Didn't know where to start

https://cloud.google.com/datalab/ took me to a billing screen

went to documentation https://cloud.google.com/datalab/overview

navigated to https://datalab.cloud.google.com/

Explore, transform, analyze and visualize your data, using Google Cloud Platform.

Select Cloud Project: My First Project ▾

[Start Datalab] [Deploy Datalab] [Manage Datalab]

Cloud Datalab will be deployed as an AppEngine application module in the selected project. The Google Compute Engine and BigQuery APIs must be enabled for the project, and you must be authorized to use the project as an owner or editor.

The following additions will be made to the project:

- An AppEngine module named "datalab". This can be managed within the AppEngine section in the Developer Console. AppEngine charges will accrue.

- A Compute Engine network named "datalab" will be added if it doesn't exist.

- A branch named "datalab" will be added to the git repository associated with the project, and is browsable within the Source section in Developer Console to contain samples, docs and your notebooks.

Cloud Datalab is being installed into the 'thematic-lore-110201' project. Note that this deployment can take more than 10 minutes while the AppEngine application is provisioned and started.

Thanks for waiting.

```
|    |    |    |    |    |    |    |    |▓▓▓|    |
```

(took about 8 minutes)

From log:

```
20151018-01-16-46 startupscript:
20151018-01-16-46 startupscript: The following components will be updated:
20151018-01-16-46 startupscript:     ---------------------------------------------------------------
20151018-01-16-46 startupscript:     | BigQuery Command Line Tool                    |     2.0.18 | < 1 MB |
20151018-01-16-46 startupscript:     | Cloud DNS Admin Command Line Interface        | 2015.04.21 | < 1 MB |
20151018-01-16-46 startupscript:     | Cloud SDK Core Libraries                      | 2015.04.21 | 1.8 MB |
20151018-01-16-46 startupscript:     | Cloud SQL Admin Command Line Interface        | 2015.04.09 | < 1 MB |
20151018-01-16-46 startupscript:     | Cloud Storage Command Line Tool               |       4.11 | 2.0 MB |
20151018-01-16-46 startupscript:     | Compute Engine Command Line Interface         | 2015.04.21 | < 1 MB |
20151018-01-16-46 startupscript:     | Compute Engine Command Line Tool (deprecated) |     1.16.5 | < 1 MB |
20151018-01-16-46 startupscript:     ---------------------------------------------------------------
20151018-01-16-46 startupscript: The following components will be installed:
20151018-01-16-46 startupscript:     ---------------------------------------------------------------
20151018-01-16-46 startupscript:     | App Engine Command Line Interface (Preview) | 2015.04.21 |  < 1 MB |
20151018-01-16-46 startupscript:     | Developer Preview gcloud Commands           | 2015.04.21 |  < 1 MB |
20151018-01-16-46 startupscript:     | gcloud app Go Extensions (Linux, x86_64)    |     1.9.18 | 26.4 MB |
20151018-01-16-46 startupscript:     | gcloud app Java Extensions                  |     1.9.18 | 92.9 MB |
20151018-01-16-46 startupscript:     | gcloud app Python Extensions                |     1.9.18 |  6.8 MB |
20151018-01-16-46 startupscript:     ---------------------------------------------------------------
20151018-01-16-46 startupscript:
20151018-01-16-46 startupscript:
```

```
:ript:  21 files changed, 10271 insertions(+)
:ript:  create mode 100644 .gitignore
:ript:  create mode 100644 Hello World.ipynb
:ript:  create mode 100644 datalab/Readme.ipynb
:ript:  create mode 100644 datalab/intro/Introduction to Notebooks.ipynb
:ript:  create mode 100644 datalab/intro/Introduction to Python.ipynb
:ript:  create mode 100644 datalab/intro/Using Datalab - Accessing Cloud Data.ipynb
:ript:  create mode 100644 datalab/intro/Using Datalab - Managing Notebooks with Git.ipynb
:ript:  create mode 100644 datalab/samples/Anomaly Detection in HTTP Logs.ipynb
:ript:  create mode 100644 datalab/samples/Conversion Analysis with Google Analytics Data.ipynb
:ript:  create mode 100644 datalab/samples/Exploring Genomics Data.ipynb
:ript:  create mode 100644 datalab/samples/Programming Language Correlation.ipynb
:ript:  create mode 100644 datalab/tutorials/BigQuery/BigQuery APIs.ipynb
:ript:  create mode 100644 datalab/tutorials/BigQuery/BigQuery Commands.ipynb
:ript:  create mode 100644 datalab/tutorials/BigQuery/Hello BigQuery.ipynb
:ript:  create mode 100644 datalab/tutorials/BigQuery/Importing and Exporting Data.ipynb
:ript:  create mode 100644 datalab/tutorials/BigQuery/SQL Parameters.ipynb
:ript:  create mode 100644 datalab/tutorials/BigQuery/SQL Query Composition.ipynb
:ript:  create mode 100644 datalab/tutorials/BigQuery/SQL and Pandas DataFrames.ipynb
:ript:  create mode 100644 datalab/tutorials/Data/Interactive Charts with Google Charting APIs.ipynb
:ript:  create mode 100644 datalab/tutorials/Storage/Storage APIs.ipynb
:ript:  create mode 100644 datalab/tutorials/Storage/Storage Commands.ipynb
:ript: remote: Storing objects: 76% (23/30)#033[K#015remote: Storing objects: 90% (27/30)#033[K#015remote: Storing objects:
K#015remote: Storing objects: 100% (30/30), done.#033[K
:ript: remote: Processing commits: 100% (1/1)#033[K#015remote: Processing commits: 100% (1/1), done.#033[K
:ript: To https://source.developers.google.com/p/thematic-lore-110201/
```

