

California State University - East Bay

MS in Business Analytics

BAN 602 Case 1 Solution

Group 5

- | | |
|-----------------------------|--------------------------------|
| 1. Aastha Tandon | atandon@horizon.csueastbay.edu |
| 2. Sukirti Bhat | sbhat2@horizon.csueastbay.edu |
| 3. Venu Madhava Reddy Kotla | vkotla@horizon.csueastbay.edu |
| 4. Shaik Ahamed | sahamed@horizon.csueastbay.edu |

Lecturer:

Somak Paul

Submission Date : August 31,2022

1 Solution

Summary of Quantitative variables							
Summary	Mean	Median	Standard Deviation	Variance	Range	Minimum	Maximum
Time Min	12.81	11.40	6.06	36.76	28.60	4.30	32.90
Pages Viewed	4.82	4.50	2.04	4.15	8.00	2.00	10.00
Amount Spent	68.13	62.15	32.34	1046.12	140.67	17.84	158.51

Table 1: Descriptive statistics

1. Time Min: Minimum and maximum time spent by a user is 4.30 and 32.90 respectively. Average time spent by a user is 12.81. Median time is 11.40. Since mean is higher than median, data is positively right skewed.
2. Pages Viewed: Minimum and maximum Pages viewed are 2 and 10 respectively. Average Pages viewed by a user is 4.82. Median Pages viewed is 4.50. Since mean is higher than median, data is positively right skewed.
3. Amount Spent: Minimum and maximum Amount spent are 17.84 and 158.51 respectively. Average Amount spent by a user is 68.13. Median Amount Spent is 62.15. Since mean is higher than median, data is positively right skewed.

For all 3 variables in Table 1, Mean is higher than Median. Amount Spent has higher variability compared to Time min and Pages Viewed hence Variance is higher.

2 Solution

1. Visual Measure:

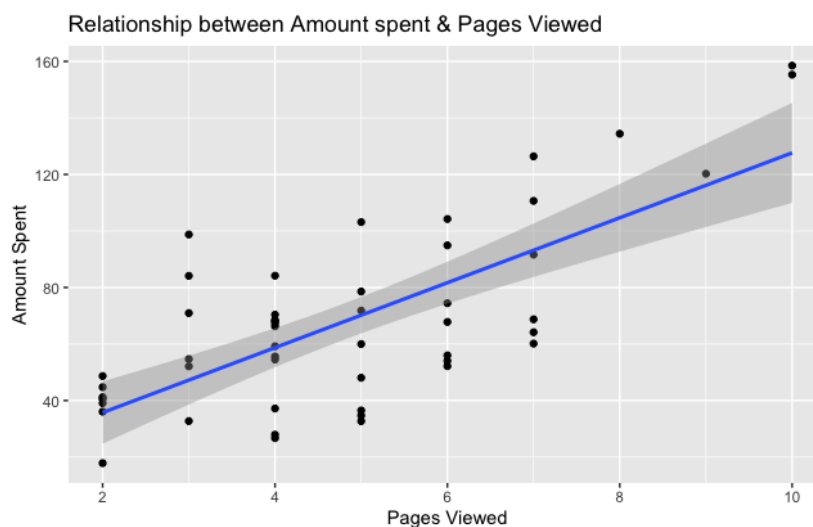


Figure 1: Scatterplot

- (a) Scatter Plot in Figure 1 indicates a positive relationship between Pages viewed and Amount spent.

- (b) Higher Pages viewed are associated with a higher Amount spent.
- (c) The relationship is not perfect since all plotted point in the scatter plot are not on a straight line.

2. Numerical Measure:

- (a) Covariance is a numerical measure which shows a linear association between two variables. Covariance of Pages Viewed and Amount spent is 47.685. Since it is a positive number, there is a positive relationship between two variables.
- (b) Correlation is another numerical measure of linear association between two variables. Correlation of Pages Viewed and Amount spent is 0.724. Since the value is near 1, it indicates strong positive linear relationship between two variables.

