

The background is a dark navy blue. It features abstract geometric line art in teal and purple. In the top-left corner, there are several interconnected lines forming a series of triangles and polygons. In the bottom-right corner, there are more geometric shapes, including a large triangle and some smaller polygons, also formed by teal and purple lines.

# AceLeraDev

# Data Science

1 - Ementa do programa

# Ementa geral do programa

/

- **Semana 1: Introdução a Ciência de Dados**
- Semana 2: Pré-processamento de Dados em Python
- Semana 3: Análise de dados exploratória
- Semana 4: Continuação Análise de dados exploratória
- Semana 5: Pensamento estatístico em Python
- Semana 6: Continuação Pensamento estatístico em Python
- Semana 7: Engenharia de Features
- Semana 8: Regressão
- Semana 9: Classificação

The background is a solid dark blue. It features abstract geometric lines in teal and purple. In the top-left corner, there are several overlapping lines forming a series of connected 'V' and inverted 'V' shapes. In the bottom-right corner, there are more lines forming a similar geometric pattern, including a large 'V' shape and some horizontal segments.

# AceLeraDev Data Science

2 - Quem sou eu ?

## RESUMO CURRICULUM

/

- Túlio Vieira de Souza
- Background Engenharia Civil
- Data Science Engineer @Avenue Code
- Pesquisador @Unilivrecoop
- Cofounder Machine Learning Experience
- <https://www.linkedin.com/in/tuliovieira/>

The background is a dark navy blue. It features abstract geometric lines in teal and purple. In the top-left corner, there are several overlapping lines forming a series of connected shapes. In the bottom-right corner, there are more lines forming a similar pattern. The main text is centered in the upper half of the image.

# AceLeraDev Data Science

3 - 0 que é ciência?

# Ciência

/

- Começa com observação;
  - Todo nosso “mundo” observado
- As observações são neutras;
  - Não interferimos no mundo ao observá-lo
- Raciocínio Lógico Indutivo
  - Proposições, conclusões e regras

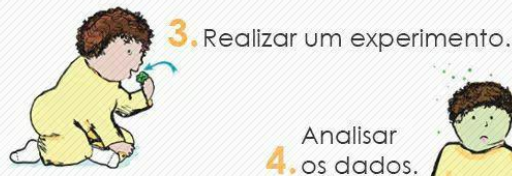
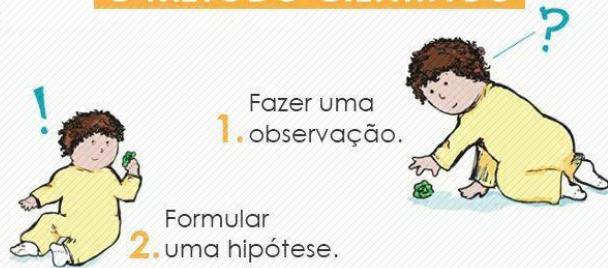
[ref.](#)

# Ciência

/

- René Descartes
  - O discurso do método
- Edgard Morin
  - Pensamento Complexo
  - O viés da variável omitida

# TODO BEBÊ CONHECE O MÉTODO CIENTÍFICO



Convidar terceiros  
para replicar os  
**6.** resultados.





# Ciência

/

- René Descartes
  - O discurso do método
- Edgard Morin
  - Pensamento Complexo
  - O viés da variável omitida

# Ciência de Dados

/

- A profissão “mais sexy” do séc. 21 ([ref](#))
- Resolver problemas interdisciplinar;
- Foco no resultado (do negócio);
- Utilizar métodos e metodologias científicas
- DON`T FOLLOW THE HYPE

[ref.](#)

The background is a dark navy blue. In the top-left corner, there are several overlapping geometric shapes made of thin lines in teal and purple. In the bottom-right corner, there are more overlapping geometric shapes, also in teal and purple, mirroring the style of the top-left decorations.

# AceLeraDev

# Data Science

4 - Big data: Um grande problema



## Velocity

Speed at which data is emanating and changes are occurring between the diverse data sets

## Velocity



## Volume

This refers to the sheer volume of data being generated every second.

## Volume



## Variety

## Variety

Can use structured as well as unstructured data.



## Veracity

## Veracity

Data reliability and trust.  
Verifying and validating the data



## Value

## Value

Having access to big data is all well and good but that's only useful if we can turn it into a value.



