Tarefa 4 - Correção tarefa 3

Aluna: Letícia Trein Medeiros

Avaliação do trabalho de Vanessa Ysla - Ocorrências atendidas pela Guarda Municipal de Curitiba.

1) Acesse o pdf de algum colega que ainda não foi avaliado, e avalie a apresentação.

a. **Que detalhes pedidos na Tarefa 03 que estão faltando ou não estão claros?**

As demandas da tarefa 3 (descrição dos dados, quantidades de tuplas, problemas encontrados na inserção, cinco perguntas e as respostas) foram cumpridas

b. **A tabela está criada no servidor e com dados?**

Sim e o caminho está presente na apresentação.

c. **Avalie todos os parâmetros, os SQLs, os resultados, e dê uma nota de 0 a 10.**

Nota inicial: Em 'anexo' encontra-se os slides originais com as anotações especificas;

De forma geral o trabalho causou uma impressão bem positiva, apresentando todas as demandas anteriores atendidas; os slides apresentam a linguagem clara (com uma exceção apontada no slide 2), visivelmente organizado e com informações bem distribuídas; as Querys apresentadas entragam corretamente o resultado esperado; Outros pontos positivos de destaque são a complexidade de algumas Querys (utilizando agrupamento, ordenação, seleção com extração em único input por exemplo); Já, sobre os pontos negativos, destaca a ausência do processo inicial de limpeza e criação da tabela e uma das questões (4) levantadas apresentou um pequeno problema no output.; Contudo de forma geral, perguntas elaboradas se mostram relevantes e interessantes.

Nota: 9/10

d. **Que contribuições este tipo de dado pode trazer para a cidade?**

Os dados utilizados podem ser muito úteis para os órgãos de administração e organização da Guarda Municipal de Curitiba, uma vez que ela abrange cada ocorrência registrada, incluindo, por exemplo, a descrição e detalhes do chamado e logradouro, possuindo informações relevantes e especificas; além disso, os atributos temporais podem ser uteis para o planejamento e prevenção de ocorrências conforme as demandas já conhecidas.

2) Considerando as funções GIS vistas em aula, cite ao menos três perguntas diferentes, com três funções diferentes que poderiam ser feitas com este dado. Outras tabelas poderiam ser correlacionadas para as perguntas.

Uma vez que a base de dados atualmente não possui coluna Geom, para conseguir aplicar as funções GIS, de forma teórica, para a elaboração das perguntas será considerado que foi georeferenciado os logradouros (schema arruamento), bairros e regionais (schema limites_legais); também considere que foram cruzados os dados das tabelas de parques e praças (schema arruamento)

Perguntas possíveis:

1. Qual é a distância média entre as ocorrências e a praças mais próximas.
2. Quantas ocorrências aconteceram em regiões que fazem fronteira com o bairro com maior número de ocorrências?
3. Quais logradouros são completamente cobertos por zonas de alta incidência de ocorrências?

Anexo: Avaliação individual por slide

CDA - T4: Banco de Dados - Avançado
GLADYS VANESSA YSLA RAMIREZ

Sistema Sigesguarda

Ocorrências cadastradas atendidas pela Guarda Municipal de Curitiba

Descrição do dataset

Base de dados em formato ".csv", contendo os dados das ocorrências cadastradas no sistema Sigesguarda atendidas pela Guarda Municipal de Curitiba, tendo como objetivo a vigilância das ações da Guarda Municipal no município.

Pontos Positivos

- Informações sobre dataset original
- Detalhamento cobertura dados
- Informações de limpeza de dados
- Caminho para arquivo original para download

Cobertura dos dados:

- Espectro temporal: De 2022 até o momento da extração (Set 2024).
- Frequência de atualização: Mensal.
- Cada linha do dataset representa um único registro realizado durante esse período.

