A. 下表為某公司的員工資料表，請採用漢明距離進行相似度的衡量

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 員工代號 | 性別 | 部門 | 薪資 | 配偶 |
| 01 | 男 | 市場部 | 24000 | 無 |
| 02 | 女 | 研發部 | 45000 | 有 |
| 03 | 男 | 會計部 | 45000 | 無 |
| 04 | 男 | 研發部 | 40000 | 無 |
| 05 | 女 | 市場部 | 24000 | 有 |
| 06 | 女 | 研發部 | 40000 | 無 |
| 07 | 男 | 市場部 | 24000 | 有 |

B. 根據上表，利用漢明距離和資訊理論來進行特性的排序，並且找出對資訊量最沒有影響的（建議可以刪除的） feature。

註：log底數請使用2

P0:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 01 | 02 | 03 | 04 | 05 | 06 | 07 |
| 01 |  | **0/4** | **2/4** | **2/4** | **2/4** | **2/4** | **2/4** |
| 02 |  |  | **1/4** | **1/4** | **2/4** | **2/4** | **1/4** |
| 03 |  |  |  | **2/4** | **0/4** | **1/4** | **1/4** |
| 04 |  |  |  |  | **0/4** | **3/4** | **1/4** |
| 05 |  |  |  |  |  | **1/4** | **3/4** |
| 06 |  |  |  |  |  |  | **0/4** |
| 07 |  |  |  |  |  |  |  |

P1: X性別

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 01 | 02 | 03 | 04 | 05 | 06 | 07 |
| 01 |  | **0/3** | **1/3** | **1/3** | **2/3** | **1/3** | **2/3** |
| 02 |  |  | **1/3** | **1/3** | **1/3** | **1/3** | **1/3** |
| 03 |  |  |  | **1/3** | **0/3** | **1/3** | **0/3** |
| 04 |  |  |  |  | **0/3** | **3/3** | **0/3** |
| 05 |  |  |  |  |  | **0/3** | **3/3** |
| 06 |  |  |  |  |  |  | **0/3** |
| 07 |  |  |  |  |  |  |  |

P2: X 部門

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 01 | 02 | 03 | 04 | 05 | 06 | 07 |
| 01 |  | **0/3** | **2/3** | **2/3** | **1/3** | **1/3** | **2/3** |
| 02 |  |  | **1/3** | **0/3** | **2/3** | **1/3** | **1/3** |
| 03 |  |  |  | **2/3** | **0/3** | **1/3** | **1/3** |
| 04 |  |  |  |  | **0/3** | **2/3** | **1/3** |
| 05 |  |  |  |  |  | **1/3** | **2/3** |
| 06 |  |  |  |  |  |  | **0/3** |
| 07 |  |  |  |  |  |  |  |

P3: X 薪資

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 01 | 02 | 03 | 04 | 05 | 06 | 07 |
| 01 |  | **0/3** | **2/3** | **2/3** | **1/3** | **1/3** | **2/3** |
| 02 |  |  | **0/3** | **1/3** | **2/3** | **2/3** | **1/3** |
| 03 |  |  |  | **2/3** | **0/3** | **1/3** | **1/3** |
| 04 |  |  |  |  | **0/3** | **2/3** | **1/3** |
| 05 |  |  |  |  |  | **1/3** | **2/3** |
| 06 |  |  |  |  |  |  | **0/3** |
| 07 |  |  |  |  |  |  |  |

P4: X配偶

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 01 | 02 | 03 | 04 | 05 | 06 | 07 |
| 01 |  | **0/3** | **1/3** | **1/3** | **2/3** | **0/3** | **3/3** |
| 02 |  |  | **1/3** | **1/3** | **1/3** | **2/3** | **0/3** |
| 03 |  |  |  | **1/3** | **0/3** | **0/3** | **1/3** |
| 04 |  |  |  |  | **0/3** | **2/3** | **1/3** |
| 05 |  |  |  |  |  | **1/3** | **2/3** |
| 06 |  |  |  |  |  |  | **0/3** |
| 07 |  |  |  |  |  |  |  |

