**Introdução/Visão geral**

Este documento foi projetado para ajudá-lo a entender como você pode usar um banco de dados do Azure para a instância PostgreSQL para permitir o cenário de negócios de seleção do site. Vamos contar com as capacidades geoespaciais robustas dentro postgresql para isso.

A seleção do local é uma função importante em uma variedade de indústrias que incluem o varejo, as telecomunicações, e o transporte/logística. Neste exemplo,vamos executar o processo de seleção do site a partir do ponto de vista de uma empresa de scooter. A empresa scooter deve decidir onde colocar sua primeira instalação de carregamento scooter em uma cidade para garantir que scooters são cobrados e prontos para ir todas as manhãs.

Uma vez que a empresa scooter implanta suas scooters, eles podem usar dados de localização das scooters para tomar decisões sobre locais subsequentes do site de carregamento. Mas fou a seleção inicial do site, eles estarão contando com informações sobre o seguinte:

* População (maior é melhor)
* Distância das estações de metrô (maior é melhor)
* Tempos médios de deslocamento (maior é melhor)
* Taxa de criminalidade (menor é melhor)

Em suma, a empresa scooter quer atender maior número de potenciais clientes com mais viagens mais longe das estações de metrô em áreas com menor criminalidade. Eles gostariam de situar suas estações de carregamento em cantos / cruzamentos não associados com rodovias para que eles possam tomar em scooters de duas ruas em vez de um.

Usaremos Nova York (NYC) para este exemplo porque temos fácil acesso a todos os dados necessários para ilustrar esse cenário para Nova York. Por favor, note que este passo a passo é fortemente inspirado e adaptado de materiais encontrados no [PostGIS Introductory Workshop](https://postgis.net/workshops/postgis-intro/) que foram publicados uma [atribuição-ShareAlike 3.0 Estados Unidos (CC BY-SA 3.0 EUA)](https://creativecommons.org/licenses/by-sa/3.0/us/)  [licença](https://creativecommons.org/licenses/by-sa/3.0/us/).

**Ferramentas e software necessários**

Para executar este passo a passo, você vai precisar do seguinte:

* Uma assinatura do Azure (opção gratuita [aqui)](https://azure.microsoft.com/free/free-account-faq/) e um banco de [dados azure para instância PostgreSQL](https://ms.portal.azure.com/#create/hub) (porfavor, note que você pode usar um único servidor e nível de serviço basic com 1 ou 2 vCores e 50GB de armazenamento - esta configuração consumirá cerca de US $ 2 ou menos por dia a partir de ou sua assinatura de avaliação paga ou gratuita se você deixá-lo funcionando 24 horas por dia).
* [PostGIS](http://postgis.net/install/) - as ferramentas do cliente incluem o PostGIS 2.0 Shapefile e DBF Loader Exporter que usaremos para importar dados geoespaciais relacionados a Nova York para nossa instância PostgreSQL
* [QGIS](https://qgis.org/en/site/forusers/download.html) - um excelente conjunto de ferramentas para visualizar dados geoespaciais do PostgreSQL e outros bancos de dados
* [pgAdmin](https://www.pgadmin.org/download/) - uma ferramenta de administração robusta para PostgreSQL

**Criando e povoando seu banco de dados nyc**

Uma vez que seu banco de dados do Azure para a instância PostgreSQL for criado, você precisará criar um banco de dados PostgreSQL nessa instância. Use o pgAdmin para se conectar ao seu Banco de Dados Do Azure para a instância PostgreSQL usando a corda de conexão do Azure e o nome de usuário/senha que você especificou durante a configuração de sua instância. Uma vez que o pgAdmin esteja conectado ao seu banco de dados do Azure para a instância PostgreSQL, clique direito no nome de sua instância, selecione Criar > Bancode dados... e criar um banco de dados chamado nyc usando os padrões.

Uma vez que seu banco de dados nyc é criado, você vai usar pgAdmin seguinte para garantir que seu banco de dados nyc tem a extensão postgis instalado. Clique à direita em seu bancode dados recém-criado nyc, selecione Criar > Extensão... e selecione postgis do campo de nome na janela pop-up "Criar - Extensão".

