

AMAZON QUICKSIGHT ハンズオン

Amazon Web Services Japan

初版 2018/06/21, 最終更新 : 2019/02/07

目次

本資料について.....	3
事前準備と環境.....	3
ハンズオンの流れ.....	3
QuickSight の利用を開始する	5
QuickSight にサインアップする	5
ハンズオン 1 : CSV ファイルをアップロードし、QuickSight で分析	8
QuickSight と SPICE.....	8
CSV ファイルをダウンロードする	8
QuickSight (SPICE)に CSV ファイルをアップロードする	9
QuickSight でコンボチャートを作成する	13
パラメータとコントロールの作成.....	18
パラメータを含んだフィルタ（絞り込み条件）の作成	21
ハンズオン 2 : 大規模データを Athena 経由でクエリし、可視化する	23
QuickSight+Athena+S3 の構成	23
Athena とサンプルデータ	24
Athena からサンプルデータをクエリする	24

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.....	32
QuickSight から Athena に接続する	36
計算フィールド (Calculated Field) の追加	40
QuickSight を利用した可視化	42
Visual 1: 航続距離が長いキャリア Top 5 (横棒グラフ)	45
Visual 2: キャンセル率が高い州 (地図)	48
Visual 3: キャンセル率が高い州 (縦棒グラフ + ドリルダウン)	49
Visual 4: Vermont のキャンセル理由 (円グラフ)	52
Visual 5: キャンセル理由 (ピボットテーブル)	54
Visual のサイズや位置の調整	54
まとめと後片付け.....	56
補足情報・参考資料.....	57

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/>

ハンズオンでは QuickSight 以外に、Amazon S3（以下 S3）、Amazon Athena（以下 Athena）を利用しますので、最低限それらを利用できる権限を持つユーザ（IAM ユーザ）を使用してください。

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。本ハンズオンでは、QuickSight は無料枠の範囲（最初の 1 ユーザのみ）、Athena については \$1 未満の費用、S3 についても無料枠内で利用できるよう設計されていますが、ハンズオン記載以上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（最初の 1 ユーザ、SPICE 1GB 分は無料）

<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（Firefox もしくは Chrome の最新版）とインターネットに繋がる環境が必要です。

ハンズオンの流れ

最初に QuickSight のアカウント（AWS アカウントとは別です）にサインアップした後で、以下の2パートのハンズオンに分かれています。順に実行していただく事で QuickSight の使い方が体験できるようになっていますが、パート1とパート2は独立しているため、どちらか一方だけを実施することも可能です。

- ハンズオン：パート1

手元のデータ（CSV ファイル）を QuickSight にアップロードしてクイックに可視化を体験するハンズオンです。

- ハンズオン：パート2

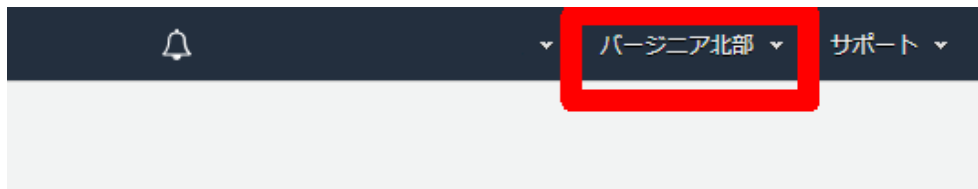
S3 上に配置された巨大データに対して、Athena を通じてクエリ（SQL 実行）できるように設定し、その環境に QuickSight でアクセスして可視化することを体験するハンズオンです。

QUICKSIGHT の利用を開始する

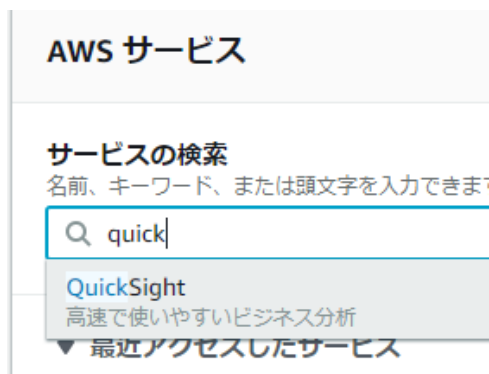
本章では QuickSight にサインアップし、利用開始できるまでを説明します。すでに QuickSight を利用されている場合は本章をスキップしてください。

QUICKSIGHT にサインアップする

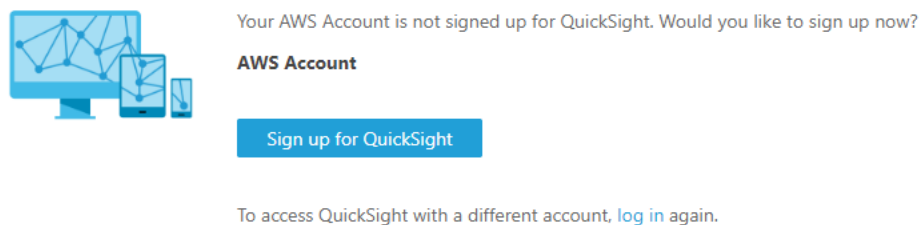
<https://console.aws.amazon.com> にアクセスし、AWS マネジメントコンソールにログインします。AWS マネジメントコンソールでは、右上のメニューから利用リージョンを「バージニア北部」に設定してください。



サービス一覧から QuickSight を選択します。検索窓で quick 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ような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。

Edition	<input checked="" type="radio"/> Standard	<input type="radio"/> Enterprise
First author with 1GB SPICE	FREE	FREE
Team trial for 60 days (4 authors)*	FREE	FREE
Additional author per month (yearly)**	\$9	\$18
Additional author per month (monthly)**	\$12	\$24
Additional readers (Pay-per-Session)	N/A	\$0.30/session (max \$5/reader/month) ****
Additional SPICE per month	\$0.25 per GB	\$0.38 per GB
Single Sign On with SAML or OpenID Connect	✓	✓
Connect to spreadsheets, databases & business apps	✓	✓
Access data in Private VPCs		✓
Row-level security for dashboards		✓
Hourly refresh of SPICE data		✓
Secure data encryption at rest		✓
Connect to your Active Directory		✓
Use Active Directory Groups ***		✓
Send email reports		✓

Standard と Enterprise どちらでも本ハンズオンは実行可能ですので好きな方を選択してください。最初の 1 ユーザはどちらのエディションでも無料で利用可能です（SPICE 利用領域は 1GB まで無料）。Standard Edition を選択しても後から Enterprise Edition にアップグレードが可能のため、ここでは Standard Edition を選択して、画面下方にある Continue を押します。

Create your QuickSight account

Edition

Standard

QuickSight account name



You will need this for you and others to sign in.

Notification email address

For QuickSight to send important notifications.

QuickSight region

US East (N. Virginia) ▼



Select a region.

> ☒ Enable autodiscovery of data and users in your Amazon Redshift, Amazon RDS and AWS IAM services.

☐ Amazon Athena
Enables QuickSight access to Amazon Athena databases

☐ Amazon S3
Enables QuickSight to auto-discover your Amazon S3 buckets

[Choose S3 buckets](#)

☐ Amazon S3 Storage Analytics
Enables QuickSight to visualize your S3 Storage Analytics data

☐ Amazon IoT Analytics
Enable QuickSight to visualize your IoT Analytics data

Finish

この画面では、QuickSight アカウントの設定を行ないます。一番上の QuickSight account name は任意の名前を付けます。

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

QuickSight region は、QuickSight の 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）。

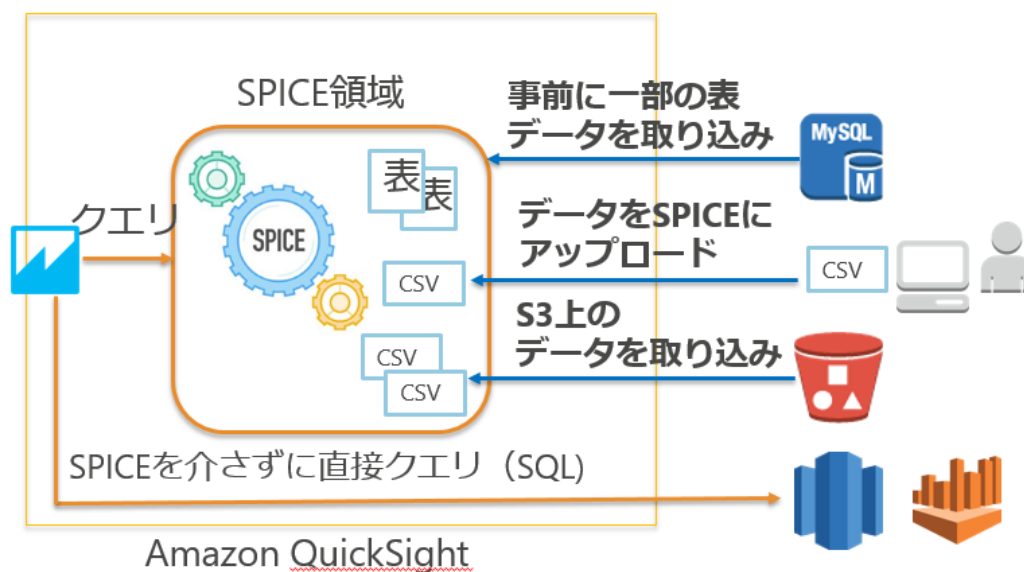
それ以外のチェックボックスは上記のように” Enable autodiscovery of…” にだけチェックし、Finish を押してください（それ以外の部分はハンズオンの途中で設定します）。

これで QuickSight が利用可能になりました。

ハンズオン 1 : CSV ファイルをアップロードし、QUICKSIGHT で分析

QUICKSIGHT と SPICE

QuickSight には SPICE というインメモリ型的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。

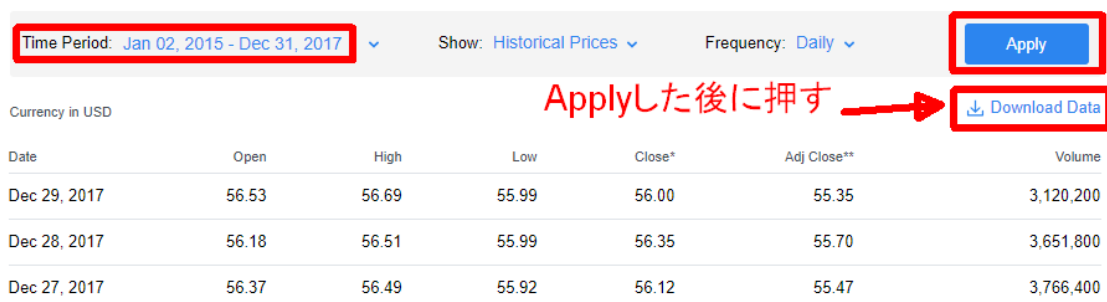
CSV ファイルをダウンロードする

このハンズオンでは、CSV ファイルの例として株価データの CSV ファイルを利用します。米 Yahoo! Finance (<https://finance.yahoo.com>) のサイトで任意の企業名を検索して、Historical Data を確認してください。例えばデルタ航空 (シンボルは DAL) は以下の URL です。

<https://finance.yahoo.com/quote/DAL/history>

ここでは 2015 年 1 月 2 日から、2017 年 12 月 31 日までの 3 年間分のデータをダウンロードします(1 月 1 日からにすると 2014 年 12 月 31 日のデータが含まれてしまいます)。Time Period で期間を設定し、Apply (適用) をクリックしてください。もしくは以下の URL をクリックしても同じ画面に移動できます。

<https://finance.yahoo.com/quote/DAL/history?period1=1420124400&period2=1514646000&interval=1d&filter=history&frequency=1d>



Time Period: Jan 02, 2015 - Dec 31, 2017 Show: Historical Prices Frequency: Daily Apply

Applyした後に押す → Download Data

Currency in USD

Date	Open	High	Low	Close*	Adj Close**	Volume
Dec 29, 2017	56.53	56.69	55.99	56.00	55.35	3,120,200
Dec 28, 2017	56.18	56.51	55.99	56.35	55.70	3,651,800
Dec 27, 2017	56.37	56.49	55.92	56.12	55.47	3,766,400

期間を Apply したら、Apply ボタンの下にある Download Data で CSV ファイルがダウンロードできますので、任意のフォルダ内に保存します。

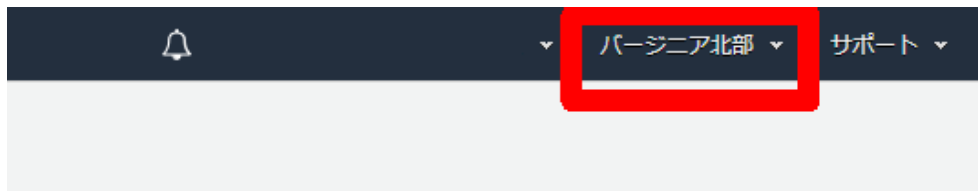
ダウンロードしたファイルを確認すると、以下のように 2015 年 1 月から 2017 年 12 月までの株価が記録されているはずです。

	A	B	C	D	E	F	G
1	Date	Open	High	Low	Close	Adj Close	Volume
2	2015/1/2	49.92	50.01	48.71	49.18	46.38771	8637300
3	2015/1/5	49.46	49.48	47.81	48.36	45.61427	10556500
4	2015/1/6	48.55	48.74	46.25	47.22	44.539	12880400
5	2015/1/7	47.48	47.59	46.51	46.97	44.30319	10516200
6	2015/1/8	47.48	48.32	47.26	48.19	45.45392	10499300

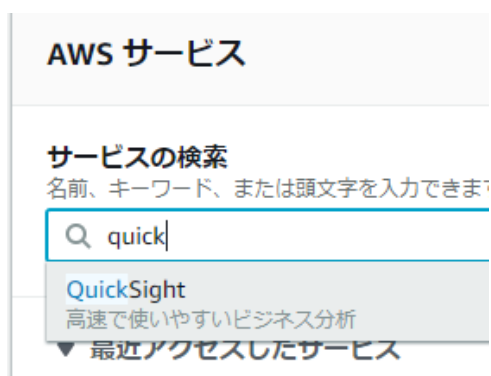
QUICKSIGHT (SPICE) に CSV ファイルをアップロードする

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

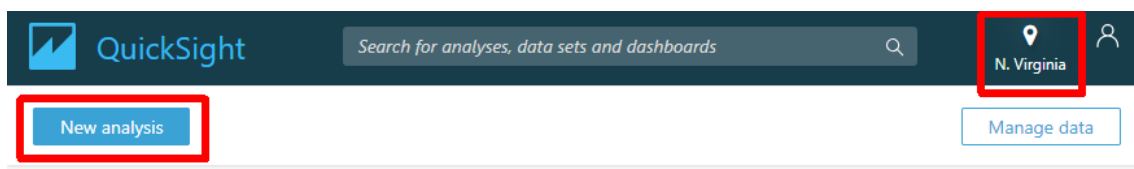
ト作成済であった場合は、そのアカウントの QuickSight region に切り替えてください（以前に QuickSight を東京リージョンで開始していた場合は東京を選択）。



