

Mini-Lesson 2: Reproducible Example (Reprex)

Richard Chen

Intro

Communicating clearly is an extremely valuable skill in all aspects of life.

Coding is no exception. **Use reprex.**

What is Reproducible Example (Reprex)?

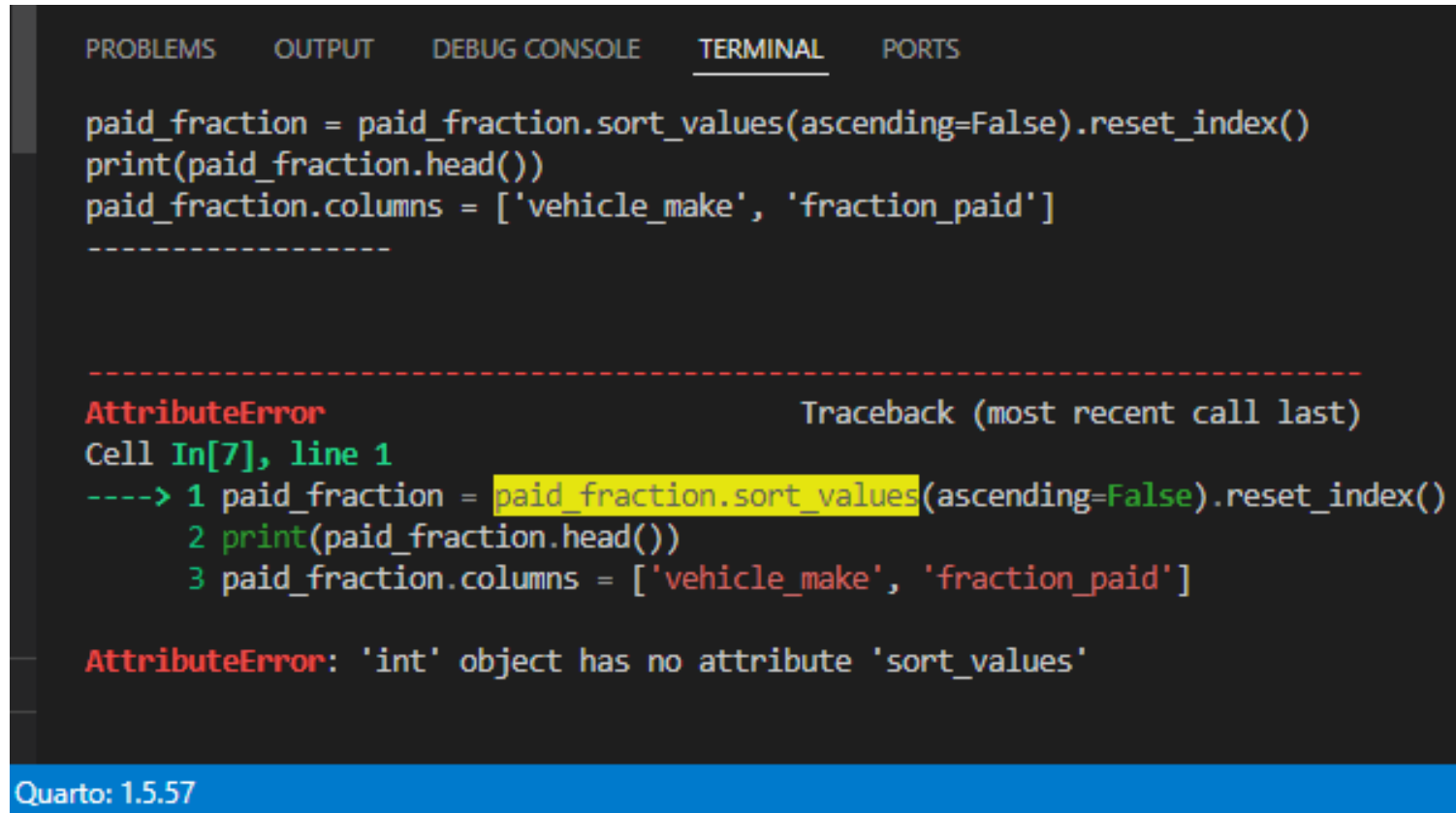
A **reproducible example (reprex)** is a simplified piece of code designed to help others understand and troubleshoot an issue quickly. Reprex involves:

- code that **actually runs**
- code that **I don't have to run**
- code that **I can easily run**

How to create reprex

1. Isolate the part of your code where the issue happens.
2. Remove unrelated parts to keep things simple.
3. Use dummy data, instead of using real files or large datasets.
4. Make sure all variables and data are defined.

Know where the error is!

