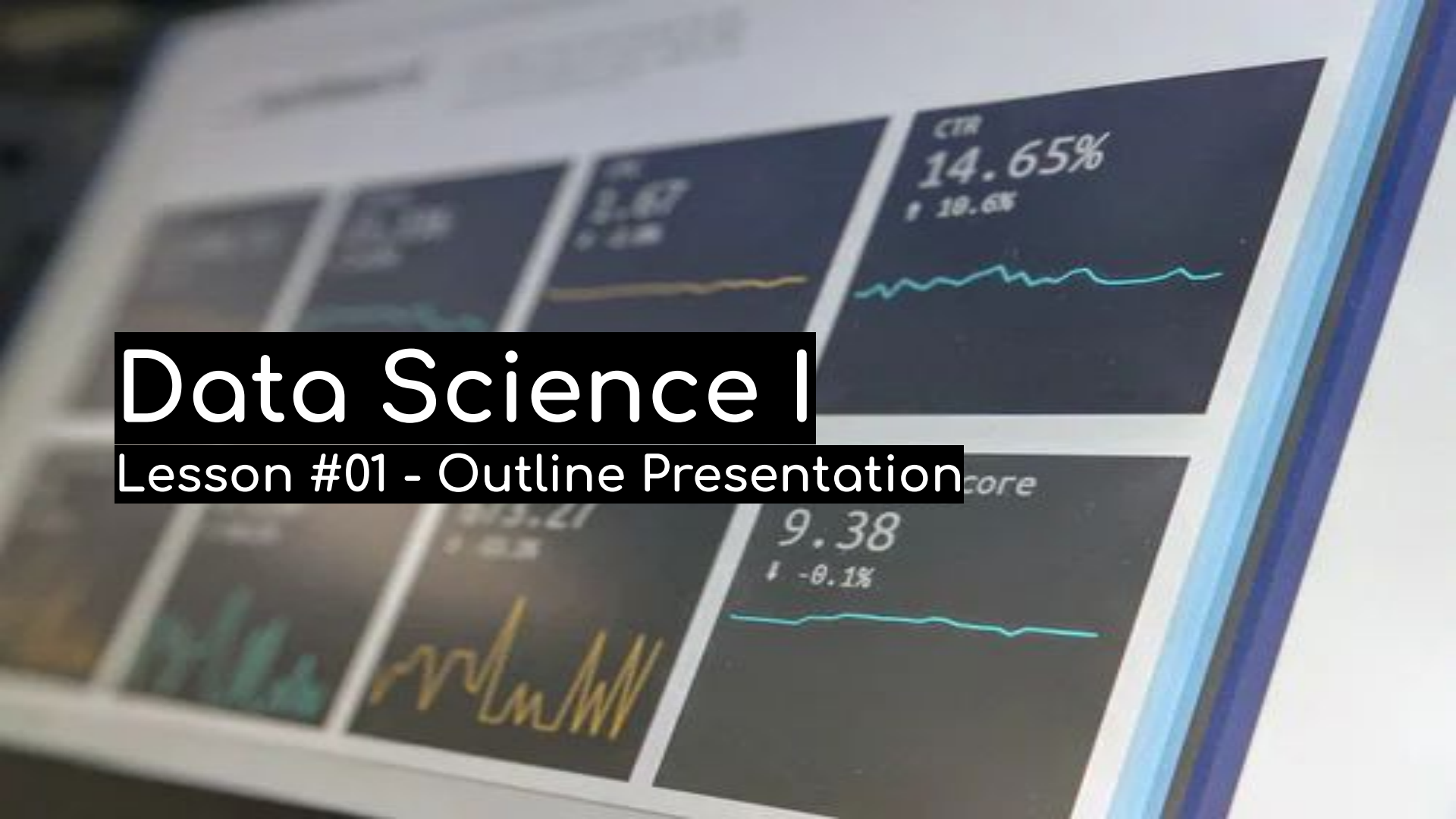
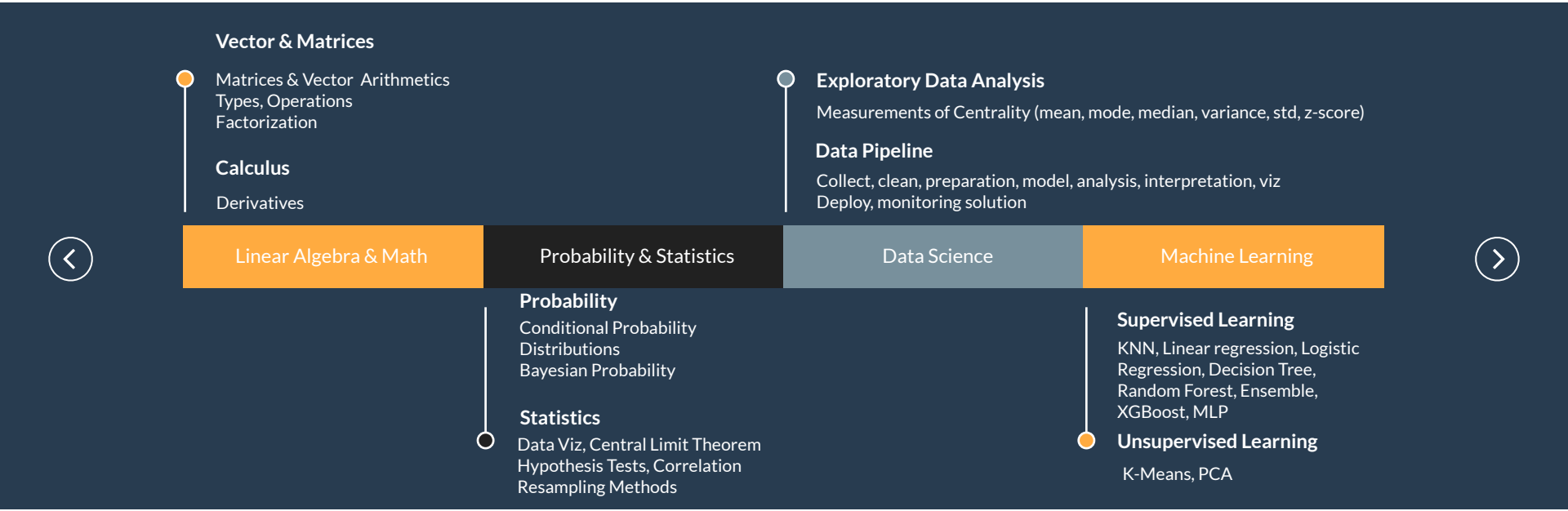


