



**CS³
MESH⁴
EOSC**

Connecting European Data



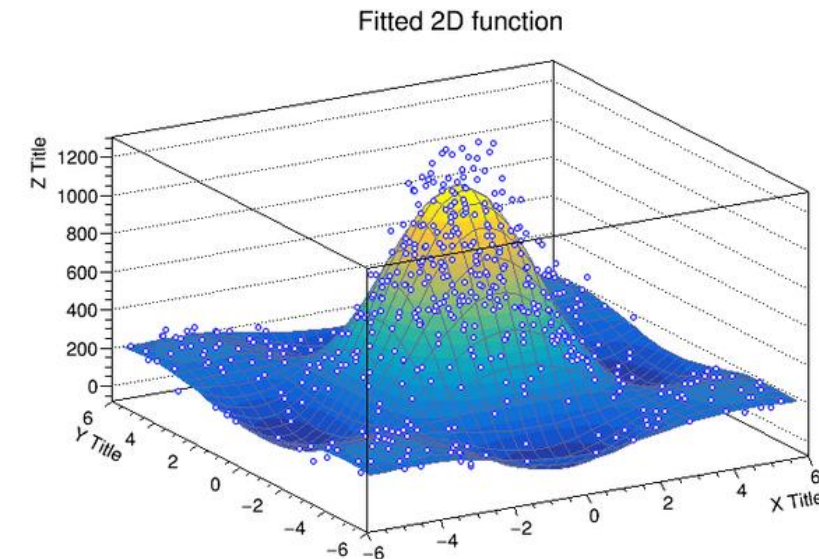
JupyterLab+ScienceMesh: Collaborative Data Science in sync-and-share environment.

Marcin Sieprawski
Head of Big Data Lab, Software Mind

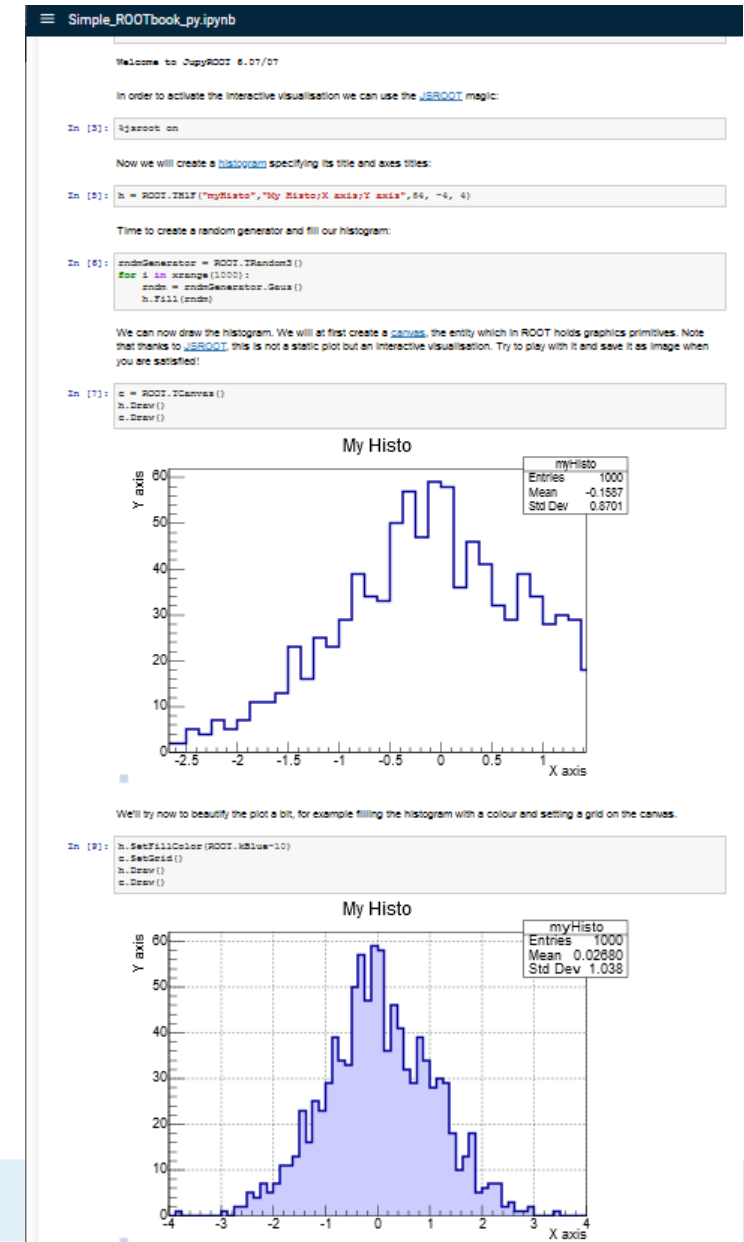


CS3MESH4EOSC has received funding from the European Union's Horizon 2020 Research and Innovation programme under **Grant Agreement No. 863353**.

- # Gartner - Critical Capabilities for Data Science and Machine Learning Platforms
 - # (13 March 2021)
 - # **By 2023, 30% of organizations will harness the collective intelligence** of their analytics communities, outperforming competitors that rely solely on centralized analytics or self-service.
 - # **By 2024, 70% of enterprises will use cloud and cloud-based AI infrastructure** to operationalize AI, thereby significantly alleviating concerns about integration and upscaling.
- # All scientific disciplines nowadays are data-driven
 - # Data analytics play an increasing role in all types of research
 - # Distributed data science environments => all fields of study
 - # A more effective collaboration between scientific institutions
- # Business: develop new products in all sectors
 - # Finance, IoT, SmartCities, energy and many others



- # A free, open-source, interactive web tool: a computational notebook
 - # combine software code, computational output, explanatory text and multimedia resources in a single document
 - # rapid uptake, an enthusiastic community of user-developers
 - # Python / R / more
 - # 10 milion public Jupyter Notebooks on GitHub (December 2020)
 - # 2.5M in September 2018, 200k in 2015
- # platform of choice for data scientists to build interactive applications and to tackle big data and AI problems
- # Replacing Business Intelligence tools





 Gallery

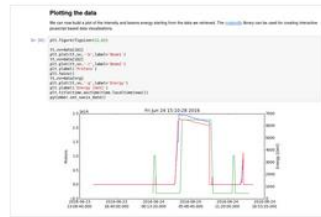
- Basic Examples
- ROOT Primer
- Accelerator Complex
- Beam Dynamics
- Machine Learning
- Apache Spark
- Outreach
- AWAKE

Accelerator Complex

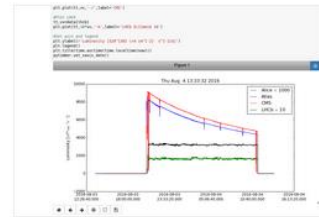
This gallery shows examples of machine studies relative to the CERN accelerators' complex.

