1 Python virtual environment

- standard libraries: venv
- third party libraries: virtualenv, pyenv, pyenv-virtualenv, virtualenvwrapper...
  this project uses **pipenv**, because
  1. this is the recommended tool for application dependency management by python.org. It combines Pipfile, pip and virtualenv into one command
  2. It replaces the requirements.txt and especially solves some issues that use requirements.txt in multiple environments such as test integration and production. This means we are able to manage virtual environments and packages only with one tool.
- commands related to use pipenv
  - install pipenv:
    `python3 -m pipenv shell --python /Library/Frameworks/Python.framework/Versions/3.11/bin/python3`
  - check the location of virtual environment: `pipenv --venv`
  - install package: `pipenv install pandas`
  - run `jupyter notebook` inside pipenv environment: `pipenv run jupyter notebook`
- AWS setup
  - install aws cli: `pipenv install awscli`
- connect aws s3 using `boto3`
  -
- pandas aggregation command
  - there are multiple ways to call an aggregation function, like list, dictionary and tuple
  - The major distinction to keep in mind is that `count` will not include `NaN` values whereas `size` will.

    ```
    agg_func_count = {'embark_town': ['count', 'nunique', 'size']}
    df.groupby(['deck']).agg(agg_func_count)
    ```