Explore, transform, analyze and visualize your data, using Google Cloud Platform.

Start using Datalab   **Manage Datalab Application**

Cloud Datalab was successfully deployed. Details can be found in the deployment log.

Feedback | Privacy Policy | Terms of Service

Hello World IPython notebook

Datalab guide: https://main-dot-datalab-dot-thematic-lore-110201.appspot.com/notebooks/datalab/Readme.ipynb

Example notebooks included

Active Notebooks

📗 Hello World.ipynb		Shutdown
📗 datalab/Readme.ipynb		Shutdown
📗 datalab/intro/Using Datalab - Accessing Cloud Data.ipynb		Shutdown
📗 datalab/tutorials/Data/Interactive Charts with Google Charting APIs.ipynb		Shutdown
📗 datalab/tutorials/BigQuery/Importing and Exporting Data.ipynb		Shutdown

🧪 **Google Cloud Datalab**   **Conversion Analysis with Google Analytics Data**   (autosaved)

📓 Notebook ▾   ⊕ Add Code ⊕ Add Markdown ⊟ Delete ⋀ Move Up ⋁ Move Down ▶ Run ▾ ⊟ Clear ▾ ⟳ Reset Session

Conversion Analysis with Google Analytics Data

This sample notebook demonstrates working with Google Analytics page views and session data exported to BigQuery.

Google Analytics offers BigQuery export as part of its premium offering. If you're a premium user, you already have the ability to export any of your analytics views to a BigQuery dataset you own. If you're not, you can use the Analytics API to retrieve and import the data used to generate the default Analytics dashboards.

The sample data used in this notebook is shares the same schema as the Google Analytics BigQuery export, but is from a sample account, and hence available publicly. However it is also tiny in terms of size. This notebook demonstrates one possible custom analytics scenario, rather than the actual data.

Related Links:

- BigQuery
- Google Analytics
- Google Charting API for data visualization

```
import gcp.bigquery as bq
```

Understanding the Hits Data

Its helpful to inspect the schema and a sample of the data we're working with.

```
%%bigquery schema --table "google.com:analytics-bigquery:LondonCycleHelmet.ga_sessions_20130910"
```

name	type	mode	description
visitorId	INTEGER	NULLABLE	
visitNumber	INTEGER	NULLABLE	
visitId	INTEGER	NULLABLE	

I want to do this with my own Google Analytics data… guess I need BigQuery?

Guess I need a dataset…. How to import? Tiny dropdown next to "My First Project"
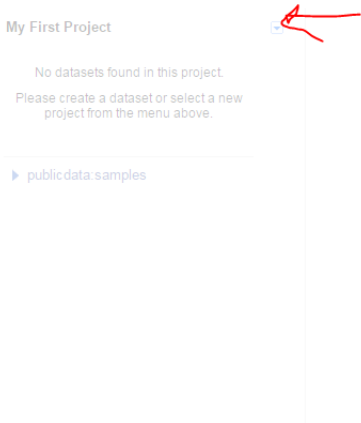
http://www.lunametrics.com/blog/2014/01/27/google-analytics-bigquery-whys-hows/

http://www.lunametrics.com/blog/2013/05/16/google-analytics-premium-bigquery/

… only allowed for Google Analytics Premium

https://support.google.com/analytics/answer/3437618?hl=en

Decided to go with Natality data from public data set built in

public data:samples [+] [-]