3 Solution

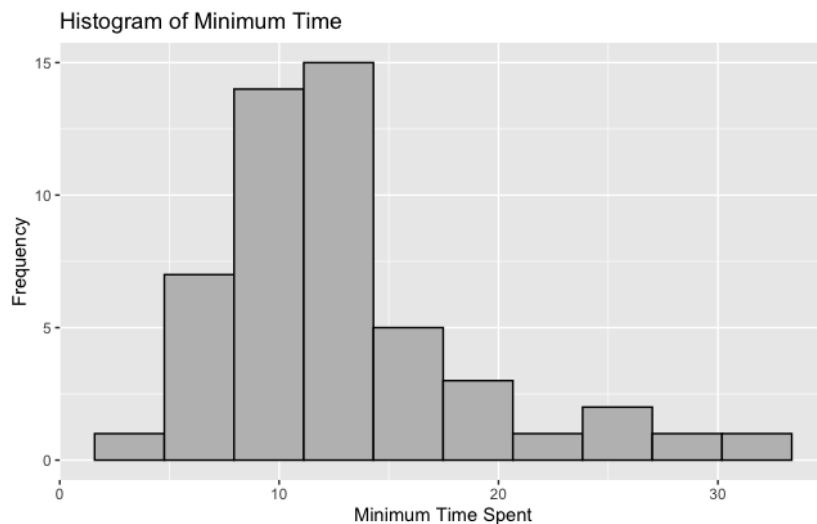


Figure 2: Histogram

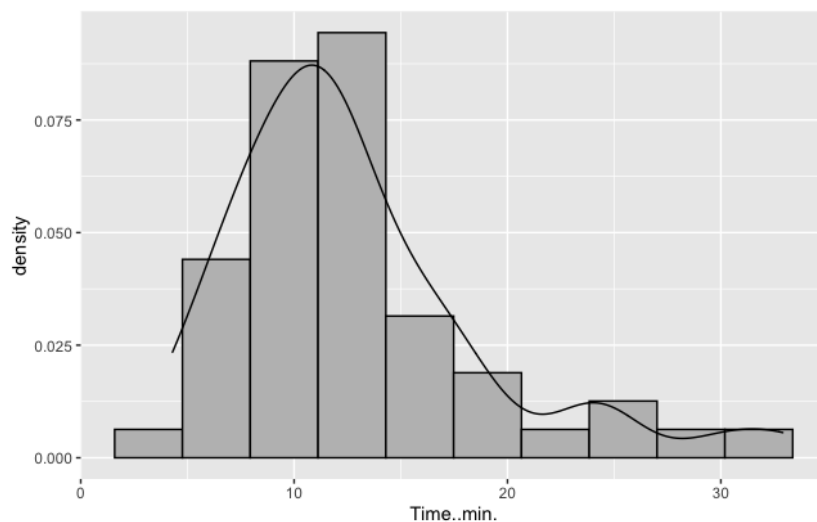


Figure 3: Histogram with Density curve

The histogram in Figure 2 shows a frequency distribution for minimum time spent on website. The height of each bar indicates count of minimum time spent in that range. Below are some observations from the histogram.

1. Minimum time spent by a user ranges from 4.30 to 32.90.
2. Most of the users are in interval of 11-14.
3. As the time spent on the website increases the frequency of the viewers decreases. Large proportion of people spend the time in the range of 5 min to 20 min.
4. The distribution is right skewed and has a longer tail to the right and can be seen with a plotted density curve in Figure 3.
5. Since the histogram is skewed right, mean (12.81) is greater than median(11.40).

4 Solution

Cross Tabulation				
	Chrome	Firefox	Others	Row Total
Fri	4	6	1	11
Mon	5	2	2	9
Sat	3	3	1	7
Sun	3	1	1	5
Thu	4	1	0	5
Tue	4	2	1	7
Wed	4	1	1	6
Col Total	27	16	7	50

Table 2: Cross Tabulation

Row Percentages				
	Chrome	Firefox	Others	Row Total
Fri	36.364%	54.545%	9.091%	100%
Mon	55.556%	22.222%	22.222%	100%
Sat	42.857%	42.857%	14.286%	100%
Sun	60.000%	20.000%	20.000%	100%
Thu	80.000%	20.000%	0.000%	100%
Tue	57.143%	28.571%	14.286%	100%
Wed	66.667%	16.667%	16.667%	100%

Table 3: Row Percentages

Observations from Table 3 :

1. On Sunday, Monday, Tuesday, Wednesday and Thursday, Chrome was mostly used by the users.
2. On Friday, Firefox was majorly used browser.

3. On Saturday, Chrome and Firefox were equally used.
4. On Monday, Wednesday and Sunday, Firefox and other browsers were equally used.
5. On Thursday, other browsers were not used.

