- Qwik start_console:
  -le code
  bq query --use_legacy_sql=false
  '
  SELECT
  weight_pounds, state, year, gestation_weeks
  FROM
  `bigquery-public-data.samples.natality`
  ORDER BY weight_pounds DESC LIMIT 10;
  '

bq mk babynames

https://labshell-service-mvrcyiow4a-uc.a.run.app

gsutil cp gs://spls/gsp072/baby-names.zip .

unzip baby-names.zip

bq load --autodetect --source_format=CSV babynames.names_2014 gs://spls/gsp072/baby-names/yob2014.txt name:string,gender:string,count:integer

Task6: Query a custom dataset

```
#standardSQL
SELECT
name, count
FROM
babynames.names_2014
WHERE
gender = 'M'
ORDER BY count DESC LIMIT 5;
```

Task 7. Test your understanding

Below is a true/false question to reinforce your understanding of this lab's concepts. Answer it to the best of your abilities.

BigQuery is a fully-managed enterprise data warehouse that enables super-fast SQL queries.

_check_True

BigQuery : Qwik Start – Ligne de commande:

-tache 1: bq show bigquery-public-data:samples.shakespeare
-tache 2: bq help
-tache 3:
'SELECT
word,
SUM(word_count) AS count
FROM
`bigquery-public-data`.samples.shakespeare
WHERE
word LIKE "%raisin%"
GROUP BY
word'

```
+--------------+-------+
| word | count |
+--------------+-------+
| praising | 8 |
| Praising | 4 |
| raising | 5 |
| dispraising | 2 |
| dispraisingly | 1 |
| raisins | 1 |
+--------------+-------+
```

'SELECT
word
FROM
`bigquery-public-data`.samples.shakespeare
WHERE
word = "huzzah"'
tache 4:
bq ls bigquery-public-data
bq mk babynames
bq ls
curl -LO http://www.ssa.gov/OACT/babynames/names.zip
ls
unzip names.zip
ls
bq load babynames.names2010 yob2010.txt name:string,gender:string,count:integer
bq load babynames.names2010 yob2010.txt name:string,gender:string,count:integer

bq ls babynames: sorti
names2010 TABLE
bq show babynames.names2010
-tache 5:
bq query "SELECT name,count FROM babynames.names2010 WHERE gender = 'F' ORDER BY count DESC LIMIT 5"
tache 6:
bq rm -r babynames

rm: remove dataset 'qwiklabs-gcp-04-363e041944bc:babynames'? (y/N) y
bq ls babynames BigQuery error in ls operation: Not found: Dataset qwiklabs-gcp-04-363e041944bc:babynames

Données météorologiques dans BigQuery - Weather Data in BigQuery

compte

projectid : qwiklabs-gcp-03-e25d805a61b3

Tâche 1. Explorer les données météorologiques

Ouvrez la console BigQuery

1. Dans Google Cloud Console, sélectionnez **Menu de navigation** > **BigQuery** .

La boîte de message **Bienvenue dans BigQuery dans Cloud Console** s'ouvre. Cette boîte de message fournit un lien vers le guide de démarrage rapide et les notes de version.

2. Cliquez sur **Terminé** .

La console BigQuery s'ouvre.

