

Eric_Hirsch_622_Final_Assignment

Predicting the Space Titanic Kaggle Competition

Eric Hirsch

12/3/2022

Contents

Summary	2
Introduction	2
Prediction using the Kaggle Spaceship Titanic Data Set	3
The Business Problem	3
Data Summary	3
Distributions	3
Missing Values	3
Multicollinearity	3
Outliers	4
Data Preparation	4
Feature Engineering	4
Modelling	4
Choosing and Testing Models	4
Hyperparameter Tuning	4
Results	4
Discussion	4
1. Data Exploration	4
A. Summary Statistics	4
Distributions	6
Look at Multicollinearity	6
Look at Missing Values	148
First Pass: Logistic regression	151
Data Preparation and Feature Engineering	153
1. Create Groups needs to be done before missing values correction since it counts group members	153

Handle Missing Variables	154
Create Cabin variables	154
Create Dummy Variables	167
Perform Logistic Regression With Features	172
Perform PCA	174
Perform Random Forest	174
Perform K-Means	174
Create a smaller database	174
Perform GBM	174
Perform SVM	175

Summary

Introduction

In machine learning, we predict target variables based on input variables. For this final exercise, we will apply various machine learning algorithms to a Kaggle data set (Spaceship Titanic) in order to predict which passengers have been transported to another dimension.

While it's tempting to throw as many algorithms at the problem as possible to see what sticks, the statistical fact is that while it is rare that a poor model will perform well on a holdout set, the chances of making false conclusions based on performance increases if we simply try one model after another. Besides, if we don't understand our model and our data, and the model becomes much more difficult to maintain.

When choosing models, we are balancing simplicity and complexity, and therefore tendencies to underfit or overfit. When the relationships in the data are simple and certain statistical conditions are met, parametric methods like OLS work well and have the advantage of being easily interpretable. If, for example, we are predicting height from weight, the relationship is simple enough that we can create a linear regression model and capture most of the variation that can be explained for these two variables.

When we increase our dimensions and/or complexity of relationships within the dataset, parametric methods are likely to underfit the data. Even in our simple height and weight example, if the relationship between height and weight varies considerably at lower weights, medium weights and higher weights, spline regression or another nonparametric technique will be necessary. As dimensions and complexity increases, we adopt techniques that are more powerful at morphing the data shape so that we can model the underlying structure, such as trees, SVM and neural nets.

Choosing the more complex algorithm will likely fit the training data better, but may be less interpretable and more subject to overfitting. With this in mind, each of these techniques has its advantages and disadvantages. In my experience with earlier datasets in this class, trees will pick up autonomous clusters in the data set better than SVMs. For example, if there were a small but significant anomalous cluster of individuals for whom height and weight were inversely related, trees will incorporate the cluster while SVMs will ignore it. Of course, clusters like this might signal a missing variable, but not all of the necessary variables will be found in any given data set. Trees may be bagged (e.g., Random Forest) or boosted (e.g. xgBoost) - either will generally perform better than a single decision tree. Because xgBoost is not a lazy learner, it will often have the upper hand in fitting the training data. On the other hand, when the relationships are more systematic and class boundaries are clear, SVMs may perform better because the kernel trick allows SVMs to radically change the data shape in order to find the class boundary. SVMs can also perform better when there is less data.

One of the biggest advantages of neural networks is that they effectively do the feature engineering for you if you can apply enough layers. They are also subject to the “double descent” phenomenon, which helps with managing underfitting. However, for a student using a home computer like myself, it’s often impractical to take advantage of these facts as the algorithm would run too long. Neural networks, like SVMs, also powerfully change the data shape in order to find class boundaries.

Accurate prediction depends not only on algorithm choice. We also need to engineer features (except possibly in very large neural nets) and tune hyperparameters. We also need to choose metrics that tell us whether or not our model is effective.

Prediction using the Kaggle Spaceship Titanic Data Set

For this exercise I’ve chosen a Kaggle Competition – the Kaggle Spaceship data set. The advantages of using this a competition data set are that we can compare our performance those of others. Achieving 90% on a holdout set in and of itself tells us nothing - we don’t know if achieving 95% would have been easy or impossible. In this competition, the 2,000 or so submitted accuracies on the leaderboard range from about 76% to 82%, which gives us a good idea of how well our model is working.

The main disadvantages of this data set are that the data is made up and the scenario a bit far-fetched. However, I wanted a data set that had a simple class as a target, as opposed to an image example, and the standard Titanic data set has been over analyzed, this was one of the few good choices.

The Business Problem

In the year 2912, the Spaceship Titanic, an interstellar passenger liner with almost 13,000 passengers on board, collided with a spacetime anomaly hidden within a dust cloud. Though the ship stayed intact, almost half of the passengers were transported to an alternate dimension. Our job is to predict which passengers were transported by the anomaly using records recovered from the spaceship’s damaged computer system.

Data Summary

The data set consists of 8693 records and 13 variables, including spending on the ship’s various amenities (VR Deck, Spa, Room Service, Food Court, Shopping Mall, cabin number, whether the individual was traveling with the group, whether the individual was a VIP, planet of origin and destination, and so on. These columns map to some degree with the original Titanic database. The target variable, Transported, is roughly equally distributed between false (4315) and true (4378).

Distributions

Missing Values 1073, or 12%, of records have missing values. The vast majority of missing values are found in the amenity expenditure columns. Oddly, the amenity expenditure rows with missing values are completely independent of each other - there are no records where more than one of these values is missing. This may be an artifact of the fact that the data is manufactured. In order to confirm that there is no systematic relationship between missing data and the target variable, we look at the Chi square between the target and a flag designating missing data. We do this for each amenity expenditure column and find no relationship between missing data and the target variable. We therefore eliminate rows with missing values for the training set. The test set, we impute the median.

Multicollinearity There is very little, even surprisingly little, multicollinearity in the database. In the case of variables that track spending on amenities this is most surprising, and may suggest that passengers were working within a budget and only spent money on the activities they liked most.

Outliers All of the spending variables are highly skewed, with very large ending occurring at the very end of the distribution. However, as most of our techniques are robust for outliers, records with extreme values remain in the database, as there is no reason to think that the spending is a data entry error or an anomalous occurrence.

Data Preparation

Feature Engineering The data set holds a number of opportunities for feature engineering. Through testing, it was found that the following new features were significant in predicting transportation. They are:

Modelling

Choosing and Testing Models

Hyperparameter Tuning

Results

Discussion

1. Data Exploration

A. Summary Statistics

```
## PassengerId      HomePlanet      CryoSleep      Cabin
## Length:8693      Length:8693      Length:8693      Length:8693
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
## Destination      Age      VIP      RoomService
## Length:8693      Min.   : 0.00  Length:8693      Min.   :    0.0
## Class :character  1st Qu.:19.00  Class :character  1st Qu.:    0.0
## Mode  :character  Median :27.00  Mode  :character  Median :    0.0
##                      Mean  :28.83                      Mean  :   224.7
##                      3rd Qu.:38.00                      3rd Qu.:   47.0
##                      Max.   :79.00                      Max.   :14327.0
##                      NA's   :179                      NA's   :181
## FoodCourt      ShoppingMall      Spa      VRDeck
## Min.   :    0.0  Min.   :    0.0  Min.   :    0.0  Min.   :    0.0
## 1st Qu.:    0.0  1st Qu.:    0.0  1st Qu.:    0.0  1st Qu.:    0.0
## Median :    0.0  Median :    0.0  Median :    0.0  Median :    0.0
## Mean   :  458.1  Mean   :  173.7  Mean   :  311.1  Mean   :  304.9
## 3rd Qu.:   76.0  3rd Qu.:   27.0  3rd Qu.:   59.0  3rd Qu.:   46.0
## Max.   :29813.0  Max.   :23492.0  Max.   :22408.0  Max.   :24133.0
## NA's   :183     NA's   :208     NA's   :183     NA's   :188
## Name      Transported
## Length:8693  Length:8693
## Class :character  Class :character
```

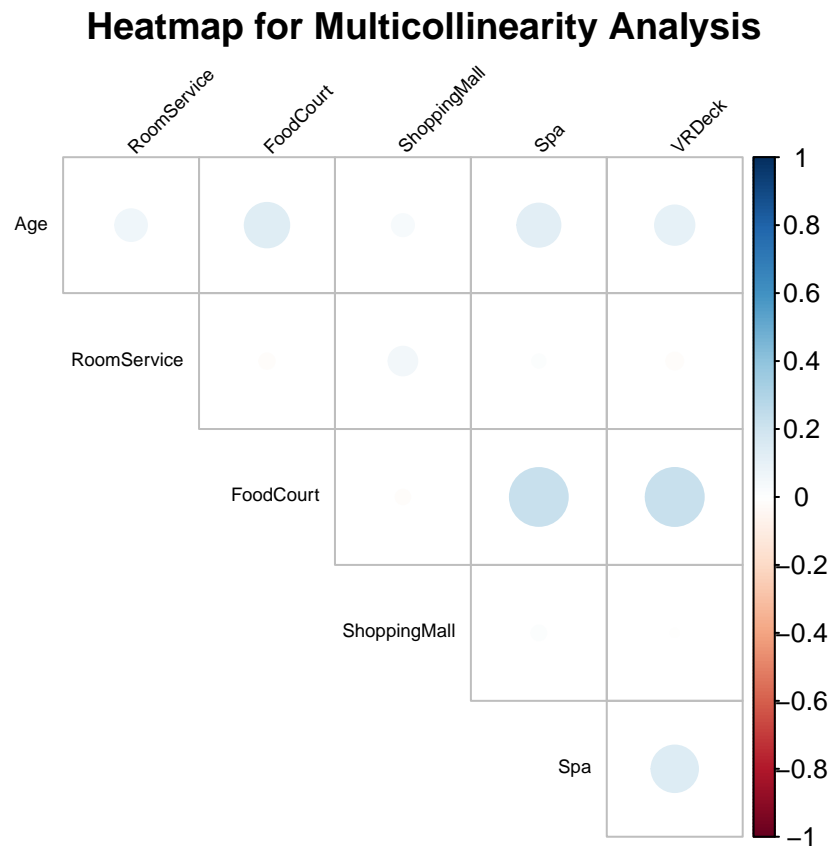
```
## Mode :character Mode :character
##
##
##
##
```

```
## 'data.frame': 8693 obs. of 14 variables:
## $ PassengerId : chr "0001_01" "0002_01" "0003_01" "0003_02" ...
## $ HomePlanet : chr "Europa" "Earth" "Europa" "Europa" ...
## $ CryoSleep : chr "False" "False" "False" "False" ...
## $ Cabin : chr "B/O/P" "F/O/S" "A/O/S" "A/O/S" ...
## $ Destination : chr "TRAPPIST-1e" "TRAPPIST-1e" "TRAPPIST-1e" "TRAPPIST-1e" ...
## $ Age : num 39 24 58 33 16 44 26 28 35 14 ...
## $ VIP : chr "False" "False" "True" "False" ...
## $ RoomService : num 0 109 43 0 303 0 42 0 0 0 ...
## $ FoodCourt : num 0 9 3576 1283 70 ...
## $ ShoppingMall: num 0 25 0 371 151 0 3 0 17 0 ...
## $ Spa : num 0 549 6715 3329 565 ...
## $ VRDeck : num 0 44 49 193 2 0 0 NA 0 0 ...
## $ Name : chr "Maham Ofracculy" "Juanna Vines" "Altark Susent" "Solam Susent" ...
## $ Transported : chr "False" "True" "False" "False" ...
```

```
## dfTrain$Transported n
## 1 False 4315
## 2 True 4378
```