# 5V'S OF BIG DATA

# Governança de Dados

- Para onde vou (em termos de negócios), como vou, quando vou?
- Que dados serão necessários nesse caminho?
- Como obtê-los, como mantê-los?
- Que áreas serão prioritárias no tratamento dos dados, baseado nas estratégias de negócios?



# AceLeraDev Data Science

5 - Papéis dentro de um projeto de dados



# Top Data Science Jobs

Data Scientist

Data Architect

Data Engineer

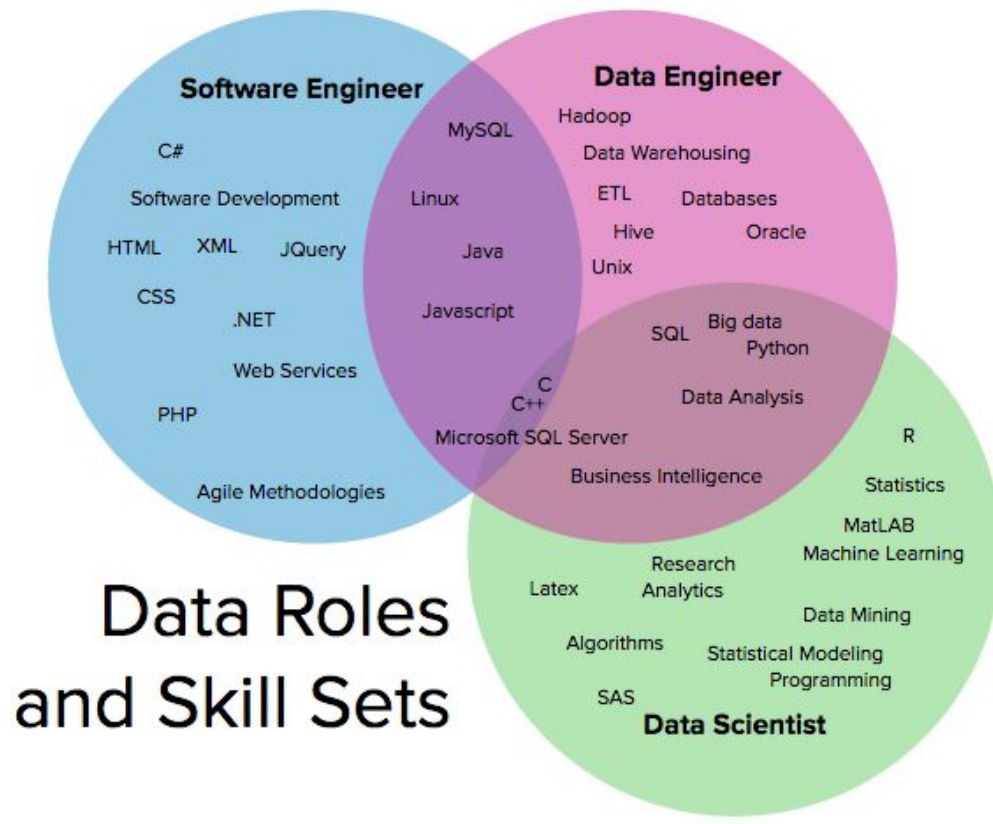
Statistician

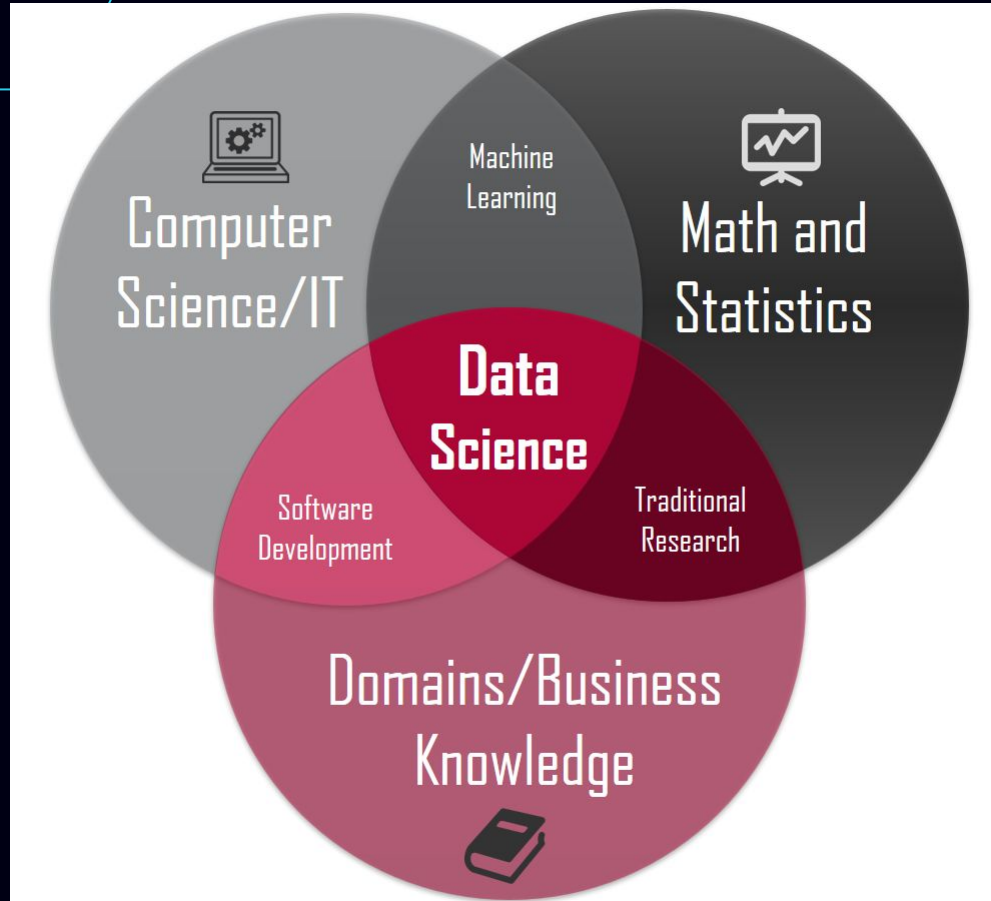
Data Science  
Manager

Machine  
Learning  
Engineer

Decision  
Scientists







The background is a dark navy blue. It features abstract geometric line art in teal and purple. In the top-left corner, there are several overlapping lines forming a series of connected 'V' and inverted 'V' shapes. In the bottom-right corner, there are more lines forming a similar geometric pattern, including a prominent 'V' shape and some horizontal segments.

# AceLeraDev Data Science

6 - Tipos de soluções de análises

# Tipos de Soluções de Análises

/

- Solução Descritiva (Descriptive Analytics)
  - Descreve o que existe;
  - Olha apenas para o **passado/atual**;
  - **“O que aconteceu ?”**

[ref](#)

# Tipos de Soluções de Análises

/

- Solução Diagnóstica
  - Descreve os **porquês**;
  - Olha apenas para o **passado/atual**;
  - **“Porque isso aconteceu ?”**

[ref](#)

# Tipos de Soluções de Análises

/

- Solução Preditiva (Predictive Analytics)
  - Descreve o que vai existir;
  - Usa o **passado** para prever/predizer o **futuro** ;
  - **“O que vai acontecer ?”**

[ref](#)

# Tipos de Soluções de Análises

/

- Solução Preditiva (Predictive Analytics)
- **Previsão:** subconjunto da predição vinculado ao tempo.
- **Predição:** Forma genérica de como um fato seria antes de sua ocorrência.

