

# Data Science C

1. **DESCRIPTION:** Teams will demonstrate a theoretical and practical knowledge of programming, with a focus on data analysis techniques and algorithms.

**TEAMS OF UP TO:** 2

**EYE PROTECTION:** None

**IMPOUND:** None

**EVENT TIME:** 50 minutes

2. **EVENT PARAMETERS:** Teams will complete both a written test and an on-site series of coding challenges in Python 3 using the Jupyter Notebook interface.
- Teams must bring writing utensils. Teams may bring a single calculator (which must not have access to the internet). Teams may bring any amount of printed or written notes.
  - The event supervisor will arrange to provide each team with a single computer which is capable of editing and running Jupyter Notebooks with Python 3. Any necessary external packages (e.g. NumPy, scikit-learn) will be installed. In the case of a virtual tournament, participants will complete the challenges using a Google Colab notebook.
  - Internet access is forbidden at all times. Violation of this rule will result in immediate disqualification.
3. **THE COMPETITION:**
- Written Test:** Teams will demonstrate a theoretical understanding of the following topics:
    - Python 3 Syntax & Programming Knowledge.** Includes variables, data types, control flow, methods/functions, classes, object-oriented programming, data structures, algorithms, and time complexity analysis.
    - Statistical Methods.** Includes common statistical distributions, central tendency and dispersion, correlation and covariance, joint/marginal/conditional distributions, visualizing data, central limit theorem, and confidence intervals/z-score.
    - Machine Intelligence.** Includes high-dimensional data, feature extraction algorithms, principal component analysis, k-means, binary classification (decision trees, k-nearest neighbors, SVM), linear regression, gradient descent, and neural networks.
  - Coding Challenges:** Teams will demonstrate a practical understanding of programming concepts.
    - Each team will be given a single notebook with several programming-based questions. Each question will have a detailed description of the problem, input, and expected output. Teams must code solutions to each problem in order to earn points.
    - Sample input and output data will be provided to teams. Teams' solutions will be scored on these sample cases, as well as on hidden test cases.
    - Topics that can be tested over include but are not limited to algorithms, data structures, data analysis techniques, and machine learning.
    - Teams will complete and submit their solutions in accordance with the event supervisor's instructions.
    - If teams are expected to use external libraries, the supervisor will provide the APIs. Examples of such libraries include math, NumPy, scikit-learn, and Pandas.
4. **SCORING:** Highest score wins.
- Final Score = Written Score (WS) + Coding Score (CS). WS will be worth 60% of the total score, whereas CS will be worth 40%.

- b. Coding challenges will be scored on correctness. Intentionally obfuscated code will receive 0 points. Any solution that hardcodes the outputs will receive 0 points.
- c. Tiebreakers will be based on (in order of priority): CS subscore, code readability (well-documented variables, organization, etc.), first written question missed.

**5. RECOMMENDED RESOURCES:**

- a. <https://www.atxscioly.org/resources> (past tests)
- b. <https://codingbat.com/python>
- c. <https://leetcode.com/problemset/all/>
- d. <https://www.codecademy.com/learn/learn-statistics-with-python>
- e. <https://towardsdatascience.com/beginners-guide-to-machine-learning-with-python-b9ff35bc9c51>
- f. [https://www.youtube.com/watch?v=aircAruvnKk&list=PLZHQObOWTODNU6R1\\_67000Dx\\_ZCJB-3pi](https://www.youtube.com/watch?v=aircAruvnKk&list=PLZHQObOWTODNU6R1_67000Dx_ZCJB-3pi)
- g. <https://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one>