

✔ Query executed successfully.

How I Developed and Utilized a (Pretty Neat) Database

By Sam Deery-Schmitt

INTRODUCTION

A1 College Prep is a tutoring organization based in Koreatown in Los Angeles, CA of which this author is an employee. A1 primarily works with high school students, although we work with middle school students and college students/graduates as well. We specialize in three areas: academic tutoring, subject test tutoring, and college admissions. Data is currently collected and stored in a number of ways, which makes it difficult to develop marketing strategies or assist students with applications without asking our administrator for specific information. Primarily, a third-party-managed software client is used to input client information about both parents and students, which is also used for payroll and billing purposes. It is difficult to use this software to determine key information about students, such as when they will graduate/graduated high school, especially if seeking information about many students at once. It is impossible to use this software for other information that is needed for marketing or college applications; for example, information about parents' educational backgrounds is not captured there. Some processes are planned on paper or spreadsheets, such as what colleges students will apply to and what college counselors will work with them, but there is not a single source of truth for this information, i.e. often a counselor will need to ask both their students and the A1 administrator and management multiple questions to determine exact information. The objective of this database is to create one location where any information about a student or their parent(s) can be stored and then retrieved by anyone in the organization, including schools a student is applying to, whether they got it, where they went (or currently go), what their parent's or parents' email address is/are, etc. Eventually the database may be expanded to include more information relevant to the business, such as a student's performance in school, payments for services, responses to internal questions ("How was A1 most helpful for you in your college search? What could have A1 done better?") and could even expand to employees.

Stakeholders

- Administrator
 - Responsibilities: input existing and past student data into the database (update student and parent information especially when taking in new students)
 - Using the database: example use—generate mailing lists unique to students and parents who meet specific parameters, such as students enrolled in standardized test services but not college admissions services
- Founder
 - Responsibilities: none, except assisting with determining what information should be captured about students and parents
 - Using the database: gain insight into success of tutoring services with specific students, for example—see how a student's test score has increased over time, what colleges they were admitted to
- Chief Academic Officer

- Responsibilities: assist with determining what information should be captured about students and parents, provide admin with data on leads, update student, parent and institution data as needed
- Using the database: gain insight into success of tutoring services, examine relationships between leads and clients, answer specific questions about students/their parents when needed
- Me (Chief Business Officer and College Counselor)
 - Responsibilities: assist with determining what information should be captured about students and parents, provide admin with data on leads, update student, parent and institution data as needed
 - Using the database: Gain insight into success of tutoring services, leverage data to create marketing and sales campaigns, answer specific questions about students/their parents when needed. Examples—be able to determine the percentage of our students who successfully apply to a given school, or a given category of schools; contact 50 former students and ask for a short quote about their time with A1; identify A1 alumni willing to work with applicants to current school
- Tutors
 - Responsibilities: none currently
 - Using the database: retrieve information needed to assist students with college applications, such as citizenship

Business Rules

A student can apply to zero or many institutions.

A student can apply in three or more ways: early action, early decision, or regular decision.

Early decision is binding; the other application methods are not.

A student can attend community college and earn a two-year degree.

A student can also transfer to another institution from a community college to pursue a bachelor's degree.

A student can apply as a freshman or transfer student.

A student can apply to undergraduate, graduate, and professional programs.

A student may or may not be a U.S. citizen.

A student's parents may or may not be U.S. citizens.

A student's parents may or may not speak English.

An institution can accept, reject, or waitlist a student.

A student's parents may be their biological parents, adoptive parents, or legal guardians.

Parents may have one or many students using A1's services.

Students and parents may have native-language names and English names.

Glossary

Client: student + parent(s)

Student: middle school, high school or college student using A1's services; also includes college graduates preparing for standardized tests and/or graduate/professional school.

Institution: college tends to be used when discussing colleges and universities; here, institution refers to any degree-conferring institution of higher education, including community colleges, 4-year colleges and universities, and professional schools.

Similarly, college application is a stand-in for any application to an institution of higher learning.

Reach school: an institution that the student is statistically unlikely to get into, but would match the student's personality and interests well

Match: a school the student has good chances of getting into and that matches the student's personality and interests; may not be as desirable for clients as reach schools.

Safety: a school the student faces near-certain odds of getting into, which may match the student's personality and interests but is not as desirable as a target or reach school.

Data Questions

What percentage of our students, on an annual basis and overall, are admitted to their: reach schools? Match schools? Either?

How many 1st generation students do we work with annually and overall?

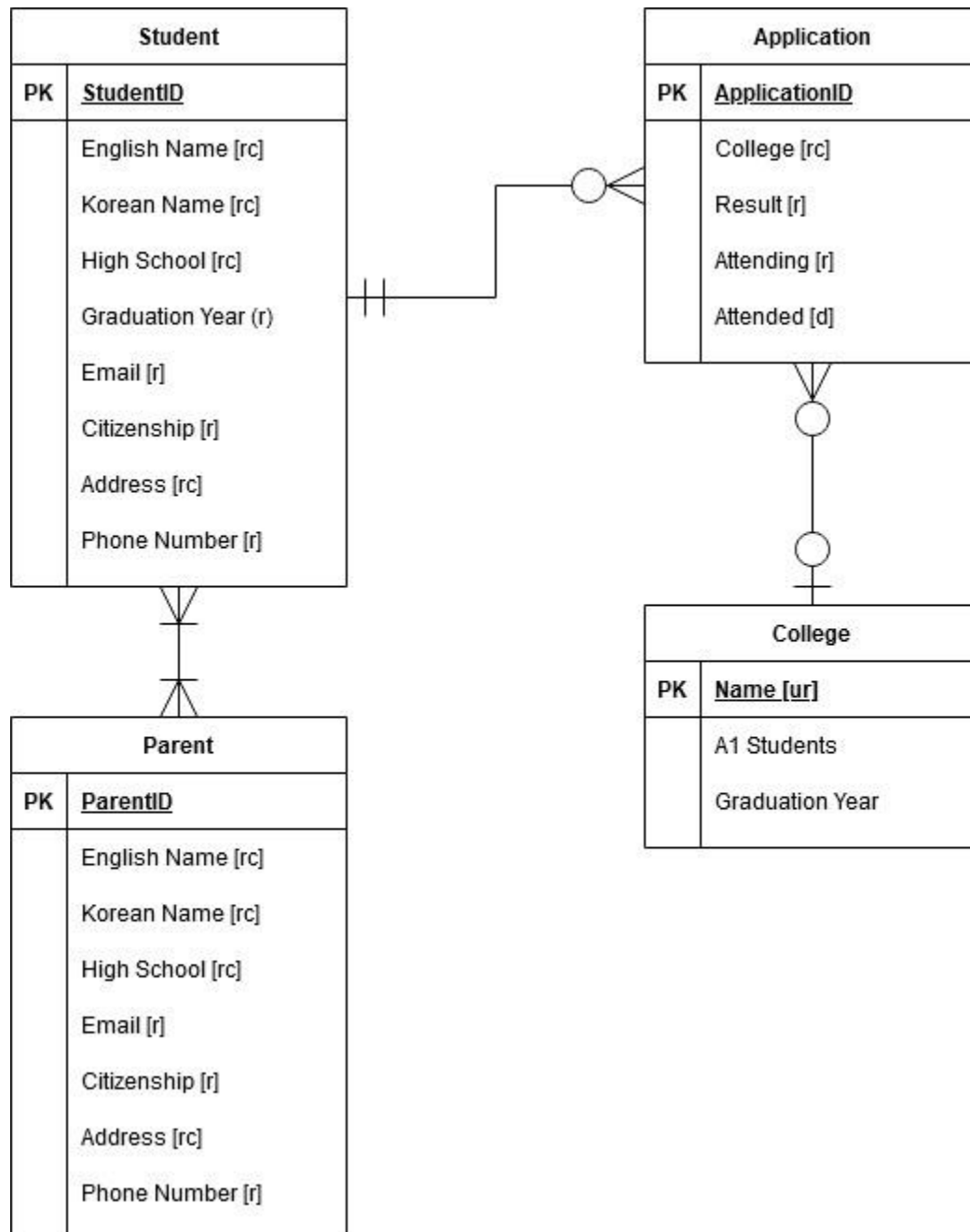
How many of our clients live in the United States outside of Los Angeles? How has this changed from year-to-year?