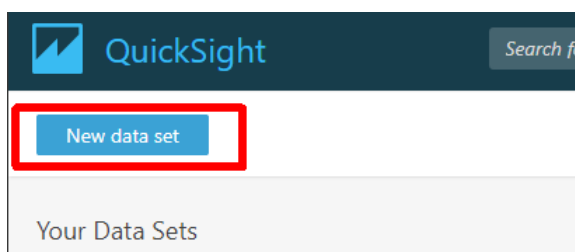
サービス一覧から QuickSight を選択します。

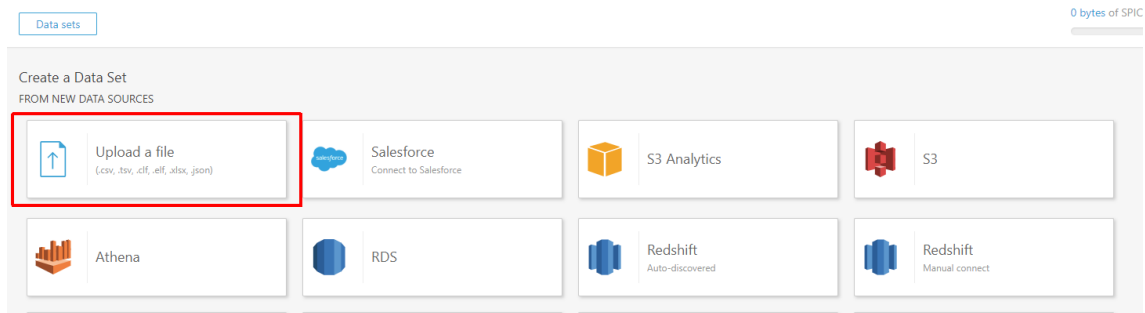


QuickSight の画面でも念の為、先程指定したリージョンが選択されていることを確認してください。確認できたら、左上の” New analysis” をクリックします。

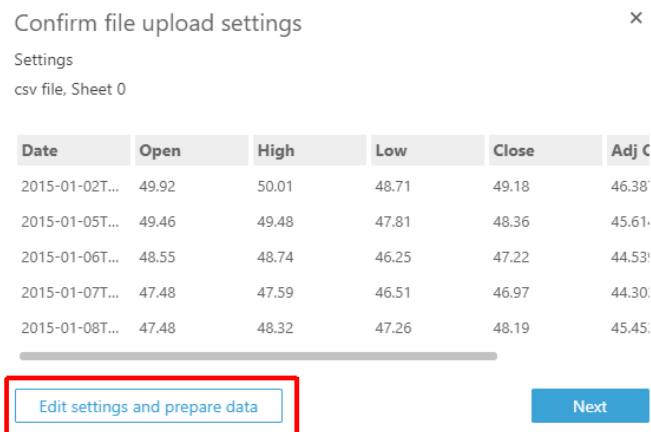


次の画面で” New data set” をクリックし、” Upload a file” をクリックします。

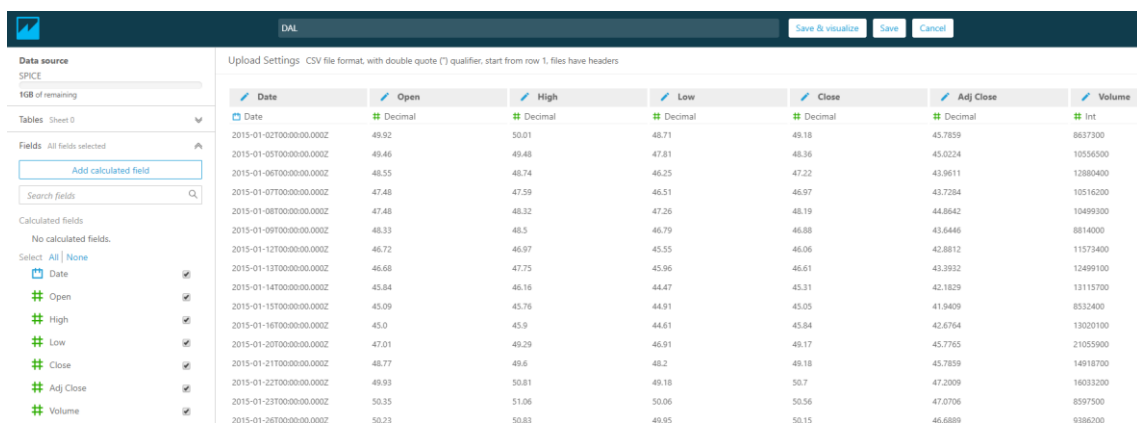




ファイル選択のダイアログで先程ダウンロードした CSV ファイルを選択します。確認画面が出ますので、そこで” Edit settings and prepare data” を選択します。

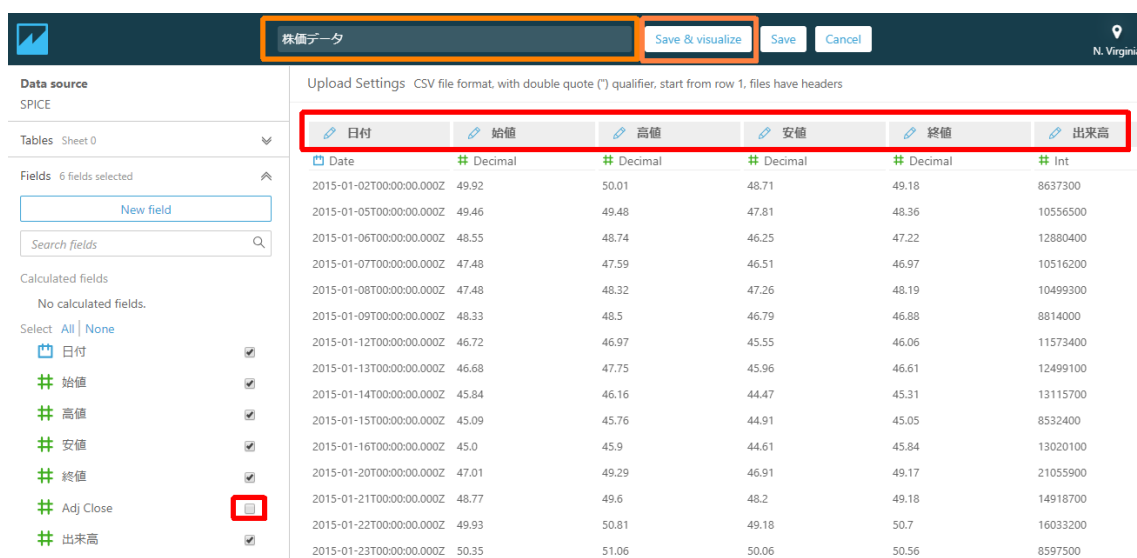


以下のような画面が表示されます。これは” Prepare” 機能で、分析の前にデータを利用しやすい形に準備するためのものです。



ここままで分析することも可能ですが、分析しやすい形になるようここで少し設定を行ないます。

- 左側には列 (Field) 一覧が並んでいますが、このチェックボックスを外すと不要な列を削除できます。今回は”Adj close” は利用しませんのでチェックを外します
- 右側には列名、列の型、データが表示されています。列名は CSV の最初の行から自動的に生成されますが、分かりやすいように別の名前をつけることが可能です。ここでは以下のように修正します (各列の鉛筆マークをクリックすると列名を変更できます)
 - Date→日付
 - Open→始値
 - High→高値
 - Low→安値
 - Close→終値
 - Volumes→出来高
- また、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「株価データ」としました。



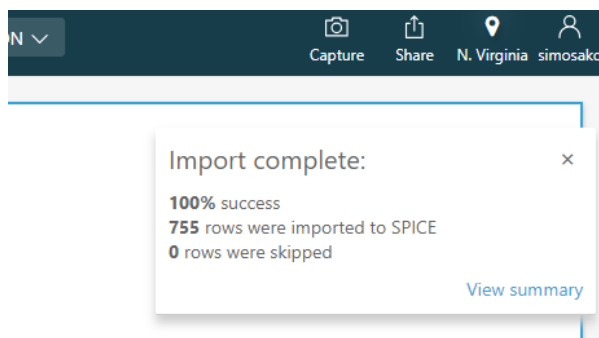
株値データ

Save & visualize Save Cancel

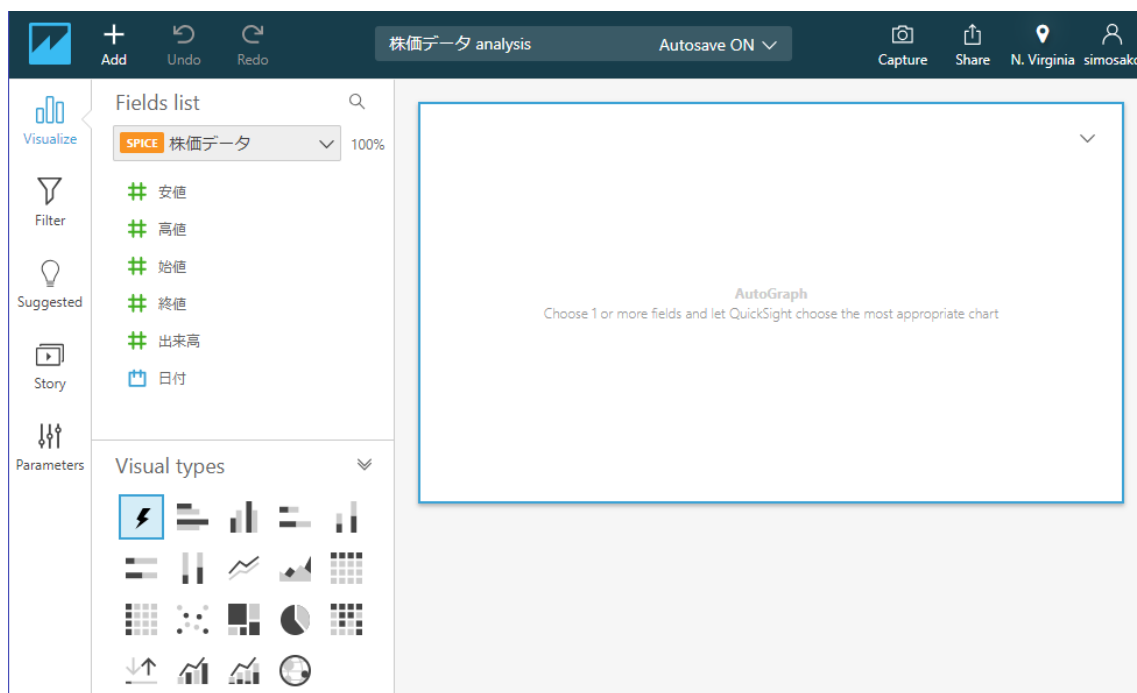
Upload Settings CSV file format, with double quote (") qualifier, start from row 1, files have headers

日付	始値	高値	安値	終値	出来高
2015-01-02T00:00:00.000Z	49.92	50.01	48.71	49.18	8637300
2015-01-05T00:00:00.000Z	49.46	49.48	47.81	48.36	10556500
2015-01-06T00:00:00.000Z	48.55	48.74	46.25	47.22	12880400
2015-01-07T00:00:00.000Z	47.48	47.59	46.51	46.97	10516200
2015-01-08T00:00:00.000Z	47.48	48.32	47.26	48.19	10499300
2015-01-09T00:00:00.000Z	48.33	48.5	46.79	46.88	8814000
2015-01-12T00:00:00.000Z	46.72	46.97	45.55	46.06	11573400
2015-01-13T00:00:00.000Z	46.68	47.75	45.96	46.61	12499100
2015-01-14T00:00:00.000Z	45.84	46.16	44.47	45.31	13115700
2015-01-15T00:00:00.000Z	45.09	45.76	44.91	45.05	8532400
2015-01-16T00:00:00.000Z	45.0	45.9	44.61	45.84	13020100
2015-01-20T00:00:00.000Z	47.01	49.29	46.91	49.17	21055900
2015-01-21T00:00:00.000Z	48.77	49.6	48.2	49.18	14918700
2015-01-22T00:00:00.000Z	49.93	50.81	49.18	50.7	16033200
2015-01-23T00:00:00.000Z	50.35	51.06	50.06	50.56	8597500

上記のように変更したら、“Save & visualize” 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 Import complete となったらデータの取り込みが完了しています。



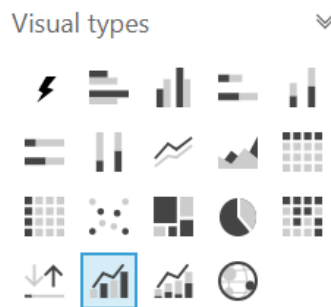
QUICKSIGHT でコンボチャートを作成する



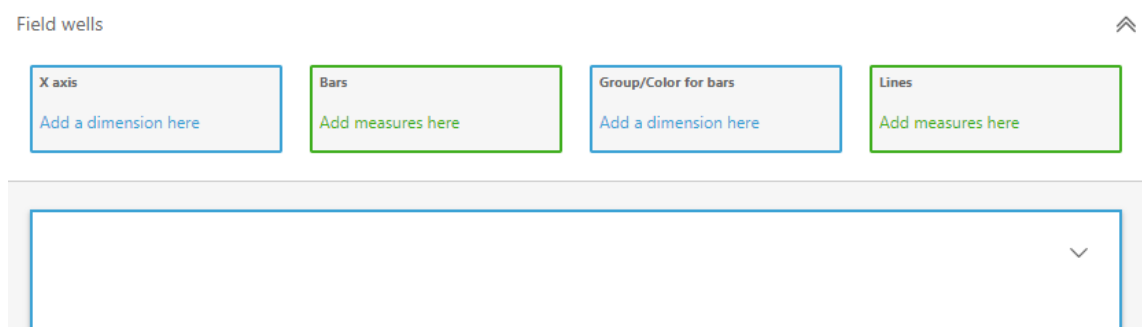
ここからは SPICE に取り込んだ CSV ファイルを可視化していきます。

画面左上に“Fields list”があり、読み込んだデータの列 (Field) が表示されています。画面右にある白い領域が“Visual”です。ここに左下にある Visual Type (グラフの種類) を設定して利用します。左上にある“+”ボタンで Visual は複数追加することが可能です。また Visual はサイズを変更したり位置を入れ替えたりすることが可能です。

