

63010805 รัชฎา แซ่มซ้าย
63010849 วรวิษ ตั้งพจน์ทวีพร

โปรแกรมที่ใช้ : MATLAB

Column 1 : Aftertaste

Properties

ตัวแปรต้น :

X-axis = Aftertaste point

จำนวนรายการทั้งหมด = 1310 แถว

ตัวแปรตาม :

Y-axis = Total.Cup.Point

Min : 6.17 points

Max : 8.67 points

Mean : 7.4033 points

Median : 7.42 points

Mode : 7.5 points

Range : 2.5 points

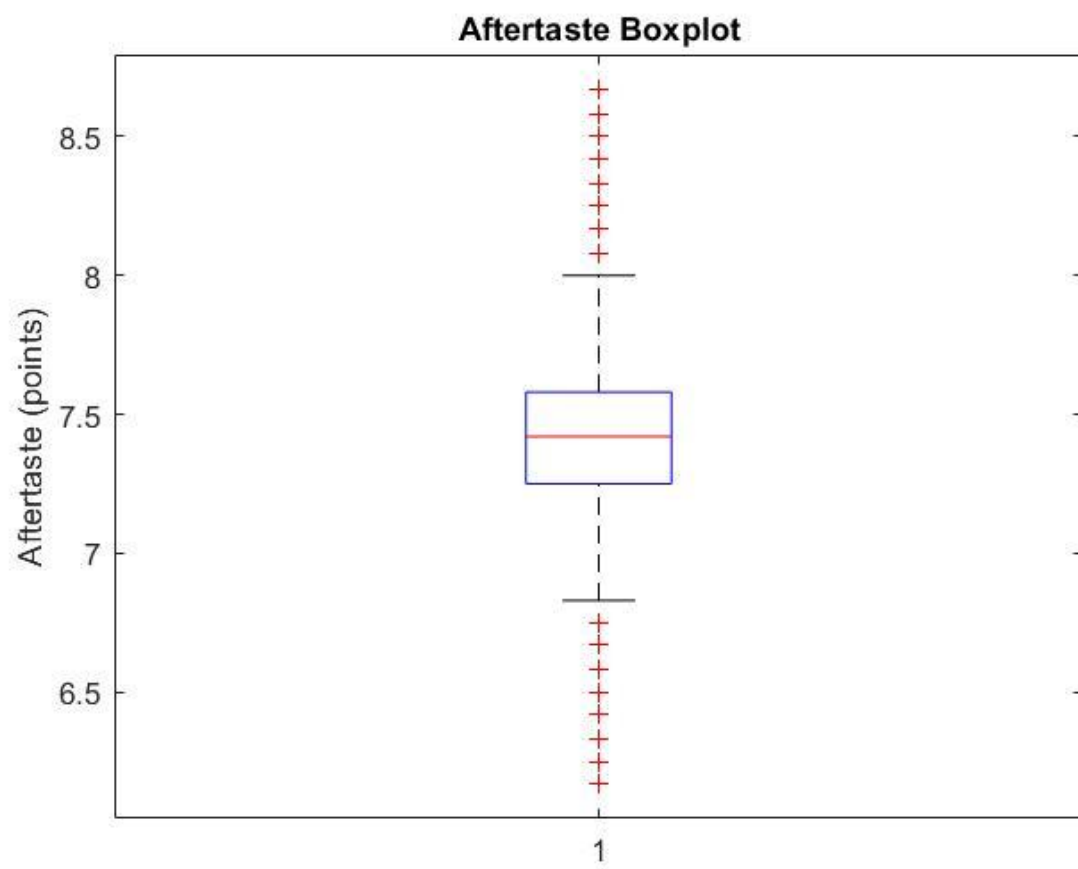
Mean_trm : 7.4209

Sample standard deviation : 0.3499 points

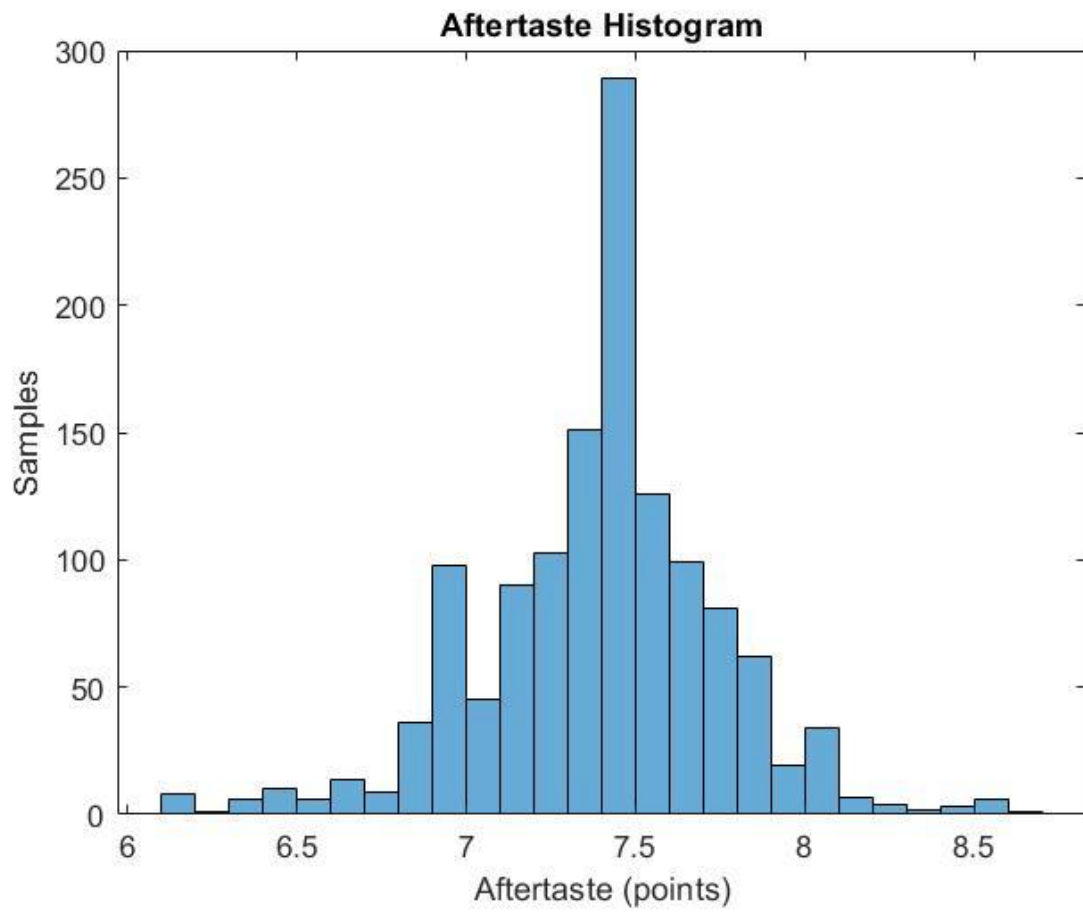
Variance : 0.1224 points

Plot

Box Plot



Histogram



Stem and Leaves

[illegible]