What are the institutions our students apply to the most? What are the institutions our students receive the most acceptances from? What are the institutions where most of our students go?

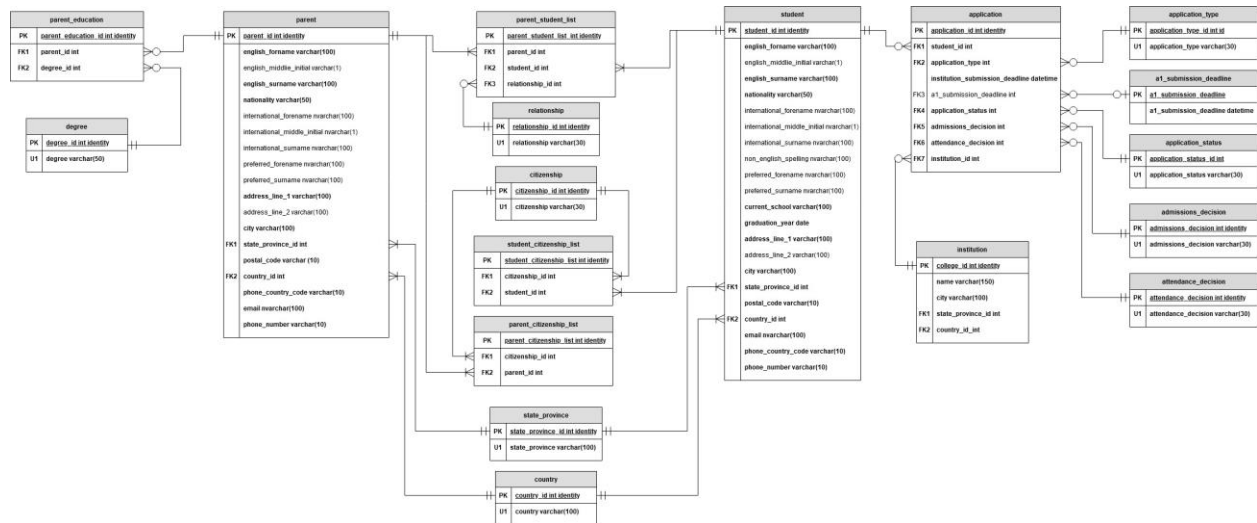
Do our international students have a better track record getting admitted than domestic students?

Do our students gain admissions to highly competitive programs?

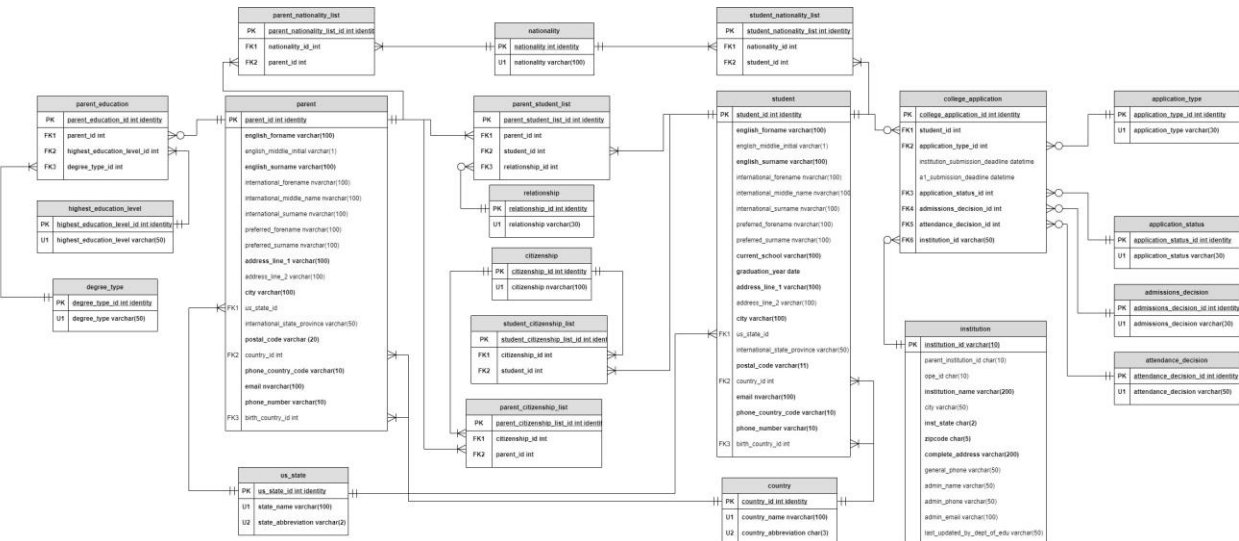
ERD



E-ERD, Deliverable 1



E-ERD, Deliverable 2



METHODS

The scripts used to create the schema, populate the database, and perform analytics against the database are available in 'scripts' folder of the github repository for this project. My personal github can be found at <https://github.com/sampds>.

