2023 U.S. Information Technology Collegiate Conference Business Analytics Competition

Sponsored by College Raptor

Do not put your name(s) or your school's name on anything that you submit.
Doing so will result in disqualification of your team.
The only identifying information you should use is your team number.

Welcome to the USITCC Business Analytics Competition. In this competition, you will be required to analyze educational data. This competition can be solved using various technology platforms such as R, Python, Excel, or Microsoft SQL. We will provide access to both the raw data (if you want to use R or Excel or Python) and to an on-site Microsoft SQL Server. Contestants may use any software, including access to the Internet, to develop and implement their model.

[R, Weka Users Only] R can be download at https://www.r-project.org/ and Weka can be downloaded at https://www.cs.waikato.ac.nz/ml/weka/downloading.html). You may also find the Rattle extension to R at https://rattle.togaware.com/rattle-install-mswindows.html.

[SQL Users Only] SQL Management Studio (SSMS) is available for free at https://docs.microsoft.com/en-us/sql/ssms/download-sql-server-management-studio-ssms. HeidiSQL is available for free at https://www.heidisql.com/download.php.

A CSV file containing the data is available on the contest server.

Notes:

1) The data files for this year's contest are small enough that contestants may find it sufficient to solve the problem on their laptops directly and may not need to use SQL
2) The data available in the CSV and XLSX files are the same as what is on the SQL server.

Data on the SQL server may be access at:

Hostname: alamo.usitcc.lab (IP address: 10.103.1.11)
SQL Username: BA
Password: Raptor2023!
Database Name: USITCC_Raptor

General Guidelines

You will be required to provide a summary of your results, as well as a complete write-up. The components of your write up should include the following.

1. Decision support analysis for each problem **with interpretation**
2. Management insights associated with each analysis
3. Provide R or Python code, Excel spreadsheets, or SQL queries used in each step

The final product should directly answer all questions and provide management insights. All elements of each question should be addressed for full credit. Of specific importance is your interpretation of results and findings.

Decision support products not included in the discussion may be placed in appendices.

Not only are we looking for correct “answers” but we are looking for professional and well-substantiated explanations. Consider that you are creating this report for the college’s leadership team and the report should be able to “stand alone” with enough details and comments which we will use for grading.

COMPETITION DESCRIPTION

You are a data analyst for College Raptor (“Raptor”). A college has experienced declining enrollment and revenues during the past few years and is debating if it should change its approach to awarding need- and merit-based grants and scholarships for admitted students. The college hired Raptor to do a “financial aid leveraging” (FAL) project to help the college understand the relationship between its “net price” and enrollment decisions for students of varying academic strengths and financial situations.

This morning, the college sent Raptor the college’s enrollment data for all admitted students from the past year. The college’s vice president of enrollment requested that Raptor complete its analysis and recommendations within four hours so she can present the findings tomorrow to the college’s president and board of trustees. You are tasked with completing the FAL project. The college’s goals are to increase its net tuition revenue by at least $1,000,000 per year, increase the number of enrolled students by at least 100 per year, and maintain the college’s academic profile of incoming students to within 0.5 ACT points from the current level. You must work quickly so Raptor can help the college achieve these goals. Good luck!

ADVICE ON COMPLETING THIS COMPETITION: Read through the background and concepts below and then start answering the competition questions. Refer to the dictionary and descriptions below as needed. The competition questions are designed to help you gain familiarity with the concepts as you progress toward more challenging problems.

Definitions and background on concepts

You are new to the Raptor team and are not (yet) an expert in FAL. Fortunately, Raptor has provided key things you will need to know to complete the FAL project.

Before we focus on FAL itself, let's define/explain some of the terms required to do FAL. For the purposes herein, we are considering only the cost/revenues for a single year. (We are not multiplying these values by four for a four-year college.)

Total Cost of Attendance (also known as Sticker Price) – The sum of annual tuition, fees, room, board, books/supplies, transportation, and personal expenses for the student for a year of college.

Net Price – This is the Total Cost of Attendance minus Grants and Scholarships. Net Price is a measure of the true cost of college from the family's perspective.

Institutional Gift Aid – This is the total amount of grants and scholarships awarded to the student by the college (institution). This will include financial aid such as Merit scholarships, Need-Based Aid, and Athletic scholarships. This does *not* include financial aid from other sources such as federal and state grants or private scholarships.

Net Tuition Revenue – This is the amount of money the college receives for educating the student. Net tuition revenue is the same as Tuition minus any Institutional Gift Aid (Merit, Need-Based, and Athletic). Federal and State grants or outside scholarships will still be counted as revenue and should therefore not be subtracted from the Net Tuition Revenue. (Think of it this way: if the college provides a scholarship or grant to the student that decreases the student's net price and the college's net tuition revenue. But if the government provides a grant to a student that decreases the student's net price but does not decrease the college's net revenue because the grant money is paid to the college. The same applies for an outside scholarship that is paid to the college on the student's behalf; it decreases the student's net price but not the college's net revenue.)