# Tipos de Soluções de Análises

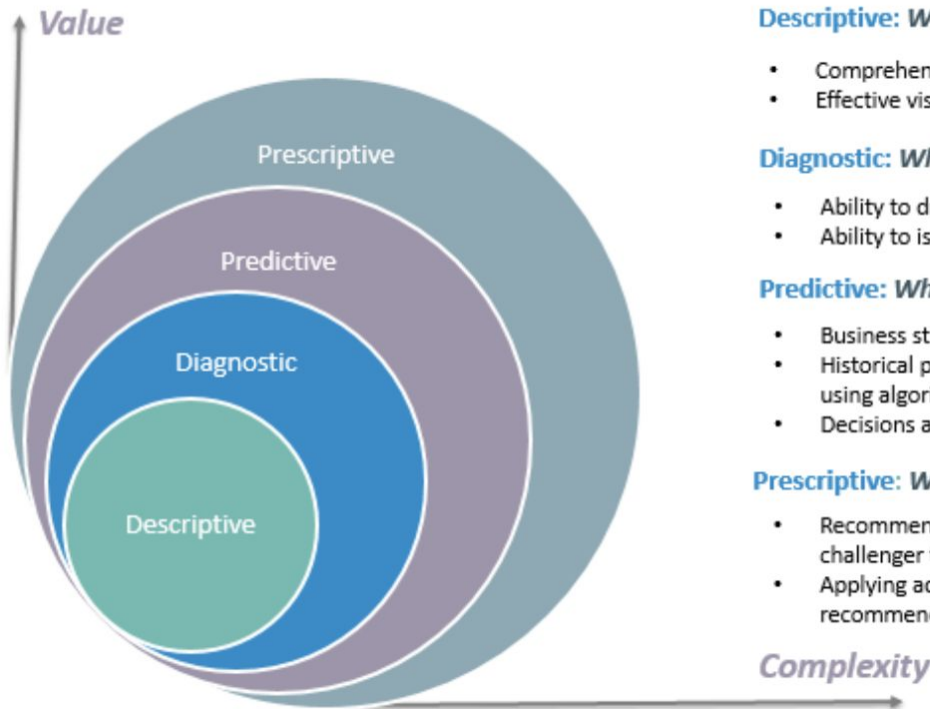
/

- Solução Prescritiva (Prescriptive Analytics)
  - Recomenda tomada de decisões
  - Usando um predição/previsão **recomenda uma ação**
  - **“O que vou fazer?”**

[ref](#)



## 4 types of Data Analytics



### What is the data telling you?

#### **Descriptive:** *What's happening in my business?*

- Comprehensive, accurate and live data
- Effective visualisation

#### **Diagnostic:** *Why is it happening?*

- Ability to drill down to the root-cause
- Ability to isolate all confounding information

#### **Predictive:** *What's likely to happen?*

- Business strategies have remained fairly consistent over time
- Historical patterns being used to predict specific outcomes using algorithms
- Decisions are automated using algorithms and technology

#### **Prescriptive:** *What do I need to do?*

- Recommended actions and strategies based on champion / challenger testing strategy outcomes
- Applying advanced analytical techniques to make specific recommendations

The slide features a dark blue background with decorative geometric lines in the corners. In the top-left, there are overlapping lines in purple and teal forming a hexagonal shape. In the bottom-right, there are similar overlapping lines in teal and purple forming a hexagonal shape.

# AceLeraDev

# Data Science

7 - Tipos de problemas

# Tipos de problemas

## Classificação binária

- Preveem um resultado binário (uma de duas classes possíveis)
- Regressão logística.
- Exemplos de problemas de classificação binária
  - "Este e-mail é spam ou não?"
  - "O cliente comprará este produto?"
  - "Este produto é um livro ou um animal de fazenda?"
  - "Esta revisão foi escrita por um cliente ou por um robô?"

# Tipos de problemas

## Classificação multiclasse

- Permitem gerar previsões para várias classes (prever um entre mais de dois resultados).
- Regressão logística multinomial.
- Exemplos de problemas multiclasse
  - "Este produto é um livro, um filme ou vestuário?"
  - "Este filme é uma comédia romântica, um documentário ou um suspense?"
  - "Qual categoria de produtos é mais interessante?"

# Tipos de problemas

## Regressão

- Modelo de regressão
- Preveem um valor numérico
- Regressão linear.
- Exemplos de problemas de regressão
  - "Qual será a temperatura em Joinville amanhã?"
  - "Quantas unidades deste produto serão vendidas?"
  - "Qual será o preço de venda desta casa?"

# Tipos de problemas

## Agrupamento / clusterização

- Modelo aprende grupos / divisões nos dados
- Retorna um grupo
- KMeans
- Exemplos de problemas de agrupamento
  - "Quanto perfis de cliente minha empresa tem?"
  - "Quanto grupos de alunos tenho na minha classe?"
  - "Qual será o preço de venda desta casa?"

# Tipos de problemas

## Sistemas de Recomendação

- Modelo sugere/ recomenda algo à um usuário
- Retorna sugestão
- Collaborative filtering / Content Based
- Exemplos de problemas de recomendação?
  - "Qual filme recomendar à um usuário baseado em seu gosto?"
  - "Qual filme recomendar à um usuário baseado no gosto de pessoas similares?"
  - "Qual imóvel comprar?"

The background is a dark navy blue. It features abstract geometric line art in teal and purple. In the top-left corner, there are several interconnected lines forming a series of triangles and polygons. In the bottom-right corner, there are more geometric shapes, including a large triangle and some smaller polygons, also formed by teal and purple lines.