デフォルトでは Visual Type が” AutoGraph” になっています。これは選択した Field によって自動的に Visual Type を選択する機能ですが、今回は Clustered bar combo chart (Visual Type の一番下、左から 2 つ目) を選択します。

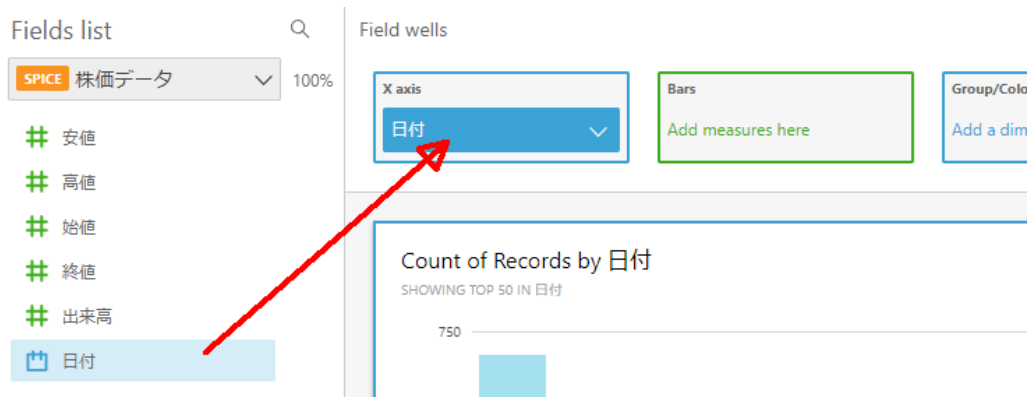


選択すると、右側の Visual の上に” Field wells” が準備されます。

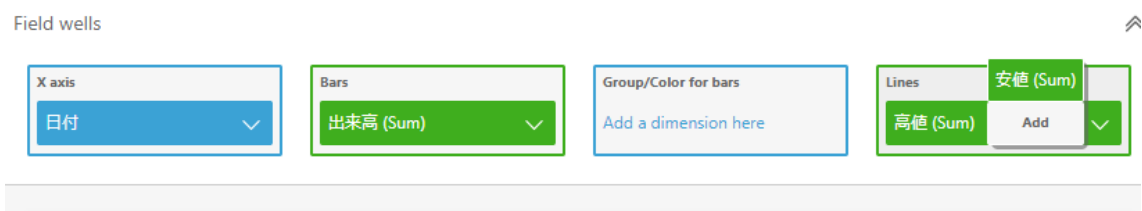


この 4 つの領域に、左から Field を Drag & Drop (D&D) することでその Field データをもとに可視化が行われます。

まず横軸 (X axis) から設定します。今回は時間にそったデータの変遷を見たいため、” 日付” Field を D&D します。

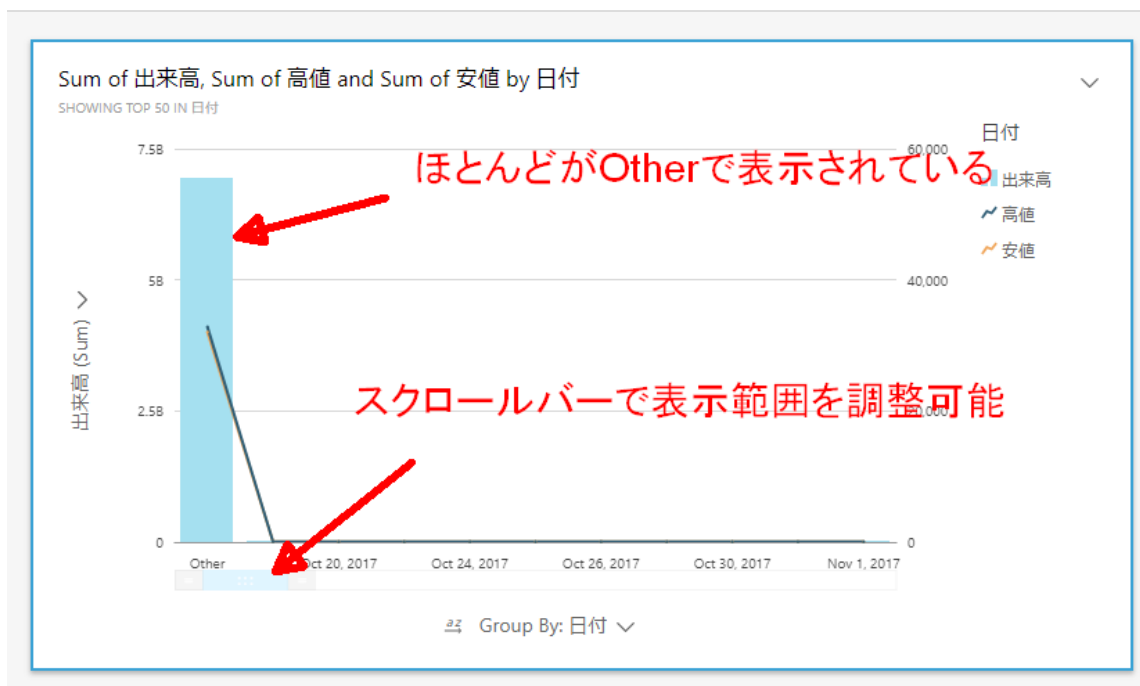


同様に” Bars”（棒グラフ）には、” 出来高” を D&D します。Group/Color の枠は今回利用しません。一番右の” Lines”（折れ線グラフ）にはまず” 高値” を D&D し、次に“安値” も D&D します。2つ目を入れる際に下図のように Add と出る位置で D&D してください。（Replace と出ている場合は最初に入れた” 高値” と入れ替わってしまいます）



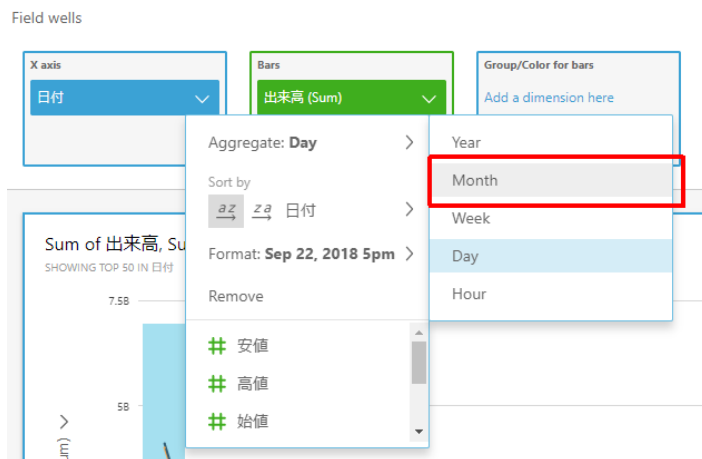
Field wells

<p>X axis</p> <p>日付 ▼</p>	<p>Bars</p> <p>出来高 (Sum) ▼</p>	<p>Group/Color for bars</p> <p>Add a dimension here</p>	<p>Lines</p> <p>高値 (Sum) ▼</p> <p>安値 (Sum) ▼</p>
---------------------------	--------------------------------	---	--



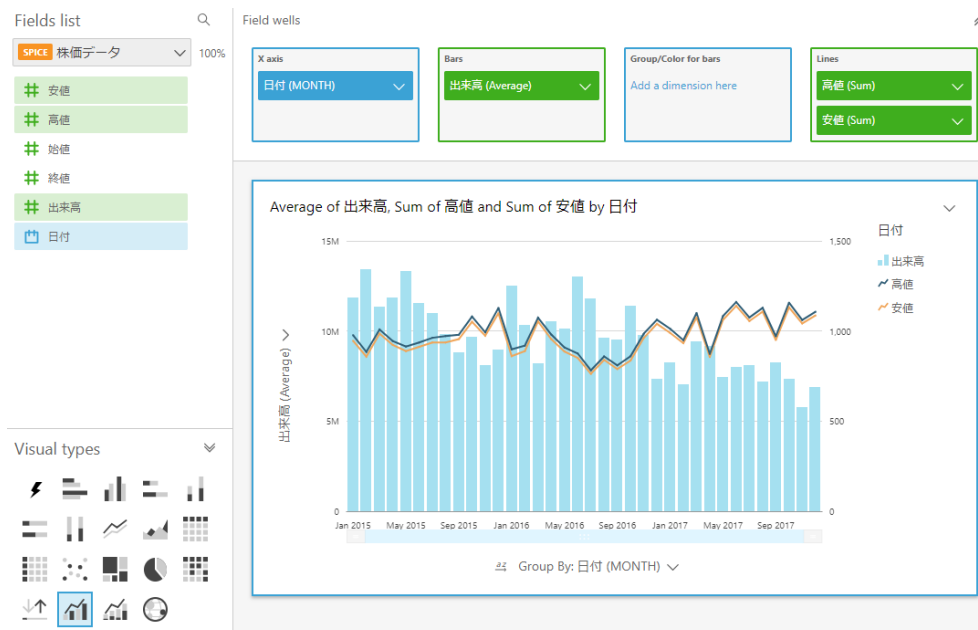
このような Visual が出力されました。この図はまだ自分が見たいデータになっていないので、少し修正します。グラフの左側に Other の棒グラフが出来ていますが、これは表示の粒度「日」単位になっており、表示すべき量が多くて画面に入り切らず、結果として入らないものが Other として積算されるためです。このような場合は週や月で集約することで全体を確認できます。

現在 X axis (日付) の粒度は、日単位で集計 (Aggregate) されているので、これを月単位にします。X axis のボックスに置いた「日付」の右側にある▼マークをクリックして、Aggregate を Month に切り替えます。



グラフが更新されます。グラフの下側にある横スクロールのバーを移動する、もしくは伸縮させることで表示範囲を調整することが可能です。

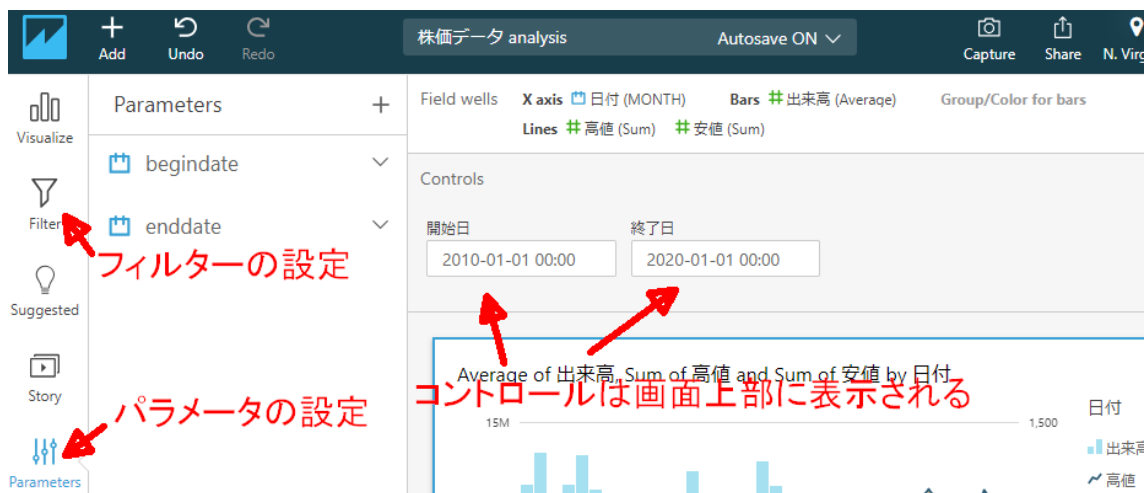
初期状態ではグラフの値が合計（Sum）になっているためデータを合計した値が表示されますが、ここで必要なのは平均値なので Average に変更する必要があります。出来高、高値、安値ともに (Sum) になっていますが、これらそれぞれで▼をクリックして、Aggregate（集計方法）を Average（平均）に変更します。以下のようなグラフが得られます。（見やすいように横スクロールバーを調整してください）



パラメータとコントロールの作成

QuickSightには、Filter(フィルタ)機能があります。Filterはデータの閲覧範囲をユーザ側で絞り込むための機能です。例えば今回は2015/1/2~2017/12/31のデータを持っていますが、そのうちの2016年だけを表示するといった事が可能です。Filterは画面左端の”Filter”で固定的に設定することが可能ですが、見たい範囲を頻繁に変更するのであれば、Parameter(パラメータ)とControl(コントロール)との組み合わせで利用するのが便利です。

パラメータは、一種の変数で。コントロールは入力用のGUIです。コントロールに入れられた値がパラメータ(変数)に保存されるよう設定できます。また、このパラメータをフィルタの絞り込み条件に適用することが可能です。



まずは変数であるパラメータと、入力枠(コントロール)作成します。今回は自分が閲覧したい期間の絞り込みをするために、開始日(begindate)と終了日(enddate)という2つのパラメータを作成します。左端の”Parameters”をクリックすると、まだパラメータが作成されていないので、何も無い状態です。ここで”Create one...”を選択します。以下のようなダイアログが表示されるので begindate パラメータを作成します。Name(名前)には begindate と入れ、Data type(型)は Datetime、Default value(デフォルト値)は、2010-01-01 00:00 等の過去の値を指定(カレンダーGUIをクリックして選択するか、日付を直接入力してリターンを押す)します。

Create new parameter ×

Use parameters to dynamically control values in your fields, filters and sheet

Name

Data type (Not alterable after creation)

Datetime ▼

Default value

2010-01-01 00:00 Set a dynamic default

Cancel Create

次の画面では、この作成したパラメータをユーザが操作できるように Control (コントロールの GUI) を紐づけます。Control をクリックします。

Parameter Added ×

Connect your parameter:

Create a filter, using the combination of parameters, new control, and a filter.

Filter

Create a new control for a filter or a calculated field.

Control

Use a parameter in a calculated field.

Calculated field

Create a URL action with parameters.

