



ELEMENTS THAT AFFECT TOURIST'S SATISFACTION

OBJECTIVE:

- To analyze which attractions influence tourist's satisfaction when traveling across Europe

RESEARCH QUESTIONS:

- What factors have a significant effect tourist's satisfaction?
- What attracts tourists?
- What variables can significantly affect the overall ratings of a trip across Europe?

About the Dataset

“This data set is populated by capturing user ratings from Google reviews. Reviews on attractions from 24 categories across Europe are considered. Google user rating ranges from 1 to 5 and average user rating per category is calculated..”

Dataset was reduced to 3724 responses due to missing information

Originally:
5456 rows and 26 columns

Reduced to:
3724 rows and 25 columns after
omitting NA/0 values and
unnecessary columns

This dataset is mostly
numerical

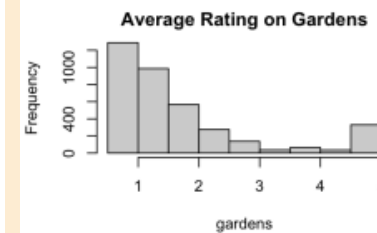
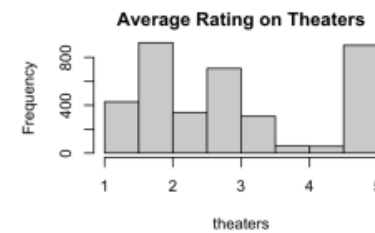
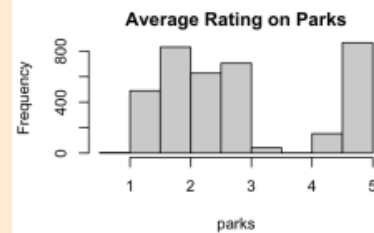
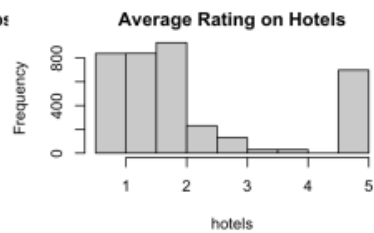
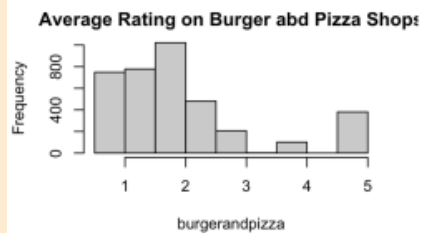
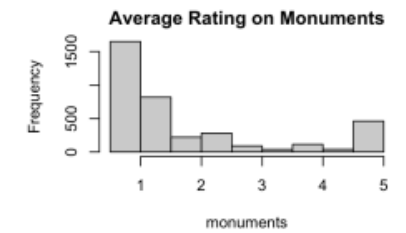
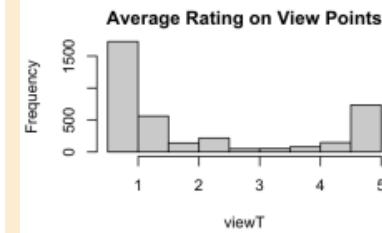
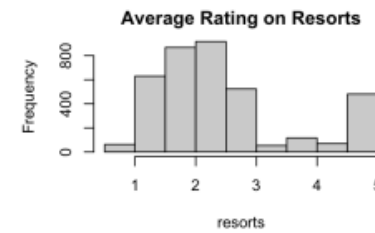
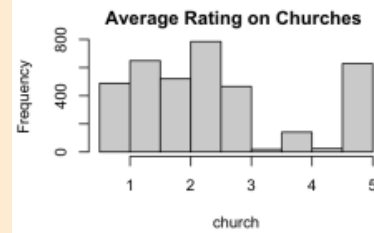
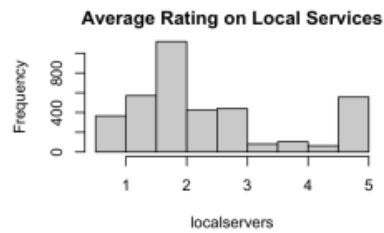
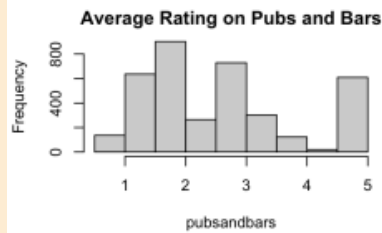
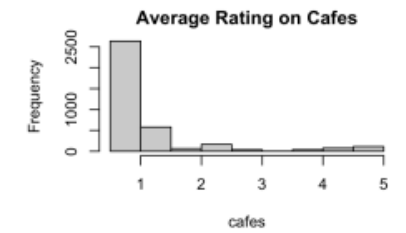
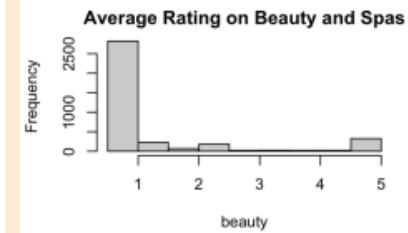
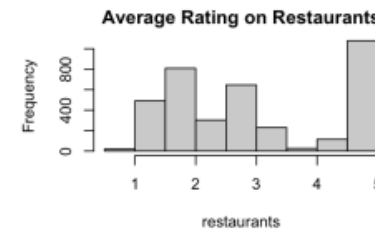
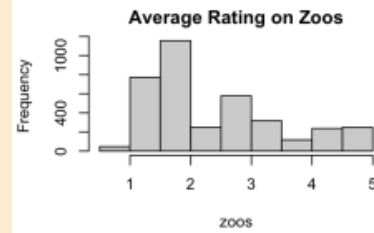
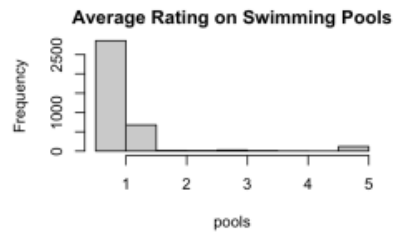
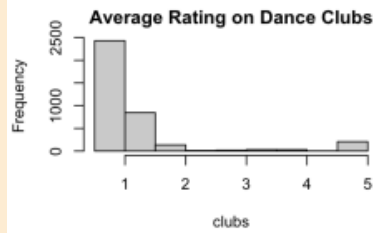
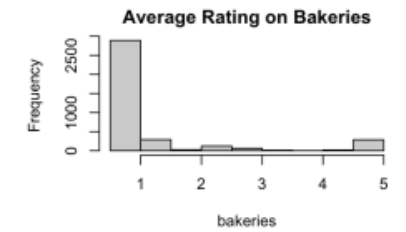
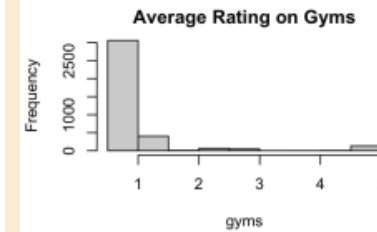
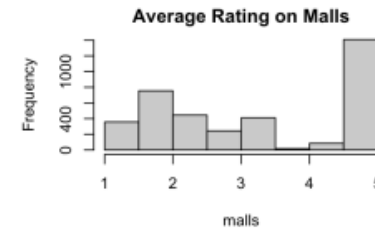
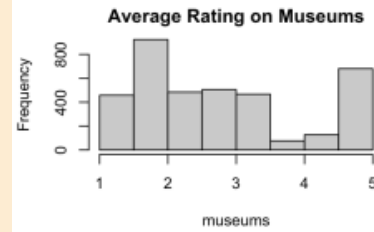
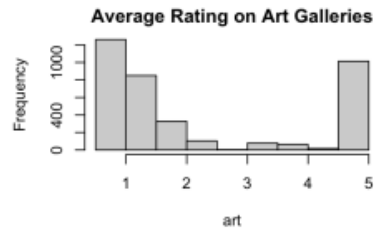
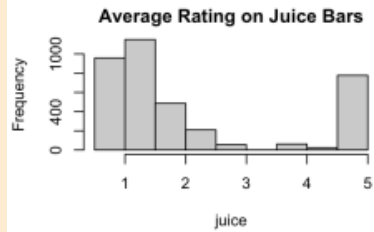
Variables