Open in  SWAN

LHC Page1



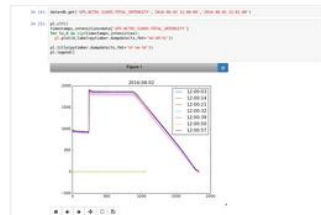
Experiments' Luminosities



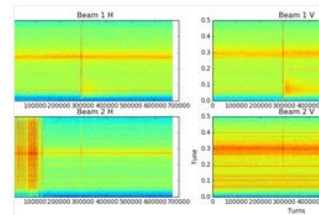
PyTimber Tutorial



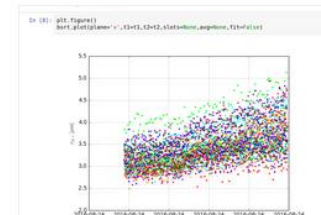
SPS Intensity

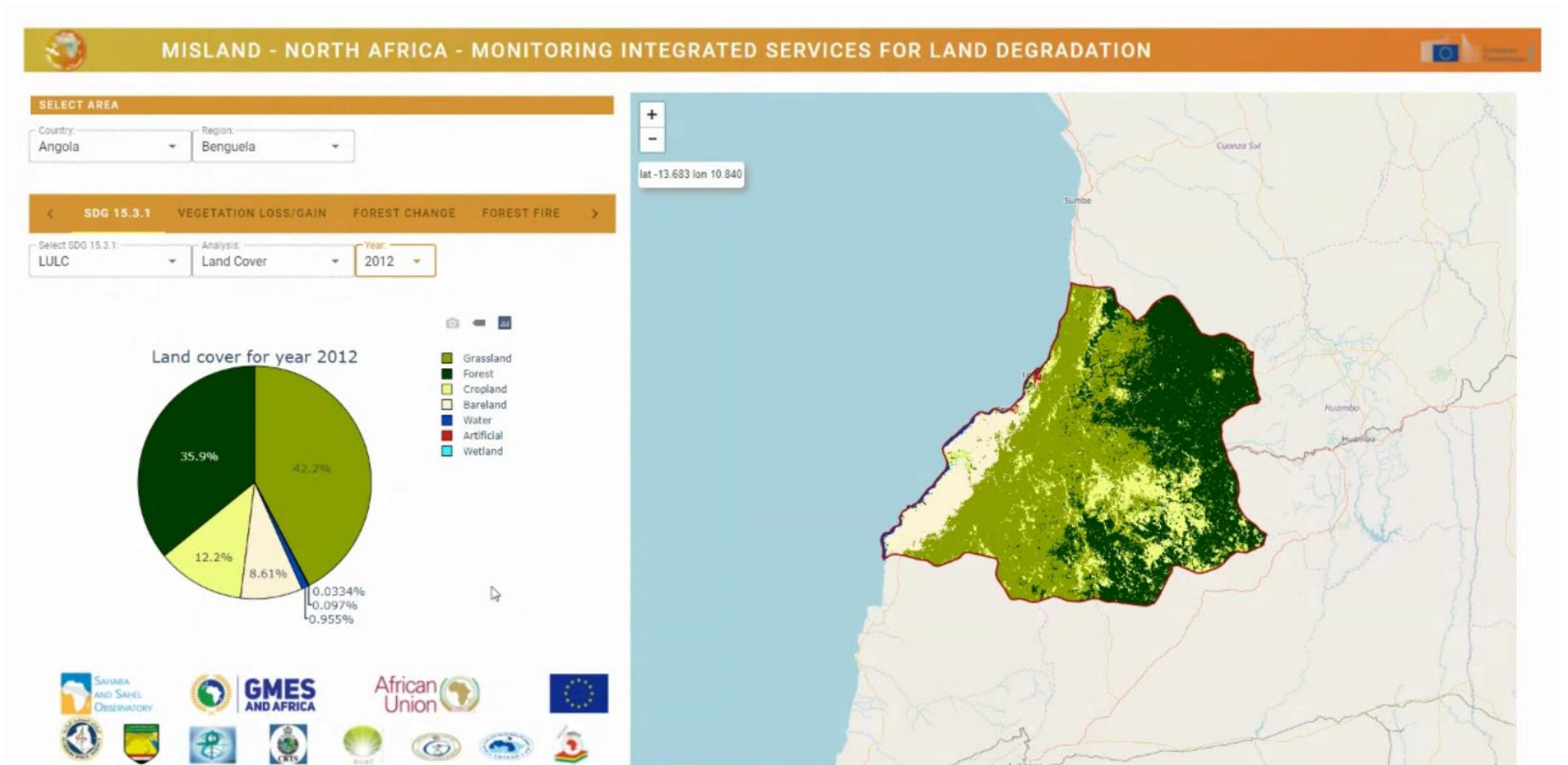


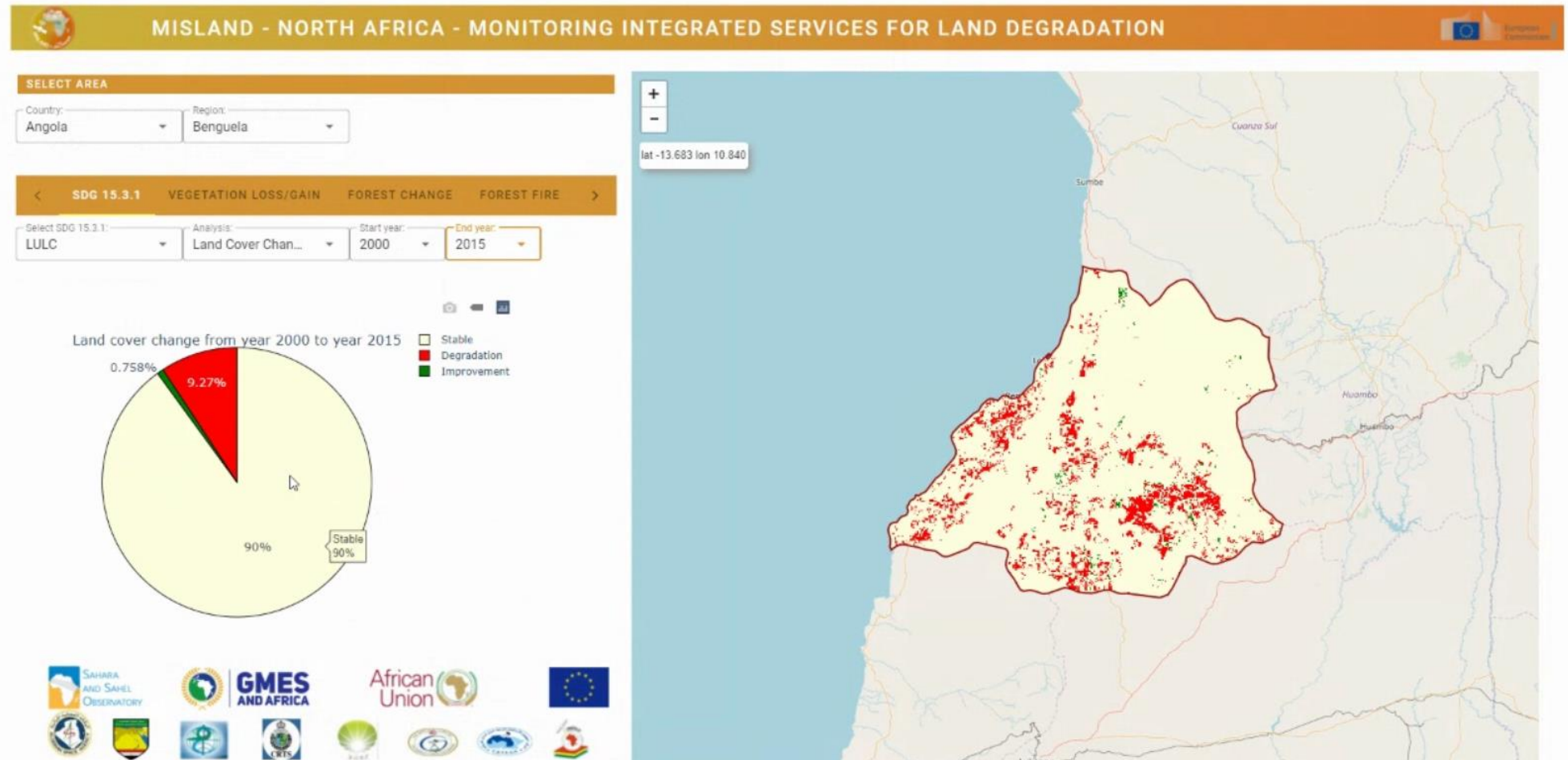
LHC BBQ Example



BSRT Example







```
File Edit View Run Kernel Tabs Settings Help
Launcher x MISLAND.ipynb x reports.py x
Code v
    rgb2hex(0,60,0),      # 7=Forest
    ]

LULC_LandCover_ColorMap = { 0: "black" }
for i in range(len(LULC_LandCover_colors)):
    LULC_LandCover_ColorMap[i+1] = LULC_LandCover_colors[i]

# Custom python function to be applied to ESA CCI Land Cover
def LandCover_MaskAndRecode(img):
    v, esa = img[0],img[1]

    # Mask-out what is outside the input polygon
    esa[v == 0] = 0

    # Recode
    esa[ ( esa>=10 ) * ( esa<=30 ) ] = 5    # Crop
    esa[ esa==40 ] = 6                    # Grass
    esa[ ( esa>=50 ) * ( esa<=100 ) ] = 7  # Forest
    esa[ ( esa>=110 ) * ( esa<=153 ) ] = 6  # Grass