Agora que as capacidades geoespaciais completas são habilitadas, você usará o aplicativo PostGIS 2.0 Shapefile e DBF Loader Exporter para carregar os quatro elementos de dados geoespaciais para este exemplo no banco de dadosde Nova York: ruas, estações de metrô, setores censitários e homicídios. Os dados de homicídios são nosso proxy para a taxa de criminalidade e usaremos os setores censitários como uma área geográfica relativamente pequena como nossa granularidade de alto nível para a seleção do local. Por favor, note que você precisará descompactar esses dados geoespaciais antes que você possa importá-los para postgresql. PostGIS tem uma excelente explicação de como usar o PostGIS 2.0 Shapefile e DBF Loader Exportador  [aqui](https://postgis.net/workshops/postgis-intro/loading_data.html).

Como você viu no link anterior, você precisará fornecer suas credenciais de login para seu banco de dados azure para instância PostgreSQL para o Exportador de Shapefile e Carregador DBF PostGIS 2.0. Seu nome de usuário e senha serão o que você especificou quando criou seu banco de dados do Azure para a instância PostgreSQL. Seu servidor será o ponto final público associado ao seu banco de dados azure para a instância PostgreSQL - por exemplo, your\_instance\_name.postgres.database.azure.com. O porto será 5432 e seu nome de banco de dados será nyc.

Depois de ter especificado suas credenciais de login, você precisará adicionar o . shp arquivos do nosso conjunto de dados geoespacial que você extraído do arquivo zip contendo esses dados. Clique no botão "Adicionar Arquivo", adicione os quatro . shp arquivos e configurar o importador para usar um valor de 26918 para o campo SRID de cada uma das entidades geoespaciais.

You vai saber que você fez isso corretamente quando a janela em seu PostGIS 2.0 Shapefile e DBF Loader Exportador se parece com este:

Um screenshot de um telefone de pilha

Descrição gerada automaticamente

Quando você configurou isso corretamente, clique no botão "Importação" para importar essas quatro entidades geoespaciais em seu banco de dados de Nova York.

Cada uma dessas quatro entidades/tabelas tem elementos geoespaciais na coluna geom e elementos relacionais "tradicionais" em outras colunas. Os characteristics geoespaciaisde cada entidade podem ser um ponto (no caso das estações de metrô e homicídios), uma/múltipla linhas(no caso das ruas), ou um/vários polígonos (no caso dos setores censitários).

Detalhes sobre essas entidades/tabelas e algumas das colunas úteis nele são descritos no dicionário de dados abaixo.

## Dicionário de Dados Geoespaciais

Colunas úteis na tabela "nyc\_streets" :

|  |  |
| --- | --- |
| Nalma | Nome comum da rua |
| T ype Type | Que tipo de rua é essa? |
| Geom G eom | Geometria multilinestring da rua |

Colunas úteis na tabela "nyc\_subway\_stations" :

|  |  |
| --- | --- |
| Nome | Nome comum da estação |
| Geom | Geometria do ponto da rua |

Colunas úteis na tabela "nyc\_homicides" :

|  |  |
| --- | --- |
| Arma | Arma usada no homicídio |
| Geom | Geometria ponto do homicídio |

Colunas úteis na tabela "nyc\_census\_dointervalo": MVA

|  |  |
| --- | --- |
| tratoid | Um códigon 11-digit que identifica exclusivamente todos os setores censitários. Por exemplo:36005000100 |
| Geom | Limite multipolygon do trato |

**Enriquecer nossoconjunto** de dados geoespaciais **com dados socioeconômicos**

Nossos dados geoespaciais já têm alguns elementos relacionais, mas precisamos combinar esses dados com dados relacionais adicionais para alcançar nosso objetivo comercial.

Para esses dados adicionais, usaremos o arquivo chamado nyc\_census\_sociodata.sql para importar dados em nosso banco de dados de Nova York. Dentro do pgAdmin,clique direito no banco de dados de Nova York e selecione ferramenta de consulta. Colar o conteúdo do nosso arquivo nyc\_census\_sociodata.sql nesta nova janela de consulta e, em seguida, executar a consulta, quer pressionando F5 em seu teclado ou clicando no botão relâmpago pequeno na barra de ferramentas.

