

LISBON
DATASCIENCE
ACADEMY

Introduction Session

Batch #7 - 2023/24

18. 11. 2023, 10 am GMT

“

The **Lisbon Data Science Academy**
exists to create and maintain a
community of excellence, with the
teaching and learning of Data
Science at its core.

”

The Lisbon Data Science Academy [mission](#)

People behind the scenes of this batch:

QA - ensure that everything works:

Ahmed Emad, Beatriz Guerner, Caitlin Hulse, Cora Dutsch, Filipa Pereira, Mohamed Gaber, Rita Carvalho

Senior instructors - specialization leads, virtual classes, AMA sessions:

Bruno Batista, Gustavo Fonseca, Maria Dominguez, João Ascensão, José Rebelo, Roberto Álvarez, Rui Azevedo, Sam Hopkins, Telmo Felgueira

Unit instructors - unit development:

Carlos Matrena, David Belo, Fábio Cruz, Francisco Costa, Gustavo Fonseca, João Gomes, João Nobre, João von Gilsa, Kim Pronk, Mafalda Vaz da Silva, Margarida Antunes, Mária Hanulová, Rodrigo Veríssimo, Rúben Belchior

Supporting roles - community, marketing, partnerships, ... :

Carla Cotas, Minh Hoang, Pedro Ved, Pratiksha Jain, Vasco Mano

Partner companies:

Daredata, LOKA, LTPlabs, Springer



The starters academy targets those who want to become entry-level data scientists no matter their background.



Helped **1 000+** students



800+ candidates



250 actually become DS students



120+ graduated in Data Science



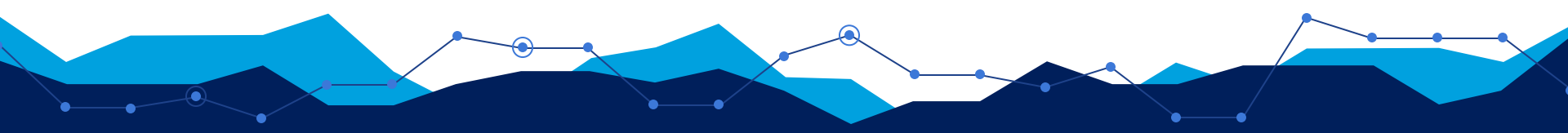
linkedin

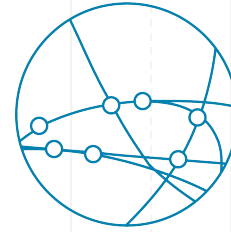
3k followers



Instagram

1k followers





LISBON
DATASCIENCE
ACADEMY

Curriculum Overview

The Lisbon Data Science Starters Academy Curriculum

	Subject	Duration	Learning units	Hackathon
Bootcamp	Virtual classes	2 half-days	19	
Specializations	S01 Binary Classification	3 weeks		x
	S02 Data Wrangling	4 weeks	3	x
	S03 Time Series	4 weeks	3	x
	S04 Text Classification	4 weeks	3	x
	S05 Recommender Systems	optional	3	
	S06 Data Science in the Real World	4 weeks	3	x
Capstone	Real world example	11 weeks		

8 months

34 learning units

5 specializations
(+ 1 optional)

5 AMA sessions

5 hackathons

1 capstone

Timeline: 19 Nov 2023 - 14 Jul 2024

Graduates will be announced on 22 Jul 2024

Binary Classification	Data Wrangling	Time Series	NLP	Recommender Systems	Data Science in the Real World	Capstone
Nov/Dec	Jan	Feb	Mar	Optional	April	May - Jul
19 SLUs	3 BLUs	3 BLUs	3 BLUs	3 BLUs	3 BLUs	1 Proposal 1 Report
Bootcamp 1 26 Nov 2023	Hackathon #1 17 Dec 2023	Hackathon #2 4 Feb 2024	Hackathon #3 3 Mar 2024	Hackathon #4 7 April 2024		Hackathon #6 5 May 2024