    esa[ ( esa>=160 ) * ( esa<=180 ) ] = 4  # Wetland
    esa[ esa==190 ] = 3                   # Urban
    esa[ ( esa>=200 ) * ( esa<=202 ) ] = 2  # Bare
    esa[ esa==210 ] = 1                   # Water
    esa[ esa > 7 ] = 0                    # Nodata including snow

    return esa

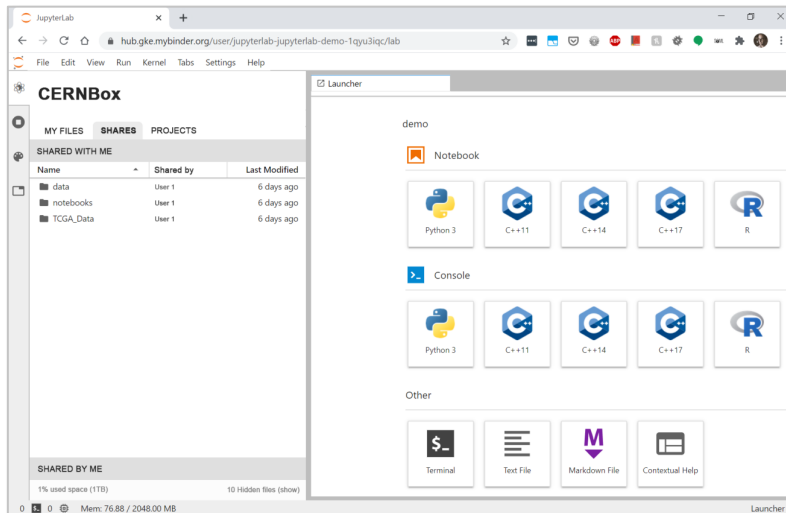
# Process a polygon and display the output raster in the map
def LULC_LandCover():

    if len(selected_geojson) > 0 and len(year.v_model) > 0:

        waitON()
        map.clear()
```

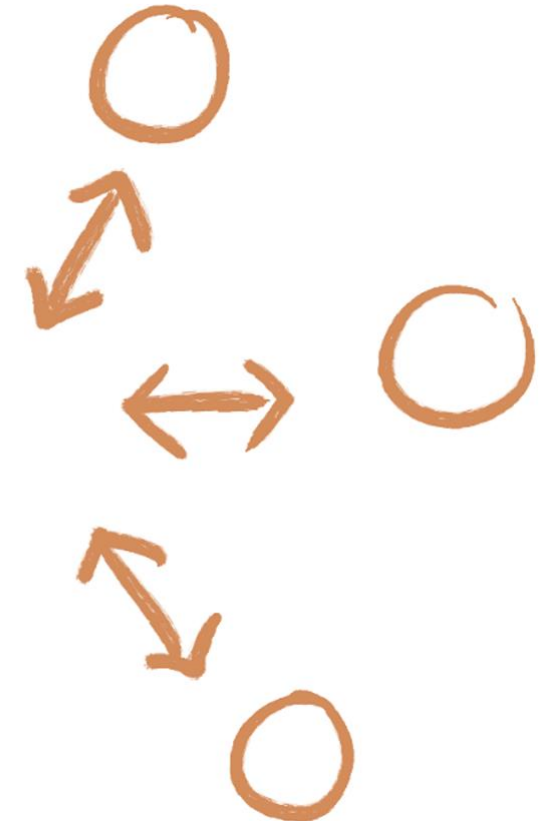

JupyterLab extension (Cs3Api4Lab)

Integration with ScienceMesh IOP (CS3 APIS)



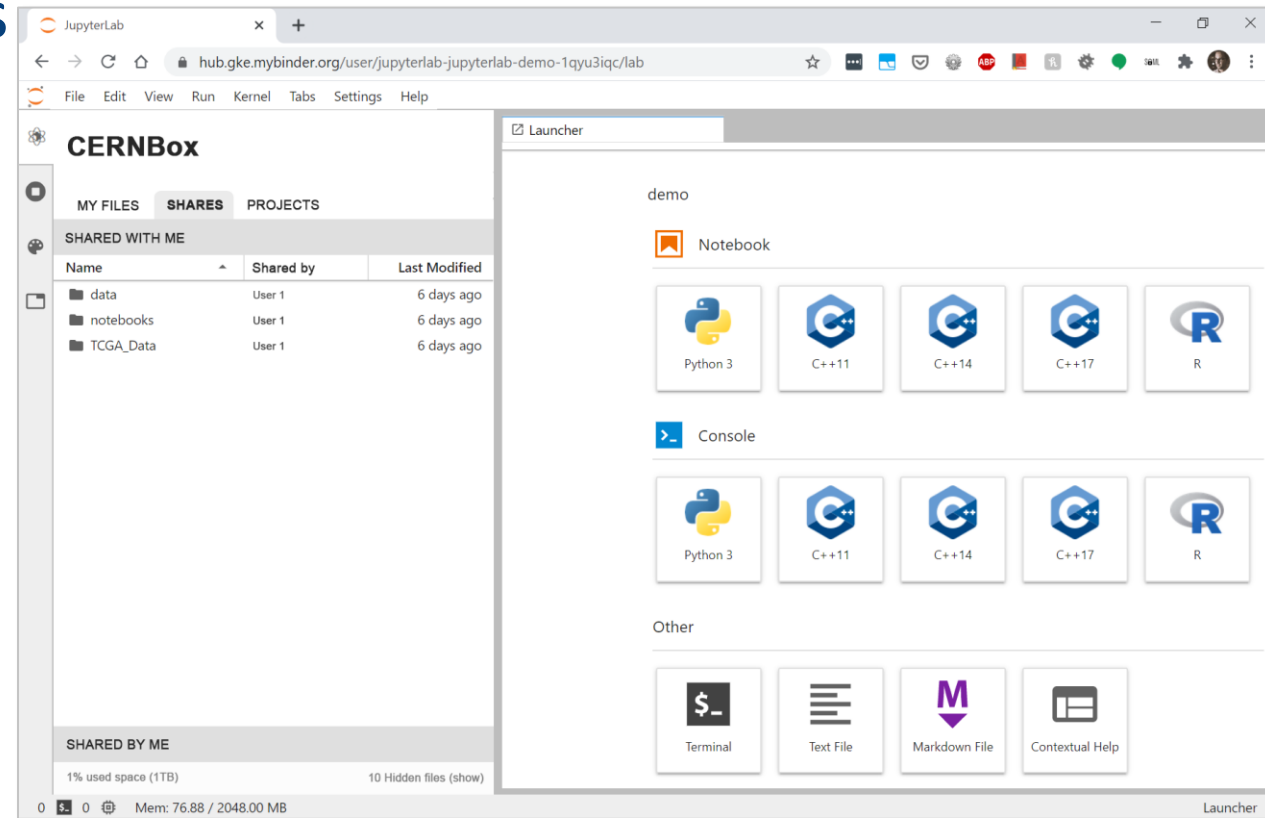
jupyterlab

CS3 APIs



JupyterLab extension (Cs3Api4Lab): Frontend

- # Full client in Lab
- # File browser – share functionalities
 - # Shared by/with tab
 - # Sharing buttons
 - # Entries in the context menus
 - # Pop-up windows: file information and sharing status
 - # Account info
- # File browsing



