

**Coders Lab**  
IT ACADEMY

# DATA SCIENCE

## HANDBOOK



**Gain data science competencies from scratch**

# Learn Data Science from scratch

According to its definition, Data Science is a research area that involves drawing conclusions based on large amounts of data and using a variety of research methods. These conclusions find application in solving multiple specific problems. It is not surprising that the significance and popularity of Data Science is rising in all aspects of our lives.

For the conclusions to be reliable, it is necessary to correctly frame the problem, find the necessary data and process them appropriately. This is why Data Analysts work according to the Data Science process. And that requires a set of competencies.

## Framing the problem

You're going to learn to ask the right questions and transform them into specific, well-defined problems. The skill useful for this involves analysis, basics of math and statistics.

### Relevant skills:

- » analysis
- » statistics
- » math



## Collecting data

With basic programming skills you'll have no problem collecting information from various sources, websites and systems. You will efficiently store the data in useful formats such as: .csv, .json or .xml and get ready to work with it.

### Relevant skills:

- » programming



## Presentation of analysis results

We'll show you how to use the market-leading tools to correctly visualize your analysis results and use data storytelling. This will make your analyses more understandable and let them be the basis for making better decisions.

### Relevant skills:

- » visualization
- » storytelling



## Cleaning and processing data

Knowledge of programming fundamentals will make it easy for you to clean the data from missing, corrupted, or error-prone values, and to process the data. These are indispensable skills for any data analyst working with big data sets.

### Relevant skills:

- » programming



## In-depth data analysis\*

At this stage, understanding of machine learning rules is helpful. Their applications include creating predictive models and in-depth analysis of the data from sources such as databases. This is a higher degree of mastery: one that follows the Data Analyst course.

### Relevant skills:

- » machine learning



## Exploring data and defining new variables

Together with programming languages, the basics of statistics will help you to identify patterns, isolate features and research the data acquired earlier. This will make your work more efficient, and most of all, effective.

### Relevant skills:

- » programming
- » statistics



\*This process stage is a part of the Premium package: the Data Scientist course

# Data Analysts in the job market

Data influences every part of our lives: industry, technology, education, marketing, and sales. As a Data Analyst you're going to use the data to perform complex analyses. The results based on your flawless analysis, and in turn the actions based on them will streamline the activity in your department, the company, and in the future, maybe the entire industry as well. This is how our world has been developing and technology is able to advance.

That's why companies never stop looking for analysts. All this is based on the data from [pracuj.pl](https://pracuj.pl) <sup>1</sup>

## 3 000+

number of monthly job offers for Data Analysts

### How do companies use data analysis?



learning about their customers (e.g. who they are, how they make decisions)



identifying opportunities and threats



detecting financial fraud



forecasting demand for their products



boosting understanding and predicting sales trends.



accurately measuring the results of their business decisions

<sup>1</sup> Data from Pracuj.pl [accessed: 02.2023]

# Who is the Data Analyst course for?



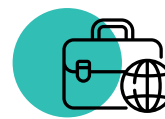
## EVERYONE AIMING TO GET DATA ANALYST JOB

with no prior experience in working with data



## EVERYONE AIMING TO DEVELOP THEIR QUALIFICATIONS

focusing on becoming more competitive in the job market



## BANKING EMPLOYEES

from the industries including:  
finance, sales, marketing,  
controlling, or logistics

Data analyst is a course that develops Data Science competencies. We've tailored it for people who have no, or very limited experience in working with data. **Each module of the course starts from the basics and provides an introduction to working with data** – from framing the problem, to presenting analysis results. Selecting the highest package: Data Scientist course, you'll also gain the knowledge regarding machine learning. All of this with specialist tools that you are going to learn during classes.

## Data Analyst's competences are especially useful in the industries such as:



banking



insurance



telecommunications



media and marketing



e-commerce



retail and services



industry



education



public administration



health service

# Course entry requirements

## To start the Data Analyst course you need to:

- know the basics of Excel (creating documents, basics of data entry and editing);
- have the fundamental understanding of charts and the data they present,
- be open to develop your mathematical, logical thinking and data-based reasoning skills.

# Data Analyst – get a future-proof job

The Data Analyst course covers areas including basics of statistics, data analysis, programming, and data visualization; and in the advanced version, also machine learning. All of this based on our proprietary programme and supervised by knowledgeable experts.

Data Analyst course includes modules covering the following:

## Introduction to Data Analysis

In a practical manner, with no prior experience, you are going to get the knowledge and skills necessary to make the first steps on your data analysis journey. You are going to learn:



the practice of using  
the data analysis pipeline



basic statistical patterns



working with databases and  
basics of data analysis



## Python – Data Analysis

Basics of programming is a key Data Science skill. This will make it easy for you to clean the data from missing, corrupted, or error-prone values. You are going to learn



automate creating  
reports with Python



obtaining data  
from external  
services and websites.



obtaining data from  
database systems (ERP,  
accounting, etc.)

