

MGMT 582

Final Project: College Review Website

4/25/2021

Group 7

Craig Mc Iver, Mai Nguyen, Erika Ergart, Stefanie Walsh, Theo Ginting

**Background**

We are consulting for Niche.com, a college review website partnering with over 1700 schools to provide a thorough and comprehensible dataset to prospective students and parents alike. Niche offers many metrics of measurements, from national university rankings to demographic breakdowns as well as the all important information regarding cost of attendance to name just a few. While Niche is a reputable resource in the American market, their website is limited to this specific country. It is probable that Niche.com receives a large amount of traffic from individuals overseas researching US schools, but the full international market is not being captured due to the limitations Niche.com currently has. In a world of ever-increasing globalization, the added value from having a more internationally inclusive database will be significant: Students from all around the world, including the US, will be better able to research all universities and the additional traffic from these new users will generate more advertising revenue.

**Introduction**

While Niche boasts an impressive website with a thorough database and intuitive UX, they are operating in a highly competitive market. Websites such as Studentsreview.com, bigfuture.collegeboard.org, as well as individual school websites all stand to diminish Niche.com’s traffic. Some of these websites boast a more global dataset by providing information on foreign universities in addition to the American domestic ones.

Our firm has been brought in to expand upon their current data set by introducing more distinguishing fields as well as to establish data on universities from abroad in order to grow their user base internationally. Offering a broader array of variables to search on in addition to introducing international universities will make Niche.com a proverbial one-stop-shop for all those researching universities. We set out to do this optimally to ensure high performance and minimize server requirements that would be put under more strain due to the higher volume of website traffic.

