**Cleaning of The Dataset**

**(Student Classification Dataset)**

**اعداد/ حمزة عباس اسحاق**

**1.1 مقدمة**

تحتوي مجموعة البيانات الخاصة بالطلاب على البيانات التي تمثل عواملًا مؤثرة في معدلاتهم ومستوياتهم الدراسية، لكن تلك المجموعة لم تكن نقية وجاهزة لبدء تحليلها والعمل عليها، من هذا المنطلق اضطررت الى تنقيتها.

**1.2 الأدوات المستخدمة**

* اللغة البرمجية بايثون.
* مكتبة بانداز.
* محرر اكواد جوبيتر او فيجيوال ستوديو كود.
  1. **الذي تم تنظيفه**

تم حذف جميع الصفوف التي توجد فيها قيم فارغة، أيضًا تم حذف الاسطر التي يوجد فيها قيم غير صحيحة وتم شرحها في هذا الجدول:

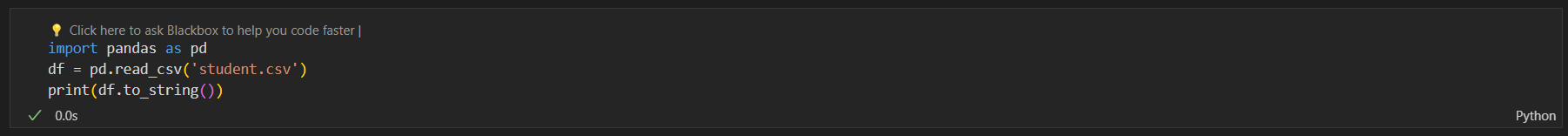
**جدول (1.1) توضيح ما تم تنظيفه**

|  |  |  |  |
| --- | --- | --- | --- |
| اسم العمود | رقم الصف | القيمة التي حُذف بسببها الصف | القيم الافتراضية للعمود |
| Student\_Age | 14 | 37 | From 18 to 25 |
| Transport | 84 | Car | Private, Bus |
| Sex | 137 | 96 | Female, Male |
| Notes | 60 | 6 | Yes, No |
| Attendance | 112 | 3 | Always, Never, or Sometimes |
| Listening\_in\_Class | 126 | 6 | Yes or No |

**1.4 الخطوات**

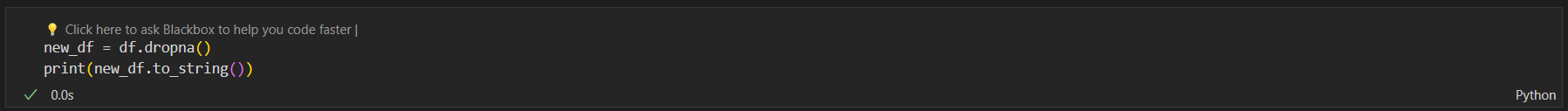
سلسة الخطوات البرمجية المتبعة حتى اكتمال التنظيف والحصول على النتيجة النهائية:

1. استدعاء المكتبة، وتحويل الملف الخاص بمجموعة البيانات الى داتا فريم، ثم طباعة المتغير.



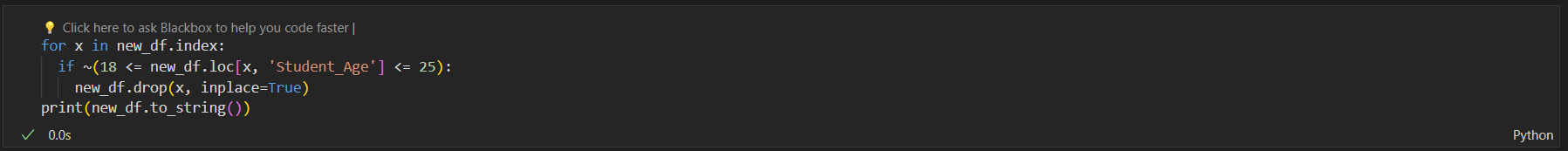
**صورة (1.1) الخطوة الأولى**

1. حذف جميع الصفوف التي تحتوي على قيم **فارغة** وطباعة الملف المحدث.



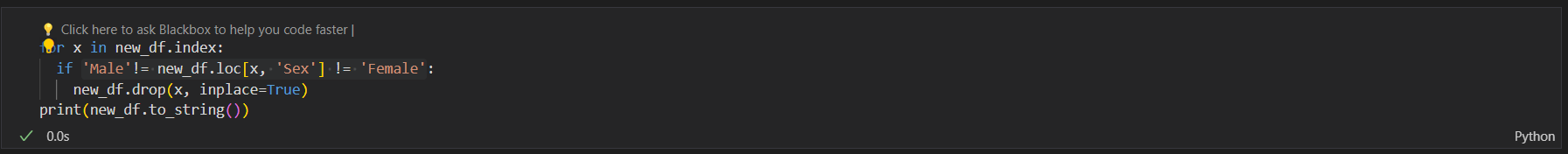
**صورة (1.2) الخطوة الثانية**

1. التأكد من كل القيم الخاصة بعمود عمر الطالب؛ حيث لا تقل عن **18** ولا تزيد عن **25**، وان وجد ما دون ذلك يتم حذف الصف، (تم حذف القيمة 37 في هذه الخطوة).



