

# college\_level\_data

March 16, 2025

## 1 College Data

```
[1]: # Data science libraries
import matplotlib.pyplot as plt
import plotly.express as px
import pandas as pd
import plotly.express as px
import seaborn as sns

# Local modules
import scripts.cccco as cccco
import scripts.college_scorecard as scorecard
```

### 1.1 1. Gather all relevant metadata on California Community Colleges using colleges endpoint of the [the CCC API](#)

#### 1.1.1 College and District Location

```
[2]: all_colleges = cccco.get_ccc_colleges()
all_districts = cccco.get_ccc_districts()
```

```
[3]: print(f"{all_colleges.shape[0]:,} rows")
all_colleges.head()
```

115 rows

```
[3]: CollegeID      CollegeName DistrictID      StreetAddress \
0      021      Cuyamaca College      020  900 Rancho San Diego Parkway
1      022      Grossmont College      020  8800 Grossmont College Drive
2      031  Imperial Valley College      030      380 East Aten Road
3      051      MiraCosta College      050      1 Barnard Drive
4      061      Palomar College      060      1140 West Mission Road

      City      County      Zip ZipPlus4      MailingAddress \
0  El Cajon  San Diego  92019      4304  900 Rancho San Diego Parkway
1  El Cajon  San Diego  92020      1799  8800 Grossmont College Drive
2  Imperial  Imperial  92251      9787      380 East Aten Road
3  Oceanside  San Diego  92056      3899      1 Barnard Drive
```

4	San Marcos	San Diego	92069	1487	1140 West Mission Road
---	------------	-----------	-------	------	------------------------

	MailingCity	MailingZip	Phone	WebsiteURL	Latitude \
0	El Cajon	92019	619.660.4000	www.cuyamaca.edu	32.744890
1	El Cajon	92020	619.644.7000	www.grossmont.edu	32.817897
2	Imperial	92251	760.352.8320	www.imperial.edu	32.825859
3	Oceanside	92056	760.757.2121	www.miracosta.edu	33.188864
4	San Marcos	92069	760.744.1150	www.palomar.edu	33.147015

	Longitude	LogoURL	District
0	-116.935229	CuyamacaCollegeLogo.jpg	None
1	-117.005640	GrossmontCollegelogo.jpg	None
2	-115.502999	ImperialValleyCollegeLogocopy.jpg	None
3	-117.301064	Mira_Costa_College_Logo_4c.png	None
4	-117.183980	PalomarCollegeLogo.jpg	None

```
[4]: print(f"{all_districts.shape[0]:,} rows")
all_districts.head()
```

73 rows

```
[4]: DistrictID          DistrictName \
0      020  Grossmont-Cuyamaca      ...
1      030   Imperial              ...
2      050   MiraCosta              ...
3      060   Palomar                ...
4      070   San Diego              ...
```

	DistrictTitle \
0	Grossmont-Cuyamaca Community College District
1	Imperial Community College District
2	MiraCosta Community College District
3	Palomar Community College District
4	San Diego Community College District

	StreetAddress	City	Zip	Phone \
0	8800 Grossmont College Drive	El Cajon	92020	619-644-7010
1	380 E. Aten Road (PO Box 158)	Imperial	92251	760-352-8320
2	None	None	None	None
3	1140 W. Mission Road	San Marcos	92069	760-744-1150
4	3375 Camino del Rio South	San Diego	92108	619-388-6500

	WebsiteURL	Boundaries \
0	www.gcccd.edu	None
1	www.imperial.edu	None
2	www.miracosta.edu	None
3	www.palomar.edu	None
4	www.sdccd.edu	None

```

                                Colleges
0  [{'CollegeID': '021', 'CollegeName': 'Cuyamaca...
1  [{'CollegeID': '031', 'CollegeName': 'Imperial...
2  [{'CollegeID': '051', 'CollegeName': 'MiraCost...
3  [{'CollegeID': '061', 'CollegeName': 'Palomar ...
4  [{'CollegeID': '071', 'CollegeName': 'San Dieg...

```

```
[5]: # Each district has a number of colleges associated with it. These colleges are
      # stored in dictionaries in the "Colleges" column.
```

```
all_districts["CollegesPerDistrict"] = all_districts["Colleges"].apply(len)
```

```
[6]: all_districts = all_districts[
      # This is Calbright College, which is the system's first online-only college
      # and doesn't have a physical campus.
      all_districts["DistrictID"] != "210"
    ]
```

```
[7]: all_districts["CollegesPerDistrict"].value_counts()
```

```
[7]: CollegesPerDistrict
1      49
2      12
3       7
4       3
9       1
Name: count, dtype: int64
```

```
[8]: # There is a clear largest district. What is it?
```

```
biggest_district = all_districts.loc[
    all_districts["CollegesPerDistrict"] == 9,
    "DistrictID"
]

all_colleges[all_colleges["DistrictID"].isin(biggest_district)]
```

```
[8]:   CollegeID      CollegeName DistrictID \
78      741      Los Angeles City College      740
79      742      Los Angeles Harbor College      740
80      743      Los Angeles Mission College      740
81      744      Los Angeles Pierce College      740
82      745      Los Angeles Southwest College      740
83      746  Los Angeles Trade-Technical College      740
84      747      Los Angeles Valley College      740
85      748      East Los Angeles College      740
```

