**Universidade Tecnológica Federal do Paraná - Campus Dois Vizinhos**

**Especialização em Ciência de Dados**

**Projeto Integrador**

**Alexandre Hannisch**

**Bruno Faustino Amorim**

**Leonardo Garcia**

**Nicolas Soffi**

**Dois Vizinhos - PR**

**2021**

# Administração e Gerenciamento de Banco de Dados

## Tabelas

Text

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Usando a base de dados de Casos de Covid utilizada no módulo do curso, criamos três tabelas. casos\_covid: armazenamento dos casos de covid

localizacao: registros das cidades

relatorio: exibi a quantidade de casos confirmados por cidade

## Índices

City: Full Text Search

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Aqui usamos o Gin, um tipo de índice especializado em buscas textuais de Full Text Search

State: Índice Bitmap

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Criamos um índice de bitmap para as consultas de baixa seletividade como o campo Estado. Nesse caso o índice bitmap é recomendado atributos discretos com até 7 tipos.

Order for Place: Índice Hash

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Parecido com o caso anterior quando utilizamos bitmap. Entretanto agora temos cerca de 100 tipos de order\_for\_place, nesse cenário recomenda-se a utilização de indice hash

Indice FK

Graphical user interface, text, application, email

Description automatically generated

Vemos que a consulta na primeira tabela está otimizada, foi feito o Index Scan devido ao indice da PK. Já na segunda tabela, foi feito o Seq Scan, ou seja, o banco está percorrendo a tabela inteira para tentar achar o registro indicado (idcasos\_covid = 100)

Graphical user interface, text, application, email

Description automatically generated

Criamos o indice na chave estrangeira para otimizar a leitura em disco e vemos que o tempo de busca caiu drasticamente

Indice Código IBGE

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Criaremos tambem mais um indice para consultas através do código do IBGE (city\_ibge\_code).

No código do IBGE os 2 primeiros numeros representam o código do estado. Por exemplo, todos os municipios

do estado do Rio de Janeiro começam com 33, assim poderiamos consultar todos os municipios pertencentes ao

Rio de Janeiro usando uma consulta como LIKE '33%'. Para isso, iremos considera-lo como um campo texto

para fazer Full Text Search.

## Trigger

Graphical user interface, text

Description automatically generated

Graphical user interface, application, table

Description automatically generated

Text

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Após a execução vemos que um automaticamente um novo registro foi incluído na tabela de relatório.

## Usuários

Simularemos a criação de um cenários com três perfis:

gestores: são responsaveis por gerenciar a inclusão e atualização de casos de covid.

analistas: podem somente ter permissão de leitura, não podendo fazer qualquer tipo de atualização ou inserção de novos registros.

administradores: todos os privilégios

Usuário: Analista

Graphical user interface, application

Description automatically generated

Fazemos login com o usuário Analista

Text, table

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

O usuário Analista consegue ler os dados nas não ter permissão para alterar ou inserir.

Usuário: Gestor

Graphical user interface, text, application, email

Description automatically generated

Fazemos login com o usuário Gestor

Table

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

O usuário Analista consegue ler, alterar e inserir.

Usuário: Administrador

Graphical user interface, text, application, email

Description automatically generated

Fazemos login com o usuário Administrador

Table

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

O usuário Administrador possui todos os privilégios.

## Transações

Como estamos tratando de casos de covid é essencial que outros usuários leiam apenas transações que forem efetivadas (commit)

Graphical user interface, text, application

Description automatically generated

Atendendo ao requisito, abrimos a transação com o usuário administrador e inserimos um novo registro.

Graphical user interface, text, application

Description automatically generated

Como a transação ainda não foi efetivada (“commitada”) o registro de id 401 não aparece para o usuário analista.

Graphical user interface, text, application, email

Description automatically generated

Efetivamos a transação do usuário administrador.

Graphical user interface, text, application

Description automatically generated

Agora o dado efetivado torna-se visível para o usuário analista. Fazendo com que todos os usuários do grupo analista visualizem somente os dados que realmente foram efetivados dentro do database.