Outlier

Interquartile range (IQR)

จาก $IQR = Q3 - Q1$

ข้อมูลจำนวน 1310 ชุด

$$\begin{aligned} Q1 &= (1/4)(n+1) \\ &= (1/4)(1310 + 1) \\ &= 327.75 \end{aligned}$$

$$\begin{aligned} \text{จะได้ ข้อมูลตำแหน่งที่ } 327.75 &= x_{327} + 0.75(x_{328} - x_{327}) \\ &= 7.25 + 0.75(7.25 - 7.25) \\ &= 7.25 \text{ points} \end{aligned}$$

$$\begin{aligned} Q3 &= (3/4)(n+1) \\ &= (3/4)(1310 + 1) \\ &= 983.25 \end{aligned}$$

$$\begin{aligned} \text{จะได้ ข้อมูลตำแหน่งที่ } 327.75 &= x_{983} + 0.25(x_{984} - x_{983}) \\ &= 7.58 + 0.25(7.58 - 7.58) \\ &= 7.58 \text{ points} \end{aligned}$$

$$\begin{aligned} \text{จะได้ } IQR &= 7.58 - 7.25 \\ &= 0.33 \end{aligned}$$

$$\begin{aligned} \text{Lower outlier ข้อมูลที่มีค่าน้อยกว่า } Q1 - 1.5(IQR) &= 7.25 - 1.5(0.33) \\ &= 6.755 \text{ points} \end{aligned}$$

$$\begin{aligned} \text{Higher outlier ข้อมูลที่มีค่ามากกว่า } Q3 + 1.5(IQR) &= 7.58 + 1.5(0.33) \\ &= 8.075 \text{ points} \end{aligned}$$