Custom actions

Close

Control のダイアログボックスでは、このパラメータがどのような名前で画面上みえるのか (Display name) と、どういう入カスタイル (Style) かを選択します。

Add control for parameter

Parameter

begindate

Display name

開始日

Style

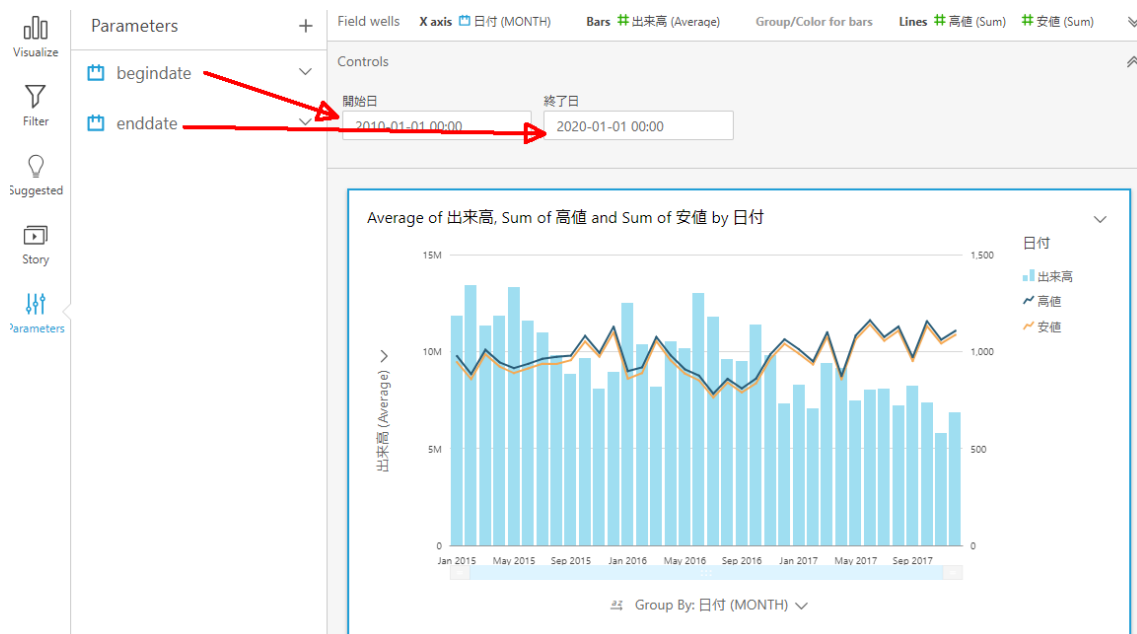
Date picker

Cancel

Add

上図のように、Display name に” 開始日” と入力し、Style は” Date picker” のままにして、Add をクリックしてください。

同様に、enddate もパラメータを作成してください。Parameters の横にある” +” を押すと新しいパラメータが追加できます。enddate という名前で、Datetime 型、デフォルト値は 2020-01-01 00:00 等の未来の時刻にしておきます。同様に enddate に紐付いたコントロールを” 終了日” として作成します。

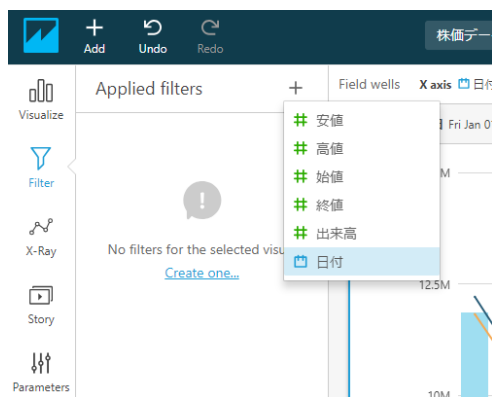


コントロールが Visual の上部に配置されました。（折り畳まれていて見えない場合は、右端の矢印マークをクリックして広げてください）。図に示したように、それぞれの入力枠（コント

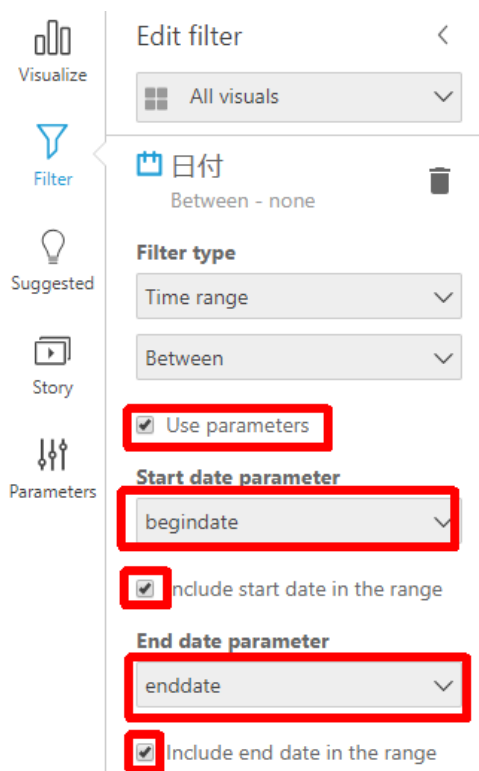
ロール) が、パラメータに結びついています。例えば利用者が開始日に入力した値は、begindate というパラメータ (変数) に入ります。

パラメータを含んだフィルタ (絞り込み条件) の作成

パラメータの準備が出来たので、パラメータに入れられた値に従って日付の範囲を絞り込むようにフィルタを作成します。左端の” Filter” をクリックして、” Create one...” を選択し、フィルタ対象である” 日付” を選択します (日付に対して絞り込みを実施するため)。



左側に” 日付” のフィルタが表示されるので、それをクリックし、以下のように設定します。

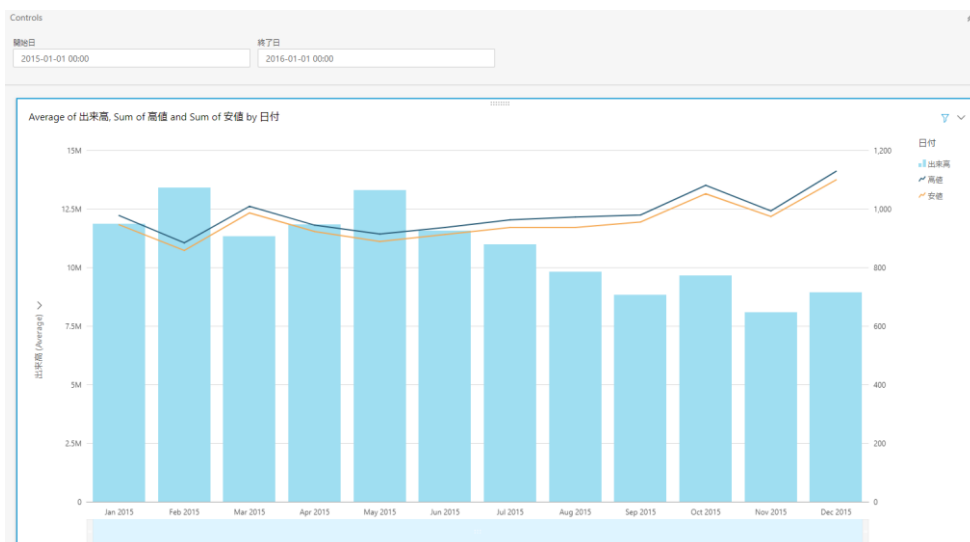


Filter typeはTime range(日付)で Between(範囲)を指定し、” Use parameters” にチェックを入れます。この時” Change the scope of this filter ?” と聞かれますので Yes を押します。

Start date parameter に begindate、End data parameter に enddate を設定します。また” Include start date in the range” と” Include end date in the range” 両方にチェックを入れます。これにより、begindate～enddate の日付の範囲でフィルタ（絞り込み）が実行されます。設定できたら Apply をクリックすることでフィルタ設定が完了します。

設定したコントロール、パラメータ、フィルタの連携が機能しているかを試してみましょう。コントロールの開始日を 2015-01-01 00:00 に、終了日を

2016-01-01 00:00 に設定してみてください。自動的に Visual が更新され、2015 年のデータのみになるのが確認できます。



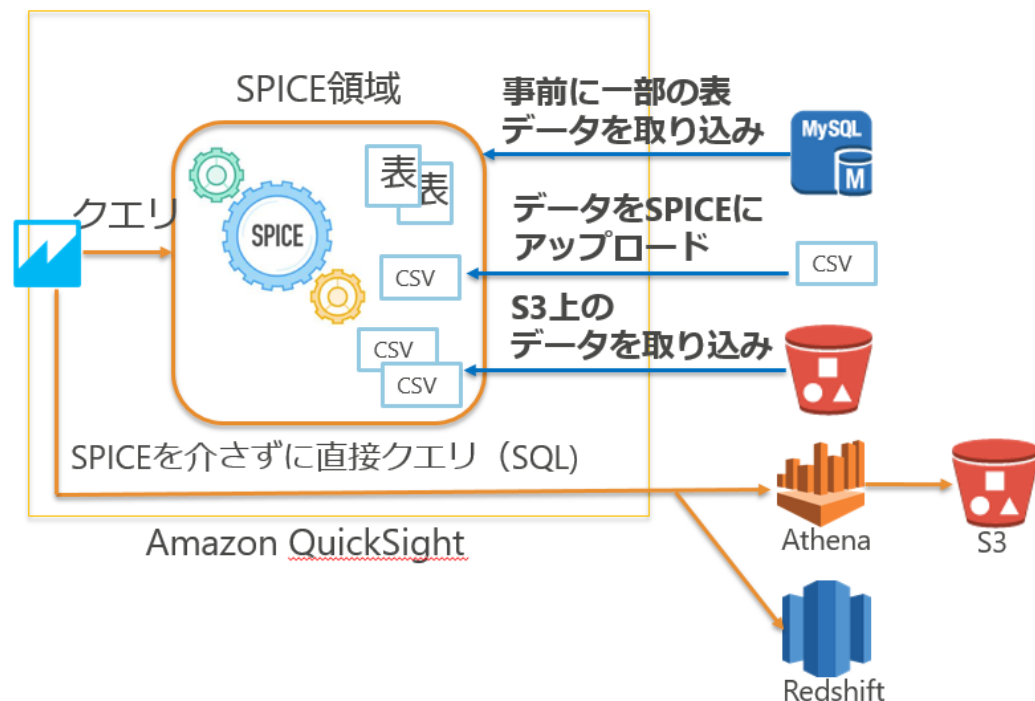
これで本ハンズオンは完了です。

ハンズオン 2 : 大規模データを ATHENA 経由でクエリし、可視化する

QUICKSIGHT + ATHENA + S3 の構成

QuickSight 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）

SPICE を使わずに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 25GB まで」という制限を超えるデータを分析可能になりますし、SPICE に取り込む必要が無い分、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。

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。

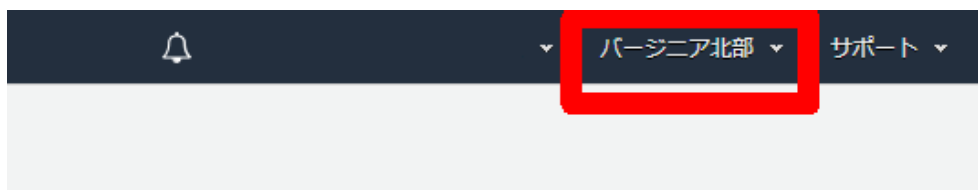
<https://aws.amazon.com/jp/athena/pricing/>

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（このハンズオン 2 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。



初めて Athena にアクセスした場合は” Get Started” ボタンを押して利用開始してください。
その場合、利用チュートリアルが開始されますので、ダイアログの右上にあるバツ(x)ボタンでチュートリアルを終了させてください。

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。まず、今回の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 TABLE を実行することで S3 のデータが表としてアクセスできるようになります。その表をまとめておく単位がデータベースです。

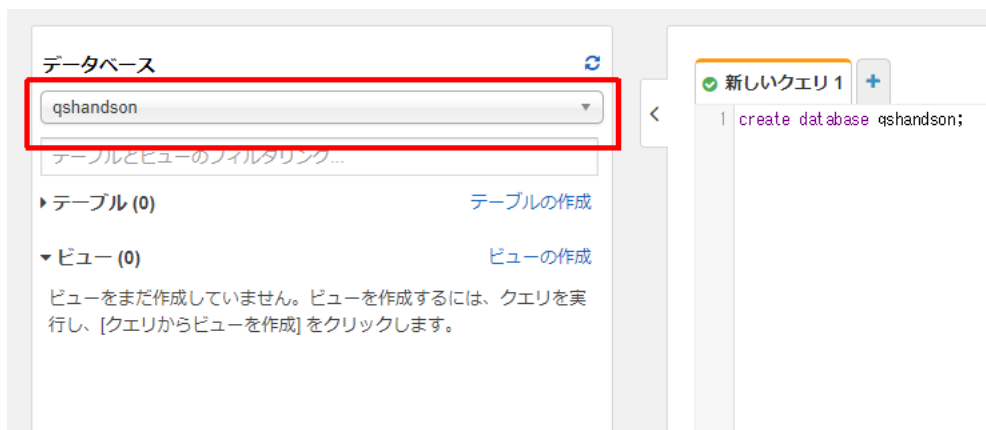
右側のペインで

```
create database qshandson;
```

と入力して” クエリの実行” をクリックします。



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” が作成され、それが選択された状態になりました。この後の操作はこの qshandson データベースに対して行われます。

まず 1 つ目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 URL に保存されています。

s3://quicksight-bi-handson/flightdata1/

flightdata1/以下には 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 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：以下の画面は、S3 バケットを保有しているユーザの画面です。ハンズオン実施者の管理コンソールからは見えませんので、以下の図はご参考までに掲載しています。AWS CLI が利用可能な環境にある場合は、`aws s3 ls s3://quicksight-bi-handson/flightdata1/` で確認可能です。）

Amazon S3 > quicksight-bi-handson / flightdata1

概要

🔍 プレフィックスを入力し、Enter キーで検索します。ESC を押してクリアします。

📁 アップロード + フォルダの作成 さらに ▾

米国東部 (バージニア北部) 🔄

表示中 1 ~ 1

<input type="checkbox"/> 名前 ↑	最終更新日時 ↑	サイズ ↑	ストレージ クラス ↑
<input type="checkbox"/> 📄 flightdata1.tsv.gz	6月 17, 2018 4:17:19 午後 GMT+0900	621.4 MB	スタンダード