Unique user id	User #	omited
Average ratings on churches	Scale of 1 to 5	'church'
Average ratings on resorts	Scale of 1 to 5	'resorts'
Average ratings on beaches	Scale of 1 to 5	'beaches'
Average ratings on parks	Scale of 1 to 5	'parks'
Average ratings on theatres	Scale of 1 to 5	'theaters'
Average ratings on museums	Scale of 1 to 5	'museums'
Average ratings on malls	Scale of 1 to 5	'malls'
Average ratings on zoo	Scale of 1 to 5	'zoos'
Average ratings on restaurants	Scale of 1 to 5	'restaurants'
Average ratings on pubs/bars	Scale of 1 to 5	'pubsandbars'
Average ratings on local services	Scale of 1 to 5	'localservs'
Average ratings on burger/pizza shops	Scale of 1 to 5	'burgerandpizza'

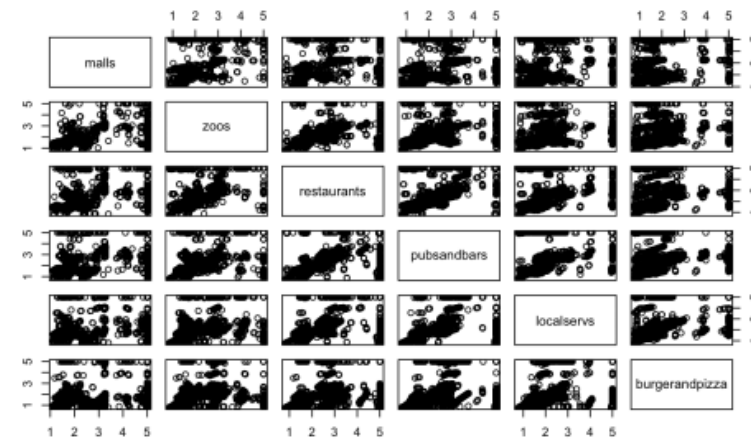
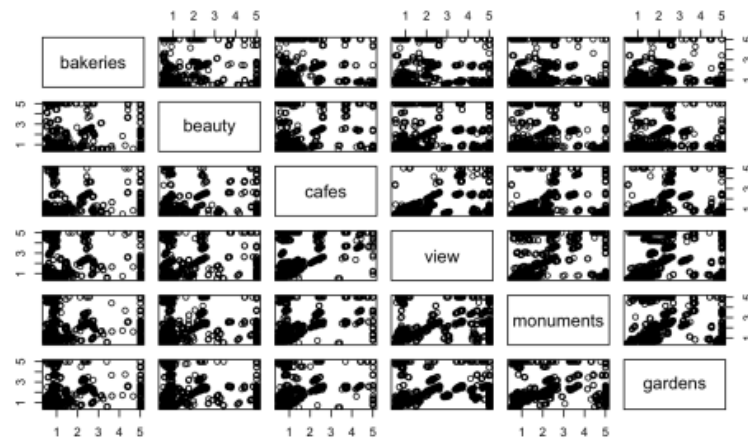
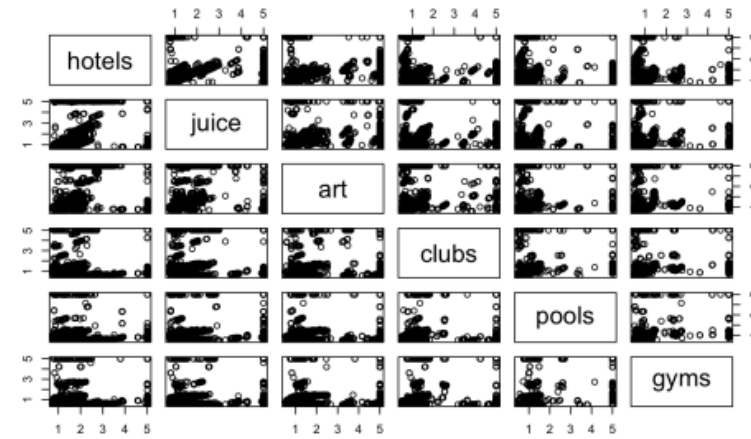
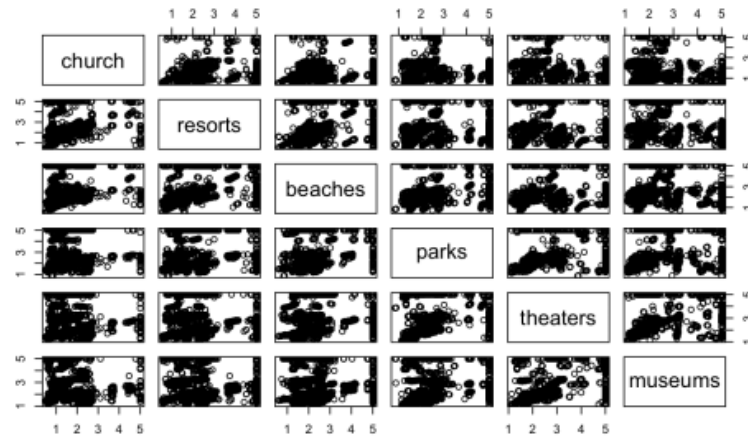
Variables

Average ratings on hotels/other lodgings	Scale of 1 to 5	'hotels'
Average ratings on juice bars	Scale of 1 to 5	'juice'
Average ratings on art galleries	Scale of 1 to 5	'art'
Average ratings on dance clubs	Scale of 1 to 5	'clubs'
Average ratings on swimming pools	Scale of 1 to 5	'pools'
Average ratings on gyms	Scale of 1 to 5	'gyms'
Average ratings on bakeries	Scale of 1 to 5	'bakeries'
Average ratings on beauty & spas	Scale of 1 to 5	'beauty'
Average ratings on cafes	Scale of 1 to 5	'cafes'
Average ratings on view points	Scale of 1 to 5	'view'
Average ratings on monuments	Scale of 1 to 5	'monuments'
Average ratings on gardens	Scale of 1 to 5	'gardens'
Overall ratings	Scale of 1 to 5	'overall'

Histograms



Pairwise Plots



Full Model

Full Model

formula =

Overall ~ church+resorts+beaches+parks+theaters+museums+
+malls+zoos+restaurants+pubsandbars+localservs+
+burgerandpizza+hotels+juice+art+clubs+pools+gyms+
+bakeries+beauty+cafes+view+monuments+gardens

