

# Data mining

Bem vindos!!!



# Na aula de hoje

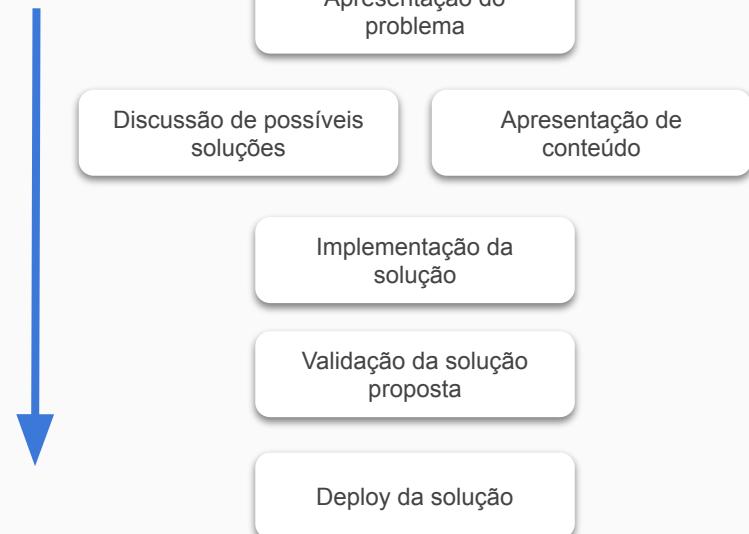
## Tópicos da aula

- Overview do curso
- Apresentação do cliente
- Apresentação...
- O que é Data Mining?
- Exemplos de Data Mining
- Importância no mercado
- Como fazemos Data Mining?
- Data Mining Life Cycle
- A linguagem Python
- Ambientes de programação
- Por onde começar?
- Exemplos

# Overview do curso

- **Aula de Introdução**
- Atacando o problema...
- Implementação da solução...
- Implementação da solução...
- Validação da solução

Aulas estudo de caso:



# Nosso trabalho final



## **Primeira possibilidade:**

Vocês podem escolher um problema real da sua empresa e consolidar uma estrutura de solução real com um diagrama mostrando cada etapa do processo da solução de IA.

Nesse caso sua nota será determinada pelo engajamento que terá na solução do problema e pela validade da solução e coesão do processo de solução desenvolvida.

# Perguntas?

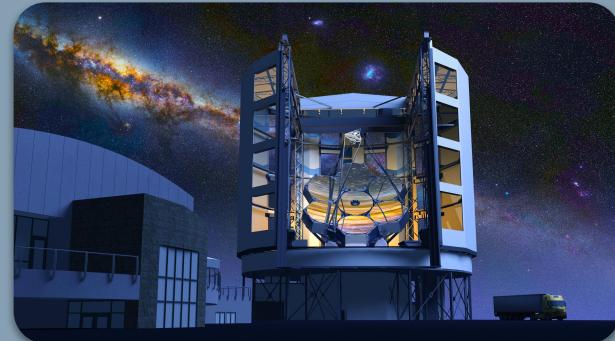
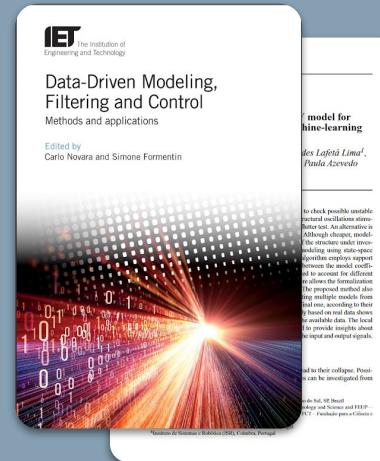
# Minha história...

## *Engenharia Eletrônica*

IMT 2015 - 2019

### Projetos:

- Wing Flutter Benchmark
- Giant Magellan Telescope
- Alertness Project
- Epidemic Models Project
- CoRoT Contributions
- PLATO



O que é  
Data Mining?

