DATASET 1: Weight of Apple in grams

Appendix 1:

Sno	Weight (g) of Apple
1	220
2	240
3	183
4	140
5	218
6	205
7	215
8	185
9	165
10	175
11	210
12	185
13	160
14	170
15	205
16	165
17	215
18	190
19	175
20	190
21	225
22	210
23	185
24	170
25	230
26	190
27	210
28	150
29	130
30	120
31	247
32	140
33	177
34	110
35	125
36	135
37	105
38	120

39	130
40	140
41	100
42	135
43	115
44	140
45	130
46	120
47	115
48	118
49	130
50	123
51	269
52	152
53	270
54	246
55	253
56	298
57	243
58	108
59	180
60	210
61	126
62	190
63	200
64	230
65	180
66	190
67	220
68	225
69	200
70	215
71	220
72	185
73	200
74	220
75	210
76	200
77	205
78	220
79	158
80	192
81	220

82	137
83	236
84	293
85	209
86	169
87	252
88	297
89	250
90	270
91	260
92	300
93	280
94	111
95	240
96	245
97	285
98	260
99	240
100	275
101	230
102	290
103	255
104	235
105	240
106	280
107	270
108	235
109	250
110	240
111	280
112	127
113	202
114	143
115	120
116	184
117	279
118	260
119	117

Sample Mean:

The mean ***19.042*** represents the average weight of the apples in the dataset, calculated by summing all the weights and dividing by the number of observations.

The mean was calculated using the formula:
=AVERAGE(DATASET1 READING)

Median:

The median is the middle value of the dataset when the values are arranged in ascending order. The median weight of the apple dataset is ***202 grams*** when all the data is arranged in ascending order, which means that half of the apples weigh less than 202 grams and the other half weigh more. The median is less affected by outliers and skewed data than the mean, offering a more robust measure of central tendency.

The median was calculated using the formula:
=MEDIAN(DATASET1 READING)

Sample Standard Deviation:

The standard deviation of ***53.88 grams*** in this dataset is a key indicator of the weights' variability, highlighting the diversity in apple sizes and providing insight into the distribution's spread around the average weight.

The Standard deviation was calculated using the formula:
=STDEV(DATASET1 READING)

MEAN	198.0420168
MEDIAN	202
STANDARD DEVIATION	53.87656875

Quartiles:

Q1 (1st Quartile): The median of the first half of the dataset, which is 150 grams for the data. This means that 25% of the apples weigh less than 151 grams.

Q1 was calculated using the formula:
=QUARTILE.EXC(B2:B120,1)

Q2 (2nd Quartile or Median): The median of the dataset, is 202 grams.

Q3 (3rd Quartile): The median of the second half of the dataset, which is 240 grams for the data. This indicates that 75% of the apples weigh less than 240 grams.

Q3 was calculated using the formula:
=QUARTILE.EXC(B2:B120,3)

QUARTILES	
Q1	150
Q2	202
Q3	240

Maximum:

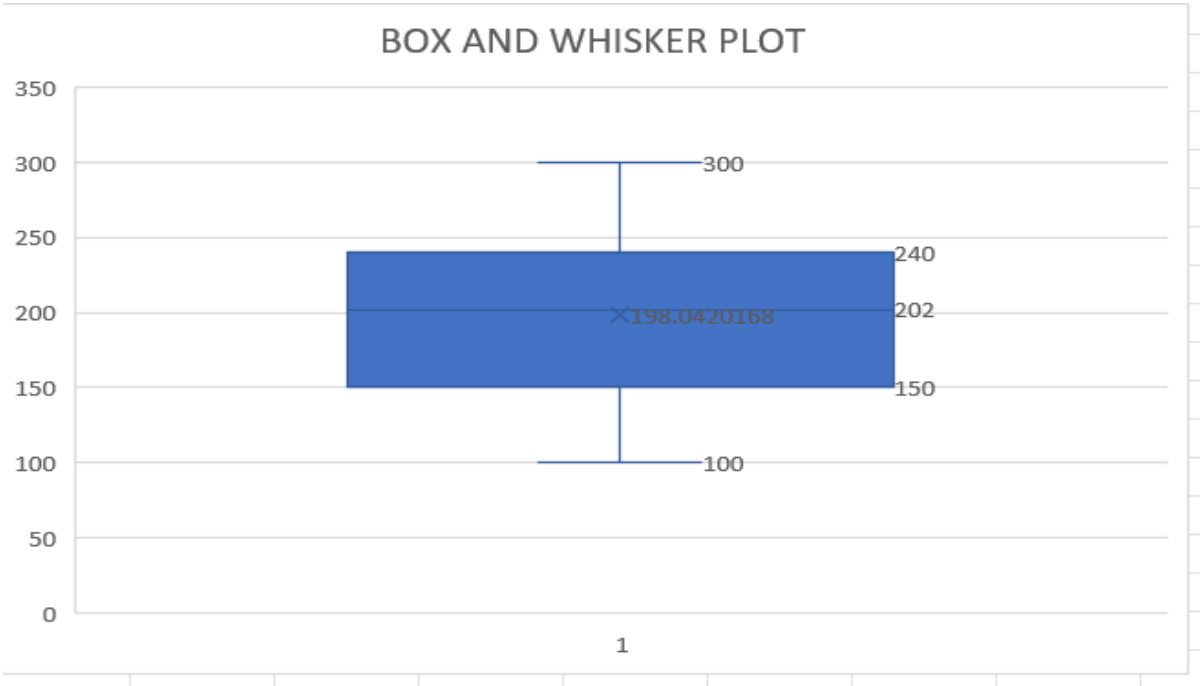
The maximum value was calculated using the formula:
=MAX(DATASET1 READING)

