



# Data Science com Python: frameworks e melhores práticas

# Quem eu sou?

---

**Marlesson Santana** - marlesson@cquantt.com



- **CDO na CQuantt** - Responsável por conduzir projetos de Ciência de Dados e Machine Learning
- **Pesquisador** na área de Deep Learning - UFG
- Community Manager do **Data Hackers** e DataTrain

**CQuantt**



**UFG**



/marlesson-santana-25019358/



/marlesson

#estatistica #machineLearning #spark  
#python #deepLearning #tensorflow  
#bigData #analytics

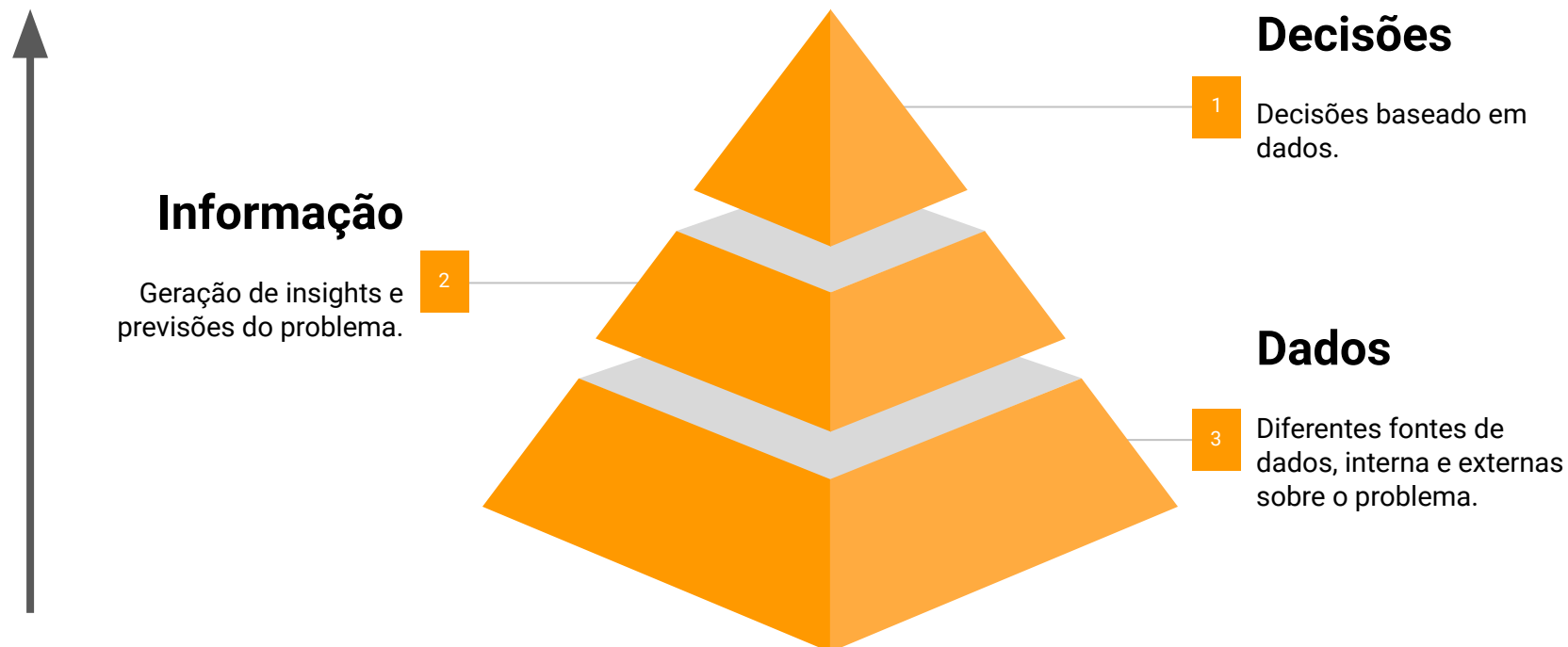
# O que é Data Science?



