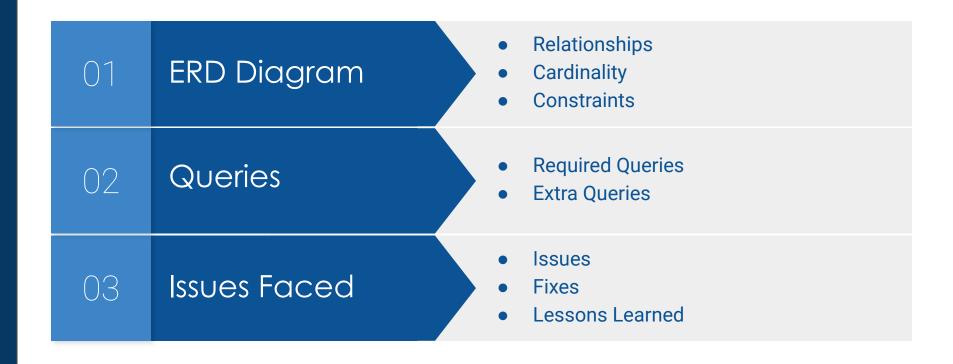
Master of Business Analytics Database and Query Creation

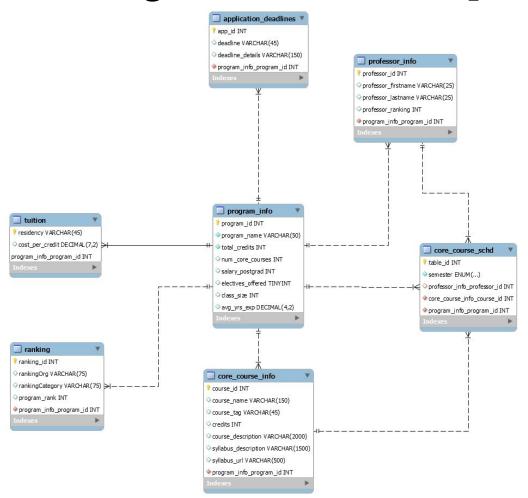
OTM422/IS722 - Group 3
Breydn Duckett
Radhika Kulkarni
Rachel Raasch,
Reid Robideaux
Darnell Sims
Corrine Vanevenhoven



Presentation Path

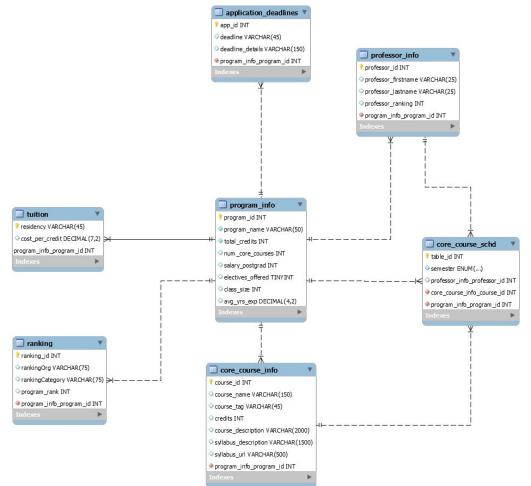


ERD Diagram - Relationships and Cardinality



- 'Program_info' Overall program information such as program name, total number of credits etc.
- A single program: Multiple core courses, application deadlines etc.
- A single course: Multiple semesters and multiple professors teaching it.

ERD Diagram - Relationships and Cardinality

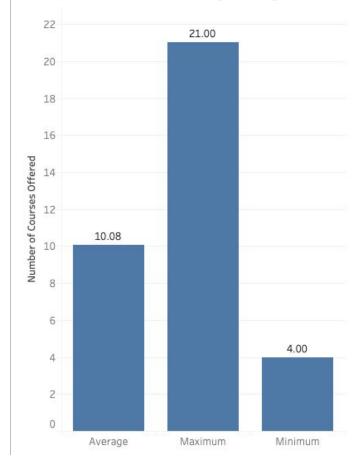


- Program, course and professor names and other details/descriptions: VARCHAR
- Table IDs, credits and other numerical attributes: INT 'Semester': ENUM
- Electives offered: TINYINT
- Application deadlines: Set to VARCHAR initially and then cast to DATE

Query 1: What is the minimum, maximum and average number of courses offered by a program?