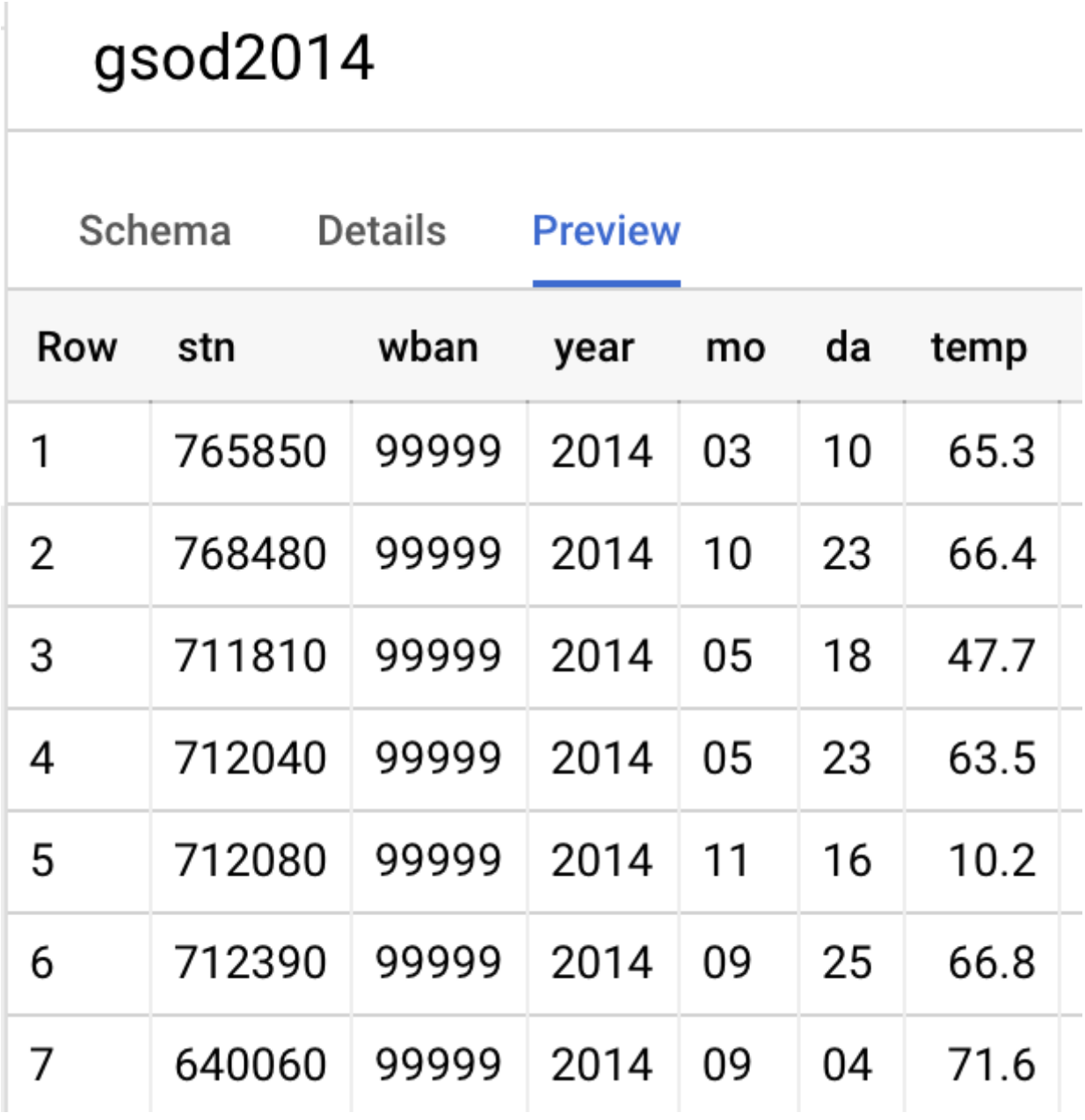
3. Dans le volet **Explorateur** , cliquez sur **+ AJOUTER** .

La fenêtre Ajouter des données s'ouvre.

4. Cliquez sur **Star un projet par son nom** sous Sources supplémentaires.
5. Entrez `bigquery-public-data` et cliquez sur **ÉTOILE** .

Dans la console BigQuery, vous voyez deux projets dans le volet Explorateur, l'un nommé **ID de votre projet** d'atelier et l'autre nommé **bigquery-public-data** .

6. Dans le volet **Explorateur** de la console BigQuery, développez l'ensemble de données **bigquery-public-data** . Dans le champ **Type de recherche,** recherchez **noaa_gsod** et sélectionnez la table **gsod2014** .
7. Dans la fenêtre Table (gsod2014), cliquez sur l' onglet **Aperçu** .



gsod2014

Schema Details Preview

Row	stn	wban	year	mo	da	temp
1	765850	99999	2014	03	10	65.3
2	768480	99999	2014	10	23	66.4
3	711810	99999	2014	05	18	47.7
4	712040	99999	2014	05	23	63.5
5	712080	99999	2014	11	16	10.2
6	712390	99999	2014	09	25	66.8
7	640060	99999	2014	09	04	71.6

8. Examinez les colonnes et certaines valeurs de données.
9. Cliquez sur **Requête > Dans un nouvel onglet** puis collez la requête suivante :

```
SELECT
  -- Create a timestamp from the date components.
  stn,
  TIMESTAMP(CONCAT(year,"-",mo,"-",da)) AS timestamp,
  -- Replace numerical null values with actual null
  AVG(IF (temp=9999.9,
```

```
      null,
      temp)) AS temperature,
  AVG(IF (wdsp="999.9",
      null,
      CAST(wdsp AS Float64))) AS wind_speed,
  AVG(IF (prcp=99.99,
      0,
      prcp)) AS precipitation
FROM
  `bigquery-public-data.noaa_gsod.gsod20*`
WHERE
  CAST(YEAR AS INT64) > 2010
  AND CAST(MO AS INT64) = 6
  AND CAST(DA AS INT64) = 12
  AND (stn="725030" OR  -- La Guardia
    stn="744860")    -- JFK
GROUP BY
  stn,
  timestamp
ORDER BY
  timestamp DESC,
  stn ASC
```