Bootcamp 2
3 Dec 2023

LDSSA learning material

individual study

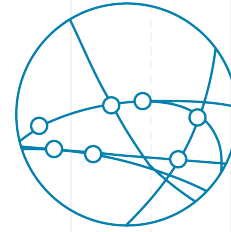


- **SLU - small learning unit**
 - 1 learning notebook
 - 1 exercise notebook
- **BLU - big learning unit**
 - 3 learning notebooks
 - 1 exercise notebook
- **Capstone**

group or interactive learning



- **hackathon**
 - group problem solving
- **virtual classes**
 - only in specialization 1
- **AMA sessions**
 - meet an experienced data scientist



LISBON
DATASCIENCE
ACADEMY

Specialization 1

Specialization 1 - basics of data science in 19 SLUs

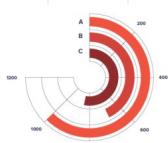
Part 1: from data exploration and cleaning to first models



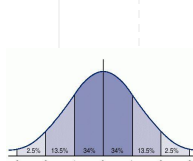
SLU 1 - Pandas



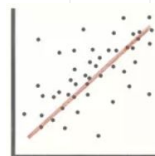
SLU 2 - Subsetting with Pandas



SLU 3 - Visualizing Data



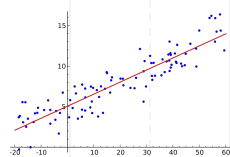
SLU 4 - Basic Statistics



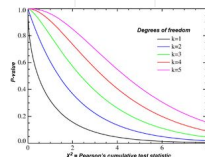
SLU 5 - Covariance & Correlation



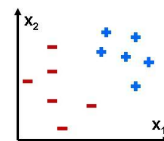
SLU 6 - Data Problems



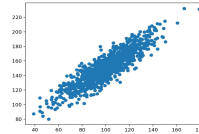
SLU 7 - Regression



SLU 8 - Metrics for Regression



SLU 9 - Classification with Logistic Regression



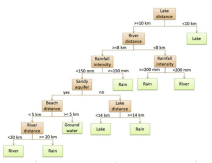
SLU 10 - Metrics for Classification

SLU 1-3: done during the admissions

SLU 32: hackathon training unit part 1, optional

Specialization 1 - basics of data science in 19 SLUs

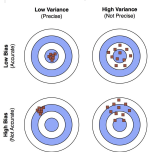
Part 2: more models, model optimization, ethics



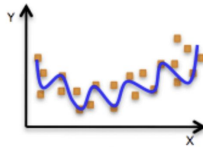
SLU 11 – Tree-Based Models



SLU 12 – Feature Engineering




SLU 13 – Bias-Variance Tradeoff & Model Selection



SLU 14 – Model Complexity & Overfitting



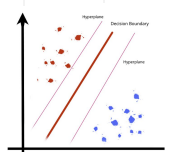
SLU 15 – Hyperparameter Tuning



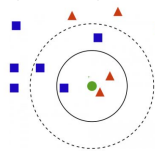
SLU 16 – Workflow



SLU 17 – Ethics & Fairness



SLU 18 – Support Vector Machines (SVM)



SLU 19 – k-Nearest Neighbors (kNN)



Mandatory



Optional

SLU 18 - 19: optional

SLU 64: hackathon training unit part 2, optional

Specialization 1 timeline

Setup week: 19 - 25 Nov 2023

- release of the learning notebooks SLU04 - 10
- prepare your computer

Week 0: 19 - 25 Nov 2023

Bootcamp part 1: 26 Nov 2023

- 3 virtual classes for SLU04 - 10
- release of the rest of SLUs

Week 1: 26 Nov - 2 Dec 2023

Bootcamp part 2: 3 Dec 2023

- 3 virtual classes for SLU11 - 17

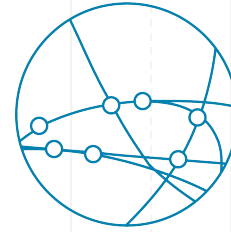
Week 2: 3 - 9 Dec 2023

Delivery deadline: 16 Dec 2023 8 pm GMT

Week 3: 10 - 17 Dec 2023

Hackathon #1: 17 Dec 2023





LISBON
DATASCIENCE
ACADEMY

Specializations 2 - 6

Specializations 2 - 6: exploring the realms of data science



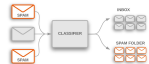
Specialization 2 Data Wrangling

BLU01 Messy Data
BLU02 Advanced wrangling
BLU03 Data Sources



Specialization 3 Time Series

BLU04 Time Series Concepts
BLU05 Classical Time Series Models
BLU06 Machine Learning for Time Series



