

AMAZON QUICKSIGHT ハンズオン

Amazon Web Services Japan

第 2 版 2020/03/11 作成, 最終更新 : 2022/1/6

目次

本資料について	4
事前準備と環境	4
ハンズオンの流れ	5
QuickSight の利用を開始する	5
QuickSight にサインアップする	6
事前準備 : CSV ファイルをアップロードし、データセットを作成する	10
QuickSight と SPICE	10
CSV ファイルを確認する	10
SPICE に 1 つ目の CSV ファイルをアップロードする	11
データセットを準備 (PREPARE) する	13
ハンズオンの進め方	18
ハンズオン 1 : 基本的な可視化とクイックフィルタ	19
垂直積み上げ棒グラフの作成	19
ドーナツグラフの作成	24
ピボットテーブルとコンディショナルフォーマットの作成	27

クイックフィルタ機能による絞り込みの実現.....	32
（オプション演習）折れ線グラフと予測機能（Enterprise Edition が必要）	35
（オプション演習）テーマを切り替える	39
ハンズオン 2：より進んだ可視化.....	40
ハンズオン 2 を始める前に	40
垂直棒グラフの作成とドリルダウンの実現.....	42
ビジュアルの調整とコピー	47
コントロールの作成とフィルタ	48
計算フィールド①：シンプルな計算フィールド	53
計算フィールド②：テーブル計算関数（runningSum）	56
（オプション）計算フィールド③：テーブル計算関数（percentDifference）	58
（オプション）計算フィールド④：テーブル計算関数（percentOfTotal）	60
ML インサイトによる異常値検出の設定	61
ML インサイトによるナラティブ	64
ML インサイトによる異常値検出の内容を確認する	66
（オプション）ハンズオン 3：Level-Aware Aggregation.....	68
Level-Aware Aggregation とは.....	68
Level-Aware Aggregation (countOver)	69
Level-Aware Aggregation (sumOver).....	71
（オプション）ハンズオン 4：大規模データを Athena 経由で可視化する	74

QuickSight + Athena + S3 の構成.....	74
本ハンズオンで発生する費用	75
Athena とサンプルデータ	76
Athena からサンプルデータをクエリする	76
Athena からパーティショニングされたサンプルデータをクエリする	84
QuickSight から Athena に接続する	89
計算フィールド(Calculated Field)の追加	93
QuickSight を利用した可視化.....	95
ビジュアル 1:航続距離が長いキャリア Top 5（水平棒グラフ）	97
ビジュアル 2:キャンセル率が高い州（地図）	101
ビジュアル 3:キャンセル率が高い州（垂直棒グラフ+ドリルダウン）	102
ビジュアル 4:Vermont のキャンセル理由（円グラフ）	106
ビジュアル 5:キャンセル理由（ピボットテーブル）	109
（オプション）ビジュアルのサイズや位置の調整	110
まとめと後片付け	111
補足情報・参考資料（QuickSight）	112
補足情報・参考資料（Athena）	113
補足情報・参考資料（S3 他）	114

本資料について

事前準備と環境

本資料は Amazon Web Services（以下 AWS）が提供する BI サービス、Amazon QuickSight（以下 QuickSight）を使ってみることで、その機能を体験するハンズオンです。

ハンズオンを実行するには **AWS のアカウントが必要**になります。まだ AWS アカウントをお持ちでない場合は、以下より登録が可能です。

<https://aws.amazon.com/jp/register-flow/>

ハンズオンの大半は QuickSight のみで構成されていますがオプションのハンズオン 4 では Amazon Athena（以下 Athena）と Amazon S3（以下 S3）も合わせて利用します。

それぞれの サービスは所定の費用がかかります。費用の詳細は以下のホームページでご確認ください。また、ハンズオン記載以外の処理を実施するとその分費用が発生する可能性がある点にご注意ください。

- QuickSight の料金

<https://aws.amazon.com/jp/quicksight/pricing/>

- Athena の料金

<https://aws.amazon.com/jp/athena/pricing/>

- S3 の料金

<https://aws.amazon.com/jp/s3/pricing/>

ハンズオンを実行するには、ブラウザ（**Chrome か Firefox の最新版を推奨**。その他 Safari 7 以降、Edge 最新バージョンもサポート）とインターネットに繋がる環境が必要です。ブラウザと QuickSight サービス間の通信は、HTTPS と WebSockets Secure(wss://)を使用しますので、ファイアウォール等で防がれていないかご確認ください。

また、利用ブラウザのプラグイン・アドオン、特にブラウザ付属の自動翻訳機能が QuickSight の動作を妨げる事がありますのでご注意ください。

ハンズオンの流れ

最初に AWS アカウントで管理コンソールにログインし、QuickSight にサインアップしてデータセットを準備します。事前準備が完了した後は以下のハンズオンを試すことができます。各ハンズオンは独立しているため、どれか 1 つだけを実施することも可能です。

- ハンズオン：パート 1

パート 1 では、棒グラフやドーナツチャート、ピボットテーブルといった基本的なビジュアルの作成方法に加え、クイックフィルタの設定や、コンディショナルフォーマット（値によって色を変える）を体験します。オプションで折れ線グラフを作成し、予測を追加する方法やテーマ（配色）切り替えを体験します。

- ハンズオン：パート 2

パート 2 ではより進んだ活用方法を体験します。コントロール（GUI）を配置してのフィルタの設定、計算フィールド（関数）の利用、ML インサイトによる異常検知と自動ナラティブの作成を体験します

- ハンズオン：パート 3（オプション）

高度な関数の使い方である Level Aware Aggregations (LAA) の利用方法を体験します。

- ハンズオン：パート 4（オプション）

S3 上に存在する大規模データに対し、Athena でクエリ可能にする方法や性能最適化の方法を体験した後に、QuickSight から Athena に接続して可視化する方法を体験します。

QUICKSIGHT の利用を開始する

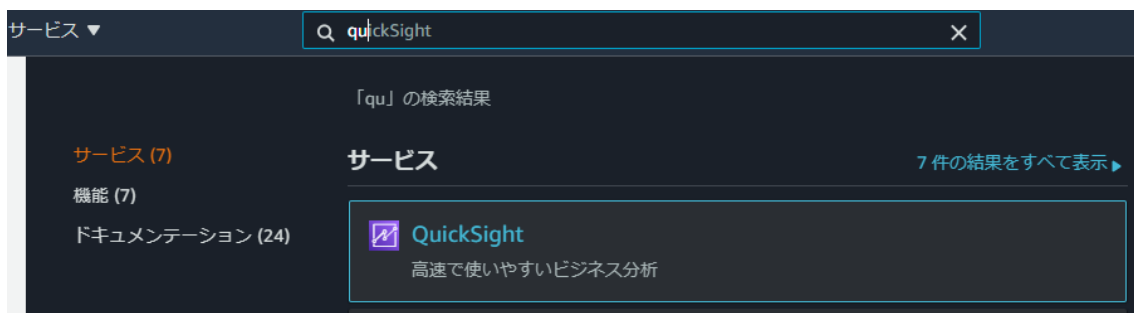
本章では QuickSight にサインアップし、利用開始できるまでを説明します。すでに QuickSight を利用されている場合は本章をスキップしてください。

QUICKSIGHT にサインアップする

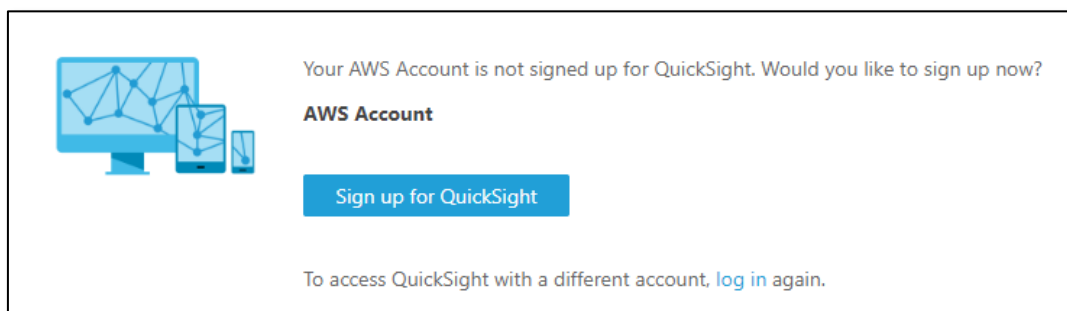
<https://console.aws.amazon.com> にアクセスし、AWS マネジメントコンソールにログインします。AWS マネジメントコンソールでは、右上のメニューから利用リージョンを「バージニア北部」に設定してください。（※本資料はすべてバージニア北部で画面ショットを作成していますが、他リージョンで QuickSight をサインアップしていてもハンズオンは実施可能です）



サービス一覧から **QuickSight** を選択します。上部の検索欄で **qu** 等と入力すると絞り込まれます。



初めて利用する場合は以下の画面が表示されますので、“**Sign up for QuickSight**”をクリックします。



クリックすると以下のエディション選択の画面が出ます。

Create your QuickSight account

Enterprise | Standard

Edition	<input checked="" type="radio"/> Enterprise	<input type="radio"/> Enterprise + Q Learn more
Team trial for 30 days (4 authors)*	FREE	FREE
Author per month (yearly)**	\$18	\$28
Author per month (monthly)**	\$24	\$34
Readers (pay-per-Session)	\$0.30 / session (max \$5)****	\$0.30 / session (max \$10)****
Additional SPICE per month	\$0.38 per GB	\$0.38 per GB
QuickSight Q regional fee	N/A	\$250 / mo / region
Personalized Q authoring workshop	N/A	Starting from \$199
Natural language query with QuickSight Q	N/A	INCLUDED
Single Sign On with SAML or OpenID Connect	✓	✓
Connect to spreadsheets, databases & business apps	✓	✓
Access data in Private VPCs	✓	✓
Row-level security for dashboards	✓	✓
Secure data encryption at rest	✓	✓
Connect to your Active Directory	✓	✓
Use Active Directory groups***	✓	✓
Send email reports	✓	✓
Embed QuickSight	✓	✓
Capacity-based pricing	✓	✓
Supported regions	Learn more	Learn more

* Trial authors are auto-converted to month-to-month subscription upon trial expiry

** Each additional author includes 10GB of SPICE capacity

*** Active Directory groups are available in accounts connected to Active Directory

**** Sessions of 30-minute duration. Total charges for each reader are capped at \$5 per month. [Conditions apply.](#)

[Continue](#)

初期状態では Enterprise Edition が選択されていますので、そのまま Continue を選択します。

なお Standard Edition を選択しても後から Enterprise Edition にアップグレード可能です。

(補足：本ハンズオンの多くは Standard Edition でも実施可能ですが、ML インサイトの予測や異常検知の部分を実施するには Enterprise Edition が必要です。また、Enterprise + Q を選択すると QuickSight Q が利用可能になりますが、別途費用が発生する点に注意してください)

Authentication method

- ☒ Use IAM federated identities & QuickSight-managed users
Authenticate with single sign-on (SAML or OpenID Connect), AWS IAM credentials, or QuickSight credentials
- ☐ Use IAM federated identities only
Authenticate with single sign-on (SAML or OpenID Connect) or AWS IAM credentials
- ☐ Use Active Directory
Authenticate with Active Directory credentials

QuickSight region

Select a region

US East (N. Virginia) ▼

US East(N.Virginia)を選択

Account info

QuickSight account name

You will need this for you and others to sign in

Enter a unique QuickSight account name

任意のアカウント名
(半角英数のみ、他の人と被らないもの)

Notification email address

For QuickSight to send important notifications

Enter account notification email address

emailアドレスを入力

QuickSight access to AWS services

Make your existing AWS data and users available in QuickSight. [Learn more](#)

IAM Role

- ☒ Use QuickSight-managed role (default)
- ☐ Use an existing role

Allow access and autodiscovery for these resources

☒ Amazon Redshift

☒ Amazon RDS

☒ IAM

☐ Amazon S3

[Select S3 buckets](#)

☐ Amazon Athena

Make sure you've chosen the right Amazon S3 buckets for QuickSight access

Aamzon Athenaの
チェックボックスは外す

☐ Amazon S3 Storage Analytics

☐ AWS IoT Analytics

☐ Amazon OpenSearch Service

☐ Amazon SageMaker

☐ Amazon Timestream

Finish

この画面では、QuickSight アカウントの設定を行ないます。QuickSight Region は、Author ユーザに付属する SPICE 容量がどのリージョンに確保されるかを指定するものです。今回は US

East(N. Virginia)を選択してください。 なおここで選択した以外のリージョンにも追加で SPICE 容量を追加する事は可能です（追加 SPICE 容量の費用が必要です）。

QuickSight account name は**任意の名前**を付けます（半角のアファベット、数字、ハイフンのみ利用可能で、他の人と被らないユニークなもの）。

Notification email address には、ご自身のメールアドレスを入れてください。

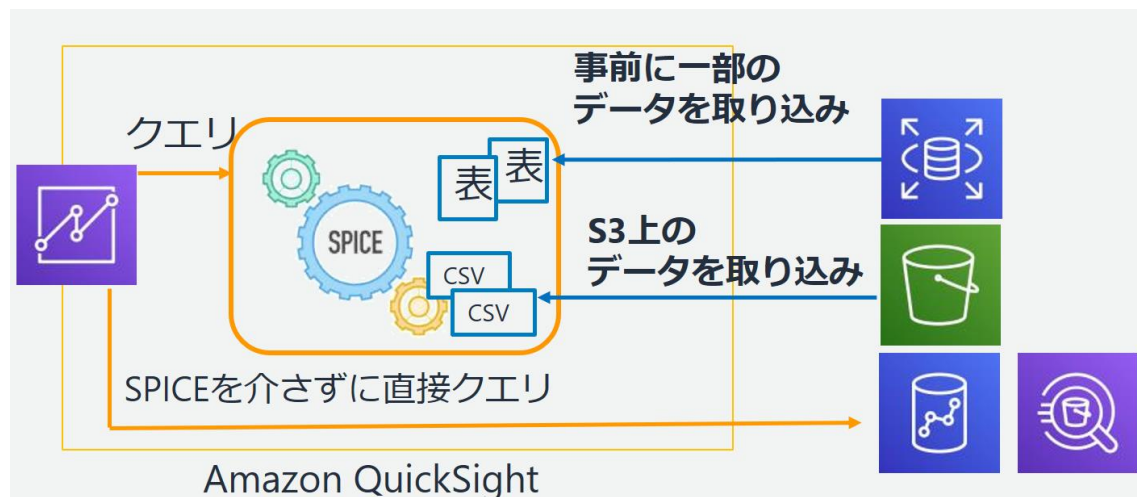
加えて上記のように Amazon Athena からはチェックを外し、Finish を押してください。少し待つと”Go to Amazon QuickSight”のボタンが表示されますので、それを押します。

事前準備：CSV ファイルをアップロードし、データセットを作成する

※本章「事前準備」で作成するデータセットは、ハンズオン 1～3 で利用しますがハンズオン 4 では使用しません。ハンズオン 4 のみ実施される場合はこの準備をスキップ可能です。

QUICKSIGHT と SPICE

QuickSight には SPICE というインメモリ型の高速データベースが内蔵されています。この SPICE の領域にデータを取り込んでおくことで、高速な BI 応答が可能になります。もちろん、SPICE を使わず、Athena や Amazon Redshift に直接クエリ（SQL）を発行して利用することも可能です。



本ハンズオンでは、CSV ファイルを SPICE 領域にアップロードし、それを可視化します。

CSV ファイルを確認する

このハンズオンで使用する CSV ファイルは、このハンズオン資料に同梱されています。もしくはハンズオンの講師が指示する URL からダウンロードしてください。

CSV ファイルは 2 つあります。Patient-Info.csv ファイルをテキストエディタ等で開くと、以下のようにカンマ区切りでデータが格納されている事が分かります。最初の行は列名です。Patient ID（患者 ID）から始まっており、Admit Date（入院日付）等の日付は、日/月/西暦

(2 桁) の形式で記録されている事が分かります。(※この CSV を EXCEL で見ると、日付部分の解釈の違いにより、以下とは異なる見え方になるので注意が必要です)

```
1 Patient ID,Admit Date,Discharge Date,Priority,Hospital,
2 3,10/14/16,10/21/16,Low,North Hospital,-213.25,38.94,3E
3 6,2/20/18,2/21/18,Not Specified,North Hospital,-4.64,2.
4 32,7/15/17,7/16/17,High,North Hospital,-128.38,8.46,8.9
5 32,7/15/17,7/16/17,High,South Hospital,-1748.56,70.89,8
6 32,7/15/17,7/17/17,High,North Hospital,-85.129,7.99,5.0
7 32,7/15/17,7/17/17,High,North Hospital,1054.82,107.53,5
8 35,10/22/17,10/23/17,Not Specified,North Hospital,60.72
9 35,10/22/17,10/24/17,Not Specified,North Hospital,48.98
```

もう一つの Assign.csv ファイルは小さいファイルです。テキストエディタ等で開くと各社員 (Account Rep) が担当している Region (地域) を表すマスター表であることが分かります。この Region 列は、Patient-Info.csv にも存在します。

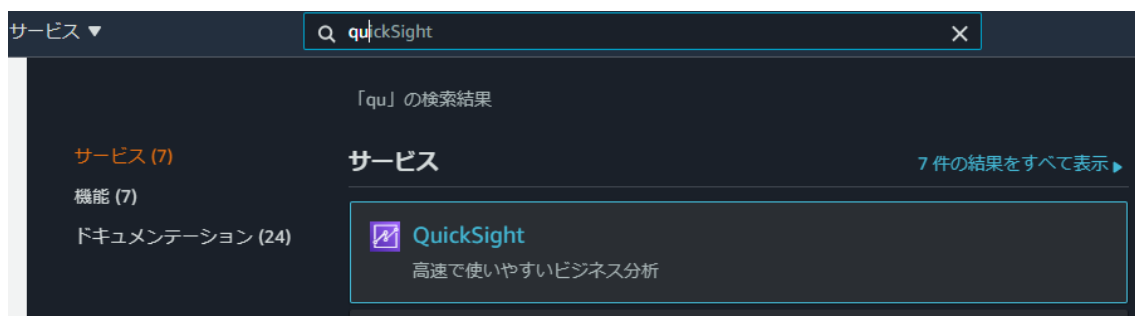
```
1 Account Rep,Region↓
2 Kevin,East↓
3 Jose,Central↓
4 Susan,South↓
5 Mark,West↓
```

SPICE に 1 つ目の CSV ファイルをアップロードする

(まだの場合は) AWS マネジメントコンソールにログインし、利用リージョンを QuickSight 登録時に QuickSight region に指定したリージョンに切り替えます。今回のハンズオンで初めて QuickSight を利用開始した場合はバージニア北部 (N. Virginia) を選択します。事前に QuickSight アカウント作成済であった場合は、そのアカウントの QuickSight region に切り替えてください (QuickSight を東京リージョンで subscribe していた場合は東京を選択)。



サービス一覧から QuickSight を選択します。検索欄で qu 等と入力すると絞り込まれます。

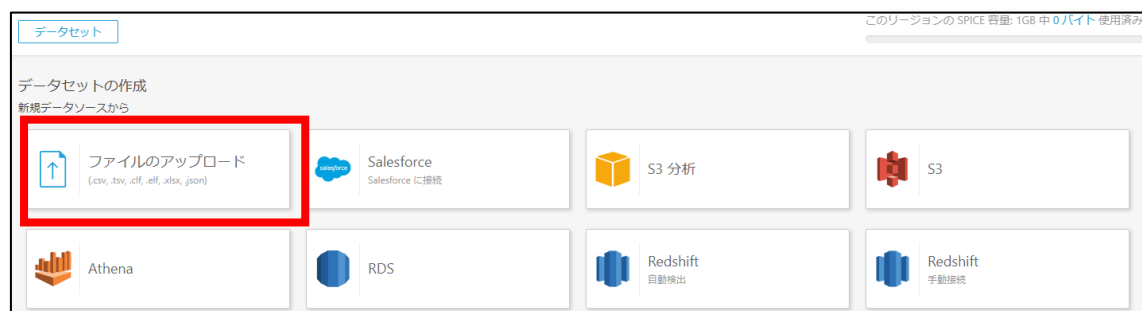


QuickSight の画面で右上にある AWS ログイン ID(IAM 名)が記載された部分をクリックし、先程指定したリージョンが選択されていることを確認してください（画面が英語表示の場合は同じところで言語を日本語に変更できます）。確認できたら、画面左にある”データセット”をクリックします。

次の画面で右上の”新しいデータセット”をクリックし、



さらに”ファイルのアップロード”をクリックします。



ファイル選択のダイアログで Patient-Info.csv の方を選択してアップロードします。確認画面が出ますので、そこで”設定の編集とデータの準備”を選択します。

ファイルのアップロード設定の確認

×

詳細はこちらをクリックして、ファイルのアップロード設定を調整します。

設定

csv ファイル、 Patient-Info.csv

Patient ID	Admit Date	Discharge ...	Priority	Hospital	Profit
3	2016-10-14...	2016-10-21...	Low	North Hospital	-213.
6	2018-02-20...	2018-02-21...	Not Specified	North Hospital	-4.64
32	2017-07-15...	2017-07-16...	High	North Hospital	-128.
32	2017-07-15...	2017-07-16...	High	South Hospital	-1748
32	2017-07-15...	2017-07-17...	High	North Hospital	-85.1

設定の編集とデータの準備

次へ

すると「準備（PREPARE）」の GUI に移動します。

データセットを準備（PREPARE）する

以下の画面は「準備（PREPARE）」ツールで、ここでは分析の準備として、データを利用しやすい形にすることができます。例えばフィールド（列）の名前を変えたり、不要なフィールドを見えなくしたり、事前に複数のデータソースをジョイン（結合）する等です。

データセット名: Patient-info

フィールド: すべてのフィールドが含まれています

計算フィールドを追加

SageMaker によるオーグメント

フィールドの検索

フォーカス: すべてのフィールド

選択: すべて | なし

除外されたフィールド: 除外されたフィールドはありません

フィルタ: フィルターが適用されていません

クエリモード: SPICE

残りの容量: 1.4GB

データ

自動プレビュー

ズーム: 100%

データを追加

データセット

Patient ID	Admit Date	Discharge ...	Priority	Hospital	Profit	Price	Cost	Revenue	Discount	Payor Seg...	Category
3	2016-10-14...	2016-10-21...	Low	North Hospital	-213.25	38.94	35.0	326.925	0.04	Personal Cost	Labor & De
6	2018-02-20...	2018-02-21...	Not Specified	North Hospital	-4.64	2.08	2.56	8.6625	0.01	Medicaid	Critical Car
32	2017-07-15...	2017-07-16...	High	North Hospital	-128.38	8.46	8.99	175.7	0.04	Medicaid	NICU
32	2017-07-15...	2017-07-16...	High	South Hospital	-1748.56	70.89	89.3	2201.75	0.09	Medicaid	Surgical
32	2017-07-15...	2017-07-17...	High	North Hospital	-85.129	7.99	5.03	200.2919	0.04	Medicaid	Step Down
32	2017-07-15...	2017-07-17...	High	North Hospital	1054.82	107.53	5.81	3510.1	0.07	Medicaid	Surgical
35	2017-10-22...	2017-10-23...	Not Specified	North Hospital	60.72	9.11	2.25	360.7	0.03	Medicaid	Surgical
35	2017-10-22...	2017-10-24...	Not Specified	North Hospital	48.987	155.99	8.99	2366.06	0.01	Medicaid	Psychology
36	2017-11-02...	2017-11-02...	Urgent	North Hospital	657.477	65.99	4.2	3105.9319	0.1	Corporate In...	Ante/Post
65	2017-03-17...	2017-03-18...	Urgent	North Hospital	1470.3	115.79	1.99	4765.9125	0.02	Medicaid	Critical Car

ここでは、以下3点の調整を行ないます。

1) 他データソースのジョイン

もう一つの CSV データをここで追加し、Region フィールド（列）を基準にジョインします。画面上方の「データを追加」をクリックし、ダイアログから「ファイルのアップロード」を選択します。ファイルは、先程ダウンロードした中にある Assign.csv を選択します。

アップロードされると以下の画面になるので、「次へ」を押します。

ファイルのアップロード設定の確認

[詳細はこちら](#) をクリックして、ファイルのアップロード設定を調整します。

設定

csv ファイル、Assign.csv

Account Rep	Region
Kevin	East
Jose	Central
Susan	South
Mark	West

次へ

アップロードした後は2つのデータの結合方法を指定します。2つの円のようなアイコン（上部）をクリックし、画面下に出る結合設定で結合（ジョイン）する列を指定します。今回は Region フィールドでジョインしたいので、両方とも Region を指定し、**結合タイプは Inner** に変更して、「適用」を押します。

クエリモード
SPICE

フィールド
計算フィールドを追加
SageMaker によるオーグメント
フィールドの検索
フォーカス
すべてのフィールド
選択 すべて | なし

データ
Patient-Info.csv
Assign.csv

自動プレビュー
データを追加

ズーム
100%

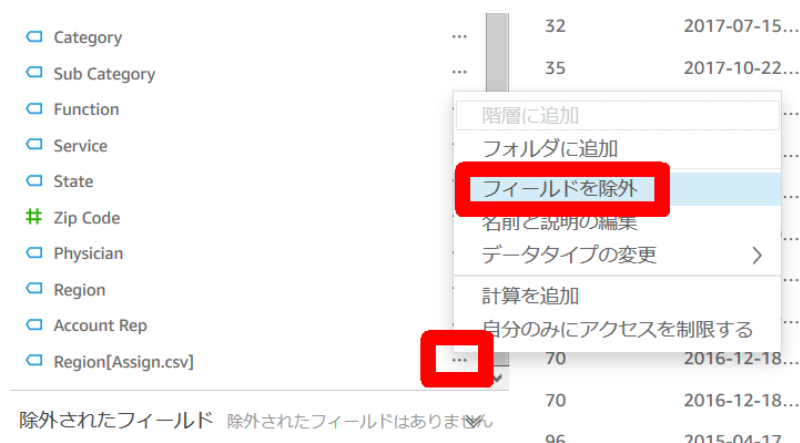
結合設定
キャンセル 適用

結合句
Patient-Info.csv
Assign.csv
Region = Region

結合タイプ
Inner Left Right Full

Some optimization settings aren't available with SPICE.

