Analyzing Pizza Brand with Customer Preference

NEEL BHAVESH NAIK
23MBD021
Advance Statistical Method

Title: Analyzing Pizza Brand with Customer Preference

Introduction:

Pizza is a universally beloved food that has captured the hearts and taste buds of people around the world. Its versatility, convenience, and endless variety of toppings make it a popular choice for meals and gatherings. To gain a deeper understanding of people's preferences and trends related to pizza consumption, a survey was conducted to explore various aspects of pizza choices and habits.

The survey aimed to uncover insights into topics such as favorite pizza, preferred Brands, frequency of pizza consumption, and ordering habits. By analyzing the survey data, we can gain valuable insights that can inform businesses in the food industry, marketers, and pizza enthusiasts alike.

This report presents an analysis of the survey data, including descriptive statistics, visualizations, and key findings. The analysis provides a snapshot of current pizza preferences and behaviors, highlighting trends and patterns that can help understand the ever-evolving landscape of pizza consumption.

Through this analysis, we hope to provide a glimpse into the world of pizza, shedding light on what makes this dish so beloved and how people's preferences shape the pizza industry.

Objective:

The objective of this study is to analyze a pizza survey to gain insights into people's preference:

Descriptive Analysis:

- Identify a Mean, Median, Mode, Variance, Standard Deviation Ratings of different variable i.e is Quality and Size of Pizza
- Identify Kurtosis and Skewness Ratings of different variable i.e is Quality and Size of Pizza

Regression Analysis:

- Can we predict the probability of Like it based on Menu list and pizza size (Logistic Regression)
- How do Size Rating, Quality rating and Pizza Size impact Restaurants (MLR)
- How closely does the linear model fit the relationship between Gender and Menu list or Restaurants (Least square method)

Hypotheses testing:

- Is there a significant association between Brands and Factor Influencing and test the hypothesis (Chi-square)
- Is there a significant difference in the mean frequency of Quality rating and restaurants test the hypothesis (ANOVA)
- Are there Mean differences in the frequency of Restaurants based on Menu list and Quality rating test with different types of tests of hypothesis (MANOVA)

Graphical Representation:

- How does a pie chart represent the distribution of pizza consumption and restaurants among survey
- How does a bar plot illustrate the frequency of menu list and the frequency of Order Methods
- How does a histogram display the frequency of menu list and restaurants among survey
- How does a boxplot show the frequency of Order method ad Menu list among survey

Contingency Table:

- Contingency table with respect to Brand and Quality Rating
- Contingency table with respect to Gender and Like it

Methodology:

Data Collection:

The data for the pizza survey can be collected using a combination of online survey. The survey can be designed to capture a wide range of information related to pizza preferences.

Survey design:
Develop a comprehensive survey questionnaire that covers various aspects of pizza. It include Multiple-choice questions to gather detailed responses.

Sampling strategy:
Determine the target population for the survey

Data Collection Methods:
Online Surveys: Distribute the survey questionnaire through online platforms to friends and family via Email and Social Media.

The data has been collected and stored into .CSV format in excel sheet

Timestamp	ID	Gender	Like_it	Frequency	Menu_list	Pizza_size	Restaurant	Factors_Ir	Satisfied	Size_Awar	Size_rating	Factors_In	Quality_Ra	Order_Me	Brand	Region_for_Brand
2024/03/04 12:10:23 PM GMT+5:30	1	Male	YES	Occasiona	Chicken	8	Dominos	Quality an	Yes	Partial awa	4	Appetite	4	Dine-in	Dominos	Quality
2024/03/04 12:13:11 PM GMT+5:30	2	Female	YES	Occasiona	Margherita	6	Dominos	Quality an	Yes	Partial awa	4	No. of pec	4	Dine-in	Dominos	Both
2024/03/04 1:41:19 PM GMT+5:30	3	Female	YES	Occasiona	Vegies	8	Dominos	Quality an	Yes	Partial awa	4	No. of pec	4	Delivery	Dominos	Both
2024/03/04 2:11:11 PM GMT+5:30	4	Female	YES	Occasiona	Chicken	8	Pizza expre	Quality an	Yes	Partial awa	4	No. of pec	4	Delivery	Dominos	Both
2024/03/04 3:48:11 PM GMT+5:30	5	Male	YES	Occasiona	Margherita	8	Dominos	Taste	Yes	Partial awa	4	Appetite	4	Dine-in	Dominos	Both
2024/03/04 3:54:52 PM GMT+5:30	6	Female	YES	Occasiona	Chicken	12	Pappa john	Taste	Yes	Yes, aware	5	Appetite	5	Delivery	Dominos	Quality
2024/03/04 4:03:28 PM GMT+5:30	7	Female	YES	Occasiona	Other	6	Dominos	Taste	Yes	Yes, aware	3	No. of pec	4	Delivery	Dominos	Both
2024/03/04 4:07:10 PM GMT+5:30	8	Female	YES	Occasiona	Chicken	6	Pizza hut	Quality an	Yes	Partial awa	5	Occasion	5	Dine-in	Dominos	Both
2024/03/04 4:09:01 PM GMT+5:30	9	Female	YES	Occasiona	Chicken	8	Dominos	Taste	Yes	Partial awa	4	No. of pec	3	Takeout	Dominos	Quality
2024/03/04 4:10:43 PM GMT+5:30	10	Male	YES	Occasiona	Other	8	Pizza hut	Quality an	Yes	Yes, aware	4	Appetite	3	Dine-in	Dominos	Both
2024/03/04 4:12:47 PM GMT+5:30	11	Female	YES	Occasiona	Chicken	8	Pizza hut	Price	Yes	Partial awa	3	No. of pec	4	Takeout	Dominos	Both
2024/03/04 4:14:09 PM GMT+5:30	12	Female	YES	Occasiona	Chicken	12	Dominos	Quality an	Yes	Yes, aware	5	No. of pec	5	Delivery	Dominos	Both
2024/03/04 4:15:32 PM GMT+5:30	13	Female	YES	Occasiona	Vegies	8	Dominos	Taste	Yes	Yes, aware	3	No. of pec	4	Delivery	Dominos	Both
2024/03/04 4:16:18 PM GMT+5:30	14	Male	YES	Twice in a	Vegies	8	Dominos	Taste	Yes	No, not aw	4	No. of pec	5	Takeout	Dominos	Quality
2024/03/04 4:25:42 PM GMT+5:30	15	Male	NO	Occasiona	Margherita	8	Pappa john	Taste	Yes	Partial awa	3	Appetite	1	Takeout	Pizza hut	Quality
2024/03/04 4:30:18 PM GMT+5:30	16	Female	YES	Occasiona	Chicken	8	Pizza hut	Quality an	Yes	Yes, aware	3	No. of pec	4	Delivery	Pizza hut	Both
2024/03/04 4:36:49 PM GMT+5:30	17	Female	YES	Occasiona	Chicken	6	Dominos	Quality an	Yes	Partial awa	3	No. of pec	3	Takeout	Dominos	Both
2024/03/04 4:44:27 PM GMT+5:30	18	Female	YES	Occasiona	Chicken	8	Dominos	Quality an	Yes	Yes, aware	4	Occasion	3	Delivery	Dominos	Both
2024/03/04 4:44:31 PM GMT+5:30	19	Male	YES	Twice in a	Vegies	6	Dominos	Taste	Yes	Yes, aware	5	Appetite	4	Delivery	Dominos	Quality