⊞ github_nested

⊞ github_timeline

⊞ gsod

⊞ **natality**

⊞ shakespeare

⊞ trigrams

⊞ wikipedia

Born 1980-1984 [?]   [Query Editor] [UDF Editor] [×]

```
1 SELECT year, month, day, wday, state, is_male, gestation_weeks, mother_age, father_age, apgar_1min, apgar_5min, weight_pounds FROM [publicdata:samples.natality]
2 where year between 1980 and 1984
3
```
10201

Destination Table: thematic-lore-110201:Natality_1980_1984.Births_1980_1984 [×]   Allow Large Results [×]

RUN QUERY   Save Query   Save View   Format Query   Show Options   Query complete (19.7s elapsed, 9.30 GB processed)   ✅

Table Details: natality   [Schema] [Details] [Query Table]

Schema

source_year	INTEGER	REQUIRED	Four-digit year of the birth. Example: 1975.
year	INTEGER	NULLABLE	Four-digit year of the birth. Example: 1975.
month	INTEGER	NULLABLE	Month index of the date of birth, where 1=January.
day	INTEGER	NULLABLE	Day of birth, starting from 1.
wday	INTEGER	NULLABLE	Day of the week, where 1 is Sunday and 7 is Saturday.
state	STRING	NULLABLE	The two character postal code for the state. Entries after 2004 do not include this value.
is_male	BOOLEAN	REQUIRED	TRUE if the child is male, FALSE if female.
child_race	INTEGER	NULLABLE	The race of the child. One of the following numbers: 1 - White 2 - Black 3 - American Indian 4 - Chinese
weight_pounds	FLOAT	NULLABLE	Weight of the child, in pounds.

Query Failed

Error: Response too large to return. Consider setting allowLargeResults to true in your job configuration. For more details, see https://cloud.google.com/bigquery/querying-data#largequeryresults

Job ID: thematic-lore-110201:job_34TVvItVB5pbFMxcx_JvNuyLivg

Destination Table		Select Table	thematic-lore-110201:Natality_1980_1984.Births_1980_1984 ✕
Write Preference		● Write if empty ○ Append to table ○ Overwrite table	
Results Size		☑ Allow Large Results ?	
Results Schema		☑ Flatten Results ?	
Query Caching		☐ Use Cached Results ?	
Query Priority		● Interactive ○ Batch ?	
UDF Source URIs		Edit ?	

RUN QUERY   Save Query   Save View   Format Query   Hide Options   Query complete (19.7s elapsed, 9.30 GB processed)

Query complete (19.7s elapsed, 9.30 GB processed)

Table Details: Births_1980_1984

Schema

year	INTEGER	NULLABLE	Describe this field...
month	INTEGER	NULLABLE	Describe this field...
day	INTEGER	NULLABLE	Describe this field...
wday	INTEGER	NULLABLE	Describe this field...
state	STRING	NULLABLE	Describe this field...
is_male	BOOLEAN	NULLABLE	Describe this field...
gestation_weeks	INTEGER	NULLABLE	Describe this field...
mother_age	INTEGER	NULLABLE	Describe this field...
father_age	INTEGER	NULLABLE	Describe this field...
apgar_1min	INTEGER	NULLABLE	Describe this field...
apgar_5min	INTEGER	NULLABLE	Describe this field...
weight_pounds	FLOAT	NULLABLE	Describe this field...

Table ID	thematic-lore-110201:Natality_1980_1984.Births_1980_1984
Table Size	1.20 GB
Number of Rows	16,704,922
Creation Time	Oct 17, 2015, 9:50:39 PM
Last Modified	Oct 17, 2015, 9:50:39 PM
Data Location	US

Preview

Table	JSON

Row	year	month	day	wday	state	is_male	gestation_weeks	mother_age	father_age	apgar_1min	apgar_5min	weight_pounds
1	1980	1	5	null	AL	true	40	35	36	8	9	7.3744626639
2	1980	1	16	null	AL	true	38	25	25	6	8	6.6248909731
3	1980	1	16	null	AL	true	45	25	21	7	8	8.000575487979999
4	1980	1	18	null	AL	false	41	21	22	9	9	7.31273323054
5	1980	1	21	null	AL	true	99	19	25	8	10	4.62529825676

First < Prev   Rows 1 - 5 of 16704922   Next > Last

Seems each bigquery query needs to be in its own ipython cell or won't run

Had some problems updating… would timeout and had to stop, close, and reopen everything.

Exported to HTML, had to replace Google charts w/static images.

Final Test Notebooks:

http://www.becomingadatascientist.com/wp-content/uploads/2015/10/datalab/First%20Datalab%20Project%20-%20Natality%201980-1984%20Pandas.html

http://www.becomingadatascientist.com/wp-content/uploads/2015/10/datalab/First%20Datalab%20Project%20-%20Natality%201980-1984%20Google%20Charts.html