Module 14

Advanced Insights

In this module we will:

- Introducing Cloud Datalab
- Cloud Datalab Notebooks and Cells
- Benefits of Cloud Datalab

Google Cloud

Although the majority of this course focused so far on using the BigQuery web UI for data analysis, I want to introduce you to another powerful tool in the toolkit of every data scientist which is the online collaborative notebook.

For those of you who have heard of Jupyter or iPython notebooks before a lot of this content will look familiar because Cloud Datalab is an open source free GCP tool that shares the same roots.

Let's discuss a little more about what it is and why you should add Cloud Datalab as the next tool to learn in your analysis toolkit.

Cloud Datalab Notebooks bring Data Science at Scale

Datalab is a tool that helps you:

1. **Explore data** with Python, SQL, R, and more
2. **Present, collaborate, and publish data insights** in a fast, simple, and cost-effective way.
3. Develop and **run machine learning models**

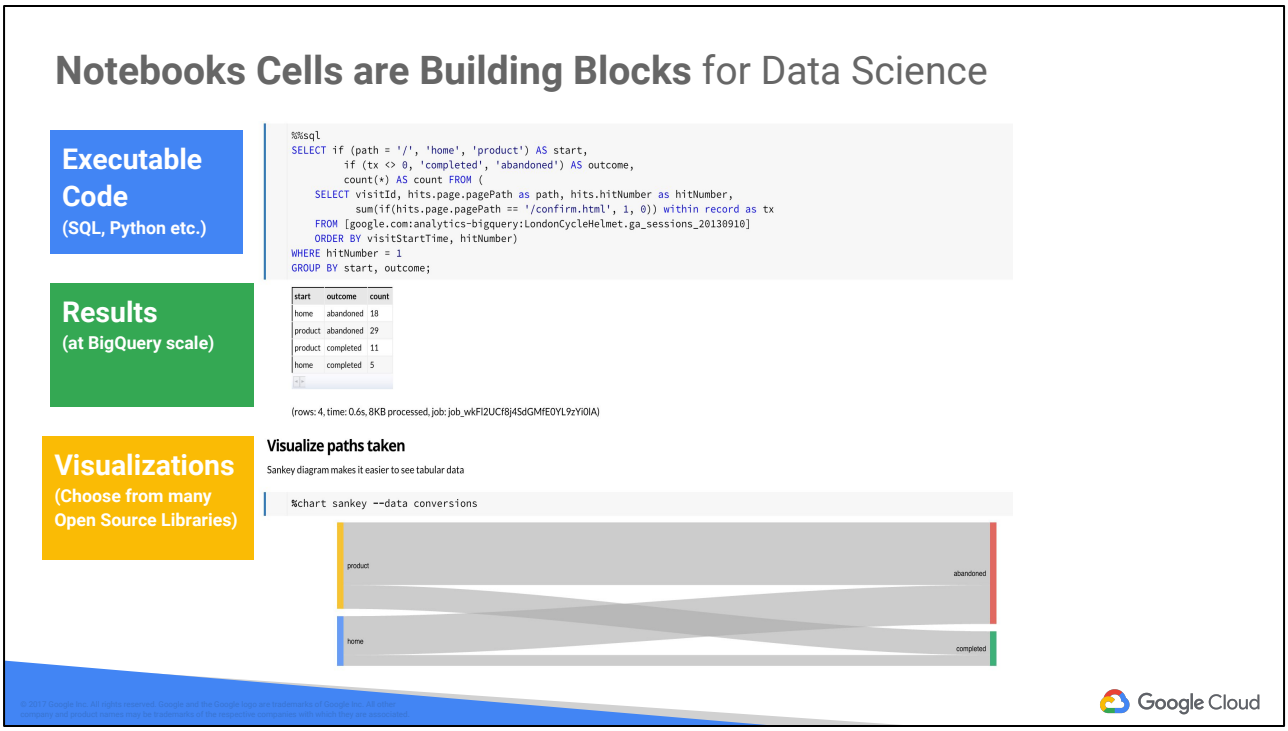
Google Cloud

Cloud Datalab is a tool for data processing and analysis.

Datalab provides a simple web interface for editing and running scripts using familiar languages, such as Python and SQL.

These scripts are deployed within notebook documents or “notebooks.”

