**始めに/概要**

このドキュメントは、Azure Database for PostgreSQL インスタンスを使用して立地選択のビジネス シナリオを有効にする方法を理解するのに役立ちます。この方法は、PostgreSQL内の堅牢な地理空間機能に依存します。

立地の選択は、小売、通信、輸送・物流など様々な業界で重要な機能です。この例では、スクーター会社の視点から立地選択のプロセスを実行します。スクーター会社は、スクーターが毎朝充電され、行く準備ができていることを確認するために、都市に最初のスクーター充電施設を置く場所を決定する必要があります。

スクーター会社がスクーターを配備すると、スクーターの位置データを使用して、その後の充電場所に関する意思決定を行うことができます。ただし、最初の立地の選択では、次の情報に依存します:

* 人口（多い方が良い）
* 地下鉄駅からの距離（遠い方が良い）
* 平均通勤時間（長い方が良い）
* 犯罪率（低い方が良い）

要するに、スクーター会社は、犯罪の少ない地域の地下鉄駅から遠く、通勤時間が長い潜在的な顧客により多くのサービスを提供したいと考えています。彼らは、彼らが1つではなく2つの通りからスクーターに乗ることができるように、高速道路に関連付けられていない曲がり角/交差点に充電ステーションを配置したいと考えています。

この例ではニューヨーク (NYC) を使用します。NYCであれば、このシナリオを説明するために必要なすべてのデータに簡単にアクセスできるからです。 このチュートリアルは、[Attribution-ShareAlike 3.0 United States (CC BY-SA 3.0 US) license](https://creativecommons.org/licenses/by-sa/3.0/us/)の下で公開されている [PostGIS Introductory Workshop](https://postgis.net/workshops/postgis-intro/)にある資料に強く触発され、適応しています。

**必要なツールとソフトウェア**

このウォークスルーを実行するには、以下が必要です:

* Azureのサブスクリプション（無償版は[ここ](https://azure.microsoft.com/free/free-account-faq/)）とAzure Database for PostgreSQLのインスタンス（1 または 2 個のvCPUと 50 GB のストレージを備えた単一のサーバーと基本サービスレベルを使用できること - このセットアップは、1 日 24 時間実行したままにしておくと、有料または無料の試用版サブスクリプションから 1 日あたり約 2 米ドル以下を消費します。）
* [PostGIS](http://postgis.net/install/) – クライアントツールにはPostGIS 2.0シェープファイルとDBFローダーエクスポータが含まれており、NYCに関連する地理空間データをPostgreSQLインスタンスにインポートするために使用します。
* [QGIS](https://qgis.org/en/site/forusers/download.html) - PostgreSQLや他のデータベースから地理空間データを視覚化するための優れたツールセット
* [pgAdmin](https://www.pgadmin.org/download/) – PostgreSQ のための堅牢な管理ツール

**NYCのデータベースを作成しデータを入れる**

Azure Database for PostgreSQL のインスタンスを作成したら、そのインスタンスに PostgreSQL データベースを作成する必要があります。pgAdmin を使用して、Azure からの接続文字列とインスタンスのセットアップ時に指定したユーザー名/パスワードを使用して、PostgreSQL インスタンスの Azure データベースに接続します。postgreSQL インスタンスの Azure データベースに pgAdmin が接続されたら、インスタンス名を右クリックし、[作成] > [データベース...デフォルトを使用して nyc という名前のデータベースを作成します。

nyc データベースが作成されたら、次に pgAdmin を使用して、nyc データベースに postgis 拡張がインストールされていることを確認します。新しく作成した nyc データベースを右クリックし、[作成] > [拡張機能] をクリックし、「作成 - 拡張」ポップアップウィンドウの「名前」フィールドから postgis を選択します。

完全な地理空間機能が有効になったので、PostGIS 2.0 シェープファイルおよび DBF ローダー エクスポータアプリケーションを使用して、この例の 4 つの地理空間データ要素 (道路、地下鉄駅、国勢調査地区、殺人) を nyc データベースに読み込みます。殺人データは犯罪率の代理であり、国勢調査地区を比較的小さな地理的領域として、立地選択の高レベルの粒度として使用します。 PostgreSQL にインポートする前に、この地理空間データを解凍する必要があることに注意してください。PostGISのPostGIS 2.0シェープファイルとDBFローダーエクスポータを使用する方法のよく出来た説明は[ここ](https://postgis.net/workshops/postgis-intro/loading_data.html)にあります。