Multiple R ²	1
Adjusted R ²	1

*this model includes all the variables

```

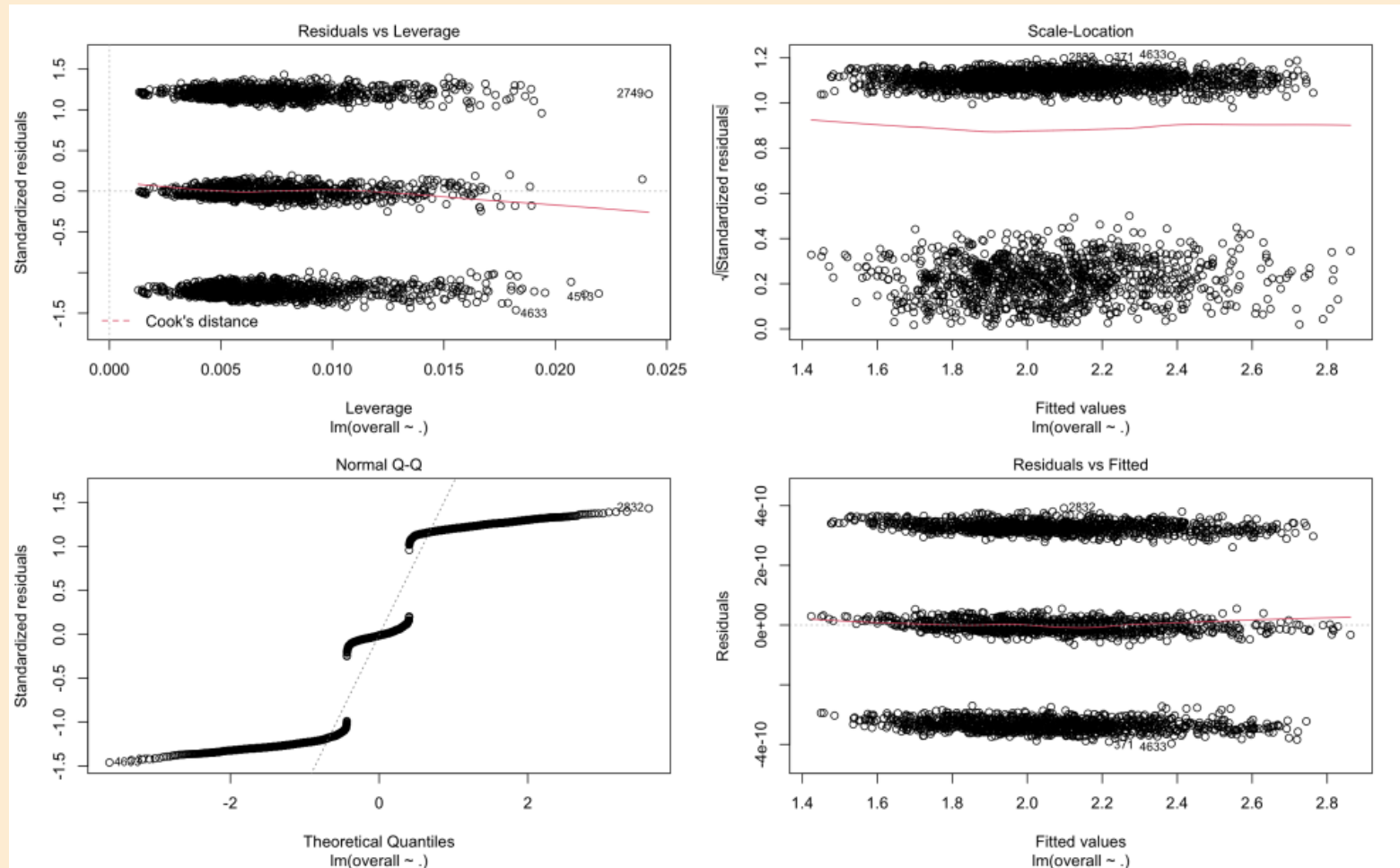
Residuals:
    Min       1Q   Median       3Q      Max
-3.972e-10 -3.220e-10 -3.730e-12  3.175e-10  3.918e-10

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.808e-11  4.385e-11  -6.400e-01   0.522
church       4.167e-02  6.765e-12  6.159e+09   <2e-16 ***
resorts      4.167e-02  3.847e-12  1.083e+10   <2e-16 ***
beaches      4.167e-02  4.729e-12  8.811e+09   <2e-16 ***
parks        4.167e-02  4.903e-12  8.499e+09   <2e-16 ***
theaters     4.167e-02  5.119e-12  8.139e+09   <2e-16 ***
museums      4.167e-02  4.982e-12  8.364e+09   <2e-16 ***
malls        4.167e-02  4.761e-12  8.752e+09   <2e-16 ***
zoos         4.167e-02  6.038e-12  6.900e+09   <2e-16 ***
restaurants  4.167e-02  4.953e-12  8.412e+09   <2e-16 ***
pubsandbars  4.167e-02  5.333e-12  7.814e+09   <2e-16 ***
localservs   4.167e-02  4.575e-12  9.107e+09   <2e-16 ***