# O que é um dado?

	carat	cut	color	clarity	depth	table	price	x	y	z
1	0.52	Ideal	D	VS2	61.4	56	1664	5.16	5.19	3.18
2	0.5	Very Good	F	SI1	62.3	60	1250	5.07	5.11	3.17
3	0.61	Ideal	G	VVS2	61.6	54	2242	5.45	5.49	3.37
4	0.36	Premium	G	VS2	62.5	58	756	4.55	4.51	2.83
5	0.7	Very Good	E	VS2	63.5	54	2889	5.62	5.66	3.58
6	0.56	Ideal	F	VS1	61.7	56	2016	5.32	5.28	3.27
7	1.19	Premium	E	I1	60.2	61	3572	6.91	6.87	4.15
8	0.52	Ideal	F	IF	60.6	57	2575	5.21	5.22	3.16
9	0.38	Ideal	E	IF	62.7	55	1433	4.61	4.67	2.91
10	0.51	Ideal	E	VS2	62.1	54	1608	5.13	5.15	3.19

Showing 1 to 10 of 1,000 entries      Previous [1](#) [2](#) [3](#) [4](#) [5](#) ... [100](#) Next



*Do grego, dado é um presente.*

Dado é qualquer que seja a informação que possa ser digitalizada e interpretada. E assim utilizada para extração de informação/conhecimento.

# O que é Data Mining?



Dado Bruto

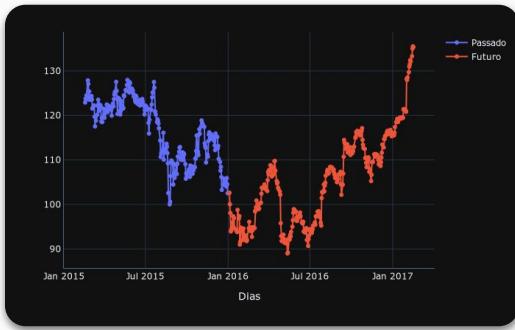


Informação  
Conhecimento

PROCESSO

	carat	cut	color	clarity	depth	table	price	x	y	z
1	0.52	Ideal	D	VS2	61.4	56	1664	5.16	5.19	3.18
2	0.5	Very Good	F	SI1	62.3	60	1250	5.07	5.11	3.17
3	0.61	Ideal	G	VVS2	61.6	54	2242	5.45	5.49	3.37
4	0.36	Premium	G	VS2	62.5	58	756	4.55	4.51	2.83
5	0.7	Very Good	E	VS2	63.5	54	2889	5.62	5.66	3.58
6	0.56	Ideal	F	VS1	61.7	56	2016	5.32	5.28	3.27
7	1.19	Premium	E	I1	60.2	61	3572	6.91	6.87	4.15
8	0.52	Ideal	F	IF	60.6	57	2575	5.21	5.22	3.16
9	0.38	Ideal	E	IF	62.7	55	1433	4.61	4.67	2.91
10	0.51	Ideal	E	VS2	62.1	54	1608	5.13	5.15	3.19

Showing 1 to 10 of 1,000 entries Previous 1 2 3 4 5 ... 100 Next



Dashboard

**\$ 500,43,00**

Actions

- Credit Card
- Invest money now
- View statistic now
- Send money now

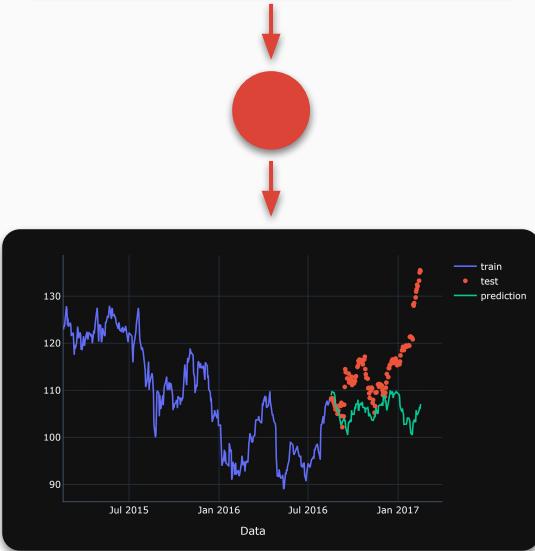
Transaction

Date	Amount	Number
03-21-2018	\$162,00	FRAK 5840 6942 5744
03-19-2018	\$32,00	FKGF 5435 4433 6532
03-17-2018	\$122,00	LSNS 7382 8374 4672
03-17-2018	\$16,00	LSNS 8759 3847 3839