これで、画面左側のフィールドに、Assign.csv にあった Account Rep と Region[Assign.csv]が追加されます。Region 列は Patient-Info にも存在し、2つある必要はないため、Region[Assign.csv]の横にある…ボタンから、「フィールドを除外」を選択します。



今回のハンズオンでは CSV ファイルと CSV ファイルをジョインしましたが、異なるデータソース間、例えば PostgreSQL データベース上の表と、MySQL データベース上の表と、CSV ファイルをジョインする事も可能です。

(補足：異なるデータソースをまたがってジョインされたデータは、SPICE 内に保存されます。SPICE の容量を消費する点に注意が必要です。)

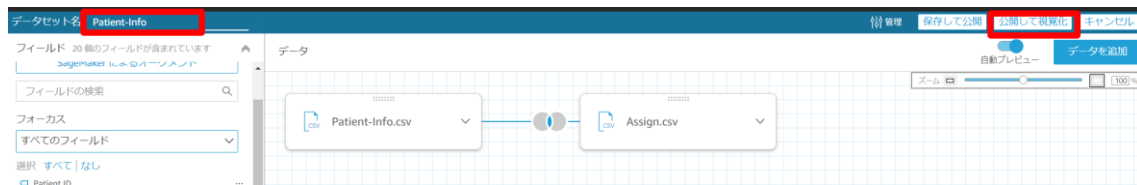
2) Patient-ID フィールドの型を修正

Patient-ID フィールドが自動的に整数(Int)型に認識されていますが、正しくは文字列なので、文字列型に変更します。右側の Patient ID フィールドの下にある「整数」と書いた部分をクリックし、文字列を選択します。



3) 名前の変更と保存

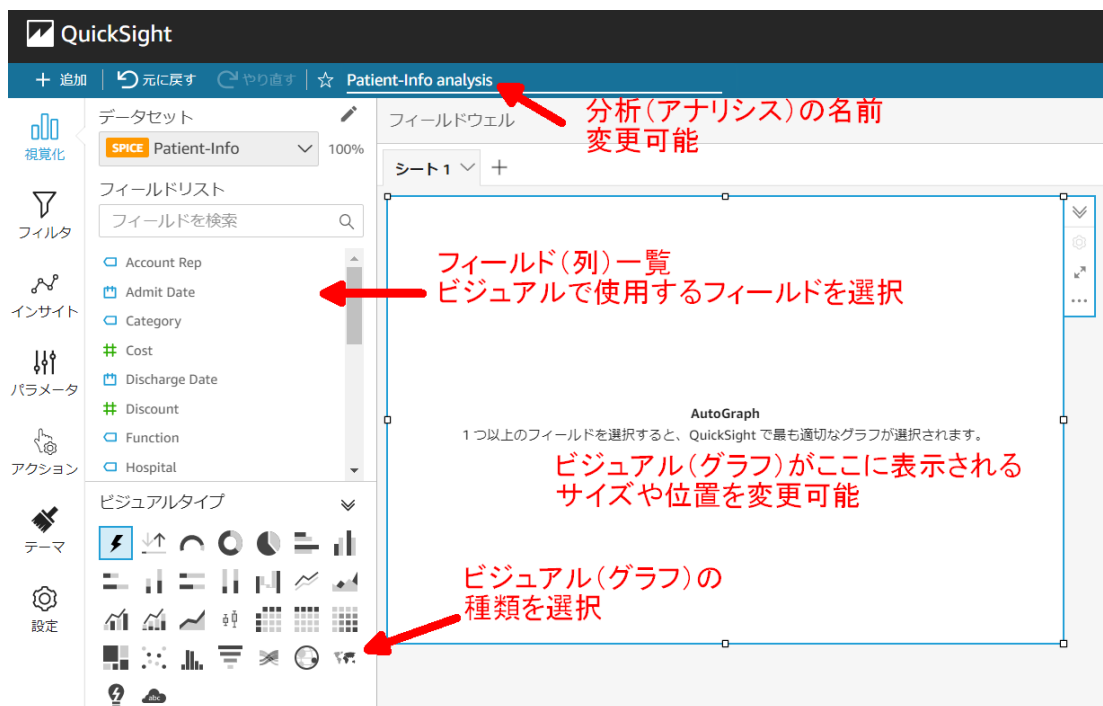
画面最上部にある枠で、このデータセットの名前をつけることが可能です。任意の文字列に変更してください。ここではデフォルトの「Patient-Info」のままとしました。



最後に、“公開して視覚化”を押し、分析画面に移動します。SPICE にデータが取り込まれるまで少し待ちます。右上の表示が”インポートの完了”になったらデータの取り込みが完了しています。（このダイアログは x を押して閉じてください） ※補足：少し待ってもインポート完了の表示が出ない場合は、ブラウザのリロードを試してください。

インポートの完了: x

100% が成功しました
8392 行が SPICE にインポートされました
0 行がスキップされました



データ取り込みが完了すると上記のような画面が表示されています。これは分析（Analysis）を作っていく画面です。左側にフィールド（列）と、ビジュアル（グラフ）の一覧が表示されています。これで準備が完了しました。

QuickSight での可視化は、まずビジュアルタイプ（グラフの種類）を選択し、その表示に必要なフィールドを選択することで可視化されます。ビジュアルは左上の「+」を押すことで複数追加できますし、サイズや位置を変更することが可能です。

操作は自動的にセーブされます。また、画面左上にある「元に戻す」ボタンで操作を取り消す（UNDO する）事が可能です。

ハンズオンの進め方

この後はハンズオン 1 に進んでください。

各ハンズオンは内容が独立しているため、セルフラーニングで実施されている方はハンズオン 1 を飛ばしてハンズオン 2 や、4 を実施していただいても問題ありません。

ハンズオン 3 から実施することも可能ですが、内容はハンズオン 2 で説明する計算フィールドの基本的な概念を理解していることを前提にしています。

また、ハンズオン 4 は Athena の利用方法についての学習にフォーカスしており、基本的な QuickSight の操作をある程度理解している事を前提にしています。まだ慣れていない方はハンズオン 1 を先に実施される事をおすすめします。

ハンズオン 1：基本的な可視化とクイックフィルタ

ハンズオンを開始するには前章の「事前準備」を完了している必要があります。

垂直積み上げ棒グラフの作成

ここからは SPICE に取り込んだデータを可視化していきます。

QuickSight

分析(アナリシス)の名前 変更可能

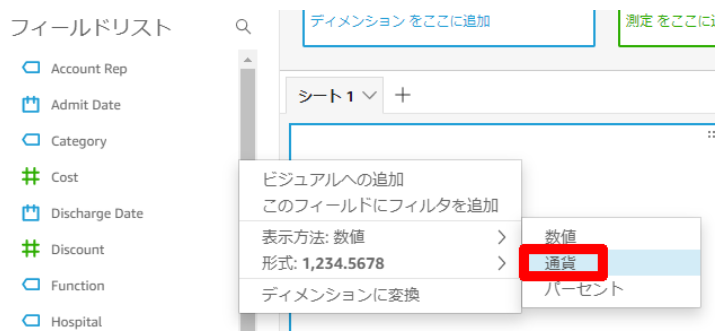
フィールド(列)一覧
ビジュアルで使用するフィールドを選択

ビジュアル(グラフ)の種類を選択

ビジュアル(グラフ)がここに表示される
サイズや位置を変更可能

ここまでの準備で、分析(Analysis)作成の画面（上記）になっており、1つ空白のビジュアル（グラフ）があります。

まず一部のフィールドの表示形式を修正します。フィールドリストにある **Cost** の上にマウスカーソルを移動すると、“…”が表示されますので、それをクリックし、表示方法の「数値」を「通貨」に変更してください。



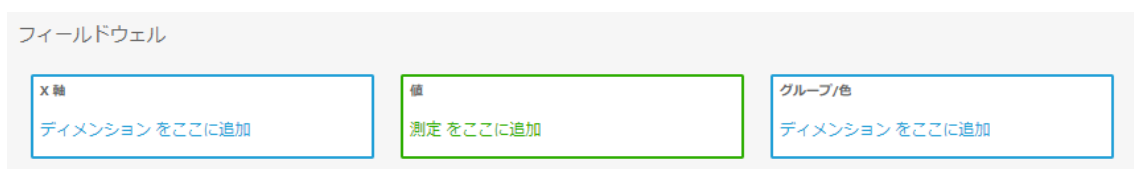
同様に、**Discount**、**Price**、**Profit**、**Revenue** も「通貨」に変更してください。


ここからビジュアルを作成していきます。

デフォルトではビジュアルタイプは"AutoGraph"になっています。これは選択したフィールドによって自動的にビジュアルタイプを選択する機能ですが、今回は垂直積み上げ棒グラフを選択します。（上から2段目、左から2番目のアイコン）

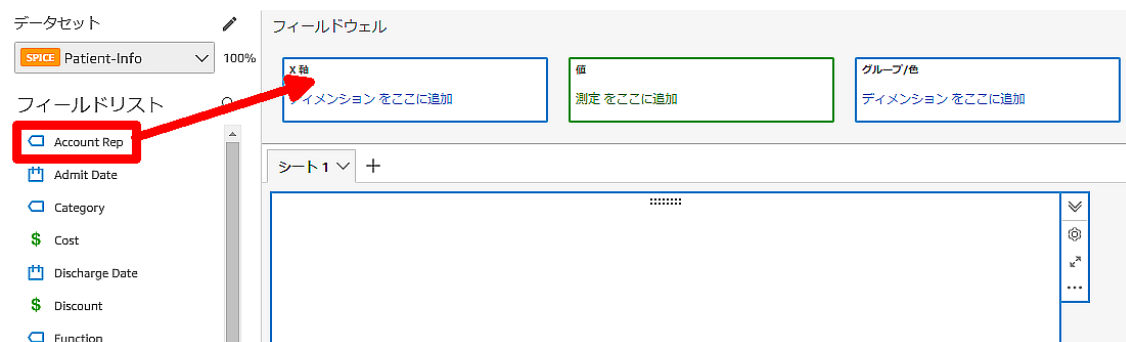


選択すると、右側のビジュアルの上に"フィールドウェル"が準備されます。

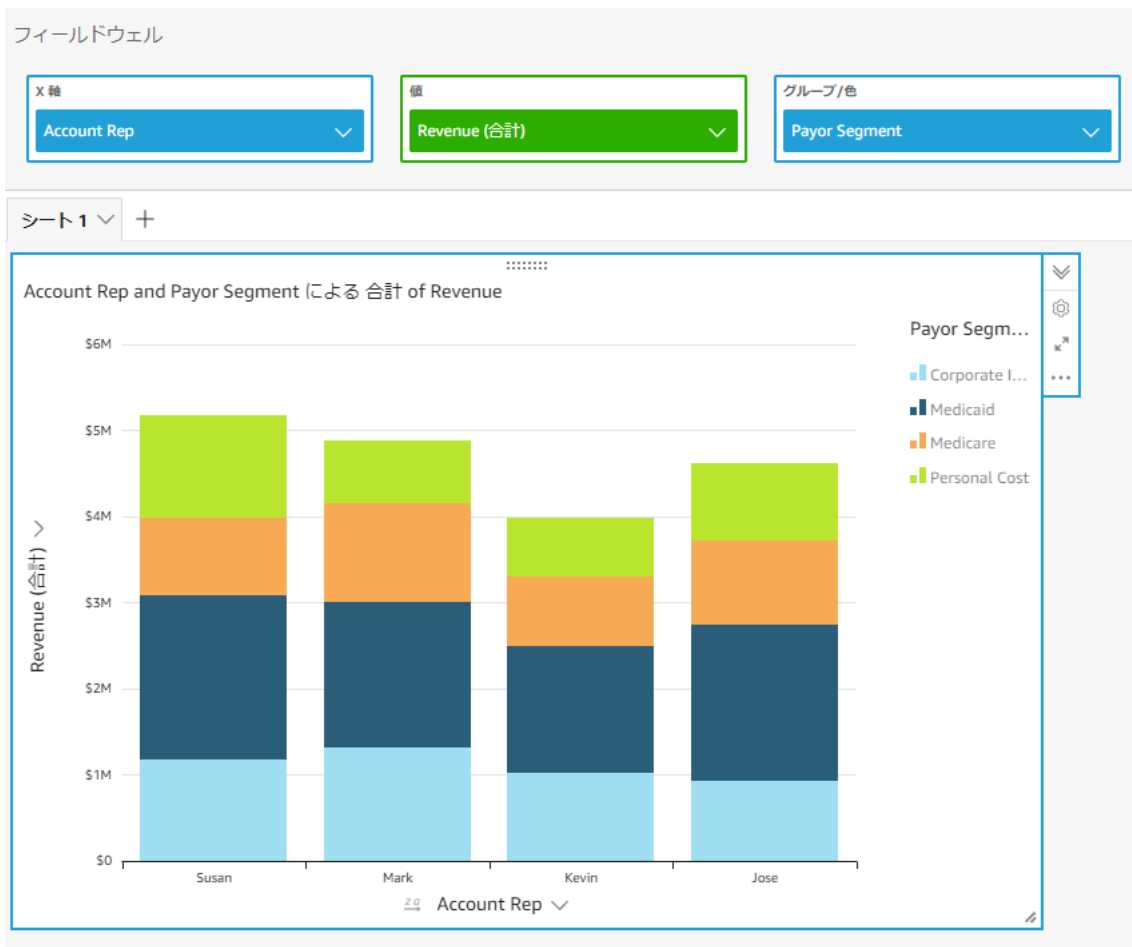


※もしフィールドウェルが上記のように表示されない場合は、右上にある  マークをクリックして、フィールドウェルを開いてください

この3つの領域に、左からフィールドを Drag & Drop（以下、D&D）することでそのフィールドのデータをもとに可視化が行われます。まず X 軸（横軸）から設定します。担当者(Account Rep)ごとの売上(Revenue)を見たいため、X 軸には Account Rep を D&D します。




同様に値には、“Revenue”（売上）を、グループ/色には“Payor Segment”（支払いセグメント）を D&D します。

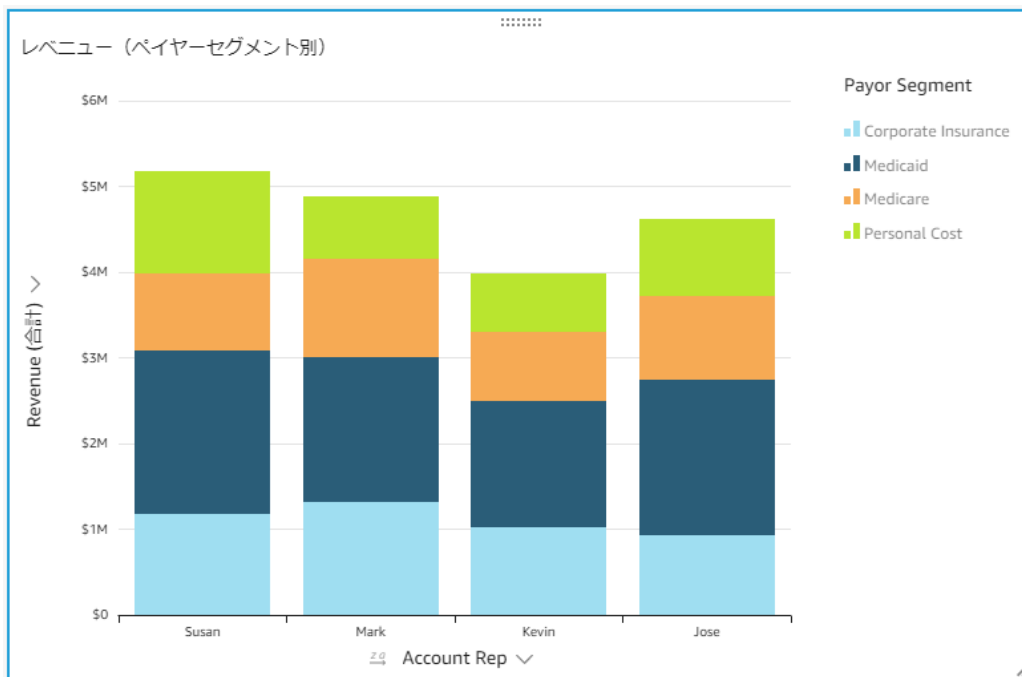


上記のようなビジュアルが出力されました。このビジュアルは担当者ごとの Revenue を合計したものを”Payor Segment”別に積み上げて表示しています。Revenue が合計されているのは、上部フィールドウェルの”値”が”Revenue(合計)”になっているためです、この集計方法を変えた場合は、その横にある”V”をクリックする事で”平均”や、”カウント”（個数）等に変えることができます。（今回は合計のままで進めます）

ビジュアル上部にある”Account Rep and Payor Segment による合計…”といった文字列は自動的に作成されたこのビジュアルのタイトルですが、変更可能です。文字列をクリックして、分かりやすい表現、例えば「レベニュー（ペイヤーセグメント別）」としてください。

また、ビジュアル左側にあるレジェンド（凡例）の領域が小さくて文字が見切れています。ビジュアル上にマウスカursorを置くと、のようなマークが表示されるので、そこをドラッグしてサイズを変えることができます。これで1つ目のビジュアルが完成しました。（下図）

このビジュアルで使っている Account Rep フィールドは、Assign.csv から、他は Patient-Info.csv から取得しています。手軽に複数のデータをジョインして可視化する事ができるのが分かります。



ドーナツグラフの作成

次はドーナツグラフを作成して、レベニュー（売上）が、どの州でどれぐらいの比率になっているかを可視化します。

左上の”+追加”->“ビジュアルを追加”を押すと新しいビジュアルが追加されます。その新しいビジュアルが選択された状態で、ビジュアルタイプとして”ドーナツグラフ”（最上段、左から4つ目）を選択してください。

フィールドウェルには、”グループ/色”と、”値”が用意されます。州ごとの売上比率を見たいので”グループ/色”には State（州）を、”値”には Revenue を D&D してください。

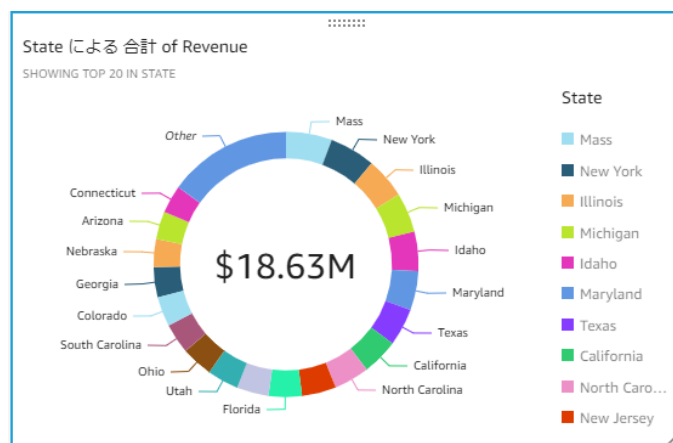
グループ/色

State

値


Revenue (合計)

以下のようなビジュアルが表示されます。



ドーナツ上になっている円の中心にあるのが、Revenue の合計です。その周辺を州ごとの売上合計が表示されていますが、ビジュアル左上に薄い字で”SHOWING TOP 20 IN STATE”と書かれているのが分かります。これは自動的に上位 20 までに絞った表示にしているという意味です。トップ 20 に入らなかった州は”Other”に積み上げられています。

あまり表示増やしても見づらくなりますが、今回はあえて全て表示させてみましょう。

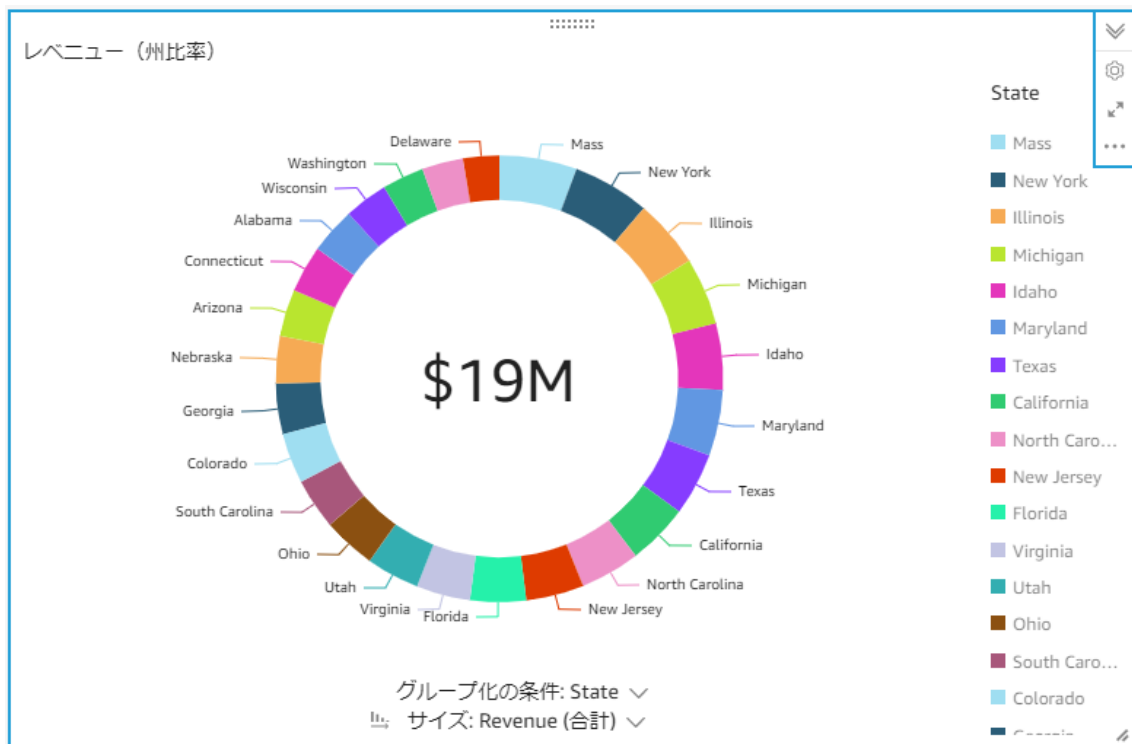
ビジュアル右側にあるツールバーからギアのマーク () をクリックすると、左側に”ビジュアルのフォーマット”ペインが表示されます。

その中から”グループ/色”を開き、”表示されたスライスの数”に 50 と入れて ENTER キーを押してください。



米国は 50 州なので、これで全てのデータが表示されます。ただし、ブラウザのサイズによっては、州の名前（ラベル）が表示しきれない場合があります。その場合はブラウザサイズを広げたり、適切なスライス数、例えば 15 に変えたりしてください。（スライス数をいくつにしても、この後のハンズオンには影響がありません）。

最初のビジュアルと同様に、ビジュアルのタイトルを”レベニュー（州比率）”に変更します。これで、以下のようなビジュアルが完成しました。



ピボットテーブルとコンディショナルフォーマットの作成

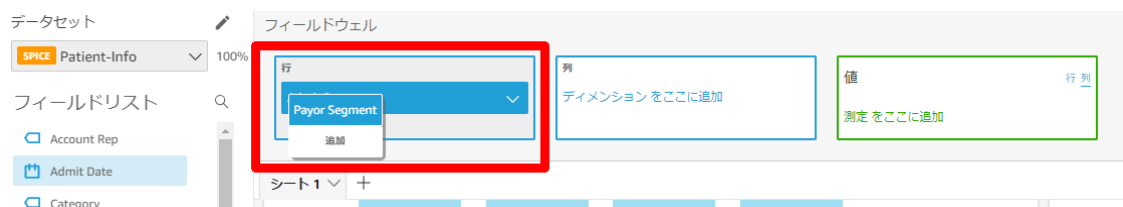
次はピボットテーブルを使って、売上や利益の詳細データを表示し、それにコンディショナルフォーマットを追加することで、利益が高いところは緑色、低いところには赤色を表示させます。

左上の”+追加”->“ビジュアルを追加”を押すと新しいビジュアルが追加されます。その新しいビジュアルが選択された状態で、ビジュアルタイプとして、“ピボットテーブル”（上から3段目段、左から5つ目）を選択してください。

フィールドウェルには、“行”、“列”、“値”が用意されます。

行にも列にも、複数のフィールドを指定して階層化できるのがピボットテーブルの特徴です。フィールドリストの Admit Date（入院日）をクリックしてください。（D&D 無しで）自動的にフィールドウェルの”行”に Admit Date が挿入されます。これはフィールドの型や定義から自動的に判断してフィールドが追加される仕組みです。続いて、Payor Segment をクリックすると、Admit Date の下に配置されます。これで階層化が実現できます。

もしくは、Admit Date の下あたりに Payor Segment を D&D する方法でもかまいません。以下の図のように、“追加”と表示される位置までドラッグしてからドロップしてください。



以下のようにフィールドウェルの”行”で2つのフィールドが階層化されているはずです。

フィールドウェル

行

Admit Date

▼

Payor Segment

▼

列

ディメンションをここに追加

値

測定をここに追加

行列

シート 1 ▼ +

Admit Date and Payor Segment による レコードのカウント

Admit Date	Payor Segment	Count
1月 1, 2015 1...	Corporate Insura...	2
	Medicaid	1
	Medicare	7
1月 2, 2015 1...	Corporate Insura...	1
	Personal Cost	1
1月 3, 2015 1...	Medicaid	3

このような方法を使って、以下の図に合わせて、行と値のフィールドウェルを設定してください。行は階層を表しているため、上下の並びが下図を同じようになるようにしてください。

また、必要に応じて見やすいサイズまでピボットテーブルの横幅を広げてみてください。（ビジュアルの右下をドラッグするとリサイズできます）

フィールドウェル

行

Admit Date

▼

Payor Segment

▼

Hospital

▼

Category

▼

列

ディメンションをここに追加

値

Revenue (合計)

▼

Profit (合計)

▼

Cost (合計)

▼

行列

シート 1 ▼ +

Admit Date, Payor Segment, Hospital, and Category による 合計 of Revenue, 合計 of Profit, and 合計 of Cost

Admit Date	Payor Segment	Hospital	Category	Revenue	Profit	Cost
1月 1, 2015 1...	Corporate In...	North Hospital	Step Down	\$906.79	\$245.97	\$5.76
			Surgical	\$235.91	-\$32.48	\$5.83
	Medicaid	South Hospital	Ante/Post Partum	\$1,555.24	-\$131.31	\$43.32
	Medicare	North Hospital	BMT	\$97.60	\$1.74	\$0.75
			Critical Care	\$2,526.43	-\$60.39	\$5.99
			Medical	\$3,437.63	\$667.45	\$5.26
			Psychologych	\$300.38	-\$154.62	\$5.68
			Surgical	\$16,001.88	-\$363.70	\$36.99
		South Hospital	Critical Care	\$816.93	-\$1,358.90	\$53.03

ピボットテーブルが構成されました。”-“マークをクリックすると各セルを閉じたり開いたりすることが可能です。

次に日付表示をカスタマイズします。フィールドウエルの”行”にある”Admit Date“のVをクリックし、形式=>その他フォーマットオプションを選択します。

左に出るデータのフォーマットペインで、”日付”を選択し、好みのフォーマット、例えば”2020年9月20日”等を選択します。これで、以下のような表示になります。

Admit Date, Payor Segment, Hospital, and Category による 合計 of Revenue, 合計 of Profit, and 合計 of Cost						
Admit Date	Payor Segment	Hospital	Category	Revenue	Profit	Cost
2015年1月1日	Corporate Ins...	North Hospital	Step Down	\$906.79	\$245.97	\$5.76
			Surgical	\$235.91	-\$32.48	\$5.83
	Medicaid	South Hospital	Ante/Post Partum	\$1,555.24	-\$131.31	\$43.32
	Medicare	North Hospital	BMT	\$97.60	\$1.74	\$0.75
			Critical Care	\$2,526.43	-\$60.39	\$5.99

最後に、コンディショナルフォーマットを設定します。これは条件に合わせて表示（色やアイコン）を変化させる機能です。ここでは、Profit（利益）が、マイナスである場合に背景が赤色に、\$100 以上の場合には緑色になるように設定します。