Minimum:

The minimum value of the weight of apples was calculated using the formula:
=MIN(DATASET1 READING)

MIN	100
MAX	300

Box Plot and Whisker Plot:



Central Box: The central box of the plot represents the middle 50% of the data, known as the interquartile range (IQR). The bottom of the box is at the first quartile (150 grams), and the top is at the third quartile (240 grams), which says that 50% of the apple weights fall between these two values.

Median: The line within the box marks the median of the dataset at 202 grams. This is the value that separates the higher half from the lower half of the dataset.

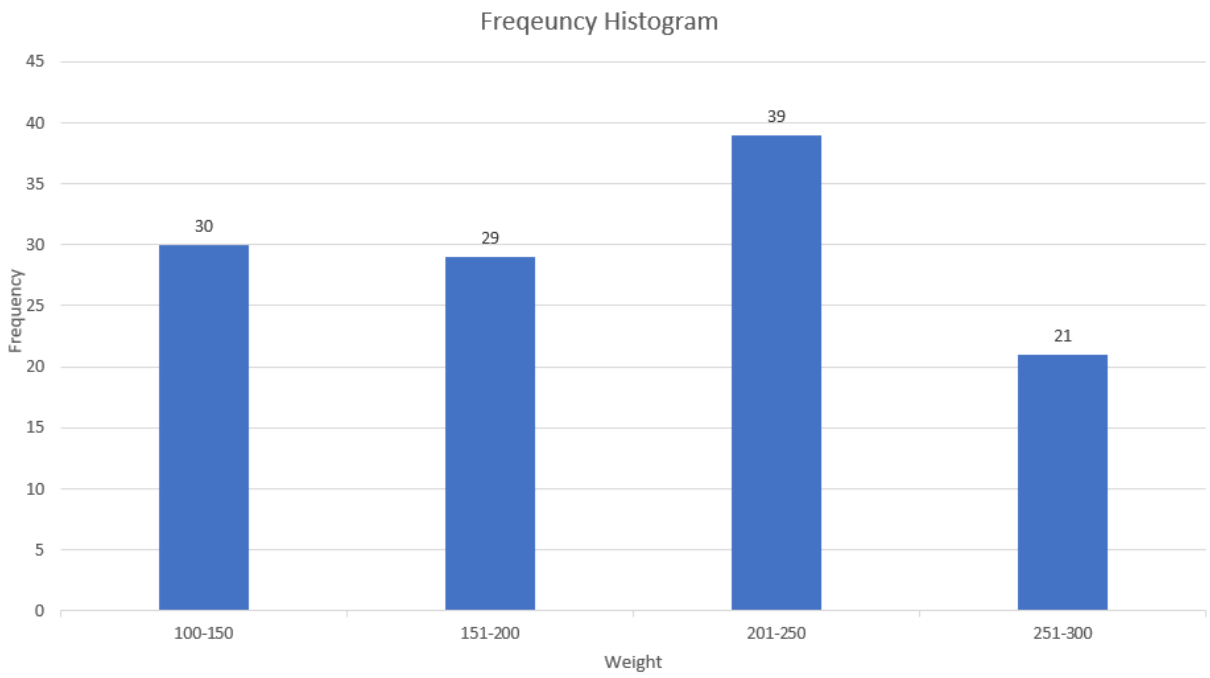
Whiskers: The "whiskers" extend from the box to the smallest and largest values within 1.5 times the IQR from the first and third quartiles, respectively. In the above plot, the lower whisker extends down to 100 grams, and the upper whisker extends up to 300 grams. These endpoints represent the typical range of the data, excluding outliers.

