A View of the California Community College System

Ivan Karpov
ikarpov@ucdavis.edu,

Erin Moffat
egmoffat@ucdavis.edu,

Cody Vig
cjvig@ucdavis.edu

March 17, 2025

The source code for this project has been made available on GitHub:
https://github.com/ccc-shopper/ccc-shopper-etl

Abstract

Data on the California Community College system were acquired from the California Community College Chancellor's Office (CCCCO), the Integrated Postsecondary Education Data System (IPEDS) via the Department of Education's College Scorecard, and California's Employment Development Department (EDD). These data provide insights into enrollment trends, program popularity, student demographics, and alignment between educational programs and labor market projections. Sentiment analysis of student perceptions from online reviews is incorporated to provide a qualitative perspective on institutional effectiveness.

The data suggest that the most popular academic programs align typical career-oriented fields such as Business Administration and Psychology. However, a significant disparity between projected labor market demands and program offerings was identified, particularly in high-growth sectors like healthcare support. Success rates across institutions and programs exhibit well-known variation, though all sampled colleges arrive at a success rate of over 60% for each program examined.

This study contributes to a better understanding of the CCC system's role in California's educational and economic landscape. By examining enrollment patterns, student success rates, and workforce alignment, subsequent systemwide research in this area can provide valuable insights for policymakers, educators, and stakeholders aiming to enhance the effectiveness of community college education in the state.

2. Introduction

The California Community College (CCC) system is the largest system of higher education in the United States, which serves over 1.8 million students across 115 colleges via 73 districts. As an essential component of California's workforce development and higher education pipeline, community college institutions provide accessible, affordable, and flexible educational opportunities, including associate degrees, career and technical education, and transfer pathways to four-year universities.

Much of the data collected by these colleges is reported to their central administrative branch, the California Community College Chancellor's Office. Additional data are sent to the Department of Education as regulatory compliance for receiving Title IV funding. Student level data are aggregated by these institutions and made publicly available. These public data offer valuable insights into student demographics, enrollment trends, academic outcomes, and institutional performance.

This paper examines key public datasets related to the CCC system, exploring their structure, accessibility, and implications for educational research. We aim to highlight trends and opportunities within the system, and to discuss how these trends align with the broader state of the labor market.

3. Data Acquisition

The California Community College Chancellor's Office (CCCCO) hosts a REST API that allows the collection of basic Community College metadata information, such as the college name, its geolocation coordinates, the district to which it belongs, and its website. The API also permits the user to query which colleges provide programs that the user specifies, either by passing the name of the program or the program's Taxonomy of Program (TOP) Code.

In order to obtain more data at a college level, CCCO Chancellor's Management Information System's Data Mart was utilized. This is an online data mart, or database, that offers aggregated metrics that each Community College reports at the end of every academic period. This resource allows us to mine popular programs and student success rates per program across colleges. Although the data mart is publicly available, its primary purpose is to provide a centralized repository for colleges to audit their data for accountability purposes. It is therefore not optimized for large queries or automation. It is an ASPX web form with complex payload parameters and necessary cookies, so writing a wrapper that pulled data from the data mart in real time was not feasible. Instead, large extracts were pulled at once, ETL processes were performed on the data, and the data were stored locally in CSV files for analysis.

Data that Community Colleges submit at the federal level are collected via the Integrated Postsecondary Education Data Survey (IPEDS) and housed at the National Center for Education Statistics (NCES), which itself is a subsidiary of the Department of Education. Important metrics include financial aid allocation information, student to faculty ratio, cost of attendance, graduation rates, and median earnings of graduates. IPEDS makes this aggregated data available via Microsoft Access database files that can be used for analysis, but these files are difficult to integrate into environments outside of the Microsoft ecosystem. Instead, we gather important subsets of this data from the Department of Education's College Scorecard tool. This resource pulls data from IPEDS and provides a documented REST API. The difficulty with this API is that the primary keys that index colleges and universities at the federal level are not the same as the primary keys that index Community Colleges at the state level. Instead, we query the College Scorecard API for all institutions that reside in California and fuzzy join them with the college names provided by the CCCCO API.

From here, we aim to collect and analyze data beyond what is reported by colleges to government agencies. Although it is subjective and highly sensitive to extreme opinion, a general qualitative consensus about the college is an important statistic for analysis. To this end, we scrape ratemyprofessors.com for each college in the Community College system and conduct a sentiment analysis on the results in an attempt to ascertain a quantitative metric for student satisfaction. To discover the student sentiments, we took each review from the Rate My Professors College Page and quantified its sentiment using the Python library TextBlob. The sentiments are calculated on a scale from 0 to 1, where a sentiment of 0 indicates a completely negative review, while 1 is a review with nothing but positive sentiment. We would obtain the sentiment for all reviews, compute the average, and repeat the process for every college. In this case, the main difficulty is adjusting the query for every community college. This is done by looking up every school in the list of community colleges and sending a post request to obtain its ID. In addition to the sentiments, we were able to obtain the average professor rating for each school as well.

Additionally, our analysis includes a comparison of recent enrollment trends in various programs and California labor market projections in jobs associated with those programs. Labor market projections, both at the state level and metropolitan statistical area level, were taken from California's Employment Development Department (EDD) Labor Market Information Division. The EDD publishes 10-year labor market projections every two years, with the latest publication being projections from 2020 to 2030 at the time of writing. This API requires the result set to be requested using SQLite syntax.

