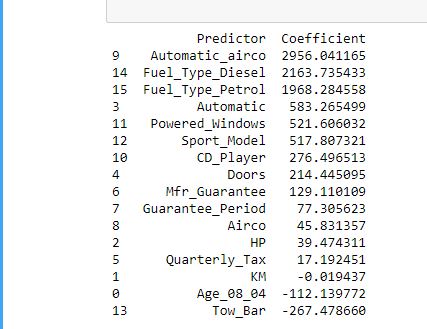
**تکلیف سوم درس مبانی داده کاوی**

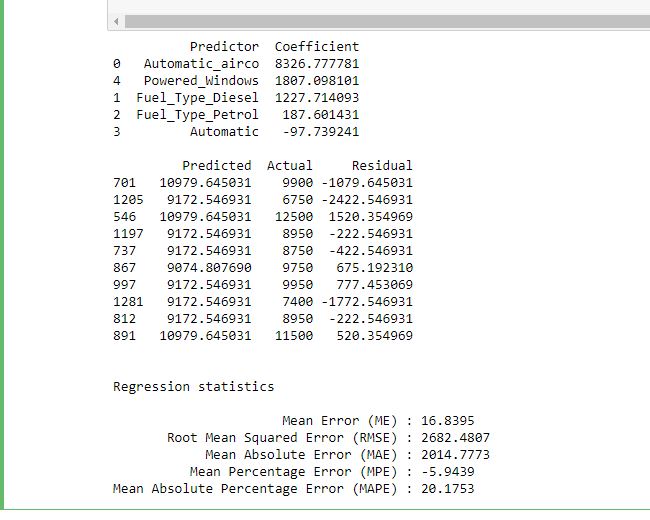
**امید رئیسی (۹۶۲۱۱۶۰۰۱۵)**

**۱)**

**الف) همان‌طور که از شکل زیر پیداست متغیر‌هایی که دارای بیشترین ضرایب می‌باشند بیشترین تأثیر را در پیشبینی قیمت اتوموبیل دارند. چهار متغیری که بیشترین تأثیر را دارند عبارت‌اند از: Automatic\_airco, Fuel\_Type, Automatic, Powered\_Windows**

