Analysing and Visualising Data with Google Cloud Datalab

University of Michigan
Google Cloud Platform provides the foundation for your data analytics and machine learning workloads:

- BigQuery
- Cloud Dataflow
- Cloud Dataproc
- Cloud Datalab
- Cloud Pub/Sub
- Cloud Machine Learning
- Vision API
- Speech API
- Natural Language API
- Translation API
Introducing Data Science Notebooks
Business and financial analysts use spreadsheets as a tool for what-if analysis

- Enter data
- Manipulate data using formulas and macros
- Visualise data using tables, graphs and charts
- Share documents with co-workers
Data scientists also need a tool to capture and document work:

- Write code
- Author queries using SQL
- Visualize data using tables and charts
- Annotate with text
- Share results with co-workers
Jupyter notebook is a Web application to support data scientists

- Write code using over 40 languages
- Access big data
- Display data widgets
- Add explanations with formatted text
- Share results via git
Create **notebook documents** in a Web browser

- Each notebook has an `.ipynb` extension
- Documents can be **organised** into **folders**
- Documents can be **saved** and **shared**
Add cells to notebooks containing code and documentation using a Web browser

- Two standard cell types
  - Python for code
  - Markdown for documentation
- Each cell is interactive
  - Write code or documentation
  - Press `<Shift> + <Enter>` to run the cell
print 'Hello from a Notebook'  # some Python code

Hello from a Notebook

Heading
=======
Second Heading
-------------

These are Markdown bullets
- First
- Second
- Third

And a link [Gartner](http://www.gartner.com)
Other cell types are created using an escape sequence

- Each cell begins either `%command` or `%%command`
  - This is called **cell magic**
- Example built-in magics

<table>
<thead>
<tr>
<th>%%bash</th>
<th>Shell script</th>
</tr>
</thead>
<tbody>
<tr>
<td>%%javascript</td>
<td>JavaScript code</td>
</tr>
<tr>
<td>%%html</td>
<td>HTML documentation</td>
</tr>
</tbody>
</table>
```bash
ls -al

total 16
drwxr-xr-x  7 root root  4096 Mar 13 03:17 ..
drwxr-xr-x  2 root root  4096 Mar 13 03:20 .ipynb_checkpoints
-rw-r--r--  1 root root 2017 Mar 13 03:31 Jupyter Demo Notebook.ipynb
```

```javascript
function hello(name) { return 'Hello ' + name; }
var result = hello('Everyone');
element.text(result);

Hello Everyone
```

```html
<br>

**Name:** Doug
```
Cloud Datalab
Google Cloud Datalab provides Jupyter notebooks with integration for Google big data and machine learning

- **Datalab software is free**
- **Create Datalab machine**
  - **Standalone** with command-line
    
    datalab create my-machinename
  - Integrated with Dataproc
    - Employ initialization script
Datalab integrates with Google Big Query

- Two mechanisms
  - Import package to access from Python code
    ```python
    import datalab.bigquery as bq
    ```
  - Employ `%sql` cell magic
```python
import datalab.bigquery as bq

%%sql -d standard -m query
SELECT
    CAST(source_year AS string) AS year,
    COUNT(is_male) AS birth_count
FROM
    `publicdata.samples.natality`
GROUP BY year
ORDER BY year DESC
LIMIT 15

bq.Query(query).to_dataframe(dialect = 'standard').set_index('year').plot(kind='bar')
```

![Bar chart showing birth count](image-url)
Datalab integrates with Google Machine Learning

- Import packages including
  
<table>
<thead>
<tr>
<th>Package</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pandas</td>
<td>High performance data analysis tools</td>
</tr>
<tr>
<td>numpy</td>
<td>Numerical extensions</td>
</tr>
<tr>
<td>tensorflow</td>
<td>Google's machine learning library</td>
</tr>
</tbody>
</table>

- Sample notebook included with Datalab provides a real-world starting point
  
  notebooks/datalab/docs/samples/TensorFlow/Machine%20Learning%20with%20Financial%20Data.ipynb
Google Datalab Demo
Resources

Product Documentation
https://cloud.google.com/datalab/

Try it yourself in a Codelab!
https://goo.gl/Z2uv7F (setup)
https://goo.gl/WpHxFA (Datalab)