Data Analysis:

Descriptive Analysis:
Summarize the data using descriptive statistics such as mean, median, mode, range, and standard deviation.

Hypothesis testing:
Formulate hypotheses based on the relationships observed in the EDA.
Use statistical tests such as chi-square test, ANOVA and MANOVA to test the hypotheses and determine if the relationship are statistically significant.

Regression Analysis:
Perform regression analysis to understand the impact of independent variables on dependent variable.
Use multiple linear regression for analyzing relationships involving multiple independent variable
Use linear regression for analyzing relationships involving one independent variable.

Graphical Representation:
Create visualization (e.g, charts, graphs, etc.) to present key findings and insight based on the survey data.

Visualization:

Bar charts:
Use bar charts to visualize the distribution of categorical variables such as menu list and order method

Pie Charts:
Pie charts can be used to show the relative proportions of different categories within a single variable.

Histogram:
Histograms are useful for visualizing the distribution of numerical variables such as Menu list and Restaurants.

Box plot:
Box plot can be used to visualize the distribution of numerical variables, such as Menu list and Restaurants.

Result and Analysis:

Objectives:

1) Identify a Mean, Median, Mode, Variance, Standard Deviation Ratings of different variable i.e is Quality and Size of Pizza

Code:

```
mean(df$Quality_Rating)
mean(df$Size_rating)
median(df$Quality_Rating)
median(df$Size_rating)
mode<-function(v){
  uniqv<-unique(v)
  uniqv[which.max(tabulate(match(v,uniqv)))]
}
mode(df$Quality_Rating)
mode(df$Size_rating)
sd(df$Quality_Rating)
sd(df$Size_rating)
var(df$Size_rating)
var(df$Quality_Rating)
```

Output:

```
> mean(df$Quality_Rating)#On Scale of  1-5 rate Quality
[1] 3.866667
> mean(df$Size_rating) #On Scale of  1-5 rate Size
[1] 3.758333
> median(df$Quality_Rating)#On Scale of  1-5 rate Quality
[1] 4
> median(df$Size_rating)#On Scale of  1-5 rate Size
[1] 4
> #Mode
> mode<-function(v){
+   uniqv<-unique(v)
+   uniqv[which.max(tabulate(match(v,uniqv)))]
+ }
> mode(df$Quality_Rating)
[1] 4
> mode(df$Size_rating)
[1] 4
> #Standard Deviation
> sd(df$Quality_Rating)
[1] 1.020243
> sd(df$Size_rating)
[1] 1.076844
```

```
> #Variance
> var(df$Size_rating)
[1] 1.159594
> var(df$Quality_Rating)
[1] 1.040896
```

Conclusion:

- Here, the mean Rating given to Quality and Size as per the survey is 3.867 for Quality and 3.758 for Size.
- Here, the median Rating given to Quality and Size as per the survey is 4 for both.
- Here, the mode Rating given to Quality and Size as per the survey is 4 for both

2) Identify Kurtosis and Skewness Ratings of different variable i.e is Quality and Size of Pizza

Code:

```
library(moments)
kurtosis(df$Quality_Rating)
kurtosis(df$Size_rating)
skewness(df$Quality_Rating)
skewness(df$Size_rating)
```

Output:

```
> #Kurtosis
> library(moments)
> kurtosis(df$Quality_Rating)
[1] 3.819809
> kurtosis(df$Size_rating)
[1] 3.436768
> #skewness
> skewness(df$Quality_Rating) #Negatively skewness
[1] -1.018454
> skewness(df$Size_rating)
[1] -0.887057
```

Conclusion:

- Here, the Kurtosis for Quality rating and Size rating as per the survey is greater than 3 so, it is leptokurtic.
- Here, the Skewness for Quality rating and Size rating as per the survey is in negative so, it is Negatively Skewed.

