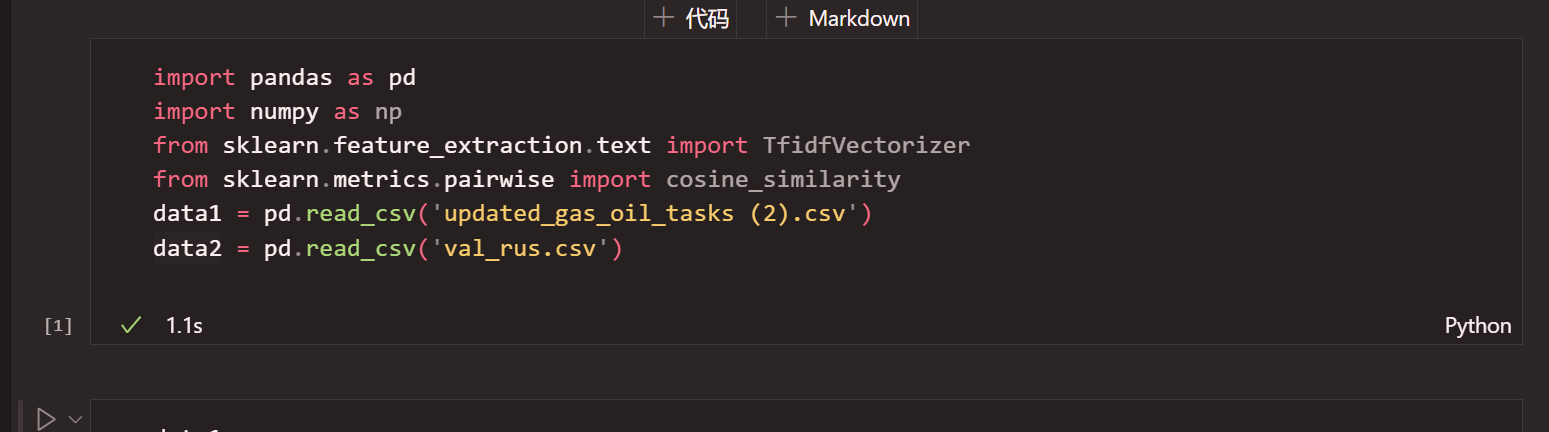
1. 读取文件
2. 1. Read the file



1. 匹配规则

（1）匹配主要是根据task\_name首字母和文本是否包含匹配；

（2）如果没找到找一个相似度最高并返回task\_id

（3）找到1个返回task\_id

（4）找到2个返回类名；