Dicionário de dados

https://mid.curitiba.pr.gov.br/dadosabertos/Sigesguarda/2024-10-01_sigesguarda_-_Base_de_Dados.csv

Total de colunas: 34

Ponto negativo:

Frase um pouco confusa

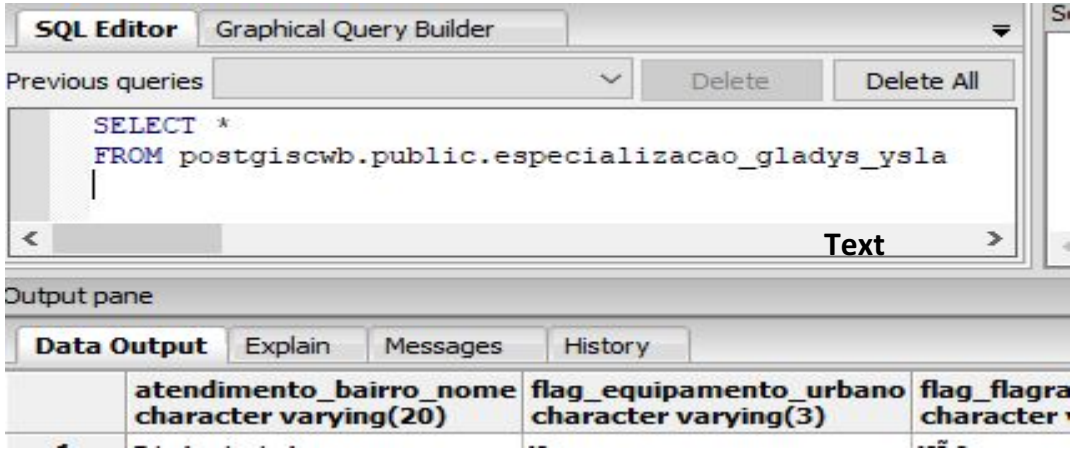
- Acharam-se muitas colunas com dados nulos (Ex: de 79168 células apenas foram preenchidas 53, 3 ou nenhuma célula). Portanto essas colunas foram não levadas para a tabela.
- Outras colunas tinham dados que já estavam contidos em outras colunas (Ex: Os dados das colunas OCORRENCIA_ANO e OCORRENCIA_MES já estão consideradas na coluna OCORRENCIA_DATA.

Foi criada a tabela “especializacao_gladys_ysla”

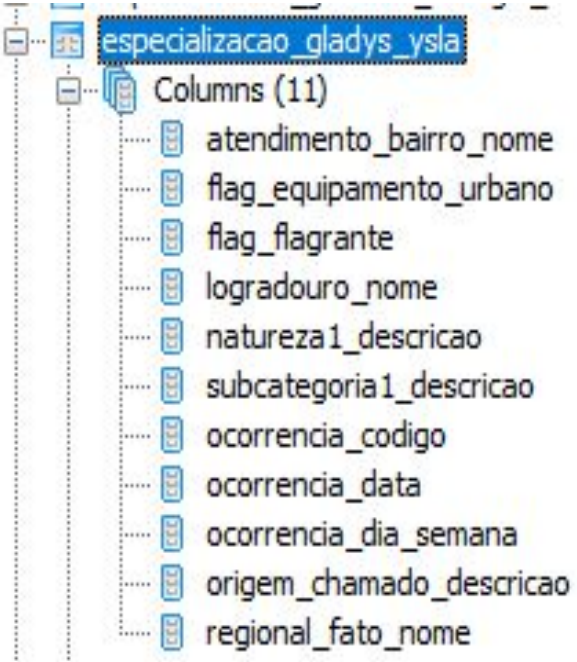
localizada em “postgiscwb.public.especializacao_gladys_ysla”

Pontos positivos:

- Nome tabela
- Caminho Servidor
-Colunas finais
-Três colunas para georeferenciamento



Finalmente, foram consideradas **11 colunas** para a criação da tabela



Dificuldades para a inserção dos dados na tabela “especializacao_gladys_ysla”

1. Formato dos dados na base original.

Foi feita a mudança do formato da coluna OCORRENCIA_DATA, para deixar no formato certo.

```
sigesguarda['OCORRENCIA_DATA'] = pd.to_datetime(sigesguarda['OCORRENCIA_DATA'])
```

Tento una OCORRENCIA_HORA, mas não deu certo, já qu o ele adicionava a data de quando está-se aplica i considerada para a construção da tabela.