表示中 1 ~ 1

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 ではデータをスキャンした容量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のが一般的なお勧めです。

```

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
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 -12.000000 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 CREATE EXTERNAL TABLE で S3 上のデータに表定義を与える必要があります。TSV で記載されたデータの各列がどのような名前、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。

今回は以下の 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。（長いので以下をコピー&ペーストして利用してください。なおダウンロードした ZIP ファイルにコピー用のテキストファイルが同梱されていますので、そちらをお使いいただく事も可能です。）

```

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://quicksight-bi-handson/flightdata1'
;

```

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,                <= 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://quicksight-bi-handson/flightdata1' <=データの位置を指定
```

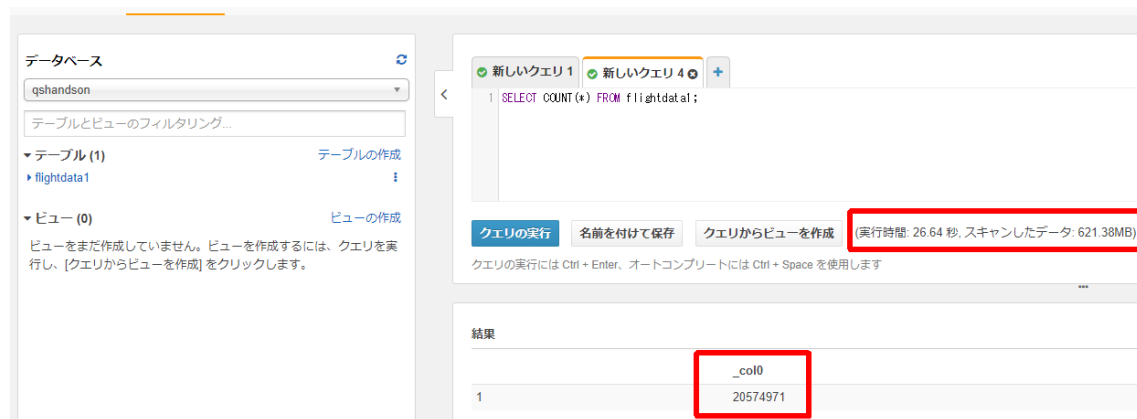
これで s3://quicksight-bi-handson/flightdata1/以下に置かれた tsv. gz ファイルが表として SQL でクエリできるようになりました。画面左側に flightdata1 という表が作成されているはずです。

The screenshot displays the Amazon Redshift console interface. On the left, the 'データベース' (Database) section shows 'qshandson' selected, with a list of tables under 'テーブル (1)'. The 'flightdata1' table is expanded, showing its columns and data types: 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), and carrier_name (string). On the right, the '新しいクエリ 1' (New Query 1) editor shows a SQL statement: '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);'. Below the editor, the 'クエリの実行' (Execute Query) button is highlighted. At the bottom, the '結果' (Results) section shows a message: 'クエリは成功しました。' (Query was successful).

“+”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 30 秒で結果が出力されます。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。



データベース: qshandson

テーブルとビューのフィルタリング...

▼ テーブル (1) テーブルの作成

▶ flightdata1

▼ ビュー (0) ビューの作成

ビューをまだ作成していません。ビューを作成するには、クエリを実行し、[クエリからビューを作成] をクリックします。

新しいクエリ 1 新しいクエリ 4 +

1 SELECT COUNT(*) FROM flightdata1;

クエリの実行 名前を付けて保存 クエリからビューを作成 (実行時間: 26.64 秒, スキャンしたデータ: 621.38MB)

クエリの実行には Ctrl + Enter、オートコンプリートには Ctrl + Space を使用します

結果

	_col0
1	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 50 秒程度時間がかかります。



新しいクエリ 1 新しいクエリ 4 新しいクエリ 5 +

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

クエリの実行 名前を付けて保存 クエリからビューを作成 (実行時間: 50.84 秒, スキャンしたデータ: 621.38MB)

クエリの実行には Ctrl + Enter、オートコンプリートには Ctrl + Space を使用します

結果

	_col0	_col1
1	2014	5819811
2	2013	5899736
3	2010	142466
4	2012	5735649
5	2011	2977309

(補足 : Athena で実行したクエリの結果セットは、ユーザの S3 バケットに自動的に保存されます。保存先はデフォルトでは `s3://aws-athena-query-results-xxxx-us-east-1/` のようなバケットになっていますが、右上の” 設定 ” ボタンから変更することも可能です。)

このように CREATE EXTERNAL TABLE を実行するだけで、データベースサーバを起動する事なく S3 上のファイルに SELECT 文が実行できる事が確認できました。しかしクエリ時間が速くありません。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は主に 2 つあります。1 つはファイルを圧縮することで、最初のサンプルデータでも Gzip で圧縮が行われていました。もう 1 つの方法がファイルを分割し、適切にパーティショニングする事です。

パーティショニングとは、値に応じてファイルを分割して保存することです。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 `/yaer=2011/` 以下に置きます。年 (yaer) 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/`

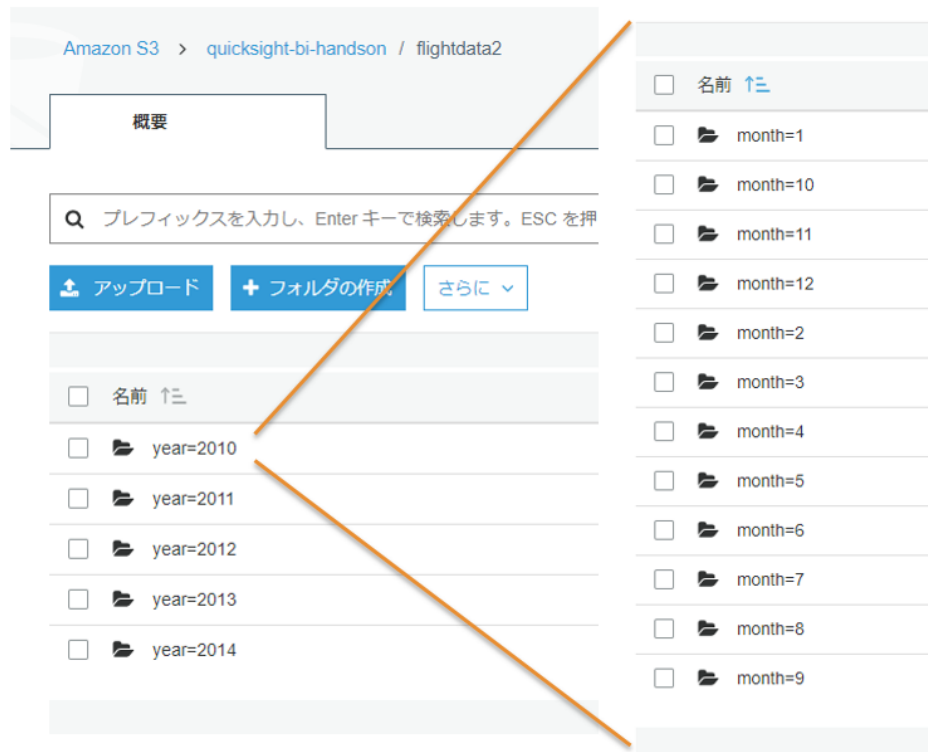
`/month=1/2010 年 1 月のデータ.gz`

`/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も 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）



このようなパーティショニングで保存されたデータが `s3://quicksight-bi-handson/flightdata2/` に用意されています。（補足：AWS CLI が利用可能な環境にある場合は、`aws s3 ls s3://quicksight-bi-handson/flightdata2/` とすることでフォルダ構成を確認できます）

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://quicksight-bi-handson/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) <=パーティションの指定が追加
ROW FORMAT DELIMITED FIELDS TERMINATED BY '¥t'
LOCATION 's3://quicksight-bi-handson/flightdata2' <=URL が変更
;
```

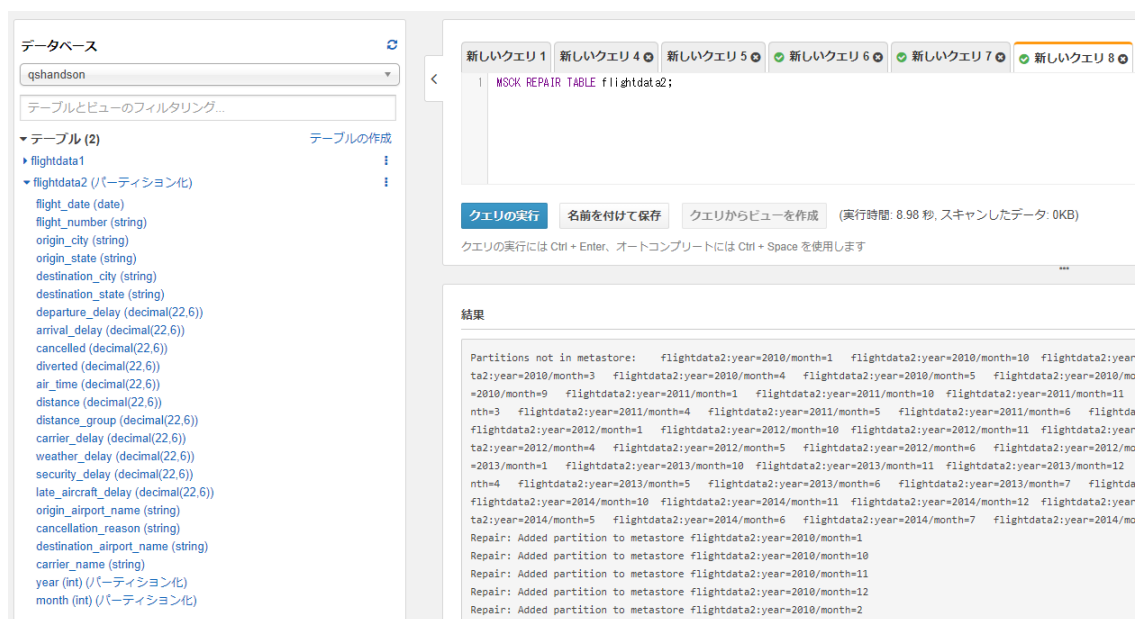
パーティションの指定が追加されたのが大きな変更点です。S3 上で/year=…/month=…/とパーティショニングしているため、それを指定しています。

CREATE EXTERNAL TABLE を実行した際に、結果として以下のような表示が出ます。

「クエリは成功しました。 テーブルにパーティションがある場合、データをクエリできるようにこのパーティションをロードする必要があります。すべてのパーティションをロードするか、パーティションを個別にロードことができます。すべてのパーティションのロード (MSCK REPAIR TABLE) コマンドを使用する場合、パーティションは Hive が理解できる形式である必要があります。」

パーティショニングされた表を CREATE EXTERNAL TABLE した場合は、パーティションがどのように存在するかを Athena に伝える必要があります。1 フォルダごとに、ALTER TABLE … ADD PARTITION 文で追加することも可能ですし、本サンプルのように Hive が理解できる形式でパーティショニングしている場合は、MSCK REPAIR TABLE コマンドを一度実行すれば全パーティションが自動的に認識されます。クエリエディタで以下のコマンドを実行します。