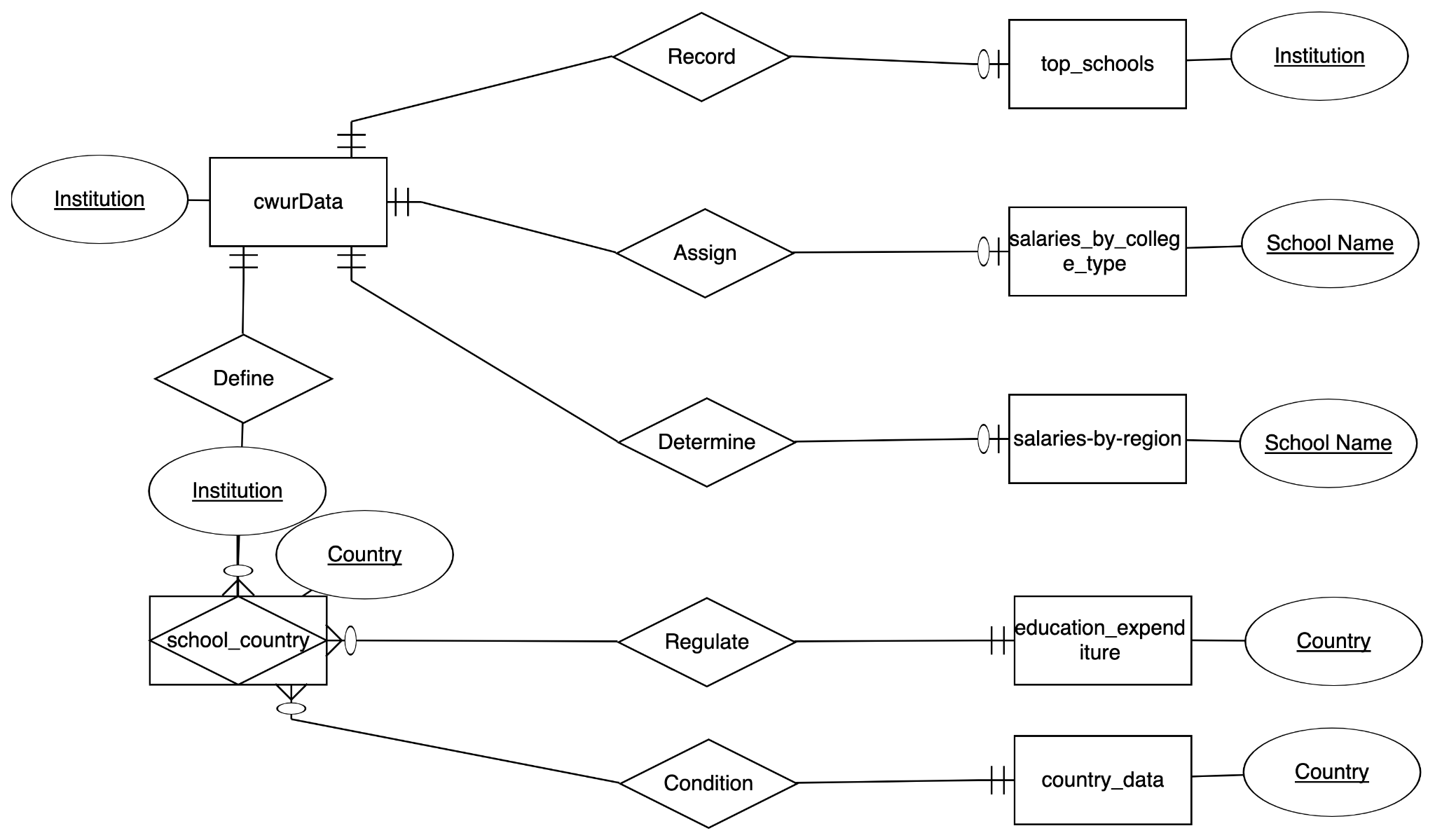
**Datasets**

The majority of our data were obtained from two Kaggle datasets listed in the Appendix that were composed of multiple tables. The first dataset, World University Rankings, allowed us to introduce the international universities that will allow Niche.com to enter a market of countless more people. World University Rankings (**cwurData**) provides many of the metrics Niche has, such as overall rankings, but it also introduces variables such as citations and patents- variables that are of interest to many highly motivated college hopefuls.

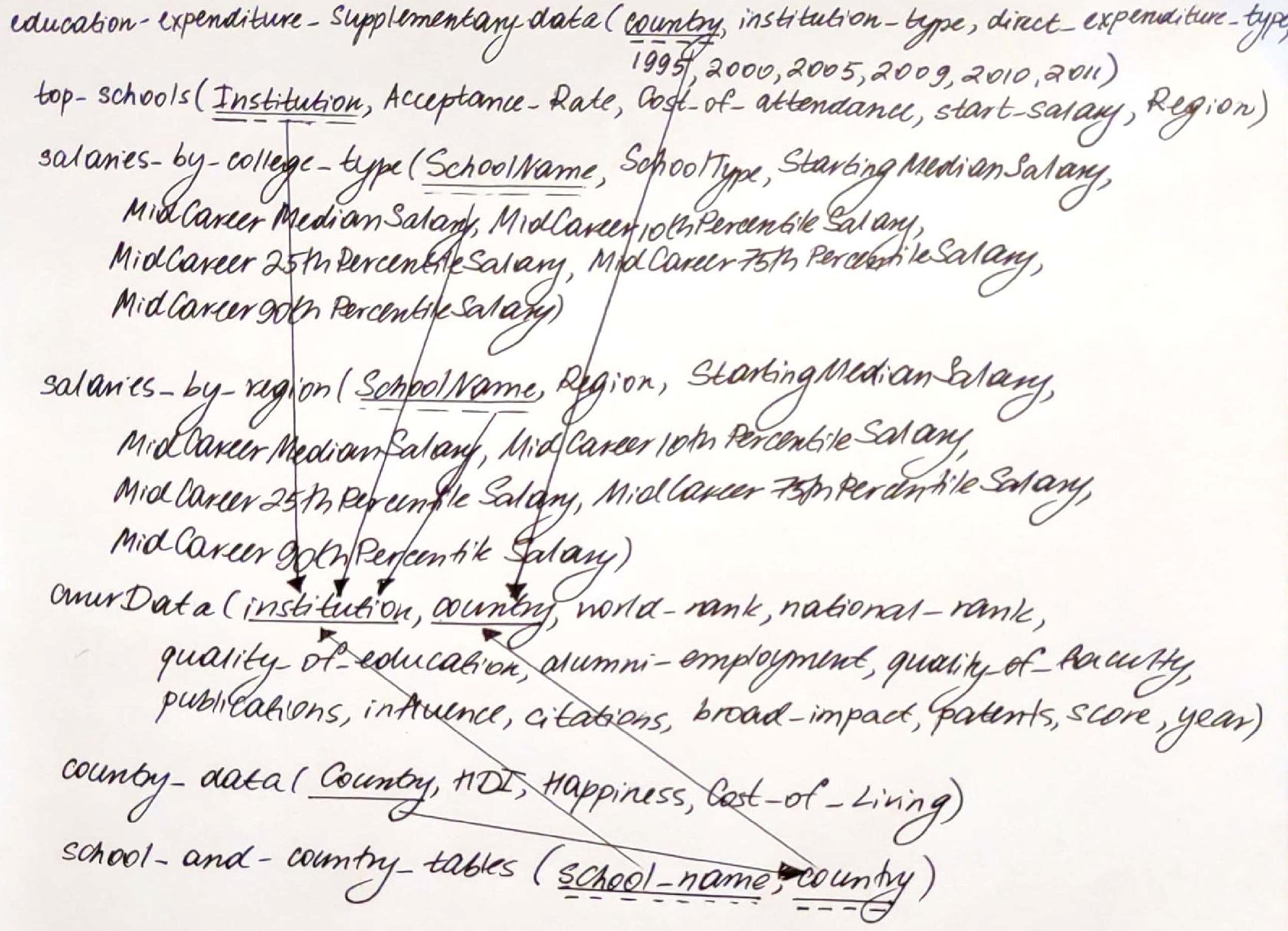
Our second dataset consists of **Salaries\_by\_Region** and **Salaries\_by\_College\_Type**, a broad dataset that breaks down payoffs from different types of schools (state, private, party, etc.), majors, and US regions. Since attending a college is one of the most important, and costly, investments people will make in their life, it is crucial to know what awaits a graduate, regarding salary.