Specialization 4 Text Classification

BLU07 Feature Extraction
BLU08 Dimensionality Reduction
BLU09 Information Extraction



Specialization 5 – optional Recommender Systems

BLU10 Non-Personalised recommender systems
BLU11 Personalised recommenders
BLU12 WorkFlow



Specialization 6 Data Science in the real world

BLU04 Basic Model Deployment
BLU05 Deployment in the real world
BLU06 Model CSI

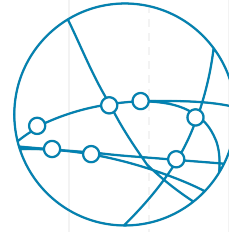
Specializations 2 - 6 timeline

- 4 weeks
- LUs released on Mondays

example for specialization 2:

Release of BLU01	Week 4: 8 - 14 Jan 2024
Release of BLU02	Week 5: 15 - 21 Jan 2024
Release of BLU03	Week 6: 22 - 28 Jan 2024
Delivery deadline: 3 Feb 2024, 8 pm GMT	Week 7: 29 Jan - 3 Feb 2024
Hackathon #2: 4 Feb 2024	





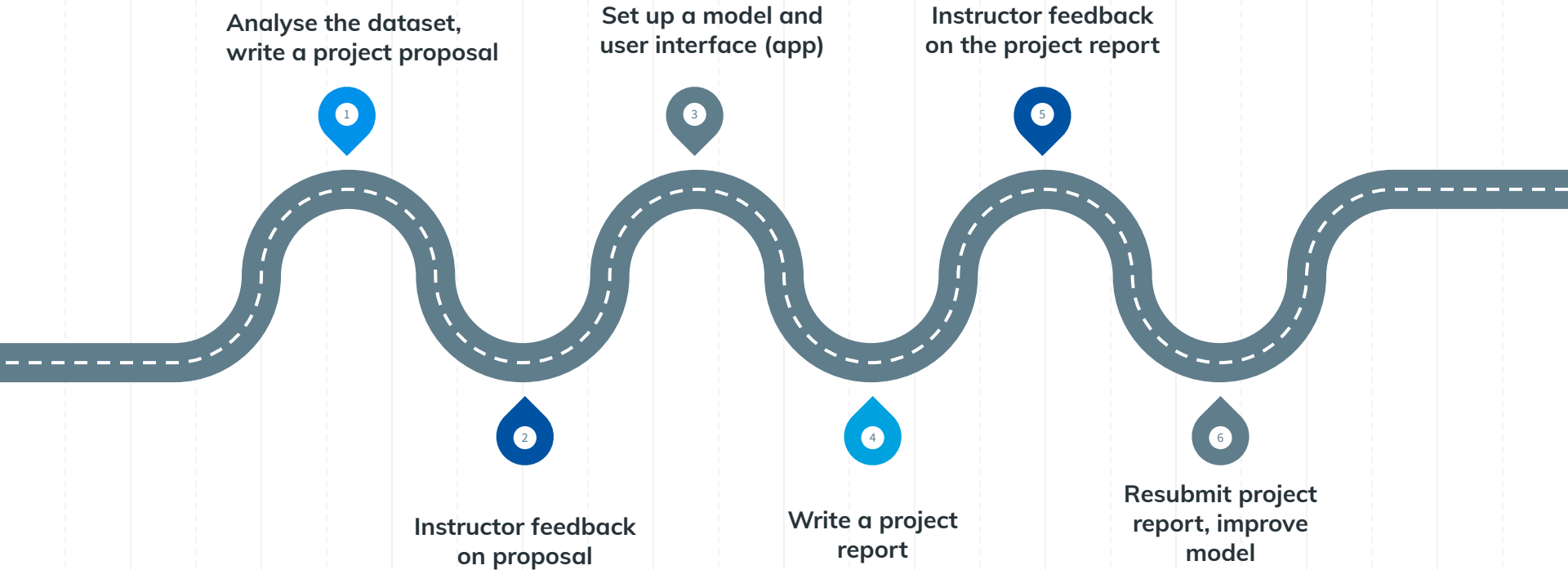
LISBON
DATASCIENCE
ACADEMY

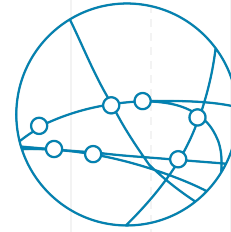
Capstone

The Capstone - a real life simulation

- **you are a data scientist**
- **you get an email and a dataset from a client**
- **you need to**
 - **define the project -> write a proposal**
 - **to analyze the dataset**
 - **set up and optimize a model**
 - **set up a user interface**
 - **write a project report**