ดังนั้น ข้อมูลที่เป็นoutlier ได้แก่คะแนน

6.17 จำนวน 8 ข้อมูล

6.25 จำนวน 1 ข้อมูล

6.33 จำนวน 6 ข้อมูล

6.42 จำนวน 4 ข้อมูล

6.5 จำนวน 6 ข้อมูล

6.58 จำนวน 6 ข้อมูล

6.67 จำนวน 14 ข้อมูล

6.75 จำนวน 9 ข้อมูล

8.08 จำนวน 7 ข้อมูล

8.17 จำนวน 7 ข้อมูล

8.25 จำนวน 4 ข้อมูล

8.33 จำนวน 2 ข้อมูล

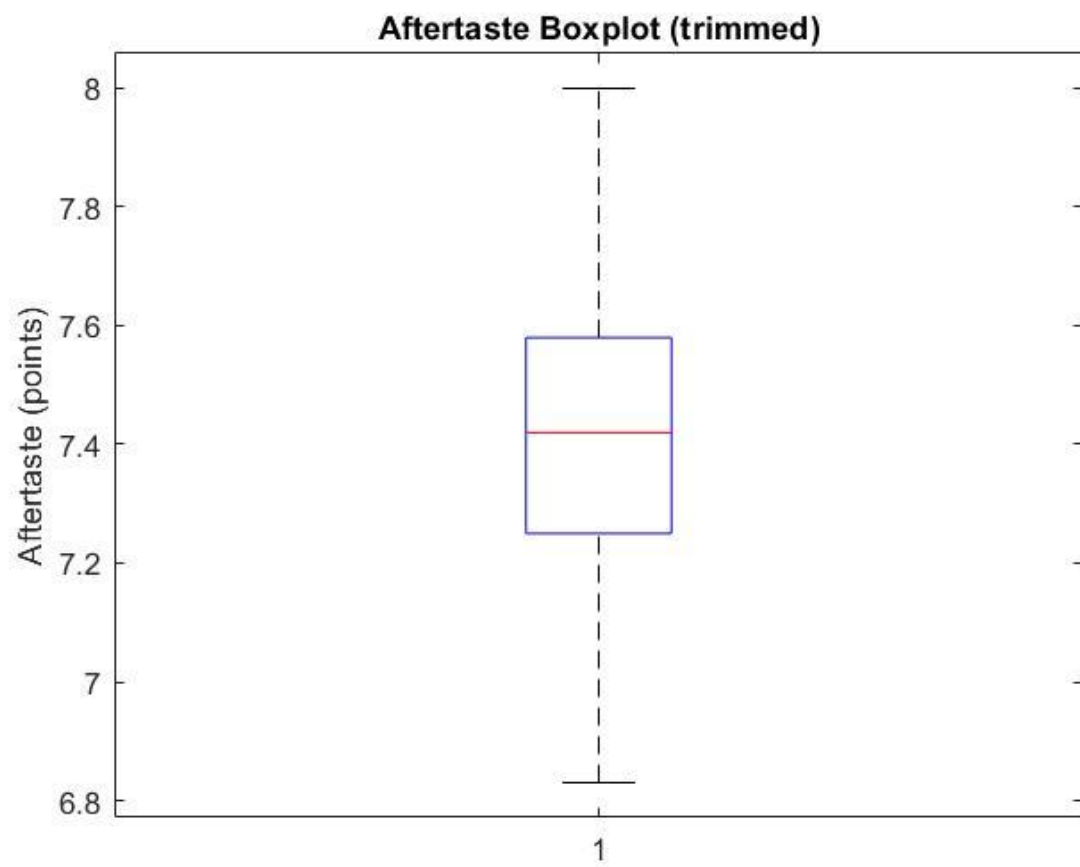
8.42 จำนวน 3 ข้อมูล

8.50 จำนวน 4 ข้อมูล

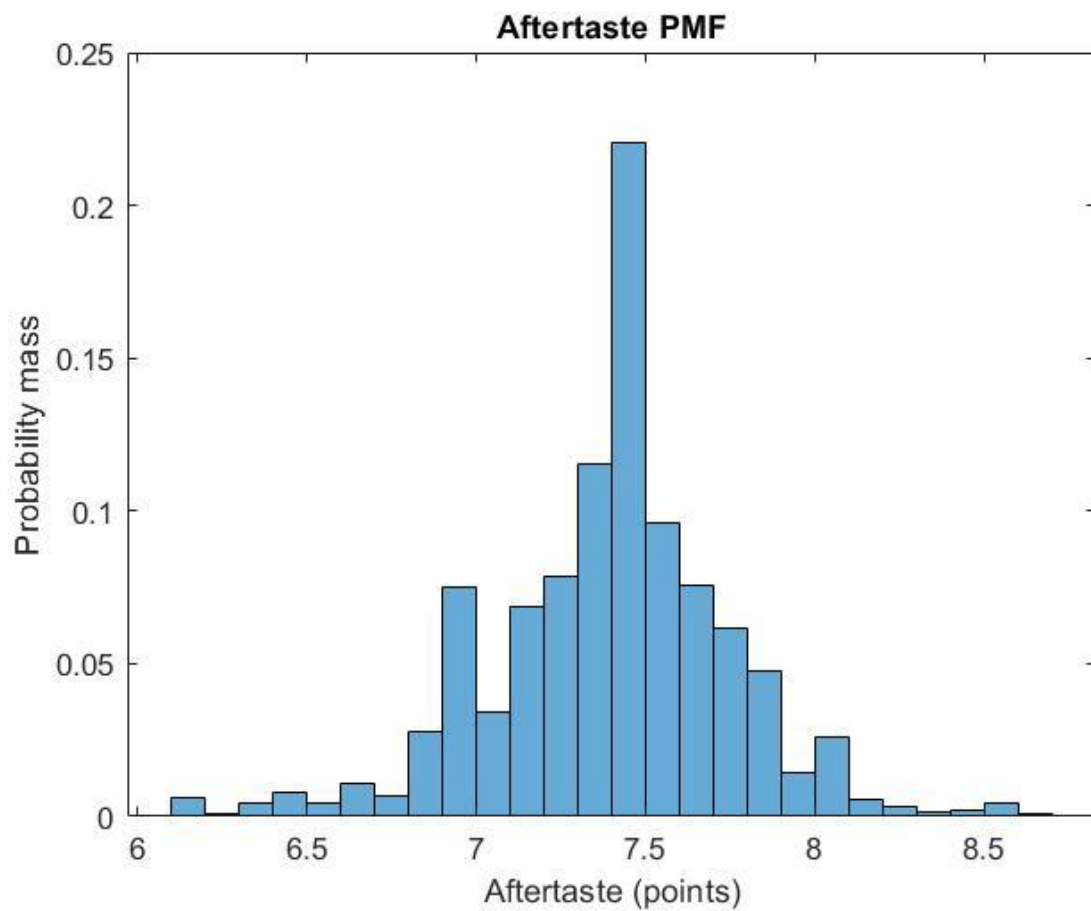
8.58 จำนวน 2 ข้อมูล

8.67 จำนวน 1 ข้อมูล

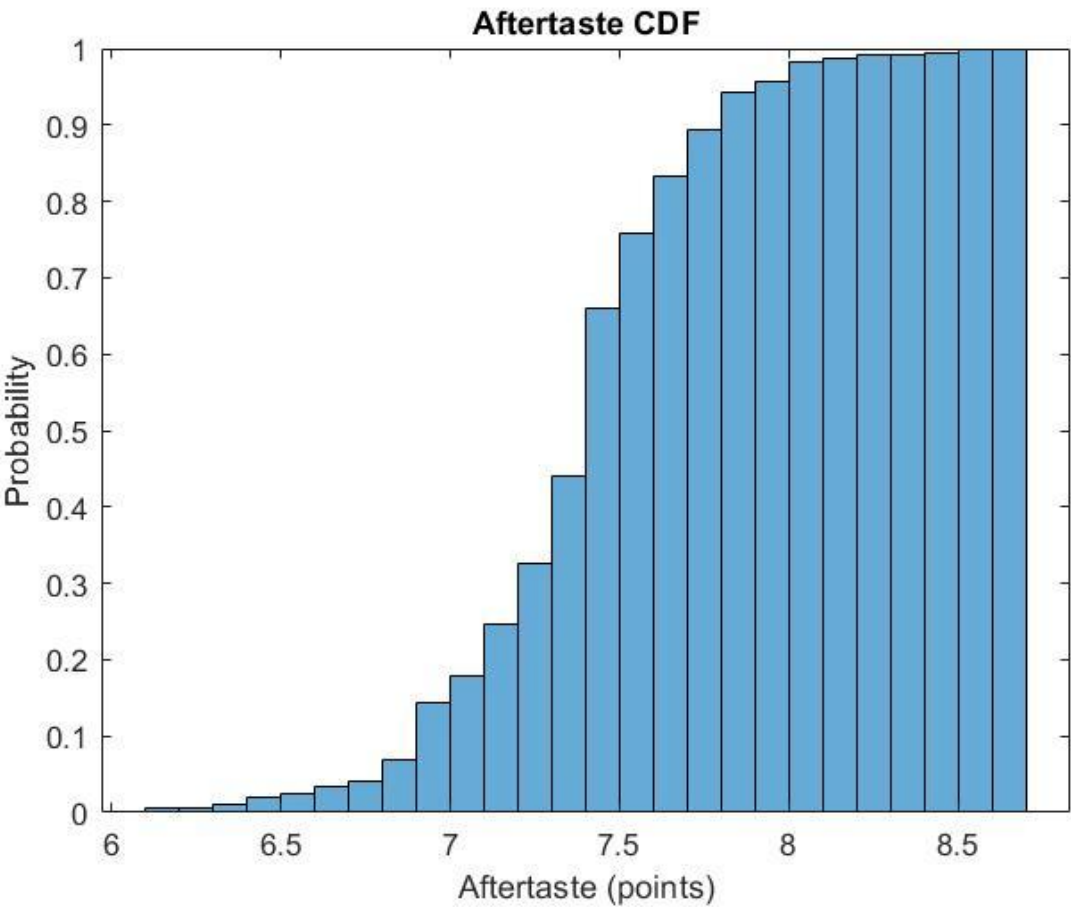
เป็นจำนวน 84 ข้อมูล นับเป็น 6.41221% ของข้อมูลทั้งหมด



Probability Mass Function



Cumulative Probability Function



Column 2 : Sweetness

Properties

ตัวแปรต้น :

X-axis = Sweetness

จำนวนรายการทั้งหมด = 1310 แถว

ตัวแปรตาม :