10. Cliquez sur **EXÉCUTER** . Regardez le résultat et essayez de déterminer ce que fait cette requête.

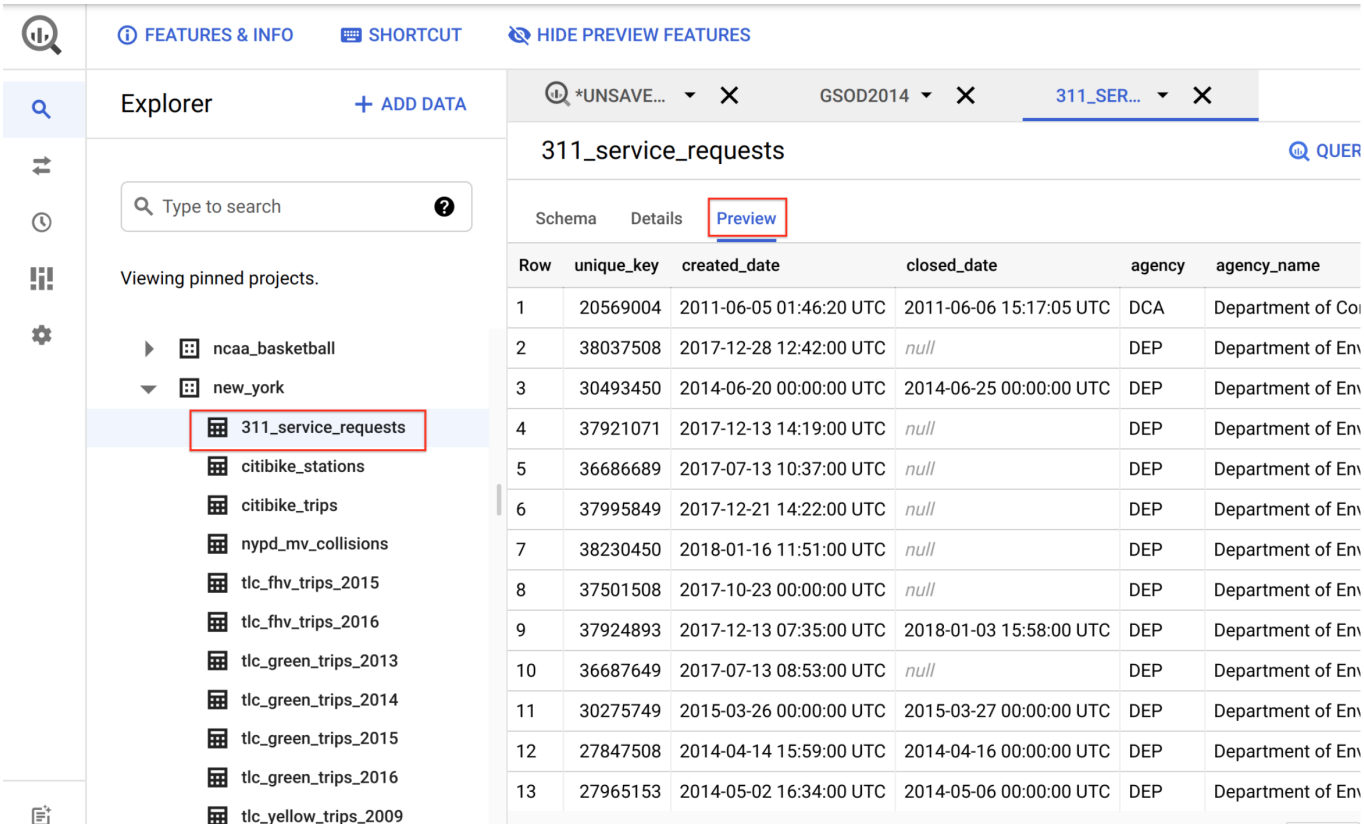
Cliquez sur **Vérifier mes progrès** ci-dessous pour vérifier que vous êtes sur la bonne voie dans cet atelier.

