

MARKETING ANALYTICS



ASSIGNMENT-5

RESTAURANT TIP ANALYSIS

GROUP-3

BANOTH KEERTHANA - 21110042
KANUGULA TRIVENI - 21110095
MADINENI DIVYA -21110115
RUDRESHWAR - 21110172

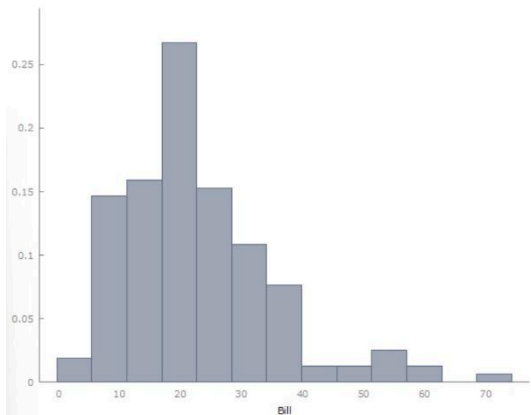
BILL AMOUNT ANALYSIS

Summary statistics, using the observations 1 - 157
for the variable 'Bill' (157 valid observations)

Mean	22.729
Median	20.220
Minimum	1.6600
Maximum	70.510
Standard deviation	12.157
C.V.	0.53488
Skewness	1.2283
Ex. kurtosis	1.9562
5% percentile	7.7540
95% percentile	50.384
Interquartile range	13.835
Missing obs.	0

From the given dataset, out of 157 observations the average bill that was made by customers is 22.79, median of the bill is 20.22,max bill made is 70.0 .

FREQUENCY DISTRIBUTION OF THE BILLS



Frequency distribution for Bill, obs 1-157
number of bins = 13, mean = 22.729, sd = 12.1572

interval	midpt	frequency	rel.	cum.
< 5.7375	2.8688	3	1.91%	1.91%
5.7375 - 11.475	8.6063	23	14.65%	16.56% *****
11.475 - 17.213	14.344	25	15.92%	32.48% *****
17.213 - 22.950	20.081	42	26.75%	59.24% *****
22.950 - 28.688	25.819	24	15.29%	74.52% *****
28.688 - 34.425	31.556	17	10.83%	85.35% ***
34.425 - 40.163	37.294	12	7.64%	92.99% **
40.163 - 45.900	43.031	2	1.27%	94.27%
45.900 - 51.638	48.769	2	1.27%	95.54%
51.638 - 57.375	54.506	4	2.55%	98.09%
57.375 - 63.112	60.244	2	1.27%	99.36%
63.112 - 68.850	65.981	0	0.00%	99.36%
>= 68.850	71.719	1	0.64%	100.00%

It shows the frequency of the bills in the particular range and you can see that most of the bill range is in 17-22 units as it is having more frequency in that range. And no bill is made in the range 63-68.

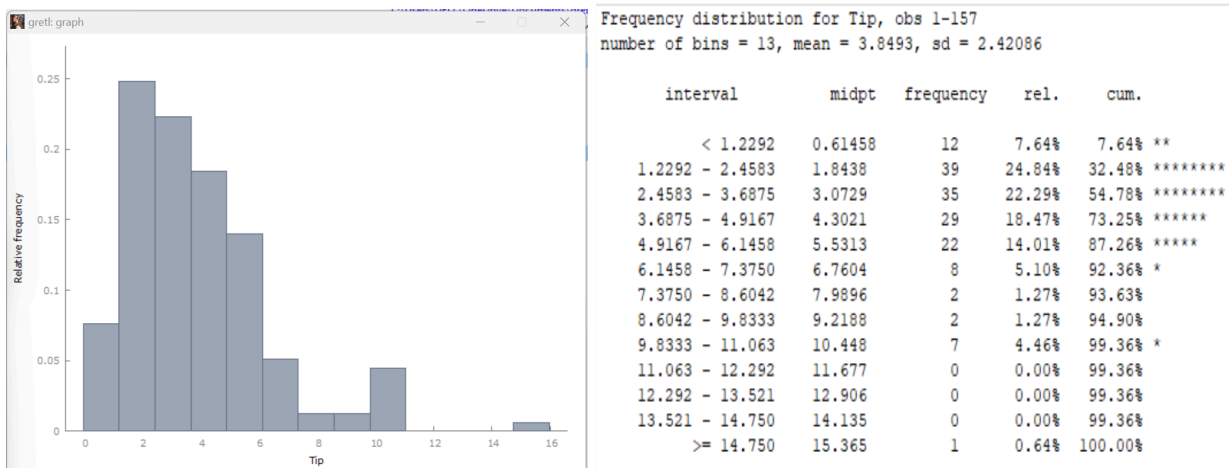
TIP AMOUNT ANALYSIS

Summary statistics, using the observations 1 - 157
for the variable 'Tip' (157 valid observations)