Expected Family Contribution (EFC) – When a student completes the Free Application for Federal Student Aid (FAFSA), the result is the U.S. government's estimate of how much money the family should be able to afford to contribute in the upcoming year for that student to attend college. That amount is known as the Expected Family Contribution or EFC. EFC varies based on factors like family income, assets, number of members in the family, state of residency, and number of students attending college simultaneously. EFC is a reasonable metric for a family's financial situation or wealth.

Financial Need – This is the greater of $0 and the full Cost of Attendance minus the Expected Family Contribution. Financial Need is a measure of how much grant and scholarship money the family would require to be able to pay for college without taking on debt. Financial Need should not be negative.

% of Need Met – In cases where Financial Need is greater than $0, this is the sum of all grants and scholarships (from all sources, including the college, outside, or government) divided by the Financial Need.

College Enrollment Data Dictionary

- StudentID – A unique identifier for each student
- ACT – ACT Test score of the prospective student (Note, if the student took the SAT instead of the ACT, we have already converted the SAT score into an ACT score, so no SAT information is provided)
- GPA – Reported GPA of the prospective student on 0 – 4.0 scale. (Note that due to weighting, some GPAs may be higher than 4.0.)
- EFC – The Expected Family Contribution of the student as determined by the Free Application for Federal Student Aid (FAFSA)
- COA (Cost of Attendance) - The full cost for a student to attend the school including tuition & fees, housing & meals, transportation costs, books & supplies, plus any personal expenses
- Tuition – The full cost of tuition to attend the school.
- Federal Aid - The amount in dollars awarded to the student in the form of federal government grants such as the Pell Grant.
- State Aid - The amount in dollars awarded to the student in the form of state government grants.
- Merit Aid – The amount in dollars awarded to the student by the institution in the form of a merit scholarship
- Need-Based Grant - The amount in dollars awarded to the student by the institution in the form of a need-based grant
- Athletic Scholarship – The amount in dollars awarded to the student by the institution in the form of an athletic scholarship. Each scholarship athlete receives a $2000 award
- Enrolled – “Y” indicates the student enrolled in the school and “N” indicates the student did not enroll

All findings should be submitted electronically to the contest server.

Descriptive statistics (20%) (Difficulty: Easy)

- (2%) What is the average GPA of the students that were admitted to this school?
- (3%) What is the average GPA and ACT of the students that ***Enrolled*** in this school?
- (3%) Find each quartile of the student EFC population
- (3%) What is the *Net Price* for each student? Create a new column to track this.
- (3%) What is the *Net Tuition Revenue* for each student? Create a new column to track this.
- (3%) What is the *% of Need Met* for each student? Create a new column to track this.
- (3%) What is the total *Net Revenue* the school received from Enrolled students?

Visualization (25%) (Difficulty: Medium)

- (10%) Build a histogram based on EFC. What does this tell us about the demographics of the data?
- (15%) Create a box plot comparing the ACT scores of students that enrolled vs did not enroll. What inferences can be drawn from the plot?

Predictive modeling and inferential statistics (25%) (Difficulty: Hard)

Split the dataset into your own training, validation, and testing sets to avoid building models that overfit the training data and will perform poorly. A typical split of a data set is 70% for training, 15% for validation, and 15% for testing.

- (10%) - Using the training data, build a logistical regression model that predicts whether a prospective student will enroll based on the prospect's GPA, EFC, and Net Price.

- (10%) – What is the "AUC" or area under the ROC (receiver operating characteristic) curve for your logistic regression model against your validation data set?

- (5%) – Evaluate your model. Why should the college's Cabinet trust your model?

Apply new awarding methodology (30%) (Difficulty: Medium)

The school wants to test how their enrollment and revenue would be affected by updating their awarding methodology. They want to standardize their % of Need Met to meet the following levels:

EFC	% Need Met
EFC < $5,000	55%
EFC < $10,000	30%
EFC < $25,000	15%
EFC >= $25,000	0%

They also want to introduce a new $1,000 grant for students with an EFC below $10,000. Finally, they want to increase Athletic scholarships for any athletes that have at least a 3.0 GPA by $500.

- (20%) Create a copy of your data and apply the new awarding methodology to recalculate Awards, Net Price, and Net Tuition Revenue.
- (10%) Use your model to predict enrollment for all students using the new awarding methodology

Extra Credit (10%)

- (3%) What is the total sum of the predicted probabilities? How does this compare to the number of students that actually enrolled?
- (3%) Multiply the updated amount of *Net Tuition Revenue* with the enrollment probability for each student. How does the sum of this compare to the sum of *Net Tuition Revenue* for enrolled students calculated earlier?
- (4%) For students that submitted an ACT, multiply their ACT score by their enrollment probability and divide it by the total sum of probabilities for students that submitted an ACT score. How does this compare to average ACT score of students that actually enrolled? Would you recommend the school implement the new financial aid leveraging plan based on their stated goals?