Problems

# Problem

I do not know how to force Python to automatically set working directory like in R. So, I have a separate script for it.

01\_Set the working directory.py

# Problem

Python does not provide table function.

In R

> table(df$age)

13 14 15 16 17 18 19 20 21

516 1318 2484 3357 3846 4476 3416 3610 3485

22 23 24 25 26 27 28 29 30

2999 2884 2550 2466 2220 2101 1986 1731 1873

Solution: I had to write it myself.

# Problem

How to save data so it will keep the info like correct categories and types specified in previous script.

So, I have chosen Parquet Format.

# Save the DataFrame to Parquet file

df.to\_parquet('data/data.parquet')

# Loading a DataFrame from a Parquet file

df = pd.read\_parquet('data/data.parquet')

The Apache Parquet project provides a standardized open-source columnar storage format for use in data analysis systems.

# Problem

How to make this expression readable:

# Group by the selected quasi-identifiers and calculate the size of each group

grouped = df.groupby(quasi\_identifiers).size().reset\_index(name='count').sort\_values(by='count', ascending=True)

Solution:

grouped = df.groupby(quasi\_identifiers

).size(

).reset\_index(name='count'

).sort\_values(by='count', ascending=True)

# Problem

Issue when installing package sdv

C:\Users\ma1187200>pip install sdv

Collecting sdv

Using cached sdv-0.3.2-py2.py3-none-any.whl.metadata (14 kB)

Collecting exrex<0.11,>=0.9.4 (from sdv)

Using cached exrex-0.10.5.tar.gz (4.8 kB)

Installing build dependencies ... done

Getting requirements to build wheel ... done

Preparing metadata (pyproject.toml) ... done

Collecting numpy<1.17,>=1.15.4 (from sdv)

Using cached numpy-1.16.6.zip (5.1 MB)

Installing build dependencies ... done

Getting requirements to build wheel ... done

Preparing metadata (pyproject.toml) ... error

error: subprocess-exited-with-error

× Preparing metadata (pyproject.toml) did not run successfully.

│ exit code: 1

╰─> [33 lines of output]

Running from numpy source directory.

<string>:394: UserWarning: Unrecognized setuptools command, proceeding with generating Cython sources and expanding templates

C:\Users\ma1187200\AppData\Local\Temp\pip-install-1vuo\_pbo\numpy\_4afdde6eedd64ec4ba79d120a8a31f4f\numpy\distutils\misc\_util.py:476: SyntaxWarning: "is" with 'str' literal. Did you mean "=="?

return is\_string(s) and ('\*' in s or '?' is s)

Traceback (most recent call last):

File "C:\Users\ma1187200\AppData\Local\Programs\Python\Python313\Lib\site-packages\pip\\_vendor\pyproject\_hooks\\_in\_process\\_in\_process.py", line 353, in <module>

main()

~~~~^^

File "C:\Users\ma1187200\AppData\Local\Programs\Python\Python313\Lib\site-packages\pip\\_vendor\pyproject\_hooks\\_in\_process\\_in\_process.py", line 335, in main

json\_out['return\_val'] = hook(\*\*hook\_input['kwargs'])

~~~~^^^^^^^^^^^^^^^^^^^^^^^^

File "C:\Users\ma1187200\AppData\Local\Programs\Python\Python313\Lib\site-packages\pip\\_vendor\pyproject\_hooks\\_in\_process\\_in\_process.py", line 149, in prepare\_metadata\_for\_build\_wheel

return hook(metadata\_directory, config\_settings)

File "C:\Users\ma1187200\AppData\Local\Temp\pip-build-env-degixvy3\overlay\Lib\site-packages\setuptools\build\_meta.py", line 377, in prepare\_metadata\_for\_build\_wheel

self.run\_setup()

~~~~~~~~~~~~~~^^

File "C:\Users\ma1187200\AppData\Local\Temp\pip-build-env-degixvy3\overlay\Lib\site-packages\setuptools\build\_meta.py", line 522, in run\_setup

super().run\_setup(setup\_script=setup\_script)

~~~~~~~~~~~~~~~~~^^^^^^^^^^^^^^^^^^^^^^^^^^^

File "C:\Users\ma1187200\AppData\Local\Temp\pip-build-env-degixvy3\overlay\Lib\site-packages\setuptools\build\_meta.py", line 320, in run\_setup

exec(code, locals())

~~~~^^^^^^^^^^^^^^^^

File "<string>", line 419, in <module>

File "<string>", line 398, in setup\_package

File "C:\Users\ma1187200\AppData\Local\Temp\pip-install-1vuo\_pbo\numpy\_4afdde6eedd64ec4ba79d120a8a31f4f\numpy\distutils\core.py", line 26, in <module>

from numpy.distutils.command import config, config\_compiler, \

...<2 lines>...

install\_clib

File "C:\Users\ma1187200\AppData\Local\Temp\pip-install-1vuo\_pbo\numpy\_4afdde6eedd64ec4ba79d120a8a31f4f\numpy\distutils\command\config.py", line 19, in <module>

from numpy.distutils.mingw32ccompiler import generate\_manifest

File "C:\Users\ma1187200\AppData\Local\Temp\pip-install-1vuo\_pbo\numpy\_4afdde6eedd64ec4ba79d120a8a31f4f\numpy\distutils\mingw32ccompiler.py", line 34, in <module>

from distutils.msvccompiler import get\_build\_version as get\_build\_msvc\_version

ModuleNotFoundError: No module named 'distutils.msvccompiler'

[end of output]

note: This error originates from a subprocess, and is likely not a problem with pip.

error: metadata-generation-failed

× Encountered error while generating package metadata.

╰─> See above for output.

note: This is an issue with the package mentioned above, not pip.

hint: See above for details.

It looks like there is an issue related to **numpy** and **distutils** during the installation of the **SDV** library. This type of error is common when a package requires specific dependencies or when trying to install with an incompatible Python version.

# Problem