Considerable difficulty was encountered when trying to reliably map occupations from the labor market projection data to programs at the Community Colleges. Occupations and classes of occupations are both indexed at the federal level by a six-digit Standard Occupational Classification (SOC) Code. Similarly, California Community College programs are indexed at the state level by a six-digit Taxonomy of Programs (TOP) Code. Because

SOC Codes are a federal index and TOP Codes are a state index, there is no universal dictionary between the two. Instead, state-level TOP codes were matched with federal-level Classification of Instructional Programs (CIP) Codes, maintained by the NCES, using a crosswalk hosted by the CCCCO. Finally, the NCES, in collaboration with the Bureau of Labor Statistics, provides and maintains a crosswalk between CIP Codes and SOC Codes that was ultimately used to map TOP Codes to SOC Codes. Neither the SOC-CIP crosswalk nor the CIP-TOP crosswalk is a one-to-one mapping, so as a result, a single community college program can map to many occupations, as is expected.

4. Data Analysis

Community College Affiliation

The Community College system exhibits a generally hierarchical structure, in which colleges report to district offices, and the district offices report to the Chancellor's Office. Some districts oversee only one college, while others contain several colleges. The distribution of district allocation to colleges is shown in Figure 1.

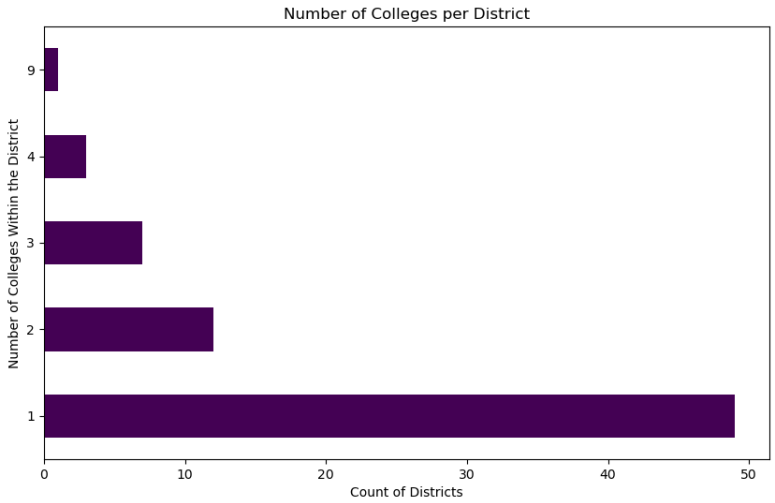

Figure 1

Of the 73 community college districts, 49 operate as single-college districts. The Los Angeles Community College District has the highest number of affiliated colleges, totaling nine. The largest multi-college districts are the Los Angeles CCD (with 9 affiliated colleges), Los Rios CCD (4 affiliated colleges), Peralta CCD (4 affiliated colleges), and State Center CCD (4 affiliated colleges). The distribution of the most recent academic year's enrollments,

according to data reported by the District to the Chancellor's Office, at the top ten largest districts in the system is shown in Figure 2.

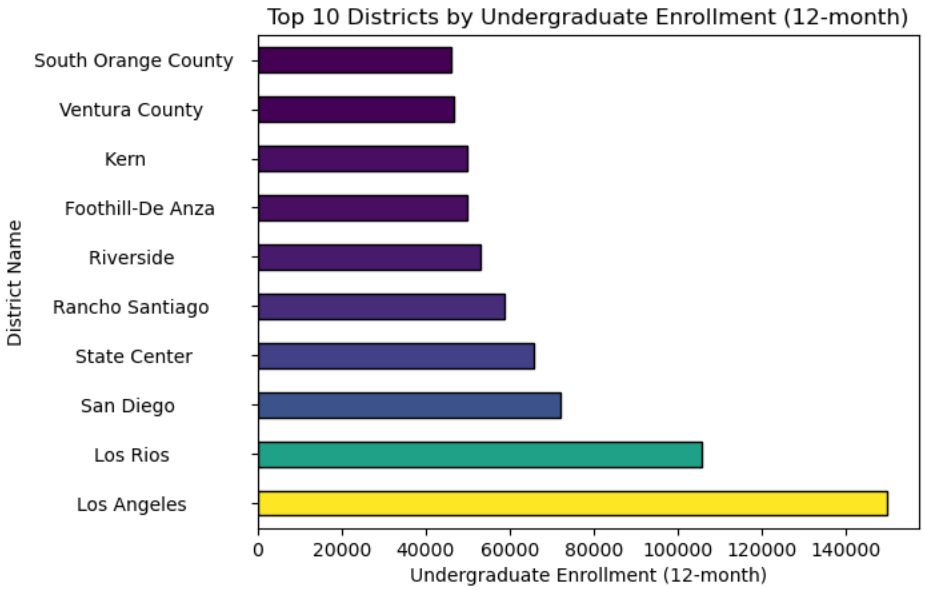

Figure 2

As would be expected, Los Angeles CCD is tightly clustered in the Los Angeles area, Los Rios is similarly clustered in Sacramento, and Peralta is located in the northern Bay Area. These findings may suggest that the size of the district is highly correlated with the size of the service area. State Center CCD, however, is centralized near Fresno, suggesting its size is likely more for the benefit of resource allocation than it is for the breadth of the service area. The full geographic distribution of colleges and districts is included in Figure 3.