86            749                            West Los Angeles College            740

	StreetAddress	City	County	Zip \
78	855 North Vermont Avenue	Los Angeles	Los Angeles	90029
79	1111 Figueroa Place	Wilmington	Los Angeles	90744
80	13356 Eldridge Avenue	Sylmar	Los Angeles	91342
81	6201 Winnetka Avenue	Woodland Hills	Los Angeles	91371
82	1600 West Imperial Highway	Los Angeles	Los Angeles	90047
83	400 West Washington Boulevard	Los Angeles	Los Angeles	90015
84	5800 Fulton Avenue	Van Nuys	Los Angeles	91401
85	1301 Avenida Cesar Chavez	Monterey Park	Los Angeles	91754
86	9000 Overland Avenue	Culver City	Los Angeles	90230

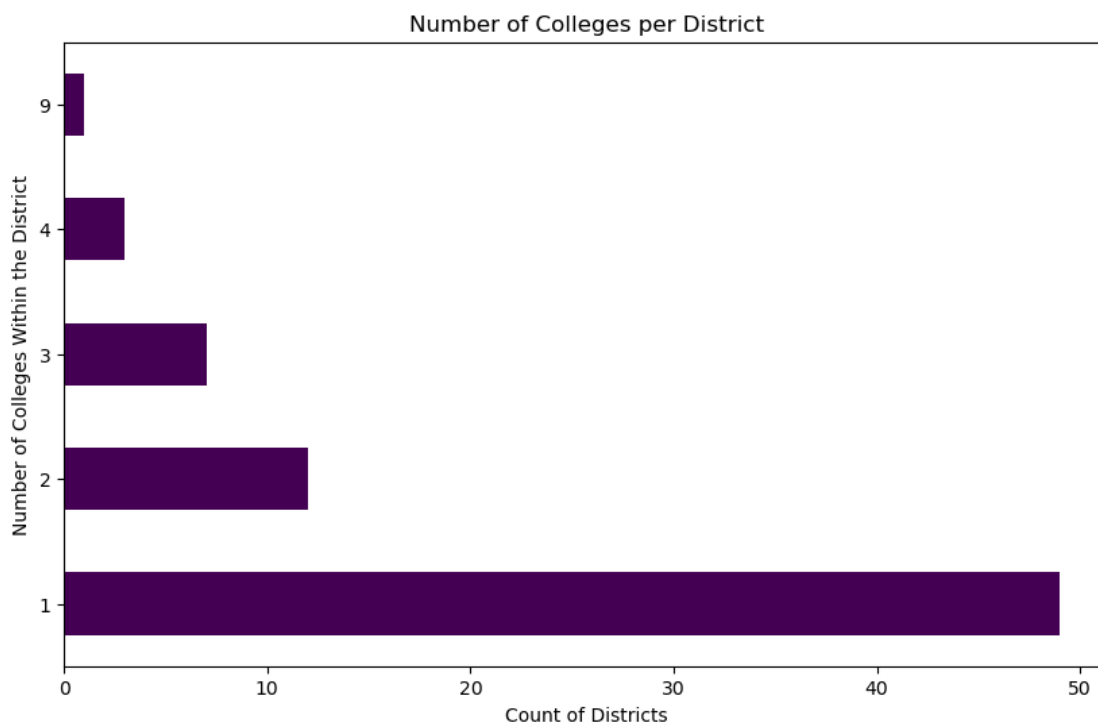
	ZipPlus4	MailingAddress	MailingCity	MailingZip \
78	3590	855 North Vermont Avenue	Los Angeles	90029
79	2397	1111 Figueroa Place	Wilmington	90744
80	3200	13356 Eldridge Avenue	Sylmar	91342
81	0001	6201 Winnetka Avenue	Woodland Hills	91371
82	4899	1600 West Imperial Highway	Los Angeles	90047
83	4108	400 West Washington Boulevard	Los Angeles	90015
84	4096	5800 Fulton Avenue	Van Nuys	91401
85	6099	1301 Avenida Cesar Chavez	Monterey Park	91754
86	5002	9000 Overland Avenue	Culver City	90230

	Phone	WebsiteURL	Latitude	Longitude \
78	323.953.4000	www.lacitycollege.edu	34.086844	-118.291864
79	310.233.4000	www.lahc.edu	33.784817	-118.281649
80	818.364.7600	www.lamission.edu	34.314604	-118.420189
81	818.719.6401	www.piercecollege.edu	34.183189	-118.571124
82	323.241.5225	www.lasc.edu	33.930808	-118.304588
83	213.763.7000	www.lattc.edu	34.033038	-118.269624
84	818.947.2600	www.lavc.edu	34.176044	-118.422335
85	323.265.8650	www.elac.edu	34.039736	-118.148754
86	310.287.4200	www.wlac.edu	34.001820	-118.391139