Two smaller datasets were also created to fill gaps in the aforementioned sets: **top\_schools** takes a look at the acceptance rate, cost of attendance for resident students, and median starting salary for graduates for the top 100 ranked universities in the world. While this data was present for the US schools, it did not exist for the international institutions and was procured through various websites. The second dataset was created to answer follow-up questions students would have when researching international schools. **Country\_data** provides the human development index rank (obtained from UN.org), happiness rankings, and cost of living rankings for the nations represented in the top 100 schools. This dataset is essential to provide since the real cost of tuition and salary data is highly dependent on the economics of a nation. Additionally, the HDI and happiness rank provide some information on what life is like for people living in these countries.

**Entity-Relationship Diagram**

****

**Relational Data Model**

****

**Explanation of Choices**

Data was imported utilizing MySQL’s import wizard capability, hence the lack of manually created tables, other DDL, and triggers.

Some basic cleaning occurred to get data in correct forms. An example of this was replacing the 0 values in the start\_salary column of the top\_schools table with NULL since 0s were in the dataset due to no data being found for certain schools.

**SELECT NULLIF(start\_salary, 0) AS start\_salary FROM top\_schools**

**Queries**

**1. Natural Language Description of Query:** Group Universities into brackets based on tuition so website users can click on links like “Affordable”; “Average”; “Expensive” to quickly see universities based on their financial restraints.

**SQL Code and Results:**

SELECT \*

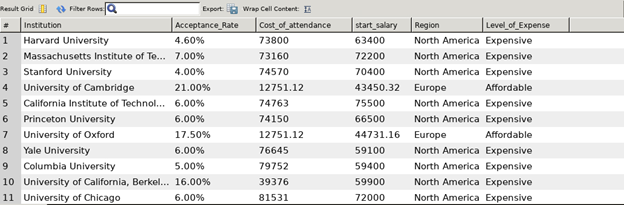
, CASE WHEN Cost\_of\_attendance between 0 and 20000 then 'Affordable'

WHEN Cost\_of\_attendance between 20000 and 35000 then 'Average'

WHEN Cost\_of\_attendance between 50000 and 10000 then 'Expensive'

END AS Level\_of\_Expense

FROM top\_schools



**Description of Business Problem Addressed:** Financial resources are hugely influential when it comes to selecting a university to attend. This query creates brackets of schools based on their cost of attendance ( “Affordable”; “Average”; “Expensive”) in order to make the searching process easier for website visitors as these 3 levels are easier to search on than continuous numerical variables in Cost\_of\_attendance.