burgerandpizza 4.167e-02  4.823e-12  8.639e+09   <2e-16 ***
hotels       4.167e-02  4.065e-12  1.025e+10   <2e-16 ***
juice        4.167e-02  3.774e-12  1.104e+10   <2e-16 ***
art          4.167e-02  3.060e-12  1.362e+10   <2e-16 ***
clubs        4.167e-02  4.763e-12  8.748e+09   <2e-16 ***
pools        4.167e-02  6.954e-12  5.992e+09   <2e-16 ***
gyms         4.167e-02  6.762e-12  6.162e+09   <2e-16 ***
bakeries     4.167e-02  4.611e-12  9.036e+09   <2e-16 ***
beauty       4.167e-02  4.114e-12  1.013e+10   <2e-16 ***
cafes        4.167e-02  6.001e-12  6.944e+09   <2e-16 ***
view         4.167e-02  3.713e-12  1.122e+10   <2e-16 ***
monuments    4.167e-02  3.966e-12  1.051e+10   <2e-16 ***
gardens      4.167e-02  4.686e-12  8.891e+09   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

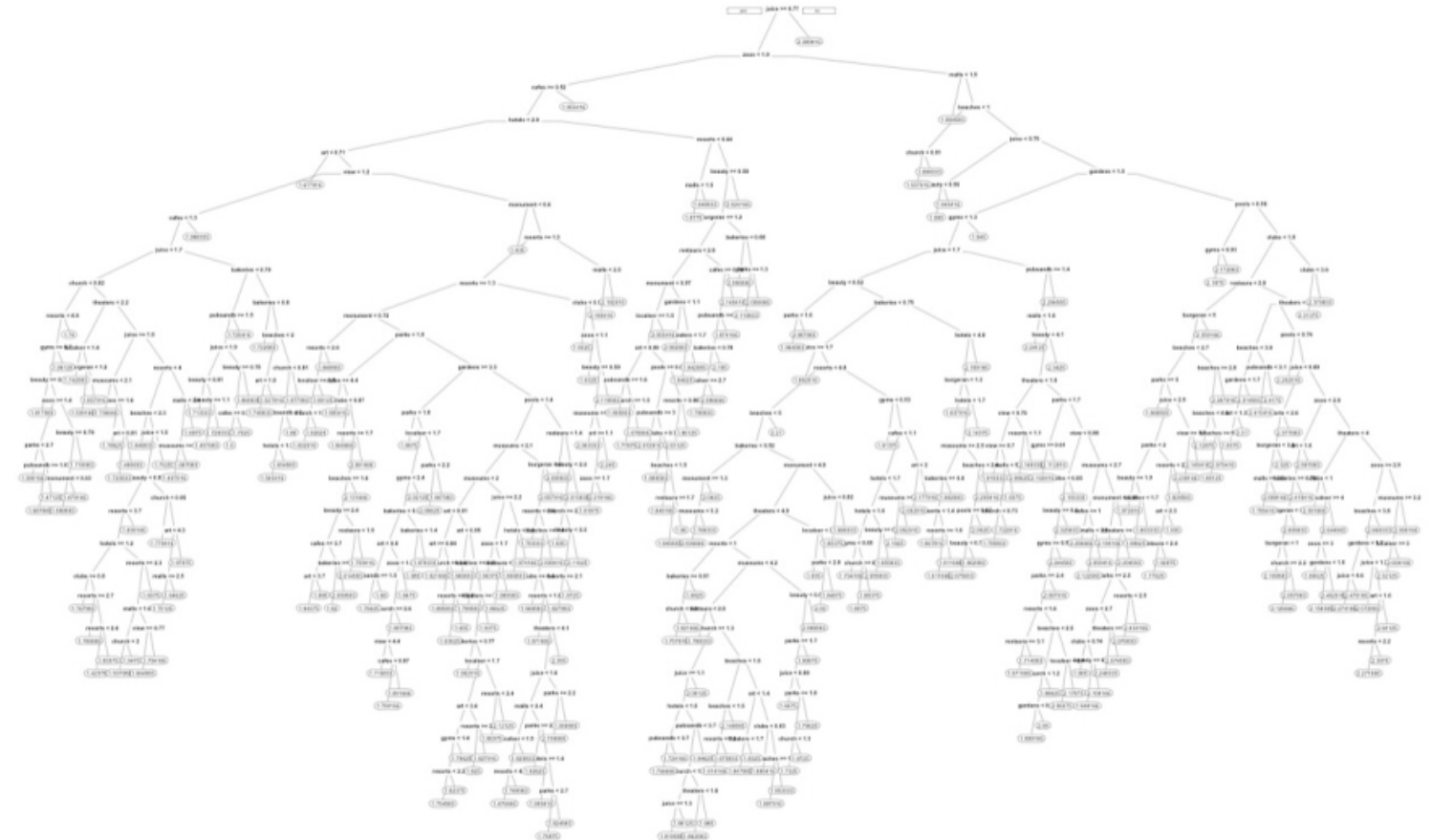
Residual standard error: 2.742e-10 on 3442 degrees of freedom
Multiple R-squared:  1,    Adjusted R-squared:  1
F-statistic: 1.135e+20 on 24 and 3442 DF, p-value: < 2.2e-16