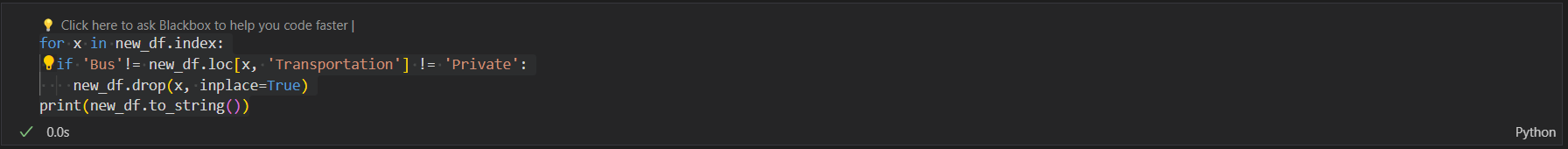
**صورة (1.3) الخطوة الثالثة**

1. التأكد من جميع القيم الخاصة بعمود جنس الطالب، وان لا تخرج القيم عن **ذكر** او **انثى**، واي بيانات اُخرى يتم حذف صفها، (تم حذف القيمة "96" في هذه الخطوة).



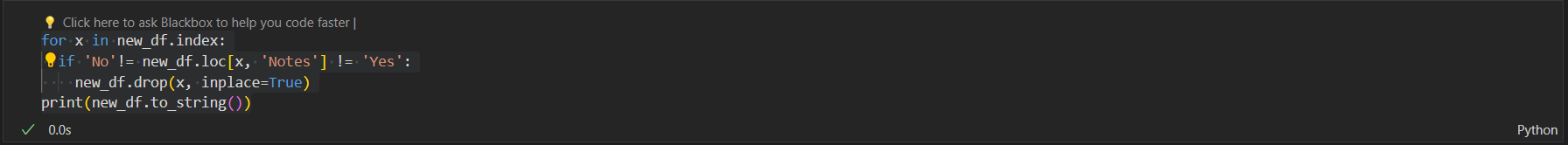
**صورة (1.4) الخطوة الرابعة**

1. التأكد من جميع القيم الخاصة بعمود طريقة تنقل الطالب، وان لا تخرج القيم عن **خاص** او **باص**، واي بيانات اُخرى يتم حذف صفها، (تم حذف القيمة "Car" في هذه الخطوة).



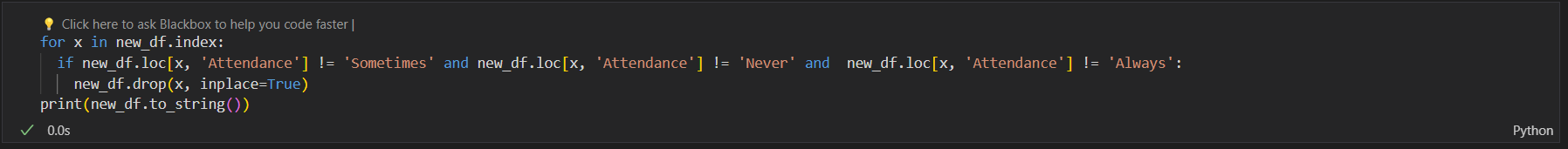
**صورة (1.5) الخطوة الخامسة**

1. التأكد من جميع القيم الخاصة بعمود هل يكتب الطالب ملاحظات، وان لا تخرج القيم عن **نعم** او **لا**، واي بيانات اُخرى يتم حذف صفها، (تم حذف القيمة "6" في هذه الخطوة).



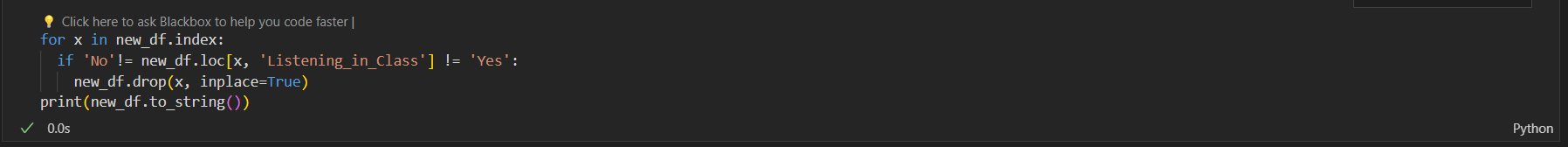
**صورة (1.6) الخطوة السادسة**

1. التأكد من جميع القيم الخاصة بعمود حضور الطالب، وان لا تخرج القيم عن **دائما** او **احيانًا** او **ابدا**، واي بيانات اُخرى يتم حذف صفها، (تم حذف القيمة "3" في هذه الخطوة).



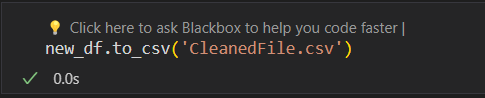
**صورة (1.7) الخطوة السابعة**

1. التأكد من جميع القيم الخاصة بعمود هل يركز الطالب اثناء الدروس، وان لا تخرج القيم عن **نعم** او **لا**، واي بيانات اُخرى يتم حذف صفها، (تم حذف القيمة "6" في هذه الخطوة).



**صورة (1.8) الخطوة الثامنة**

1. تحويل الداتا فريم المنقى الى ملف سي اس في.



**صورة (1.9) الخطوة التاسعة**