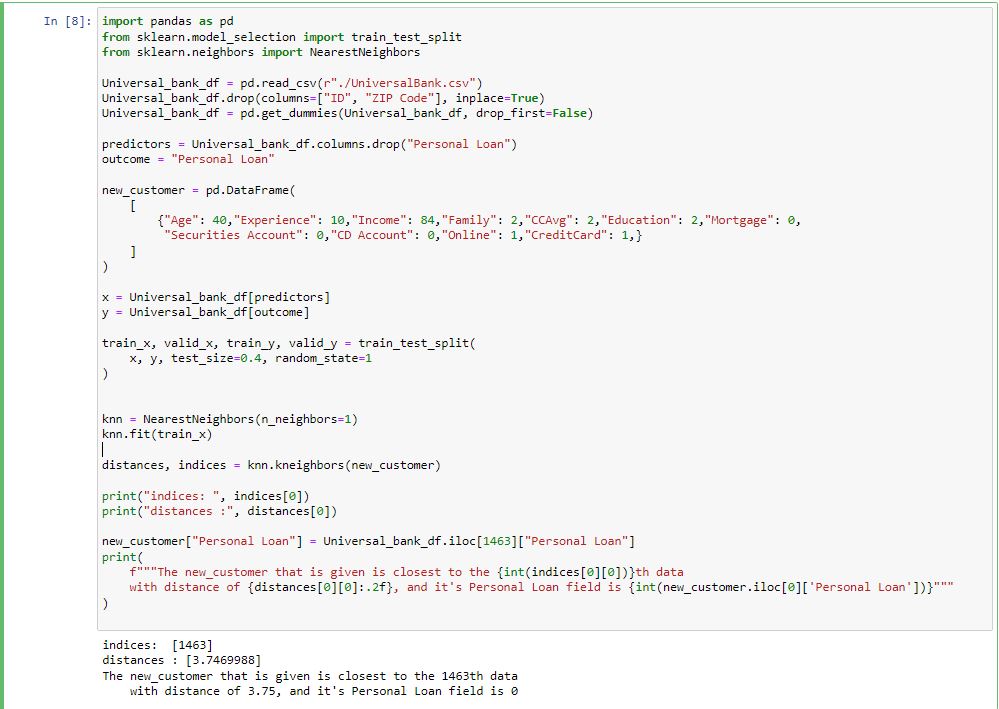
****

**ب) با استفاده از متغیر‌هایی که در بالا نام بردیم دوباره مدل رگرسیون خود را می‌سازیم و سپس با استفاده از داده‌های اعتبارسنجی مدل خود را ارزیابی می‌کنیم. (به دلیل کاهش تعداد متغیر‌ها میزان خطا بالا رفته است.)**

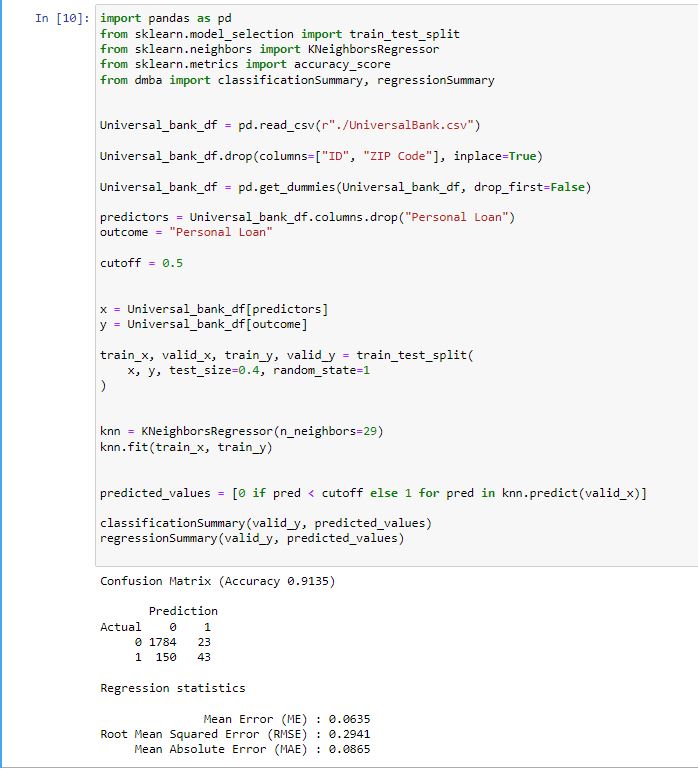
****

**۲)**

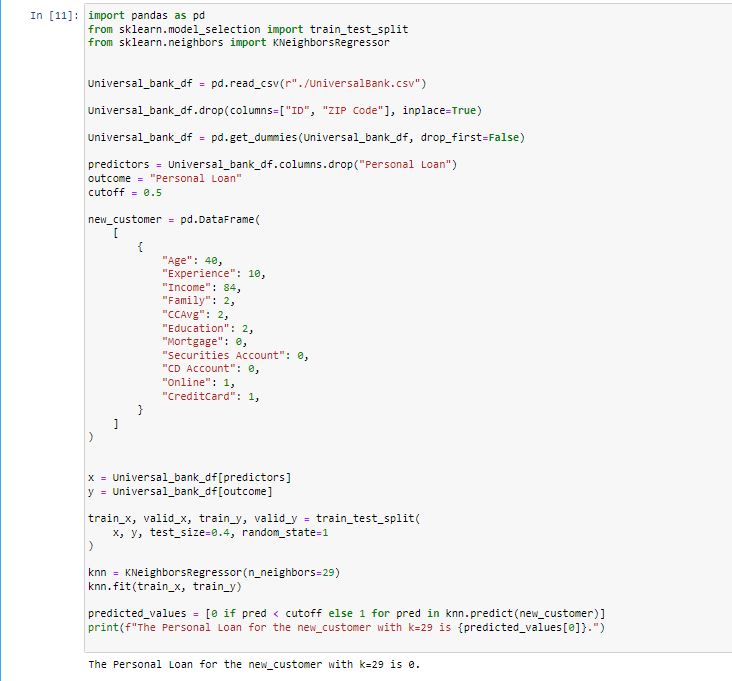
**الف) با توجه به اینکه k=1 می‌باشد پس مقدار Personal Loan نزدیک‌ترین همسایه برای داده جدید نیز انتخاب می‌شود که مقدار آن 0 می‌باشد و این یعنی مشتری جدید هم پیشنهاد بانک را نخواد پذیرفت.**