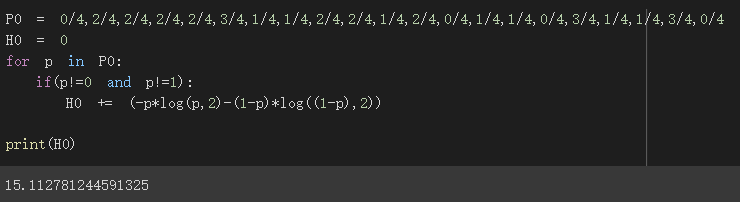
P0 = 0/4,2/4,2/4,2/4,2/4,3/4,1/4,1/4,2/4,2/4,1/4,2/4,0/4,1/4,1/4,0/4,3/4,1/4,1/4,3/4,0/4

P1 = 0/3,1/3,1/3,2/3,1/3,2/3,1/3,1/3,1/3,1/3,1/3,1/3,0/3,1/3,0/3,0/3,3/3,0/3,0/3,3/3,0/3

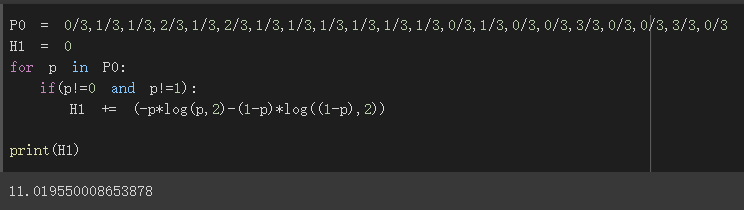
P2 = 0/3,2/3,2/3,1/3,1/3,2/3,1/3,0/3,2/3,1/3,1/3,2/3,0/3,1/3,1/3,0/3,2/3,1/3,1/3,2/3,0/3

P3 = 0/3,2/3,2/3,1/3,1/3,2/3,0/3,1/3,2/3,2/3,1/3,2/3,0/3,1/3,1/3,0/3,2/3,1/3,1/3,2/3,0/3

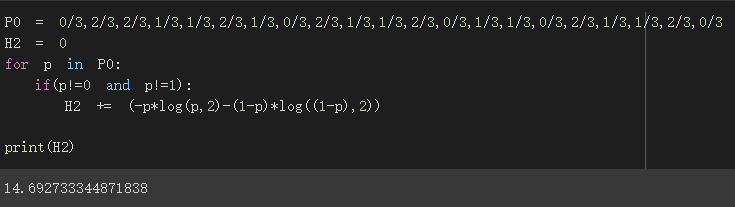
P4 = 0/3,1/3,1/3,2/3,0/3,3/3,1/3,1/3,1/3,2/3,0/3,1/3,0/3,0/3,1/3,0/3,2/3,1/3,1/3,2/3,0/3