**2. Natural Language Description of Query:** Create a filter condition to get a list of EU universities since they often have benefits for EU citizens regardless of countries. This will be attractive to Europeans and will grow our user base in this region.

**SQL Code and Results:**

SELECT \* FROM cwurData

INNER JOIN top\_schools

ON top\_schools.Institution = cwurData.institution

WHERE country in Austria", "Belgium", "Bulgaria", "Croatia", "Cyprus",

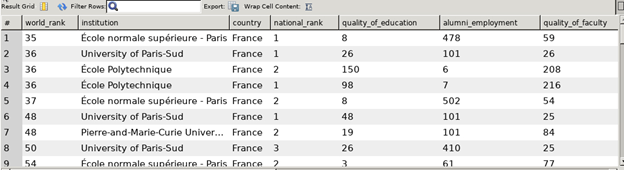
"Czech Republic", "Denmark", "Estonia", "Finland", "France", "Germany",

"Greece", "Hungary", "Iceland", "Ireland", "Italy", "Lithuania",

"Netherlands", "Norway", "Poland", "Portugal", "Romania", "Serbia",

"Slovenia", "Spain", "Sweden")

ORDER BY world rank DESC



Note: Joined Data not Captured on Screenshot. It is to the right.

**Description of Business Problem Addressed:**  Since we are expanding business operations abroad, it is important to consider international unions that have an influence on our variables. In this case, admissions into EU universities is easier and more affordable for EU citizens so a page on the website can be developed with the SQL code above to show students from the EU where they are eligible to attend as “in-state” residents.

**3. Natural Language Description of the Query:** How does happiness index affect college quality of education?

**SQL Code and Results:**

SELECT c.quality\_of\_education,c.institution, c.country, c1.Happiness

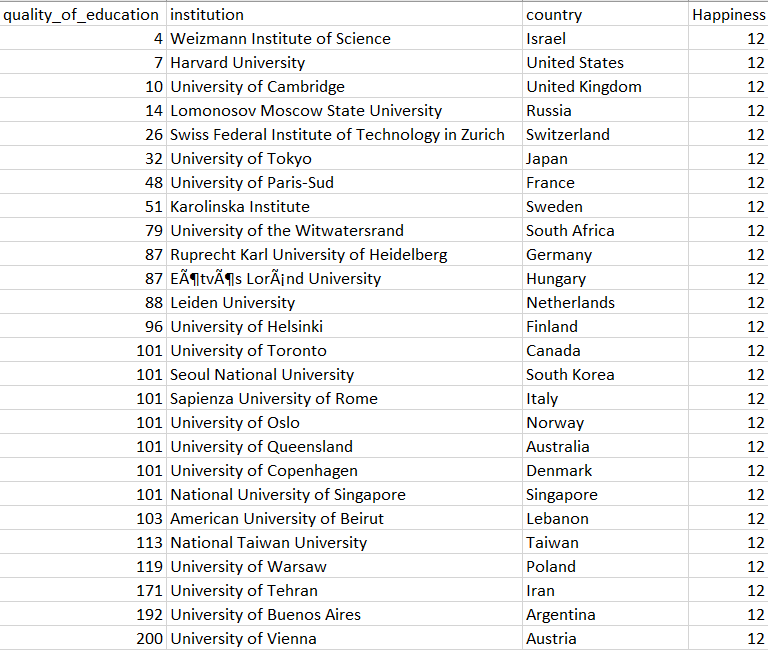
FROM cwurData as c

INNER JOIN country\_data as c1

GROUP BY country

HAVING min(quality\_of\_education)

ORDER BY quality\_of\_education



**Description of Business Problem Addressed:** From the output, it seems that the best quality of education coincides with the maximum happiness index, which is 12. This might be explained by the fact that being depressed actually decreases academic performance according to Dr. Heiligenstein's research. The decreased academic performance is explained by the fact that students miss class, spend less time dedicating themselves to studying, and miss out on interpersonal collaboration to solve problems.