Frequency Table Distribution table:

A frequency table in the context of the "Weight (g) of Apple" dataset categorizes the apples based on their weights and counts how many apples fall into each weight category.

RANGE	Frequency
100-150	30
151-200	29
201-250	39
251-300	21

Frequency Histogram:



DATASET 2: Weight of Apple in grams

Appendix 2:

Sno	Person	Entry Time (UC subway)	Interval
1	P1	10:00:34	0
2	P2	10:01:04	30
3	P3	10:02:10	66
4	P4	10:02:41	31
5	P5	10:03:09	28
6	P6	10:03:20	11
7	P7	10:04:01	41
8	P8	10:05:37	96
9	P9	10:06:19	42
10	P10	10:06:31	12
11	P11	10:07:46	75
12	P12	10:07:54	8
13	P13	10:08:12	18
14	P14	10:10:00	108
15	P15	10:10:46	46
16	P16	10:12:49	123
17	P17	10:12:55	6
18	P18	10:14:20	85
19	P19	10:14:38	18
20	P20	10:16:15	97
21	P21	10:16:35	20
22	P22	10:16:56	21
23	P23	10:17:01	5
24	P24	10:17:56	55
25	P25	10:18:02	6
26	P26	10:19:44	102
27	P27	10:21:07	83
28	P28	10:21:15	8
29	P29	10:22:43	88
30	P30	10:24:38	115
31	P31	10:24:55	17
32	P32	10:25:00	5
33	P33	10:25:28	28
34	P34	10:26:25	57
35	P35	10:28:05	100

36	P36	10:30:06	121
37	P37	10:31:39	93
38	P38	10:32:35	56
39	P39	10:33:01	26
40	P40	10:33:47	46
41	P41	10:34:01	14
42	P42	10:34:07	6
43	P43	10:34:21	14
44	P44	10:34:22	1
45	P45	10:37:58	216
46	P46	10:39:23	85
47	P47	10:39:51	28
48	P48	10:40:33	42
49	P49	10:40:35	2
50	P50	10:40:54	19
51	P51	10:42:36	102
52	P52	10:42:38	2
53	P53	10:42:48	10
54	P54	10:43:32	44
55	P55	10:43:33	1
56	P56	10:44:55	82
57	P57	10:45:34	39
58	P58	10:45:47	13
59	P59	10:47:04	77
60	P60	10:48:08	64
61	P61	10:48:24	16
62	P62	10:48:39	15
63	P63	10:50:05	86
64	P64	10:51:13	68
65	P65	10:51:32	19
66	P66	10:51:39	7
67	P67	10:51:44	5
68	P68	10:52:32	48
69	P69	10:52:51	19
70	P70	10:53:22	31
71	P71	10:55:42	140
72	P72	10:56:25	43
73	P73	10:57:24	59
74	P74	10:57:41	17
75	P75	10:59:16	95
76	P76	10:59:21	5
77	P77	11:02:52	211
78	P78	11:03:18	26

79	P79	11:04:03	45
80	P80	11:04:44	41
81	P81	11:04:50	6
82	P82	11:05:43	53
83	P83	11:06:33	50
84	P84	11:06:54	21
85	P85	11:08:37	103
86	P86	11:11:37	180
87	P87	11:13:46	129
88	P88	11:14:34	48
89	P89	11:15:14	40
90	P90	11:15:51	37
91	P91	11:15:55	4
92	P92	11:16:38	43
93	P93	11:17:38	60
94	P94	11:19:58	140
95	P95	11:20:43	45
96	P96	11:20:59	16
97	P97	11:21:15	16
98	P98	11:21:27	12
99	P99	11:21:51	24
100	P100	11:24:11	140

Sample Mean:

The mean ***50.17*** seconds suggests that, on average, each entry into the subway occurs about 50 seconds after the previous one

The mean was calculated using the formula:
=AVERAGE(DATASET1 READING)

Median:

The median interval of ***40.5*** seconds indicates that half of the entry intervals are less than 40.5 seconds, and the other half are more, highlighting the midpoint of interval distribution.

