Data Analytics / Visualization (2020)

Contest Overview and Components

The Data Analytics and Visualization competition is composed of a progressive set of tasks typically found within analytics workflows. This may include tasks for:

data acquisition, data cleaning, wrangling, data integration, simple data analysis (e.g., outlier detection, correlation/collinearity detection, dimension reduction, summary statistics, and skewness), statistical model building, clustering, geospatial analysis, data visualization, visual analytics, and problem solving. The tasks are designed so that a student with minimal experience in any of these areas can learn from materials introduced along the way.

Students will be given the problem statement at the beginning of the contest time. Time will be allowed to read the problem statement and ask any questions in a common session. Further questions will be permitted at any point during the contest. At the end of the contest period, the judging team will inspect each student's progress for grading. The contest is divided into multiple tasks, and those that complete the most tasks correctly gain more points. Teams may want to have relevant tools (e.g., statistical and visualization software) preloaded on their machines prior to the competition.

Team Composition

Individuals or teams of 2 students can register for this competition.

Skills

All analytics skills and techniques needed to do well in this contest will be presented as you go. Common techniques (equally weighted) include acquiring data, handling missing values, finding outliers, identifying collinearity between variables, determining skewed attributes, finding patterns, finding clusters, plotting geo-temporal points on a map, recommendation, and decision making. Experience with programming and statistical analysis is preferred, but a motivated student with no prior experience can learn to complete multiple tasks within this contest. Students wishing to be successful should have an inquisitive nature, willingness to learn new techniques, and a drive to succeed.

For the competition, points will be given for successful and correct completion of tasks. In case of a tie, the winners' partial, incomplete, or incorrect submissions will be analyzed and partial points will be given for attempted but incorrect challenges. In the event of multiple students completing the entire contest 100% correctly, the student to complete the contest first will be given the win.

Scoring

The rubric for scoring will be distributed with the problem statement at the beginning of the contest time. General guidelines are in the Contest Overview above.

Schedule

See contest schedule for contest time and place. The time will include contest overview, contest work, and turn in. Check in will begin 20 minutes before the contest begins.

Resources

Each team must have at least 1 computer to complete your project.

Tableau (free software available to students) is highly recommended for this contest. Students are recommended to download, install, and register the visualization software ahead of time. This step will save some time at the beginning of the contest (especially in receiving a free student license code) but is not required as the installation may only take a few minutes. Alternative analytics or visualization software that students may find useful include Excel, R, Python, SAS, SPSS, Weka, and/or Power BI. Students are encouraged to use any software and online resources they wish, including Google search, to learn and complete the challenges.