3) Can we predict the probability of Like it based on Menu list and pizza size (Logistic Regression)

   Code:

```
df$Like_it <- ifelse(df$Like_it == "YES", 1, 0)
logistic=glm(Like_it~Menu_list+Pizza_size,data=df,family=binomial)
logistic
prediction=predict(logistic,type="response")
head(prediction)
table(df$Like_it,prediction>0.5)
accuracy=mean((prediction>0.5)==df$Like_it)
accuracy
```

   Output:

```
> #Logistic Regression
> df$Like_it <- ifelse(df$Like_it == "YES", 1, 0)
> logistic=glm(Like_it~Menu_list+Pizza_size,data=df,family=binomial)
> logistic

Call:  glm(formula = Like_it ~ Menu_list + Pizza_size, family = binomial,
    data = df)

Coefficients:
       (Intercept)  Menu_listMargherita    Menu_listMushroom       Menu_listOther   Menu_listPepperoni
            2.5300              -1.2015              -2.5221              -1.8063              13.7540
    Menu_listVegies           Pizza_size
           -1.1108               0.1763

Degrees of Freedom: 119 Total (i.e. Null);  113 Residual
Null Deviance:      53.37
Residual Deviance: 47.41        AIC: 61.41
```

```
> prediction=predict(logistic,type="response")
> head(prediction)
        1         2         3         4         5         6
0.9809245 0.9157552 0.9442358 0.9809245 0.9392632 0.9904832
```

```
> table(df$Like_it,prediction>0.5)

    TRUE
  0    7
  1  113
> accuracy=mean((prediction>0.5)==df$Like_it)
> accuracy
[1] 0.9416667
```

Conclusion:

- As per the logistic regression used on above variables it has predicted the accuracy score with probability 0.941667 which means the model is a good fit

4) How do Size Rating, Quality rating and Pizza Size impact Restaurants (MLR)
Code:

```
df$Restaurants=factor(df$Restaurants,levels = c("Dominos","Pizza express","Pizza
hut","Pappa johns","Other"),
                labels = c("1","2","3","4","5"))
df$Restaurants=as.numeric(df$Restaurants)
set.seed(123)
model=lm(formula = Restaurants~Size_rating+Quality_Rating+Pizza_size,data=df)
summary(model)
set.seed(123)
model2=lm(formula = Restaurants~Quality_Rating+Pizza_size,data=df)
summary(model2)
model3=lm(formula = Restaurants~Pizza_size,data=df)
summary(model3)
```

Output:

```
> #MLR multiple linear Regression:
> #Back Ward Selection
> df$Restaurants=factor(df$Restaurants,levels = c("Dominos","Pizza express","Pizza hut","Pappa johns","Other"),
+                       labels = c("1","2","3","4","5"))
> df$Restaurants=as.numeric(df$Restaurants)
> set.seed(123)
> model=lm(formula = Restaurants~Size_rating+Quality_Rating+Pizza_size,data=df)
> summary(model)

Call:
lm(formula = Restaurants ~ Size_rating + Quality_Rating + Pizza_size,
    data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.4433 -1.0913 -0.8946  0.9861  3.2782 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)  
(Intercept)     1.61419    0.69095   2.336   0.0212 *
Size_rating     0.07741    0.13758   0.563   0.5748  
Quality_Rating -0.09543    0.14617  -0.653   0.5151  
Pizza_size      0.06865    0.07317   0.938   0.3501  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.419 on 116 degrees of freedom
Multiple R-squared:  0.01208,	Adjusted R-squared:  -0.01347 
F-statistic: 0.4728 on 3 and 116 DF,  p-value: 0.7018
```

```
> model2=lm(formula = Restaurants~Quality_Rating+Pizza_size,data=df)
> summary(model2)

Call:
lm(formula = Restaurants ~ Quality_Rating + Pizza_size, data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.4385 -1.0781 -0.9284  0.9829  3.0716 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)  
(Intercept)     1.72348    0.66114   2.607   0.0103 *
Quality_Rating -0.06103    0.13238  -0.461   0.6456  
Pizza_size      0.07484    0.07212   1.038   0.3016  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.415 on 117 degrees of freedom
Multiple R-squared:  0.009385,	Adjusted R-squared:  -0.007549 
F-statistic: 0.5542 on 2 and 117 DF,  p-value: 0.576


> model3=lm(formula = Restaurants~Pizza_size,data=df)
> summary(model3)

Call:
lm(formula = Restaurants ~ Pizza_size, data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.3470 -1.0848 -0.9537  0.9152  3.0463 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept)  1.56051    0.55684   2.802  0.00593 **
Pizza_size   0.06554    0.06901   0.950  0.34422   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.41 on 118 degrees of freedom
Multiple R-squared:  0.007585,	Adjusted R-squared:  -0.0008253 
F-statistic: 0.9019 on 1 and 118 DF,  p-value: 0.3442
```

Conclusion:

- As for Multiple linear regression the variable required to be in numeric type so the above code is to convert categorical value into numeric.
- As per the summary provided above in Multiple Linear regression, we remove the variable whose probability value is higher, as per the above we will remove Size_rating because it has 0.5748 p value which is maximum.
- As per the above we will remove Quality_Rating because it has 0.6456 p value which is maximum
- So, as per the above output Pizza_size is more suitable with the Restaurants this is called Backward Selection Method.