前のリンクで説明したように、PostGIS 2.0 シェープファイルおよび DBF ローダー エクスポータにAzure Database for PostgreSQL のインスタンスのログイン資格情報を指定する必要があります。ユーザー名とパスワードは、Azure Database for PostgreSQL のインスタンスを作成したときに指定した内容になります。サーバー ホストは、PostgreSQL インスタンス用の Azure データベースに関連付けられたパブリック エンドポイント、例えばyour\_instance\_name. postgres.database.azure.comになります。ポートは 5432 で、データベース名は nyc になります。

ログイン資格情報を指定したら、このデータを含む zip ファイルから抽出した地理空間データセットから .shp ファイルを追加する必要があります。 [ファイルの追加] ボタンをクリックし、4 つの .shp ファイルを追加し、各地理空間エンティティの SRID フィールドに値 26918 を使用するようにインポータを構成します。

PostGIS 2.0 シェープファイルと DBF ローダー エクスポータのウィンドウが次のように表示されたら、正しく動作したことがわかります:

A screenshot of a cell phone

Description automatically generated

これを正しく構成したら、[インポート] ボタンをクリックして、これら 4 つの地理空間エンティティを nyc データベースにインポートします。

これら 4 つのエンティティ/テーブルには、geom 列に地理空間要素があり、他の列には "従来の" リレーショナル要素があります。各エンティティの地理空間特性は、ポイント (地下鉄の駅や殺人の場合)、1/複数のライン (道路の場合)、または 1/複数のポリゴン (国勢調査地区の場合) のいずれかです。

これらのエンティティ/テーブルとその中の有用な列の一部の詳細については、以下のデータディクショナリで説明します。

## Geospatial Data Dictionary

Useful columns in the “nyc\_streets” table:

|  |  |
| --- | --- |
| Name | Common name of the street |
| Type | What kind of street is this? |
| Geom | MultiLinestring geometry of the street |

Useful columns in the “nyc\_subway\_stations” table:

|  |  |
| --- | --- |
| Name | Common name of the station |
| geom | Point geometry of the street |

Useful columns in the “nyc\_homicides” table:

|  |  |
| --- | --- |
| weapon | Weapon used in the homicide |
| geom | Point geometry of the homicide |

Useful columns in the “nyc\_census\_tracts” table:

|  |  |
| --- | --- |
| tractid | An 11-digit code that uniquely identifies every census tract. Eg: 36005000100 |
| geom | MultiPolygon boundary of the tract |

**社会経済データを使用した地理空間データセットの充実**

地理空間データには既にいくつかのリレーショナル要素がありますが、ビジネス目標を達成するには、このデータを追加のリレーショナル データと組み合わせる必要があります。

この追加データについては、nyc\_census\_sociodata.sql という名前のファイルを使用して、nyc データベースにデータをインポートします。 pgAdmin 内で、nyc データベースを右クリックし、クエリ ツールを選択します。nyc\_census\_sociodata.sql ファイルの内容をこの新しいクエリ ウィンドウに貼り付け、キーボードの F5 キーを押すか、ツールバーの小さな稲妻ボタンをクリックしてクエリを実行します。

このスクリプトは、各国勢調査地区の社会経済的属性を持つ新しいリレーショナル テーブルを作成し、nyc\_census\_tracts テーブルの外部キーとして機能する tractid という名前の列を含めます。さらに、このスクリプトは、人口、平均通過時間、犯罪率など、関心のある特性のほとんどによって各国勢調査地区にスコアを付けるアルゴリズムを形成する 3 つのビューを作成します。まずは4つ目の特性(地下鉄駅からの距離)に対処するために、これらの特性とスコアをしばらく脇に置きます。

**地理空間データの視覚化**

地理空間データの視覚化を開始するには、以前にダウンロードしてインストールした QGIS アプリケーションを開きます。QGIS内で、新しいプロジェクトを開始し、ブラウザ領域でPostGISノードを右クリックし、[新しい接続...]を選択します。

A screenshot of a social media post

Description automatically generated

表示されるウィンドウ (以下の強調表示されたフィールド) に、Azure Database for PostgreSQL インスタンスの適切な資格情報と情報を入力します。次に示すように、フレンドリ名を使用する必要があります。PostgreSQL インスタンスの Azure データベースに接続するには SSL が必要であることに注意してください。

A screenshot of a computer

Description automatically generated

認証が完了すると、国勢調査地区、殺人、通り、地下鉄の 4 つの地理空間エンティティ/テーブルが nyc データベースに表示されます。 次に示すように、4 つすべてを選択し、それぞれを右クリックして[選択したレイヤーをプロジェクトに追加]を選択します。