# Preparação Pedagógica

|  |
| --- |
| **PLANO DE ENSINO** |

|  |  |
| --- | --- |
| **CURSO** | Treinamento análise de dados sobre os casos covid |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **DISCIPLINA/UNIDADE CURRICULAR** | **CÓDIGO** | **PERÍODO** |  | **CARGA HORÁRIA (qtde. aulas)** |
| **Análise da dados de casos da Covid com Python** | **AN86R** |  |  | **12** |
|  |

|  |  |
| --- | --- |
| **PRÉ-REQUISITO** | Nenhum |

|  |
| --- |
| **OBJETIVOS**  Banco de dados PostgreSql. Índices. Permissões de acesso. Analisar os dados de casos de covid em 5 cidades, a fim de obter e comparar informações entre as cidades. |

|  |
| --- |
| **EMENTA**  Criação de índices no banco de dados PostgreSql. Gerenciar permissões de acesso ao banco de dados. Ler dados dos casos de covid utilizando python. Preparação e agrupamento dos dados. Exibição de gráficos utilizando os dados. |

|  |  |  |
| --- | --- | --- |
| **CONTEÚDO PROGRAMÁTICO** | | |
| **ITEM** | **EMENTA** | **CONTEÚDO** |
| 1 | Índices no PostgreSql | Aplicação de índices em tabelas do banco de dados |
| 2 | Gerenciar permissões | Gerenciamento de permissões de grupos e usuários |
| 3 | Biblioteca psycopg2 | Acesso e manipulação do banco de dados utilizando python |
| 4 | Biblioteca Pandas | Manipulação de DataFrames com python |
| 5 | Biblioteca Matplotlib | Exibição de gráficos utilizando python. |

|  |  |
| --- | --- |
| **PROFESSORES** | **TURMA** |
| Nicolas Soffi | N07 |

|  |  |  |
| --- | --- | --- |
| **PROGRAMAÇÃO E CONTEÚDOS DAS AULAS (PREVISÃO)** | | |
| **Dia/Mês ou Semana** | **Conteúdo das Aulas** | **Número de Aulas** |
| 08/07 | Apresentação e criação de índices com PostgreSql | 2 |
| 13/07 | Criação de índices e gerenciamento de permissões no PostgreSql | 2 |
| 15/07 | Utilização e manipulação do PostgreSql com Python | 2 |
| 20/07 | Manipulação de dados com Python usando o Pandas | 2 |
| 22/07 | Criação e exibição de gráficos com Python | 2 |
| 27/07 | Atividade | 2 |

|  |
| --- |
| **PROCEDIMENTOS DE ENSINO** |
| **AULAS TEÓRICAS** |
| As aulas serão realizadas de forma interativa, com discussões sobre o tema. Os exemplos serão feitos utilizando recursos como computador e o código será testado com os alunos. |
| **AULAS PRÁTICAS** |
| Treinamento do projeto de análise dos dados de casos de covid por meio da criação de código utilizando a linguagem de programação python e o banco de dados PostgreSql. Realização de atividades de fixação em grupo, com a orientação do professor. |
| **PROCEDIMENTOS DE AVALIAÇÃO** |
| Nota final = nota da atividade em grupo  A nota final do treinamento será obtida pelo trabalho realizado em grupo.  O trabalho tem por objetivo a fixação do treinamento realizado. |

|  |
| --- |
| **REFERÊNCIAS** |
| **Referências Básicas:**  Design patterns: Elements of reusable object-oriented software,  Erich Gamma, Richard Helm, Ralph Johnson, John Vlissides, Addison-Wesley, 1995, ISBN 0-201-63361-2. |
| **Referências Complementares:**  John Hunt. A Beginners Guide to Python 3 Programming. Undergraduate Topics in ComputerScience. Springer, 2019. |

|  |
| --- |
| **ORIENTAÇÕES GERAIS** |
| Não haverá reprovações, visto que o objetivo deste treinamento é capacitar os alunos a analisarem os casos de covid. |

|  |
| --- |
| **PLANO DE AULA** |

|  |  |
| --- | --- |
| **IDENTIFICAÇÃO** | |
| **Disciplina:** | Análise da dados de casos da Covid com Python |
| **Professor:** | Nicolas Soffi |

|  |
| --- |
| **TEMA DA AULA** |
| Manipulação de dados utilizando a biblioteca Pandas com Python. |

|  |
| --- |
| **PRÉ-REQUISITOS** |
| Conhecimentos básicos da linguagem Python. |

|  |
| --- |
| **OBJETIVOS** |
| Ao final da aula, espera-se que os alunos tenham aprendido como manipular dados utilizando o Python. |

|  |
| --- |
| **CONTEÚDO PROGRAMÁTICO - APRESENTAÇÃO** |
| Apresentação da biblioteca Pandas;  Filtragem de dados;  Agrupamentos;  Seleção de métricas;  Criação de novas métricas; |