	LogoURL	District
78	Los Angeles CC-Logo.jpg	None
79	LosAngelesHarborCollegelogo.jpg	None
80	LosAngelesMClogocopy.jpg	None
81	LosAngelesPierceCollegeLogo.jpg	None
82	Los Angeles Southwest-logo.jpg	None
83	LosAngelesTTlogo.jpg	None
84	Los Angeles Valley College-logo.jpg	None
85	EastLosAngelesCollegeLogo.jpg	None
86	WLAC logo-color.jpg	None

```
[9]: ax = all_districts["CollegesPerDistrict"].value_counts().plot(
    kind="barh",
    title="Number of Colleges per District",
    xlabel="Count of Districts",
    ylabel="Number of Colleges Within the District",
    figsize=(10, 6),
    rot=0,
    cmap='viridis'
)

#plt.savefig("./figures/number_of_colleges_per_district.png")
plt.show()
```



```
[10]: # If a district has only one college, we'll call it a "Single College District"

all_districts.loc[
    all_districts["Colleges"].apply(len) == 1, "DistrictName"
] = "Single College District"
```

```
[11]: all_districts["DistrictName"].value_counts()
```

```
[11]: DistrictName
Single College District
```

49

Grossmont-Cuyamaca	1
Riverside	1
South Orange County	1
Rancho Santiago	1
North Orange County	1
Coast	1
Los Angeles	1
Ventura County	1
Yosemite	1
West Hills	1
State Center	1
Kern	1
West Valley-Mission	1
Chabot-Las Positas	1
San José-Evergreen	1
Foothill-De Anza	1
San Mateo County	1
Peralta	1
Contra Costa	1
Yuba	1
Los Rios	1
San Diego	1
San Bernardino	1

Name: count, dtype: int64

```
[12]: # Append district data to the college dataset.
```

```
all_colleges = pd.merge(
    all_colleges,
    all_districts.drop(
        columns=[
            "StreetAddress",
            "City",
            "Zip",
            "Phone",
            "Boundaries",
            "Colleges",
        ]
    ),
    on='DistrictID',
    how='left',
    suffixes=('', 'District')
)

all_colleges.head()
```

```
[12]: CollegeID      CollegeName DistrictID      StreetAddress \
0      021      Cuyamaca College      020 900 Rancho San Diego Parkway
1      022      Grossmont College      020 8800 Grossmont College Drive
2      031      Imperial Valley College      030      380 East Aten Road
3      051      MiraCosta College      050      1 Barnard Drive
4      061      Palomar College      060      1140 West Mission Road
```

```
      City      County      Zip ZipPlus4      MailingAddress \
0  El Cajon  San Diego  92019  4304  900 Rancho San Diego Parkway
1  El Cajon  San Diego  92020  1799  8800 Grossmont College Drive
2  Imperial  Imperial  92251  9787      380 East Aten Road
3  Oceanside  San Diego  92056  3899      1 Barnard Drive
4  San Marcos  San Diego  92069  1487      1140 West Mission Road
```

```
      MailingCity ...      Phone      WebsiteURL      Latitude      Longitude \
0  El Cajon ...  619.660.4000  www.cuyamaca.edu  32.744890 -116.935229
1  El Cajon ...  619.644.7000  www.grossmont.edu  32.817897 -117.005640
2  Imperial ...  760.352.8320  www.imperial.edu  32.825859 -115.502999
3  Oceanside ...  760.757.2121  www.miracosta.edu  33.188864 -117.301064
4  San Marcos ...  760.744.1150  www.palomar.edu  33.147015 -117.183980
```

```
      LogoURL District \
0      CuyamacaCollegeLogo.jpg      None
1      GrossmontCollegelogo.jpg      None
2  ImperialValleyCollegeLogocopy.jpg      None
3      Mira_Costa_College_Logo_4c.png      None
4      PalomarCollegeLogo.jpg      None
```

```
      DistrictName \
0  Grossmont-Cuyamaca      ...
1  Grossmont-Cuyamaca      ...
2      Single College District
3      Single College District
4      Single College District
```

```
      DistrictTitle WebsiteURLDistrict \
0  Grossmont-Cuyamaca Community College District      www.gcccd.edu
1  Grossmont-Cuyamaca Community College District      www.gcccd.edu
2      Imperial Community College District      www.imperial.edu
3      MiraCosta Community College District      www.miracosta.edu
4      Palomar Community College District      www.palomar.edu
```

```
CollegesPerDistrict
0      2
1      2
2      1
3      1
```