****

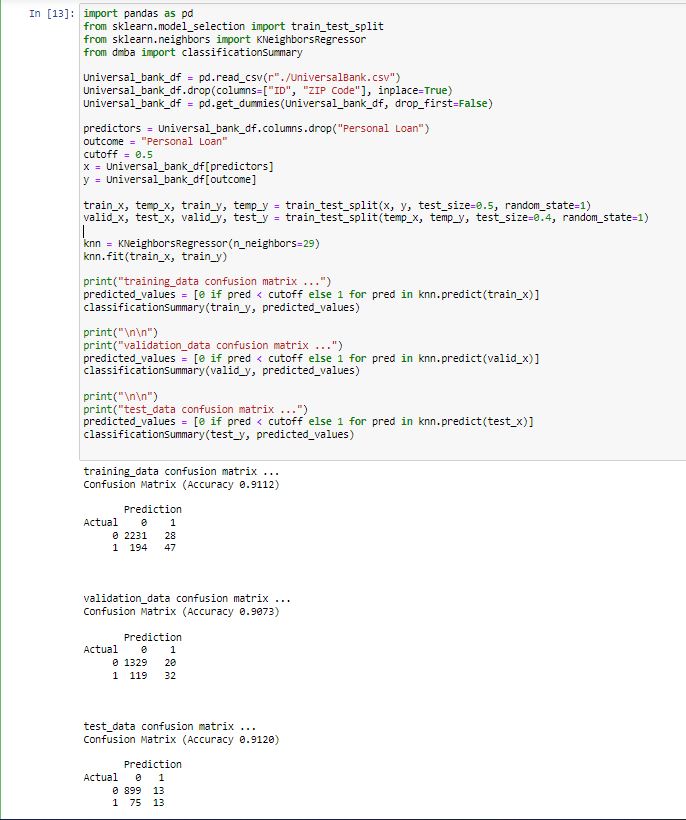
**ب) برای پیدا کردن این موازنه باید مقدار بهترین K را برای این مسأله پیدا کنیم که طبق عکس مقدار آن برابر با 29 می‌باشد. (k از 1 تا 50 فرض شده است.)**