Y-axis = Total.Cup.Point

Min : 1.33

Max : 10

Mean : 9.9108

Median : 10

Mode : 10

Range: 8.67

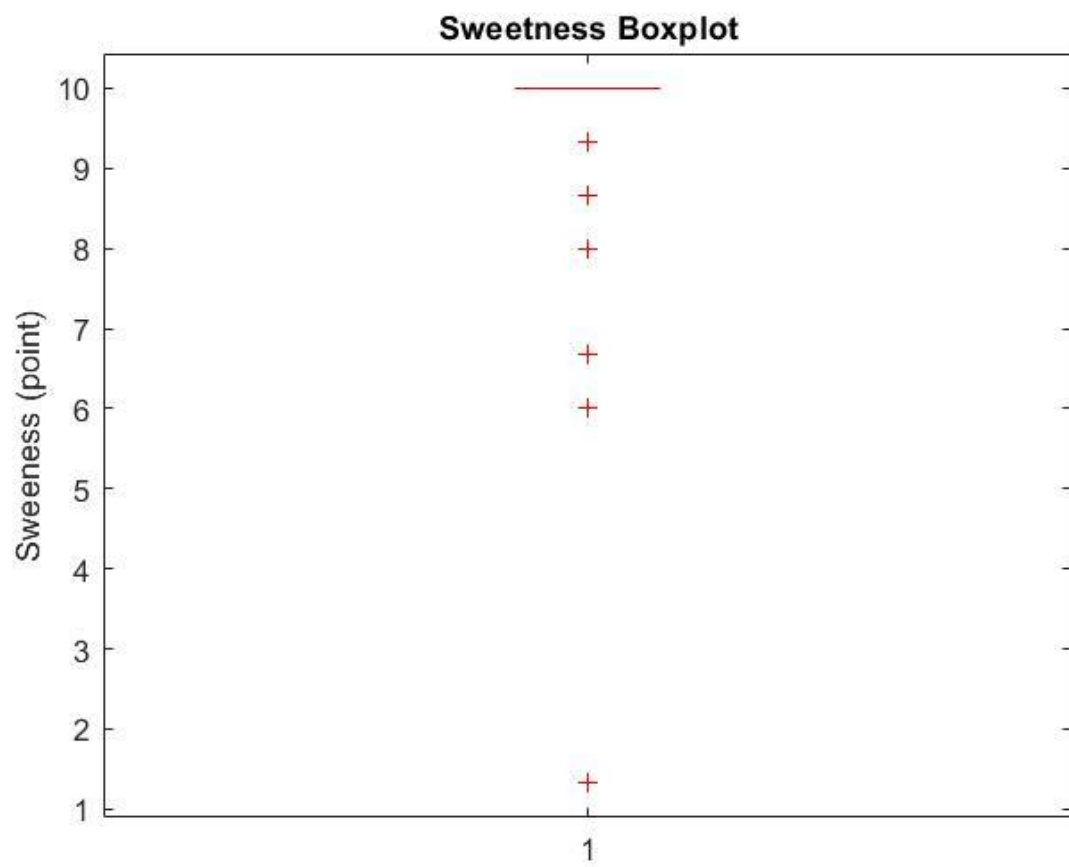
Sample standard deviation : 0.455

C.V. : 0.207

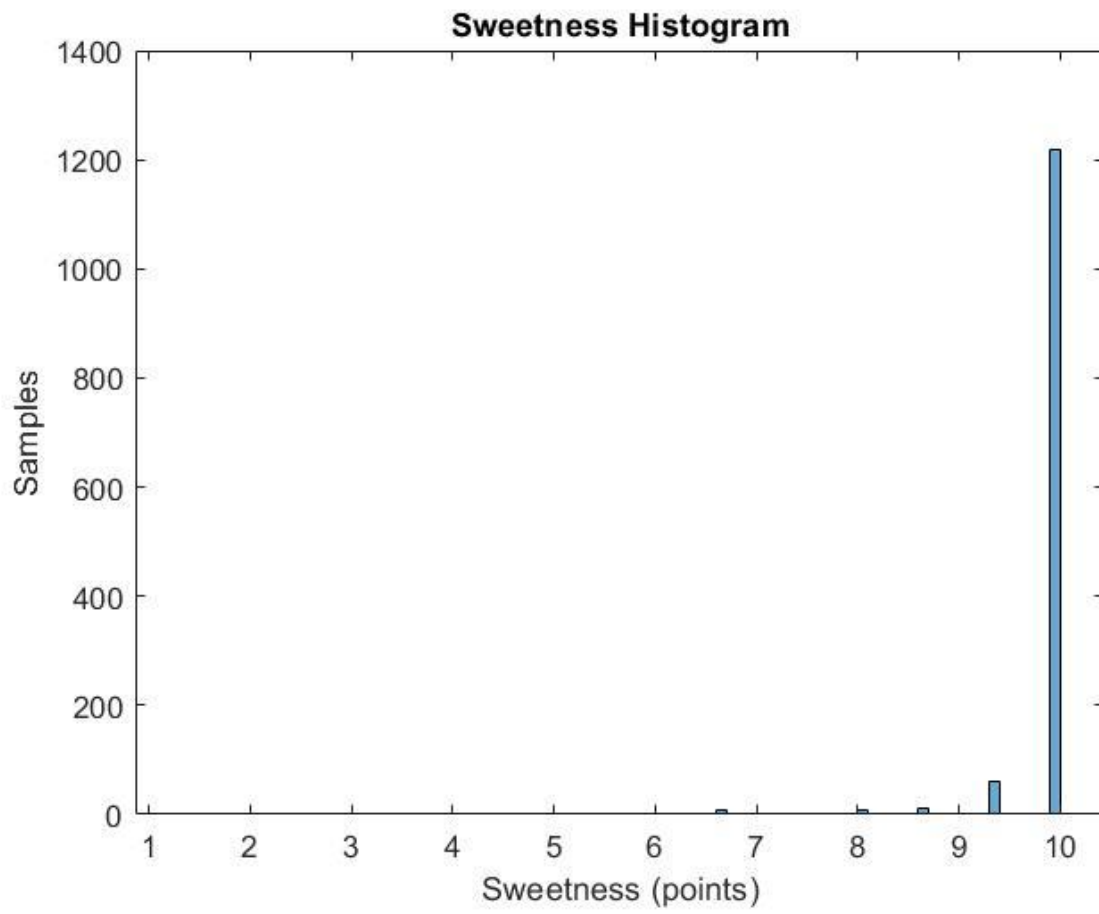
Mean_trm = 10

Plot

Box Plot

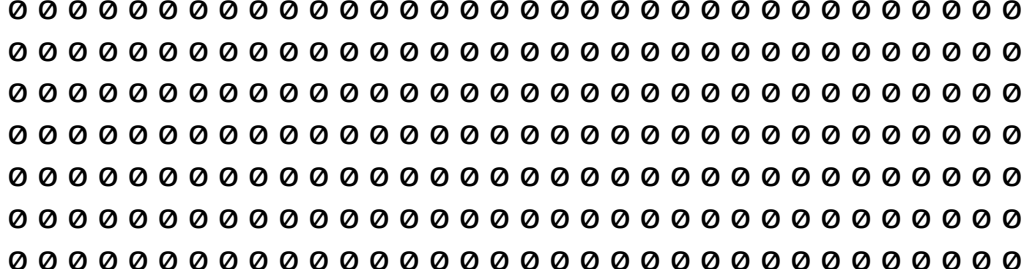


Histogram



Stem and Leaves

[illegible]

	
--	--

Outlier

Interquartile range (IQR)

จาก $IQR = Q3 - Q1$

ข้อมูลจำนวน 1310 ชุด

$$\begin{aligned} Q1 &= (1/4)(n+1) \\ &= (1/4)(1310 + 1) \\ &= 327.75 \end{aligned}$$

$$\begin{aligned} \text{จะได้ ข้อมูลตำแหน่งที่ } 327.75 &= x_{327} + 0.75(x_{328} - x_{327}) \\ &= 10 + 0.75(10 - 10) \\ &= 10 \text{ points} \end{aligned}$$

$$\begin{aligned} Q3 &= (3/4)(n+1) \\ &= (3/4)(1310 + 1) \\ &= 983.25 \end{aligned}$$

$$\begin{aligned} \text{จะได้ ข้อมูลตำแหน่งที่ } 327.75 &= x_{983} + 0.25(x_{984} - x_{983}) \\ &= 10 + 0.25(10 - 10) \\ &= 10 \text{ points} \end{aligned}$$

$$\begin{aligned} \text{จะได้ } IQR &= 10 - 10 \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Lower outlier ข้อมูลที่มีค่าน้อยกว่า } Q1 - 1.5(IQR) &= 10 - 1.5(0) \\ &= 10 \text{ points} \end{aligned}$$

$$\begin{aligned} \text{Higher outlier ข้อมูลที่มีค่ามากกว่า } Q3 + 1.5(IQR) &= 10 + 1.5(0) \\ &= 10 \text{ points} \end{aligned}$$

