**Lab 6.2 KMeans Clustering (3/6/2022)**

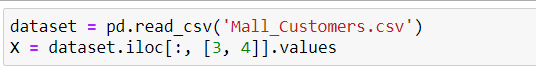
**ລະຫັດນັກສຶກສາ: 205Q0010.19**

**ຊື່ ແລະ ນາມສະກຸນ: ທ້າວ ນູຊົ່ວ ເຮີ**

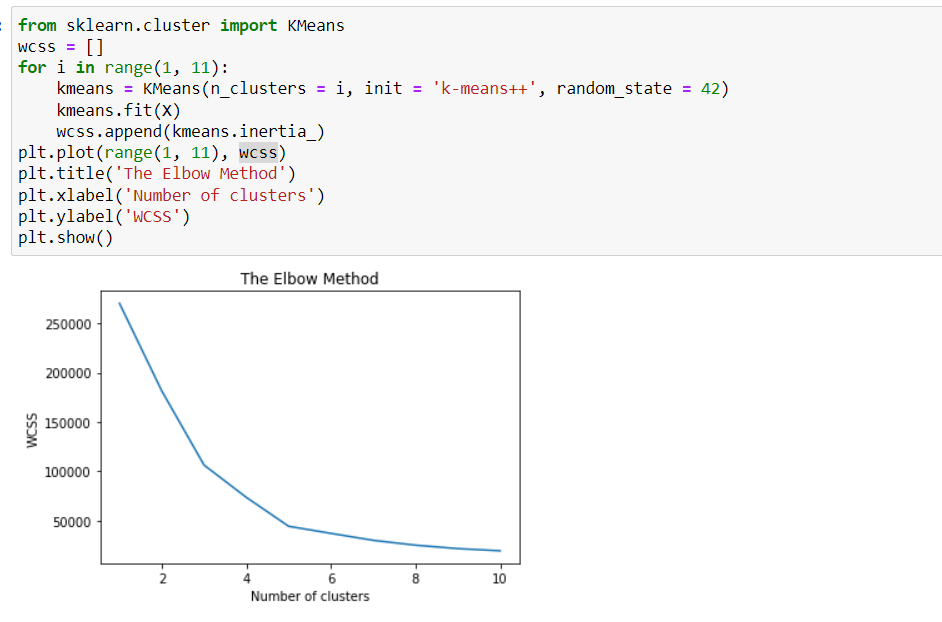
ຈົ່ງຕອບຄຳຖາມຕໍ່ໄປນີ້ໃຫ້ສຳເລັດດ້ວຍການນຳໃຊ້ຄຳສັ່ງຂອງ Python:

**ພາກທີ 1**

* 1. ຈາກຊຸດຂໍ້ມູນ Mall\_Customers.csv. ຈົ່ງທຳການເລືອກຄຸນລັກສະນະ (Features): Age ແລະ Annual Income (k$).



1.2 ຈົ່ງທຳການກຳນົດຄ່າຂອງ K ດ້ວຍການນຳໃຊ້ sklearn.cluster ແລະ Kmeans ເພື່ອຄິດໄລ່ຄ່າ wcss ພ້ອມ ແຕ້ມເສັ້ນສະແດງຂອງ Elbow Method



1.3. ຈົ່ງສ້າງແບບຈຳຮອງ Kmeans ເພື່ອແບ່ງກຸ່ມຊຸດຂໍ້ມູນດ້ວຍຕາມຄ່າຂອງ K

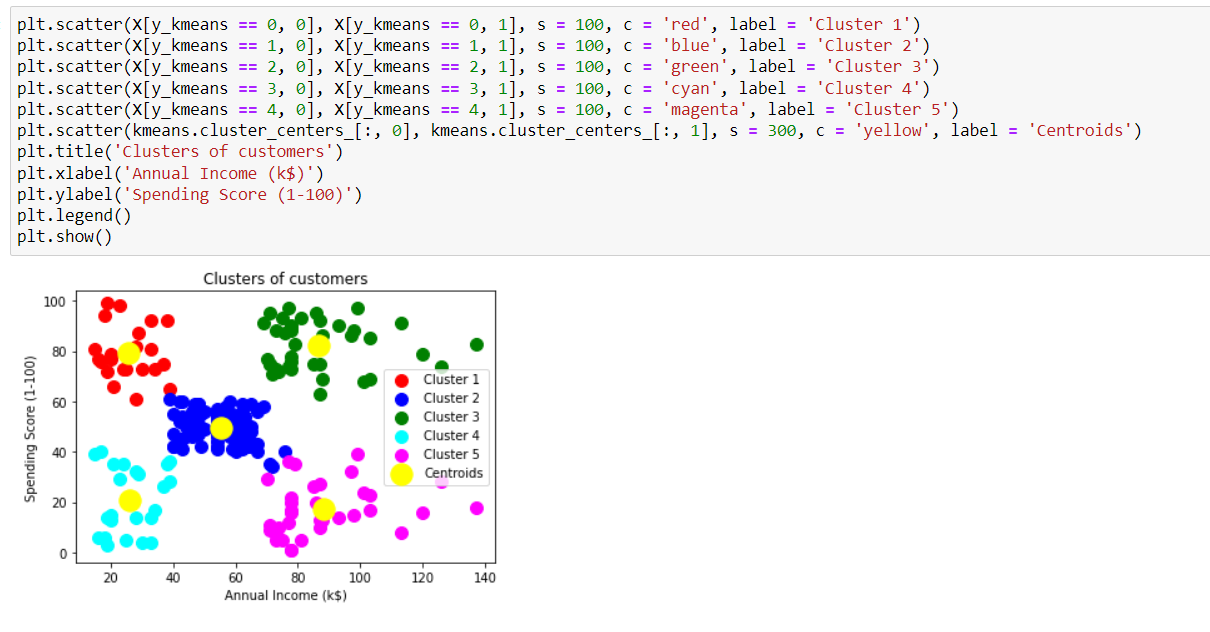
ໃນຂໍ້1.2.