Status

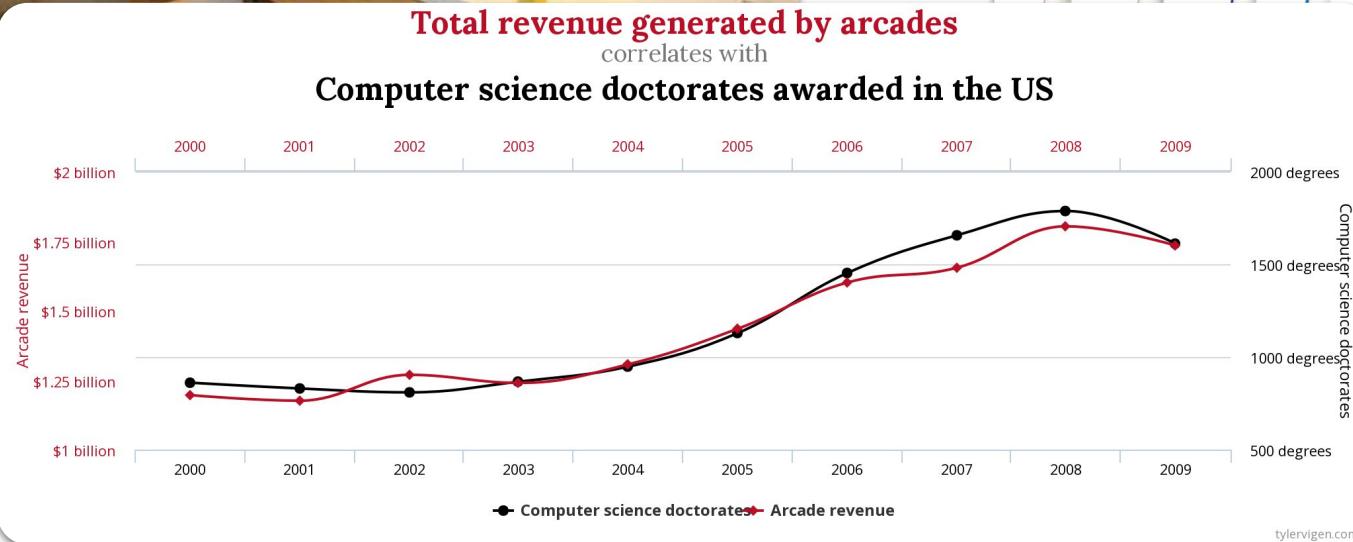
Today \$768,00

MIN \$768,00 MAX \$768,00



Exemplos de Data Mining

# Spurious Correlations

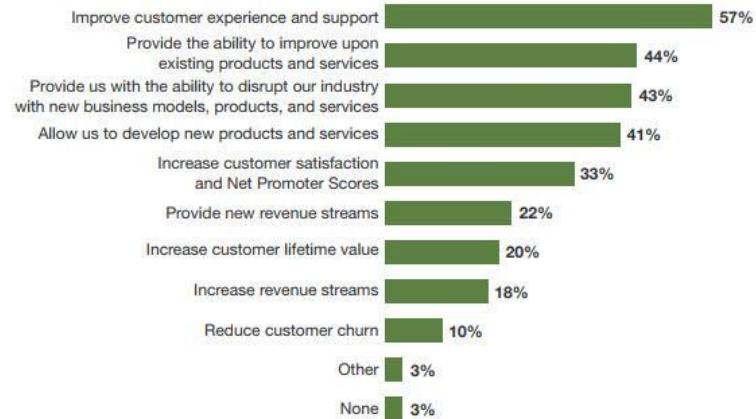


# Importância de mercado



**“What are the biggest strategic/growth benefits AI will contribute to your organization?”**

(Please select up to three)