```


Full Model Diagnostic Plots



Full Model Decision Tree



Reduced Model 1

Reduced Model 1

formula =

overall~parks+theaters+museums+malls+zoos+restaurants+

+pubsandbars+localservs+hotels+juice

Multiple R ²	0.3524
Adjusted R ²	0.3506

```
Call:
lm(formula = overall ~ parks + theaters + museums + malls + zoos +
    restaurants + pubsandbars + localservs + hotels + juice,
    data = TrainDF)

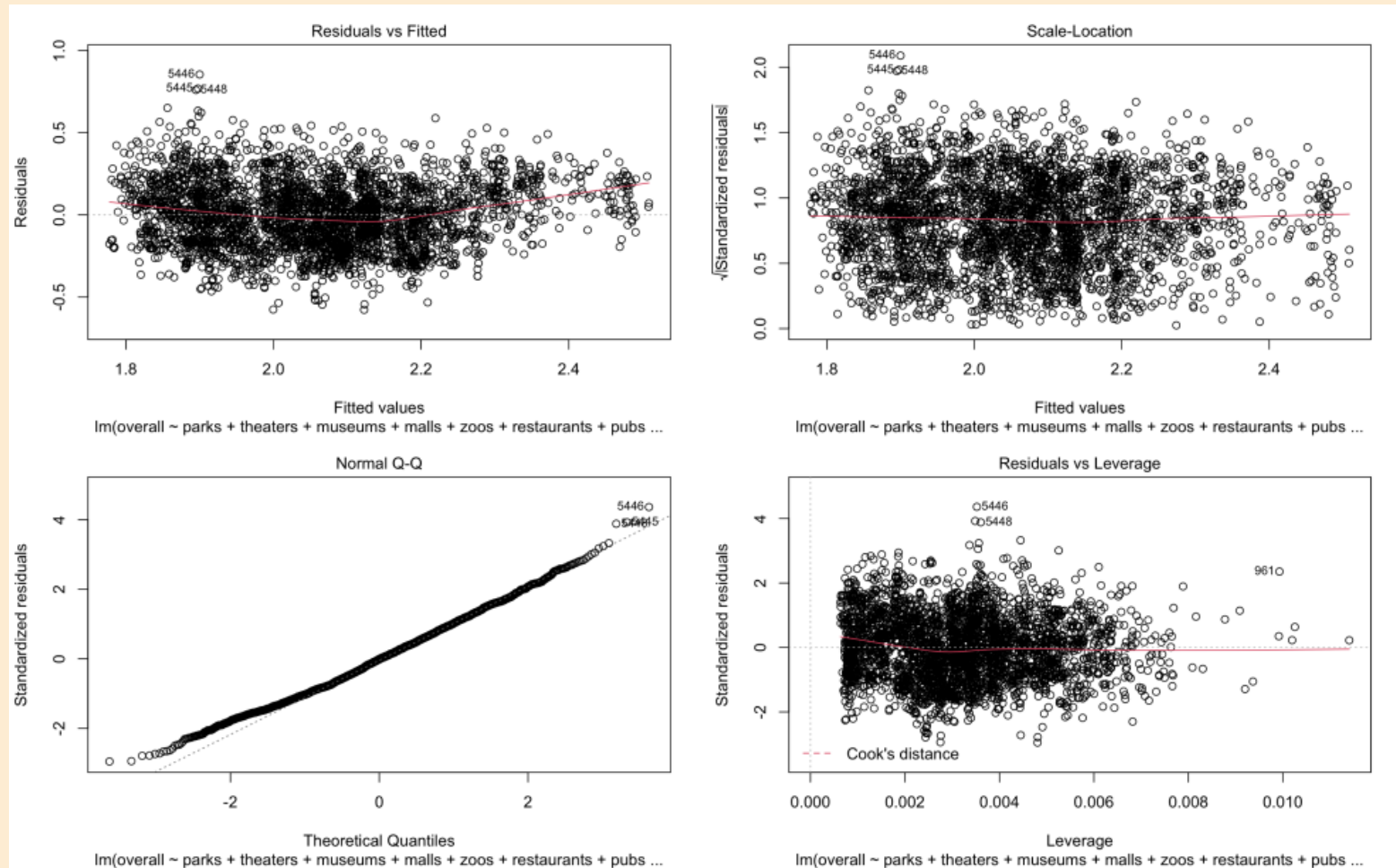
Residuals:
    Min       1Q   Median       3Q      Max
-0.57734 -0.14958 -0.00244  0.13146  0.85384

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.547617   0.014817 104.452 < 2e-16 ***
parks         0.040956   0.003304  12.396 < 2e-16 ***
theaters      0.026314   0.003541   7.430 1.36e-13 ***