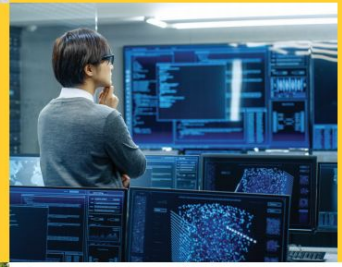
Data Science I

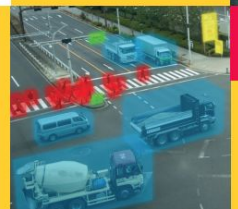
Lesson #01 - Outline Presentation

Your Pathway



AI Career Pathways



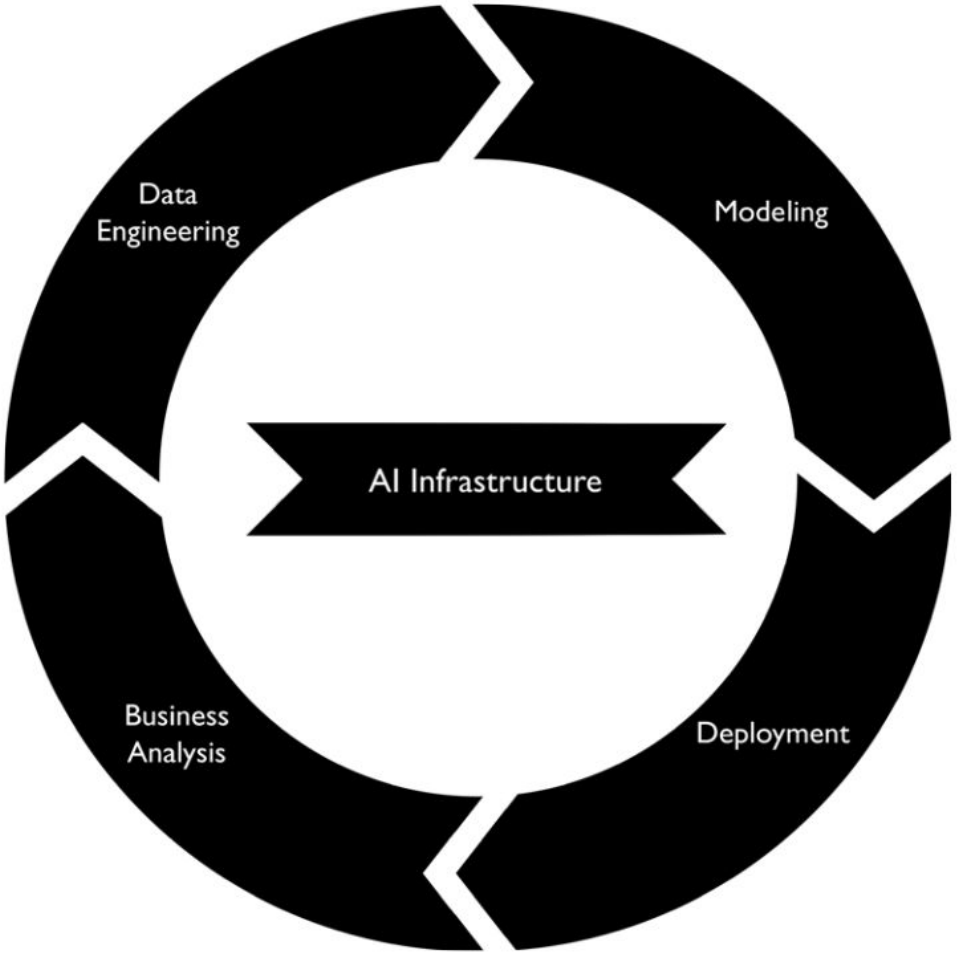


Put yourself (data scientist, machine learning engineers, software engineers, ...) on the right track. [Ref.]

