEICHTMAN-LEVINE FAMILY FOUNDATION ENVIRO/Los Angeles
# ALLIANCE MARINE - INNOVATION AND TECHNOLOGY 6-12 C/Los Angeles
# ALLIANCE OUCHI-O'DONOVAN 6-12 COMPLEX/Los Angeles
# OPTIONS FOR YOUTH - DUARTE/Los Angeles
# OPTIONS FOR YOUTH-ACTION/Los Angeles
# CHAMPS - CHARTER HIGH OF ARTS-MULTIMEDIA AND PERFORMIN
# CALIFORNIA PACIFIC CHARTER - SAN DIEGO/San Diego
# CALIFORNIA PACIFIC CHARTER- LOS ANGELES/Los Angeles
# In UC data:
# ALLIANCE GERTZ-RESSLER HIGH SCHOOL/Los Angeles
# ALLIANCE LEICHTMAN-LEVINE ESHS/Los Angeles
# ALLIANCE MIT 6-12 COMPLEX/Los Angeles
# ALLIANCE OUCHI-O'DONOVAN 6-12/Los Angeles
# CONNECTING WATERS - EAST BAY/Stanislaus
# CONNECTING WATERS-CTRL VALLEY/Stanislaus
# HIGH TECH HIGH - NORTH COUNTY/San Diego
# HIGH TECH HIGH LA/Los Angeles

```

```

In [10]: # testSchoolDetails
testSchoolDetails = {} # Key is ID/SCHOOL/County/CITY/District/ZIP/ExpandedSchool. Value is number of records with this key, typically 2 for Math + ELA.
for index, row in dfTests.iterrows():
    school = row["School Name"].upper()

    # Generate stdSchool, which is tokenized and used to compute expandedSchool.
    if school in testParents:
        stdSchool = testParents[school]
    else:
        stdSchool = re.sub(r"\s+", "INC$", "", school)
        stdSchool = re.sub(r"\s+", "", stdSchool) # Remove commas
        stdSchool = re.sub(r"([A-Z]): ", r"\1 - ", stdSchool) # Remove colon (CAASPP-specific)
        stdSchool = re.sub(r"(:ALTERNATIVE|CONTINUATION|\\(ALTERNATIVE\\)|\\(CONTINUATION\\))$", "", stdSchool) # Remove trailing ALTERNATIVE/CONTINUATION (CAASPP-specific)
        stdSchool = re.sub(r"\^(.*)\\((.*)\\)", r"\2 \1", stdSchool) # Deal with text in parentheses, e.g. "LINCOLN (ABRAHAM) HIGH" -> "ABRAHAM LINCOLN HIGH"

    # Expand abbreviations
    tokens = stdSchool.split()
    expandedTokens = []
    expanded = False
    for token in tokens:
        if token in abbreviations:
            expandedTokens.append(abbreviations[token])
            expanded = True
        else:
            noPunctToken = re.sub(r"\.$", "", token) # Remove periods from abbreviations: not used in UC data
            expandedTokens.append(noPunctToken)
    expandedSchool = " ".join(expandedTokens)

    key = "|".join([row["School Code"], school, code2County[row["County Code"]], code2City[row["School Code"]], row["District Name"], code2ZIP[row["School Code"]], expandedSchool])
    if row["Total Students Enrolled"] == "*":
        nStudents = 0
    else:
        nStudents = int(row["Total Students Enrolled"])
    if key in testSchoolDetails:
        testSchoolDetails[key][0] += 1
        testSchoolDetails[key][1] += nStudents
    else:
        testSchoolDetails[key] = [1, nStudents]

print(len(testSchoolDetails))

```

2715

```

In [11]: # Create testSchools{} and testStdSchools{} structures that will be used for matching. If testID == "multiple", the match will be skipped.
testID2Details = {}
testRawSchoolID2Details = {} # Key is school (not expanded and no county)
testSchools = {} # Key is expandedSchool/county
testSchool2Details = {} # Key is expandedSchool/county
testStdSchools = {} # Key is stdSchool/county: drop trailing "SENIOR HIGH SCHOOL" etc. Value is CAASPP school ID or "multiple" if it is not a unique key.
testStdSchool2Details = {} # Key is stdSchool/county. Value is \n-jointed set of testSchoolDetails keys (id, school, county, city, district, zip, expandedSchool).
for key in testSchoolDetails:
    (id, school, county, city, district, zip, expandedSchool) = key.split("|")

    if id in testID2Details:
        print("Duplicate ID")
    else:
        testID2Details[id] = key

    if school in testRawSchool2Details:
        testRawSchool2Details[school] = "multiple"
    else:
        testRawSchool2Details[school] = key

    schoolCounty = "|".join([expandedSchool, county])
    if schoolCounty in testSchools:

```

```

    testSchools[schoolCounty] = "multiple"
else:
    testSchools[schoolCounty] = id
    testSchool2Details[schoolCounty] = key

# Further standardization, starting with expandedSchool
stdSchool = re.sub(r" SCHOOL$", "", expandedSchool)
stdSchool = re.sub(r" (:UNION|(:JUNIOR|SENIOR )*(?:HIGH|SECONDARY)$", "", stdSchool)
stdSchool = re.sub(r"(CHARTER|HIGH|MIDDLE) SCHOOL", r"\1", stdSchool)
stdSchool = re.sub(r" CENTER|CENTRAL|CHARTER", "CTR", stdSchool)
stdSchool = re.sub(r" COLLEGE|COLLEGiate", "COLLEG", stdSchool)
stdSchool = re.sub(r" TECHNICAL|TECHNOLOGY", "TECH", stdSchool)
stdSchool = re.sub(r" [1-9K]-12\b", "", stdSchool) # Delete "K-12" etc.
stdSchool = re.sub(r" *-* ", " ", stdSchool) # Replace hyphen with a single space
schoolCounty = "|".join([stdSchool, county])
if schoolCounty in testStdSchools:
    testStdSchools[schoolCounty] = "multiple"
    testStdSchool2Details[schoolCounty] = "\n".join([testStdSchool2Details[schoolCounty], key])
else:
    testStdSchools[schoolCounty] = id
    testStdSchool2Details[schoolCounty] = key

print(len(testSchools))
print(len(testStdSchools))

```

2711
2708

In [12]: # Find duplicate stdSchool names. Especially look for duplicate stdSchool names that result from normal and ALTERNATIVE/CONTINUATION schools with the same base name in the same ZIP code.

```

for schoolCounty in sorted(testStdSchools):
    if testStdSchools[schoolCounty] == "multiple":
        print(schoolCounty)
        print(testStdSchool2Details[schoolCounty])
        print()

```

ACADEMY OF INNOVATION|Riverside
0135574|ACADEMY OF INNOVATION|Riverside|NORCO|Corona-Norco Unified|92860|ACADEMY OF INNOVATION
3330503|ACADEMY OF INNOVATION|Riverside|HEMET|Hemet Unified|92543|ACADEMY OF INNOVATION

ALESSANDRO|Riverside
3330065 |ALESSANDRO HIGH|Riverside|HEMET|Hemet Unified|92543|ALESSANDRO HIGH
0105940 |ALESSANDRO|Riverside|MORENO VALLEY|Moreno Valley Unified|92553|ALESSANDRO

JORDAN|Los Angeles
1934470|JORDAN HIGH|Los Angeles|NORTH/LONG BEACH|Long Beach Unified|90805|JORDAN HIGH
1934454|JORDAN HIGH|Los Angeles|GREENMEAD/LOS ANGELES|Los Angeles Unified|90002|JORDAN HIGH

MONTEREY|Los Angeles
1933332|MONTEREY HIGH (CONTINUATION)|Los Angeles|BURBANK|Burbank Unified|91506|MONTEREY HIGH
1931989|MONTEREY CONTINUATION|Los Angeles|LOS ANGELES|Los Angeles Unified|90022|MONTEREY

SAN ANTONIO|Los Angeles
1931807|SAN ANTONIO HIGH (CONTINUATION)|Los Angeles|LA VERNE|Claremont Unified|91711|SAN ANTONIO HIGH
1930551|SAN ANTONIO CONTINUATION|Los Angeles|SOTO/HUNTINGTON PARK|Los Angeles Unified|90255|SAN ANTONIO

SAVA SACRAMENTO ACADEMIC AND VOCATIONAL ACADEMY|Sacramento
0137281|SAVA - SACRAMENTO ACADEMIC AND VOCATIONAL ACADEMY|Sacramento|LAGUNA/ELK GROVE|SAVA - Sacramento Academic and Vocational|95758|SAVA - SACRAMENTO ACADEMIC AND VOCATIONAL ACADEMY
0137406|SAVA - SACRAMENTO ACADEMIC AND VOCATIONAL ACADEMY|Sacramento|COLONIAL/SACRAMENTO|SAVA - Sacramento Academic and Vocational|95820|SAVA - SACRAMENTO ACADEMIC AND VOCATIONAL ACADEMY
0114272|SAVA - SACRAMENTO ACADEMIC AND VOCATIONAL ACADEMY|Sacramento|NORTH NATOMAS/SACRAMENTO|SAVA - Sacramento Academic and Vocational|95834|SAVA - SACRAMENTO ACADEMIC AND VOCATIONAL ACADEMY

In [13]: #### Read the UC admission data

```

# The Count field indicates what is being counted: "Adm", "App", or "Enr".
directory = "UCAdmissions"
files = {"FR ENR_data.All.csv": "all",
         "FR ENR_data.Berkeley.csv": "Berkeley",
         "FR ENR_data.Davis.csv": "UC Davis",
         "FR ENR_data.Irvine.csv": "UCI",
         "FR ENR_data.LosAngeles.csv": "UCLA",
         "FR ENR_data.Merced.csv": "UC Merced",
         "FR ENR_data.Riverside.csv": "UCR",
         "FR ENR_data.SanDiego.csv": "UCSD",
         "FR ENR_data.SantaBarbara.csv": "UCSB",
         "FR ENR_data.SantaCruz.csv": "UCSC"}
dfs = []
for filename in files.keys():
    dfOnefile = pd.read_csv("./".join([directory, filename]), delimiter=",")
    dfOnefile[["Campus"]] = files[filename]
    dfs.append(dfOnefile)
dfUC = pd.concat(dfs, ignore_index=True)
dfUC[["City"]] = dfUC[["City"]].str.upper() # Convert City to uppercase
print(dfUC.shape[0])
print(dfUC)

```

79224

	Calculation1	School	City	County/State/ Territory	Count	uad_uc_ethn_7_cat	Pivot	Field	Values	Campus
0	A B MILLER HIGH SCHOOL05944	A B MILLER HIGH SCHOOL	FONTANA	San Bernardino	Enr	Hispanic/ Latinx	24	all		
1	A B MILLER HIGH SCHOOL05944	A B MILLER HIGH SCHOOL	FONTANA	San Bernardino	Adm	Hispanic/ Latinx	56	all		
2	A B MILLER HIGH SCHOOL05944	A B MILLER HIGH SCHOOL	FONTANA	San Bernardino	App	Hispanic/ Latina	66	all		
3	A B MILLER HIGH SCHOOL05944	A B MILLER HIGH SCHOOL	FONTANA	San Bernardino	Adm	African American	7	all		
4	A B MILLER HIGH SCHOOL05944	A B MILLER HIGH SCHOOL	FONTANA	San Bernardino	App	African American	9	all		
...
79219	YUCAIPA SENIOR HIGH SCHOOL053820	YUCAIPA SENIOR HIGH SCHOOL	YUCAIPA	San Bernardino	Adm	Hispanic/ Latinx	6	UCSC		
79220	YUCAIPA SENIOR HIGH SCHOOL053820	YUCAIPA SENIOR HIGH SCHOOL	YUCAIPA	San Bernardino	App	Hispanic/ Latinx	11	UCSC		
79221	YUCAIPA SENIOR HIGH SCHOOL053820	YUCAIPA SENIOR HIGH SCHOOL	YUCAIPA	San Bernardino	Adm	All	16	UCSC		
79222	YUCAIPA SENIOR HIGH SCHOOL053820	YUCAIPA SENIOR HIGH SCHOOL	YUCAIPA	San Bernardino	App	All	23	UCSC		
79223	YUCCA VALLEY HIGH SCHOOL053818	YUCCA VALLEY HIGH SCHOOL	YUCCA VALLEY	San Bernardino	Adm	All	3	UCSC		

[79224 rows x 8 columns]

In [14]: *### Read the UC admission GPA data*

```
directory = "UCAdmissions"
files = {"FR_GPA_by_Year_data.All.csv": "all",
         "FR_GPA_by_Year_data.Berkeley.csv": "Berkeley",
         "FR_GPA_by_Year_data.Davis.csv": "UC Davis",
         "FR_GPA_by_Year_data.Irvine.csv": "UCI",
         "FR_GPA_by_Year_data.LosAngeles.csv": "UCLA",
         "FR_GPA_by_Year_data.Merced.csv": "UC Merced",
         "FR_GPA_by_Year_data.Riverside.csv": "UCR",
         "FR_GPA_by_Year_data.SanDiego.csv": "UCSD",
         "FR_GPA_by_Year_data.SantaBarbara.csv": "UCSB",
         "FR_GPA_by_Year_data.SantaCruz.csv": "UCSC"}

dfs = []
for filename in files.keys():
    dfOneFile = pd.read_csv("./"+filename, delimiter="\t")
    dfOneFile["Campus"] = files[filename]
    dfs.append(dfOneFile)
dfGPA = pd.concat(dfs, ignore_index=True)
dfGPA["City"] = dfGPA["City"].str.upper() # Convert City to uppercase
print(dfGPA.shape[0])
print(dfGPA)
```

12183

	Calculation1	School	City	County/State/Country	App	GPA	Adm	GPA	Enrl	GPA	Campus
0	A B MILLER HIGH SCHOOL05944	A B MILLER HIGH SCHOOL	FONTANA	San Bernardino	3.75	3.83	3.96	all			
1	ABLE CHARTER054651	ABLE CHARTER	STOCKTON	San Joaquin	3.77	3.79	NaN	all			
2	ABRAHAM LINCOLN HIGH SCHOOL051520	ABRAHAM LINCOLN HIGH SCHOOL	LOS ANGELES	Los Angeles	3.69	3.88	3.94	all			
3	ABRAHAM LINCOLN HIGH SCHOOL052910	ABRAHAM LINCOLN HIGH SCHOOL	SAN FRANCISCO	San Francisco	3.82	3.95	4.03	all			
4	ABRAHAM LINCOLN HIGH SCHOOL053075	ABRAHAM LINCOLN HIGH SCHOOL	SAN JOSE	Santa Clara	3.94	4.04	4.16	all			
...
12178	IGNACIO VALLEY HIGH SCHOOL050663	IGNACIO VALLEY HIGH SCHOOL	CONCORD	Contra Costa	3.52	3.87	NaN	UCSC			
12179	YORBA LINDA HIGH SCHOOL054321	YORBA LINDA HIGH SCHOOL	YORBA LINDA	Orange	3.80	4.02	NaN	UCSC			
12180	YOSEMITE HIGH SCHOOL052176	YOSEMITE HIGH SCHOOL	OAKHURST	Madera	3.99	4.07	NaN	UCSC			
12181	YUBA CITY HIGH SCHOOL053815	YUBA CITY HIGH SCHOOL	YUBA CITY	Sutter	3.71	4.04	NaN	UCSC			
12182	YUCAIPA SENIOR HIGH SCHOOL053820	YUCAIPA SENIOR HIGH SCHOOL	YUCAIPA	San Bernardino	3.86	4.02	NaN	UCSC			

[12183 rows x 8 columns]

In [15]: *# Compare ucCounties, entitiesCounties*

```
ucCounties = dfUC[["County/State/ Territory"]].value_counts()
for county in ucCounties.sort_index().index:
    print(county, ucCounties[county], entitiesCounties[county])
```

Alameda 4127 401
Amador 54 17
Butte 277 101
Calaveras 68 28
Colusa 85 25
Contra Costa 2874 306
Del Norte 21 17
El Dorado 332 90
Fresno 1982 382
Glenn 50 34
Humboldt 221 118
Imperial 291 85
Inyo 44 29
Kern 1500 318
Kings 281 82
Lake 74 49
Lassen 16 34
Los Angeles 21068 2290
Madera 276 86
Marin 662 100
Mariposa 22 16
Mendocino 150 74
Merced 671 124
Modoc 14 13
Mono 44 15
Monterey 770 166
Napa 347 46
Nevada 95 45
Orange 6325 677
Placer 1001 158
Plumas 25 16
Riverside 4562 545
Sacramento 2945 401
San Benito 93 41
San Bernardino 3709 618
San Diego 6680 829
San Francisco 1159 127
San Joaquin 1928 263
San Luis Obispo 429 96
San Mateo 1819 210
Santa Barbara 751 150
Santa Clara 4293 460
Santa Cruz 569 99
Shasta 199 119
Sierra 1 7
Siskiyou 57 66
Solano 799 113
Sonoma 861 227
Stanislaus 1040 223
Sutter 227 59
Tehama 35 54
Trinity 4 28
Tulare 820 238
Tuolumne 42 40
Ventura 1853 238
Yolo 458 71
Yuba 124 48

```
In [16]: # Dictionary of closest UC for each county created using https://www.universityofcalifornia.edu/about-us/information-center/k-12-directory
countyCampus = {
    "all": "all",
    "Alameda": "Berkeley",
    "Amador": "UC Davis",
    "Butte": "UC Davis",
    "Calaveras": "UC Merced",
    "Colusa": "UC Davis",
    "Contra Costa": "Berkeley",
    "Del Norte": "UC Davis",
    "El Dorado": "UC Davis",
    "Fresno": "UC Merced",
    "Glenn": "UC Davis",
    "Humboldt": "UC Davis",
    "Imperial": "UCSD",
    "Inyo": "UC Merced",
    "Kern": "UCSB",
    "Kings": "UC Merced",
    "Lake": "UC Davis",
    "Lassen": "UC Davis",
    "Los Angeles": "UCLA",
    "Madera": "UC Merced",
    "Marin": "Berkeley",
    "Mariposa": "UC Merced",
    "Mendocino": "UC Davis",
    "Merced": "UC Merced",
```

```

"Modoc": "UC Davis",
"Mono": "UC Merced",
"Monterey": "UCSC",
"Napa": "Berkeley",
"Nevada": "UC Davis",
"Orange": "UCI",
"Placer": "UC Davis",
"Plumas": "UC Davis",
"Riverside": "UCR",
"Sacramento": "UC Davis",
"San Benito": "UCSC",
"San Bernardino": "UCR",
"San Diego": "UCSD",
"San Francisco": "Berkeley",
"San Joaquin": "UC Davis",
"San Luis Obispo": "UCSB",
"San Mateo": "Berkeley",
"Santa Barbara": "UCSB",
"Santa Clara": "UCSC",
"Santa Cruz": "UCSC",
"Shasta": "UC Davis",
"Sierra": "UC Davis",
"Siskiyou": "UC Davis",
"Solano": "UC Davis",
"Sonoma": "Berkeley",
"Stanislaus": "UC Merced",
"Sutter": "UC Davis",
"Tehama": "UC Davis",
"Trinity": "UC Davis",
"TuLare": "UC Merced",
"Tuolumne": "UC Merced",
"Ventura": "UCSB", # Would be UCLA by max count of high schools
"Yolo": "UC Davis",
"Yuba": "UC Davis",
}

```

```
In [17]: # Understand "All" in the UC data: it seems to be the count of unique students as opposed to the sum across campuses
for index, row in dfUC.iterrows():
    if (row["School"] == "CANYON CREST ACADEMY" or row["School"] == "GOMPERS PREPARATORY ACADEMY") and row["uad_uc_ethn_7_cat"] == "All" and (row["Count"] == "App" or row["Count"] == "Adm"):
        print("".join(map(str, row.values())))

```

```

CANYON CREST ACADEMY50893|CANYON CREST ACADEMY|SAN DIEGO|San Diego|Adm|All|332|all
CANYON CREST ACADEMY50893|CANYON CREST ACADEMY|SAN DIEGO|San Diego|App|All|438|all
GOMPERS PREPARATORY ACADEMY54353|GOMPERS PREPARATORY ACADEMY|SAN DIEGO|San Diego|Adm|All|88|all
GOMPERS PREPARATORY ACADEMY54353|GOMPERS PREPARATORY ACADEMY|SAN DIEGO|San Diego|App|All|101|all
CANYON CREST ACADEMY50893|CANYON CREST ACADEMY|SAN DIEGO|San Diego|Adm|All|34|Berkeley
CANYON CREST ACADEMY50893|CANYON CREST ACADEMY|SAN DIEGO|San Diego|App|All|327|Berkeley
GOMPERS PREPARATORY ACADEMY54353|GOMPERS PREPARATORY ACADEMY|SAN DIEGO|San Diego|Adm|All|3|Berkeley
GOMPERS PREPARATORY ACADEMY54353|GOMPERS PREPARATORY ACADEMY|SAN DIEGO|San Diego|App|All|22|Berkeley
CANYON CREST ACADEMY50893|CANYON CREST ACADEMY|SAN DIEGO|San Diego|Adm|All|91|UC Davis
CANYON CREST ACADEMY50893|CANYON CREST ACADEMY|SAN DIEGO|San Diego|App|All|346|UC Davis
GOMPERS PREPARATORY ACADEMY54353|GOMPERS PREPARATORY ACADEMY|SAN DIEGO|San Diego|App|All|14|UC Davis
CANYON CREST ACADEMY50893|CANYON CREST ACADEMY|SAN DIEGO|San Diego|Adm|All|104|UCI
CANYON CREST ACADEMY50893|CANYON CREST ACADEMY|SAN DIEGO|San Diego|App|All|352|UCI
GOMPERS PREPARATORY ACADEMY54353|GOMPERS PREPARATORY ACADEMY|SAN DIEGO|San Diego|Adm|All|27|UCI
GOMPERS PREPARATORY ACADEMY54353|GOMPERS PREPARATORY ACADEMY|SAN DIEGO|San Diego|App|All|68|UCI
CANYON CREST ACADEMY50893|CANYON CREST ACADEMY|SAN DIEGO|San Diego|Adm|All|33|UCLA
CANYON CREST ACADEMY50893|CANYON CREST ACADEMY|SAN DIEGO|San Diego|App|All|380|UCLA
GOMPERS PREPARATORY ACADEMY54353|GOMPERS PREPARATORY ACADEMY|SAN DIEGO|San Diego|Adm|All|7|UCLA
GOMPERS PREPARATORY ACADEMY54353|GOMPERS PREPARATORY ACADEMY|SAN DIEGO|San Diego|App|All|67|UCLA
CANYON CREST ACADEMY50893|CANYON CREST ACADEMY|SAN DIEGO|San Diego|Adm|All|75|UC Merced
CANYON CREST ACADEMY50893|CANYON CREST ACADEMY|SAN DIEGO|San Diego|App|All|78|UC Merced
GOMPERS PREPARATORY ACADEMY54353|GOMPERS PREPARATORY ACADEMY|SAN DIEGO|San Diego|Adm|All|10|UC Merced
GOMPERS PREPARATORY ACADEMY54353|GOMPERS PREPARATORY ACADEMY|SAN DIEGO|San Diego|App|All|10|UC Merced
CANYON CREST ACADEMY50893|CANYON CREST ACADEMY|SAN DIEGO|San Diego|Adm|All|198|UCR
CANYON CREST ACADEMY50893|CANYON CREST ACADEMY|SAN DIEGO|San Diego|App|All|225|UCR
GOMPERS PREPARATORY ACADEMY54353|GOMPERS PREPARATORY ACADEMY|SAN DIEGO|San Diego|Adm|All|50|UCR
GOMPERS PREPARATORY ACADEMY54353|GOMPERS PREPARATORY ACADEMY|SAN DIEGO|San Diego|App|All|59|UCR
CANYON CREST ACADEMY50893|CANYON CREST ACADEMY|SAN DIEGO|San Diego|Adm|All|63|UCSD
CANYON CREST ACADEMY50893|CANYON CREST ACADEMY|SAN DIEGO|San Diego|App|All|384|UCSD
GOMPERS PREPARATORY ACADEMY54353|GOMPERS PREPARATORY ACADEMY|SAN DIEGO|San Diego|Adm|All|40|UCSD
GOMPERS PREPARATORY ACADEMY54353|GOMPERS PREPARATORY ACADEMY|SAN DIEGO|San Diego|App|All|99|UCSD
CANYON CREST ACADEMY50893|CANYON CREST ACADEMY|SAN DIEGO|San Diego|Adm|All|127|UCSD
CANYON CREST ACADEMY50893|CANYON CREST ACADEMY|SAN DIEGO|San Diego|App|All|363|UCSB
GOMPERS PREPARATORY ACADEMY54353|GOMPERS PREPARATORY ACADEMY|SAN DIEGO|San Diego|Adm|All|16|UCSB
GOMPERS PREPARATORY ACADEMY54353|GOMPERS PREPARATORY ACADEMY|SAN DIEGO|San Diego|App|All|29|UCSB
CANYON CREST ACADEMY50893|CANYON CREST ACADEMY|SAN DIEGO|San Diego|Adm|All|175|UCSC
CANYON CREST ACADEMY50893|CANYON CREST ACADEMY|SAN DIEGO|San Diego|App|All|275|UCSC
GOMPERS PREPARATORY ACADEMY54353|GOMPERS PREPARATORY ACADEMY|SAN DIEGO|San Diego|Adm|All|23|UCSC
GOMPERS PREPARATORY ACADEMY54353|GOMPERS PREPARATORY ACADEMY|SAN DIEGO|San Diego|App|All|37|UCSC

```

```
In [18]: # ucSchoolDetails
# Key is ID|SCHOOL|County|CITY|ExpandedSCHOOL
# Value is [ALLApp, ALLAdm, ALLEnr, Found, FoundID], where:
```

```

# ALLApp is the number of applications in the ALL category
# Found is a code indicating whether and how we found this school in the CAASPP data with values:
#   0 => not found.
#   1 => found using initial key match.
#   2 => found after standardization.
#   3 => found by ChatGPT.
ucSchoolDetails = {}
count2i = {"App": 0,
           "Adm": 1,
           "Enr": 2}
for index, row in dfUC.iterrows():
    schoolID = row["Calculation1"]
    code = re.sub(r"^\.*?(\d{5})$", r"\1", schoolID)
    if not re.fullmatch(r"5\d{4}", code): # Five digit code, and the first digit is always 5.
        print(schoolID, code)

    # Expand abbreviations
    school = row["School"]
    stdSchool = re.sub(r"([A-Z])&([A-Z])", r"\1 & \2", school) # R&D -> R & D
    tokens = stdSchool.split()
    expandedTokens = []
    expanded = False
    for token in tokens:
        if token in abbreviations:
            expandedTokens.append(abbreviations[token])
            expanded = True
        else:
            expandedTokens.append(token)
    expandedSchool = " ".join(expandedTokens)

    key = "|".join([code, school, row["County/State/ Territory"], row["City"], expandedSchool])
    if key not in ucSchoolDetails:
        ucSchoolDetails[key] = [0, 0, 0, ""]
    if row["uad_ue_ethn_7_cat"] == "All":
        ucSchoolDetails[key][count2i[row["Count"]]] += row["Pivot Field Values"]

print(len(ucSchoolDetails))

```

1420

```

In [19]: ucSchools = {} # Key is SCHOOL/County
ucSchool2stdSchool = {}
ucStdSchools = {} # Format is same as ucSchools but school name is standardized: drop trailing "SENIOR HIGH SCHOOL" etc.
ucStdSchool2Details = {}
for key in ucSchoolDetails:
    (id, school, county, city, expandedSchool) = key.split("|")

    schoolCounty = "|".join([expandedSchool, county])
    if schoolCounty in ucSchools:
        ucSchools[schoolCounty] += 1
    else:
        ucSchools[schoolCounty] = 1

    # Deal with standardization
    stdSchool = re.sub("SCHOOL$", "", expandedSchool)
    stdSchool = re.sub("(?:(JUNIOR )?SENIOR )*(?:HIGH|SECONDARY)$", "", stdSchool)
    stdSchool = re.sub("(CHARTER|HIGH|MIDDLE SCHOOL", r"\1", stdSchool)
    stdSchool = re.sub("CENTER|CENTRAL|CHARTER", "CTR", stdSchool)
    stdSchool = re.sub("COLLEGE|COLLEGiate", "COLLEG", stdSchool)
    stdSchool = re.sub("TECHNICAL|TECHNOLOGY", "TECH", stdSchool)
    stdSchool = re.sub("[1-9K]-12\b", "", stdSchool) # Delete "K-12" etc.
    stdSchool = re.sub("\-* ", " ", stdSchool) # Replace hyphen with a single space
#    stdSchool3 = re.sub(r"INST OF.*\/", " INSTITUTE/", stdSchool)
    ucSchool2stdSchool[expandedSchool] = stdSchool
    schoolCounty = "|".join([stdSchool, county])
    if schoolCounty in ucStdSchools:
        ucStdSchools[schoolCounty] += 1
        ucStdSchool2Details[schoolCounty] = "\n".join([ucStdSchool2Details[schoolCounty], key])
    else:
        ucStdSchools[schoolCounty] = 1
        ucStdSchool2Details[schoolCounty] = key

print(len(ucSchools))
print(len(ucStdSchools))

```

1418
1417

```

In [20]: # Find duplicate school names
for schoolCounty in sorted(ucStdSchools):
    if ucStdSchools[schoolCounty] > 1:
        print(ucStdSchool2Details[schoolCounty])

```

```

50412|SWEETWATER SECONDARY SCHOOL|San Diego|CHULA VISTA|SWEETWATER SECONDARY SCHOOL
52105|SWEETWATER UNION HIGH SCHOOL|San Diego|NATIONAL CITY|SWEETWATER UNION HIGH SCHOOL
54539|WONDERFUL COLLEGE PREP ACADEMY|Kern|DELANO|WONDERFUL COLLEGE PREPARATORY ACADEMY
55584|WONDERFUL COLLEGE PREP ACADEMY|Kern|LOST HILLS|WONDERFUL COLLEGE PREPARATORY ACADEMY
51505|WOODROW WILSON HIGH SCHOOL|Los Angeles|LONG BEACH|WOODROW WILSON HIGH SCHOOL
51840|WOODROW WILSON HIGH SCHOOL|Los Angeles|LOS ANGELES|WOODROW WILSON HIGH SCHOOL

In [21]: # Try to find each UC school in the set of CAASPP schools, recording which ones we found.
# We will go through ucSchoolDetails. If we find a matching CAASPP school:
# * Store its ID in the value in ucSchoolDetails, else set the value to "".
# * Record that this CAASPP school has been "used".
# Duplicate UC schools and duplicate CAASPP matching schools are both skipped and will be handled through ChatGPT / Lookup.psv
nFound = 0
matchedTestSchools = {} # CAASPP schools that matched a UC school. Key is ??
tokensNotMatched = {}
usedTestIDs = {}
for key in ucSchoolDetails:
    (id, school, county, city, expandedSchool) = key.split("|")
    schoolCounty = "|".join([expandedSchool, county])
    stdSchool = ucSchool2StdSchool[expandedSchool]
    stdSchoolCounty = "|".join([stdSchool, county])

    # Try various types of heuristic matching & record the results. Then we'll try ChatGPT on the ones that do not unambiguously match.
    # Heuristic matches are based on school name and county only: straight match, match with "HIGH SCHOOL" (etc.) truncated from both.
    if ucSchools[stdSchoolCounty] > 1:
        continue # Two schools in UC data with same standardized name, same county
    if schoolCounty in testSchools:
        testID = testStdSchools[stdSchoolCounty] # The use of testStdSchools rather than testSchools is intentional: more conservative rejection of testID == "multiple".
        if testID == "multiple":
            continue
        if testID in usedTestIDs:
            print("Previously used 1:", testID) # Never happens
            continue
        ucSchoolDetails[key][3] = 1
        ucSchoolDetails[key][4] = testID
        usedTestIDs[testID] = "|".join([key, "1"])
        nFound += 1
    elif stdSchoolCounty in testStdSchools:
        testID = testStdSchools[stdSchoolCounty]
        if testID == "multiple":
            continue
        if testID in usedTestIDs:
            print("Previously used 2:", testID) # Never happens
            continue
        ucSchoolDetails[key][3] = 2
        ucSchoolDetails[key][4] = testID
        usedTestIDs[testID] = "|".join([key, "2"])
        nFound += 1
    else:
        for token in stdSchool.split():
            if token in tokensNotMatched:
                tokensNotMatched[token] += 1
            else:
                tokensNotMatched[token] = 1
print(nFound)

```

1132

```

In [22]: # Explore
query = r"^\^ENTERPRISE|^\^MONTEREY|^\^ROSELAND|^\^SAN ANTONIO|^\^SWEETWATER"
print("\nUC Details")
for key in ucSchoolDetails:
    (id, school, county, city, expandedSchool) = key.split("|")
    if re.search(query, school):
        print(key, ucSchoolDetails[key])

print("\nCAASPP")
for schoolCounty in sorted(testSchools):
    if re.search(query, schoolCounty):
        key = testSchoolDetails[schoolCounty]
        print(key, testSchoolDetails[key])

```

```

UC Details
52580|ENTERPRISE HIGH SCHOOL|Shasta|REDDING|ENTERPRISE HIGH SCHOOL [114, 52, 22, 2, '4532750']
52050|MONTEREY HIGH SCHOOL|Monterey|MONTEREY|MONTEREY HIGH SCHOOL [566, 241, 56, 2, '2732808']
50902|MONTEREY TRAIL HIGH SCHOOL|Sacramento|ELK GROVE|MONTEREY TRAIL HIGH SCHOOL [828, 331, 111, 2, '0105916']
54850|ROSELAND COLLEGIATE PREP|Sonoma|SANTA ROSA|ROSELAND COLLEGIATE PREPARATORY [111, 49, 13, 0, '']
54044|ROSELAND UNIVERSITY PREP|Sonoma|SANTA ROSA|ROSELAND UNIVERSITY PREPARATORY [163, 68, 16, 0, '']
50412|SWEETWATER SECONDARY SCHOOL|San Diego|CHULA VISTA|SWEETWATER SECONDARY SCHOOL [8, 6, 6, 0, '']
52105|SWEETWATER UNION HIGH SCHOOL|San Diego|NATIONAL CITY|SWEETWATER UNION HIGH SCHOOL [647, 266, 60, 0, '']

```

```

CAASPP
1030469|ENTERPRISE ALTERNATIVE|Fresno|CLOVIS|Clovis Unified|93611|ENTERPRISE ALTERNATIVE [2, 50]
1033422|ENTERPRISE HIGH|Fresno|KERMAN|Kerman Unified|93630|ENTERPRISE HIGH [2, 50]
4532750|ENTERPRISE HIGH|Shasta|REDDING|Shasta Union High|96082|ENTERPRISE HIGH [2, 492]
2730232|MONTEREY COUNTY HOME CHARTER|Monterey|JOHN STEINBECK|SALINAS|Monterey County Office of Education|93901|MONTEREY COUNTY HOME CHARTER [2, 88]
1933332|MONTEREY HIGH (CONTINUATION)|Los Angeles|BURBANK|Burbank Unified|91506|MONTEREY HIGH [2, 87]
2732808|MONTEREY HIGH|Monterey|MONTEREY|Monterey Peninsula Unified|93940|MONTEREY HIGH [2, 680]
2766092|MONTEREY PENINSULA UNIFIED DISTRICT LEVEL PROGRAM|Monterey|MONTEREY|Monterey Peninsula Unified|93940|MONTEREY PENINSULA UNIFIED DISTRICT LEVEL PROGRAM [2, 0]
0112896|MONTEREY PENINSULA UNIFIED SCHOOL DISTRICT COMMUNI|Monterey|SEASIDE|Monterey Peninsula Unified|93955|MONTEREY PENINSULA UNIFIED SCHOOL DISTRICT COMMUNI [1, 0]
0105916|MONTEREY TRAIL HIGH|Sacramento|ELK GROVE|Elk Grove Unified|95624|MONTEREY TRAIL HIGH [2, 1136]
1931989|MONTEREY CONTINUATION|Los Angeles|Los Angeles|Los Angeles Unified|98082|MONTEREY [2, 0]
0101923|ROSELAND CHARTER|Sonoma|ROSELAND|SANTA ROSA|Roseland Charter|95407|ROSELAND CHARTER [2, 350]
1931807|SAN ANTONIO HIGH (CONTINUATION)|Los Angeles|LA VERNE|Claremont Unified|91711|SAN ANTONIO HIGH [2, 74]
4935862|SAN ANTONIO HIGH (CONTINUATION)|Sonoma|CASA GRANDE/PETALUMA|Petaluma Joint Union High|94952|SAN ANTONIO HIGH [2, 54]
1930551|SAN ANTONIO CONTINUATION|Los Angeles|SOTO/HUNTINGTON PARK|Los Angeles Unified|98255|SAN ANTONIO [2, 80]
3738226|SWEETWATER HIGH|San Diego|NATIONAL CITY|Sweetwater Union High|91950|SWEETWATER HIGH [2, 1279]
3768411|SWEETWATER UNION HIGH DISTRICT LEVEL PROGRAM|San Diego|CHULA VISTA|Sweetwater Union High|91911|SWEETWATER UNION HIGH DISTRICT LEVEL PROGRAM [2, 622]

```

```

In [23]: # Output Test Schools to CSV files (one per county) to upload to ChatGPT
files = {}
for county in ucCounties.sort_index().index:
    countyNoSpace = re.sub(r" ", "", county)
    filename = "".join(["latestOutput/referenceSchools", countyNoSpace, ".csv"])
    files[county] = open(filename, "w")

for key in testSchoolDetails: # Key is ID/SCHOOL/County/CITY/District/ZIP/ExpandedSCHOOL
    (id, school, county, city, district, zip, expandedSchool) = key.split("|")
    if county not in files:
        continue
    file = files[county]
    file.write("|\n".join([school, city, district, zip, id]) + "\n")
for county in files:
    files[county].close()

```

```

In [24]: # Output all Test Schools to a single CSV file to upload to ChatGPT to match UC schools not found on the first pass
with open("latestOutput/referenceSchoolsAll.csv", "w") as file:
    for key in testSchoolDetails: # Key is ID/SCHOOL/County/CITY/District/ZIP/ExpandedSCHOOL
        (id, school, county, city, district, zip, expandedSchool) = key.split("|")
        file.write("|\n".join([school, city, district, zip, id]) + "\n")

```