A screenshot of a VS Code terminal window. The terminal has tabs for PROBLEMS, OUTPUT, DEBUG CONSOLE, TERMINAL (which is active), and PORTS. The code being executed is:

```
paid_fraction = paid_fraction.sort_values(ascending=False).reset_index()
print(paid_fraction.head())
paid_fraction.columns = ['vehicle_make', 'fraction_paid']
-----
```

 Below the code, a red dashed line separates it from the error message. The error is an **AttributeError** with the message: **AttributeError: 'int' object has no attribute 'sort_values'**. The traceback shows the error occurred in Cell In[7], line 1, at the first line of code: `paid_fraction = paid_fraction.sort_values(ascending=False).reset_index()`. The `sort_values` method call is highlighted in yellow in the original image. At the bottom of the terminal window, a blue bar indicates the Quarto version: 1.5.57.

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

paid_fraction = paid_fraction.sort_values(ascending=False).reset_index()
print(paid_fraction.head())
paid_fraction.columns = ['vehicle_make', 'fraction_paid']
-----

-----
AttributeError                                Traceback (most recent call last)
Cell In[7], line 1
----> 1 paid_fraction = paid_fraction.sort_values(ascending=False).reset_index()
      2 print(paid_fraction.head())
      3 paid_fraction.columns = ['vehicle_make', 'fraction_paid']

AttributeError: 'int' object has no attribute 'sort_values'

Quarto: 1.5.57
```

Your terminal or output panel has all the information.

Can't see your terminal? Go to Terminal > New Terminal in VSCode menu bar.

Bad example