File Edit View Run Kernel Tabs Settings Help

cs3Api4Lab

My Files Shares Projects

Filter files by name

Name	Last Modified
analysis_2	an hour ago
MyShares	11 hours ago
completely_different.ipynb	11 hours ago
simple_visualization_1.ipynb	6 minutes ago

simple_visualization_1.ipynb

```
[1]: import pandas as pd
import io
import csv
import matplotlib.pyplot as plt

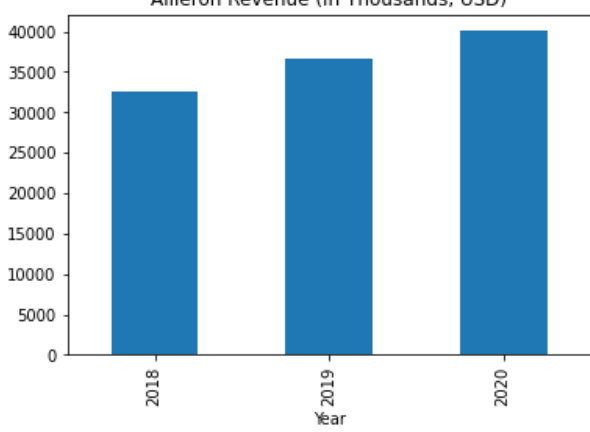
[2]: data_string = '''
Year→EV→Revenue→EBITDA→Net Income→Total Assets
2018→35623→32579→3866→1388→36397
2019→29646→36636→2715→13000→35333
2020→36340→40054→5665→2418→43703
'''

[3]: ailleron_financials_summary = pd.read_csv(io.StringIO(data_string), sep='\t', quoting=csv.QUOTE_NONE, index_col='Year');
ailleron_financials_summary.dropna(inplace=True);

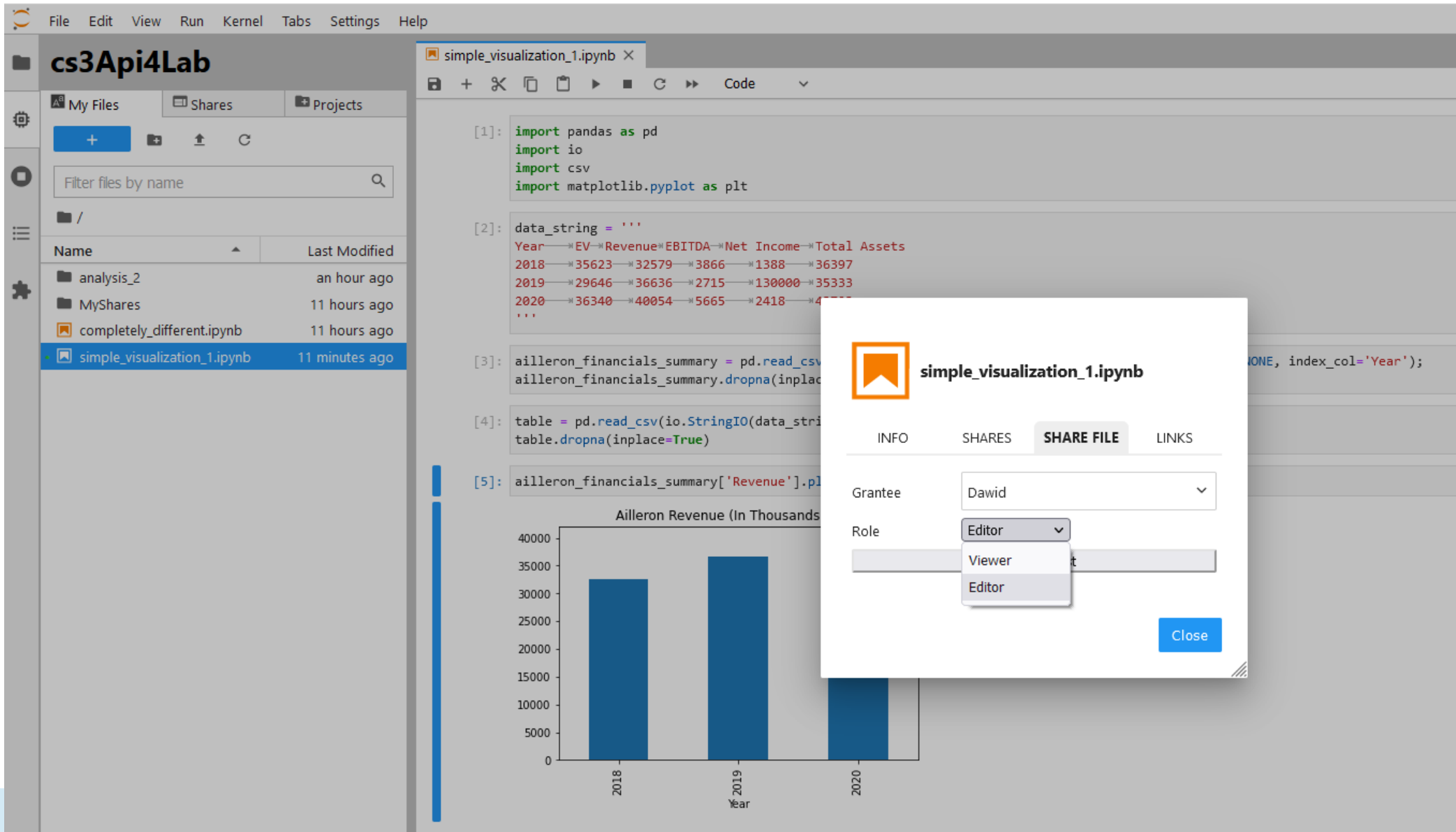
[4]: table = pd.read_csv(io.StringIO(data_string), sep=',', quoting=csv.QUOTE_NONE)
table.dropna(inplace=True)

[5]: ailleron_financials_summary['Revenue'].plot.bar(title="Ailleron Revenue (In Thousands, USD)");
```

Ailleron Revenue (In Thousands, USD)



Year	Revenue
2018	35623
2019	29646
2020	36340



The screenshot displays the **cs3Api4Lab** web interface. On the left, a file explorer shows a list of files, with **simple_visualization_1.ipynb** selected. The main area shows the notebook content, including code cells and a bar chart titled "Ailleron Revenue (In Thousands)".