museums      -0.006392   0.003434  -1.861  0.0628 .
malls        -0.013598   0.003149  -4.318 1.62e-05 ***
zoos          0.042822   0.004159  10.296 < 2e-16 ***
restaurants   0.002469   0.003304   0.747  0.4550
pubsandbars   0.027810   0.003672   7.573 4.66e-14 ***
localservs    0.025633   0.003044   8.420 < 2e-16 ***
hotels        0.036431   0.002735  13.321 < 2e-16 ***
juice         0.029169   0.002541  11.481 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

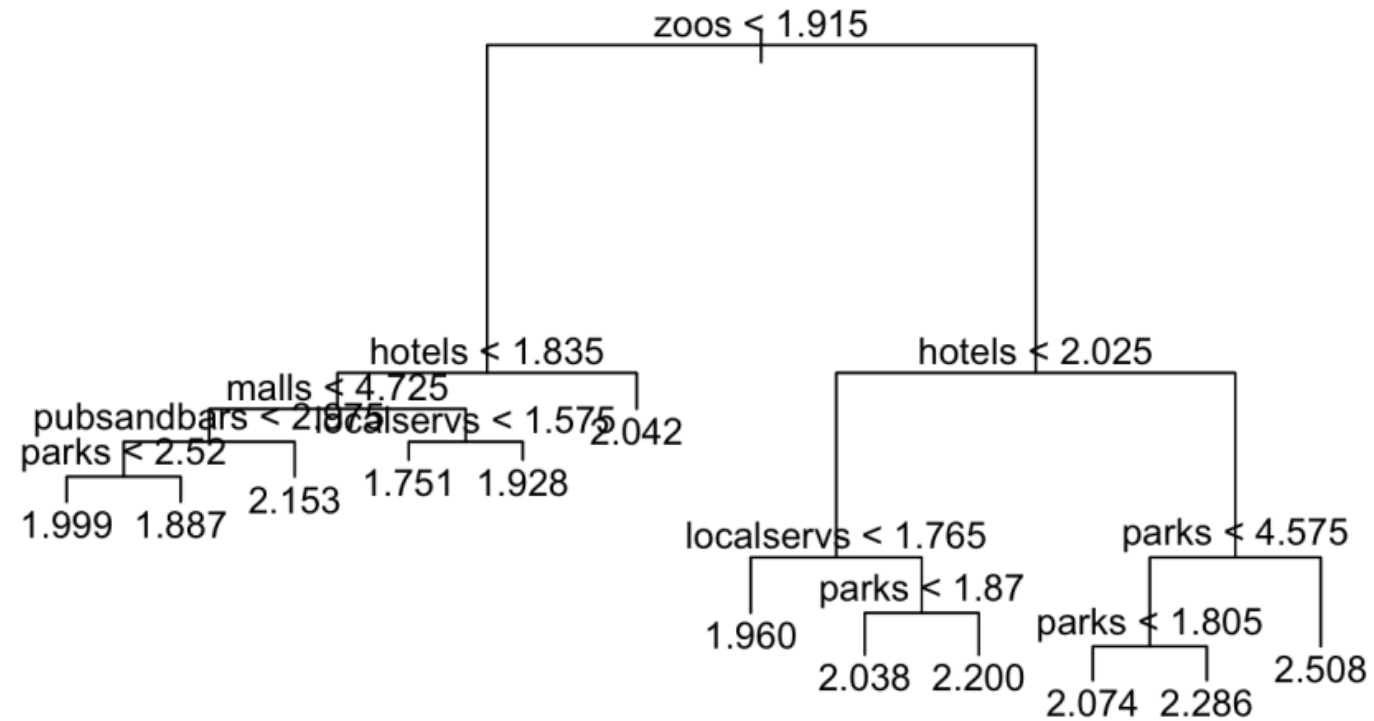
Residual standard error: 0.196 on 3456 degrees of freedom
Multiple R-squared:  0.3524,    Adjusted R-squared:  0.3506
F-statistic: 188.1 on 10 and 3456 DF,  p-value: < 2.2e-16
```