ดังนั้น ข้อมูลที่เป็น outlier ได้แก่คะแนน 1.33 จำนวน 1 ข้อมูล
6.00 จำนวน 3 ข้อมูล

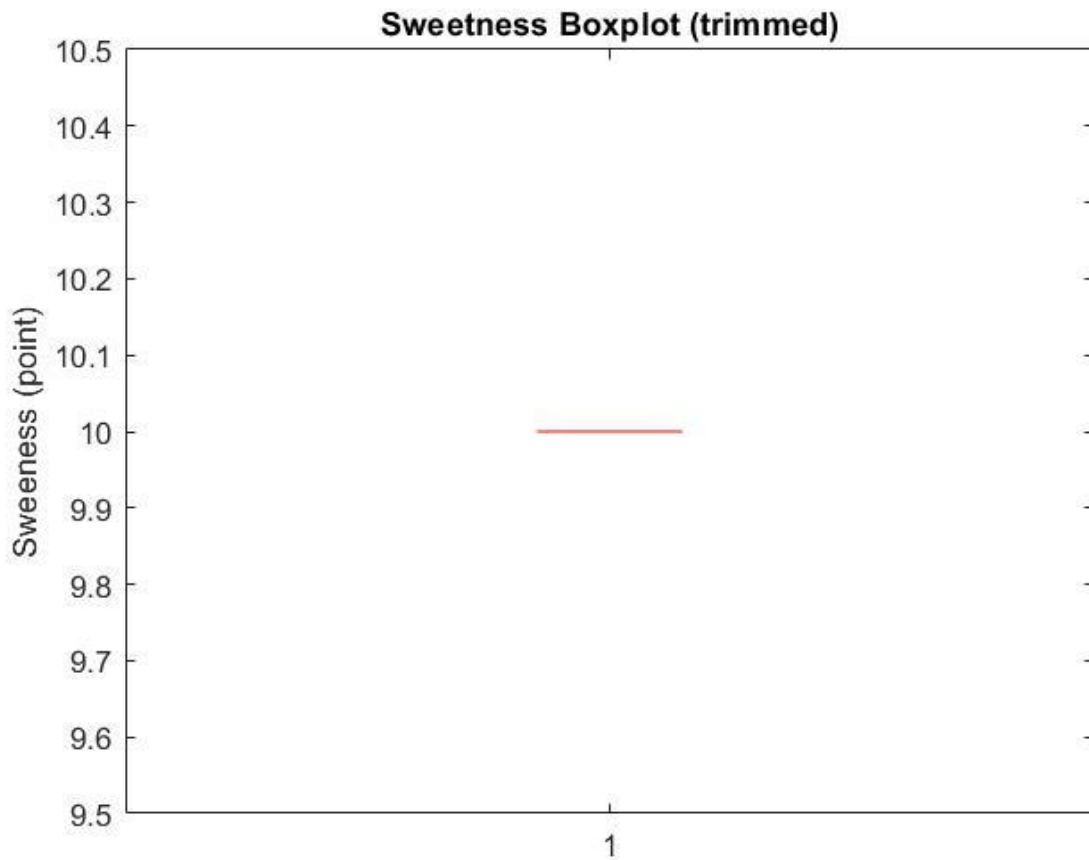
6.67 จำนวน 7 ข้อมูล

8.00 จำนวน 8 ข้อมูล

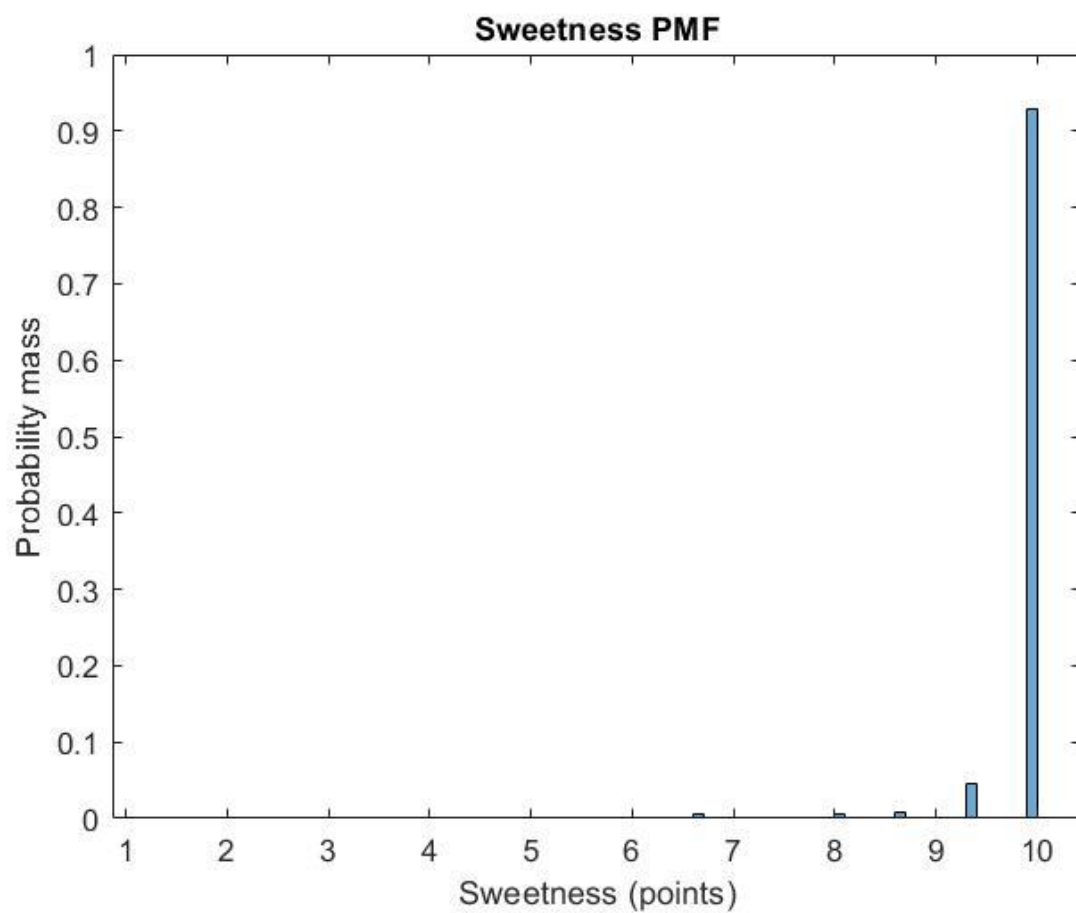
8.67 จำนวน 12 ข้อมูล

9.33 จำนวน 61 ข้อมูล

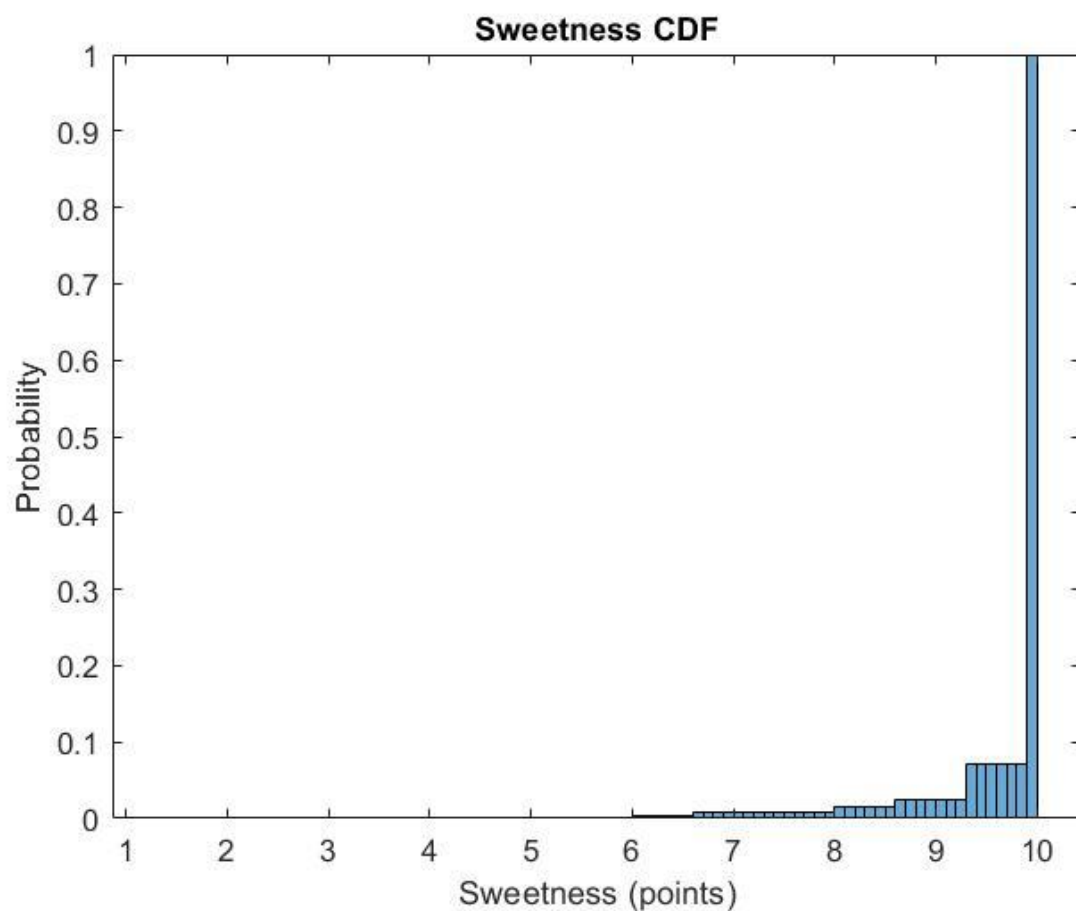
เป็นจำนวน 92 ข้อมูล นับเป็น 7.02290% ของข้อมูลทั้งหมด



