

Ciência de Dados

WEBINAR 01 – Apresentação da disciplina



UNIVERSIDADE
CANDIDO
MENDES

EAD ■

Ciência de Dados

Unidade 01



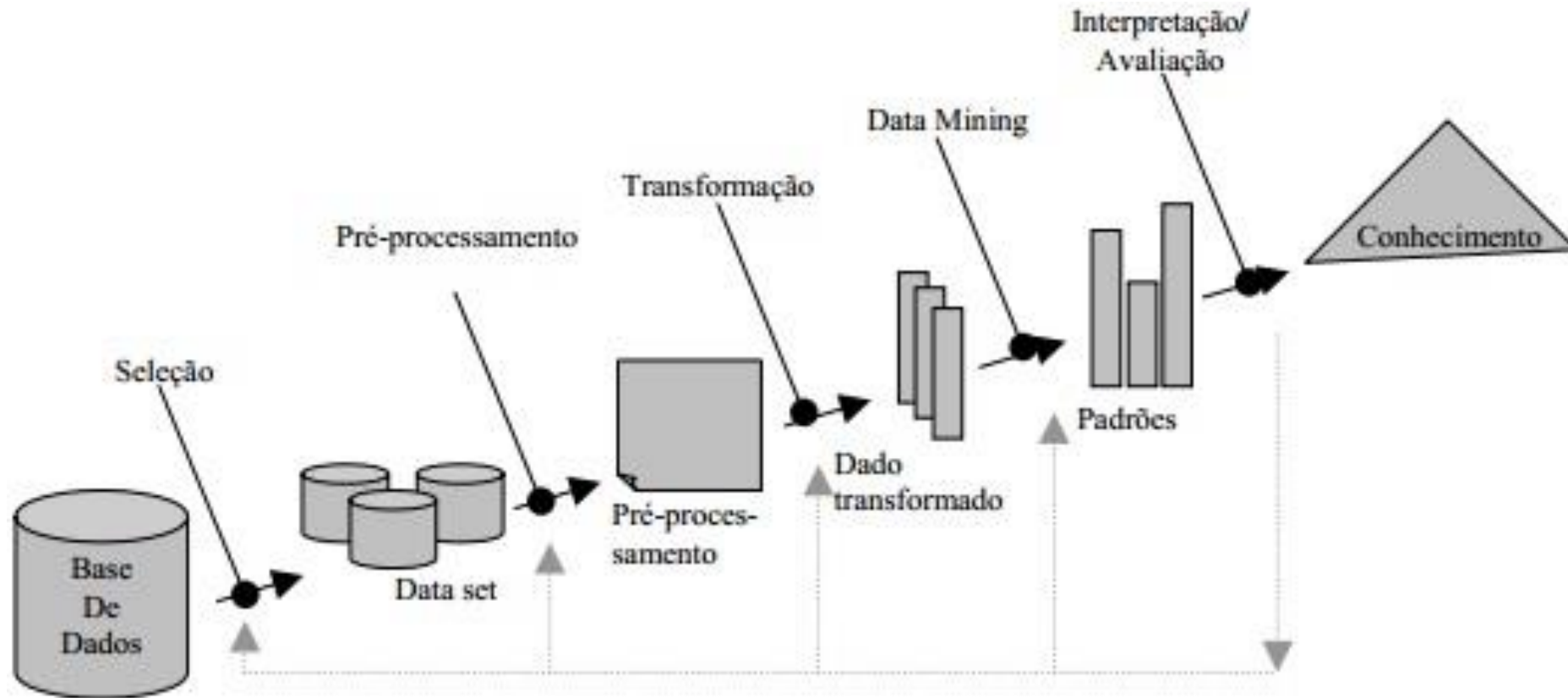
UNIVERSIDADE
CANDIDO
MENDES

EAD ■

■ Roadmap

- Introdução a Big Data e Ciência de Dados:
 - Os 6 Vs do Big Data:
 - Volume, Velocidade, Variedade, Veracidade, Valor e Variabilidade;
 - Tipos de dados: estruturados, semiestruturados e não estruturados;
- Ciclo de vida dos dados:
 - DMBok:
 - Data Management Body of Knowledge;
 - Governança de dados;
 - Extract, Transform and Load (ETL):
 - Extração de dados, transformação e limpeza -> preparação necessária para o KDD (knowledge-discovery in databases)
 - O que são Outliers?

