NYC_Italian_restaurant

Saahil Shroff

2022-12-12

Find Cheap Restaurants

Loading the data and performing basic sanity checks.

```
##            Restaurant Food Decor Service East Cheap
## 1 Daniella Ristorante   22    18      20    0     0
## 2  Tello's Ristorante   20    19      19    0     1
## 3          Biricchino   21    13      18    0     1
## 4             Bottino   20    20      17    0     0
## 5          Da Umberto   24    19      21    0     0
## 6            Le Madri   22    22      21    0     0
```

```
## Dimensions on the loaded dataframe: 168 6
```

```
## Are there any NULLs present in the loaded data frame? FALSE
```

```
## Logistic regression is suitable for predicting if a restaurant is cheap or not. It is beca
use logistic regression is used for classification tasks since it is a simple and effective w
ay to predict a binary outcome. The model uses logistic function to model the relationship be
tween a binary dependent variable (cheap variable in this case) and one or more independent v
ariables (such as decor, service) to predict the probability that an instance belongs to a ce
rtain class. Linear regression, on the other hand, is used for predicting continuous values,
and would have been suitable for predictiong price of a meal at a restaurant.
```

Building the model - estimating the model and interpreting the results.

Factoring food,decor,and service using the following categories:

0-4: 0;

5-9: 1;

10-14: 2;

15-19: 3;

20-24: 4;

25 =<: 5.

```
##            Restaurant Food Decor Service East Cheap split_food split_decor
## 1 Daniella Ristorante   22    18      20    0     0          4           3
## 2  Tello's Ristorante   20    19      19    0     1          3           3
## 3          Biricchino   21    13      18    0     1          4           2
## 4             Bottino   20    20      17    0     0          3           3
## 5          Da Umberto   24    19      21    0     0          4           3
## 6            Le Madri   22    22      21    0     0          4           4
##   split_service
## 1             3
## 2             3
## 3             3
## 4             3
## 5             4
## 6             4
```

Creating a logistic model:

```
##
## Call:
## glm(formula = Cheap ~ factor(split_food) + factor(split_decor) +
##     factor(split_service) + factor(East), family = "binomial",
##     data = nyc_restaurants)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1241  -0.6388  -0.0002   0.5551   2.0149
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)               19.8570  3578.8109   0.006  0.99557
## factor(split_food)4       -1.5214     0.4942  -3.078  0.00208 **
## factor(split_decor)2     -17.0968  3578.8109  -0.005  0.99619
## factor(split_decor)3     -19.2063  3578.8109  -0.005  0.99572
## factor(split_decor)4     -36.7229  3801.8392  -0.010  0.99229
## factor(split_service)3    -0.2894     1.2971  -0.223  0.82342
## factor(split_service)4    -0.6929     1.4010  -0.495  0.62089
## factor(East)1             -0.3256     0.4291  -0.759  0.44801
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 225.12  on 167  degrees of freedom
## Residual deviance: 147.96  on 160  degrees of freedom
## AIC: 163.96
##
## Number of Fisher Scoring iterations: 17
```

```
## My assumption was that all the variables - Food, Decor, Service, and East will play a role to determine if a restaurant is cheap or not. In my experience, if the food, decor, service a re top-class and it is situated in a posh locality, there are high chances for the restaurant to be expensive. However, looking at the model, only top-class food (4/5) is statistically si gnificant towards prediciting if a restaurant is cheap or not, which goes against my common o bservation and experience.
```

Adding two new places with the following scores and locations to make predictions

```
##           Restaurant Food Decor Service East
## 1 Assagio Ristorante   23    17      22    0
## 2             Altura   18    15      24    1
```

Getting the dataframe ready for prediction:

```
##           Restaurant Food Decor Service East split_food split_decor
## 1 Assagio Ristorante   23    17      22    0          4           3
## 2             Altura   18    15      24    1          3           2
##   split_service
## 1             4
## 2             4
```

Prediction and Inference:

```
##           Restaurant normalized_prediction
## 1 Assagio Ristorante                     0
## 2             Altura                     1
```

```
## Thus, according to the logistic-regression model's predicition, Altura is a cheap Italian restaurant, while Assagio Ristorante is an expensive Italian restaurant.
```