5) How closely does the linear model fit the relationship between Gender and Menu list or Restaurants (Least square method)

Code:

```
df$Gender.=factor(df$Gender.,levels = c("Male","Female"),labels=c("1","2"))
df$Gender.=as.numeric(df$Gender.)
str(df$Menu_list)
df$Menu_list=factor(df$Menu_list,levels=c("Margherita","Mushroom ",
                   "Other","Pepperoni","Vegies","Chicken"),
              labels = c("1","2","3","4","5","6"))
df$Menu_list=as.numeric(df$Menu_list)
result2=lm(Gender.~Restaurants,data=df)
summary(result2)
res1=resid(result2)
plot(density(res1))
```

Output:

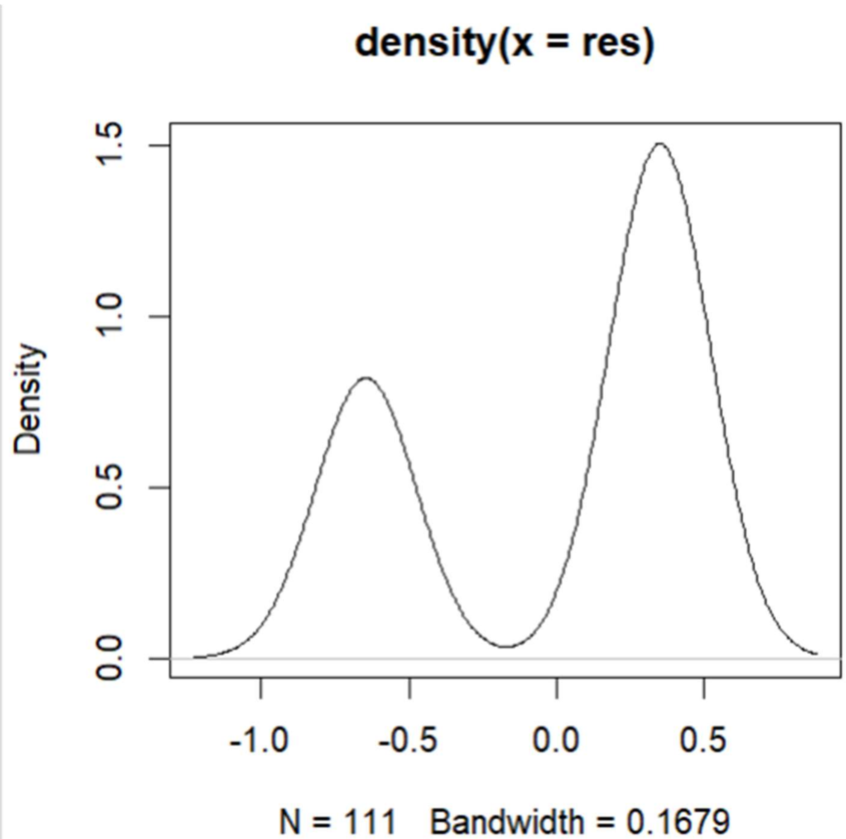
```
> #Using Regression Method Least square method
> df$Gender.=factor(df$Gender.,levels = c("Male","Female"),labels=c("1","2"))
> df$Gender.=as.numeric(df$Gender.)
> df$Menu_list=factor(df$Menu_list,levels=c("Margherita","Mushroom ",
+                             "Other","Pepperoni","Vegies","Chicken"),
+                     labels = c("1","2","3","4","5","6"))
> df$Menu_list=as.numeric(df$Menu_list)
> result=lm(Gender.~Menu_list,data=df)
> result

Call:
lm(formula = Gender. ~ Menu_list, data = df)

Coefficients:
(Intercept)    Menu_list
    1.74075     -0.01947
```

```
> summary(result)

Call:
lm(formula = Gender. ~ Menu_list, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-0.7213 -0.6239  0.3177  0.3761  0.3761

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.74075    0.13213  13.174   <2e-16 ***
Menu_list   -0.01947    0.02622  -0.743    0.459
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4805 on 109 degrees of freedom
  (9 observations deleted due to missingness)
Multiple R-squared:  0.005034,	Adjusted R-squared:  -0.004094
F-statistic: 0.5515 on 1 and 109 DF,  p-value: 0.4593
```

```
> res=resid(result)
> head(res)
         1          2          3          4          5          6 
-0.6239137  0.2787255  0.3566141  0.3760863 -0.7212745  0.3760863 
> plot(density(res))
```

Conclusion:

- As for the least squared method variable required to be in numeric type so the above code is to convert categorical value into numeric.
- Here, we find the linear regression between Gender and Menu list on line y=b+ax Equation will be: Y=1.74075-0.01947x
- As per the density plot show the residuals present into data with Number of elements with errors and their Bandwidth.
- These are the residuals which are in the data as per least squared method.