```
MSCK REPAIR TABLE flightdata2;
```



これで flightdata2 表にクエリが可能になりました。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 YAEER(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;

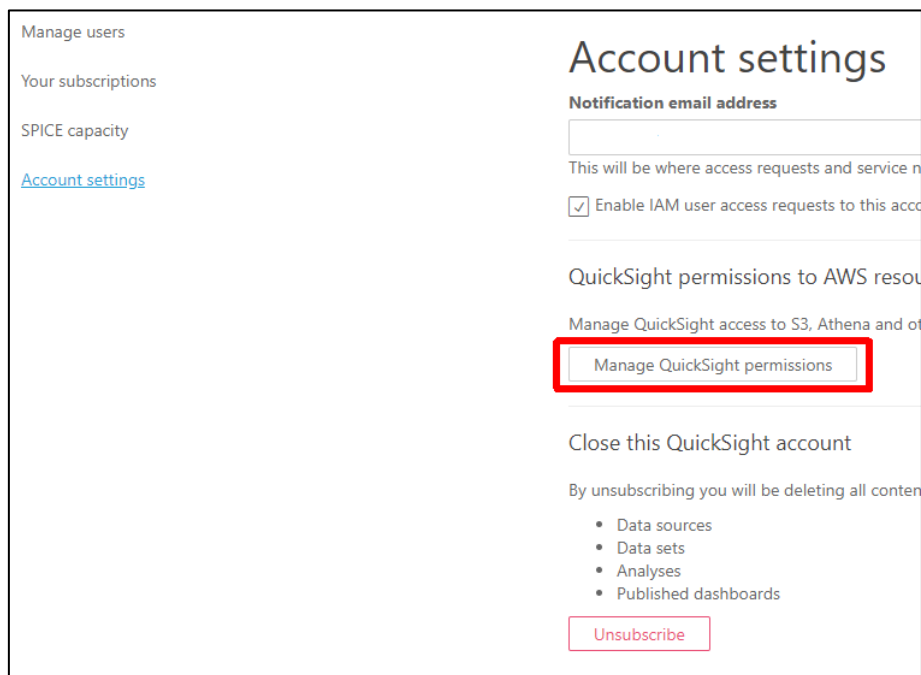
パーティションとして指定した year はそのまま列のように使用できるため、上記のような書き方が可能になります。こうすることで速度改善だけでなく、2) のようなケースでは、スキャンするデータは year=2010/以下に限定できるため、スキャン対象のデータサイズを減らすことが可能です。これにより利用費用も削減可能です。

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り（”AWS”ロゴをクリック）、QuickSight の画面に移動します。QuickSight 画面の右上で”N.Virginia”が選択されている事を確認してください。

まず、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。画面右上端の自分の ID が書いてある人の形をしたアイコン→”Manage QuickSight”を選択し、左のメニューから”Account Settings”を選択します。この画面で”Manage QuickSight permissions”を選択します。



Edit QuickSight read-only access to AWS resources

> ☒ Enable autodiscovery of data and users in your Amazon Redshift, Amazon RDS and AWS IAM services.

☒ Amazon Athena
Enables QuickSight access to Amazon Athena databases

Please ensure the right Amazon S3 buckets are also enabled for QuickSight.

☒ Amazon S3 (4 buckets) [Choose S3 buckets](#)
Enables QuickSight to auto-discover your Amazon S3 buckets

☐ Amazon S3 Storage Analytics
Enables QuickSight to visualize your S3 Storage Analytics data

☐ Amazon IoT Analytics
Enable QuickSight to visualize your IoT Analytics data

QuickSight からアクセス可能なリソースを設定する画面が表示されますので、上図のようにまず Amazon Athena、次に Amazon S3 にチェックを入れてください。すると以下の画面が表示されます。

Select Amazon S3 buckets

S3 buckets linked to QuickSight account S3 buckets you can access across AWS

Select AWS S3 buckets to give QuickSight access permission to the bucket.

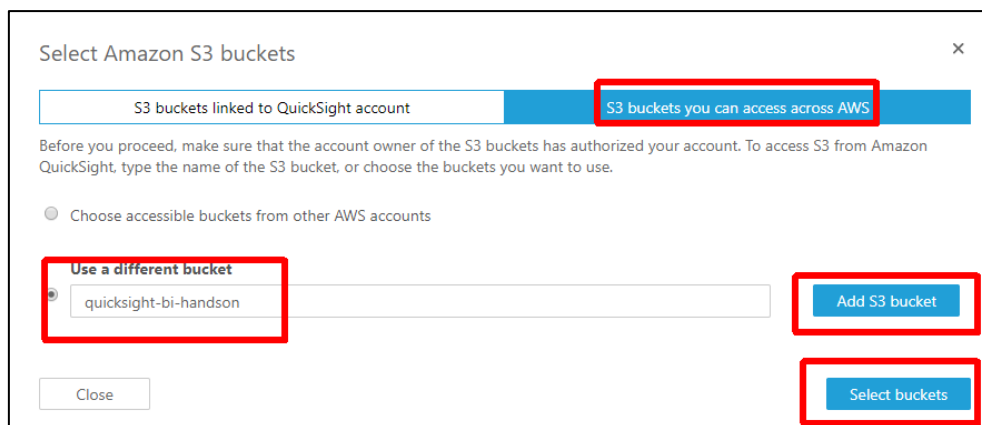
☒ Select all

☒ aws-athena-query-results-...-us-east-1

☒ aws-athena-query-results-us-east-1-...

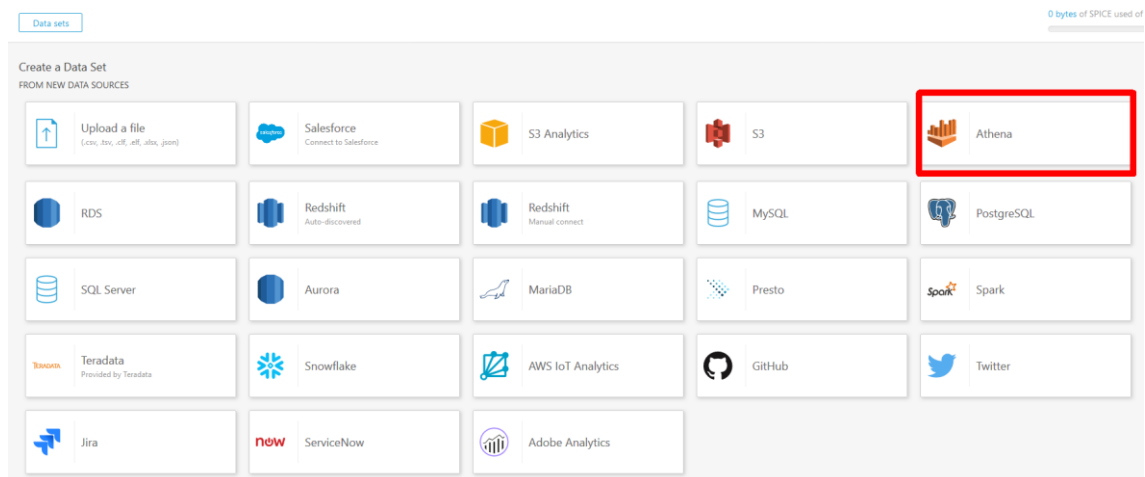
自分が保持する S3 バケット一覧が出ますので、Athena の SQL 実行結果が出力される S3 バケットにチェックを入れます（もしくは”Select All” にチェックを入れます）。次に”S3 buckets you can access across AWS” を押します。

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です。” Use a different bucket” に quicksight-bi-handson 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）、” Add S3 bucket” で追加します。

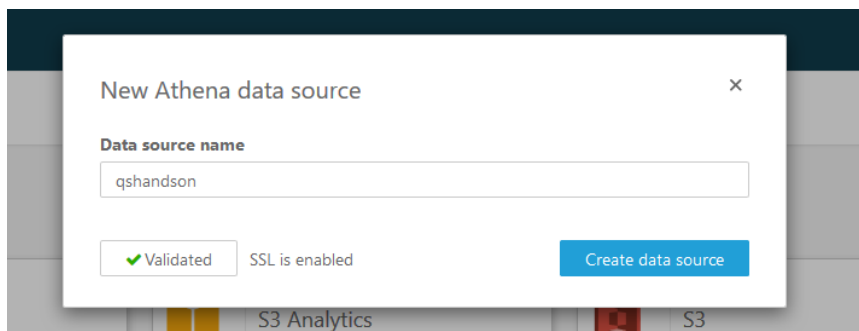


追加できたら、” Select buckets” を押して確定させ、次の画面で” Apply” を押します。設定が完了したら、左上の QuickSight アイコンをクリックしてメイン画面に戻ります。

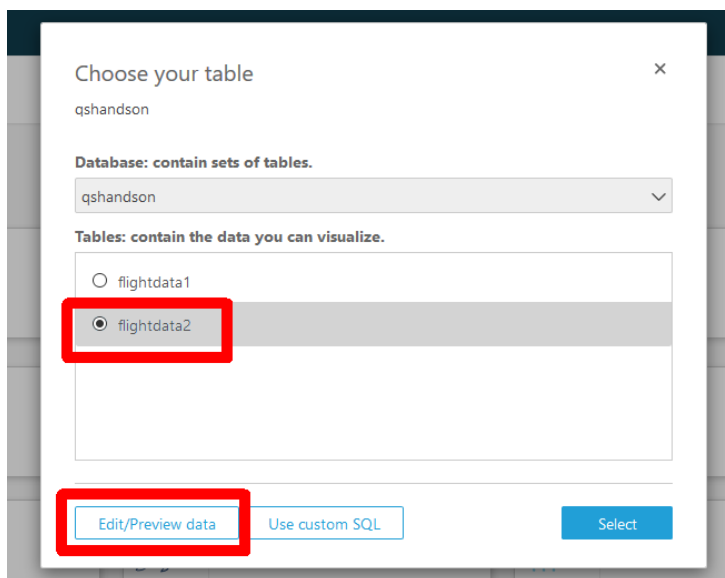
次に Athena 経由で flightdata2 表へアクセスするための” データセット” を作成します。右上の” Manage Data” をクリックし、次に左上の” New data set” をクリックします。データセットに登録可能なデータソース一覧が表示されますので、Athena をクリックします。



ダイアログで Data source name 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” Create data source” をクリックします。



Athenaに接続し、データベース名一覧が表示されるので qshandson を選択し、その中にある flightdata2 を選択して” Edit/Preview data” 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

Edit/Preview data 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、フィールドの別名（日本語）に変更したりする事が可能ですが、このハンズオンでは触れません（ハンズオン 1 を参照）。

ここでは、計算フィールド（Calculated Field）の追加を行ないます。Calculated Field は、元のデータに無い擬似的な列を計算して QuickSight 側で追加する機能です。（詳細は以下の URL を参照）

<https://docs.aws.amazon.com/quicksight/latest/user/working-with-calculated-fields.html>

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 “New field” をクリックします。

The screenshot shows the Amazon QuickSight interface for editing the 'flightdata2' dataset. On the left, the 'Fields' section is expanded, and the 'New field' button is highlighted with a red rectangle. Below it, a search bar for fields is visible. Further down, the 'Calculated fields' section shows 'No calculated fields.' and a 'Select' dropdown with 'All' and 'None' options. A list of fields is shown with checkboxes: flight_date, flight_number, origin_city, origin_state, and destination_city, all of which are checked. The main area on the right displays a preview of the data table with columns for flight_date, flight_number, origin_city, and origin_state. The data rows show flight information for 2010-01-01T... and 2010-01-03T....