The Capstone - a real life simulation





LISBON
DATASCIENCE
ACADEMY

Graduation Criteria

Graduation criteria:

Specialization 1:

- Participate in bootcamp presentations
- Deliver SLU04 - SLU17 with at least 16/20
- SLU18 and SLU19 are optional
- Participate in the hackathon #1

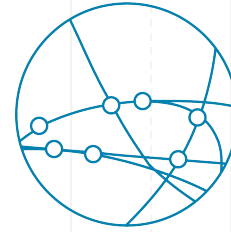
Specializations 2, 3, 4, 6:

- Deliver all BLUs with at least 16/20
- Participate in 3 out of 4 remaining hackathons

Capstone

- Deliver the project proposal and report
- Deliver a functional API





LISBON
DATASCIENCE
ACADEMY

Time Commitment

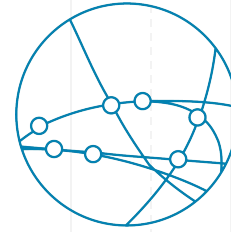
Time commitment

- **5 - 10 hours a week** on average

It can be more in the busiest times during

- **Specialization 1**
- **Capstone**





LISBON
DATASCIENCE
ACADEMY

How to ask for help

Situation: *You're doing the learning units at home and you have a question.*

What most students will want to do: ask the instructor of the learning unit via a direct message on Slack

What students should do:

- I. Check if the answer is in the learning materials.
- II. *StackoverGoogle* (especially if it is a Python question)
- III. Check if someone asked the same question in the appropriate slack channel
- IV. Ask the question in the appropriate slack channel

Advantage of asking in public channels vs. asking by DM:

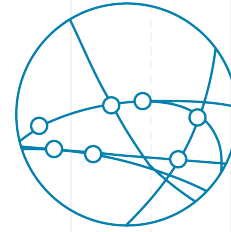
- more people will see your message
- higher chances of getting a good and quick answer

Do not share solutions or code on Slack!

- when asking a question, describe your problem
- when answering a question to help a colleague, give a hint

If everything else fails or you're shy and would rather talk to an instructor, please go ahead!





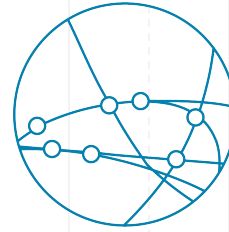
LISBON
DATASCIENCE
ACADEMY

Code of Conduct

- 1. Academic integrity is expected of everyone**
- 2. No one shall provide answers to assignments, unless they have been officially provided**
- 3. No discrimination, harassment or bullying**
- 4. Be respectful and follow professional conduct**



<https://ldssa.github.io/wiki/About%20us/Code-of-Conduct/>



LISBON
DATASCIENCE
ACADEMY

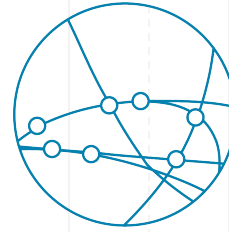
Refund Policy

- **You are eligible for a full refund within 2 weeks after the start date (until 4 Dec 2023), no questions asked**
- **Why 2 weeks? It's enough time to get a good view of our teaching style**
- **That being said, feel free to reach out to any instructor whenever you feel like something is bothering you**



[https://ldssa.github.io/wiki/Starters%20Academy%20\(LDSSA\)/07-Refund-Policy/](https://ldssa.github.io/wiki/Starters%20Academy%20(LDSSA)/07-Refund-Policy/)





LISBON
DATASCIENCE
ACADEMY

Important places



Important places to be

- <https://ldssa.github.io/wiki/> - LDSA wiki, the source of (most) truth
- [LDSSA batch7 Slack](#) - our communication tool
- batch7 calendar - link shared in slack
- <https://github.com/LDSSA/batch7-students> - GitHub repository where the material will be released
- <https://portal.lisbondatascience.org/> - portal for unit submission

Question time