WORKERA

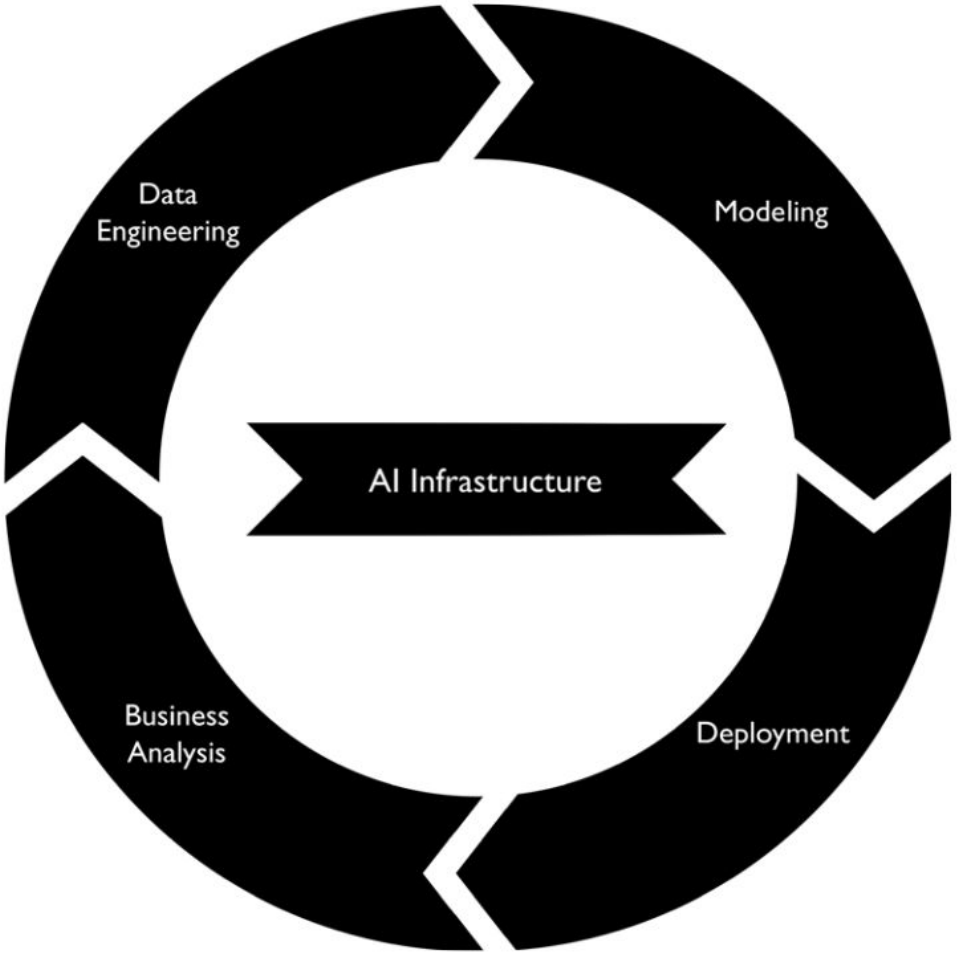
a deeplearning.ai company

Developing an AI project development life cycle involves five distinct tasks

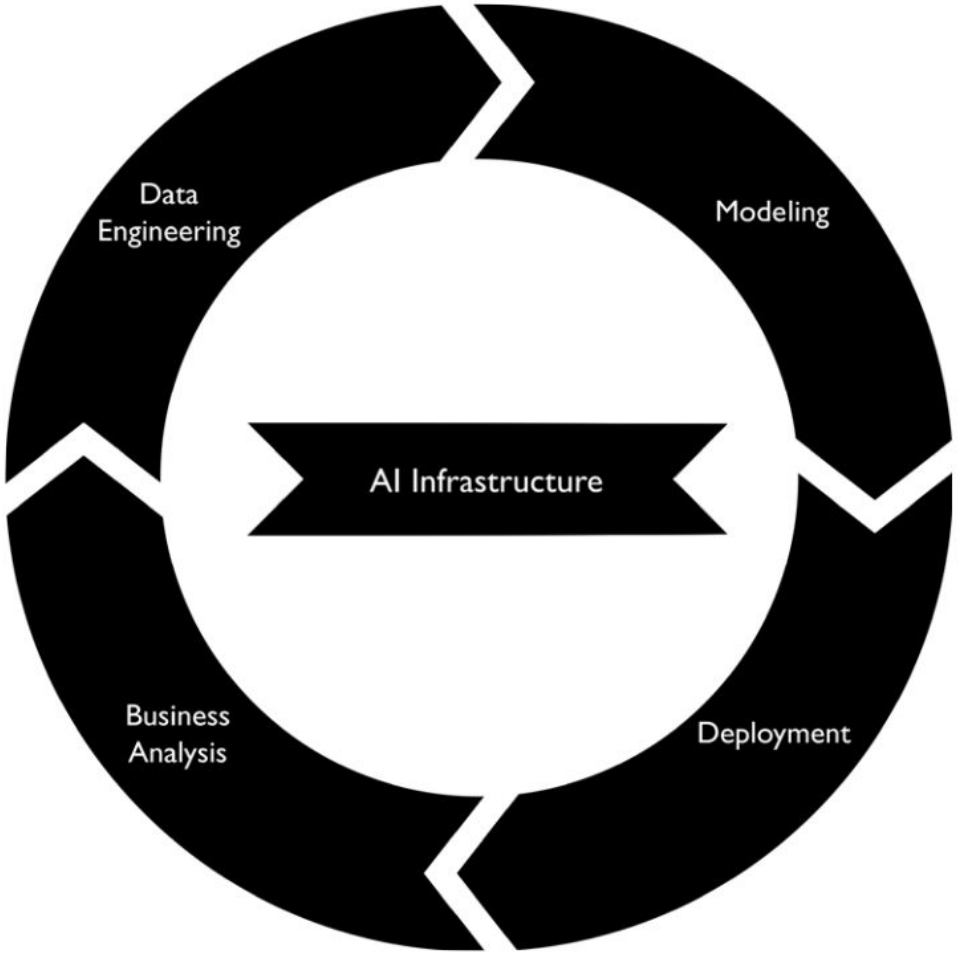


- **Data engineering:** People responsible for data engineering prepare data and transform data into formats that other team members can use.