OCORRENCIA_HORA
2024-10-10 13:12:00

Pontos positivos:

- Código de correção do formato da data
- Motivo da retirada da coluna Ocorrencia_Hora

Pontos negativos:

- Caberia aqui colocar os scripts de limpeza descritos anteriormente
- Poderia ter o script da criação da tabela

2.

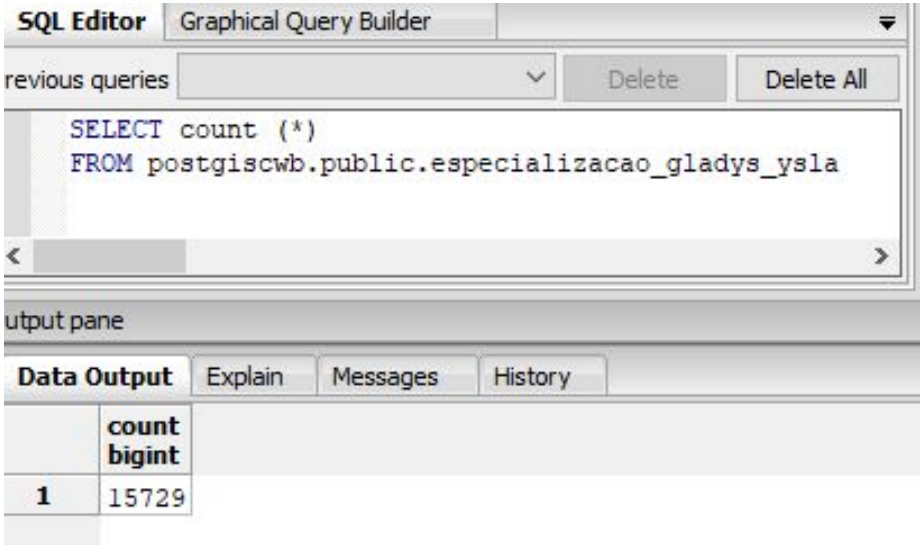
Foi u no chave primária. No momento da inserção, o sis erro, já que achava células duplicadas. Se procedeu a descartar essas tuplas, já que ao colocar a coluna OCORRENCIA_CODIGO como chave primária, os valores de cada célula devem ser únicos.

Pontos positivos:

- Query de contagem correta
- Output com número de tuplas >10k

Quantidade de tuplas:

15729



1. Quantas ocorrências há por dia da semana?

Previous queries

```
SELECT ocorrencia_dia_semana, COUNT(*) AS total_ocorrencias
FROM public.especializacao_gladys_ysla
GROUP BY ocorrencia_dia_semana
ORDER BY total_ocorrencias DESC;
```

Output pane

Data Output | Explain | Messages | History

	ocorrencia_dia_semana character varying(15)	total_ocorrencias bigint
1	Quinta	2514
2	Sexta	2362
3	Quarta	2321
4	Terça	2317
5	Sábado	2167
6	Segunda	2141
7	Domingo	1907

- Destaca-se a **quinta** como o dia de maiores ocorrências atendidas.