# Aplicações de Data Science

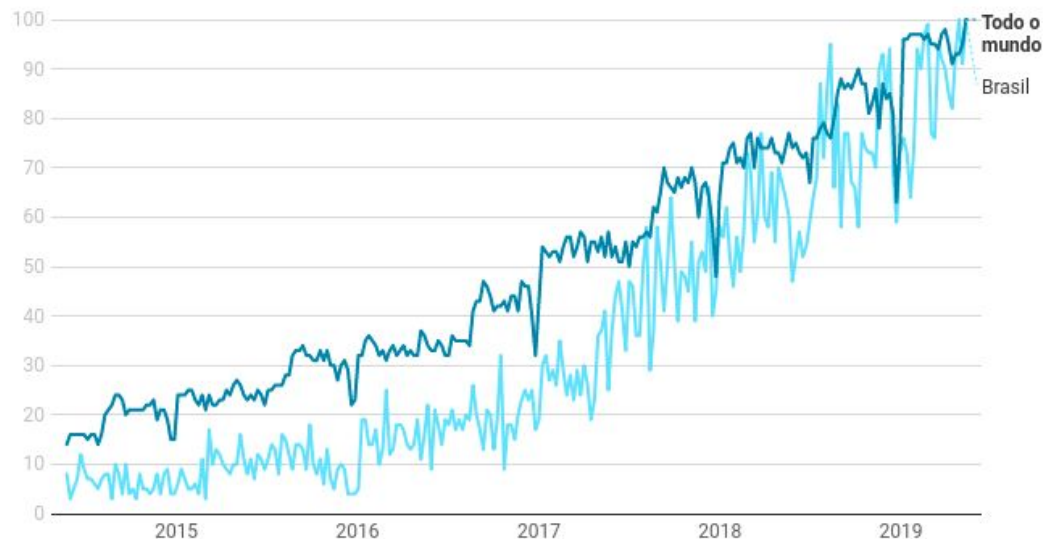
---

A área de Data Science é **transversal** dentro de qualquer empresa e atende todas os setores (financeiro, logístico, comunicação..).