Base: 598 business and technology professionals

Source: Forrester's Q2 2016 Global State Of Artificial Intelligence Online Survey

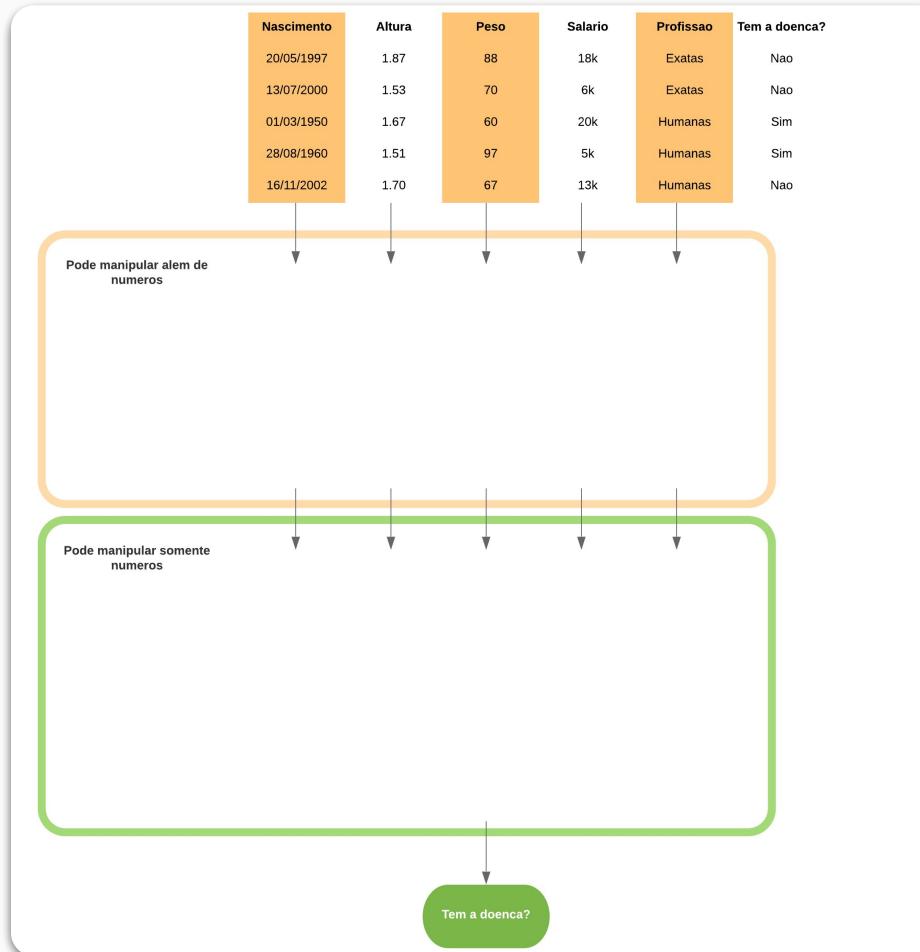
# Importância de mercado

*“57% dos executivos de empresas acreditam que o ganho mais significante da Inteligência Artificial e Machine Learning será devido a melhora da experiência de usuário e suporte ao cliente.”*

E como fazemos  
Data Mining?