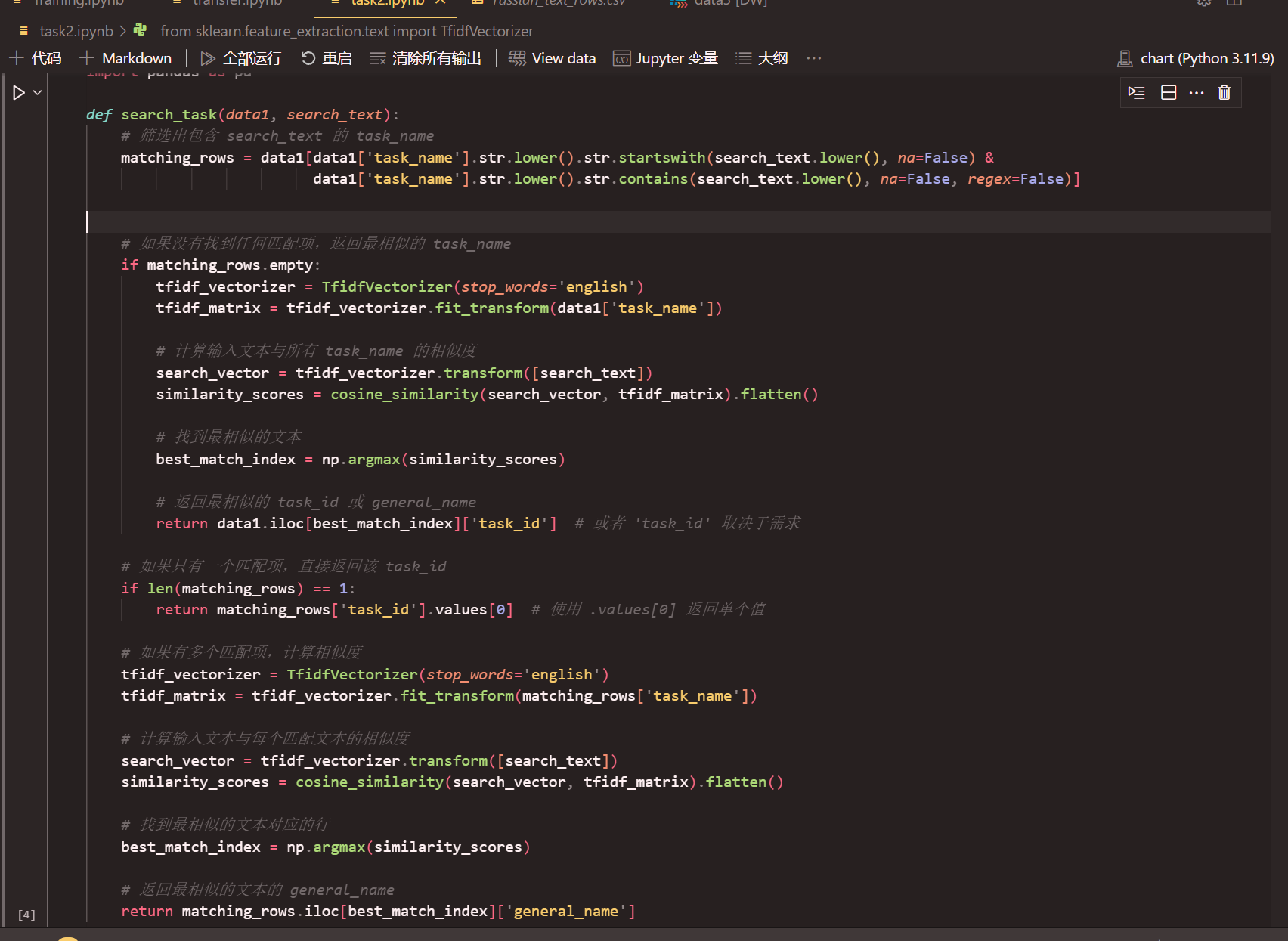
2. Matching rules

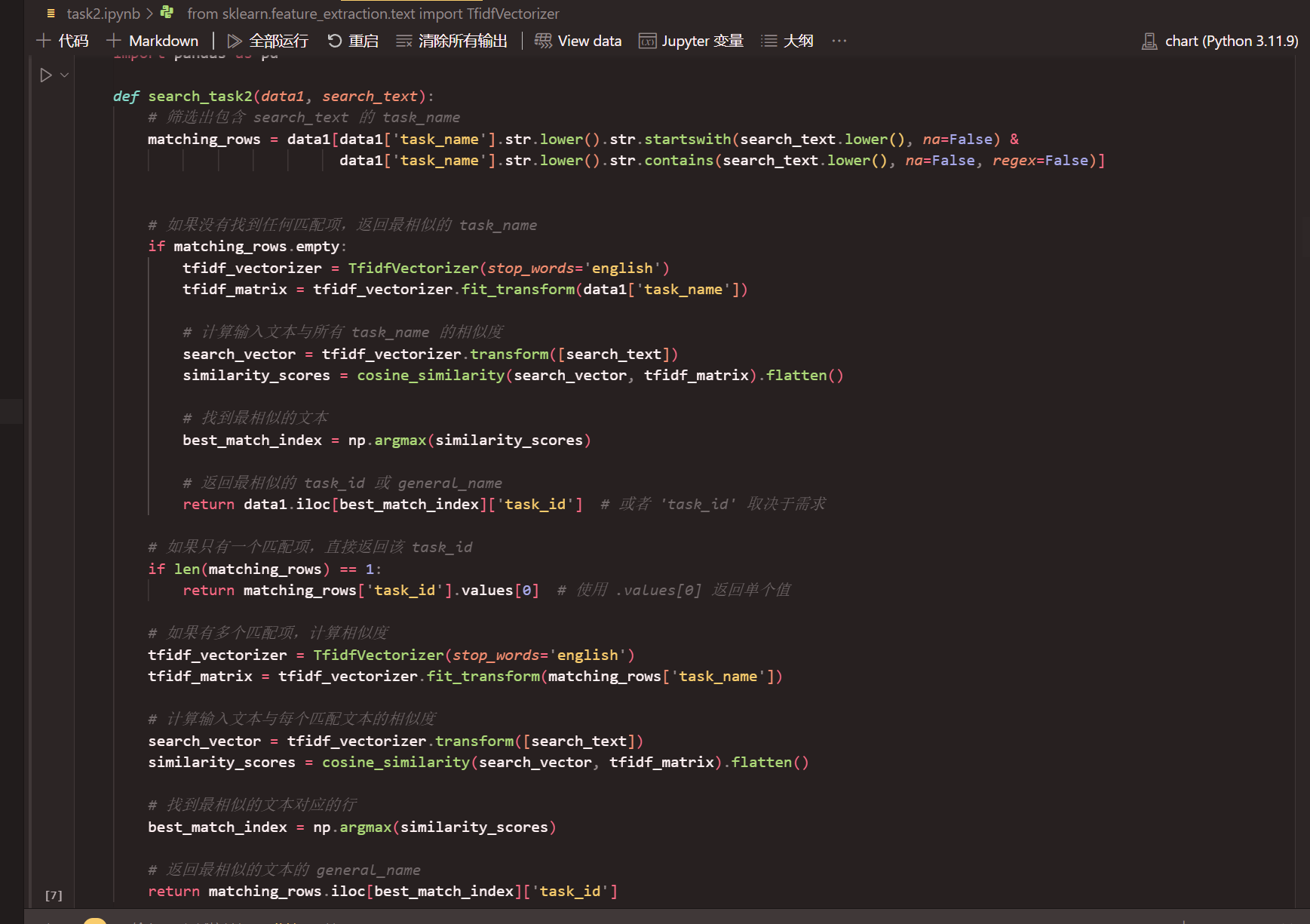
(1) Matching is mainly based on whether the first letter of task\_name and the text contain a match;