# Qual o principal motivo do crescimento de DS nos últimos anos?

## Popularidade das buscas por 'data science'

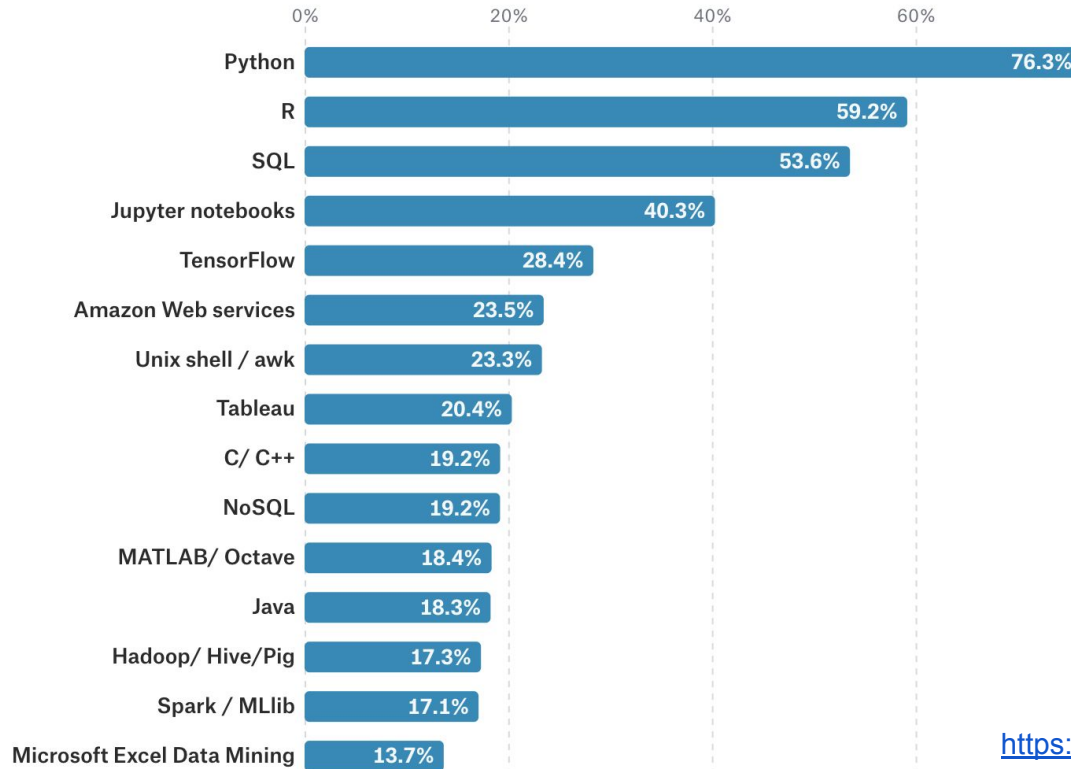




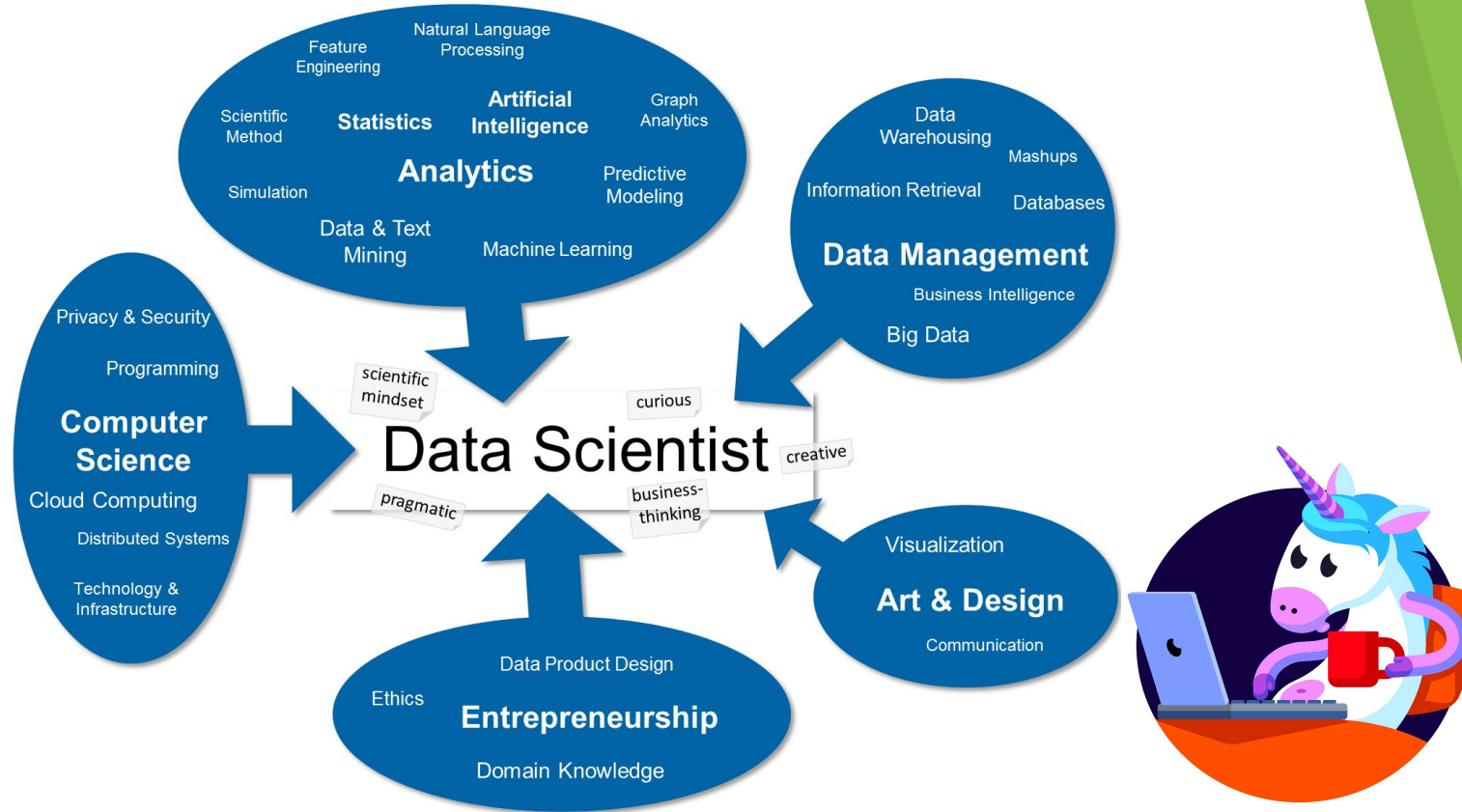
**VS**



# Popularidade das tecnologias usadas DS



# Qual o perfil do Data Scientist?





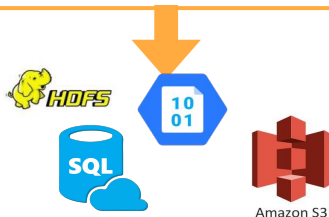
# Profissionais de Dados

## Engenheiro de Dados



- Coleta de dados
- Infraestrutura de armazenamento
- Qualidade dos dados

#sql, #nosql, #ETL #bigdata  
#python #cloud



## Cientista de Dados



- Análise estatística e ML
- Criação de Modelos
- Visualização dos resultados
- Geração de insights