DISCUSSION

✓ Query executed successfully.

I can safely run the entire script without any problems. All the tables have the records they should, and my views and procedures all work.

There are many areas I want to improve, such as the actual efficiency of the database, which would could be optimized by checking the actual maximum lengths of records, converting many varchar's to chars, and building indexes.

I also want to properly use an ORM to populate the table, instead of using stored procedures and R.

But this is a great starting point. I have a central source of truth for A1 College Prep, which is something that will help us become much more efficient and deliver better service.

Just the simple views and procedures I created to look at large amounts of demographic data at once will be immensely helpful when it is time to help students fill out many college applications again.

Keep in mind, as you saw, none of these students or parents are real people.

Here are the results of EXEC StudentDemographicDashboard
And EXEC ParentDashboard

Results Messages

	Parent	Student	RelationshipToStudent
1	Marquita Wellard	Tedd Dessent	Legal Guardian
2	Allin Beadon	Ivan Dickons	Mother (Biological)
3	Cate Markussen	Devin Harker	Legal Guardian
4	Quentin Prium	Manyo Orange	Legal Guardian
5	Dorine Bleything	Ede Stoakes	Aunt
6	Julianne Landsman	Judas Giriardelli	Mother (Biological)
7	Vanna Devita	Lemmy Gheorghe	Father (Biological)
8	Miran Mates	Teodorico Van Arsdall	Legal Guardian

	EnglishName	InternationalName	PreferredName
1	Tedd Dessent	NULL	NULL
2	Ivan Dickons	欣妍 章	NULL
3	Devin Harker	NULL	NULL
4	Manyo Orange	NULL	NULL
5	Ede Stoakes	NULL	NULL
6	Judas Giriardelli	NULL	NULL
7	Lemmy Gheorghe	军卿 劳	NULL
8	Teodorico Van Arsdall	NULL	NULL

	Student	Citizenship	
1	Tedd Dessent	United States of America	
2	Ivan Dickons	United States of America	
3	Devin Harker	United States of America	
4	Manyo Orange	Korea, Republic of	
5	Ede Stoakes	United States of America	
6	Judas Giriardelli	United States of America	
7	Lemmy Gheorghe	United States of America	
8	Teodorico Van ...	Korea, Republic of	

	Student	Nationality
1	Tedd Dessent	South Korean
2	Ivan Dickons	American

Results Messages

Parent	Student	RelationshipToStudent	
1 Marquita Wellard	Tedd Dessent	Legal Guardian	
2 Allin Beadon	Ivan Dickons	Mother (Biological)	
3 Cate Markusen	Devin Harker	Legal Guardian	
4 Quentin Prium	Mayjo Orange	Legal Guardian	
5 Dorine Bleything	Ede Stoakes	Aunt	
6 Julianne Landsman	Judas Giriardelli	Mother (Biological)	
7 Vanna Devita	Lemmy Gheorghe	Father (Biological)	
8 Miran Mates	Teodorico Van Arsdall	Legal Guardian	

EnglishName	InternationalName	PreferredName
1 Marquita Wellard	NULL	NULL
2 Allin Beadon	NULL	NULL
3 Cate Markusen	NULL	NULL
4 Quentin Prium	NULL	NULL
5 Dorine Bleything	NULL	NULL
6 Julianne Lands...	NULL	NULL
7 Vanna Devita	NULL	NULL
8 Miran Mates	茹雪 茹	NULL

Parent	HighestEducationLevel	DegreeType
1 Homer Mussilli	Graduated from trade school or community college	Associate's (AA, AS)
2 Beauregard Durbyn	Graduated from trade school or community college	Associate's (AA, AS)
3 Pearl Jeacop	Graduated from trade school or community college	Associate's (AA, AS)
4 Saul Lecordier	Graduated from trade school or community college	Associate's (AA, AS)
5 Rhianna Seefus	Graduated from trade school or community college	Associate's (AA, AS)
6 Ellary Couves	Graduated from trade school or community college	Associate's (AA, AS)
7 Duky Fowler	Graduated from trade school or community college	Associate's (AA, AS)
8 Lionello Allawy	Graduated from trade school or community college	Associate's (AA, AS)

Parent	Nationality
1 Marquita Wellard	South Korean
2 Allin Beadon	American
3 Cate Markusen	American

These two procedures combine several views that, when seen together, allow the user to do quality control on potentially all student applications at once. We don't really have anything like that at the moment.

But what I'm really excited about is the analysis I'm going to be able to perform on the performance of our students, and, in turn, the efficacy of our services.

This simple View I named CollegeAppDashboard demonstrates the power of this database:

Results		Messages					
	Student	HSGradYear	School	Ranking	SubmissionStatus	Decision	Attendance
1	Melina Farrin	2020	California Institute of Technology	9	Submitted	Accepted	Decided to attend
2	Melina Farrin	2020	Columbia University in the City of New York	3	Submitted	Accepted	Declined to attend
3	Melina Farrin	2020	Cornell University	18	Submitted	Accepted	Declined to attend
4	Lilyan Boldecke	2020	Duke University	12	Submitted	Accepted	Declined to attend
5	Maryjo Orange	2019	Georgetown University	23	Submitted	Accepted	Decided to attend
6	Lilyan Boldecke	2020	Georgetown University	23	Submitted	Accepted	Declined to attend
7	Ted Burgh	2019	Harvard University	2	Submitted	Accepted	Declined to attend
8	Maryjo Orange	2019	Massachusetts Institute of Technology	4	Submitted	Accepted	Declined to attend
9	Melina Farrin	2020	Northwestern University	9	Submitted	Accepted	Declined to attend
10	Ted Burgh	2019	University of California - Los Angeles	20	Submitted	Accepted	Decided to attend
11	Maryjo Orange	2019	University of California, Berkeley	22	Submitted	Accepted	Declined to attend
12	Lilyan Boldecke	2020	University of Notre Dame	19	Submitted	Accepted	Decided to attend
13	Maryjo Orange	2019	University of Pennsylvania	8	Submitted	Accepted	Declined to attend
14	Ted Burgh	2019	Washington University in St Louis	16	Submitted	Accepted	Declined to attend