**ج) **

**د) مشتری جدید با مشخصات داده شده پیشنهاد بانک را نخواهد پذیرفت.**

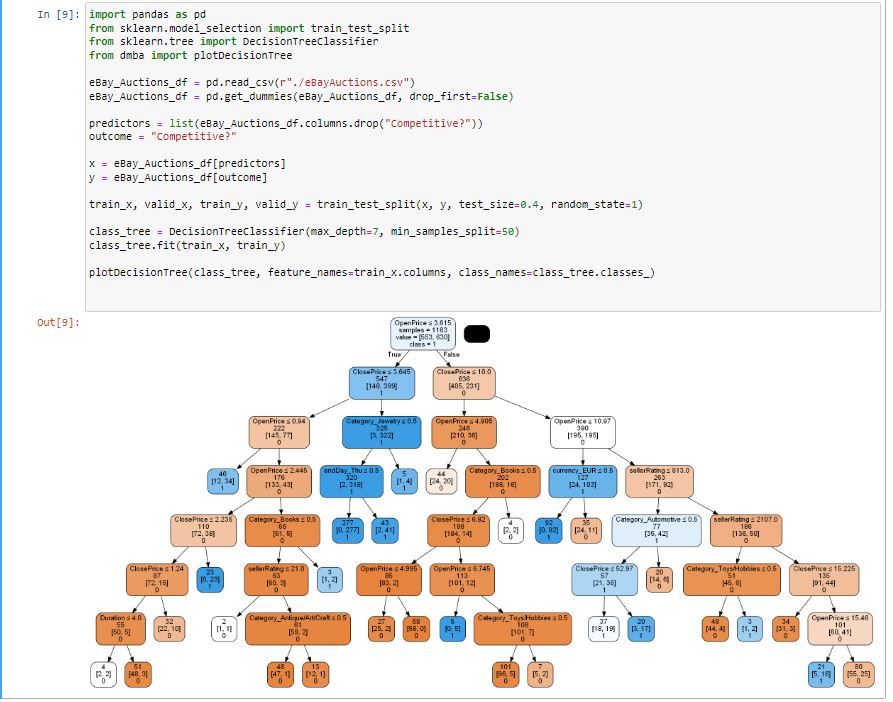
****

**ه) با توجه به اینکه تعداد داده‌های آموزشی کاهش یافته‌اند میزان دقت در داده‌‌های آموزشی نیز کاهش یافته است، اما به طور کلی چون با هر بار پیشبینی داده‌های جدیدی به اضافه شده و دقت در همسایگی افزایش می‌یابد میزان خطا در داده‌های آزمایشی از میزان خطا در داده‌های آموزشی و اعتبار‌سنجی کمتر است و این یک ویژگی کلی از الگوریتم KNN است به این صورت که هر چه پیشبینی‌ها افزایش یابد میزان دقت نیز افزایش می‌یابد.**

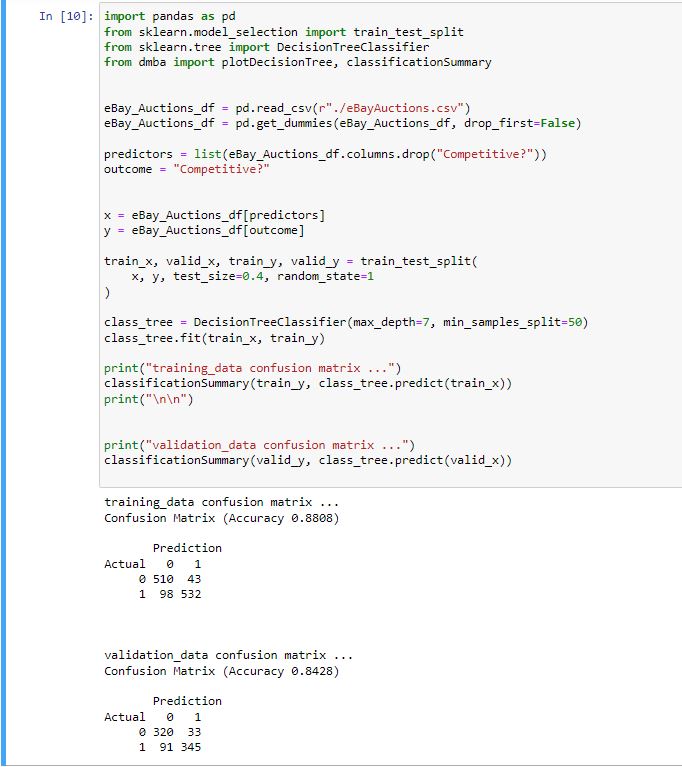
****

**۳)**

**الف)**

****

**ب) طبق شکل زیر میزان دقت برای داده‌‌های آموزشی و اعتبار‌سنجی به ترتیب 88 و 84 درصد می‌باشد که نرخ قابل قبولی نیست و می‌توان درختی با عملکرد بهتری ساخت.**

