

BigDataCovid

Repositório: <https://github.com/illiamw/BigDataCovid>

Entrega de projeto para avaliação na disciplina Mineração a partir de Grandes Bases de Dados - SCC0244 junto ao docente **Caetano Traina Jr.** e monitor **Erikson Júlio de Aguiar**, Universidade de São Paulo - USP.

Composição do repositório

Analise.pdf: Resolução das questões propostas para a aplicação dos conceitos da disciplina, com trechos de SQL, resultados de consultas, ilustrações e demais representações pertinentes.

./bases: Todas as bases analisadas acompanhando descrição de conteúdo em arquivo de texto (*descricao.csv*) como dicionário das bases utilizadas.

./scripts: Todos os comandos executados para compor a resolução das questões propostas no enunciado do projeto segundo a nomenclatura *NumeroQuestao* __ *NumeroItem* _Descrição.sql (por exemplo, 2_2EstatisticasEspecificas.sql).

./scripts/Resultados: Alguns resultados extensos que não convêm espaço no arquivo **Analise.pdf**, e seguindo a nomenclatura *NumeroQuestao* __ _*NumeroItem* _Descrição.sql (por exemplo, 2_2PositivosPorIdade.csv).

Reprodução e softwares envolvidos

O gerenciamento da base de dados foi realizadas no PostgreSQL versão 14, e as bases de dados carregadas e tratadas, são originadas do repositório FAPESP COVID-19 DataSharing/BR contendo conjunto de bases de dados de instituições de saúde referente a complicações de covid, no qual o hospital analisado foi **Hospital Sírio-Libanês**. E para questões específicas foram utilizadas as tabelas **exames** dos demais hospitais: Beneficência Portuguesa de São Paulo, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, Grupo Fleury, Hospital Israelita Albert Einstein e Hospital Sírio-Libanês.

Execução de UPDATES e CREATES

É necessário executar os seguintes scripts na ordem, pois possuem modificações de tabelas

1_1inicializacao.sql

1_2AvaliacaoLimpeza.sql

5_Hemograma.sql

8_replaceCovid.sql