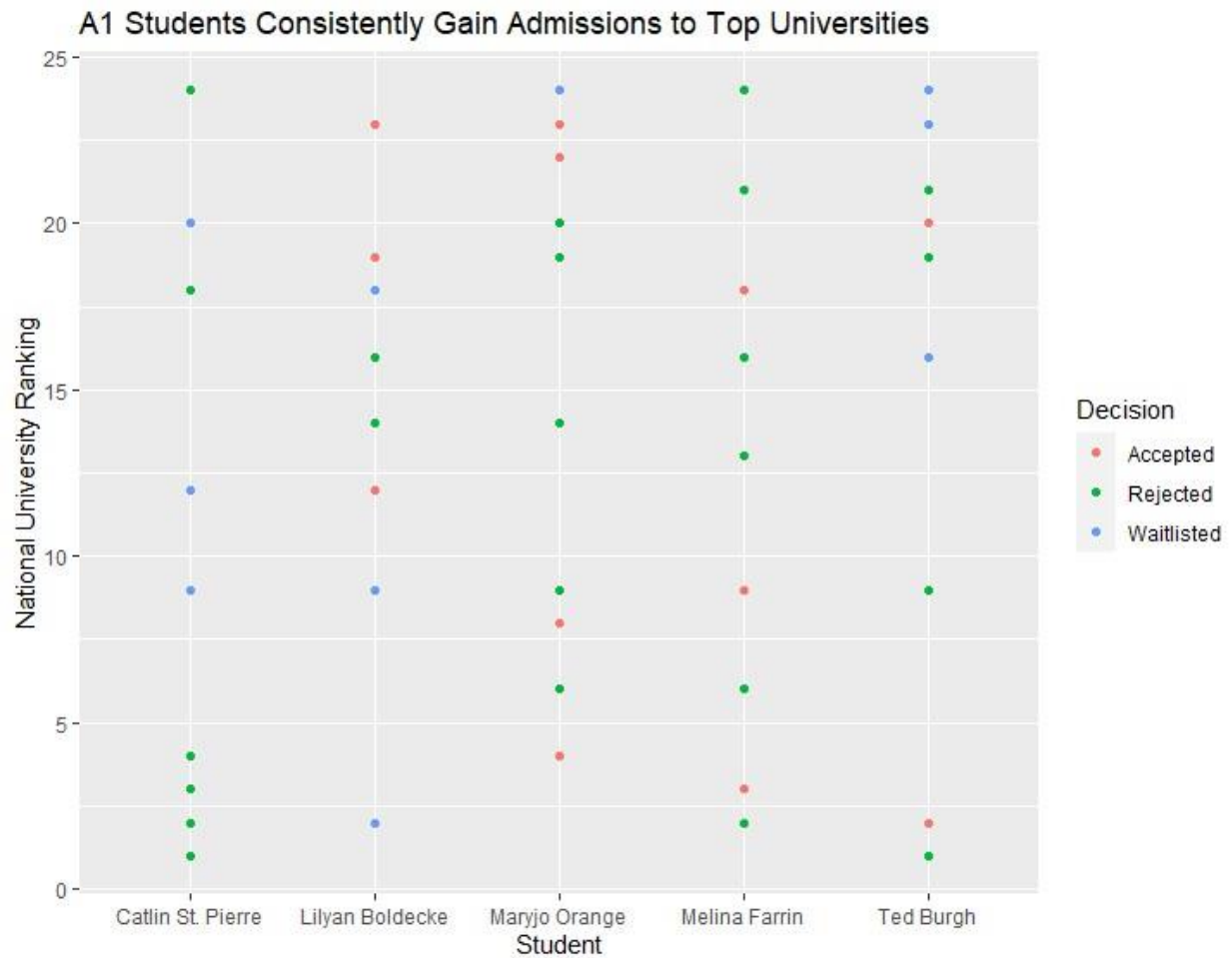
With one keystroke, I can all current, pending, and past admissions decisions. I can answer many of my business questions just by running some COUNT(s) on this View. I can say with certainty how many students were admitted to what schools in what years. I can see who went (or goes) to what college, although that information needs to be updated annually, and “decided to attend” and “declined to attend” are too similar right now, but this is good information to have on hand. I can also track how we're performing over time to find areas for improvement: is there a certain set of schools we need to learn more about, take some trips to?

Here's the same view as seen in R Studio, accessed using the odbc library.

	Student	HSGradYear	School	Ranking	SubmissionStatus	Decision	Attendance
26	Ted Burgh	2019	Harvard University	2	Submitted	Accepted	Declined to attend
37	Melina Farrin	2020	Columbia University in the City of New York	3	Submitted	Accepted	Declined to attend
29	Maryjo Orange	2019	Massachusetts Institute of Technology	4	Submitted	Accepted	Declined to attend
44	Maryjo Orange	2019	University of Pennsylvania	8	Submitted	Accepted	Declined to attend
3	Melina Farrin	2020	California Institute of Technology	9	Submitted	Accepted	Decided to attend
18	Melina Farrin	2020	Northwestern University	9	Submitted	Accepted	Declined to attend
42	Lilyan Boldecke	2020	Duke University	12	Submitted	Accepted	Declined to attend
33	Ted Burgh	2019	Washington University in St Louis	16	Submitted	Accepted	Declined to attend
40	Melina Farrin	2020	Cornell University	18	Submitted	Accepted	Declined to attend
20	Lilyan Boldecke	2020	University of Notre Dame	19	Submitted	Accepted	Decided to attend
7	Ted Burgh	2019	University of California - Los Angeles	20	Submitted	Accepted	Decided to attend
5	Maryjo Orange	2019	University of California, Berkeley	22	Submitted	Accepted	Declined to attend
12	Lilyan Boldecke	2020	Georgetown University	23	Submitted	Accepted	Declined to attend
11	Maryjo Orange	2019	Georgetown University	23	Submitted	Accepted	Decided to attend

