Goskills

Introduction to Data Science

GoSkills online course syllabus

Skill level
Beginner

Lessons
34

Pre-requisites
No prior experience needed

Video duration
2h 1m

Estimated study time
2h 1m

Instructor
Lavanya Vijayan

Introduction

1 Beginning your data science exploration
Data science is a vast field, with plenty of opportunity and exciting innovation happening.

Defining Data Science

2 Demystifying data science
Data Science is rapidly increasing in popularity and demand, and is a valuable skill as both a career or skill within an existing role.

3 The value of data science
Data Science can be used across numerous fields and offers important benefits to the world around you.

4 Defining the data science life cycle
Data scientists follow a specific workflow.

Starting with Data Design

5 Reducing bias with probability sampling
Data design, the process of data collection, is important in data science.

6 Using non-probability sampling
You can also collect data with non-probability sampling techniques.

Utilizing Computational Tools

7 Comparing Python and R

Two of the most popular computing languages for data science are currently Python and R.

8 Setting up your Jupyter environment

You'll want to set up your data science projects to be successful, and using Jupyter notebook is a great way to do so.

Structuring Your Tabular Data

9 Defining tabular data

Datasets can be structured in many ways, but they're easiest to work with when structured in a table.

10 Reading tabular data

Once you have access to a dataset, you will need to interact with it and read the data most quickly and efficiently.

11 Interpreting tabular data

Being able to read data effectively is only half the battle - you also want to be able to analyze the data for insights.

12 Gathering insights

Tabular data manipulation and drawing conclusions from data is a crucial component of data science.

13 Answering specific questions

The goal of data science is to identify and answer specific questions.

Using Exploratory Data Analysis

14 Defining exploratory data analysis

Conducting exploratory data analysis (EDA) is the next crucial stage in the data science life cycle.

15 Recognizing statistical data types

Statistical data types, including numerical and categorical data, are at the core of most data science operations.

16 Distinguishing properties of data

EDA involves determining the key properties of the data you have.

Cleaning Your Data

17 Explaining data cleaning

Data cleaning is a crucial stage in the data science life cycle because it ensures you're working with data that is accurate and organized.

18 Questions to guide data cleaning

Before you dive in and start cleaning your data, you'll want to gain some baseline information to help you navigate this process.

Using Data Visualization

19 Demystifying data visualization

Once you determine the granularity, scope, temporality, and faithfulness of your data, you'll want to connect the relationships among your data.

20 Visualizing your qualitative data

Different types of visualization correspond to different types of data.

21 Visualizing your quantitative data

Different types of visualization correspond to different types of data.

Using Inference and Statistical Analysis

22 Defining inference

Inference is also an integral part of the last stage in the data science life cycle - this is when you put the question you want answered to the test.

23 Designing a hypothesis test

Hypothesis testing is a helpful method you can use to identify if the results you're seeing in the data are meaningful.

24 Creating a permutation

Hypothesis testing allows data scientists to make informed conclusions based on the data that they observe.

25 Conducting a permutation test

If you have two or more samples of data, you'll find that using a permutation test will be helpful for you to prove your hypothesis.

26 Bootstrapping a confidence interval

You can use a confidence interval to test your hypothesis, or estimate.

Using Prediction in Data Science

27 Defining prediction for data science

When you make a prediction about your dataset, you can test against it.

28 Navigating classification

Classification is an important machine learning technique you can use when working with data.

29 Recognizing the k-NN algorithm

k-NN or k-Nearest Neighbor is a common Data Science algorithm.

30 Implementing k-Nearest Neighbors

k-Nearest Neighbors is a great algorithm to use, but navigating it with an example can be helpful to fully grasp the concept.

31 Navigating regression

Regression is all about exploring relationships, which is often what evaluating data involves.

32 Checking assumptions of regression

You can use linear regression to help you predict the value of one variable using the value of another.

33 Implementing linear regression

Linear regression uses a dependent and independent variable to help you test and form the relationships within your data.

Conclusion

34 Next Steps

Thanks for watching this course!