****

IF (OpenPrice <= 0.94) AND (ClosePrice <=3.645) THEN Class=0

IF (OpenPrice <= 2.445) AND (OpenPrice > 0.94) AND (ClosePrice <= 1.24) AND (Duration > 4) THEN Class=0

IF (OpenPrice <= 2.445) AND (OpenPrice > 0.94)AND (ClosePrice <= 2.235) AND (ClosePrice > 1.24) THEN Class=0

IF (OpenPrice <= 2.445) AND (OpenPrice > 0.94)AND (ClosePrice <= 3.645) AND (ClosePrice > 2.235) THEN Class=1

IF (OpenPrice <=3.615) AND (OpenPrice > 2.445)AND (ClosePrice <= 3.645) AND (Category\_Books <= 0.5) AND (sellerRating >21.0) AND (endDay\_fri <= 0.5) THEN Class=0

IF (OpenPrice <=3.615) AND (OpenPrice > 2.445)AND (ClosePrice <= 3.645) AND (Category\_Books <= 0.5) AND (sellerRating >21.0) AND (endDay\_fri > 0.5) THEN Class=0

IF (OpenPrice <=3.615) AND (OpenPrice > 2.445)AND (ClosePrice <= 3.645) AND (Category\_Books > 0.5) THEN Class=1

IF (OpenPrice <=3.615) AND (ClosePrice > 3.645) AND (Category\_Jewerly <= 0.5) AND (endDay\_Thu <= 0.5) THEN Class=1

IF (OpenPrice <=3.615) AND (ClosePrice > 3.645) AND (Category\_Jewerly <= 0.5) AND (endDay\_Thu > 0.5) THEN Class=1

IF (OpenPrice <=3.615) AND (ClosePrice > 3.645) AND (Category\_Jewerly > 0.5) THEN Class=1

IF (OpenPrice > 3.615) AND (ClosePrice <= 6.82) AND (Category\_Books <=0.5) THEN Class=0

IF (OpenPrice > 3.615) AND (ClosePrice <= 10.0) AND (Category\_Books <=0.5) AND (Close\_Price > 6.82) AND (Open\_Price <= 6.745)THEN Class=1

IF (OpenPrice > 3.615) AND (ClosePrice <= 10.0) AND (Category\_Books <=0.5) AND (Close\_Price > 6.82) THEN Class=0

IF (OpenPrice > 3.615) AND (ClosePrice > 10.0) AND (OpenPrice <= 10.97) AND (currency\_EUR <= 0.5) THEN Class=1

IF (OpenPrice > 3.615) AND (ClosePrice > 10.0) AND (OpenPrice <= 10.97) AND (currency\_EUR > 0.5) THEN Class=0

IF (OpenPrice > 3.615) AND (ClosePrice > 10.0) AND (sellerRating <= 813.0) AND (Category\_Automative <=0.5) THEN Class=1

IF (OpenPrice > 3.615) AND (ClosePrice > 10.0) AND (sellerRating <= 813.0) AND (Category\_Automative >0.5) THEN Class=0

IF (OpenPrice > 3.615) AND (ClosePrice > 10.0) AND (sellerRating > 813.0) AND (sellerRating <= 2107.0) AND (Category\_Toys <= 0.5) THEN Class=0

IF (OpenPrice > 3.615) AND (ClosePrice > 10.0) AND (sellerRating > 813.0) AND (sellerRating <= 2107.0) AND (Category\_Toys > 0.5) THEN Class=1

IF (OpenPrice > 3.615) AND (ClosePrice > 10.0) AND (sellerRating > 813.0) AND (ClosePrice <= 15.225) THEN Class=0