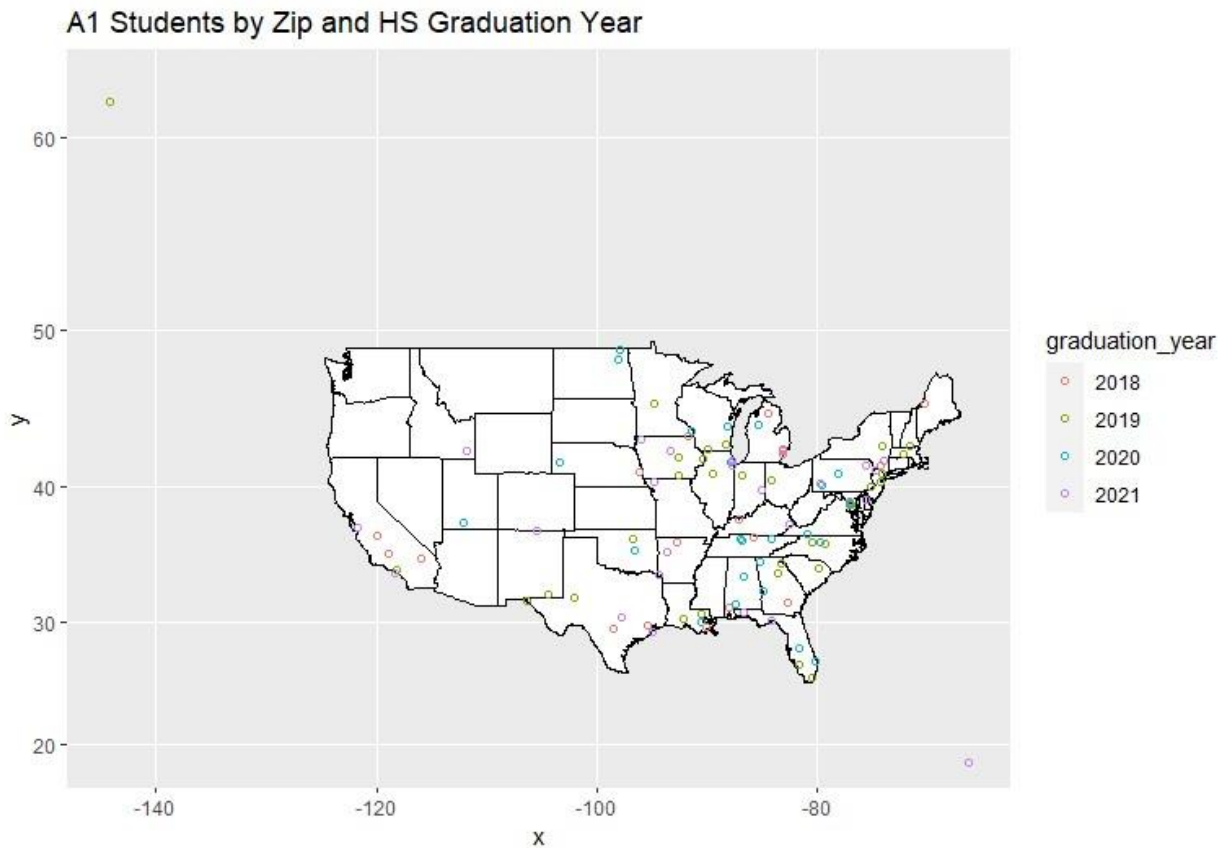
With our actual student data, I'll be able to create some powerful marketing materials that will demonstrate the efficacy of our services, just by tapping into the database.

The following chart is just a jumping-off point, but it does a good enough job summing up the facts: the fake students I randomly generated are pretty competitive.



For the initial analysis, I only inserted 50 records into the college_application table, but I just as easily could have inserted 5,000 and then I'd have some much more interesting visuals to create.

The customer analysis I can do with my database is promising, as well. I did some simple map plotting and found some interesting results:



At first, I almost deleted the aberrant record so this chart would be more visually appealing, but then it hit me: in recent years, we've only had one student in Alaska. We need to change that.

This visual is useful and helps me think about A1's business in a way I haven't been able to before. Granted, these aren't our clients, but if they were, I'd recommend figuring out why we haven't worked with anyone in the Western US for several years, and spend more time trying to land clients from out West instead of places where we are clearly well-known. Hopefully, looking at our data through a similar visualization will help me make some strategic choices.

Ultimately, I see this database growing more and more useful not only as we add more data about our students to it, but as we use it to monitor and respond to industry trends.

CONCLUSION

Gleaning insight from the real database won't take — in fact, if I can track down historical data from past students, even if it's just some of the colleges they were accepted to, I'll immediately be able to use that information to inform our efforts as we begin another admissions cycle.

The hardest part will be convincing the other stakeholders that the database approach will work in the long run, especially because we don't have anything like that already. But I think I will be

able to successfully use Microsoft Access to get other people at A1 personally invested in using the database. Simple forms, like the one on the next page, will make collecting new client information really easy, and I can create robust security within the database to ensure the data stay secure.

Parent Intake

english_forename	<input type="text" value="Marquita"/>
english_surname	<input type="text" value="Wellard"/>
international_forename	<input type="text"/>
international_surname	<input type="text"/>
preferred_forename	<input type="text"/>
email	<input type="text" value="mwellard0@economist.com"/>

Because I've built a fairly well-normalized database, I put combo boxes all over the place on these forms, and I can show the other stakeholders how they can use them to really "make the database their own." For instance, a form like this is perfect for having up during a webinar, but maybe someone else wants a more standardized one for the office when talking to potential clients on the phone.

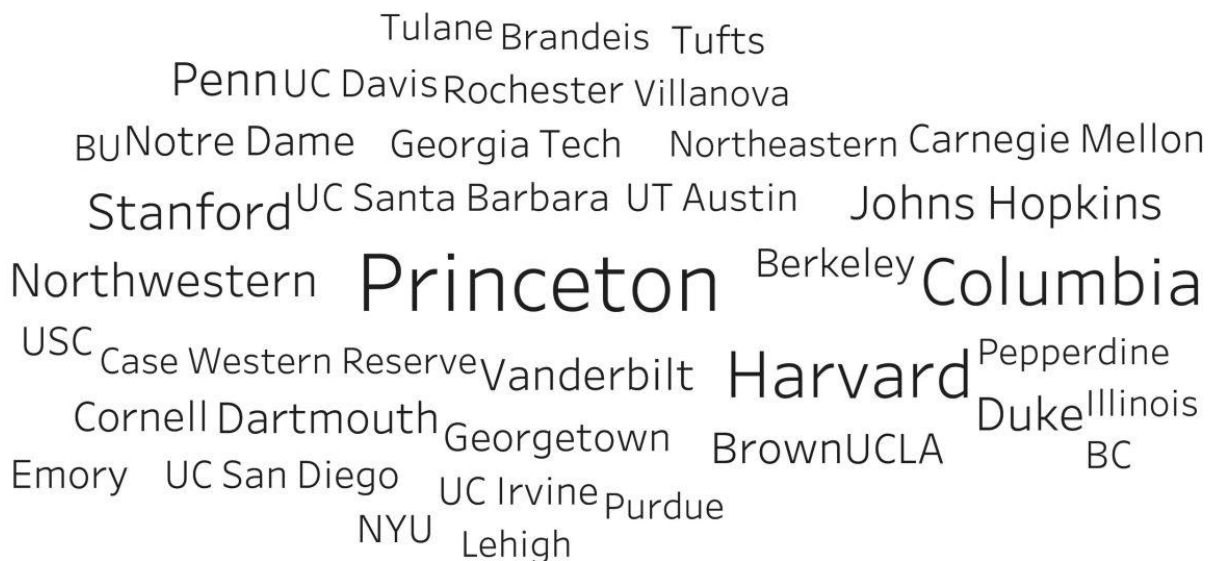
I will have to carefully structure intermediate tables to ensure someone doesn't accidentally wipe our main records, but I'm confident that Access will help the other non-technical users of the database get significant value out of it.