ビジュアル右端ツールバーの”…”から、”条件付き書式設定”をクリックします。

列の選択では色変更の対象とする”Profit”を選択し、次に”背景色を追加する”を選択します。

フォーマットフィールドでは、”条件-値”を”次より小さい”にし、値は0にします。その横にある、色の指定では赤色を指定します。これでマイナスの値のときに背景が赤色になります。

続いて、値を指定した部分の下にある”+条件を追加する”をクリックします。

すると、追加の条件が指定できるようになるため、そこでは”条件#2 - 値”に、”次以上”を指定し、値には100を、色は緑色を指定し、最後に”適用”を押します。

これで以下のようなピボットテーブルが実現できました。タイトルここでは”レベニュー/プロフィット/コスト表”に変更しています。

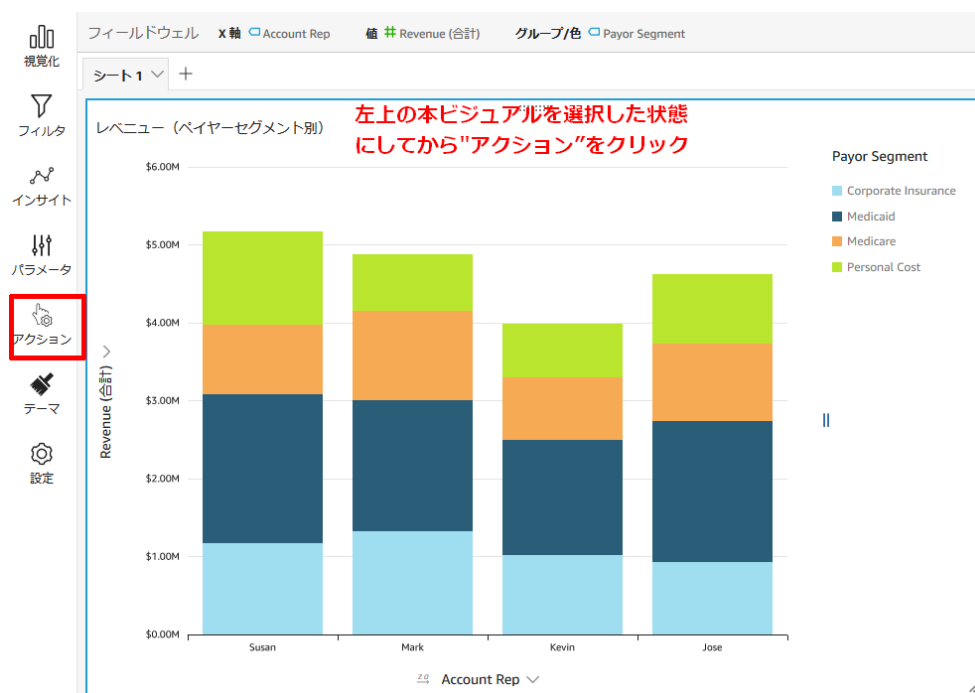
レベニュー/プロフィット/コスト表						
Admit Date	Payor Segment	Hospital	Category	Revenue	Profit	Cost
2015年1月1日	Corporate Ins...	North Hospital	Step Down	\$906.79	\$245.97	\$5.76
			Surgical	\$235.91	-\$32.48	\$5.83
	Medicaid	South Hospital	Ante/Post Partum	\$1,555.24	-\$131.31	\$43.32
	Medicare	North Hospital	BMT	\$97.60	\$1.74	\$0.75
			Critical Care	\$2,526.43	-\$60.39	\$5.99
			Medical	\$3,437.63	\$667.45	\$5.26
			Psychologych	\$300.38	-\$154.62	\$5.68
			Surgical	\$16,001.88	-\$363.70	\$36.99
		South Hospital	Critical Care	\$816.93	-\$1,358.90	\$53.03
2015年1月2日	Corporate Ins...	Downtown H...	Surgical	\$1,090.60	-\$342.91	\$35.00
	Personal Cost	North Hospital	Medical	\$225.45	-\$111.80	\$4.69

クイックフィルタ機能による絞り込みの実現

次はクイックフィルタを分析全体に適用します。フィルタとは、条件を付けてデータを絞り込む機能のことで、自分が見たいデータ範囲だけを表示させるために利用します。

QuickSight でフィルタを作成する方法は複数ありますが、ここでは手軽に作ることができるクイックフィルタを作成します。クイックフィルタは、あるビジュアルの上でクリックした際に、そのクリックされた値に応じて他ビジュアルがフィルタされる（絞り込まれる）機能です。

今回は一番初めに作った左上のビジュアルへのクリックをトリガーにフィルタしたいので、まず左上の「レベニュー（ペイヤーセグメント別）」ビジュアルを選択した状態にしてから（選択するとそのビジュアルは青い線で囲まれます）、画面左端にある”アクション”のアイコンをクリックします。



表示された”アクション”ペインで、”クイック作成”の下にある”同じシートのビジュアルをフィルタ処理する”を選択します。すると、”アクション 1”が作成されるので、その右側にある”V” => 編集をクリックします。

図のようなアクションの設定ペインが表示されます。

アクションを編集

アクション名

アクション 1

アクティベーション ①

☒ 選択

☐ メニューオプション

アクションタイプ

フィルタアクション

フィルタ範囲

☐ すべてのフィールド

☒ 選択されたフィールド

☒ Account Rep

☐ Payor Segment

ターゲットビジュアル

☒ すべてのビジュアル

☐ ビジュアルを選択する

☒ レベニュー (州比率)

☒ レベニュー/プロフィット/コスト表

保存

閉じる

削除

”アクティベーション”は、選択（クリックするとすぐフィルタされる）と、メニューオプション（一旦メニューが出る）から選択できます。今回はデフォルトの”選択”のまま使用します。

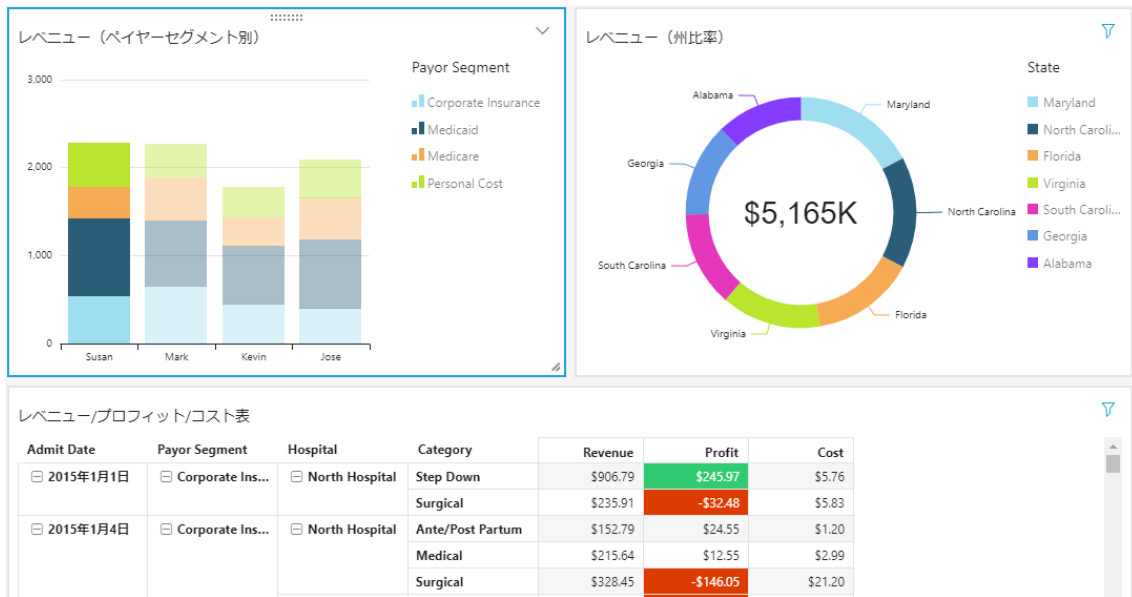
アクションタイプはフィルタアクション（フィルタする）と URL アクション（別の URL に遷移する）が選択できます。こちらもデフォルトのフィルタアクションのままにします。

フィルタ範囲は、クリックされた際、どのフィールドの値を元にフィルタするかを選択します。今回は”選択されたフィールド”を指定し、”Account Rep”だけにチェックを入れます。これにより、グラフをクリックした際に、その Account Rep だけの情報になるようフィルタされます。

もしデフォルトの”すべてのフィールド”を選択した場合は、クリックした位置の Account Rep & Payor Segment 両方の条件を満たすデータだけに絞り込まれます。

最後のターゲットビジュアルは、このフィルタがどのビジュアルに反映されるかを指定します。今回はデフォルトであるすべてのビジュアルのままにし、”保存”を押します。

これでクイックフィルタは完成です。動作確認のために、左上の”レベニュー（ペイヤーセグメント別）”の棒グラフから、Susan の棒グラフをクリックしましょう。すると、Susan 以外の棒グラフは半透明のような色に替わり、ドーナツグラフの数字や、ピボットテーブルの数字にも変化があるのが確認できるでしょう。



つまり、クリックした部分の Account Rep (Susan) でクイックフィルタが起動し、ドーナツグラフやピボットテーブルが Susan のデータだけになるよう絞り込まれた状態になっています。再度 Susan の棒グラフをクリックすると、フィルタが解除されます。他の人でもフィルタが効くか確認してください。

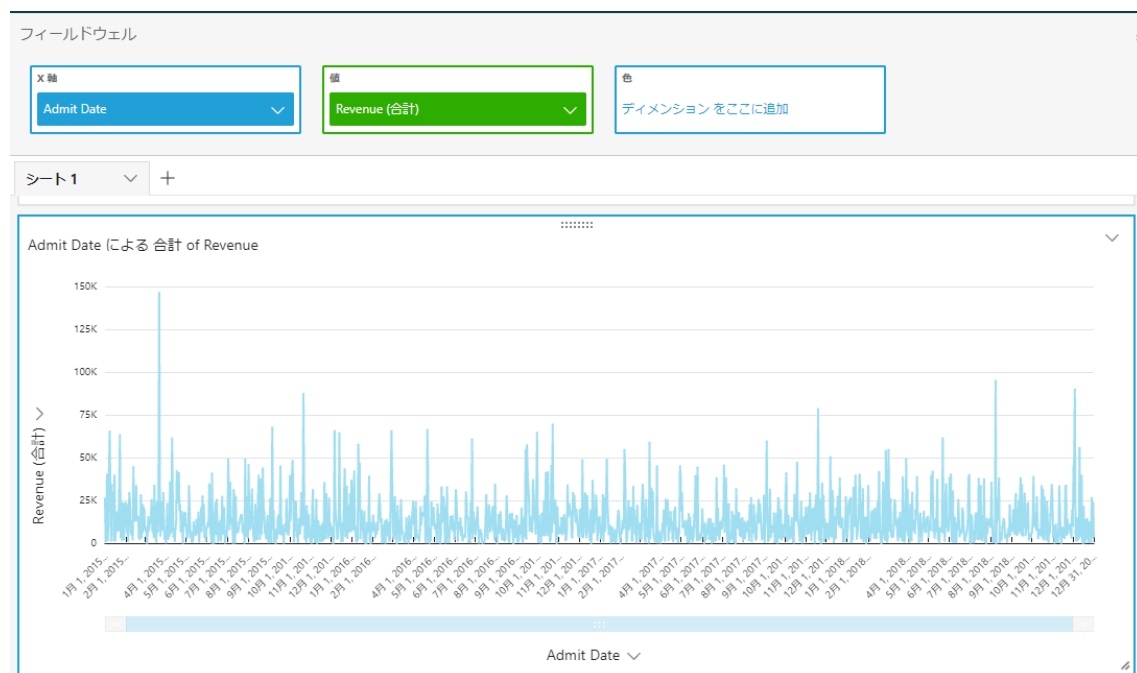
これでハンズオン 1 は完了です。時間が余った方は以下のオプション演習を実施していただくか、ハンズオン 2 に進んでください。

（オプション演習）折れ線グラフと予測機能（ENTERPRISE EDITION が必要）

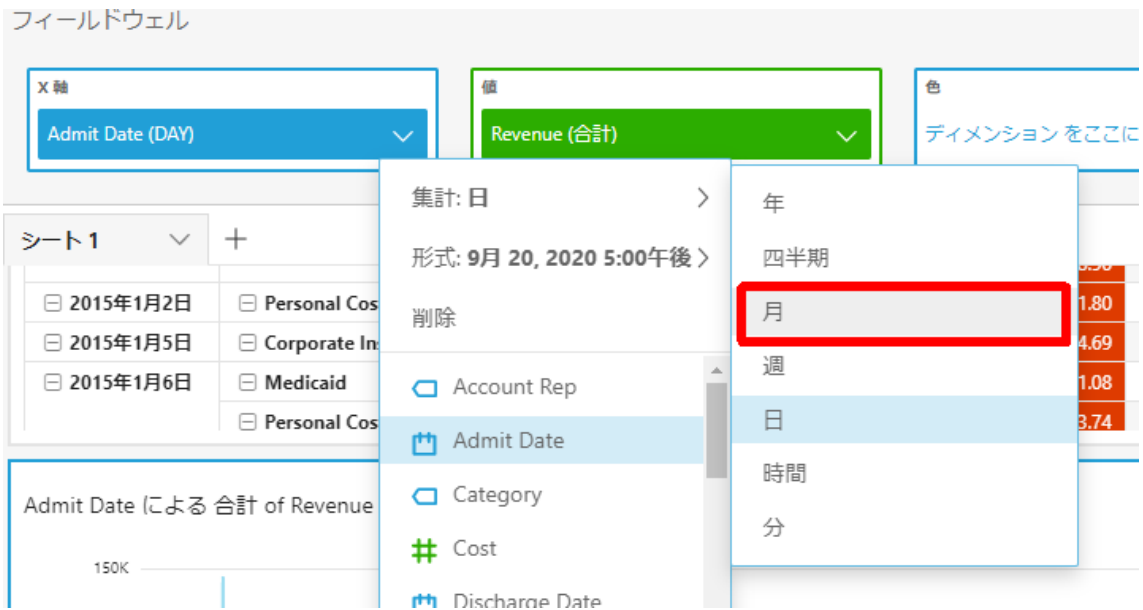
レベニュー（売上）の折れ線グラフを作成し、QuickSight の ML インサイト機能で 6 ヶ月先の売上の予想を作成します。

左上の”+追加”->“ビジュアルを追加”を押すと新しいビジュアルが追加されます。その新しいビジュアルが選択された状態で、ビジュアルタイプとして、“折れ線グラフ”（上から 2 段目段、左から 6 つ目）を選択してください。また、ビジュアルの右下隅をドラッグして、横幅を大きくしてください。

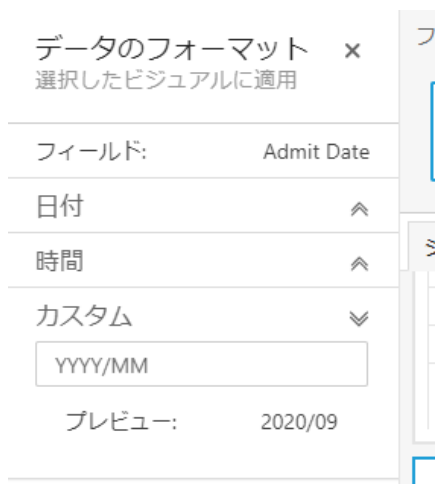
フィールドウェルには、“X 軸”、“値”、“色”が用意されます。今回は日時の経過によるレベニュー推移を表示するので、“X 軸”には Admit date を、“値”には Revenue を D&D します。（D&D せず、クリックするだけでも適切なフィールドウェルに設定されます）



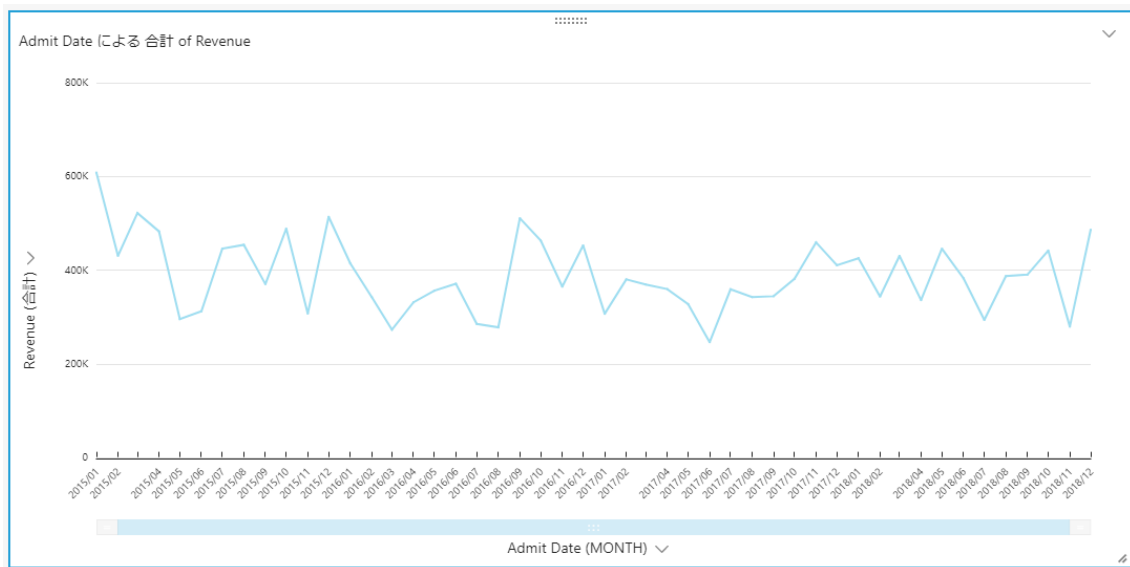
図のように、日付ごとのレベニューが表示されますが、粒度が細かすぎて見づらいため、月ごとのサマリーに変更します。フィールドウェルの X 軸にある Admit Date の”V”をクリックして、集計=>月 を選択します。



同様に、Admit Date の”V”をクリックして、形式=>その他のフォーマットオプションを選択します。左ペインで、”日付”ではカスタムを選択、その下の方にある”カスタム”で YYYY/MM と入力します。



これで、以下のような表示になります。



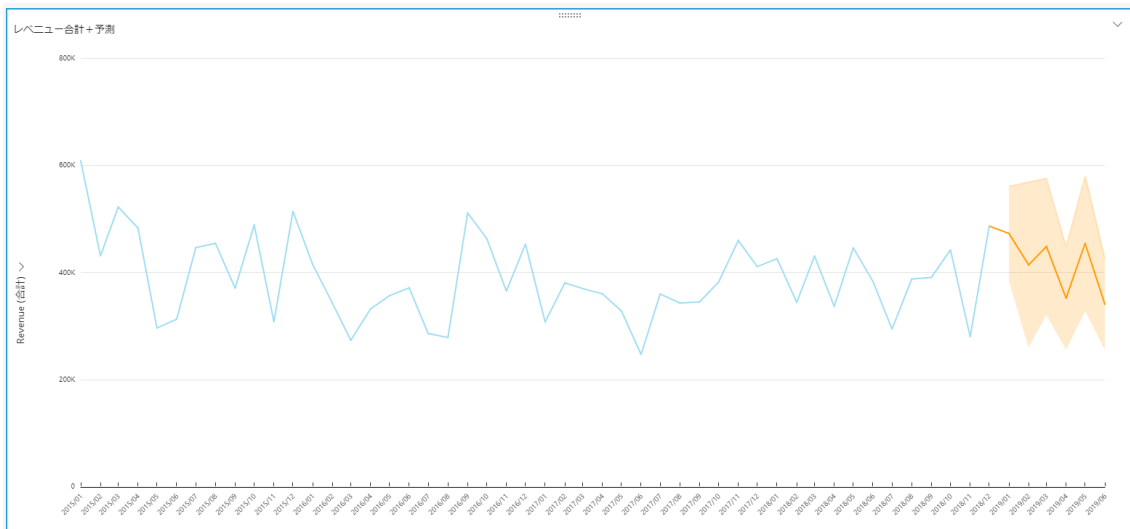
これで、レベニューの月ごとの推移が表示できましたので、これに予測を追加します。予測は QuickSight の ML（機械学習）インサイトの機能の 1 つで、与えられたデータから予測値を出力します。

この予測に使うモデルは、時系列に変化するデータに適した学習済のもので、ユーザ側では調整できません。しかし、QuickSight には Amazon SageMaker との連携機能が用意されており、これを利用すると、ユーザ独自に学習させたモデルを使った分析が可能です（本ハンズオンでは SageMaker との連携は扱いません）。

ではビジュアルの右上にある“…”から”予測を追加”をクリックします。

左ペインに予測プロパティが表示されます。ここで”期間を進める”が 14 になっていますが、今回は半年先までの予測がほしいのでここを 6 にし、適用をクリックします。

最後にビジュアルのタイトルを”レベニュー合計+予測”として、以下のようなビジュアルが完成しました。



上図で、青色の線で表示されているのが、実際の値、オレンジ色の線で表示されているのが先程設定した予測値です。あくまで機械的な予測ですので、絶対ではありません。また薄いオレンジ色で塗られた領域は、予測の「幅」を示しています。

(オプション演習) テーマを切り替える

QuickSightにはテーマ機能があります。これは画面の見た目を指定しておいたパレット（色の設定）に合わせて変えるものです。デフォルトでいくつかプリセットパレットが用意されているので、そちらに変更してみましょう。

画面左端の、“テーマ”アイコンをクリックして、“Starter themes”の下にある”Midnight”と書かれたテーマの”・・・”ボタン=>”適用”を押してください。画面の色合いやビジュアル間のボーダー等が変更されるのが確認できます。テーマの別名で保存を選択すると独自のテーマを作成可能になりますので、試してみてください。



ハンズオン 2：より進んだ可視化

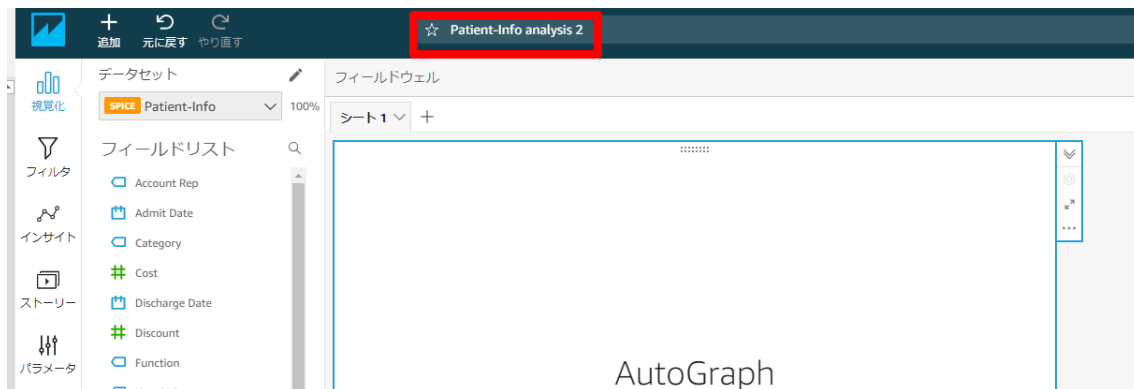
ハンズオン 2 を始める前に

ハンズオン 2 でも、ハンズオン 1 と同じデータセットを使用します。まだデータセットを作成していない方は、「事前準備」の章を確認して準備をしてください。

データセットが準備できたら、新しい分析（Analysis）を作成します。QuickSight トップ画面から（トップ画面に戻るには、画面左上の QuickSight アイコンをクリックします）、左上の”新しい分析”をクリックし、「事前準備」の章で作成済のデータセット（Patient-Info）をクリック=>分析の作成をクリックし、新規の分析を作成します。



新しい分析では、一番上にある分析のタイトルを任意の名前に変更してください（デフォルトのままだと、ハンズオン 1 で作成したものと同名前で区別が付きにくくなるため）。ここでは”Patient-Info analysis 2”としました。



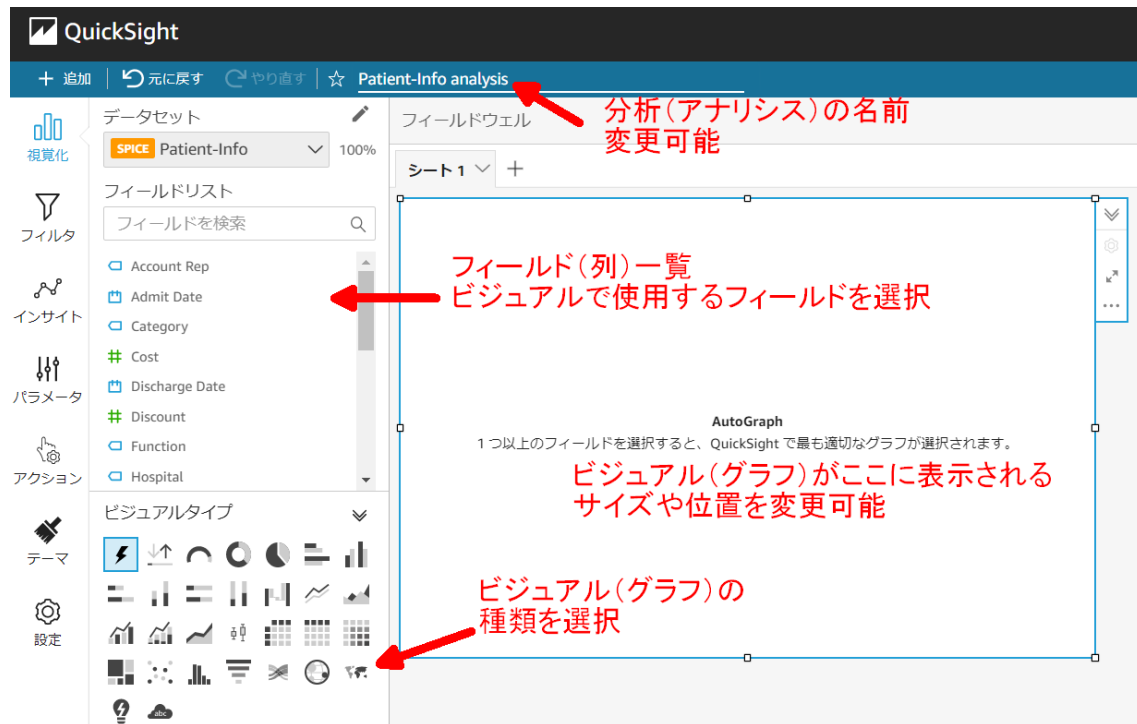
右上に“インポートの完了”ダイアログが出ている場合は x を押して閉じてください。



では、ハンズオン 2 を開始しましょう。

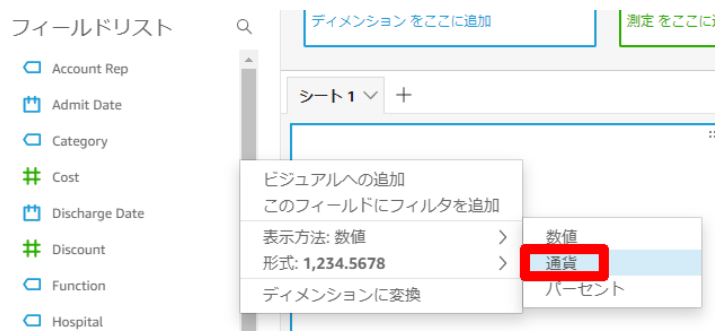
垂直棒グラフの作成とドリルダウンの実現

新規にビジュアルを作成すると以下のような画面が表示されます。



これは分析（Analysis）を作っていく画面です。左側にフィールド（列）と、ビジュアル（グラフ）の一覧が表示されています。まずビジュアルタイプ（グラフの種類）を選択し、それに必要なフィールドを選択することで可視化されます。ビジュアルは左上の「+」を押すことで複数追加できますし、サイズや位置を変更することが可能です。