(2) If not found, find the one with the highest similarity and return task\_id

(3) If 1 is found, return task\_id

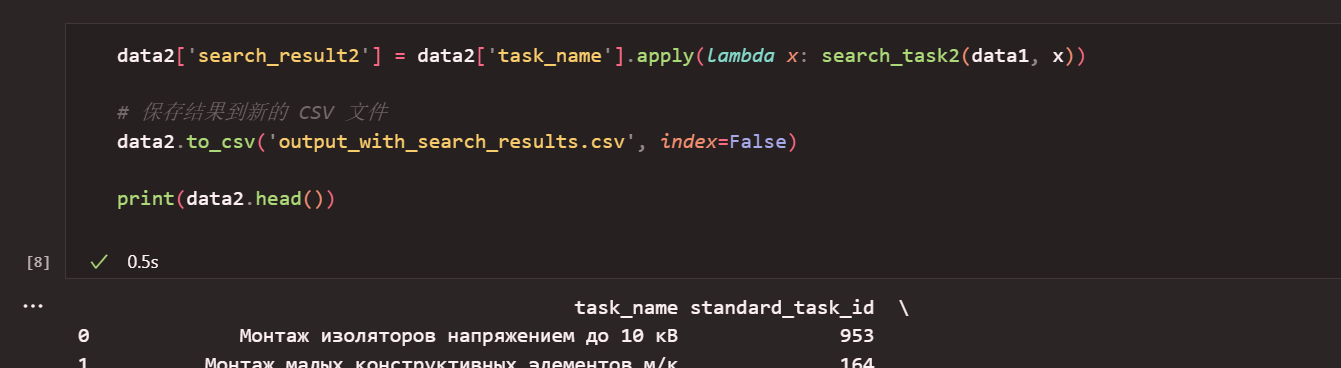
(4) If 2 are found, return class name;