I very much enjoy embarking on this project, and I learned quite a lot. I would now like to credit my sources of data, without which I would not have been able to build this database.

After completing this project, I populated the database with our actual student data and used it to create marketing materials and develop descriptive analytics. Here are some of the visualizations I created for our website, <https://a1educationalconsulting.org>. I designed the entire website, including writing copy, using GoDaddy.

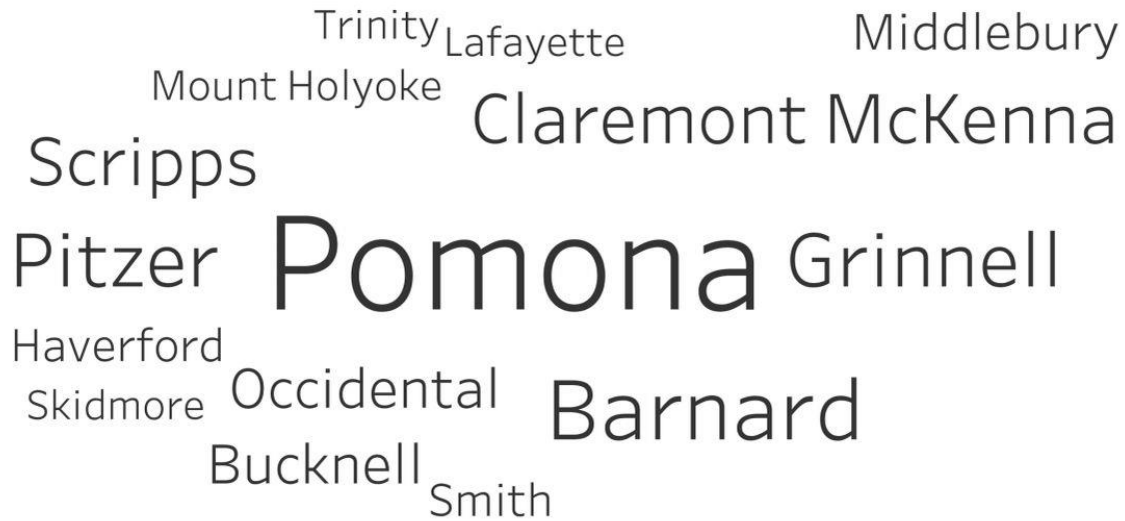
100% Transfer Rate into USC and the UC Schools

A1 students have been accepted to 40 of the top 50 national universities, per the 2022 U.S. News & World Report's Best National Universities.



Results Page: Overview

A1 students have been accepted to 16 of the top 50 national liberal arts colleges, per the 2022 U.S. News & World Report's Best National Liberal Arts Colleges.