SELECT MIN(num_courses_offered) as min_courses_offered, MAX(num_courses_offered) as max_courses_offered, AVG(num_courses_offered) as avg_courses_offered FROM (select program_info_program_id as program_id,count(course_tag) as num_courses_offered from core_course_info group by program_id) as numCoursesTable;

Average, Maximum, and Minimum Number of Core Courses Offered by a Program



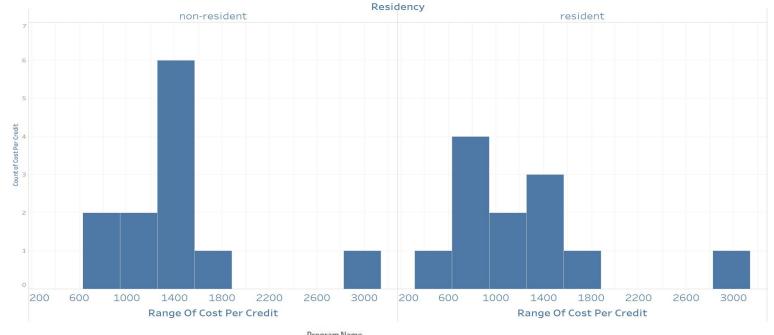
Query 2: Which are the schools that offer courses that cover "SQL"? What are the course names for these courses?

| College | Course |
|------------------------------------|--|
| Bentley University | Data Management and SQL Analytics |
| California Polytechnic Institute | Computing and Machine Learning for Economics |
| California Polytechnic Institute | Data Management for Business Analytics |
| Saint Mary's College of California | Data Management Technology |
| William & Mary | Big Data |

select program_info.program_name, core_course_info.course_name from program_info inner join core_course_info on program_info.program_id = core_course_info.program_id where core_course_info.course_description like '%sql%';

Query 3: What is the distribution of the maximum program cost

per credit? What is the distribution of the maximum program cost per credit?



| | Program Name | | | | | | | | | | | |
|--------------|--|-------------|------------|-------------------|------------|-------------------|------------------|---------------|---------------|----------------|-------------|----------------|
| | AND THE RESERVE OF THE PARTY OF | Benedictine | Bentley | California | Rutgers | Saint Mary's | University of | University of | University of | University of | Wake Forest | |
| Residency | Babson College | College | University | Polytechnic Insti | University | College of Califo | Illinois-Chicago | Maryland | Massachusetts | Michigan-Dearb | University | William & Mary |
| non-resident | 1,853 | 1,000 | 1,540 | 695 | 915 | 1,002 | 1,493 | 1,548 | 1,528 | 2,835 | 1,408 | 1,380 |
| resident | 1,853 | 1,000 | 1,540 | 695 | 383 | 1,002 | 728 | 717 | 739 | 2,835 | 1,408 | 1,380 |

select program_name as 'Name Of College', residency as "Residency Status",cost_per_credit as "Cost of Credit" from program info join tuition on program info.program id = tuition.program info program id order by residency;

Query 4: What is the number of 3 credit courses that each school offers in the Spring

| Program ID | Spring 3 credit core courses offered |
|------------|--------------------------------------|
| 1 | 3 |
| 2 | 7 |
| 4 | 4 |
| 5 | 4 |
| 6 | 2 |
| 8 | 3 |
| 11 | 11 |

Missing Programs:

- Program 10 offers 2 credit courses
- Program 3, 7, 10, and 12 did not have available data
- Program 9 offers 4 credit courses

select s.program_info_program_id as programID, count(distinct i.course_id) as numCourses from core_course_schd as s inner join core_course_info as i on s.core_course_info_course_id = i.course_id where s.semester = 'Spring' and i.credits = 3 group by s.program_info_program_id;

Query 5: Which is the university with the lowest rating in your data with also offers electives?