**4. Natural Language Description of the Query:** Which region results in the highest starting median salary?

**SQL Code and Results:**

SELECT s1.SchoolName, s1.Region,

max(s1.StartingMedianSalary) as MaxMedianSalaryBySchoolType,

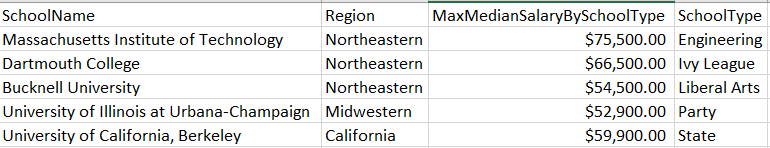
s2.SchoolType

FROM salaries\_by\_region as s1

INNER JOIN salaries\_by\_college\_type as s2

ON s1.SchoolName = s2.SchoolName

GROUP BY SchoolType



**Description of Business Problem Addressed:** From the output, it's possible to observe that the schools with the highest median salary are located in the Northeastern region. This could be explained by the fact that the first 13 colonies were formed in the Northeastern region, leading to the emergence of the first schools. The longer the schools existed, the more time there was for the teaching staff to improve on their teaching, thus leading to earning a more valuable brand name and resulting in higher median salaries for their alumni.

**5. Natural Language Description of the Query:** Is having a higher rate of international students and smaller student to staff ratio related to higher Mid-Career 90th Percentile Salary?

**SQL Code and Results:**

SELECT s1.SchoolName, s1.SchoolType,

s1.MidCareer90thPercentileSalary, t1.student\_staff\_ratio,

T1.international\_students

FROM salaries\_by\_college\_type as s1

INNER JOIN timesData as t1

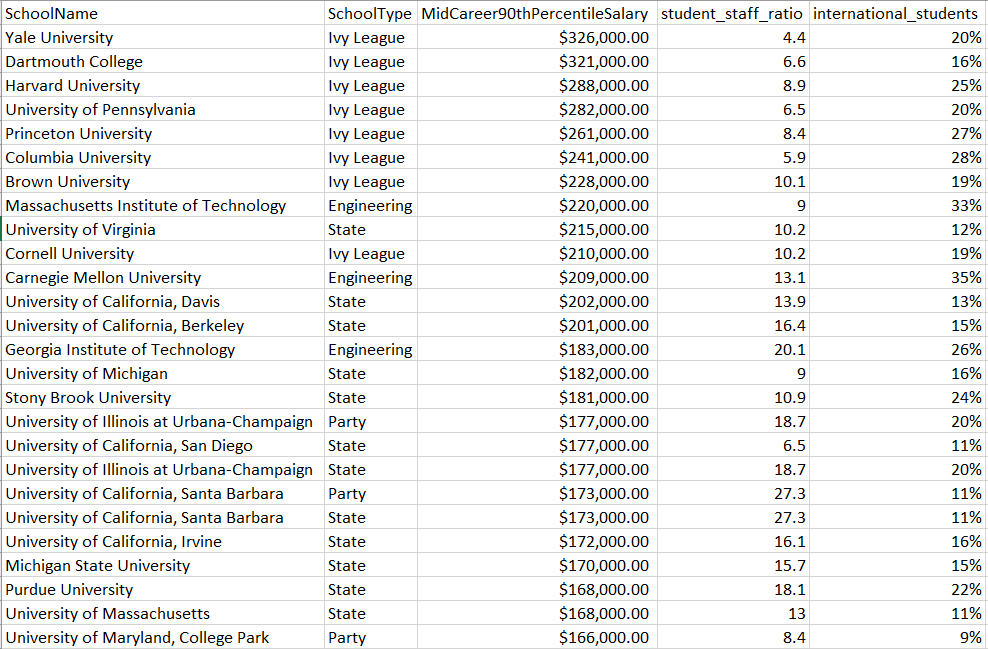
on s1.SchoolName = t1.university\_name

WHERE s1.MidCareer90thPercentileSalary != 'N/A'

and s1.SchoolName != 'University of Wisconsin'

GROUP BY s1.SchoolName, s1.SchoolType

ORDER BY s1.MidCareer90thPercentileSalary desc



**Description of Business Problem Addressed:** After exploring this topic, it was discovered that having a lower student to faculty ratio and higher percentage of international students leads to higher midcareer 90th percentile salary. The lower student to faculty ratio helps professors to develop more personal relationships with students, which not only results into professors answering all of the students' questions, but also in professors assisting their students place into higher paying jobs. The higher international student percentage can be explained by the fact that in the diverse environment there are more opportunities to learn how to approach a problem solution from multiple perspectives, which enriches the student experience and results into improved learning tactic, thus placing students in a better position.