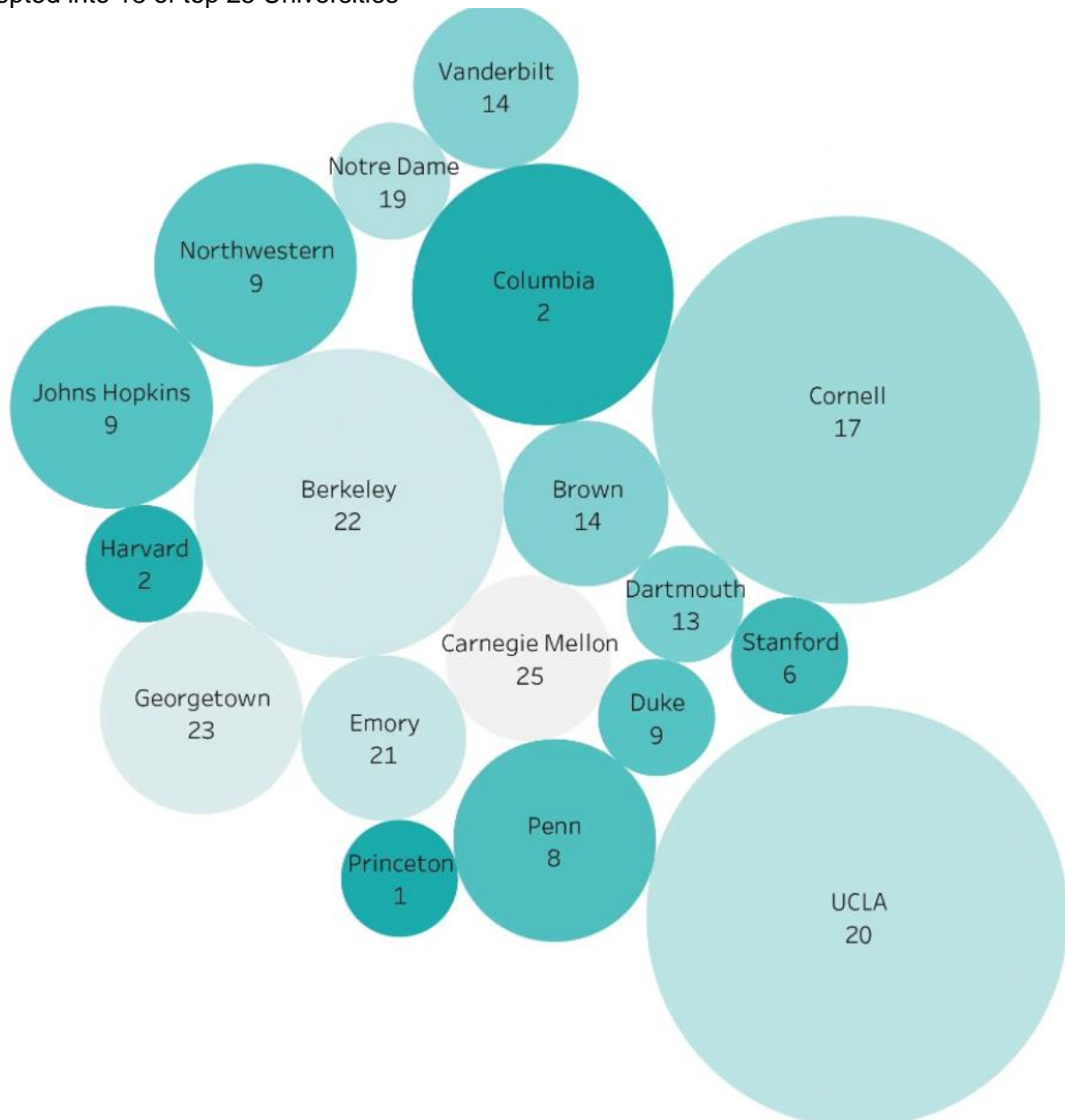


A word cloud of liberal arts colleges. The words are arranged in a roughly circular pattern, with 'Pomona' being the largest and most central. Other colleges include 'Barnard', 'Claremont McKenna', 'Grinnell', 'Middlebury', 'Lafayette', 'Trinity', 'Mount Holyoke', 'Scripps', 'Pitzer', 'Haverford', 'Skidmore', 'Occidental', 'Bucknell', and 'Smith'.

Trinity Lafayette Middlebury
Mount Holyoke Claremont McKenna
Scripps
Pitzer Pomona Grinnell
Haverford
Skidmore Occidental Barnard
Bucknell Smith

Results Page: Overview

Accepted into 18 of top 25 Universities



Acceptances include 7 of the 8 Ivy League schools.

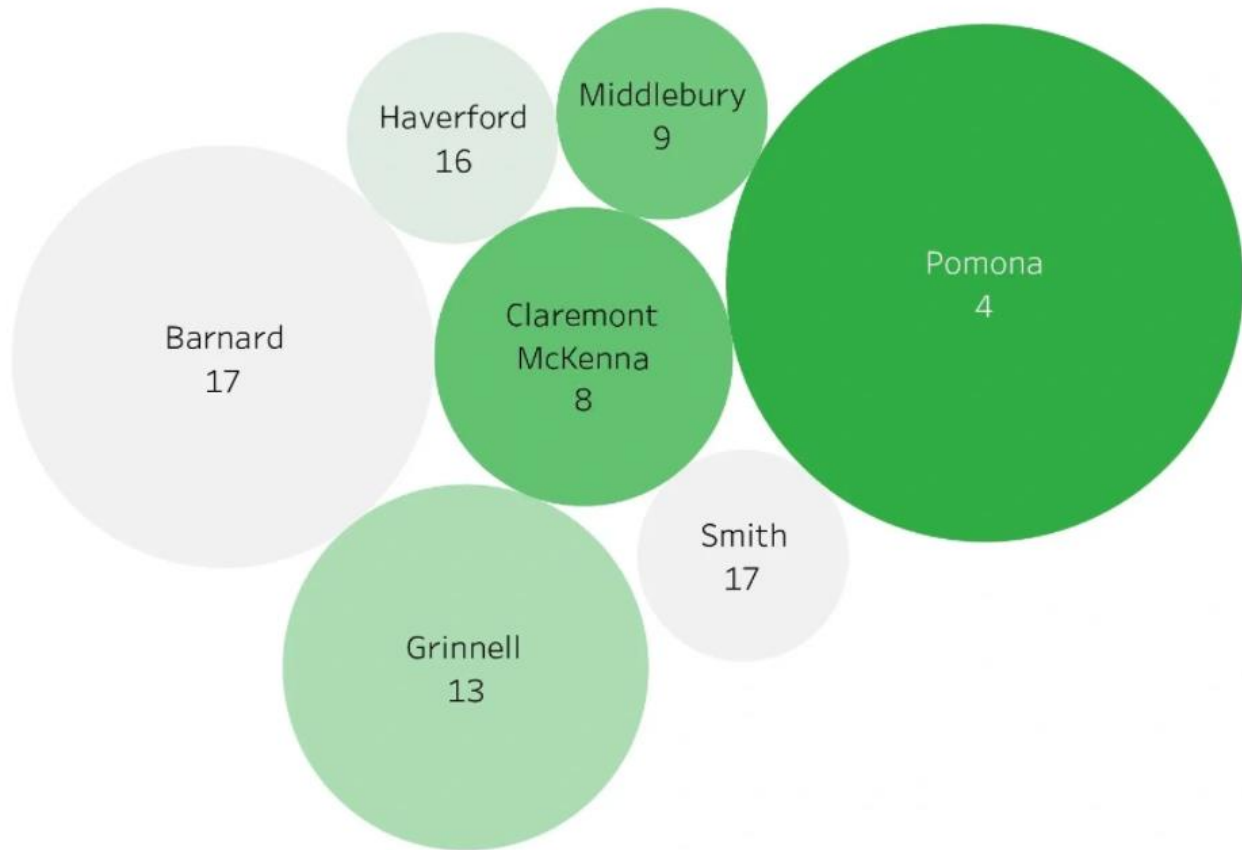
Labels & colors indicate ranking according to the 2022 U.S. News & World Report's Best National Universities.

Bubble size indicates relative number of acceptances.

A1 student data is from 2014 - present.