Figure 4 shows the enrollment distribution across the system. The enrollment distribution is highly right-skewed due to unconventionally large colleges, like East Los Angeles College (49,266 reported enrollments in the 2023-2024 academic year) and American River College (41,538 enrollments in the same period). Generally, the modal enrollment is around 12,000. The largest colleges tend to be in large metropolitan areas like Los Angeles and Sacramento, while the smallest colleges tend to be in sparsely populated regions in northern California, like College of the Siskiyous (which reported 932 students) in Weed, CA, and Lassen Community College (which reported 1,431) in Susanville, CA.

Major Popularity Analysis

Since current majors are not listed for students, one way to examine what majors people are interested in is to examine program awards, in order to view what programs were completed.

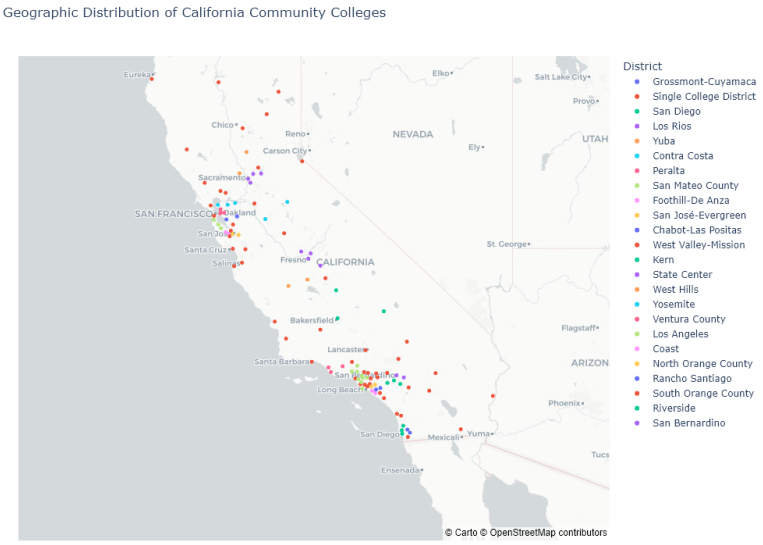

Figure 3: An interactive version of this plot is available on the GitHub repository.

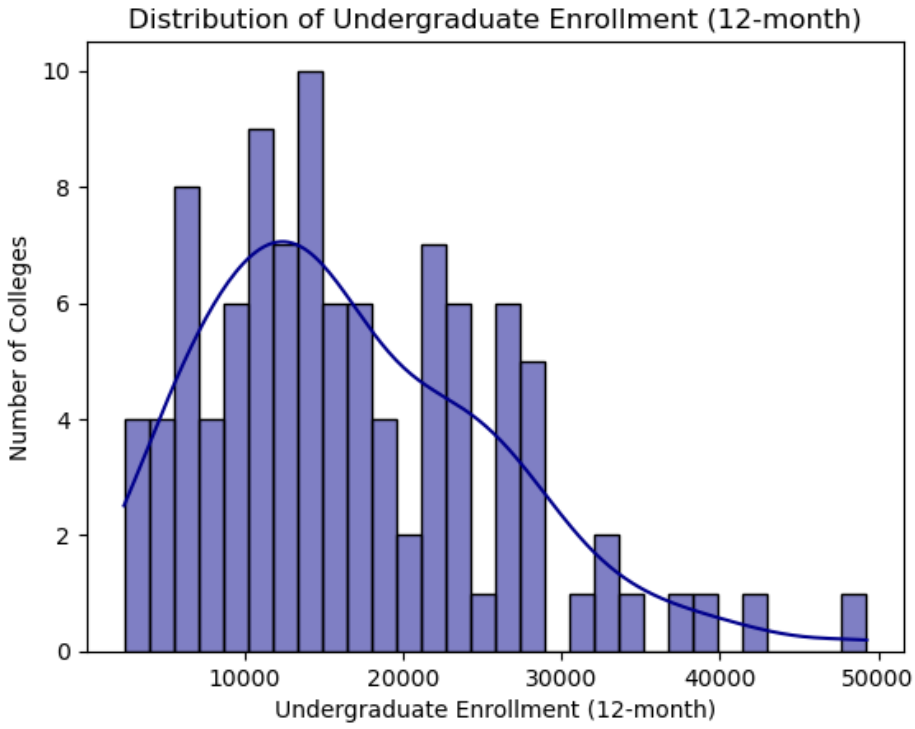

Figure 4

Program awards within the CCCCO Data Mart represent Associate degrees, certificates and noncredit awards. In 2023-2024, across the state of California, the most popular program

of study that students received awards for was "Transfer Studies", which is indicative of interest in continuing education through a Bachelor's degree before entering the job market. In addition, the second most popular program award, Liberal Arts and Sciences, is also representative of students' desire to transfer. Clovis Community College (n.d.) explains this program allows students to earn credits for transfer while determining or exploring their field of study. Exclusion of these two programs allows for a better understanding of the specific fields of study students are interested in.