[5 rows x 21 columns]

```
[13]: fig = px.scatter_mapbox(
    all_colleges,
    lat="Latitude",
    lon="Longitude",
    hover_name="CollegeName",
    zoom=5,
    width=1200,
    height=800,
    color="DistrictName",
    title="Geographic Distribution of California Community Colleges",
    labels={
        "DistrictName": "District",
    },
)

fig.update_layout(
    mapbox_style="carto-positron",
    template="plotly_white",
)

fig.update_traces(marker=dict(size=6))

fig.write_html("./docs/california_community_colleges.html")
fig.show()
```

```
[14]: # Rank the districts by college count.

mc_district = all_districts.loc[all_districts["CollegesPerDistrict"] > 1,
    ↪ "DistrictID"]
all_colleges[all_colleges["DistrictID"].isin(mc_district)]["DistrictTitle"].
    ↪ value_counts()
```

```
[14]: DistrictTitle
Los Angeles Community College District      9
Los Rios Community College District          4
Peralta Community College District          4
State Center Community College District      4
Kern Community College District              3
Ventura County Community College District    3
Riverside Community College District         3
Contra Costa Community College District      3
San Mateo County Community College District  3
Coast Community College District             3
```



San Diego Community College District	3
South Orange County Community College District	2
Rancho Santiago Community College District	2
North Orange County Community College District	2
Grossmont-Cuyamaca Community College District	2
Yosemite Community College District	2
West Hills Community College District	2
West Valley-Mission Community College District	2
Chabot-Las Positas Community College District	2
San José-Evergreen Community College District	2
Foothill-De Anza Community College District	2
Yuba Community College District	2
San Bernardino Community College District	2

Name: count, dtype: int64

## 1.2 2. Gather all colleges in the CCC system that offer a degree or certificate in the programs of interest using programs endpoint [the CCC API](#)

```
[15]: # Example of use case
```

```
programs = [
    "Database",
    "Computer Science",
    "Data Science",
]
```

```
[16]: cccco_programs = pd.DataFrame()
```

```
for program in programs:
    df = cccco.get_ccc_programs(program)
    df["Program Query"] = program
    cccco_programs = pd.concat([cccco_programs, df])
```

```
[17]: print(f"{cccco_programs.shape[0]:,} rows")
cccco_programs.sample(10)
```

207 rows

```
[17]:
```

	CollegeID	CollegeName	ProgramAward	CreditType	\
2	551	San Joaquin Delta College	C	C	
43	731	Glendale Community College	T	C	
126	749	West Los Angeles College	T	C	
15	963	Norco College	S	C	
0	233	Sacramento City College	M	C	
22	422	Foothill College	S	C	
27	592	Modesto Junior College	S	C	
25	343	Laney College	N	C	
4	749	West Los Angeles College	N	C	

5	231	American River College	S	C
---	-----	------------------------	---	---

		Title	TopCode	Program Query
2		Computer Science	070200	Computer Science
43		Computer Science	070600	Computer Science
126	Computer Science	Information Technology	070200	Computer Science
15		Data Science	070730	Data Science
0		Data Science	070200	Data Science
22		Computer Science	070600	Computer Science
27		Computer Science	070600	Computer Science
25	Data Science	Fundamentals	070100	Data Science
4		Data Science	070200	Data Science
5		CIS: Database Management	070720	Database

```
[18]: cccco_programs["Program Query"].value_counts()
```

```
[18]: Program Query
Computer Science    153
Data Science       28
Database           26
Name: count, dtype: int64
```

### 1.3 3. Merge the colleges endpoint data with the programs endpoint data (one row per college per program)

```
[19]: # Merge College Metadata with Program Data
# One row per college per program

college_data = pd.merge(
    all_colleges,
    cccco_programs.drop(columns=["CollegeName"]),
    on="CollegeID",
    how="right",
)

college_data.sort_values(by=["CollegeID", "Program Query"], inplace=True)

print(f"{college_data.shape[0]:,} rows")
college_data.head()
```

207 rows

```
[19]: CollegeID      CollegeName DistrictID \
71      021      Cuyamaca College      020
72      022      Grossmont College      020
178     022      Grossmont College      020
190     022      Grossmont College      020
73      031  Imperial Valley College      030
```

	StreetAddress	City	County	Zip	ZipPlus4	\
71	900 Rancho San Diego Parkway	El Cajon	San Diego	92019	4304	
72	8800 Grossmont College Drive	El Cajon	San Diego	92020	1799	
178	8800 Grossmont College Drive	El Cajon	San Diego	92020	1799	
190	8800 Grossmont College Drive	El Cajon	San Diego	92020	1799	
73	380 East Aten Road	Imperial	Imperial	92251	9787	

	MailingAddress	MailingCity	...	District	\
71	900 Rancho San Diego Parkway	El Cajon	...	None	
72	8800 Grossmont College Drive	El Cajon	...	None	
178	8800 Grossmont College Drive	El Cajon	...	None	
190	8800 Grossmont College Drive	El Cajon	...	None	
73	380 East Aten Road	Imperial	...	None	

	DistrictName	\
71	Grossmont-Cuyamaca	...
72	Grossmont-Cuyamaca	...
178	Grossmont-Cuyamaca	...
190	Grossmont-Cuyamaca	...
73	Single College District	