```

In [25]: # Output UC schools that don't have a match yet to enter in Copilot Bing or ChatGPT UI
for targetCounty in ucCounties.sort_index().index:
    print("\n" + targetCounty)
    for key in ucSchoolDetails:
        (id, school, county, city, expandedSchool) = key.split("|")
        if county == targetCounty and ucSchoolDetails[key][4] == "":
            print("|\n".join([school, city]))

```

Alameda
ASPIRE GOLDEN STATE PREP ACADEMY|OAKLAND
ASPIRE LIONEL WILSON COLG PREP|OAKLAND
BAY AREA TECHNOLOGY-BAYTECH|OAKLAND
ENVISION ACAD ARTS/TECHNOLOGY|OAKLAND
IMPACT ACAD ARTS/TECHNOLOGY|HAYWARD
LATITUDE 37.8 HS MERRITT COLLEGE|OAKLAND
LEADERSHIP PUBLIC SCH HAYWARD|HAYWARD
LODESTAR LIGHTHOUSE COMM CHARTER SC|OAKLAND
OAKLAND MILITARY INST COL PREP|OAKLAND
SOJOURNER TRUTH ACADEMY|OAKLAND

Amador

Butte
INSPIRE SCHOOL OF ARTS AND SCI|CHICO

Calaveras

Colusa

Contra Costa
ASPIRE RICHMOND CA COLG PREP|RICHMOND
CLAYTON VALLEY HIGH SCHOOL|CONCORD
CONTRA COSTA SCH PRFRMING ARTS|WALNUT CREEK
LEADERSHIP PUBLIC SCH RICHMOND|RICHMOND

Del Norte

DEL NORTE COUNTY HIGH SCHOOL|CRESCENT CITY

El Dorado

CHARTER COMMUNITY SCHOOL|PLACERVILLE
JOHN ADAMS ACADEMY - EL DORADO|EL DORADO HILLS
OCEAN GROVE CHARTER SCHOOL|PLACERVILLE
SKY MOUNTAIN CHARTER SCHOOL|PLACERVILLE

Fresno

CLOVIS ONLINE SCHOOL|CLOVIS
DUNCAN POLYTECHNICAL HS|FRESNO
FARBER SCH OF IND STUDIES & ONLINE|FRESNO
FLOYD B BUCHANAN HIGH SCHOOL|CLOVIS
PHILLIP J PATINO SCH ENTREPREN|FRESNO
THEODORE ROOSEVELT HIGH SCHOOL|FRESNO

Glenn

Humboldt
NORTHCOAST PREP & PERF ARTS A|ARCATA

Imperial

Inyo

MCFARLAND HIGH SCHOOL|MCFARLAND
ROSAMOND HIGH SCHOOL|ROSAMOND
SHERMAN E BURROUGHS HS|RIDGECREST
WONDERFUL COLLEGE PREP ACADEMY|DELANO
WONDERFUL COLLEGE PREP ACADEMY|LOST HILLS

Kings

Lake

Lassen

Los Angeles
ACADEMIC LEADERSHIP COMMUNITY|LOS ANGELES
ACADEMIES EDUCATION/EMPOWERMT|CARSON
ACADEMIES OF ANTELOPE VALLEY|LANCASTER
ACADEMY OF SCIENTIFIC EXPLORAT|SAN FERNANDO
ALAIN LEROY LOCKE COLLEGE PREP|LOS ANGELES
ALLIANCE CINDY/BILL SIMON TECH|LOS ANGELES
ALLIANCE COLLING FAMILY COLRDY|HUNTINGTON PARK
ALLIANCE GERTZ-RESSLER HIGH SC|LOS ANGELES
ALLIANCE JUDY BURTON TECH ACAD|LOS ANGELES
ALLIANCE LEICHTMAN-LEVINE ESHS|LOS ANGELES
ALLIANCE M&B STERN MATH SCI SCH|LOS ANGELES
ALLIANCE MARGARET M BLOOMFIELD|HUNTINGTON PARK
ALLIANCE MIT 6-12 COMPLEX|SUN VALLEY
ALLIANCE OUCHI-O'DONOVAN 6-12|LOS ANGELES
ALLIANCE PATTI & PETER ACADEMY|LOS ANGELES
ALLIANCE PIERA BARBAGLIA HSA|LOS ANGELES
ALLIANCE RENEE & MEYER LUSKIN|LOS ANGELES

ALLIANCE SMIDT TECHNOLOGY HIGH|LOS ANGELES
AMBASSADOR SCHL GLOBAL LDRSHP|LOS ANGELES
ANAHUACALMECAC INTRL UNIV PREP|LOS ANGELES
ANIMO CTY CHAMPIONS CHARTER HS|INGLEWOOD
ANIMO JACKIE ROBINSON CHR SCH|LOS ANGELES
ANIMO LEADERSHIP CHARTER HS|INGLEWOOD
ANIMO PAT BROWN CHARTER HS|LOS ANGELES
ANIMO S LOS ANGELES CHARTER HS|LOS ANGELES
ANIMO WATTS COLLEGE PREP ACDMY|LOS ANGELES
ARTS & ENTERTAINMENT VIRTUAL ACAD|CARSON
ASPIRE OLLIN UNIV PREP ACDMY|HUNTINGTON PARK
BARACK OBAMA GLOBAL PREP ACADEMY|LOS ANGELES
BIRMINGHAM COMMUNITY HS|LAKE BALBOA
BRIQ COLLEGE PREP|LOS ANGELES
CA SCH ARTS SAN GABRIEL VALLEY|DUARTE
CALIFORNIA ACAD MATH & SCIENCE|CARSON
CAMINO NUEVO HIGH SCHOOL DALZE|LOS ANGELES
CATCH PREP CHARTER HIGH|LOS ANGELES
CESAR CHAVEZ LA ARTES MAGNET|SAN FERNANDO
CHARTER HS ARTS MEDIA PERFORM|VAN NUYS
CHATSWORTH HIGH SCHOOL|CHATSWORTH
CITY HONORS HIGH SCHOOL|INGLEWOOD
CLARK MAGNET HIGH SCHOOL|LA CRESCENTA
CLEVELAND HIGH SCHOOL|RESEDA
COLLEGE BRIDGE ACADEMY|LOS ANGELES
COLLEGiate CHARTER HIGH SCHOOL|LOS ANGELES
CRENSHAW HIGH SCHOOL|LOS ANGELES
CRITICAL DESIGN AND GAMING SCH|LOS ANGELES
DANIEL PEARL MAGNET HIGH SCHL|VAN NUYS
DAV STARR JORDAN HIGH SCHOOL|LONG BEACH
DAVID STARR JORDAN HS|LOS ANGELES
DIEGO RIVERA COMM/TECH SCHOOL|LOS ANGELES
DISCOVERY CHARTER PREP SCHOOL|SYLMAR
DOWNTOWN MAGNETS HIGH SCHOOL|LOS ANGELES
DR RICHARD VLADOVIC HARBOR ACA|WILMINGTON
EARLY COLLEGE ACADEMY|LOS ANGELES
EAST COLLEGE PREP|LOS ANGELES
EAST LA PERFORMING ARTS MAGNET|LOS ANGELES
EAST LA RENAISSANCE ACADEMY|LOS ANGELES
EDWARD R ROYBAL LEARNING CNTR|LOS ANGELES
ENGInIERING & TECH ACAD AT THS|LOS ANGELES
ENVIRONMENTAL CHARTER HS|LAWNDALE
ESPERANZA COLLEGE PREP|LOS ANGELES
FELICITAS GONZALO MENDEZ HS|LOS ANGELES
FREMONT ACD ENGINEERING/DESIGN|POMONA
GAHR HIGH SCHOOL|CERRITOS
GIRLS ACADEMIC LEADERSHIP ACAD|LOS ANGELES
GREEN DESIGN STEAM ACADEMY|LOS ANGELES
HIGH TECH HIGH LA|VAN NUYS
HILDA L SOLIS LEARNING ACDMY|LOS ANGELES
HUMANITAS ACADEMY ART AND TECH|LOS ANGELES
HUMANITIES AND ARTS ACADEMY-LA|HARBOR CITY
HUNTINGTON PARK INST APLD MED|HUNTINGTON PARK
ILEAD EXPLORATION HYBRID|ACTON
ILEAD ONLINE CHARTER SCHOOL|ACTON
INTERNATIONAL STUDIES LEARNING|SOUTH GATE
JOHN BURROUGHS HIGH SCHOOL|BURBANK
JOHN GLENN HIGH SCHOOL|NORWALK
JOHN MARSHALL FUNDAMENTAL SCH|PASADENA
JUAN RODRIGUEZ CABRILLO HS|LONG BEACH
KING DREW MAGNET HS MED/SCI|LOS ANGELES
LA CENTER ENRICHED STUDIES|LOS ANGELES
LAKE BALBOA COLLEGE PREP K-12|VAN NUYS
LEADERSHIP&PUBLIC SVC VIRTUAL ACA|LOS ANGELES
LEGACY HIGH SCHOOL VAPA|SOUTH GATE
LENNOX MATH SCIENCE TECH ACAD|INGLEWOOD
LIBRA ACADEMY AT LINDA MARQUEZ HS|HUNTINGTON PARK
LINDA MARQUEZ HS-SOC JUSTICE|HUNTINGTON PARK
LONG BEACH POLYTECHNIC HS|LONG BEACH
LOS ANGELES ACAD ARTS ENTRPRSE|LOS ANGELES
LOS ANGELES SCH GLOBAL STUDIES|LOS ANGELES
MATH SCIENCE TECHNOLOGY MAGNET|LOS ANGELES
MAYWOOD CTR ENRICHED STUDIES|MAYWOOD
MIGUEL CONTRERAS LRNG COMPLEX|LOS ANGELES
MT SAC EARLY COLG ACD W COVINA|WEST COVINA
NARBONNE HIGH SCHOOL|HARBOR CITY
OPPORTUNITIES LRN-WILLIAM HART|VALENCIA
ORTHOPAEDIC HOSP MED MAGNET HS|LOS ANGELES
PERFORMING ARTS COMMUNITY SCH|LOS ANGELES
PUBLIC SERVICE COMMUNITY SCHOOL|LOS ANGELES
PUC COMMUNITY CHARTER EARLY COLL HS|SYLMAR
PUC EARLY COLG ACADEMY LEADERS|LOS ANGELES
PUC TRIUMPH CHARTER ACADEMY HS|SYLMAR
RAMON C CORTINES SCH OF VISUAL|LOS ANGELES

RENAISSANCE HIGH SCHOOL ARTS|LONG BEACH
RESEDA HIGH SCHOOL|RESEDA
ROBERT A MILLIKAN HIGH SCHOOL|LONG BEACH
SANTA CLARITA VALLEY INTERNATI|CASTAIC
SATO ACADEMY OF MATH & SCIENCE|LONG BEACH
SCH SOCIAL JUSTICE-CONT LEARN CTR|LOS ANGELES
SCHOOL OF EXTENDED EDU OPTIONS|POMONA
SCHOOL VISUAL ARTS HUMANITIES|LOS ANGELES
SHERMAN OAKS CTR ENRCHED STDs|TARZANA
SOAR HIGH SCHOOL|LANCASTER
SOCIAL JUSTICE HUMANITAS ACAD|SAN FERNANDO
SOCIAL JUSTICE LDRSHP @ TORRES|LOS ANGELES
SONIA SOTOMAYOR ARTS SCI ACAD|LOS ANGELES
SOUTH LA COLLEGE PREP|LOS ANGELES
STEAM LEGACY HIGH SCHOOL CMPLX|SOUTH GATE
STEAM VIRTUAL ACADEMY|LOS ANGELES
STEM ACADEMY OF BOYLE HEIGHTS|LOS ANGELES
STEM ACADEMY OF HOLLYWOOD|HOLLYWOOD
SYLMAR BIOTECH HLTH & ENG MGMT|SYLMAR
SYLMAR HIGH SCHOOL|SYLMAR
TECHNOLOGY PREPARATORY ACADEMY|SAN FERNANDO
TRACY HIGH SCHOOL|CERRITOS
UNIVERSITY HIGH SCHOOL|LOS ANGELES
UNIVERSITY PATHWAYS PUB SVC AC|LOS ANGELES
UP MEDICAL MAGNET ACADEMY|LOS ANGELES
USC HYBRID HIGH COLLEGE PREP|LOS ANGELES
USC MEDIA & ENGINEERING MAGNET SCH|LOS ANGELES
VALLEY ACADEMY ARTS & SCIENCES|GRANADA HILLS
VALLEY OAKS CTR ENRICHED STDs|SUN VALLEY
VAUGHN NEXT CENTURY LEARN CTR|SAN FERNANDO
VIEW PARK PREP ACCEL CHARTER HS|LOS ANGELES
VILLAGE ACADEMY HIGH SCHOOL|POMONA
WARREN HIGH SCHOOL|DOWNEY
WESTBROOK ACADEMY|SOUTH GATE
WESTCHESTER ENRICHED SCIENCE MAGNET|LOS ANGELES
WILLIAM HOWARD TAFT HS|WOODLAND HILLS
WILLIAM J PETE KNIGHT HIGH SCH|PALMDALE
WOODROW WILSON HIGH SCHOOL|LONG BEACH
WOODROW WILSON HIGH SCHOOL|LOS ANGELES
WORKMAN HIGH SCHOOL|CITY OF INDUSTRY

Madera

Marin

Mariposa

Mendocino

Merced

Modoc

Mono

Monterey

KING CITY JOINT UNION HS|KING CITY

Napa

Nevada

SIERRA ACADEMY EXPEDITION LRNG|NEVADA CITY

Orange

ADVANCED LRNG ACDMY EARLY COLG|SANTA ANA
CA CONNECTIONS ACADEMY CAPISTRANO|SAN JUAN CAPISTRANO
CALIFORNIA PACIFIC CHARTER LA|NEWPORT BEACH
EPIC CHARTER SCHOOL - ANAHEIM|ANAHEIM
HECTOR GODINEZ FUNDAMENTAL HS|SANTA ANA
LEGACY COLLEGE PREP|SANTA ANA
ORANGE COUNTY HS OF THE ARTS|SANTA ANA
SAN JOAQUIN HIGH SCHOOL|IRVINE

Placer

JOHN ADAMS ACADEMY|ROSEVILLE

Plumas

Riverside

ACADEMY OF INNOVATION|HEMET
ELITE ACADEMIC ACADEMY-LUCERNE|TEMECULA
HYBRID SCHOOL OF INNOVATION|NORCO
JOHN F KENNEDY MIDDLE COLG HS|NORCO
MARTIN LUTHER KING HIGH SCHOOL|RIVERSIDE

METHOD SCHOOLS|MURRIETA
RIVERSIDE POLY HIGH SCHOOL|RIVERSIDE

Sacramento
ANTELOPE HIGH SCHOOL|ANTELOPE
ASPIRE ALEXANDER TWILIGHT SCND|SACRAMENTO
ENCINA HIGH SCHOOL|SACRAMENTO
HEALTH PROFESSIONS HIGH SCH|SACRAMENTO
MET SACRAMENTO HIGH SCHOOL|SACRAMENTO
NATOMAS CHARTER SCHOOL PERFORM|SACRAMENTO
REX & MARGARET FORTUNE ECHS|ELK GROVE
SACRAMENTO HIGH SCHOOL|SACRAMENTO
SCH ENGINEERING & SCIENCES|SACRAMENTO
UMOJA INTERNATIONAL ACADEMY|SACRAMENTO

San Benito
SAN BENITO COUNTY EVENING HIGH|HOLLISTER

San Bernardino
A B MILLER HIGH SCHOOL|FONTANA
ACADEMY OF CAREERS AND EXPLORA|HELENDALE
COBALT INSTITUTE MATH/SCIENCE|VICTORVILLE
ENCORE JR/SR HIGH SCHOOL|HESPERIA
EXCELSIOR CHARTER SCHOOLS|VICTORVILLE
GORMAN LEARNING CENTER|REDLANDS
GRAND TERRACE HIGH SCHOOL|GRAND TERRACE
OPTIONS FOR YOUTH - ACTON|FONTANA
OPTIONS FOR YOUTH - SAN BERNAR|SAN BERNARDINO
SAGE OAK CHARTER SCHOOL KEPPEL|REDLANDS

San Diego
AUDEO CHARTER SCHOOL|SAN DIEGO
BAYFRONT CHARTER HIGH SCHOOL|CHULA VISTA
CHARTER SCHOOL OF SAN DIEGO|SAN DIEGO
CHULA VISTA LRNA CMNTY CHARTER|CHULA VISTA
COASTAL ACADEMY HS OCEANSIDE|OCEANSIDE
DEL LAGO ACADEMY|ESCONDIDO
GUAJOME PARK ACADEMY|VISTA
HEALTH SCIENCES HIGH/MID COLL|SAN DIEGO
HERBERT HOOVER HIGH SCHOOL|SAN DIEGO
JAMES MADISON HIGH SCHOOL|SAN DIEGO
JULIAN CHARTER SCHOOL LIVE|JULIAN
KEARNY SCH BIOMED SCIENCE TECH|SAN DIEGO
LEARNING CHOICE - CHULA VISTA|CHULA VISTA
LEARNING CHOICE ACADEMY-SAN DIEGO|SAN DIEGO
LITERACY FIRST CHARTER SCHOOLS|ALPINE
OFARRELL CHARTER SCHOOL|SAN DIEGO
PATRICK HENRY HIGH SCHOOL|SAN DIEGO
SAN DIEGO CITY SCHOOLS|SAN DIEGO
SAN DIEGO MET|SAN DIEGO
SAN DIEGO SCHL CREATE/PERFORM|SAN DIEGO
SAN DIEGUITO ACADEMY|ENCINITAS
SCHL ENTREPRENEURSHIP & TECH|SAN DIEGO
SWEETWATER SECONDARY SCHOOL|CHULA VISTA
SWEETWATER UNION HIGH SCHOOL|NATIONAL CITY

San Francisco
ACADEMY-SAN FRAN @ MCATEER|SAN FRANCISCO
GALILEO ACAD OF SCIENCE & TECH|SAN FRANCISCO
PHILLIP & SALA BURTON ACAD HS|SAN FRANCISCO
RUTH ASAWA SCHOOL OF THE ARTS|SAN FRANCISCO
THURGOOD MARSHALL ACADEMIC HS|SAN FRANCISCO

San Joaquin
AMOS ALONZO STAGG HIGH SCHOOL|STOCKTON
ASPIRE BENJAMIN HOLT COLG PREP|STOCKTON
CALIFORNIA CNCTS MONTEREY BAY|RIPON
CALIFORNIA CONNECTIONS ACD RPN|RIPON
MILLENNIUM HIGH SCHOOL|TRACY
TRACY JOINT UNION HIGH SCHOOL|TRACY
VISTA OAKS CHARTER SCHOOL|LODI
WEBER INST OF APPLIED SCI TECH|STOCKTON

San Luis Obispo

San Mateo
SUMMIT PREPARATORY CHARTER HIG|REDWOOD CITY
SUMMIT PUBLIC SCHOOLS SHASTA|DALY CITY

Santa Barbara
SANTA BARBARA SENIOR HIGH SCHO|SANTA BARBARA

Santa Clara
ADRIAN C WILCOX HIGH SCHOOL|SANTA CLARA

DOWNTOWN COLG PREP EL PRIMERO|SAN JOSE
DOWNTOWN COLLEGE PREP ALUM ROC|SAN JOSE
DR TJ OWENS GILROY EARLY COLG|GILROY
HENRY T GUNDERSON HIGH SCHOOL|SAN JOSE
ROBERTO CRUZ LEADERSHIP ACDMY|SAN JOSE
UNIVERSITY PREP ACADEMY|SAN JOSE
W C OVERFELT HIGH SCHOOL|SAN JOSE

Santa Cruz
PACIFIC COLLEGiate SCHOOL|SANTA CRUZ
SANTA CRUZ ALTERNATIVE|SANTA CRUZ

Shasta
WEST VALLEY HIGH SCHOOL|COTTONWOOD

Sierra

Siskiyou
TULELAKE HIGH SCHOOL|TULELAKE

Solano
BUCKINGHAM COLLEGiate CHARTER|VACAVILLE
RIO VISTA HIGH SCHOOL|RIO VISTA

Sonoma
ROSELAND COLLEGiate PREP|SANTA ROSA
ROSELAND UNIVERSITY PREP|SANTA ROSA

Stanislaus
CONNECTING WATERS - EAST BAY|WATERFORD
CONNECTING WATERS-CTRL VALLEY|MODESTO
JOHANSEN HIGH SCHOOL|MODESTO

Sutter

Tehama

Trinity

Tulare
ELEANOR ROOSEVELT COMM LRNG CN|VISALIA
SUMMIT CHARTER COLLEGiate ACAD|PORTERVILLE
UNIVERSITY PREPATORY HIGH SCH|VISALIA
VISALIA TECHNICAL EDU CENTER|VISALIA

Tuolumne

Ventura
ARCHITECTURE CONSTRUCT ENG HS|CAMARILLO
CALIFORNIA VIRTUAL ACAD SAN DIEGO|SIMI VALLEY
CALIFORNIA VIRTUAL ACAD SAN MATEO|SIMI VALLEY
CALIFORNIA VIRTUAL ACAD SUTTER|SIMI VALLEY
CALIFORNIA VIRTUAL ACADEMY LA|SIMI VALLEY
CALIFORNIA VIRTUAL ACDM SONOMA|SIMI VALLEY
CALIFORNIA VIRTUAL-SAN JOAQUIN|SIMI VALLEY
COMPASS CHARTER SCHLS OF LA|THOUSAND OAKS
COMPASS CHARTER SCHOOL OF YOLO|THOUSAND OAKS
HIGH SCHOOL AT MOORPARK COLL|MOORPARK

Yolo
DAVIS SCHOOL INDEPENDENT STUDY|DAVIS
DELTA HIGH SCHOOL|CLARKSBURG

Yuba
MARYSVILLE CHARTER ACAD ARTS|MARYSVILLE

Here is the prompt: The attached file contains one row per reference high school. Each row consists of pipe-delimited values schoolName|cityName|districtName|ZIP|ID. Below is a second list of high schools. Each line consists of schoolName|cityName. The names in the two lists do not match exactly due to abbreviations and other variations. For each line below, print the input line followed by "!" as a delimiter followed by the line for the best match school from the reference file. Include all elements from the reference file including districtName, ZIP, and ID. If Copilot Bing indicates "Message exceeds 10240 characters", shorten the list in the query.

In [26]: # Below, we read mapping of UC school name to CAASPP school name from Lookup.psv, which was iteratively created.

```
# Lines with the following formats contain mapping, others are ignored by this program.  
# Left || right: mapping found by ChatGPT 1st pass (Looking only within one county).  
# Left ||| right: mapping found by ChatGPT 2nd pass (Looking across counties).  
# Left -- right: manual delete.  
# Left ++ right: manual add.  
# Left = "/" delimited (ucSchool, ucCity). This is the ucSchoolDetails key.  
# right = "/" delimited (testSchool, testCity, testDistrict, testZIP, testID)  
  
# Error checking includes:  
# * Is the format of the line correct?  
# * Does the testID exist?  
# * Is the testSchool generated by ChatGPT associated with that testID?  
# * Is uckey a duplicate, meaning that ChatGPT generated the left side of the line incorrectly?  
# * Is testID a duplicate, meaning that ChatGPT matched the same CAASPP school to two different UC schools?
```

```
In [27]: # Create lookup table to get UC key from ChatGPT output
ucSchoolCity2Key = {}
for key in ucSchoolDetails:
    (id, school, county, city, expandedSchool) = key.split("|")
    schoolCity = "|".join([school, city])
    if schoolCity in ucSchoolCity2Key:
        # Check for duplicates: there are none
        print(schoolCity, key, ucSchoolCity2Key[schoolCity])
        continue
    ucSchoolCity2Key[schoolCity] = key

In [30]: # Read ChatGPT output from Lookup.csv and update ucSchoolDetails, which has the following format:
# Key is ID/SCHOOL/County/CITY/ExpandedSCHOOL
# Value is [ALLApp, ALLAdm, ALLEnr, Found, FoundID]
f = open('lookup.csv', 'rt', encoding='utf8')
for line in f:
    if re.match(r"^\#", line):
        continue # Skip records that were commented out
    if not re.search(r"^\|", line): # Only the ChatGPT output lines have this format
        continue
    cleanline = line.rstrip()
    (ucline, testline) = cleanline.split(" || ")
    if re.match(r"No ", testline): # May be "No exact match", "No close match", etc.
        continue
    key = ucSchoolCity2Key[ucline] # Get the ucSchoolDetails key
    items = testline.split("|") # Parse testLine to get testID
    if len(items) != 5:
        nitems = len(items)
        print(f"Bad testLine format {nItems} items\n    UC: {key}\n    Found: {testLine}") # Happens once. Seems Like the output was actually generated by an LLM!
        continue
    (testSchool, testCity, testDistrict, testZIP, testID) = items
    testSchool = testSchool.upper() # ChatGPT generated titlecase ~20 times

    if testID not in testID2Details:
        print(f"Bad testLine ID\n    UC: {key}\n    Found: {testLine}")
        continue
    (idTRef, schoolTRef, countyTRef, cityTRef, districtTRef, zipTRef, expandedSchoolTRef) = testID2Details[testID].split("|")
    if testSchool != schoolTRef:
        actualLine = "|".join([schoolTRef, cityTRef, districtTRef, zipTRef, idTRef])
        if testSchool in testRawSchool2Details:
            if testRawSchool2Details[testSchool] == "multiple":
                sNameLine = "Multiple CAASPP schools with this found school name"
            else:
                (idSName, schoolSName, countySName, citySName, districtSName, zipSName, expandedSchoolSName) = testRawSchool2Details[testSchool].split("|")
                sNameLine = "|".join([schoolSName, citySName, districtSName, zipSName, idSName])
        else:
            sNameLine = "Found school name is hallucinated"
        print(f"School mismatch\n    UC: {key}\n    Found: {testLine}\n    FoundN: {sNameLine}\n    Actual: {actualLine}")
        continue

    if ucSchoolDetails[key][3] > 0:
        print(f"Previously found\n    UC: {key}") # Should not happen because we didn't search using items we previously found.
        continue
    if testID in usedTestIDs:
        print(f"Previously used\n    UC: {key}\n    PrevUC: {usedTestIDs[testID]}\n    Found: {testLine}")
        continue

    # Found a new item
    ucSchoolDetails[key][3] = 3
    ucSchoolDetails[key][4] = testID
    usedTestIDs[testID] = "|".join([key, "3"])
f.close()
```

Bad testLine format 4 items

UC: 53914|ACADEMY OF INNOVATION|Riverside|HEMET|ACADEMY OF INNOVATION
 Found: Academy of Innovation|Hemet Unified|92543|3330503

Previously used

UC: 51443|SAGE OAK CHARTER SCHOOL KEPPEL|San Bernardino|REDLANDS|SAGE OAK CHARTER SCHOOL KEPPEL
 PrevUC: 51664|SAGE OAK CHARTER|San Bernardino|REDLANDS|SAGE OAK CHARTER|1
 Found: SAGE OAK CHARTER|LUGONIA|REDLANDS|Sage Oak Charter|92373|0136069

Previously used

UC: 54643|CALIFORNIA CONNECTIONS ACD RPN|San Joaquin|RIPON|CALIFORNIA CONNECTIONS ACADEMY RPN
 PrevUC: 54952|CALIFORNIA CNCTS MONTEREY BAY|San Joaquin|RIPON|CALIFORNIA CNCTS MONTEREY BAY|3
 Found: CALIFORNIA CONNECTIONS ACADEMY NORTHERN CALIFORNIA|STOCKTON|California Connections Academy Northern|95366|0125849

Previously used

UC: 54044|ROSELAND UNIVERSITY PREP|Sonoma|SANTA ROSA|ROSELAND UNIVERSITY PREPARATORY
 PrevUC: 54858|ROSELAND COLLEGIATE PREP|Sonoma|SANTA ROSA|ROSELAND COLLEGIATE PREPARATORY|3
 Found: ROSELAND CHARTER|ROSELAND|SANTA ROSA|Roseland Charter|95407|0101923

School mismatch

UC: 54057|LOS ANGELES SCH GLOBAL STUDIES|Los Angeles|LOS ANGELES|LOS ANGELES SCHOOL GLOBAL STUDIES
 Found: SCHOOL OF GLOBAL STUDIES AT CONTRERAS LEARNING COM|FOY/LOS ANGELES|Los Angeles Unified|90017|0112029
 FoundN: Found school name is hallucinated
 Actual: SCHOOL OF BUSINESS AND TOURISM AT CONTRERAS LEARNI|FOY/LOS ANGELES|Los Angeles Unified|90017|0112029

School mismatch

UC: 54053|MIGUEL CONTRERAS LRNG COMPLEX|Los Angeles|LOS ANGELES|MIGUEL CONTRERAS LEARNING COMPLEX
 Found: MIGUEL CONTRERAS LEARNING COMPLEX GREEN DESIGN STEAM|FOY/LOS ANGELES|Los Angeles Unified|90017|0124503
 FoundN: Found school name is hallucinated
 Actual: DIEGO RIVERA LEARNING COMPLEX GREEN DESIGN STEAM A|LOS ANGELES|Los Angeles Unified|90001|0124503

School mismatch

UC: 54542|SOCIAL JUSTICE HUMANITAS ACAD|Los Angeles|SAN FERNANDO|SOCIAL JUSTICE HUMANITAS ACADEMY
 Found: SOCIAL JUSTICE HUMANITAS ACADEMY AT ESTEBAN TORRES|HAZARD/LOS ANGELES|Los Angeles Unified|90063|0122358
 FoundN: Found school name is hallucinated
 Actual: HUMANITAS ACADEMY OF ART AND TECHNOLOGY AT ESTEBAN|HAZARD/LOS ANGELES|Los Angeles Unified|90063|0122358

Previously used

UC: 54444|SOCIAL JUSTICE LDRSHP @ TORRES|Los Angeles|LOS ANGELES|SOCIAL JUSTICE LEADERSHIP - TORRES
 PrevUC: 54695|SCH SOCIAL JUSTICE-CONT LEARN CTR|Los Angeles|LOS ANGELES|SCHOOL SOCIAL JUSTICE-CONT LEARN CTR|3
 Found: SOCIAL JUSTICE LEADERSHIP ACADEMY MAGNET AT ESTEBA|HAZARD/LOS ANGELES|Los Angeles Unified|90063|0122366

School mismatch

UC: 51505|WOODROW WILSON HIGH SCHOOL|Los Angeles|LONG BEACH|WOODROW WILSON HIGH SCHOOL
 Found: WOODROW WILSON SENIOR HIGH|LONG BEACH|Long Beach Unified|90804|1939875
 FoundN: WOODROW WILSON SENIOR HIGH|EL SERENO/LINCOLN HEIGHTS/LOS ANGELES|Los Angeles Unified|90032|1939859
 Actual: WILSON HIGH|LONG BEACH|Long Beach Unified|90804|1939875

Previously used

UC: 52842|SAN DIEGO CITY SCHOOLS|San Diego|SAN DIEGO|SAN DIEGO CITY SCHOOLS
 PrevUC: 53903|SAN DIEGO HIGH SCHOOL|San Diego|SAN DIEGO|SAN DIEGO HIGH SCHOOL|2
 Found: SAN DIEGO HIGH|SAN DIEGO|San Diego Unified|92101|0107219

```
In [31]: # Explore the above issues
query = r"SIERRA VISTA"
desiredZIP = ""
desiredTestID = ""
print("UC")
for key in ucSchoolDetails:
    (ucID, school, county, city, expandedSchool) = key.split("|")
    if re.search(query, school, re.IGNORECASE):
        print(key, ucSchoolDetails[key])

print("\nCAASPP")
for key in testSchoolDetails:
    (testID, school, county, city, district, zip, expandedSchool) = key.split("|")
    if re.search(query, school, re.IGNORECASE) or zip == desiredZIP or testID == desiredTestID:
        if testID in usedTestIDs:
            print(key, testSchoolDetails[key], " *** ", usedTestIDs[testID])
        else:
            print(key, testSchoolDetails[key])

UC
50228|SIERRA VISTA HIGH SCHOOL|Los Angeles|BALDWIN PARK|SIERRA VISTA HIGH SCHOOL [540, 243, 58, 2, '1938166']

CAASPP
1938166|SIERRA VISTA HIGH|Los Angeles|IRWINDALE|BALDWIN PARK|Baldwin Park Unified|91706|SIERRA VISTA HIGH [2, 789] *** 50228|SIERRA VISTA HIGH SCHOOL|Los Angeles|BALDWIN PARK|SIERRA VISTA HIGH SCHOOL|2
1995653|SIERRA VISTA HIGH (ALTERNATIVE)|Los Angeles|WHITTIER|Whittier Union High|90605|SIERRA VISTA ALTERNATIVE HIGH [2, 220]
0130708|SIERRA VISTA CHARTER HIGH|Tulare|Tulare Joint Union High|93274|SIERRA VISTA CHARTER HIGH [2, 126]
5431119|SIERRA VISTA HIGH (CONTINUATION)|Tulare|DINUBA|Dinuba Unified|93618|SIERRA VISTA HIGH [2, 96]
```

```
In [32]: # Read the 2nd pass ChatGPT output (searching across the whole state rather than by county) and update ucSchoolDetails
f = open('lookup.psv', 'rt', encoding='utf8')
for line in f:
    if not re.search(r" ||| ", line): # Only 2nd pass ChatGPT output Lines have this format
        continue
    cleanline = line.rstrip()
    (ucLine, testLine) = cleanline.split(" ||| ")
    if re.match(r"No", testLine): # May be "No exact match", "No close match", etc.
        continue
    key = ucSchoolCity2Key[ucLine] # Get the ucSchoolDetails key
    items = testLine.split("|") # Parse testLine to get testID
    if len(items) != 5:
        nItems = len(items)
```

```

print(f"Bad testLine format {nItems} items\n  UC: {key}\n  Found: {testLine}") # Happens once. Seems like the output was actually generated by an LLM!
continue
(testSchool, testCity, testDistrict, testZIP, testID) = items
testSchool = testSchool.upper() # ChatGPT generated titlecase ~20 times

if testID not in testID2Details:
    print(f"Bad testLine ID\n  UC: {key}\n  Found: {testLine}")
    continue
(idTRef, schoolTRef, countyTRef, cityTRef, districtTRef, zipTRef, expandedSchoolTRef) = testID2Details[testID].split("|")
if testSchool != schoolTRef:
    actualLine = "|".join([schoolTRef, cityTRef, districtTRef, zipTRef, idTRef])
    if testSchool in testRawSchool2Details:
        if testRawSchool2Details[testSchool] == "multiple":
            sNameLine = "Multiple CAASPP schools with this found school name"
        else:
            (idSName, schoolSName, countySName, citySName, districtSName, zipSName, expandedSchoolSName) = testRawSchool2Details[testSchool].split("|")
            sNameLine = "|".join([schoolSName, citySName, districtSName, zipSName, idSName])
    else:
        sNameLine = "Found school name is hallucinated"
    print(f"School mismatch\n  UC: {key}\n  Found: {testLine}\n  FoundN: {sNameLine}\n  Actual: {actualLine}")
    continue

if ucSchoolDetails[key][3] > 0:
    print(f"Previously found\n  UC: {key}") # Should not happen because we didn't search using items we previously found.
    continue
if testID in usedTestIDs:
    print(f"Previously used\n  UC: {key}\n  PrevUC: {usedTestIDs[testID]}\n  Found: {testLine}")

# Found a new item
ucSchoolDetails[key][3] = 4
ucSchoolDetails[key][4] = testID
usedTestIDs[testID] = "|".join([key, "4"])

f.close()

```

School mismatch

```

UC: 51799|ELITE ACADEMIC ACADEMY-LUCERNE|Riverside|TEMECULA|ELITE ACADEMIC ACADEMY-LUCERNE
Found: Elite Academic Academy - Lucerne|TEMECULA|Elite Academic Academy - Lucerne|92590|0136978
FoundN: ELITE ACADEMIC ACADEMY - LUCERNE|TEMECULA|Elite Academic Academy - Lucerne|92590|0136960
Actual: ELITE ACADEMIC ACADEMY - MOUNTAIN EMPIRE|TEMECULA|Elite Academic Academy - Mountain Empire|92590|0136978

School mismatch
UC: 54250|ANTELOPE HIGH SCHOOL|Sacramento|ANTELOPE|ANTELOPE HIGH SCHOOL
Found: Antelope High|ANTELOPE|Roseville Joint Union High|95843|0113118
FoundN: ANTELOPE HIGH|ELKHORN/SACRAMENTO|Roseville Joint Union High|95843|0116459
Actual: DOUGHERTY VALLEY HIGH|SAN RAMON|San Ramon Valley Unified|94582|0113118

School mismatch
UC: 54846|ILEAD EXPLORATION HYBRID|Los Angeles|ACTON|ILEAD EXPLORATION HYBRID
Found: iLEAD Hybrid|ACTON|iLEAD Hybrid|93510|0137315
FoundN: iLEAD HYBRID|ACTON|iLEAD Hybrid|93510|0131987
Actual: KIPP NAVIGATE COLLEGE PREP|FOOTHILL/SAN JOSE|SBE - KIPP Navigate College Prep|95127|0137315

```

In [33]:

```

# Delete incorrect matches
f = open('lookup.psv', 'rt', encoding='utf8')
for line in f:
    if not re.search(r"\-\-", line):
        continue
    cleanline = line.rstrip()
    (ucline, testline) = cleanline.split(" -- ")
    key = ucSchoolCity2Key[ucline] # Get the ucSchoolDetails key
    items = testline.split("|") # Parse testline to get testID
    (testSchool, testCity, testDistrict, testZIP, testID) = items

    if testID not in usedTestIDs:
        print(f"testID not used\n  UC: {key}\n  Delete: {testLine}")
        continue

    # Delete this item from the "matched" records
    usedWithKey = re.sub(r"\|[1-9]\$","", usedTestIDs[testID])
    if usedWithKey != key:
        print(f"Wrong usedWithKey {key} vs {usedWithKey}")
        continue

    ucSchoolDetails[key][3] = 0
    ucSchoolDetails[key][4] = ""
    del usedTestIDs[testID]

f.close()

```

In [34]:

```

# Add fixed/corrected matches
f = open('lookup.psv', 'rt', encoding='utf8')
for line in f:
    if not re.search(r"\+\+\+", line):
        continue
    cleanline = line.rstrip()

```

```

        (ucLine, testLine) = cleanline.split(" + ")
key = ucSchoolCity2Key[ucLine] # Get the ucSchoolDetails key
items = testLine.split("|") # Parse testLine to get testID
(testSchool, testCity, testDistrict, testZIP, testID) = items
testSchool = testSchool.upper() # ChatGPT generated titlecase ~20 times

if testID not in testID2Details:
    print(f"Bad testLine ID\n  UC: {key}\n    Found: {testLine}")
    continue
(idRef, schoolTRef, countyTRef, cityTRef, districtTRef, zipTRef, expandedSchoolTRef) = testID2Details[testID].split("|")
if testSchool != schoolTRef:
    actualLine = "|".join([schoolTRef, cityTRef, districtTRef, zipTRef, idRef])
    if testSchool in testRawSchool2Details:
        if testRawSchool2Details[testSchool] == "multiple":
            sNameLine = "Multiple CAASPP schools with this found school name"
        else:
            (idSName, schoolSName, countySName, citySName, districtSName, zipSName, expandedSchoolSName) = testRawSchool2Details[testSchool].split("|")
            sNameLine = "|".join([schoolSName, citySName, districtSName, zipSName, idSName])
    else:
        sNameLine = "Found school name is hallucinated"
    print(f"School mismatch\n  UC: {key}\n    Found: {testLine}\n    FoundN: {sNameLine}\n    Actual: {actualLine}")
    continue

if ucSchoolDetails[key][3] > 0:
    print(f"Previously found\n  UC: {key}")
    continue
if testID in usedTestIDs:
    print(f"Previously used\n  UC: {key}\n  PrevUC: {usedTestIDs[testID]}\n    Found: {testLine}")
    continue

# Found a new item
ucSchoolDetails[key][3] = 5
ucSchoolDetails[key][4] = testID
usedTestIDs[testID] = "|".join([key, "5"])
f.close()

```

In [37]: # How many UC records did we find vs not find?