6) Is there a significant association between Brands and Factor Influencing and test the hypothesis (Chi-square)

Code:

```
H0='There is an association between Brand and Factor_Influencing'
H1='There is not association between Brand and Factor_Influencing'
res1=chisq.test(table(df$Brand, df$Factors_Influencing))
pvalue=res1$p.value
los=0.05
if(pvalue>los)
{
 print("Accept H0")
 print(H0)
}else{
 print("Cannot Accept H0")
 print(H1)
}
```

Output:

```
> # Chi-square test for independence between Brand and Factor_Influence?
> H0='There is an association between Brand and Factor_Influencing'
> H1='There is not association between Brand and Factor_Influencing'
> res1=chisq.test(table(df$Brand, df$Factors_Influencing))
Warning message:
In chisq.test(table(df$Brand, df$Factors_Influencing)) :
  Chi-squared approximation may be incorrect
> res1

	Pearson's Chi-squared test

data:  table(df$Brand, df$Factors_Influencing)
X-squared = 1.0379, df = 2, p-value = 0.5951

> pvalue=res1$p.value
> pvalue
[1] 0.5951408
> los=0.05
> if(pvalue>los)
+ {
+   print("Accept H0")
+   print(H0)
+ }else{
+   print("Cannot Accept H0")
+   print(H1)
+ }
[1] "Accept H0"
[1] "There is an association between Brand and Factor_Influencing"
```

Conclusion:

- Chi-Square method check the independence between Brand and factor Influencing to customer for buying the Pizza .

- As per to perform test we denoted Ho i.e Null hypothesis and H1 i.e Alternative hypothesis.
- After that we will compare the pvalue with level of significance which is by default 0.05 i.e 5%
- If the pvalue is greater than level of significance than, we accept H0 otherwise reject the H0 and accept H1.

7) Is there a significant difference in the mean frequency of Quality rating and restaurants test the hypothesis (ANOVA)

Code:

```
H0="Mean Quality ratingis Equal to all the Restaurants"
H1="Mean Quality rating is not Equal to all the Restaurants"
anova_result <- aov(Quality_Rating ~ Restaurants, data =df)
res=summary(anova_result)
p_value=res[[1]][1,5]
los=0.05
if(p_value>los)
{
 print("Accept HO")
 print(H0)
}else{
 print("Do not Accept H0")
 print(H1)
}
```

Output:

```
> #ANNOVA using Quality_rating and Restaurants
> H0="Mean Quality ratingis Equal to all the Restaurants"
> H1="Mean Quality rating is not Equal to all the Restaurants"
> anova_result <- aov(Quality_Rating ~ Restaurants, data =df)
> res=summary(anova_result)
> res
             Df Sum Sq Mean Sq F value Pr(>F)
Restaurants   1   0.03  0.0332   0.032  0.859
Residuals   118 123.83  1.0494
> p_value=res[[1]][1,5]
> p_value
[1] 0.859187
> los=0.05
> if(p_value>los)
+ {
+   print("Accept HO")
+   print(H0)
+ }else{
+   print("Do not Accept H0")
+   print(H1)
+ }
[1] "Accept HO"
[1] "Mean Quality ratingis Equal to all the Restaurants"
```

Conclusion:

- ANNOVA testing is used to test the impact of Quality_Rating on Restaurants. In this there is only one dependent and one independent variable.
- As per to perform test we denoted Ho i.e Null hypothesis and H1 i.e Alternative hypothesis.
- After that we will compare the pvalue with level of significance which is by default 0.05 i.e 5%
- If the pvalue is greater than level of significance than, we accept H0 otherwise reject the H0 and accept H1.

8) Are there Mean differences in the frequency of Restaurants based on Menu list and Quality rating test with different types of tests of hypothesis (MANOVA)

Code:

```
H0='Mean of Quality_Rating, Size_Rating and Restaurants are same'
H1='Mean of Quality_Rating, Size_Rating and Restaurants are different'
df$Menu_list=factor(df$Menu_list,levels=c("Margherita","Mushroom
","Other","Pepperoni","Vegies","Chicken"),
          labels = c("1","2","3","4","5","6"))
df$Menu_list=as.numeric(df$Menu_list)
manovatest=manova(cbind(df$Quality_Rating,df$Menu_list)~df$Restaurants)
manovatest
#Using Wilks Test Method
wilks=summary(manovatest,test="Wilks")
wilks
los=0.05
pvalue1=wilks$stats["df$Restaurants","Pr(>F)"]
if(pvalue1>los)
{
  print("Accept H0")
  print(H0)
}else{
  print("Cannot Accept H0")
  print(H1)
}
#Using Roy Test Method
roy=summary(manovatest,test="Roy")
roy
los=0.05
pvalue2=roy$stats["df$Restaurants","Pr(>F)"]
if(pvalue2>los)
{
  print("Accept H0")
  print(H0)
```

```
}else{
  print("Cannot Accept H0")
  print(H1)
}
```