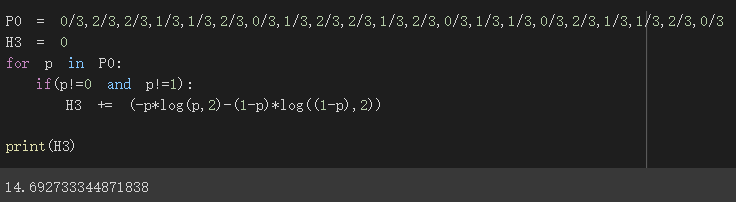
H0



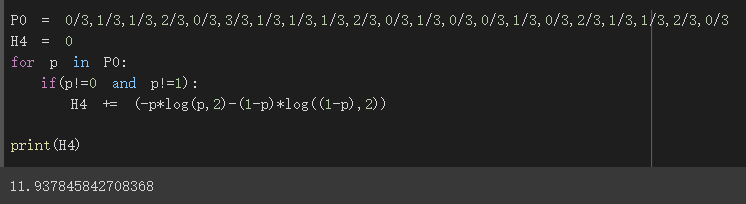
H1= 



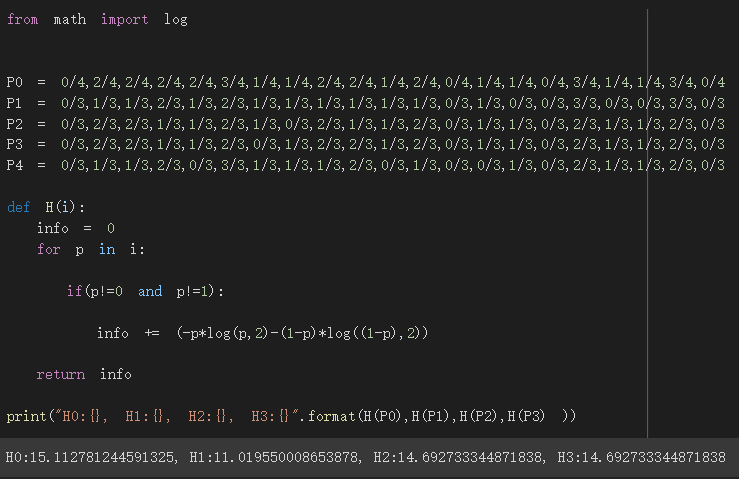
H2



H3

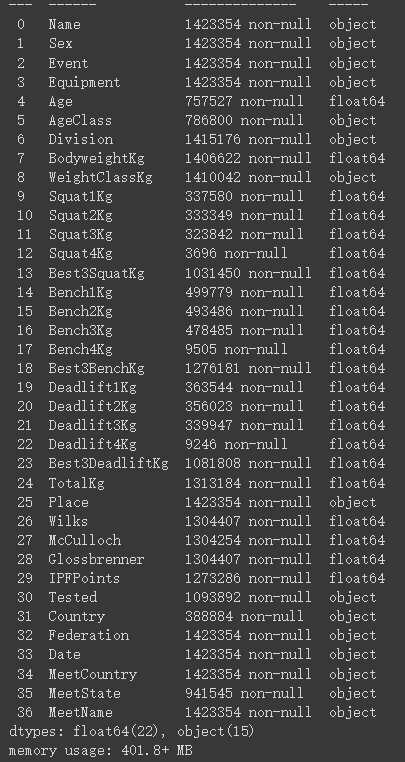
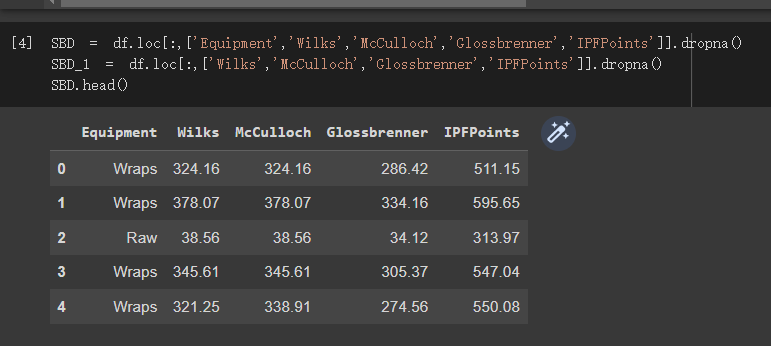
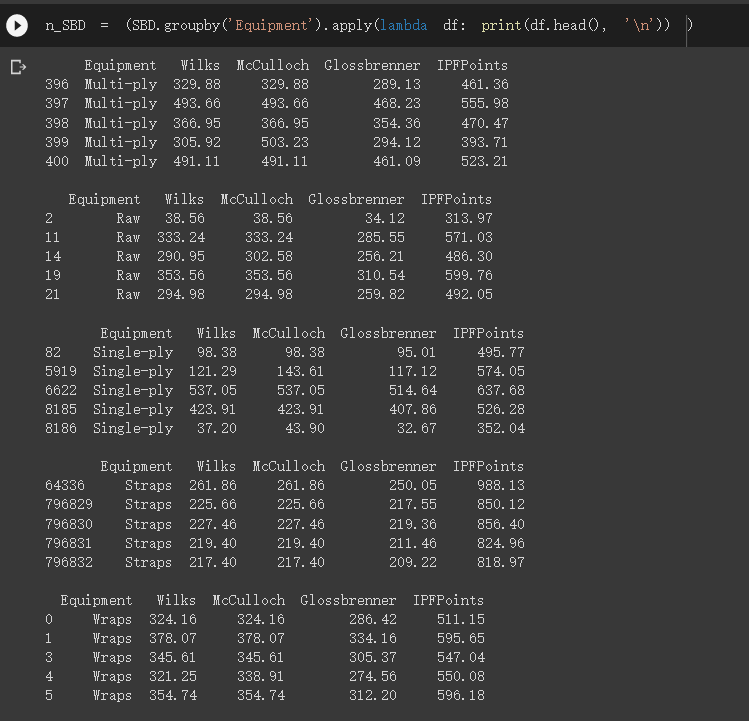


