Data Science with Python Syllabus

High School - One Month (20-25 hours)

Course Overview and Goals

Industries of all types are hiring data scientists to analyze and highlight the hidden patterns in data. This course equips students with the essential skills of a data scientist which include data collection, cleanup, transformation, analysis, and visualization. Students will write algorithms, tell data stories, and build statistical models using Python libraries. They will use the same tools that data scientists use to draw meaningful insights and solve organizational problems.

Learning Environment

This course utilizes a blended classroom approach. The content is fully web-based, with students writing and running code in the browser. Each module of the course is broken down into lessons. Lessons consist of video tutorials, short quizzes, example programs to explore, and written programming exercises.

Programming Environment

Students write and run Python programs in the browser using the CodeHS editor.

Prerequisites

The Data Science course is designed for intermediate computer science students with at least some knowledge of programming (not language specific) and an interest in computer science. The course is highly visual, dynamic, and interactive, and engaging.

Supplemental Project

There is an optional, supplemental project included in this course. If students have prior Python knowledge they should be able to complete this module in the one-month course.

More Information

Browse the content of this course at https://codehs.com/course/15387

Course Breakdown

Module 1: The Data Science Life Cycle (3-4 weeks/15-20 hours)

Students will learn and apply the process of the data science life cycle. This includes asking statistical questions, collecting or obtaining reliable raw data, analyzing the data using measures of central tendency and spread and interpreting, and summarizing the results.

Objectives / Topics Covered	- What is Data Science? - Gathering Data - Quantitative/Qualitative - Exploring Data Using Python

	- Modules and Libraries - Using the Pandas Library - Series - Measures of Central Tendency - Measures of Spread - DataFrames - Selecting Columns - Using Functions
Example Assignments / Projects	**Mini-Project:** Students will go through the first two steps of the data cycle using data of their choosing. - Ask Questions: Formulate a statistical question that can be answered with data. - Consider Data: Collect or find data that will aid in answering your question. - Analyze Data: Perform statistical analysis, run calculations and/or create data displays to identify patterns and relationships - Interpret Data: Answer questions and summarize the results. **Hot Dog Plots:** Use the correct Python functions to create a boxplot of the data. Using the graph, determine the summary statistics and the spread. **Roller Coaster Rankings:** Define a function that will compute a score for each roller coaster. Use this function to store the results in a new column. **Student Test Scores:** Create a function that finds the maximum test score between test one and test two for each student. Create a function that finds the maximum test score between all three tests for each student. Decide which calculations, along with these two new columns, can help you answer the original statistical question? Explore and further analyze your data until you come to a conclusion.

Supplemental: Data Science for Change (1-2 weeks/5-10 hours)

Students will use and analyze data to better understand a problem, measure the scope of a problem, or understand how people are affected by the problem. They will learn more about cleaning a dataset and filtering by column, rows, and conditions.

Objectives / Topics Covered	- What is Big Data? - Cognitive Bias - Importing and Filtering `loc` `iloc` - By a condition - Data Cleaning - Dropping Data - Fixing Data Types - Exploring with Data Visualizations
Example Assignments / Projects	**Project - Data Science for Change:** Students will run through the data science life cycle with the intent to use data to better understand a problem, to measure the scope of a problem, and to understand how people are

	affected by the problem. ● **Instagram Filters:** This dataset consists of popular Instagram accounts and their number of followers (in millions). Use conditional filtering to print the rows where followers are greater than 230 (million), print only the account columns of those from the United States, print the account and followers columns of the row with the maximum number of followers. ● **Book Conditions**: This data was acquired from the Google Books store. It includes the title of each book, the author(s), the rating (from 1-5), the total voters, the price, the publisher, the page count, and the date the book was published. Print the title, rating, and voters columns for books that have a rating of 4.0 or higher and over 9000 voters. ● **Cleaning Book Data:** Check the data types of the dataset. Do they look okay? Do they need to be changed at all? Permanently drop the publisher and published_date columns. Print the shape of the data and check for duplicate rows. How many are there? Permanently drop duplicate rows from the dataset. Determine the number of missing values in the dataset. What would be the best decision for dealing with the missing values? Make the call and change the data.

Supplemental: What's Next?

Students will explore the next chapter in learning about data science and the careers that are available and growing.

Objectives / Topics Covered	● Data Science Pathways ● Artificial Intelligence ● Python Programming