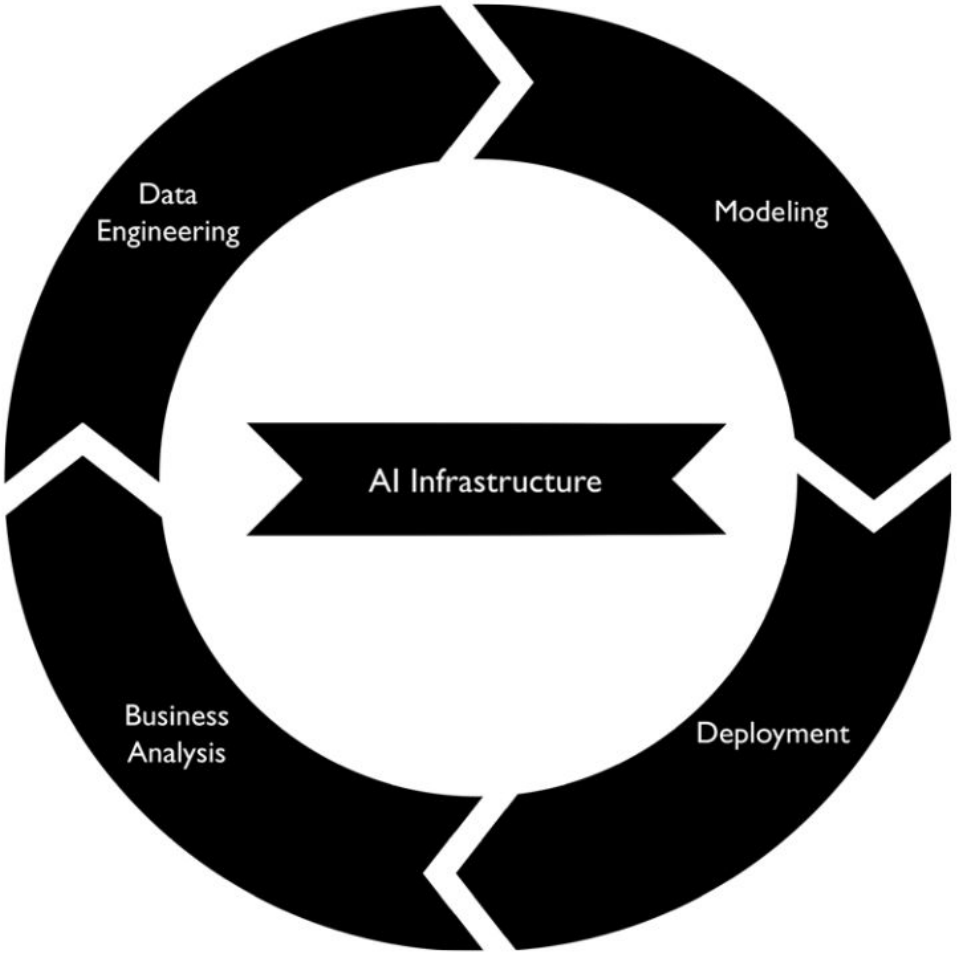
- **Modeling:** People assigned to modeling look for patterns in data that can help a company predict outcomes of various decisions, identify business risks and opportunities, or determine cause-and-effect relationships.