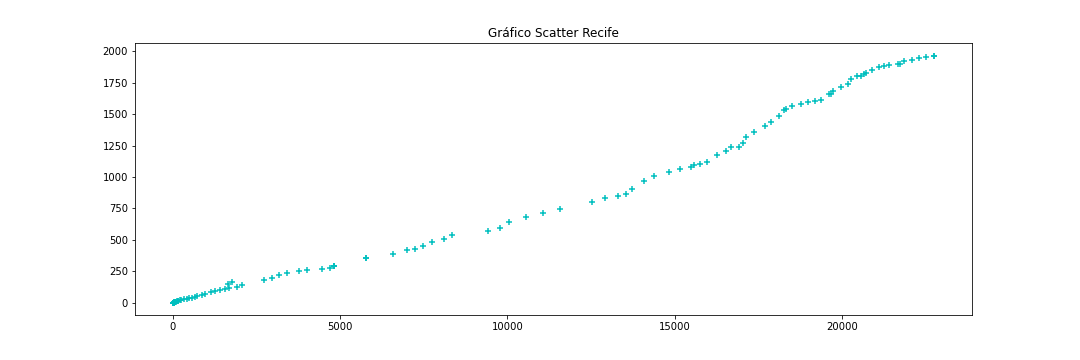
|  |
| --- |
| **DESENVOLVIMENTO DO TEMA** |
| A aula será realizada em passos, levando em consideração o conteúdo programático. Será feita discussões e demonstrações práticas em cada passo dos conteúdos. |

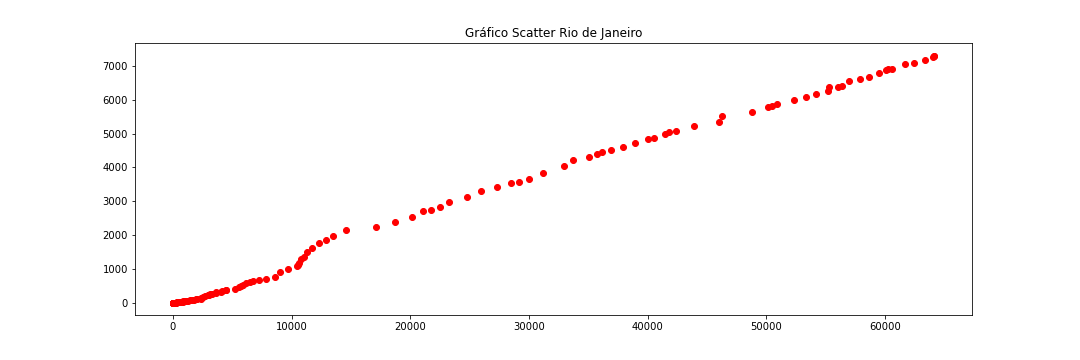
|  |
| --- |
| **RECURSOS DIDÁTICOS** |
| Será utilizado computadores para a execução do código a fim de demonstrações práticas do tema abordado. |

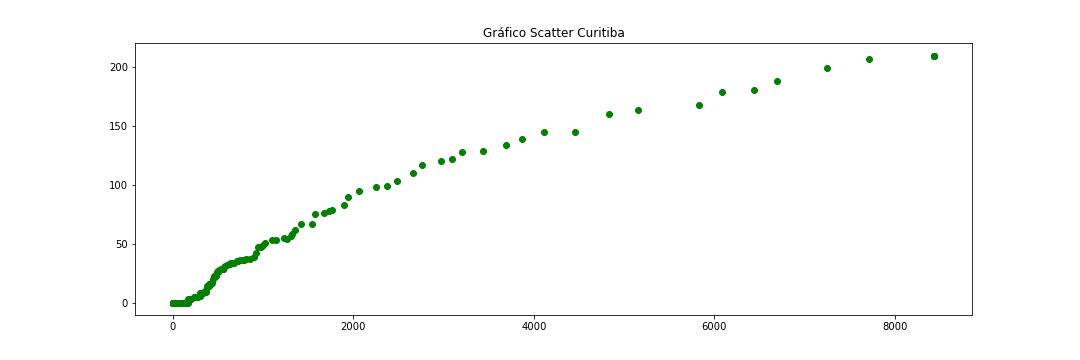
|  |
| --- |
| **AVALIAÇÃO** |
| Nota final = nota da atividade final em grupo  A nota final do treinamento será obtida pelo trabalho realizado em grupo.  O trabalho tem por objetivo a fixação do treinamento realizado. |