H4

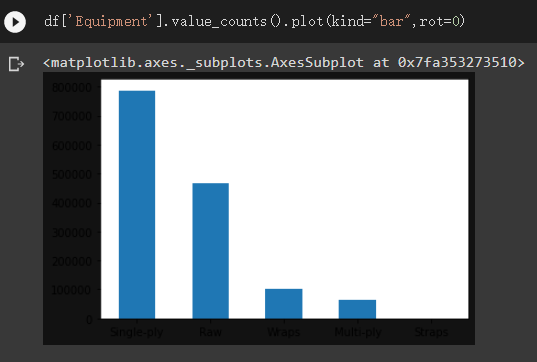


**H1 > H2 = H3 > H0**

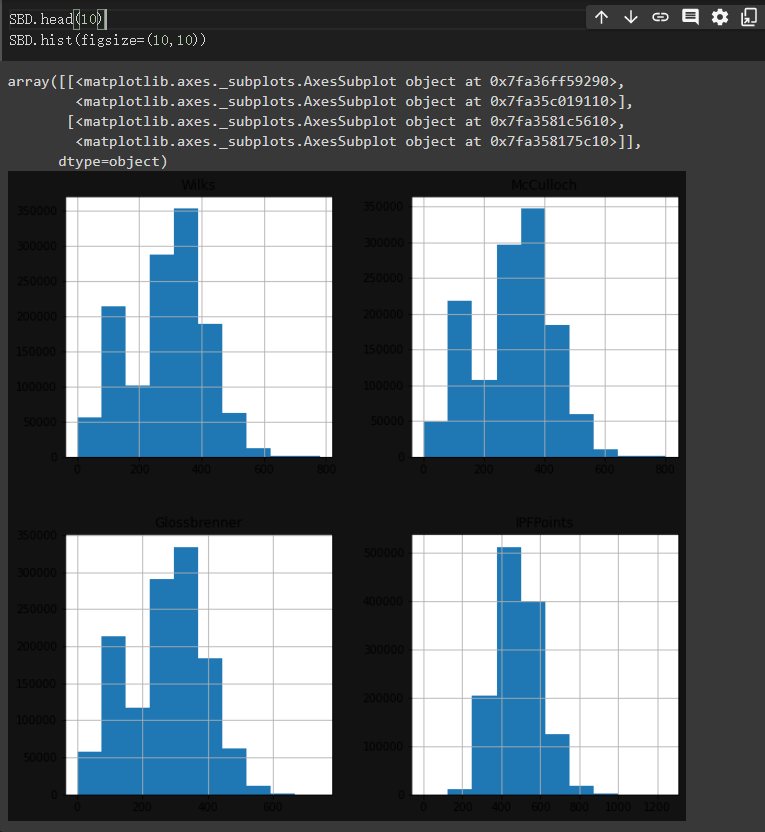
****

****

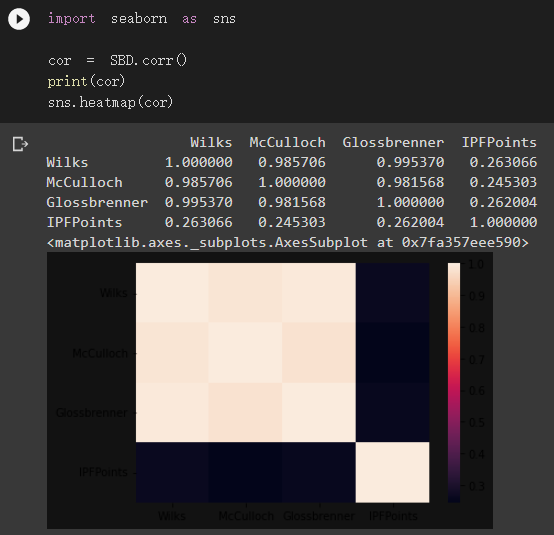