Este script cria uma nova tabela relacional com atributos socioeconômicos para cada setor censitário e inclui uma coluna chamada tractid que atua como uma chave estrangeira para a tabela nyc\_census\_tracts. Além disso, este script cria três visualizações que juntos formam um algoritmo que pontua cada setor censitário pela maioria das características com as quais nos preocupamos: população, tempo médio de trânsito e taxa de criminalidade. Vamos definir essas características e pontuações de lado por um tempo para lidar com a quarta característica (distância das estações de metrô) em primeiro lugar.

**Visualizando nossos dados geoespaciais**

Para começar a visualizar nossos dados geoespaciais, abra o aplicativo QGIS que você baixou e instalou anteriormente. Dentro do QGIS, inicie um novo projeto,clique direito no nó PostGIS na área do navegador e selecione Nova Conexão...

Uma captura de tela de um post de mídia social

Descrição gerada automaticamente

Entrada na janela resultante (campos destacados abaixo) as credenciais e informações apropriadas para o seu banco de dados azure para a instância PostgreSQL. Você provavelmente vai querer usar um nome amigável,como mostradoabaixo. Por favor, note que o SSL é obrigado a se conectar ao seu banco de dados do Azure para a instância PostgreSQL.

Um screenshot de um computador

Descrição gerada automaticamente

Depois de ter autenticado, você vai ver asnossas quatro entidades geoespaciais / tabelas no banco de dados nyc: setores censitários, homicídios, ruase estações demetrô. Selecione todos os quatro,clique direito em qualquer um deles, e selecione "Adicionar camadas selecionadas ao projeto", como mostrado abaixo.

Uma captura de tela de um post de mídia social

Descrição gerada automaticamente

Uma vez que os dados geoespaciais são carregados no projeto QGIS, podemos ativar/desligar camadas. A partir do padrão de ver todas as camadas na visualização resultante, desmarcar a caixa ao lado de cídos de homiliaparaque estejamos apenas vendo setores censitários, ruas e estações de metrô, como mostrado abaixo.

Um screenshot de um telefone de pilha

Descrição gerada automaticamente

**Manipular dados geoespaciais e revisualizar a saída**

Agora que nossa configuração inicial no QGIS está completa, realizaremos várias operações em pgAdmin. Você pode se lembrar que nosso cenário de negócios envolve a seleção de um local em um cruzamento de ruas para que nosso site tenha acesso a mais de uma rua de cada vez. Podemos criar a tabela de interseção executando um script na ferramenta de consulta pgAdmin. Em pgAdmin,clique direito em seu banco de dados nyc e selecionar ferramenta de consulta ... ter um lugar para colar/executar o seguinte roteiro:

CREATE TABLE public.nyc\_intersections

(

gid SERIAL PRIMARY KEY,

street1 character varying(200) COLLATE pg\_catalog."default",

street2 character varying(200) COLLATE pg\_catalog."default",

geom geometry NOT NULL

)

WITH (

OIDS = FALSE

)

TABLESPACE pg\_default;

/\* calculate all the intersections in NYC based on the geometries of the streets \*/

INSERT INTO nyc\_intersections (street1, street2, geom)

SELECT s1.name, s2.name, ST\_Intersection(s1.geom, s2.geom) AS intersection\_geom

FROM nyc\_streets AS s1

JOIN nyc\_streets AS s2

ON ST\_Intersects(s1.geom, s2.geom)

WHERE

s1.name IS NOT NULL AND

s2.name IS NOT NULL AND

s1.name <> s2.name

ORDER BY s1.name, s2.name;

/\* Remove duplicate intersections \*/

DELETE FROM nyc\_intersections

WHERE gid NOT IN (

SELECT gid FROM (

SELECT MIN(gid) as gid, geom

FROM nyc\_intersections

GROUP BY geom

) AS uniqueintersections);

Este script creates ponto geografias dos cruzamentos das linhas que são ruas. Porque nós igualmente estamos executando uma junção cartesiana entre dois exemplos da tabela das ruas, igualmente elimina interseções duplicadas - se nós temos a interseção da 1st Avenida e da rua de Allen, nós não precisamos igualmente a interseção da rua de Allen e da 1st Avenida.

Depois de criar seus cruzamentos, retorne ao QGIS. Para visualizar os cruzamentos recém-criados no QGIS, clique à direita na conexão PostGIS que você originalmente criou e seleciona "Atualização". Você deve ver uma nova entidade geoespacial chamada nyc\_intersections ao lado das quatro entidades originais que vimos antes. Clique à direita em nyc\_intersections eelegem "Adicionar Camada ao Projeto". Com cerca de 43.000 cruzamentos não duplicados, a visualização QGIS deve se encher de novos pontos que cobrem quase completamente as outras entidades geoespaciais.