Analysis of the other program awards helps generate a better understanding of the types of fields students may be pursuing. In order to conduct an analysis, we assessed current results that summarize the number of awards across each college per program. A bar chart follows this text, illustrating the most frequent awards distributed in 2023-2024, excluding "Transfer Studies" as well as Liberal Arts and Sciences. The most frequent program awards for more specific programs include Business Administration; Biological and Physical Sciences (and Mathematics); and Psychology, General. These awards reveal that students are completing these programs at a higher level, surpassing award counts for other programs, such as Sociology, by more than 10,000. It is of interest to mention that the most common program awards illustrate student interest across a variety of fields, from fields related to the Social Sciences, to fields related to other specializations such as Healthcare or Accounting. This information is helpful in gaining an understanding of what fields community college students across California are currently most interested in.

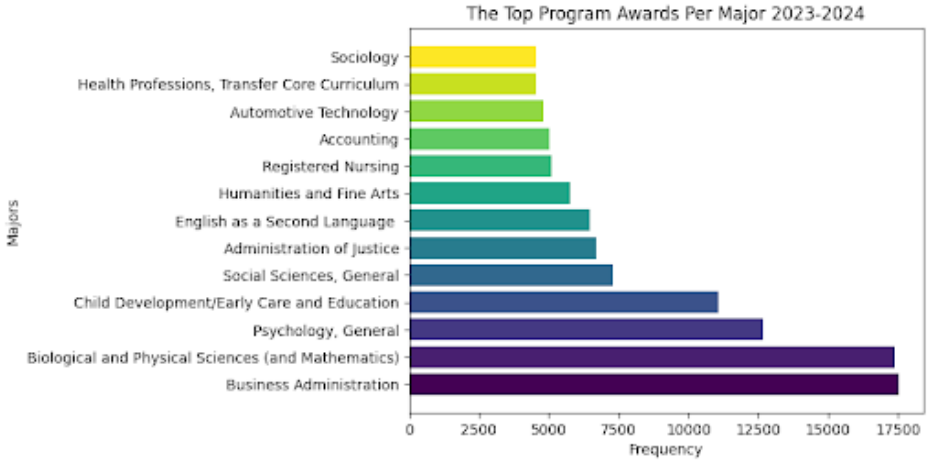

Figure 5

Success Rate: A View of Specific Program Level Success Across Colleges

In order to assess program offerings, success rate is one metric that can be pictured across colleges. Gathering current program relevant data was of interest as current information

is important for students who are considering a program. The success rate per college is an important factor to visualize; however, success rate is only one variable. We are not making statistical conclusions based on the success rate as success rate can be determined by the number of students. If less students enroll in the class then the success rate will be impacted highly by one student not succeeding in the course. The CCCCO Data Mart calculates success rate as the number of students who succeeded in a course per term divided by the number enrolled in it. Overall success rate was visualized per program across Fall and Spring 2024, where we calculated the success across both of these semesters or quarters. This method allows us to gain a more general understanding of success across colleges. In addition, due to a variety of types of instruction, success rate was assessed for "Non Distance Education Methods Total", which incorporates information for students in the specific areas of the colleges we are assessing, which allows information to be gained for students planning on attending a specific college in person. In addition, to visualize how success changes across colleges, we can do a case study of the three most popular specific program awards mentioned above.

Business Administration Course Success Rate

Students most commonly pursued Business Administration program awards, so it was of particular importance to gather student success information for this program. In order to gain an understanding of success rate, the colleges with the highest success for this program were visualized. All but a few colleges had over a 60 percent success rate for Business Administration, with approximately 77 percent success rate for the top 30 success rates per college. The five colleges with the highest percentage of success had around a 90 or higher percent success rate. These colleges include: Columbia College, San Joaquin Delta College, Mendocino College, Laney College and Victor Valley College.

Biology, General Course Success Rate

The Biological and Physical Sciences (and Mathematics) program award focuses on three key areas of study. In order to visualize course level success for this program across colleges, the Biology, General course success rate was compiled. The distribution of success rates across colleges was above 60 percent for most colleges, and the top thirty success rates was at around 80 percent. The 5 colleges for these two terms with the highest success rates were Lassen College, De Anza College, Palo Verde CC, West Hills Coalinga College, and Sierra College.

Psychology, General Course Success Rate

Psychology is another program that students are receiving awards for at a higher level. The data was again calculated for Fall and Spring 2024, and there was again above a 60 percent success rate for almost all colleges, other than two. For this term, the overall success rate was highest for a different set of colleges: West Hills Coalinga College, Oxnard College, College of the Siskiyous, De Anza College, and Columbia College. Success rate was approximately 80 percent and above for the top thirty success rates per college for this program, similar to the Biology, General course.