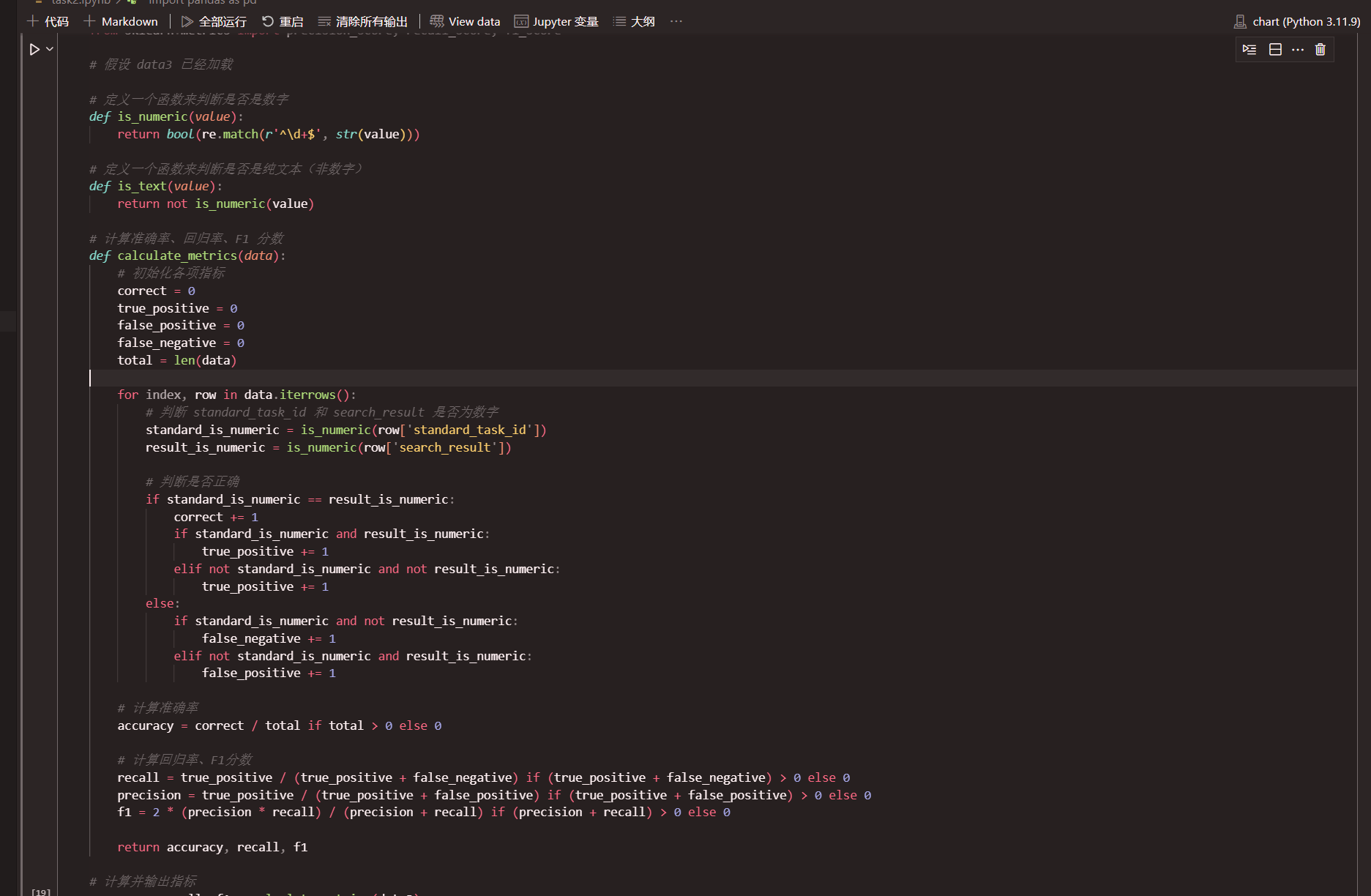


重复上述工作，但把多个也返回task\_id

Repeat the above steps, but return multiple task\_ids as well.