#estatística #python #r  
#machinelearning #dataviz



## Engenheiro de Machine Learning



- Operacionalizar modelos
- Implementar modelos escaláveis
- Integração

#matemática #machinelearning  
#cloud #MLOps

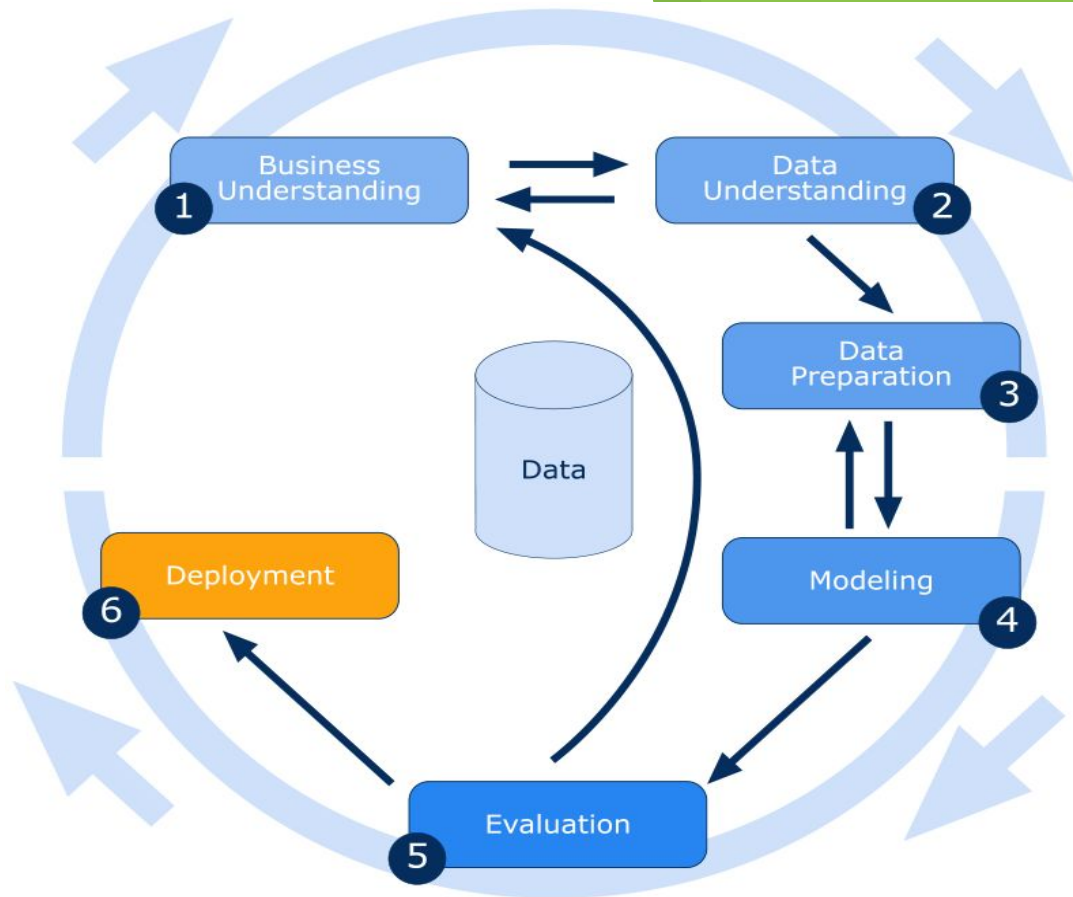


# Projetos de Data Science, como gerenciar?

KDD

**CRISP-DM**

SEMMA



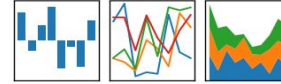
# Como vender Data Science dentro da minha empresa?



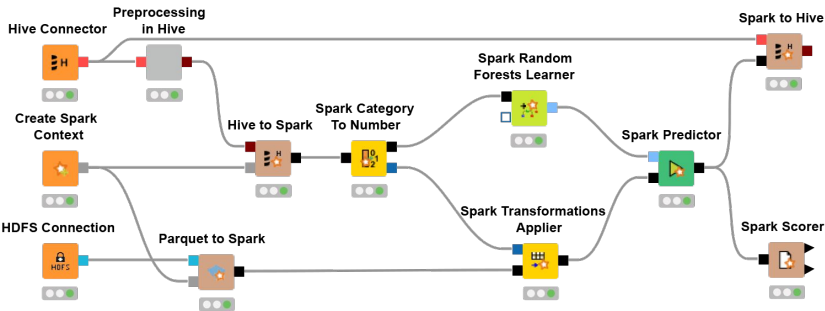
# Ferramentas e Frameworks



pandas  
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



# Ferramentas e Frameworks



```
from __future__ import absolute_import, division, print_function, unicode_literals

# Install TensorFlow
try:
    # %tensorflow_version only exists in Colab.
    %tensorflow_version 2.x
except Exception:
    pass

import tensorflow as tf
```

Load and prepare the [MNIST dataset](#). Convert the samples from integers to floating-point numbers:

```
mnist = tf.keras.datasets.mnist

(x_train, y_train), (x_test, y_test) = mnist.load_data()
x_train, x_test = x_train / 255.0, x_test / 255.0
```

Build the [tf.keras.Sequential](#) model by stacking layers. Choose an optimizer and loss function for training:

```
model = tf.keras.models.Sequential([
    tf.keras.layers.Flatten(input_shape=(28, 28)),
    tf.keras.layers.Dense(128, activation='relu'),
    tf.keras.layers.Dropout(0.2),
    tf.keras.layers.Dense(10, activation='softmax')
])

model.compile(optimizer='adam',
              loss='sparse_categorical_crossentropy',
              metrics=['accuracy'])
```

Como começar na área de dados?

kaggle

alura





**Marlesson Santana**

[marlesson@cquantt.com](mailto:marlesson@cquantt.com)

@marlesson

Obrigado por ter vindo!