
2018 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 and income data. This competition can be solved using various technology platforms such as R, Excel, or Microsoft SQL. We will provide access to both the raw data (if you want to use R or Excel) and to an on-site Microsoft SQL Server. Contestants may use any software, including access to the Internet, to develop and implement their model.

[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>. And HeidiSQL is available for free at <https://www.heidisql.com/download.php>.

[R, Weka Users Only] R can download it at <https://www.r-project.org/> and Weka can be downloaded at <https://www.cs.waikato.ac.nz/ml/weka/downloading.html>)

CSV and XSLX files containing the data are available on the contest server (printed on your check-in ticket).

Data on the SQL server may be access at:

IP Address: 10.103.1.11

SQL Username: BA

Password: Raptor2018!

Database Name: USITCC_Raptor

Note that the data available in the CSV and XLSX files are the same as what is on the SQL server.

The competition will use the following datasets:

- Education data from the US Department of Education
 - General Description - <https://collegescorecard.ed.gov/data/>
 - <https://ed-public-download.app.cloud.gov/downloads/Most-Recent-Cohorts-All-Data-Elements.csv> (full, raw data)
 - <https://collegescorecard.ed.gov/assets/CollegeScorecardDataDictionary.xlsx> (Data Dictionary)
 - [SQL Users only] Tables: most-recent-cohorts-all-data-elements, most-recent-cohorts-all-data-elementsX. Because this data has more columns than SQL Server can handle, the data is divided into two tables. The tables may be joined on the UnitId, which is the Department of Education's unique institution identifier.

- Income data
 - This is the income data for 2015, which is aggregated for all tax returns for each ZIP code; it is not segmented by income level.
 - <https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-2015-zip-code-data-soi> (See <https://www.irs.gov/pub/irs-soi/15zpallnoagi.csv>)
 - (Data Dictionary available at <https://www.irs.gov/pub/irs-soi/15zpdoc.doc>)
 - [SQL Users only] Table: zpallnoagi

- [SQL Users only] Aside from most-recent-cohorts-all-data-elements, most-recent-cohorts-all-data-elementsX, and zpallnoagi, you should disregard all other tables in the database.

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 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 company's executive team and the report should be able to “stand alone” with enough details and comments which we will use for grading.

COMPETITION DESCRIPTION

You work as a marketing analyst for a large retailer that operates convenience stores at colleges throughout the country. This year, the firm plans to launch new stores at three college campuses. Your task is to analyze and suggest locations for where to locate the new stores. The Chief Marketing Officer (CMO), your boss's boss, has suggested six potential locations. You are to review and rank the CMO's suggested locations. You want to do your best work because this project is politically sensitive and you will be evaluated for a large promotion and raise based on the success of the new stores.

List of suggested college locations from the CMO:

1. Trinity University (San Antonio, TX)
2. Rice University (Houston, TX)
3. Harvard University (Cambridge, MA)
4. Marquette University (Milwaukee, WI)
5. Grinnell College (Grinnell, IA)
6. Harvey Mudd College (Claremont, CA)

In considering the best locations for the convenience store, you will analyze (1) the income levels near the college, (2) the ACT test scores of students who attend each college, (3) the starting salary of students who graduate from each college, and (4) the total number of students on each campus. Note that colleges with strong metrics in all areas may *not* be the best locations for new stores because there will likely be too much competition.

The meeting with the executive team is in four hours so you must work fast!

To accomplish this project, you will use data available from the U.S. Department of Education to evaluate colleges by their academic profile, size, and median early career salaries of their graduates. You will also use information published by the Internal Revenue Service (IRS) to estimate the average income by ZIP code. Once you have organized and are comfortable with this data, you will analyze the CMO's suggested store locations. Finally, you will justify the reasons why the company should accept your recommendations.

Definitions:

- "Median Earnings" means "median earnings of students working and not enrolled 6 years after entry"
- "College Size" means "enrollment of undergraduate certificate/degree-seeking students"

All findings should be submitted electronically to the contest server

Descriptive statistics (25%)

- (3%) For Harvard University, what are the Median Earnings?
- (3%) Find the median ACT score for Harvey Mudd College.
- (4%) Find the College Size for Rice University.
- (5%) From the IRS data, what is the average household income for the ZIP code at which Trinity University is located?
- (10%) How many colleges in the U.S. meet all of the following characteristics: grant predominantly bachelor's degrees; have a Carnegie Code Classification (CCBASIC) of 14-23; have a College Size of at least 300 students; report ACT scores; and report Median Earnings?

[Note: the answer to this question is a single number. This question is designed to help you create a filter so you will have 'clean' data for the remaining questions. **Use the filter you create here for all remaining questions.**]

Comparative/inferential statistics (30%)

For the colleges you identified in the filter above:

- (3%) Plot a scatter plot of median ACT score vs. Median Earnings.
- (3%) Plot a scatter plot of school size vs. Median Earnings.
- (4%) Plot a scatter plot of average household income at the college's ZIP code (IRS) vs. median earnings
- (5%) For the three plots above, *visually identify* which data is most correlated to the median earnings?
- (15%) Calculate the z-scores for all colleges by median ACT, College Size, and Median Earnings. Comment about the nature of any outliers you discover.

[continued on next page]

Predictive modeling (45%)

Your boss wants you to consider the profitability of 10 existing convenience stores when recommending where to open new locations. First, complete the table below.

- (10%) - Data Preparation - Complete the Table Below

| College | Median ACT z-Score | College Size z-Score | Median Earnings z-Score | Average household income in ZIP Code - z-Score |
|---|--------------------|----------------------|-------------------------|--|
| Bryn Mawr College | | | | |
| University of Wisconsin-Madison | | | | |
| Pace University-New York | | | | |
| Pennsylvania State University-Penn State Schuylkill | | | | |
| Xavier University of Louisiana | | | | |
| Stanford University | | | | |
| Kenyon College | | | | |
| Florida Atlantic University | | | | |
| Fordham University | | | | |
| Stephen F Austin State University | | | | |

- (20%) - Create a predictive model

Your boss provided the report below of the profitability of your company's convenience stores for 10 existing locations. Build a model that predicts the profitability of a store based on the z-scores for each college's median ACT, College Size, Median Earnings, and IRS average household income for the college's ZIP code.

| College | Profitability % (negative means loss) |
|---|---------------------------------------|
| Bryn Mawr College | 10.95% |
| University of Wisconsin-Madison | 24.62% |
| Pace University-New York | 17.76% |
| Pennsylvania State University-Penn State Schuylkill | -20.05% |
| Xavier University of Louisiana | -11.30% |
| Stanford University | -4.22% |
| Kenyon College | 3.59% |
| Florida Atlantic University | 16.38% |
| Fordham University | 23.48% |
| Stephen F Austin State University | 6.00% |

- (15%) - Use your predictive model

Using the model you created, predict the profitability of the locations below

| College | Predicted Store Profitability % |
|--------------------------------------|---------------------------------|
| Trinity University (San Antonio, TX) | |
| Rice University (Houston, TX) | |
| Harvard University (Cambridge, MA) | |
| Marquette University (Milwaukee, WI) | |
| Grinnell College (Grinnell, IA) | |
| Harvey Mudd College (Claremont, CA) | |

- (Extra Credit 10%) - Using your predictive model recommend 3 college locations to open up new convenience stores. Justify why your answers are better than those of the CMO.