The median was calculated using the formula:
=MEDIAN(DATASET1 READING)

Standard Deviation:

A standard deviation of ***46.32*** seconds in the data indicates a wide range of intervals between people entering the subway, pointing to significant variability around the average interval.

The Standard deviation was calculated using the formula:
=STDEV(DATASET1 READING)

MEAN	50.17
MEDIAN	40.5
STD DEVIATION	46.32233464

Quartiles:

Q1 (1st Quartile)- *15.25 seconds*: This is the '25th percentile.' It means that 25% of the intervals between entries are less than or equal to 15.25 seconds. A quarter of the people entering the subway do so within 15.25 seconds of the previous person.

Q1 was calculated using the formula:
=QUARTILE.EXC(B2:B120,1)

Q2 (2nd Quartile or Median) - *40.5 seconds*: This is the '50th percentile. Half of the intervals are less than or equal to 40.5 seconds.

Q3 (3rd Quartile*)* *- 80.75 seconds*: This is the '75th percentile.' It indicates that 75% of the intervals are less than or equal to 80.75 seconds, and the remaining 25% are more.

Q3 was calculated using the formula:
=QUARTILE.EXC(B2:B120,3)

QUARTILES	
Q1	15.25
Q2	40.5
Q3	80.75

Maximum:

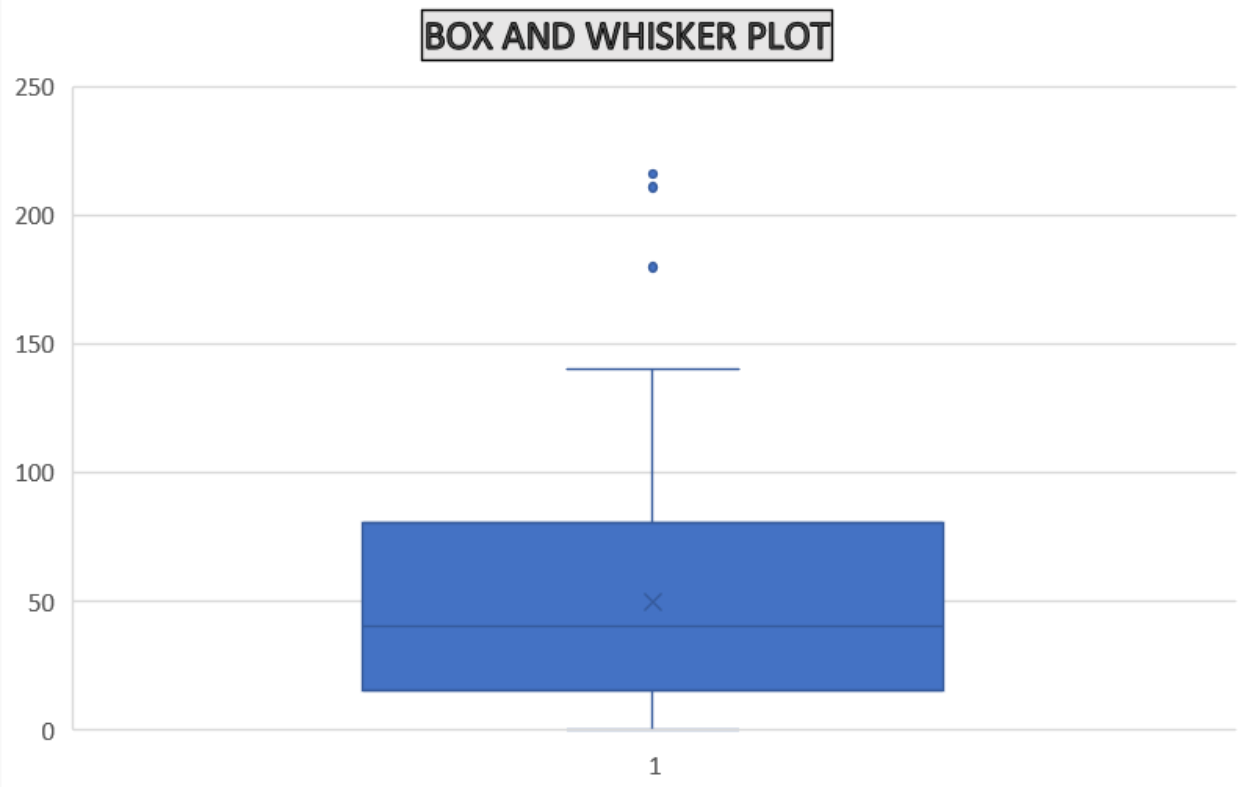
The maximum value of the person entering the UC Subway was calculated using the formula:
=MAX(DATASET1 READING)

Minimum:

The minimum value of the person entering the UC Subway was calculated using the formula:
=MIN(DATASET1 READING)

MIN	0
MAX	216

Box Plot and Whisker Plot:



First Quartile (Q1): The lower edge of the box, approximately at 15.25 seconds, signifies the first quartile. This value illustrates that 25% of the entry intervals are of this duration or less, reflecting a rapid succession of individuals entering the subway station.