*this model only includes the 10 variables with the least left skewness

Reduced Model 1 Diagnostic Plots



Reduced Model 1 Decision Tree



Reduced Model 2

Reduced Model 2

formula =

overall~resorts+museums+malls+restaurants+pubsandbars+hotel
s+art+clubs+cafes+view+monuments+localservs

Multiple R ²	0.6353
Adjusted R ²	0.634

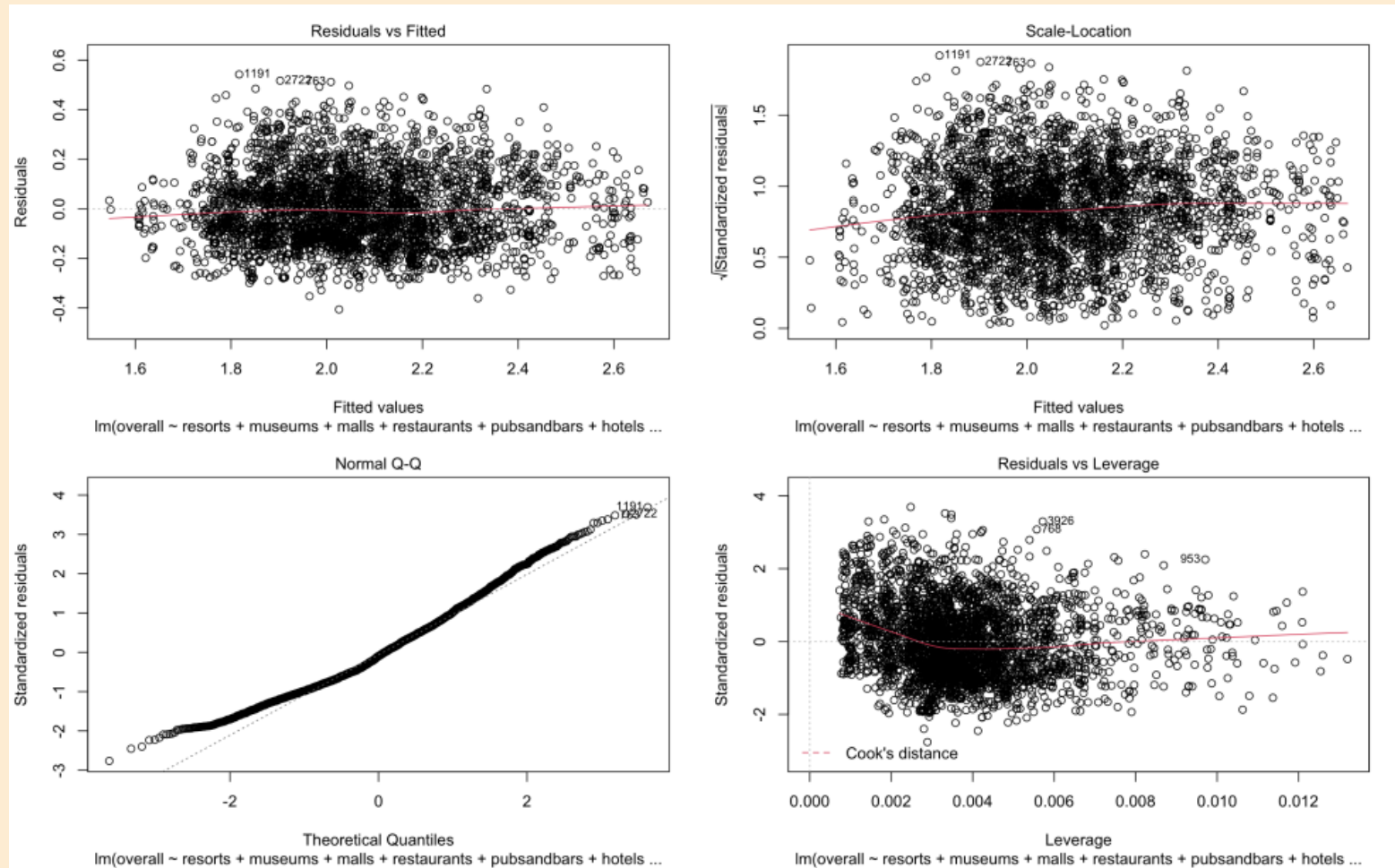
```
Residuals:
    Min       1Q   Median       3Q      Max
-0.40570 -0.11053 -0.01474  0.09212  0.54224

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.977145   0.015980  61.146 < 2e-16 ***
resorts      0.059899   0.001928  31.067 < 2e-16 ***
museums      0.047285   0.002354  20.090 < 2e-16 ***
malls        0.027176   0.002430  11.185 < 2e-16 ***
restaurants  0.008884   0.002436   3.647 0.00027 ***
pubsandbars  0.035848   0.002631  13.626 < 2e-16 ***
hotels       0.072028   0.001888  38.157 < 2e-16 ***
art          0.032683   0.001560  20.948 < 2e-16 ***
clubs        0.052612   0.002476  21.246 < 2e-16 ***
cafes        0.030643   0.003068   9.988 < 2e-16 ***
view         0.043770   0.001859  23.548 < 2e-16 ***
monuments    0.054393   0.002010  27.063 < 2e-16 ***
localservs   0.035887   0.002362  15.196 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