Results Page: Overview

Accepted into 7 of top 25 Liberal Arts Colleges



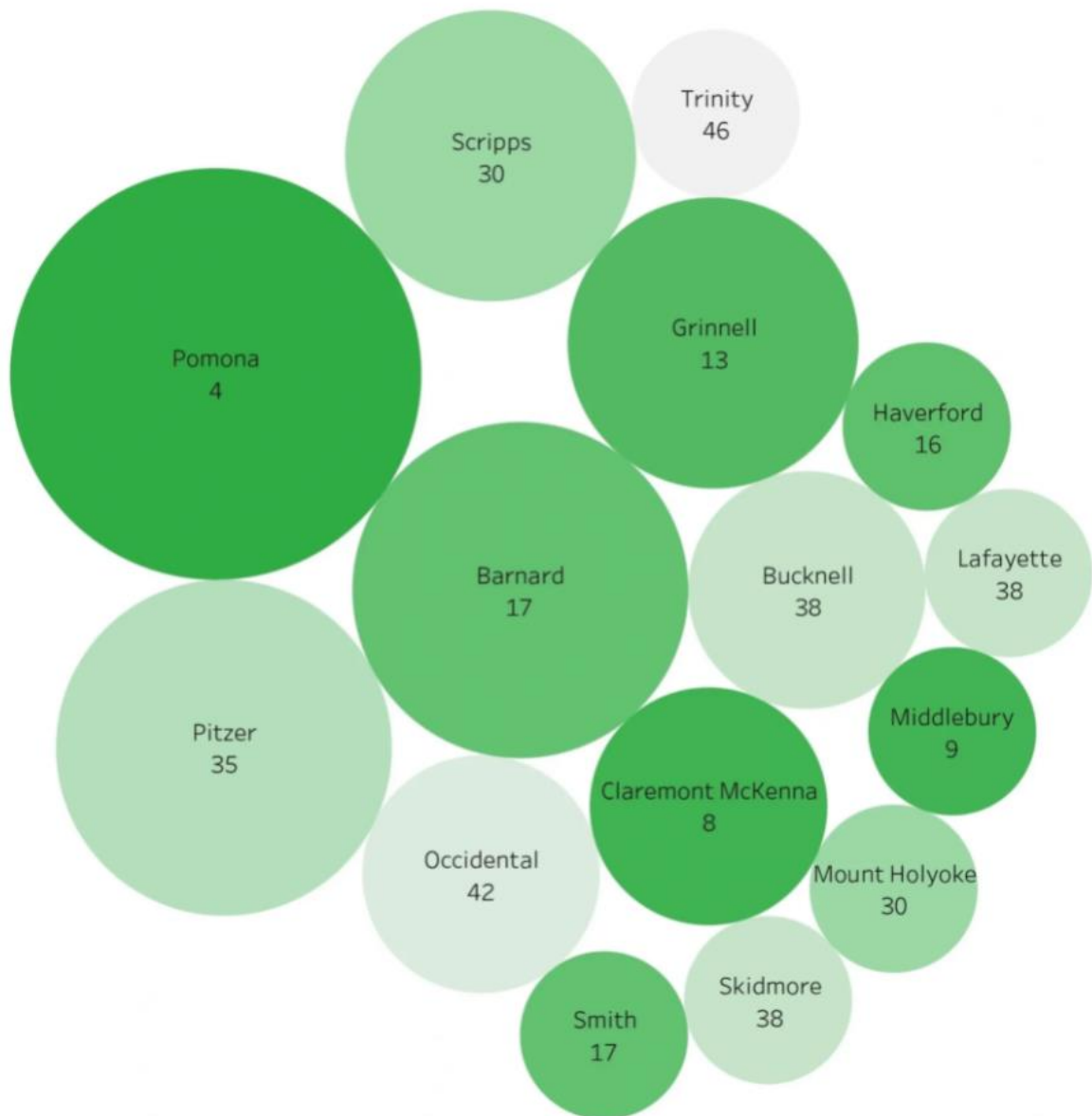
Labels & color indicate ranking according to the 2022 U.S. News & World Report's Best National Liberal Arts Colleges.

Bubble size indicates relative number of acceptances.

A1 student data is from 2014 – present

Results Page: Overview

Accepted into 16 of top 50 Liberal Arts Colleges



Labels & color indicates ranking according to the 2022 U.S. News & World Report's Best National Liberal Arts Colleges.

Bubble size indicates relative number of acceptances.

A1 student data is from 2014 - present.

Results Page: Details

A1 Students Earn Admission to 40 of the top 50 Universities

Princeton	1
Columbia	2
Harvard	2
Stanford	6
Penn	8
Duke	9
Johns Hopkins	9
Northwestern	9
Dartmouth	13
Brown	14
Vanderbilt	14
Cornell	17
Notre Dame	19
UCLA	20
Emory	21
Berkeley	22
Georgetown	23
Carnegie Mellon	25
USC	27
NYU	28
Tufts	28
UC Santa Barbara	28
Rochester	34
UC San Diego	34
BC	36
UC Irvine	36
Georgia Tech	38
UC Davis	38
UT Austin	38
Brandeis	42
BU	42
Case Western Reser..	42
Tulane	42
Illinois	47
Lehigh	49
Northeastern	49
Pepperdine	49
Purdue	49
Villanova	49

Top-50 acceptances include every ranked UC school.

Labels and colors indicate ranking according to the 2022 U.S. News & World Report's Best National Universities.

A1 student data is from 2014 - present.

Results Page: Details

A1 Students Earn Admission to 18 of the top 25 Universities

Princeton	1
Columbia	2
Harvard	2
Stanford	6
Penn	8
Duke	9
Johns Hopkins	9
Northwestern	9
Dartmouth	13
Brown	14
Vanderbilt	14
Cornell	17
Notre Dame	19
UCLA	20
Emory	21
Berkeley	22
Georgetown	23
Carnegie Mellon	25

Top-25 acceptances include 7 of the 8 Ivy League schools.

Labels and colors indicate ranking according to the 2022 U.S. News & World Report's Best National Universities.

A1 student data is from 2014 - present.

Results Page: Details

A1 Students Earn Admission to 16 of the top 50 Liberal Arts Colleges

Pomona	4
Claremont McKenna	8
Middlebury	9
Grinnell	13
Haverford	16
Barnard	17
Smith	17
Mount Holyoke	30
Scripps	30
Pitzer	35
Bucknell	38
Lafayette	38
Skidmore	38
Occidental	42
Trinity	46

Top-50 acceptances include 4 of the 5 undergraduate Claremont Colleges.

Labels and colors indicate ranking according to the 2022 U.S. News & World Report Best National Liberal Arts Colleges.

A1 student data is from 2014 - present.

Results Page: Details

A1 Students Earn Admission to 7 of the top 25 Liberal Arts Colleges

Pomona	4
Claremont McKenna	8
Middlebury	9
Grinnell	13
Haverford	16
Barnard	17
Smith	17

Top-25 acceptances include West Coast, Midwest, and East Coast colleges.

Labels and colors indicate ranking according to the 2022 U.S. News & World Report Best National Liberal Arts Colleges.

A1 student data is from 2014 - present.

DATA SOURCES

Nationalities list: [List of nationalities in YAML, CSV and TXT format](#)

Countries, territories, and regions list: [World countries available in multiple languages, in CSV, JSON, PHP and SQL formats, with associated alpha-2, alpha-3, and numeric codes as defined by the ISO 3166 standard, and with national flags included](#)

US States list: [List-of-US-States/states.csv at master · jasonong/List-of-US-States · GitHub](#)

All US nationally accredited institutions of higher learning: [DAPIP | Homepage](#)