	DistrictTitle	WebsiteURL	District	\
71	Grossmont-Cuyamaca Community College District	www.gcccd.edu		
72	Grossmont-Cuyamaca Community College District	www.gcccd.edu		
178	Grossmont-Cuyamaca Community College District	www.gcccd.edu		
190	Grossmont-Cuyamaca Community College District	www.gcccd.edu		
73	Imperial Community College District	www.imperial.edu		

	CollegesPerDistrict	ProgramAward	CreditType	\
71	2	X	C	
72	2	X	C	
178	2	S	C	
190	2	N	C	
73	1	X	C	

	Title	TopCode	\
71	Computer Science	070600	
72	Computer Science	070600	
178	University Studies: Mathematics and Natural Sc...	490200	
190	Data Science	070700	
73	Computer Science	070600	

	Program Query
71	Computer Science
72	Computer Science
178	Computer Science

```
190     Data Science
73     Computer Science
```

```
[5 rows x 26 columns]
```

#### 1.4 4. Merge the college data with the DataMart program data (one row per college per program)

```
[20]: pass # Not implemented in this notebook
```

#### 1.5 5. Merge the College Scorecard data

```
[21]: scorecard.get_scorecard_by_college("Solano")
```

```
[21]:          school_name  latest_student_size \
0  Solano Community College                7213

      latest_student_enrollment_undergrad_12_month \
0                                           11932

      latest_student_demographics_over_23_at_entry \
0                                           0.5

      latest_student_demographics_first_generation \
0                                           0.484509

      latest_student_demographics_median_hh_income \
0                                           72436

      latest_student_demographics_student_faculty_ratio \
0                                           25

      latest_student_FAFSA_applications
0                                2251
```

```
[22]: import time

scorecard_data = pd.DataFrame()

for name, city in all_colleges[["CollegeName", "City"]].drop_duplicates().
    .itertuples(index=False):
    time.sleep(0.25)
    try:
        scorecard_info = scorecard.get_scorecard_by_college(name, city, "CA")
        scorecard_data = pd.concat([scorecard_data, scorecard_info],
    .ignore_index=True)
    except Exception as e:
```

```

print(f"Error with {name}: {e}")

# Merge the scorecard data with the all_colleges DataFrame
all_colleges = pd.merge(
    all_colleges,
    scorecard_data,
    left_on="CollegeName",
    right_on="school_name",
    how="left"
).drop(columns=["school_name"])

print(f"{all_colleges.shape[0]:,} rows")
all_colleges.head()

```

Error with Lake Tahoe Community College: "None of [Index(['school\_name'], dtype='object')] are in the [columns]"

Error with Los Angeles Valley College: "None of [Index(['school\_name'], dtype='object')] are in the [columns]"

115 rows

```

[22]: CollegeID      CollegeName DistrictID      StreetAddress \
0      021      Cuyamaca College      020  900 Rancho San Diego Parkway
1      022      Grossmont College      020  8800 Grossmont College Drive
2      031  Imperial Valley College      030      380 East Aten Road
3      051      MiraCosta College      050      1 Barnard Drive
4      061      Palomar College      060      1140 West Mission Road

```

```

      City      County      Zip ZipPlus4      MailingAddress \
0  El Cajon  San Diego  92019    4304  900 Rancho San Diego Parkway
1  El Cajon  San Diego  92020    1799  8800 Grossmont College Drive
2  Imperial  Imperial  92251    9787      380 East Aten Road
3  Oceanside  San Diego  92056    3899      1 Barnard Drive
4  San Marcos  San Diego  92069    1487      1140 West Mission Road

```

```

      MailingCity ...      DistrictTitle \
0  El Cajon ...  Grossmont-Cuyamaca Community College District
1  El Cajon ...  Grossmont-Cuyamaca Community College District
2  Imperial ...      Imperial Community College District
3  Oceanside ...      MiraCosta Community College District
4  San Marcos ...      Palomar Community College District

```

```

      WebsiteURLDistrict CollegesPerDistrict  latest_student_size \
0      www.gcccd.edu      2      6527.0
1      www.gcccd.edu      2      10994.0
2      www.imperial.edu      1      6952.0
3      www.miracosta.edu      1      10094.0

```

4	www.palomar.edu	1	15057.0
---	-----------------	---	---------

	latest_student_enrollment_undergrad_12_month \	
0	12428.0	
1	18249.0	
2	10185.0	
3	16847.0	
4	26222.0	

	latest_student_demographics_over_23_at_entry \	
0	0.47	
1	0.39	
2	0.26	
3	0.43	
4	0.44	

	latest_student_demographics_first_generation \	
0	0.610113	
1	0.550389	
2	0.609207	
3	0.509714	
4	0.538365	

	latest_student_demographics_median_hh_income \	
0	61734.0	
1	60424.0	
2	45460.0	
3	68755.0	
4	68227.0	

	latest_student_demographics_student_faculty_ratio \	
0	23.0	
1	19.0	
2	31.0	
3	19.0	
4	21.0	

	latest_student_FAFSA_applications	
0	2901.0	
1	5463.0	
2	3798.0	
3	4602.0	
4	5316.0	