## SQL – Data Analysis

Thanks to in-depth understanding of SQL you will quickly find the specific data in the database and get more time for what's important: careful analysis. You are going to learn



creating efficient  
and complex  
queries in SQL,



additional methods  
of formatting  
the output dataset,



rules  
of working with  
relational databases



## Data Visualization

In the final course you'll master data storytelling and creating clear visuals. Thanks to this, you will understand data better and start drawing better conclusions and making the right decisions. You are going to learn



data storytelling  
and creating  
data-driven narration



using libraries  
and data visualization  
tools



creating ready-to-use  
analytical dashboards

## Machine Learning\*

You'll get the knowledge on how to choose, apply, and verify the correct machine learning model to solve a particular business problem. You are going to learn



NLP – natural language  
processing



advanced machine learning  
models



for regression models



\*The module is available in the Data Scientist Premium Package

# Data Analyst course program



## BLOCK 0: Introduction to Data Analysis

### Introduction to Excel

- Excel installation and configuration
- Discussing Excel functionalities in the data analysis process
- Arithmetic operations with Excel

### Introduction to Data Analysis

- Visualization and data analysis – introduction and inspiration
- Data analysis process
- Processing data
- Data exploration and mathematical tools
- In-depth data analysis and elements of statistics
- Data presentation and analysis
- Sample datasets and their scope

### Data Science process: step 1. Framing the problem

- Framing the problem

### Data Science process: step 2. Collecting data

- Collecting data
- Key performance indicators

### Data Science process: step 3. Cleaning and processing data

- Cleaning data
- Processing data

### Statistics in data analysis

- Basics of descriptive statistics

### Data Science process: step 4. Exploring data

- Exploring data
- Analysis of variables in a data set
- Filtering and sorting data
- Pivot tables

### Data Science process: step 5. In-depth data analysis

- Introduction to in-depth data analysis — machine learning
- Linear regression
- Linear correlation
- Other types of curve fitting

### Homework

- Problem analysis

### Data Science process: step 6. Communicating results

- Data storytelling
- Introduction to data visualization
- Examples of good and bad visualizations
- Types of charts
- Creating visualizations
- Moving visualization to PowerPoint

### Introduction to A/B tests

- Introduction to A/B tests
- Examples of business applications of A/B tests
- First A/B test

## Module exam

### Final project

- Introduction to the final project
- Collecting data
- Cleaning and processing data
- Exploratory data analysis
- In-depth data analysis
- Communicating results
- Presentations of final projects



# BLOCK 1: Python – Data Analysis

## Prework - programming basics

- Introduction to Python
- Data types in Python
- Python standard library

## Python

- Functions
- Lists
- Tuples
- Strings
- Objects
- Files
- Exceptions
- Libraries

## SQL

- Basic operations
- PostgreSQL
- Relationships
- Functions

## JSON and API

- JSON
- API
- Authentication

## Pandas

- Filtering
- Processing data
- Grouping data
- Merge
- Datetime
- Pivot
- OpenPyXL

## Web scraping

- Introduction to HTML
- Web scraping in Python
- BS Element

## Data visualization

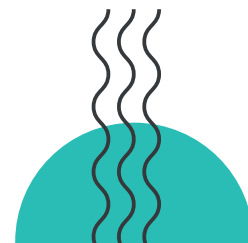
- Data visualization - creating plots

## Generating PDFs

- Creating PDFs
- Creating a document through a story
- Tables

## Module exam

## Final project







## BLOCK 2: SQL – Data Analysis

### Pework - databases

- Types of databases
- UML basics
- Database installation and preparing the work environment
- Getting familiar with a database
- Viewing data (SQL basics)
- Basics of AND, OR operator logic
- History of databases

### SQL basics

- SQL basics
- Basics of Boolean logic and algebra
- Additional clauses
- Operations on sets
- Subqueries
- Tables
- Manipulating database records

### Relationships

- Relationships and their types
- Joins and their types
- Joins in use
- Delete cascade
- Indexes
- Other types of joins, good practices for joins and subqueries

### Data analysis

- Data analysis
- View
- Grouping data
- Additional grouping functions
- Window functions
- Working with datetime
- Order of operations in SQL
- Rollup, query plan.

### Procedural SQL

- Procedural SQL
- Introduction to procedural SQL
- Flow control
- Transactions
- SQL built-in functions
- Loops
- Cursors
- JSON
- Data normalization
- Notebook for SQL



### Module exam

### Final project







## BLOCK 3: Data Visualization

### Prework – Introduction to data visualization

- Introduction and preparation
- Introduction to data visualization
- Visual encoding
- Color
- Datawrapper
- Data types and operations

### Seaborn/Plotly

- Theory of data visualization
- Color maps
- Seaborn
- Plotly
- Basics of visualization in specific libraries
- Plotly charts in Google Slides
- Bokeh basics

### Maps

- Projections, coordinates, cartograms
- Introduction to libraries
- Geopandas, Geometry, GeoSeries, GeoDataFrame osm
- Geojson, Shapefile
- Mapbox