Mean	3.8493
Median	3.3500
Minimum	0.25000
Maximum	15.000
Standard deviation	2.4209
C.V.	0.62891
Skewness	1.4892
Ex. kurtosis	2.9239
5% percentile	1.0000
95% percentile	10.000
Interquartile range	2.9500
Missing obs.	0

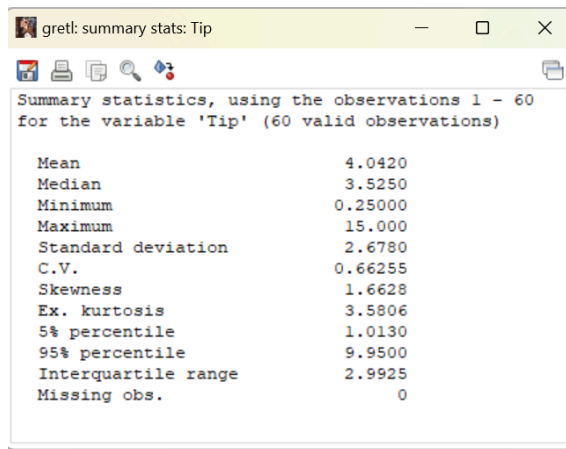
The above is tip analysis , the average tip that was given by customers is 3.8, median of the tip 3.3 ,max and min are 15 and 0.25 respectively.

FREQUENCY DISTRIBUTION OF THE TIPS

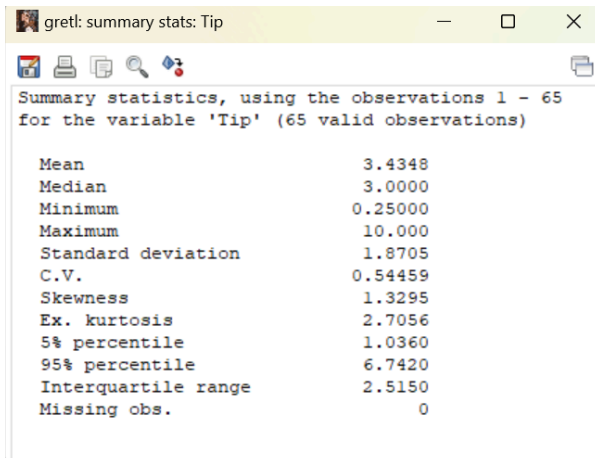


The next shows the frequency of the tips in the particular range and you can see that most of the tip range is in 1.2-2.4 units as it is having more frequency in that range. And no tip is made in the range 11-14.7.