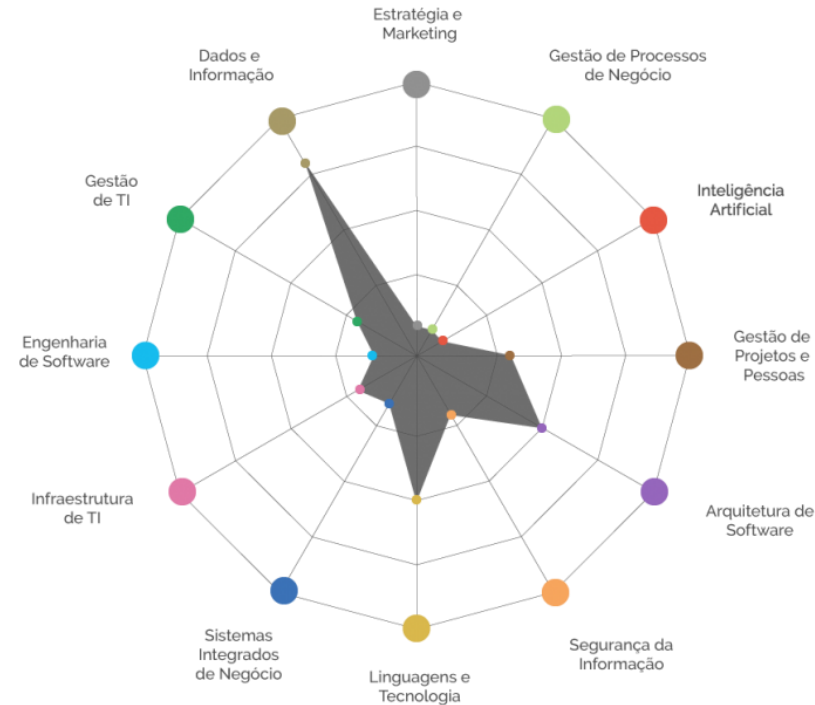
■ Roadmap



■ Roadmap

■ Diversos papéis em projetos de Big Data e Ciência de Dados:

- DBA;
- Engenheiro de Dados;
- Analista de Dados;
- Estatístico;
- **Cientista de Dados;**
- DevOps;



Ciência de Dados

Unidade 02



UNIVERSIDADE
CANDIDO
MENDES

EAD ■

■ Roadmap

- Técnicas utilizadas em Big Data e Ciências de Dados:
 - O que é aprendizado de máquina;
 - Tipos de aprendizado de máquina;
- Ferramentas utilizadas em Big Data e Ciência de Dados:
 - Python – Utilização de Bibliotecas:
 - Pandas e Numpy;
 - Matplotlib, Seaborn e Plotly;
 - Scikit-learn e Statsmodel;
 - Anaconda;
 - Google Colaboratory;
- Estatística descritiva aplicada a projetos de Big Data:
 - Entendendo seus dados;
 - Estatística descritiva utilizando Python;

IMPORTANTE: NECESSÁRIO CONHECIMENTO BÁSICO DE PROGRAMAÇÃO EM PYTHON -> ESTUDEM

Ciência de Dados

Unidade 03



UNIVERSIDADE
CANDIDO
MENDES

EAD ■

■ Roadmap

- Introdução ao Aprendizado de Máquina:
 - Tipos de aprendizado de máquina por tipo de problema a ser solucionado;
 - Determinando o melhor algoritmo a ser aplicado para resolução de um determinado problema;
- Aprendizados não supervisionados:
 - Problemas de agrupamentos:
 - Entendendo o k-means – teoria;
 - Aplicação do k-means – prática;
 - Utilizando Python para resolução de problemas de agrupamentos utilizando o k-means;

Ciência de Dados

Unidade 04



UNIVERSIDADE
CANDIDO
MENDES

EAD ■

■ Roadmap

■ Aprendizados supervisionados:

■ Problemas de classificação e regressão

- Tipos de algoritmos CART (Classification and Regression Trees);
- O que é entropia?
- Entendendo algoritmos CART – teoria;
- Aplicação de algoritmos CART – prática;
- Utilizando Python para resolução de problemas de classificação e regressão utilizando algoritmos CART;