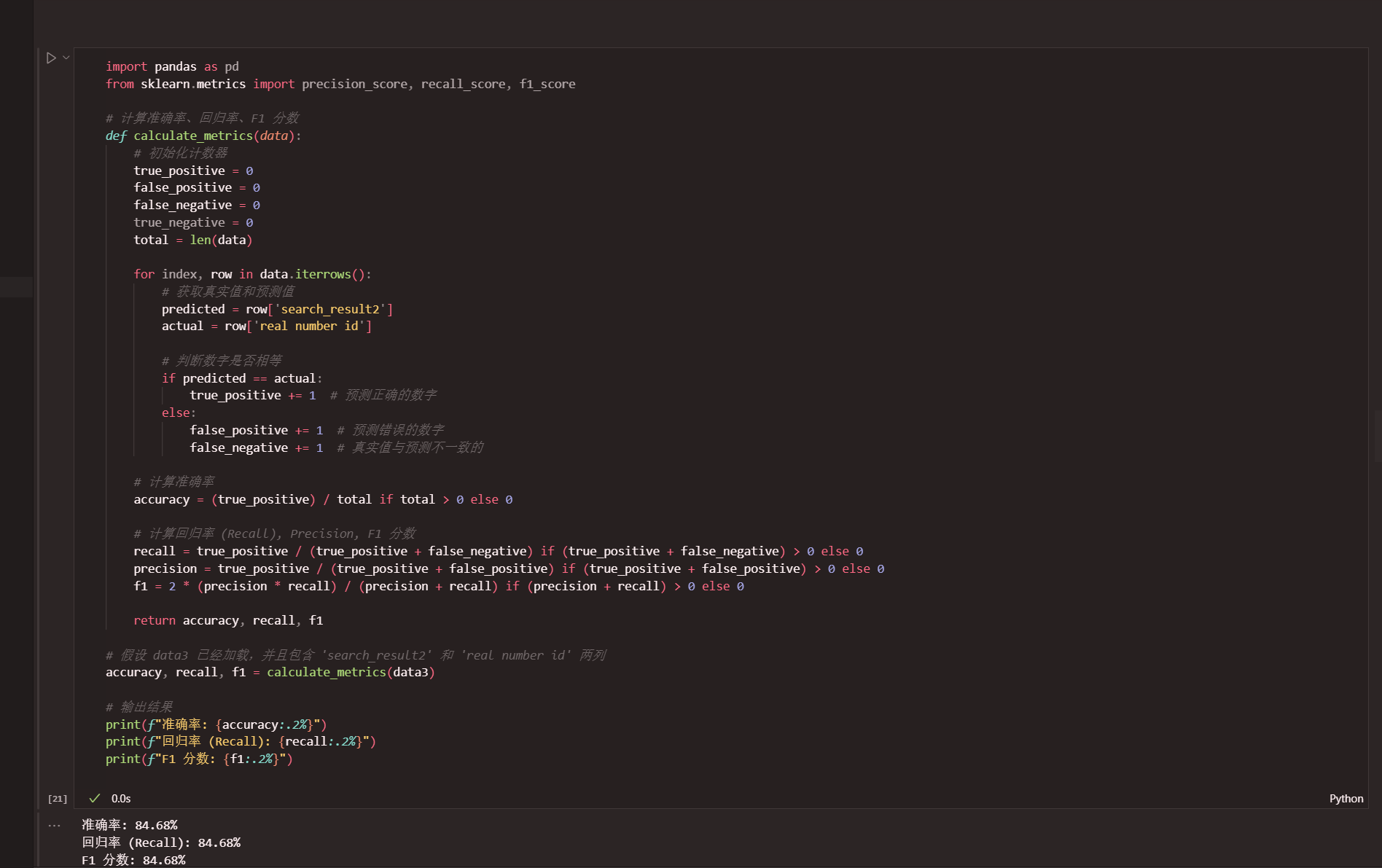


推理结果

Reasoning results

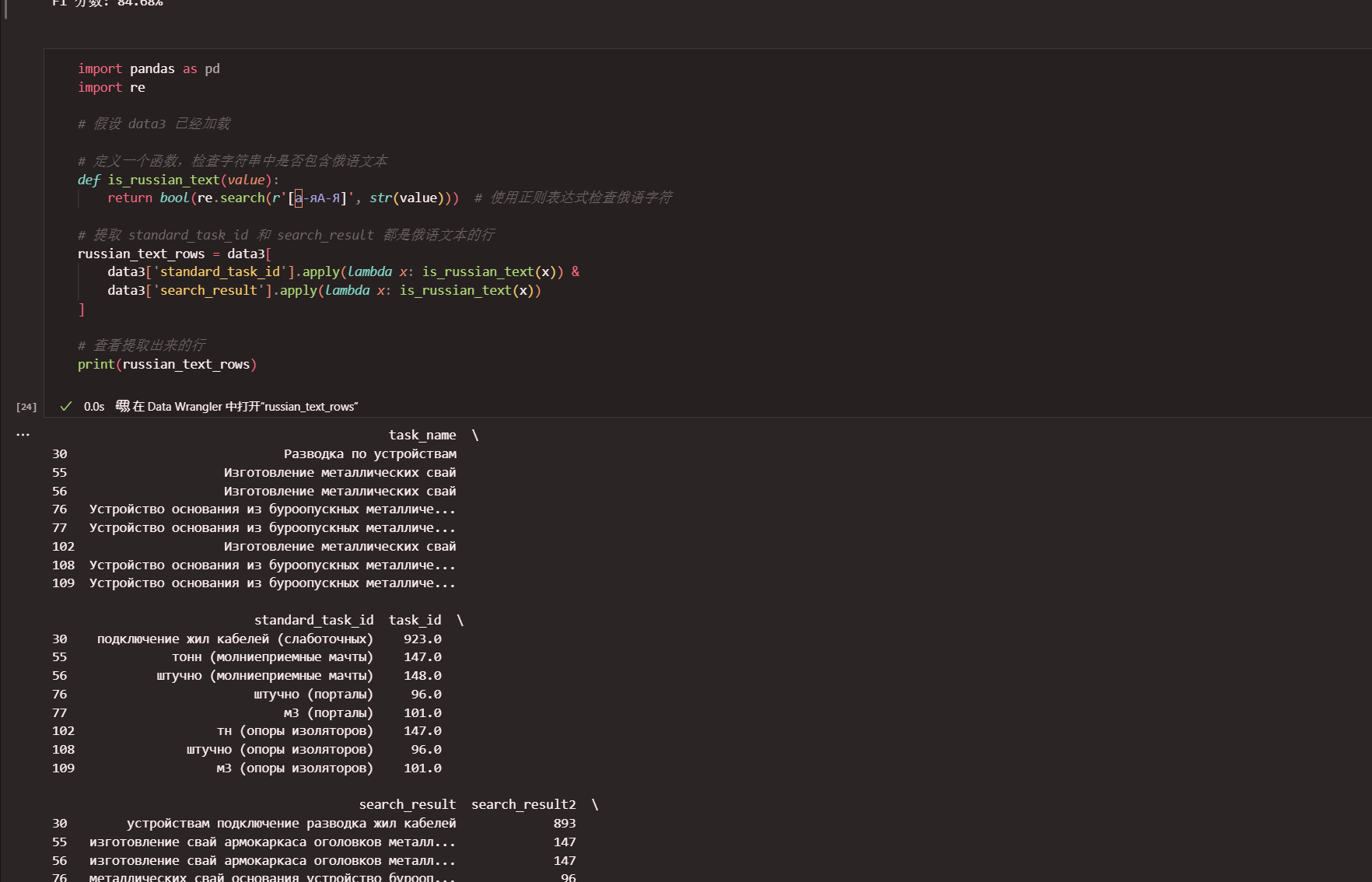
计算是否都为数字正确率输出3个指标

Calculate whether all the numbers are correct and output 3 indicators



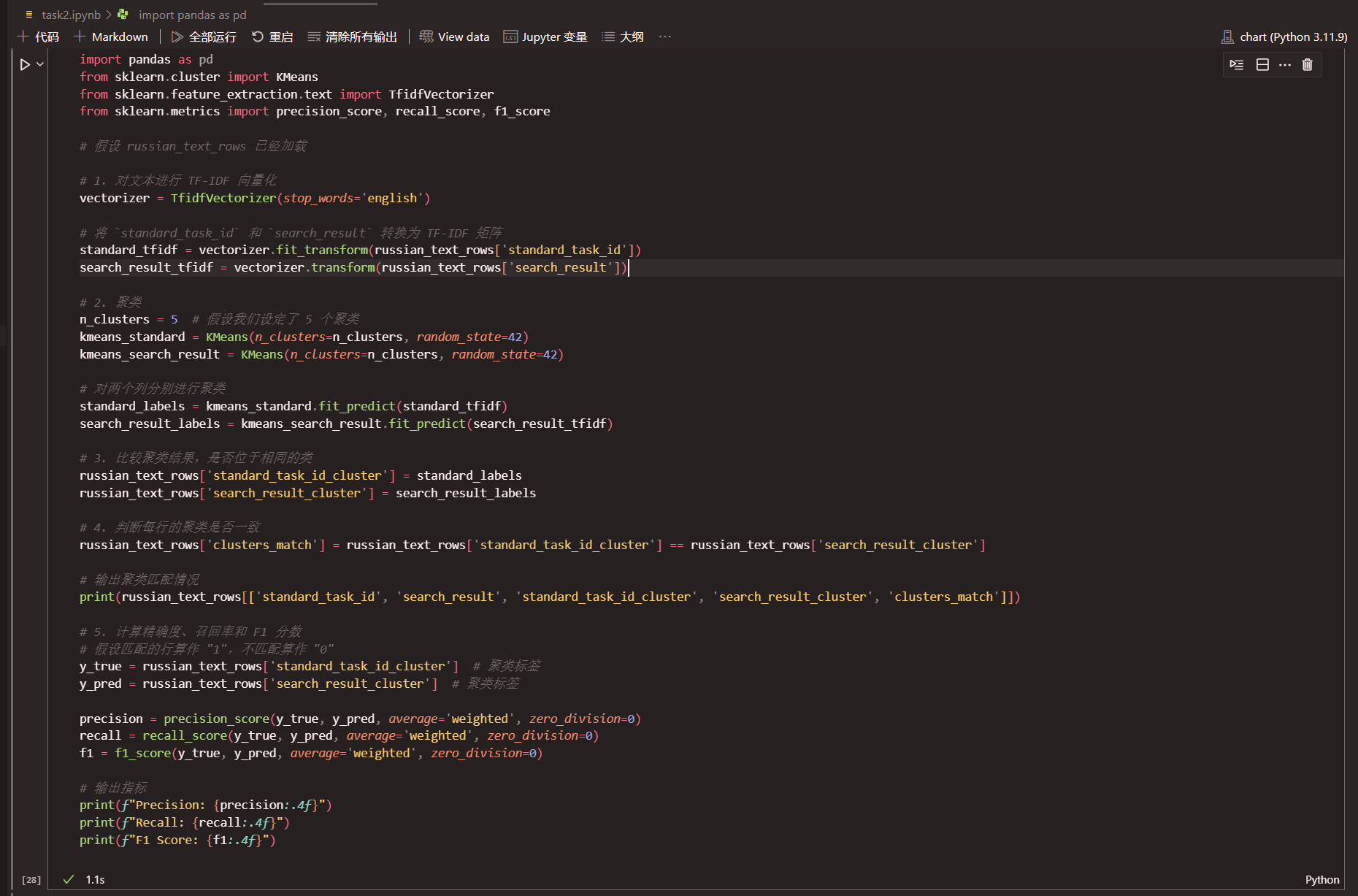
1. 计算纯数字正确率（所有行都包括）

7. Calculate the accuracy of pure numbers (all rows are included)



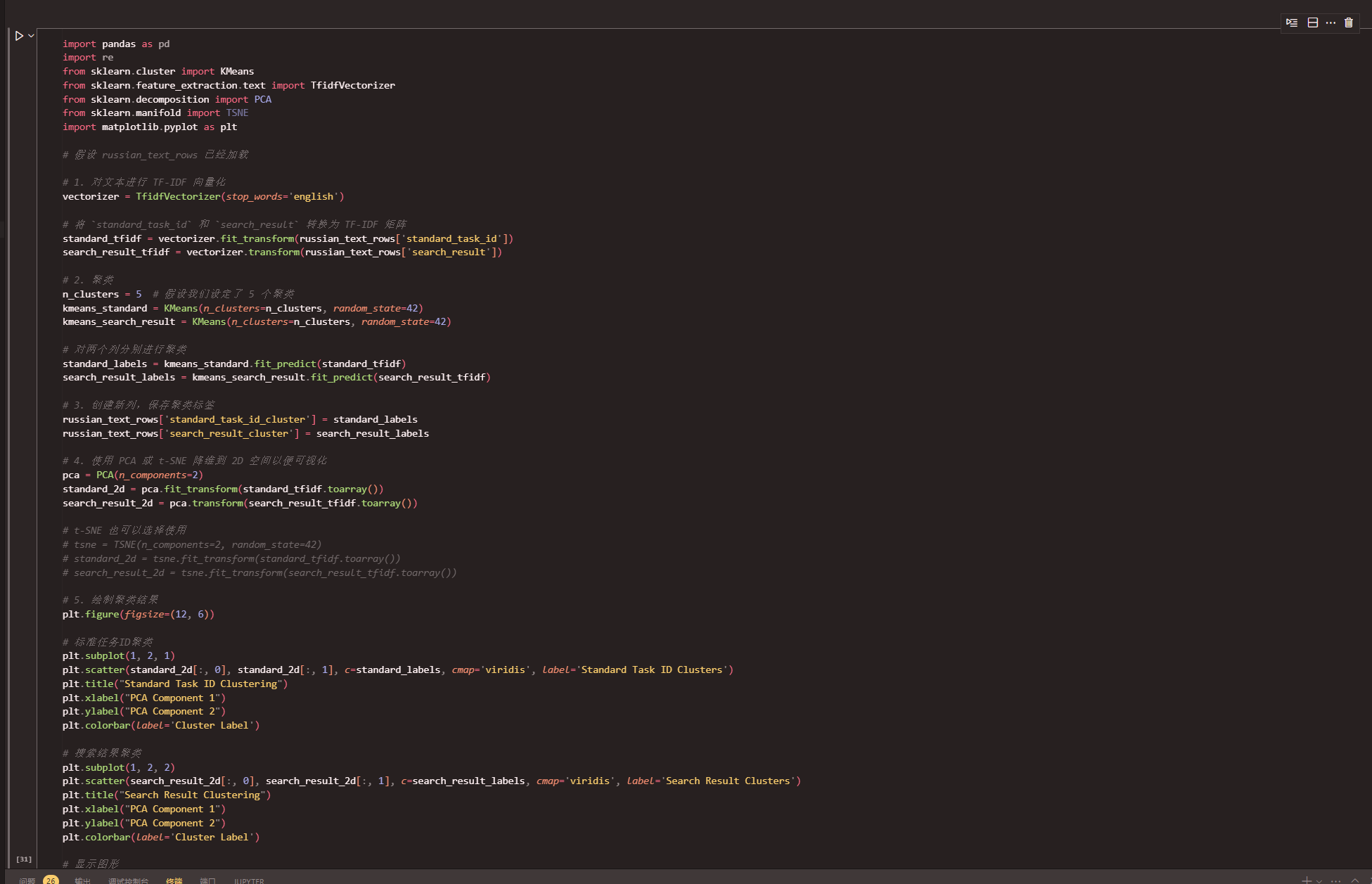
1. 都是类名调出来

8. All are called out by class name