### Dash, graphs

- Networkx and graphs
- What is a dashboard?
- Introduction to dash
- Dash components
- Dash datatable
- Dash callback and app lifecycle
- Editable data table
- Interacting with a dashboard
- Dash Cytoscape

### Dashboards, Data storytelling

- Data storytelling
- Dashboard
- Tooltip
- Multipage app
- Bootstrap styles
- Dash and SQL
- Dash and Maps

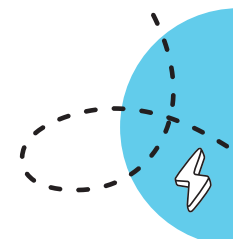
### Tableau

- What is Tableau?
- Preparing the work environment
- First project
- Creating a dashboard
- Use scenarios
- Integrating Tableau and Mapbox

### Module exam

### Final project

### Data Analyst course final exam





## BLOCK 4: Machine Learning (available in the Premium Package)

### Introduction to Machine Learning

- Course introduction
- Environment configuration and requirements
- Python revision (numpy, pandas, visualization libraries)
- Statistics in machine learning
- Introduction to machine learning
- How to correctly prepare data for the model?

### Regression

- Regression
- Linear regression
- Regularization in linear regression model
- Polynomial regression
- Regression problem using a decision tree
- How to determine the quality of a regression model? Regression models metrics

### Classification

- K-nearest neighbors
- Logistic regression
- Classification problem using a decision tree
- SVM (Support Vector Machine)
- How to determine the quality of a classification model? Classification models metrics

### Homework

- How to take your notebook from Colab to Kaggle?
- Linear regression with regularization
- Decision tree (regression)
- Logistic regression
- SVM

### Basics of neural networks and NLP

- Basics of neural networks
- Basics of NLP

### Advanced machine learning models

- Random forests
- Boosting and bagging
- Neural networks
- Day 3 summary

### NLP – natural language processing

- How to work with textual data
- Bag of words
- TF-IDF
- Word2Vec
- BERT

### Homework

- How to improve the model performance: hyperparameter optimization
- Random forests
- XGBoost
- Neural networks

### Unsupervised learning

- Unsupervised learning
- Dimensionality reduction
- Clustering
- Anomaly detection
- Day 5 summary

### Final project, exam

- Final project
- Final results presentation
- Exam
- Course summary

### Machine Learning: Module exam

### Final project

### Data Scientist course final exam

# Lecturers and mentors

You are never alone during a Coders Lab course. Our hand-picked lecturers and mentors are with you. You can judge us by the effects of their work. That's why we selected experienced data analysts who will teach you their expertise and support you throughout the course.



## Who is a lecturer?

During your live Virtual Classroom sessions you'll be constantly in touch with your lecturer. Depending on the length and scope of the course, different subject-matter experts may take up the role of the lecturer.

### Lecturer's tasks include:



teaching practical  
classes



providing support  
and answering in-class  
questions



monitoring learning  
progress  
during the course



## Who is a mentor?

Mentor is a content specialist in your course: the same from start to finish. You can consult them on all of the course content on Slack. A mentor can also be the lecturer of the whole or a part of the course.

### Mentor's tasks include:



introducing you  
to the profile  
and plan of the course



supporting you during  
Pework and answering  
your questions on Slack



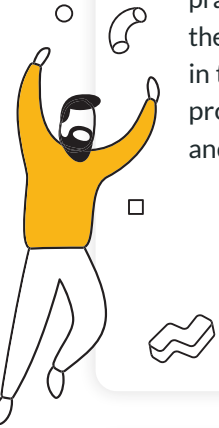
supporting you  
throughout the course  
and looking after your  
progress

# Why is the Coders Lab course worth it?

1

## We prioritize practice

In our courses, from the very beginning you gain practical skills. The program is designed to cover the skill in demand on the job market. Moreover, in the course you are going to create your own project that will be the basis of your portfolio and your job-market showcase.



2

## We are supporting you every step of the way.

Career advisors who walk you through the process of choosing and buying the course is just a beginning. Next, you will get support from lecturers and mentors. They will introduce you to the world of programming, software testing or data analysis. Depending on the package you choose, you can also count on our help with your CV, portfolio and preparing for job interviews.

3

## You learn in a group

Every course participant brings something to the table. Personal experience, knowledge or a unique perspective on the subject. Thanks to this every workshop is different but always enhancing. It is also a way to forge new relationships and support in the job market later.

4

## We know the requirements of the job market.

We carefully track the changes in the market and analyze the needs of potential employers. We see who they look for, what technologies they use, and what they do every day. That's how we know what skills to prioritize in our classes.

5

## You gain the most in-demand competencies

Employees dealing with a broad scope of data analysis are getting recruited by world-class companies such as: IBM, Amazon, Microsoft, Facebook, Google, and Apple. Demand for data analysis specialists is growing. That's why if you develop your Big Data competencies you'll be matching the job market needs.

6

## Online sessions are live

In our courses classes are held in Virtual Classrooms on a group communication platform (Zoom). Thanks to it, the lecturer and other course participants are right there with you. And you can learn from anywhere in the world.