Distributions

Look at Multicollinearity



```
## [[1]]
##           Age RoomService  FoodCourt ShoppingMall      Spa
## Age      1.00000000  0.06721473  0.13187705  0.032528456  0.12395659
## RoomService 0.06721473  1.00000000 -0.01565917  0.054897946  0.01165098
## FoodCourt   0.13187705 -0.01565917  1.00000000 -0.01388049  0.22326101
## ShoppingMall 0.03252846  0.05489795 -0.0138805  1.000000000  0.01444154
## Spa         0.12395659  0.01165098  0.22326101  0.014441541  1.00000000
## VRDeck      0.10342448 -0.01886955  0.22441645 -0.004451394  0.14436460
##           VRDeck
## Age      0.103424482
## RoomService -0.018869546
## FoodCourt   0.224416452
## ShoppingMall -0.004451394
## Spa         0.144364598
## VRDeck      1.000000000
##
## [[2]]
##           col1 col2 correlation
## 1           <NA> <NA>          <NA>
## 2           <NA> <NA>          <NA>
## 3           <NA> <NA>          <NA>
## 4           <NA> <NA>          <NA>
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

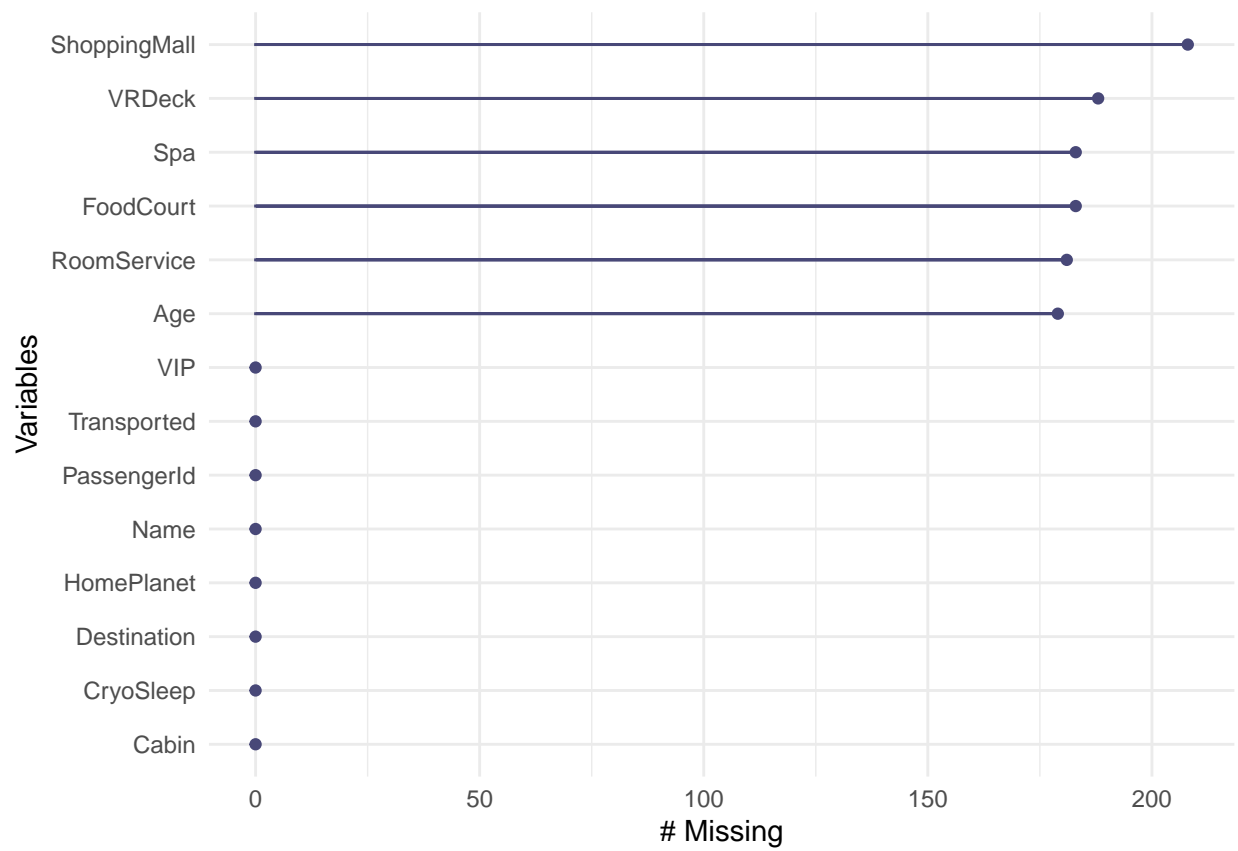
[illegible]

[illegible]

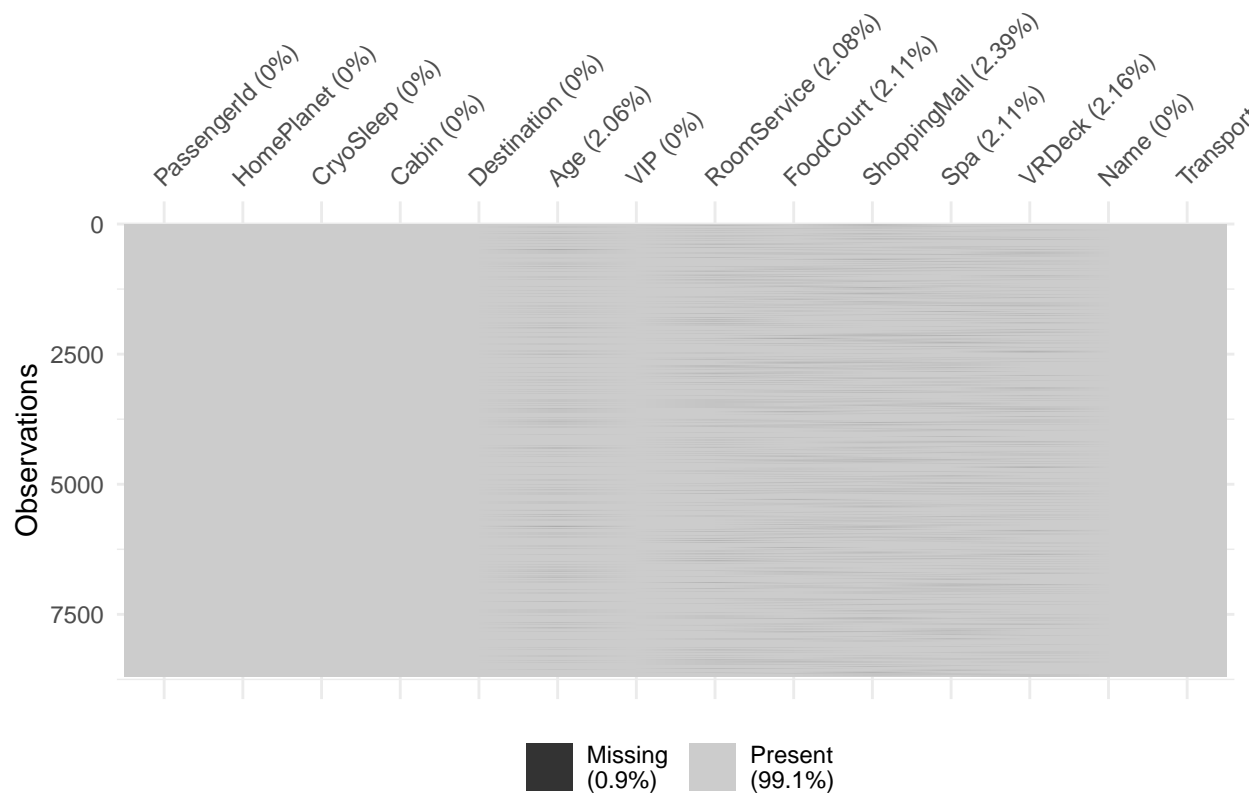
```
## 7619      <NA> <NA>      <NA>
## 7620      <NA> <NA>      <NA>
## 7621 No Values    0        0
```

Look at Missing Values

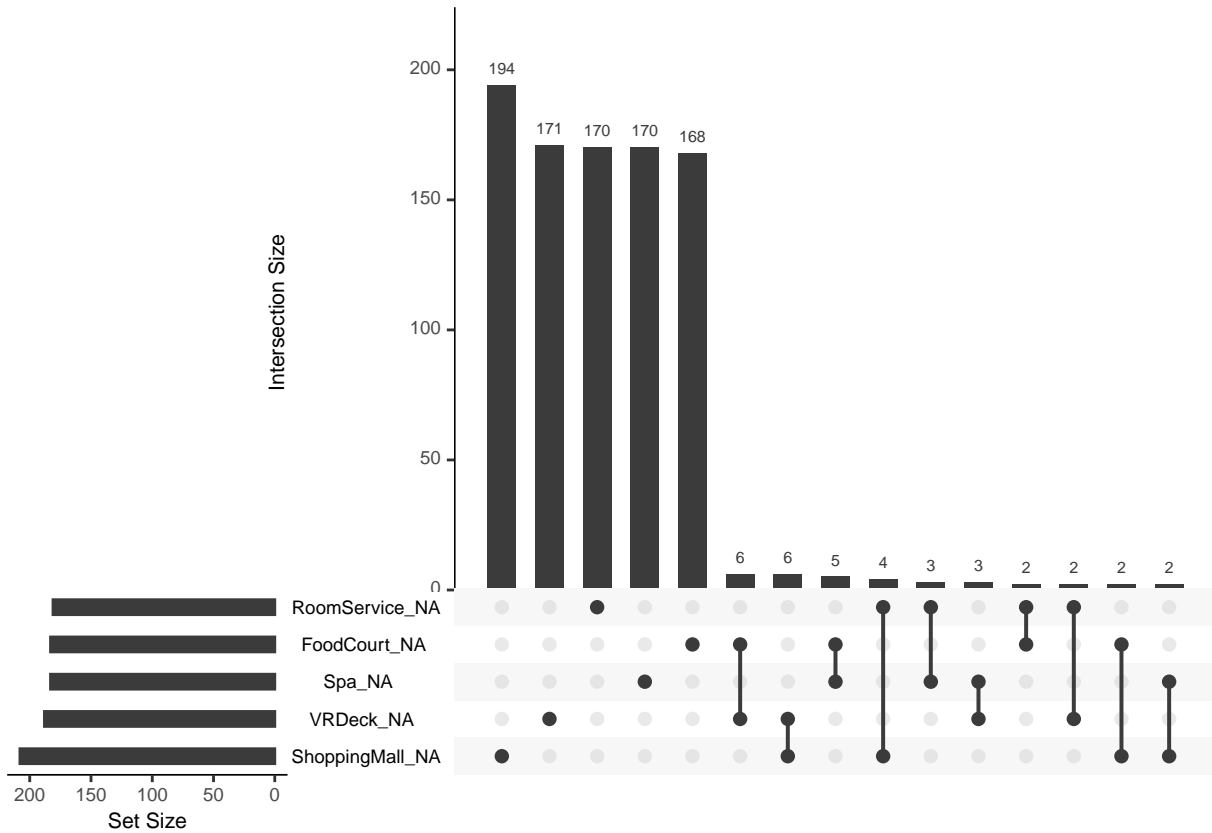
```
## [[1]]
```