The code in the notebook is as follows:

```
[1]: import pandas as pd
import io
import csv
import matplotlib.pyplot as plt

[2]: data_string = '''
Year→EV→Revenue→EBITDA→Net Income→Total Assets
2018→35623→32579→3866→1388→36397
2019→29646→36636→2715→130000→35333
2020→36340→40054→5665→2418→40000
'''

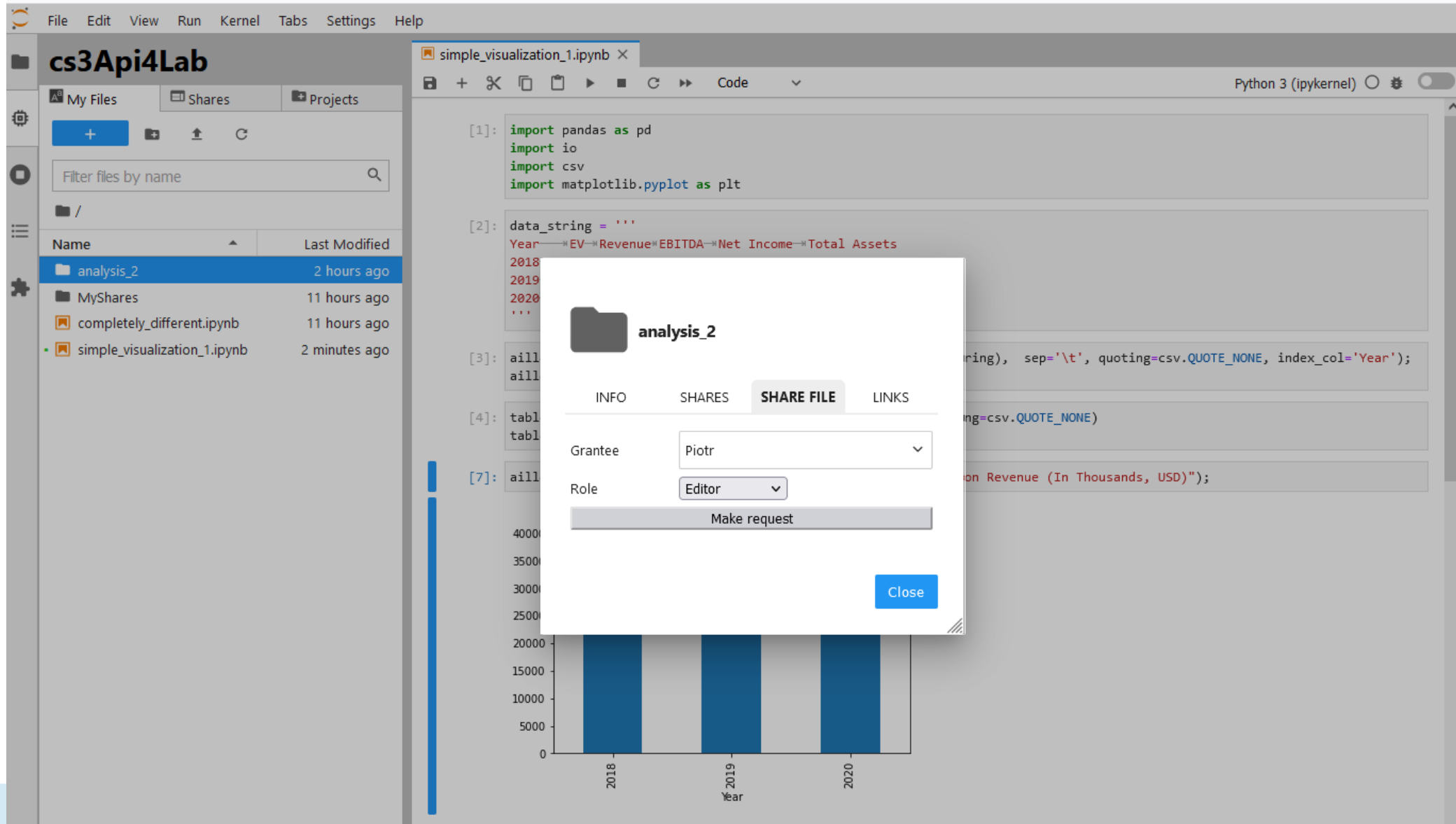
[3]: ailleron_financials_summary = pd.read_csv(io.StringIO(data_string))
ailleron_financials_summary.dropna(inplace=True)

[4]: table = pd.read_csv(io.StringIO(data_string))
table.dropna(inplace=True)

[5]: ailleron_financials_summary['Revenue'].plot(kind='bar')
```

The bar chart shows revenue for the years 2018, 2019, and 2020. The y-axis ranges from 0 to 40,000. The bars represent revenue values of approximately 35,623 for 2018, 29,646 for 2019, and 36,340 for 2020.

A sharing dialog box is open over the chart, titled "simple_visualization_1.ipynb". It has tabs for "INFO", "SHARES", "SHARE FILE", and "LINKS". The "SHARE FILE" tab is active, showing a "Grantee" dropdown set to "Dawid" and a "Role" dropdown with "Editor" selected. A "Close" button is at the bottom right.



The screenshot shows a JupyterLab environment with a file browser on the left and a code editor on the right. The file browser shows a folder named 'analysis_2' selected. The code editor contains Python code for data analysis and visualization. A dialog box is open over the code editor, showing the sharing options for the 'analysis_2' folder. The dialog has tabs for 'INFO', 'SHARES', 'SHARE FILE', and 'LINKS'. The 'SHARE FILE' tab is active, showing a 'Grantee' dropdown set to 'Piotr' and a 'Role' dropdown set to 'Editor'. A 'Make request' button is visible at the bottom of the dialog. The background code editor shows the following code:

```
[1]: import pandas as pd
import io
import csv
import matplotlib.pyplot as plt

[2]: data_string = '''
Year→EV→Revenue→EBITDA→Net Income→Total Assets
2018
2019
2020
...

[3]: aill
aill

[4]: tabl
tabl

[7]: aill

4000
3500
3000
2500
20000
15000
10000
5000
0
2018 2019 2020
Year
```


File Edit View Run Kernel Tabs Settings Help

cs3Api4Lab

My Files Shares Projects

Share by Me

Filter files by name

/

Name	Last Modified
analysis_2	7 hours ago
simple_visualization_1.ipynb	7 hours ago

Share with Me

Filter files by name

/

Name	Last Modified
shared_folder	6 hours ago

simple_visualization_1.ipynb X

Python 3 (ipykernel)

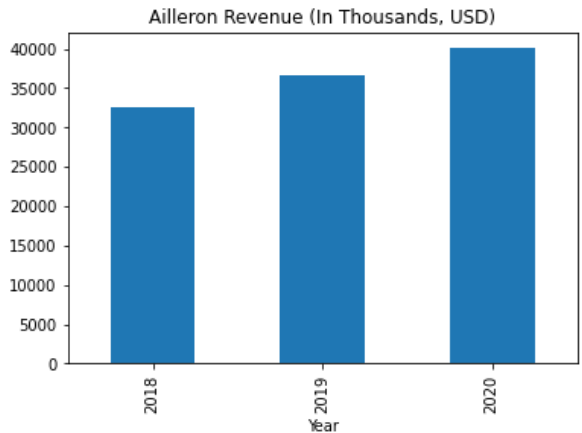
```
[1]: import pandas as pd
import io
import csv
import matplotlib.pyplot as plt

[2]: data_string = '''
Year→EV→Revenue→EBITDA→Net Income→Total Assets
2018→35623→32579→3866→1388→36397
2019→29646→36636→2715→130000→35333
2020→36340→40054→5665→2418→43703
'''

[3]: ailleron_financials_summary = pd.read_csv(io.StringIO(data_string), sep='\t', quoting=csv.QUOTE_NONE, index_col='Year');
ailleron_financials_summary.dropna(inplace=True);

[4]: table = pd.read_csv(io.StringIO(data_string), sep=',', quoting=csv.QUOTE_NONE)
table.dropna(inplace=True)

[7]: ailleron_financials_summary['Revenue'].plot.bar(title="Ailleron Revenue (In Thousands, USD)");
```



Year	Revenue
2018	32579
2019	36636
2020	40054

File Edit View Run Kernel Tabs Settings Help

cs3Api4Lab

My Files Shares Projects

Share by Me

Filter files by name

/

Name	Last Modified
shared_folder	6 hours ago

Share with Me

Filter files by name

/

Name	Last Modified
analysis_2	7 hours ago
simple_visualization_1.ipynb	7 hours ago

Launcher

Notebook

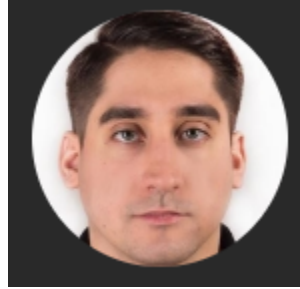
Python 3 (ipykernel)

Console

Python 3 (ipykernel)

Other

Terminal Text File Markdown File Show Contextual Help



Dawid

File Edit View Run Kernel Tabs Settings Help

cs3Api4Lab

My Files Shares Projects

Share by Me

Filter files by name

/

Name	Last Modified
shared_folder	6 hours ago

Share with Me

Filter files by name

/

Name	Last Modified
analysis_2	7 hours ago
simple_visualization_1.ipynb	7 hours ago

simple_visualization_1.ipynb X Python 3 (ipykernel)

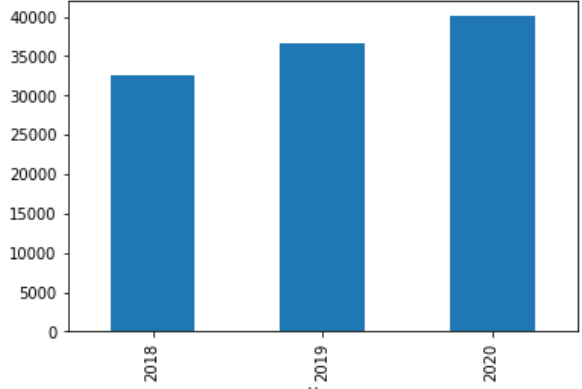
```
[1]: import pandas as pd
import io
import csv
import matplotlib.pyplot as plt

[2]: data_string = '''
Year→EV→Revenue→EBITDA→Net Income→Total Assets
2018→35623→32579→3866→1388→36397
2019→29646→36636→2715→130000→35333
2020→36340→40054→5665→2418→43703
...
'''
[3]: ailleron_financials_summary = pd.read_csv(io.StringIO(data_string), sep='\t', quoting=csv.QUOTE_NONE, index_col='Year');
ailleron_financials_summary.dropna(inplace=True);

[4]: table = pd.read_csv(io.StringIO(data_string), sep=',', quoting=csv.QUOTE_NONE)
table.dropna(inplace=True)

[7]: ailleron_financials_summary['Revenue'].plot.bar(title="Ailleron Revenue (In Thousands, USD)");
```