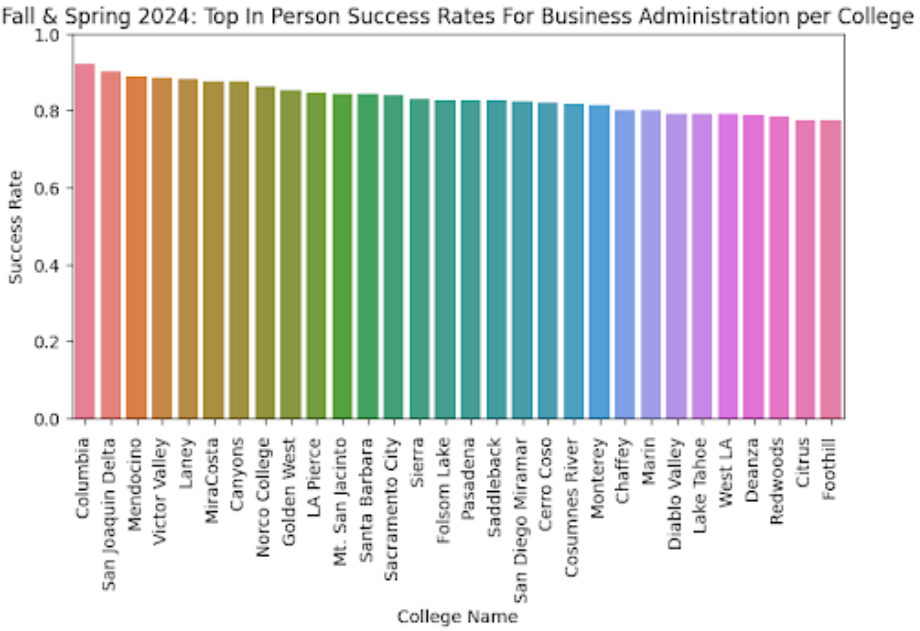

Figure 6

Student Demographics Across Selected High Success Rate Colleges

In order to assess the demographics across the colleges with the top 5 success rates per the courses we discussed previously, information on student counts can be viewed using the CCCCO Data Mart. Data was collected from the most recent year, 2023-2024. Student counts varied greatly across the colleges this year, from counts of more than 60,000 students to counts of less than 10,000. De Anza College, Sierra College, and Victor Valley College had tens of thousands more students than Columbia College or Lassen College.

Ethnicity was another category that varied across these colleges. For instance, Sierra College had a higher proportion of White Non-Hispanic students than Oxnard or De Anza College, despite all three colleges having high student counts in comparison to the other

selected colleges. Oxnard College and Victor Valley College also have higher counts of Hispanic students. In comparison across all of these colleges, African American students do not represent a large proportion of the students, but Laney college appears to have a higher proportion of African American students compared to the other colleges.

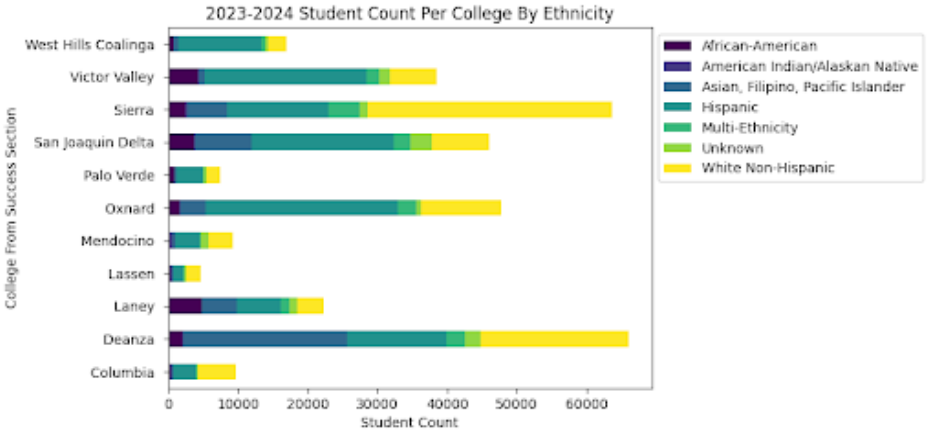

Figure 7: An interactive version of ethnicity is available here.

Gender differences are also visible across these selected colleges. Most colleges appear to have a higher ratio of female students, but Palo Verde and Lassen College both appear to have a higher proportion of male students instead of female students. Non-binary students don't represent a large proportion of students across any of the selected colleges, and it is notable that San Joaquin Delta College appears to have less non-binary students recorded than Oxnard College, despite having similar student counts for this year.

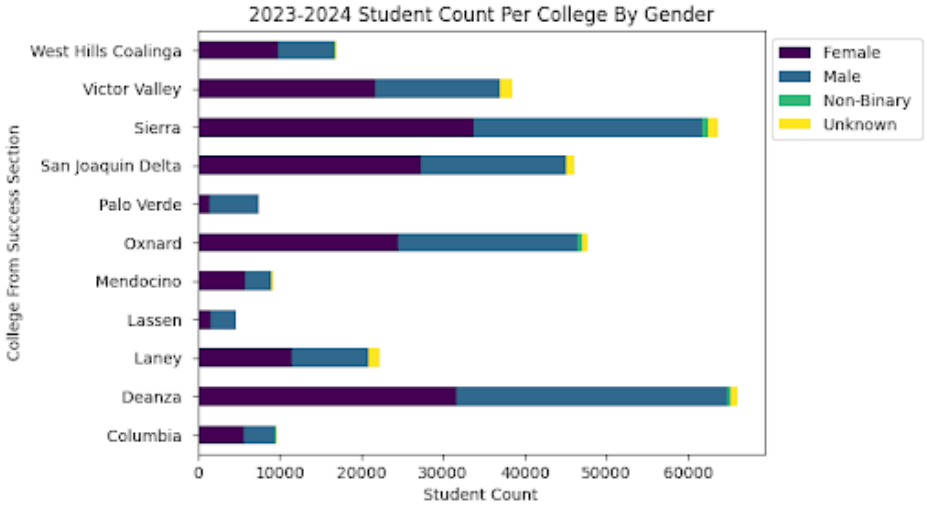

Figure 8: An interactive version of gender is available here.