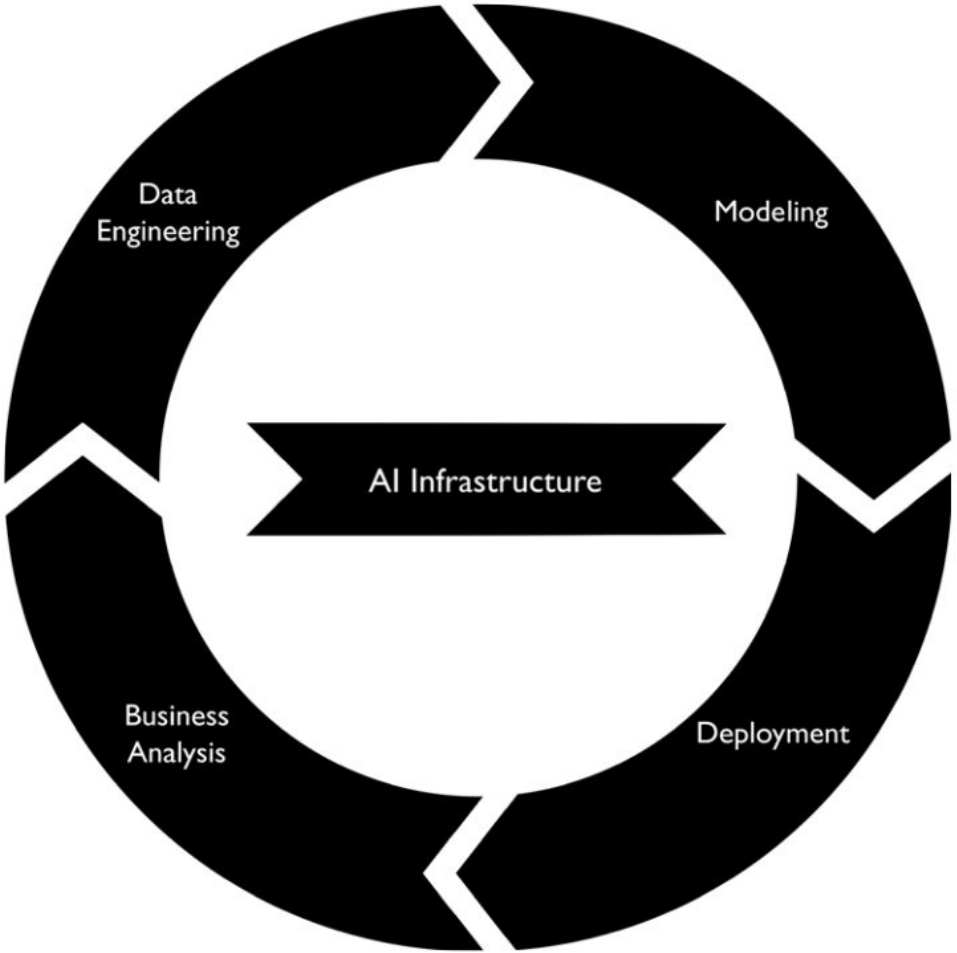
- **Deployment:** People in charge of deployment take a stream of data, combine it with a model, and test the integration before putting the model into production.

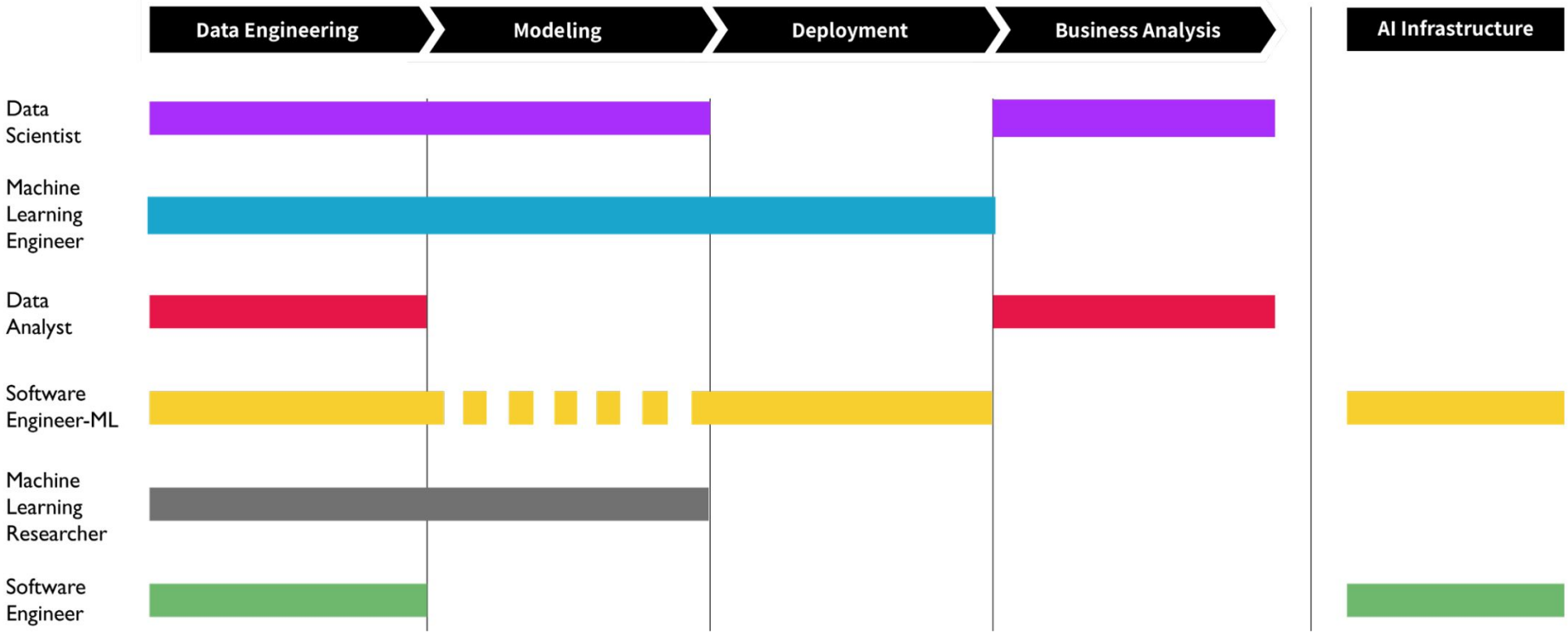


- **Business analysis:** Team members responsible for business analysis evaluate a deployed model's performance and business value and adjust accordingly to maximize benefit or abandon unproductive models



- **AI infrastructure:** People who work in AI infrastructure build and maintain reliable, fast, secure, and scalable software systems to help people working in data engineering, modeling, deployment and business analysis.