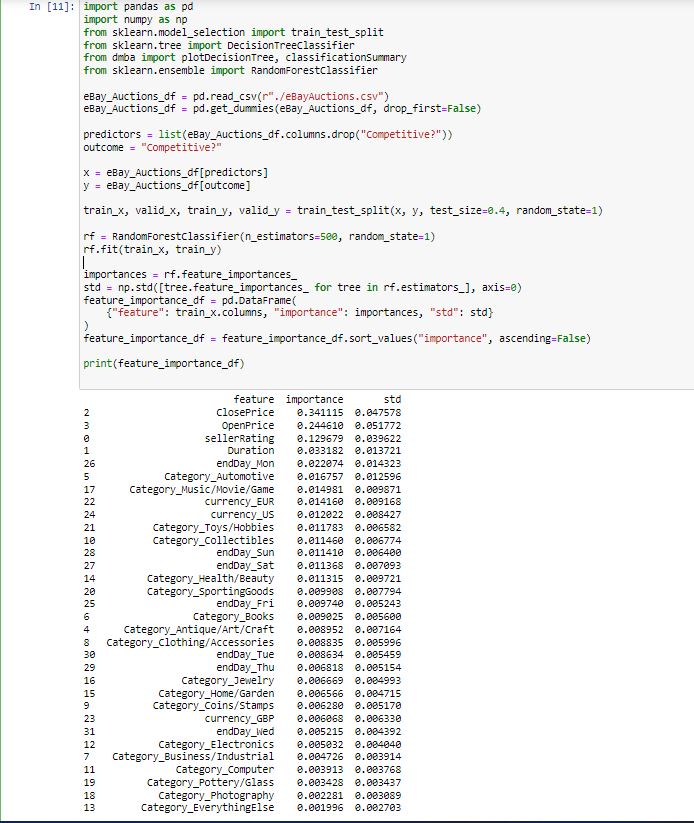
IF (OpenPrice > 3.615) AND (ClosePrice > 10.0) AND (sellerRating > 813.0) AND (OpenPrice <=15.46) THEN Class=1

IF (OpenPrice > 3.615) AND (ClosePrice > 10.0) AND (sellerRating > 813.0) THEN Class=0

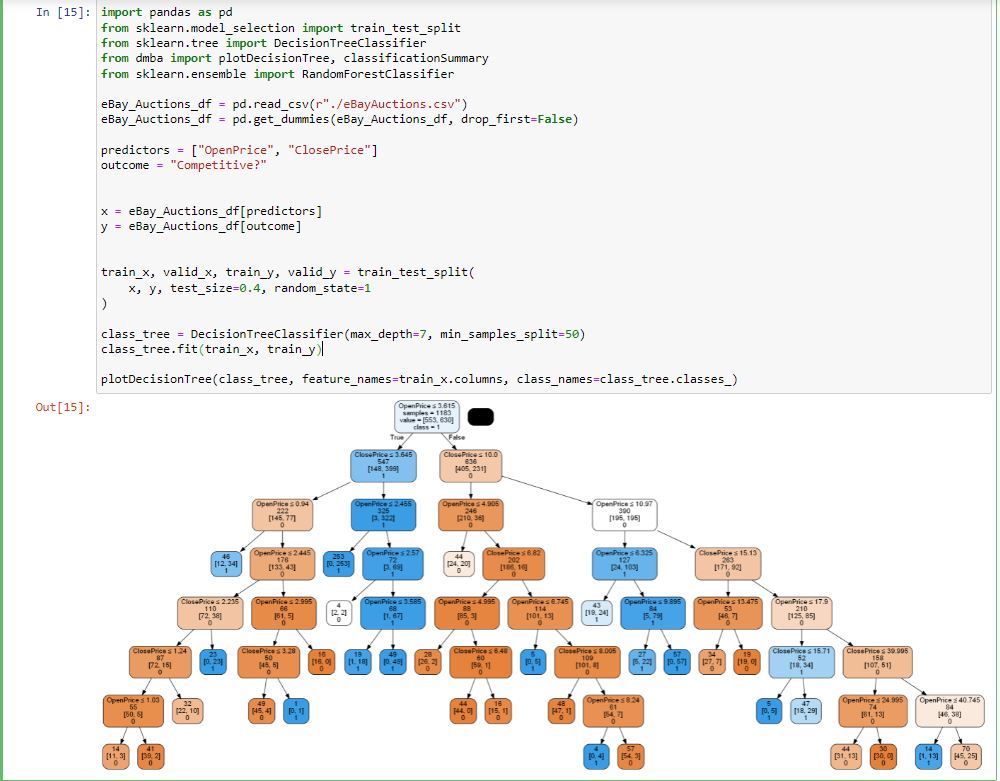
**ج) با توجه به درخت چند قانون زیر وجود دارد:**