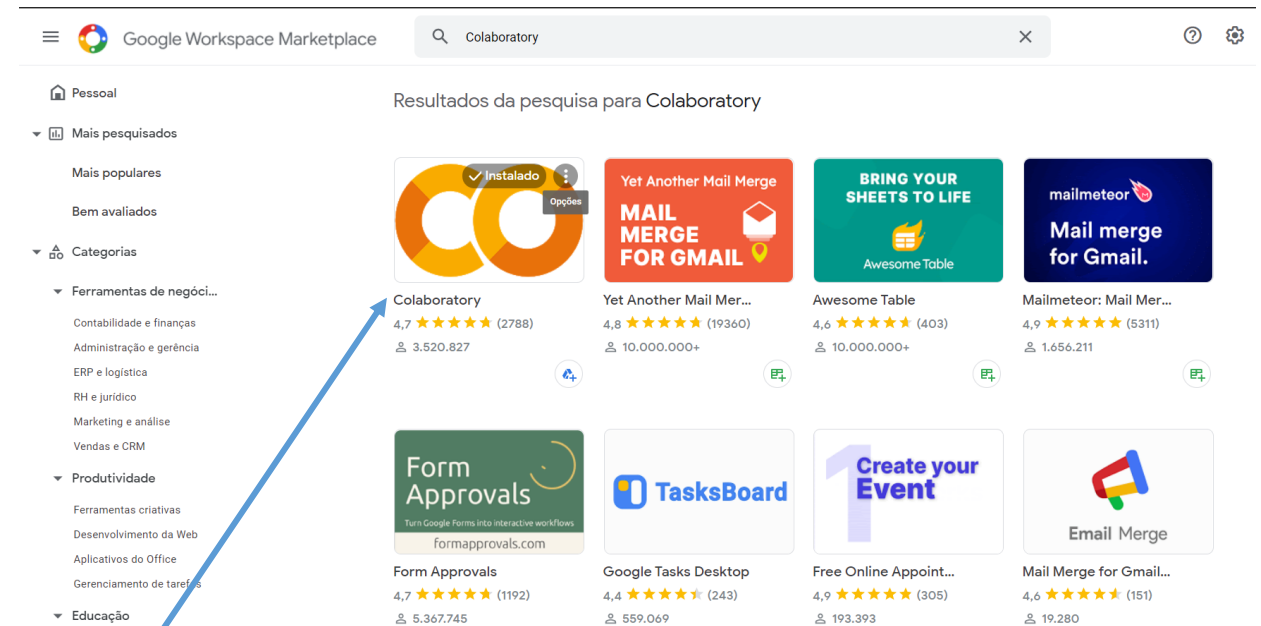
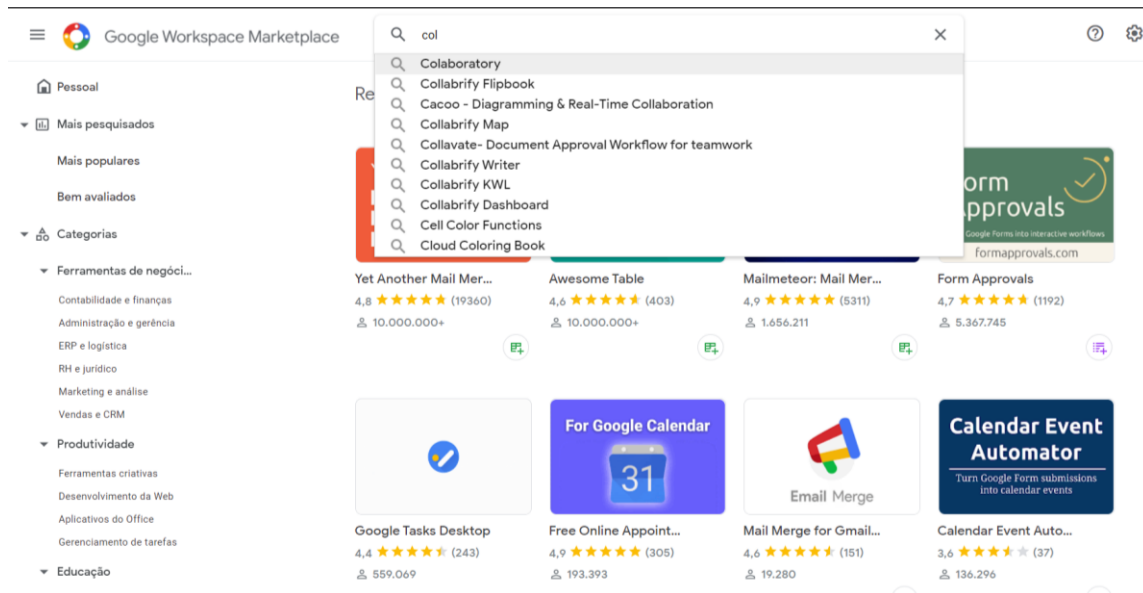


■ Roadmap

■ OBSERVAÇÕES:

- Favor estudar Python para ter um mínimo de conhecimento da linguagem;
- Estudem Pandas utilizando Python;
- Tragam suas dúvidas para os webinaries;
- Não deixem de utilizar os fóruns para resolução de dúvidas também;
- Se desejarem programar em seus computadores, instalem o Anaconda;
- Caso tenham dúvidas de como fazer, baixem o Anaconda em: www.anaconda.com e irei auxiliá-los na instalação;
- Para aqueles que não puderem utilizar seus computadores para programar, utilizem o Google Colaboratory;

Roadmap



Clique em Instalar (3 pontinhos)



UNIVERSIDADE
CANDIDO
MENDES

EAD ■