```

nCount = 0
yCount = 0
for key in ucSchoolDetails:
    values = ucSchoolDetails[key]
    if values[3] > 0:
        yCount += 1
    else:
        nCount += 1
print(yCount, nCount)

```

1413 7

In [38]: # Which schools from UC records didn't we find?

```

print("Couldn't find the following schools from UC records:")
for key in ucSchoolDetails:
    values = ucSchoolDetails[key]
    if values[3] > 0:
        continue
    print(key, ucSchoolDetails[key])

```

Couldn't find the following schools from UC records:

```

54053|MIGUEL CONTRERAS LRNG COMPLEX|Los Angeles|LOS ANGELES|MIGUEL CONTRERAS LEARNING COMPLEX [95, 36, 4, 0, '']
54850|ROSELAND COLLEGIATE PREP|Sonoma|SANTA ROSA|ROSELAND COLLEGIATE PREPARATORY [111, 49, 13, 0, '']
52842|SAN DIEGO CITY SCHOOLS|San Diego|SAN DIEGO|SAN DIEGO CITY SCHOOLS [0, 6, 6, 0, '']
54695|SCH SOCIAL JUSTICE-CONT LEARN CTR|Los Angeles|LOS ANGELES|SCHOOL SOCIAL JUSTICE-CONT LEARN CTR [169, 59, 0, 0, '']
54542|SOCIAL JUSTICE HUMANITAS ACAD|Los Angeles|SAN FERNANDO|SOCIAL JUSTICE HUMANITAS ACADEMY [321, 122, 24, 0, '']
54889|UMOJA INTERNATIONAL ACADEMY|Sacramento|SACRAMENTO|UMOJA INTERNATIONAL ACADEMY [73, 40, 3, 0, '']
51687|WESTBROOK ACADEMY|Los Angeles|SOUTH GATE|WESTBROOK ACADEMY [78, 16, 0, 0, '']

```

In [39]: # Suspicious if UC nApplications > CAASPP nStudents. ucSchoolDetails sums across 9 campuses. testSchoolDetails sums across 2 test types.

```

for key in ucSchoolDetails:
    values = ucSchoolDetails[key]
    if values[3] == 0:
        continue # No matching CAASPP record found
    testID = values[4]
    testKey = testID2Details[testID]
    nStudents = testSchoolDetails[testKey][1]
    if values[0] / 9 > nStudents / 2:
        print(key, ucSchoolDetails[key])
        print(testKey, testSchoolDetails[testKey])

```

In [40]: # *** Manual Exploration ***

```

# MIGUEL CONTRERAS LRNG COMPLEX
# query = "contreras/esteban/torres"
# desiredZIP = "90063"
# desiredTestID = ""

```

```

# query = "ANTELOPE|ELITE ACADEMIC ACADEMY|envision|iLead|walnutwood"
# desiredZIP = ""
# desiredTestID = ""

# SWEETWATER SECONDARY SCHOOL|San Diego|CHULA VISTA may be https://www.altussouthbay.com/about/Locations 1655 Broadway, Chula Vista, CA 91911, which may be 3731155|OPTIONS SECONDARY|San Diego|CHULA VISTA|Sweetwater Union High|91911. So this mapping se
# query = "sweetwater|altus"
# desiredZIP = ""
# desiredTestID = "3731155"

# Figure out the two STEM ACADEMY UC records and correct in Lookup.psv
query = "HOLLYWOOD|STEM ACADEMY|BOYLE HEIGHTS"
desiredZIP = ""
desiredTestID = ""
print("UC")
for key in ucSchoolDetails:
    (ucID, school, county, city, expandedSchool) = key.split("|")
    if re.search(query, school, re.IGNORECASE):
        print(key, ucSchoolDetails[key])

print("\nCAASPP")
for key in testSchoolDetails:
    (testID, school, county, city, district, zip, expandedSchool) = key.split("|")
    if re.search(query, school, re.IGNORECASE) or zip == desiredZIP or testID == desiredTestID:
        if testID in usedTestIDs:
            print(key, testSchoolDetails[key], " *** ", usedTestIDs[testID])
        else:
            print(key, testSchoolDetails[key])

UC
51615|HOLLYWOOD HIGH SCHOOL|Los Angeles|LOS ANGELES|HOLLYWOOD HIGH SCHOOL [460, 192, 49, 2, '1934033']
52145|NORTH HOLLYWOOD HIGH SCHOOL|Los Angeles|NORTH HOLLYWOOD|NORTH HOLLYWOOD HIGH SCHOOL [1453, 616, 186, 2, '1936350']
51614|ODYSSEY STEM ACADEMY|Los Angeles|LAKEWOOD|ODYSSEY STEM ACADEMY [522, 170, 20, 1, '0136705']
54784|RIVERSIDE STEM ACADEMY|Riverside|RIVERSIDE|RIVERSIDE STEM ACADEMY [287, 120, 44, 1, '0131359']
54780|STEM ACADEMY OF BOYLE HEIGHTS|Los Angeles|LOS ANGELES|STEM ACADEMY OF BOYLE HEIGHTS [54, 27, 0, 5, '0129536']
54624|STEM ACADEMY OF HOLLYWOOD|Los Angeles|HOLLYWOOD|STEM ACADEMY OF HOLLYWOOD [489, 223, 44, 3, '0125989']

CAASPP
0125989|STEM ACADEMY AT BERNSTEIN HIGH|Los Angeles|HOLLYWOOD/LOS ANGELES|Los Angeles Unified|90028|STEM ACADEMY AT BERNSTEIN HIGH [2, 284] *** 54624|STEM ACADEMY OF HOLLYWOOD|Los Angeles|HOLLYWOOD|STEM ACADEMY OF HOLLYWOOD|3
0129536|BOYLE HEIGHTS SCIENCE, TECHNOLOGY, ENGINEERING AND|Los Angeles|LUGO/LOS ANGELES|Los Angeles Unified|90023|BOYLE HEIGHTS SCIENCE TECHNOLOGY ENGINEERING AND [2, 76] *** 54780|STEM ACADEMY OF BOYLE HEIGHTS|Los Angeles|LOS ANGELES|STEM ACADEMY OF BOYLE HEIGHTS|5
1932821|BOYLE HEIGHTS CONTINUATION|Los Angeles|BOYLE HEIGHTS/LOS ANGELES|Los Angeles Unified|90033|BOYLE HEIGHTS [2, 30]
1934033|HOLLYWOOD SENIOR HIGH|Los Angeles|HOLLYWOOD/LOS ANGELES|Los Angeles Unified|90028|HOLLYWOOD SENIOR HIGH [2, 514] *** 51615|HOLLYWOOD HIGH SCHOOL|Los Angeles|LOS ANGELES|HOLLYWOOD HIGH SCHOOL|2
1936350|NORTH HOLLYWOOD SENIOR HIGH|Los Angeles|CHANDLER/NORTH HOLLYWOOD|Los Angeles Unified|91601|NORTH HOLLYWOOD SENIOR HIGH [2, 1086] *** 52145|NORTH HOLLYWOOD HIGH SCHOOL|Los Angeles|NORTH HOLLYWOOD|NORTH HOLLYWOOD HIGH SCHOOL|2
0136705|ODYSSEY STEM ACADEMY|Los Angeles|LAKEWOOD|Paramount Unified|90712|ODYSSEY STEM ACADEMY [2, 202] *** 51614|ODYSSEY STEM ACADEMY|Los Angeles|LAKEWOOD|ODYSSEY STEM ACADEMY|1
0131359|RIVERSIDE STEM ACADEMY|Riverside|RIVERSIDE|RIVERSIDE STEM ACADEMY [2, 108] *** 54784|RIVERSIDE STEM ACADEMY|Riverside|RIVERSIDE|RIVERSIDE STEM ACADEMY|1
0138040|AEROSTEM ACADEMY|Sutter|MARYSVILLE|AeroSTEM Academy|95991|AEROSTEM ACADEMY [2, 8]
```

In [41]: # Create a map of UC ID to CAASPP ID

```

ucID2testID = {}
for key in ucSchoolDetails:
    (ucID, school, county, city, expandedSchool) = key.split("|")
    testID = ucSchoolDetails[key][4]
    if testID != "":
        ucID2testID[ucID] = testID
print(len(ucID2testID))

1413
```

In [43]: ##### Create a UC dataframe with acceptance rate by school and campus. Include CAASPP school ID for subsequent join. Do not include UC records without a CAASPP match.

```

#           Calculation1          School      City County/State/ Territory Count uad_uc_ethn_7_cat Pivot Field Values Campus
# 0       A B MILLER HIGH SCHOOL50944   A B MILLER HIGH SCHOOL  FONTANA  San Bernardino  Enr Hispanic/ Latinx      24  all
# 1       A B MILLER HIGH SCHOOL50944   A B MILLER HIGH SCHOOL  FONTANA  San Bernardino  Adm Hispanic/ Latinx      56  all
# 2       A B MILLER HIGH SCHOOL50944   A B MILLER HIGH SCHOOL  FONTANA  San Bernardino  App Hispanic/ Latinx      66  all
# 3       A B MILLER HIGH SCHOOL50944   A B MILLER HIGH SCHOOL  FONTANA  San Bernardino  Adm African American      7  all
# ...
# 79222  YUCAIPA SENIOR HIGH SCHOOL53820  YUCAIPA SENIOR HIGH SCHOOL  YUCAIPA  San Bernardino  App      ALL      23  UCSC
# 79223  YUCCA VALLEY HIGH SCHOOL53818  YUCCA VALLEY HIGH SCHOOL  YUCCA VALLEY  San Bernardino  Adm      ALL      3  UCSC
nRows = 0
for index, row in dfUC.iterrows():
    schoolID = row["Calculation1"]
    ucID = re.sub(r"^\.*?(\d{5})$", r"\1", schoolID) # Five digit code, and the first digit is always 5. Already checked that this always succeeds above.
    if ucID not in ucID2testID:
        continue # Don't have CAASPP data for this school
    testID = ucID2testID[ucID]
    if row["uad_uc_ethn_7_cat"] != "All":
        continue # Only deal with all ethnicities combined
    school  = row["School"]
    city    = row["City"]
    county  = row["County/State/ Territory"]
    countType = row["Count"]
    countVal = row["Pivot Field Values"]
    campus  = row["Campus"]

    if nRows == 0:
        dfLeft = pd.DataFrame({{"UC_ID": [ucID], "School": [school], "City": [city], "County": [county], "uad_uc_ethn_7_cat": ["All"], "Campus": [campus], "countyCampus": [countyCampus[county]], "App": [0], "Adm": [0], "Enr": [0], "CAASPP_ID": [testID]}}
```

```

nRows += 1
matchingIndex = dfLeft.query(f'UC_ID == "{ucID}" and Campus == "{campus}"').index
if len(matchingIndex) == 0:
    newRow = pd.DataFrame({ "UC_ID": [ucID], "School": [school], "City": [city], "County": [county], "uad_uc_ethn_7_cat": ["All"], "Campus": [campus], "countyCampus": [countyCampus[county]], "App": [0], "Adm": [0], "Enr": [0], "CAASPP_ID": [testID] })
    dfLeft = pd.concat([dfLeft, newRow], ignore_index=True)
nRows += 1
matchingIndex = dfLeft.query(f'UC_ID == "{ucID}" and Campus == "{campus}"').index
iLeftRow = matchingIndex[0]

dfLeft.at[iLeftRow, countType] += countVal

print(len(dfLeft))

```

12410

```

In [44]: # Merge in the GPA data
dfLeft["App GPA"] = 0.0
dfLeft["Adm GPA"] = 0.0
dfLeft["Enr GPA"] = 0.0
for index, row in dfGPA.iterrows():
    schoolID = row["Calculation"]
    ucID = re.sub(r'^.*?(\d{5})$', r"\1", schoolID) # Five digit code, and the first digit is always 5
    if ucID not in ucID2testID:
        continue # Don't have CAASPP data for this school

    campus = row["Campus"]
    matchingIndex = dfLeft.query(f'UC_ID == "{ucID}" and Campus == "{campus}"').index
    if len(matchingIndex) == 0:
        continue
    iLeftRow = matchingIndex[0]

    appGPA = row["App GPA"]
    admGPA = row["Adm GPA"]
    enrGPA = row["Enr GPA"]
    if appGPA == appGPA: # Check for NaN
        dfLeft.at[iLeftRow, "App GPA"] = appGPA
    if admGPA == admGPA:
        dfLeft.at[iLeftRow, "Adm GPA"] = admGPA
    if enrGPA == enrGPA:
        dfLeft.at[iLeftRow, "Enr GPA"] = enrGPA

```

```

In [45]: # If fewer than 5 students applied, nApp = 0 in the UC data.
# If fewer than 3 students were admitted, nAdm = 0 in the UC data.
# It happens that nApp = 0 and nAdm = 3 or 4.
# Delete all rows with nApp == 0 or nAdm == 0.
dfLeft = dfLeft[dfLeft["App"] != 0]
print(len(dfLeft))
dfLeft = dfLeft[dfLeft["Adm"] != 0]
print(len(dfLeft))

```

12153

10609

```

In [46]: # Compute the admission rate and enrolment rate
dfLeft["AdmRate"] = dfLeft["Adm"] / dfLeft["App"]
dfLeft["EnrRate"] = dfLeft["Enr"] / dfLeft["Adm"]

```

```

In [47]: # Explore
print(dfLeft[dfLeft["Adm"] == 3]) # Smaller values are set to 0 by UC
print(dfLeft[dfLeft["Enr"] == 3]) # Smaller values are set to 0 by UC

```

	UC_ID	School	City	County	uad_uc_ethn_7_cat	Campus	countyCampus	App	Adm	Enr	CAASPP_ID	App GPA	Adm GPA	Enrl GPA	AdmRate	EnrRate
117	55034	AUDIO VALLEY CHARTER SCHOOL	MORENO VALLEY	Riverside	All	all	UCR	5	3	0	0140780	3.66	0.0	0.0	0.600000	0.0
157	50363	BORREGO SPRINGS HIGH SCHOOL	BORREGO SPRINGS	San Diego	All	all	UCSD	8	3	0	0730407	3.76	0.0	0.0	0.375000	0.0
189	52512	CALIFORNIA PACIFIC CHARTER LA	NEWPORT BEACH	Orange	All	all	UCI	5	3	0	0138800	3.21	0.0	0.0	0.600000	0.0
195	54278	CALIFORNIA VIRTUAL ACDM SONOMA	SIMI VALLEY	Ventura	All	all	UCSB	5	3	0	0107284	3.48	0.0	0.0	0.600000	0.0
199	55003	CAMBRIDGE VIRTUAL ACADEMY	ANAHEIM	Orange	All	all	UCI	5	3	0	0140707	3.95	0.0	0.0	0.600000	0.0
...
12257	53485	TAFT UNION HIGH SCHOOL	TAFT	Kern	All	UCSC	UCSB	7	3	0	1535905	3.49	0.0	0.0	0.428571	0.0
12264	54681	TECHNOLOGY PREPARATORY ACADEMY	SAN FERNANDO	Los Angeles	All	UCSC	UCLA	6	3	0	0124404	3.72	0.0	0.0	0.500000	0.0
12287	53315	TRACY INDEPENDENT STUDY CHARTR	TRACY	San Joaquin	All	UCSC	UC Davis	5	3	0	0139949	3.57	0.0	0.0	0.600000	0.0
12352	53704	WATERFORD HIGH SCHOOL	WATERFORD	Stanislaus	All	UCSC	UC Merced	12	3	0	5030259	3.30	0.0	0.0	0.250000	0.0
12390	53790	WINTERS HIGH SCHOOL	WINTERS	Yolo	All	UCSC	UC Davis	8	3	0	5738505	3.56	0.0	0.0	0.375000	0.0

[671 rows x 16 columns]

	UC_ID	School	City	County	uad_uc_ethn_7_cat	Campus	countyCampus	App	Adm	Enr	CAASPP_ID	App GPA	Adm GPA	Enrl GPA	AdmRate	EnrRate
1	54651	ABLE CHARTER	STOCKTON	San Joaquin	All	all	UC Davis	14	10	3	0126755	3.77	3.79	0.0	0.714286	0.300000
6	54247	ACADEMIA AVANCE CHARTER SCHOOL	LOS ANGELES	Los Angeles	All	all	UCLA	12	10	3	0109926	3.58	3.67	0.0	0.833333	0.300000
51	54527	ALLIANCE TENNENBAUM FAMILY TECH HS	LOS ANGELES	Los Angeles	All	all	UCLA	8	8	3	0121293	3.88	3.88	0.0	1.000000	0.375000
106	54260	ASPIRE GOLDEN STATE PREP ACADEMY	OAKLAND	Alameda	All	all	Berkeley	29	14	3	0118224	3.32	3.53	0.0	0.482759	0.214286
116	53199	AUDEO CHARTER SCHOOL	SAN DIEGO	San Diego	All	all	UCSD	5	4	3	3731395	3.82	0.00	0.0	0.800000	0.750000
...
12360	53521	WEST HIGH SCHOOL	TORRANCE	Los Angeles	All	UCSC	UCLA	82	57	3	1939602	3.92	4.11	0.0	0.695122	0.052632
12389	53787	WINDSOR HIGH SCHOOL	WINDSOR	Sonoma	All	UCSC	Berkeley	25	16	3	4930228	3.73	4.05	0.0	0.640000	0.187500
12391	51705	WISH ACADEMY HIGH SCHOOL	LOS ANGELES	Los Angeles	All	UCSC	UCLA	13	10	3	0135632	3.88	4.08	0.0	0.769231	0.300000
12395	52598	WOODCREEK HIGH SCHOOL	ROSEVILLE	Placer	All	UCSC	UC Davis	28	17	3	3130176	3.82	4.04	0.0	0.607143	0.176471
12407	53815	YUBA CITY HIGH SCHOOL	YUBA CITY	Sutter	All	UCSC	UC Davis	40	26	3	5139001	3.71	4.04	0.0	0.650000	0.115385

[1034 rows x 16 columns]

```
In [48]: #### PRIVATE SCHOOLS
# Read the UC admission data for California private schools
directory = "UCAdmissions"
files = {"FR ENR_private.All.csv": "all",
         "FR ENR_private.Berkeley.csv": "Berkeley",
         "FR ENR_private.Irvine.csv": "UCI",
         "FR ENR_private.LosAngeles.csv": "UCLA",
         "FR ENR_private.Merced.csv": "UC Merced",
         "FR ENR_private.Riverside.csv": "UCR",
         "FR ENR_private.SanDiego.csv": "UCSD",
         "FR ENR_private.SantaBarbara.csv": "UCSB",
         "FR ENR_private.SantaCruz.csv": "UCSC"}
dfs = []
for filename in files.keys():
    dfOneFile = pd.read_csv("./".join([directory, filename]), delimiter="\t", encoding="utf-16")
    dfOneFile[("Campus")] = files[filename]
    dfs.append(dfOneFile)
dfPrivate = pd.concat(dfs, ignore_index=True)
dfPrivate[("City")] = dfPrivate[("City")].str.upper()
print(dfPrivate.shape[0])
print(dfPrivate)
```

	Calculation1	School	City	County/State/ Territory	Count	All	African American	American Indian	Hispanic/ Latino	Pacific Islander	Asian	White	Domestic	Unknown	Int'l	Campus
0	A1 COLLEGE PREP55050	A1 COLLEGE PREP	LOS ANGELES	Los Angeles	App	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	all	
1	A1 COLLEGE PREP55050	A1 COLLEGE PREP	LOS ANGELES	Los Angeles	Adm	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	all	
2	A1 COLLEGE PREP55050	A1 COLLEGE PREP	LOS ANGELES	Los Angeles	Enr	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	all	
3	ACACIAWOOD SCHOOLS50381	ACACIAWOOD SCHOOL	ANAHEIM	Orange	App	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	all	
4	ACACIAWOOD SCHOOLS50381	ACACIAWOOD SCHOOL	ANAHEIM	Orange	Adm	3.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	all	
...	
12760	XAVIER COLLEGE PREPARATORY HSS54103	XAVIER COLLEGE PREPARATORY HS	PALM DESERT	Riverside	Adm	18.0	NaN	NaN	4.0	NaN	3.0	7.0	NaN	NaN	UCSC	
12761	XAVIER COLLEGE PREPARATORY HSS54103	XAVIER COLLEGE PREPARATORY HS	PALM DESERT	Riverside	Enr	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	UCSC	
12762	YORK SCHOOLS52053	YORK SCHOOL	MONTEREY	Monterey	App	19.0	NaN	NaN	5.0	NaN	NaN	7.0	NaN	NaN	UCSC	
12763	YORK SCHOOLS52053	YORK SCHOOL	MONTEREY	Monterey	Adm	7.0	NaN	NaN	NaN	NaN	NaN	3.0	NaN	NaN	UCSC	
12764	YORK SCHOOLS52053	YORK SCHOOL	MONTEREY	Monterey	Enr	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	UCSC	

[12765 rows x 15 columns]

```
In [49]: # Create a UC dataframe with acceptance rate by school and campus for ***private schools***
nRows = 0
for index, row in dfPrivate.iterrows():
    schoolID = row["Calculation1"]
    ucID = re.sub(r"^\.*?(\d{5})$|^\*\1", "", schoolID) # Five digit code
    testID = ""
    school = row["School"]
    city = row["City"]
    county = row["County/State/ Territory"]
    countType = row["Count"]
    countVal = row["All"]
    campus = row["Campus"]

    if nRows == 0:
        dfLeftPrivate = pd.DataFrame({("UC_ID": [ucID], "School": [school], "City": [city], "County": [county], "uad_uc_ethn_7_cat": ["All"], "Campus": [campus], "countyCampus": [countyCampus[county]], "App": [0], "Adm": [0], "Enr": [0], "CAASPP_ID": [0]})})
        nRows += 1
    matchingIndex = dfLeftPrivate.query(f'UC_ID == {ucID} and Campus == {campus}'.index
```

```

if len(matchingIndex) == 0:
    newRow = pd.DataFrame({ "UC_ID": [ucID], "School": [school], "City": [city], "County": [county], "uad_uc_ethn_7_cat": ["All"], "Campus": [campus], "countyCampus": [countyCampus[county]], "App": [0], "Adm": [0], "Enr": [0], "CAASPP_ID": [testID] })
    dfLeftPrivate = pd.concat([dfLeftPrivate, newRow], ignore_index=True)
    nrows += 1
    matchingIndex = dfLeftPrivate.query(f'UC_ID == "{ucID}" and Campus == "{campus}"').index
    leftRow = matchingIndex[0]

    dfLeftPrivate.at[leftRow, countType] += countVal

print(len(dfLeftPrivate))

# If too few students applied, App = NaN in the UC data. Delete those rows.
dfLeftPrivate = dfLeftPrivate[dfLeftPrivate["App"] == dfLeftPrivate["App"]]
print(len(dfLeftPrivate))

# If only 1 or 2 students were admitted, Adm is set to NaN in the UC data. Delete those rows.
dfLeftPrivate = dfLeftPrivate[dfLeftPrivate["Adm"] == dfLeftPrivate["Adm"]]
print(len(dfLeftPrivate))

# Compute the admission rate and enrolment rate for each record
dfLeftPrivate["AdmRate"] = dfLeftPrivate["Adm"] / dfLeftPrivate["App"]
dfLeftPrivate["EnrRate"] = dfLeftPrivate["Enr"] / dfLeftPrivate["Adm"]

4255
2573
2070

```

```
In [50]: # Classify private schools as religious vs other
dfLeftPrivate["Religious"] = False
dfLeftPrivate.loc[dfLeftPrivate["School"].str.contains(r"\b(?:CATHOLIC|CHRISTIAN|SAINT|OUR LADY)\b|^(?:ROCK ACADEMY)$", regex=True), "Religious"] = True
```

```
In [51]: # Compute admission statistics for private high schools: all vs local vs secular vs local secular
selectCampus = "UCSD"
dfLeftPrivateCampus = dfLeftPrivate[(dfLeftPrivate["Campus"] == selectCampus)]
dfLeftPrivateLocalCampus = dfLeftPrivate[(dfLeftPrivate["Campus"] == selectCampus) & (dfLeftPrivate["countyCampus"] == dfLeftPrivate["Campus"])]
dfLeftPrivateCampusNonRelig = dfLeftPrivateCampus[dfLeftPrivateCampus["Religious"] == False]
dfLeftPrivateLocalCampusNonRelig = dfLeftPrivateLocalCampus[dfLeftPrivateLocalCampus["Religious"] == False]

print(dfLeftPrivateLocalCampusNonRelig)
print()

nApp = dfLeftPrivateCampus["App"].astype(int).sum()
nAdm = dfLeftPrivateCampus["Adm"].astype(int).sum()
admRate = nAdm / nApp
print(f"{selectCampus} private schools {nApp} {nAdm} {admRate:.1%}")

nApp = dfLeftPrivateLocalCampus["App"].astype(int).sum()
nAdm = dfLeftPrivateLocalCampus["Adm"].astype(int).sum()
admRate = nAdm / nApp
print(f"{selectCampus} local private schools {nApp} {nAdm} {admRate:.1%}")

nApp = dfLeftPrivateCampusNonRelig["App"].astype(int).sum()
nAdm = dfLeftPrivateCampusNonRelig["Adm"].astype(int).sum()
admRate = nAdm / nApp
print(f"{selectCampus} secular private schools {nApp} {nAdm} {admRate:.1%}")

nApp = dfLeftPrivateLocalCampusNonRelig["App"].astype(int).sum()
nAdm = dfLeftPrivateLocalCampusNonRelig["Adm"].astype(int).sum()
admRate = nAdm / nApp
print(f"{selectCampus} local secular private schools {nApp} {nAdm} {admRate:.1%}")
```

UC_ID	School	City	County	uad_uc_ethn_7_cat	Campus	countyCampus	App	Adm	Enr	CAASPP_ID	AdmRate	EnrRate	Religious
3007	50470 ARMY AND NAVY ACADEMY	CARLSBAD	San Diego	All	UCSD	UCSD	20.0	6.0	4.0	0.30000	0.666667	NaN	False
3030	51330 BISHOPS SCHOOL	LA JOLLA	San Diego	All	UCSD	UCSD	68.0	15.0	NaN	0.220588	NaN	False	
3049	54785 CAMBRIDGE SCHOOL	SAN DIEGO	San Diego	All	UCSD	UCSD	12.0	8.0	NaN	0.666667	NaN	False	
3115	52843 FRANCIS W PARKER SCHOOL	SAN DIEGO	San Diego	All	UCSD	UCSD	58.0	8.0	NaN	0.137931	NaN	False	
3159	53036 HIGH BLUFF ACADEMY	RANCHO SANTA FE	San Diego	All	UCSD	UCSD	13.0	3.0	NaN	0.230769	NaN	False	
3189	51333 LA JOLLA COUNTRY DAY SCHOOL	LA JOLLA	San Diego	All	UCSD	UCSD	51.0	11.0	3.0	0.215686	0.272727	NaN	False
3269	50449 PACIFIC RIDGE SCHOOL	CARLSBAD	San Diego	All	UCSD	UCSD	51.0	11.0	NaN	0.215686	NaN	False	

UCSD private schools 10014 2146 21.4%
 UCSD local private schools 712 152 21.3%
 UCSD secular private schools 7343 1548 21.1%
 UCSD local secular private schools 273 62 22.7%

```
In [122... #### Back to CAASPP
# Explore
dfExplore = dfTests[dfTests["Percentage Standard Met and Above"] != ""]
print(dfExplore[(dfExplore["Test ID"] == "2") & (dfExplore["Percentage Standard Met and Above"].astype(float) >= 82)])
```

County Code	District Code	District Name	School Code	School Name	Type	ID	Test Year	Test Type	Test ID	Student Group	ID	Grade	Total Students	Enrolled Total Students	Total Students Tested	Total Students Test
35675	01	61176	Fremont Unified	0135244	Mission San Jose High	7	2024	B	2	1	11		418		410	
410	2759.3		ABC Unified	1931880	Whitney (Gretchen) High	7	2024	B	2	1	11		176		176	
763410	19	64212	Long Beach Unified	1995539	California Academy of Mathematics and Science	7	2024	B	2	1	11		170		170	
176	2803.1		Renaissance Arts Academy	0101683	Renaissance Arts Academy	9	2024	B	2	1	11		35		35	
952958	19	64725	Los Angeles Unified	0102921	Dr. Richard A. Vladovic Harbor Teacher Preparation	7	2024	B	2	1	11		113		113	
170	2773.5		Los Angeles Unified	0133736	Science Academy STEM Magnet	7	2024	B	2	1	11		43		43	
982948	19	64733	San Marino Unified	1937754	San Marino High	7	2024	B	2	1	11		203		194	
35	2776.4		Nevada Joint Union High	0112367	William & Marian Ghidotti High	7	2024	B	2	1	11		36		34	
985924	19	64733	Anaheim Union High	3030616	Oxford Academy	7	2024	B	2	1	11		210		207	
113	2728.5		San Dieguito Union High	0106328	Canyon Crest Academy	7	2024	B	2	1	11		553		544	
1063230	19	64733	Fremont Union High	4334421	Lynbrook High	7	2024	B	2	1	11		404		398	
43	2810.7		Fremont Union High	4334462	Monta Vista High	7	2024	B	2	1	11		438		424	
1369324	19	64964	Gilroy Unified	0114496	Dr. TJ Owens Gilroy Early College Academy	7	2024	B	2	1	11		70		70	
194	2746.1		Los Gatos-Saratoga Union High	4337762	Saratoga High	7	2024	B	2	1	11		293		283	
1741155	29	66357														
34	2711.0															
1775438	30	66431														
207	2756.5															
2888317	37	68346														
544	2748.3															
3367766	43	69468														
398	2792.6															
3367948	43	69468														
424	2767.6															
3369460	43	69484														
70	2718.7															
3382544	43	69534	Los Gatos-Saratoga Union High	4337762												
283	2741.3															

Count Standard Not Met Overall Total	Percentage Exceeded Count Standard Met Exceeded Percentage Standard Met Count Standard Met Percentage Standard Met and Above Count Standard Met and Above Percentage Standard Nearly Met Count Standard Nearly Met Percentage Standard Nearly Met Count Standard Nearly Met Percentage Standard Not Met
35675	69.27
16	410
763410	88.64
1	176
952958	75.88
2	170
982948	80.00
1	35
985924	58.41
2	113
1063230	95.35
0	43
1369324	67.01
9	194
1741155	38.24
1	34
1775438	69.08
5	207
2888317	63.60
32	544
3367766	81.91
12	398
3367948	72.17
18	424
3369460	50.00
3	70
3382544	64.31
22	283

```
In [52]: print(dfTests[dfTests["School Name"].str.contains(r"Middle College|Canyon Crest|Torrey Pines", regex=True)])
```

County Code	District Code	District Name	School Code	School Name	Type	ID	Test Year	Test Type	Test ID	Student Group	ID	Grade	Total Students	Enrolled Total Students	Tested Total Students	Test
ed with Scores	Mean Scale Score	\	61796	West Contra Costa Unified	0730291	Middle College High	7	2024	B	1	1	11	72	72	72	72
277980	07	2689.6	61796	West Contra Costa Unified	0730291	Middle College High	7	2024	B	2	1	11	72	72	72	72
72	07	2638.0	62166	Fresno Unified	0108555	Design Science Middle College High	7	2024	B	1	1	11	55	55	55	55
388051	10	2668.7	62166	Fresno Unified	0108555	Design Science Middle College High	7	2024	B	2	1	11	55	55	55	55
55	2642.2	62265	Kings Canyon Joint Unified	0126292	Reedley Middle College High	10	2024	B	1	1	11	56	56	56	56	
432610	10	2701.7	62265	Kings Canyon Joint Unified	0126292	Reedley Middle College High	10	2024	B	2	1	11	56	56	56	56
432676	10	2616.5	63982	Lemoore Middle College High	0110205	Lemoore Middle College High	9	2024	B	1	1	11	58	58	58	58
719814	16	2680.9	63982	Lemoore Middle College High	0110205	Lemoore Middle College High	9	2024	B	2	1	11	58	58	58	58
719904	16	2621.4	64733	Los Angeles Unified	0102913	Middle College High	7	2024	B	1	1	11	78	78	78	78
78	2666.4	64733	Los Angeles Unified	0102913	Middle College High	7	2024	B	2	1	11	78	78	78	78	
985708	19	2635.7	10306	Unity Middle College High	0133959	Unity Middle College High	9	2024	B	1	1	11	12	12	12	12
12	2537.2	10306	Unity Middle College High	0133959	Unity Middle College High	9	2024	B	2	1	11	12	12	12	12	
12	2463.6	66670	Santa Ana Unified	3030608	Middle College High	7	2024	B	1	1	11	93	93	93	93	
93	2707.7	66670	Santa Ana Unified	3030608	Middle College High	7	2024	B	2	1	11	93	93	93	93	
1923992	30	2650.8	66993	Beaumont Unified	0141366	Beaumont Middle College High	7	2024	B	1	1	11	21	21	21	21
2083588	33	2628.5	66993	Beaumont Unified	0141366	Beaumont Middle College High	7	2024	B	2	1	11	21	21	21	21
2083652	33	2618.3	67249	San Jacinto Unified	0142117	San Jacinto Middle College High	7	2024	B	1	1	11	10	10	10	10
10	*	*	67249	San Jacinto Unified	0142117	San Jacinto Middle College High	7	2024	B	2	1	11	10	10	10	10
2211426	33	2692.6	67876	San Bernardino City Unified	3631090	Middle College High	7	2024	B	1	1	11	71	70	70	70
70	2666.6	67876	San Bernardino City Unified	3631090	Middle College High	7	2024	B	2	1	11	71	70	70	70	
2753221	37	2691.7	68130	Grossmont Union High	3731536	Grossmont Middle College High	7	2024	B	1	1	11	23	22	22	22
22	2680.6	68130	Grossmont Union High	3731536	Grossmont Middle College High	7	2024	B	2	1	11	23	21	21	21	
21	2795011	37	68296	Poway Unified	0141390	Poway to Palomar Middle College High	7	2024	B	1	1	11	42	38	38	38
38	2649.6	68296	Poway Unified	0141390	Poway to Palomar Middle College High	7	2024	B	2	1	11	42	38	38	38	
38	2603.3	68338	Health Sciences High and Middle College	0114462	Health Sciences High and Middle College	9	2024	B	1	1	11	155	153	153	153	
153	2571.7	68338	Health Sciences High and Middle College	0114462	Health Sciences High and Middle College	9	2024	B	2	1	11	155	153	153	153	
153	2512.1	68338	San Diego Unified	0118315	East Village Middle College High	7	2024	B	1	1	11	29	28	28	28	
28	2594.9	68338	San Diego Unified	0118315	East Village Middle College High	7	2024	B	2	1	11	29	28	28	28	
28	2688223	37	68346	San Dieguito Union High	0106328	Canyon Crest Academy	7	2024	B	1	1	11	553	545	545	545
545	2715.6	68346	San Dieguito Union High	0106328	Canyon Crest Academy	7	2024	B	2	1	11	553	544	544	544	
2888317	37	68346	San Dieguito Union High	3730033	Torrey Pines High	7	2024	B	1	1	11	640	622	622	622	
544	2748.3	68346	San Dieguito Union High	3730033	Torrey Pines High	7	2024	B	2	1	11	639	621	621	621	
2888697	37	2670.5	68346	San Dieguito Union High	3730033	Torrey Pines High	7	2024	B	2	1	11	639	621	621	621
621	2650.7	68346	Oxnard Union High	0136804	Oxnard Middle College High	7	2024	B	1	1	11	48	48	48	48	
620	2633.7	72546	Oxnard Union High	0136804	Oxnard Middle College High	7	2024	B	2	1	11	48	48	48	48	
3046939	39	2690.9	72546	Washington Unified	0135939	Washington Middle College High	10	2024	B	1	1	11	21	21	21	21
71	2651.0	72694	Washington Unified	0135939	Washington Middle College High	10	2024	B	2	1	11	21	21	21	21	

Percentage Standard Exceeded Count Standard Exceeded Percentage Standard Met Count Standard Met Percentage Standard and Above Count Standard Met and Above Percentage Standard Nearly Met Count Standard Nearly Met Percentage Standard Not Met

		Count	Standard	Not Met	Overall	Total					
277980 0	72		55.56	40	36.11	26	91.67	66	8.33	6	0.00
278066 14	72		20.83	15	31.94	23	52.78	38	27.78	20	19.44
388051 3	55		50.91	28	29.09	16	80.00	44	14.55	8	5.45
388127 10	55		27.27	15	34.55	19	61.82	34	20.00	11	18.18
432610 0	56		62.50	35	32.14	18	94.64	53	5.36	3	0.00
432676 14	56		8.93	5	46.43	26	55.36	31	19.64	11	25.00
719814 3	57		52.63	30	35.09	20	87.72	50	7.02	4	5.26
719904 13	58		13.79	8	39.66	23	53.45	31	24.14	14	22.41
985708 3	78		48.72	38	37.18	29	85.90	67	10.26	8	3.85
985772 8	78		12.82	10	42.31	33	55.13	43	34.62	27	10.26
1750030 4	12		8.33	1	16.67	2	25.00	3	41.67	5	33.33
1750086 11	12		0.00	0	8.33	1	8.33	1	0.00	0	91.67
1923914 1	93		63.44	59	31.18	29	94.62	88	4.30	4	1.08
1923992 9	93		21.51	20	34.41	32	55.91	52	34.41	32	9.68
2083588 2	21		28.57	6	38.10	8	66.67	14	23.81	5	9.52
2083652 7	21		19.05	4	38.10	8	57.14	12	9.52	2	33.33
2211426 *	*		*	*	*	*	*	*	*	*	*
2211478 *	*		*	*	*	*	*	*	*	*	*
2590486 1	70		58.57	41	31.43	22	90.00	63	8.57	6	1.43
2590562 7	70		31.43	22	34.29	24	65.71	46	24.29	17	10.00
2753221 2	22		59.09	13	22.73	5	81.82	18	9.09	2	9.09
2753275 2	21		42.86	9	28.57	6	71.43	15	19.05	4	9.52
2795011 4	38		42.11	16	28.95	11	71.05	27	18.42	7	10.53
2795879 13	38		13.16	5	28.95	11	42.11	16	23.68	9	34.21
2820000 40	153		19.61	30	28.10	43	47.71	73	26.14	40	26.14
2820092 99	153		5.88	9	11.11	17	16.99	26	18.30	28	64.71
2821514 4	28		53.57	15	25.00	7	78.57	22	7.14	2	14.29
2821594 10	28		17.86	5	25.00	7	42.86	12	21.43	6	35.71
2888223 13	545		67.16	366	24.77	135	91.93	501	5.69	31	2.39
2888317 32	544		63.60	346	18.93	103	82.54	449	11.58	63	5.88
2888697 35	621		48.63	302	33.01	205	81.64	507	12.72	79	5.64
2888799 133	620		31.61	196	28.23	175	59.84	371	18.71	116	21.45
3046939 1	71		69.01	49	28.17	20	97.18	69	1.41	1	1.41
3047023 5	71		39.44	28	33.80	24	73.24	52	19.72	14	7.04
3456862 2	19		31.58	6	36.84	7	68.42	13	21.05	4	10.53
3456934 6	18		33.33	6	16.67	3	50.00	9	16.67	3	33.33
3950228 0	48		52.08	25	39.58	19	91.67	44	8.33	4	0.00
3950296 9	48		22.92	11	25.00	12	47.92	23	33.33	16	18.75
4018838 0	21		52.38	11	33.33	7	85.71	18	14.29	3	0.00
4018906 3	21		14.29	3	42.86	9	57.14	12	28.57	6	14.29

In [53]: # Test for conditions that shouldn't exist in the data
print(dfTests[dfTests["Test_Year"] != "2024"])