[5 rows x 28 columns]

```
[23]: # Top 5 largest colleges by annual enrollment
```

```
all_colleges.sort_values("latest_student_enrollment_undergrad_12_month",
↪ascending=False).head()
```

```
[23]:
```

	CollegeID	CollegeName	DistrictID	StreetAddress	\
85	748	East Los Angeles College	740	1301 Avenida Cesar Chavez	
17	231	American River College	230	4700 College Oak Drive	
98	871	Santa Ana College	870	1530 W. 17th Street	
52	521	Bakersfield College	520	1801 Panorama Drive	
88	781	Santa Monica College	780	1900 Pico Boulevard	

	City	County	Zip	ZipPlus4	MailingAddress	\
85	Monterey Park	Los Angeles	91754	6099	1301 Avenida Cesar Chavez	
17	Sacramento	Sacramento	95841	4286	4700 College Oak Drive	
98	Santa Ana	Orange County	92706	3398	1530 W. 17th Street	
52	Bakersfield	Kern County	93305	1299	1801 Panorama Drive	
88	Santa Monica	Los Angeles	90405	1628	1900 Pico Boulevard	

	MailingCity	...	DistrictTitle	\
85	Monterey Park	...	Los Angeles Community College District	
17	Sacramento	...	Los Rios Community College District	
98	Santa Ana	...	Rancho Santiago Community College District	
52	Bakersfield	...	Kern Community College District	
88	Santa Monica	...	Santa Monica Community College District	

	WebsiteURL	District	CollegesPerDistrict	latest_student_size	\
85	www.laccd.edu		9	23499.0	
17	www.losrios.edu		4	24054.0	
98	www.rscdd.edu		2	20309.0	
52	www.kccd.edu		3	18098.0	
88	www.smc.edu		1	20376.0	

	latest_student_enrollment_undergrad_12_month	\
85	49266.0	
17	41538.0	
98	38996.0	
52	37017.0	
88	34036.0	

	latest_student_demographics_over_23_at_entry	\
85	0.31	
17	0.52	
98	0.32	
52	0.36	
88	0.25	

```

latest_student_demographics_first_generation \
85      0.685203
17      0.514148
98      0.696058
52      0.630508
88      0.501535

latest_student_demographics_median_hh_income \
85      45627.0
17      59196.0
98      69027.0
52      46274.0
88      54172.0

latest_student_demographics_student_faculty_ratio \
85      29.0
17      27.0
98      31.0
52      28.0
88      26.0

latest_student_FAFSA_applications
85      7896.0
17      7428.0
98      3832.0
52      8128.0
88      7297.0

```

[5 rows x 28 columns]

[24]: *# Top 5 smallest colleges by annual enrollment*

```

all_colleges[
  all_colleges["latest_student_enrollment_undergrad_12_month"].notna()
].sort_values("latest_student_enrollment_undergrad_12_month", ascending=True).
  head()

```

```

[24]:   CollegeID      CollegeName DistrictID \
15      181  College of the Siskiyous      180
11      131  Lassen Community College      130
64      591      Columbia College      590
103     911  Barstow Community College      910
54      523   Porterville College      520

      StreetAddress      City      County      Zip \
15      800 College Avenue      Weed      Siskiyou  96094
11      478-200 Hwy 139  Susanville      Lassen  96130

```



64	11600 Columbia College Drive	Sonora	Tuolumne	95370
103	2700 Barstow Road	Barstow	San Bernardino	92311
54	100 East College Avenue	Porterville	Tulare	93257

	ZipPlus4	MailingAddress	MailingCity	...	\
15	2899	800 College Avenue	Weed	...	
11	3000	PO Box 3000	Susanville	...	
64	8518	11600 Columbia College Drive	Sonora	...	
103	6699	2700 Barstow Road	Barstow	...	
54	5901	100 East College Avenue	Porterville	...	

	DistrictTitle	WebsiteURLDistrict	\
15	Siskiyou Joint Community College District	www.siskiyous.edu	
11	Lassen Community College District	www.lassencollege.edu	
64	Yosemite Community College District	www.yosemite.edu	
103	Barstow Community College District	www.barstow.edu	
54	Kern Community College District	www.kccd.edu	