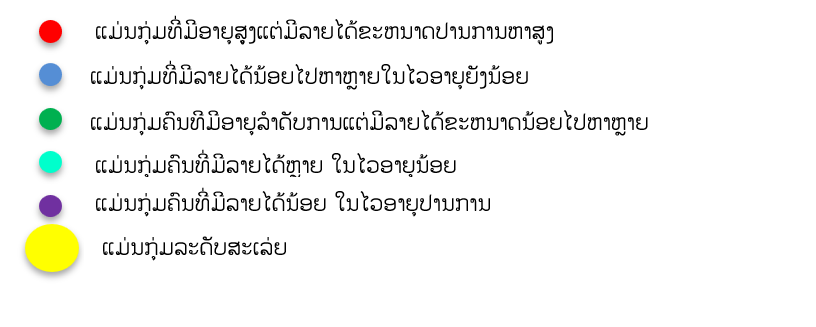


1.4 ຈົ່ງແບ່ງກຸ່ມຊຸດຂໍ້ມູນໃນຂໍ້ທີ 1.1 ດ້ວຍແບບຈຳຮອງ (Algorithm) ຂໍ້ 1.3



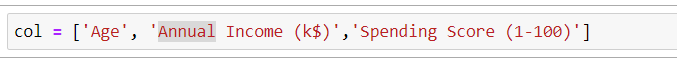
1.5. ຈົ່ງທຳການສຳຫຼວດ (visualizing) ກຸ່ມຂໍ້ມູນທີ່ຖືກແບ່ງດ້ວຍ scatter plot ພ້ອມອະທິບາຍຜົນ.



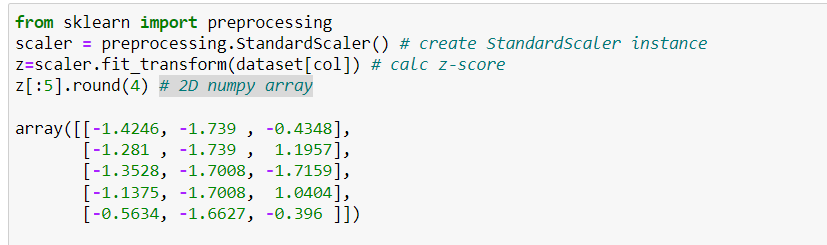


**ພາກທີ 2**

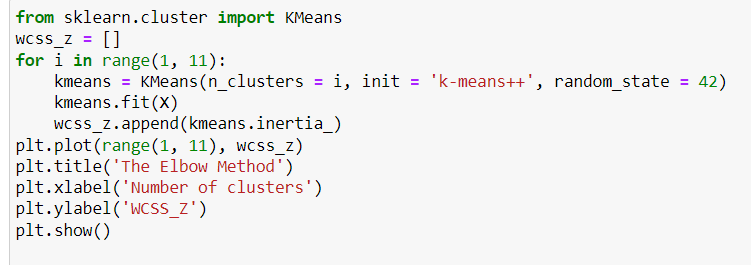
2.1 ຈາກຊຸດຂໍ້ມູນ Mall\_Customers.csv. ຈົ່ງທຳການເລືອກຄຸນລັກສະນະ (Features): Age, Annual Income(k$) ແລະ Spending Score (1-100)

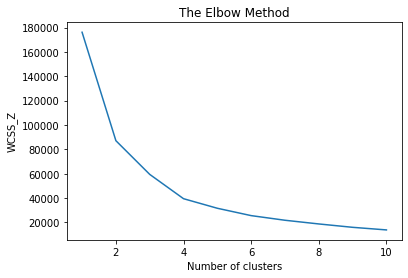


2.2. ຈົ່ງທຳການກະກຽມຂໍ້ມູນດ້ວຍ preprocessing.StandardScaler() ຂອງຄຸນລັກສະນະຂອງຂໍ້ມູນໃນຂໍ້2.1.



