Data Analysis with Python

by IBM

About this Course
Analyzing data with Python is an essential skill for Data Scientists and Data Analysts. This course will take you from the basics of data analysis with Python to building and evaluating data models.

Topics covered include:
- collecting and importing data
- cleaning, preparing & formatting data
- data frame manipulation
- summarizing data
- building machine learning regression models
- model refinement
- creating data pipelines

You will learn how to import data from multiple sources, clean and wrangle data, perform exploratory data analysis (EDA), and create meaningful data visualizations. You will then predict future trends from data by developing linear, multiple, polynomial regression models & pipelines and learn how to evaluate them.

In addition to video lectures you will learn and practice using hands-on labs and projects. You will work with several open source Python libraries, including Pandas and Numpy to load, manipulate, analyze, and visualize cool datasets. You will also work with scipy and scikit-learn, to build machine learning models and make predictions.

If you choose to take this course and earn the Coursera course certificate, you will also earn an IBM digital badge.

Show less

Taught by: Joseph Santarcangelo, Ph.D., Data Scientist at IBM
IBM Developer Skills Network

Basic Info	Course 7 of 12 in the IBM Data Science Specialization
Level	Intermediate
Commitment	This course requires approximately two hours a week for six weeks.
Language	English, **Subtitles:** Arabic, French, Bengali, Ukrainian, Chinese (Simplified), Greek, Italian, Portuguese (Brazil), Dutch, Korean, Oriya, German, Pashto, Urdu, Russian, Thai, Indonesian, Swedish, Turkish, Azerbaijani, Spanish, Dari, Hindi, Japanese, Kazakh, Hungarian, Polish
How To Pass	Pass all graded assignments to complete the course.
User Ratings	Average User Rating 4.7

Syllabus

Module 1

Importing Data Sets

In this module, you will learn how to understand data and learn about how to use the libraries in Python to help you import data from multiple sources. You will then learn how to perform some basic tasks to start exploring and analyzing the imported data set.

6 videos, 1 reading

1. **Video:** Course Introduction
2. **Video:** Understanding the Data
3. **Video:** Python Packages for Data Science
4. **Video:** Importing and Exporting Data in Python
5. **Video:** Getting Started Analyzing Data in Python
6. **Video:** Accessing Databases with Python
7. **Reading:** Lesson Summary
8. **Graded Assignment:** Practice Quiz: Importing Data Sets
9. **App Item:** Lab: Importing Data Sets - Used Cars Pricing
10. **Ungraded Plugin:** Overview: Laptop Pricing Data Set
11. **App Item:** Lab: Importing Datasets - Laptop Pricing
12. **Ungraded Plugin:** Module 1 Cheat Sheet: Importing Data Sets

Show less

Graded: Graded Quiz: Importing Data Sets

Module 2

Data Wrangling

In this module, you will learn how to perform some fundamental data wrangling tasks that, together, form the pre-processing phase of data analysis. These tasks include handling missing values in data, formatting data to standardize it and make it consistent, normalizing data, grouping data values into bins, and converting categorical variables into numerical quantitative variables.

6 videos, 1 reading

1. **Video:** Pre-processing Data in Python
2. **Video:** Dealing with Missing Values in Python
3. **Video:** Data Formatting in Python
4. **Video:** Data Normalization in Python
5. **Video:** Binning in Python
6. **Video:** Turning Categorical Variables into Quantitative Variables in Python
7. **Reading:** Lesson Summary
8. **Graded Assignment:** Practice Quiz: Data Wrangling
9. **App Item:** Lab: Data Wrangling - Used Cars Pricing
10. **App Item:** Lab: Data Wrangling - Laptop Pricing
11. **Ungraded Plugin:** Module 2 Cheat Sheet: Data Wrangling

Show less

Graded: Graded Quiz: Data Wrangling

Module 3

Exploratory Data Analysis

In this module, you will learn what is meant by exploratory data analysis, and you will learn how to perform computations on the data to calculate basic descriptive statistical information, such as mean, median, mode, and quartile values, and use that information to better understand the distribution of the data. You will learn about putting your data into groups to help you visualize the data better, you will learn how to use the Pearson correlation method to compare two continuous numerical variables, and you will learn how to use the Chi-square test to find the association between two categorical variables and how to interpret them.