Probability Density Function



Cumulative Probability Function



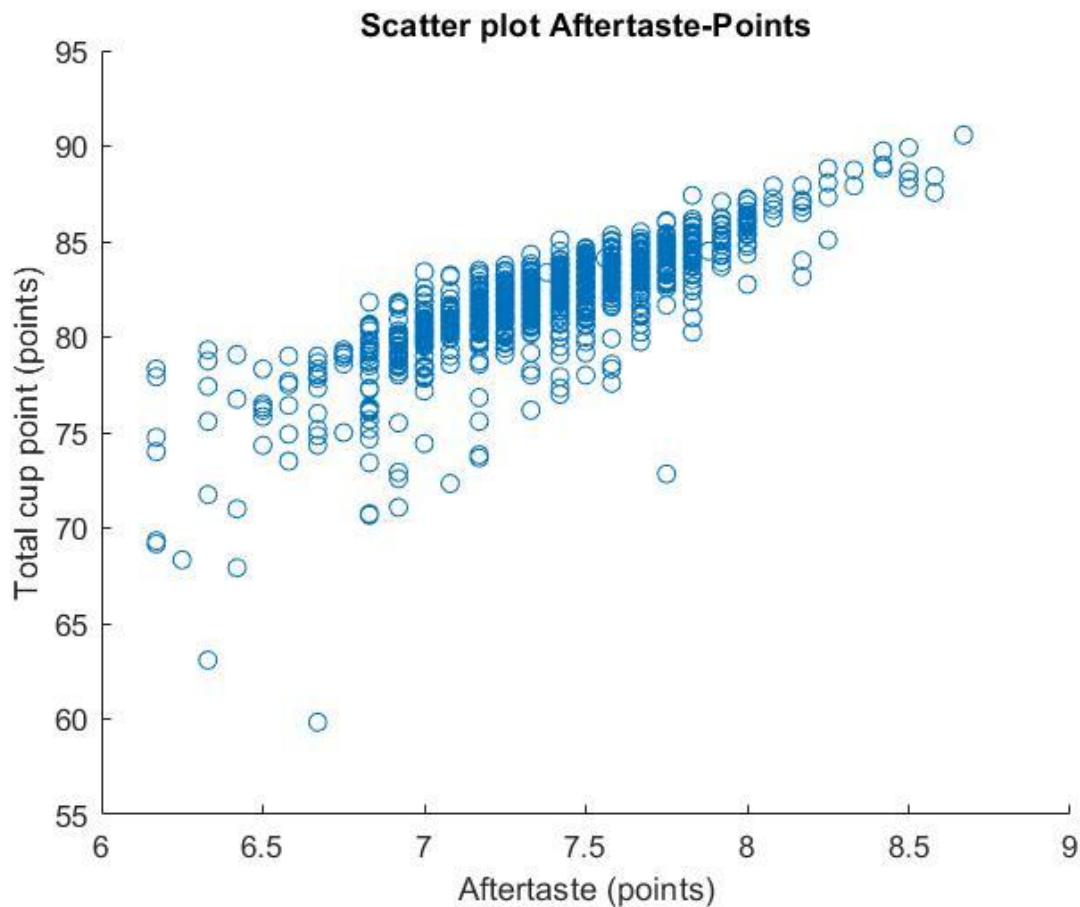
Scatter plot

ส่วนที่ 1 : Scatter plot Aftertaste-Points

ตัวแปรต้น : Aftertaste

ตัวแปรตาม : Total cup point

เนื่องจาก Aftertaste เป็นตัวแปรหนึ่งซึ่งมีผลต่อ Total cup point ตามเกณฑ์การประเมิน