Data Engineering
Modeling
Deployment
Business Analysis
AI Infrastructure
Data Scientist
Machine Learning Engineer
Data Analyst
Software Engineer-ML
Machine Learning Researcher
Software Engineer

Defining the Skills

Mathematics
People with mathematics skills demonstrate the ability to solve problems using linear algebra (for instance, matrix vector operations, eigenvalues, eigenvectors, and combinatorics), calculus (derivatives, integrals, and so on) and mathematical functions (simple functions, min/max/argmin/argmax, and so on).

Algorithmic coding
People with algorithmic coding skills demonstrate the ability to understand algorithms written with code, implement classic algorithms like sorting and search, and use classic data structures like trees, dictionaries and arrays.

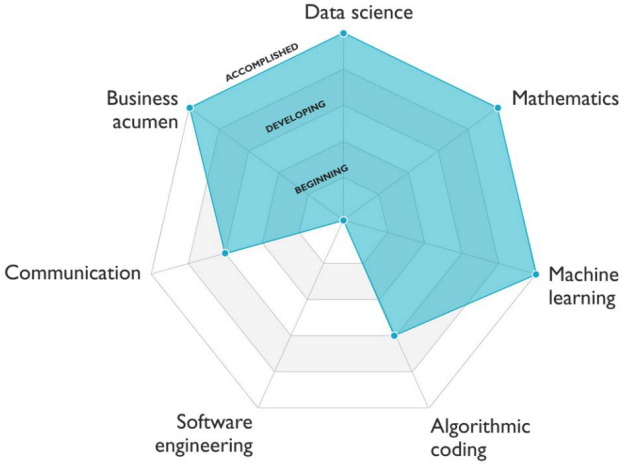
Software engineering
People with software engineering skills demonstrate the ability to use a variety of computer science and software methods such as object-oriented programming, internet protocols, HTTP requests, agile/scrum methodologies, databases, version control (such as Git), containers, and unit testing.

Machine learning
People with machine learning skills demonstrate the ability to use classic machine learning models (for example, PCA, K-means, K-NNs, SVM, Logistic Regression, Linear Regression, and Decision Tree learning), methods to train them (such as initialization, optimization, regularization, and hyperparameter tuning), and techniques to strategize machine learning projects.

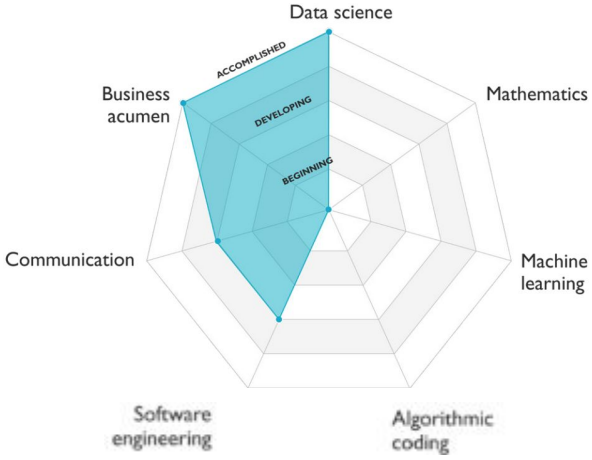
Deep learning
People with deep learning skills demonstrate the ability to use classic deep learning models (such as fully connected networks, convolutional neural networks, recurrent neural networks, and layers), methods to train them (such as initialization, regularization, optimization, and transfer learning), and techniques to strategize deep learning projects.

Data science
People with data science skills demonstrate the ability to use probabilities (including distributions, conditional probabilities, independence, Bayes theorem, etc.), statistics (including hypothesis testing, bias/variance tradeoffs, mean, variance, and mode) and data analysis (including preprocessing, visualization and metrics such as accuracy, R-squared, residuals, precision, and recall).