**6. Natural Language Description of Query:** Identifying institutions that have return on investment between starting salary and cost of attendance that are higher than the average return on investment among top schools across different regions.

**SQL Code and Results:**

SELECT \*, (start\_salary-Cost\_of\_attendance)/Cost\_of\_attendance\*100 AS ROI

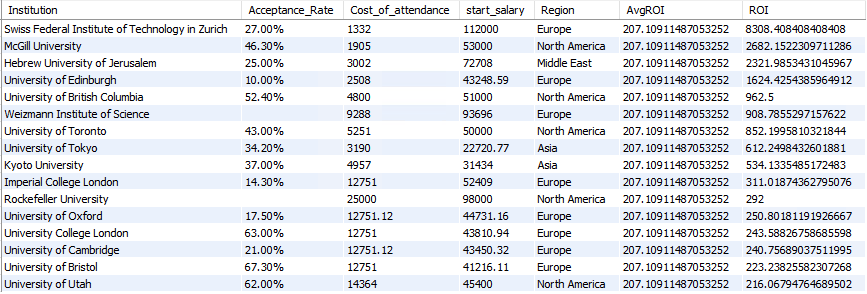
FROM top\_schools,

(SELECT AVG((start\_salary-Cost\_of\_attendance)/Cost\_of\_attendance\*100) AS AvgROI

FROM top\_schools) AS Average

WHERE (start\_salary-Cost\_of\_attendance)/Cost\_of\_attendance\*100 >= Average.AvgROI

ORDER BY ROI DESC;



**Description of Business Problem Addressed:** The ratio between starting salary and cost of attendance is an important factor when it comes to choosing a prospective institution. As it is shown here, interestingly, European universities tend to have the highest return on investment due to many subsidies for college tuition followed by North American universities. It should also be taken into consideration that acceptance rates for top ROI schools are among average universities but not the Ivy League. Thus, it can be a great deciding factor for students who are targeting good ROI without the fierce competition to get accepted.

**7. Natural Language Description of Query:** We chose the variables institution, average score over the years available, and average rank of quality of education over the years available. Then, we grouped by institution to get averages for each institution and ordered the results by the highest average college score.

