

# EDx Data Science for Construction, Architecture and Engineering

Launch 3 Syllabus

Sept. 30, 2020 - Mar. 3, 2021



**NUS**  
National University  
of Singapore



Link:

<https://www.edx.org/course/Data-Science-for-Construction-Architecture-and-Engineering>

Instructor: Dr Clayton Miller

Assistants (alphabetical order): Ananya Joshi, Charlene Tan, Chun Fu, James Zhan, Mahmoud Abdelrahman, Matias Quintana, Miguel Martin, and Vanessa Neo.

This course introduces data science skills targeting applications in the design, construction, and operations of buildings. You will learn practical coding within this context with an emphasis on basic Python programming and the Pandas library.

<b>Overview</b>	<b>2</b>
<b>Section 1: Introduction to Course and Python Fundamentals</b>	<b>3</b>
<b>Section 2: Introduction to the Pandas Data Analytics Library and Design Phase Application Example</b>	<b>4</b>

<b>Section 3: Pandas Analysis of Time-Series Data from IoT and Construction Phase Application Example</b>	<b>5</b>
<b>Section 4: Statistics and Visualization Basics and Operations Phase Application Example</b>	<b>6</b>
<b>Section 5: Introduction to Machine Learning for the Built Environment</b>	<b>7</b>
<b>EDx Grading and Evaluation Policy</b>	<b>8</b>
<b>ECTS Credit Equivalent</b>	<b>9</b>

## Overview

The building industry is exploding with data sources that impact the energy performance of the built environment and health and well-being of occupants. Spreadsheets just don't cut it anymore as the sole analytics tool for professionals in this field. Participating in mainstream data science courses might provide skills such as programming and statistics, however, the applied context to buildings is missing, which is the most important part for beginners.

This course focuses on the development of data science skills for professionals specifically in the built environment sector. It targets architects, engineers, construction and facilities managers with little or no previous programming experience. An introduction to data science skills is given in the context of the building life cycle phases. Participants will use large, open data sets from the design, construction, and operations of buildings to learn and practice data science techniques.

Essentially this course is designed to add new tools and skills to supplement spreadsheets. Major technical topics include data loading, processing, visualization, and basic machine learning using the Python programming language, the Pandas data analytics and sci-kit learn machine learning libraries, and the web-based Colaboratory environment. In addition, the course will provide numerous learning paths for various built environment-related tasks to facilitate further growth.

The following table outlines the Sections of this course:

Section 1	Introduction to Course and Python Fundamentals	<ul style="list-style-type: none"> <li>- Overview of the course, tools, and data</li> <li>- Overview of the phases of the building industry - design, construction, operations</li> <li>- Key Python concepts</li> <li>- Introduction to Variables and Control Statements</li> <li>- Python Data Structures</li> <li>- An overview of the Real-World Data Science Applications</li> <li>- Two exercises with discussion support and two quizzes (Verified Track)</li> </ul>
Section 2	Introduction to the Pandas Data Analytics Library and Design Phase Application Example	<ul style="list-style-type: none"> <li>- Overview of the Design Phase of Buildings</li> <li>- Introduction to the Pandas data analysis library</li> <li>- Basics of Pandas data structures</li> <li>- Common Pandas functions</li> <li>- Applications for designers focused on EnergyPlus, Revit, Rhino, spatial analytics, generative design</li> <li>- Two exercises with discussion support and two quizzes (Verified Track)</li> </ul>
Section 3	Pandas Analysis of Time-Series Data from IoT and Construction Phase Application Example	<ul style="list-style-type: none"> <li>- Overview of the Construction Phase of Buildings</li> <li>- Overview of time-series data and its challenges</li> <li>- Deep dive in the Pandas functions specific to time-series data</li> <li>- Applications overview for project management, building management system (BMS) data analysis, and digital construction such as robotic fabrication.</li> <li>- Two exercises with discussion support and two quizzes (Verified Track)</li> </ul>
Section 4	Statistics and Visualization Basics and Operations Phase Application Example	<ul style="list-style-type: none"> <li>- Overview of the Operations Phase of Buildings</li> <li>- Discussion of occupant satisfaction in buildings</li> <li>- Pandas statistical aggregations and visualization techniques in Pandas</li> <li>- An introduction to the Seaborn visualization library</li> <li>- Applications for operations focused on occupant satisfaction with a focus on spatial optimization and feedback collection</li> <li>- Two exercises with discussion support and two quizzes (Verified Track)</li> </ul>
Section 5	Introduction to Machine Learning for the Built Environment	<ul style="list-style-type: none"> <li>- A high level overview of how prediction can impact the building industry</li> <li>- Implementation of prediction, classification, and clustering using the sci-kit learn library</li> <li>- Application to smart meter and occupant comfort data for these techniques</li> <li>- Applications focused on machine learning in the building industry and example of a Kaggle competition</li> <li>- Concluded with suggestions on more in-depth Python, Data Science, and Statistics courses on EDx</li> <li>- Two exercises with discussion support and two quizzes (Verified Track)</li> </ul>