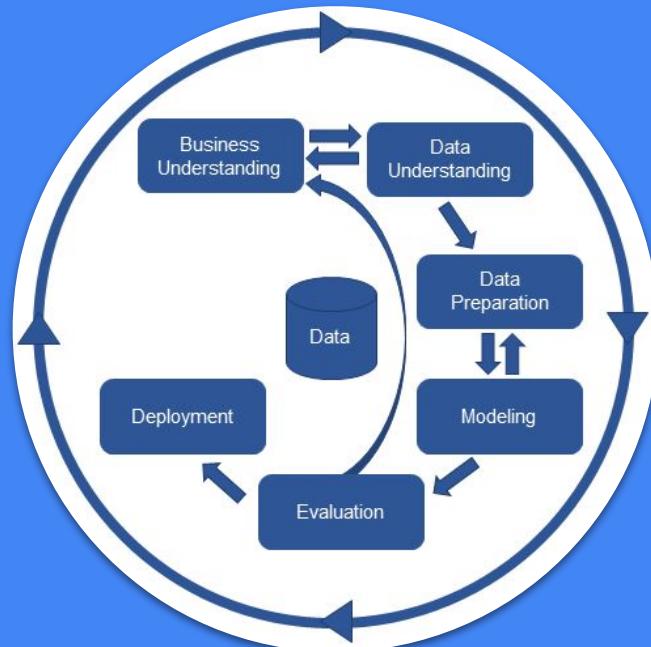
Nascimento	Altura	Peso	Salario	Profissao	Tem a doença?
20/05/1997	1.87	88	18k	Exatas	Nao
13/07/2000	1.53	70	6k	Exatas	Nao
01/03/1950	1.67	60	20k	Humanas	Sim
28/08/1960	1.51	97	5k	Humanas	Sim
16/11/2002	1.70	67	13k	Humanas	Nao

Como, baseado nos dados da tabela, e seguindo às regras, conseguimos criar um processo para saber se a pessoa tem uma doença?