まず一部のフィールドの表示形式を修正します。フィールドリストにある **Cost** の上にマウスカーソルを移動すると、「V」が表示されますので、それをクリックし、表示方法の「数値」を「通貨」に変更してください。すると Cost の左側にあるアイコンが\$マークに変わります。



同様に、**Discount**、**Price**、**Profit**、**Revenue** も「通貨」に変更してください。

ここからビジュアルを作成していきます。画面に空白のビジュアルが1つ表示されているので、まずは左下のビジュアルタイプ一覧から**垂直棒グラフ**（最上段、一番右）を選択してください。すると以下のような画面になります。



QuickSight での可視化の基本はビジュアルを追加した後に、フィールドをフィールドウェル上に設定していくことです。

※上図のようにフィールドウェルがビジュアルの上部に表示されない場合は、まだビジュアルタイプを選択していないか、もしくは折り畳まれている場合です。その場合は「フィールドウェル」と書かれている部分をクリックして広げてください。

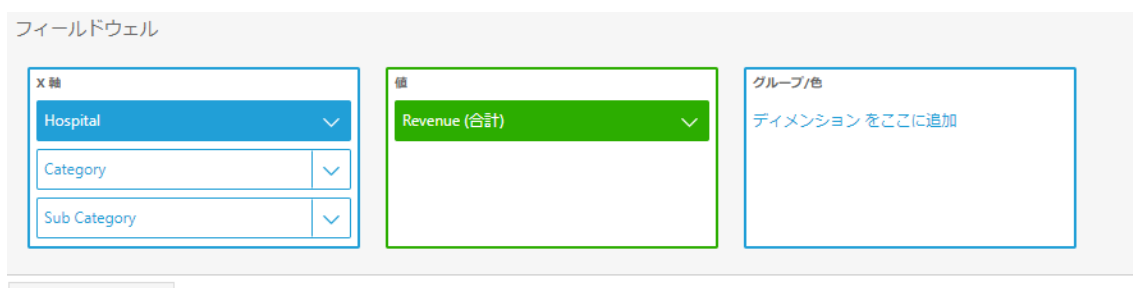
今回フィールドウェルには、“X 軸”、“値”、“グループ/色”が用意されます。ここでは、病院（Hospital）ごとの利益（Revenue）を見るので、まず、フィールドリストの Hospital をクリックしてください。自動的にフィールドウェルの“X 軸”に Hospital が挿入されます。これはフィールドの型や定義から自動的に判断してフィールドが追加される仕組みです。続いて、Revenue をクリックすると、“値”に Revenue(合計)として挿入されます。

これで病院ごとの Revenue の合計が得られました。これにドリルダウン（階層）を付けていきます。今回は Hospital の下階層として、Category を置き、そのさらに下に Sub Category を配置します。

フィールドリストから Category ドラッグ&ドロップ（D&D）で、X 軸にすでにある Hospital の下に移動させます。このとき、下図のように“ドリルダウンレイヤーの追加”と表示される場所でドロップしてください。

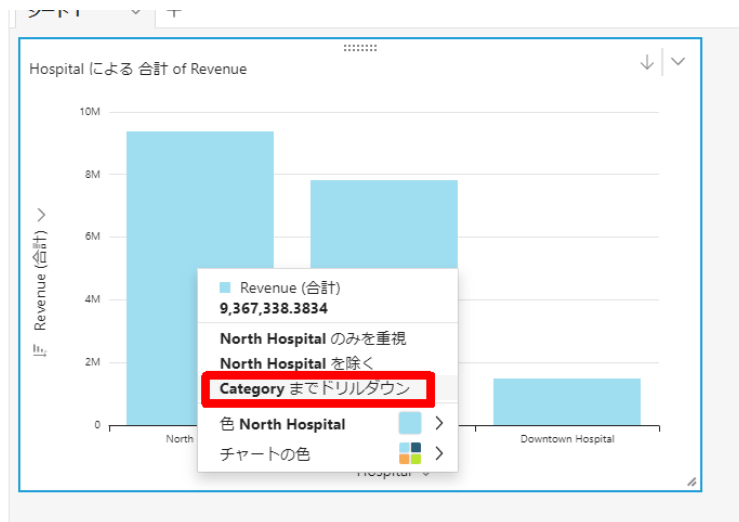


同様に、Category の下に、Sub Category をドリルダウンレイヤーとして D&D してください。結果、フィールドウェルは以下のようになります。

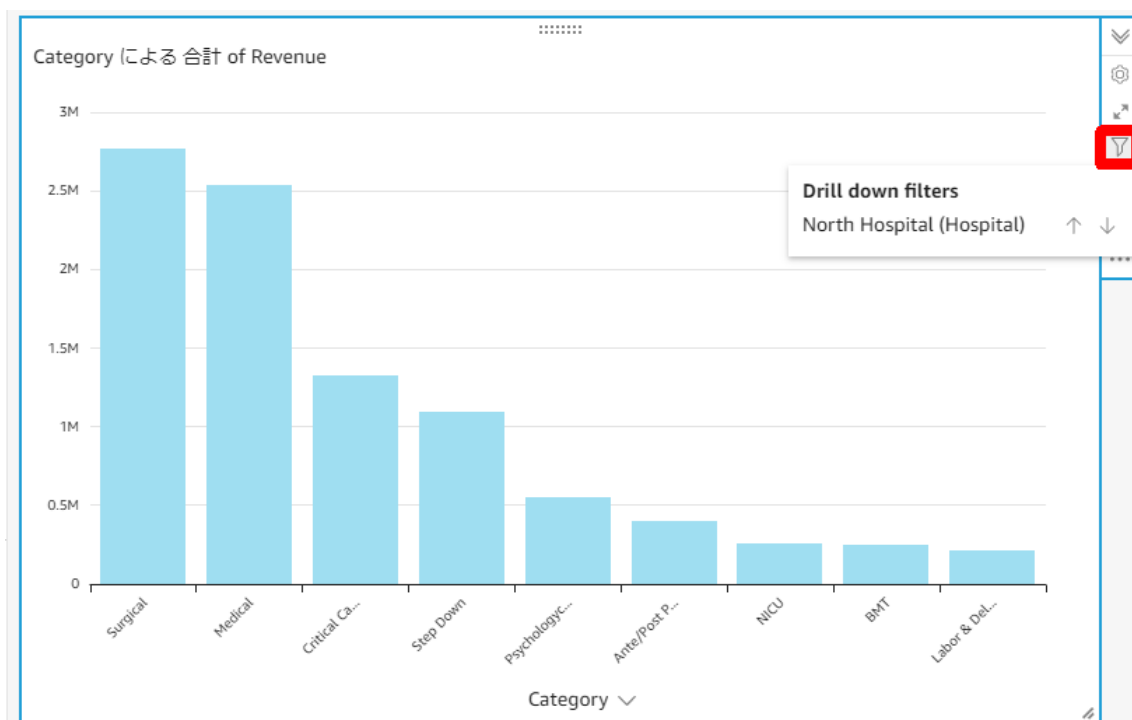


これでドリルダウンの階層が設定できましたので、実際に動かしてみましょう。

グラフの North Hospital をクリックし、メニューの”Category までドリルダウン”をクリックします。

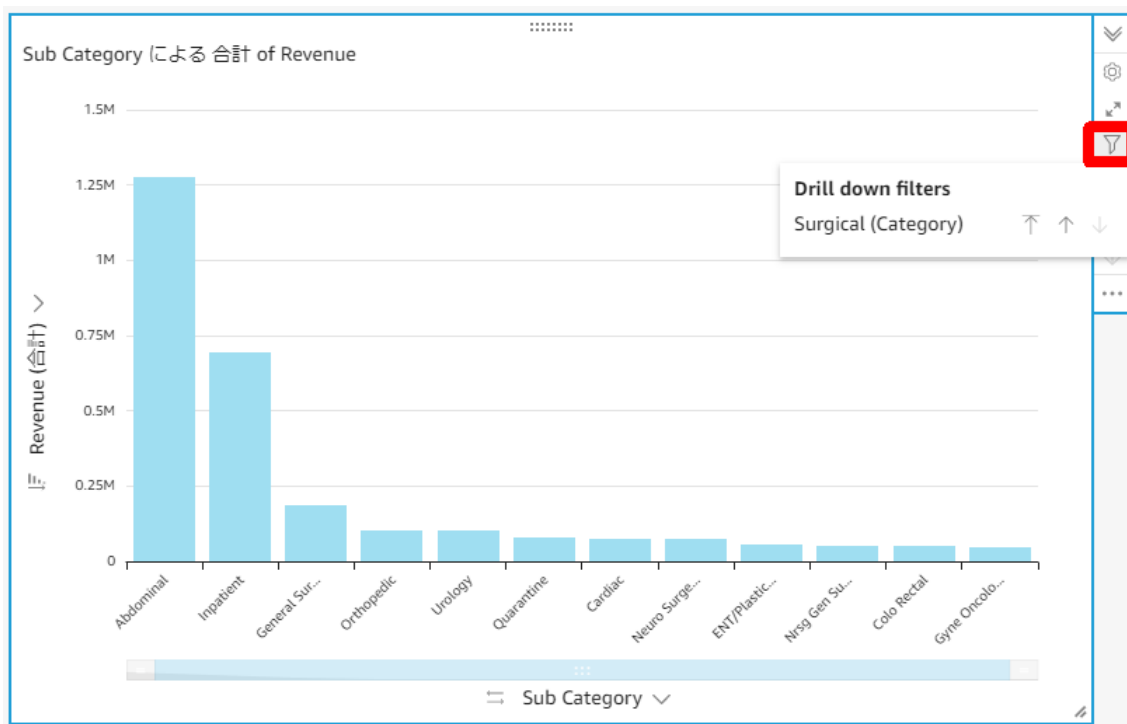


すると以下のように、”North Hospital”の中での Category 別の Revenue の表示に替わります。

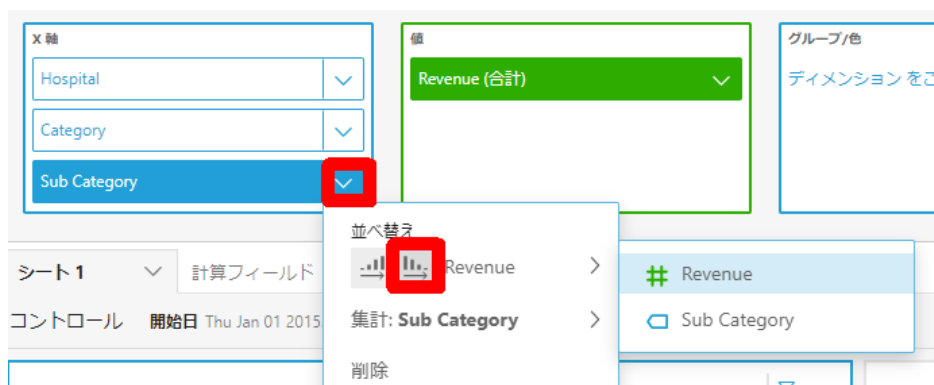


ビジュアル右側の斗マークをクリックすると、“North Hospital(Hospital)”と表示されます。これが North Hospital だけの範囲でドリルダウンが行われていることを示しています。

さらに一番左に表示されている”Surgical”をクリックし、“Sub Category までドリルダウン”をクリックします。これで、North Hospital において、Surgical 内での Sub Category ごとの Revenue が表示されます。



※補足上記の順でデータが並ばなかった場合は、下図のように Sub Category の並べ替えを大きい順に変更してください

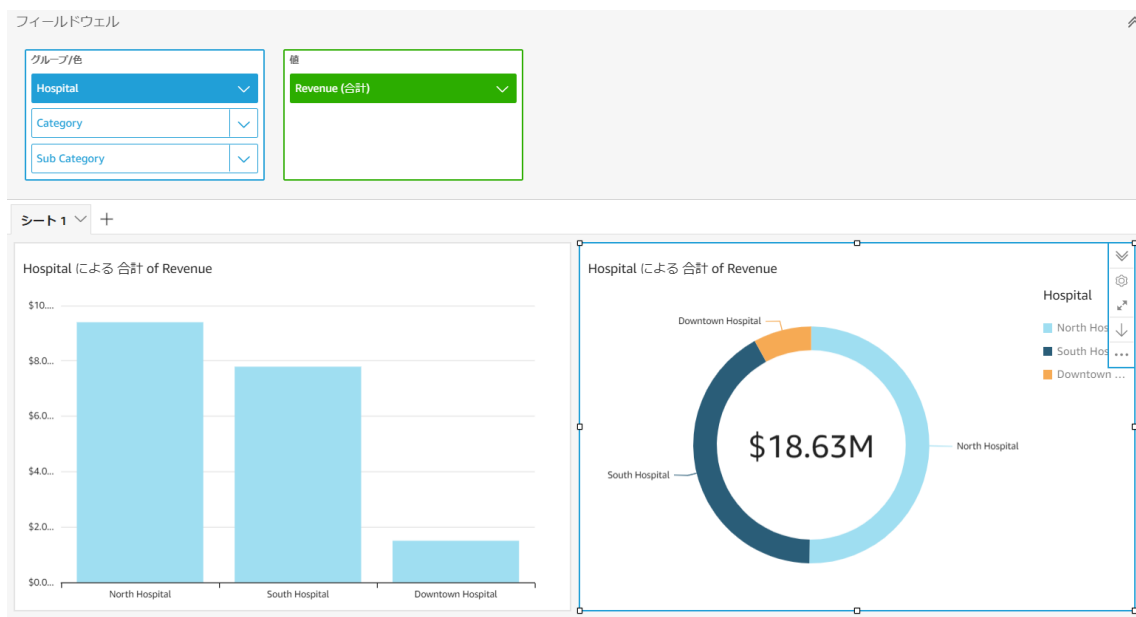


ドリルアップをするには、ビジュアル右にある上向きの矢印（↑）をクリックします。また、
↑を押すと最上位（Hospital）までドリルアップします。↑をクリックしてください。

ビジュアルの調整とコピー

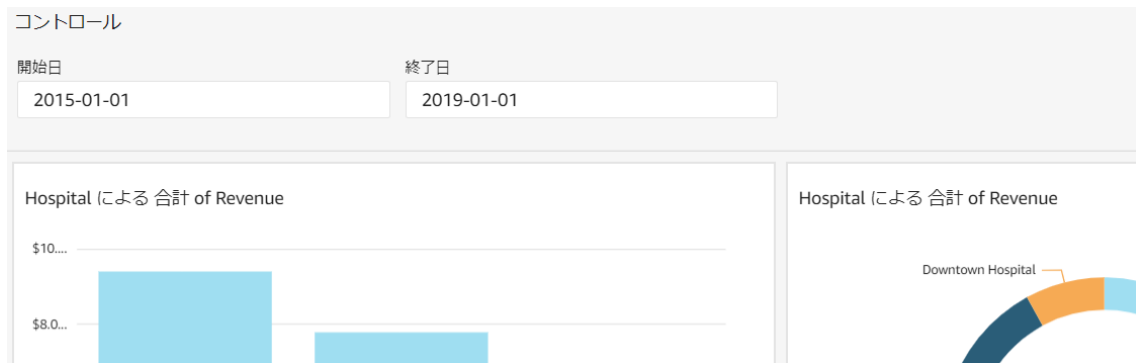
次にこの垂直棒グラフをコピーして別のビジュアルに変更します。ビジュアルの右側にある“…”をクリックして、“ビジュアルの複製”をクリックすると複製されます。

この新しいビジュアルが選択された状態（新しく作られたビジュアルが青色線で囲まれている状態）で、左下のビジュアルタイプ一覧からドーナツグラフ(最上段、左から4つ目)を選択してください。以下のようにフィールドウェルの設定はそのままにビジュアルが変更されます。



コントロールの作成とフィルタ

ここでは、ここまで作成したビジュアルに対して、ユーザが操作できるコントロールという GUI パーツ（下図）を画面上に作成し、データをフィルタ（絞り込み）可能にします。



コントロールは、Reader ユーザ（分析を作成する権限を持たない、読み取りのみのユーザ）でも操作できるため、コントロールを付けた分析を公開することで、インタラクティブなダッシュボードが実現できます。以下のような流れで作成します。

- ① 最初に**パラメータ**を作成します。これは変数のようなもので、ユーザが指定した値を保存する領域です
- ② 次に、①で作ったパラメータにリンクさせた、**コントロール**（GUI）を作成します。これにより GUI で入力した値がパラメータ（変数）に格納されます
- ③ 最後に、②で作成したパラメータを条件に指定した**フィルタ**を作成します。これにより GUI の操作がフィルタ条件に反映され、絞り込みが行われます

まずパラメータから作成します。画面左端の“パラメータ”をクリックし、“作成...”を選択します。パラメータには英数字の名前を付ける必要があります。今回は BeginDate としました。また、データタイプは“日時”にし、時間詳細度は“日”のままにし、静的デフォルト値には 2015-01-01 を指定し、“作成”を押します。

新しいパラメータを作成

×

パラメータを使用して、フィールド、フィルタ、シートの値を動的に制御します

名前
BeginDate

データタイプ (作成後は変更不可)
日時

時間詳細度
日

デフォルトの日付
2015-01-01

Dynamic default
動的デフォルト値を設定

キャンセル

作成

作成すると、続いてパラメータの接続先を選択するダイアログが出るので、コントロールを選択します。

パラメータが追加されました

パラメータを接続

パラメータ、新しいコントロール、フィルタの組み合わせを使用してフィルタを作成します。

フィルタまたは計算済みフィールドに向けた新しいコントロールを作成します。

計算済みフィールドでパラメータを使用します。

パラメータで URL アクションを作成します

閉じる

フィルタ

コントロール

計算フィールド

カスタムアクション

以下のダイアログでは、ユーザに見せるコントロールの名前を付けます。ここでは日本語を含む文字列を使用できます。今回は”開始日”としました。スタイルは規定の”日付選択ツール”のままにし、”追加”を押します。

パラメータにコントロールを追加 ×

パラメータ
BeginDate

表示名
開始日

スタイル
日付選択ツール

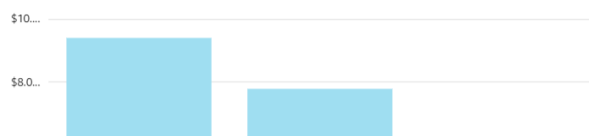
キャンセル 追加

これで、ビジュアルの上方に”開始日”のコントロールが作成されます。（コントロールと書かれている部分をクリックして開いてください）

コントロール

開始日
2015-01-01

Hospital による 合計 of Revenue



これで、1つ目のパラメータとコントロールは作成できました。日付を範囲でフィルタ（絞り込み）するには、もう一つ、終了日のパラメータとコントロールが必要です。左ペイン「パラメータ」の右側にある”+”を押し、以下の条件でもう一つ作成してください。

- パラメータ名：EndDate
- データタイプ：日時型
- 時間詳細度：日
- デフォルトの日付：2019-01-01
- コントロールの名前：終了日

以下ようになります。

パラメータ	+	フィールドウェル	グループ/色	Hospital	値	Revenue (合計)
BeginDate	▽	シート 1	+			
EndDate	▽	コントロール				
		開始日		終了日		
		2015-01-01		2019-01-01		

これで2つのコントロール（開始日と終了日）が作成され、入力された値はそれぞれに紐付いたパラメータ（BeginDate と EndDate）に入るようになりました。次にパラメータを条件にしたフィルタを作成します。画面左端から”フィルタ”を選択、”作成...”を選択して、フィールドは”Admit Date”を選択します。

フィルタの編集

☒ 該当するすべてのビジュアル

Admit Date
 次の間 - なし

フィルタタイプ
 日付と時刻の範囲
 次の間

☒ パラメータを使用

開始日のパラメータ
 BeginDate

☒ 開始日を含める

終了日のパラメータ
 EndDate

☐ 終了日を含める

または

左側ペインに”Admit Date”のフィルタが作成されるのでクリックして開き、左図のように設定します。

一番上にある選択で”該当するすべてのビジュアル”を指定。これでフィルタが Analysis 全体のビジュアルに適用されます

フィルタタイプは日付と時刻の範囲と次の間を選択し、パラメータを使用にチェックを入れます。

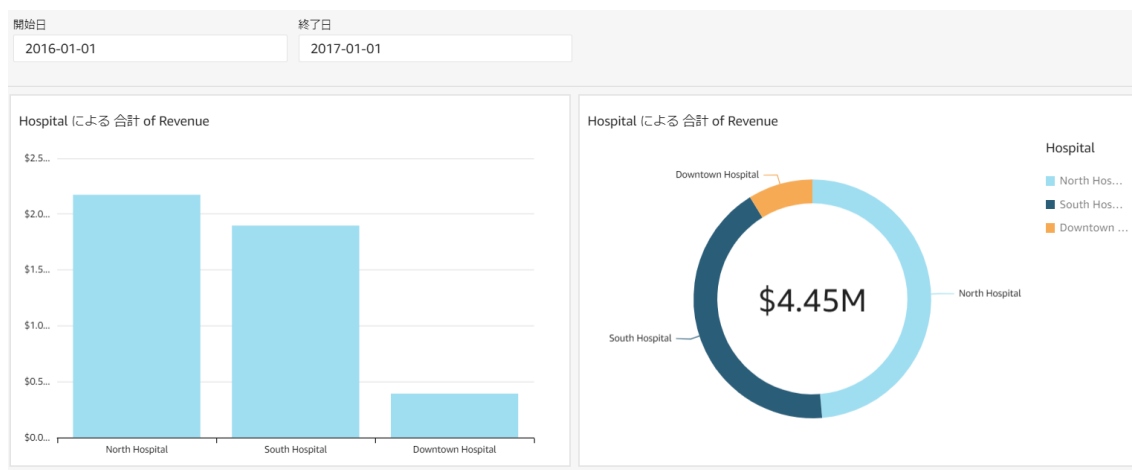
開始日のパラメータには BeginDate を指定し、”開始日を含める”にチェックを入れます。チェックを入ると、指定した値を含んだ範囲指定になります。（＜と≦の違い）

終了日のパラメータには EndDate を指定し、”終了日を含める”のチェックは入れません。

設定できたら、”適用”をクリックします。

これで、コントロール（GUI）で入れた値がパラメータ（変数）に入り、それによりフィルタが実施される準備が整いました。開始日や終了日を変更し、2つのビジュアルがどう変わるか

を確認してください。（例えば、開始日を 2016-01-01 に、終了日を 2017-01-01 に指定すると、2016 年のみのデータに絞り込まれます）



[補足]本ハンズオンでは、パラメータを先に作って、コントロールを結びつけ、それをフィルタに反映させるという方法を説明しました。これはパラメータの値を URL から指定する等応用が利く方法ですが、前述のように手順が多くなります。

パラメータが不要な場合はより簡単な手順で作成する方法も用意されており、先にフィルタを作成し、そこからすぐにコントロールを作成することが可能です（フィルターコントロール）。 https://docs.aws.amazon.com/ja_jp/quicksight/latest/user/filter-controls.html

計算フィールド①：シンプルな計算フィールド

ここからは、計算フィールドの機能を使用した分析を作成します。計算フィールド(Calculated Field)とは、擬似的な列の機能です。データソースには無かった列を追加することで、より高度な分析が可能になります。

計算フィールドは、作成中の分析(Analyze)上に追加する方法と、データセット自体に追加する方法の2種類があります。本ハンズオンでは前者を使用します。後者のデータセット自体への追加の場合、利用できる関数等に制限がありますが、計算可能な部分は事前に計算しておき、表示時に毎回計算する必要を無くせるというメリットがあります。

まずはシンプルな計算フィールドを作成します。データを四半期(クォーター、Quarter)単位で分析したいと思っても、データセットにはクォーターフィールド(列)がありません。そこで計算フィールドで擬似的にクォーターフィールドを作成します。

日付自体は Admit Date に存在するため、その日付に extract 関数を使ってクォーターを抽出します。extract は日付から任意のフォーマットで文字列を抽出する関数で、抽出フォーマットを第1引数に、日付フィールドを第2引数に取ります。以下のように使用するとクォーター番号が抽出できます。

```
extract("Q",{Admit Date})
```

これを計算フィールドとして定義します。画面左上の"+追加"から、"計算フィールドを追加"を選択し、計算フィールド名に **Quarter**、計算式に上記式を入力し、"作成"を押します。

計算フィールドを編集する

関数リスト

- abs
- addDateTime
- avg
- avgIf
- avgOver
- ceil
- coalesce
- concat
- count

フィールドリスト

- Account Rep
- Admit Date
- Category
- Cost
- Discharge Date
- Discount
- Function
- Hospital

計算フィールド名

Quarter

計算式

extract("Q",{Admit Date})

作成

これで Quarter という名前で計算フィールドが作成されました。フィールドリストにも Quarter というフィールドが表示されているはずです。この Quarter フィールドを使ってピボットテーブルを作成します。（※もし正しく入力しても計算フィールド作成時にエラーが出る場合は一度ブラウザでリロードをしてから再度作成して見てください）

新しいピボットテーブルを作成する前に、シートを切り替えます。シート名（シート 1 等）の横にある“+”を押すと、新しいシートが作成されます。



シートにはそれぞれ任意の名前を付けることができます。新しいシートの中に 1 つビジュアルが作成されていますので、ビジュアルタイプからピボットテーブル（上から 3 段目、左から 5 つ目）を選択します。ピボットテーブルのフィールドウェルには、以下のようになるよう、クリックや D&D でフィールドを設定してください。

フィールドウェル

行

- Admit Date
- Quarter

列

ディメンションをここに追加

値

- Revenue (合計)

次にフィールドウェル上で、Admit Date の右側にある”V”をクリックして表示を調整します。
集計を”年”に設定し、さらに形式は”2020”のように年だけの表示にしてください。

以下ようになります。

フィールドウェル

行
Admit Date (YEAR) ✓
Quarter ✓

列
ディメンションをここに追加

値
Revenue (合計) ✓

シート 1 計算フィールド ✓ +

Admit Date	Quarter	Revenue
2015	1	\$1,561,877.72
	2	\$1,092,500.97
	3	\$1,271,644.94
	4	\$1,311,380.01
2016	1	\$1,032,065.62
	2	\$1,060,424.57
	3	\$1,076,119.85
	4	\$1,282,178.57

年単位で Admit Date が表示され、その横に Quarter フィールドが第 1 四半期から第 4 四半期まで表示されている事が分かります。値として四半期毎の Revenue(合計)が表示されています。

計算フィールド②：テーブル計算関数（RUNNINGSUM）

作成したピボットテーブルでは四半期毎の Revenue がありますが、これに加えて、各四半期での**その年の“累計”**を表示するようにします。つまり、第2四半期は第1四半期+第2四半期の値を、第3四半期には、第1四半期+第2四半期+第3四半期の値という累計を新たな値として追加します。