```
##
## [[2]]
```



```
##  
## [[3]]
```



```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(dfMissingRecordsFlagAny)
## X-squared = 0.15887, df = 1, p-value = 0.6902

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(dfMissingRecordsFlag_SPA)
## X-squared = 0.0098187, df = 1, p-value = 0.9211

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(dfMissingRecordsFlag_FoodCourt)
## X-squared = 0.89665, df = 1, p-value = 0.3437

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(dfMissingRecordsFlag_VRDeck)
## X-squared = 0.1728, df = 1, p-value = 0.6776

##
```

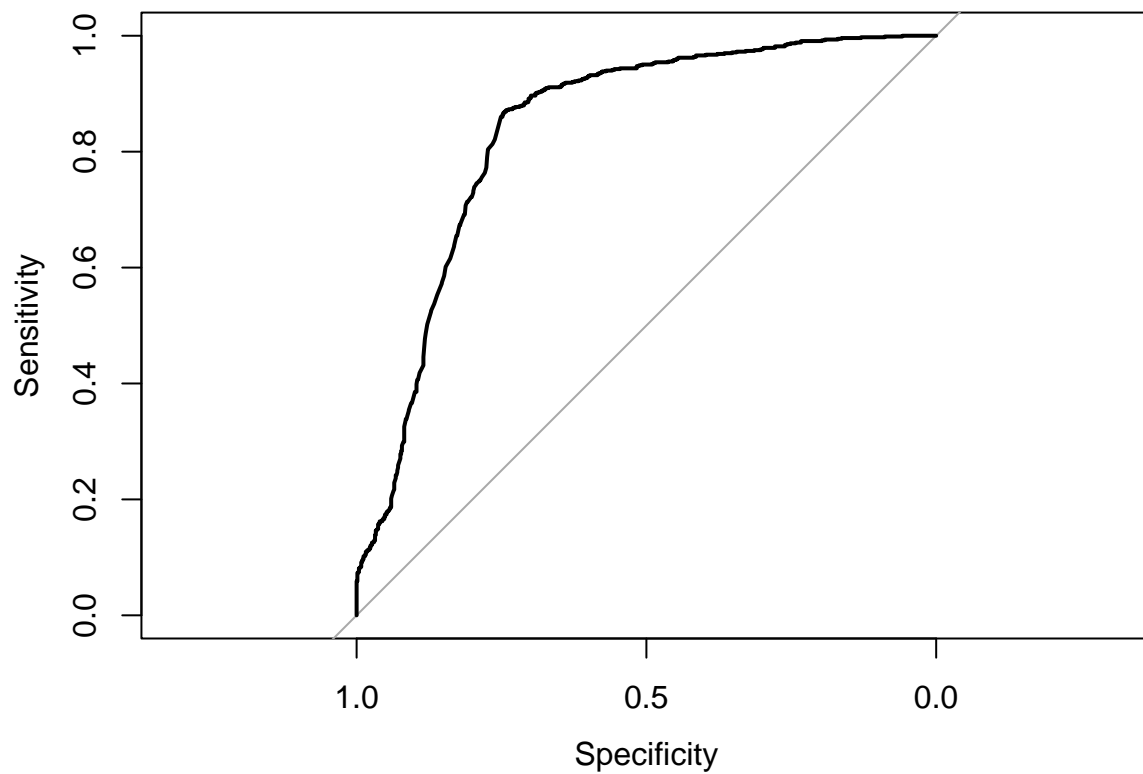
```
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: table(dfMissingRecordsFlag_ShoppingMall)
## X-squared = 1.5072, df = 1, p-value = 0.2196

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: table(dfMissingRecordsFlag_RoomService)
## X-squared = 1.3229, df = 1, p-value = 0.2501
```

First Pass: Logistic regression

```
##
## Call:
## glm(formula = fla, family = "binomial", data = train_reg)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4181  -0.8483   0.0219   0.8739   3.9956
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  7.026e-01  6.629e-02  10.598 < 2e-16 ***
## Age          2.523e-03  2.134e-03   1.182  0.23703
## RoomService -2.102e-03  1.061e-04 -19.815 < 2e-16 ***
## FoodCourt    7.383e-04  4.489e-05  16.448 < 2e-16 ***
## ShoppingMall 1.888e-04  6.179e-05   3.056  0.00224 **
## Spa          -2.358e-03  1.272e-04 -18.544 < 2e-16 ***
## VRDeck       -2.138e-03  1.175e-04 -18.196 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 8456.2  on 6099  degrees of freedom
## Residual deviance: 6302.3  on 6093  degrees of freedom
## (855 observations deleted due to missingness)
## AIC: 6316.3
##
## Number of Fisher Scoring iterations: 7
##
## Confusion Matrix and Statistics
##
##              Reference
## Prediction    0    1
##              0 504  68
##              1 251 697
##
##              Accuracy : 0.7901
##              95% CI : (0.7688, 0.8104)
##      No Information Rate : 0.5033
##      P-Value [Acc > NIR] : < 2.2e-16
```

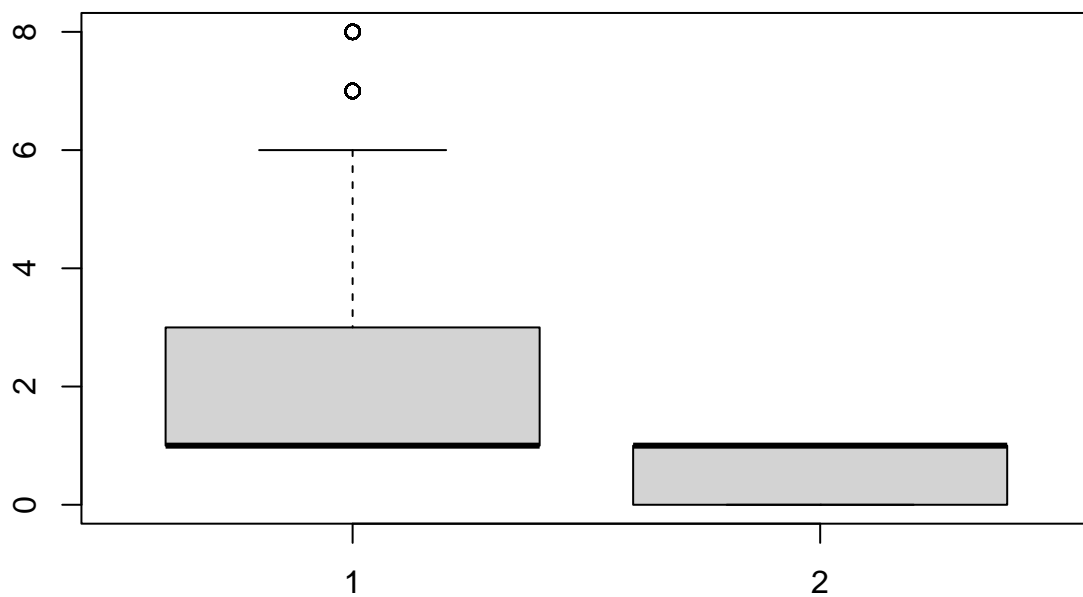
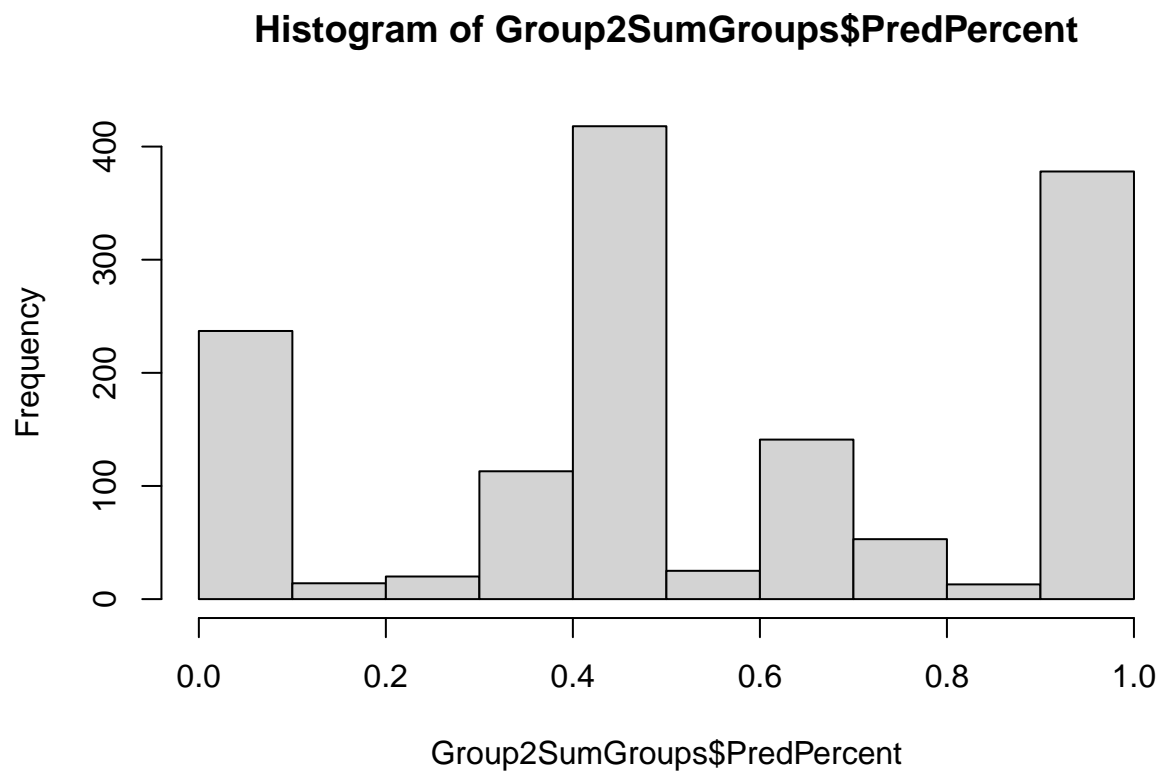
```
##
##           Kappa : 0.5796
##
## Mcnemar's Test P-Value : < 2.2e-16
##
##           Sensitivity : 0.6675
##           Specificity : 0.9111
##           Pos Pred Value : 0.8811
##           Neg Pred Value : 0.7352
##           Prevalence : 0.4967
##           Detection Rate : 0.3316
##           Detection Prevalence : 0.3763
##           Balanced Accuracy : 0.7893
##
##           'Positive' Class : 0
##
```