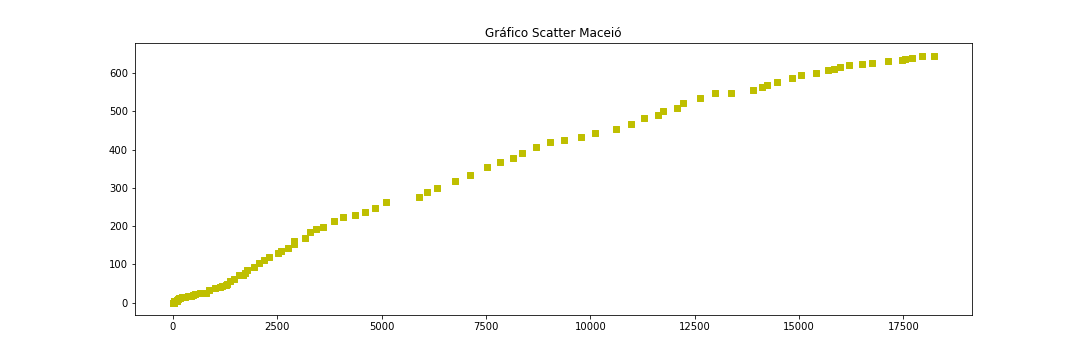
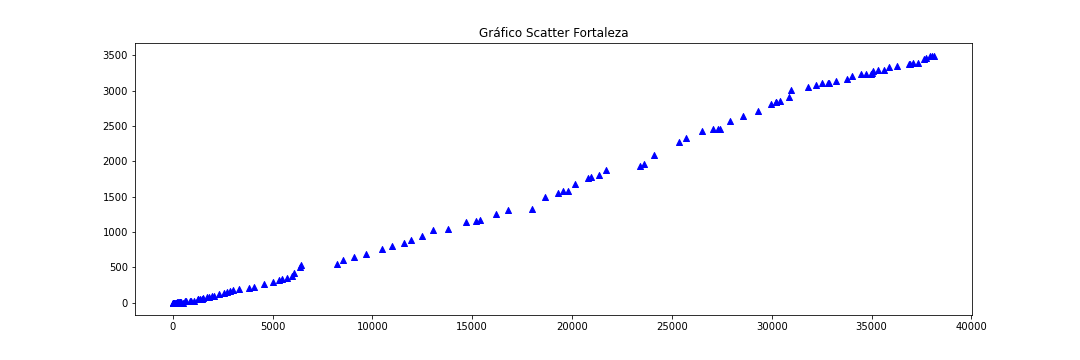
|  |
| --- |
| **BIBLIOGRAFIA** |
| Python para análise de dados,  Tratamento de dados com Pandas, NumPy e IPython [Wes McKinney](https://www.google.com.br/search?hl=pt-BR&gbpv=1&printsec=frontcover&q=inauthor:%22Wes+McKinney%22&tbm=bks&sa=X&ved=2ahUKEwiUo_eysubxAhUNrJUCHSPBBEMQmxMoADAIegQIEhAC), 2011, ISBN 9788575226476. |

# Linguagens de Programação para Ciência de Dados







Chart, bar chart

Description automatically generatedA piece of paper with writing on it

Description automatically generated with medium confidence

Graphical user interface, text, application, email

Description automatically generated

**Análise dos gráficos**

Os gráficos de dispersões criados verificam se a correlação entre o aumento do número de casos confirmados com o aumento do número de mortes.

Todas as cidades selecionadas apresentam um gráfico linear quase que perfeito. Os gráficos dão indícios que a uma correlação entre o aumento do número de casos confirmados com o aumento de mortos. Algumas situações podem explicar essa correlação. Pode ocorrer dos hospitais ficarem cheios podendo faltar leitos para pessoas que estão com casos graves, falta de equipamentos de respiração, além disso pode ter um aumento da demanda para a equipe de saúde que precisara ser redimensionada para atender todos os pacientes.

No gráfico de barras conseguimos ver a volumetria do número total de novas mortes.

Dentre as cidades selecionadas Rio de Janeiro é a cidade com o maior número de novas mortes o dobro da segunda maior que é Fortaleza. Dentre as cidades do Brasil com a maior população o Rio de Janeiro está em segundo lugar ficando atrás apenas de São Paulo, o que justifica o alto número de novas mortes, pois a Covid 19 quanto maior o contato e aglomeração maior é o contágio. Fortaleza tem o segundo maior índice de novas mortes, podemos ver que quanto maior a população da cidade maior é o número de mortos.

No gráfico onde se compartilha o eixo x que é a data, observamos que Curitiba ao manter por mais tempo o distanciamento social vemos que o aumento súbito de mortes ocorre no final do gráfico enquanto cidades que não mantiveram um distanciamento social firme como Fortaleza e Recife tiveram esse aumento antes, já Maceió e Rio de Janeiro foram mais medianos no comparativo das 5 cidades.