Sentiment Analysis of California Community Colleges

After retrieving sentiments from each college, we found the 5 colleges with the highest and lowest sentiment. The college that students felt the happiest about was Fullerton College with a sentiment score of 0.3875, followed by West Los Angeles College, San Diego Miramar College, Porterville College, and Pasadena City College. The college that students disliked the most was Moreno Valley College with a sentiment of just 0.3431, followed by Mission College, Ohlone College, Merced College, and Irvine Valley College.

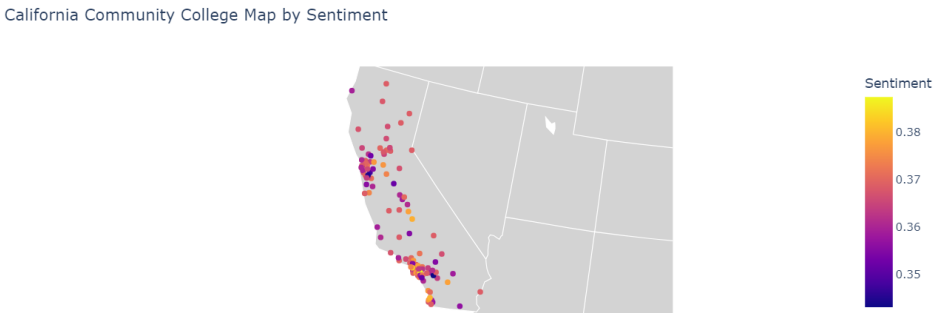

Figure 9

Figure 9 maps all of the community colleges in California, and is color-coded based on sentiment.

We also were interested in determining whether there is a link between the average professor rating at the community college and the overall sentiment. However, we found almost no correlation between these 2 variables, as seen in the graph in Figure 10 (the horizontal axis is the sentiment and the vertical axis is the average professor rating from 1 to 5.)

We found out that the top 5 schools based on rating were mostly different than the top 5 schools based on sentiment. The only college that appears in both lists is Porterville College. With an average sentiment of 0.3795 and an average rating of 4.0795, it is the best community college choice based on these 2 criteria.

Enrollment Trends vs Labor Market Projections

Every two years, the State of California's Employment Development Department publishes a 10-year projection of occupational and industrial labor market changes. These data are aggregated at the metropolitan statistical area (MSA) level. Colleges were partitioned into MSAs by matching them based on the county their main campus is located. In 18 of the

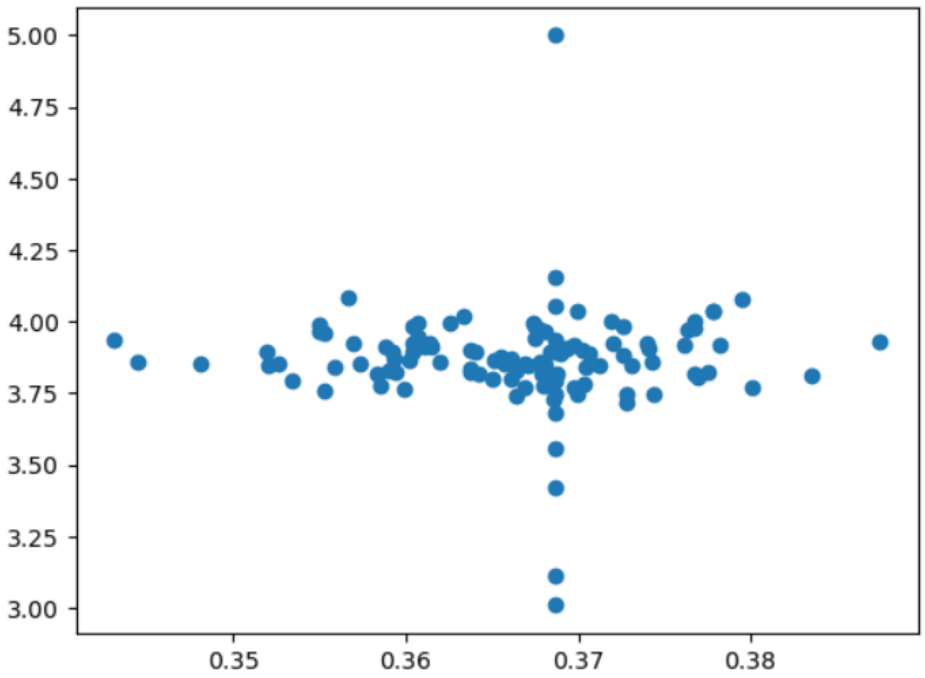

Figure 10

29 metropolitan service areas in California, home health and personal care aides occupations were projected to see the largest overall growth out of all occupations. Of note, no community colleges in any of these key MSAs offer a program that directly correlates with this occupation. This may in part be because these colleges prefer to offer programs with more breadth, but such programs are not uncommon elsewhere. In fact, there are 64 programs across 12 colleges that have a TOP Code that is correlated with either home health or personal care aides, as shown in Figure 11. Prospective students who have an interest in this field would be best served by attending one of the colleges in the Los Angeles-Long Beach-Glendale metropolitan district.

