

Automate Solution for Data Dump Files

- 1) Open Anaconda Navigator → JupyterLab, and open your notebook file
- 2) Install the scheduler `#pip install jupyter_scheduler`
- 3) Restart the Kernel and closed all. And try to open again.
- 4) See the below figure-1, click to open the highlighted icon, [Create Notebook Jobs] file will open
- 5) Setup the cron timestamp for scheduler.

The screenshot shows a JupyterLab notebook titled 'credit_risk_data.ipynb'. The interface includes a toolbar with various icons, and a code editor with the following content:

```
[17]: # install notebook scheduler
      #pip install jupyter_scheduler

[40]: import pandas as pd

[41]: # Lets assume import data core system archived files that are not support in current data models
      df = pd.read_csv('C:\Users\Lenovo\Documents\Jupyter Notebook\credit_risk_dataset.csv')
      df.head()

[42]: from datetime import datetime

      # current datetime and format it as 'MM-DD-YYYY HH-MM-SS'
      timestamp = datetime.now().strftime('%m-%d-%Y %H-%M-%S')
      # define filename with timestamp
      filename = f'exported_data_{timestamp}.csv'

[43]: output_path = f'C:\Users\Lenovo\Documents\Jupyter Notebook\output\{filename}'
      df.to_csv(output_path, index=False)
      df.head()
```

Below the code, two data tables are displayed. The first table shows the first 5 rows of the 'credit_risk_dataset.csv' file. The second table shows the first 5 rows of the 'exported_data_{timestamp}.csv' file, which is identical to the first table.

	person_age	person_income	person_home_ownership	person_emp_length	loan_intent	loan_grade	loan_amnt	loan_int_rate	loan_status	loan_percent_income	cb_person_default_on_file	cb_p
0	22	59000	RENT	123.0	PERSONAL	D	35000	16.02	1	0.59	Y	
1	21	9600	OWN	5.0	EDUCATION	B	1000	11.14	0	0.10	N	
2	25	9600	MORTGAGE	1.0	MEDICAL	C	5500	12.87	1	0.57	N	
3	23	65500	RENT	4.0	MEDICAL	C	35000	15.23	1	0.53	N	
4	24	54400	RENT	8.0	MEDICAL	C	35000	14.27	1	0.55	Y	

	person_age	person_income	person_home_ownership	person_emp_length	loan_intent	loan_grade	loan_amnt	loan_int_rate	loan_status	loan_percent_income	cb_person_default_on_file	cb_p
0	22	59000	RENT	123.0	PERSONAL	D	35000	16.02	1	0.59	Y	
1	21	9600	OWN	5.0	EDUCATION	B	1000	11.14	0	0.10	N	
2	25	9600	MORTGAGE	1.0	MEDICAL	C	5500	12.87	1	0.57	N	
3	23	65500	RENT	4.0	MEDICAL	C	35000	15.23	1	0.53	N	
4	24	54400	RENT	8.0	MEDICAL	C	35000	14.27	1	0.55	Y	

At the bottom of the notebook, the status bar shows 'Mode: Edit', 'Ln 2, Col 32', and the file name 'credit_risk_data.ipynb'.

Figure-1

- 6) Setup every 3 minutes(able to adjust times (hours, days, weeks etc.) to run the dump files `['*/3 * * * *']` in figure-2. Reference: [Crontab file \(mcron 1.2.2\) \(gnu.org\)](https://www.gnu.org/software/cron/)
- 7) In figure-3, lastly you can see the result of the automated results.

The screenshot shows the 'Edit Job Definition' page for a job named 'credit_risk'. The page has a breadcrumb 'Job Definitions / credit_risk / Edit'. It indicates the job definition was updated at '1/2/2024, 6:53:27 PM'. Under 'Input file snapshot', the file 'credit_risk_data.ipynb' is listed. A 'Schedule' section shows a 'Custom schedule' dropdown, a cron expression of '*/3 * * * *', and a note 'Every 3 minutes'. A 'Time zone' dropdown is set to 'America/New_York'. At the bottom are 'Cancel' and 'Save Changes' buttons.

Figure-2

The screenshot shows the 'Notebook Jobs' table in the application. On the left, a file browser shows the 'output /' directory containing multiple 'exported_data' CSV files. The main table lists the jobs, all of which are 'Completed'.

Job name	Input file	Output files	Created at	Status	Actions
credit_risk	credit_risk_data.ipynb		1/2/2024, 7:34:04 PM	Completed	X
credit_risk	credit_risk_data.ipynb		1/2/2024, 7:21:04 PM	Completed	X
credit_risk	credit_risk_data.ipynb		1/2/2024, 7:18:03 PM	Completed	X
credit_risk	credit_risk_data.ipynb		1/2/2024, 7:15:03 PM	Completed	X
credit_risk	credit_risk_data.ipynb		1/2/2024, 7:12:03 PM	Completed	X
credit_risk	credit_risk_data.ipynb		1/2/2024, 7:09:02 PM	Completed	X
credit_risk	credit_risk_data.ipynb		1/2/2024, 7:06:02 PM	Completed	X
credit_risk	credit_risk_data.ipynb		1/2/2024, 7:03:02 PM	Completed	X
credit_risk	credit_risk_data.ipynb		1/2/2024, 7:00:02 PM	Completed	X
credit_risk	credit_risk_data.ipynb		1/2/2024, 6:57:02 PM	Completed	X
credit_risk	credit_risk_data.ipynb	HTML	1/2/2024, 6:54:01 PM	Completed	X

Figure-3