こういった“ある期間ごとの集計”の機能は、QuickSight では**テーブル計算関数**として提供されており、累計には runningSum 関数が用意されています。（※関数一覧等、ドキュメントへのリンクは資料末尾にあります）

runningSum は第1引数に、集計する値を（この場合は Sum(Revenue)）、第2引数にはソート順の指定、第3引数にはパーティション（今回は年単位）を指定します。

runningSum(sum(Revenue),[{Admit Date} ASC],[{Admit Date}])

今回は上記の計算式で、**YearToDateSales** という名前の計算フィールドを作成してください。（※フィールド名は大文字小文字を区別します。この後、この計算フィールドを別の計算フィールド内で使用するため、名前の打ち間違いに注意してください）

フィールドリストにある YearToDateSales の上にマウスカーソルを移動すると、“V”が表示されますので、クリックし、表示方法を「通貨」に設定します。



変更できたら、以下のように値としてフィールドウェルに追加します。ピボットテーブルは以下のようになります。

フィールドウェル

行

Admit Date (YEAR) ▼

Quarter ▼

列

ディメンションをここに追加

値

Revenue (合計) ▼

YearToDateSales (カスタム) ▼

シート1 計算フィールド ▼ +

Admit Date and Quarter による 合計 of Revenue and Yeartodatesales

Admit Date	Quarter	Revenue	YearToDateSales
2015	1	\$1,561,877.72	\$1,561,877.72
	2	\$1,092,500.97	\$2,654,378.69
	3	\$1,271,644.94	\$3,926,023.63
	4	\$1,311,380.01	\$5,237,403.64
2016	1	\$1,032,065.62	\$1,032,065.62
	2	\$1,060,424.57	\$2,092,490.20
	3	\$1,076,119.85	\$3,168,610.04
	4	\$1,282,178.57	\$4,450,788.61
2017	1	\$1,058,493.45	\$1,058,493.45
	2	\$935,728.42	\$1,994,221.88
	3	\$1,048,166.83	\$3,042,388.70
	4	\$1,252,884.08	\$4,295,272.78
2018	1	\$1,200,943.33	\$1,200,943.33

YearToDate 列がその四半期までの Revenue の積算になっているのが確認できます。

(オプション) 計算フィールド③: テーブル計算関数

(PERCENTDIFFERENCE)

ピボットテーブルに、昨年同四半期との比較を追加します。例えば 2018 年第 3 四半期時点の累積レベニューは、2017 年第 3 四半期時点と比較して何%の違いがあるかといった計算です。こういった計算にはテーブル計算関数の percentDifference を使用します。

percentDifference は第 1 引数に、比較する値、第 2 引数にソート基準、第 3 引数にインデックスを取ります。インデックスとは、相対的にどの位置にある値を比較対象にするかを指します。以下の式を見てください。

percentDifference(YearToDateSales, [{Admit Date} ASC], -4, [])

インデックスに -4 と書かれています。つまり **4 つ前**のデータを比較対象にします。今回のピボットテーブルは四半期単位で行を構成しているため、4 つ前は 1 年前という事になります。

※補足: このように各テーブル計算関数はデフォルトでは、**表示された時点のデータの並びに依存した形**で動作しているという点に注意してください。つまり、-4 が 1 年前のデータを指すのは、あくまで今回のピボットテーブルが四半期単位で集約されているためです。表示や集約の結果に依存せずに演算を行う機能としては LAA (Level-Aware Aggregation) が用意されています。LAA はオプションのハンズオンで解説しています。

ではこれまでと同様に、上記の式を **YTD % Difference** という名前の計算フィールドとして定義し、今回は比率なので、表示方法を”パーセント”に変更した上で、フィールドウェルの”値”の一番下に追加します。

以下のような表示になります。

行	列	値
Admit Date (YEAR) ▼	ディメンションをここに追加	Revenue (合計) ▼
Quarter ▼		YearToDateSales (カスタム) ▼
		YTD % Difference (カスタム) ▼

シート 1

計算フィールド ▼

+

Admit Date and Quarter による 合計 of Revenue, Yeartodatesales, and Ytd % Difference				
Admit Date	Quarter	Revenue	YearToDateSales	YTD % Difference
2015	1	\$1,561,877.72	\$1,561,877.72	
	2	\$1,092,500.97	\$2,654,378.69	
	3	\$1,271,644.94	\$3,926,023.63	
	4	\$1,311,380.01	\$5,237,403.64	
2016	1	\$1,032,065.62	\$1,032,065.62	-33.92%
	2	\$1,060,424.57	\$2,092,490.20	-21.16%
	3	\$1,076,119.85	\$3,168,610.04	-19.29%
	4	\$1,282,178.57	\$4,450,788.61	-15.01%
2017	1	\$1,058,493.45	\$1,058,493.45	2.56%
	2	\$935,728.42	\$1,994,221.88	-4.69%
	3	\$1,048,166.83	\$3,042,388.70	-3.98%
	4	\$1,252,884.08	\$4,295,272.78	-3.49%
2018	1	\$1,200,943.33	\$1,200,943.33	13.45%
	2	\$1,166,292.13	\$2,367,235.46	18.7%
	3	\$1,073,458.86	\$3,440,694.32	13.09%
	4	\$1,209,204.94	\$4,649,899.27	8.25%

2015 年は、その前の年がデータに存在しないため表示されていませんが、2016 年からは前年同四半期時点での累積としての差が計算されています。

(オプション) 計算フィールド④：テーブル計算関数 (PERCENTOFTOTAL)

ピボットテーブルに、各四半期のレベニューが、その年の売上の何%を占めているかの情報を追加します。この場合、関数 `percentOfTotal` が利用できます。第 1 引数に演算式、第 2 引数にパーティショニングの基準（この例では、1 年ごと）を指定します。

`percentOfTotal(sum(Revenue),[{Admit Date}])`

これまでと同様に上記の式を、**PercentOfTotalSales** という計算フィールド名で定義し、表示方法を”パーセント”に変更した上で”値”のフィールドウェル一番下に追加してください。以下のようになります。PercentOfTotalSales 列で、各四半期がその年のレベニューの何%を占めているかが表示できました。

The screenshot shows a pivot table configuration interface with three panels: Rows (行), Columns (列), and Values (値). The Rows panel contains 'Admit Date (YEAR)' and 'Quarter'. The Columns panel is empty with a placeholder 'ディメンション をここに追加'. The Values panel contains 'Revenue (合計)', 'YearToDateSales (カスタム)', 'YTD % Difference (カスタム)', and 'PercentOfTotalSales (カスタム)'. Below the configuration is a table titled 'Admit Date and Quarter による 合計 of Revenue, Yeartodatesales, Ytd % Difference, and Percentoftotalsales'.

Admit Date	Quarter	Revenue	YearToDateSales	YTD % Difference	PercentOfTotalSales
2015	1	\$1,561,877.72	\$1,561,877.72		29.82%
	2	\$1,092,500.97	\$2,654,378.69		20.85%
	3	\$1,271,644.94	\$3,926,023.63		24.28%
	4	\$1,311,380.01	\$5,237,403.64		25.03%
2016	1	\$1,032,065.62	\$1,032,065.62	-33.92%	23.18%
	2	\$1,060,424.57	\$2,092,490.20	-21.16%	23.82%
	3	\$1,076,119.85	\$3,168,610.04	-19.29%	24.17%
	4	\$1,282,178.57	\$4,450,788.61	-15.01%	28.8%
2017	1	\$1,058,493.45	\$1,058,493.45	2.56%	24.64%
	2	\$935,728.42	\$1,994,221.88	-4.69%	21.78%
	3	\$1,048,166.83	\$3,042,388.70	-3.98%	24.4%
	4	\$1,252,884.08	\$4,295,272.78	-3.49%	29.16%
2018	1	\$1,200,943.33	\$1,200,943.33	13.45%	25.82%
	2	\$1,166,292.13	\$2,367,235.46	18.7%	25.08%
	3	\$1,073,458.86	\$3,440,694.32	13.09%	23.08%
	4	\$1,209,204.94	\$4,649,899.27	8.25%	26%

ML インサイトによる異常値検出の設定

ここからは、QuickSight の ML(機械学習)インサイトの機能のうち、異常値検出とナラティブを体験します。

まず ML インサイトのハンズオン用に、シートの横にある”+”ボタンを押し、別のシートを作成します。シートには任意の名前を付けることができます。



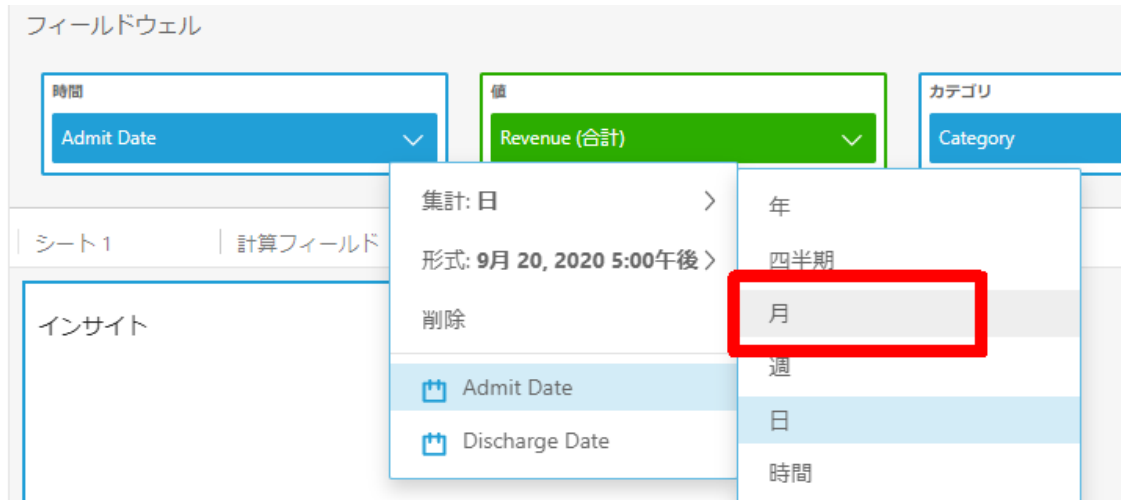
新しいシートには、空白のビジュアルが 1 つ用意されていますが、ここでは使わないため、ビジュアルの右の”…”を押し、”削除”を選択してください。

まず、異常検知の計算を設定します。この**異常値検出の機能を利用するにはユーザ料金とは別に料金が必要**になります。 (<https://aws.amazon.com/jp/quicksight/pricing/>)

画面左上の”+追加”から、”インサイトを追加”を選択し、”異常値検出”を選択し、”選択”を押します。”時間”、”値”、”カテゴリ”のフィールドウェルが用意されますので、それぞれ、以下のようにフィールドを配置してください。



この際、“時間”の Admit Date は以下のように”集計:月”を必ず選択してください。（月単位にしないと、計算量が増えてしまうため）



次にインサイトの”今すぐ始める”を押し、設定画面に移動します。

設定はほとんどデフォルトのままで問題ありませんが、”上位の寄与要因”をクリックして、設定を変更します。これは、異常値を検出した際、その異常がどのフィールドの組み合わせによって起こったのかという計算（寄与率の算出）を行うものです。対象とするフィールドは最大4つまで指定できるので、ここでは以下のように、Patient ID、Payer Segment、Physician、Service の4つを選択します。

スケジュールオプション 毎月 (09:58 (Asia/Tokyo) に 2021/03/31 を開始) ▼

上位の寄与要因 ^

フィールドを選択
異常の背後にある主要な要因を見つけます。最大 4 つのフィールドを選択できます。

Patient ID, Payor Segment, Physician, Service ▼

フィールドを検索 🔍

Hospital	<input type="checkbox"/>
Patient ID	<input checked="" type="checkbox"/>
Payor Segment	<input checked="" type="checkbox"/>
Physician	<input checked="" type="checkbox"/>
Priority	<input type="checkbox"/>
Region	<input type="checkbox"/>
Service	<input checked="" type="checkbox"/>
State	<input type="checkbox"/>
Sub Category	<input type="checkbox"/>

上記のようにチェックをいれて、右上の”保存”を押して元の画面に戻り、インサイトの画面にある”今すぐ実行”を押すと、異常値を検出するための計算が始まります。今回の設定の場合、演算には約 2-3 分間かかります。これを待つ間に、もう一つのナラティブ機能を確認します。

ML インサイトによるナラティブ

ナラティブは、インサイトを図ではなく人間の言葉や箇条書きで表現する機能です。

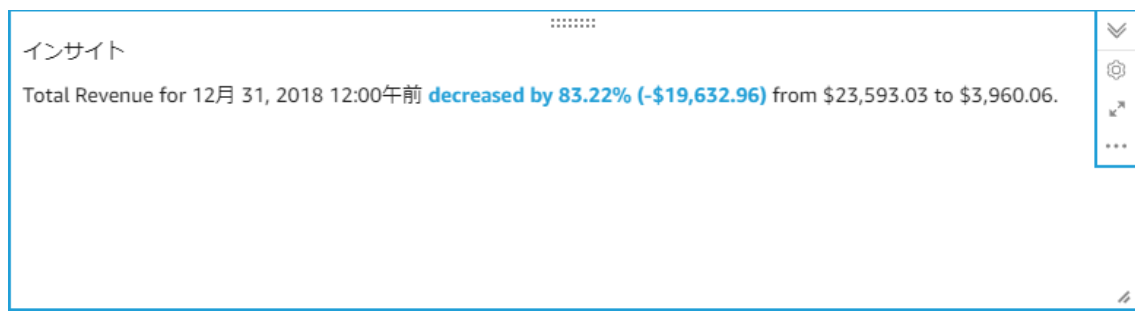
画面左上の“+”から“インサイトを追加”を選択し、“前期比”を選択し、“選択”を押します。これは、一定期間前と比較した現在の状況を説明するナラティブ（文章）です。

フィールドウェルが表示されるので、以下のように設定してください。

フィールドウェル

時間 Admit Date	値 Revenue (合計)	カテゴリ ディメンションをここに追加
------------------	-------------------	-----------------------

すると、以下のような結果が得られます。



このサンプルデータは最新の日時が 2018 年 12 月 31 日のため、31 日と、その前日と比較した数字が文章で表示されているのが分かります。これがナラティブ機能です。

ナラティブはデフォルトで多数のプリセットが用意されていますが、自分でも文章をカスタマイズする事が可能です。インサイト右の“...”をクリック⇒“説明をカスタマイズ”を選択してください。ナラティブエディタが表示されます。

説明を編集 ①

コードを挿入 12/パラグラフ

Total PeriodOverPeriod.metricField.name for PeriodOverPeriod.currentTimeValue.formattedValue If PeriodOverPeriod.percentDifference.value >= 0 increased by If PeriodOverPeriod.percentDifference.value < 0 decreased by PeriodOverPeriod.percentDifference.formattedAbsoluteValue (PeriodOverPeriod.absoluteDifference.formattedValue) from PeriodOverPeriod.previousMetricValue.formattedValue to PeriodOverPeriod.currentMetricValue.formattedValue .

計算
パラメータ
関数

+ 計算を追加

GUI のリッチテキストエディタで、中に式や IF 文が入っているのが確認できます。ナラティブの文章はこのような形でロジックを組むことで実現されています。

今回は編集せず、右上の”キャンセル”を押して、前の画面に戻ります。

ML インサイトによる異常値検出の内容を確認する

先に実行しておいた異常値検出の計算が完了すると以下のような画面になっているはずです。

インサイト

The top anomaly detected on 12月 1, 2018 12:00午前 was:

- **Total Revenue** for **Labor & Delivery** at \$54,246.84, which was **higher** than the expected \$7,824.15

[異常の探索](#)

インサイトでは1つの異常（Anomaly）が12月に発見されたとあり、内容としては Labor & Delivery において、想定（予想）の 7,824 よりずっと高い（異常に高い）54,236 を記録したと書いてあります。このように、フィールドの色々な組み合わせから、異常と思われる値を探し出すのが異常値検出の目的です。

インサイトの左下にある”異常の探索”を押して、詳細画面に移動します。



右のグラフは異常値を表しています。計算時に”月単位”にしたため、異常値の計算は月あたり一回ずつになっています。Labor & Delivery は 12 月に急に増加しているのがこのグラフからも分かります。

また、左側ペインには、寄与率が掲載されています。この異常な増加の要因となったのが、Patient ID 40032 と、7427 であること等、影響が大きかったものが自動的に記載されています。

内容を確認したら、上部の”分析に戻る”で前の画面に戻ります。

これでハンズオン 2 は完了です。

(オプション) ハンズオン 3 : LEVEL-AWARE AGGREGATION

LEVEL-AWARE AGGREGATION とは

本章では、Level-Aware Aggregation(LAA)機能を使った高度な分析を体験します。LAA は関数の計算（集計等）を行う**タイミング**を指定する機能です。ハンズオン 2 とは独立した内容ですが、ハンズオン 2 で説明した計算フィールドの基礎を知っている事を前提にしています。

ハンズオン 2 内のテーブル計算関数のハンズオンで、関数が「ビジュアルが表示されたタイミング」に依存して動いていることが確認しましたが、このような関数が実行されるタイミングを調整する機能が LAA です。（LAA が利用できる関数一覧については、ドキュメントを確認してください。資料末尾にドキュメントへのリンクがあります）

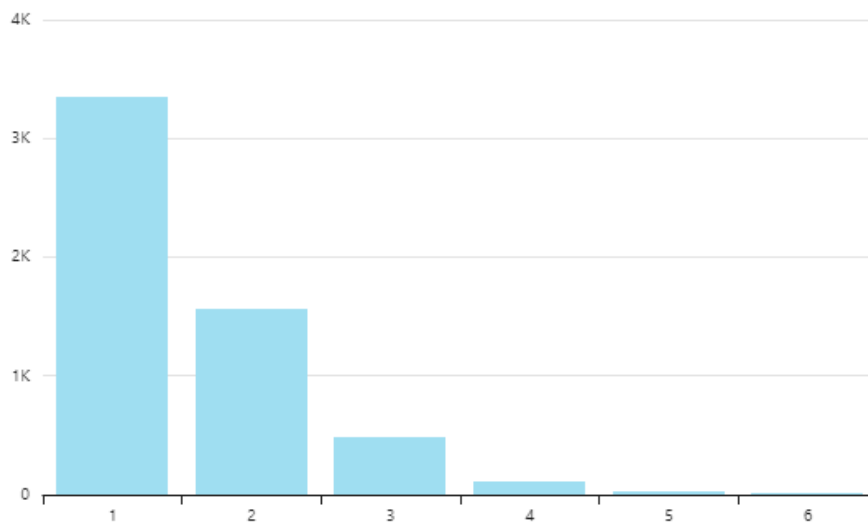
以下のタイミングを指定可能です。

- PRE_FILTER – データセットをフィルタする前に計算
- PRE_AGG – 集計をビジュアルに適用する前に計算
- POST_AGG_FILTER – ビジュアルが表示されるタイミングで計算（ここまで見てきた挙動であり、関数の**デフォルト動作**）

これがどのような場合に役に立つのかをハンズオンで確認します。LAA 用のハンズオンを開始する前に、シートを切り替えてください。シート名の横にある”+”を押すと、新しいシートが作成されます。

LEVEL-AWARE AGGREGATION (COUNTOVER)

以下の図のように、1 回入院した患者の総数は XX 名、2 回入院した患者の総数は XX 名というように、入院回数ごとの患者数をカウントしたいとします。この場合 X 軸にあたる「入院回数」はデータに無いため、入院患者数を集計する前に、X 軸を計算しておく必要があります。このようなケースでは PRE_AGG を指定して計算フィールドを作成します。



以下の内容で計算フィールドを作成してください

名前: **AdmittancePerPatient**

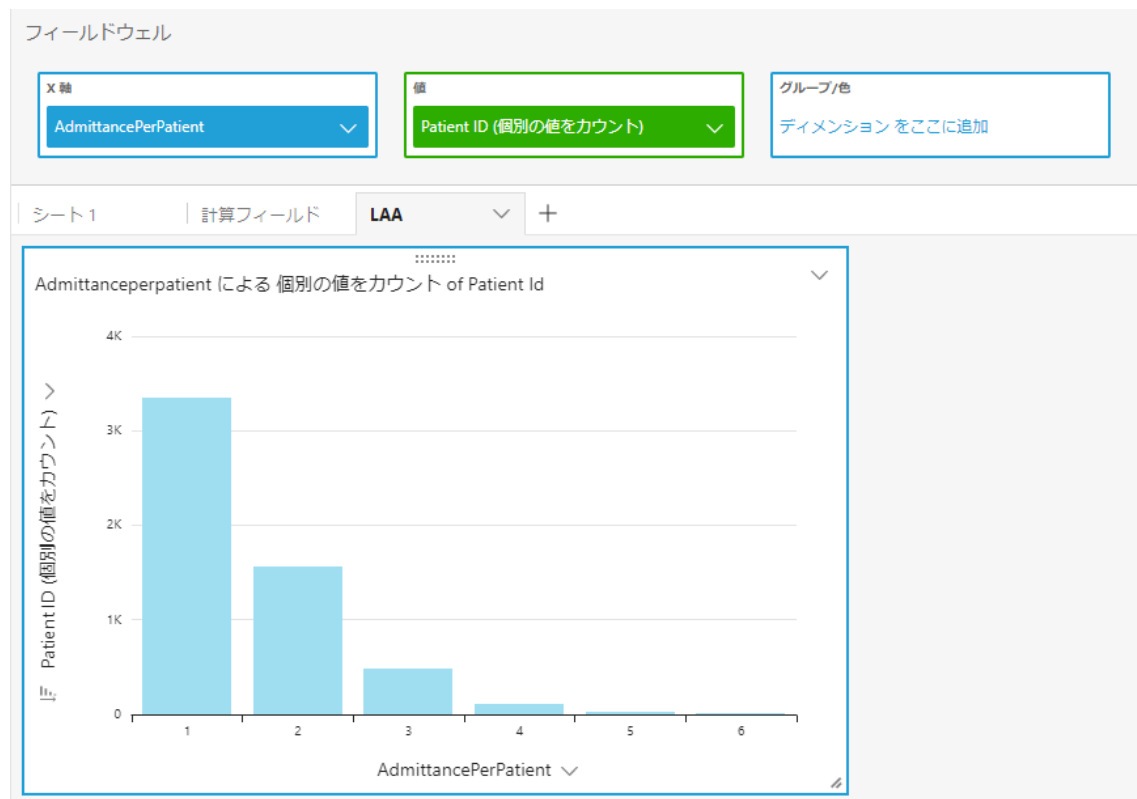
計算式: `countOver(({Admit Date})),[{Patient ID}],PRE_AGG)`

`countOver` の第 1 引数は、カウントの対象ですが、今回は入院日のフィールドにしています。第 2 引数がパーティショニングで、上記では Patient ID（患者 ID）を指定していますので、患者 ID ごとに入院回数がカウントされます。第 3 引数で、PRE_AGG を指定しています。

AdmittancePerPatient フィールドが作成できたら、垂直棒グラフ（最上段、一番右）ビジュアルを作成し、フィールドウェルの X 軸に AdmittancePerPatient を、値に、Patient ID を置きま

す。Patient ID は集計で”個別の値をカウント”を選択してください（SQL でいうところの distinct 演算）。

以下のようになれば完成です。



LEVEL-AWARE AGGREGATION (SUMOVER)

以下の図のように、病院(Hospital)ごとのレベニュー (Revenue) を集計したとします。

Hospital	Revenue
Downtown Hospital	1,480,027.1607
North Hospital	9,367,338.3834
South Hospital	7,785,998.76

上記の表をフィルタして Revenue を構成する患者(Patient)の中で、生涯 (通算) で\$10,000 以上支払っている Patient だけに絞りたい場合はどうすれば良いでしょうか。

上記の表を可視化する時点ではすでにデータは集約されてしまっていますので、データを集約する前(PRE_AGG)で、Patient ごとの通算 Revenue を計算する必要があります。

まずは上記の表を表示します。新しいビジュアルを作成し、ピボットテーブル (上から 3 目、左から 5 目) を選択します。行には”Hospital”を、値には”Revenue“を設定し、Revenue は通貨に指定します。

フィールドウェル

行 Hospital	列 ディメンションをここに追加	値 Revenue (合計)
---------------	--------------------	-------------------

Hospital による 合計 of Revenue

Hospital	Revenue
Downtown Hospital	\$1,480,027.16
North Hospital	\$9,367,338.38
South Hospital	\$7,785,998.76

次に、Patient 毎の総支払い表す計算フィールドを LAA で定義します。

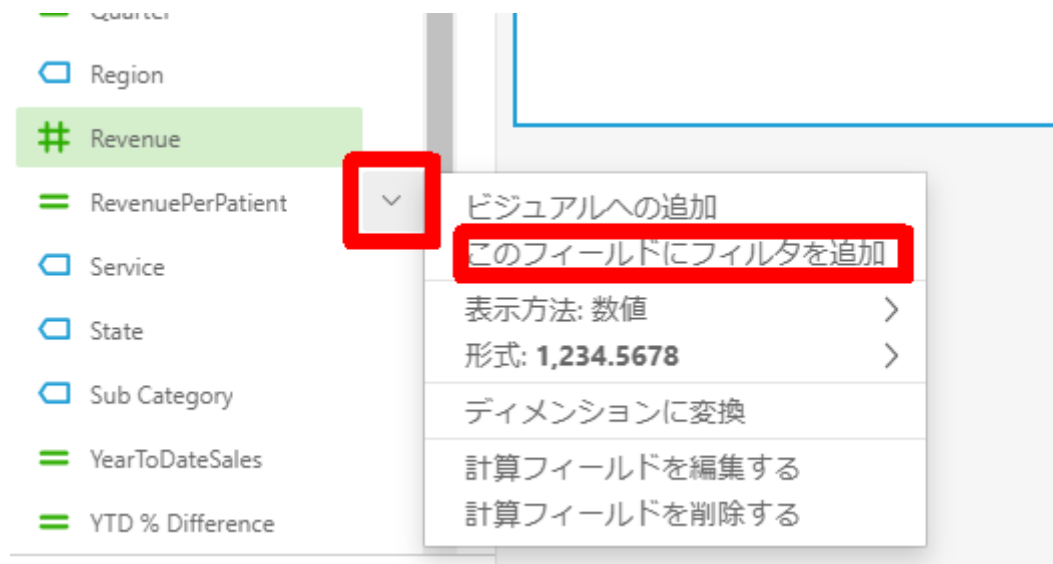
以下の内容で計算フィールドを作成してください。

名前： **RevenuePerPatient**

計算式： **sumOver(Revenue,[{Patient ID}],PRE_AGG)**

このようにすることで、Revenue が Patient ID 毎に合計されます。これをビジュアルのフィルタ条件に設定します。

先程作成したピボットテーブルが選択された状態で、フィールドリストの”RevenuePerPatient”の上にマウスカーソルを移動させ、横にある”V”をクリックし、“このフィールドにフィルタを追加”をクリックします。