Column Percentages			
	Chrome	Firefox	Others
Fri	14.815%	37.500%	14.286%
Mon	18.519%	12.500%	28.571%
Sat	11.111%	18.750%	14.286%
Sun	11.111%	6.250%	14.286%
Thu	14.815%	6.250%	0.000%
Tue	14.815%	12.500%	14.286%
Wed	14.815%	6.250%	14.286%
Column Total	100%	100%	100%

Table 4: Column Percentages

Observations from Table 4 :

1. Chrome was majorly used on Monday.
2. Chrome was equally used on Tuesday, Wednesday, Thursday and Friday (14.815%).
3. Chrome was least used on Saturday and Sunday.
4. Firefox was majorly used on Friday.
5. Firefox was equally used on Monday and Tuesday (12.5%).
6. Firefox was least used on Wednesday, Thursday and Sunday.
7. Other browser was majorly used on Monday and not used at all on Thursday.
8. On the remaining days, other browser were equally used (14.286%).

5 Solution

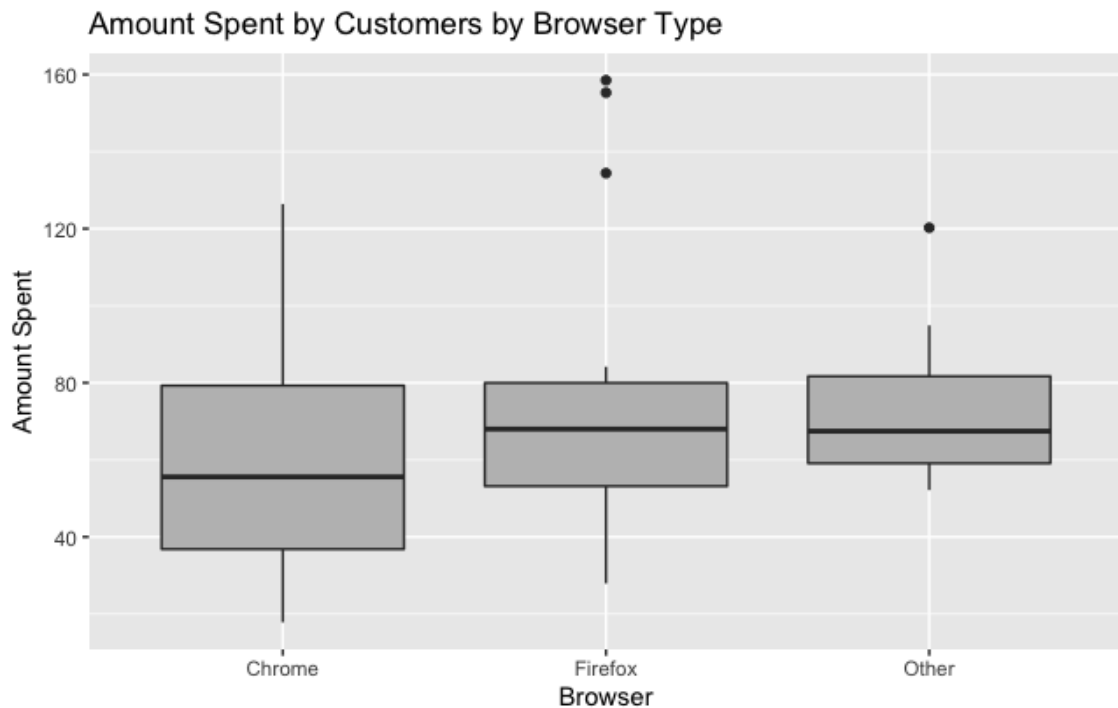


Figure 4: Boxplot

Quartiles of Amount Spent			
	Q1	Q2	Q3
Chrome	36.825	55.580	79.275
Firefox	53.160	67.985	79.978
Other	59.100	67.440	81.675

Table 5: Quartiles of Amount Spent for different browser

As per Box plot in Figure 4 , there are 3 outliers present in Firefox, 1 in other browser types and no outlier in Chrome.

Table 5 shows the first, second and third Quartiles of amount spent for each of the three browsers.

1. Interquartile Range (IQR) of Chrome is 42.450 (Q3-Q1)
2. Interquartile Range (IQR) of Firefox is 26.817
3. Interquartile Range (IQR) of Others is 22.575

IQR of Chrome is the highest followed by Firefox. Other browsers have least IQR. Q3 of Chrome, Firefox and other browsers are very near to each other.