```
##
## Call:
## roc.default(response = dfPred_raw$class, predictor = dfPred_raw$predict_reg,      plot = TRUE)
##
## Data: dfPred_raw$predict_reg in 755 controls (dfPred_raw$class 0) < 765 cases (dfPred_raw$class 1).
## Area under the curve: 0.8373
```

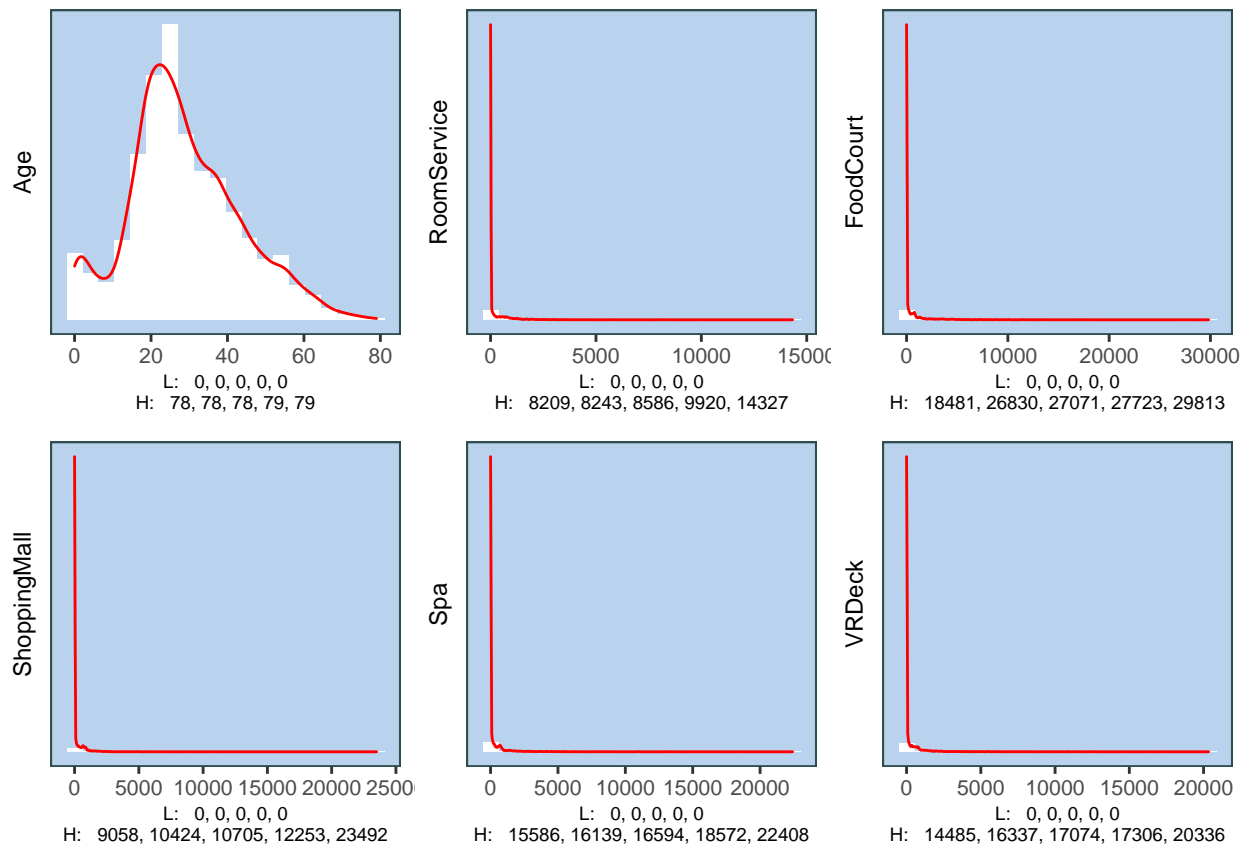

Data Preparation and Feature Engineering

1. Create Groups needs to be done before missing values correction since it counts group members

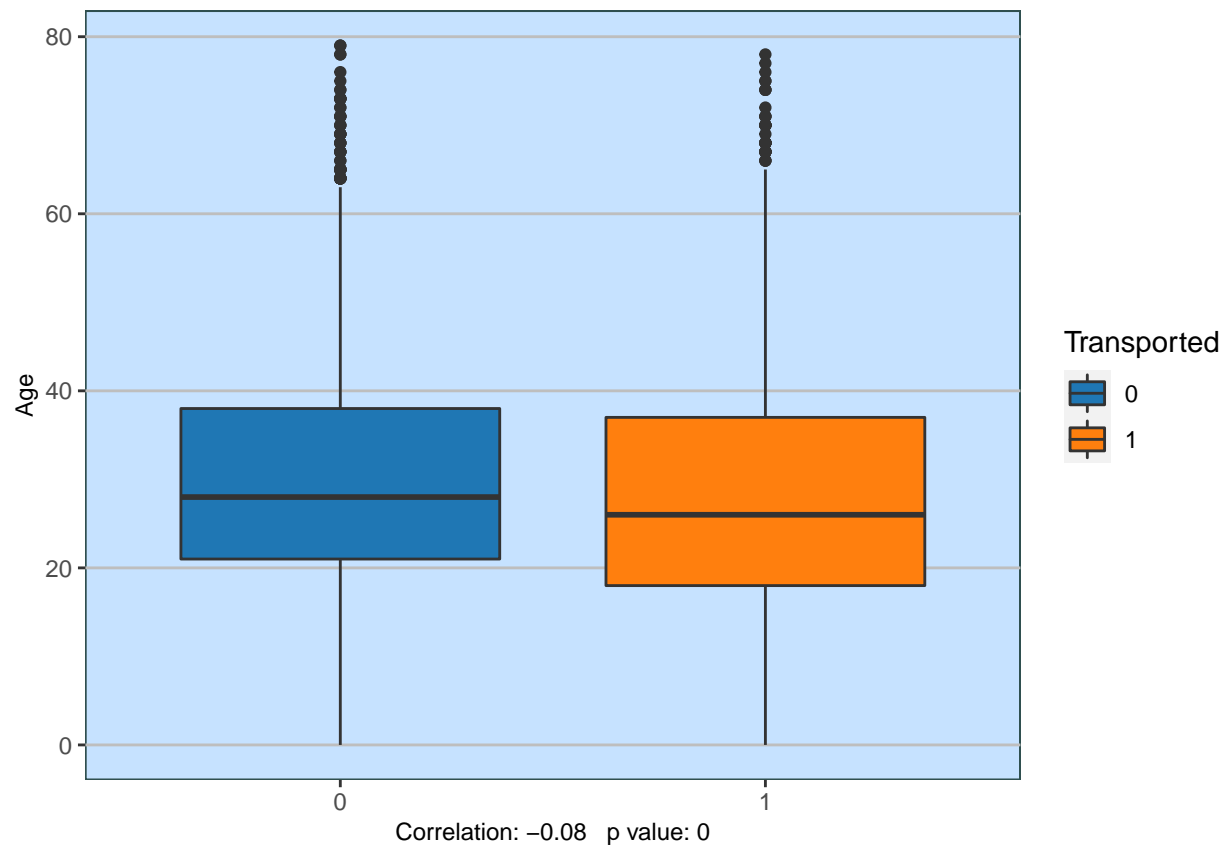


Handle Missing Variables

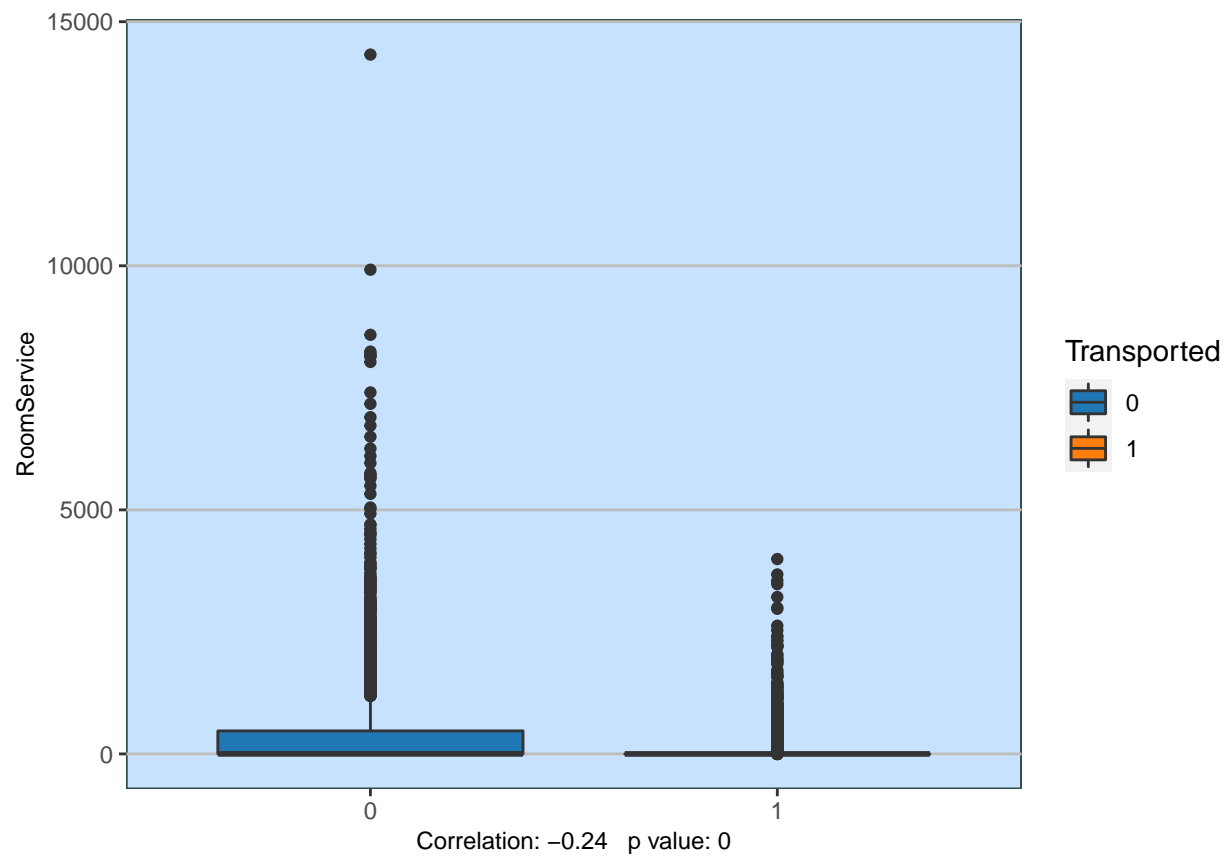
Create Cabin variables



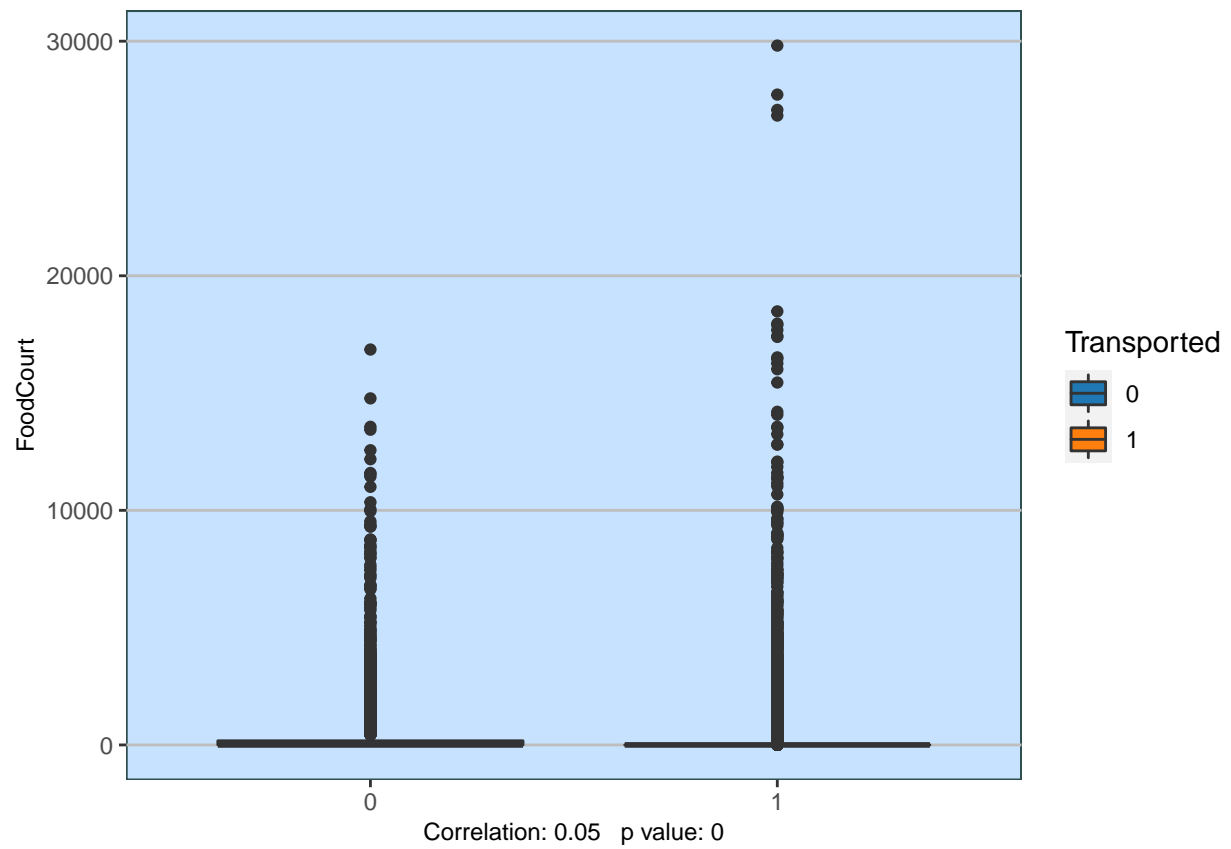
```
## [[1]]
```



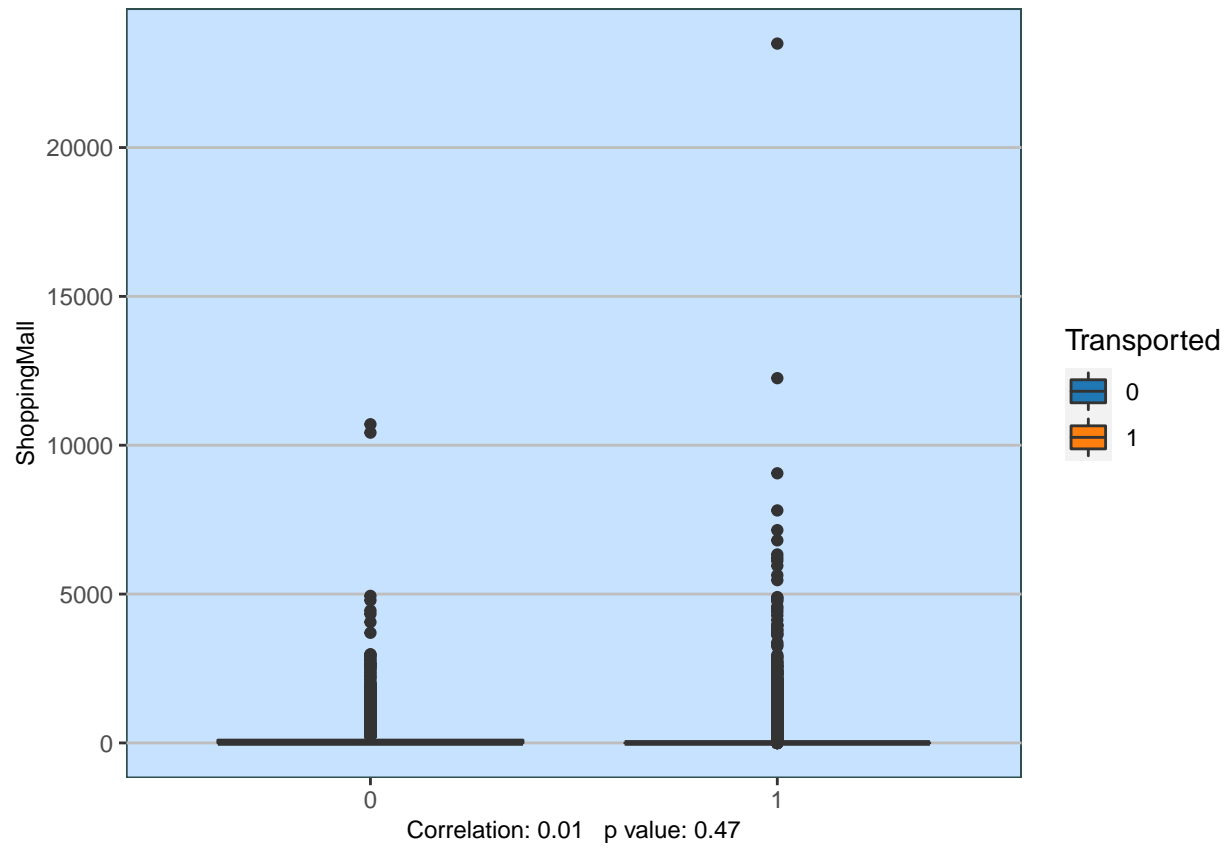
```