****

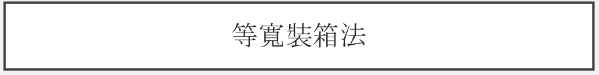
****

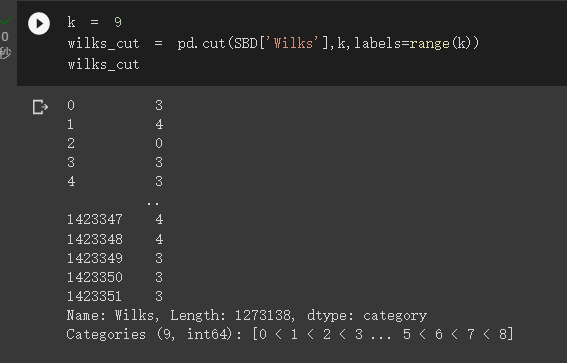
****

****

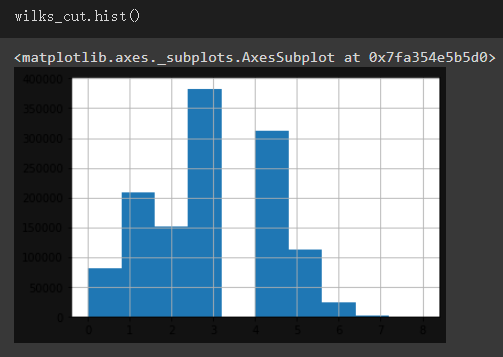
****

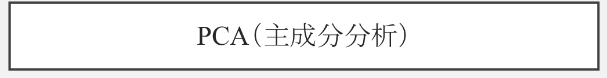