flight_date	flight_number	origin_city	origin_state
2010-01-01T...	2745	El Paso, TX	Texas
2010-01-03T...	1528	Charlotte, NC	North Carolina

クリックすると calculated field を追加するダイアログが表示されるので、Calculated field name に distance_km と入力し、その下の Formula(式)に、`{distance}*1.61` と入力します ({} も入力する必要があります)。マイルを 1.61 倍することで Km に変換するという式です。

Calculated field name

distance_km

Formula

{distance}*1.61

入力したら” Create” を押します。左側のフィールドリストの一番上に distance_km が追加されたのを確認して、上部の” Save & visualize” を押します。

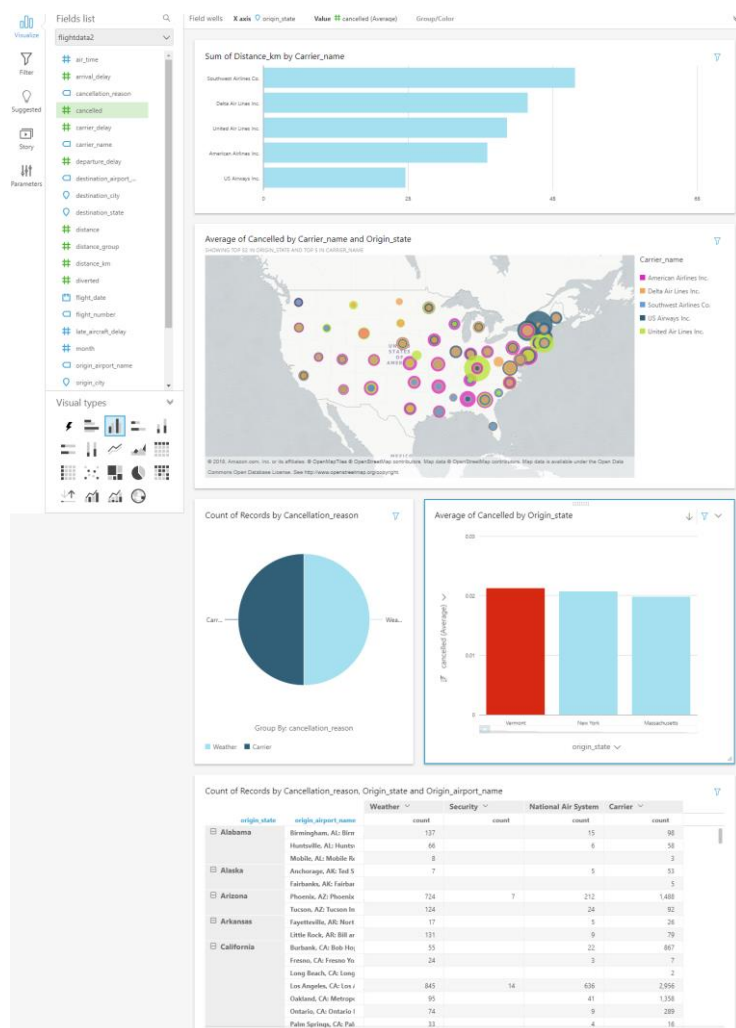
Save & visualize

Save

Cancel

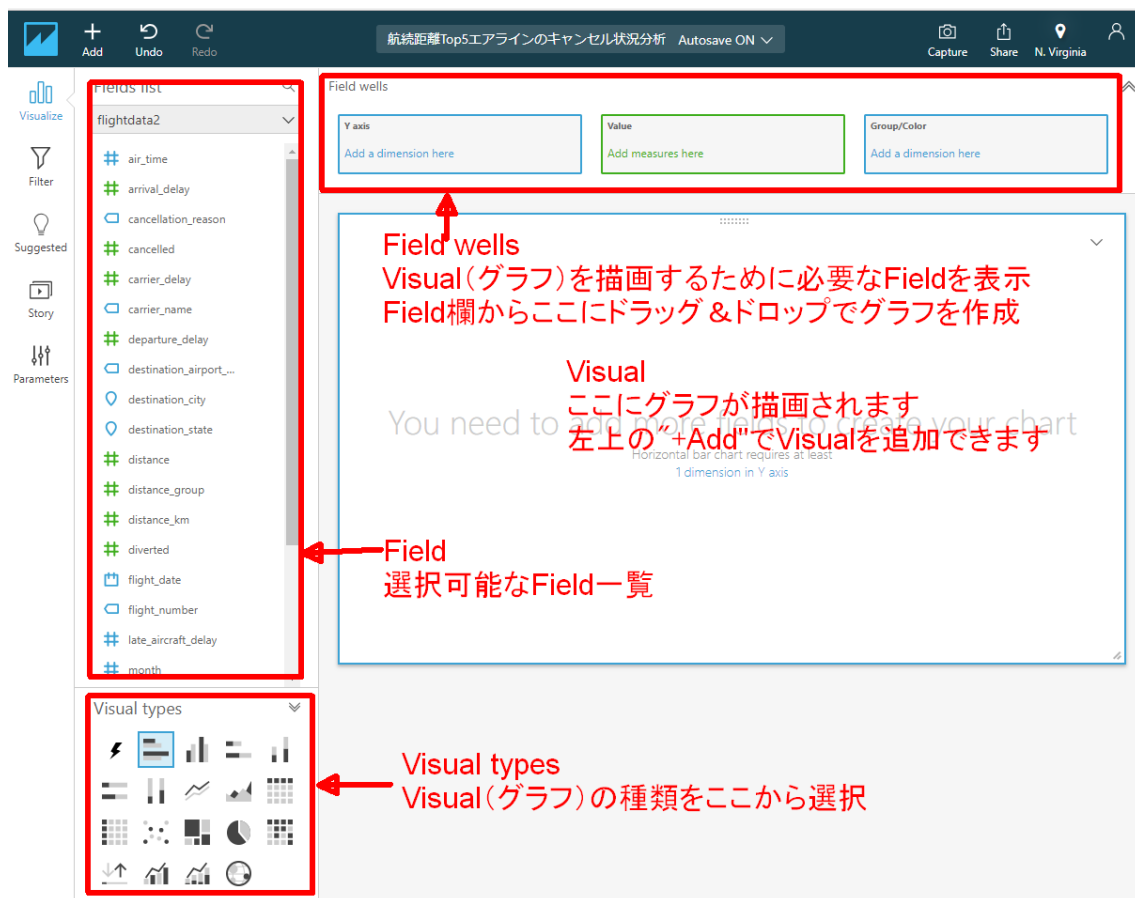
QUICKSIGHT を利用した可視化

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 Visual (図) を組み合わせて、以下のような画面を作成します。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 Analyze(分析)を作成直後は以下のような画面になります。まず、タイトルを任意の名前に変更してください。QuickSight ではデフォルトで Autosave (自動セーブ)が ON になっているため、保存の操作を行わなくても常に変更が保存されます。例ではタイトルを“航続距離 Top5 エアラインのキャンセル状況分析”としました。

画面に空白の Visual が表示されていますので、まずは左下の Visual Type 一覧から”Horizontal bar chart”（最上段、左から2つ目）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、まず Visual を追加し、Visual Type を選択した後に、Field を Field wells へドラッグ&ドロップ（以下 D&D）していくことです。Field wells は図が成り立つだけに必要な Field さえ指定すれば、全てを埋める必要はありません。（上図のように Field wells が表示されない場合は、まだ Visual Type 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” Field wells” と書かれている部分をクリックして広げてください）

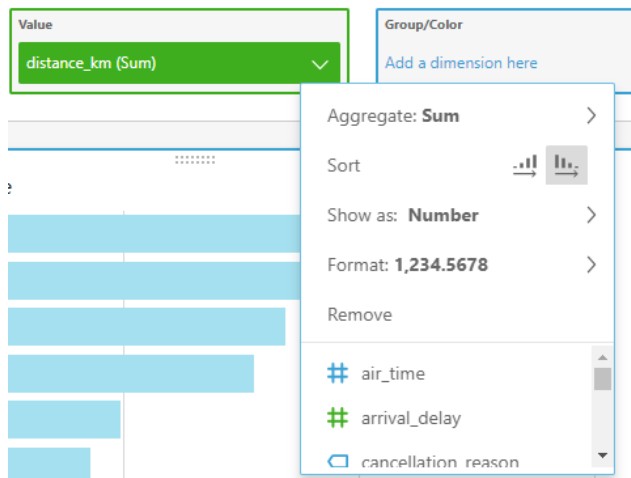
VISUAL 1: 航続距離が長いキャリア TOP 5 (横棒グラフ)

航続距離が長いキャリア Top5 を横棒グラフで可視化します。

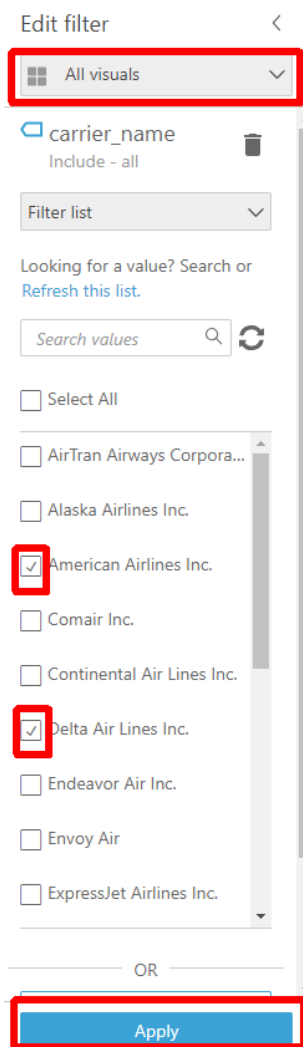
Visual type で” Horizontal bar chart” (最上段、左から2つ目) が選択されている事を確認して、FieldからはY axis (Y 軸) には carrier_name (キャリア名) を、Value (値) には distance_km (航続距離の Km 表記) を選択します。すると以下のような画面になります。



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。



Value はデフォルト（初期状態）では Sum（合計値）として集計されています。value に置いた distance_km の” v” マークをクリックすると、上記画面のように Sum 以外の集計（Average 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。（今回は変更しません）



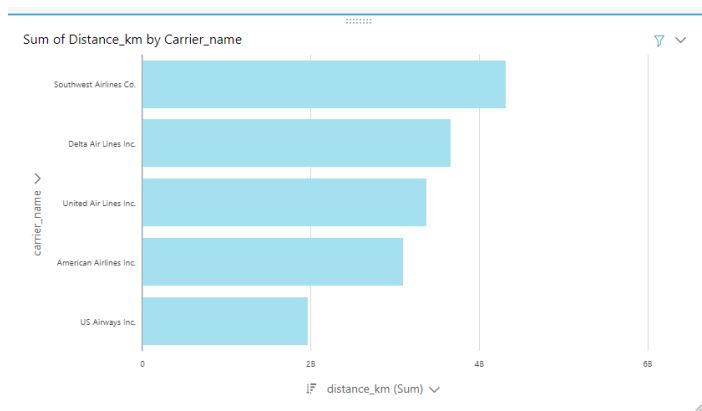
グラフを見ると、1 位から順に、Southwest Airlines, Delta Air Lines, United Air Lines, American Airlines, US Airways が上位 5 社ですので、以下ではこの 5 社に絞って分析を進めることにします。QuickSight では Filter 機能を利用すると可視化されるデータの範囲を絞り込むことが可能です。

左側の Filter アイコンをクリックし、Create one... を選択します。表示される一覧から、carrier_name を選択します。（キャリア名で絞り込むため）

carrier_name という Filter が作成されるので、クリックをして開きます。まず上部にある” Only this visual” をクリックし、” All visuals” に変更します。Filter での絞り込みは 1 つの Visual に対して適用することもできますし、今回のように分析全体の Visual に共通して適用することも可能です。

次に、リストから 上位 5 社全てにチェック を入れて、最後に” Apply” を押します。

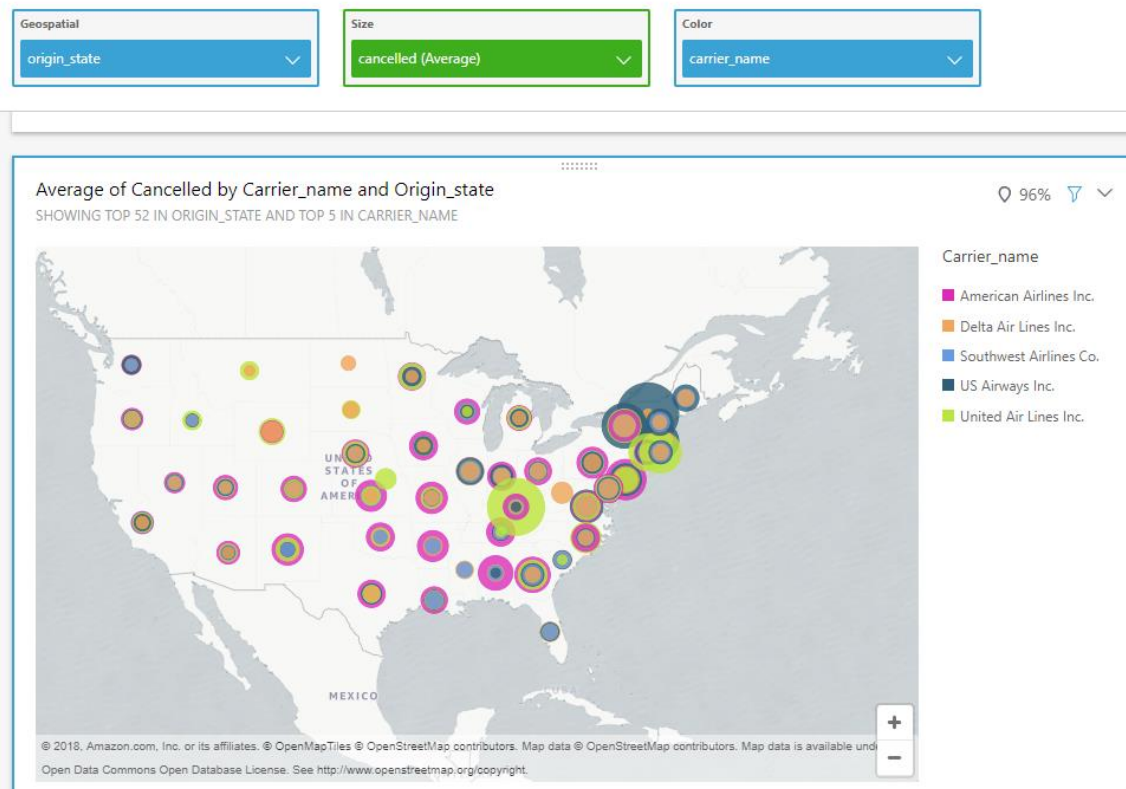
Apply をすると、Visual が 5 社に絞り込まれるのが確認できます。この後複数の Visual を追加しますが、ここで設定した Filter は共通して適用されます。



VISUAL 2: キャンセル率が高い州（地図）

Top 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” +Add” ->” Add Visual” で新しい Visual を追加し、Visual type は” Points on map”（最下段、左から4つ目）を選択します。

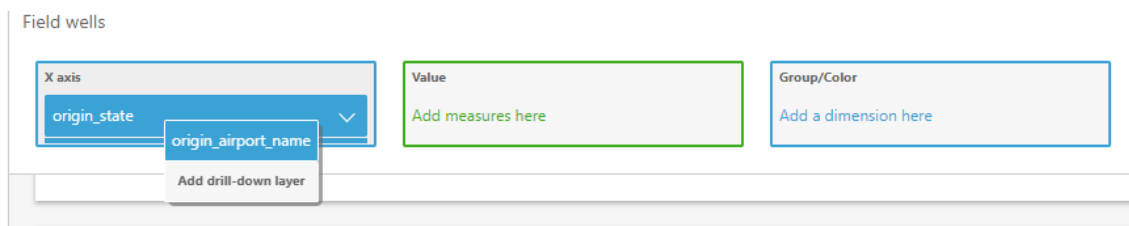
Field wells の Geospatial（地理情報）には、origin_state(出発州)、Size には cancelled、Color には carrier_name を D&D してください。また、cancelled は集計を Average に変更してください。以下のようなグラフが作成されます。Carrier_name には5つのキャリアしか表示されませんが、これは Visual 1 で作成したフィルタがこの Visual にも効いているためです。



VISUAL 3: キャンセル率が高い州（縦棒グラフ+ドリルダウン）

Visual 2 と同様にキャンセルが多い州を縦棒グラフで可視化します。グラフには州→エアポート名のドリルダウンを設定します。左上の” +Add” ->” Add Visual” で新しい Visual を追加し、Visual type は” Vertical bar chart” （最上段、左から 3 つ目）を選択します。

Field wells の X axis には origin_state を D&D します。追加でさらに origin_state の下に” add drill-down layer” と表示されるように、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。



続けて、Value には Cancelled を D&D し、集約を (Average) に変更します。以下のような結果になります。

Field wells

X axis

origin_state ▼

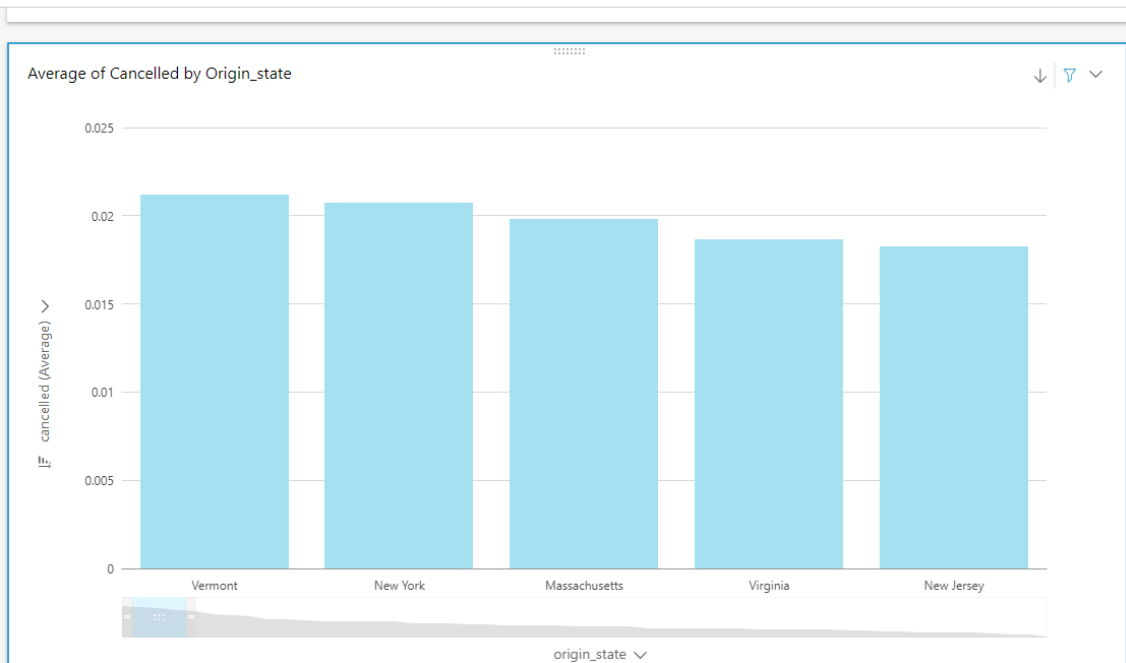
origin_airport_name ▼

Value

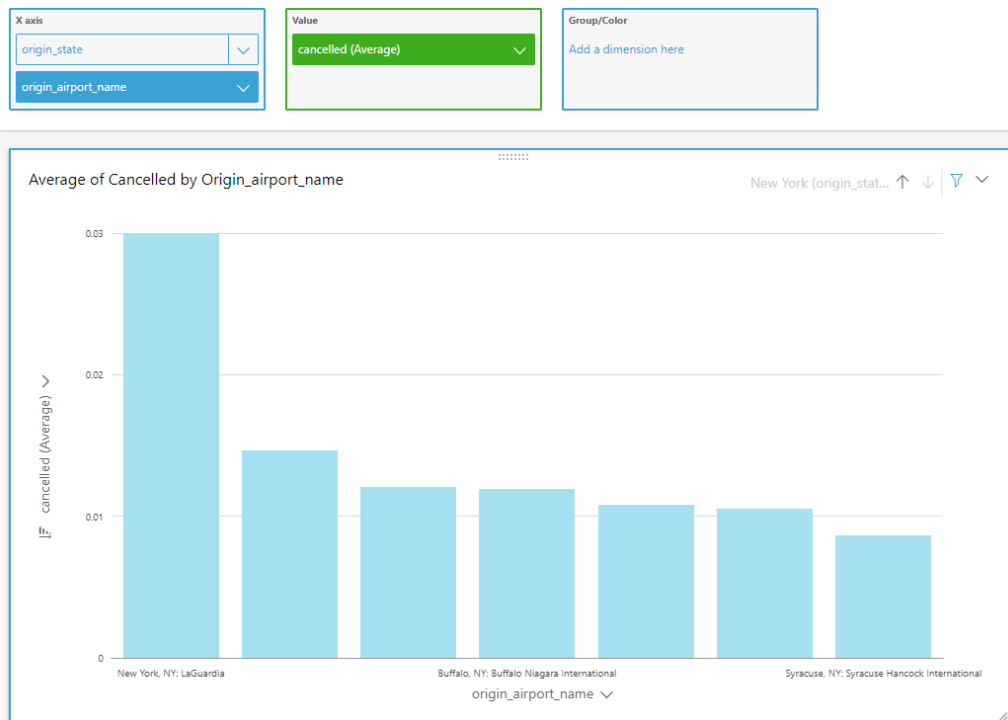
cancelled (Average) ▼

Group/Color