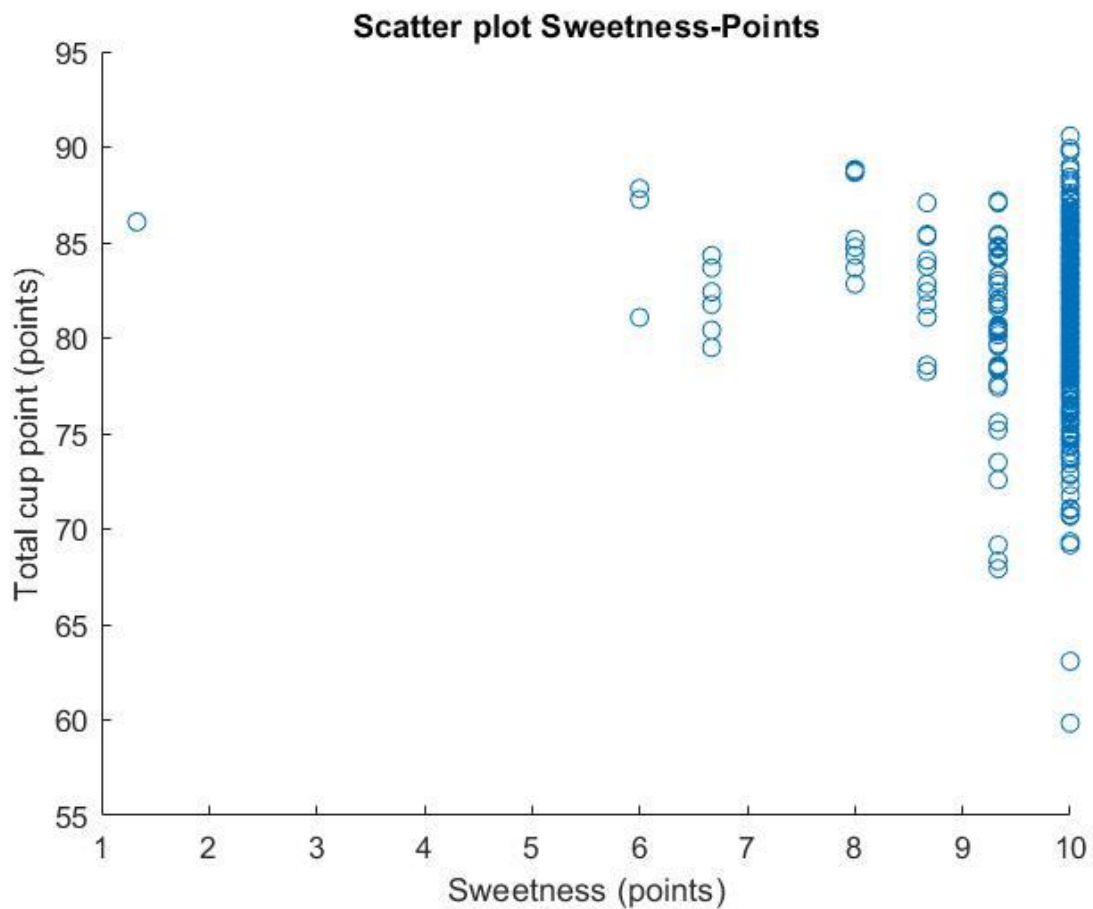


ส่วนที่ 2 : Scatter plot Sweetness-Points

ตัวแปรต้น : Sweetness

ตัวแปรตาม : Total cup point

เนื่องจาก Sweetness เป็นตัวแปรหนึ่งซึ่งมีผลต่อ Total cup point
ตามเกณฑ์การประเมิน



บทวิเคราะห์ข้อมูลจากกราฟ

ส่วนที่ 1 Aftertaste

จาก Scatter plot Aftertaste-Points ข้างต้น มีข้อมูล Aftertaste เป็นตัวแปรต้น และ Total cup point เป็นตัวแปรตาม แสดงให้เห็นว่า ทั้งสองข้อมูลมีแนวโน้มแปรผันตรงไปในทิศทางเดียวกัน เนื่องจากด้วย Aftertaste เป็นตัวแปรหนึ่งซึ่งส่งผลต่อคะแนนคุณภาพของกาแฟ (Total cup points) โดยส่วนมากกาแฟสายพันธุ์อาราบิกานำมาเข้าสู่กระบวนการรับรอง ได้รับคะแนนด้าน Aftertaste อยู่ที่ประมาณ 7-8 คะแนน สังเกตได้จากการความหนาแน่นของข้อมูลใน Scatter plot, Boxplot, Stem and leaves และอื่นๆ ในขณะเดียวกัน มีบางกลุ่มให้ Aftertaste ที่ดีกว่า ซึ่งอาจเป็นผลจากตัวแปรในรายละเอียดอื่นๆ ที่เกี่ยวข้อง เช่น ด้านการดูแล พื้นที่เพาะปลูก เป็นต้น

ส่วนที่ 2 Sweetness

จาก Scatter plot Sweetness-Points จะกำหนดให้ Sweetness เป็นตัวแปรต้น แล้วให้ Total cup point เป็นตัวแปรตามเพื่อวิเคราะห์ถึงอิทธิพลของตัวแปร Sweetness จะสามารถเปลี่ยนแปลงคะแนนได้มากน้อยเพียงใด ซึ่งจากทุกๆ กราฟที่ใช้แสดงผล Sweetness จะเห็นได้ว่า โดยรวมๆ แล้ว จากชุดข้อมูลจะได้คะแนน Sweetness ค่อนข้างสูงและมีความถี่จากกราฟมาก อยู่ที่ 10 คะแนน เป็นคะแนนสูงสุด ซึ่งจะมีบางส่วนที่คะแนนลดลงตามมาเป็นลำดับๆ สามารถให้คำอธิบายอย่างคร่าวๆว่า Sweetness มีผลต่อการให้คะแนนมาก และสาเหตุที่ทำให้ Sweetness มีคะแนนกระจุกกันที่ 8-10 เนื่องจากการวัดระดับกาแฟ ผู้ต้องการการรับรองจาก CQI (Coffee-

Quality Institute) จะนำกาแฟที่ดีที่สุดมาประเมิน ซึ่ง Sweetness จะขึ้นอยู่กับคุณภาพการคั่วกาแฟ การคั่วจะทำให้เกิดสีน้ำตาลคาราเมลซึ่ง หากใช้เวลาในการคั่วนานขึ้น ซูโครสในเมล็ดกาแฟ และสารอินทรีย์ชนิดอื่นๆ จะถูกย่อยสลายมากขึ้น นำไปสู่สารประกอบคาราเมลที่ละเอียดและซับซ้อน ทำให้รสชาติหอม หวานละเอียด และกระจายตัวได้ดี ความหวานของกาแฟจึงเพิ่มมากขึ้น และได้รับคะแนนสูงจากการประเมินได้