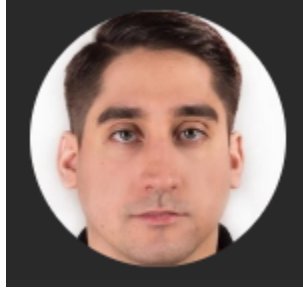
Ailleron Revenue (In Thousands, USD)



Year	EV	Revenue	EBITDA	Net Income	Total Assets
2018	35623	32579	3866	1388	36397
2019	29646	36636	2715	130000	35333
2020	36340	40054	5665	2418	43703

← Insert new data

2021 (Q1-Q3)	38040	52526	7714	2291	63740
--------------	-------	-------	------	------	-------



Dawid

File Edit View Run Kernel Tabs Settings Help

cs3Api4Lab

My Files Shares Projects

Share by Me

/

Name	Last Modified
analysis_2	7 hours ago
simple_visualization_1.ipynb	8 hours ago

Share with Me

/

Name	Last Modified
shared_folder	7 hours ago

simple_visualization_1.ipynb X Python 3 (ipykernel)

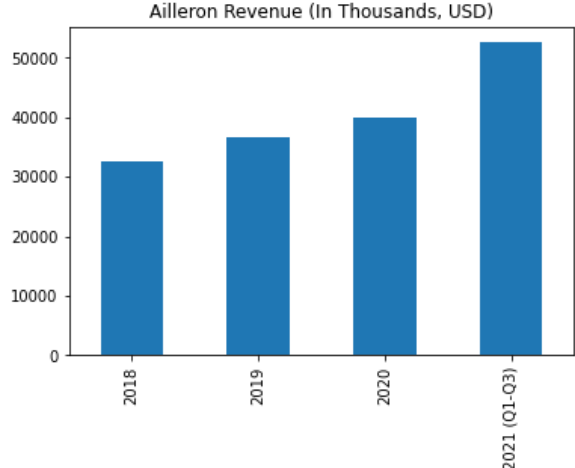
```
[1]: import pandas as pd
import io
import csv
import matplotlib.pyplot as plt

[11]: data_string = '''
Year→EV→Revenue→EBITDA→Net Income→Total Assets
2018→35623→32579→3866→1388→36397
2019→29646→36636→2715→130000→35333
2020→36340→40054→5665→2418→43703
2021 (Q1-Q3)→38040→52526→7714→2291→63740'''

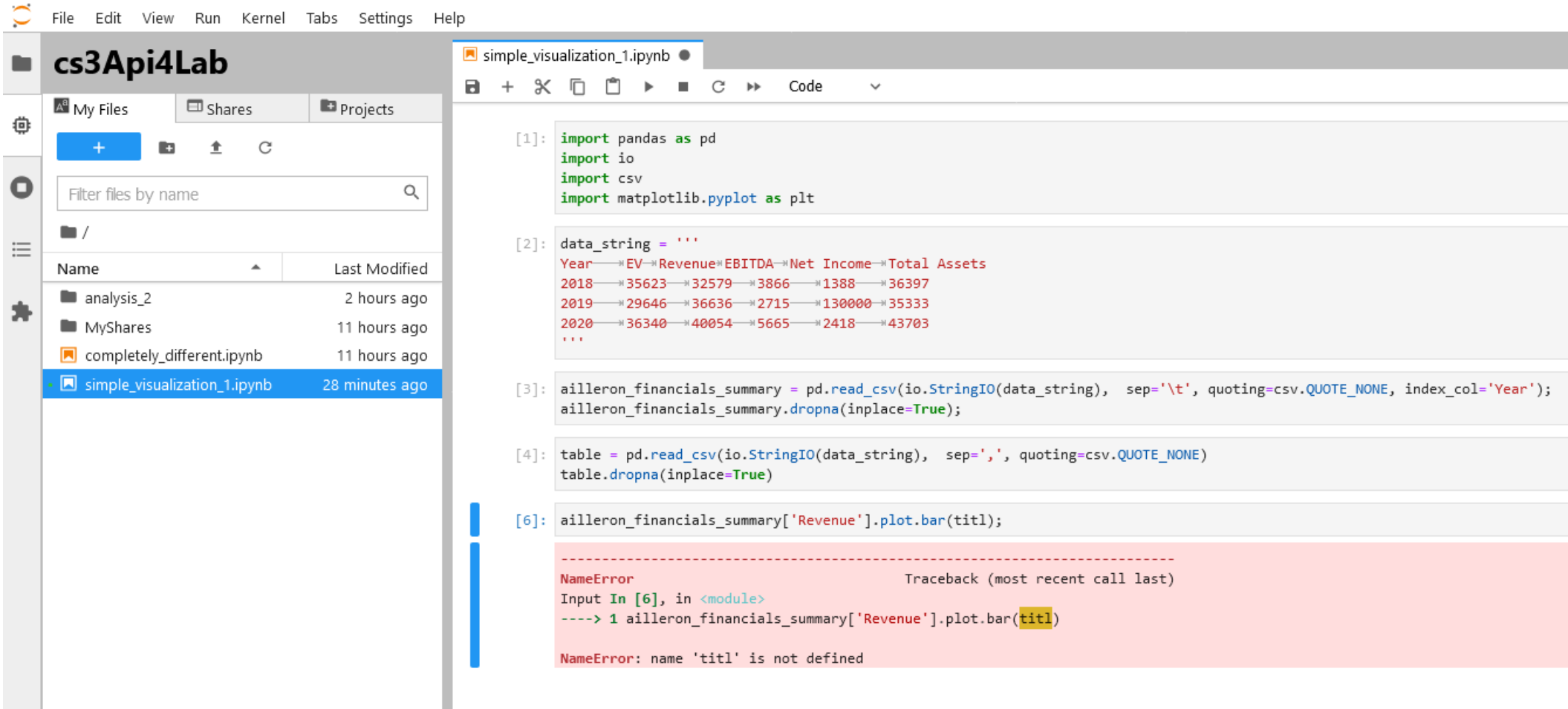
[12]: ailleron_financials_summary = pd.read_csv(io.StringIO(data_string), sep='\t', quoting=csv.QUOTE_NONE, index_col='Year');
ailleron_financials_summary.dropna(inplace=True);

[13]: table = pd.read_csv(io.StringIO(data_string), sep=',', quoting=csv.QUOTE_NONE)
table.dropna(inplace=True)

[14]: ailleron_financials_summary['Revenue'].plot.bar(title="Ailleron Revenue (In Thousands, USD)");
```



Year	Revenue
2018	32579
2019	36636
2020	40054
2021 (Q1-Q3)	52526



File Edit View Run Kernel Tabs Settings Help

cs3Api4Lab

My Files Shares Projects

Filter files by name

Name	Last Modified
analysis_2	2 hours ago
MyShares	11 hours ago
completely_different.ipynb	11 hours ago
simple_visualization_1.ipynb	28 minutes ago

```
[1]: import pandas as pd
import io
import csv
import matplotlib.pyplot as plt

[2]: data_string = '''
Year→EV→Revenue→EBITDA→Net Income→Total Assets
2018→35623→32579→3866→1388→36397
2019→29646→36636→2715→130000→35333
2020→36340→40054→5665→2418→43703
...

[3]: ailleron_financials_summary = pd.read_csv(io.StringIO(data_string), sep='\t', quoting=csv.QUOTE_NONE, index_col='Year');
ailleron_financials_summary.dropna(inplace=True);

[4]: table = pd.read_csv(io.StringIO(data_string), sep=',', quoting=csv.QUOTE_NONE)
table.dropna(inplace=True)

[6]: ailleron_financials_summary['Revenue'].plot.bar(titl);

-----
NameError                                Traceback (most recent call last)
Input In [6], in <module>
----> 1 ailleron_financials_summary['Revenue'].plot.bar(titl)

NameError: name 'titl' is not defined
```