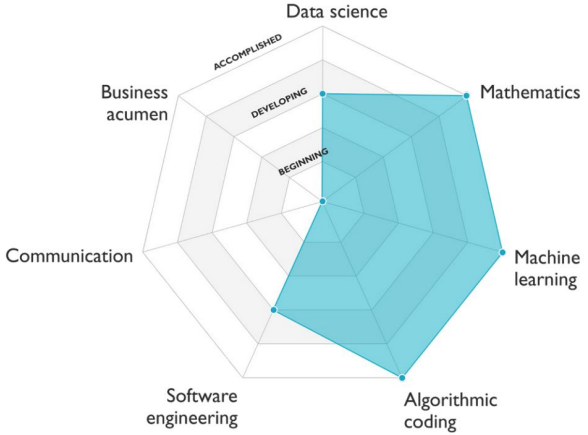
Data Scientist



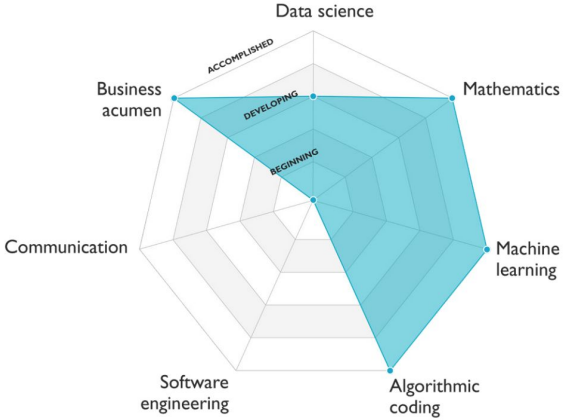
Data Analyst



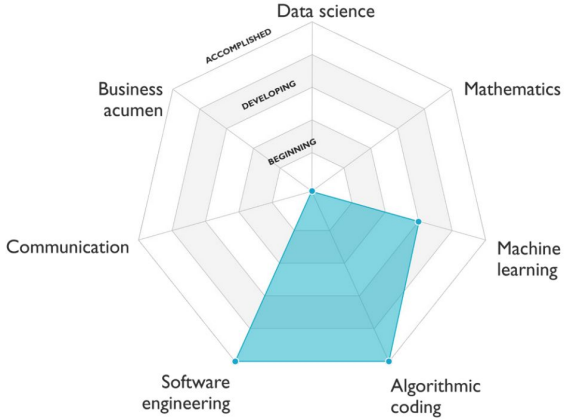
Machine Learning Engineer



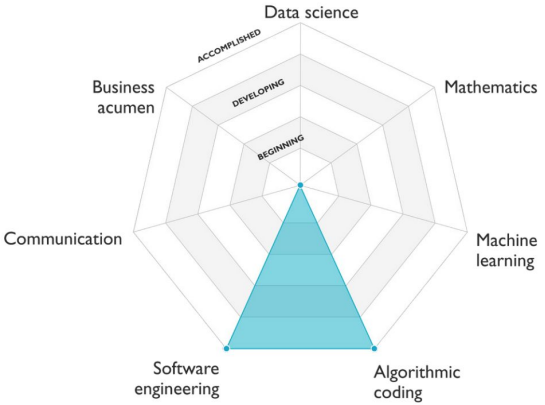
Machine Learning Researcher

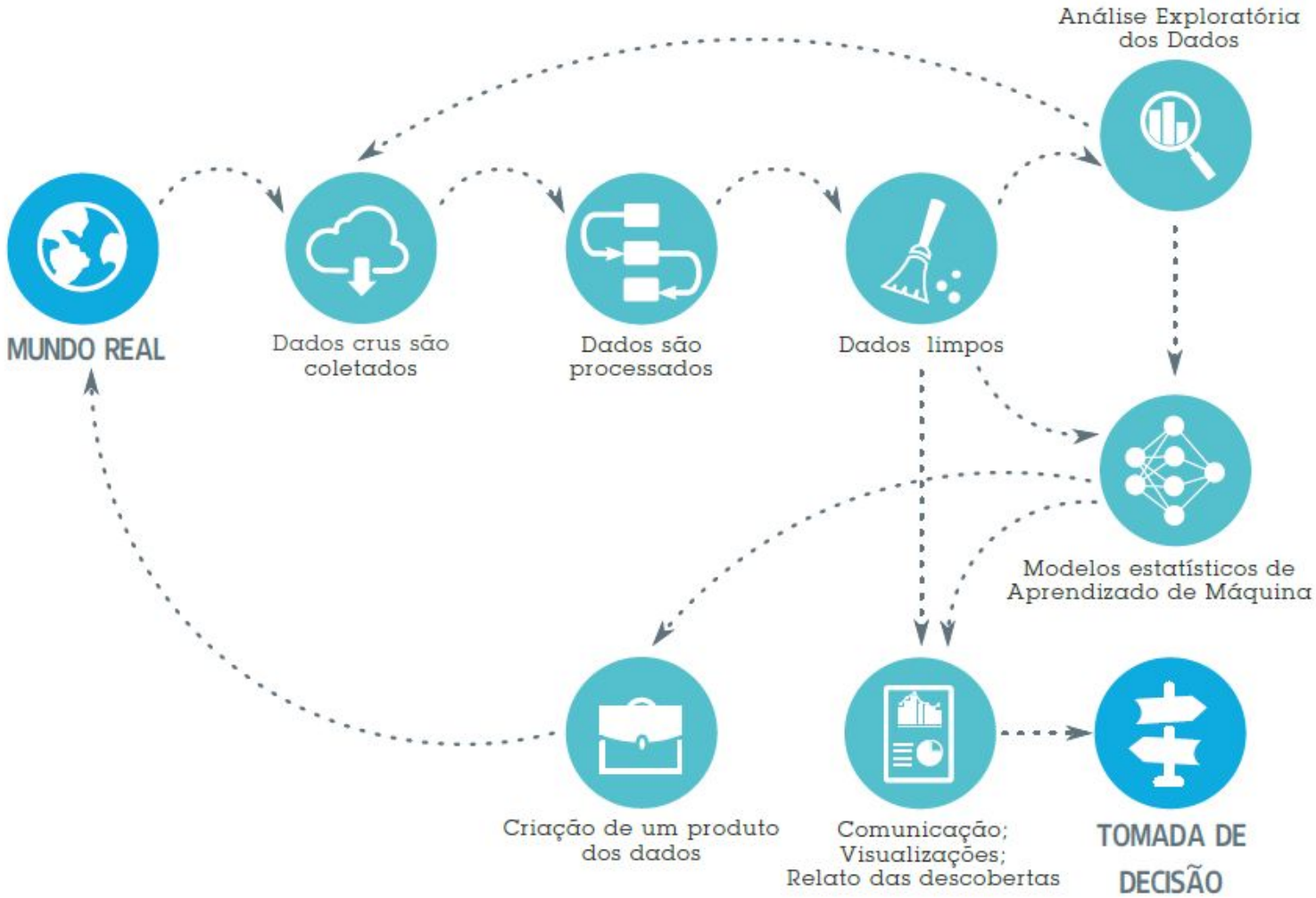


Software Engineer- Machine Learning



Software Engineer





Gisliany Alves et al. Uma abordagem orientada a dados para a criação de um indicador de habitabilidade baseado na API da UBER