	CollegesPerDistrict	latest_student_size	\
15	1	932.0	
11	1	1431.0	
64	2	1801.0	
103	1	2496.0	
54	3	2757.0	

	latest_student_enrollment_undergrad_12_month	\
15	2421.0	
11	2885.0	
64	3304.0	
103	3615.0	
54	5062.0	

	latest_student_demographics_over_23_at_entry	\
15	0.39	
11	0.41	
64	0.51	
103	0.61	
54	0.35	

	latest_student_demographics_first_generation	\
15	0.446043	
11	0.549383	
64	0.534121	
103	0.531070	
54	0.630058	

	latest_student_demographics_median_hh_income	\
--	--	---

15	45814.0
11	52031.0
64	53192.0
103	50748.0
54	42464.0

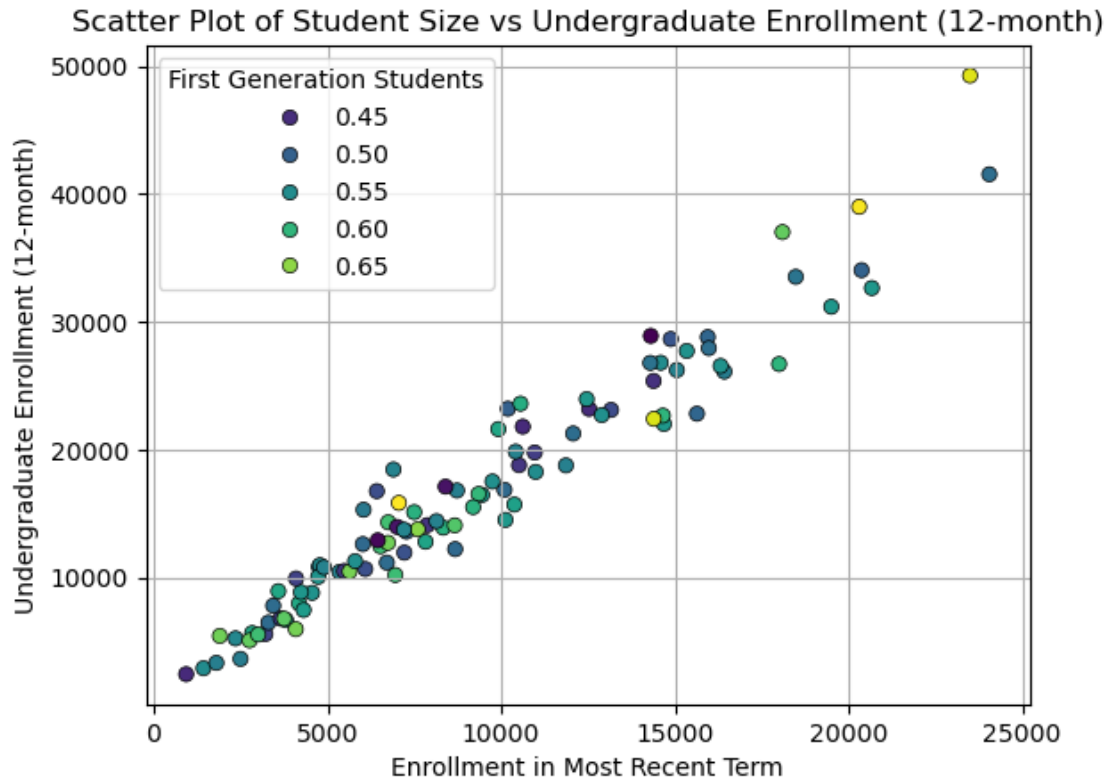
	latest_student_demographics_student_faculty_ratio \
15	13.0
11	13.0
64	14.0
103	21.0
54	24.0

	latest_student_FAFSA_applications
15	810.0
11	383.0
64	877.0
103	1973.0
54	1845.0