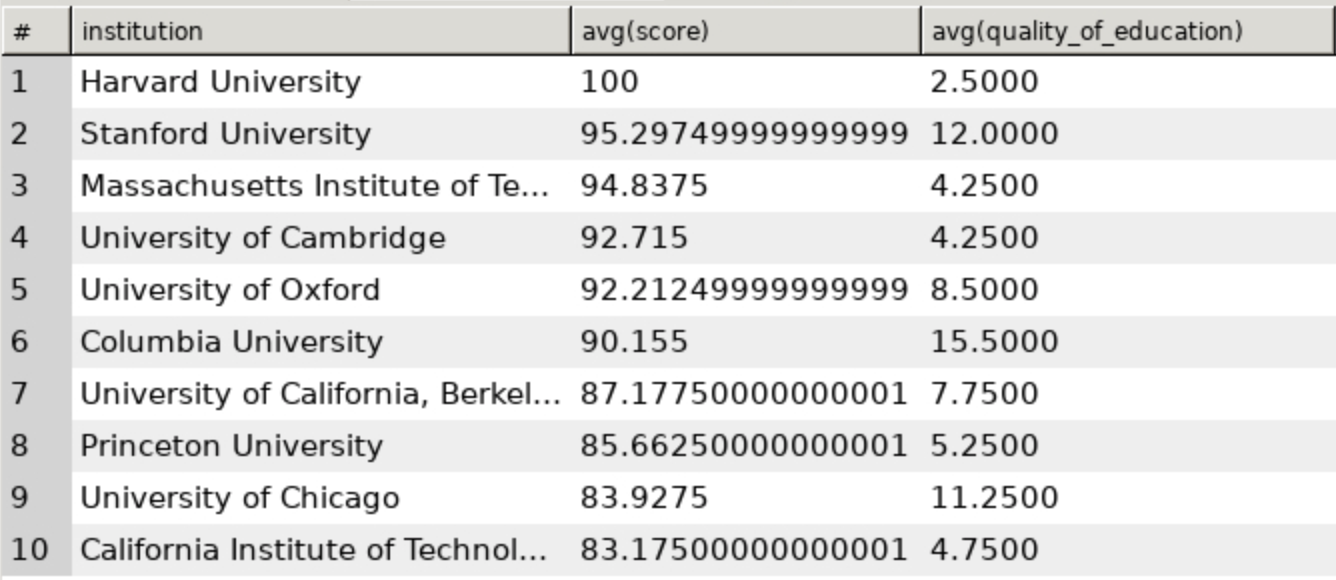
**SQL Code and Results:**

SELECT institution, avg(score), avg(quality\_of\_education)

FROM `582\_clean\_schema`.cwurData

GROUP BY institution

ORDER BY avg(score) desc

limit 10;

**Description of Business Problem Addressed:** The task was to look at the impact of quality of education on the final college score to see if there was any apparent relationship. This could help in determining which scoring component is the most critical to the final score. From the table, looking solely at the top 10 in average total score over each year available, we can see that there doesn’t seem to be a strong relationship between quality of education and average score for schools that are strong overall. The number 2 school Stanford even is ranked 12th on average for quality of education and Columbia at number 6 is ranked 12th on average. Thus, we suggest the client look at other factors to see if there is a stronger relationship elsewhere.

**8. Natural Language Description of Query:** We selected the country, the institution type (public, private, total), and the expenditure in 2011. Then, we grouped by country and type to get the education expenditure value for each country and school type combination. After that, we ordered it by expenditure amount to see which countries spend the most money on education.

**SQL Code and Results:**

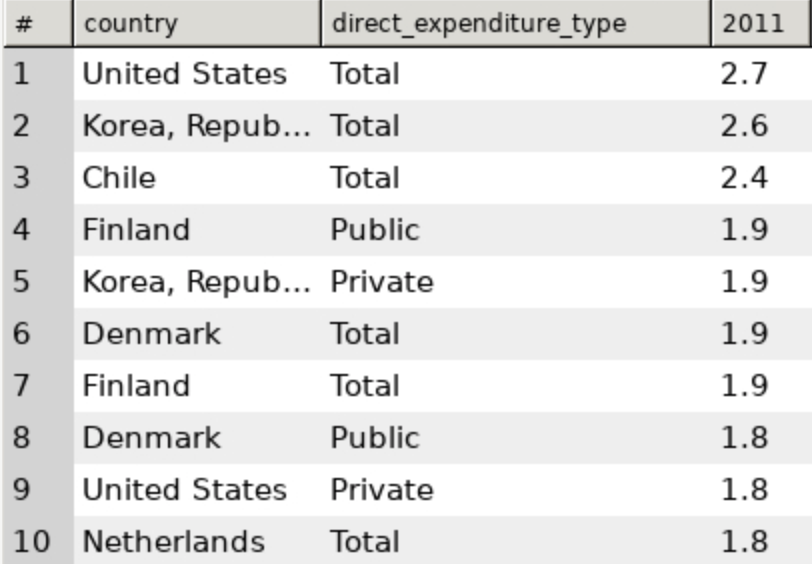
SELECT country, direct\_expenditure\_type, `2011`

FROM education\_expenditure\_supplementary\_data

WHERE `2011` < 2011 AND `2011` IS NOT NULL #this was added because the first row was column titles

GROUP BY country, direct\_expenditure\_type

ORDER BY `2011` desc ;



**Description of Business Problem Addressed:** The task was to examine education expenditure by type of institution and country to see if there are any noticeable trends. While the US does come in first as expected in total education expenditure, it is interesting to note countries like Korea, Finland, and Denmark have high expenditures given the size of their country. With this information, our client could decide to focus some exploration into schools in those countries, assuming that higher spending on education leads to better education outcomes. Starting with these countries could help our client expand its college reviews to schools outside the US. The analysis would be further benefitted by including population size as a variable to get expenditure per capita.

**9. Natural Language Description of Query:** We first joined the rankings table and the school and country table together to get a country for each of the schools. Then we found the average score by country and ordered them from highest to lowest scores.