## Section 1: Introduction to Course and Python Fundamentals

In this introduction, an overview of key Python concepts is covered as well as the motivating factors for building industry professionals to learn to code. The NZEB at the NUS School of Design and Environment is introduced as an example of a building that uses various data science-related technologies in its design, construction, and operations.

Introduction and Orientation Videos	Introduction Video - Overview of Course	Video
	Data Sets Overview	Video
	Tools and Techniques Overview	Video
	Course and Section 1 High Level Overview	Video
Python Fundamentals - Variables and Operators	Link to the Colab Notebooks and Discussion Board	Discussion
	Colab Introduction Video by Google	Video
	Introduction to Colaboratory	Video
	Starting with Notebooks	Video
	Python Variables and Operators	Video
	Comparison Operators	Video
Python Fundamentals - Variables and Operators Feedback and Evaluation (Verified Track Users Only)	Exercise 1.1 Explanation Video	Video
	Exercise 1.1 and Discussion Board	Exercise
	Quiz 1.1 - Python Fundamentals I	Quiz
Python Fundamentals - Control Statements, Functions, and Libraries	Control Statements	Video
	Data Types	Video
	Python Functions	Video
	Python Libraries	Video
	Python Data Structures	Video
Python Fundamentals - Control Statements, Functions, and Libraries Feedback and Evaluation (Verified Track Users Only)	Exercise 1.2 Explanation Video	Video
	Exercise 1.2 and Discussion Board	Exercise
	Quiz 1.2 - Python Fundamentals II	Quiz
Real-World Data Science Tools and Applications	Explanation of Real-World Data Science Tools and Applications	Video
	Introduction to Built Environment Data Science Characters	Video
	SDE4 Case Study Introduction	Video

## Section 2: Introduction to the Pandas Data Analytics Library and Design Phase Application Example

The foundational functions of Pandas are demonstrated in the context of the integrated design process through the processing of data from parametric EnergyPlus models. Further future learning path examples are introduced for the Design Phase including building information modelling (BIM) using Revit or Rhino, spatial analytics, and building performance modelling Python libraries.

Design Phase Data Science Overview	Design Phase of the Building Industry	Video
	Introduction to the Pandas Library and its Functions	Video
	Section 2 High Level Overview	Video
Pandas Library Fundamentals	Link to the Colab Notebooks and Discussion Board	Discussion
	Introduction to Pandas Library	Video
	Introduction to Case Study	Video
	Pandas DataFrame Object	Video
	Pandas File Input Functions	Video
	Selecting Subsets of Data from DataFrames	Video
	Introduction to Plotting with Pandas	Video
Pandas Library Fundamentals Feedback and Evaluation (Verified Track Users Only)	Exercise 2.1 Explanation Video	Video
	Exercise 2.1 and Discussion Board	Exercise
	Quiz 2.1 - Pandas Fundamentals I	Quiz
Application Example - Simulation Data Processing	Pandas Concatenate Function to Combine Data	Video
	Making a New Column and Function	Video
	Comparison of Design Options	Video
	Debugging Problems and using StackOverflow	Video
Application Example - Simulation Data Processing Feedback and Evaluation (Verified Track Users Only)	Exercise 2.2 Explanation Video	Video
	Exercise 2.2 and Discussion Board	Exercise
	Quiz 2.2 - Pandas Fundamentals II	Quiz
Design Phase Real-world Data Science Tools and Applications	Design Phase Data Science Characters	Video
	Rhino and Grasshopper Platform	Video
	Generative Design	Video
	Ladybug Tools for Performance Simulation	Video
	Revit Python Shell Introduction	Video
	EnergyPlus Automation	Video
	Designbuilder and IESVE Python	Video
	SDE4 Case Study: Integrated Design Process	Video
	Design Phase Discussion	Discussion