The most popular programs among community college students were determined to be those in the biological and physical sciences (and mathematics). This category is so broad that it would provide little insight to examine the labor market projections for this field. Instead, we examine occupations in the next most popular program award, psychology.

The largest percent change projected for psychology-related occupations is an 18.2% change for Clinical, Counseling, and School Psychologists in the Kings County area. This is a deceptive metric, as this corresponds to a growth of only 20 jobs. The largest projected numerical growth is from 2,790 jobs to 5,260 jobs (a growth of 470) for the same occupation in the Los Angeles-Long Beach-Glendale metropolitan district. Gathering the total numeric change for all psychology-related occupations and partitioning by MSA, we arrive at the distribution in Figure 12.

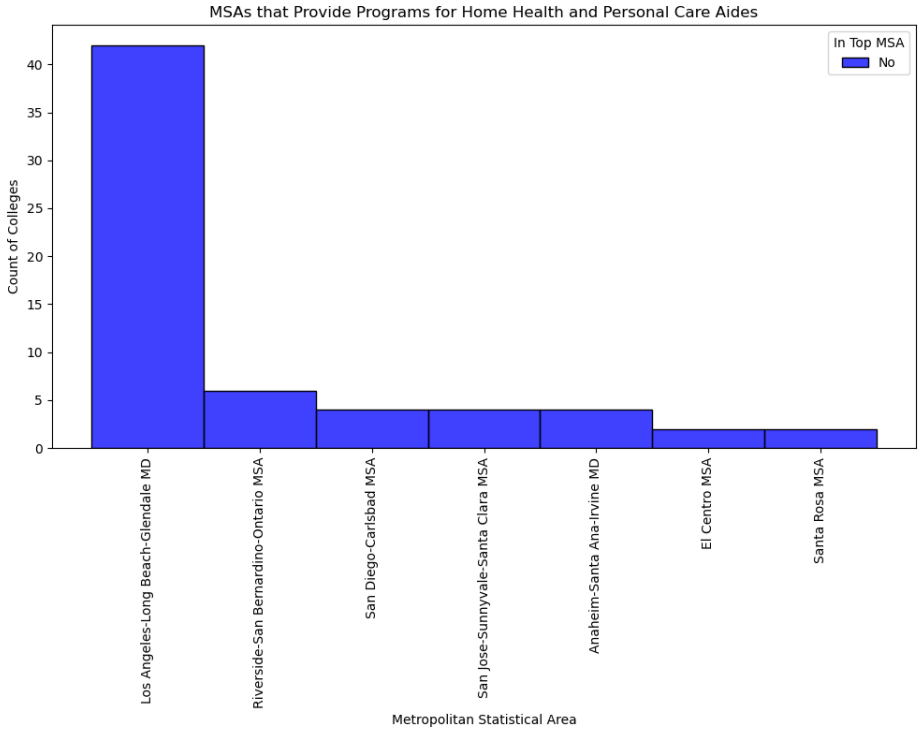

Figure 11

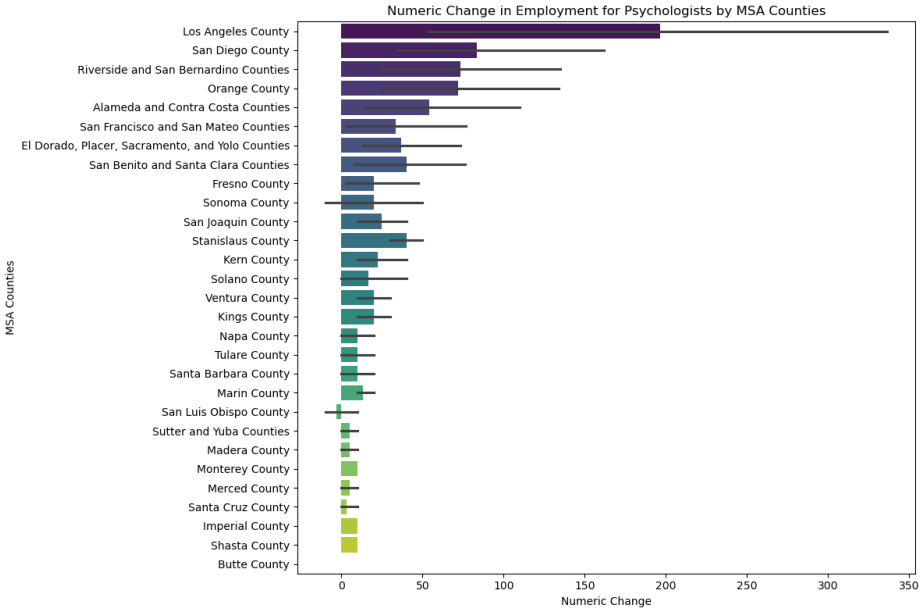

Figure 12

Prospective students with an interest in psychology would be best served by attending a college in Southern California and perhaps seeking an internship during their time in the

area. Of course, students who have a career interest in psychology do not tend to get their terminal degree from a community college, so there are additional factors to consider, but there does appear to be a stronger correlation between enrollment trends and labor market projections in Southern California schools than there is in Northern California schools.