```

print(dfTests[dfTests["Test Type"] != "B"])
print(dfTests[dfTests["Student Group ID"] != "1"])
print(dfTests[dfTests["Grade"] != "11"])

Empty DataFrame
Columns: [County Code, District Code, District Name, School Code, School Name, Type ID, Test Year, Test Type, Test ID, Student Group ID, Grade, Total Students Enrolled, Total Students Tested, Total Students Tested with Scores, Mean Scale Score, Percent age Standard Exceeded, Count Standard Exceeded, Percentage Standard Met, Count Standard Met, Percentage Standard Met and Above, Count Standard Met and Above, Percentage Standard Nearly Met, Count Standard Nearly Met, Percentage Standard Not Met, Count Standard Not Met, Overall Total]
Index: []
Empty DataFrame
Columns: [County Code, District Code, District Name, School Code, School Name, Type ID, Test Year, Test Type, Test ID, Student Group ID, Grade, Total Students Enrolled, Total Students Tested, Total Students Tested with Scores, Mean Scale Score, Percent age Standard Exceeded, Count Standard Exceeded, Percentage Standard Met, Count Standard Met, Percentage Standard Met and Above, Count Standard Met and Above, Percentage Standard Nearly Met, Count Standard Nearly Met, Percentage Standard Not Met, Count Standard Not Met, Overall Total]
Index: []
Empty DataFrame
Columns: [County Code, District Code, District Name, School Code, School Name, Type ID, Test Year, Test Type, Test ID, Student Group ID, Grade, Total Students Enrolled, Total Students Tested, Total Students Tested with Scores, Mean Scale Score, Percent age Standard Exceeded, Count Standard Exceeded, Percentage Standard Met, Count Standard Met, Percentage Standard Met and Above, Count Standard Met and Above, Percentage Standard Nearly Met, Count Standard Nearly Met, Percentage Standard Not Met, Count Standard Not Met, Overall Total]
Index: []
Empty DataFrame
Columns: [County Code, District Code, District Name, School Code, School Name, Type ID, Test Year, Test Type, Test ID, Student Group ID, Grade, Total Students Enrolled, Total Students Tested, Total Students Tested with Scores, Mean Scale Score, Percent age Standard Exceeded, Count Standard Exceeded, Percentage Standard Met, Count Standard Met, Percentage Standard Met and Above, Count Standard Met and Above, Percentage Standard Nearly Met, Count Standard Nearly Met, Percentage Standard Not Met, Count Standard Not Met, Overall Total]
Index: []

```

In [54]: # 2024: Create a CAASPP dataframe with pass rate by school. Filter out:

```

# * Rows with "Total Students Tested with Scores" = "" or "Total Students Tested with Scores" < 50
# * Select English Language Arts/Literacy ("Test ID" = 1) or Mathematics ("Test ID" = 2)
keepColumns = ["County Code", "District Code", "District Name", "School Code", "School Name", "Type ID", "Test ID",
              "Total Students Enrolled", "Total Students Tested", "Total Students Tested with Scores", "Mean Scale Score",
              "Count Standard Exceeded", "Count Standard Met", "Count Standard Met and Above", "Count Standard Nearly Met", "Count Standard Not Met"]
dfRight = dfTests[keepColumns] # 5428 rows
dfRight = dfRight[dfRight["Total Students Tested with Scores"] != ""] # 5016 rows remain
dfRight = dfRight[dfRight["Total Students Tested with Scores"].astype(int) >= 50] # 3105 rows remain
print(len(dfRight))

# Add columns for PassRate, County name, countyCampus
dfRight["PassRate"] = dfRight["Count Standard Met and Above"].astype(int) / dfRight["Total Students Tested with Scores"].astype(int)
dfRight["CAASPP County Name"] = dfRight["County Code"].map(code2County)
dfRight["CAASPP County Campus"] = dfRight["CAASPP County Name"].map(countyCampus)

```

3105

In [55]: # 2025: Create a CAASPP dataframe with pass rate by school. Filter out:

```

# * Rows with "Total Students Tested with Scores" = "" or "Total Students Tested with Scores" < 50
# * Select English Language Arts/Literacy ("Test ID" = 1) or Mathematics ("Test ID" = 2)
keepColumns = ["County Code", "District Code", "District Name", "School Code", "School Name", "Type ID", "Test ID",
              "Total Students Enrolled", "Total Students Tested", "Total Students Tested with Scores", "Mean Scale Score",
              "Count Standard Exceeded", "Count Standard Met", "Count Standard Met and Above", "Count Standard Nearly Met", "Count Standard Not Met"]
dfRight2025 = dfTests2025[keepColumns]
dfRight2025 = dfRight2025[dfRight2025["Total Students Tested with Scores"] != ""]
dfRight2025 = dfRight2025[dfRight2025["Total Students Tested with Scores"].astype(int) >= 50]
print(len(dfRight2025))

# Add columns for PassRate, County name, countyCampus
dfRight2025["PassRate"] = dfRight2025["Count Standard Met and Above"].astype(int) / dfRight2025["Total Students Tested with Scores"].astype(int)
dfRight2025["CAASPP County Name"] = dfRight2025["County Code"].map(code2County)
dfRight2025["CAASPP County Campus"] = dfRight2025["CAASPP County Name"].map(countyCampus)

```

3107

In [56]: ##### Read the Schools Table that includes a LCFF+ indicator. The "CDS Code" in this table concatenates the "County Code", "District Code", and "School Code" fields of the CAASPP table.

```

filename = "CA_Schl_Table_EthPct.csv"
dfSchoolsTable = pd.read_csv(filename, delimiter="\t", dtype=str, encoding="latin1")
# dfSchoolsTable = dfSchoolsTable[dfSchoolsTable["Highest Grade"] == "12"] # What is highest grade of "San Jacinto Leadership Academy - Magnet 1", "Norton ... Magnet", etc.?

for index, row in dfSchoolsTable.iterrows():
    school = row["School Name"]
    truncSchool = school[:50].strip()
    district = row["District"]
    if district != district: # NaN
        continue
    key = "|".join([truncSchool, district])
    dfSchoolsTable.at[index, "key"] = key
    dfSchoolsTable.at[index, "Trunc School Name"] = truncSchool
print(dfSchoolsTable)

```

School Name	School and City and Zip Code	County	District	Highest Grade	CDS Code	LCFF Plus School	t UC Feeder	High School	t Closest UC Campus	SAPEP Programs Serving School		
0 21st Century Learning Institute	Street \ 939 East 10th Street	Riverside	Beaumont Unified	12	33669930129882	N	N	UCR	(none)	21st Century Learnin		
1 'O Me-nok Learning Center	Klamath,95548-0065	Del Norte	Del Norte County Unified	6	08618206005417	N	N	UCD	(none)	'O Me-nok Lea		
2 rning Center	300 Minot Creek Road	Orange	Cypress Elementary	6	30664806027767	N	N	UCI	(none)	A. E. Arnol		
3 d Elementary	A. E. Arnold Elementary,Cypress,90630-2724	Orange	Garden Grove Unified	6	30665226028211	N	N	UCI	(none)	A. G. Coo		
4 k Elementary	A. G. Cook Elementary,Garden Grove,92844-2819	Orange	Rialto Unified	5	36678506036669	Y	N	UCR	(none)	A. H. Morga		
5 n Elementary	A. H. Morgan Elementary,Rialto,92376-3665	San Bernardino	Rialto Unified	12	3667850603630530	Y	Y	UCR	(none)	Zoe Barnum High,Eureka,95501-4074		
6 ...	216 West Harris Street	Humboldt	Eureka City Schools	12	12755151232057	Y	N	UCD	(none)	Zoe Barnum High,Eureka,95501-4074		
7 Barnum High	Zoe Barnum High,Eureka,95501-4074	San Bernardino	Rialto Unified	12	3667850603630530	Y	Y	UCR	(none)	Zupanic Virtua		
8 tual Academy	266 West Randall Avenue									l Academy Rialto		
9 key	City Zip Code	ATP Code	Students Served	Masked Enrollment	Total %	African American %	American Indian %	Chicanx/Latinx %	Asian & Pac. Isl. %	White %	Other/Unknown Geodesic Distance	Calpads Upp
0 Beaumont	Trunc School Name											
1 Unified	21st Century Learning Institute	NaN	NaN	668	8.1%	NaN	55.4%	3.9%	26.9%	5.7%	21	1 21st Century Learning Institute Beaumont
2 Unified	Klamath	95548-0065	NaN	190	NaN	58.9%	14.7%	NaN	14.7%	11.6%	240	1 'O Me-nok Learning Center Del Norte County
3 mentary	'O Me-nok Learning Center	NaN	NaN	1,270	5.5%	NaN	35.0%	32.4%	22.2%	4.6%	18	1 A. E. Arnold Elementary Cypress Ele
4 mentary	Cypress	90630-2724	NaN	NaN	694	NaN	20.2%	71.2%	4.6%	4.0%	11	1 A. G. Cook Elementary Garden Grove
5 mentary	A. E. Arnold Elementary	NaN	NaN	1,018	12.4%	NaN	82.3%	1.0%	1.8%	2.2%	11	1 A. H. Morgan Elementary Rialto
6
7 10393	Elk Grove	95757-6262	NaN	2,070	6.5%	NaN	15.7%	56.6%	11.9%	9.3%	22	0 Zehnder Ranch Elementary Elk Grove
8 Unified	Zehnder Ranch Elementary	NaN	NaN	1-5	1,798	30.3%	NaN	61.6%	3.4%	1.3%	13	1 Zela Davis Ha
9 whorne	Hawthorne	90250-6228	NaN	NaN	1,514	0.9%	NaN	88.9%	2.6%	5.8%	79	1 Zephyr Lane Elementary Fairfax Ele
10 mentary	Zela Davis	93307-3180	NaN	NaN	150	NaN	9.3%	20.0%	4.0%	50.7%	200	1 Zoe Barnum High Eureka City
11 10396	Bakersfield	93307-3180	NaN	NaN	1,196	10.5%	NaN	83.1%	1.0%	2.8%	8	1 Zupanic Virtual Academy Rialto
12 Unified	Zephyr Lane Elementary	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN		
13 10397	Eureka	95501-4074	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN		
14 Unified	Zoe Barnum High	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN		
15 10397	Rialto	92376-6926	0 053967	NaN	NaN	NaN	NaN	NaN	NaN	NaN		
16 Unified	Zupanic Virtual Academy	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN		

[10398 rows x 26 columns]

In [57]: # 2024: For each CAASPP school, find matching Schools Table records and add the LCFF Plus indicator to dfRight

```
dfRight["LCFF Plus School t"] = "?" # Add a new column with default value for all rows
nNone = 0
nOne = 0
nOne = 0
nMore = 0

for index, row in dfRight.iterrows():
    cdsCode = ''.join([row["County Code"], row["District Code"], row["School Code"]])
    school = row["School Name"]
    district = row["District Name"]
    dfResult = dfschoolsTable[dfschoolstable["CDS Code"] == cdsCode]
    if len(dfResult) == 1: # Exactly one match
        nOne += 1
        dfRight.at[index, "LCFF Plus School t"] = dfResult.iloc[0]["LCFF Plus School t"]
        # if dfResult.iloc[0]["Highest Grade"] != "12":
        #     print(dfResult.iloc[0])
        # The following schools have "Highest Grade" = 10:
        # * San Jacinto Leadership Academy - Magnet
        # * Norton Science and Language Academy
    elif len(dfResult) > 1: # More than one match
        nMore += 1
    else: # No school/district match
        print("NO MATCH:", school, row["Test ID"])
        nNone += 1

print(nNone, nOne, nMore)
```

NO MATCH: Los Angeles Unified District Level Program 1
 NO MATCH: Los Angeles Unified District Level Program 2
 NO MATCH: Sweetwater Union High District Level Program 1
 NO MATCH: Sweetwater Union High District Level Program 2
 4 3101 0

```
In [58]: # 2025: For each CAASPP school, find matching Schools Table records and add the LCFF Plus indicator to dfRight2025
dfRight2025["LCFF Plus School t"] = "?" # Add a new column with default value for all rows
nNone = 0
nOne = 0
nMore = 0
for index, row in dfRight2025.iterrows():
    cdsCode = ''.join([row["County Code"], row["District Code"], row["School Code"]])
    school = row["School Name"]
    district = row["District Name"]
    result = dfSchoolsTable[dfSchoolsTable["CDS Code"] == cdsCode]
    if len(result) == 1: # Exactly one match
        nOne += 1
        dfRight2025.at[index, "LCFF Plus School t"] = result.iloc[0]["LCFF Plus School t"]
    elif len(result) > 1: # More than one match
        nMore += 1
    else: # No school/district match
        print("NO MATCH:", school, row["Test ID"])
        nNone += 1

print(nNone, nOne, nMore)

NO MATCH: Los Angeles Unified District Level Program 1
NO MATCH: Los Angeles Unified District Level Program 2
NO MATCH: Sweetwater Union High District Level Program 1
NO MATCH: Sweetwater Union High District Level Program 2
4 3103 0
```

```
In [59]: print(dfRight[dfRight["School Name"].str.contains(r"Canyon Crest|Torrey Pines|Gompers|Preuss", regex=True)])
```

ard Met Count Standard Met and Above \\\	County Code District Code	District Name	School Code	School Name Type ID Test ID Total Students Enrolled Total Students Tested Total Students with Scores Mean Scale Score Count Standard Exceeded Count Stand									
				9	1	158	157	157	2603.9	39			
2821922 55	37 94	68338 Gompers Preparatory Academy	0119610 Gompers Preparatory Academy	9	1	158	157	157	2603.9	39			
2822143 11	37 14	68338 Gompers Preparatory Academy	0119610 Gompers Preparatory Academy	9	2	157	156	156	2502.8	3			
2833602 38	37 90	68338 Preuss School UCSD	3731189 Preuss School UCSD	9	1	117	116	116	2652.5	52			
2833810 28	37 54	68338 Preuss School UCSD	3731189 Preuss School UCSD	9	2	117	117	117	2617.7	26			
2888223 135	37 501	68346 San Dieguito Union High	0106328 Canyon Crest Academy	7	1	553	545	545	2715.6	366			
2888317 103	37 449	68346 San Dieguito Union High	0106328 Canyon Crest Academy	7	2	553	544	544	2748.3	346			
2888697 205	37 507	68346 San Dieguito Union High	3730033 Torrey Pines High	7	1	640	622	621	2670.5	302			
2888799 175	37 371	68346 San Dieguito Union High	3730033 Torrey Pines High	7	2	639	621	620	2650.7	196			
Count Standard Nearly Met Count Standard Not Met PassRate CAASPP County Name CAASPP County Campus LCFF Plus School t													
2821922	40	23 0.598726	San Diego	UCSD	Y								
2822143	41	101 0.089744	San Diego	UCSD	Y								
2833602	19	7 0.775862	San Diego	UCSD	Y								
2833810	29	34 0.461538	San Diego	UCSD	Y								
2888223	31	13 0.919266	San Diego	UCSD	N								
2888317	63	32 0.825368	San Diego	UCSD	N								
2888697	79	35 0.816425	San Diego	UCSD	N								
2888799	116	133 0.598387	San Diego	UCSD	N								

```
In [60]: print(dfRight2025[dfRight2025["School Name"].str.contains(r"Canyon Crest|Torrey Pines|Gompers|Preuss", regex=True)])
```

County Code	District Code	District Name	School Code	School Name	Type	ID	Test ID	Total Students Enrolled	Total Students Tested	Total Scores	Mean Scale Score	Count Standard Exceeded	Stand
ard Met Count Standard Met and Above \													
2810944	37	68338	Gompers Preparatory Academy	0119610	Gompers Preparatory Academy	9	1	177	176	176	2621.2	43	
77		120											
2811152	37	68338	Gompers Preparatory Academy	0119610	Gompers Preparatory Academy	9	2	177	175	175	2543.5	6	
25		31											
2822873	37	68338	Preuss School UCSD	3731189	Preuss School UCSD	9	1	115	114	113	2666.6	50	
43		93											
2823091	37	68338	Preuss School UCSD	3731189	Preuss School UCSD	9	2	115	114	114	2626.5	28	
28		56											
2878175	37	68346	San Dieguito Union High	0106328	Canyon Crest Academy	7	1	572	561	561	2724.8	404	
117		521											
2878267	37	68346	San Dieguito Union High	0106328	Canyon Crest Academy	7	2	572	559	559	2756.7	370	
113		483											
2878653	37	68346	San Dieguito Union High	3730033	Torrey Pines High	7	1	605	582	581	2669.4	289	
175		464											
2878751	37	68346	San Dieguito Union High	3730033	Torrey Pines High	7	2	605	584	584	2648.0	202	
123		325											
Count Standard Nearly Met Count Standard Not Met PassRate CAASPP County Name CAASPP County Campus LCFF Plus School t													
2810944	40	16	0.681818	San Diego	UCSD	Y							
2811152	63	81	0.177143	San Diego	UCSD	Y							
2822873	17	3	0.823009	San Diego	UCSD	Y							
2823091	23	35	0.491228	San Diego	UCSD	Y							
2878175	27	13	0.928699	San Diego	UCSD	N							
2878267	38	38	0.864043	San Diego	UCSD	N							
2878653	73	44	0.798623	San Diego	UCSD	N							
2878751	117	142	0.556507	San Diego	UCSD	N							

In [61]: # For each CAASPP school, find matching Schools Table records based on name and district

```
# nOne = 0
# nNone = 0
# nMore1 = 0
# nMore2 = 0
# nBadDistrict = 0
# for index, row in dfRight.iterrows():
#     if row["Test ID"] != "2":
#         continue
#     school = row["School Name"]
#     district = row["District Name"]
#     key = "|".join([school, district])
#     result1 = dfSchoolsTable[dfSchoolsTable["key"] == key]           # Match school and district names
#     result2 = dfSchoolsTable[dfSchoolsTable["Trunc School Name"] == school] # Match school name only
#     if len(result1) == 1: # Exactly one school|district match
#         nOne += 1
#     elif len(result1) > 1: # More than one school/district match
#         nMore1 += 1
#     else: # No school/district match
#         if len(result2) == 1: # Exactly one school match => district name must differ
#             nBadDistrict += 1
#             # if nOne + nBadDistrict < 10:
#             #     print(district, "||", result2.iloc[0]["District"])
#         elif len(result2) > 1: # More than one school match
#             nMore2 += 1
#             matchingDistricts = " | ".join(result2["District"])
#             matchingCodes = " | ".join(result2["CDS Code"])
#             print("BAD DISTRICT:", school, "||", district, "||", matchingDistricts, "||", row["County Code"], row["District Code"], row["School Code"], "||", matchingCodes)
#         else: # No school match
#             nNone += 1
#             print("NO MATCH:", school)

# print()
# print(nOne, nBadDistrict, "nNone:", nNone, "nMore1:", nMore1, "nMore2:", nMore2)
```

In [62]: # UC San Diego partners with 13 high schools across San Diego County through its TRIO Outreach Programs.

```
# These programs are designed for first-generation and/or low-income students who are motivated to attend college.
# Students receive year-round support including tutoring, academic advising, career exposure, college application help, and summer programs.
# UCSD's outreach goal is to help students successfully matriculate into higher education, including UC campuses, CSU campuses, community colleges, and prestigious universities nationwide.
trioSchoolNames = {
    "Chula Vista High": 1,
    "Claremont High": 1,
    "El Cajon Valley High": 1,
    "Hoover High": 1,
    "Lincoln High": 1,
    "Mar Vista High": 1,
    "Mission Bay High": 1,
    "Monte Vista High": 1,
    "Mount Miguel High": 1,
    "Mountain Empire High": 1,
    "San Diego High": 1,
    "Southwest High": 1,
    "Sweetwater Union High": 1,
}
```

```

for index, row in dfRight.iterrows():
    if row["Test ID"] != "2" or row["County Code"] != "37":
        continue
    cdsCode = "".join([row["County Code"], row["District Code"], row["School Code"]])
    schoolName = row["School Name"]
    if schoolName in trioSchoolNames:
        print(f"(cdsCode): \"TRIO\", # {schoolName}")

```

```

"37681303731692": "TRIO", # El Cajon Valley High
"37681303734548": "TRIO", # Monte Vista High
"37681303734761": "TRIO", # Mount Miguel High
"37682133734878": "TRIO", # Mountain Empire High
"37683380107219": "TRIO", # San Diego High
"37683380114025": "TRIO", # Lincoln High
"37683383731213": "TRIO", # Clairemont High
"37683383732997": "TRIO", # Hoover High
"37683383734431": "TRIO", # Mission Bay High

In [63]: # Still need to find:
# Chula Vista High School
# Mar Vista High School
# Southwest High School
# Sweetwater Union High School
query = r"Chula Vista|Mar Vista|Southwest|Sweetwater"
for index, row in dfRight.iterrows():
    if row["Test ID"] != "2" or row["County Code"] != "37":
        continue
    cdsCode = "".join([row["County Code"], row["District Code"], row["School Code"]])
    schoolName = row["School Name"]
    if re.search(query, schoolName, re.IGNORECASE):
        print(f"(cdsCode): \"TRIO\", # {schoolName}")

"37680236115778": "TRIO", # Chula Vista Learning Community Charter
"37684113730124": "TRIO", # Southwest Senior High
"37684113731064": "TRIO", # Chula Vista Senior High
"37684113733953": "TRIO", # Mar Vista Senior High
"37684113738226": "TRIO", # Sweetwater High
"37684113768411": "TRIO", # Sweetwater Union High District Level Program
"37764710114678": "TRIO", # High Tech High Chula Vista

```

```

In [64]: # Dictionaries for category column to highlight specific schools in plots
highlight = {
    "19647251995539": "CAMS", # CALIFORNIA ACADEMY OF MATH & SCIENCE
    "37683460106328": "CCA", # CANYON CREST ACADEMY
    "#": "37682960118935": "Norte", # Del Norte High
    "#": "37683463730033": "Torrey", # Torrey Pines High
    "#": "37683380119610": "Partner", # GOMPERS PREPARATORY ACADEMY
    "#": "37683383731189": "Partner", # PREUSS SCHOOL UCSD
    "#": "37683380107086": "KEARNY", # Kearny School of Biomedical Science and Technology
    "#": "37681303731692": "TRIOy", # EL Cajon Valley High
    "#": "37681303734548": "TRIOy", # Monte Vista High
    "#": "37681303734761": "TRIOy", # Mount Miguel High
    "#": "37682133734878": "TRIOh", # Mountain Empire High
    "#": "37683380107219": "TRIOh", # San Diego High
    "#": "37683380114025": "TRIOy", # Lincoln High
    "#": "37683383731213": "TRIOh", # Clairemont High
    "#": "37683383732997": "TRIOy", # Hoover High
    "#": "37683383734431": "TRIOh", # Mission Bay High
    "#": "37684113730124": "TRIOy", # Southwest Senior High
    "#": "37684113731064": "TRIOy", # Chula Vista Senior High
    "#": "37684113733953": "TRIOy", # Mar Vista Senior High
    "#": "37684113738226": "TRIOy", # Sweetwater High
    "#": "", #
}

customPalette = {
    "CAMS": "purple",
    "CCA": "black",
    "Partner": "green",
    "KEARNY": "yellow",
    "TRIOh": "darkred",
    "TRIOy": "red",
    "LCFF+": "grey",
    "": "#1f77b4",
}

categoryText = {
    "CAMS": "California Academy of Math & Science (not LCFF+)",
    "CCA": "Canyon Crest Academy (not LCFF+)",
    "Partner": "UCSD Partners Gompers & Preuss (LCFF+)",
    "KEARNY": "Kearny School of Biomedical Science & Tech (LCFF+)",
    "TRIOh": "UCSD TRIO Outreach Program (not LCFF+)",
    "TRIOy": "UCSD TRIO Outreach Program (LCFF+)",
    "LCFF+": "Other LCFF+",
    "": "Other not LCFF+",
}

```

```

}
categoryTextNarrow = {
    "CAMS": "CA Acad Ma&Sci\n(n(not LCFF+)",
    "CCA": "CCA\n(n(not LCFF+)",
    "Partner": "UCSD Partners\n(LCFF+)",
    "KEARNY": "Kearny Biomed\n(LCFF+)",
    "TRIOy": "UCSD TRIO\nLCFF+",
    "TRIOn": "UCSD TRIO\nnot LCFF+",
    "LCFF+": "Other\nLCFF+",
    " ": "Other\nnot LCFF+",
}

```

```

In [65]: # Add a new column 'Category'
dfRight["Category"] = " " # Initialize with " " for all rows. " " sorts before strings beginning with letters.
for index, row in dfRight.iterrows():
    cdsCode = "".join([row["County Code"], row["District Code"], row["School Code"]])
    if cdsCode in highlight:
        dfRight.at[index, "Category"] = highlight[cdsCode]
    if row["Test ID"] == "2":
        print(cdsCode, row["School Name"], row["LCFF Plus School t"])
    elif row["LCFF Plus School t"] == "Y":
        dfRight.at[index, "Category"] = " LCFF+"

```

19647251995539 California Academy of Mathematics and Science N
37681303731692 El Cajon Valley High Y
37681303734548 Monte Vista High Y
37681303734761 Mount Miguel High Y
37682133734878 Mountain Empire High N
37683380107088 Kearny School of Biomedical Science and Technology Y
37683380107219 San Diego High N
37683380114025 Lincoln High Y
37683380119610 Gompers Preparatory Academy Y
37683383731189 Preuss School UCSD Y
37683383731213 Clairemont High N
37683383732997 Hoover High Y
37683383734431 Mission Bay High N
37683460106328 Canyon Crest Academy N
37684113730124 Southwest Senior High Y
37684113731064 Chula Vista Senior High Y
37684113733953 Mar Vista Senior High Y
37684113738226 Sweetwater High Y

```
In [66]: print(dfRight)
```

County	Code	District	Code	District	Name	School	Code	School	Name	Type	ID	Test	ID	Total	Students	Enrolled	Total	Students	Tested	Total	Students	Tested	with Scores	Mean Scale	Score	Count	Standard	Exceeded
Count	Standard	Met	\																									
11469		01	61119		Alameda Unified		0130229		Alameda High		7	1		467		440			440		2647.1		181					
157																												
11569		01	61119		Alameda Unified		0130229		Alameda High		7	2		466		438			438		2623.7		131					
109																												
11791		01	61119	Alameda Community Learning Center			0130609	Alameda Community Learning Center			9	1		53		51			51		2556.0		8					
10																												
11999		01	61119	Alameda Community Learning Center			0130609	Alameda Community Learning Center			9	2		53		50			50		2515.0		3					
10																												
12220		01	61119		Alameda Unified		0132142		Encinal Junior/Senior High		7	1		239		218			218		2627.4		69					
77																												
...	
4039195		58	72736	Marysville Joint Unified			5830096	South Lindhurst Continuation High			7	2		140		128			128		2439.4		0					
1																												
4039649		58	72736	Marysville Joint Unified			5835202		Marysville High		7	1		231		223			223		2542.9		29					
56																												
4039751		58	72736	Marysville Joint Unified			5835202		Marysville High		7	2		231		227			227		2501.7		5					
17																												
4050422		58	72769	Wheatland Union High			5838305		Wheatland Union High		7	1		240		237			237		2582.1		48					
80																												
4050524		58	72769	Wheatland Union High			5838305		Wheatland Union High		7	2		240		237			237		2545.9		14					
41																												

Count	Standard	Met and Above	Count	Standard	Nearly Met	Count	Standard	Not Met	PassRate	CAASPP	County	Name	CAASPP	County	Campus	LCFF	Plus School	t	Category
11469		338			52				50	0.768182	Alameda	Berkeley		N					
11569		240			72				126	0.547945	Alameda	Berkeley		N					
11791		18			20				13	0.352941	Alameda	Berkeley		N					
11999		13			8				29	0.260000	Alameda	Berkeley		N					
12220		146			39				33	0.669725	Alameda	Berkeley		N					
...	
4039195		1			10				117	0.007812	Yuba	UC Davis	Y	LCFF+					
4039649		85			65				73	0.381166	Yuba	UC Davis	N						
4039751		22			58				147	0.096916	Yuba	UC Davis	N						
4050422		128			59				50	0.540084	Yuba	UC Davis	Y	LCFF+					
4050524		55			62				120	0.232068	Yuba	UC Davis	Y	LCFF+					

[3105 rows x 21 columns]

```
In [67]: # Overall CAASPP stats, not grouped by school types
dfRightByYear = {
    "2024": dfRight,
    "2025": dfRight2025
}
for year in ("2024", "2025"):
    df = dfRightByYear[year]
    for selectTestID in ("2", "1"):
        filteredRight = df[df["Test ID"] == selectTestID]
        nStudents = filteredRight["Total Students Enrolled"].astype(int).sum()
        nStudentsMet = filteredRight["Count Standard Met and Above"].astype(int).sum()
        nStudentsNotMet = filteredRight["Count Standard Nearly Met"].astype(int).sum() + filteredRight["Count Standard Not Met"].astype(int).sum()
        rateStudentsMet = nStudentsMet / nStudents
        print(f'{year} {selectTestID} {nStudents} {nStudentsMet} {nStudentsNotMet} {rateStudentsMet:.1%}')
2024 Math          438903 119026 296830 27.1%
2024 English (ELA) 439205 236153 180330 53.8%
2025 Math          433498 129025 284181 29.8%
2025 English (ELA) 433722 239913 173986 55.3%
```

```
In [68]: # Compute CAASPP statistics independent of UC applications (2024). Some info is buried in statsCAASPP() output below.
filteredRight = dfRight[dfRight["Test ID"] == "2"]
nStudentsAll = filteredRight["Total Students Enrolled"].astype(int).sum()
filteredRight = filteredRight[filteredRight["PassRate"] <= 0.25]
nStudents25 = filteredRight["Total Students Enrolled"].astype(int).sum()
rate = nStudents25 / nStudentsAll
print(f'CAASPP Pass <= 25% Schools {rate:.1%} {nStudents25} {nStudentsAll}')
CAASPP Pass <= 25% Schools 52.4% 229828 438903
```

```
In [69]: def statsCAASPP(df, year):
    # Fraction of schools < 25%
    # Fraction of students in schools < 25%
    # Fraction of LCFF+ schools
    # Fraction of students in LCFF+ schools
    for selectTestID in ("2", "1"):
        filteredRight = df[df["Test ID"] == selectTestID]

        nSchools = len(filteredRight)
        nStudents = filteredRight["Total Students Enrolled"].astype(int).sum()

        dfLowCAASPP = filteredRight[filteredRight["PassRate"] <= 0.25]
        nSchoolsLowCAASPP = len(dfLowCAASPP)
        nStudentsLowCAASPP = dfLowCAASPP["Total Students Enrolled"].astype(int).sum()
        rateSchoolsLowCAASPP = nSchoolsLowCAASPP / nSchools
        rateStudentsLowCAASPP = nStudentsLowCAASPP / nStudents

        dfHighCAASPP = filteredRight[filteredRight["PassRate"] >= 0.75]
        nSchoolsHighCAASPP = len(dfHighCAASPP)
        nStudentsHighCAASPP = dfHighCAASPP["Total Students Enrolled"].astype(int).sum()
        rateSchoolsHighCAASPP = nSchoolsHighCAASPP / nSchools
        rateStudentsHighCAASPP = nStudentsHighCAASPP / nStudents

        dfLCFF = filteredRight[filteredRight["LCFF Plus School t"] == "Y"]
        nSchoolsLCFF = len(dfLCFF)
        nStudentsLCFF = dfLCFF["Total Students Enrolled"].astype(int).sum()
        rateSchoolsLCFF = nSchoolsLCFF / nSchools
        rateStudentsLCFF = nStudentsLCFF / nStudents
        topSchoolLCFF = dfLCFF.loc[dfLCFF["PassRate"].idxmax(), "School Name"]

        dfLCFFLowCAASPP = dfLCFF[dfLCFF["PassRate"] <= 0.25]
        nSchoolsLCFFLowCAASPP = len(dfLCFFLowCAASPP)
        nStudentsLCFFLowCAASPP = dfLCFFLowCAASPP["Total Students Enrolled"].astype(int).sum()
        rateSchoolsLCFFLowCAASPP = nSchoolsLCFFLowCAASPP / nSchoolsLCFF
        rateStudentsLCFFLowCAASPP = nStudentsLCFFLowCAASPP / nStudentsLCFF

        dfLCFFHighCAASPP = dfLCFF[dfLCFF["PassRate"] >= 0.75]
        nSchoolsLCFFHighCAASPP = len(dfLCFFHighCAASPP)
        nStudentsLCFFHighCAASPP = dfLCFFHighCAASPP["Total Students Enrolled"].astype(int).sum()
        rateSchoolsLCFFHighCAASPP = nSchoolsLCFFHighCAASPP / nSchoolsLCFF
        rateStudentsLCFFHighCAASPP = nStudentsLCFFHighCAASPP / nStudentsLCFF

        dfNotLCFF = filteredRight[filteredRight["LCFF Plus School t"] == "N"]
        nSchoolsNotLCFF = len(dfNotLCFF)
        nStudentsNotLCFF = dfNotLCFF["Total Students Enrolled"].astype(int).sum()
        rateSchoolsNotLCFF = nSchoolsNotLCFF / nSchools
        rateStudentsNotLCFF = nStudentsNotLCFF / nStudents

        dfNotLCFFLowCAASPP = dfNotLCFF[dfNotLCFF["PassRate"] <= 0.25]
        nSchoolsNotLCFFLowCAASPP = len(dfNotLCFFLowCAASPP)
        nStudentsNotLCFFLowCAASPP = dfNotLCFFLowCAASPP["Total Students Enrolled"].astype(int).sum()
        rateSchoolsNotLCFFLowCAASPP = nSchoolsNotLCFFLowCAASPP / nSchoolsNotLCFF
        rateStudentsNotLCFFLowCAASPP = nStudentsNotLCFFLowCAASPP / nStudentsNotLCFF
```

```

dfNotLCFFHighCAASPP = dfNotLCFF[dfNotLCFF["PassRate"] >= 0.75]
nSchoolsNotLCFFHighCAASPP = len(dfNotLCFFHighCAASPP)
nStudentsNotLCFFHighCAASPP = dfNotLCFFHighCAASPP["Total Students Enrolled"].astype(int).sum()
rateSchoolsNotLCFFHighCAASPP = nSchoolsNotLCFFHighCAASPP / nSchoolsNotLCFF
rateStudentsNotLCFFHighCAASPP = nStudentsNotLCFFHighCAASPP / nStudentsNotLCFF

ratioSchoolsFailLCFF = rateSchoolsLowCAASPP / rateSchoolsLCFF
ratioStudentsFailLCFF = rateStudentsLowCAASPP / rateStudentsLCFF

print(f"\n{year} {testID2Name[selectTestID]:20} CAASPP LCFF LCFF Not LCFF")
print(f"          nLow nHigh rLow rHigh nLCFF nNot rLCFF rNot Low/LCFF rLow rHigh rLow rHigh topSchoolLCFF")
print(f"All    Counties nSchools: {nSchools:6} {nSchoolsLowCAASPP:6} {nSchoolsHighCAASPP:6} {rateSchoolsLowCAASPP:5.1%} {rateSchoolsHighCAASPP:5.1%} {nSchoolsLCFF:6} {nSchoolsNotLCFF:6} {rateSchoolsLCFF:5.1%} {rateSchoolsNotLCFF:5.1%} {rat")
print(f"All    Counties nStudents: {nStudents:6} {nStudentsLowCAASPP:6} {nStudentsHighCAASPP:6} {rateStudentsLowCAASPP:5.1%} {rateStudentsHighCAASPP:5.1%} {nStudentsLCFF:6} {nStudentsNotLCFF:6} {rateStudentsLCFF:5.1%} {rateStudentsNotLCFF:5

for selectCampus in campuses:
    filteredRight = df[df["Test ID"] == selectTestID]
    filteredRight = filteredRight[filteredRight["CAASPP County Campus"] == selectCampus]

    nSchools = len(filteredRight)
    nStudents = filteredRight["Total Students Enrolled"].astype(int).sum()

    dfLowCAASPP = filteredRight[filteredRight["PassRate"] <= 0.25]
    nSchoolsLowCAASPP = len(dfLowCAASPP)
    nStudentsLowCAASPP = dfLowCAASPP["Total Students Enrolled"].astype(int).sum()
    rateSchoolsLowCAASPP = nSchoolsLowCAASPP / nSchools
    rateStudentsLowCAASPP = nStudentsLowCAASPP / nStudents

    dfHighCAASPP = filteredRight[filteredRight["PassRate"] >= 0.75]
    nSchoolsHighCAASPP = len(dfHighCAASPP)
    nStudentsHighCAASPP = dfHighCAASPP["Total Students Enrolled"].astype(int).sum()
    rateSchoolsHighCAASPP = nSchoolsHighCAASPP / nSchools
    rateStudentsHighCAASPP = nStudentsHighCAASPP / nStudents

    dfLCFF = filteredRight[filteredRight["LCFF Plus School t"] == "Y"]
    nSchoolsLCFF = len(dfLCFF)
    nStudentsLCFF = dfLCFF["Total Students Enrolled"].astype(int).sum()
    rateSchoolsLCFF = nSchoolsLCFF / nSchools
    rateStudentsLCFF = nStudentsLCFF / nStudents
    topSchoolLCFF = dfLCFF.loc[dfLCFF["PassRate"].idxmax(), "School Name"]

    dfLCFFLowCAASPP = dfLCFF[dfLCFF["PassRate"] <= 0.25]
    nSchoolsLCFFLowCAASPP = len(dfLCFFLowCAASPP)
    nStudentsLCFFLowCAASPP = dfLCFFLowCAASPP["Total Students Enrolled"].astype(int).sum()
    rateSchoolsLCFFLowCAASPP = nSchoolsLCFFLowCAASPP / nSchoolsLCFF
    rateStudentsLCFFLowCAASPP = nStudentsLCFFLowCAASPP / nStudentsLCFF

    dfLCFFHighCAASPP = dfLCFF[dfLCFF["PassRate"] >= 0.75]
    nSchoolsLCFFHighCAASPP = len(dfLCFFHighCAASPP)
    nStudentsLCFFHighCAASPP = dfLCFFHighCAASPP["Total Students Enrolled"].astype(int).sum()
    rateSchoolsLCFFHighCAASPP = nSchoolsLCFFHighCAASPP / nSchoolsLCFF
    rateStudentsLCFFHighCAASPP = nStudentsLCFFHighCAASPP / nStudentsLCFF

    dfNotLCFF = filteredRight[filteredRight["LCFF Plus School t"] == "N"]
    nSchoolsNotLCFF = len(dfNotLCFF)
    nStudentsNotLCFF = dfNotLCFF["Total Students Enrolled"].astype(int).sum()
    rateSchoolsNotLCFF = nSchoolsNotLCFF / nSchools
    rateStudentsNotLCFF = nStudentsNotLCFF / nStudents

    dfNotLCFFLowCAASPP = dfNotLCFF[dfNotLCFF["PassRate"] <= 0.25]
    nSchoolsNotLCFFLowCAASPP = len(dfNotLCFFLowCAASPP)
    nStudentsNotLCFFLowCAASPP = dfNotLCFFLowCAASPP["Total Students Enrolled"].astype(int).sum()
    rateSchoolsNotLCFFLowCAASPP = nSchoolsNotLCFFLowCAASPP / nSchoolsNotLCFF
    rateStudentsNotLCFFLowCAASPP = nStudentsNotLCFFLowCAASPP / nStudentsNotLCFF

    dfNotLCFFHighCAASPP = dfNotLCFF[dfNotLCFF["PassRate"] >= 0.75]
    nSchoolsNotLCFFHighCAASPP = len(dfNotLCFFHighCAASPP)
    nStudentsNotLCFFHighCAASPP = dfNotLCFFHighCAASPP["Total Students Enrolled"].astype(int).sum()
    rateSchoolsNotLCFFHighCAASPP = nSchoolsNotLCFFHighCAASPP / nSchoolsNotLCFF
    rateStudentsNotLCFFHighCAASPP = nStudentsNotLCFFHighCAASPP / nStudentsNotLCFF

    ratioSchoolsFailLCFF = rateSchoolsLowCAASPP / rateSchoolsLCFF
    ratioStudentsFailLCFF = rateStudentsLowCAASPP / rateStudentsLCFF

print(f"\n{selectCampus:9} Counties nSchools: {nSchools:6} {nSchoolsLowCAASPP:6} {nSchoolsHighCAASPP:6} {rateSchoolsLowCAASPP:5.1%} {rateSchoolsHighCAASPP:5.1%} {nSchoolsLCFF:6} {nSchoolsNotLCFF:6} {rateSchoolsLCFF:5.1%} {rateSchoolsNotLCFF:5.1%} {rat")
print(f"\n{selectCampus:9} Counties nStudents: {nStudents:6} {nStudentsLowCAASPP:6} {nStudentsHighCAASPP:6} {rateStudentsLowCAASPP:5.1%} {rateStudentsHighCAASPP:5.1%} {nStudentsLCFF:6} {nStudentsNotLCFF:6} {rateStudentsLCFF:5.1%} {rateStudentsNotLCFF:5.1%} {rat
print("")
```

In [70]: # Get CAASPP statistics independent of UC applications for 2024 and 2025
 statsCAASPP(dfRight, 2024)
 statsCAASPP(dfRight2025, 2025)

UCR	Counties nStudents:	61526	5799	7003	9.4%	11.4%	32221	29305	52.4%	47.6%	0.18
UCSB	Counties nSchools:	99	14	8	14.1%	8.1%	42	57	42.4%	57.6%	0.33
UCSB	Counties nStudents:	31314	1774	1775	5.7%	5.7%	14427	16887	46.1%	53.9%	0.12
UCSC	Counties nSchools:	91	10	21	11.0%	23.1%	29	62	31.9%	68.1%	0.34
UCSC	Counties nStudents:	27367	860	7516	3.1%	27.5%	6553	20814	23.9%	76.1%	0.13
UCSD	Counties nSchools:	133	14	27	10.5%	20.3%	49	83	36.8%	62.4%	0.29
UCSD	Counties nStudents:	37684	1720	8691	4.6%	23.1%	12407	25007	32.9%	66.4%	0.14