[5 rows x 28 columns]

```
[25]: sns.scatterplot(
    data=all_colleges,
    x="latest_student_size",
    y="latest_student_enrollment_undergrad_12_month",
    hue="latest_student_demographics_first_generation",
    palette="viridis",
    edgecolor="black"
)
plt.title("Scatter Plot of Student Size vs Undergraduate Enrollment (12-month)")
plt.xlabel("Enrollment in Most Recent Term")
plt.ylabel("Undergraduate Enrollment (12-month)")
plt.legend(title="First Generation Students")
plt.grid(True)

#plt.savefig('./figures/student_size_vs_undergraduate_enrollment.png',
#            bbox_inches='tight')
plt.show()
```



```
[ ]: y_mean = all_colleges.
    ↳groupby("latest_student_demographics_student_faculty_ratio")["
        "latest_student_enrollment_undergrad_12_month"
    ].mean()
y_std = all_colleges.
    ↳groupby("latest_student_demographics_student_faculty_ratio")["
        "latest_student_enrollment_undergrad_12_month"
    ].std()

plot_data = pd.DataFrame(
    {
        "Student Faculty Ratio": y_mean.index,
        "Mean Enrollment": y_mean.values,
        "Enrollment Std Dev": y_std.values,
    }
)

# Plot with error bars
fig = px.scatter(
    plot_data,
    x="Student Faculty Ratio",
```

```

    y="Mean Enrollment",
    error_y="Enrollment Std Dev",
    color_continuous_scale="viridis",
    title="Mean Enrollment vs Student Faculty Ratio",
    labels={"Student Faculty Ratio": "Student Faculty Ratio", "Mean Enrollment":
↪ "Mean Enrollment"},
    template="gridon"
)

fig.update_traces(marker=dict(size=6, line=dict(width=2,
↪ color="DarkSlateGrey")))

fig.show()

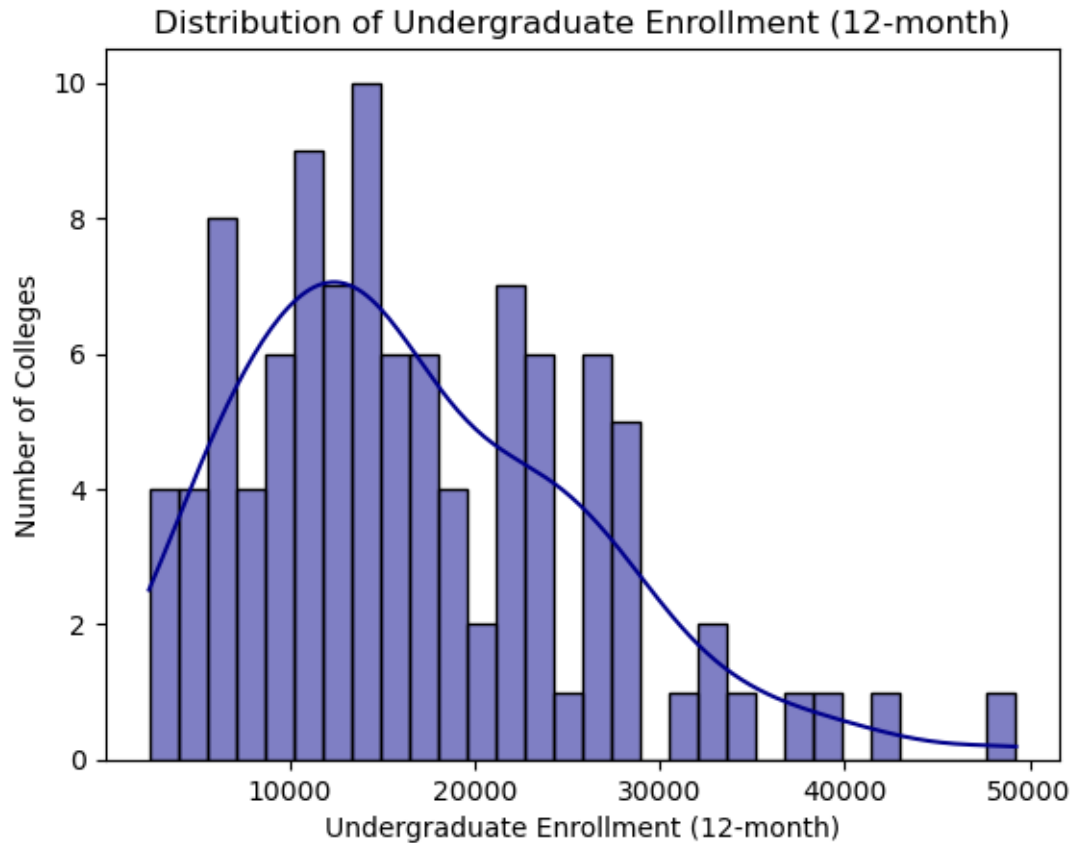
```

```

[27]: sns.histplot(
    data=all_colleges,
    x="latest_student_enrollment_undergrad_12_month",
    bins=30,
    kde=True,
    color="darkblue",
    edgecolor="black",
)
plt.title("Distribution of Undergraduate Enrollment (12-month)")
plt.xlabel("Undergraduate Enrollment (12-month)")
plt.ylabel("Number of Colleges")

plt.savefig('./figures/undergraduate_enrollment_distribution.png',
↪ bbox_inches='tight')
plt.show()

```



```
[ ]: # Rank districts by sum("latest_student_enrollment_undergrad_12_month")

all_districts = cccco.get_ccc_districts()

district_enrollment = all_colleges.groupby("DistrictID")[
    "latest_student_enrollment_undergrad_12_month"
].sum().reset_index()

district_enrollment = pd.merge(
    district_enrollment,
    all_districts[["DistrictID", "DistrictName"]],
    on="DistrictID",
    how="left"
)

to_plot = district_enrollment.set_index("DistrictName")[
    "latest_student_enrollment_undergrad_12_month"
].sort_values(ascending=False).head(10)

# Create a color scale based on the size of the bars
```

```

norm = plt.Normalize(to_plot.min(), to_plot.max())
sm = plt.cm.ScalarMappable(cmap="viridis", norm=norm)
sm.set_array([])

ax = to_plot.plot(
    kind="barh",
    title="Top 10 Districts by Undergraduate Enrollment (12-month)",
    xlabel="Undergraduate Enrollment (12-month)",
    color=plt.cm.viridis(norm(to_plot)),
    edgecolor="black",
)

ax.set_ylabel("District Name")
ax.set_yticklabels(ax.get_yticklabels(), ha='right', position=(0.3, 0))

fig = ax.get_figure()
fig.savefig('./figures/top_10_districts_by_enrollment.png', bbox_inches='tight')
plt.show()

```