9.分类指标计算；

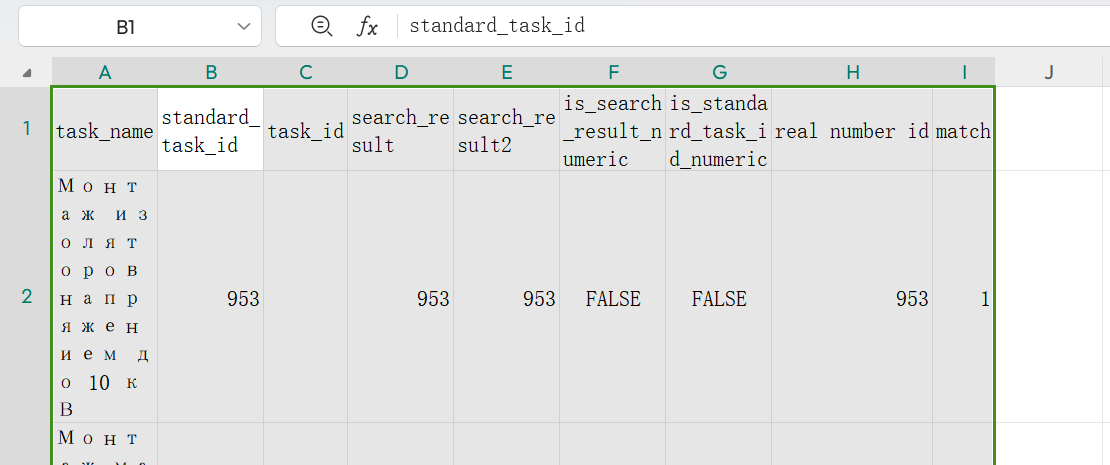
9.Calculation of classification indicators;



聚类结果画图

Plotting clustering results;

1. data3解释



Task\_name 初始

Stadard—— 验证集初始

Task\_id 验证集初始，代表找不到了返回类名task id

Sararch——result 返回预测结果

Sararch——result2 返回预测结果（纯数字版）

Is（2列） ———— 是否为数字的中间变量；

Match 数字是否匹配的中间变量

Task\_name initial

Stadard—— validation set initial