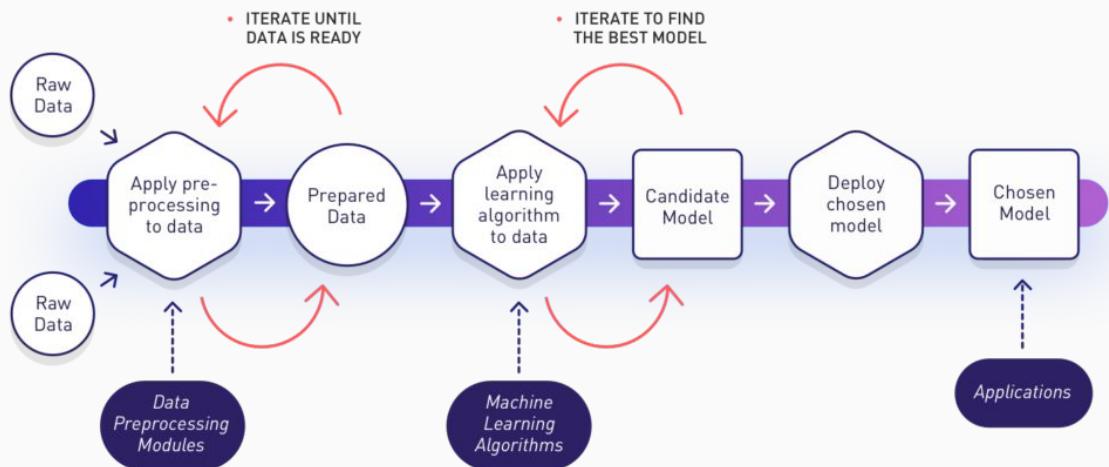


# Life Cycle

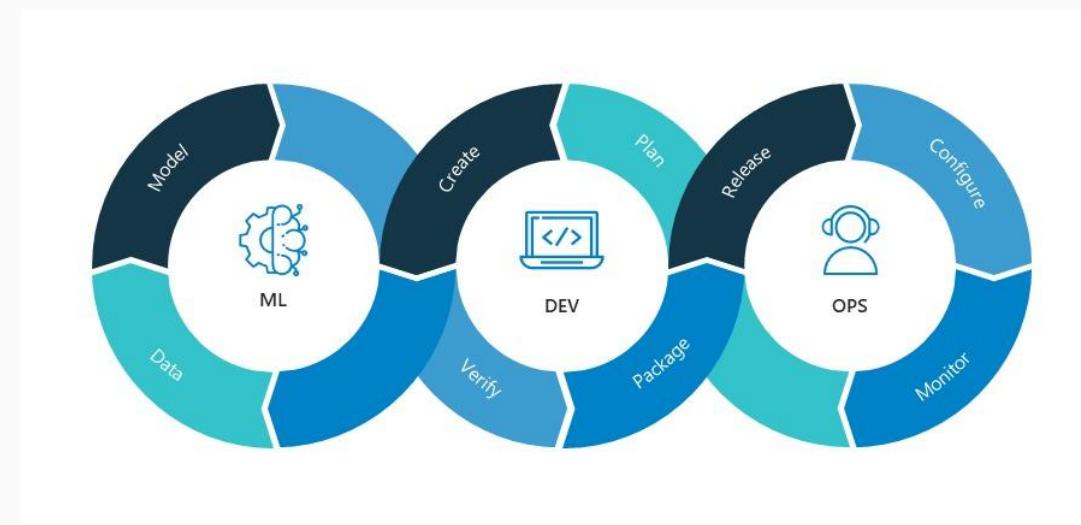
Método CRISP-DM



# Uma outra forma de visualizar



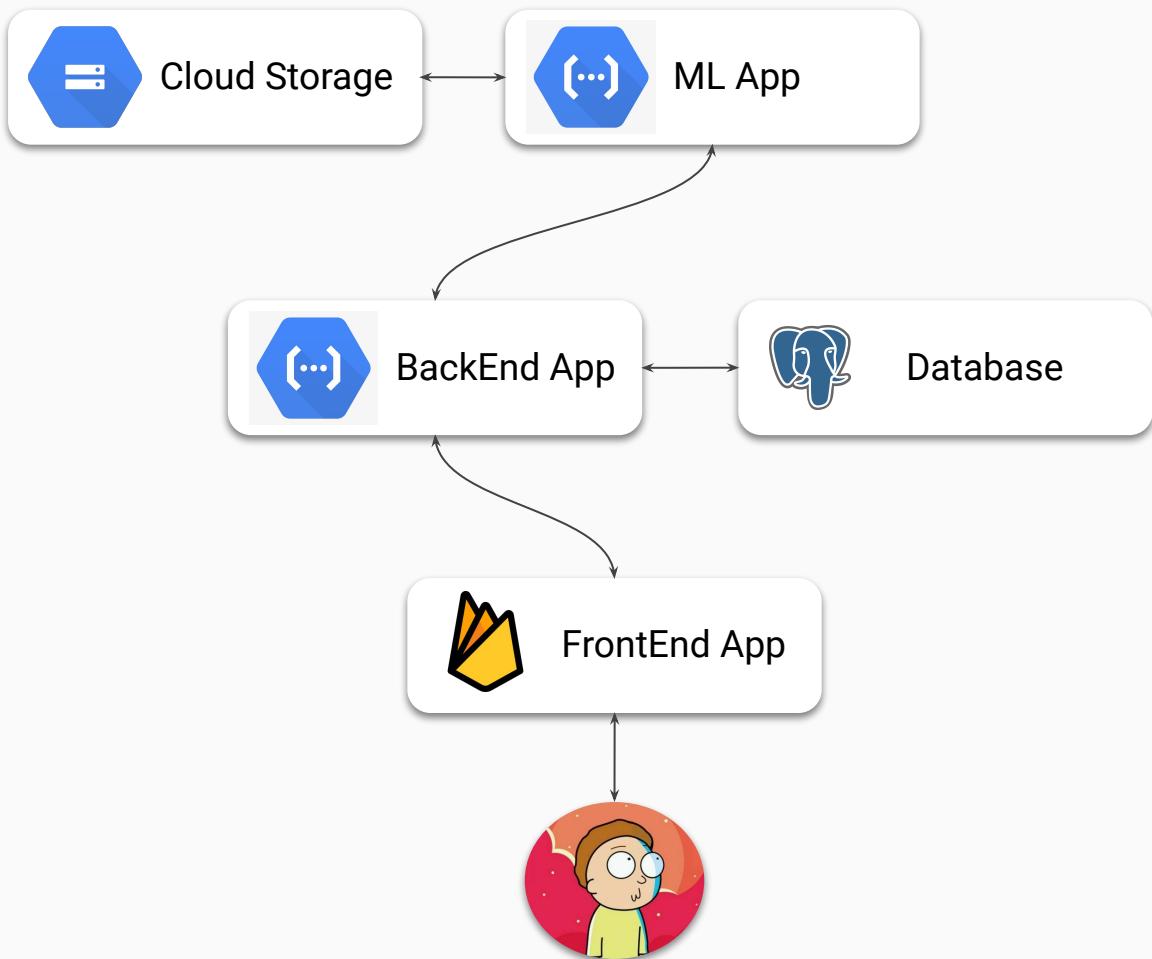
# Do ponto de vista de software



# Software Solution

# Qual a cara de uma solução de software?

- Base de dados
- Cloud
  - Infraestrutura
  - Arquitetura
  - Segurança
- Aplicação de BackEnd
  - Autorização
  - Autenticação
  - Funcionalidades básicas
  - APIs do modelo
  - APIs da aplicação