Show less

5 videos, 1 reading

1. **Video:** Exploratory Data Analysis
2. **Video:** Descriptive Statistics
3. **Video:** Groupby in Python
4. **Ungraded Plugin:** Creating Different Types of Plots in Python
5. **Video:** Correlation
6. **Video:** Correlation - Statistics
7. **Ungraded Plugin:** Chi-Square Test for Categorical Variables
8. **Reading:** Lesson Summary
9. **Graded Assignment:** Practice Quiz: Exploratory Data Analysis
10. **App Item:** Lab: Exploratory Data Analysis - Used Car Pricing
11. **App Item:** Lab: Exploratory Data Analysis - Laptop Pricing
12. **Ungraded Plugin:** Module 3 Cheat Sheet: Exploratory Data Analysis

Show less

Graded: Graded Quiz: Exploratory Data Analysis

Module 4

Model Development

In this module, you will learn how to define the explanatory variable and the response variable and understand the differences between the simple linear regression and multiple linear regression models. You will learn how to evaluate a model using visualization and learn about polynomial regression and pipelines. You will also learn how to interpret and use the R-squared and the mean square error measures to perform in-sample evaluations to numerically evaluate our model. And lastly, you will learn about prediction and decision making when determining if our model is correct.

Show less

6 videos, 1 reading

1. **Video:** Model Development
2. **Video:** Linear Regression and Multiple Linear Regression
3. **Video:** Model Evaluation using Visualization
4. **Video:** Polynomial Regression and Pipelines
5. **Video:** Measures for In-Sample Evaluation
6. **Video:** Prediction and Decision Making
7. **Reading:** Lesson Summary
8. **Graded Assignment:** Practice Quiz: Model Development
9. **App Item:** Lab: Model Development - Used Car Pricing
10. **App Item:** Lab: Model Development - Laptop Pricing
11. **Ungraded Plugin:** Module 4 Cheat Sheet: Model Development

Show less

Graded: Graded Quiz: Model Development

Module 5

Model Evaluation and Refinement

In this module, you will learn about the importance of model evaluation and discuss different data model refinement techniques. You will learn about model selection and how to identify overfitting and underfitting in a predictive model. You will also learn about using Ridge Regression to regularize and reduce standard errors to prevent overfitting a regression model and how to use the Grid Search method to tune the hyperparameters of an estimator.

4 videos, 1 reading

1. **Video:** Model Evaluation and Refinement
2. **Video:** Overfitting, Underfitting and Model Selection
3. **Ungraded Plugin:** Introduction to Ridge Regression
4. **Video:** Ridge Regression
5. **Video:** Grid Search
6. **Reading:** Lesson Summary
7. **Graded Assignment:** Practice Quiz: Model Evaluation and Refinement
8. **App Item:** Lab: Model Evaluation and Refinement - Used Cars Pricing
9. **App Item:** Lab: Model Evaluation and Refinement - Laptop Pricing
10. **Ungraded Plugin:** Module 5 Cheat Sheet: Model Evaluation and Refinement

Show less

Graded: Graded Quiz: Model Evaluation and Refinement

Module 6

Final Assignment

Congratulations! You have now completed all the modules for this course. In this last module, you will complete the final assignment that will be graded by your peers. In this final assignment, you will assume the role of a Data Analyst working at a real estate investment trust organization who wants to start investing in residential real estate. You will be given a dataset containing detailed information about house prices in the region based on a number of property features, and it will be your job to analyze and predict the market price of houses given that information.

Show less

5 readings

1. **Ungraded Plugin:** Practice Project Overview
2. **App Item:** Practice Project - Data Analytics for Insurance Cost Data Set
3. **Reading:** Final Project Scenario
4. **App Item:** Lab for Final Project - Data Analytics for House Pricing Data Set
5. **Reading:** Cheat Sheet: Data Analysis for Python
6. **Reading:** IBM Digital Badge
7. **Reading:** Congratulations and Next Steps
8. **Reading:** Thanks from the Course Team

Show less

Graded: Submit your Project and Review Others

Graded: Final Exam

View Less

How It Works

General

How do I pass?

To earn your Certificate, you'll need to earn a passing.

More

Peer-graded assignments

Peer-graded assignments require you and your classmates to grade each other's work.

More

Course 7 of Specialization

Prepare for a career as a data scientist
Build job-ready skills – and must-have AI skills – for an in-demand career. Earn a credential from IBM. No prior experience required.

IBM Data Science
IBM

Learn More

View the course in catalog

Related Courses

Data Science Challenge
Coursera Project Network

Foundations of Data Science: K-Means Clustering in Python
University of London, Goldsmiths, University of London