```
1 paid_fraction = paid_fraction.sort_values(ascending=False).reset_index()  
2 print(paid_fraction.head())  
3 paid_fraction.columns = ['vehicle_make', 'fraction_paid']
```

- `paid_fraction` is not defined (we don't know what it contains)
- `print()` is not necessary
- Not clear where (and what) the error is

Good example (using repr)

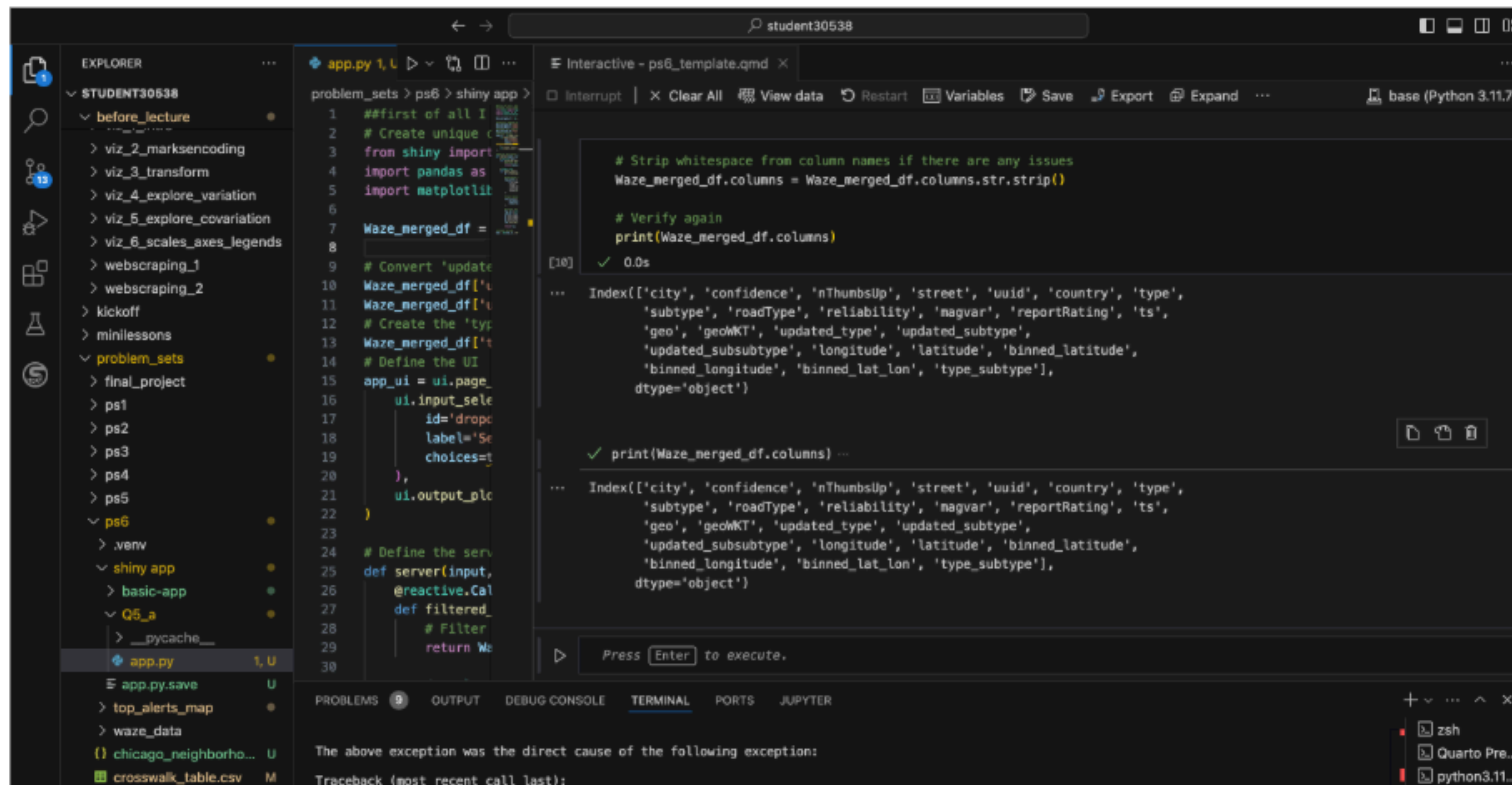
```
1 import pandas as pd
2 df = pd.DataFrame({ "vehicle_make": ["LEXU", "FORD"],
3                       "ticket_queue": ["Paid", "Paid"]})
4
5 ticket_freq = df['vehicle_make'].value_counts()
6 paid_tickets = df[df['ticket_queue'] == 'Paid'].groupby('vehicle_make').size()
7 paid_fraction = dict(paid_tickets / ticket_freq)
8
9 # The following line produces error:
10 # AttributeError: 'dict' object has no attribute 'sort_values'
11 paid_fraction = paid_fraction.sort_values(ascending=False).reset_index()
```

- all variables and data are defined
- focus on where the bug is
- contain only necessary things
- **BUT, doesn't mean you copy the entire code!**

Bad Example (from previous class)

App one is literally not working for me on the most stupid thing, it keeps failing to recognize the "updated-type" and "updated_subtype" columns, even though I printed both of them and they exist, failing to create a dropdown menu.

I have tried everything: changing to strings, creating a list, even GPT does not know what is wrong!



The screenshot shows a JupyterLab environment with a file explorer on the left, a code editor in the center, and a terminal at the bottom. The file explorer shows a project structure with folders like 'before_lecture', 'viz_2_marksencoding', 'viz_3_transform', 'viz_4_explore_variation', 'viz_5_explore_covariation', 'viz_6_scales_axes_legends', 'webscraping_1', 'webscraping_2', 'kickoff', 'minilessons', 'problem_sets', 'final_project', 'ps1', 'ps2', 'ps3', 'ps4', 'ps5', 'ps6', '.venv', 'shiny app', 'basic-app', 'Q5_a', '__pycache__', 'app.py', 'top_alerts_map', 'waze_data', 'chicago_neighborhoods', and 'crosswalk_table.csv'. The code editor shows a Python script for a Shiny app. The script includes comments and code for creating a dropdown menu. The terminal shows the output of the script, which includes a list of column names. The output is as follows:

```
[10] ✓ 0.0s
... Index(['city', 'confidence', 'nThumbsUp', 'street', 'uuid', 'country', 'type',
'subtype', 'roadType', 'reliability', 'magvar', 'reportRating', 'ts',
'geo', 'geomKT', 'updated_type', 'updated_subtype',
'updated_subsubtype', 'longitude', 'latitude', 'binned_latitude',
'binned_longitude', 'binned_lat_lon', 'type_subtype'],
dtype='object')
✓ print(Waze_merged_df.columns) ...
... Index(['city', 'confidence', 'nThumbsUp', 'street', 'uuid', 'country', 'type',
'subtype', 'roadType', 'reliability', 'magvar', 'reportRating', 'ts',
'geo', 'geomKT', 'updated_type', 'updated_subtype',
'updated_subsubtype', 'longitude', 'latitude', 'binned_latitude',
'binned_longitude', 'binned_lat_lon', 'type_subtype'],
dtype='object')
```

The terminal also shows a traceback message: "The above exception was the direct cause of the following exception: Traceback (most recent call last):".

Reproducible Example (Reprex)

Good Example

```
1 import pandas as pd
2 df = pd.DataFrame({"id":[1,2,3], "score":[10,20,30]})
3 # This line will create an error ValueError: Cannot set a DataFrame with mu
4 # df["flag"] = df[df["score"] > 15]
```

- all variables and data are defined
- focus on where the bug is
- contain only necessary things

Thank you!