```
In [71]: def createPlot(ax, dfLeft, dfRight, selectCampus, selectCounties, xAxis="PassRate", yAxis="AdmRate", filterMetrics=[], selectTest="Math", selectMinApps=20, selectClassSize=50, title="auto", legend="N", compact="N"):
    # selectCampus = "UCSD"
    # selectCounties = "San Diego"
    # selectTest = "Math" # "English", "ELA", or "Math"
    # xAxis = "PassRate" # "PassRate", "App GPA"
    # yAxis = "AdmRate" # "AdmRate", "Adm GPA", "App GPA", "EnrRate", "SeniorRate"
    # selectClassSize = 100 # Records with <50 were already filtered in dfRight
    # title = "auto" # "" for no title, "auto", or a string for the title

    if selectTest in ("English", "ELA"):
        selectTestID = "1"
    else:
        selectTestID = "2" # Math

    # Filter the dataframes
    filteredLeft = dfLeft[dfLeft["Campus"] == selectCampus]
    filteredLeft = filteredLeft[filteredLeft["App"] >= selectMinApps]
    if selectCounties == "Local":
        filteredLeft = filteredLeft[filteredLeft["CountyCampus"] == selectCampus]
    elif selectCounties != "All":
        filteredLeft = filteredLeft[filteredLeft["County"] == selectCounties]
    if xAxis == "App GPA" or yAxis == "App GPA":
        filteredLeft = filteredLeft[filteredLeft["App GPA"] >= 3.0]
    if yAxis == "Adm GPA":
        filteredLeft = filteredLeft[filteredLeft["Adm GPA"] >= 3.0]

    filteredRight = dfRight[dfRight["Test ID"] == selectTestID]
    filteredRight = filteredRight[filteredRight["Total Students Tested with Scores"].astype(int) >= selectClassSize]

    # Join
    joined = pd.merge(filteredLeft, filteredRight, left_on="CAASPP_ID", right_on="School Code", how="inner")

    joined["SeniorRate"] = joined["Adm"].astype(int) / joined["Total Students Enrolled"].astype(int) # Fraction of seniors admitted
    joined["ApplyRate"] = joined["App"].astype(int) / joined["Total Students Enrolled"].astype(int) # Fraction of seniors who applied

    # Filter out specified rows
    for spec in filterMetrics:
        filterMetric = spec[0]
        filterOp = spec[1]
        filterVal = spec[2]
        if filterOp == "<=":
            joined = joined[joined[filterMetric] <= filterVal]
        elif filterOp == ">=":
            joined = joined[joined[filterMetric] >= filterVal]
        elif filterOp == "=":
            joined = joined[joined[filterMetric] == filterVal]

    dfSorted = joined.sort_values(by="Category")

    # Scatterplot with bubble sizes and weighted linear fit.
    # Seaborn's scatterplot returns a Matplotlib Axes object (ax), which we customize.
    sns.scatterplot(
        x=xAxis,
        y=yAxis,
        size="ApplyRate", # Bubble size is application rate
        sizes=(20, 300), # min/max point sizes
        hue="Category",
        palette=customPalette,
        data=dfSorted,
        legend=False,
        alpha=0.6,
        ax=ax
    )

    # Format axes as percentages
    if re.search("Rate", xAxis):
        ax.xaxis.set_major_formatter(PercentFormatter(xmax=1, decimals=0)) # denominator=1 means 1.0 → 100%
    if re.search("Rate", yAxis):
        ax.yaxis.set_major_formatter(PercentFormatter(xmax=1, decimals=0))

    # Set axis labels
    if compact == "N":
        axisFontSize=11
        labelpad = 1
    else:
```

```

axisFontSize=12
labelPad = 2
if xAxis == "PassRate":
    ax.set_xlabel(f"CAASPP {selectTest} Pass Rate", fontsize=axisFontSize, labelpad=labelPad)
elif xAxis == "App GPA":
    ax.set_xlabel(f"GPA of Applicants to {selectCampus}", fontsize=axisFontSize, labelpad=labelPad)
elif xAxis == "ApplyRate":
    ax.set_xlabel(f"Fraction of Seniors who Apply to {selectCampus}", fontsize=axisFontSize, labelpad=labelPad)
if yAxis == "AdmRate":
    if compact == "y":
        ax.set_ylabel(f"{selectCampus} Adm. Rate", fontsize=axisFontSize, labelpad=labelPad)
    else:
        ax.set_ylabel(f"{selectCampus} Admission Rate", fontsize=axisFontSize, labelpad=labelPad)
elif yAxis == "Adm GPA":
    ax.set_ylabel(f"GPA of Admitted Applicants to {selectCampus}", fontsize=axisFontSize, labelpad=labelPad)
elif yAxis == "App GPA":
    ax.set_ylabel(f"GPA of Applicants to {selectCampus}", fontsize=axisFontSize, labelpad=labelPad)
elif yAxis == "EnrRate":
    ax.set_ylabel(f"{selectCampus} Enrollment Rate", fontsize=axisFontSize, labelpad=labelPad)
elif yAxis == "SeniorRate":
    ax.set_ylabel(f"{selectCampus} nAdmitted/nSeniors", fontsize=axisFontSize, labelpad=labelPad)

# Set plot title
if title == "auto":
    if selectCounties in ("All", "Local"):
        ax.set_title(f"High Schools in {selectCounties} Counties")
    else:
        ax.set_title(f"High Schools in {selectCounties} County")
elif title != "":
    ax.set_title(title)

# Compute the weighted regression separately and then plot the line.
X = joined[xAxis]
y = joined[yAxis]
if yAxis == "Adm GPA":
    weights = joined["Adm"] # Weight is number of students admitted
elif yAxis == "EnrRate":
    weights = joined["Adm"] # Weight is number of students admitted
elif yAxis == "SeniorRate":
    weights = joined["ApplyRate"] # Weight is fraction of seniors who applied
    # weights = joined["Total Students Enrolled"].astype(float) # Weight is number of students
else:
    weights = joined["App"].astype(float) # Weight is number of applications
X_const = sm.add_constant(X) # Add constant for intercept

# Weighted least squares regression
modelWLSq = sm.WLS(y, X_const, weights=weights)
results = modelWLSq.fit()

# Get fitted line
x_vals = np.linspace(X.min(), X.max(), 100)
y_vals = results.predict(sm.add_constant(x_vals))

# Add the regression line to the plot
# If the p-value is small (commonly < 0.05), we reject the null hypothesis that the slope = 0 => there is evidence of a trend between X and Y.
slopeRange = results.conf_int().loc[xAxis]
angle0 = math.degrees(math.atan(slopeRange[0]))
angle1 = math.degrees(math.atan(slopeRange[1]))
angleDiff = abs(angle1 - angle0)
# if results.rsquared > 0.01 or results.pvalues[xAxis] < 0.05:
if angleDiff < 30:
    sns.lineplot(x=x_vals, y=y_vals, color="blue", ax=ax)

if legend == "Y":
    ax.legend([],[], frameon=False) # Remove Seaborn's automatic legend

# Build custom Legend handles
legendElements = []
for category in categoryText:
    legendElements.append(Line2D([0], [0], marker='o', color='w', label=categoryText[category], markerfacecolor=customPalette[category], markersize=10))

# Add custom legend
ax.legend(handles=legendElements, bbox_to_anchor=(1.05, 1), loc="upper left")

return

```

```

In [72]: def createPlots(dfLeft, dfRight, selectCounties, xAxis="PassRate", yAxis="AdmRate", filterMetrics=[], selectTest="Math", selectMinApps=20, selectClassSize=50, title=""):
    # Create a 3x3 grid of subplots
    fig, axes = plt.subplots(3, 3, figsize=(12, 12))

    # axes is a 2D numpy array of Axes objects axes[row, col]. Pass each ax object to createPlot().
    createPlot(axes[0, 0], dfLeft, dfRight, "Berkeley", selectCounties, xAxis, yAxis, filterMetrics, selectTest, selectMinApps, selectClassSize, "")
    createPlot(axes[0, 1], dfLeft, dfRight, "UC Davis", selectCounties, xAxis, yAxis, filterMetrics, selectTest, selectMinApps, selectClassSize, "")
    createPlot(axes[0, 2], dfLeft, dfRight, "UCI", selectCounties, xAxis, yAxis, filterMetrics, selectTest, selectMinApps, selectClassSize, "")

```

```

createPlot(axes[1, 0], dfLeft, dfRight, "UCLA", selectCounties, xAxis, yAxis, filterMetrics, selectTest, selectMinApps, selectClassSize, "")
createPlot(axes[1, 1], dfLeft, dfRight, "UC Merced", selectCounties, xAxis, yAxis, filterMetrics, selectTest, selectMinApps, selectClassSize, "")
createPlot(axes[1, 2], dfLeft, dfRight, "UCR", selectCounties, xAxis, yAxis, filterMetrics, selectTest, selectMinApps, selectClassSize, "")

createPlot(axes[2, 0], dfLeft, dfRight, "UCSB", selectCounties, xAxis, yAxis, filterMetrics, selectTest, selectMinApps, selectClassSize, "")
createPlot(axes[2, 1], dfLeft, dfRight, "UCSC", selectCounties, xAxis, yAxis, filterMetrics, selectTest, selectMinApps, selectClassSize, "")
createPlot(axes[2, 2], dfLeft, dfRight, "UCSD", selectCounties, xAxis, yAxis, filterMetrics, selectTest, selectMinApps, selectClassSize, "")

fig.suptitle(title, fontsize=16)
plt.tight_layout() # Adjust layout so plots don't overlap
plt.show()

```

```

In [73]: def createPlots10(dfLeft, dfRight, selectCounties, xAxis="PassRate", yAxis="AdmRate", filterMetrics=[], selectTest="Math", selectMinApps=20, selectClassSize=50, title=""):
    # Create a 3x3 grid of subplots
    fig, axes = plt.subplots(3, 3, figsize=(10, 12))

    # axes is a 2D numpy array of Axes objects axes[row, col]. Pass each ax object to createPlot().
    createPlot(axes[0, 0], dfLeft, dfRight, "Berkeley", selectCounties, xAxis, yAxis, filterMetrics, selectTest, selectMinApps, selectClassSize, "", compact="Y")
    createPlot(axes[0, 1], dfLeft, dfRight, "UC Davis", selectCounties, xAxis, yAxis, filterMetrics, selectTest, selectMinApps, selectClassSize, "", compact="Y")
    createPlot(axes[0, 2], dfLeft, dfRight, "UCI", selectCounties, xAxis, yAxis, filterMetrics, selectTest, selectMinApps, selectClassSize, "", compact="Y")
    createPlot(axes[1, 0], dfLeft, dfRight, "UCLA", selectCounties, xAxis, yAxis, filterMetrics, selectTest, selectMinApps, selectClassSize, "", compact="Y")
    createPlot(axes[1, 1], dfLeft, dfRight, "UC Merced", selectCounties, xAxis, yAxis, filterMetrics, selectTest, selectMinApps, selectClassSize, "", compact="Y")
    createPlot(axes[1, 2], dfLeft, dfRight, "UCR", selectCounties, xAxis, yAxis, filterMetrics, selectTest, selectMinApps, selectClassSize, "", compact="Y")
    createPlot(axes[2, 0], dfLeft, dfRight, "UCSB", selectCounties, xAxis, yAxis, filterMetrics, selectTest, selectMinApps, selectClassSize, "", compact="Y")
    createPlot(axes[2, 1], dfLeft, dfRight, "UCSC", selectCounties, xAxis, yAxis, filterMetrics, selectTest, selectMinApps, selectClassSize, "", compact="Y")
    createPlot(axes[2, 2], dfLeft, dfRight, "UCSD", selectCounties, xAxis, yAxis, filterMetrics, selectTest, selectMinApps, selectClassSize, "", compact="Y")

    legendAxis = axes[4, 1]
    legendAxis.axis("off")
    legendAxis.legend([],[], frameon=False) # Remove Seaborn's automatic legend
    legendElements = [] # Build custom legend handles
    for category in categoryText:
        legendElements.append(Line2D([0], [0], marker='o', color='w', label=categoryText[category], markerfacecolor=customPalette[category], markersize=10))
    legendAxis.legend(handles=legendElements, bbox_to_anchor=(-0.1, 1), loc="upper left") # Add custom legend

    fig.suptitle(title, fontsize=16)
    plt.tight_layout() # Adjust layout so plots don't overlap
    plt.show()

```

```

In [74]: # Compare UCSD and UCI in a pair of plots
def createPlots2(dfLeft, dfRight, selectCounties, xAxis="PassRate", yAxis="AdmRate", filterMetrics=[], selectTest="Math", selectMinApps=20, selectClassSize=50, title="", height=6, width=12, campusTitles="N", legend="N", titleGap=0.96):
    # Create a 1x2 grid of subplots
    fig, axes = plt.subplots(1, 2, figsize=(width, height))

    if campusTitles == "Y":
        title1 = "UCI"
        title2 = "UCSD"
    else:
        title1 = ""
        title2 = ""

    createPlot(axes[0], dfLeft, dfRight, "UCI", selectCounties, xAxis, yAxis, filterMetrics, selectTest, selectMinApps, selectClassSize, title1)
    createPlot(axes[1], dfLeft, dfRight, "UCSD", selectCounties, xAxis, yAxis, filterMetrics, selectTest, selectMinApps, selectClassSize, title2, legend)

    fig.suptitle(title, fontsize=16, y=titleGap) # titleGap specifies gap between title and plots. Use 0.93-0.96.
    plt.tight_layout() # Adjust layout so plots don't overlap
    plt.show()

```

```

In [75]: def computeStats(dfLeft, dfRight, selectCampus, selectCounties, metrics=["App", "Adm"], filterMetrics=[["PassRate", ">=", 0.0], ["PassRate", "<=", 0.2]], selectTest="Math", selectMinApps=20, selectMinAdms=0, selectClassSize=0):
    intMetrics = ["Total Students Enrolled"]
    if selectTest in ("English", "ELA"):
        selectTestId = "1"
    else:
        selectTestId = "2" # Math

    # Filter the dataframes
    filteredLeft = dfLeft[dfLeft["Campus"] == selectCampus]
    filteredLeft = filteredLeft[filteredLeft["App"] >= selectMinApps]
    filteredLeft = filteredLeft[filteredLeft["Adm"] >= selectMinAdms]
    if selectCounties == "Local":
        filteredLeft = filteredLeft[filteredLeft["countyCampus"] == selectCampus]
    elif selectCounties != "All":
        filteredLeft = filteredLeft[filteredLeft["County"] == selectCounties]

    if "App GPA" in metrics:
        filteredLeft = filteredLeft[filteredLeft["App GPA"] >= 3.0]
    if "Adm GPA" in metrics:
        filteredLeft = filteredLeft[filteredLeft["Adm GPA"] >= 3.0]

    filteredRight = dfRight[dfRight["Test ID"] == selectTestId]
    filteredRight = filteredRight[filteredRight["Total Students Tested with Scores"].astype(int) >= selectClassSize]

    # Join
    joined = pd.merge(filteredLeft, filteredRight, left_on="CAASPP_ID", right_on="School Code", how="inner")

```

```

# Calculate additional metrics
joined[["SeniorRate"]] = joined[["Adm"]].astype(int) / joined[["Total Students Enrolled"]].astype(int) # Fraction of seniors admitted
joined[["ApplyRate"]] = joined[["App"]].astype(int) / joined[["Total Students Enrolled"]].astype(int) # Fraction of seniors who applied

# Filter out specified rows
for spec in filterMetrics:
    filterMetric = spec[0]
    filterOp     = spec[1]
    filterVal   = spec[2]
    if filterOp == "<":
        joined = joined[joined[filterMetric] <= filterVal]
    elif filterOp == "<=":
        joined = joined[joined[filterMetric] < filterVal]
    elif filterOp == ">":
        joined = joined[joined[filterMetric] >= filterVal]
    elif filterOp == ">=":
        joined = joined[joined[filterMetric] > filterVal]
    elif filterOp == "=":
        joined = joined[joined[filterMetric] == filterVal]

# Compute the stats
results = {}
for metric in metrics:
    if metric in intMetrics:
        results[metric] = joined[metric].astype(int).sum()
    else:
        results[metric] = joined[metric].sum()

return results

```

```

In [76]: # Explore overall UC application stats vs. UCSD vs. filter out schools with few applications to avoid small counts set to 0 issue
# Compare "all" to sum across campuses. The latter is smaller due to small counts set to 0 issue.
campus = "all"
result = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[], selectMinApps=0)
print(f'{campus}:20 {result["App"][:6]} {result["Adm"][:5]} {result["Enr"][:5]}')

sumEnr = 0
for campus in campuses:
    result = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[], selectMinApps=0)
    sumEnr += result["Enr"]
    print(f'{campus}:20 {result["App"][:6]} {result["Adm"][:5]} {result["Enr"][:5]}')
print(sumEnr)
print()

campus = "all"
result = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[], selectMinApps=20)
print(f'{campus}:20 {result["App"][:6]} {result["Adm"][:5]} {result["Enr"][:5]}')

sumEnr = 0
for campus in campuses:
    result = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[], selectMinApps=20)
    sumEnr += result["Enr"]
    print(f'{campus}:20 {result["App"][:6]} {result["Adm"][:5]} {result["Enr"][:5]}')
print(sumEnr)

all      115530 82315 37381
Berkeley 54891 8399 3741
UC Davis 57779 21159 4108
UCI      73802 16857 4139
UCLA     71057 6836 3882
UC Merced 23832 21440 1283
UCR      43719 32935 4424
UCSB     62929 20927 3004
UCSC     49132 30725 3048
UCSD     74272 20223 4762
32391

all      113577 80882 36858
Berkeley 53412 8031 3706
UC Davis 53396 19108 3916
UCI      72331 16418 4126
UCLA     70837 6780 3871
UC Merced 16076 15857 1110
UCR      39603 30015 4351
UCSB     59023 19045 2876
UCSC     44208 27716 2946
UCSD     71732 19158 4679
31581

In [77]: print("Acceptance Rate      App% Adm% Enr%")
for campus in campuses:
    # Denominator: All schools
    results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[])

```

```

appAll = results["App"]
admAll = results["Adm"]
enrAll = results["Enr"]

# CAASPP Pass Low vs high
results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[["PassRate", "<=", 0.25]])
app = results["App"]
adm = results["Adm"]
enr = results["Enr"]
rate = float(adm) / float(app)
appRate = app / appAll
admRate = adm / admAll
enrRate = enr / enrAll
print(f"CAASPP Pass <= 25% {campus:9} {rate:5.1%} {adm:5} {app:5} {appRate:5.1%} {admRate:5.1%} {enrRate:5.1%}")
rate0 = rate

results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[["PassRate", ">=", 0.75]])
app = results["App"]
adm = results["Adm"]
enr = results["Enr"]
if app > 0:
    rate = float(adm) / float(app)
else:
    rate = 0
rateRatio = rate0 / rate
print(f"CAASPP Pass >= 75% {campus:9} {rate:5.1%} {adm:5} {app:5} {rateRatio:.2f}")

# LCFF+ Y vs N
results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[["LCFF Plus School t", "==", "Y"]], selectMinApps=10)
app = results["App"]
adm = results["Adm"]
enr = results["Enr"]
rate = float(adm) / float(app)
appRate = app / appAll
admRate = adm / admAll
enrRate = enr / enrAll
print(f"LCFF+ Y {campus:9} {rate:5.1%} {adm:5} {app:5} {appRate:5.1%} {admRate:5.1%} {enrRate:5.1%}")
rate0 = rate

results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[["LCFF Plus School t", "==", "N"]], selectMinApps=10)
adm = results["Adm"]
app = results["App"]
enr = results["Enr"]
rate = float(adm) / float(app)
rateRatio = rate0 / rate
print(f"LCFF+ N {campus:9} {rate:5.1%} {adm:5} {app:5} {rateRatio:.2f}")

# LCFF+ = "Y"
results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[["PassRate", "<=", 0.25], ["LCFF Plus School t", "==", "Y"]])
app = results["App"]
adm = results["Adm"]
enr = results["Enr"]
rate = float(adm) / float(app)
appRate = app / appAll
admRate = adm / admAll
enrRate = enr / enrAll
print(f"LCFF+, Pass <= 25% {campus:9} {rate:5.1%} {adm:5} {app:5} {appRate:5.1%} {admRate:5.1%} {enrRate:5.1%}")
rate0 = rate

results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[["PassRate", ">=", 0.50], ["LCFF Plus School t", "==", "Y"]])
app = results["App"]
adm = results["Adm"]
enr = results["Enr"]
if app > 0:
    rate = float(adm) / float(app)
    rateRatio = rate0 / rate
else:
    rate = 0
    rateRatio = 0
print(f"LCFF+, Pass >= 50% {campus:9} {rate:5.1%} {adm:5} {app:5} {rateRatio:.2f}")

# LCFF+ = "N"
results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[["PassRate", "<=", 0.25], ["LCFF Plus School t", "==", "N"]])
app = results["App"]
adm = results["Adm"]
enr = results["Enr"]
rate = float(adm) / float(app)
appRate = app / appAll
admRate = adm / admAll
enrRate = enr / enrAll
print(f"Not LCFF+, Pass <= 25% {campus:9} {rate:5.1%} {adm:5} {app:5} {appRate:5.1%} {admRate:5.1%} {enrRate:5.1%}")
rate0 = rate

results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[["PassRate", ">=", 0.75], ["LCFF Plus School t", "==", "N"]])

```

```
app = results["App"]
adm = results["Adm"]
enr = results["Enr"]
if app > 0:
    rate = float(adm) / float(app)
else:
    rate = 0
rateRatio = rate0 / rate
print(f"Not LCF+{Pass} >= 75% {campus:9} {rate:5.1%} {adm:5} {app:5} {rateRatio:.2f}")
print()
```

			Acceptance Rate			App%	Adm%	Enr%
CAASPP Pass <= 25%	Berkeley	16.5%	1949	11809		22.1%	24.3%	18.1%
CAASPP Pass >= 75%	Berkeley	14.5%	538	3722	1.14			
LCFF+ Y	Berkeley	17.4%	1884	10815		20.2%	23.5%	14.2%
LCFF+ N	Berkeley	14.7%	6484	43997	1.18			
LCFF+, Pass <= 25%	Berkeley	16.6%	1166	7014		13.1%	14.5%	10.0%
LCFF+, Pass >= 50%	Berkeley	10.5%	29	277	1.59			
Not LCFF+, Pass <= 25%	Berkeley	16.3%	783	4795		9.0%	9.7%	8.1%
Not LCFF+, Pass >= 75%	Berkeley	14.5%	538	3722	1.13			
CAASPP Pass <= 25%	UC Davis	42.4%	5242	12368		23.2%	27.4%	28.8%
CAASPP Pass >= 75%	UC Davis	23.5%	882	3747	1.80			
LCFF+ Y	UC Davis	43.3%	4779	11049		20.7%	25.0%	16.5%
LCFF+ N	UC Davis	34.8%	16046	46114	1.24			
LCFF+, Pass <= 25%	UC Davis	43.2%	2830	6551		12.3%	14.8%	11.7%
LCFF+, Pass >= 50%	UC Davis	44.6%	78	175	0.97			
Not LCFF+, Pass <= 25%	UC Davis	41.5%	2412	5817		10.9%	12.6%	17.2%
Not LCFF+, Pass >= 75%	UC Davis	23.5%	882	3747	1.76			
CAASPP Pass <= 25%	UCI	22.4%	4511	20114		27.8%	27.5%	24.0%
CAASPP Pass >= 75%	UCI	25.5%	1039	4078	0.88			
LCFF+ Y	UCI	23.6%	4810	20401		28.2%	29.3%	28.2%
LCFF+ N	UCI	22.5%	11994	53273	1.05			
LCFF+, Pass <= 25%	UCI	22.7%	3121	13777		19.0%	19.0%	18.1%
LCFF+, Pass >= 50%	UCI	29.6%	141	476	0.76			
Not LCFF+, Pass <= 25%	UCI	21.9%	1390	6337		8.8%	8.5%	6.0%
Not LCFF+, Pass >= 75%	UCI	25.5%	1039	4078	0.86			
CAASPP Pass <= 25%	UCLA	18.2%	1859	18152		25.6%	27.4%	25.5%
CAASPP Pass >= 75%	UCLA	8.5%	350	4109	1.20			
LCFF+ Y	UCLA	18.5%	1819	17386		24.5%	26.8%	25.4%
LCFF+ N	UCLA	9.3%	5014	53665	1.12			
LCFF+, Pass <= 25%	UCLA	18.2%	1220	11973		16.9%	18.0%	16.0%
LCFF+, Pass >= 50%	UCLA	18.1%	40	398	1.01			
Not LCFF+, Pass <= 25%	UCLA	18.3%	639	6179		8.7%	9.4%	9.5%
Not LCFF+, Pass >= 75%	UCLA	8.5%	350	4109	1.21			
CAASPP Pass <= 25%	UC Merced	91.4%	4923	5388		33.5%	32.7%	51.4%
CAASPP Pass >= 75%	UC Merced	96.8%	1293	1336	0.94			
LCFF+ Y	UC Merced	91.3%	6740	7386		45.9%	44.8%	45.9%
LCFF+ N	UC Merced	94.3%	13045	13832	0.97			
LCFF+, Pass <= 25%	UC Merced	91.5%	3464	3786		23.6%	23.0%	36.4%
LCFF+, Pass >= 50%	UC Merced	96.2%	50	52	0.95			
Not LCFF+, Pass <= 25%	UC Merced	91.1%	1459	1602		10.0%	9.7%	15.0%
Not LCFF+, Pass >= 75%	UC Merced	96.8%	1293	1336	0.94			
CAASPP Pass <= 25%	UCR	67.8%	8175	12060		30.5%	27.2%	29.4%
CAASPP Pass >= 75%	UCR	82.6%	2086	2545	0.83			
LCFF+ Y	UCR	67.5%	9785	14384		36.3%	32.3%	27.7%
LCFF+ N	UCR	79.6%	22349	28089	0.85			
LCFF+, Pass <= 25%	UCR	66.7%	6232	9339		23.6%	20.8%	21.3%
LCFF+, Pass >= 50%	UCR	72.9%	186	255	0.91			
Not LCFF+, Pass <= 25%	UCR	71.4%	1943	2721		6.9%	6.5%	8.1%
Not LCFF+, Pass >= 75%	UCR	82.6%	2086	2545	0.87			
CAASPP Pass <= 25%	UCSB	37.8%	4689	12407		21.0%	24.6%	19.5%
CAASPP Pass >= 75%	UCSB	29.9%	1172	3915	1.26			
LCFF+ Y	UCSB	38.3%	4628	12087		20.5%	24.3%	11.8%
LCFF+ N	UCSB	31.7%	15900	50183	1.21			
LCFF+, Pass <= 25%	UCSB	38.2%	2656	6953		11.8%	13.9%	7.3%
LCFF+, Pass >= 50%	UCSB	31.9%	74	232	1.20			
Not LCFF+, Pass <= 25%	UCSB	37.3%	2033	5454		9.2%	10.7%	12.2%
Not LCFF+, Pass >= 75%	UCSB	29.9%	1172	3915	1.25			
CAASPP Pass <= 25%	UCSC	59.4%	5999	10099		22.8%	21.6%	16.0%
CAASPP Pass >= 75%	UCSC	62.7%	1946	3106	0.95			
LCFF+ Y	UCSC	54.9%	5702	10378		23.5%	20.6%	8.6%
LCFF+ N	UCSC	64.5%	24424	37873	0.85			
LCFF+, Pass <= 25%	UCSC	55.4%	3212	5794		13.1%	11.6%	5.7%
LCFF+, Pass >= 50%	UCSC	61.6%	77	125	0.90			
Not LCFF+, Pass <= 25%	UCSC	64.7%	2787	4305		9.7%	10.1%	10.3%
Not LCFF+, Pass >= 75%	UCSC	62.7%	1946	3106	1.03			
CAASPP Pass <= 25%	UCSD	37.1%	6695	18055		25.2%	34.9%	31.8%
CAASPP Pass >= 75%	UCSD	16.8%	791	4177	2.21			
LCFF+ Y	UCSD	42.1%	7381	17532		24.4%	38.5%	35.1%
LCFF+ N	UCSD	22.5%	12675	56429	1.87			
LCFF+, Pass <= 25%	UCSD	41.5%	4723	11371		15.9%	24.7%	22.9%
LCFF+, Pass >= 50%	UCSD	46.5%	171	368	0.89			
Not LCFF+, Pass <= 25%	UCSD	29.5%	1972	6684		9.3%	10.3%	8.9%
Not LCFF+, Pass >= 75%	UCSD	16.8%	791	4177	1.76			

```
In [78]: # Campus LowAccept% HighAccept% Ratio NotLCFF+LowAccept% NotLCFF+HighAccept% Ratio
print("          Public Schools      Non-LCFF+ Schools      LCFF+ Schools      Private")
print("Campus     All  LowMath HighMath Ratio  LowMath HighMath Ratio  LowMath HighMath Ratio  All")
for campus in campuses:
    # Denominators
    results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[])
    app = results["App"]
    adm = results["Adm"]
    rateAll = float(adm) / float(app)

    # CAASPP Pass Low vs high
    results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[["PassRate", "<=", 0.25]])
    app = results["App"]
    adm = results["Adm"]
    rateLow = float(adm) / float(app)

    results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[["PassRate", ">", 0.75]])
    app = results["App"]
    adm = results["Adm"]
    rateHigh = float(adm) / float(app)
    rateRatio = rateHigh / rateLow

    # LCFF+ = "N"
    results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[["PassRate", "<=", 0.25], ["LCFF Plus School t", "==", "N"]])
    app = results["App"]
    adm = results["Adm"]
    rateLowNotLCFF = float(adm) / float(app)

    results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[["PassRate", ">", 0.75], ["LCFF Plus School t", "==", "N"]])
    app = results["App"]
    adm = results["Adm"]
    rateHighNotLCFF = float(adm) / float(app)
    rateRatioNotLCFF = rateHighNotLCFF / rateLowNotLCFF

    # LCFF+ = "Y"
    results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[["PassRate", "<=", 0.25], ["LCFF Plus School t", "==", "Y"]])
    app = results["App"]
    adm = results["Adm"]
    rateLowLCFF = float(adm) / float(app)

    results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[["PassRate", ">", 0.50], ["LCFF Plus School t", "==", "Y"]])
    app = results["App"]
    adm = results["Adm"]
    rateHighLCFF = float(adm) / float(app)
    rateRatioLCFF = rateHighLCFF / rateLowLCFF

    dfLeftPrivate[campus] = dfLeftPrivate[dfLeftPrivate["Campus"] == campus]
    app = dfLeftPrivateCampus["App"].astype(int).sum()
    adm = dfLeftPrivateCampus["Adm"].astype(int).sum()
    ratePrivate = adm / app

    print(f'{campus}: {rateAll:.5f} {rateLow:.5f} {rateHigh:.5f} {rateRatio:.2f} {rateLowNotLCFF:.5f} {rateHighNotLCFF:.5f} {rateRatioNotLCFF:.2f} {rateLowLCFF:.5f} {rateHighLCFF:.5f} {rateRatioLCFF:.2f} {ratePrivate:.1f}')


    Public Schools      Non-LCFF+ Schools      LCFF+ Schools      Private
Campus     All  LowMath HighMath Ratio  LowMath HighMath Ratio  LowMath HighMath Ratio  All
Berkeley   15.0% 16.5% 14.5% 0.88   16.3% 14.5% 0.89   16.6% 10.5% 0.63   18.2%
UC Davis   35.8% 42.4% 23.5% 0.56   41.5% 23.5% 0.57   43.2% 44.6% 1.03   36.9%
UCI        22.7% 22.4% 25.5% 1.14   21.9% 25.5% 1.16   22.7% 29.6% 1.31   23.1%
UCLA       9.6% 10.2% 8.5% 0.83   10.3% 8.5% 0.82   10.2% 10.1% 0.99   11.2%
UC Merced  93.7% 91.4% 96.8% 1.06   91.1% 96.8% 1.06   91.5% 96.2% 1.05   92.5%
UCR        75.8% 67.8% 82.0% 1.21   71.4% 82.0% 1.15   66.7% 72.9% 1.09   76.1%
UCSB       32.3% 37.8% 29.9% 0.79   37.3% 29.9% 0.80   38.2% 31.9% 0.84   29.0%
UCSC       62.7% 59.4% 62.7% 1.05   64.7% 62.7% 0.97   55.4% 61.6% 1.11   64.7%
UCSD       26.7% 37.1% 16.8% 0.45   29.5% 16.8% 0.57   41.5% 46.5% 1.12   21.4%
```

```
In [79]: # What fraction of admitted students enroll?
print(" Enrollment Rate by Math CAASPP Group")
print("Campus     All  Low    Mid    High")
for campus in campuses:
    results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[], selectMinAdms=10)
    app = results["App"]
    adm = results["Adm"]
    enr = results["Enr"]
    rateAll = float(enr) / float(adm)

    results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[["PassRate", "<=", 0.25]], selectMinAdms=10)
    app = results["App"]
    adm = results["Adm"]
    enr = results["Enr"]
    rateLow = float(enr) / float(adm)

    results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[["PassRate", ">", 0.25], ["PassRate", "<", 0.75]], selectMinAdms=10)
    app = results["App"]
    adm = results["Adm"]
```

```

enr = results["Enr"]
rateMed = float(enr) / float(adm)

results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[["PassRate", ">=", 0.75]], selectMinAdms=10)
app = results["App"]
adm = results["Adm"]
enr = results["Enr"]
rateHigh = float(enr) / float(adm)

print(f"\n{campus}:9 {rateAll:5.1%} {rateLow:5.1%} {rateMed:5.1%} {rateHigh:5.1%}")

```

Campus	All	Low	Mid	High
Berkeley	53.3%	46.4%	53.4%	61.7%
UC Davis	21.1%	23.0%	20.8%	14.9%
UCI	27.1%	25.5%	27.7%	26.4%
UCLA	60.5%	63.4%	61.1%	50.5%
UC Merced	7.4%	11.6%	5.6%	3.4%
UCR	14.5%	15.7%	14.4%	10.7%
UCSB	15.4%	12.5%	16.8%	11.7%
UCSC	18.7%	8.8%	11.8%	7.6%
UCSD	25.4%	23.6%	26.6%	22.2%

In [81]: # Check composition of UC Campus student body: what fraction of applicants and enrolled students from each school perf type.