Task\_id validation set initial, represents the class name task id that cannot be found

Sararch——result returns the prediction result

Sararch——result2 returns the prediction result (pure digital version)

Is (2 columns) ———— whether it is a digital intermediate variable;

Match whether the digital matches the intermediate variable

1. russ——test 解释

都是类名结果返回

standard\_task\_id 真实结果

search\_result 我们聚类结果

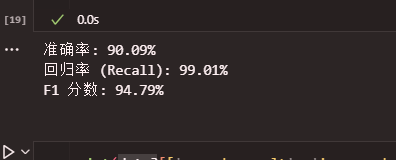
All the results are returned by class name

standard\_task\_id real result

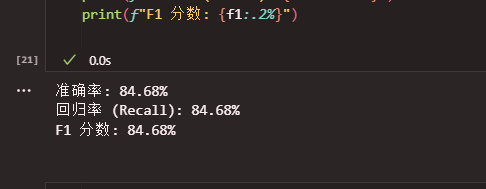
search\_result our clustering result

11.分别是否区分文本，类名判断结果；数字匹配结果，和文本聚类结果

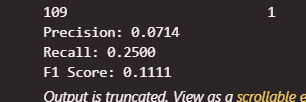
Whether to distinguish text, class name judgment results; digital matching results, and text clustering results



12.



13.



聚类分布图不推荐使用

Cluster distribution plots are not recommended.

