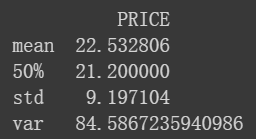
AI二 B10905001 張啟玥 巨量資料分析 作業一

第一題

請計算此資料中房價的平均數、中位數、變異數與標準差



平均數:22.532806

中位數:21.200000

變異數:84.586723

標準差:9.197104

第二題

請計算此資料中房價超過20(PRICE TYPE=2)與低於20(PRICE TYPE=1)的個數及比例



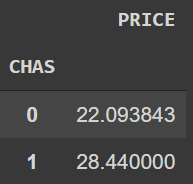
房價低於20(PRICE TYPE=1)的個數為215 比例為0.424901

房價超過20(PRICE TYPE=2)的個數為291 比例為0.575099

* 代表房價有一半以上超過20

第三題

請計算是否臨近Carles River(CHAS=0 or CHAS=1)地區房價之差異



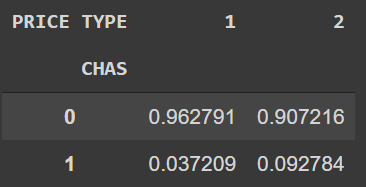
不臨近Carles River(CHAS=0)的平均房價為22.093843

臨近Carles River(CHAS=1)的平均房價為28.44

* 代表臨近Carles River的房價平均下來比較高

第四題

請計算是否臨近Carles River(CHAS=0 or CHAS=1)地區房價高低之差異



* 以房價高低的角度來看

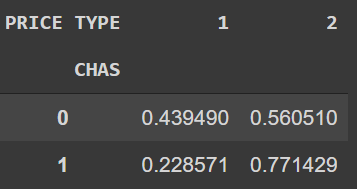
房價低於20(PRICE TYPE=1):

0.962791不臨近Carles River(CHAS=0)，0.037209臨近Carles River(CHAS=1)

房價高於20(PRICE TYPE=2):

0.907216不臨近Carles River(CHAS=0)，0.092784臨近Carles River(CHAS=1)

* 代表無論房價高低，不臨近Carles River(CHAS=0)的房子比較多



* 以是否臨近Carles River的角度來看

臨近Carles River(CHAS=1):

0.439490房價低於20(PRICE TYPE=1)，0.560512房價高於20(PRICE TYPE=2)

不臨近Carles River(CHAS=0):

0.2228571房價低於20(PRICE TYPE=1)，0.771429房價高於20(PRICE TYPE=2)

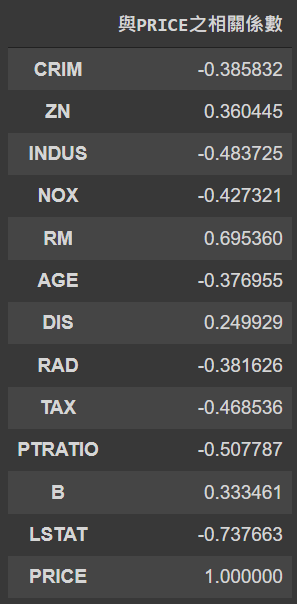
* 代表無論是否臨近Carles River，房價高於20(PRICE TYPE=2)的房子比較多

第五題

請分析哪些量化因素與房價具有較高的關聯性

須先扣除掉類別，留下量化因素

被扣除掉的類別有三個:Unnamed: 0、CHAS、PRICE TYPE，剩下的為量化因素



有正相關:ZN、RM、DIS、B

有負相關:CRIM、INDUS、NOX、AGE、RAD、TAX、PTRATIO、LSTAT

自己和自己比較的相關係數為1:PRICE

其中相關係數的絕對值大於0.5的較有參考價值:RM(0.695360)、PTRATIO(-0.507787)、LSTAT(-0.737663)

* 代表每戶平均房間數(RM)和房價有高的正向關聯性
* 代表人口地位較低的百分比(LSTAT)和房價有較高的負向關聯性