Explorer les données météorologiques

Vérifier mes progrès

Tâche 2. Explorer les données sur les plaintes des citoyens de New York

1. Dans le volet Explorateur de la console BigQuery, sélectionnez le projet **bigquery-public-data** nouvellement ajouté , dans le champ **Type de recherche,** recherchez l' ensemble de données **new_york_311** , puis sélectionnez la table **311_service_requests** .
2. Cliquez ensuite sur l' onglet **Aperçu** . Votre console devrait ressembler à ce qui suit :



3. Examinez les colonnes et certaines valeurs de données.
4. Si l'éditeur a été fermé, cliquez sur l'icône " **+** " (Créer une requête SQL).
5. Collez ce qui suit dans la requête **EDITOR** :

```
SELECT
  EXTRACT(YEAR
  FROM
    created_date) AS year,
  complaint_type,
  COUNT(1) AS num_complaints
FROM
  `bigquery-public-data.new_york.311_service_requests`
GROUP BY
  year,
  complaint_type
ORDER BY
  num_complaints DESC
```

6. Cliquez sur **EXÉCUTER** .
7. Examinez les résultats pour déterminer quelles sont les plaintes les plus courantes. Vous tenterez de déterminer si ces plaintes sont liées aux conditions météorologiques dans une partie ultérieure de cet atelier.

Cliquez sur **Vérifier mes progrès** ci-dessous pour vérifier que vous êtes sur la bonne voie dans cet atelier.

Explorez les données sur les plaintes des citoyens de New York

Vérifier mes progrès

Tâche 3. Sauvegarde d'un nouveau tableau de données météorologiques

1. Dans le volet Explorateur de la console BigQuery, cliquez sur les **trois points** à côté de votre `Project ID`, puis sélectionnez **Créer un ensemble de données** .
2. Dans la boîte de dialogue Créer un ensemble de données, définissez l' **ID de l'ensemble de données** sur `demos` et laissez les autres options à leurs valeurs par défaut.
3. Cliquez sur **Créer un ensemble de données** . Votre projet dispose désormais d'un ensemble de données nommé `demos`.
4. Cliquez sur l'icône **« + »** (Créer une requête SQL), puis exécutez la requête suivante :

```
SELECT
  -- Create a timestamp from the date components.
  timestamp(concat(year,"-",mo,"-",da)) as timestamp,
  -- Replace numerical null values with actual nulls
  AVG(IF (temp=9999.9, null, temp)) AS temperature,
  AVG(IF (visib=999.9, null, visib)) AS visibility,
  AVG(IF (wdsp="999.9", null, CAST(wdsp AS Float64))) AS wind_speed,
  AVG(IF (gust=999.9, null, gust)) AS wind_gust,
  AVG(IF (prcp=99.99, null, prcp)) AS precipitation,
  AVG(IF (sndp=999.9, null, sndp)) AS snow_depth
FROM
  `bigquery-public-data.noaa_gsod.gsod20*`
WHERE
  CAST(YEAR AS INT64) > 2008
  AND (stn="725030" OR  -- La Guardia
       stn="744860")    -- JFK
GROUP BY timestamp
```

5. Dans la section ÉDITEUR de requêtes, cliquez sur **Plus** > **Paramètres de requête** .
6. Dans la boîte de dialogue Paramètres de requête, définissez les champs suivants. Laissez tous les autres à leur valeur par défaut.

`Destination:` sélectionnez **Définir une table de destination pour les résultats de la requête**

`Dataset:` Tapez **des démos** et sélectionnez votre ensemble de données.

`Table Id` : Tapez **nyc_weather**

`Results size` : cochez **Autoriser les résultats volumineux (pas de limite de taille)**

7. Cliquez sur **ENREGISTRER**
8. Cliquez sur **EXÉCUTER** .

Les résultats sont maintenant enregistrés dans l'ensemble de données que vous avez créé (démos).

9. Revenez à **Plus > Paramètres de requête** et, dans l'option Destination, sélectionnez **Enregistrer les résultats de la requête dans une table temporaire** . Cela supprime l'ensemble de données de démonstration en tant que destination pour les requêtes futures.
10. Cliquez sur **ENREGISTRER** pour fermer les paramètres de requête.

Cliquez sur **Vérifier mes progrès** ci-dessous pour vérifier que vous êtes sur la bonne voie dans cet atelier.

Sauvegarde d'un nouveau tableau de données météo

Vérifier mes progrès

Tâche 4. Trouver une corrélation entre les ensembles de données

Une forte corrélation, mesurée par le `CORR function`, indique une relation étroite et cohérente entre deux variables. À mesure que la valeur d'une variable augmente, la valeur de l'autre variable a également tendance à augmenter (corrélation positive) ou à diminuer (corrélation négative) de manière prévisible. Une forte corrélation est généralement considérée comme une valeur supérieure ou égale à 0,7, en termes absolus. Cela signifie que les changements dans une variable peuvent expliquer au moins 49 % des changements dans l'autre variable.