Output:

```
> H0='Mean of Quality_Rating, Size_Rating and Restaurants are same'
> H1='Mean of Quality_Rating, Size_Rating and Restaurants are different'
> df$Menu_list=factor(df$Menu_list,levels=c("Margherita","Mushroom ","Other","Pepperoni","Vegies","Chicken"),
+                     labels = c("1","2","3","4","5","6"))
> df$Menu_list=as.numeric(df$Menu_list)
> manovatest=manova(cbind(df$Quality_Rating,df$Menu_list)~df$Restaurants)
> manovatest
Call:
   manova(cbind(df$Quality_Rating, df$Menu_list) ~ df$Restaurants)

Terms:
                df$Restaurants Residuals
resp 1                  4.7040  103.9987
resp 2                  6.2361  329.6558
Deg. of Freedom              4       106

Residual standard errors: 0.9905148 1.763508
Estimated effects may be unbalanced
9 observations deleted due to missingness
```

```
> #Using Wilks Test Method
> wilks=summary(manovatest,test="Wilks")
> wilks
                Df   Wilks approx F num Df den Df Pr(>F)
df$Restaurants   4 0.94108  0.80928      8    210 0.5951
Residuals      106
> los=0.05
> #For Pvalue
> pvalue1=wilks$stats["df$Restaurants","Pr(>F)"]
> pvalue1
[1] 0.5950757
> if(pvalue1>los)
+ {
+   print("Accept H0")
+   print(H0)
+ }else{
+   print("Cannot Accept H0")
+   print(H1)
+ }
[1] "Accept H0"
[1] "Mean of Quality_Rating, Size_Rating and Restaurants are same"
```

```
> #Using Roy Test Method
> roy=summary(manovatest,test="Roy")
> roy
               Df      Roy approx F num Df den Df Pr(>F)
df$Restaurants   4 0.057688   1.5287      4    106 0.1991
Residuals      106
> los=0.05
> #For Pvalue
> pvalue2=roy$stats["df$Restaurants","Pr(>F)"]
> pvalue2
[1] 0.1991059
> if(pvalue2>los)
+ {
+   print("Accept H0")
+   print(H0)
+ }else{
+   print("Cannot Accept H0")
+   print(H1)
+ }
[1] "Accept H0"
[1] "Mean of Quality_Rating, Size_Rating and Restaurants are same"
```

Conclusion:

- MANOVA testing is used test the impact of Two independent variable on one depending variable.
- This can be done by using various testing type but here I have used only two types of testing that is Wilks and Roy and compare their probability value with level of significance.
- As per to perform test we denoted H0 i.e Null hypothesis and H1 i.e Alternative hypothesis.
- If the pvalue is greater than level of significance than, we accept H0 otherwise reject the H0 and accept H1.

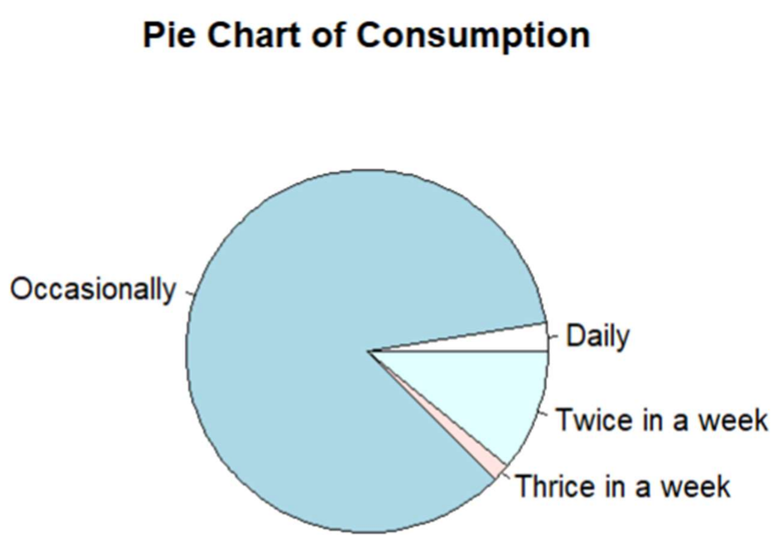
9) How does a pie chart represent the distribution of pizza consumption among survey

Code:

```
table(df$Frequency)
pie(table(df$Frequency),labels = c('Daily','Occasionally','Thrice in a week','Twice in a week'),
   main = "Pie Chart of Consumption")
```

Output:

```
> table(df$Frequency)

           Daily     Occasionally Thrice in a week  Twice in a week
               3              102                2               13
> pie(table(df$Frequency),labels = c('Daily','Occasionally','Thrice in a week','Twice in a week'),
+     main = "Pie Chart of Consumption")
```



Conclusion:

As per the above visualization more number of customer preferred to consume pizza Occasionally.