* **اگر قیمت پایانی کمتر از 3.645 و قیمت شروع بیشتر از 0.94 و کمتر از3.615 باشد اکثر کالا‌های حراجی رقابتی نبوده‌اند.**
* **اگر قیمت شروع کمتر از 3.615 و قیمت نهایی بیشتر از 3.645 باشد آنگاه تقریبا تمام کالا‌ها رقابتی بوده‌اند.**
* **اگر قیمت شروع بیشتر از 3.615 و قیمت نهایی کمتر از 10 باشد آنگاه تقریبا تمام کالا‌ها رقابتی نبوده‌اند.**

**د) با استفاده از الگوریتم درخت‌‌های تصادفی متغیر‌های که بیشترین تفکیک را ایجاد می‌کنند شناسایی کرده و سپس با استفاده از آن متغیر‌ها دوباره درخت تصمیم خود را ایجاد می‌کنیم.**

****

**همانطور که می‌بینیم متغیر‌های ClosedPrice, OpenPrice بیشترین اهمیت را دارند.**

****

def tree(OpenPrice, ClosePrice):

*if* OpenPrice <= 3.615000009536743:

*if* ClosePrice <= 3.6449999809265137:

*if* OpenPrice <= 0.9399999976158142:

*return* [[12. 34.]]

*else*:  *# if OpenPrice > 0.9399999976158142*

*if* OpenPrice <= 2.4450000524520874:

*if* ClosePrice <= 2.2350000143051147:

*if* ClosePrice <= 1.2400000095367432:

*if* OpenPrice <= 1.0300000309944153:

*return* [[11.  3.]]

*else*:  *# if OpenPrice > 1.0300000309944153*

*return* [[39.  2.]]

*else*:  *# if ClosePrice > 1.2400000095367432*

*return* [[22. 10.]]

*else*:  *# if ClosePrice > 2.2350000143051147*

*return* [[ 0. 23.]]

*else*:  *# if OpenPrice > 2.4450000524520874*

*if* OpenPrice <= 2.9950000047683716:

*if* ClosePrice <= 3.2799999713897705:

*return* [[45.  4.]]

*else*:  *# if ClosePrice > 3.2799999713897705*

*return* [[0. 1.]]

*else*:  *# if OpenPrice > 2.9950000047683716*

*return* [[16.  0.]]

*else*:  *# if ClosePrice > 3.6449999809265137*

*if* OpenPrice <= 2.4550000429153442:

*return* [[  0. 253.]]

*else*:  *# if OpenPrice > 2.4550000429153442*

*if* OpenPrice <= 2.5700000524520874:

*return* [[2. 2.]]

*else*:  *# if OpenPrice > 2.5700000524520874*

*if* OpenPrice <= 3.584999918937683:

*return* [[ 1. 18.]]

*else*:  *# if OpenPrice > 3.584999918937683*

*return* [[ 0. 49.]]

*else*:  *# if OpenPrice > 3.615000009536743*

*if* ClosePrice <= 10.0:

*if* OpenPrice <= 4.9049999713897705:

*return* [[24. 20.]]

*else*:  *# if OpenPrice > 4.9049999713897705*

*if* ClosePrice <= 6.819999933242798:

*if* OpenPrice <= 4.994999885559082:

*return* [[26.  2.]]

*else*:  *# if OpenPrice > 4.994999885559082*

*if* ClosePrice <= 6.479999780654907:

*return* [[44.  0.]]

*else*:  *# if ClosePrice > 6.479999780654907*

*return* [[15.  1.]]