## Section 3: Pandas Analysis of Time-Series Data from IoT and Construction Phase Application Example

Time-series analysis Pandas functions are demonstrated in the Construction Phase through the analysis of hourly IoT data from electrical energy meters. Further future learning path examples are introduced for the Construction Phase including project management, building management system (BMS) data analysis, and digital construction such as robotic fabrication.

Construction Phase Data Science Overview	Construction Phase of the Building Industry	Video
	Pandas for Time Series Data	Video
	Section 3 High Level Overview	Video
Analysis of Time-Series IoT Data	Link to the Colab Notebooks and Discussion Board	Discussion
	Introduction to the Building Data Genome Project	Video
	Loading Time-Series Data using Pandas	Video
	Time-Series Objects in Pandas DataFrame	Video
	Time-Series Frequency Resampling Function	Video
	Time-Series Analysis Truncation Function	Video
	Visualizing Patterns in Time-Series Data	Video
Analysis of Time-Series IoT Data Feedback and Evaluation (Verified Track Users Only)	Exercise 3.1 Explanation Video	Video
	Exercise 3.1 and Discussion Board	Exercise
	Quiz 3.1 - Time-Series Pandas I	Quiz
Normalization and Weather Data Analysis	Building Energy Data Normalization Process	Video
	Energy Comparison of Group of Buildings	Video
	Removing Outliers in Time-Series Weather Data	Video
	Merging Weather and Energy Data to Understand Te	Video
Normalization and Weather Data Analysis Feedback and Evaluation (Verified Track Users Only)	Exercise 3.2 Explanation Video	Video
	Exercise 3.2 and Discussion Board	Exercise
	Quiz 3.2 - Time-Series Pandas II	Quiz
Construction Phase Real-world Data Science Tools and Applications	Construction Phase Data Science Characters	Video
	Building Commissioning Process	Video
	Getting Time-Series IoT/BMS Data from REST API	Video
	BRICK Metadata Schema for Buildings	Video
	Digital Fabrication and Robotics	Video
	SDE4 Case Study: Energy and Comfort Systems	Video
	Construction Phase Discussion	Discussion

## Section 4: Statistics and Visualization Basics and Operations Phase Application Example

Various statistical aggregations and visualization techniques using Pandas and the Seaborn library are demonstrated on Operations Phase occupant comfort data from the ASHRAE Thermal Comfort Database II. Further future learning path examples are introduced for the Operations Phase including energy auditing, IoT analysis, and occupant detection and reinforcement learning.

Operations Phase Data Science Overview	Operations Phase of the Building Industry	Video
	Pandas Visualization and Statistics	Video
	Section 4 High Level Overview	Video
Pandas Statistics, Reshaping, and Visualization Functions	Link to the Colab Notebooks and Discussion Board	Discussion
	Introduction to Thermal Comfort Data	Video
	Pandas Statistical Descriptor Functions	Video
	Pandas Aggregation Functions	Video
	Pivoting and Reshaping Functions	Video
	Grouping Functions	Video
	Box Plot Visualizations	Video
Pandas Statistics, Reshaping, and Visualization Functions Feedback and Evaluation (Verified Track Users Only)	Scatterplot and Histogram Visualizations	Video
	Exercise 4.1 Explanation Video	Video
	Exercise 4.1 and Discussion Board	Exercise
Analysis of Thermal Comfort Data in Building Operations	Quiz 4.1 - Pandas Stats and Viz I	Quiz
	What makes people feel comfortable?	Video
	What personal factors contribute to comfort?	Video
Analysis of Thermal Comfort Data in Building Operations Feedback and Evaluation (Verified Track Users Only)	Does the location of the study have an impact?	Video
	Exercise 4.2 Explanation Video	Video
	Exercise 4.2 and Discussion Board	Exercise
Operations Phase Real-world Data Science Tools and Applications	Quiz 4.2 - Pandas Stats and Viz II	Quiz
	Operations Phase Data Science Characters	Video
	WELL Certification and CBE Survey	Video
	Occupant Feedback Case Study: Changi Airport	Video
	SDE4 Case Study: Learning Trail and Cozie Smartwatch Feedback	Video
	SDE4 Case Study: SpaceMatch - Space optimization and allocation	Video
Operations Phase Discussion	Discussion	