Add a dimension here

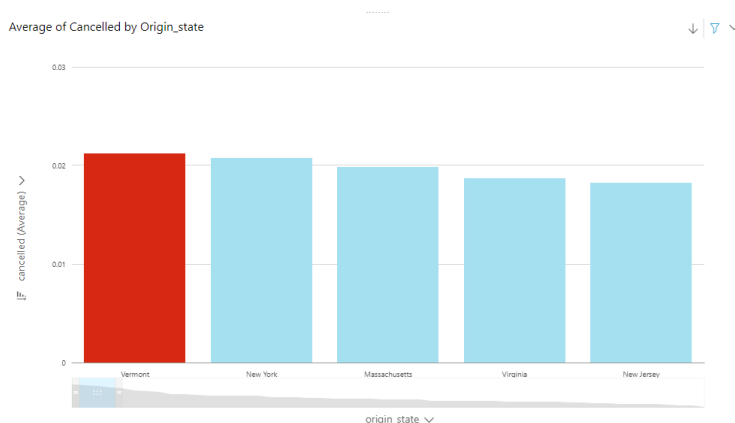


X axis で、origin_state の下位フィールドとして origin_airport_name を設定しましたので、ドリルダウンの動きを試してみます。New York のバーをクリック（左クリック）すると、Drill down to origin_airport_name と表示されるのでそこをクリックします。



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。再度グラフをクリックして、Drill up to origin_state をクリックすると、州レベルの画面に戻ります。

Vermont が一番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、Color Vermont から赤色等、目立つ色に変更してください。

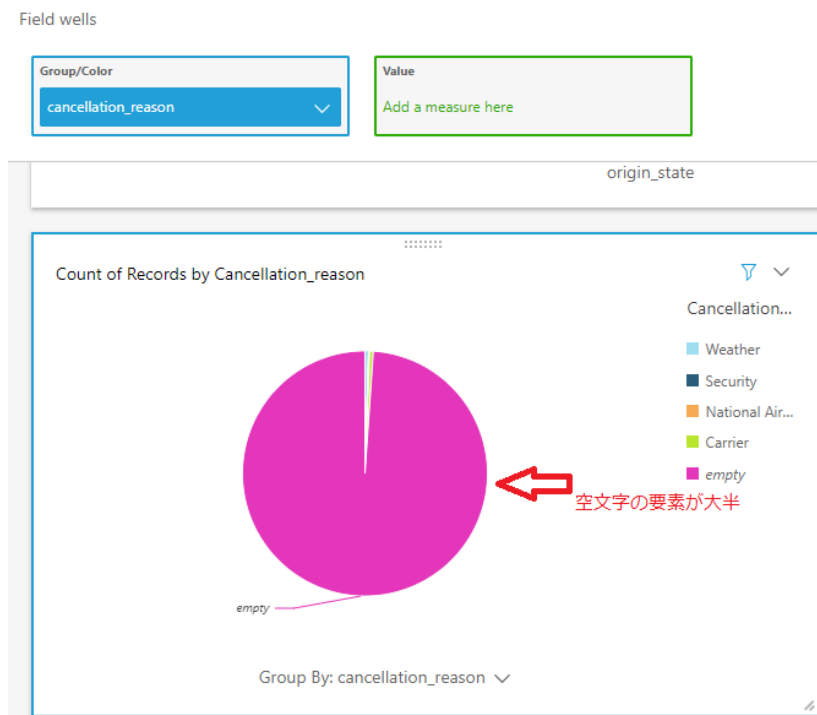


VISUAL 4: VERMONT のキャンセル理由（円グラフ）

キャンセル率が高い Vermont 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” +Add” ->” Add Visual” で新しい Visual を追加し、Visual type は” Pie chart”（上から三段目、左から 4 つ目）を選択します。Field wells の Group/Color に cancellation_reason（キャンセル理由）を D&D します。

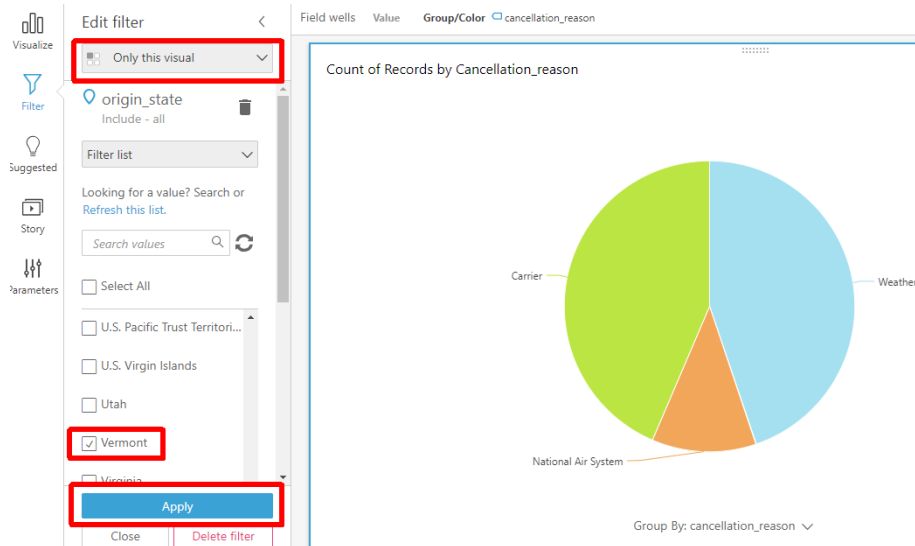
作成直後のグラフでは以下のように空文字「」が大半を占めています。これは cancellation_reason の列は、キャンセルされない時は空文字が入るためです。今回はキャンセルされた時の理由なので、この空文字をグラフから削除します。



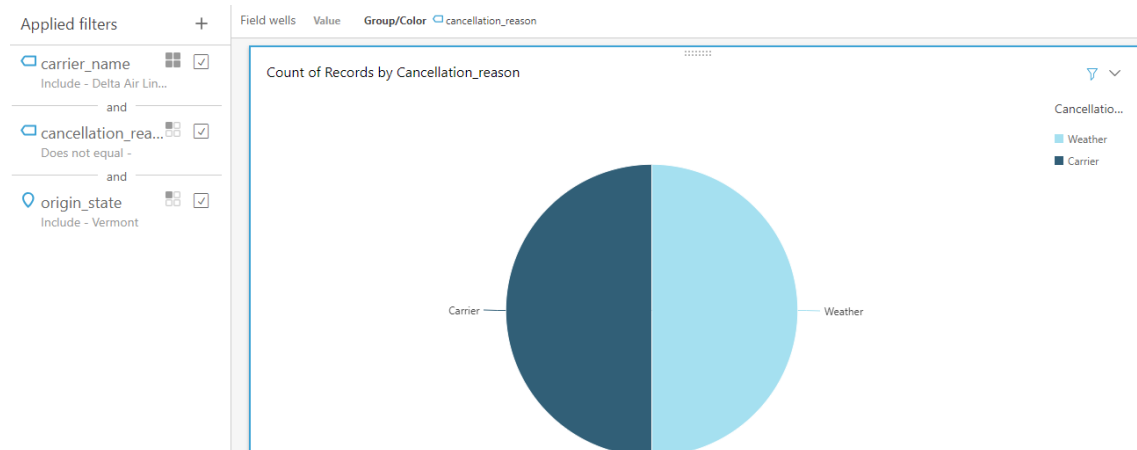
円グラフの空文字部分をクリックして Exclude（除外）を選択してください。Exclude を選択するとその値を除外する条件のフィルタが設定されます。

また、今は全地域のデータが表示されているため 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” Filter” を選択し、” +” で Filter を追加します（※最上部の” +” ではなく、Applied Filters という文字の横にある” +” です）。

Filter する Field として origin_state を選択し、Vermont だけ含めるように設定します。上部で” Only this visual” になっている事を確認し、Apply をクリックします。



結果として以下のような Visual が作成されます。Vermont でのキャンセル理由は Weather（天候）と Carrier（航空会社）がほぼ半分ずつを占めることが分かります。



VISUAL 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” +Add” ->” Add Visual” で新しい Visual を追加し、Visual type は” Pivot table”（上から 2 段目、左から 5 つ目）を選択します。Field wells の Rows に origin_state を D&D し、origin_airport_name をその下に ADD します。また、Columns には cancellation_reason を D&D します。さらに、Filter として cancellation_reson で空文字以外を選択して Apply します。

☐ Select All

☐

☒ Carrier

☒ National Air System

☒ Security

☒ Weather

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ます。

Rows

origin_state
origin_airport_name

Columns

cancellation_reason

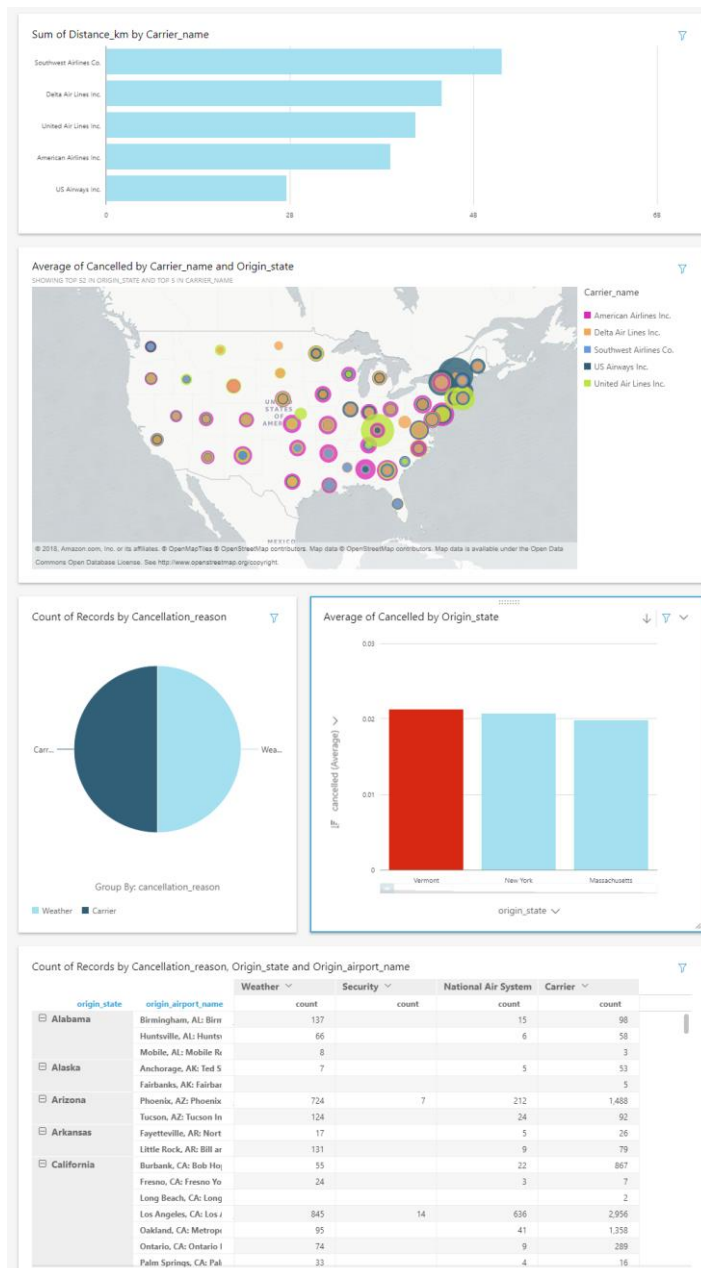
Values

Add measures here

		Weather	Security	National Air System	Carrier
origin_state	origin_airport_name	count	count	count	count
Alabama	Birmingham, AL: Birr	137		15	98
	Huntsville, AL: Hunts	66		6	58
	Mobile, AL: Mobile R	8			3
Alaska	Anchorage, AK: Ted S	7		5	53
	Fairbanks, AK: Fairbar				5
Arizona	Phoenix, AZ: Phoenix	724	7	212	1,488
	Tucson, AZ: Tucson In	124		24	92
Arkansas	Fayetteville, AR: Nort	17		5	26
	Little Rock, AR: Bill ar	131		9	79
California	Burbank, CA: Bob Ho	55		22	867
	Fresno, CA: Fresno Yo	24		3	7

VISUAL のサイズや位置の調整

各 Visual 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。見やすいように調整してください



これで本ハンズオンは完了です。

まとめと後片付け

QuickSight について

QuickSight 1 ユーザ分は無料で利用できますので、そのまま使っていただいても費用は発生しません。削除する場合は QuickSight の画面でリージョンを N. Virginia に変更した後

右上のユーザアイコン→Manage QuickSight→Account Settings を選択し、Unsubscribe 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）

Athena について

Athena ではクエリ単位の費用のため、クエリを実行しない限り費用は発生しません。作成した表定義は、DROP TABLE 文で、データベース定義は DROP DATABASE 文で削除可能です。

例)

```
drop table flightdata1;
```

```
drop table flightdata2;
```

```
drop database qshandson;
```

S3 について

Athena で実行したクエリ結果は S3 に保存されています。S3 には 5GB 分の無料枠があるため、それを超えない範囲では費用が発生しません。不要な場合は S3 の管理コンソールからバケットやオブジェクトを削除することが可能です。

補足情報・参考資料

■Amazon QuickSight

- ホームページ <https://aws.amazon.com/jp/quicksight/>
- ドキュメント <https://aws.amazon.com/jp/documentation/quicksight/>
- サービス説明資料 <https://www.slideshare.net/AmazonWebServicesJapan/20180228-aws-black-belt-online-seminar-quicksight-89889593>

■Amazon Athena

- ホームページ <https://aws.amazon.com/jp/athena/>
- 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>

■Amazon S3

- ホームページ <https://aws.amazon.com/jp/s3/>
- 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

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/>