```

Composition by Math CAASPP Group
print("          nApplications    %Apps    nEnrolled    %Enr")
print("Campus      All   Low   High   All   Low   High   Low   Mid   High")
for campus in campuses:
    # Denominators
    results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[], selectMinApps=0)
    appAll = results["App"]
    adm = results["Adm"]
    enrAll = results["Enr"]

    # CAASPP Pass Low to High
    results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[["PassRate", "<=", 0.25]], selectMinApps=0)
    appLow = results["App"]
    adm = results["Adm"]
    enrLow = results["Enr"]

    results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[["PassRate", ">", 0.25], ["PassRate", "<", 0.75]], selectMinApps=0)
    appMid = results["App"]
    adm = results["Adm"]
    enrMid = results["Enr"]

    results = computeStats(dfLeft, dfRight, campus, "All", metrics=["App", "Adm", "Enr"], filterMetrics=[["PassRate", ">=", 0.75]], selectMinApps=0)
    appHigh = results["App"]
    adm = results["Adm"]
    enrHigh = results["Enr"]

    appHighRate = float(appHigh) / float(appAll)
    enrLowRate = float(enrLow) / float(enrAll)
    enrMidRate = float(enrMid) / float(enrAll)
    enrHighRate = float(enrHigh) / float(enrAll)

    print(f"\n{campus}:9 {appAll:5} {appLow:5} {appHigh:4} {appHighRate:5.1%} {enrAll:4} {enrLow:4} {enrHigh:4} {enrLowRate:5.1%} {enrMidRate:5.1%} {enrHighRate:5.2%}")

    Composition by Math CAASPP Group
    print("          nApplications    %Apps    nEnrolled    %Enr")
    print("Campus      All   Low   High   All   Low   High   Low   Mid   High")

```

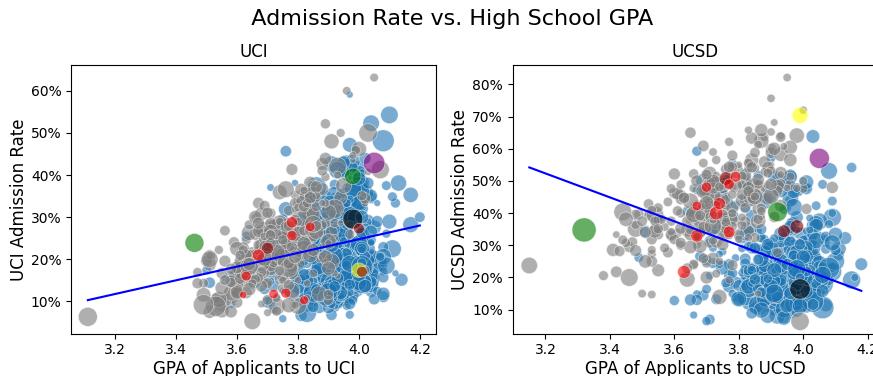
Campus	All	Low	High	All	Low	High	Low	Mid	High	
Berkeley	54891	12897	3722	6.8%	3741	700	330	18.7%	72.5%	8.82%
UC Davis	57779	15523	3764	6.5%	4108	1270	131	30.9%	65.9%	3.19%
UCI	73802	21334	4078	5.5%	4139	1002	269	24.2%	69.3%	6.50%
UCLA	71057	18303	4109	5.8%	3882	991	173	25.5%	70.0%	4.46%
UC Merced	23032	9287	1407	6.1%	1283	676	44	52.7%	43.9%	3.43%
UCR	43719	14711	2585	5.9%	4424	1332	223	30.1%	64.9%	5.04%
UCSB	62929	15342	3915	6.2%	3004	666	136	22.2%	73.3%	4.53%
UCSC	49132	13705	3123	6.4%	3048	531	147	17.4%	77.8%	4.82%
UCSD	74272	20091	4177	5.6%	4762	1542	151	32.4%	64.4%	3.17%

In [82]: ##### Conclusion plots

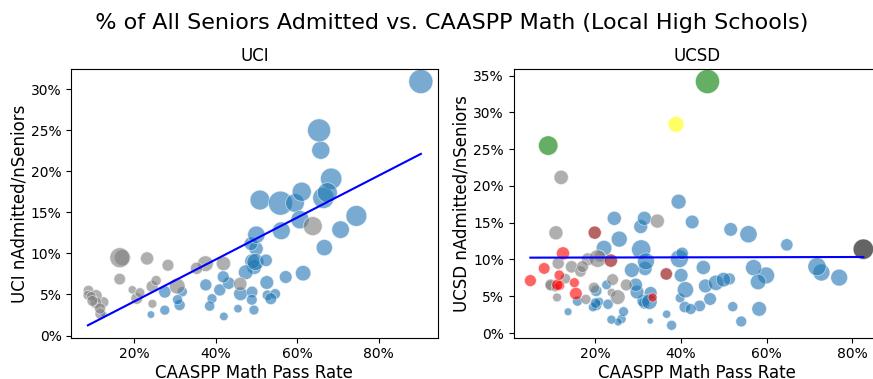
```

# createPlots2(dfLeft, dfRight, "All",
createPlots2(dfLeft, dfRight, "All",   height=4, width=13.49, campusTitles="Y", title="Admission Rate vs CAASPP Math, All High Schools
                                         , Legend="Y")
createPlots2(dfLeft, dfRight, "All",   height=4, width=9,   campusTitles="Y", title=" Admission Rate vs. High School GPA")

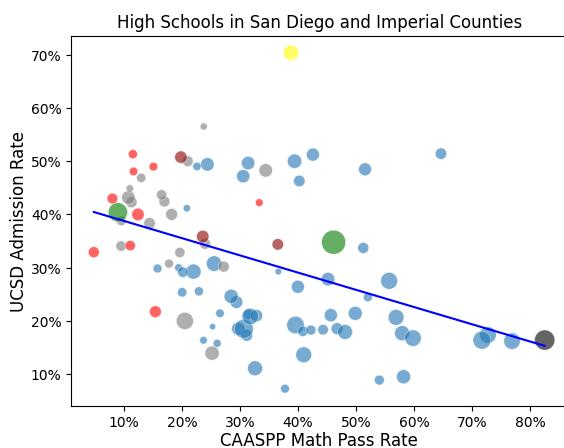
```

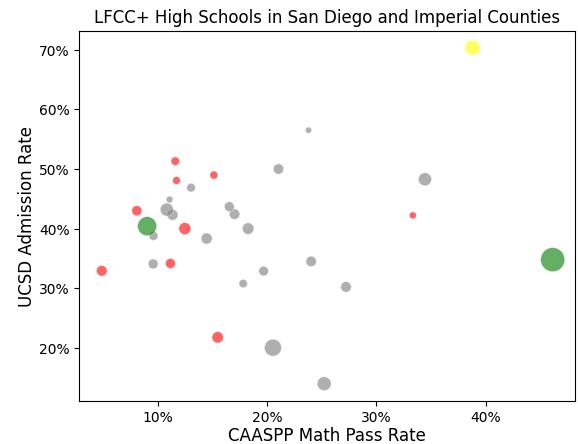


```
In [83]: createPlots2(dfLeft, dfRight, "Local", yAxis="SeniorRate", selectMinApps=0, height=4, width=9, campusTitles="Y", title=" % of All Seniors Admitted vs. CAASPP Math (Local High Schools)")
```

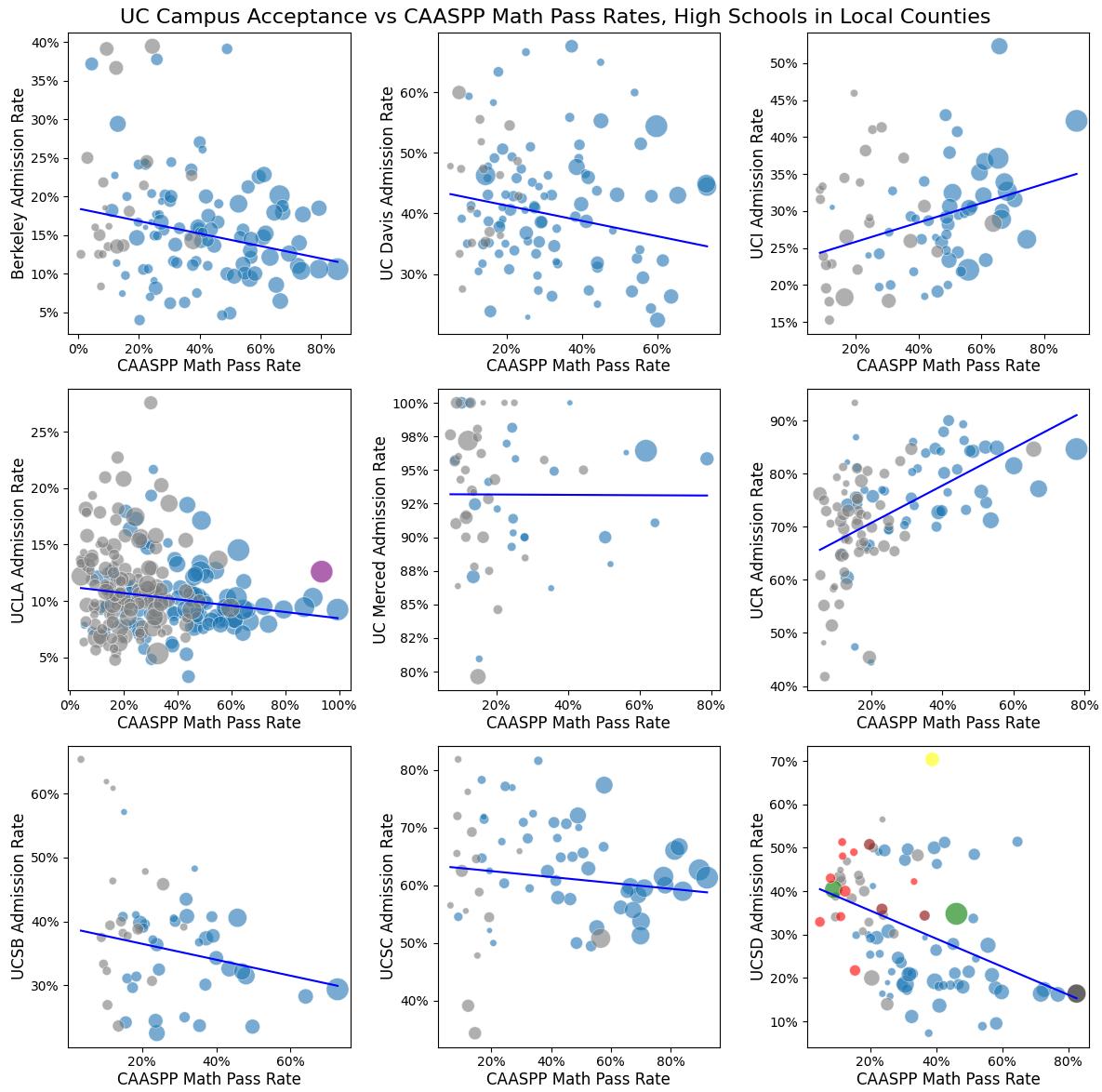


```
In [84]: fig, ax = plt.subplots() # Create a Matplotlib Axes object
createPlot(ax, dfLeft, dfRight, "UCSD", "Local", title="High Schools in San Diego and Imperial Counties")
plt.show()
fig, ax = plt.subplots() # Create a Matplotlib Axes object
createPlot(ax, dfLeft, dfRight, "UCSD", "Local", filterMetrics=[["LCFF Plus School t", "==", "Y"]], title="LFCC+ High Schools in San Diego and Imperial Counties")
plt.show()
```

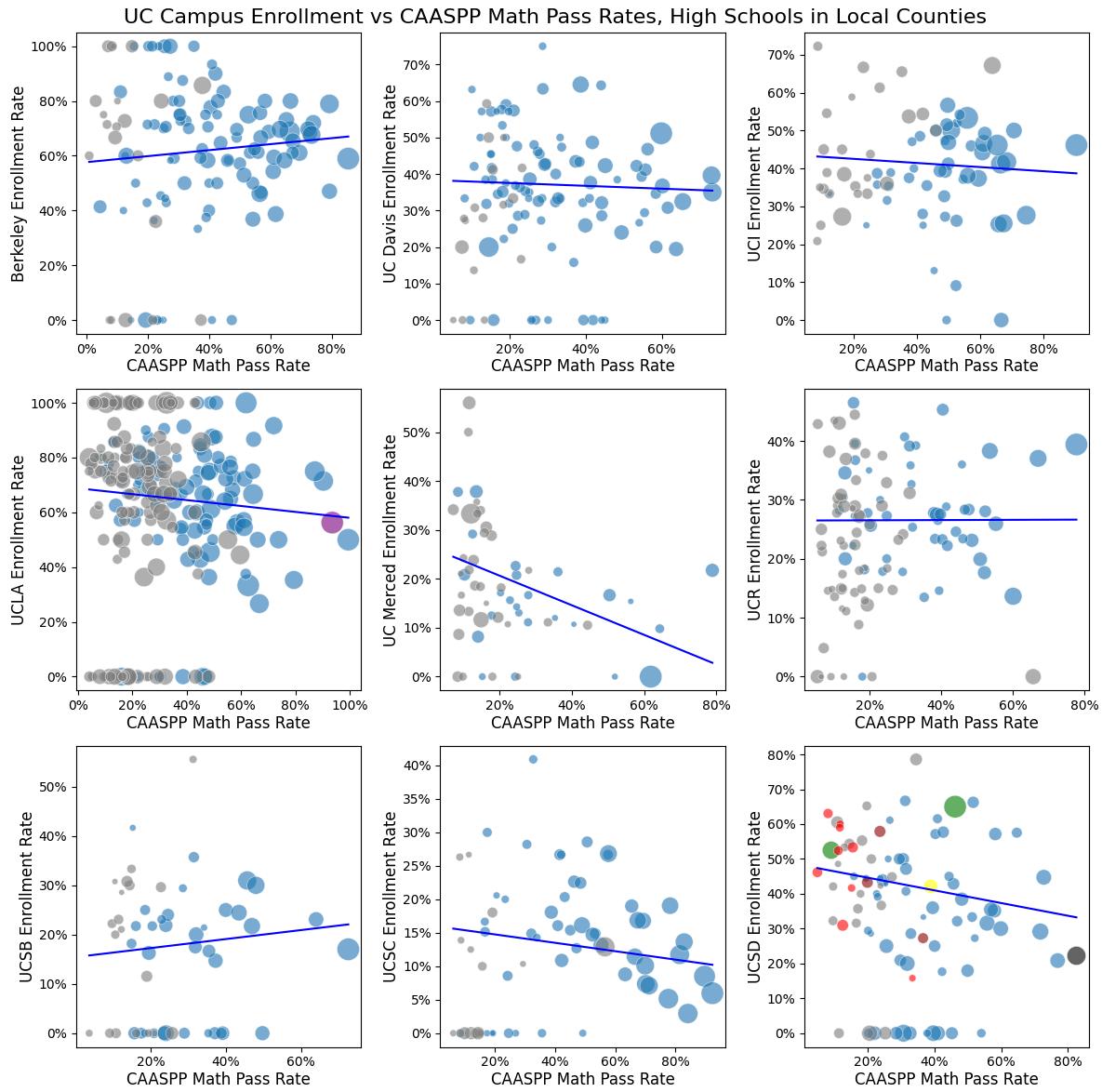




In [85]: `createPlots(dfLeft, dfRight, "Local", title="UC Campus Acceptance vs CAASPP Math Pass Rates, High Schools in Local Counties")`



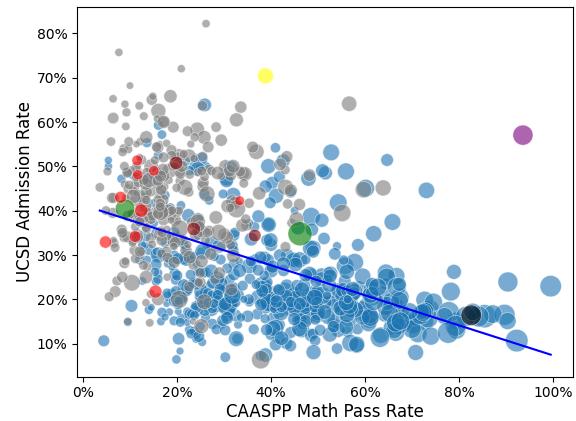
```
In [86]: createPlots(dfLeft, dfRight, "Local", yAxis="EnrRate", title="UC Campus Enrollment vs CAASPP Math Pass Rates, High Schools in Local Counties")
```



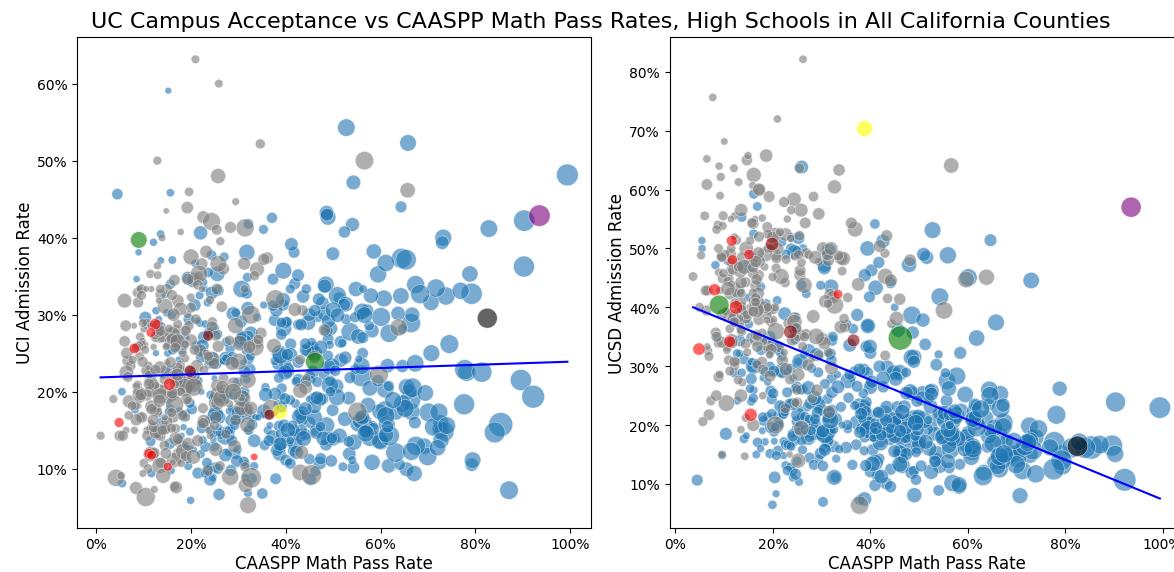
```
In [87]: fig, ax = plt.subplots() # Create a Matplotlib Axes object
createPlot(ax, dfLeft, dfRight, "UCSD", "All", title="")

# plt.Legend([],[], frameon=False) # Remove Seaborn's automatic Legend
# legendElements = [] # Build custom Legend handles
# for category in categoryTextNarrow:
#     legendElements.append(Line2D([0], [0], marker='o', color='w', label=categoryTextNarrow[category], markerfacecolor=customPalette[category], markersize=10))
# plt.legend(handles=legendElements, bbox_to_anchor=(1.0, 1), loc="upper left") # Add custom Legend

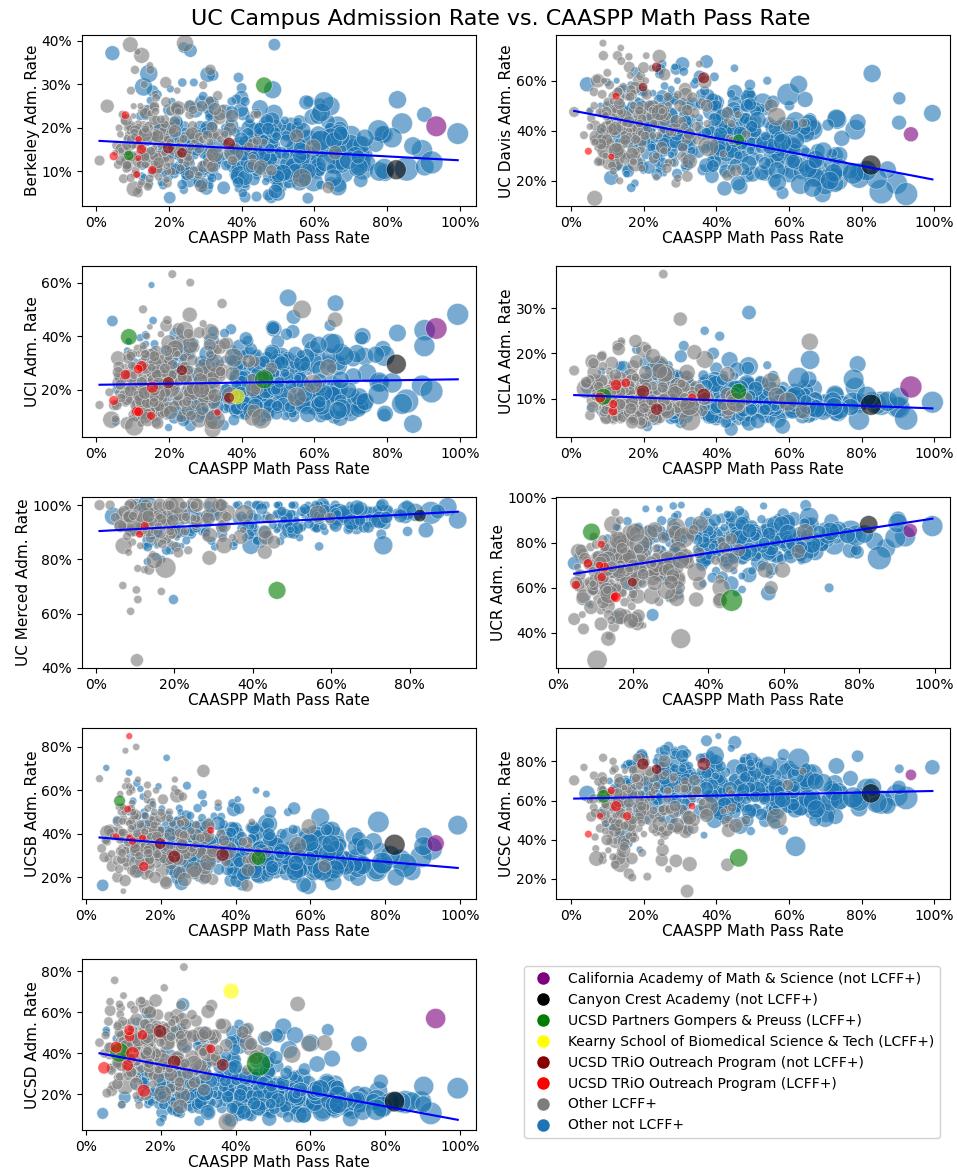
plt.show()
```



```
In [88]: createPlots2(dfLeft, dfRight, "All", title="UC Campus Acceptance vs CAASPP Math Pass Rates, High Schools in All California Counties")
```

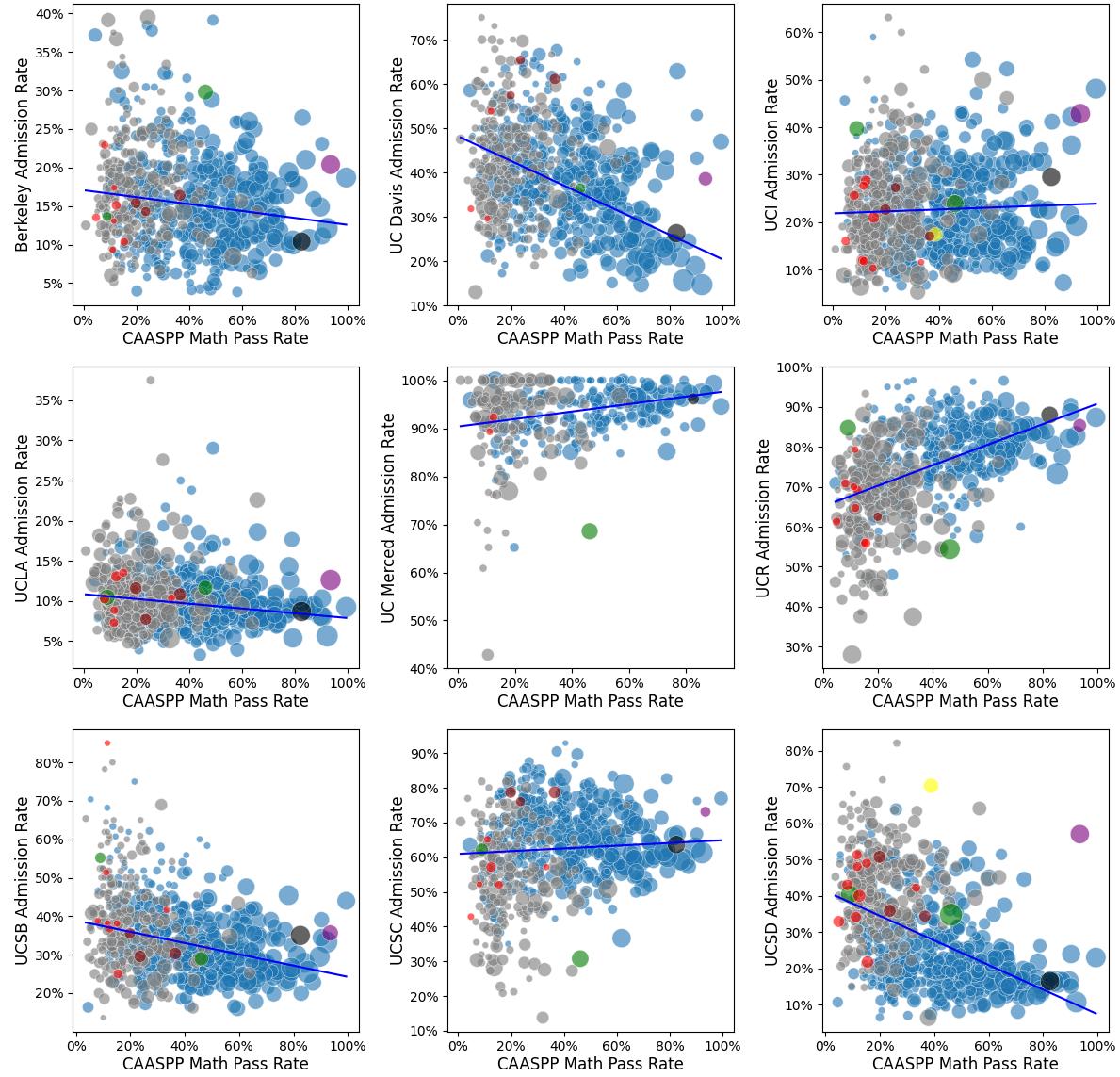


```
In [89]: # 2x5
createPlots10(dfLeft, dfRight, "All", title="UC Campus Admission Rate vs. CAASPP Math Pass Rate")
```



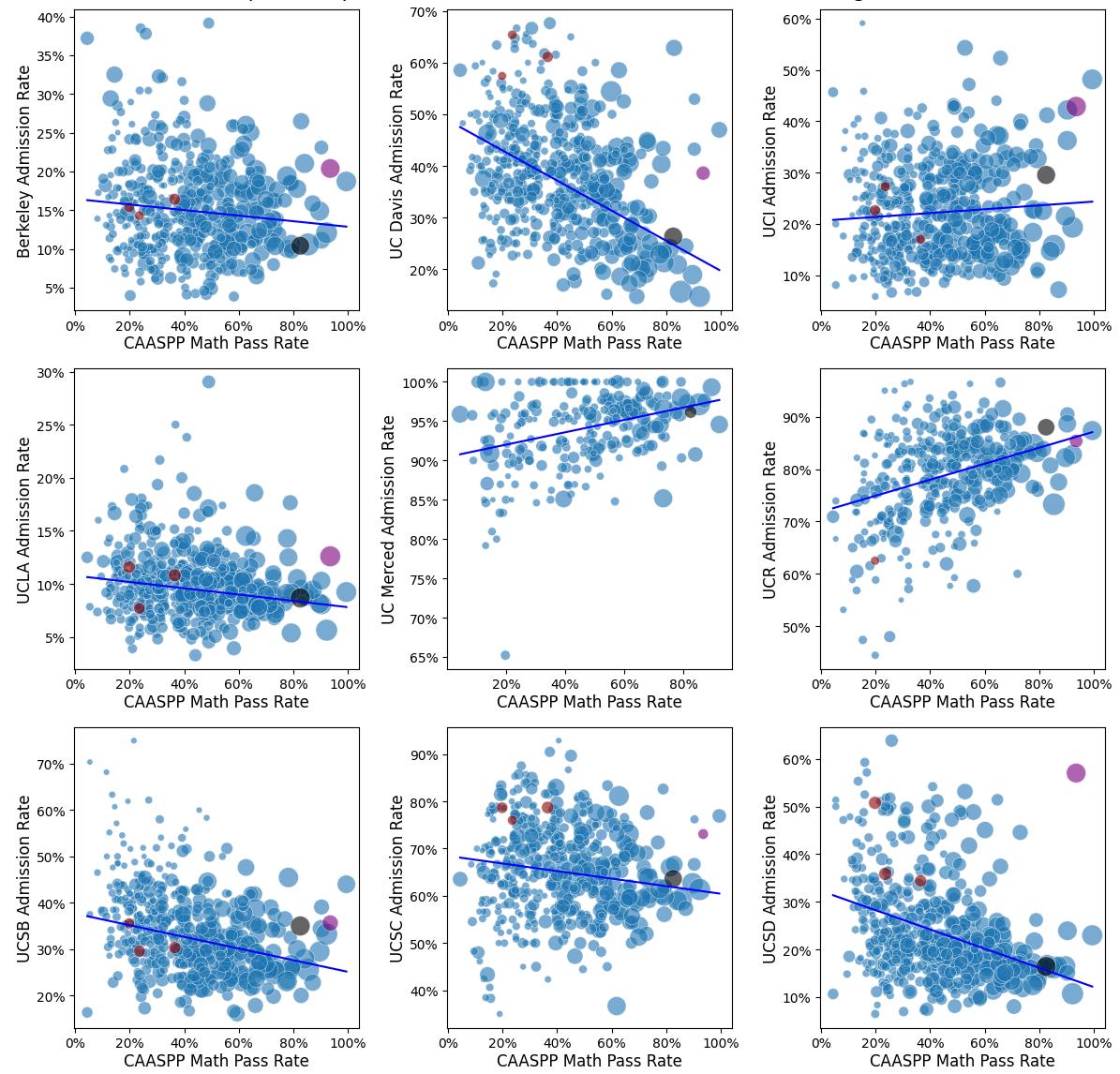
```
In [90]: # 3x3
createPlots(dfLeft, dfRight, "All",
            title="UC Campus Acceptance vs CAASPP Math Pass Rates, High Schools in All California Counties")
```

UC Campus Acceptance vs CAASPP Math Pass Rates, High Schools in All California Counties

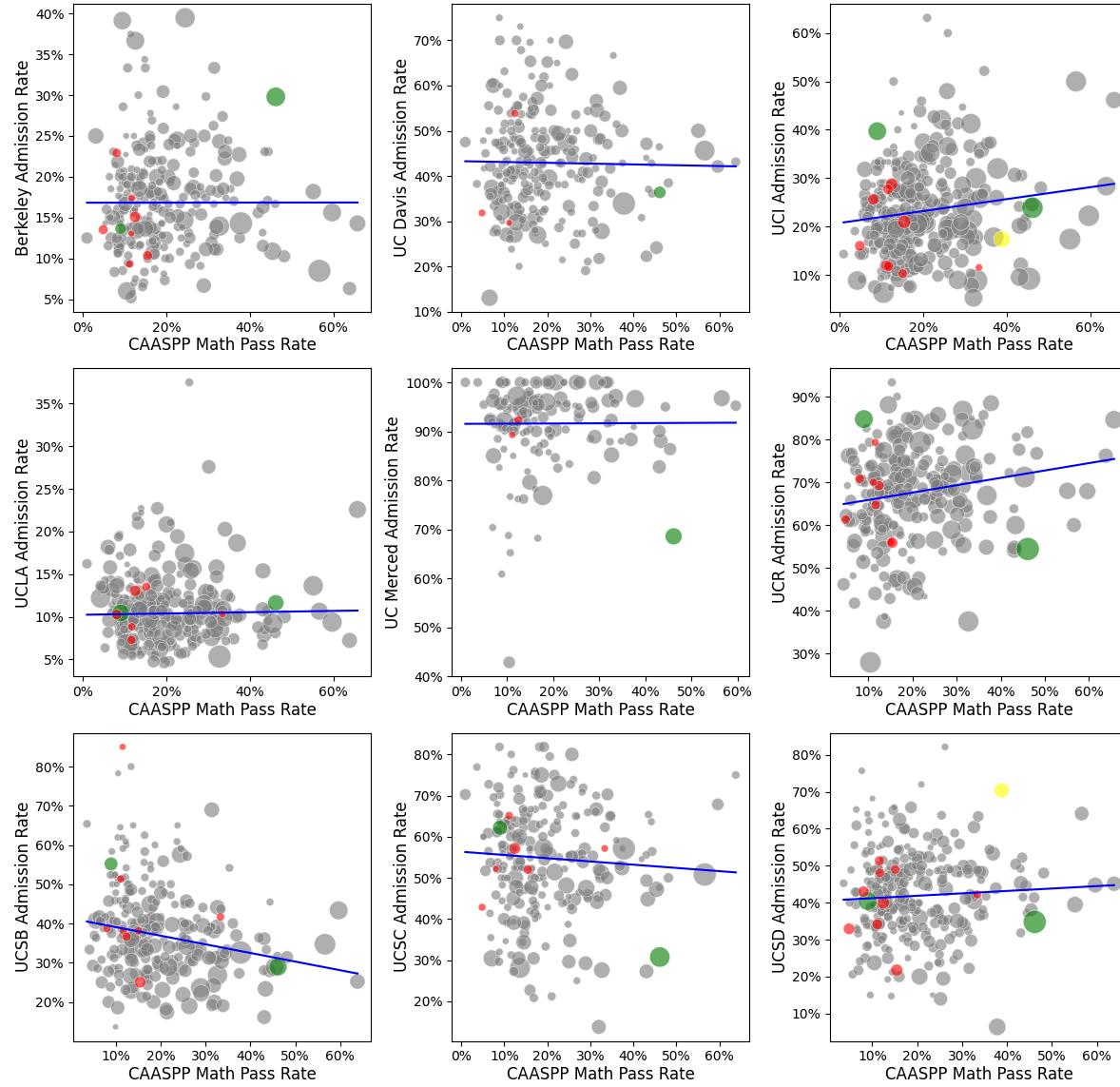


```
In [91]: # Look separately at LCFF+ schools and only non-LCFF+ schools
createPlots(dfLeft, dfRight, "All", filterMetrics=[["LCFF Plus School t", "==", "N"]], title="UC Campus Acceptance vs CAASPP Math Pass Rates, No LCFF+ High Schools")
createPlots(dfLeft, dfRight, "All", filterMetrics=[["LCFF Plus School t", "==", "Y"]], title="UC Campus Acceptance vs CAASPP Math Pass Rates, Only LCFF+ High Schools")
```

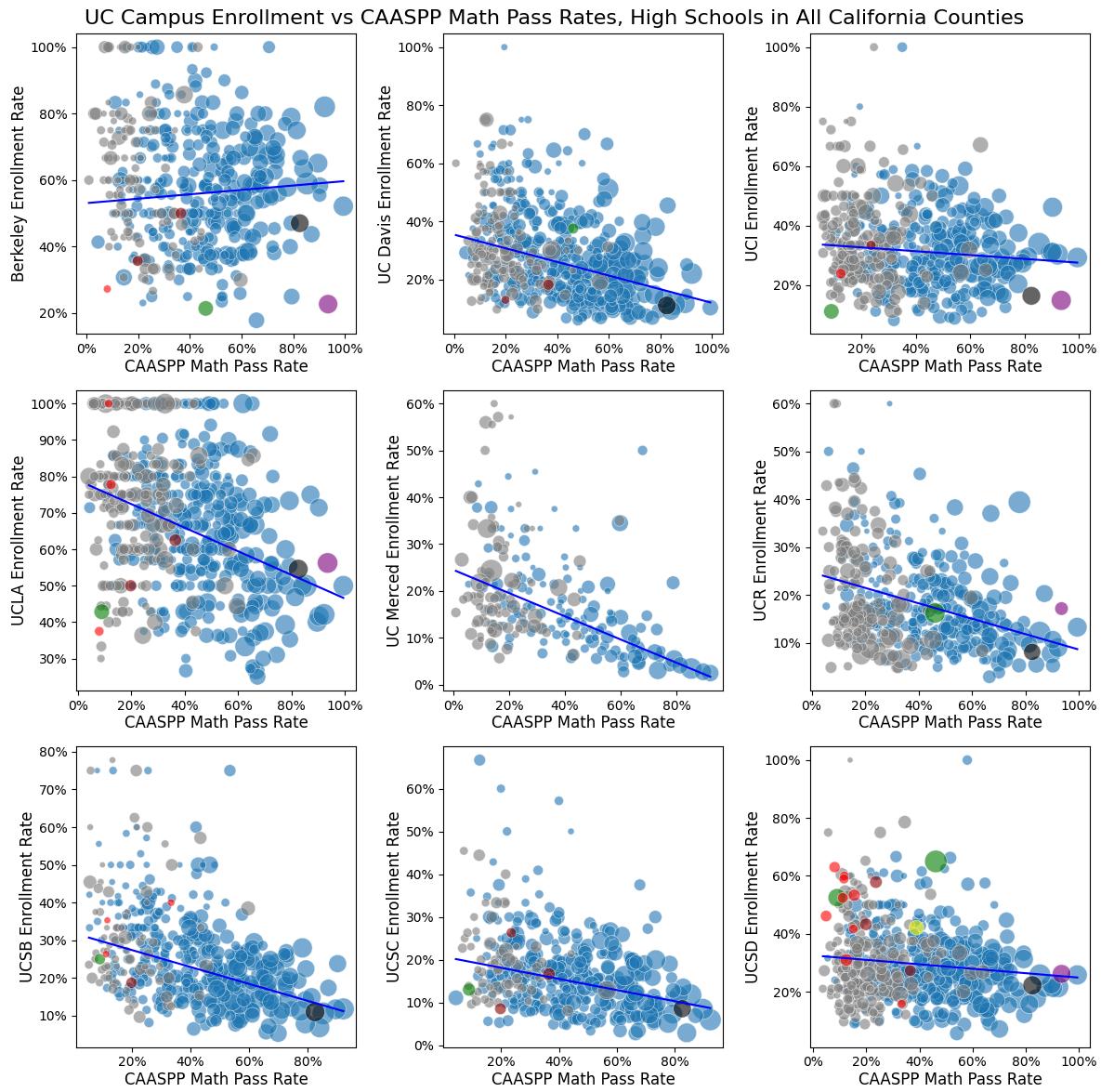
UC Campus Acceptance vs CAASPP Math Pass Rates, No LCFF+ High Schools



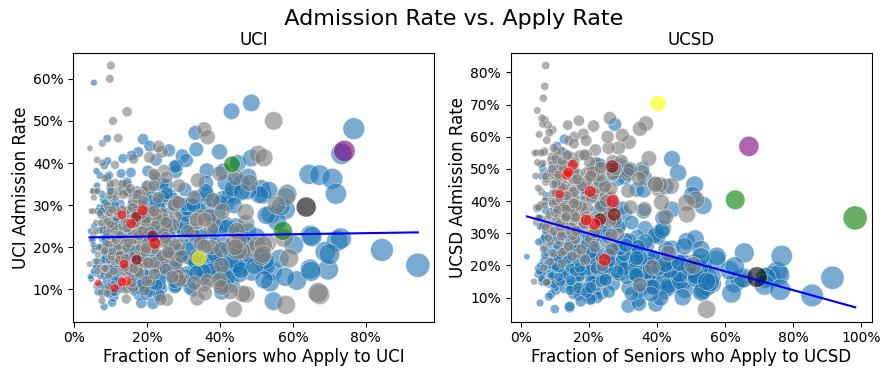
UC Campus Acceptance vs CAASPP Math Pass Rates, Only LCFF+ High Schools



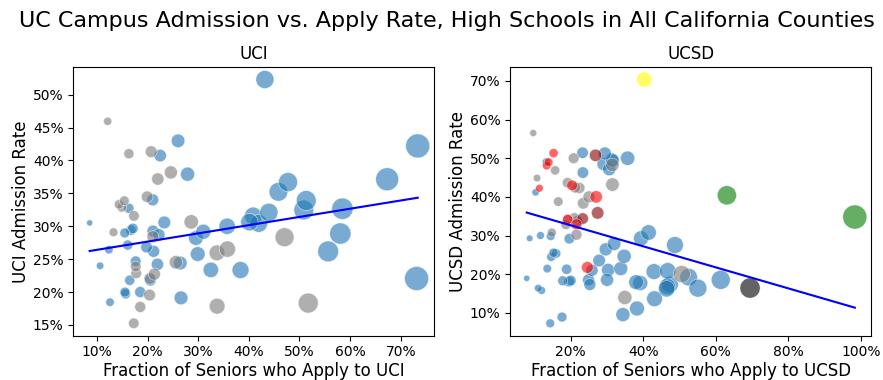
```
In [92]: createPlots(dfLeft, dfRight, "All", yAxis="EnrRate", filterMetrics=[[{"Enr": ">=", "3"}], selectMinApps=0, title="UC Campus Enrollment vs CAASPP Math Pass Rates, High Schools in All California Counties")
```