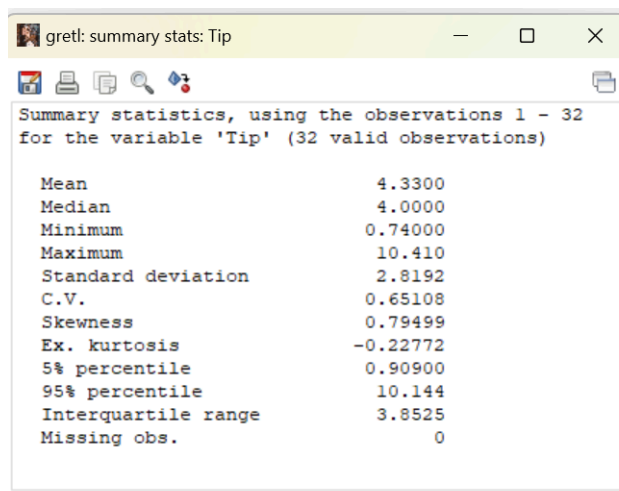
Analysis of Tip on basis of server



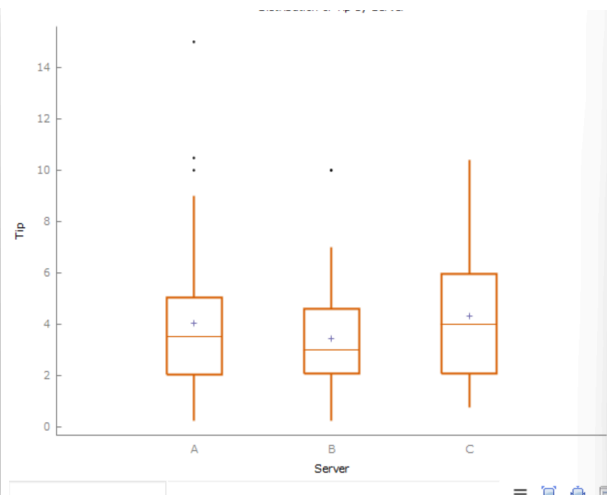
Tip summary statistics for Server- A



Tip summary statistics for Server- B



Tip summary statistics for Server- C



Distribution of Tip by Server

This graph shows the average tip amounts for each server, We consider Tip Amounts on Y axis and Individual server on X- axis. Each bar represents the average tip amount for a particular server. The outliers shown as the dots outside the main range represent exceptionally high tips received by a server on certain occasions.

When we observe the three plots, server B has more frequency compared than the other two. But coming to the tip analysis he was getting less tips than the others.

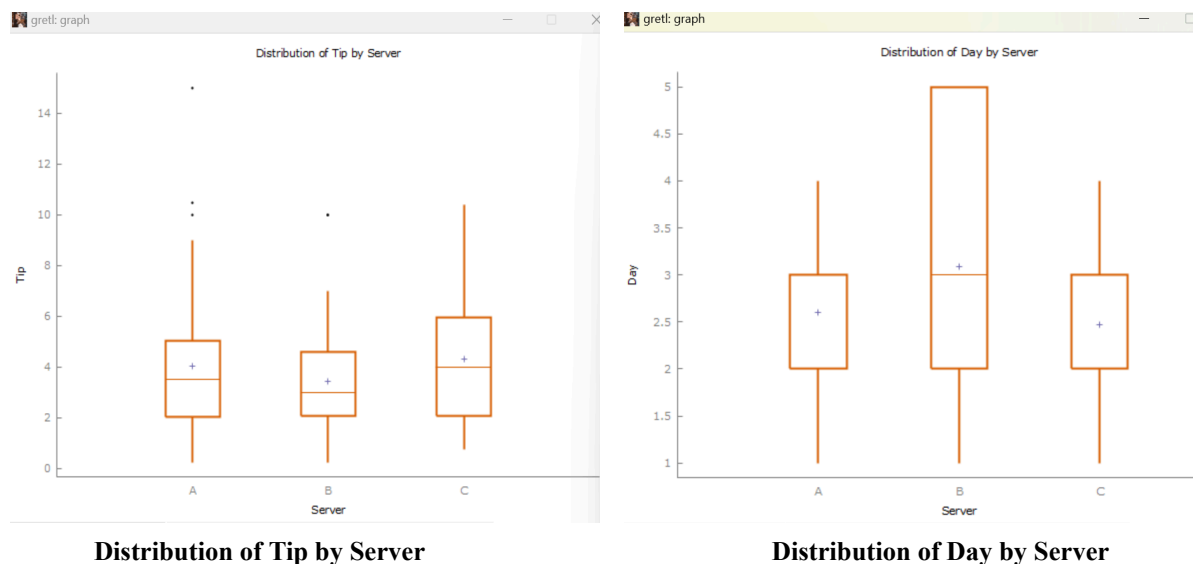
We can see that Server C has a higher median value, indicating that they receive more tips on average. Based on our analysis, we can conclude that Server C performs best in terms of tip

earnings. This could be due to their ability to connect with customers, provide better service, and have good communication skills.

There are several factors that could impact a server's tip earnings. For example:

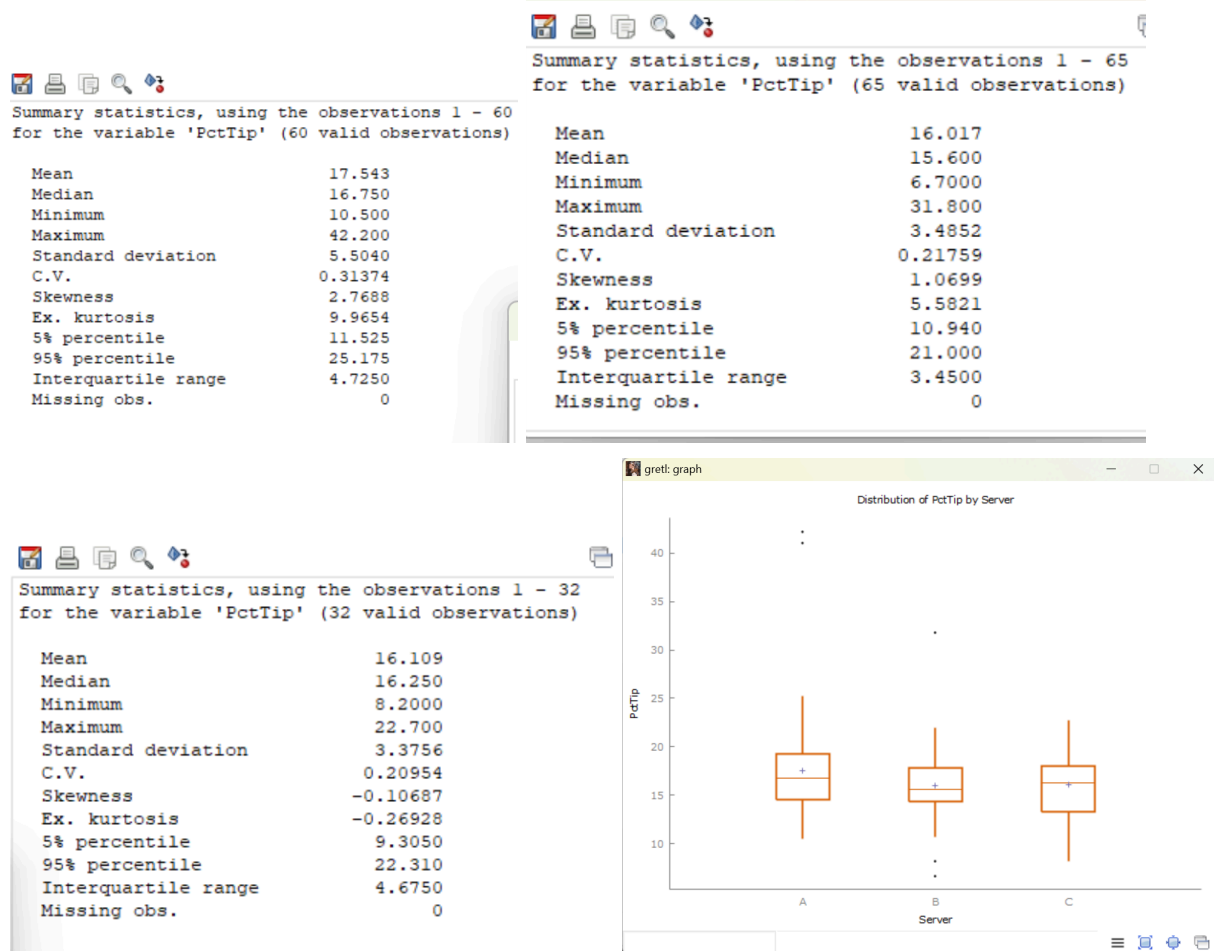
- * Ability to connect with customers
- * Quality of service
- * Appearance and communication skills
- * Time of shift worked