## Section 5: Introduction to Machine Learning for the Built Environment

This concluding section gives an overview of the motivations and opportunities for the use of prediction in the built environment. Prediction, classification, and clustering using the sci-kit learn library is demonstrated on electrical meter and occupant comfort data. The course is concluded with suggestions on more in-depth Python, Data Science, and Statistics courses on EDx.

Overview of Introduction to ML for Buildings	Machine Learning and Prediction for Buildings	Video
	Python ML Library - Scikit Learn Introduction	Video
	Section 5 High Level Overview	Video
ML Introduction and Unsupervised Clustering	Link to the Colab Notebooks and Discussion Board	Discussion
	Introduction to Scikit-Learn Library	Video
	Unsupervised Machine Learning Introduction	Video
	Daily Load Profile Generation	Video
	k-Means Clustering of Daily Load Profiles	Video
ML Introduction and Unsupervised Clustering Feedback and Evaluation (Verified Track Users Only)	Exercise 5.1 Explanation Video	Video
	Exercise 5.1 and Discussion Board	Exercise
	Quiz 5.1 - Introduction to ML I	Quiz
Supervised Machine Learning for Regression and Classification	Introduction to Regression	Video
	Regression for Hourly Electricity Prediction	Video
	Introduction to Classification Prediction	Video
	Random Forest Machine Learning Model for Classification	Video
	Classification Performance Metrics and Visualization	Video
Supervised Machine Learning for Regression and Classification Feedback and Evaluation (Verified Track Users Only)	Exercise 5.2 Explanation Video	Video
	Exercise 5.2 and Discussion Board	Exercise
	Quiz 5.2 - Introduction to ML II	Quiz
Machine Learning Real-world Data Science Tools and Applications	Machine Learning Characters	Video
	Kaggle Machine Learning Platform Overview	Video
	ASHRAE Great Energy Predictor Competition	Video
	ML Application Discussion	Discussion
	Conclusion and Tips to Keep Learning	Video
	Final Discussion Board	Discussion

# EDx Grading and Evaluation Policy

For Verified Track users, this course is graded through two quizzes from each section. To achieve the Verified Certificate, the participant will need to have a grade of 75% or higher. There are 10 total quizzes and the lowest two will be dropped for the final score. Each quiz has 10 points - 7 are multiple-choice questions and 3 points are available from the exercises.

## Overall Grade Range

Your overall grading scale for student final grades



## Assignment Types

Categories and labels for any exercises that are gradable

<b>Assignment Type Name</b> <input type="text" value="Quiz"/> <small>The general category for this type of assignment, for example, Homework or Midterm Exam. This name is visible to learners.</small>	<b>Abbreviation</b> <input type="text" value="Quiz"/> <small>This short name for the assignment type (for example, HW or Midterm) appears next to assignments on a learner's Progress page.</small>	
<b>Weight of Total Grade</b> <input type="text" value="100"/> <small>The weight of all assignments of this type as a percentage of the total grade, for example, 40. Do not include the percent symbol.</small>	<b>Total Number</b> <input type="text" value="10"/> <small>The number of subsections in the course that contain problems of this assignment type.</small>	<b>Number of Droppable</b> <input type="text" value="2"/> <small>The number of assignments of this type that will be dropped. The lowest scoring assignments are dropped first.</small>

## ECTS Credit Equivalent

Upon completion of the Verified Track version of this course, a participant will have completed approximately 25-30 hours of online video instruction and hands-on exercise with support from myself and a teaching assistant team. This course is equivalent to 1 ECTS credit upon completion of the Verified Certificate track that includes 10 evaluation quizzes that test the understanding of the participant.

*The Verified Certificate from EDx with name and dates must be included with this document to show full participation in the course.*

Please email me at [clayton@nus.edu.sg](mailto:clayton@nus.edu.sg) for clarifications.

Dr. Clayton Miller  
Asst. Professor  
National University of Singapore