作成されたフィルタで、以下のように 10,000 以上の値を残すように設定します。

フィルタの編集

このビジュアルのみ

RevenuePerPatient

次と等しい - なし

集計なし

次以上

☐ パラメータを使用

10000

Null を含む

これを適用すると、最終的な結果が得られます。

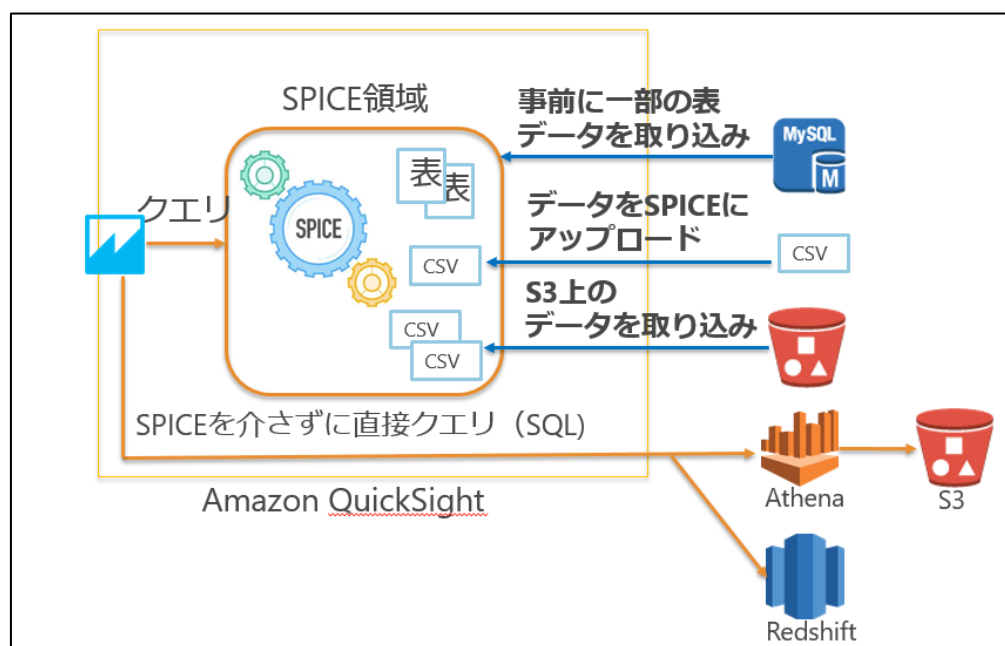
Hospital による 合計 of Revenue	
Hospital	Revenue
Downtown Hospital	\$439,359.30
North Hospital	\$3,600,224.28
South Hospital	\$4,838,198.60

(オプション) ハンズオン 4：大規模データを ATHENA 経由で可視化する

QUICKSIGHT + ATHENA + S3 の構成

QuickSight では手元のファイルを SPICE にアップロードして分析したり、S3 上のファイルを同様に SPICE に取り込んで分析したりすることが可能です。一方で、データソースが SQL を受け付けるものであれば、SPICE に取り込まずに直接 SQL でアクセスして可視化をすることも出来ます（下図の”SPICE を介さずに直接クエリ”の部分）。

直接データソースにアクセスすることで、SPICE の 1 データセットあたりの最大容量を超えるサイズを分析可能になりますし、事前に取り込む必要が無いためにデータソースをリアルタイムに反映した分析が可能になります。



本ハンズオンで発生する費用

本ハンズオンでは Athena を利用して S3 上のファイルに SQL で分析できるようにした上で、QuickSight を使って可視化を行ないます。

QuickSight はサインアップから 30 日間は 4 ユーザまで無料で試用できますが、**Athena の利用および Athena のクエリ結果が保存される S3 の利用には所定の費用が発生します。**

本ハンズオンの範囲であれば少額（1 米ドル程度、もしくはそれ未満）で利用できるよう設計されています。料金の詳細については補足情報にリンクが記載されている料金についてのページを確認してください。

また、ハンズオン終了後にはまとめと後片付けの章を確認し、不要なリソースをクリーンアップするようにしてください。

ATHENA と サンプル データ

Athena は S3 上に置かれたファイルに SQL でクエリを実行できるようにするサービスです。大規模分析にも耐えられる性能をサーバレス（ユーザがサーバを管理する必要がないサービス）で提供しています。Athena はクエリ単位で費用が発生し、1TB のデータ・スキャンあたり \$5 です（資料執筆時点）。詳細は以下 URL を確認してください。

<https://aws.amazon.com/jp/athena/pricing/>

Athena でのクエリ性能は、S3 上にどのようにデータを配置するかによって大きく変化します。本章ではデータの配置が異なる 2 種類のデータそれぞれに Athena からアクセスし、速度差を体験します。

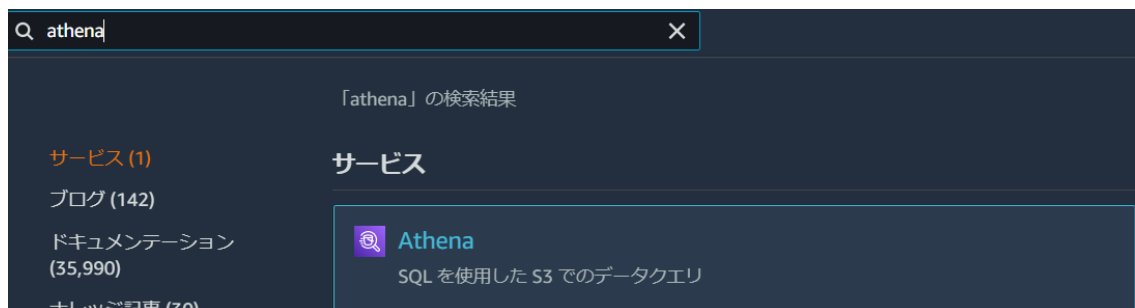
利用するデータは、米国交通統計局 (<https://www.bts.gov/>) が公開している米国国内線の発着データを使い、2010 年～2014 年の範囲で必要な部分を TSV（タブ文字で区切られたテキストファイル）で保存したものです。

ATHENA から サンプル データを クエリ する

<https://console.aws.amazon.com> にアクセスし、AWS マネジメントコンソールにログインします。AWS マネジメントコンソールでは、利用リージョンを「バージニア北部」に設定してください（本ハンズオンは必ずバージニア北部リージョンで実施する必要があります）。



サービス一覧から **Athena** を選択します。



Athena にアクセスし、左側のメニュー（ハンバーガーメニュー）からクエリエディタを選択してください。

すると左側にデータベース一覧、右側にクエリエディタ（ここに SQL を記入して実行する）が表示されます。もしくは、クエリエディタの画面になっていない場合は画面左のメニューからクエリエディタを選択してください。

（補足：もし管理コンソールの設定が日本語であるにも関わらず Athena のコンソールが英語で表示された場合は、画面左下のメニューで言語を一度”English(US)”に切り替えていただいた後に、再度日本語に切り替えていただくと日本語表示が得られます。）

この際、画面上方に“**最初のクエリを実行する前に、Amazon S3 でクエリ結果の場所を設定する必要があります。**”と表示されている場合は、Athena の結果出力用の S3 バケットを指定する必要があります。”詳細を表示”をクリックするか、もしくは画面上部の設定をクリックしてください。



設定メニューの中に「管理」ボタンがありますので、管理ボタンを押します。ここではクエリ結果を保存する場所になる S3 バケット名を指定します。

設定を管理

クエリの結果の場所と暗号化

クエリ結果の場所

表示

S3 を参照

☐ クエリ結果を暗号化

キャンセル

保存

例にあるように s3://ユニークなバケット名/としてください。このバケット名は、世界で唯一である必要がありますので、他の人と重複しない名前を指定してください。使用できる文字は半角のアルファベット、数字、ハイフンだけです。

記入できたら「保存」を押して元の画面に戻りますので、「エディタ」を選択してください。

最初に、本ハンズオンで利用するデータベースを作成します。Athena は S3 上のデータに対して CREATE EXTERNAL TABLE で表を定義することで S3 上のデータにアクセスできるようになりますが、その表をまとめておく単位がデータベースです。

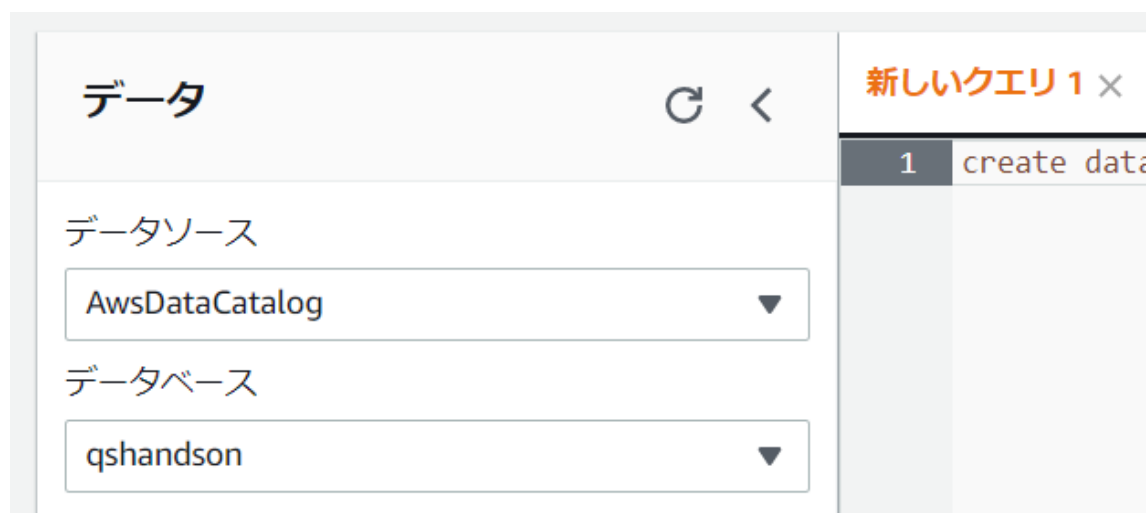
右側のペインで

```
create database qshandson;
```

と入力して”クエリの実行”をクリックします。（押せない場合は一度ブラウザをリロードしてみてください）もしくは、CTRL+Enter（Cmd+Enter）を押してもクエリを実行可能です。



成功すると左側のデータベース一覧から qshandson が選択できるようになるので、qshandson を選択します。



これでデータベース”qshandson”がデフォルトとして選択された状態になりました。この後の SQL 操作はこの qshandson データベースに対して行われます。

ここで S3 上のデータにアクセスできるように CREATE EXTERNAL TABLE を実行します。データは以下の S3 バケットに用意されています。

s3://ee-assets-prod-us-east-1/modules/aba5d66e8cd44378ba70a0427830efd5/v1/flightdata1/

このバケットには gzip 圧縮した TSV ファイルが保存されています。参考までに以下にそのスクリーンショットを示します。サイズは圧縮後で約 **621MB** ほどあるのがわかります（非圧縮の状態、約 815MB）。この 1 ファイルに 5 年間分の全データが含まれています。（補足：ハンズオン用の S3 ファイルは管理用の AWS アカウントに存在するため、ハンズオン実施者の AWS アカウントの S3 には存在しません。管理コンソールから確認するには以下の URL にアクセスしてください。

<https://console.aws.amazon.com/s3/buckets/ee-assets-prod-us-east-1?prefix=modules/aba5d66e8cd44378ba70a0427830efd5/v1/flightdata1/>

AWS CLI や AWS Cloud Shell が利用可能な環境にある場合は、`aws s3 ls s3://ee-assets-prod-us-east-1/modules/aba5d66e8cd44378ba70a0427830efd5/v1/flightdata1/` でも確認可能です。）



AWS CLI / AWS Cloud Shell での確認例)

```
[cloudshell-user@ip-10-0-57-40 ~]$ aws s3 ls s3://ee-assets-prod-us-east-1/modules/aba5d66e8cd44378ba70a0427830efd5/v1/flightdata1/
2022-01-06 09:59:30 651566887 flightdata1.tsv.gz
[cloudshell-user@ip-10-0-57-40 ~]$
```

この gz ファイルは、展開すると以下のようにタブで区切られた巨大なテキストファイルになっています。Athena ではスキャンしたデータサイズで費用が発生するので、費用削減のためにも、速度向上のためにも圧縮してファイルを格納する事が推奨されます。

1	2010-01-01	2745	El Paso, TX Texas	Phoenix, AZ Arizona	-6.000000	-5.000000	0.000000	0.000000	5
2	2010-01-03	1528	Charlotte, NC	North Carolina Philadelphia, PA	Pennsylvania	35.000000	41.000000	0	
3	2010-01-03	560	Dallas/Fort Worth, TX	Texas Newark, NJ	New Jersey	119.000000	100.000000	0.000000	0
4	2010-01-04	4234	Chicago, IL Illinois	Memphis, TN Tennessee	9.000000	-8.000000	0.000000	0.000	
5	2010-01-05	373	New York, NY	New York Charlotte, NC	North Carolina	28.000000	1.000000	0.000000	
6	2010-01-06	4004	Chicago, IL Illinois	New Orleans, LA Louisiana	60.000000	48.000000	0.000000	0	
7	2010-01-06	2340	Seattle, WA Washington	Chicago, IL Illinois	-1.000000	-3.000000	0.000000	0.000	
8	2010-01-07	5694	Greer, SC	South Carolina Chicago, IL	Illinois	38.000000	69.000000	0.000000	0
9	2010-01-07	2539	Richmond, VA	Virginia Houston, TX	Texas	-19.000000	-6.000000	0.000000	0.000
10	2010-01-07	495	Cleveland, OH	Ohio Las Vegas, NV	Nevada	35.000000	13.000000	0.000000	0.000000
11	2010-01-08	5311	Cincinnati, OH	Kentucky Omaha, NE	Nebraska	3.000000	35.000000	0.000000	0
12	2010-01-08	3776	Albuquerque, NM	New Mexico El Paso, TX	Texas	4.000000	-4.000000	0.000000	0.000
13	2010-01-09	2468	Hartford, CT	Connecticut Orlando, FL	Florida	-4.000000	-7.000000	0.000000	0.000
14	2010-01-10	171	Boston, MA	Massachusetts San Francisco, CA	California	-6.000000	-43.000000	0.000000	
15	2010-01-10	1322	Boston, MA	Massachusetts St. Louis, MO	Missouri	-5.000000	-19.000000	0.000000	
16	2010-01-11	252	Los Angeles, CA	California Miami, FL	Florida	229.000000	239.000000	0.000000	0.000000
17	2010-01-12	6456	Boston, MA	Massachusetts Cincinnati, OH	Kentucky	0.000000	-30.000000	0.000000	
18	2010-01-13	6769	San Francisco, CA	California Redding, CA	California	5.000000	2.000000	0.000000	
19	2010-01-13	7129	Washington, DC	Virginia Boston, MA	Massachusetts	-3.000000	-4.000000	0.000000	
20	2010-01-13	2499	Houston, TX	Texas Albuquerque, NM	New Mexico	-3.000000	-13.000000	0.000000	0.000
21	2010-01-13	3054	Detroit, MI	Michigan Newark, NJ	New Jersey	-7.000000	-38.000000	0.000000	0.000
22	2010-01-13	3241	Dallas/Fort Worth, TX	Texas Longview, TX	Texas	-13.000000	-12.000000	0.000000	
23	2010-01-15	471	Boston, MA	Massachusetts Fort Myers, FL	Florida	-3.000000	-13.000000	0.000000	0.000

Athena で S3 上のデータにアクセスをする場合、最初に S3 上のデータに表定義を与える必要があります。AWS Glue のクローラー機能を使うことで自動的に定義することが可能ですが、今回はハンズオンですので CREATE EXTERNAL TABLE を手動で実行します。

CREATE EXTERNAL TABLE では、TSV で記載されたデータの各列がどういう名前で、どのようなデータ型を持っているのかを列挙して定義します。

以下の CREATE EXTERNAL TABLE を右側のクエリペインで実行してください。前の CREATE DATABASE を書いたクエリエディタの内容を消して使っても良いですし、“+”を押すと新しいクエリエディタのタブが追加されます。（長いので以下よりコピー&ペーストして利用してください。）

```

CREATE EXTERNAL TABLE flightdata1(
    flight_date date,
    flight_number string,
    origin_city string,
    origin_state string,
    destination_city string,
    destination_state string,
    departure_delay decimal(22, 6),
    arrival_delay decimal(22, 6),
    cancelled decimal(22, 6),
    diverted decimal(22, 6),
    air_time decimal(22, 6),
    distance decimal(22, 6),
    distance_group decimal(22, 6),
    carrier_delay decimal(22, 6),
    weather_delay decimal(22, 6),
    security_delay decimal(22, 6),
    late_aircraft_delay decimal(22, 6),
    origin_airport_name string,
    cancellation_reason string,
    destination_airport_name string,
    carrier_name string)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '¥t'
LOCATION 's3://ee-assets-prod-us-east-
1/modules/aba5d66e8cd44378ba70a0427830efd5/v1/flightdata1';

```

※注：最後の LOCATION の行は表記上折り返しがありますが、実際は 1 行です。

実行した CREATE EXTERNAL TABLE の意味は以下の通りです。

```
CREATE EXTERNAL TABLE flightdata1 (<= flightdata1 という名前の表を作成
    flight_date date,           <= ファイルの左から順に、列名と型を指定
    flight_number string,       (以下同様)
    : (中略)
    carrier_name string)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '¥t' <=TAB(¥t) 区切りを指定
LOCATION 's3://... (略) .../flightdata1' <=S3 上のデータの位置を指定
```

これで S3 に置かれた tsv.gz ファイルが表として SQL でクエリできるようになりました。画面左側に flightdata1 という表が作成されているはずです。



右側のクエリエディタで“+”で新しいクエリエディタを開き、以下の SELECT 文を実行します。「1)」の部分は入力する必要はありません。

1) SELECT COUNT(*) FROM flightdata1;

約 22 秒で結果が出力されました。この SQL はデータの行数を数えるもので、20,574,971 件（約 2000 万件）のデータがある事がわかります。実行時間やスキャンしたデータサイズがクエリエディタの下に表示されますので、そこで確認してください。

The screenshot shows the Amazon Athena console interface. At the top, there's a tab bar with '新しいクエリ 1' (New Query 1) selected. Below it, the SQL query 'SELECT COUNT(*) FROM flightdata1;' is entered. The console shows the query has completed successfully. A red box highlights the execution details: '完了' (Completed), 'キュー内の時間: 0.288 秒' (Time in queue: 0.288 seconds), '実行時間: 21.686 秒' (Execution time: 21.686 seconds), and 'スキャンしたデータ: 621.38 MB' (Scanned data: 621.38 MB). Another red box highlights the result table header '_col0' and the single result value '20574971'.

クエリエディタを“+”で追加して、以下の2つの SELECT 文も実行してください。

2) SELECT COUNT(*) FROM flightdata1 WHERE YEAR(flight_date)=2010;

3) SELECT YEAR(flight_date), COUNT(*) FROM flightdata1 GROUP BY YEAR(flight_date);

これらは YEAR()関数で flight_date から年を抜き出し、それで WHERE や GROUP BY を実行するものです。それぞれ 45～50 秒程度時間がかかります。また、スキャンしたデータサイズはどのクエリでも約 **621MB** であることも分かります。

このように Athena を使うことで、データベースサーバを構築する事なく S3 上のファイルにクエリ (SELECT) を実行できました。しかしクエリにやや時間が掛かっています。これは大きいデータが単体のファイルとして保存されているためです。ファイルを分割し、適切にパーティションすることでこれを改善します。

ATHENA からパーティショニングされたサンプルデータをクエリする

Athena でファイル配置によってクエリ速度を改善する方法として重要なのはファイルを圧縮することです。ここまで試したサンプルデータでも gzip 圧縮が行われていました。もう一つの重要な方法がパーティショニングです。(※Athena がサポートする圧縮やパーティショニング、パフォーマンスについての資料は資料末尾の「補足資料」に記載)

パーティショニングとは、値に応じて別のディレクトリに保存することです。Athena では Apache Hive ディレクトリと同じ書式で分割・保存することが推奨されています。これはフォルダ名（正確には Prefix ですが、ここではフォルダとします）に、分割対象の列名と値を入れる方式です。例えば年（year）でファイルを分割する場合は以下のような構造になります。

/year=2010/2010 年のデータ.gz

/year=2011/2011 年のデータ.gz

:

2010 年のデータは/year=2010/以下に置き、2011 年のデータは/year=2011/以下に置きます。年（year）に加えて月(month)でパーティショニングする場合は、以下のようになります。

/year=2010/

/month=1/2010 年 1 月のデータ.gz

/month=2/2010 年 2 月のデータ.gz

:

このように、year=..のサブフォルダとして month=..を作成して分割していきます。

ここで利用する 2 つ目のサンプルデータは上記のように year と month でパーティショニングされた形でデータが保存されています。（データの内容は最初のサンプルと同じです）

データは s3://ee-assets-prod-us-east-

1/modules/aba5d66e8cd44378ba70a0427830efd5/v1/flightdata2/ に用意されています。GUI で確認する場合は、以下の URL にアクセスしてください。

<https://console.aws.amazon.com/s3/buckets/ee-assets-prod-us-east-1?prefix=modules/aba5d66e8cd44378ba70a0427830efd5/v1/flightdata2/>

AWS CLI や Cloud Shell が利用可能な環境にある場合は、`aws s3 ls s3://ee-assets-prod-us-east-1/modules/aba5d66e8cd44378ba70a0427830efd5/v1/flightdata2/` 等と実行することで以下の内容を確認できます。

```
[cloudshell-user@ip-10-0-57-40 ~]$  
[cloudshell-user@ip-10-0-57-40 ~]$ aws s3 ls s3://ee-assets-prod-us-east-1/modules/aba5d66e8cd44378ba70a0427830efd5/v1/flightdata2/  
PRE year=2010/  
PRE year=2011/  
PRE year=2012/  
PRE year=2013/  
PRE year=2014/  
[cloudshell-user@ip-10-0-57-40 ~]$ aws s3 ls s3://ee-assets-prod-us-east-1/modules/aba5d66e8cd44378ba70a0427830efd5/v1/flightdata2/year=2012/  
PRE month=1/  
PRE month=10/  
PRE month=11/  
PRE month=12/  
PRE month=2/  
PRE month=3/  
PRE month=4/  
PRE month=5/  
PRE month=6/  
PRE month=7/  
PRE month=8/  
PRE month=9/  
[cloudshell-user@ip-10-0-57-40 ~]$
```

ではパーティショニングされたデータに対して `CREATE EXTERNAL TABLE` を実行して表を定義します。以下の SQL を Athena から実行して、`flightdata2` 表を定義してください。

```

CREATE EXTERNAL TABLE flightdata2(
    flight_date date,
    flight_number string,
    origin_city string,
    origin_state string,
    destination_city string,
    destination_state string,
    departure_delay decimal(22,6),
    arrival_delay decimal(22,6),
    cancelled decimal(22,6),
    diverted decimal(22,6),
    air_time decimal(22,6),
    distance decimal(22,6),
    distance_group decimal(22,6),
    carrier_delay decimal(22,6),
    weather_delay decimal(22,6),
    security_delay decimal(22,6),
    late_aircraft_delay decimal(22,6),
    origin_airport_name string,
    cancellation_reason string,
    destination_airport_name string,
    carrier_name string)
PARTITIONED BY (year int, month int)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '¥t'
LOCATION 's3://ee-assets-prod-us-east-
1/modules/aba5d66e8cd44378ba70a0427830efd5/v1/flightdata2' ;

```

最初のサンプルの CREATE EXTERNAL TABLE とほぼ同じですが、以下の部分が変更されています。

```
CREATE EXTERNAL TABLE flightdata2( <= 表名は flightdata2
    flight_date date,
    : (中略)
    carrier_name string)
PARTITIONED BY (year int, month int) <=パーティションされていると指定
ROW FORMAT DELIMITED FIELDS TERMINATED BY '¥t'
LOCATION 's3://... (中略) .../flightdata2' <=URL が変更
```

パーティションの指定が追加されたのが大きな変更点です。S3 上では/year=.../month=.../とパーティショニングしているため、それを指定しています。

パーティションを含む表定義を作成した場合、パーティションどのようなプレフィックス（ディレクトリ）が S3 上に存在するかを Athena に確認させる必要があります。以下のコマンドをクエリエディタで実行してください。

```
msck repair table flightdata2;
```

少し待つと完了し、パーティションの情報が確認され、flightdata2 表にクエリが可能になります。flightdata1 で実行したクエリを flightdata2 にも実行してみます。

1) SELECT COUNT(*) FROM flightdata2;

結果は同じですが、速度が大きく改善したのが確認できると思います。これはファイルが分割され、並行処理が可能になったためです。同様に残りの SQL も実行して時間を確認しますが、この際に YEAR(flight_date)としていた部分は以下のように year に置き換えます。

2) SELECT COUNT(*) FROM flightdata2 WHERE year=2010;

3) SELECT year, count(*) from flightdata2 GROUP BY year;

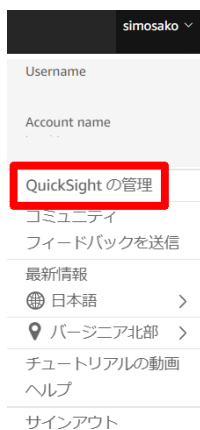
パーティションとして指定した year や month は列として使用できるため、上記のような書き方が可能になります。こうすることで、2)のようなケースでは、スキャンするデータは year=2010/以下に限定できるため、スキャン対象のデータサイズを減らすことが可能です。スキャン量の削減により、クエリにかかる費用も削減されます。（補足：1や3でスキャン量が増えているのは、細かい単位でファイルを分割しすぎているために圧縮率が下がり、総サイズが増加した影響です。一般的には分割した方が速度、費用の両面でメリットがあります。）

これで Athena 側での準備は完了です。この次の章では QuickSight からこの flightdata2 のデータを可視化します。

QUICKSIGHT から ATHENA に接続する

管理コンソールのトップ画面に戻るか（"AWS"ロゴをクリック）、上部の検索窓から QuickSight を選択してください。QuickSight の画面では右上のユーザ名(ID)をクリックし、"バージニア北部"リージョンが選択されている事を確認してください。

最初に flightdata2 のデータが置かれている S3 バケットへのアクセス設定を QuickSight に追加します。画面右上端のユーザ名をクリック→"QuickSight の管理"を選択します。




左側メニューから"セキュリティとアクセス権限"を選択し、右側で"Manage"を選択します。

QuickSight から接続可能なサービスを設定する画面が表示されますので、**Amazon Athena** にチェックを入れます。

Allow access and autodiscovery for these resources

- ☒  Amazon Redshift
- ☒  Amazon RDS
- ☒  IAM
- ☐  Amazon S3
S3 バケットを選択する

☒  Amazon Athena

Make sure you've chosen the right Amazon S3 buckets for QuickSight access

以下のようなダイアログが表示されます。これは QuickSight から Athena に接続する場合、Athena の結果セットが保存されるバケットにアクセスを許可する必要があるためです。”次へ”を押します。

Amazon Athena permissions

×

QuickSight needs permissions to access Amazon S3 buckets or AWS lambda functions used by Athena