*else*:  *# if ClosePrice > 6.819999933242798*

*if* OpenPrice <= 6.744999885559082:

*return* [[0. 5.]]

*else*:  *# if OpenPrice > 6.744999885559082*

*if* ClosePrice <= 8.005000114440918:

*return* [[47.  1.]]

*else*:  *# if ClosePrice > 8.005000114440918*

*if* OpenPrice <= 8.239999771118164:

*return* [[0. 4.]]

*else*:  *# if OpenPrice > 8.239999771118164*

*return* [[54.  3.]]

*else*:  *# if ClosePrice > 10.0*

*if* OpenPrice <= 10.96999979019165:

*if* OpenPrice <= 6.325000047683716:

*return* [[19. 24.]]

*else*:  *# if OpenPrice > 6.325000047683716*

*if* OpenPrice <= 9.894999980926514:

*return* [[ 5. 22.]]

*else*:  *# if OpenPrice > 9.894999980926514*

*return* [[ 0. 57.]]

*else*:  *# if OpenPrice > 10.96999979019165*

*if* ClosePrice <= 15.130000114440918:

*if* OpenPrice <= 13.474999904632568:

*return* [[27.  7.]]

*else*:  *# if OpenPrice > 13.474999904632568*

*return* [[19.  0.]]

*else*:  *# if ClosePrice > 15.130000114440918*

*if* OpenPrice <= 17.90000057220459:

*if* ClosePrice <= 15.710000038146973:

*return* [[0. 5.]]

*else*:  *# if ClosePrice > 15.710000038146973*

*return* [[18. 29.]]

*else*:  *# if OpenPrice > 17.90000057220459*

*if* ClosePrice <= 39.9950008392334:

*if* OpenPrice <= 24.994999885559082:

*return* [[31. 13.]]

*else*:  *# if OpenPrice > 24.994999885559082*

*return* [[30.  0.]]

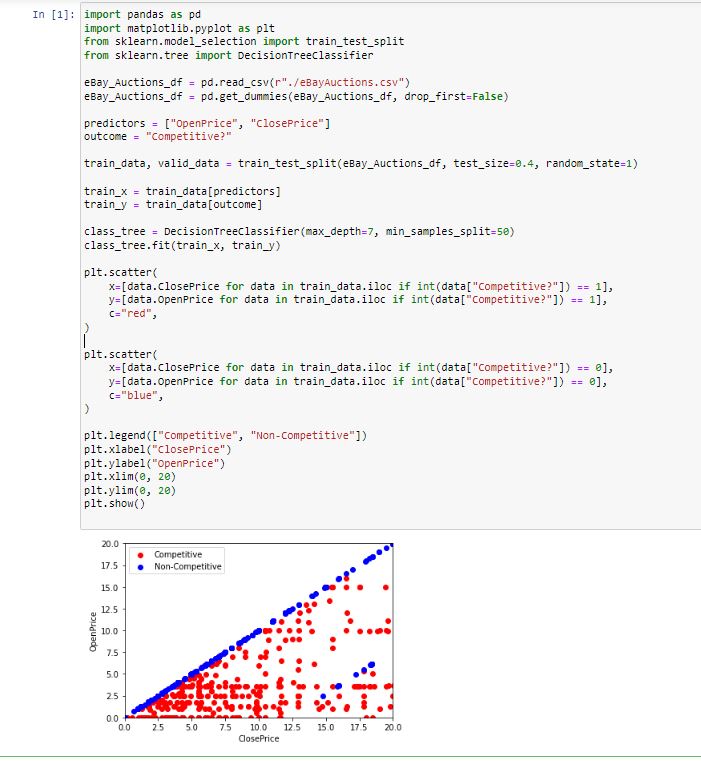
*else*:  *# if ClosePrice > 39.9950008392334*

*if* OpenPrice <= 40.7450008392334:

*return* [[ 1. 13.]]

*else*:  *# if OpenPrice > 40.7450008392334*

*return* [[45. 25.]]

**ه) **