##  
## [[2]]
```



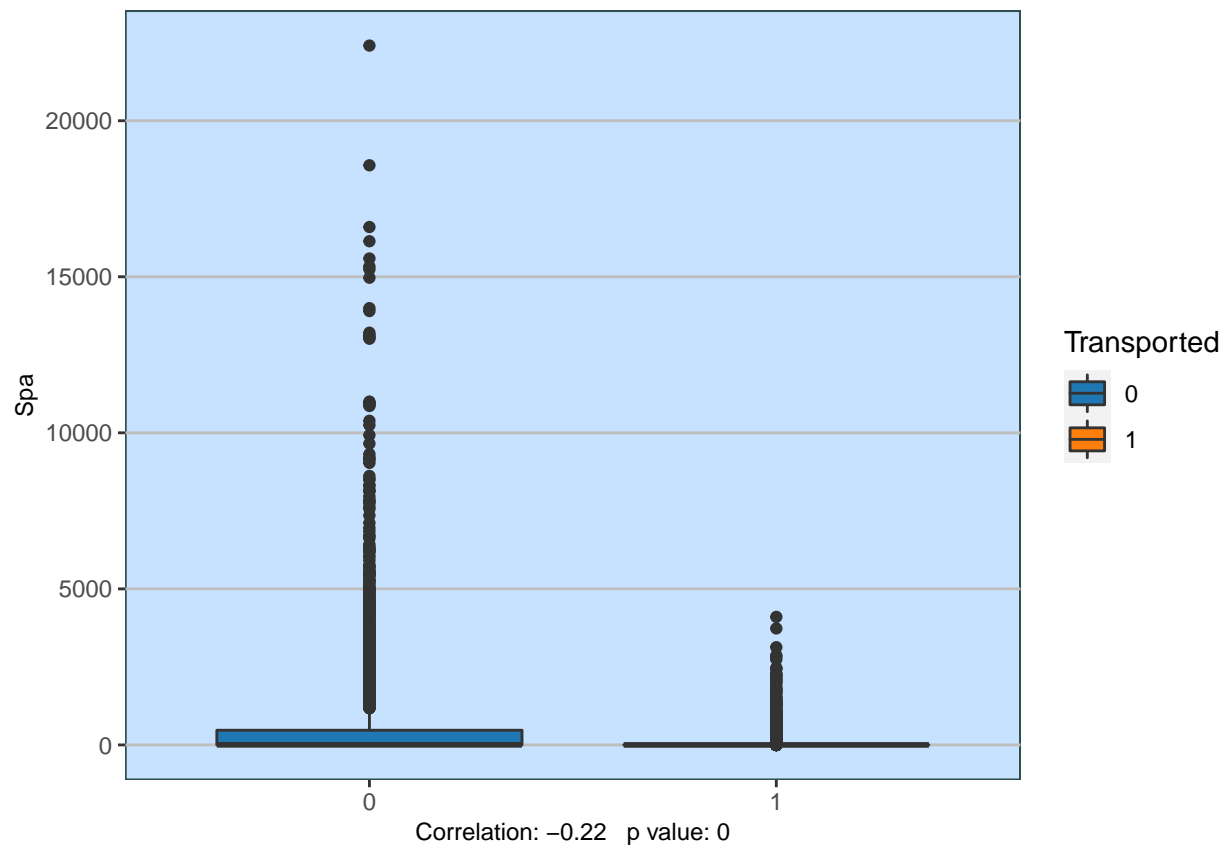
```
##  
## [[3]]
```



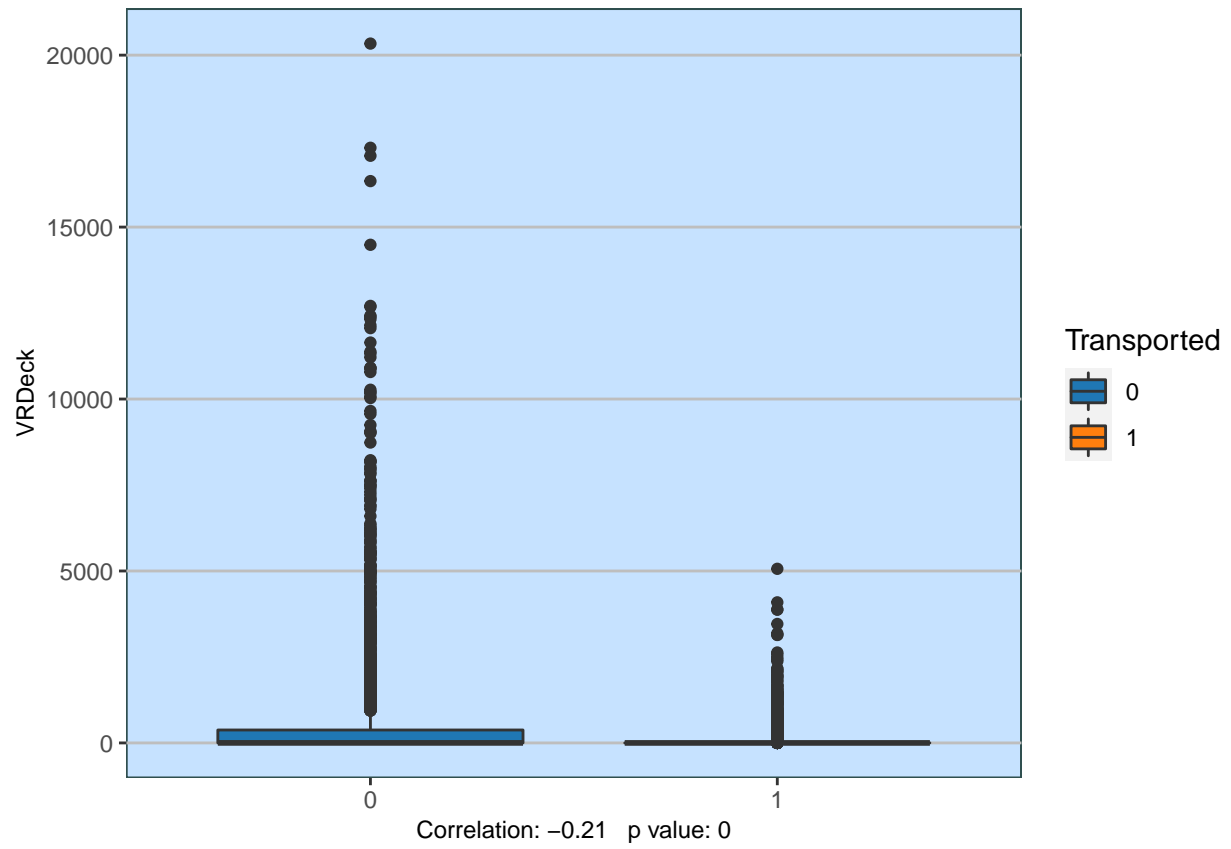
```
##  
## [[4]]
```



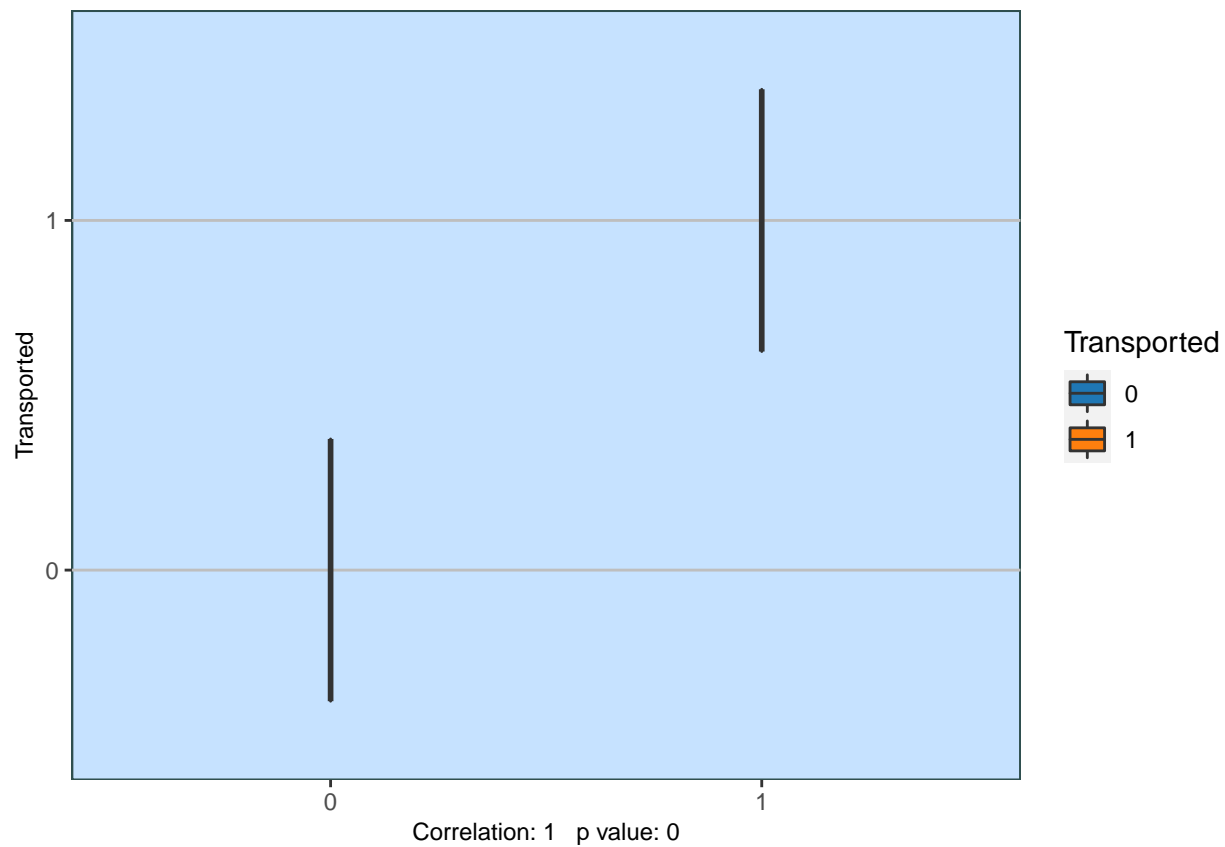
```
##  
## [[5]]
```



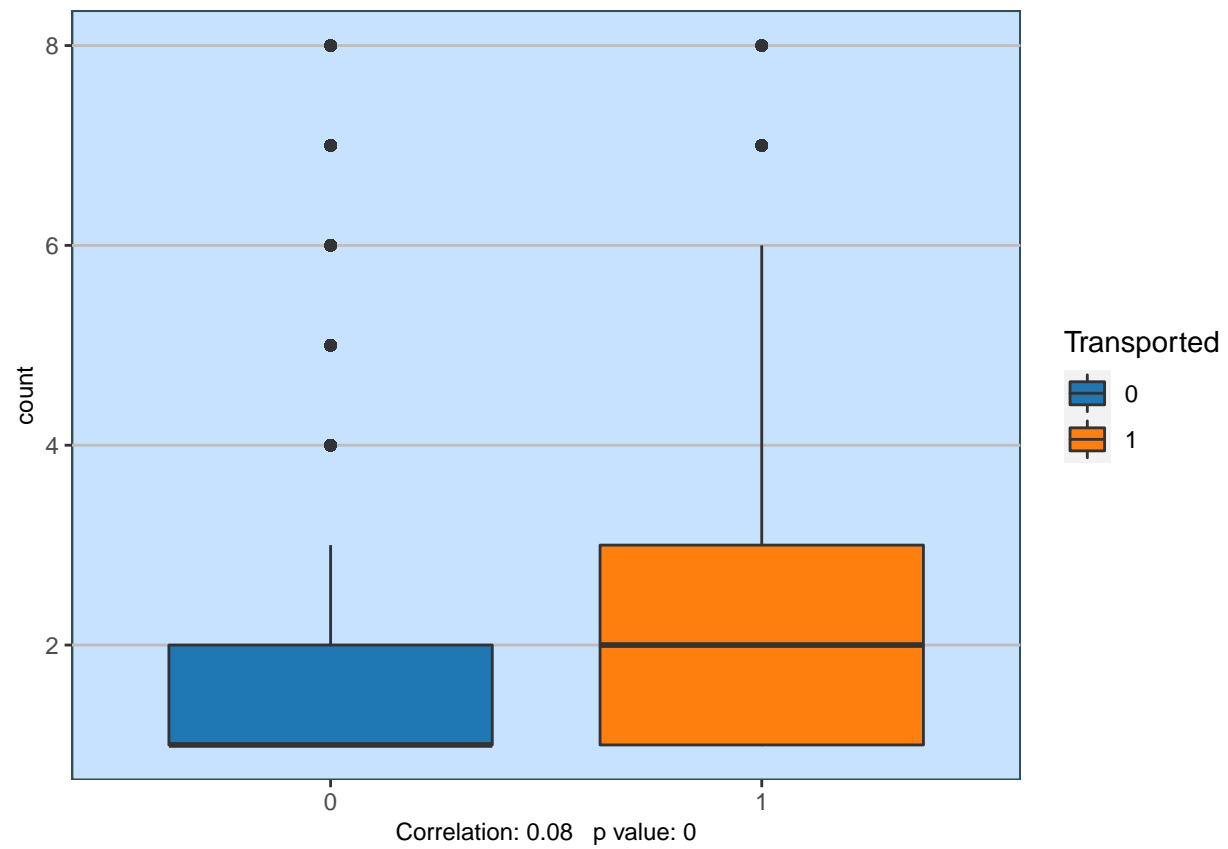
```
##  
## [[6]]
```



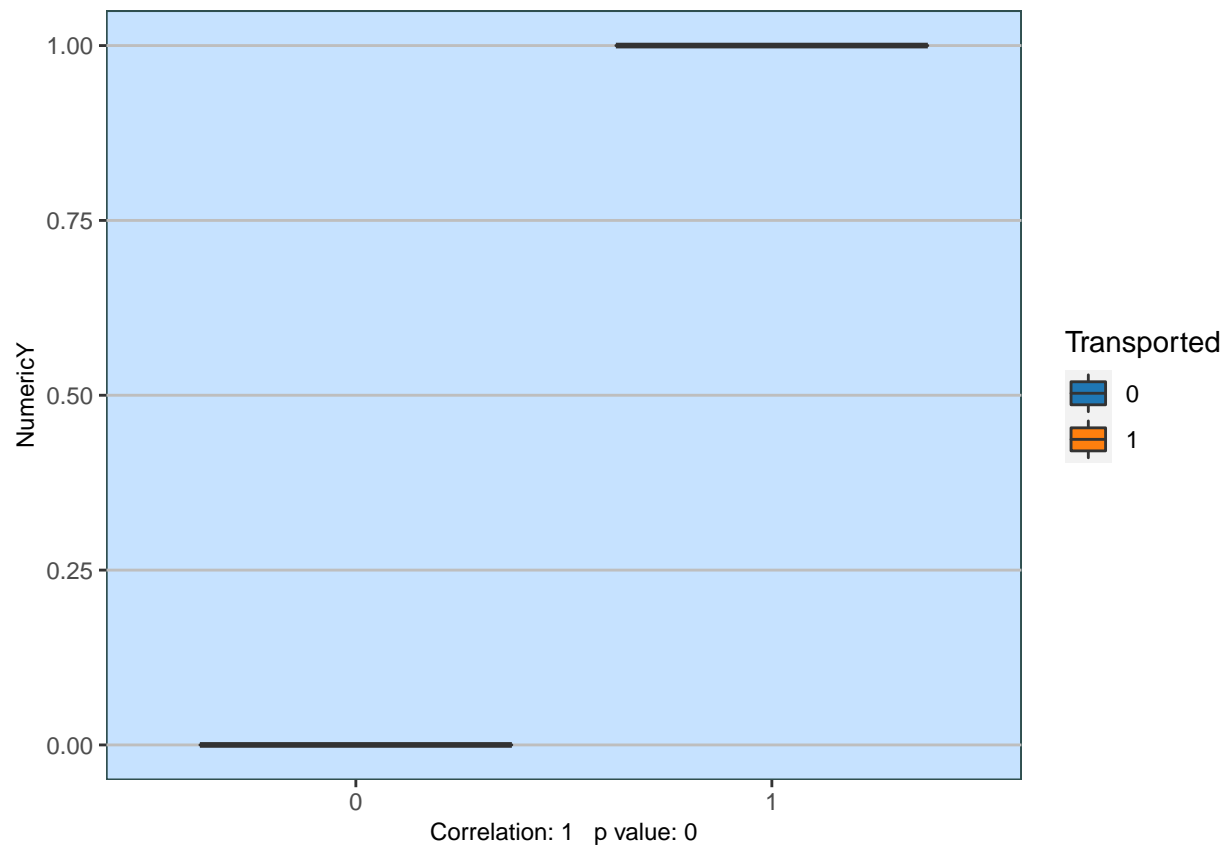
```
##  
## [[7]]
```

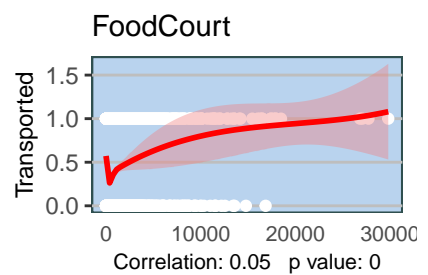
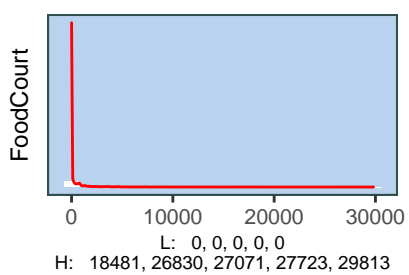
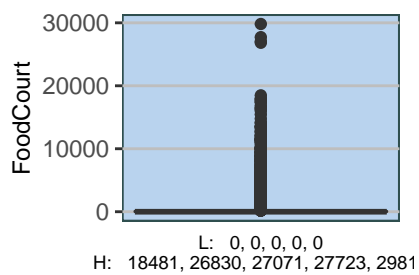
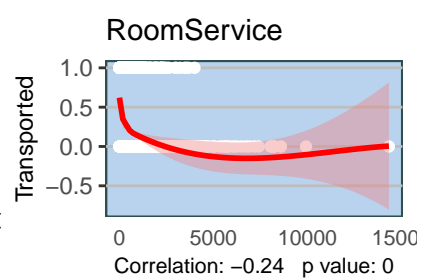
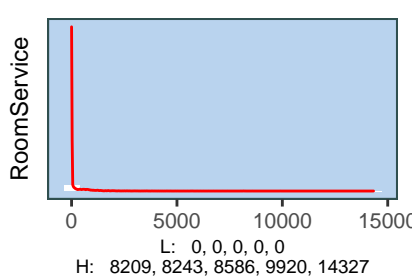
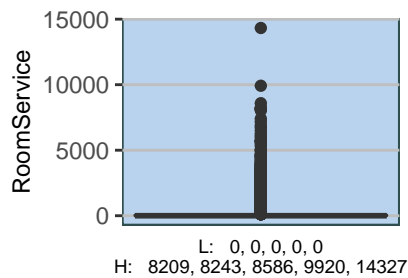
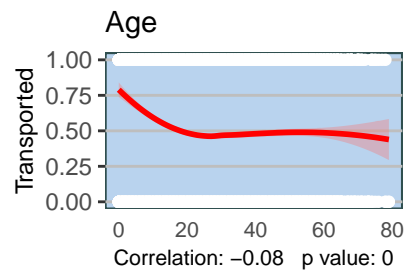
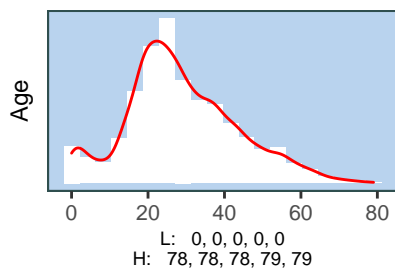