```
In [93]: # Acceptance rate vs fraction who apply
createPlots2(dfLeft, dfRight, "All", xAxis="ApplyRate", height=4, width=9, campusTitles="Y", title=" Admission Rate vs. Apply Rate", titleGap=0.93)
```



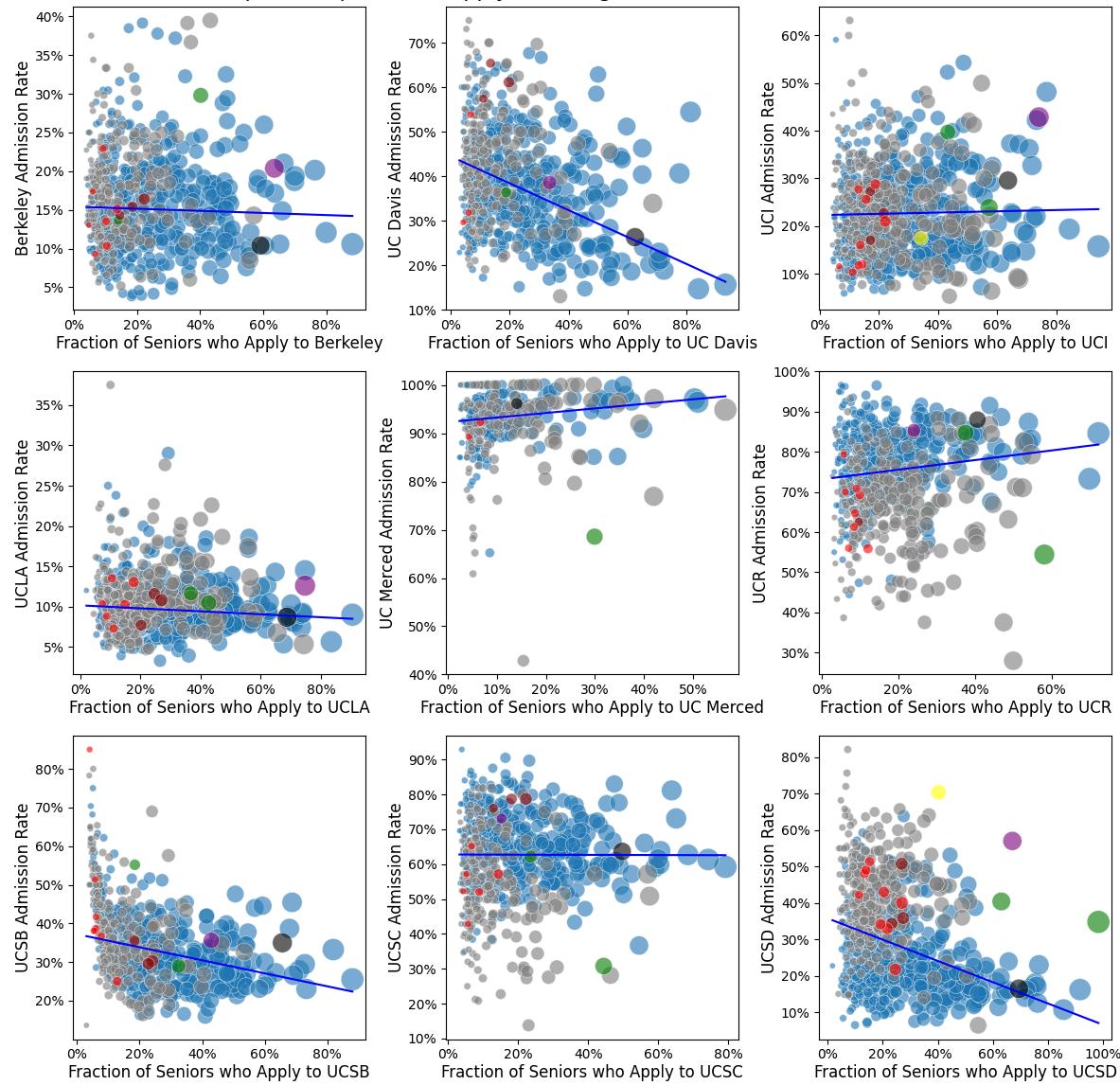
```
In [94]: createPlots2(dfLeft, dfRight, "Local", xAxis="ApplyRate", height=4, width=9, campusTitles="Y", title="UC Campus Admission vs. Apply Rate, High Schools in All California Counties")
```

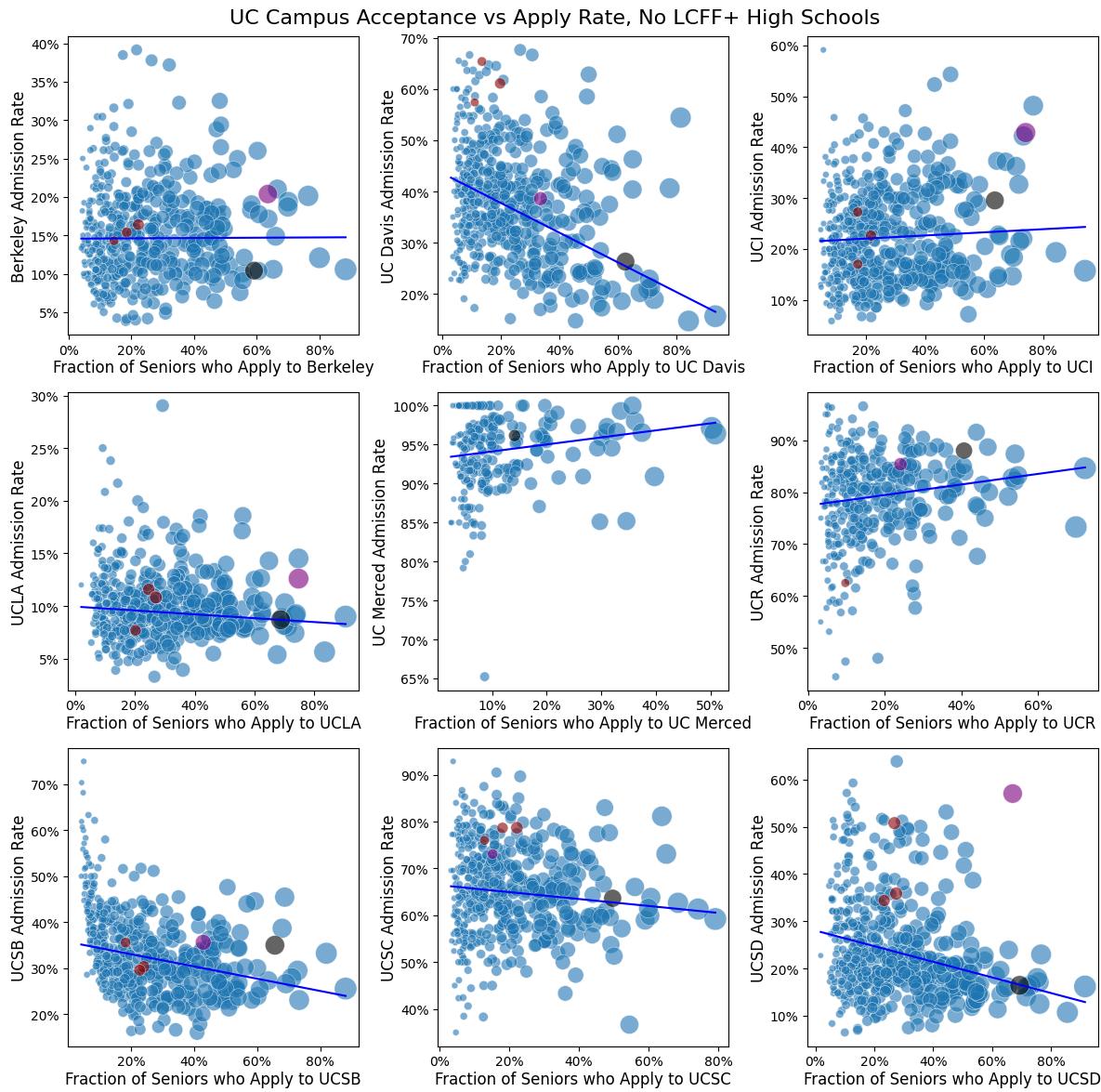


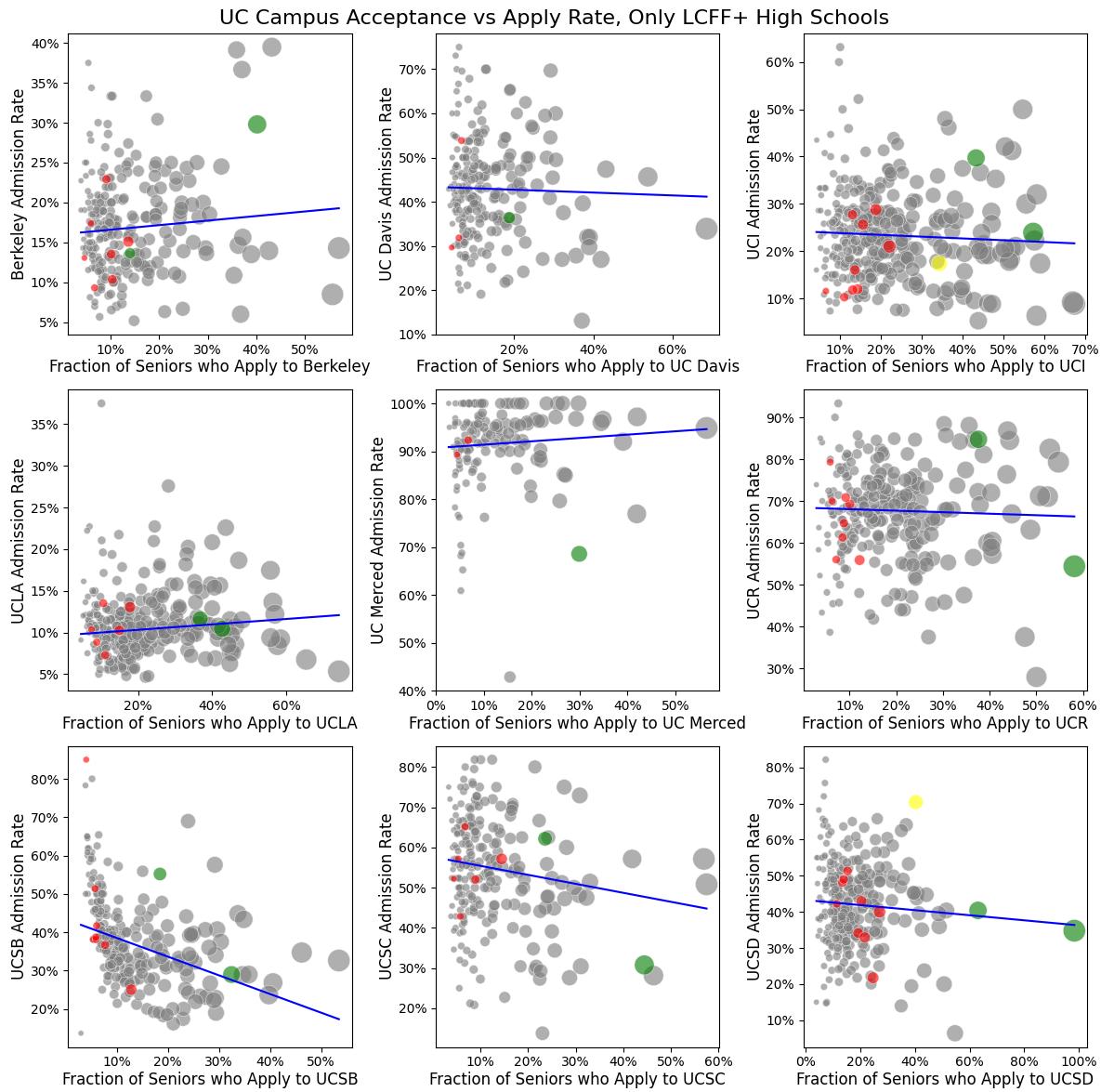
```
In [95]: # Acceptance rate vs fraction who apply
createPlots(dfLeft, dfRight, "All", xAxis="ApplyRate",
            title="UC Campus Acceptance vs Apply Rate, High Schools in All California Counties")

createPlots(dfLeft, dfRight, "All", xAxis="ApplyRate", filterMetrics=[["LCFF Plus School t", "==", "N"]], title="UC Campus Acceptance vs Apply Rate, No LCFF+ High Schools")
createPlots(dfLeft, dfRight, "All", xAxis="ApplyRate", filterMetrics=[["LCFF Plus School t", "==", "Y"]], title="UC Campus Acceptance vs Apply Rate, Only LCFF+ High Schools")
```

UC Campus Acceptance vs Apply Rate, High Schools in All California Counties

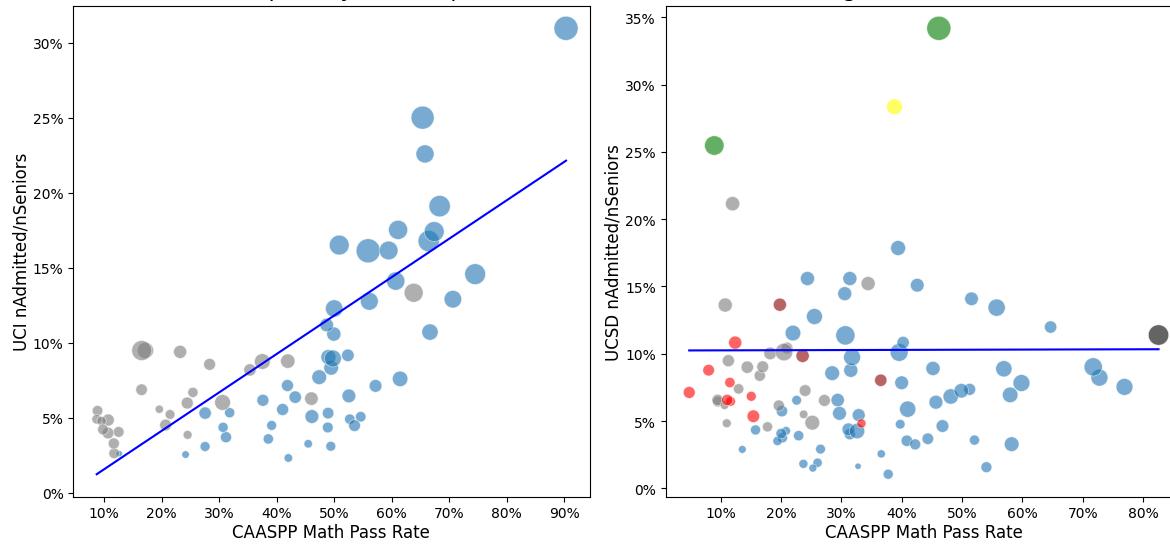






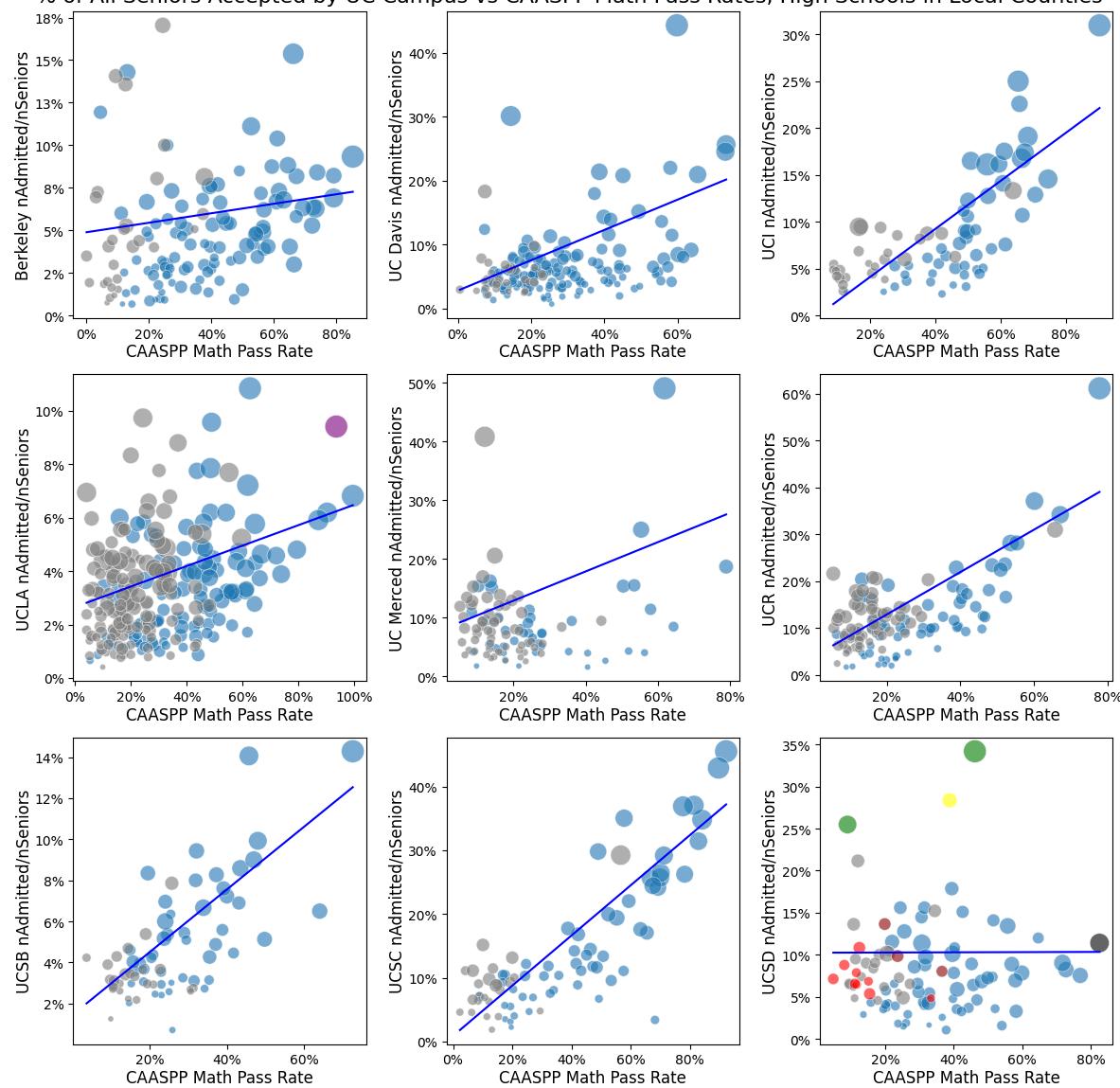
```
In [96]: createPlots2(dfLeft, dfRight, "Local", yAxis="SeniorRate", selectMinApps=0, title="% of All Seniors Accepted by UC Campus vs CAASPP Math Pass Rates, High Schools in Local Counties")
```

% of All Seniors Accepted by UC Campus vs CAASPP Math Pass Rates, High Schools in Local Counties

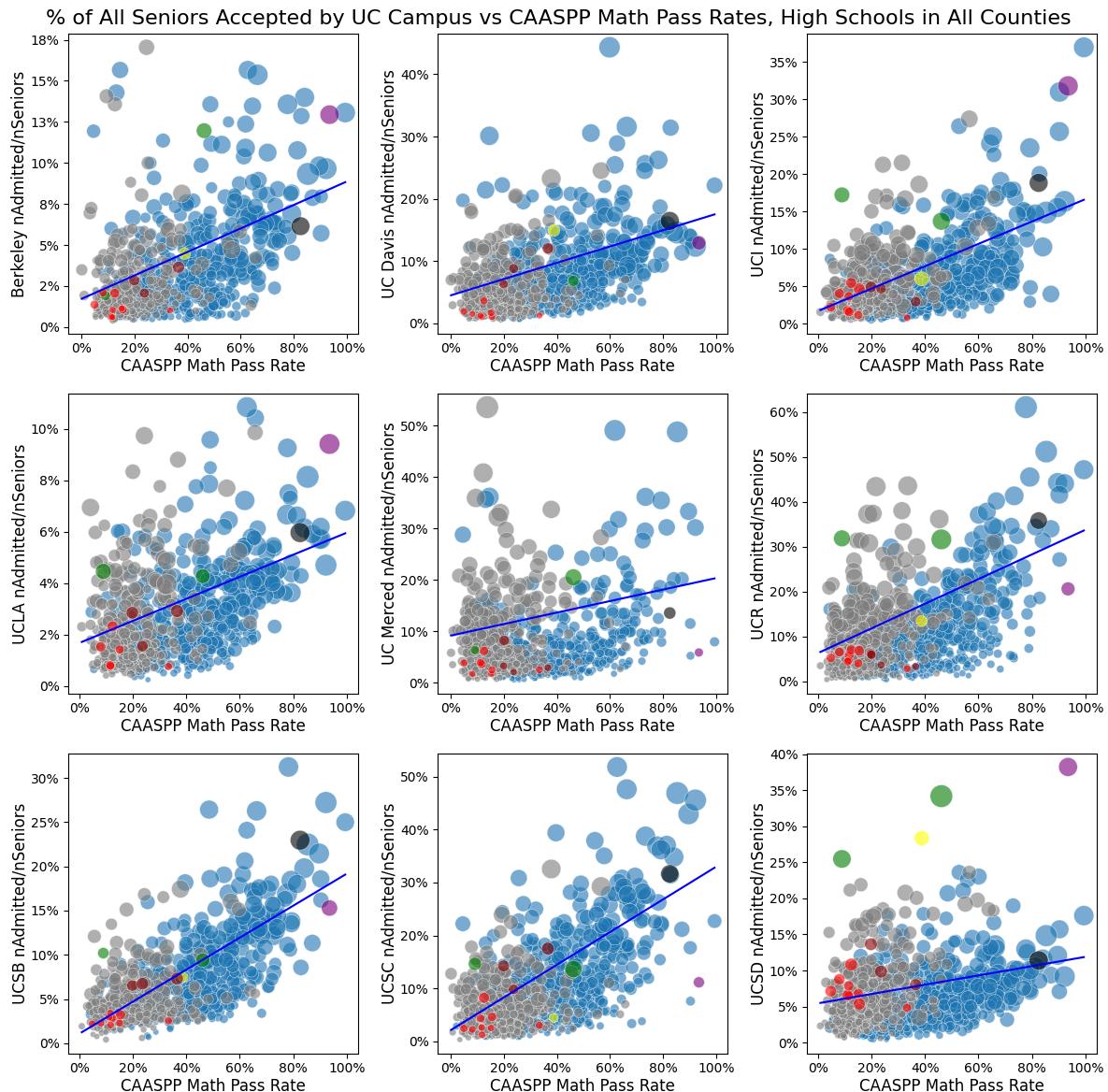


```
In [97]: createPlots(dfLeft, dfRight, "Local", yAxis="SeniorRate", selectMinApps=0, title="% of All Seniors Accepted by UC Campus vs CAASPP Math Pass Rates, High Schools in Local Counties")
```

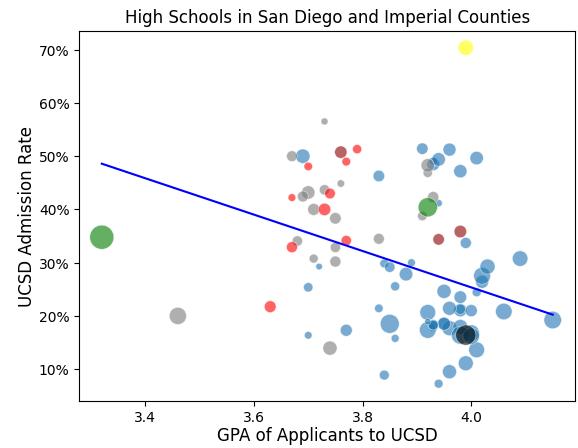
% of All Seniors Accepted by UC Campus vs CAASPP Math Pass Rates, High Schools in Local Counties



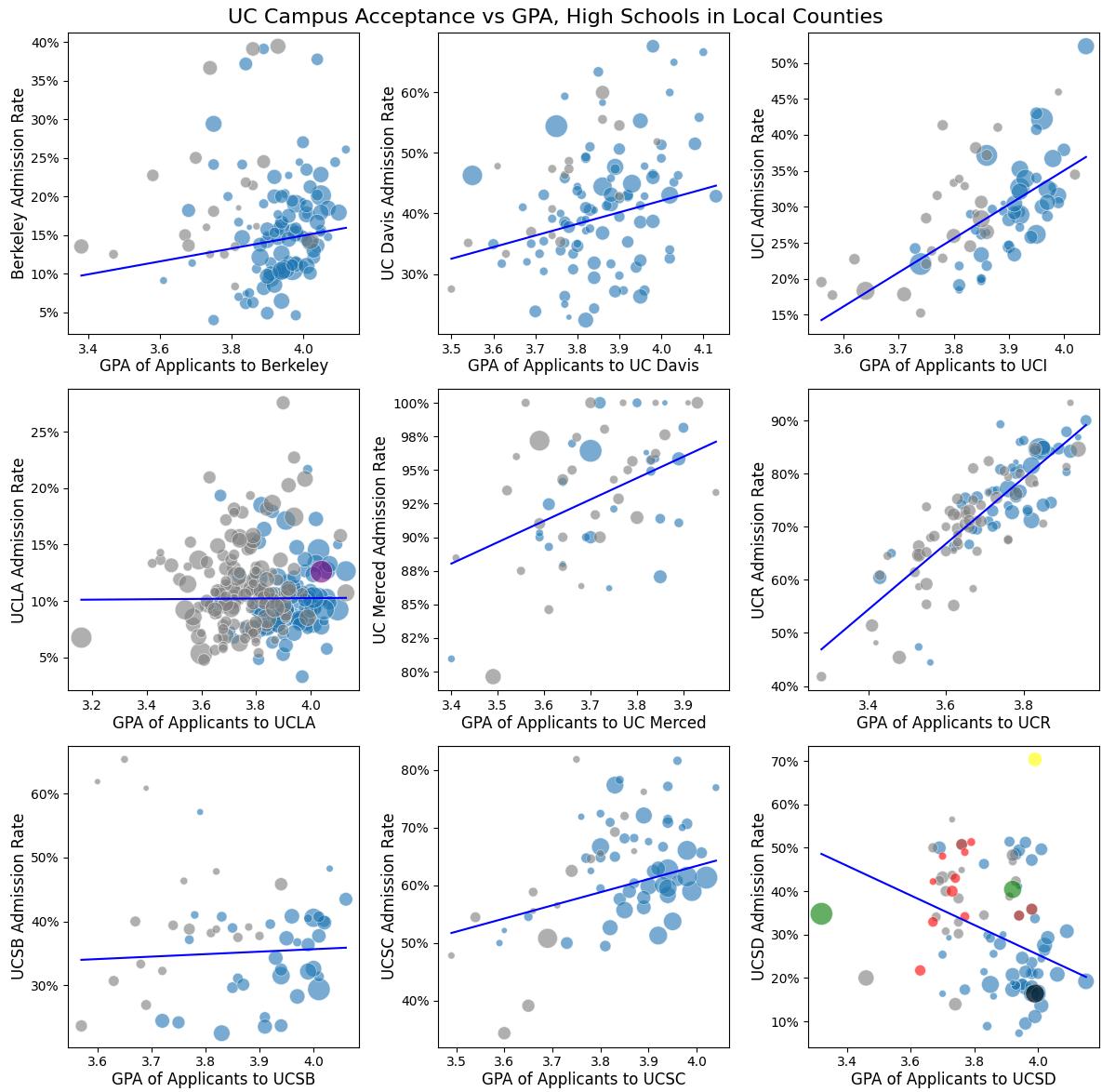
```
In [98]: createPlots(dfLeft, dfRight, "All", yAxis="SeniorRate", selectMinApps=0, title="% of All Seniors Accepted by UC Campus vs CAASPP Math Pass Rates, High Schools in All Counties")
```



```
In [99]: fig, ax = plt.subplots() # Create a Matplotlib Axes object
createPlot(ax, dfLeft, dfRight, "UCSD", "Local", xAxis="App GPA", title="High Schools in San Diego and Imperial Counties")
plt.show()
```

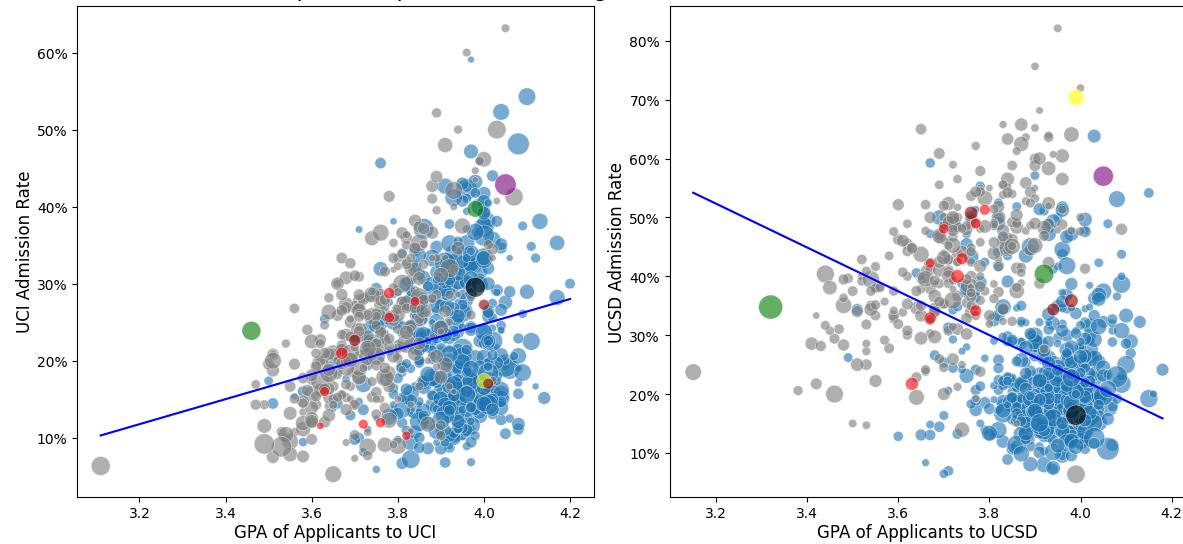


```
In [100]: createPlots(dfLeft, dfRight, "Local", xAxis="App GPA", title="UC Campus Acceptance vs GPA, High Schools in Local Counties")
```



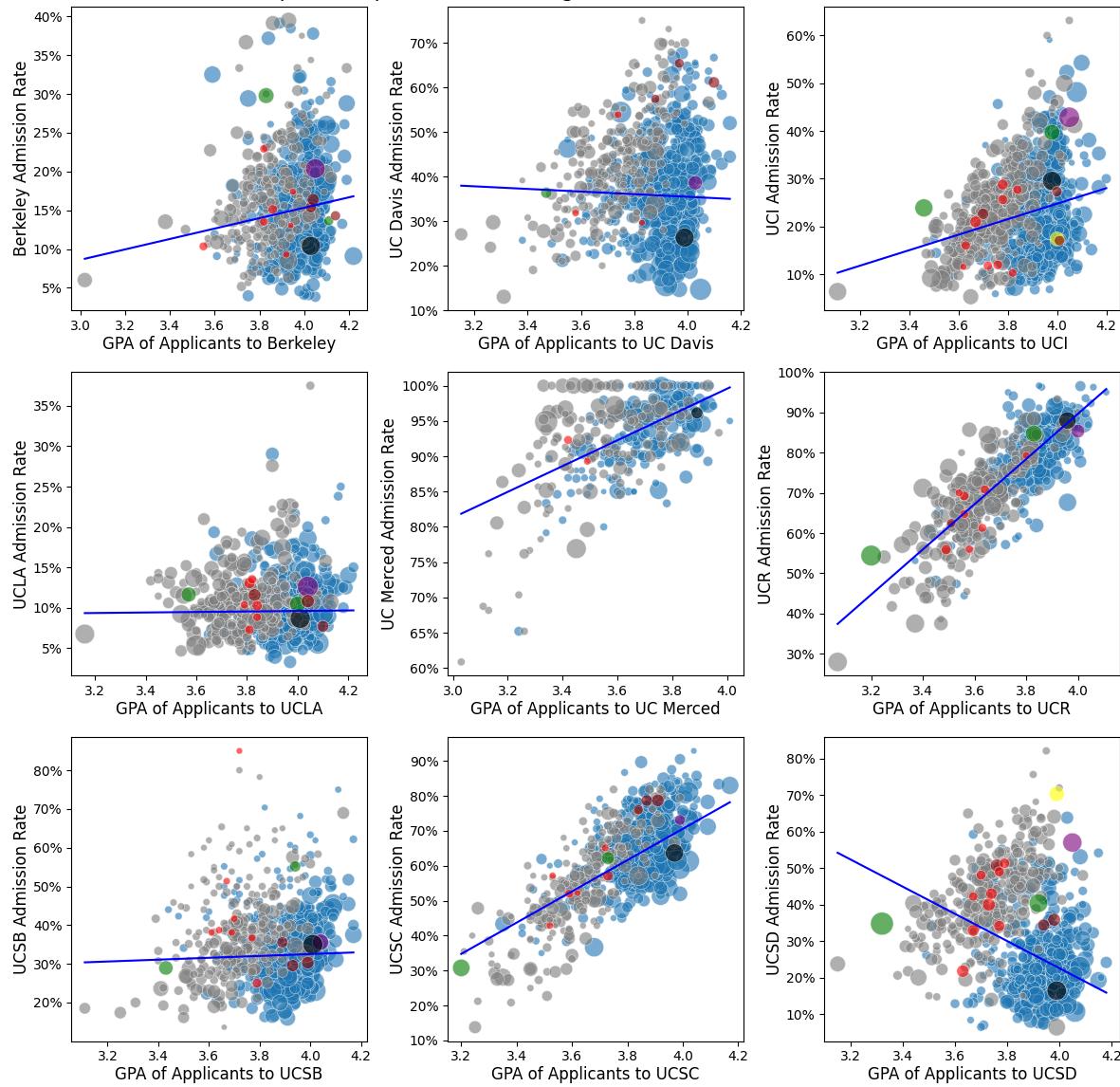
```
In [101]: createPlots2(dfLeft, dfRight, "All", xAxis="App GPA", title="UC Campus Acceptance vs GPA, High Schools in All California Counties")
```

UC Campus Acceptance vs GPA, High Schools in All California Counties

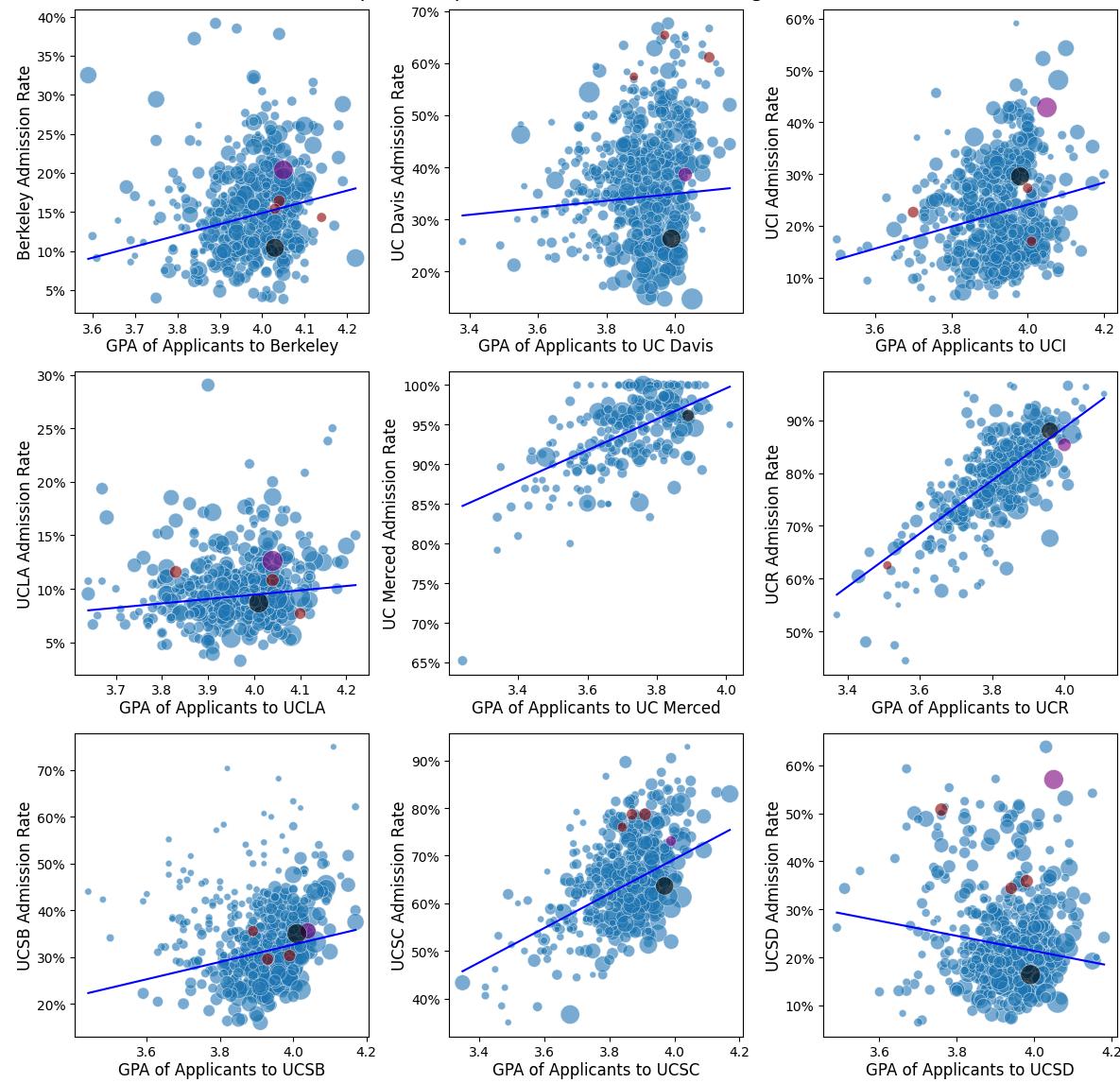