- Aplicação de FrontEnd



# Escopo do curso

# Como o python roda?

Todo código python precisa de um ambiente como base para sua arquitetura e pacotes!

## Virtual Environment



# O que é um ambiente virtual?



Todo ambiente virtual tem os seguintes elementos:

- Bibliotecas de desenvolvimento
- **Interpretador da linguagem**
- Capacidade de rodar scripts
- Gerenciadores de pacotes

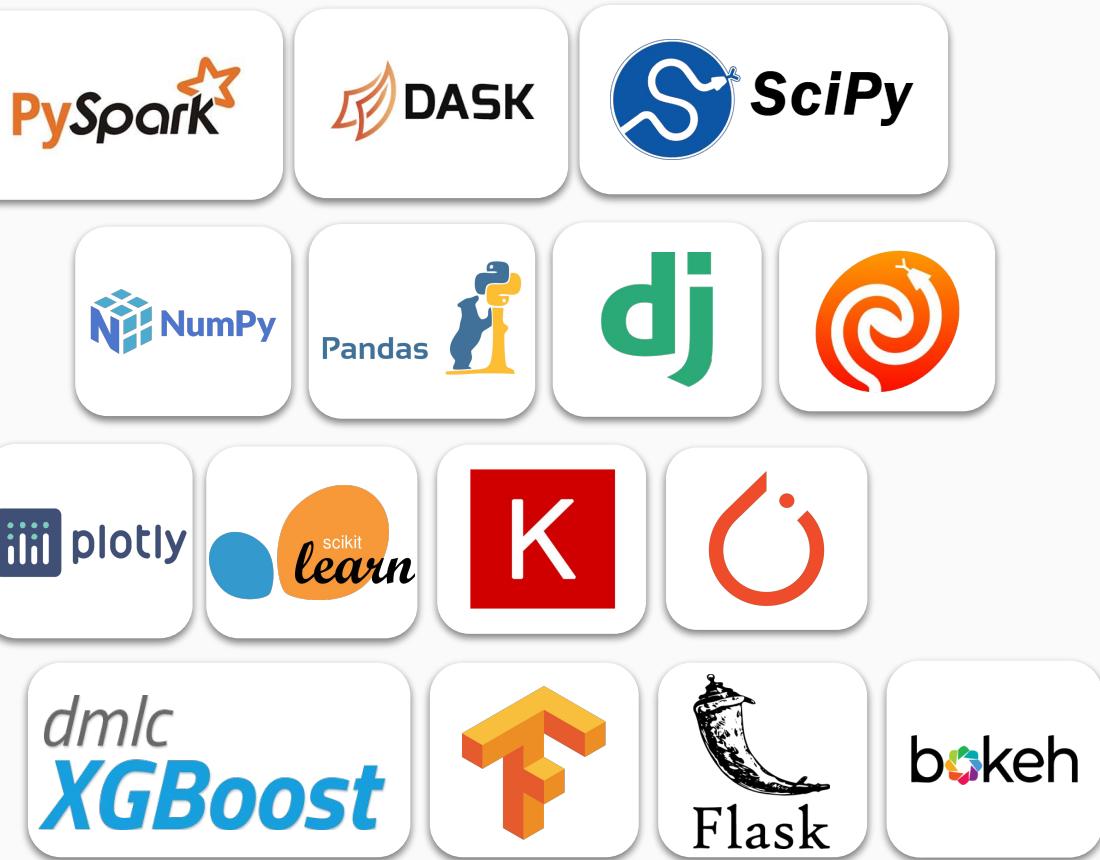
Comandos para gerenciar ambientes virtuais:

- Criar e deletar ambientes
- Editar ambientes
- Ativar e desativar ambientes
- Criar resumos do ambiente
- Exportar e importar ambientes
- Gerenciar todo o seu desenvolvimento

# O que é o pip?

Gerenciador de pacotes do python.

O que é um  
gerenciador de  
pacotes?



# Aplicações

- [Desenvolvimento Web](#)
- Desenvolvimento de Games
- [Arquitetura Cloud](#)
- [Arquitetura de Banco de Dados](#)
- [Automatização de tarefas](#)
- [Reverse Engineering](#)
- [Aplicações embarcadas](#)
- Projetos CAD (Blender)
- Aplicações Business
- Desktop GUI
- Aplicações com Audio e Video
- [Web Scraping](#)
- [Machine Learning e AI](#)
- [Data Analysis e Data Visualization](#)



# E como utilizar os ambientes?



co

kaggle™

ANACONDA®

pipenv

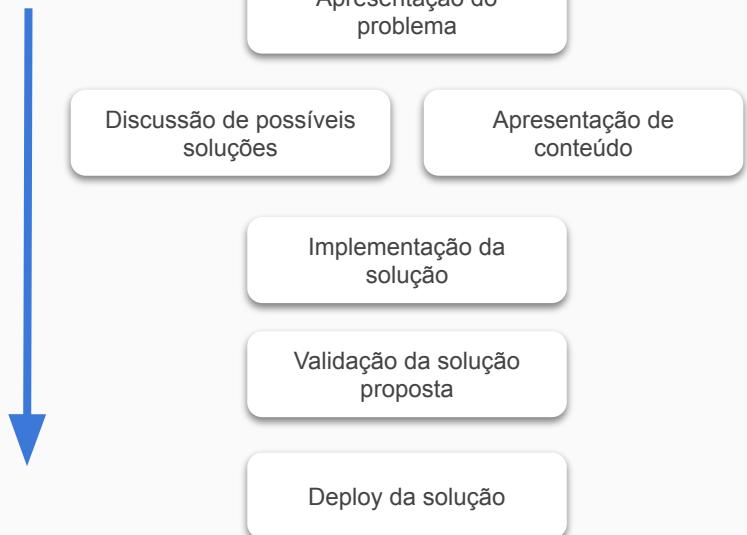
pyenv

virtualenv

# Overview do curso

- **(18/08) Aula de Introdução**
- (25/08) Técnicas de Data Mining
- Estudo de caso
- Aula para *Trabalho*
- Web Scraping
- Estudo de caso
- Problema real: Elo7
- Trabalho Final

Aulas estudo de caso:



# Nosso trabalho final



## **Primeira possibilidade:**

Vocês podem trazer um problema real de vocês para resolver dentro do curso, ou algum problema que tenha se interessado do Kaggle. Na próxima semana vamos sentar e definir o problema que você vai resolver. Tente trazer pelo menos três diferentes problemas que acham interessante para podermos discutir. Dentre esses problemas, você pode incluir algum de sua empresa, ou problemas que você conhece e deseja resolver.

## **Segunda possibilidade:**

Eu vou preparar alguns problemas para trazer na próxima semana e vocês resolverem. E vocês podem escolher dentre alguns exemplos qual mais lhe satisfaz e resolver ele até o final do curso.

Muito obrigado!  
Dúvidas?