**SQL Code and Results:**

SELECT s1.country, avg(c1.score)

FROM school\_and\_country\_table as s1

JOIN cwurData as c1

ON s1.school\_name = c1.institution

GROUP BY country

ORDER BY avg(c1.score) desc

LIMIT 10;



**Description of Business Problem Addressed:** The task was to understand the average score for universities in each country to see which countries had the best schools on average. From the results, we can see that the United States leads, but we see Singapore as a country with stronger academics. Since Singapore is less common for US high school students to study at, whether for a degree or just for a study abroad, it might make sense for our client to promote Singapore as an up-and-coming education center.

**10. Natural Language Description of Query:** We had to join three tables together: one table with the schools and scores, one table with the school name and countries, and one table with the countries and the expenditures. After those three tables were joined, we selected for the country, average score, and 2011 expenditure where expenditure type is the total spent and where we group by the country and order the results by the average scores.

**SQL Code and Results:**

SELECT t1.country, avg(c1.score), t1.2011

FROM education\_expenditure\_supplementary\_data as t1

JOIN school\_and\_country\_table as s1

ON t1.country = s1.country

JOIN cwurData as c1

ON s1.school\_name = c1.institution

WHERE t1.`2011` < 2011 AND t1.direct\_expenditure\_type = 'Total'

GROUP BY country

ORDER BY avg(c1.score) desc

LIMIT 10

****

**Description of Business Problem Addressed:** The task was to understand if there was a relationship between total education expenditure for a country in 2011 and the average scores of schools in that country. In this case, we do find that there might be a small correlation between expenditure and score, but it seems like the average is around 1.3 to 1.9. What is interesting to note is that Chile spends the second most money on education but is ranked 24 out of 26. Thus, it is important for our client to not always assume that expenditure is equal to outcome.

**11. Natural Language Description of Query:** In order for Masters or PhD students to choose universities with good research resources, we list universities with number of publications and patents gained per year and see how it relates to quality of education and quality of faculty in the same year.

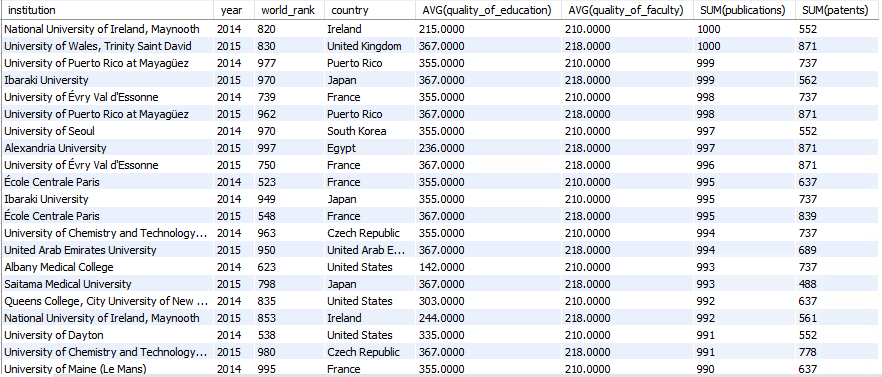
**SQL Code and Results:**

SELECT institution, year, world\_rank, AVG(quality\_of\_education), AVG(quality\_of\_faculty), SUM(publications), SUM(patents)

FROM cwurData

GROUP BY institution, year

ORDER BY SUM(publications) DESC;

****

**Description of Business Problem Addressed:** This table shows that the ranking of the university does not necessarily result in high quality of education and more publications getting published. Instead, the quality of education and quality faculty directly affects the number of publications and patents that schools achieve. This is a good reference for research scholars when it comes to choosing a research university that fits their needs.

**Appendix**

Sources:

Primary Data sets:

World University Rankings

<https://www.kaggle.com/mylesoneill/world-university-rankings?select=cwurData.csv>

Salaries by College Type

<https://www.kaggle.com/wsj/college-salaries?select=salaries-by-college-type.csv>

Secondary Sources:

<https://www.kornferry.com/about-us/press/Recent-College-Grads-Can-Expect-to-Make-the-Highest-Salaries-in-Germany-and-U.S>.

HDR from the UN for HDI

Cost of living: https://www.expatistan.com/cost-of-living/country/ranking