Ensuite, vous comparerez le nombre de plaintes reçues et la température quotidienne à l'aide de la fonction CORR .

1. Créez une requête SQL, **"+"** et exécutez la requête suivante :

```
SELECT
  descriptor,
  sum(complaint_count) as total_complaint_count,
  count(temperature) as data_count,
```

```
  ROUND(corr(temperature, avg_count),3) AS corr_count,
  ROUND(corr(temperature, avg_pct_count),3) AS corr_pct
From (
SELECT
  avg(pct_count) as avg_pct_count,
  avg(day_count) as avg_count,
  sum(day_count) as complaint_count,
  descriptor,
  temperature
FROM (
  SELECT
    DATE(timestamp) AS date,
    temperature
  FROM
    demos.nyc_weather) a
  JOIN (
  SELECT x.date, descriptor, day_count, day_count / all_calls_count as
pct_count
  FROM
    (SELECT
      DATE(created_date) AS date,
      concat(complaint_type, ": ", descriptor) as descriptor,
      COUNT(*) AS day_count
    FROM
      `bigquery-public-data.new_york.311_service_requests`
    GROUP BY
      date,
      descriptor)x
    JOIN (
      SELECT
        DATE(timestamp) AS date,
        COUNT(*) AS all_calls_count
      FROM `demos.nyc_weather`
      GROUP BY date
    )y
  ON x.date=y.date
)b
ON
  a.date = b.date
GROUP BY
  descriptor,
  temperature
)
GROUP BY descriptor
HAVING
  total_complaint_count > 5000 AND
```

```
  ABS(corr_pct) > 0.5 AND
  data_count > 5
ORDER BY
  ABS(corr_pct) DESC
```

Les résultats indiquent que les plaintes liées au chauffage sont négativement corrélées à la température (c'est-à-dire plus d'appels de chauffage les jours froids) et que les appels concernant des arbres morts sont positivement corrélés à la température (c'est-à-dire plus d'appels les jours chauds).

Comparez ensuite le nombre de plaintes et la vitesse du vent avec la fonction CORR.

2. Cliquez sur l'icône **« + »** (Créer une requête SQL) et exécutez la requête suivante :

```
SELECT
  descriptor,
  sum(complaint_count) as total_complaint_count,
  count(wind_speed) as data_count,
  ROUND(corr(wind_speed, avg_count),3) AS corr_count,
  ROUND(corr(wind_speed, avg_pct_count),3) AS corr_pct
From (
SELECT
  avg(pct_count) as avg_pct_count,
  avg(day_count) as avg_count,
  sum(day_count) as complaint_count,
  descriptor,
  wind_speed
FROM (
  SELECT
    DATE(timestamp) AS date,
    wind_speed
  FROM
    demos.nyc_weather) a
  JOIN (
  SELECT x.date, descriptor, day_count, day_count / all_calls_count as
pct_count
  FROM
    (SELECT
      DATE(created_date) AS date,
      concat(complaint_type, ": ", descriptor) as descriptor,
      COUNT(*) AS day_count
    FROM
      `bigquery-public-data.new_york.311_service_requests`
```

```
    GROUP BY
      date,
      descriptor)x
    JOIN (
      SELECT
        DATE(timestamp) AS date,
        COUNT(*) AS all_calls_count
      FROM `demos.nyc_weather`
      GROUP BY date
    )y
  ON x.date=y.date
)b
ON
  a.date = b.date
GROUP BY
  descriptor,
  wind_speed
)
GROUP BY descriptor
HAVING
  total_complaint_count > 5000 AND
  ABS(corr_pct) > 0.5 AND
  data_count > 5
ORDER BY
  ABS(corr_pct) DESC
```

3. Notez que les colonnes Corr sont toutes deux négatives pour les plaintes liées au bruit. Avez-vous une hypothèse sur la raison pour laquelle les plaintes concernant le bruit diminuent les jours venteux ? Les coefficients sont-ils statistiquement suffisants ?

Cliquez sur **Vérifier mes progrès** ci-dessous pour vérifier que vous êtes sur la bonne voie dans cet atelier.

Toutes nos félicitations!

Dans cet atelier, vous avez pu interroger les données sans configurer de clusters, ni créer d'index, etc. Vous avez également pu mélanger les deux ensembles de données et corréler les résultats, ce qui vous a donné des informations intéressantes.