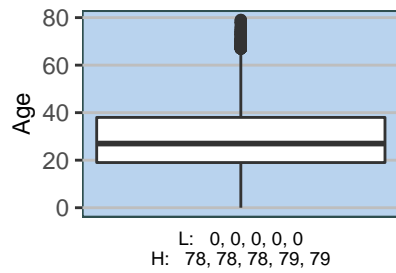



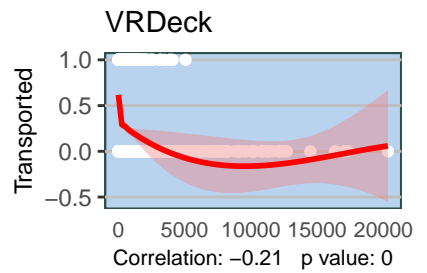
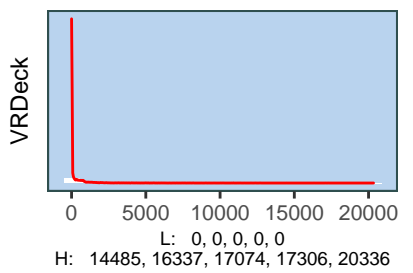
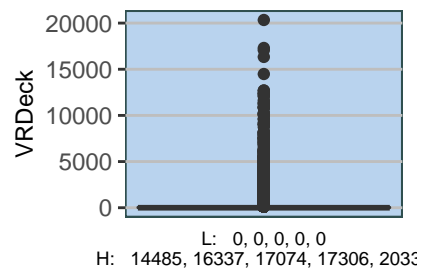
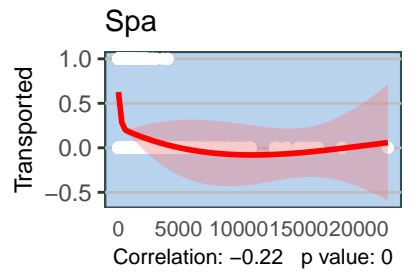
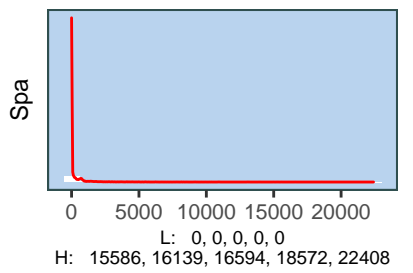
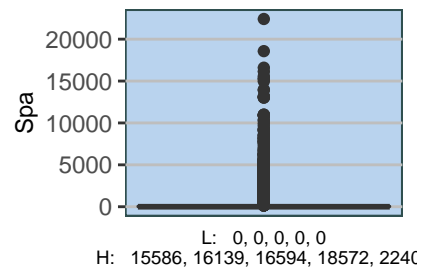
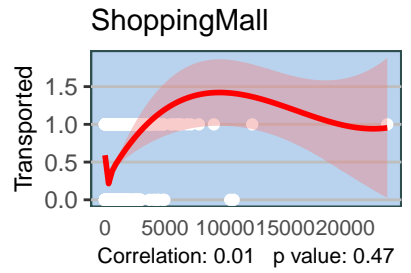
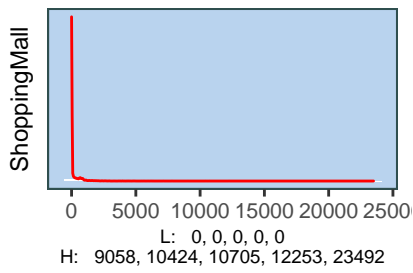
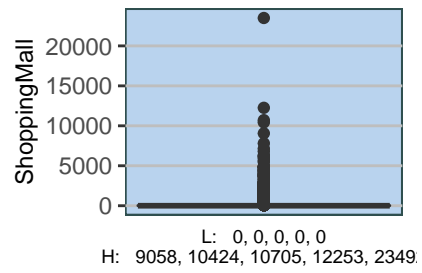
```
##  
## [[8]]
```

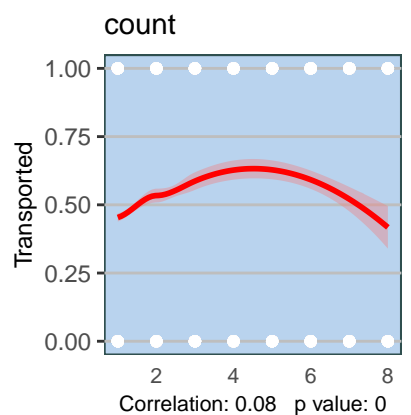
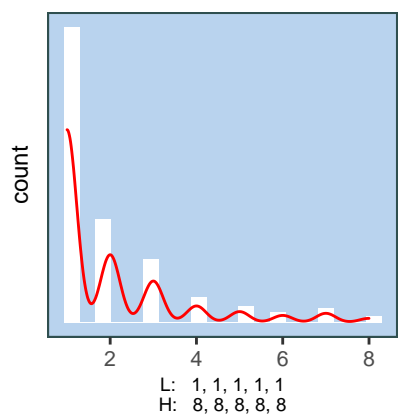
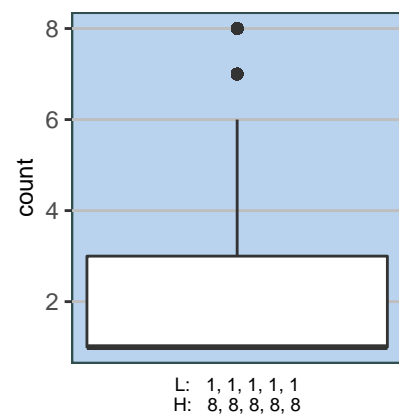
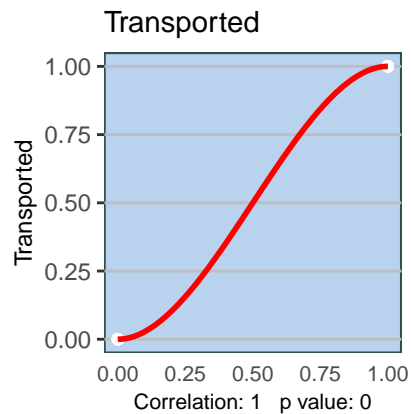
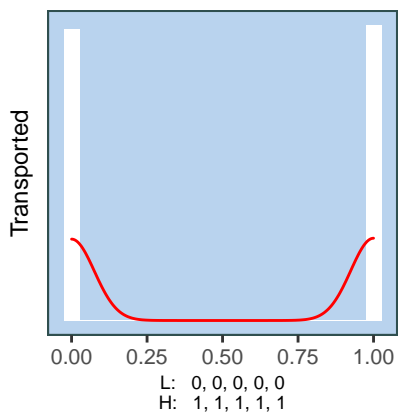
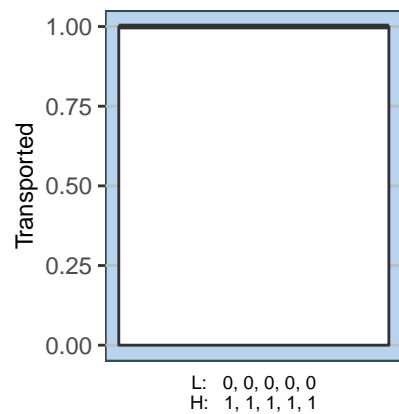