Pontos Positivos:

- Utilização de agrupamento, orgenação e função contagem

- Relevancia da pergunta: Pode ser útil para planejamento semanal de escala dos profissionais envolvidos

2. Qual é o bairro com mais ocorrências?

Previous queries

```
SELECT atendimento_bairro_nome, COUNT(*) AS total_ocorrencias
FROM public.especializacao_gladys_ysla
GROUP BY atendimento_bairro_nome
ORDER BY total_ocorrencias DESC
LIMIT 1;
```

Output pane

Data Output | Explain | Messages | History

	atendimento_bairro_nome character varying(20)	total_ocorrencias bigint
1	Centro	1619

- Observa-se que o **Centro** é o bairro com maior quantidade de ocorrências cadastradas.

Pontos Positivos:

- Utilização de funções de agrupamento, ordenação, contagem e limite
- Relevancia da pergunta: Entender qual bairro precise de uma demanda maior de esforço policial e vigilancia

3. Quais ocorrências foram feitas no logradouro 'Rua Mariano Torres'?

Pontos Positivos:

- Utilização de função Where, seleção especifica de colunas de interesse

- Relevancia: registro temporal e físico de cada ocorrencia para rápida consulta.

Previous queries

```
SELECT ocorrencia_codigo, ocorrencia_data, logradouro_nome
FROM public.especializacao_gladys_ysla
WHERE logradouro_nome = 'Rua Mariano Torres';
```

Output pane

Data Output | Explain | Messages | History

	ocorrencia_codigo character varying(15)	ocorrencia_data date	logradouro_nome character varying(70)
1	01839/2022	2022-12-28	Rua Mariano Torres
2	01947/2022	2022-12-29	Rua Mariano Torres
3	00274/2023	2023-01-04	Rua Mariano Torres
4	01364/2023	2023-01-14	Rua Mariano Torres
5	03013/2023	2023-01-27	Rua Mariano Torres
6	04114/2023	2023-02-04	Rua Mariano Torres
7	04517/2023	2023-02-07	Rua Mariano Torres

4. Qual é o total de ocorrências em cada mês do ano?

Previous queries

```
SELECT EXTRACT(MONTH FROM ocorrencia_data) AS mes, COUNT(*) AS total_ocorrencias
FROM public.especializacao_gladys_ysla
GROUP BY EXTRACT(MONTH FROM ocorrencia_data)
ORDER BY mes;
```

Output pane

Data Output | Explain | Messages | History

	mes double precision	total_ocorrencias bigint
1	1	2750
2	2	2994
3	3	4161
4	4	3691
5	5	561
6	12	1572

Pontos Positivos:

- Seleção especifica do mês dentro de uma variável de data, contagem de ocorrencia com agrupamento seletivo e ordenado.

- Relevancia: Ter um panorama temporal

Pontos negativos:

- A filtragem retornou apenas metade dos meses, potencialmente houve uma falha da amostragem pela limpeza dos dados
- Talvez essa pergunta não possa ser bem responda com a base no estado atual

- Destaca-se **março** como o mês de maiores ocorrências cadastradas

5. Quantas ocorrências foram registradas para cada descrição da Natureza 1 (descrição do chamado) e qual foi o bairro com o maior número dessas ocorrências?

Previous queries | Delete | Delete All

```
SELECT naturezal_descricao, atendimento_bairro_nome, COUNT(*) AS total_ocorrencias
FROM public.especializacao_gladys_ysla
WHERE naturezal_descricao IS NOT NULL
GROUP BY naturezal_descricao, atendimento_bairro_nome
ORDER BY naturezal_descricao, total_ocorrencias DESC;
```

Output pane

Data Output | Explain | Messages | History

	natureza1_descricao character varying(100)	atendimento_bairro_nome character varying(20)	to bi
1	Achado	Centro Cívico	
2	Achado	Centro	
3	Achado	Boqueirão	
4	Achado	Cidade Industrial	
5	Achado	Atuba	

Pontos Positivos:

- Seleção de colunas especificas, contagem de ocorrencias, seleção sem valores nulos, dois agrupamentos e ordenação com dois paramentros.

Relevancia: Pode entender melhor o perfil de ocorrencias de cada bairro e melhor direcionamento de esforços policiais adequados ao tipo de demanda