**好用的解析地址的cpca模块**

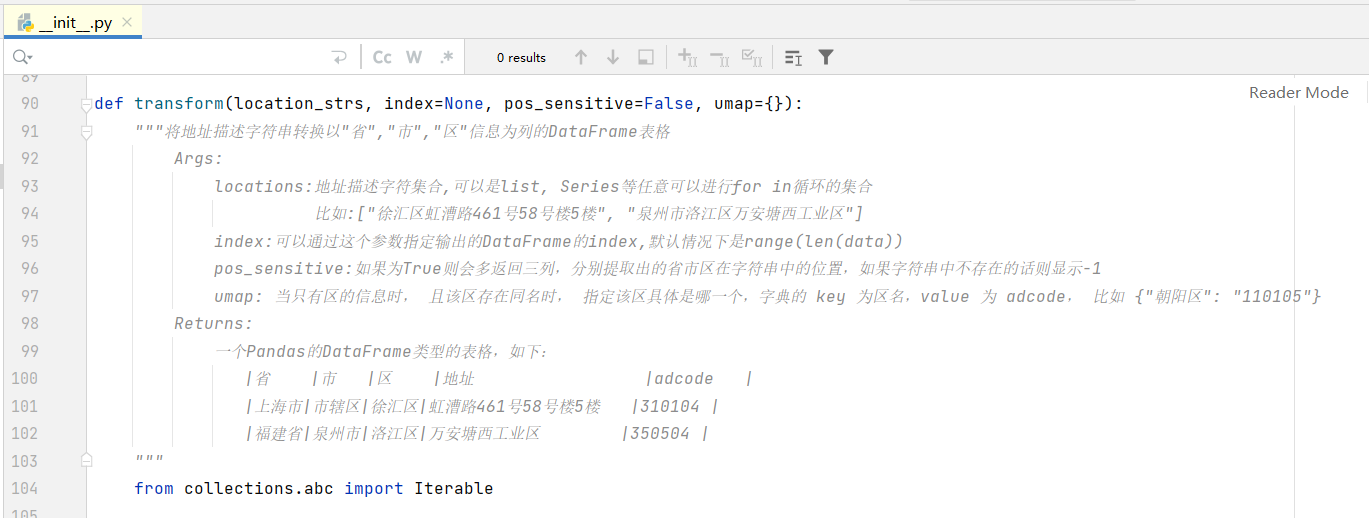
最近在参加行里建模比赛的时候，需要对两列关于地址的特征进行处理，相必大家也会遇到这样的需求，为此我这里记录下来，下次再使用的时候就信手拈来了。

1. **cpca模块介绍**

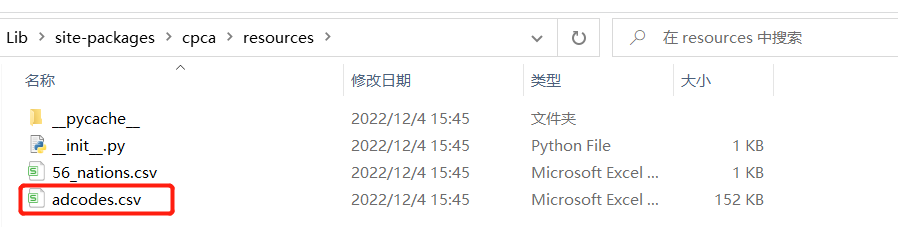
简单的说就是一个用于识别简体中文字符串中省，市和区的模块

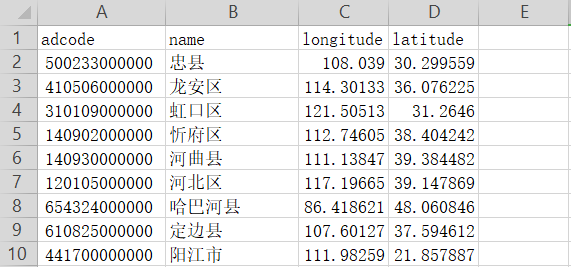
1. **主要函数**

在该模块中最为主要的函数为transform，其实现如下所示：



阅读代码可以发现，作者使用Lib\site-packages\cpca\resources\adcodes.csv下的编码文件：



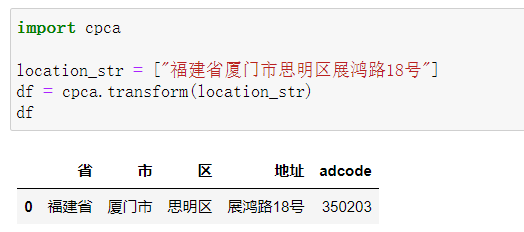


1. **案例**

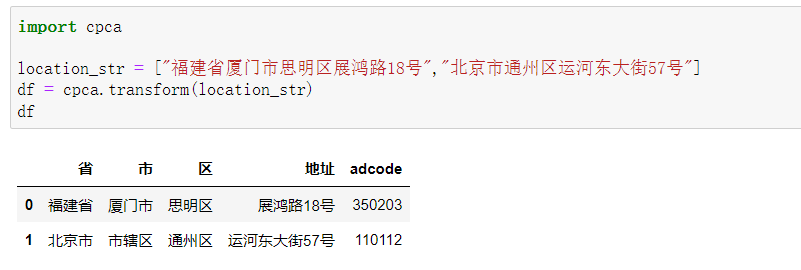
现在我们来实现一样实际的用法，假设我们现在需要提取一个字符串中的信息：

**福建省厦门市思明区展鸿路18号**

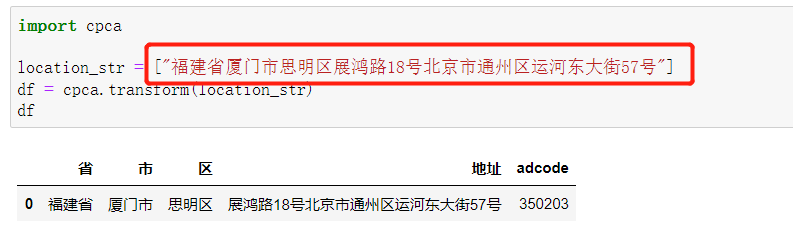
解析代码如下：



注意，这里的location\_str是一个列表结果的，如果有多个可以这样：

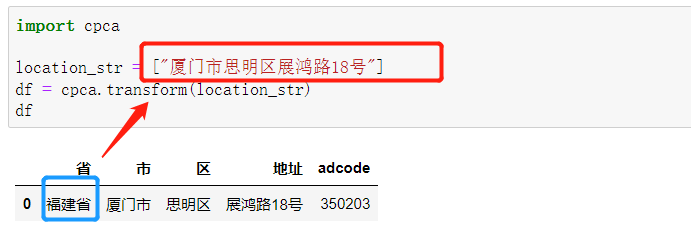


唯一需要注意的是，在进行解析的时候如果将两个省或者市，拼接在一起会出现这样的效果：



这个时候解析得到的省以第一个为主。

如果我们不提供省，看看其能不能匹配出来：



可以发现其可以正确的匹配出来，还是非常方便的。