Median (Q2): The central line within the box represents the median interval at 40.5 seconds. This indicates the middle value of the dataset, where half of the observed entry intervals fall below this figure and the other half exceed it.

Third Quartile (Q3): 80.75 seconds. This indicates that 75% of the intervals are this amount or shorter. The distance between the first and third quartiles, known as the interquartile range, encapsulates the central 50% of the data, providing a sense of the interval concentration.

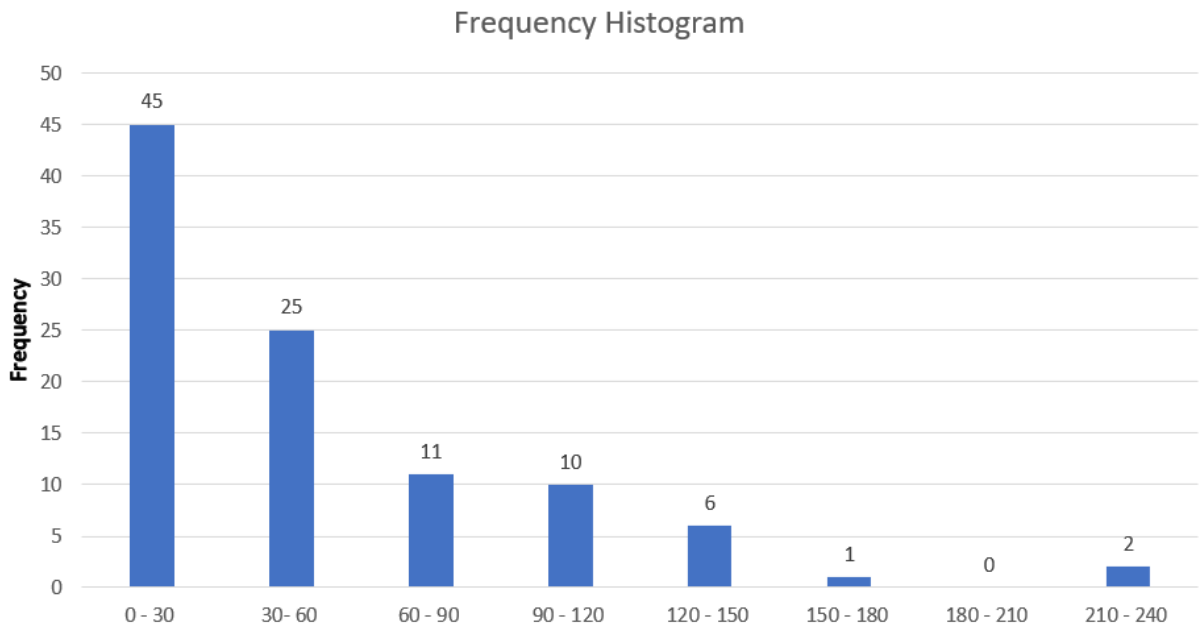
Whiskers: The lower whisker stretches down to the shortest interval, while the upper whisker extends to the longest interval within 1.5 times the interquartile range from the box.

Outliers: Points depicted above the upper whisker represent outlier intervals.

Frequency Table Distribution table:

Range	Frequency
0 - 30	45
30- 60	25
60 - 90	11
90 - 120	10
120 - 150	6
150 - 180	1
180 - 210	0
210 - 240	2

Frequency Histogram:



The X-axis (horizontal axis) represents the range of intervals between subway entries.
The Y-axis (vertical axis) shows the frequency, which is the number of times intervals within a specific range occurred.

The histogram here represents the frequency of intervals at which people enter a subway station. Each bar shows how many times people entered within a certain time range.

Conclusion:

Dataset 1: Follow an exponential distribution as the most frequent weight class of weight of Apple is skewed towards the higher end (201-250 grams) rather than the center.

Dataset 2: Follow an exponential distribution because the average time between subway entries is longer than the most common (median) time, suggesting that while most people enter in quick succession, there are a few longer waits that push the average up.