Porque queremos considerar apenas cruzamentos seguros, não incluindo uma rodovia que são 500 metros ou mais de uma estação de metrô,podemos substituir o script pgAdmin anterior com o script a seguir.

/\* Remove from consideration intersections that are within 500 meters of a subway station \*/

DELETE FROM nyc\_intersections

WHERE gid IN(

SELECT nyc\_intersections.gid

FROM nyc\_intersections

JOIN nyc\_subway\_stations

ON ST\_DWithin(nyc\_intersections.geom, nyc\_subway\_stations.geom, 500)

);

/\* Delete those intersections that include one or more highways or highway links) \*/

DELETE FROM nyc\_intersections

WHERE gid IN (

SELECT nyc\_intersections.gid

FROM nyc\_intersections

JOIN (SELECT \* FROM nyc\_streets WHERE type IN ('motorway', 'motorway\_link')) nyc\_streets

ON ST\_DWithin(nyc\_intersections.geom, nyc\_streets.geom, 10)

);

Running este script em pgAdmin irá remover mais de 16.000 cruzamentos de consideração. Podemos ver o impacto dessa manipulação retornando ao QGIS e atualizando nossa visão dos cruzamentos. Para fazer isso, clique direito na lista de nyc\_intersections na janela de camadas do QGIS e selecione "Atualização da Camada SQL... "como mostrado abaixo.

Uma imagem de um mapa

Descrição gerada automaticamente

Clique no botão Atualização no canto inferior direito da janela db manager resultante para atualizar os dados de interseção,como mostrado abaixo:

Uma captura de tela de um post de mídia social

Descrição gerada automaticamente

Quando qgis acabamentos refrescante a camada de cruzamentos, você vai ver que agora há uma zona tampão em torno de paradas de metrô. No caso de as estações de metrô e os cruzamentos têm pontos de cor semelhantes associados a eles, tsua zonatampão torna-se facilmente visível se você alternar a visibilidade das estações de metrô.

**Manipulação final de dados**

Agora que removemos da consideração os cruzamentos que envolvem rodovias ou estão muito perto das estações de metrô, podemos voltar a reduzir nossas escolhas com base na população, tempos médios de deslocamento e taxa de criminalidade. Você pode se lembrar que uma das opiniões que criamos anteriormente pontuações setores censitários por essas variáveis - podemos ver o top 10 iDs do setor censitário, executando a seguinte consulta em pgAdmin:

SELECT tractid

FROM vw\_nyc\_site\_selection\_data

LIMIT 10

Os resultados desta consulta nos permitem construir uma consulta que exclui todas as interseções que não estão dentro ou perto desta lista top 10 de setores censitários:

/\* delete those intersections that are NOT WITHIN 500 METERS of the top 10 census tracts \*/

DELETE FROM nyc\_intersections

WHERE gid NOT IN (

SELECT nyc\_intersections.gid

FROM nyc\_intersections

JOIN nyc\_census\_tracts

ON ST\_DWithin(nyc\_intersections.geom, nyc\_census\_tracts.geom, 500)

WHERE tractid IN ('36005046201', '36061013600', '36061015400', '36061013800', '36061006600', '36061013400', '36061024500', '36061015700', '36061013200', '36081045500')

)

Depois de executar esta última consulta em pgAdmin,você pode atualizar a camada nyc\_intersections novamente no QGIS para ver que reduzimos nossas considerações significativamente,até pouco mais de350 cruzamentos que atendem aos nossos critérios. Pode não ser possível refinar automaticamente mais do que isso, porque teríamos de avaliar também a disponibilidade de imóveis, preços de aluguel / compra, etc. No entanto, temos muito rapidamente eliminado mais de 99% dos cruzamentos em Nova York através de uma combinação de consultas tradicionais e geoespaciais, como mostrado abaixo.

Um fim acima de um mapa

Descrição gerada automaticamente

Para mais detalhes sobre dados geoespaciais e consultas,consulte a oficina [introdução ao PostGIS](https://postgis.net/workshops/postgis-intro/) que inspirou e informou este passo a passo.