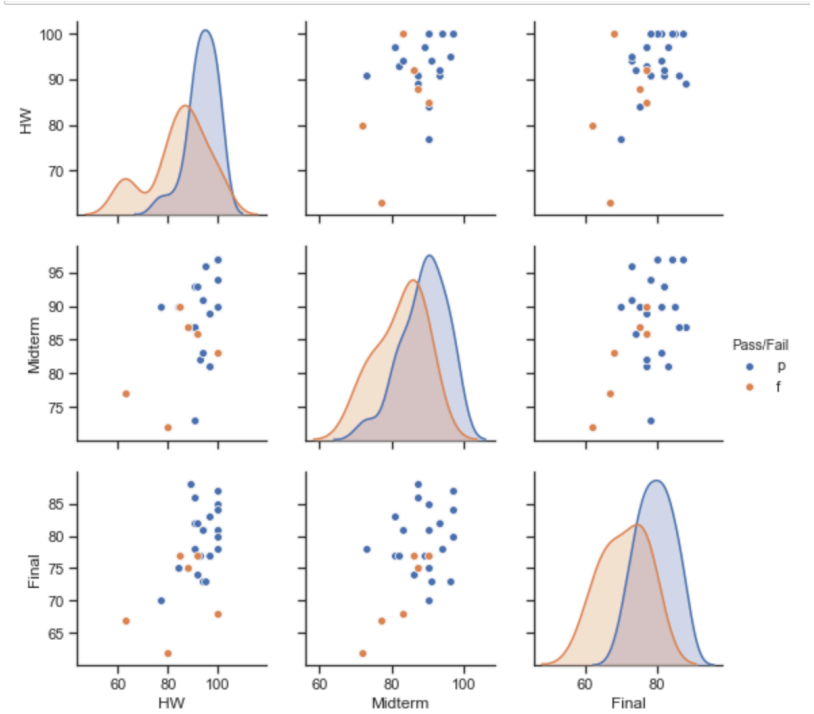
Phuong Ngo
INF550
September 23, 2019

1. Size is y, price is x
    a. The most specific hypothesis is: 270<x<460, 180<y<860
    b. The most general hypothesis is: 150<x<630 , 140<y<480
    c. Examples:
        i. A family home is a middle class home when their house's price is 300K and the size is 450sqm (11, 300, 450, Y)
        ii. False positive: 12, 550, 400, N
        iii. False negative: 13, 400, 100, Y
    d. It won't change the hypothesis. Even though it's outside of the most general hypothesis it can be an outlier case.
2. The following scatterplots shows the relationship between two features against the binary class (pass/fail). From the plots, it seems final/homework would be good models to use.



3.
    a. Find min. support count: It's 2 since the min support requirement is 40%. Since we have 5 transactions(t), we have 0.4*5=2
    b. C1 {t(A): 5, t(B):3, t(C): 5, t(D): 3, t(E):2} compare to min support count 2
    L1 {t(A): 5, t(B):3, t(C): 5, t(D): 3, t(E):2}
    c. C2 {t(AB):3, t(AC): 5, t(AD): 4, t(AE):2, t(BC): 3, t(BD): 2, t(BE): 1, t(CD): 4, t(CE): 2, t(DE): 2} rule out t(BE): 1

L2 {t(AB):3, t(AC): 5, t(AD): 4, t(AE):2, t(BC): 3, t(BD): 2, t(CD): 4, t(CE): 2, t(DE): 2}

d. C3 {t(ABC):3, t(ABD): 2, t(ACD):4, t(ACE): 2, t(ADE): 2, t(BCD): 2, t(CDE): 2} use Prune step along with min support count to rule out more subsets. Since BE is not frequent, we rule out any that have BE as a subset
   L3 {t(ABC):3, t(ABD): 2, t(ACD):4, t(ACE): 2, t(ADE): 2, t(BCD): 2, t(CDE): 2}
e. C4 {t(ABCD), t(ABCE), t(BCDE), t(ABDE), t(ACDE)} but with Prune and join step we have C4 {t(ABCD): 2, t(ACDE): 2}
   L4 {t(ABCD): 2, t(ACDE): 2}
f. All frequent item sets are: (A), (B), (C), (D), (E), (AB), (AC), (AD), (AE), (BC), (BD), (CD), (CE), (DE), (ABC), (ABD), (ACD), (ACE), (ADE), (BCD), (CDE), (ABCD), (ACDE)