5. Discussion

One of the research topics we pursued was to determine what students are currently studying, and this was determined using program award outcomes. The results of program outcomes per major across colleges pave the road for us to gain an understanding of the topics that students are interested in. For the purposes of analysis, program outcomes were summarized into one numerical unit of program awards per major; however, there is more depth to these results, as each numerical unit can include a certain number of varied degrees. For instance, the total number of Biology program awards can include A.S.-T degrees, A.S. degrees, and certificates. Further investigation could reveal whether colleges have more popularity in a certain type of degree or certificate within these program awards. Also, the most popular program award was "Transfer Studies".

The goal of the California Community College system is to provide transfer preparation, career-technical education, and remediation. The analysis herein is largely tailored to the second and third of these core principles, but arguably the largest cohort of community college students are the ones whose educational goals are to transfer to a four-year institution. Future work could be dedicated to analyzing the pipelines between the community colleges and the UCs and CSUs. These data were omitted for the purposes of this document because they are generally not available at the same level of disaggregation as the data we have chosen to examine. The University of California does host a Tableau workbook that contains data on transfer majors broken down by transfer institution and home institution, and while the CSUs do make available some disaggregated transfer data, it is not disaggregated in a form that is useful to our analysis.

The National Student Clearinghouse is a nonprofit organization that focuses on student journey statistics and tracks student enrollment and graduation at all institutions that participate in their data surveys. While they do not make much of this data publicly available, any college that participates can request this data for any subset of students that have attended their institution. As a result, many community colleges do have data on where their students go when they graduate, but many of them choose not to publish their data because it is not frequently used for institutional accountability purposes. Nevertheless, some institutions do make this data available, and that data can be used to supplement the analysis provided here or facilitate new research in this area.

Additionally, it is of importance to understand student outcomes. Success rate is one factor that could influence a student's decision to maintain or stay in a course, but it is impacted by outside factors. Our report took a generalized look at student success across just a few courses in order to visualize student success, but statistical claims can't be made. Findings showed success rate varied across colleges, but there are other factors that were not taken into account. Success rate is determined by many factors such as race, gender, and first-generation status (Deshonta & Slate, 2017). Although our analysis focused on overall student success, further research should be done to visualize success in specific courses by gender and ethnicity, as taking a general approach loses important specific information in the process.

We visualized student headcount across the campuses with high success rates and saw that the counts of students varied across campuses as well as genders and ethnicities. The US Census APIs have information that could allow comparison of the gender and ethnicity of students and faculty to the gender and ethnicity distribution of the service area (although the service area of a college is not always exactly a county, so the mapping may not be easy).

6. Conclusion

Overall, the extraction of data was incorporated from several key sources: the CCCO REST API, the CCCO Data Mart, the Department of Education's College Scorecard, and ratemyprofessors.com. These key sources allowed information about California's community colleges to be compiled and explored. The geographic location of the colleges and districts allowed an understanding to be gained of community colleges across California, where a visualization then showed how the size of colleges was associated with the size of their service area. Specifically, Los Angeles district had the highest number of community colleges. The most frequently distributed program awards across the California Community Colleges in 2023-2024 illustrate the types of programs students may currently be interested in, where Business Administration was found to be the most popular program award, after exclusion of two majors determined to provide information on transfer desire instead of specific topic studied. California's Employment Development Department (EDD) Labor Market Information Division allowed projections to be made for occupations, including psychology occupations, which was a frequently pursued program by community college students. It was found that areas in Southern California had higher projections for job growth. In addition, Rate My Professor allowed an analysis of student opinions to be made, where there was variation across colleges based on a sentiment analysis. Overall, we have researched information on California Community College's locations and service areas, enrollment, programs of study and job market information, which has allowed insight to be gained.

7. References

Clovis Community College. (n.d.) Liberal Arts.
Retrieved March 3, 2025,
From: https://www.clovis.edu/liberalarts/index.aspx

Deshonta L. H., & Slate J. R. (2017). *Differences in GPA by Gender and Ethnicity/Race as a Function of First-Generation Status for Community College Students Global Journal of Human-Social Science Research*, 17(3), 1-5.

California Community Colleges Chancellor's Office. (n.d.).
https://datamart.cccco.edu/datamart.aspx

California Employment Development Department. (n.d.). Labor Market Information.
https://labormarketinfo.edd.ca.gov/

CCCO Api. (n.d.).
https://api.cccco.edu/

CSU The California State University. (n.d.). Reports and Analytics.
https://www.calstate.edu/data-center/institutional-research-analyses/Pages/reports-and-analytics.aspx

National Student Clearinghouse. Powering the Future of Education.
https://www.studentclearinghouse.org/

Rate My Professors. (n.d).
https://www.ratemyprofessors.com/

United States Census Bureau. (n.d.).
https://www.census.gov/data/developers.html

University of California. (n.d.). Transfers by Major.
https://www.universityofcalifornia.edu/about-us/information-center/transfers-major

U.S. Department of Education. (n.d.). College Scorecard.
https://collegescorecard.ed.gov/