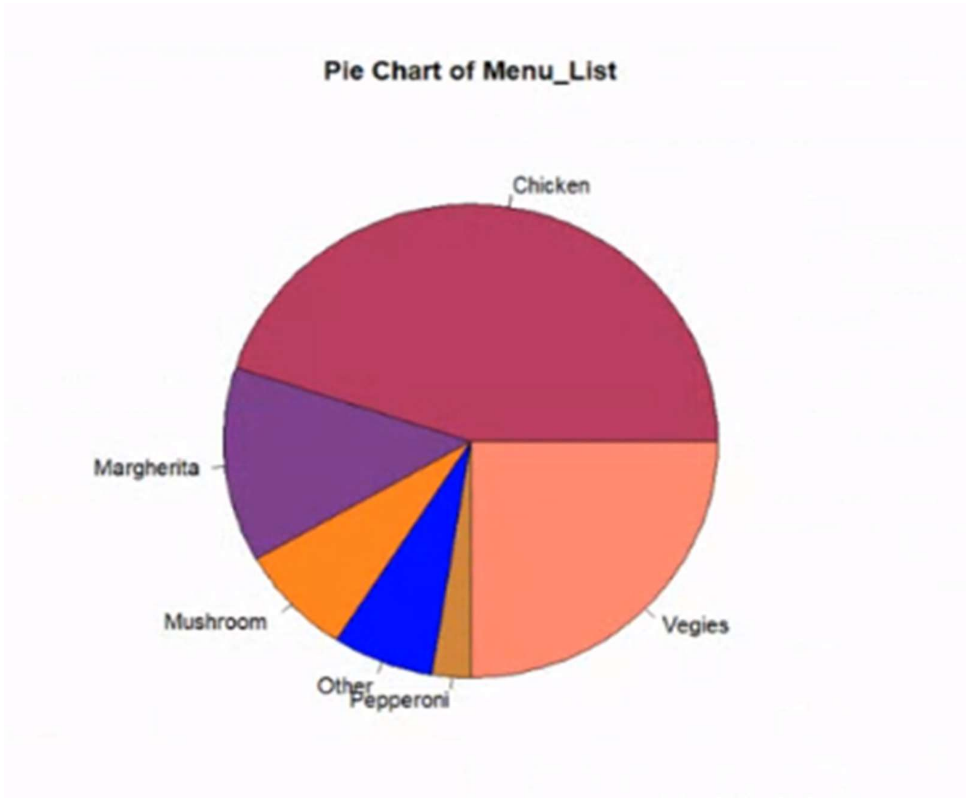
10) How does a pie chart represent the distribution of Restaurants among survey

Code:

```
table(df$Restaurants)
pie(table(df$Restaurants),col =
c('maroon','mediumorchid4','chocolate1','peru','salmon'),
   labels = c('Dominos','Others','Pappa Johns','Pizza express','Pizza Hut'),
   main = "Pie Chart of Restaurants")
```

Output:

```
> table(df$Restaurants)

      Dominos         Other    Pappa johns Pizza express     Pizza hut
           70            14              2             1            33
> pie(table(df$Restaurants),col = c('maroon','mediumorchid4','chocolate1','peru','salmon'),
+     labels = c('Dominos','Others','Pappa Johns','Pizza express','Pizza Hut'),
+     main = "Pie Chart of Restaurants")
```



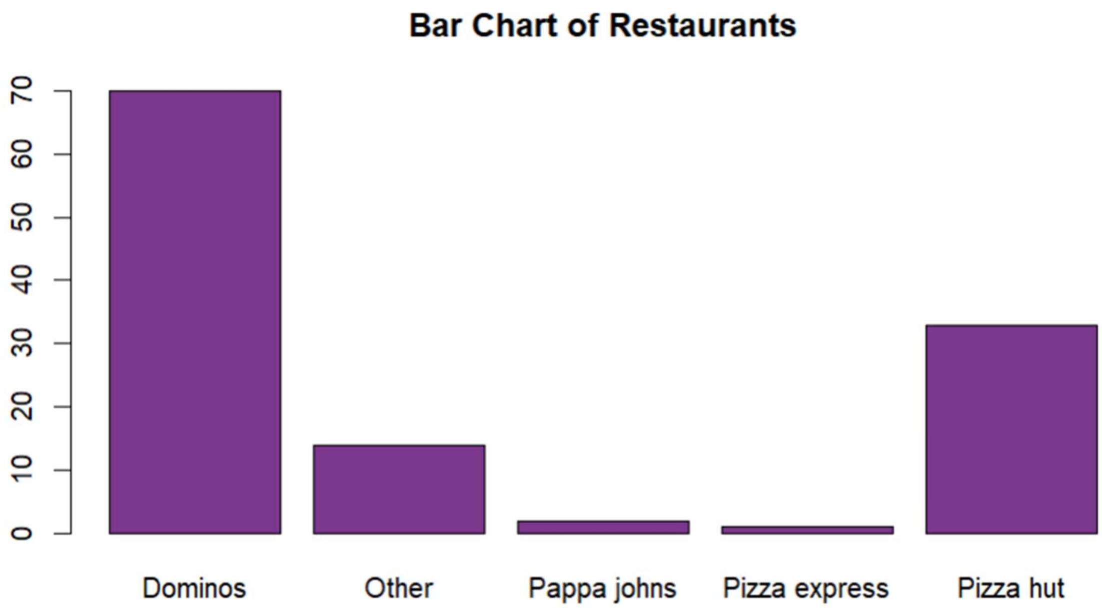
Conclusion:
As the above figure Customers likes Chicken pizza more whereas, least number of Customer preferred Pepperoni

11) How does a bar plot illustrate the frequency of Restaurants
Code:

```
barplot(table(df$Restaurants), col="mediumorchid4", main = "Bar Chart of Restaurants")
```

Output:

```
> barplot(table(df$Restaurants), col="mediumorchid4", main = "Bar Chart of Restaurants")
```