キャンセル

Skip

次へ

すると以下のような S3 バケット選択のダイアログが表示されます。

QuickSight アカウントにリンクされている S3 バケット	AWS 全体でアクセスできる S3 バケット
QuickSight がアクセスできるようにするバケットを選択します。	
選択したバケットにはデフォルトで読み取り専用アクセス許可が付与されますが、Athena Workgroup 機能には書き込みアクセス許可を付与する必要があります。	
<input checked="" type="checkbox"/> すべて選択	
S3 バケット	Athena Workgroup の書き込みアクセス許可
<input checked="" type="checkbox"/> aws-athena-quicksight-20220106	<input type="checkbox"/>

画面には自アカウント内の S3 バケット一覧が出ますので、上記のように自分の **Athena** の設定で、結果出力用に指定したバケットにチェックをいれます（もしくは”すべて選択”にチェックを入れます）。次に右上の”AWS 全体でアクセスできる S3 バケット“をクリックします。

この画面は自分の AWS アカウント以外の S3 バケットにアクセスする必要がある際の設定なので **必ず実施してください**。”別のバケットを使用”に ee-assets-prod-us-east-1 と記入し（こちらで準備したデータが置かれたバケット名です）、”S3 バケットの追加”で追加します。

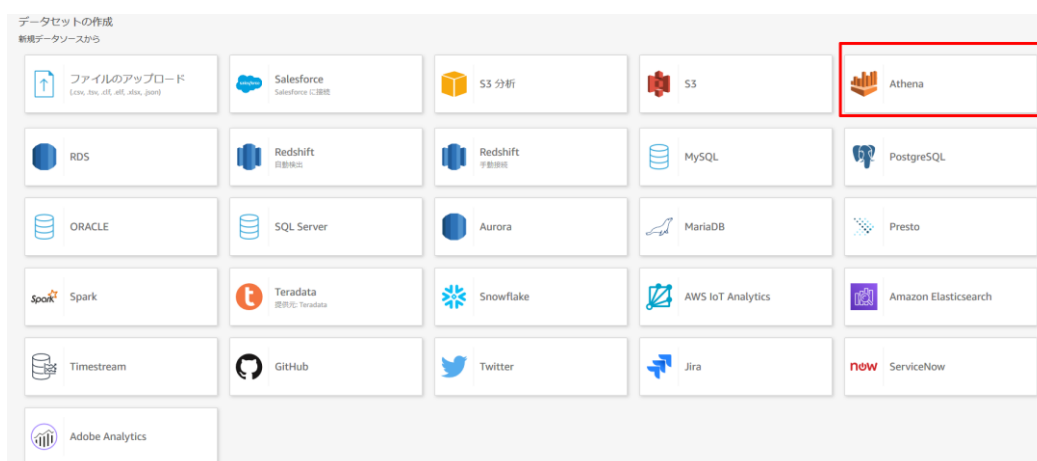
☒ 別のバケットを使用

S3 バケットの追加

追加できたら、”完了”を押して確定します。

サービスへのアクセス一覧の画面に戻るので”保存”を押します。設定が完了したら、左上の QuickSight アイコンをクリックしてメイン画面に戻ります。

次に Athena 経由で flightdata2 表へアクセスするための”データセット”を作成します。左側の”データセット”をクリックし、次に右上の”新しいデータセット”をクリックします。データセットに登録可能なデータソース一覧が表示されますので、**Athena** をクリックします。



ダイアログでデータソース名を聞かれるので任意の名前を入力します。ここでは”qshandson”と入力して”データソースを作成”をクリックします。



少し待つテーブル選択のダイアログが表示されるので、まずデータベースに **qshandson** を選択します。すると qshandson 内の表一覧が表示されるので、その中にある flightdata2 を選択して”データの編集/プレビュー”をクリックします。（注：必ず flightdata2 の方を選択してください。flightdata1 の方を選択すると、Athena でのクエリに時間を要するためレスポンスが悪化します。）



※補足：Athena 接続時にエラー（“Something went wrong”）が出る場合は、以下のドキュメントを参照して、設定を確認してください。

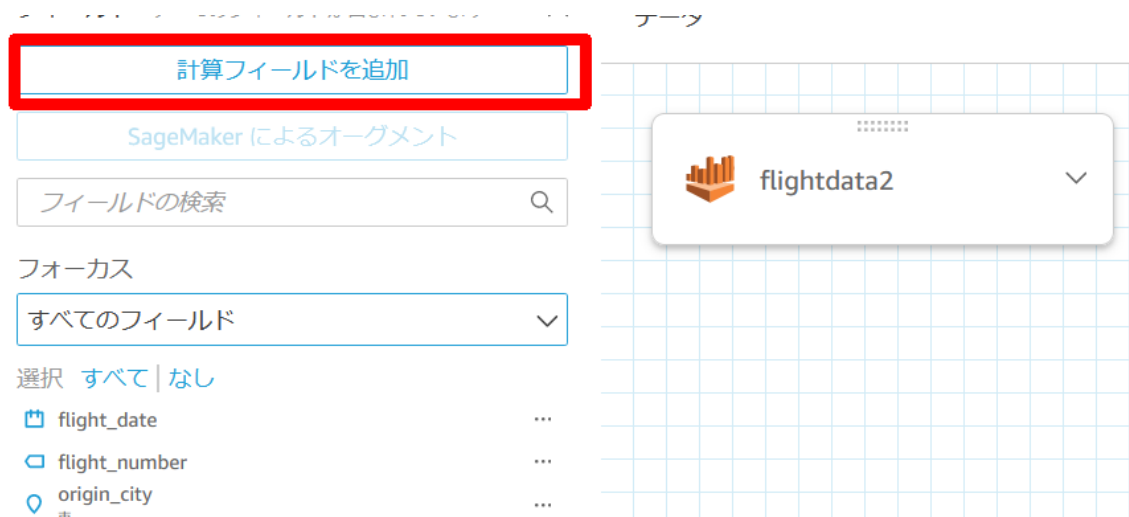
https://docs.aws.amazon.com/ja_jp/quicksight/latest/user/troubleshoot-athena.html

計算フィールド (CALCULATED FIELD) の追加

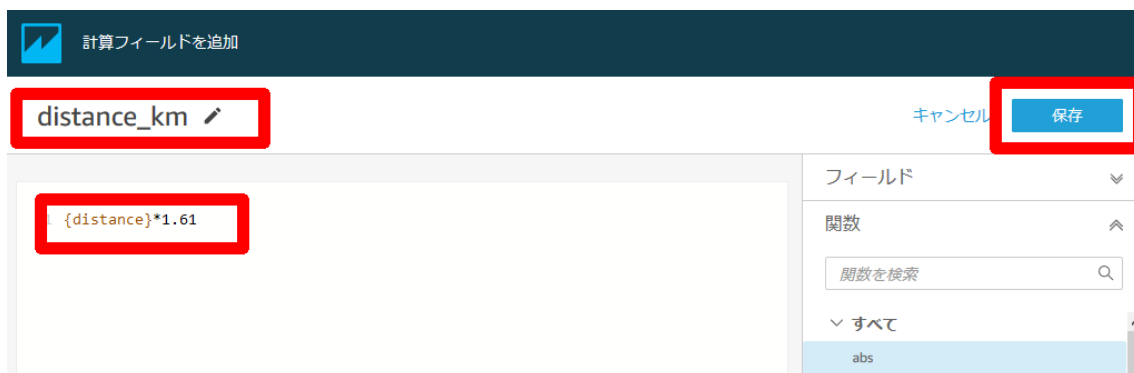
データの編集/プレビューを選択すると、以下のようにデータセットを調整するための画面が表示されます。この画面は必要ないフィールド（列）を削除したり、フィールド名を変更したりする事が可能ですが、ここでは触れません（事前準備の章を参照）。

ここでは、データセット内に計算フィールド（Calculated Field）を追加します。計算フィールドは、元のデータには無い、擬似的な列を QuickSight 側で追加する機能です。

ここでは、飛行距離を Km(キロメートル)で表示する distance_km フィールドを追加します。distance フィールドがすでにあり、ここにマイルで数値が入っているのでこれを Km に換算します。画面左側の“計算フィールドを追加”をクリックします。



クリックすると計算フィールドを追加する画面が表示されるので、名前に **distance_km** と入力し、その下の計算式に、**{distance}*1.61** と入力します。マイルを 1.61 倍することで Km に変換するという式です。入力後、ウィンドウ右の”保存”ボタンを押します。



左側の計算フィールドの欄に distance_km が追加されたのを確認してください。

※オプションとして、ここで画面左下から直接クエリか SPICE かを選択が可能です。本稿では直接クエリのままで進めますが、バージニア北部での SPICE 容量に余裕がある場合はここで SPICE を選択することで Athena 上のデータを SPICE に取り込むこともできます

クエリモード

☐ SPICE ☒ 直接クエリ

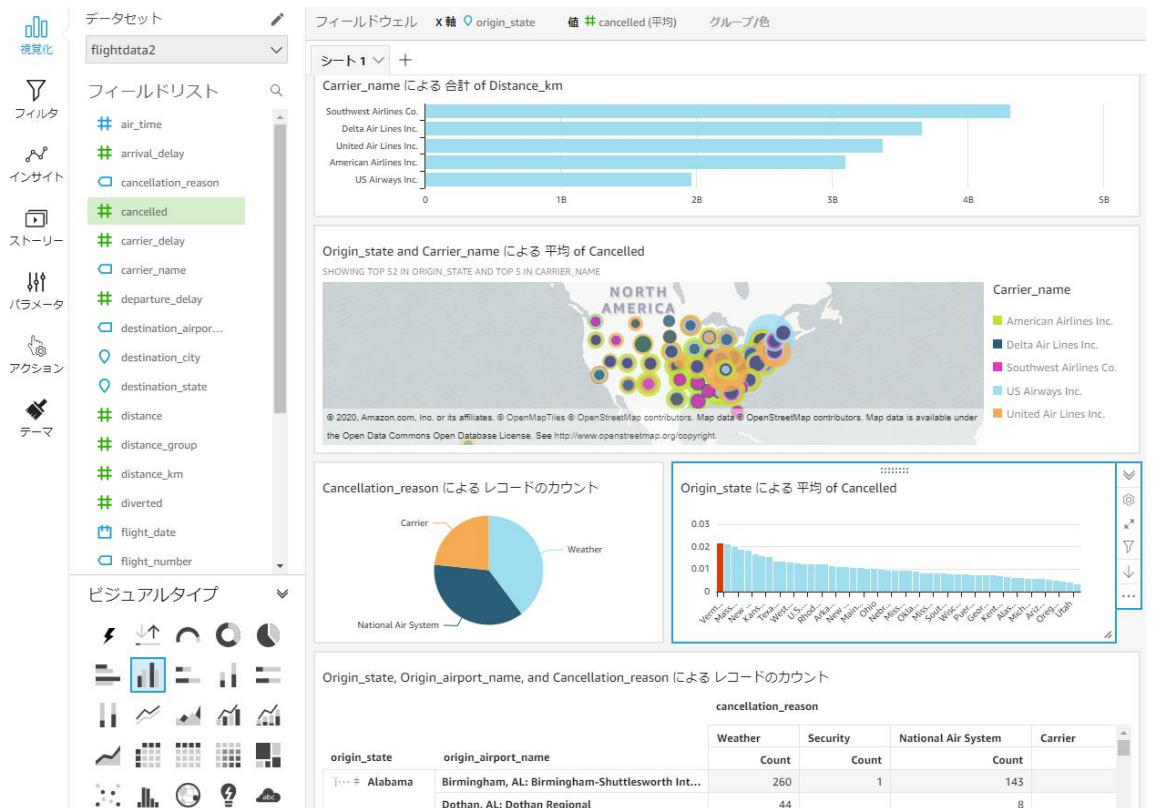
上部の”公開して視覚化”を押します。

[補足] 準備（Prepare）での計算フィールド作成

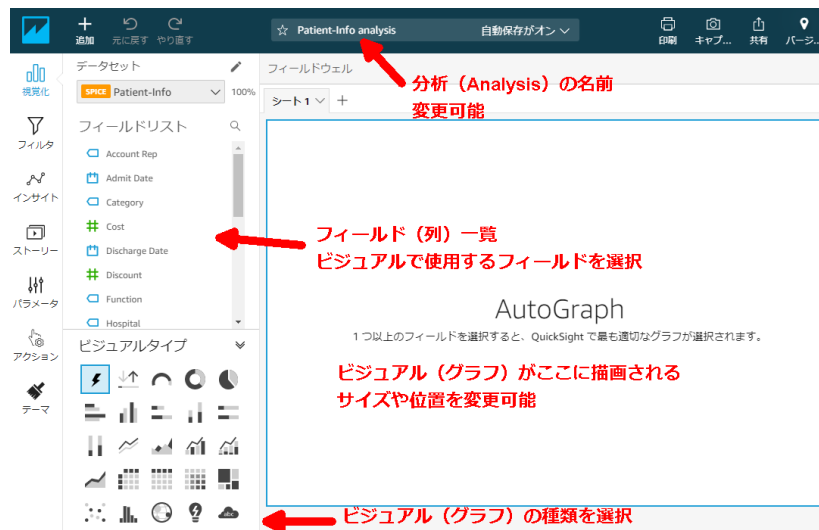
ここでは準備（Prepare）で計算フィールドを追加しましたが、ハンズオン 2 では Analysis 作成画面に入ってから計算フィールドを追加する方法を紹介しています。準備で計算フィールドを作成することで、画面描画前に可能な部分は計算しておくことが可能になり、実行が高速になるメリットがあります。一方で、テーブル関数等の高度な関数やパラメータ（変数）を取る関数定義は Analysis 画面でしか作成できません。

QUICKSIGHT を利用した可視化

データソースが準備できたので、QuickSight で可視化を行ないます。5つのビジュアル（図）を組み合わせて、以下のような画面を作成していきます。これは、あくまで可視化の一例ですので、ハンズオン内容を参考に自由に作成してみてください。



QuickSight で新しい分析（Analyze）を作成直後は以下のような画面になります。

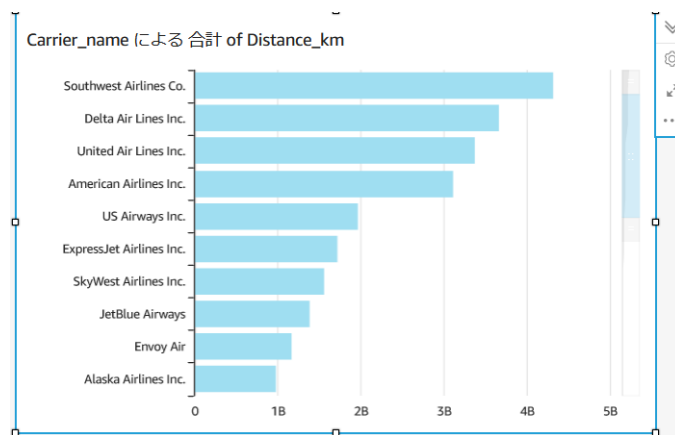


まず、画面最上部にある分析の名前を任意の名前に変更してください。例えば“航続距離 Top5 エアラインのキャンセル状況分析”等です。

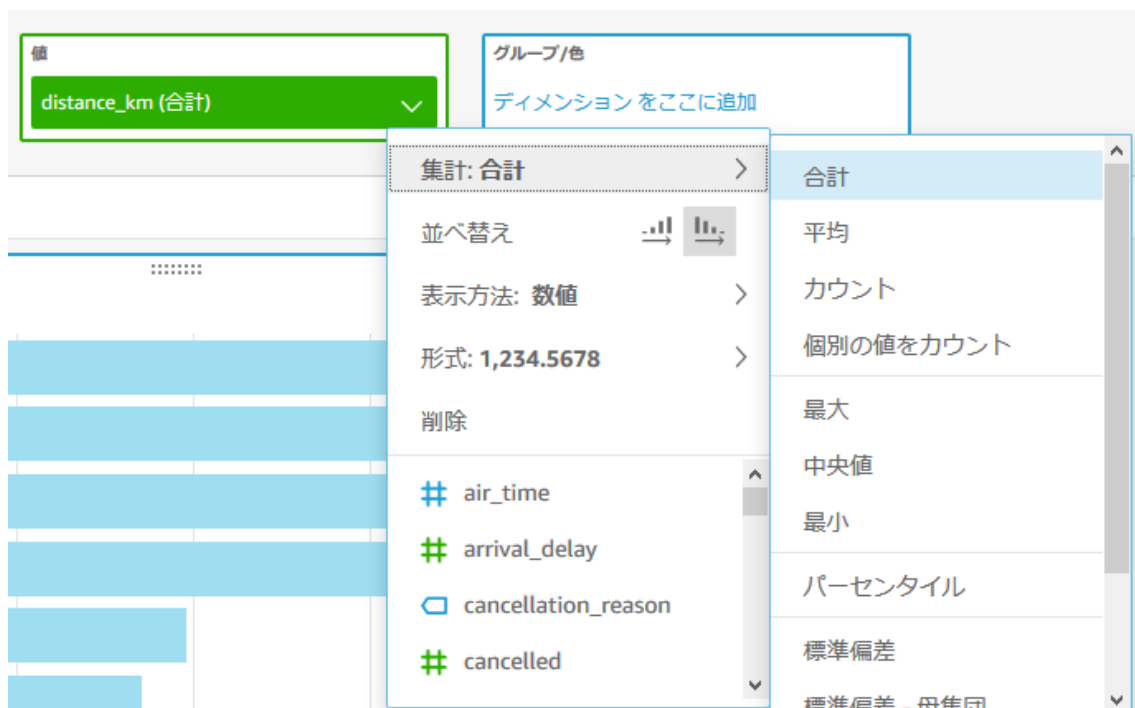
ビジュアル 1:航続距離が長いキャリア TOP 5（水平棒グラフ）

航続距離が長いキャリアトップ5を水平棒グラフ（横棒グラフ）で可視化します。

画面左下のビジュアルタイプで**水平棒グラフ**（最上段、左から6つ目）を選択し、フィールドウェルにある Y 軸には carrier_name（キャリア名）を、値には distance_km（航続距離の Km 表記）を、左側のフィールドからドラッグ・アンド・ドロップ（D&D）で配置します。少し待つと以下のような画面になります。



右側のスクロールバーの長さを変えたり移動させたりする事で表示範囲を変更させる事が可能です。



フィールドウェアの”値”はデフォルト（初期状態）では合計（SUM）として集計されています。値に置いた distance_km の”▼”をクリックすると、上記画面のように合計以外の集計（平均値、パーセンタイル等）に変更したり、並び順（小さい順か、大きい順か）を変更したりできます。（今回はそのままで良いので変更しません）

フィルタの編集

☒ 該当するすべてのビジュアル

☒ carrier_name
上 - なし

フィルタタイプ

☒ 上位と下位のフィルタ

☒ 上 ☐ 下

☐ パラメータを使用

上部を表示

5

による

☒ distance_km (合計)

+ タイブレイク

または

[フィルタ条件を追加](#)

注意: フィルタの組み合わせには制限があります。
[詳細はこちら](#)

グラフを見ると、1 位から順に、Southwest Airlines, Delta Air Lines, United Air Lines, American Airlines, US Airways が上位 5 社ですので、以下ではこの 5 社に絞って分析を進めることにします。QuickSight ではフィルタ機能を利用すると可視化されるデータの範囲を絞り込むことが可能です。

左側の「フィルタ」をクリックし、「1 つ作成...」を選択します。表示される一覧から、carrier_name を選択します。

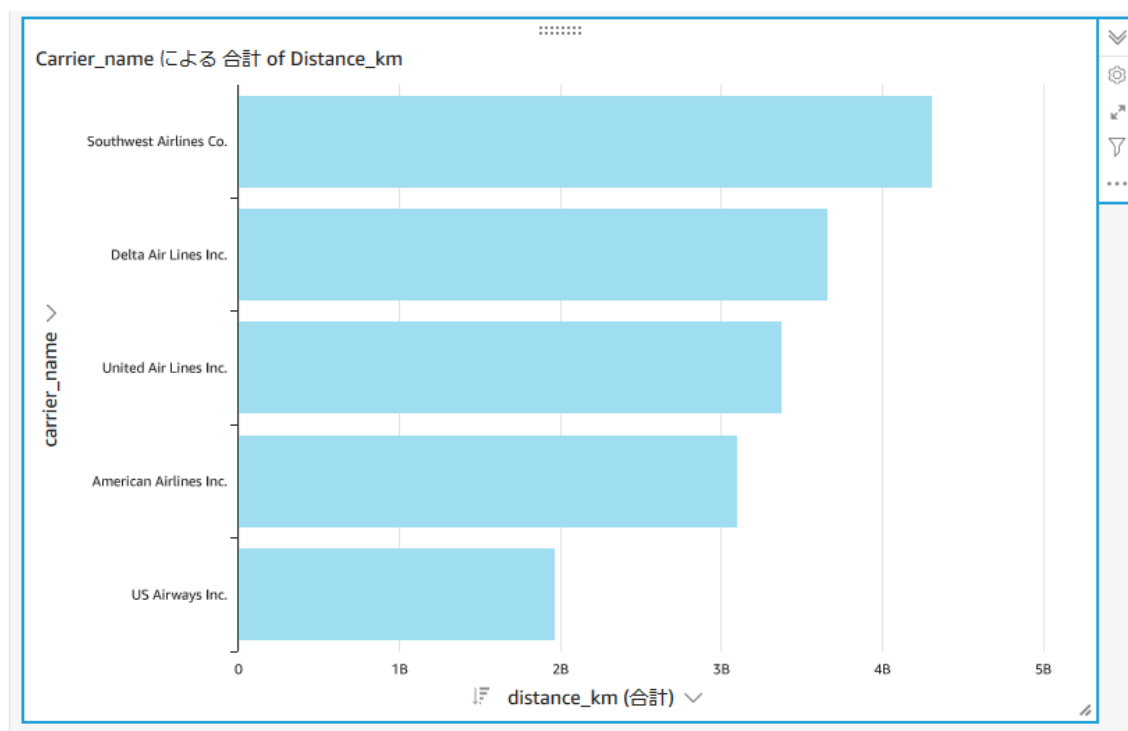
carrier_name というフィルタが作成されるので、クリックをして開きます。

左図に示すように上部にある「このビジュアルのみ」をクリックし、「該当するすべてのビジュアル」に変更します。フィルタでの絞り込みは 1 つのビジュアルに対して適用することもできますし、今回のように、分析に含まれるすべてのビジュアルに共通して適用することも可能です。

次に、トップ 5 社に絞り込む条件を指定します。会社名を 1 つずつリストから選択することも可能ですが、それでは今後データが変わった時にトップ 5 に一致しなくなる可能性があります。ここではデータの上位 N 個もしくは下位 N 個に絞り込む

ために「フィルタタイプ」で「上位と下位のフィルタ」を選択します。その下のチェックボックスで「上」を選択し、上部を表示には「5」を入れ、その下で「distance_km(合計)」を指定します。これにより distance_km 合計の上位 5 社のみ画面に表示されるようになります。

“適用”を押すと、ビジュアルが5社に絞り込まれるのが確認できます。この後で複数のビジュアルを追加しますが、carrier_name フィールドについては、ここで設定したフィルタが共通して適用されます。



これで1つ目のビジュアルが完成しました。

ビジュアル 2: キャンセル率が高い州（地図）

トップ5社が利用する空港で、どの州からの出発便にキャンセルが多いかを地図で可視化します。左上の”+追加”->“ビジュアルを追加”で新しいビジュアルを追加し、ビジュアルタイプは”地図上のポイント”（上から4段目、左から6つ目）を選択します。

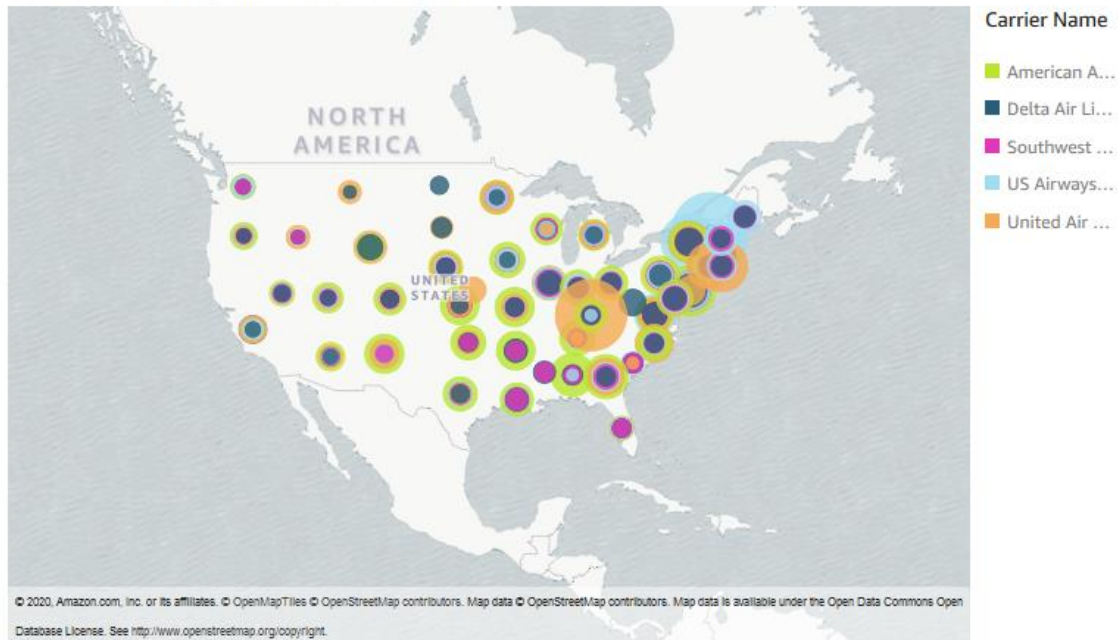
フィールドウェアの Geospatial（地理情報）には、origin_state(出発州)、Size には cancelled、Color には carrier_name を D&D してください。フィールドウェアの cancelled の右にある”v”をクリックし、集計を平均に変更します。これで以下のようなビジュアルが作成されます。Carrier_name としては5つのキャリアしか表示されませんが（右端のレジェンド参照）、これはビジュアル 1 で作成したフィルタがこのビジュアルにも効いているためです。

フィールドウェア

Geospatial origin_state	Size cancelled (平均)	Color carrier_name
-----------------------------------	-------------------------------	------------------------------

Origin_state and Carrier_name による 平均 of Cancelled

SHOWING TOP 102 IN ORIGIN_STATE AND TOP 5 IN CARRIER_NAME

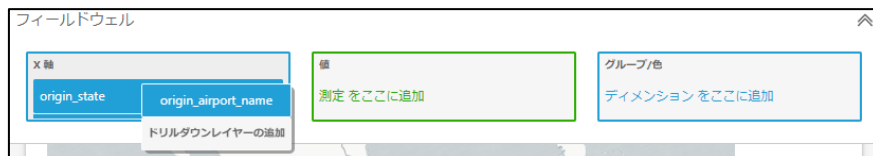


ビジュアル 3: キャンセル率が高い州（垂直棒グラフ + ドリルダウン）

ビジュアル 2 と同様にキャンセルが多い州を確認しますが、ここでは棒グラフで可視化します。また州名(origin_state)とエアポート名(origin_airport_name)の間にドリルダウン（上位/下位の関係）を設定します。