Calendar

Unit 01

Início	Fim	Descrição
18/02/2020	18/02/2020	Course Outline
20/02/2020	20/02/2020	IMD Talk
27/02/2020	27/02/2020	Fundamentals of Data Science
03/03/2020	03/03/2020	Introduction to Pandas
05/03/2020	05/03/2020	Exploring data with Pandas
10/03/2020	10/03/2020	Data Cleaning Basics
12/03/2020	12/03/2020	<<Project 01>>
17/03/2020	17/03/2020	Data Aggregation
19/03/2020	19/03/2020	Combinning Data with Pandas
24/03/2020	24/03/2020	Transforming Data with Pandas
26/03/2020	26/03/2020	Working with String in Pandas
31/03/2020	31/03/2020	Regular Expression
02/04/2020	02/04/2020	Working with missing and duplicate data
07/04/2020	07/04/2020	<< Project 02>> <<end of unit 01>>

Syllabus:

1. Pandas Fundamentals
2. Data Cleaning & Analysis

Grades

Notebooks - 10%
Project 01 - 30%
Project 02 - 40%
Datacamp (20k XP) - 20%

Calendar

Unit 02

14/04/2020	14/04/2020		Exploratory Data Analysis
16/04/2020	16/04/2020		Exploratory Data Analysis Cont.
23/04/2020	23/04/2020		Exploratory Data Analysis Cont.
28/04/2020	28/04/2020		Exploratory Data Analysis Cont.
30/04/2020	30/04/2020		<< Project 03>>
05/05/2020	05/05/2020		<< Project 03>>
07/05/2020	07/05/2020		Exploratory Spatial Data Analysis
12/05/2020	12/05/2020		Exploratory Spatial Data Analysis Cont.
14/05/2020	14/05/2020		Exploratory Spatial Data Analysis Cont.
19/05/2020	19/05/2020		<< Project 04>>
21/05/2020	21/05/2020		<< Project 04>> <<end of unit 02

Syllabus:

1. Exploratory Data Analysis
2. Exploratory Spatial Data Analysis

Grades

Notebooks - 10%
Project 03 - 40%
Project 04 - 40%
Datacamp (10k XP) - 10%

Calendar

Unit 03

26/05/2020	26/05/2020		End-To-End Project for Data Science
28/05/2020	28/05/2020		End-To-End Project for Data Science
02/06/2020	02/06/2020		<<metroind conference>>
04/06/2020	04/06/2020		<<metroind conference>>
09/06/2020	09/06/2020		Pitch Ideas (ignite talk)
16/06/2020	16/06/2020		Final Project
18/06/2020	18/06/2020		Final Project
23/06/2020	23/06/2020		Final Project
25/06/2020	25/06/2020		Final Project
30/06/2020	30/06/2020		Final Project
02/07/2020	02/07/2020		Conference

Syllabus:

1. End-to-End Project
2. Final Project
3. Conference << poster >>

Grades

Pitch Idea (10%)
Final Project (90%)

References

References

https://arxiv.org/pdf/2002.04803.pdf

Article

Machine Learning in Python: Main developments and technology trends in data science, machine learning, and artificial intelligence

Sebastian Raschka [1,*†], Joshua Patterson [2], and Corey Nolet [3,4]

[1] University of Wisconsin-Madison, Department of Statistics; sraschka@wisc.edu
[2] NVIDIA; joshuap@nvidia.com
[3] NVIDIA; cnolet@nvidia.com
[4] University of Maryland, Baltimore County, Dep. of Comp Science & Electrical Engineering; coreyn1@umbc.edu
* Correspondence: sraschka@wisc.edu
† Current address: 1300 University Ave, Medical Sciences Building, Madison, WI 53706, USA

Received: date; Accepted: date; Published: date

Abstract: Smarter applications are making better use of the insights gleaned from data, having an impact on every industry and research discipline. At the core of this revolution lies the tools and the methods that are driving it, from processing the massive piles of data generated each day to learning from and taking useful action. Deep neural networks, along with advancements in classical ML and scalable general-purpose GPU computing, have become critical components of artificial intelligence, enabling many of these astounding breakthroughs and lowering the barrier to adoption. Python continues to be the most preferred language for scientific computing, data science, and machine learning, boosting both performance and productivity by enabling the use of low-level libraries and clean high-level APIs. This survey offers insight into the field of machine learning with Python, taking a tour through important topics to identify some of the core hardware and software paradigms that have enabled it. We cover

CAREER COLUMN · 08 FEBRUARY 2019

How to use Twitter to further your research career

The social-media platform is often a tool for procrastination, says Jet-Sing M. Lee. But what else can it be?

Jet-Sing M. Lee

Andrew Trask @iamtrask · 47m

Googling Tip: when you want to learn some ML concept, google "python <concept name> from scratch".

This will often lead to a blogpost with:

- intuitive explanation
- a toy implementation
- reference to more formal papers

which you can unpack in that order :)

#100DaysOfMLCode

3 21 182

#100DaysOfDataScienceCode
#ufrn
#imd
#brazil