Analysis of Tip by Server by Day



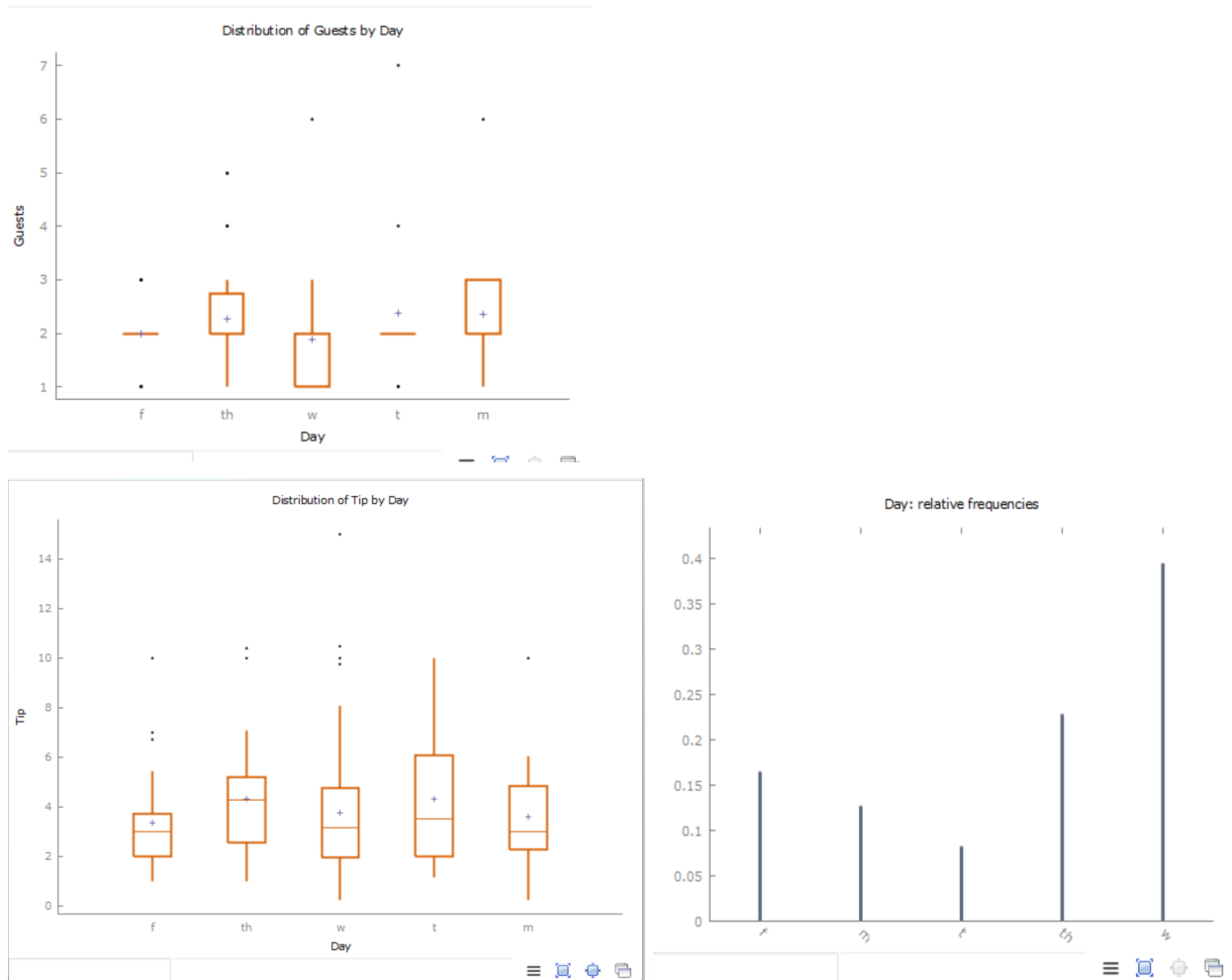
The mean value of tip is highest for server C, which is 4.33. The maximum value of tip is 15 and was received by server A. The minimum value of tip is 0.25 and was received by both servers A and B. Server C has received a larger range of tip amounts than A or B. Server C has served a minimum number of customers (32) but has highest mean value of tip which indicates that he may be the best server among the three. Server B was working for more no.of days, but he was not getting more tips compared than the other servers. Server B is getting lower tips due to less effective service and working during the less busy times. Management might consider adjusting Server B's shifts to more profitable times if the goal is to increase their tip earnings. Like Evening shifts, for instance, might attract customers who tend to tip more generously. Comparing Server B's performance on different days with other servers could help affect their tip earnings.

Percentage Tip of Each Server



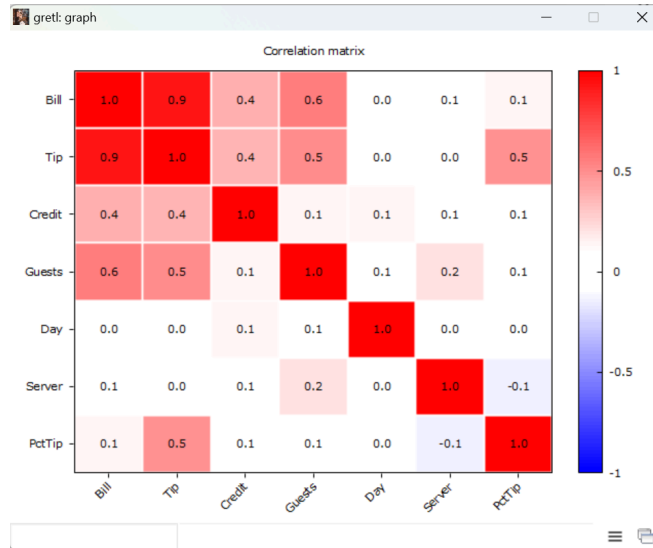
This represents the %tip(w.r.t bill) values received by each server, this indicates that on an average server A is receiving high %tip(having frequency 60). But server C is also receiving almost the same on an average with only 32 time serving. so we cannot conclude that server A is the best, just looking at the average tips received by it. There might be instances A or C that are the best servers. But we can say that server a, c is better than b(having high frequency but less avg value) in terms of receiving tip%.