Images from: https://cloud.google.com/datalab/

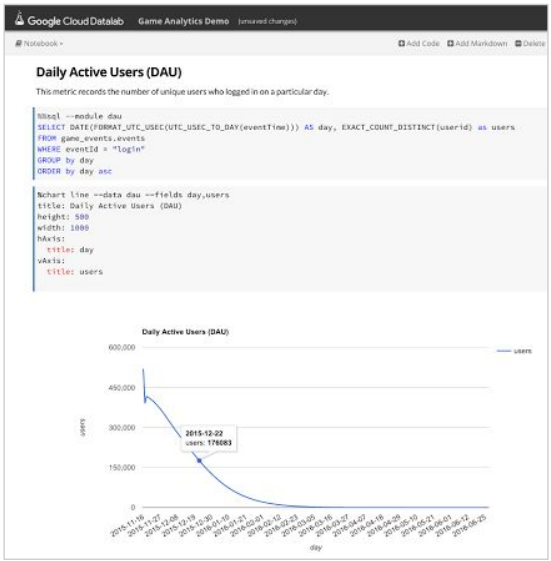


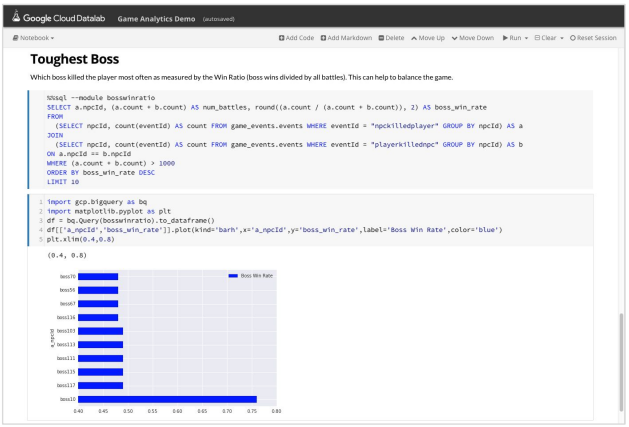
This screenshot is part of a notebook.

This notebook analyzes Google Analytics Premium data in BigQuery.

You can see it contains text, executable code, results and graphics.

Cloud Datalab - Interactive **Collaborative Data Notebooks**





Google Cloud

Optional demo using:
https://codelabs.developers.google.com/codelabs/mlimmersion-data-analysis/#1

https://cloud.google.com/bigquery/docs/visualize-datalab

https://cloud.google.com/blog/big-data/2016/08/new-version-of-cloud-datalab-jupyter-meets-tensorflow-cloud-meets-local-deployment

Image:
https://cloudplatform.googleblog.com/2016/01/build-a-mobile-gaming-analytics-platform.html

Notebooks are documents that contain:

- Computer code
- Human readable rich-text elements
- Equations
- Graphs
- Data visualizations

For humans, there are descriptions of data, analysis, code, and results, such as text, figures, and tables.

For computers, there are executable computer files that can be run to perform data analysis and samples of data.

Notebooks are easily shared with others.

Take a few minutes to explore a gallery of interesting notebooks and TensorFlow notebooks.

Benefits of Cloud Datalab for Data Scientists

- Collaborate with **documentation**, **code, and results in one place**.
- Copy **existing notebook examples** from the open source community (Datalab is based on the open-source project Jupyter)
- **Enables scalable analysis on** Google BigQuery, Cloud Machine Learning Engine, Google Compute Engine, **and** Google Cloud Storage

Datalab Quickstart

Google Cloud

Additional background: https://cloud.google.com/datalab/

- Notebooks are great for collaboration because they combine documentation, code, and results in natural sections called **cells**.
- Notebooks make sharing the results of data analysis across teams easy.
- Google Cloud Datalab's notebooks are based on the open-source project Jupyter, which has support in the open-source community. There are many open-source notebooks that you can use to get started on many projects at no cost.
- Datalab notebooks provide a single package that covers many common Google Cloud Platform customer scenarios. No need to assemble it yourself.

Summary: Use notebooks to collaborate and perform advanced analysis with your data

Explore, document, and visualize your data in one central notebook

Access additional tools like visualization libraries, machine learning APIs, python/R code and more

Present and collaborate with your peers through version controlled notebooks

Google Cloud

As we covered, Cloud Datalab notebooks are much like the BigQuery web UI with the addition that you can add multiple queries in notebooks cells and add markdown for your commentary.

Cloud Datalab is heavily used with machine learning APIs (like tensorflow). Generally Data Scientists will preprocess your data in BigQuery and then explore and build ML models with it in Cloud Datalab.

Lastly, notebooks are meant to be shared and presented with other Data Scientists. Your peers will be able to follow the same steps you took in your analysis and provide helpful feedback.

Image (notebook) cc0:
https://pixabay.com/en/coffee-pen-notebook-work-book-2306471/
Image (datalab) cc0: from: https://cloud.google.com/datalab/
Image (present) cc0: from: https://cloud.google.com/datalab/