R e PostgreSQL para Cência de Dados

JOSÉ DE JESUS FILHO

Sobre o autor:

1 – Jurimetrista: ciência de dados aplicada ao direito.

2 – Assessor do Ministério Público do Estado de São Paulo.

3 – Blog: https://rpg.consudata.com.br

4 - Pacote rpsql: https://github.com/jjesusfilho/rpsql

Porque R para ciência de dados:

1 – Porta de entrada para para programação.

2 – Alto número de pacotes estatísticos.

3 - Excelente IDE (Rstudio).

4 – Integra bem com a maioria dos SGBDs.

5 – Excelente cliente do PostgreSQL.

Porque PostgreSQL para ciência de dados:

1 – Alto nível organizacional.

2 – Integra bem com R e Python.

3 – Extensibilidade.

4 – Multiplicidade de tipos.

5 – 60 a 80% do tempo do cientista de dados é limpar dados para futura análise.

6 - Alta capacidade de absorver novos tipos de dados.

Possibilidades com R ou Python:

1 – Análise não supervisionada.

2 – Análise supevisionada.

3 – Ex. Kmeans, regressão linear, regressão logística, árvore dedecisão.

Possibilidades PostgreSQL sozinho

1 – Sumários estatísticos

2 – Análise de sobrevivência

3 – Deep learning.

4 – NLP: busca textual com word-embeddings