****

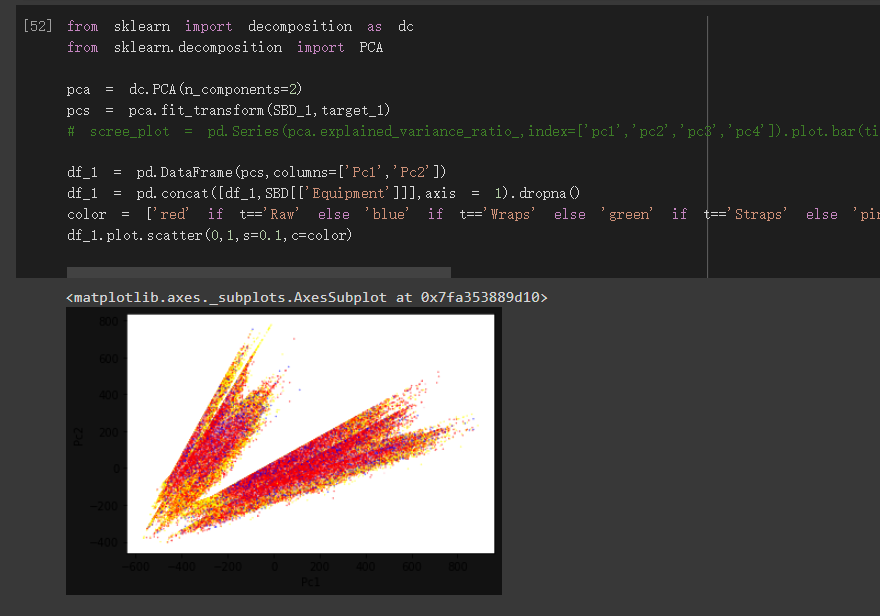
****

****

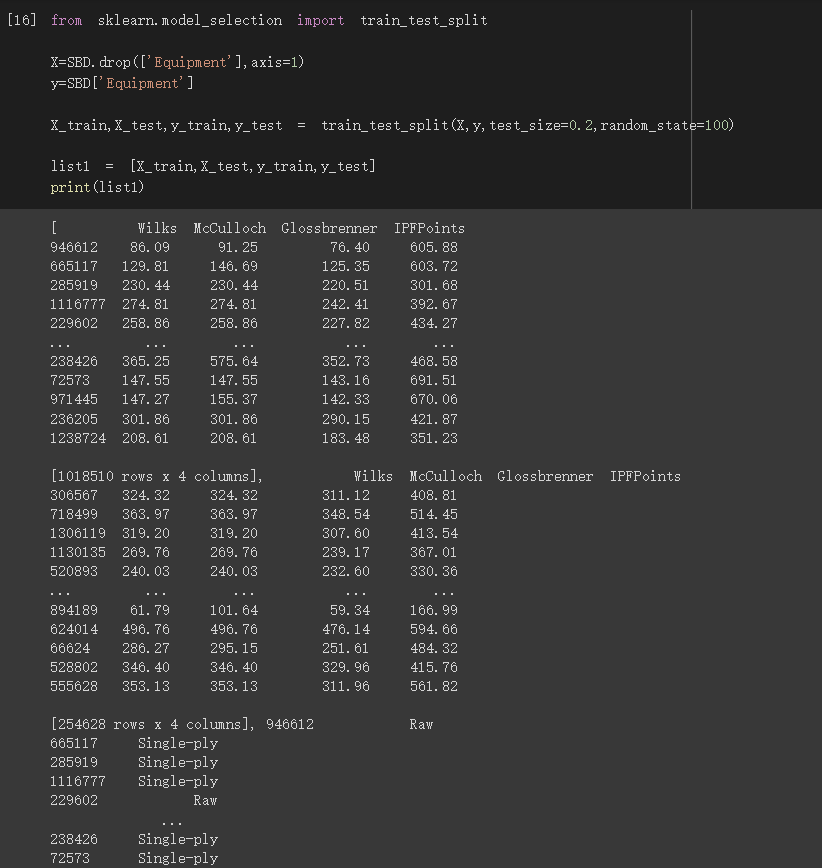
****

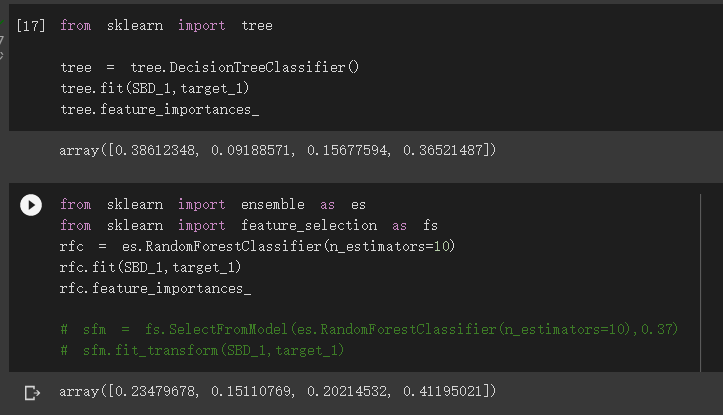
****

****

****

****

****

****