```
##
## [[9]]
```



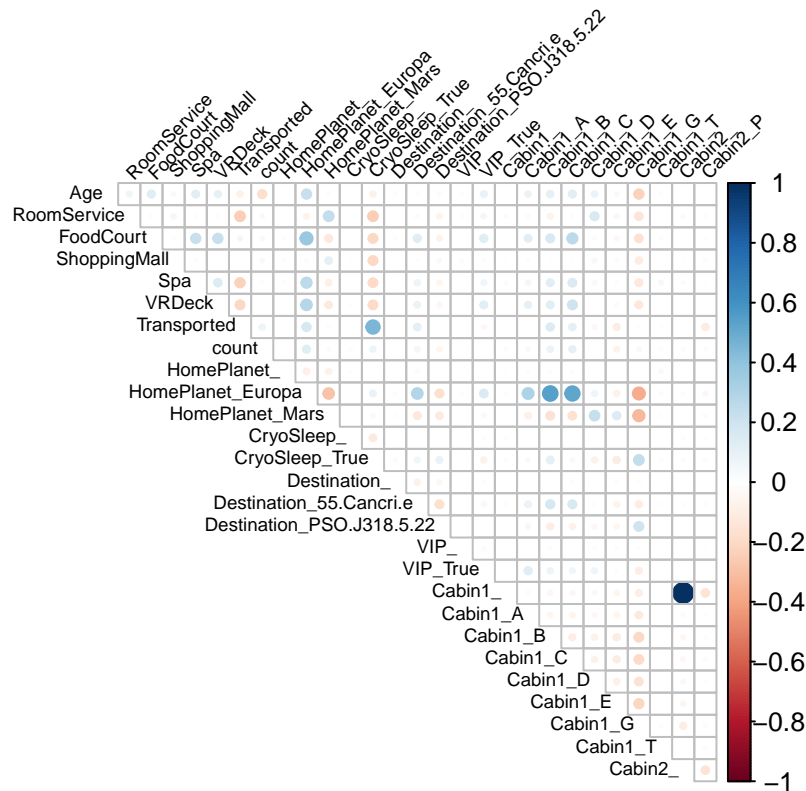






Create Dummy Variables

Heatmap for Multicollinearity Analysis