File Edit View Run Kernel Tabs Settings Help

cs3Api4Lab

My Files Shares Projects

Filter files by name

/ analysis_2 /

Name	Last Modified
simple_visualization_2.ipynb	12 minutes ago

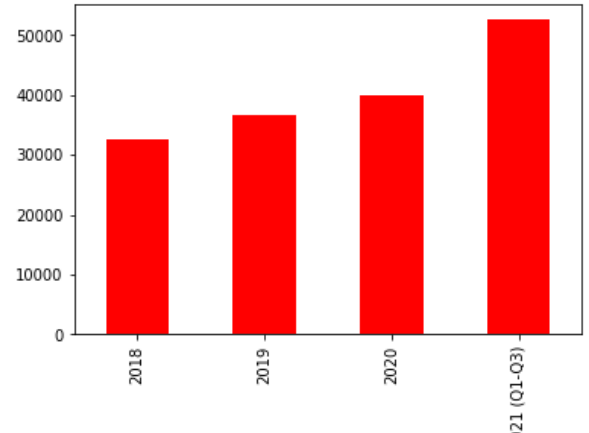
simple_visualization_2.ipynb Python 3 (ipykernel)

```
[3]: ailleron_financials_summary = pd.read_csv(io.StringIO(data_string), sep='\t', quoting=csv.QUOTE_NONE, index_col='Year');
```

```
[4]: ailleron_financials_summary
```

	EV	Revenue	EBITDA	Net Income	Total Assets
Year					
2018	35623	32579	3866	1388	36397
2019	29646	36636	2715	130000	35333
2020	36340	40054	5665	2418	43703
2021 (Q1-Q3)	38040	52526	7714	2291	63740

```
[6]: ailleron_financials_summary['Revenue'].plot.bar(title="Ailleron Revenue (In Thousands, USD)", color='r');
```



File Edit View Run Kernel Tabs Settings Help

cs3Api4Lab

My Files Shares Projects

Share by Me

Filter files by name

/

Name	Last Modified
analysis_2	42 minutes ago
simple_visualization_1.ipynb	an hour ago

Share with Me

Filter files by name

/ ... / dawid / shared_folder /

Name	Last Modified
visualization.ipynb	27 minutes ago
visualization.ipynb.marcin.2022-01-25_19_35_50.conflict.ipynb	2 minutes ago
visualization.ipynb.piotr.2022-01-25_19_09_53.conflict	11 minutes ago


simple_visualization_2.ipynb Python 3 (ipykernel)

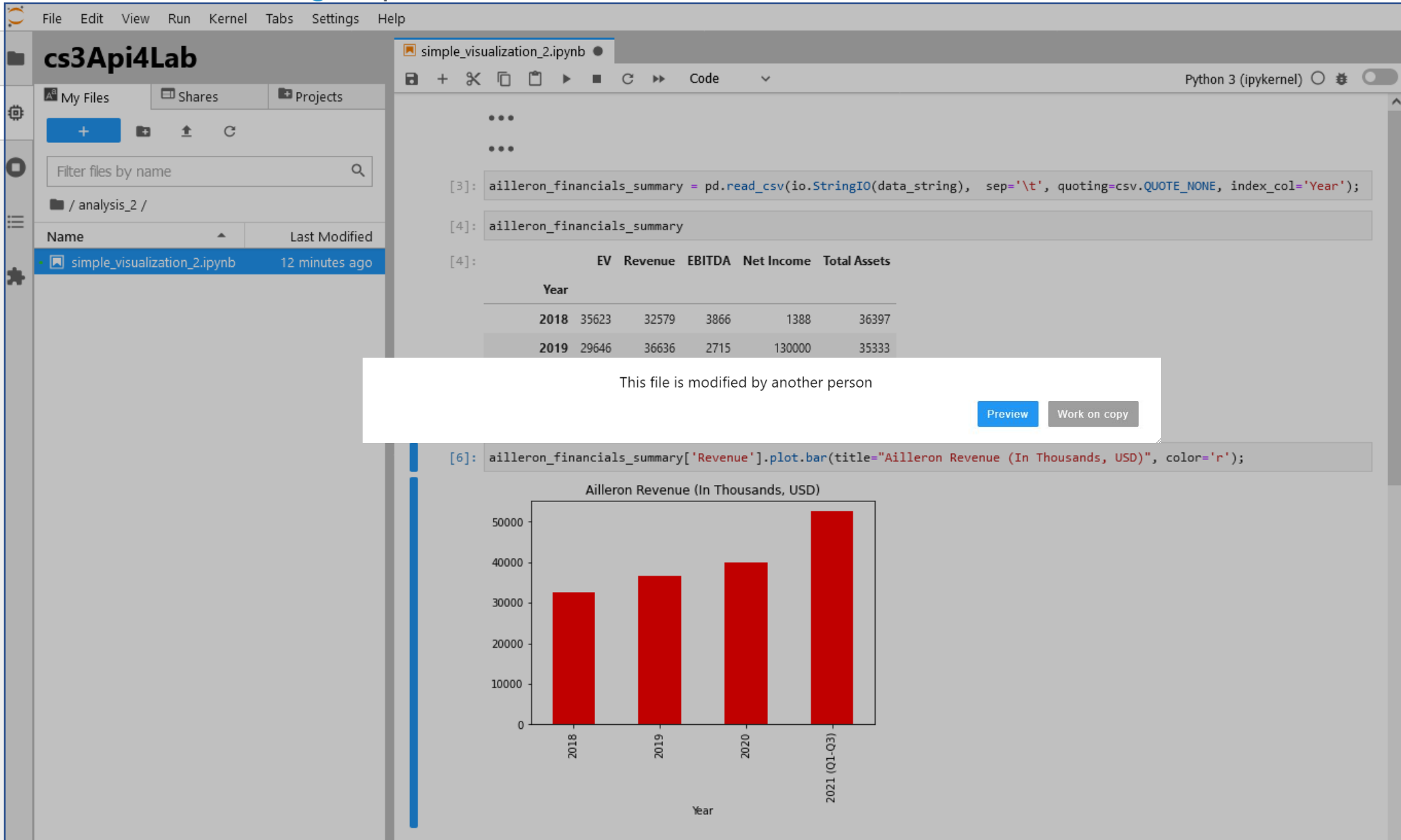
```
[3]: ailleron_financials_summary = pd.read_csv(io.StringIO(data_string), sep='\t', quoting=csv.QUOTE_NONE, index_col='Year');
```

```
[4]: ailleron_financials_summary
```

	EV	Revenue	EBITDA	Net Income	Total Assets
Year					
2018	35623	32579	3866	1388	36397
2019	29646	36636	2715	130000	35333
2020	36340	40054	5665	2418	43703
2021 (Q1-Q3)	38040	52526	7714	2291	63740

```
[6]: ailleron_financials_summary['Revenue'].plot.bar(title="Ailleron Revenue (In Thousands, USD)", color='r');
```





The screenshot shows a JupyterLab environment with a file browser on the left and a notebook in the center. The notebook contains Python code for reading a CSV file and plotting a bar chart of revenue. A white dialog box with the text "This file is modified by another person" is overlaid on the notebook, with "Preview" and "Work on copy" buttons.