Analysis of Percentage Tip with respect to server



From the above the average tip is high on tuesday and also the number of guests are less. In the above we can see that having less no. of working days on tuesday and also no. of guests less but the tip is being received more compared to others. So any server working on this day will be benefiting in terms of tip by serving for fewer people.

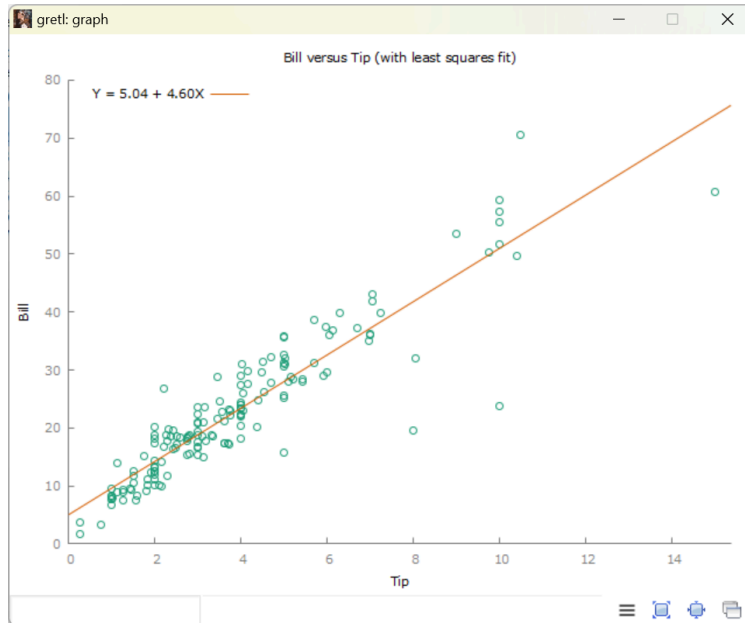
CORRELATION BETWEEN VARIABLES



The above represents the correlation (represents the strength and direction of linearity) between one variable and the other. High positive value indicates that when one increases the other increases (directly proportional), 0 represents they are independent and high negative value indicates that when one increases the other decreases strongly (inversely proportional).

For instance, (tip bills) have a value of 0.9 which indicates that with the increase in bill there will also be an increase in tip (vice versa), as they are highly linearly directly proportionally related. Bill, tip, server, %tip are not at all depending on the day as they are having value of 0 and independent of each other.

BILLS VS TIPS



Understanding the Scatter Plot:

This scatter plot illustrates the relationship between the total bill amount and the tip left at the restaurant. Each point represents a single observation (a customer's bill and corresponding tip). The orange line is a regression line, which shows the best-fit linear relationship between the two variables.

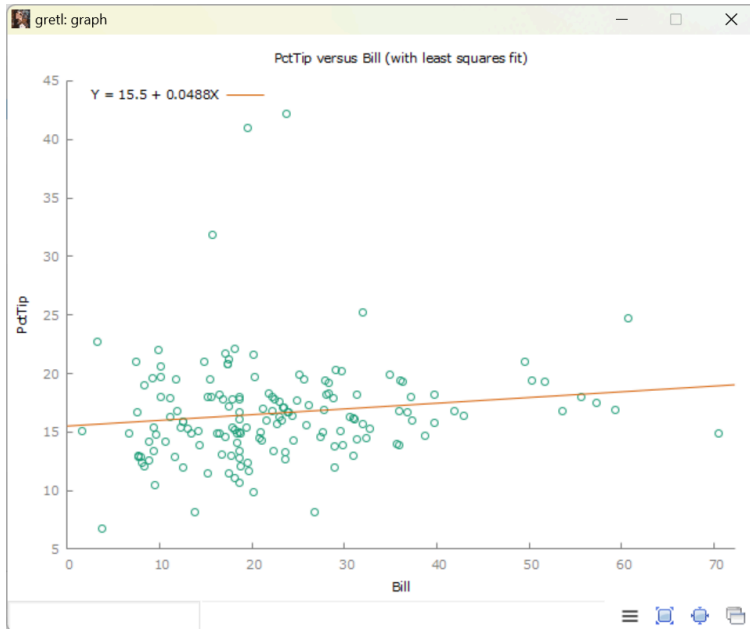
Key Observations:

Positive Correlation: The scatter plot shows a clear positive correlation between the bill amount and the tip. This means that as the bill amount increases, the tip tends to increase as well.

Regression Line: The regression line indicates the overall trend in the data. The equation of the line ($Y = 5.04 + 4.60X$) shows that for every unit increase in the bill amount (X), the tip (Y) is expected to increase by 4.60 units.

Scatter Around the Line: While there is a general trend, there is also some scatter around the regression line. This indicates that other factors besides the bill amount may influence the tip amount.

BILLS VS PERCENTAGE TIPS



Understanding the Scatter Plot:

This scatter plot illustrates the relationship between the total bill amount and the percentage tip left at a restaurant. Each point represents a single observation (a customer's bill and corresponding percentage tip). The orange line is a regression line, which shows the best-fit linear relationship between the two variables.

Key Observations:

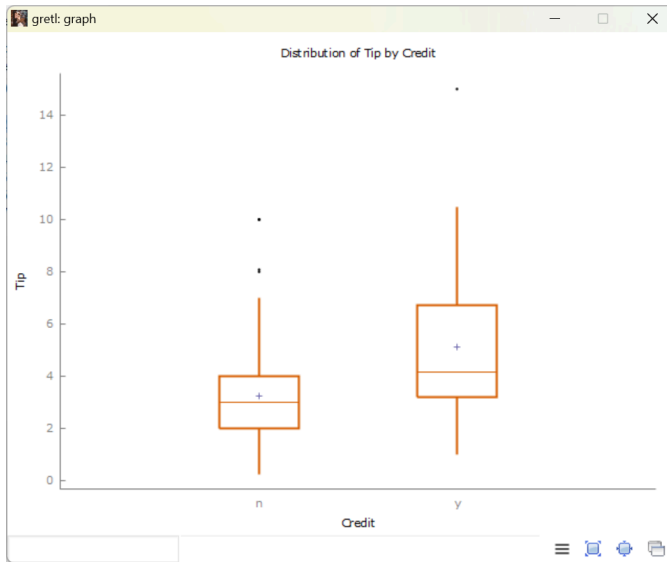
Weak Positive Correlation: The scatter plot shows a weak positive correlation between the bill amount and the percentage tip. This means that there's a slight tendency for larger bills to have higher percentage tips, but the relationship is not very strong.

Regression Line: The regression line indicates the overall trend in the data. The equation of the line ($Y = 15.5 + 0.0488X$) shows that for every unit increase in the bill amount (X), the percentage tip (Y) is expected to increase by 0.0488 units.

Scatter around line: There is a significant amount of scatter around the regression line, indicating that other factors besides the bill amount have a stronger influence on the percentage tip.

By analyzing these aspects, you can gain a better understanding of the factors that influence tipping behavior and potentially make predictions about future percentage tips.

ANALYSIS OF TIPS ON BASIS OF CREDIT



This box plot visually represents the distribution of tips based on whether a customer used credit (Y) or not (n). Each box represents a group, and the lines within the box show the median, quartiles, and potential outliers.

Key Observations:

Median Tip: The median tip for customers who used credit (Y) appears to be slightly higher than those who didn't (n).

Interquartile Range (IQR): The IQR, which represents the spread of the middle 50% of the data, seems to be similar for both groups. This suggests that the variability in tips around the median is comparable.

Outliers: There are a few potential outliers on the higher end for both groups, indicating some customers left exceptionally large tips.

Distribution: The overall shape of the box plots suggests that the distribution of tips is somewhat skewed to the right for both groups, with a longer tail on the right side. This means there are a few larger tips that pull the mean to the right.

Overall, the analysis suggests that the tip amount is primarily influenced by the bill amount and the number of guests. Other factors like day, server, and credit do not seem to have a significant impact on the tip.