```
In [102]: createPlots(dfLeft, dfRight, "All", xAxis="App GPA", title="UC Campus Acceptance vs GPA, High Schools in All California Counties")
# Look separately at LCFF+ schools and non-LCFF+ schools
createPlots(dfLeft, dfRight, "All", xAxis="App GPA", filterMetrics=[["LCFF Plus School t", "==", "N"]], title="UC Campus Acceptance vs GPA, No LCFF+ High Schools")
createPlots(dfLeft, dfRight, "All", xAxis="App GPA", filterMetrics=[["LCFF Plus School t", "==", "Y"]], title="UC Campus Acceptance vs GPA, Only LCFF+ High Schools")
```

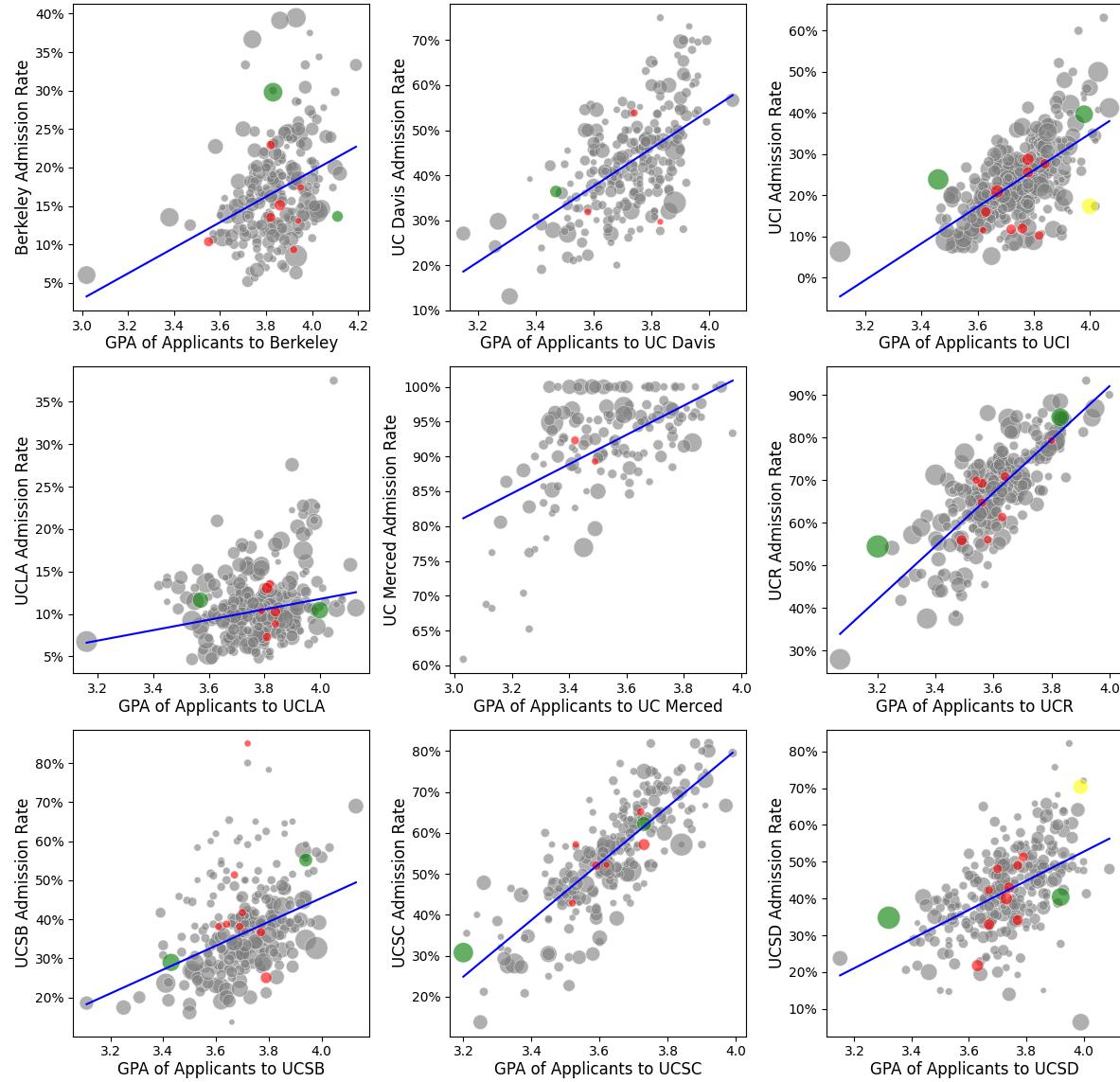
UC Campus Acceptance vs GPA, High Schools in All California Counties



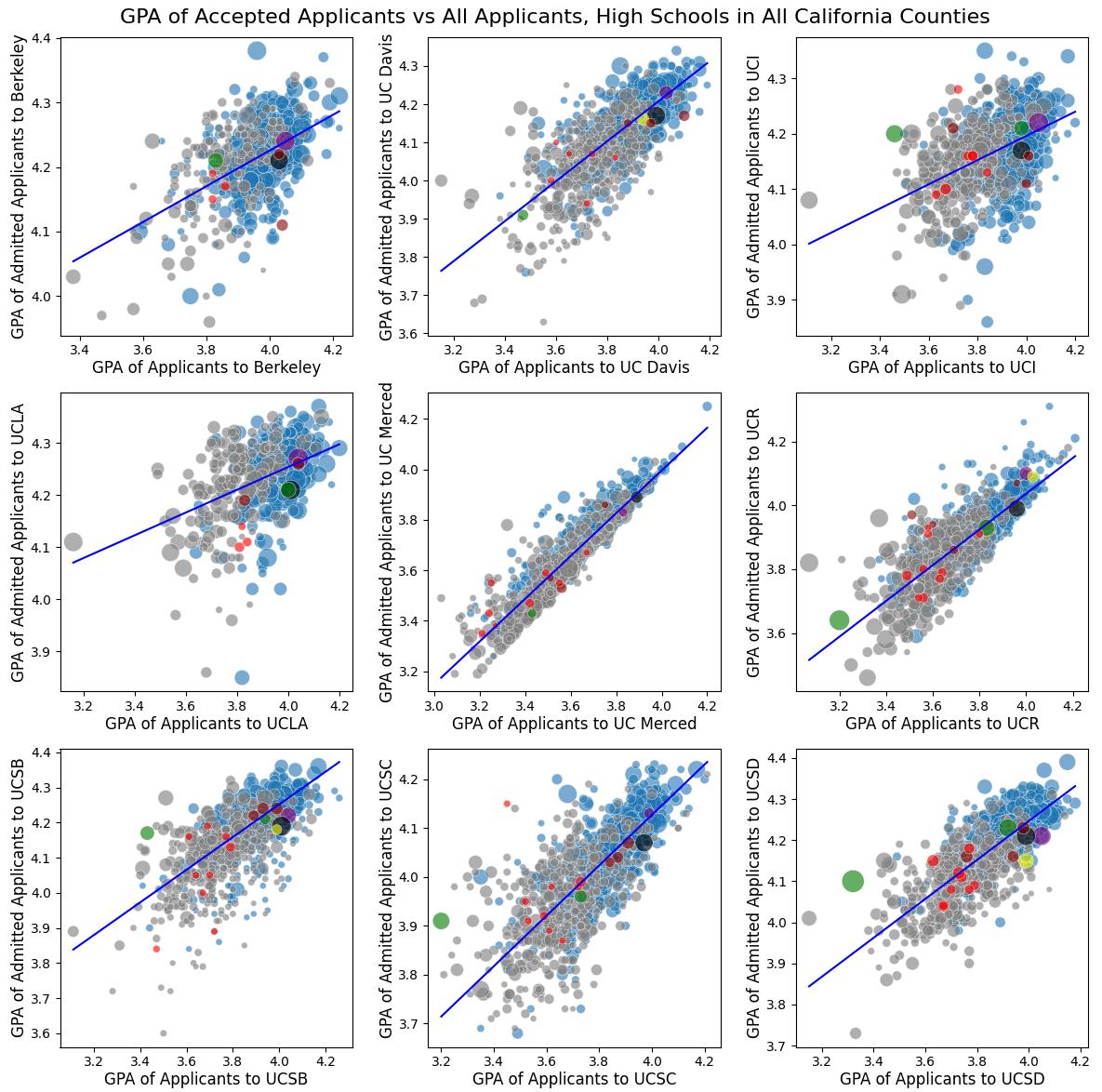
UC Campus Acceptance vs GPA, No LCFF+ High Schools



UC Campus Acceptance vs GPA, Only LCFF+ High Schools

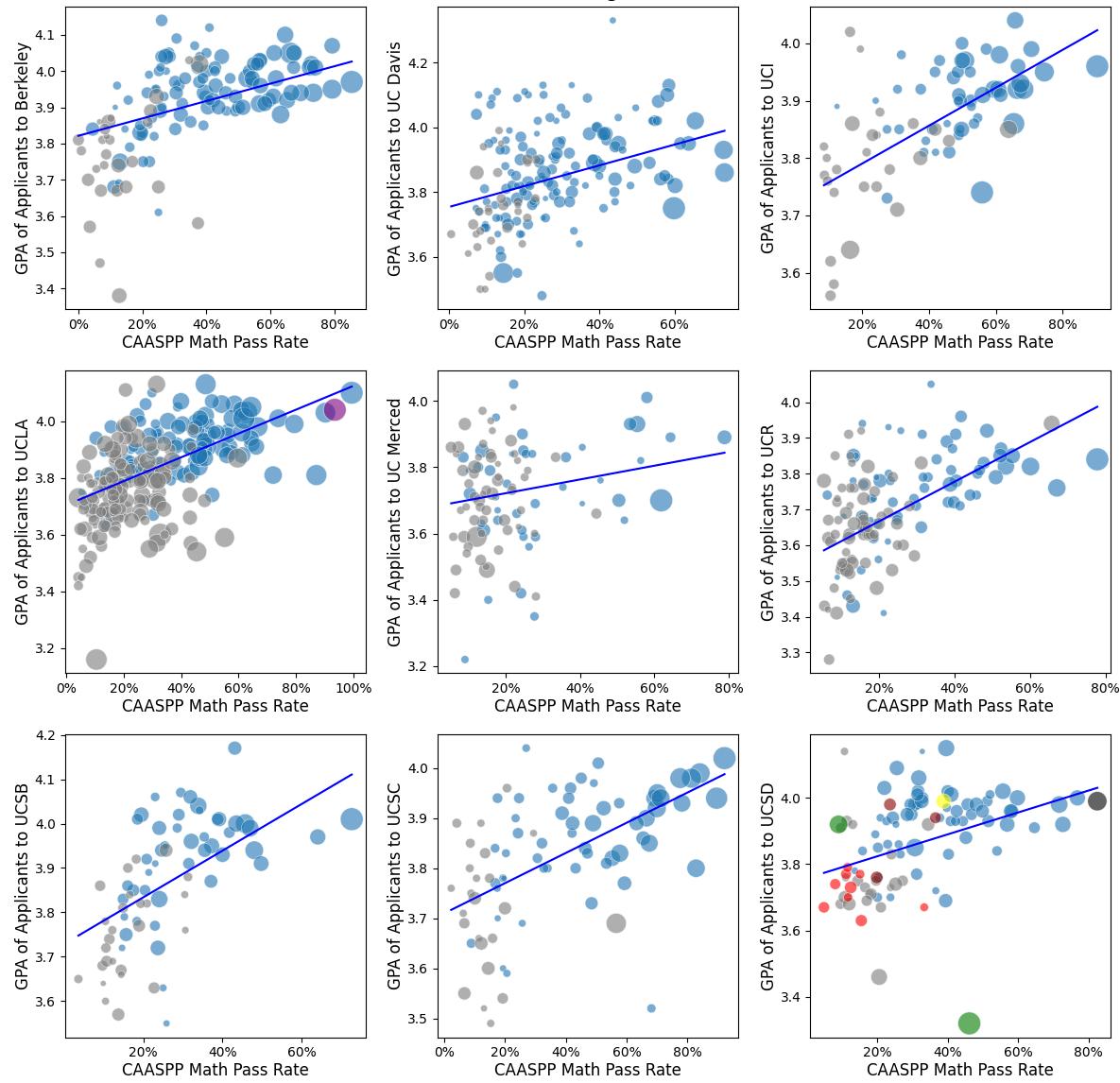


```
In [103]: createPlots(dfLeft, dfRight, "All", xAxis="App GPA", yAxis="Adm GPA", selectMinApps=0, title="GPA of Accepted Applicants vs All Applicants, High Schools in All California Counties")
```



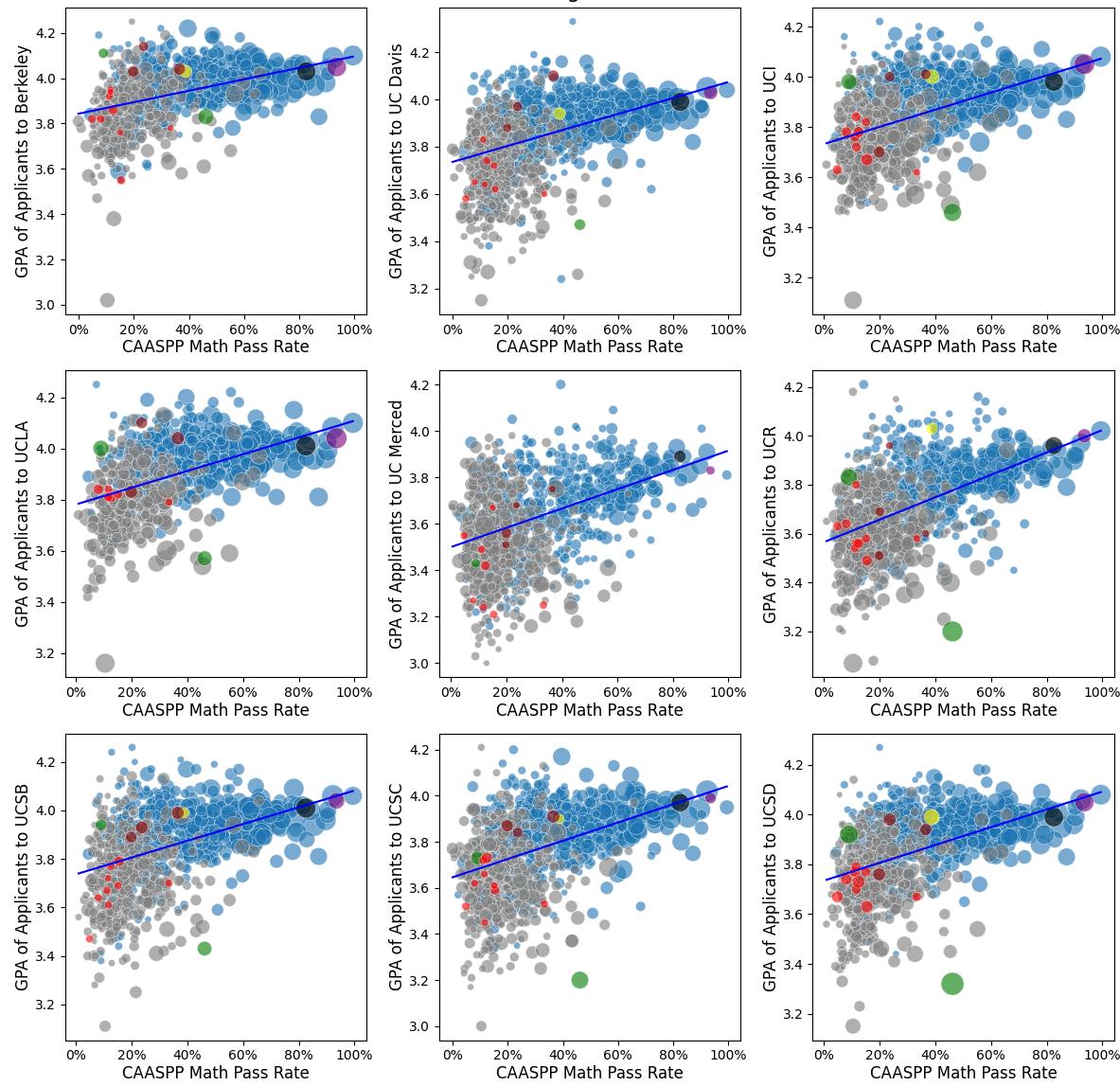
```
In [104]: createPlots(dfLeft, dfRight, "Local", yAxis="App GPA", selectMinApps=0, title="GPA vs CAASPP Math Pass Rates, High Schools in Local Counties")
```

GPA vs CAASPP Math Pass Rates, High Schools in Local Counties

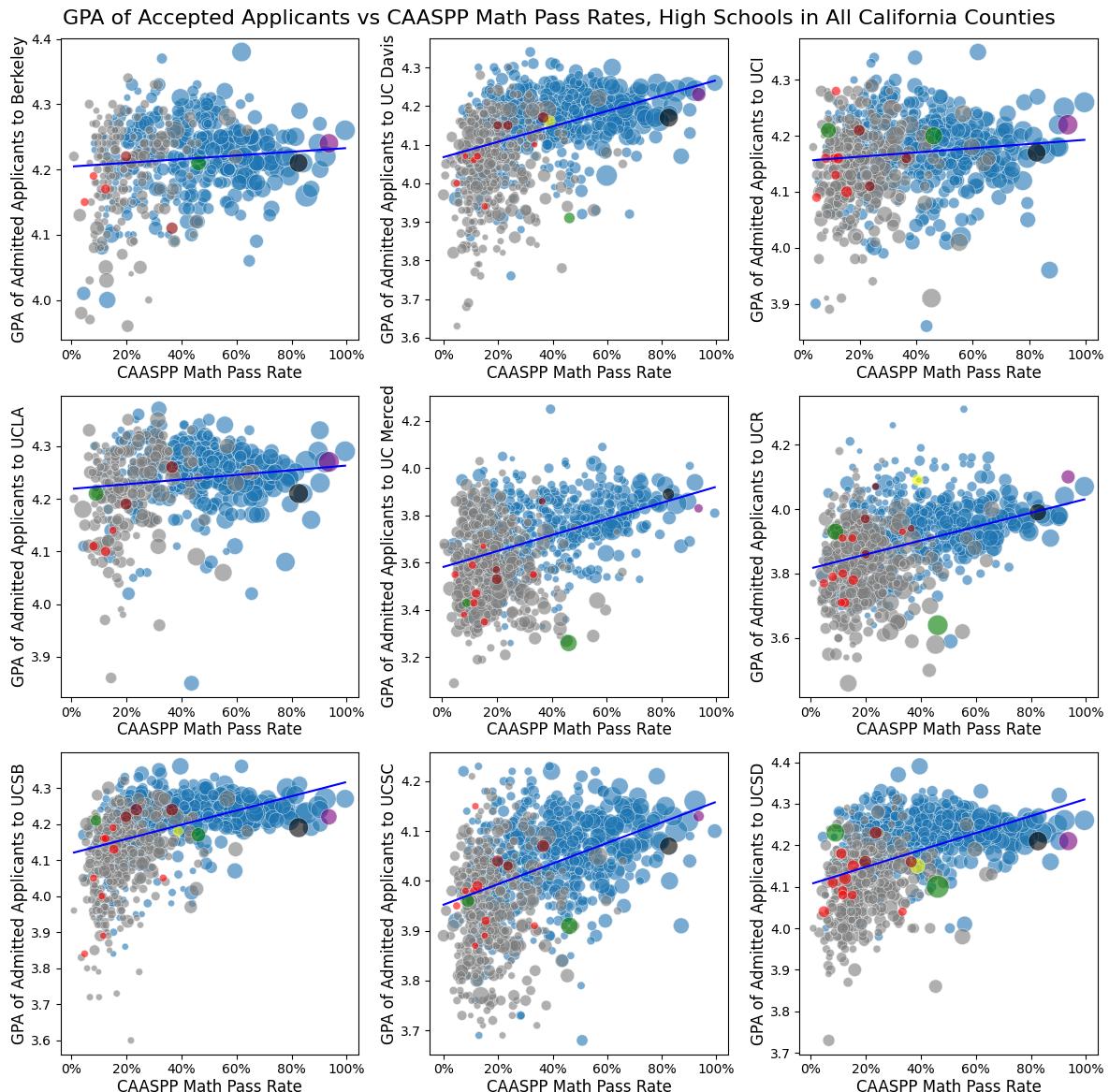


In [105]: `createPlots(dfLeft, dfRight, "All", yAxis="App GPA", selectMinApps=0, title="GPA vs CAASPP Math Pass Rates, High Schools in All California Counties")`

GPA vs CAASPP Math Pass Rates, High Schools in All California Counties



```
In [106]: createPlots(dfLeft, dfRight, "All", yAxis="Adm GPA", selectMinApps=0, title="GPA of Accepted Applicants vs CAASPP Math Pass Rates, High Schools in All California Counties")
```



In [107]:

```
# Find high schools where UCSD accepts a greater number of students than pass the CAASPP
joined = pd.merge(dfLeft, dfRight, left_on="CAASPP_ID", right_on="School Code", how="inner")
joined = joined[(joined["Campus"] == "UCSD") & (joined["Test ID"] == "2")]
joined = joined[joined["Adm"].astype(int) > joined["Count Standard Met and Above"].astype(int)]
for index, row in joined.iterrows():
    nAdm = row["Adm"]
    nMet = int(row["Count Standard Met and Above"])
    ratio = float(nAdm / nMet)
    print(f'{row["School Name"]}: {nMet} {nAdm} {ratio:.1f} {row["AdmRate"]:.0%}'')
```

Ambassador-Global Leadership	Los Angeles	4	7	1.8	47%
Animo Jackie Robinson High	Los Angeles	22	30	1.4	62%
Aspire Golden State College Preparatory Academy	Alameda	4	5	1.2	33%
Aspire Pacific Academy	Los Angeles	3	11	3.7	65%
Crawford High	San Diego	29	38	1.3	43%
Jordan High	Los Angeles	11	12	1.1	43%
Ednivate - East College Prep	Los Angeles	12	13	1.1	57%
East Los Angeles Renaissance Academy at Esteban E.	Los Angeles	7	8	1.1	73%
Fremont High	Alameda	2	6	3.0	38%
Gompers Preparatory Academy	San Diego	14	40	2.9	40%
Diego Rivera Learning Complex Green Design STEAM A	Los Angeles	8	14	1.8	61%
Hoover High	San Diego	40	46	1.1	43%
Humanitas Academy of Art and Technology at Esteban	Los Angeles	6	9	1.5	69%
King-Chavez Community High	San Diego	6	11	1.8	65%
Lincoln High	San Diego	16	26	1.6	33%
McClymonds High	Alameda	2	4	2.0	40%
MetWest High	Alameda	2	6	3.0	50%
Mira Monte High	Kern	14	19	1.4	45%
Rancho Dominguez Preparatory	Los Angeles	13	15	1.2	47%
Marshall (Thurgood) High	San Francisco	2	3	1.5	50%
University Preparatory Value High	Los Angeles	17	19	1.1	50%
West Shores High	Riverside	4	6	1.5	40%
Williams Junior/Senior High	Colusa	6	7	1.2	64%

```
In [111]: # Explore: Local schools with UCSD admission rate >= 0.5
joined = pd.merge(dfLeft, dfRight, left_on="CAASPP_ID", right_on="School Code", how="inner")
joined = joined[(joined["Campus"] == "UCSD") & (joined["countyCampus"] == "UCSD") & (joined["Test ID"] == "2")]
joined["SeniorRate"] = joined["Adm"].astype(int) / joined["Total Students Enrolled"].astype(int) # Fraction of seniors admitted
print(joined[joined["AdmRate"] >= 0.5])
```

UC_ID	School Name	School	City	County	uad_uc_ethn_7_cat	Campus	countyCampus	App	Adm	Enr	CAASPP_ID	App GPA	Adm GPA	Enrl GPA	AdmRate	EnrRate	County	Code	District	Code	District Name	School Code
14279 50430	CALIPATRIA HIGH SCHOOL	CALIPATRIA	Imperial	All	UCSD	UCSD	8	5	0	1332501	4.14	4.26	0.00	0.625000	0.000000	13	63107		Calipatria Unified	1332501		
Calipatria High																						
14341 50810	CENTRAL UNION HIGH SCHOOL	EL CENTRO	Imperial	All	UCSD	UCSD	46	26	11	1333004	3.73	4.04	3.90	0.565217	0.423077	13	63115		Central Union High	1333004		
Central Union High																						
14381 50850	CHULA VISTA SENIOR HIGH SCHOOL	CHULA VISTA	San Diego	All	UCSD	UCSD	76	39	23	3731064	3.79	4.09	4.10	0.513158	0.589744	37	68411		Sweetwater Union High	3731064		
Chula Vista Senior High																						
14559 50579	EASTLAKE HIGH SCHOOL	CHULA VISTA	San Diego	All	UCSD	UCSD	203	104	60	3730843	3.96	4.22	4.18	0.512315	0.576923	37	68411		Sweetwater Union High	3730843		
Eastlake High																						
14629 50880	ESCONDIDO HIGH SCHOOL	ESCONDIDO	San Diego	All	UCSD	UCSD	88	44	22	3732062	3.67	4.12	4.12	0.500000	0.500000	37	68106		Escondido Union High	3732062		
Escondido High																						
14803 51350	HELIX HIGH SCHOOL	LA MESA	San Diego	All	UCSD	UCSD	142	73	42	3732732	3.91	4.19	4.16	0.514085	0.575342	37	68130		Helix High	3732732		
Helix High																						
14963 53908	KEARNY SCH BIOMED SCIENCE TECH	SAN DIEGO	San Diego	All	UCSD	UCSD	27	19	8	0107086	3.99	4.15	4.14	0.703704	0.421053	37	68338		San Diego Unified	0107086		
y School of Biomedical Science and Technology																						
14967 54507	KING CHAVEZ COMMUNITY HIGH SCH	SAN DIEGO	San Diego	All	UCSD	UCSD	17	11	6	0118851	3.68	4.00	3.97	0.647059	0.545455	37	68338		King-Chavez Community High	0118851		
King-Chavez Community High																						
15353 54629	O'FARRELL CHARTER SCHOOL	SAN DIEGO	San Diego	All	UCSD	UCSD	50	25	9	6061964	3.69	4.17	4.23	0.500000	0.360000	37	68338		The O'Farrell Charter	6061964		
The O'Farrell Charter																						
15619 53903	SAN DIEGO HIGH SCHOOL	SAN DIEGO	San Diego	All	UCSD	UCSD	132	67	29	0107219	3.76	4.16	4.17	0.507576	0.432836	37	68338		San Diego Unified	0107219		
San Diego High																						

Type ID	Test ID	Total Students	Enrolled	Total Students	Tested	Total Students	Tested	with Scores	Mean Scale Score	Count Standard Exceeded	Count Standard Met	Count Standard Met and Above	Count Standard Nearly Met	Count Standard Not Met	PassRate	CA	ASPP County Name	CAASPP County Campus	\\
14279	7	2		81		75		75	2525.4		0	8		8		26		41	0.106667
Imperial			UCSD			467		467	2539.7		24	87		111		120		236	0.237687
14341	7	2				473		467											
Imperial			UCSD																
14381	7	2				496		476											
San Diego			UCSD																
14559	7	2				689		674											
San Diego			UCSD																
14629	7	2				422		418											
San Diego			UCSD																
14803	9	2				609		587											
San Diego			UCSD																
14963	7	2				67		67											
San Diego			UCSD																
14967	9	2				52		50											
San Diego			UCSD																
15353	9	2				140		137											
San Diego			UCSD																
15619	7	2				491		361											
San Diego			UCSD																

LCFF Plus School t Category	SeniorRate
14279	Y Lcff+
0.061728	
14341	Y Lcff+
0.054968	
14381	Y Tr1oY
0.078629	
14559	N
0.150943	
14629	Y LCFF+
0.104265	
14803	N
0.119869	
14963	Y KEARNY
0.283582	
14967	Y LCFF+
0.211538	
15353	N
0.178571	
15619	N TR1on
0.136456	

UC_ID	School	City	County	uad_uc_ethn_7_cat	Campus	countyCampus	App	Adm	Enr	CAASPP_ID	App GPA	Adm GPA	Enrl GPA	AdmRate	EnrRate	County	Code	District	Code	District Name	School Code	School Nam
14289 50893	CANYON CREST ACADEMY	SAN DIEGO	San Diego	All	UCSD	UCSD	384	63	14	0106328	3.99	4.21	4.14	0.164062	0.222222	37	68346		San Dieguito Union High	0106328	Canyon Crest Academ	
y	7	2				553																
14493 54351	DEL NORTE HIGH SCHOOL	SAN DIEGO	San Diego	All	UCSD	UCSD	296	48	10	0118935	4.00	4.25	4.16	0.162162	0.208333	37	68296		Poway Unified	0118935	Del Norte Hig	

Type ID	Test ID	Total Students	Enrolled	Total Students	Tested	Total Students	Tested	with Scores	Mean Scale Score	Count Standard Exceeded	Count Standard Met	Count Standard Met and Above	Count Standard Nearly Met	Count Standard Not Met	PassRate	CA	ASPP County Name	CAASPP County Campus	LC	
14289				544		2748.3		346		103			449		63		32	0.825368	San Diego	UCSD
N	CCA	0.113924																		
14493				593		2711.9		310		146			456		69		68	0.768971	San Diego	UCSD
N		0.075353																		

```
In [113]: print(joined[joined["SeniorRate"] >= 0.2])
```

UC_ID	School	City	County	uad_uc_ethn_7_cat	Campus	countyCampus	App	Adm	Enr	CAASPP_ID	App GPA	Adm GPA	Enrl GPA	AdmRate	EnrRate	County	Code	District	Code	District Name	School Code
School Name \																					
14743 54353	GOMPERS PREPARATORY ACADEMY	SAN DIEGO	San Diego	All	UCSD	UCSD	99	40	21	0119610	3.92	4.23	4.20	0.404040	0.525000	37	68338	Gompers Preparatory Academy	0119610		
Gompers Preparatory Academy																					
14963 53908	KEARNY SCH BIOMED SCIENCE TECH	SAN DIEGO	San Diego	All	UCSD	UCSD	27	19	8	0107086	3.99	4.15	4.14	0.703704	0.421053	37	68338	San Diego Unified	0107086 Kearny		
School of Biomedical Science and Technology																					
14967 54507	KING CHAVEZ COMMUNITY HIGH SCH	SAN DIEGO	San Diego	All	UCSD	UCSD	17	11	6	0118851	3.68	4.00	3.97	0.647059	0.545455	37	68338	King-Chavez Community High	0118851		
King-Chavez Community High																					
15479 51479	PREUSS SCHOOL	UCSD	LA JOLLA	San Diego	All	UCSD	115	40	26	3731189	3.32	4.10	4.03	0.347826	0.650000	37	68338	Preuss School	UCSD 3731189		
Preuss School	UCSD																				

Type	ID	Test	ID	Total Students	Enrolled	Total Students	Tested	Total Students	Tested	with Scores	Mean	Scale Score	Count	Standard Exceeded	Count Standard Met	Count Standard Met and Above	Count Standard Nearly Met	Count Standard Not Met	PassRate	CA
ASPP	County	Name	CAASPP	County	Campus	\														
14743	9	2		157		156		156		2502.8			3	11		14		41	101	0.089744
San Diego				UCSD																
14963	7	2			67	67		67		2578.1			8	18		26		13	28	0.388060
San Diego				UCSD																
14967	9	2			52	50		50		2508.7			3	3		6		13	31	0.120000
San Diego				UCSD																
15479	9	2			117	117		117		2617.7			26	28		54		29	34	0.461538
San Diego				UCSD																

LCFF Plus School t	Category	SeniorRate
14743	Y Partner	0.254777
14963	Y KEARNY	0.283582
14967	Y LCFF+	0.211538
15479	Y Partner	0.341880

```
In [114]: print(joined[joined["AdmRate"] >= 0.5])
```

UC_ID	School Name \	School	City	County	uad_uc_ethn_7_cat	Campus	countyCampus	App	Adm	Enr	CAASPP_ID	App GPA	Adm GPA	Enrl GPA	AdmRate	EnrRate	County	Code	District	Code	District Name	School Code
14279 50430	CALIPATRIA HIGH SCHOOL	CALIPATRIA	Imperial		All	UCSD		UCSD	8	5	0	1332501	4.14	4.26	0.00	0.625000	0.000000	13	63107		Calipatria Unified	1332501
Calipatria High																						
14341 50810	CENTRAL UNION HIGH SCHOOL	EL CENTRO	Imperial		All	UCSD		UCSD	46	26	11	1333004	3.73	4.04	3.90	0.565217	0.423077	13	63115		Central Union High	1333004
Central Union High																						
14381 50580	CHULA VISTA SENIOR HIGH SCHOOL	CHULA VISTA	San Diego		All	UCSD		UCSD	76	39	23	3731064	3.79	4.09	4.10	0.513158	0.589744	37	68411		Sweetwater Union High	3731064
Chula Vista Senior High																						
14559 50579	EASTLAKE HIGH SCHOOL	CHULA VISTA	San Diego		All	UCSD		UCSD	203	104	60	3730843	3.96	4.22	4.18	0.512315	0.576923	37	68411		Sweetwater Union High	3730843
Eastlake High																						
14629 50880	ESCONDIDO HIGH SCHOOL	ESCONDIDO	San Diego		All	UCSD		UCSD	88	44	22	3732062	3.67	4.12	4.12	0.500000	0.500000	37	68106		Escondido Union High	3732062
Escondido High																						
14803 51350	HELIX HIGH SCHOOL	LA MESA	San Diego		All	UCSD		UCSD	142	73	42	3732732	3.91	4.19	4.16	0.514085	0.575342	37	68130		Helix High	3732732
Helix High																						
14963 53908	KEARNY SCH BIOMED SCIENCE TECH	SAN DIEGO	San Diego		All	UCSD		UCSD	27	19	8	0107086	3.99	4.15	4.14	0.703704	0.421053	37	68338		San Diego Unified	0107086
y School of Biomedical Science and Technology																						
14967 54507	KING CHAVEZ COMMUNITY HIGH SCH	SAN DIEGO	San Diego		All	UCSD		UCSD	17	11	6	0118851	3.68	4.00	3.97	0.647059	0.545455	37	68338		King-Chavez Community High	0118851
King-Chavez Community High																						
15353 54629	OFARRELL CHARTER SCHOOL	SAN DIEGO	San Diego		All	UCSD		UCSD	50	25	9	6061964	3.69	4.17	4.23	0.500000	0.360000	37	68338		The O'Farrell Charter	6061964
The O'Farrell Charter																						
15619 53903	SAN DIEGO HIGH SCHOOL	SAN DIEGO	San Diego		All	UCSD		UCSD	132	67	29	0107219	3.76	4.16	4.17	0.507576	0.432836	37	68338		San Diego Unified	0107219
San Diego High																						

Type ID	Test ID	Total Students	Enrolled	Total Students	Tested	Total Students	Tested	with Scores	Mean Scale Score	Count Standard Exceeded Count	Standard Met	Count Standard Met	Count Standard Met and Above	Count Standard Nearly Met	Count Standard Met	Count Standard Not Met	PassRate	CA			
ASPP County Name	CAASPP County	Campus \																			
14279 7	2	81		75		75		2525.4		0		8		8		26		41	0.106667		
Imperial			UCSD		473		467		2539.7		24		87		111		120		236	0.237687	
14341 7	2		UCSD		496		476		2497.4		11		44		55		107		314	0.115546	
Imperial																					
14381 7	2	496		476		476		2601.4		108		179		287		190		197	0.425816		
San Diego			UCSD		689		674														
14559 7	2		UCSD		422		418		414		34		53		87		90		237	0.210145	
San Diego																					
14629 7	2	609		587		583		2666.5		190		187		377		137		69	0.646655		
San Diego			UCSD		67		67		2578.1		8		18		26		13		28	0.388060	
14803 9	2		UCSD		52		50		50		2508.7		3		3		6		13	0.120000	
San Diego																					
14963 7	2	140		137		137		2585.9		13		41		54		37		46	0.394161		
San Diego			UCSD		361		358		2517.5		18		53		71		71		216	0.198324	
15619 7	2		UCSD																		

LCFF Plus School t Category	SeniorRate
14279 Y	LCFF+ 0.061728
14341 Y	LCFF+ 0.054968
14381 Y	TRioY 0.078629
14559 N	0.150943
14629 Y	LCFF+ 0.104265
14803 N	0.119869
14963 Y	KEARNY 0.283582
14967 Y	LCFF+ 0.211538
15353 N	0.178571
15619 N	TRion 0.136456

In [115...]	print(joined[joined["PassRate"] >= 0.7])	UC_ID	School	City	County	uad_uc_ethn_7_cat	Campus	countyCampus	App	Adm	Enr	CAASPP_ID	App GPA	Adm GPA	Enrl GPA	AdmRate	EnrRate	County	Code	District	Code	District Name	School Code	School Nam
		14289 50893	CANYON CREST ACADEMY	SAN DIEGO	San Diego	All	UCSD		384	63	14	0106328	3.99	4.21	4.14	0.164062	0.222222	37	68346		San Dieguito Union High	0106328	Canyon Crest Academ	
		y 7	2	553																				
		14493 54351	DEL NORTE HIGH SCHOOL	SAN DIEGO	San Diego	All	UCSD		296	48	10	0118935	4.00	4.25	4.16	0.162162	0.208333	37	68296		Poway Unified	0118935	Del Norte Hig	
		h 7	2	637																				
		h 15719 52834	SCRIPPS RANCH HIGH	SAN DIEGO	San Diego	All	UCSD		219	38	17	3730884	3.92	4.24	4.22	0.173516	0.447368	37	68338		San Diego Unified	3730884	Scripps Ranch Hig	
		h 16031 52986	WESTVIEW HIGH SCHOOL	SAN DIEGO	San Diego	All	UCSD		293	48	14	3731528	3.98	4.23	4.20	0.163823	0.291667	37	68296		Poway Unified	3731528	Westview Hig	
		h 7	2	532																				

FF Plus School t Category	SeniorRate	Total Students Tested	Total Students Tested with Scores	Mean Scale Score	Count Standard Exceeded Count	Standard Met	Count Standard Met and Above	Count Standard Nearly Met	Count Standard Met	Count Standard Not Met	PassRate	CAASPP County Name	CAASPP County Campus LC	
14289 N	CCA 0.113924	544	2748.3	346	103		449		63		32	0.825368	San Diego	UCSD
14493 N	0.075353	593	2711.9	310	146		456		69		68	0.768971	San Diego	UCSD
15719 N	0.082251	455	2692.1	189	142		331		73		51	0.727473	San Diego	UCSD
16031 N	0.090226	495	2702.1	246	109		355		86		54	0.717172	San Diego	UCSD