สรุป

จากข้อมูลวิเคราะห์และกราฟข้างต้น Aftertaste คือคะแนนสัมผัสหลังดื่ม ส่วน Sweetness คือคะแนนโดยขึ้นจากสัมผัสความหวาน ซึ่งสามารถจำแนกได้จากรสชาติ ในการจัดมาตรฐานกาแฟต้องการความพิถีพิถันมากเป็นพิเศษ ในเบื้องต้น ถึงแม้ว่ามีข้อมูลหลายชุดได้รับ 10 คะแนนในส่วน Sweetness แต่เนื่องจากการประเมินประกอบด้วยตัวแปรอื่นๆ ด้วย จึงทำให้ตัวแปรตาม (Total cup point) มีค่าหลากหลาย ตั้งแต่คะแนนน้อย ไปยังคะแนนมาก ซึ่ง Aftertaste เป็นหนึ่งในเกณฑ์คะแนนเช่นกัน โดย Aftertaste มีคะแนนกระจายเป็นรูปกราฟทรงระฆังคว่ำ คะแนนส่วนมากอยู่ในช่วง 7-8 คะแนน พอให้เห็นแนวโน้มที่เป็นไปได้ของคะแนนรวมซึ่งแปรผันตรงกันกับคะแนน Aftertaste อย่างไรก็ตาม Total cup point ยังประกอบไปด้วยคะแนนของลักษณะพิเศษอื่นๆ ของกาแฟด้วย ทำให้แนวโน้มของคะแนนที่ได้รับแตกต่างจากคะแนนเพียง 2 ลักษณะข้างต้น

Source code

Twā Aftertaste.m

```
A = csvread("arabica_aftertaste.csv");

length(A)
minA = min(A);
maxA = max(A);
meanA = mean(A);
medA = median(A);
stdA = std(A);
modeA = mode(A);
varA = var(A);
pisaia = range(A);

figure
histogram(A)
title("Aftertaste Histogram")
xlabel("Aftertaste (points)")
ylabel("Samples")

figure
boxplot(A)
title("Aftertaste Boxplot")
ylabel("Aftertaste (points)")

figure
histogram(A, 'Normalization', 'probability') %pmf
title("Aftertaste PMF")
xlabel("Aftertaste (points)")
```

```
ylabel("Probability mass")
```

```
figure
```

```
histogram(A, 'Normalization', 'cdf') %cdf
```

```
title("Aftertaste CDF")
```

```
xlabel("Aftertaste (points)")
```

```
ylabel("Probability")
```

```
stemleafplot(A*10)
```

```
%box plot trimmed
```

```
A_trimmed = csvread("arabica_aftertaste_trimmed.csv");
```

```
meanA_trm = mean(A_trimmed);
```

```
figure
```

```
boxplot(A_trimmed)
```

```
title("Aftertaste Boxplot (trimmed)")
```

```
ylabel("Aftertaste (points)")
```

Twā Sweetness.m

```
S = csvread("arabica_sweetness.csv");
```

```
length(S)
```

```
minS = min(S);
```

```
maxS = max(S);
```

```
means = mean(S);
```

```
medS = median(S);
```

```
stdS = std(S);
```

```
modeS = mode(S);
```

```
varS = var(S);
```



```
pisaiS = range(S);
```

```
figure
```

```
histogram(S)
```

```
title("Sweetness Histogram")
```

```
xlabel("Sweetness (points)")
```

```
ylabel("Samples")
```

```
figure
```

```
boxplot(S)
```

```
title("Sweetness Boxplot")
```

```
ylabel("Sweeness (point)")
```

```
figure
```

```
histogram(S, 'Normalization', 'probability') %pmf
```

```
title("Sweetness PMF")
```

```
xlabel("Sweetness (points)")
```

```
ylabel("Probability mass")
```

```
figure
```

```
histogram(S, 'Normalization', 'cdf') %cdf
```

```
title("Sweetness CDF")
```

```
xlabel("Sweetness (points)")
```

```
ylabel("Probability")
```

```
stemleafplot(S*10)
```

```
%boxplot trimmed
```

```
S_trimmed = csvread("arabica_sweetness_trimmed.csv");
```

```
meanS_trm = mean(S_trimmed);
```

```

figure
boxplot(S_trimmed)
title("Sweetness Boxplot (trimmed)")
ylabel("Sweeness (point)")

%Scatter plot
A= csvread("arabica_aftertaste.csv");
pt= csvread("arabica_total.cup.point.csv");
figure
scatter(S,pt)
title("Scatter plot Sweetness-Points")
xlabel("Sweetness (points)")
ylabel("Total cup point (points)")
figure
scatter(A,pt)
title("Scatter plot Aftertaste-Points")
xlabel("Aftertaste (points)")
ylabel("Total cup point (points)")

```

Twā stemleafplot.m

(<https://www.mathworks.com/matlabcentral/fileexchange/30217-stem-and-leaf-plot>)

```

function stemleafplot(v,p)

if ~isnumeric(v); error 'Input V must be numeric'; end
if ~exist('p','var'); p = 0; elseif isempty(p); p = 0; end
if ~isnumeric(p); error 'Input P must be an integer'; end
p = round(p);

```

```
% Condition V
```

```
v = v(~isnan(v));
```

```
v = v(:);
```

```
v = roundn(v,p);
```

```
% Organize stems and leaves
```

```
allstems = floor(v./10^(p+1));
```

```
allleaves = round(abs(v./10^p));
```

```
nstems = allstems(allstems<0)+1;    % Negative stems
```

```
nstems = nstems(:);
```

```
pstems = allstems(~(allstems<0));    % Positive stems
```

```
pstems = pstems(:);
```

```
nleaves = allleaves(allstems<0);    % Negative leaves
```

```
nleaves = nleaves(:);
```

```
pleaves = allleaves(~(allstems<0)); % Negative leaves
```

```
pleaves = pleaves(:);
```

```
dig = ceil(max(log10(abs(allstems))))+1;    % Max # of digits in stem
```

```
form = strcat(['%' num2str(dig+1) 'i']);    % Format string for SPRINTF
```

```
% Plot negative stems
```

```
if ~isempty(nstems)
```

```
    for ii = min(nstems(:)):0
```

```
        strstem = sprintf(form,ii);
```

```
        if ii==0; strstem(end-1:end) = '-0'; end
```

```
        strleaves = sprintf('%2i',mod(sort(nleaves(nstems==ii)),10));
```

```
        s = strcat([strstem ' . |' strleaves]);
```

```
        disp(s)
```

```

    end % NSTEMS
end % IF

% Plot positive stems
if ~isempty(pstems)
    for ii = 0:max(pstems(:))
        strstem = sprintf(form,ii);
        strleaves = sprintf('%2i',mod(sort(pleaves(pstems==ii)),10));
        s = strcat([strstem ' ' strleaves]);
        disp(s)
    end % PSTEMS
end % IF

end % MAIN

```