# AceLeraDev

# Data Science

8 - Tipos de aprendizado



# Tipos de Aprendizados

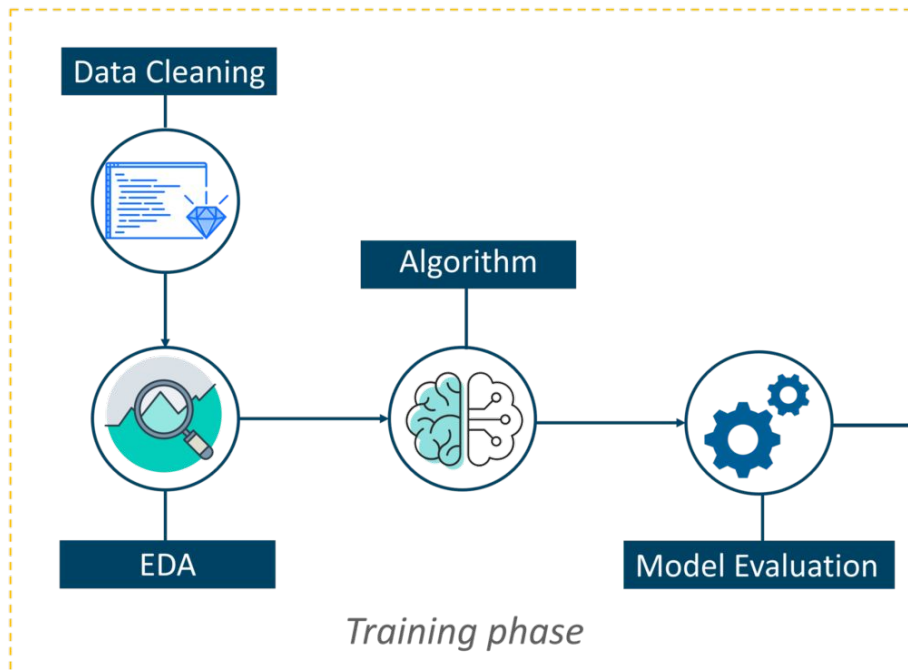
/

- **Aprendizado supervisionado:** São apresentadas ao computador exemplos de entradas e saídas desejadas, fornecidas por um "professor". O objetivo é aprender uma regra geral que mapeia as entradas para as saídas.

## Labelled Data



Known output



## Labelled Output



Class 'Jerry'



Class 'Tom'

# Tipos de aprendizado

/

- **Aprendizado não supervisionado:** Nenhum tipo de etiqueta é dado. Encontrar estrutura nas entradas fornecidas sozinho
- Pode ser um objetivo em si mesmo (descobrir novos padrões nos dados) ou um meio para atingir um fim.

Unlabelled Data



Unknown output

Data Cleaning



Algorithm



EDA

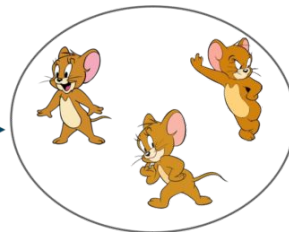


Model Evaluation



*Understand patterns & discover output*

Unlabelled Output



*Clusters formed based on feature similarity*

The background is a dark navy blue. It features abstract geometric line art in teal and purple. In the top-left corner, there are several overlapping lines forming a series of connected shapes, including a hexagon and a parallelogram. In the bottom-right corner, there are more geometric shapes, including a large parallelogram and a smaller hexagon, also formed by overlapping lines.

# AceLeraDev Data Science

9 - Maturidade com analytics

# Maturidade dos dados



## Raw Data

Data bring low value to the company.



## Information

Data start to bring value to the company but the reports are not automated.



## Business Intelligence

Clear vision about the past, present and defined data governance.



## Advanced Analytics

Deep knowledge about data and predictive analysis.

# Maturidade analítica



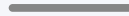
**Descriptive**

What happened?



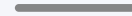
**Diagnostic**

Why did it happen?



**Predictive**

What will happen?



**Prescriptive**

What we should do?

