Dummy data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Names of the restaurants** | **Sales per day** | **Area in sq meters** | **Population in hundreds** | **Distance from Station in km** |
| A | 40,000 | 1,000.00 | 100.00 | 1 to 1.5 km |
| B | 35,000 | 700.00 | 200.00 | 1.6 to 3 km |
| C | 37,000 | 800.00 | 150.00 | 1 to 1.5 km |
| D | 40,000 | 1,050.00 | 100.00 | 1 to 1.5 km |
| E | 32,000 | 550.00 | 180.00 | 1.6 to 3 km |
| F | 40,000 | 600.00 | 300.00 | 0.5 to 1 km |
| G | 30,000 | 500.00 | 200.00 | 1.6 to 3 km |
| H | 45,000 | 1,200.00 | 90.00 | 0.5 to 1 km |
| I | 37,000 | 850.00 | 130.00 | 1 to 1.5 km |
| J | 41,000 | 1,000.00 | 100.00 | 0.5 to 1 km |
| L | 35,000 | 700.00 | 200.00 | 0.5 to 1 km |
| M | 37,000 | 800.00 | 150.00 | 0.5 to 1 km |
| N | 40,000 | 1,050.00 | 100.00 | 1.6 to 3 km |
| O | 32,500 | 550.00 | 180.00 | 1.6 to 3 km |
| P | 35,000 | 600.00 | 300.00 | 1 to 1.5 km |
| Q | 31,000 | 500.00 | 200.00 | 1.6 to 3 km |
| R | 45,000 | 1,200.00 | 90.00 | 0.5 to 1 km |
| S | 37,000 | 850.00 | 130.00 | 1 to 1.5 km |
| T | 40,000 | 1,000.00 | 100.00 | 0.5 to 1 km |
| U | 35,000 | 700.00 | 200.00 | 1 to 1.5 km |
| V | 36,000 | 800.00 | 150.00 | 1.6 to 3 km |
| W | 44,000 | 1,050.00 | 100.00 | 0.5 to 1 km |
| X | 32,500 | 550.00 | 180.00 | 1 to 1.5 km |
| Y | 35,500 | 600.00 | 300.00 | 1.6 to 3 km |
| Z | 30,000 | 500.00 | 200.00 | 1.6 to 3 km |
| A1 | 45,000 | 1,200.00 | 90.00 | 0.5 to 1 km |
| A2 | 37,000 | 850.00 | 130.00 | 1 to 1.5 km |

Output of Regression Analysis used for site selection

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model Summary** | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | .977a | .954 | .948 | 1019.687 |
| a. Predictors: (Constant), distance\_km, population\_hundreds, area\_sq\_feet | | | | |

* Value of R square is 0.954 which signifies that regression model explains 95.4 % of variance.
* In simple words independent variables explains 95.4 % variance or 95.4% percentage relationship explained by model between independent and dependent variable
* Value of R square should always be higher
* **Anova** explains whether regression model is significant. In other word if there is significant relationship between independent and dependent variables
* If p value is less than 0.05, it means it is significant

As p value is 0.000 which is less than 0.05 , it means it is significant and we can conclude that there is significant relationship between independent and dependent variable.

* Data that is provided is good to fit for the regression model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 495715136.006 | 3 | 165238378.669 | 158.920 | .000b |
| Residual | 23914493.624 | 23 | 1039760.592 |  |  |
| Total | 519629629.630 | 26 |  |  |  |
| a. Dependent Variable: sale\_per\_day | | | | | | |
| b. Predictors: (Constant), distance\_km, population\_hundreds, area\_sq\_feet | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | T | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | 13151.519 | 2001.204 |  | 6.572 | .000 |
| area\_sq\_meters | 21.952 | 1.793 | 1.140 | 12.246 | .000 |
| population\_hundreds | 27.408 | 5.507 | .397 | 4.977 | .000 |
| distance\_km | 976.274 | 312.686 | .182 | 3.122 | .005 |
| a. Dependent Variable: sale\_per\_day | | | | | | |

Regression Model based on given data

Y = a + B1 X1 + B2 X2 + B3 X3

**Sales = 13151.519 + 21.952 x Area in square feet + 27.408 x Population in Hundreds + 976.274 x Distance in Kms**

**Factor analysis is data reduction technique which is used to convert many variable into few factors without losing any important data**

**KMO Test Barlletes Test**

**KMO Test Talks about sample adequacy**

**Wether sample is adequate enough to run factor analysis**

**It should be more than o.6**

|  |  |  |
| --- | --- | --- |
| **KMO and Bartlett's Test** | | |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | .618 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 194.483 |
| df | 45 |
| Sig. | .000 |
|  |  |  |

Factor extraction talks about variable variance explained by the factor

Since my mon feel safe is variable has low factor extraction value. Hence it will not be linked with any of the factors

|  |  |  |
| --- | --- | --- |
| **Communalities** | | |
|  | Initial | Extraction |
| Indian railways are always on time | 1.000 | .933 |
| seats are always comfortable | 1.000 | .934 |
| i like food offered by railways | 1.000 | .835 |
| auto promotion to higher class if seats are available | 1.000 | .943 |
| my friends and familly like travelling by trains | 1.000 | .462 |
| coaches condition is very good | 1.000 | .922 |
| i get benefitis of frequenty travelling | 1.000 | .971 |
| it suits my time schedule | 1.000 | .955 |
| my mon feel safe when i travell by train | 1.000 | .193 |
| travelling by train suits my life style | 1.000 | .923 |
| Extraction Method: Principal Component Analysis. | | |

Total Variance Explained tables

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Component | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
| Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 3.177 | 31.775 | 31.775 | 3.041 | 30.408 | 30.408 |
| 2 | 3.050 | 30.499 | 62.274 | 3.030 | 30.296 | 60.703 |
| 3 | 1.845 | 18.447 | 80.720 | 2.002 | 20.017 | 80.720 |
| 4 |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |
| 6 |  |  |  |  |  |  |
| 7 |  |  |  |  |  |  |
| 8 |  |  |  |  |  |  |
| 9 |  |  |  |  |  |  |
| 10 |  |  |  |  |  |  |

1 factor

2 factor

3 factor

|  |  |  |  |
| --- | --- | --- | --- |
| **Rotated Component Matrixa** | | | |
|  | Component | | |
| 1 | 2 | 3 |
| Indian railways are always on time | .954 | -.004 | .153 |
| seats are always comfortable | .037 | .090 | .962 |
| i like food offered by railways | .912 | .037 | -.052 |
| auto promotion to higher class if seats are available | -.062 | .965 | .096 |
| my friends and familly like travelling by trains | .578 | .149 | -.325 |
| coaches condition is very good | .959 | -.040 | .021 |
| i get benefitis of frequenty travelling | -.028 | .985 | -.005 |
| it suits my time schedule | -.077 | .175 | .958 |
| my mon feel safe when i travell by train | -.184 | -.389 | -.086 |
| travelling by train suits my life style | -.016 | .956 | .097 |
| Extraction Method: Principal Component Analysis.  Rotation Method: Varimax with Kaiser Normalization. | | | |
| a. Rotation converged in 4 iterations. | | | |

Highly loaded variables are clubbed together to form factor