**Data Engineering Research Assistantship Position/Research Collaboration Opportunity**

We are seeking MS/PhD/advanced undergraduate students for a part-time data engineering research assistant position, that can also lead to collaboration on a research paper. The candidate will be responsible for extracting data from large “game” datasets coming from a digital platform that teaches children coding, which consist of a complex nested structure. The datasets include game-level information, user interactions, and stage-specific data. For each state in the game, there is an associated action sequence and a corresponding code generated in XML format for each level.

**Tasks:**

* Extract, organize, and process game data that includes game, user, and stage information.
* Navigate and manipulate complex nested data structures to derive insights/construct requested variables from action sequences and XML data.
* Use Python and relevant libraries (e.g., pandas) to analyze and clean large datasets.
* Manage data using Excel for reporting and documentation purposes.
* Engage in data analysis tasks such as clustering from extracted data.

**Required Skills:**

* Strong understanding of data structures and algorithms.
* Proficiency in Python programming.
* Experience working with nested data structures.
* Familiarity with pandas for data manipulation and analysis.
* Ability to work with Excel for data presentation and reporting.

Experience with XML data parsing and processing, and prior exposure to game development or working with game-related data would be a plus.

The position provides an excellent opportunity for students to gain hands-on experience in data engineering and game data analysis.

This position can be also viewed as part of a **“senior design project”** for undergraduate students, whereby the student can work on a project related to the data, such as preparing an “auto extractor” with a graphical interface to present variables from the dataset, construct clusters of coding behavior, or apply machine learning algorithms to study the predictive power of game-level interactions on specific outcomes.

Master’s/PhD students who are interested in the position and in using part of the data as a thesis chapter are welcome to talk to the PI, Prof. Seda Ertac ([sertac@ku.edu.tr](mailto:sertac@ku.edu.tr)).