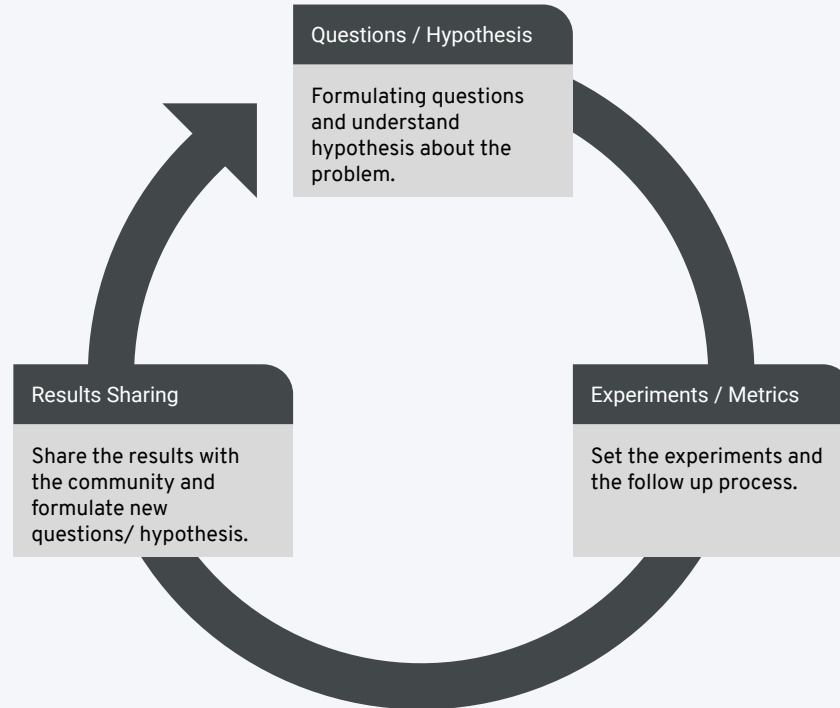
The background is a dark navy blue. It features abstract geometric lines in teal and purple. In the top-left corner, there are several overlapping lines forming a series of connected 'V' and inverted 'V' shapes. In the bottom-right corner, there are more lines forming a similar geometric pattern, including a large 'V' shape and some horizontal segments.

# AceLeraDev Data Science

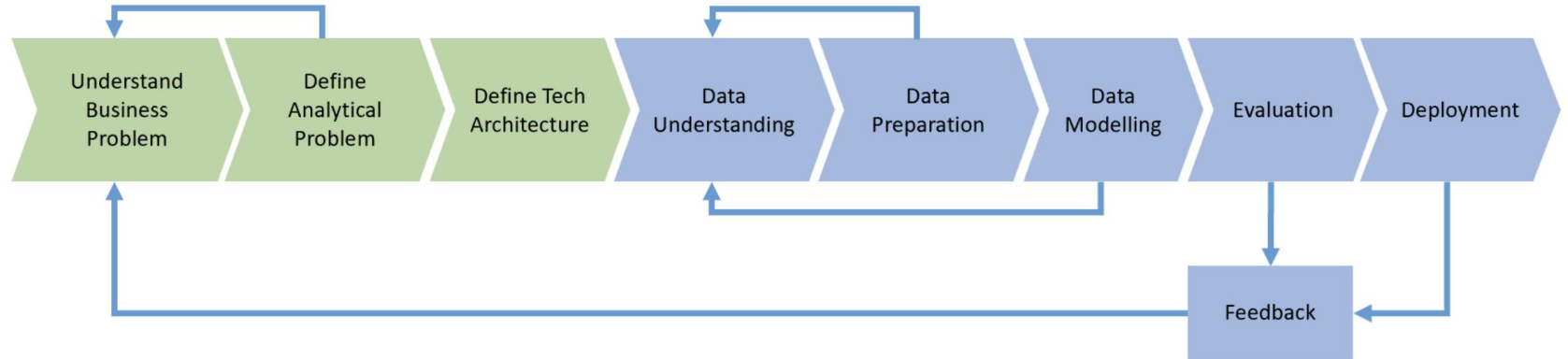
10 - O processo de ciência de dados



# Método científico



# Processo



[ref](#)



# Preliminary Phase

## Understand Business Problem

- Ask many questions as needed
- Immersion into the problem before bring the solution

## Define Analytical Problem

- Translate business needs into a data solution and plan how to delivery.
- Define business KPIs to evaluate model success.

## Define Tech Architecture

- Define what technology will be used to solve business problem.
- Clarify to the customer what we are going to do.

# Exploratory Phase

## Data Understanding

- Explore correlations between variables.
- Hypothesis testing.

## Data Processing

- Feature Selection.
- Feature Engineering.

## Data Modelling

- Set up different experiences and best metrics.
- Find the best fit model aligned to business needs.

# Production Phase

## Evaluation

- Model explainability and results communication.
- Check model performance with different scenarios.

## Deployment

- Deploy the model in project architecture.
- Assure the quality of solution.

## Feedback

- A/B test with real world.
- Monitoring model metrics and business KPI's.

# Tópicos da próxima semana:

/

- Manipulando DataFrames com Pandas
- Tipos de dados mais comuns
- Dados faltantes e imputação de dados
- Conversão de dados
- Normalização de dados