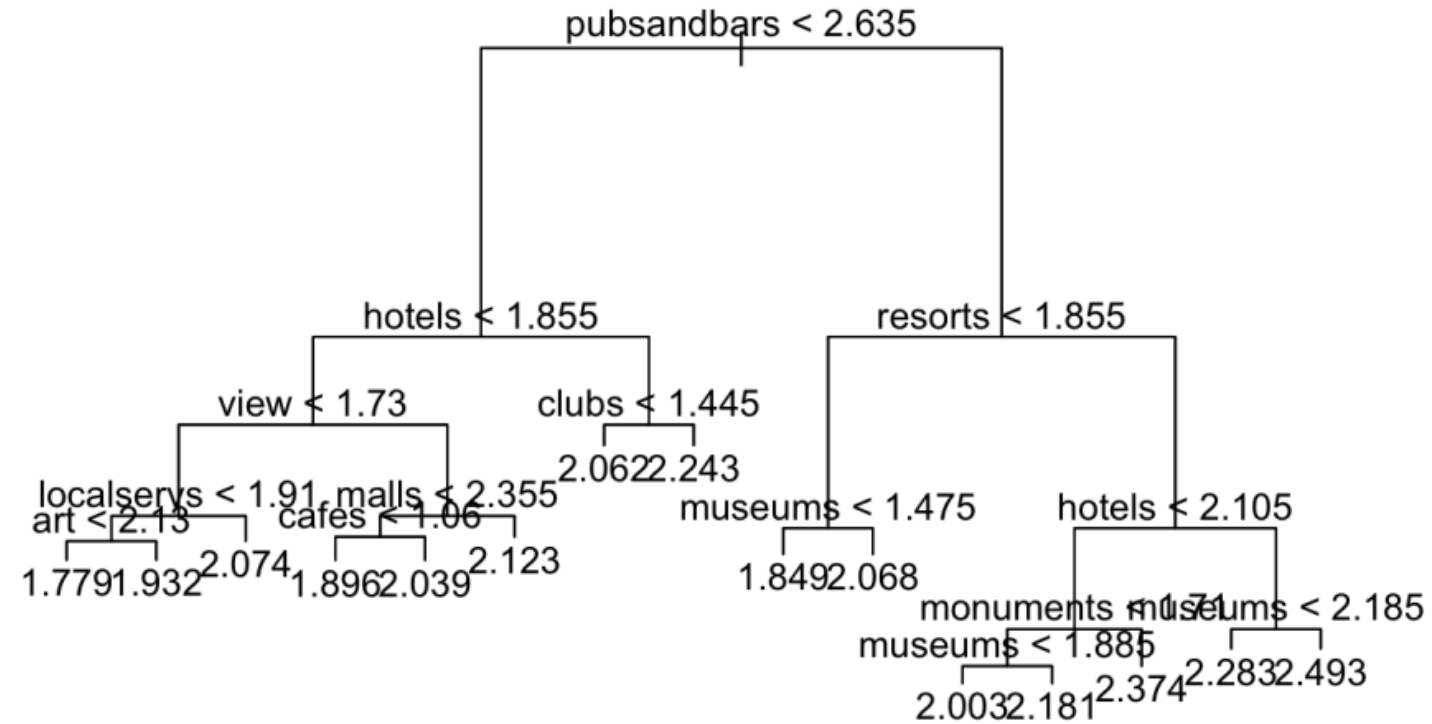
Residual standard error: 0.1471 on 3454 degrees of freedom
Multiple R-squared:  0.6353,    Adjusted R-squared:  0.634
F-statistic: 501.4 on 12 and 3454 DF,  p-value: < 2.2e-16
```

*this model includes 12 variables that were selected by making an educative guess

Reduced Model 2 Diagnostic Plots



Reduced Model 2 Decision Tree



Comparing Models

	Full Model	Reduced Model 1	Reduced Model 2
Multiple R ²	1 (overfitted)	0.3524	0.6353
Adjusted R ²	1 (overfitted)	0.3506	0.634

*Reduced Model 2 does a better job at predicting overall satisfaction than Full Model and Reduced Model 1. Reduced model 2 has a larger Multiple R² value and Adjusted R² value

CONCLUSION

The current study investigates the attractions that influence tourist's satisfaction when traveling across Europe. The result indicates that the most significant attractions affecting the overall satisfaction are resorts, museums, malls, restaurants, pubs and bars, hotels, art galleries, clubs, cafes, viewpoints, monuments, and local services

LIMITATIONS

Because I did not collect the data myself, I don't know how the data was collected or if it is trustworthy.

SOURCES

<https://www.kaggle.com/datasets/ishbhgtzms/travel-review-ratings>

THANK YOU!