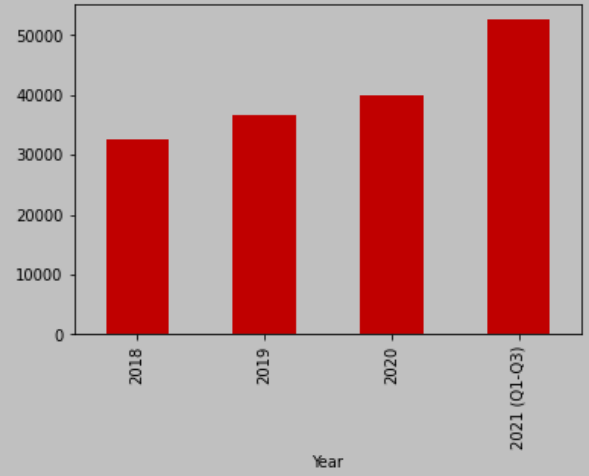
```
[3]: ailleron_financials_summary = pd.read_csv(io.StringIO(data_string), sep='\t', quoting=csv.QUOTE_NONE, index_col='Year');
```

```
[4]: ailleron_financials_summary
```

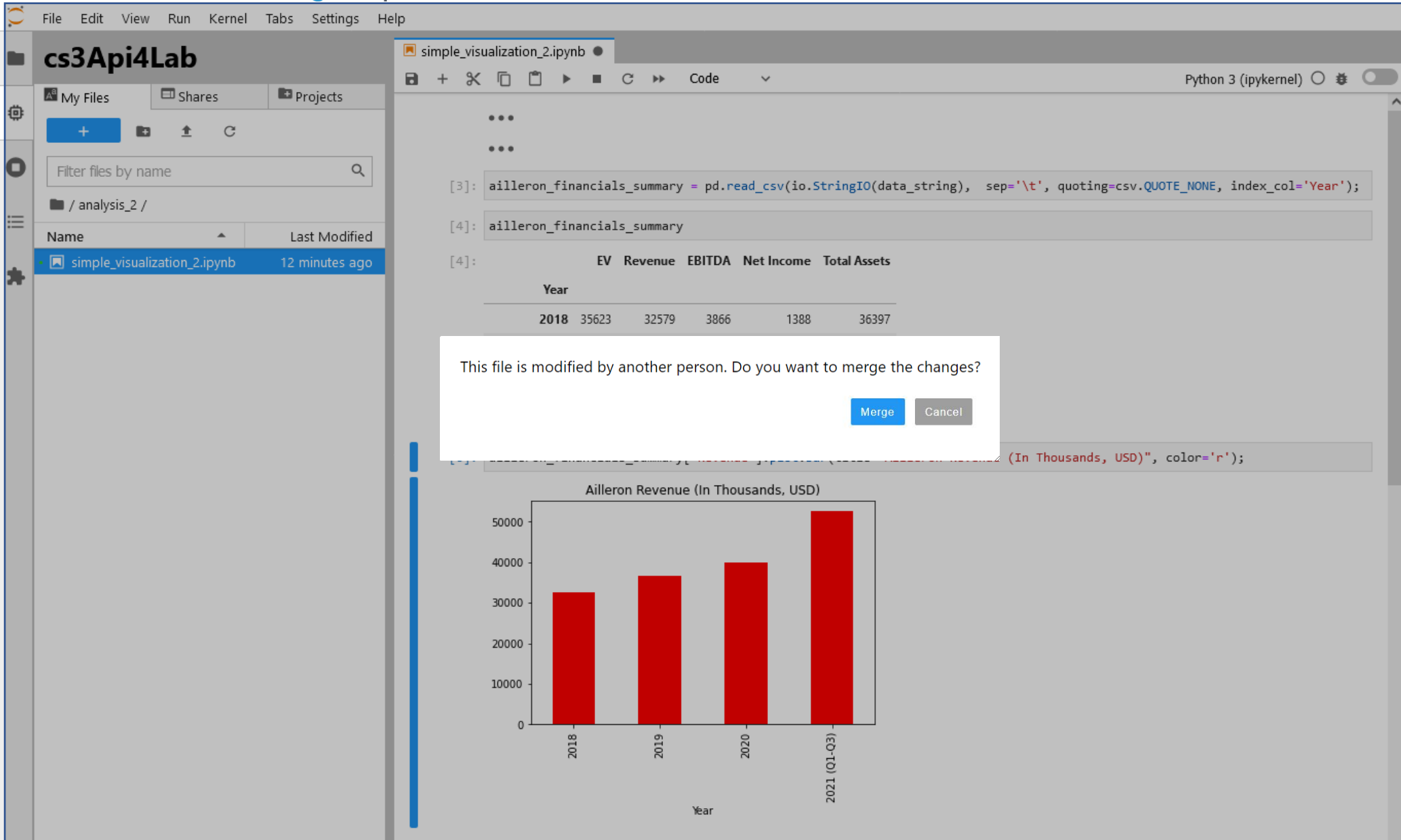
	EV	Revenue	EBITDA	Net Income	Total Assets
Year					
2018	35623	32579	3866	1388	36397
2019	29646	36636	2715	130000	35333

```
[6]: ailleron_financials_summary['Revenue'].plot.bar(title="Ailleron Revenue (In Thousands, USD)", color='r');
```

Ailleron Revenue (In Thousands, USD)



Year	Revenue (In Thousands, USD)
2018	32579
2019	36636
2020	40000
2021 (Q1-Q3)	50000



The screenshot shows a JupyterLab environment with a file browser on the left and a code editor on the right. The code editor displays a Python script that reads a CSV file and visualizes the data as a bar chart. A dialog box is overlaid on the code editor, asking for confirmation to merge changes from another user.

Code Editor Content:

```
[3]: ailleron_financials_summary = pd.read_csv(io.StringIO(data_string), sep='\t', quoting=csv.QUOTE_NONE, index_col='Year');  
[4]: ailleron_financials_summary  
[4]:
```

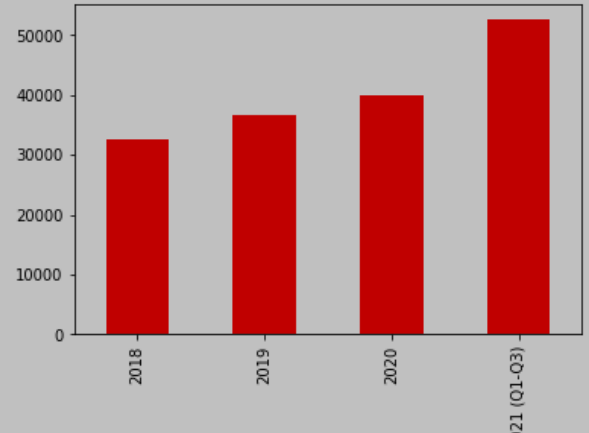
Year	EV	Revenue	EBITDA	Net Income	Total Assets
2018	35623	32579	3866	1388	36397

Dialog Box:

This file is modified by another person. Do you want to merge the changes?

Buttons: Merge, Cancel

Figure: Ailleron Revenue (In Thousands, USD)



Year	Revenue (In Thousands, USD)
2018	32579
2019	35623
2020	3866
2021 (Q1-Q3)	50000

Done

- # New implementation of file browser
- # User information (look-up users, share with a user by name)
- # Locking mechanism for concurrent updating of notebooks
- # JupyterLab 3
- # Concurrent updating of notebooks
- # Unification share APIs (shares, OCM shares)

Current

- # Testing in HEP and Earth Observation use cases

Next

- # Concurrent updating of notebooks: MERGE
- # Mount the file system, to allow local access from the kernel

MVP

github.com/sciencemesh/cs3api4lab

** ready to check it out*

Case studies:





**CS³
MESH⁴
EOSC**

Connecting European Data

Thank you!
Discover more on...

 [cs3mesh4eosc.eu](https://www.cs3mesh4eosc.eu)

 [company/cs3mesh4eosc](https://www.linkedin.com/company/cs3mesh4eosc)

 [CS3org](https://twitter.com/CS3org)

 [CS3MESH4EOSC Project](https://www.youtube.com/channel/UCHKcZEKmqXjCvc3MLFjFxbw)

<https://www.youtube.com/channel/UCHKcZEKmqXjCvc3MLFjFxbw>



CS3MESH4EOSC has received funding from the European Union's Horizon 2020 Research and Innovation programme under **Grant Agreement No. 863353**.