2.3 ຈົ່ງທຳການກຳນົດຄ່າຂອງ K ດ້ວຍການນຳໃຊ້ sklearn.cluster ແລະ Kmeans ເພື່ອຄິດໄລ່ຄ່າ wcss ພ້ອມ ແຕ້ມເສັ້ນສະແດງຂອງ Elbow Method





2.4. ຈົ່ງສ້າງແບບຈຳຮອງ Kmeans ເພື່ອແບ່ງກຸ່ມຊຸດຂໍ້ມູນດ້ວຍຕາມຄ່າຂອງ K ໃນຂໍ້2.3



2.5. ຈົ່ງແບ່ງກຸ່ມຊຸດຂໍ້ມູນໃນຂໍ້ທີ 2.2 ດ້ວຍແບບຈຳຮອງໃນຂໍ້ 2.4



2.5. ຈົ່ງຊອກຫາເມັດກາງ cluster\_centers\_ ຂອງການແບ່ງກຸ່ມຊຸດຂໍ້ມູນ



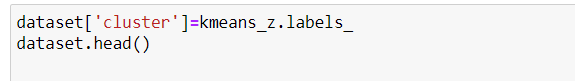
array([[ 31.9589, 72.9589],

[ 55.8148, 51.7778],

[ 39. , 106.5 ],

[ 30.3469, 29.2653]])

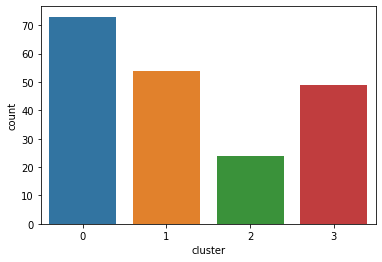
2.6. ຈົ່ງສະແດງນຳເບີກຸ່ມ (labels\_) ໃສ່ໃນຊຸດຂໍ້ມູນໃນຮູບແບບ dataframe



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| CustomerID | Gener | Age | Annual Income(K$) | Spending Score(1-100) |
| 1 | Male | 19 | 15 | 39 |
| 2 | Male | 21 | 15 | 39 |
| 3 | Fmale | 20 | 16 | 6 |
| 5 | Fmale | 23 | 16 | 77 |

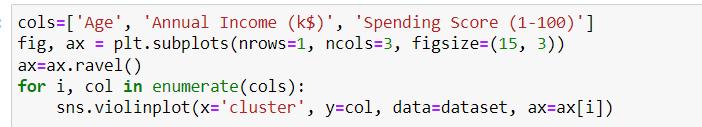
2.7. ຈົ່ງສະແດງກຸ່ມຂໍ້ມູນດ້ວຍ sns.countplot(), ພ້ອມອະທິບາຍຜົນ.

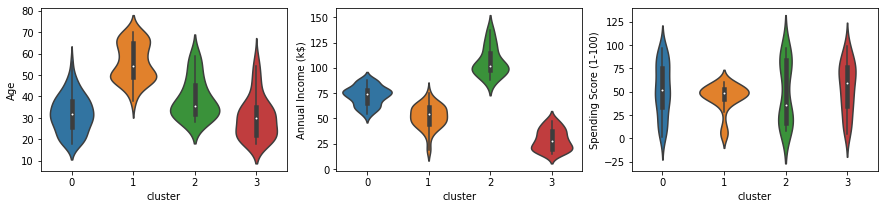




* ໃນຮູບຈະສະແດງໃຫ້ເປັນເຖິງການຂະຫຍາຍຕົວຂອງກຸ່ມວ່າເປັນຊະນິດທີມີຈຳນວນລາຍຮັບຫຼາຍຫຼື່ນ້ອງແລະ ຈັດຢູ່ໃນລຳດັບໃດ

2.8. ຈົ່ງສະແດງກຸ່ມຂໍ້ມູນດ້ວຍ sns.violinplot(), ພ້ອມອະທິບາຍຜົນ.

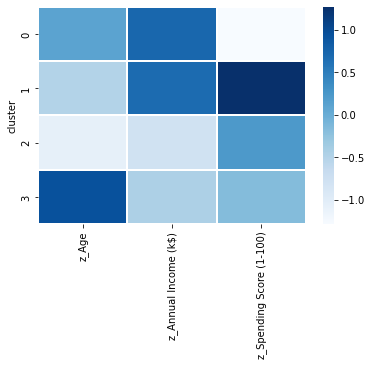




* ໃນຮູບທີ1 : ແມ່ນຈະສະແດງເຖິງອາຍຸ
* ໃນຮູບທີ່2 : ແມ່ນຈະສະແດງເຖິງລາຍໄດ້ຂອງແຕ່ລະກຸ່ມ
* ໃນຮູບທີ່3 : ແມ່ນຈະສະແດງເຖິງຄະ​ແນນ​ການ​ໃຊ້​ຈ່າຍແຕ່ລະກຸ່ມ

2.9. ຈົ່ງສະແດງກຸ່ມຂໍ້ມູນດ້ວຍ sns.heatmap(), ພ້ອມອະທິບາຍຜົນ.





ໃນຮຸບນີ້ແມ່ນຈະສະແດງແຕ່ລະແຖວໂດຍລະອຽບເຊີ່ງວ່າຖ້າແຖວໃດມີສີທີ່ແຈ້ງແມ່ນກຸ່ມທີ່ມີຈຳນວນຫຼາຍໃນກຸ່ມນັ້ນຕາມລຳດັບລົງມາ