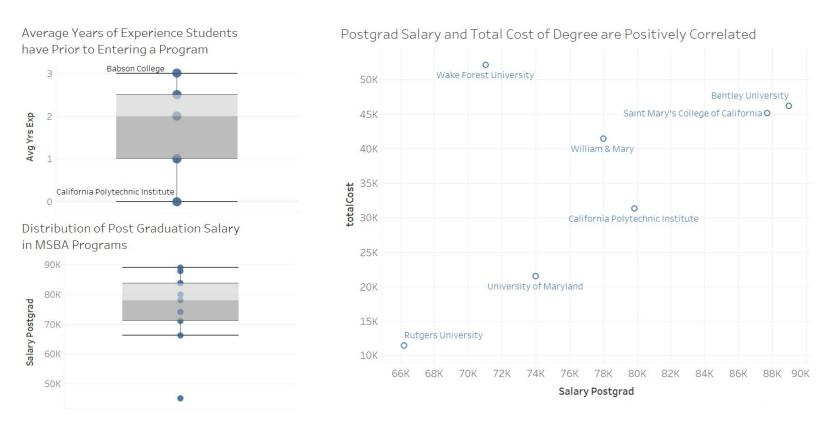
| Program Name | Rating Organization | Ranking |
|----------------|--------------------------|---------|
| Babson College | Us News and World Report | 44 |
| Babson College | Niche.com | 22 |

University of Massachusetts - Boston, University of Illinois - Chicago and Benedictine College were unranked in both ranking organizations.

(select program_name as "Lowest Ranked School", rankingOrg as Organization,min(ranking.program_rank) as Ranking from program_info join ranking on program_info.program_id = ranking.program_info_program_id where electives_offered = 1 and rankingorg = 'US News and World Report')
union

(select program_name as "Lowest Ranked School", rankingOrg as Organization,min(ranking.program_rank) as Ranking from program_info join ranking on program_info.program_id = program_info_program_id where electives_offered = 1 and rankingorg = 'niche'):

Extra Query: What are the distributions of various program stats?



select program_name, class_size **from** program_info **where** class_size **is not null order by** class_size **desc;**** this code is repeated to find the distribution of avg. years experience and post-graduation salary

Extra Query: What are the programs first and last application deadlines?

| Program: | First Deadline | Last Deadline |
|------------------------------------|----------------|---------------|
| Babson College | 2020-01-11 | 2020-10-26 |
| California Polytechnic Institute | 2020-02-01 | 2021-06-01 |
| William & Mary | 2020-03-01 | 2020-12-01 |
| University of Michigan-Dearborn | 2020-04-01 | 2020-12-01 |
| University of Massachusetts-Boston | 2020-08-01 | 2021-05-01 |
| Wake Forest University | 2020-08-01 | 2021-06-15 |
| Bentley University | 2020-10-01 | 2021-07-01 |
| University of Maryland | 2020-10-15 | 2021-01-15 |
| Rutgers University | 2020-11-15 | 2021-01-20 |
| Benedictine College | 2020-12-01 | 2021-01-29 |
| Saint Mary's College of California | 2020-12-01 | 2021-05-06 |
| University of Illinois-Chicago | 2020-12-15 | 2021-07-01 |

select program_info.program_name, **min**(STR_TO_DATE(a.deadline,'%m/%d/%Y')) **as** FirstDeadline, **max** (STR_TO_DATE (deadline,'%m/%d/%Y')) **as** LastDeadline **from** application_deadlines **as** a **left join** program_info on program_info_program_id **eroup by** program_info_program_id **order by** FirstDeadline, LastDeadline;

Issues Faced

| Issue Faced | Fixes | Lessons Learned |
|---|---|---|
| Websites for some colleges in our data set didn't provide a lot of information on classes besides the class name | Google the class names Search for the most recent syllabi to get the information we needed | Data collection isn't black and white, there is a gray area. Sometimes information from a 2017 syllabus is the best you can do |
| Primary key and null value questions in our ERD | Explain to Yash our reasoning and rationale Listen to his opinion | In group projects and in the working world, it is okay to ask questions Better to ask a question and improve the outcome |
| Populating 'core_course_schd' table in the database: Could not add data and ran into a 'cannot update/add a foreign key' error | Recreated tables keeping in mind the order in which foreign key attributes are created Created separate PK column for 'core_course_schd' table | It is important to create tables and attributes in the correct order so that data can be seamlessly entered |

Thank You!

Questions?