左上の“+追加”->“**ビジュアルを追加**”で新しいビジュアルを追加し、ビジュアルタイプは“垂直棒グラフ”（最上段、一番右）を選択します。

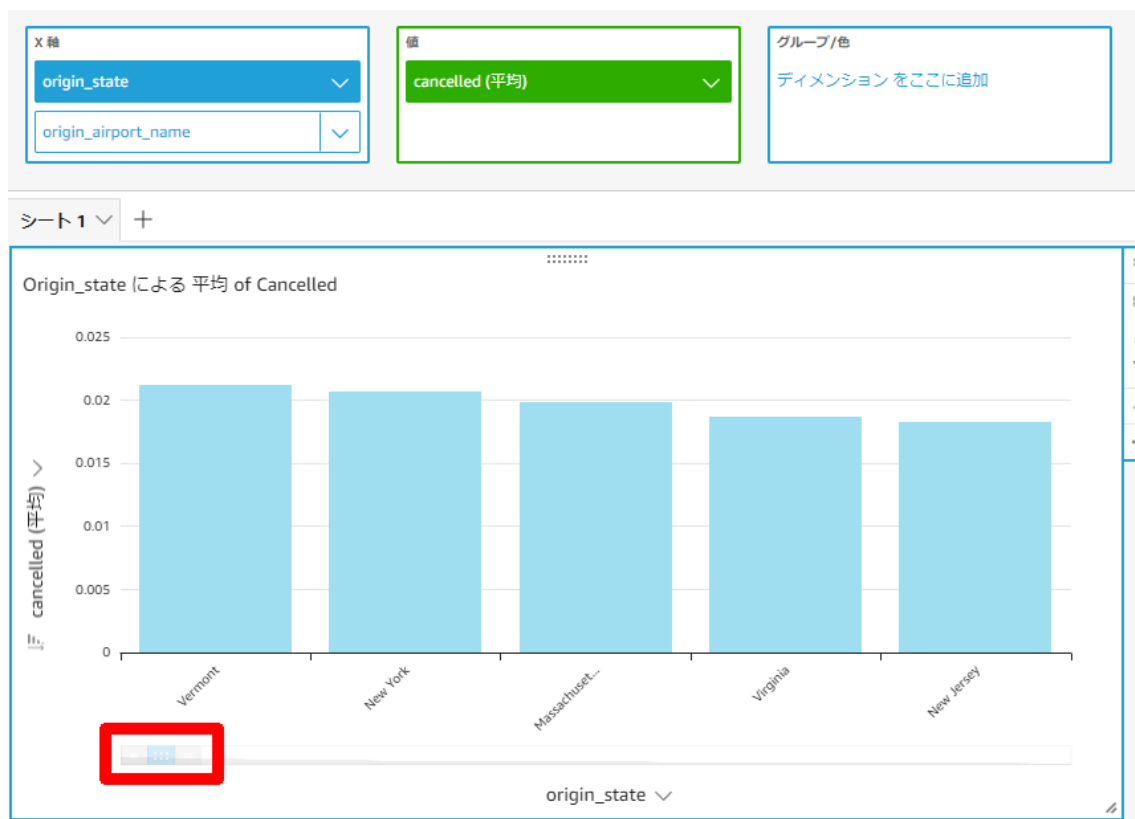
フィールドウェルの X 軸には origin_state を D&D します。追加でさらに origin_state の下に”ドリルダウンレイヤーの追加”と表示される位置で origin_airport_name を D&D します。（下図）これでドリルダウンが設定されます。



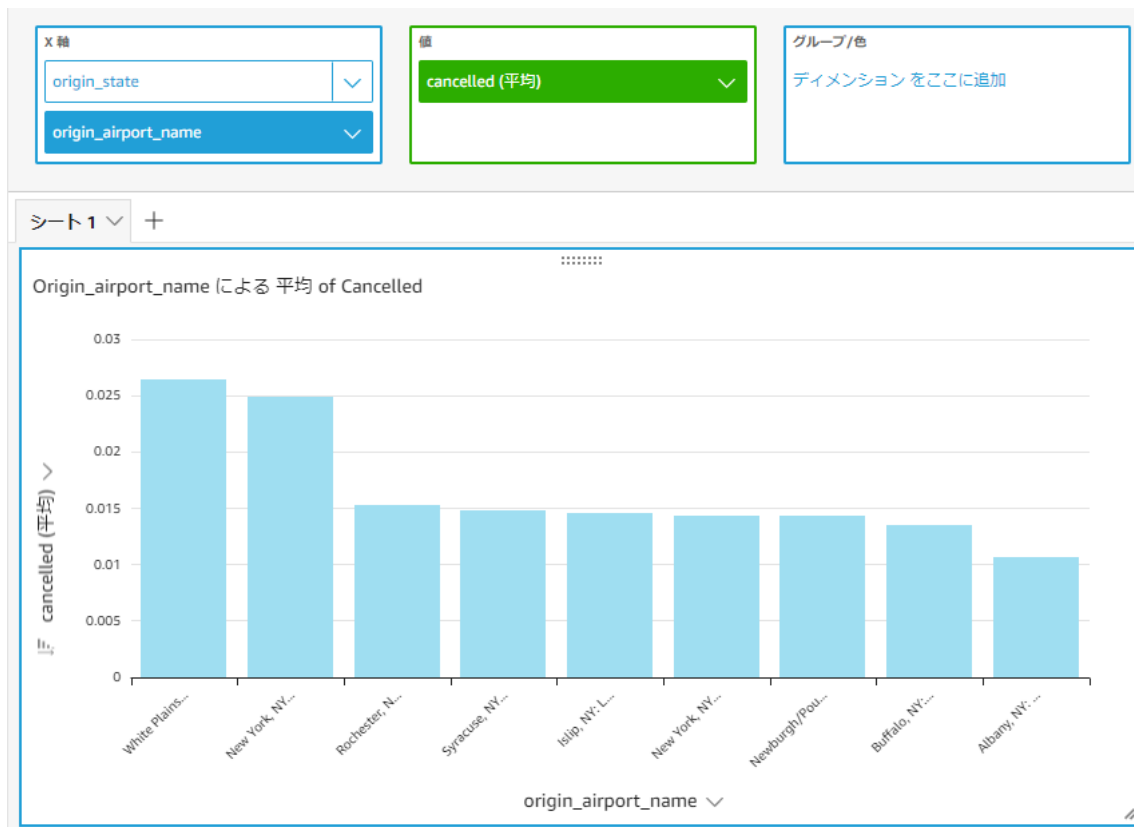
値には Cancelled を D&D し、“v”をクリックして集計を平均に変更します。棒グラフが小さすぎて見づらい場合はスライダー（以下の赤色枠部分）で表示サイズを狭める事が可能です。



下図のように、表示範囲を小さくしてみてください。

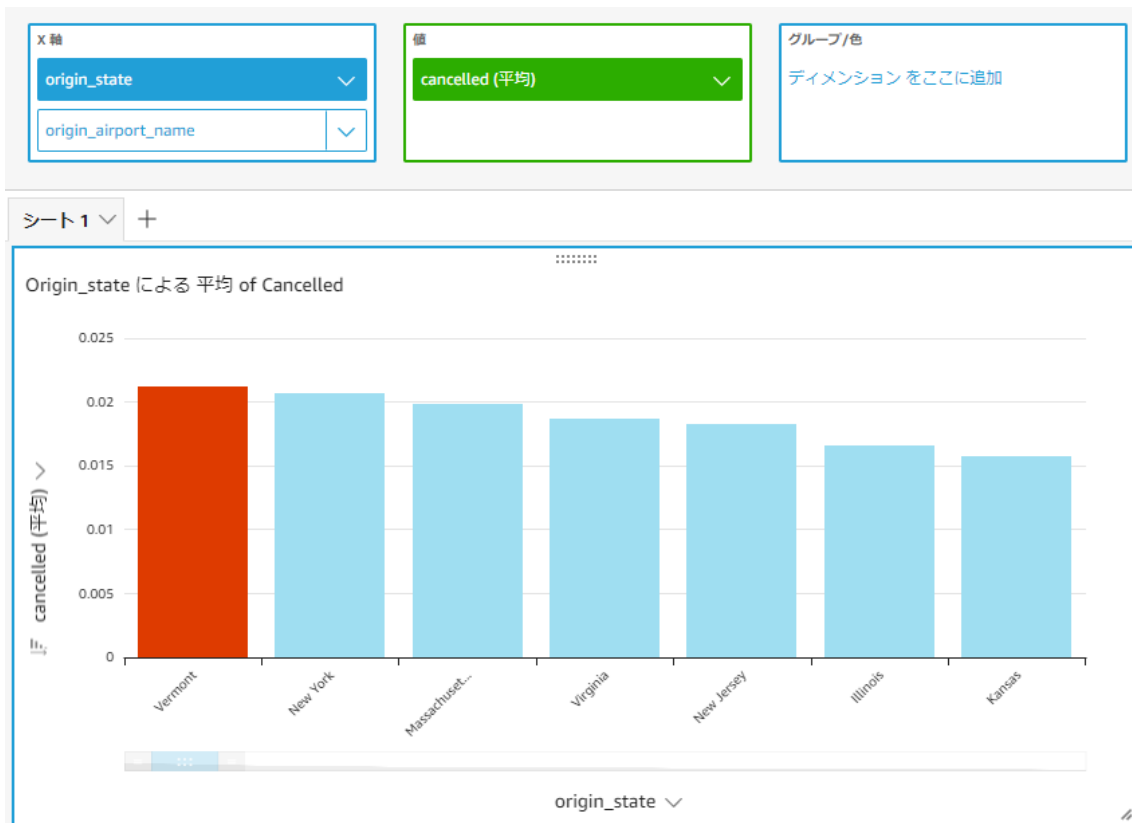


X 軸で、origin_state の下位に origin_airport_name を設定したので、ドリルダウンの動きを試してみます。New York のバーをクリック（左クリック）すると、“origin_airport_name までドリルダウン”と表示されるのでそこをクリックします。



上図のように、New York という州レベルから、New York に存在するエアポート名レベルにドリルダウンされたグラフが表示されました。ビジュアルの右の「↑」をクリックするか、再度グラフをクリックして、origin_state までドリルアップをクリックすると、州レベルの画面に戻ります。

州レベルに戻ったら、一番左の Vermont 州が最もキャンセル率が高いので、目立つようにマークしておきます。Vermont のグラフをクリックして、“色 Vermont”から赤色等、目立つ色に変更してください。画面下部の表示幅を変更するスライダー（水色部分）により表示領域を見やすく調整してください。

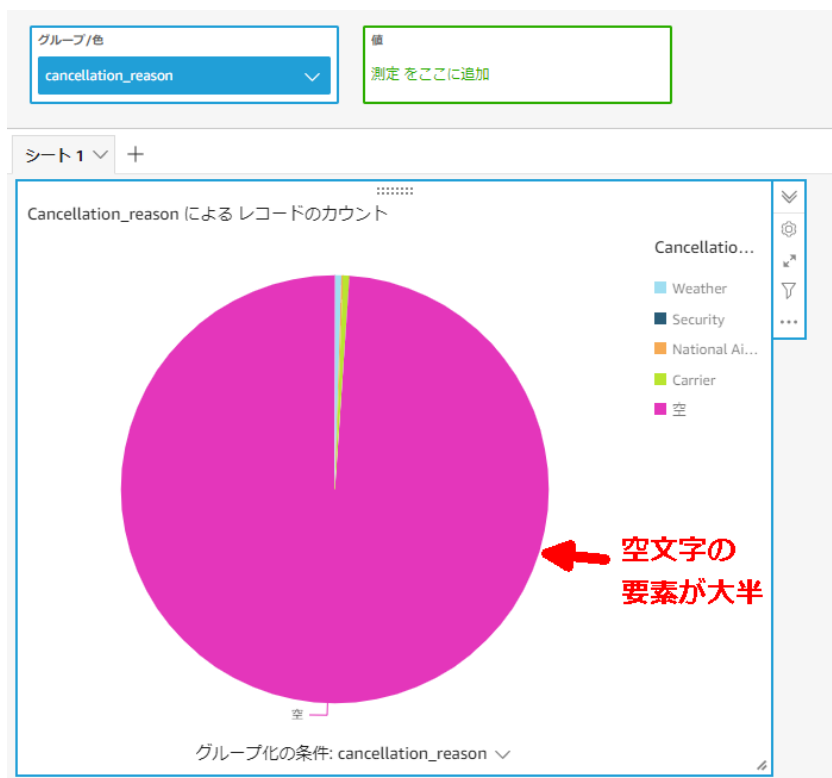


ビジュアル 4:VERMONT のキャンセル理由（円グラフ）

キャンセル率が高い Vermont 州のキャンセル理由を可視化する円グラフを作成します。

左上の”+追加”->“ビジュアルを追加”で新しいビジュアルを追加し、ビジュアルタイプは円グラフ（最上段、左から5つ目）を選択します。フィールドウェルのグループ/色に cancellation_reason（キャンセル理由）を D&D します。

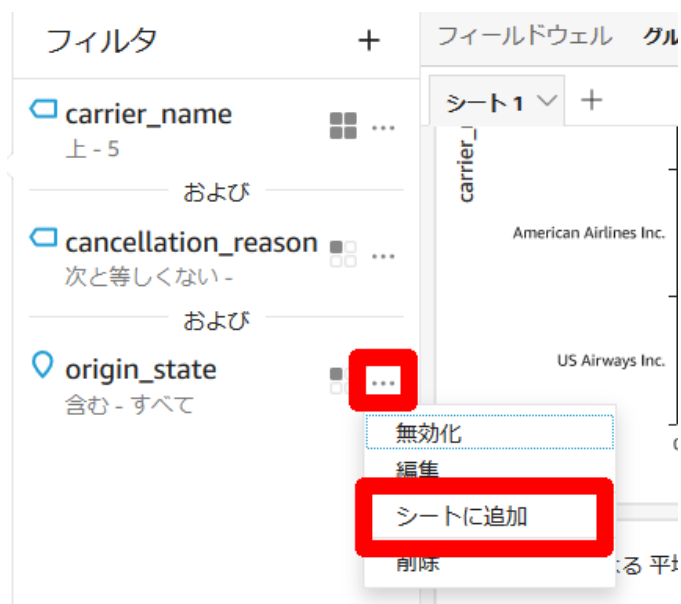
作成直後のグラフでは以下のように空文字（"空"）が大半を占めています。これは cancellation_reason の列は、キャンセルされない時は空文字が入るためです。今回はキャンセルされた時の理由なので、この空文字をグラフから削除します。



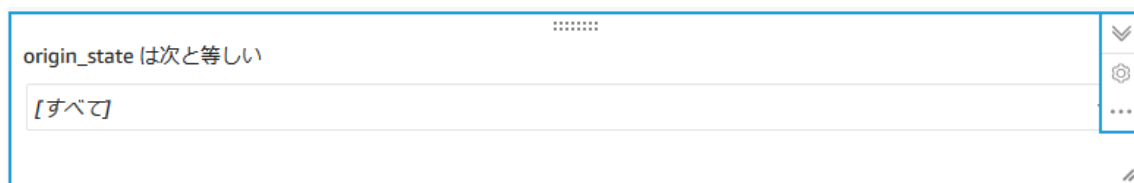
円グラフの空文字部分をクリックして“空を除く”を選択してください。これを選択するとその値を除外する条件のフィルタが自動的に作成されます。フィルタはこの円グラフのビジュアルだけに有効になっています。

また、今は全地域のデータが表示されていますが、これを Vermont のみの表示のためにフィルタを追加で設定します。左側で”フィルタ”アイコンを選択し、（最上部の”+”ではなく）”フィルタ”という文字の横にある”+”を押してフィルタを追加します。

フィルタリングするフィールドに origin_state を選択します。するとフィルタが作成されますので、その右側にある…を押して、”シートに追加”を選択します。

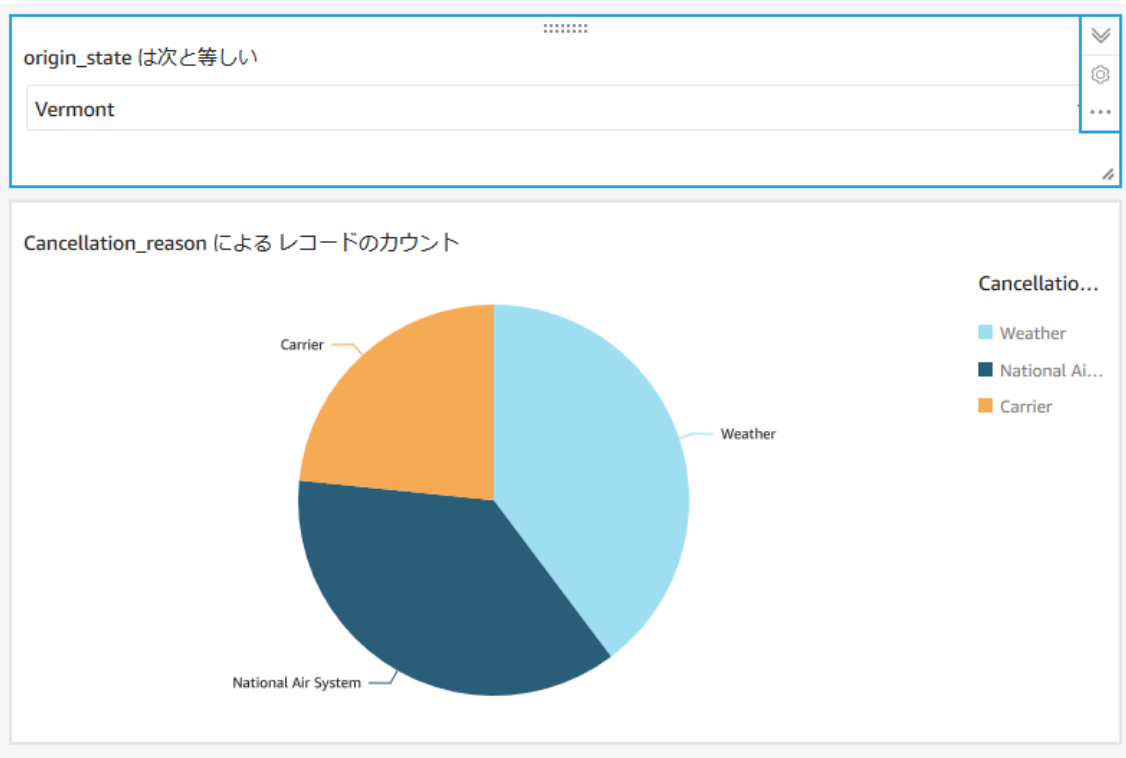


すると、新しく以下のようなフィルタコントロール（フィルタに紐付いた GUI）が作成されます。ドラッグして円グラフの上に置くなど、見やすい位置に配置してください。



このフィルタコントロールで[すべて]のチェックを外し、”Vermont”だけを選択すると円グラフが Vermont のデータだけに絞り込まれます。

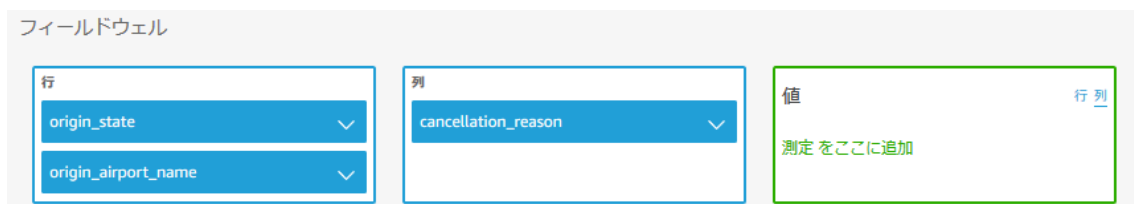
結果として以下のようなビジュアルが作成されます。Vermont でのキャンセル理由は Weather（天候）、National Air System（航空管制システム）、Carrier(航空会社)が約 1/3 ずつを占めていることが分かります。



ビジュアル 5: キャンセル理由（ピボットテーブル）

各州でのキャンセル理由を数値で把握するため、ピボットテーブルを作成します。

左上の”+追加”->“ビジュアルを追加”で新しいビジュアルを追加し、ビジュアルタイプは”ピボットテーブル”（上から3段目、左から5つ目）を選択します。フィールドウェルの行に origin_state を D&D し、origin_airport_name をその下に追加します。列には cancellation_reason を D&D します。



ピボットテーブルが描画されますが、一番右にある列(“empty”)は不要な列なので、これを表示データから外します。Empty の列名をクリックし、“空を除く”をクリックします。

最終的な画面は以下のようになり、各州・各エアポート別のキャンセル件数をピボットテーブルで確認できるようになりました。

フィルタ

carrier_name

上 - 5

および

cancellation_reason

次と等しくない -

フィールドウェル

行

origin_state

origin_airport_name

列

cancellation_reason

値

測定をここに追加

シート 1

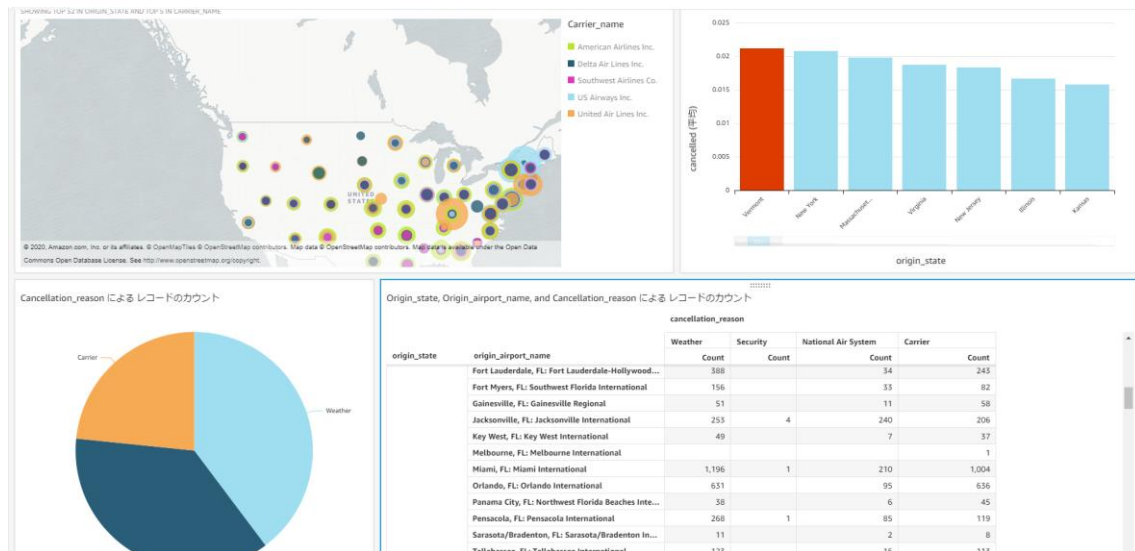
+

Origin_state, Origin_airport_name, and Cancellation_reason によるレコードのカウント

		cancellation_reason			
		Weather	Security	National Air...	Carrier
origin_state	origin_airport_n...	Count	Count	Count	Count
Alabama	Birmingham, AL:...	260	1	143	173
	Dothan, AL: Dot...	44		8	31
	Huntsville, AL: H...	195		158	201
	Mobile, AL: Mobi...	245	2	102	98
	Montgomery, AL:...	155	1	22	84
Alaska	Anchorage, AK: T...	5		4	33
	Fairbanks, AK: Fa...				4

(オプション) ビジュアルのサイズや位置の調整

各ビジュアルは右下をドラッグすることでサイズを変更したり、位置を調整したりすることが可能です。見やすい位置や大きさに調整してください



これで本ハンズオンは完了です。お疲れ様でした！

まとめと後片付け

QuickSight の費用について

QuickSight はサインアップから 30 日間は無料で試用できますが、ハンズオン 2 において「異常値検出」を作成した場合、その異常値の演算には無料枠はありませんので費用が発生します。例えばスケジュールを「毎日」に設定した場合は毎日演算が自動実行されます。異常値検出用に作成したインサイトの右にある“…”メニューから「削除」を選択することで削除できます。

QuickSight の利用全体を完全に削除する場合は QuickSight のトップ画面でリージョンをバージニア北部に変更した後、右上のユーザアイコン→QuickSight の管理→アカウントの設定を選択し、“サブスクリプション解除”で削除が可能です。（作成したデータセットや SPICE に投入したデータは全て削除されます）

S3 の費用について

S3 にはデータを保存している量に応じた費用とアクセスに応じた費用が発生します。今回ハンズオンでは直接 S3 に保存をしていませんが、パート 4 で Athena セットアップして使用すると、そのクエリ結果が S3 に保存される点には注意してください。

Athena の費用について

Athena は利用しなければ（クエリを実行しなければ）費用は発生しません。ただし、クエリの結果を保存した S3 バケットには保存費用が発生しますので、必要なければバケットを削除してください。

補足情報・参考資料（QUICKSIGHT）

- ホームページ <https://aws.amazon.com/jp/quicksight/>
- 料金 <https://aws.amazon.com/jp/quicksight/pricing/>
- ドキュメント <https://aws.amazon.com/jp/documentation/quicksight/>
- ハンズオンダウンロード（販売管理編ハンズオン等）
<https://aws.amazon.com/jp/blogs/news/amazon-quicksight-handson-202006/>
- サービス全体の説明資料
 - <https://www.slideshare.net/AmazonWebServicesJapan/20180228-aws-black-belt-online-seminar-quicksight-89889593>
 - <https://www.slideshare.net/AmazonWebServicesJapan/20180801-aws-black-belt-online-seminar-amazon-quicksight>
 - <https://www.slideshare.net/AmazonWebServicesJapan/20200204-aws-black-belt-online-seminar-amazon-quicksight>
- 計算フィールド、表計算関数
 - <https://aws.amazon.com/jp/blogs/news/advanced-analytics-with-table-calculations-in-amazon-quicksight/>
- Level-Aware Aggregation(LAA)関連
 - <https://aws.amazon.com/jp/blogs/news/create-advanced-insights-using-level-aware-aggregations-in-amazon-quicksight/>

■Amazon Athena

- ホームページ <https://aws.amazon.com/jp/athena/>
- 料金 <https://aws.amazon.com/jp/athena/pricing/>
- ドキュメント <https://aws.amazon.com/jp/documentation/athena/>
- サービス説明資料 <http://www.slideshare.net/AmazonWebServicesJapan/aws-black-belt-online-seminar-2017-amazon-athena>
- Athena パフォーマンス Tips <https://aws.amazon.com/jp/blogs/news/top-10-performance-tuning-tips-for-amazon-athena/>
- Athena がサポートする圧縮形式
https://docs.aws.amazon.com/ja_jp/athena/latest/ug/compression-formats.html
- Athena のパーティショニング
https://docs.aws.amazon.com/ja_jp/athena/latest/ug/partitions.html

補足情報・参考資料（S3 他）

■Amazon S3

- ホームページ <https://aws.amazon.com/jp/s3/>
- 料金 <https://aws.amazon.com/jp/s3/pricing/>
- ドキュメント <https://aws.amazon.com/jp/documentation/s3/>
- サービス説明資料 <https://www.slideshare.net/AmazonWebServicesJapan/aws-black-belt-online-seminar-2017-amazon-s3>

■その他の AWS サービスについて

その他、AWS の各種サービスについての資料は以下にまとまっています。

<https://aws.amazon.com/jp/aws-jp-introduction/>