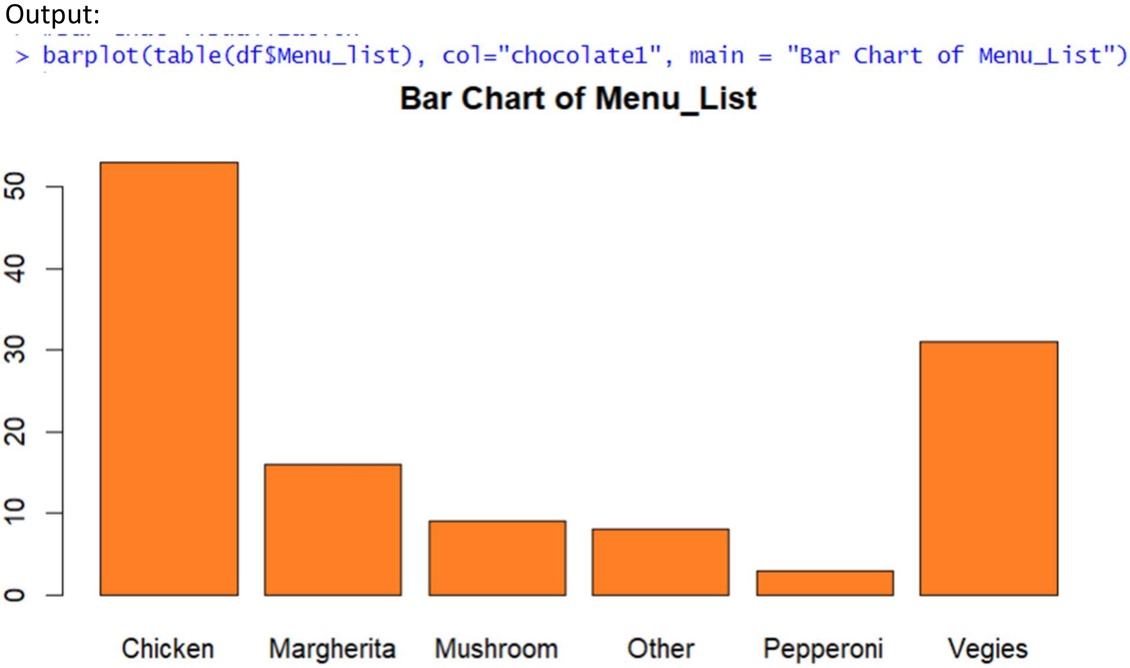


Conclusion:
As the bar graph shows the More customers like to visit Dominos to eat pizza whereas, Pappa Johns and Pizza express as the less frequency

12) How does a bar plot illustrate the frequency of menu list
Code:

```
barplot(table(df$Menu_list), col="chocolate1", main = "Bar Chart of Menu_List")
```

Output:



Conclusion:
As the bar graph shows the More customers like to eat chicken pizza whereas, Pepperonias the less frequency

13) How does a histogram display the frequency of restaurants among survey

Code:

```
hist(table(df$Restaurants),col='chocolate1')
```

Output:

```
> hist(table(df$Restaurants),col='chocolate1')
```

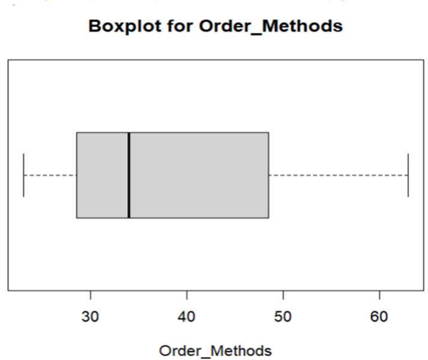
14) How does a boxplot show the frequency of Order method among survey

Code:

```
boxplot(table(df$Order_Methods),horizontal=TRUE, xlab =
"Order_Methods",main="Boxplot for Order_Methods")
```

Output:

```
> boxplot(table(df$Order_Methods),horizontal=TRUE, xlab = "Order_Methods",main="Boxplot for Order_Methods")
```



Conclusion:
A per the figure shown above the line lie between 30-35 which shows the average value for the skewness the curve can be form is Positively skewed

15) Contingency table with respect to Brand and Quality Rating
Code:
contingency_table=table(df$Brand,df$Quality_Rating)
contingency_table
Output:

```
> contingency_table=table(df$Brand,df$Quality_Rating)
> contingency_table

             1  2  3  4  5
  Dominos    4  6 14 37 21
  Pizza hut  1  1  6 18 12
```

Conclusion:
Here the table shows the relationship between Brand and their Rating which tells that Dominos has highly rated as compared to Pizza Hut

16) Contingency table with respect to Gender and Like it
Code:
contingency_table=table(df$Gender.,df$Like_it)
contingency_table
Output:

```
> contingency_table=table(df$Gender.,df$Like_it)
> contingency_table

          NO YES
  Female   4  74
  Male     3  39
```

Conclusion:
Here the table shows the relationship between Gender and their Liking which tells that Female has more number of liking as compared to Male.

- Summarizing the analysis:
  In Conclusion I would like to say that data on pizza survey which I collected through google form gives me an opportunity to perform Various analysis on Data. From Data Collecting to Data Visualization.

  In, this entire process I have used various Statistical method like descriptive analysis, Hypotheses testing, Predictive Analysis which helps me to take various decision to the dataset

  The analysis of the pizza survey helped me to understand people's preference and habits related to pizza like which brand they preferred and what menu most people like to eat, What factor influence the customers to buys the pizza and So on.

  Overall, the survey findings provide valuable insights for businesses in the food industry, marketers, and pizza enthusiasts, helping them understand the preferences and behaviors of consumers in the pizza market.