A screenshot of a social media post

Description automatically generated

地理空間データがQGISプロジェクトにロードされたら、レイヤーのオン/オフを切り消すことができます。結果のビジュアライゼーション内のすべてのレイヤーが表示されるデフォルトから、以下に示すように国勢調査地区、道路、地下鉄の駅のみが表示されるように、殺人の横にあるチェックボックスをオフにします。

A screenshot of a cell phone

Description automatically generated

**地理空間データの操作と出力の再視覚化**

QGISでの初期設定が完了したので、pgAdminでいくつかの操作を実行します。ビジネス シナリオでは、道路の交差点で場所を選択し、立地が一度に複数の道路にアクセスできるようにすることを思い出してください。pgAdmin クエリ ツールでスクリプトを実行して、交差テーブルを作成できます。pgAdmin で、nyc データベースを右クリックし、[クエリ ツール...次のスクリプトを貼り付ける/実行する場所を用意します:

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);

このスクリプトは、道路であるラインの交点からポイント地理を作成します。また、道路テーブルの2つのインスタンス間でデカルト結合を実行しているので、重複した交差点も排除されます 。1st AvenueとAllen Streetの交差点がある場合、Allen Streetと1st Avenueの交差点も必要ありません。

交差点を作成したら、QGISに戻ります。 QGISで新しく作成された交差点を視覚化するには、最初に作成したPostGIS接続を右クリックし、「更新」を選択します。前に見た元の 4 つのエンティティと共に、nyc\_intersections という名前の新しい地理空間エンティティが表示されます。nyc\_intersectionsを右クリックし、「プロジェクトにレイヤーを追加」を選択します。約43,000の非重複交差で、QGISビジュアライゼーションは、他の地理空間エンティティをカバーする新しいポイントで、ほぼ完全にいっぱいになるはずです。

地下鉄の駅から500メートル以上離れた高速道路を含まない安全な交差点のみを考慮したいので、前のpgAdminスクリプトを次のスクリプトに置き換えることができます。

/\* 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)

);

pgAdmin でこのスクリプトを実行すると、16,000 を超える交差が考慮から削除されます。私たちはQGISに戻り、交差点のビューをリフレッシュすることにより、この操作の影響を見ることができます。これを行うには、QGISのレイヤーウィンドウにnyc\_intersectionsのリストを右クリックし、「SQLレイヤーを更新...」を以下に示すように選択します。

A screenshot of a map

Description automatically generated

次に示すように、表示される [DB マネージャ] ウィンドウの右下隅にある [更新] ボタンをクリックして、交差点データを更新します:

A screenshot of a social media post

Description automatically generated

QGISが交差点レイヤーの更新を終了すると、地下鉄の停留所の周りにバッファゾーンがあることがわかります。 地下鉄の駅と交差点に同様の色のドットが関連付けられている場合、地下鉄の駅の表示/非表示を切り替えると、このバッファ ゾーンが簡単に表示されます。

**データの最終的な操作**

高速道路を含む交差点や地下鉄の駅に近すぎる交差点を考慮に入れたので、人口、平均通勤時間、犯罪率に基づいて選択肢を絞り込むことができます。以前に作成したビューの 1 つが、これらの変数によって国勢調査地区をスコア付けしたことを思い出すかもしれません - pgAdmin で次のクエリを実行すると、上位 10 個の国勢調査地区 ID を確認できます:

SELECT tractid

FROM vw\_nyc\_site\_selection\_data

LIMIT 10

このクエリの結果を使用すると、この国勢調査地区の上位 10 リストに含まれていない、または近くにいないすべての交差を除外するクエリを作成できます:

/\* 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')

)

pgAdmin でこの最後のクエリを実行した後、QGISでnyc\_intersectionsレイヤーを再度更新して、基準を満たす350以上の交差点まで、考慮事項が大幅に絞り込まれたことを確認できます。物件の利用可能性、賃貸・購入価格などを評価する必要があるため、これ以上自動的に改良できない場合があります。しかし、以下に示すように、従来のクエリと地理空間クエリの組み合わせを使用して、ニューヨーク市の交差点の 99% 以上を非常に迅速に排除しました。

A close up of a map

Description automatically generated

地理空間データとクエリの詳細については、このチュートリアルにインスピレーションを与えてくれた [PostGIS ワークショップの概要](https://postgis.net/workshops/postgis-intro/)を参照してください。