```
## [[1]]
##               Age RoomService   FoodCourt ShoppingMall
## Age           1.000000e+00  0.067214729  0.1318770457  0.032528456
## RoomService    6.721473e-02  1.000000000 -0.0156591729  0.054897946
## FoodCourt      1.318770e-01 -0.015659173  1.0000000000 -0.013888049
## ShoppingMall   3.252846e-02  0.054897946 -0.0138880494  1.000000000
## Spa            1.239566e-01  0.011650978  0.2232610114  0.014441541
## VRDeck         1.034245e-01 -0.018869546  0.2244164524 -0.004451394
## Transported    -7.653567e-02 -0.243334550  0.0510971132  0.008355925
## count         -1.739584e-01 -0.037147880  0.0325000352 -0.036571960
## HomePlanet_    1.011034e-03 -0.002205652 -0.0187494949 -0.016143892
## HomePlanet_Europa 2.246484e-01 -0.066859396  0.3759716833 -0.020994259
## HomePlanet_Mars 1.759243e-02  0.244230887 -0.1287904721  0.115128658
## CryoSleep_    -4.320066e-03 -0.005369292  0.0007767556 -0.005943743
## CryoSleep_True -7.478279e-02 -0.242706394 -0.2076388794 -0.206021976
## Destination_   1.325820e-02 -0.001463149 -0.0101150392 -0.005179735
## Destination_55.Cancr.i.e 1.969871e-02 -0.023228674  0.1339443567 -0.012656951
## Destination_PSO.J318.5.22 -2.795538e-02 -0.062144978 -0.0671260770 -0.032547920
## VIP_          -1.125434e-02  0.013177044 -0.0004801842 -0.023038426
## VIP_True       9.545633e-02  0.054169693  0.1374411877  0.014015755
## Cabin1_        4.374327e-05  0.026071871  0.0015919359 -0.008469795
## Cabin1_A       8.161195e-02 -0.024665996  0.1217369504 -0.014962472
## Cabin1_B       1.098803e-01 -0.067013280  0.1610910417 -0.006406563
## Cabin1_C       1.298787e-01 -0.017701241  0.2630768213  0.005529648
```

## Cabin1_D	8.108310e-02	0.164896444	0.0234282140	0.047195347
## Cabin1_E	3.120303e-02	0.046047738	-0.0418386506	0.027179696
## Cabin1_G	-2.219691e-01	-0.144860818	-0.1539558785	-0.108004978
## Cabin1_T	1.281369e-02	-0.004171434	0.0154076505	-0.006507588
## Cabin2_	4.374327e-05	0.026071871	0.0015919359	-0.008469795
## Cabin2_P	-9.964511e-03	0.011070578	-0.0190615723	0.021020625
##	Spa	VRDeck	Transported	count
## Age	0.123956590	0.1034244821	-0.076535669	-0.173958369
## RoomService	0.011650978	-0.0188695464	-0.243334550	-0.037147880
## FoodCourt	0.223261011	0.2244164524	0.051097113	0.032500035
## ShoppingMall	0.014441541	-0.0044513939	0.008355925	-0.036571960
## Spa	1.000000000	0.1443645979	-0.221510554	0.016832765
## VRDeck	0.144364598	1.0000000000	-0.206999865	0.009199261
## Transported	-0.221510554	-0.2069998647	1.000000000	0.079548785
## count	0.016832765	0.0091992613	0.079548785	1.000000000
## HomePlanet_	-0.010006351	0.0010233920	-0.002581189	-0.002655827
## HomePlanet_Europa	0.268337083	0.2852800045	0.176115162	0.131610791
## HomePlanet_Mars	-0.089397610	-0.1140965301	0.015205983	-0.020004958
## CryoSleep_	-0.002498412	-0.0143004987	-0.002560917	0.017696049
## CryoSleep_True	-0.198410732	-0.1950264972	0.454190618	0.077020673
## Destination_	0.009622819	-0.0092307479	-0.005372507	-0.009230657
## Destination_55.Cancr.i.e	0.077283074	0.0870261718	0.117779394	0.058544410
## Destination_PSO.J318.5.22	-0.054668779	-0.0474684228	0.002782008	-0.055420227
## VIP_	-0.005958410	0.0001752033	0.004294763	-0.004812183
## VIP_True	0.064125908	0.1240703921	-0.040820185	0.002847841
## Cabin1_	0.020918223	-0.0080009536	0.003477566	0.024558510
## Cabin1_A	0.058471514	0.1039959994	-0.002506483	0.032833709
## Cabin1_B	0.126550367	0.1267516662	0.140017430	0.080496563
## Cabin1_C	0.154352399	0.2034852762	0.111692402	0.102590293
## Cabin1_D	0.039250469	-0.0012852894	-0.040070045	-0.007542590
## Cabin1_E	-0.012847825	-0.0229009826	-0.099436095	-0.068702695
## Cabin1_G	-0.125929996	-0.1292098296	0.017553957	0.037241380
## Cabin1_T	0.043450976	0.0095575980	-0.011585224	-0.014921495
## Cabin2_	0.020918223	-0.0080009536	0.003477566	0.024558510
## Cabin2_P	-0.009267596	0.0047888862	-0.107483648	-0.019226335
##	HomePlanet_	HomePlanet_Europa	HomePlanet_Mars	
## Age	0.0010110337	0.224648403	0.017592427	
## RoomService	-0.0022056516	-0.066859396	0.244230887	
## FoodCourt	-0.0187494949	0.375971683	-0.128790472	
## ShoppingMall	-0.0161438923	-0.020994259	0.115128658	
## Spa	-0.0100063510	0.268337083	-0.089397610	
## VRDeck	0.0010233920	0.285280004	-0.114096530	
## Transported	-0.0025811889	0.176115162	0.015205983	
## count	-0.0026558272	0.131610791	-0.020004958	
## HomePlanet_	1.0000000000	-0.089553774	-0.078693600	
## HomePlanet_Europa	-0.0895537739	1.000000000	-0.288010349	
## HomePlanet_Mars	-0.0786936003	-0.288010349	1.000000000	
## CryoSleep_	-0.0141809085	0.005058453	0.004634628	
## CryoSleep_True	0.0008512312	0.090843863	0.032520764	
## Destination_	0.0009239070	-0.018503138	0.010189914	
## Destination_55.Cancr.i.e	-0.0208727999	0.290758338	-0.122512790	
## Destination_PSO.J318.5.22	-0.0052511145	-0.164479574	-0.111652038	
## VIP_	-0.0075820926	-0.013807325	0.004799192	
## VIP_True	0.0042595224	0.153266608	0.036898412	

## Cabin1_	0.0097871335	0.017711791	-0.006803652
## Cabin1_A	-0.0125257484	0.302815541	-0.088751051
## Cabin1_B	-0.0137316255	0.540672141	-0.159372844
## Cabin1_C	-0.0111090220	0.522166888	-0.154144824
## Cabin1_D	-0.0034544585	0.074514516	0.236502368
## Cabin1_E	0.0069351436	-0.077247687	0.143746103
## Cabin1_G	0.0037840985	-0.366959257	-0.322458159
## Cabin1_T	0.0339381390	0.026742754	-0.011529178
## Cabin2_	0.0097871335	0.017711791	-0.006803652
## Cabin2_P	-0.0046099767	-0.043971419	0.025889174
##	CryoSleep_	CryoSleep_True	Destination_
## Age	-0.0043200660	-0.0747827895	0.013258199
## RoomService	-0.0053692922	-0.2427063941	-0.001463149
## FoodCourt	0.0007767556	-0.2076388794	-0.010115039
## ShoppingMall	-0.0059437433	-0.2060219762	-0.005179735
## Spa	-0.0024984119	-0.1984107324	0.009622819
## VRDeck	-0.0143004987	-0.1950264972	-0.009230748
## Transported	-0.0025609172	0.4541906176	-0.005372507
## count	0.0176960486	0.0770206734	-0.009230657
## HomePlanet_	-0.0141809085	0.0008512312	0.000923907
## HomePlanet_Europa	0.0050584525	0.0908438633	-0.018503138
## HomePlanet_Mars	0.0046346281	0.0325207639	0.010189914
## CryoSleep_	1.0000000000	-0.1163128358	-0.006152675
## CryoSleep_True	-0.1163128358	1.0000000000	-0.029427509
## Destination_	-0.0061526748	-0.0294275094	1.0000000000
## Destination_55.Cancr i.e	-0.0019331181	0.0656617423	-0.075291965
## Destination_PSO.J318.5.22	-0.0077834004	0.0934633321	-0.046800545
## VIP_	0.0022713635	0.0080904969	0.000923907
## VIP_True	-0.0137767523	-0.0789547792	-0.010638483
## Cabin1_	0.0137652590	0.0246333615	-0.004711704
## Cabin1_A	0.0010019635	-0.0339520098	-0.004584605
## Cabin1_B	-0.0043970725	0.1192297419	0.001000367
## Cabin1_C	0.0075371196	0.0308843529	-0.018954096
## Cabin1_D	-0.0016852007	-0.0737728088	0.004952677
## Cabin1_E	-0.0043664514	-0.1086280317	0.019851644
## Cabin1_G	-0.0016805735	0.2430686055	-0.015836300
## Cabin1_T	-0.0036845223	-0.0165798083	-0.003366969
## Cabin2_	0.0137652590	0.0246333615	-0.004711704
## Cabin2_P	0.0125727615	-0.0250219948	-0.008333514
##	Destination_55.Cancr i.e	Destination_PSO.J318.5.22	
## Age	0.019698711	-0.027955383	
## RoomService	-0.023228674	-0.062144978	
## FoodCourt	0.133944357	-0.067126077	
## ShoppingMall	-0.012656951	-0.032547920	
## Spa	0.077283074	-0.054668779	
## VRDeck	0.087026172	-0.047468423	
## Transported	0.117779394	0.002782008	
## count	0.058544410	-0.055420227	
## HomePlanet_	-0.020872800	-0.005251114	
## HomePlanet_Europa	0.290758338	-0.164479574	
## HomePlanet_Mars	-0.122512790	-0.111652038	
## CryoSleep_	-0.001933118	-0.007783400	
## CryoSleep_True	0.065661742	0.093463332	
## Destination_	-0.075291965	-0.046800545	

## Destination_55.Cancr.i.e	1.000000000	-0.163250407		
## Destination_PSO.J318.5.22	-0.163250407	1.000000000		
## VIP_	0.002420478	0.006636435		
## VIP_True	0.049181710	-0.004201270		
## Cabin1_	0.014446398	-0.004465816		
## Cabin1_A	0.089120629	-0.050893440		
## Cabin1_B	0.175685973	-0.096187767		
## Cabin1_C	0.162716297	-0.078165730		
## Cabin1_D	0.013789175	-0.046456554		
## Cabin1_E	-0.066908537	-0.051857441		
## Cabin1_G	-0.102883341	0.206325439		
## Cabin1_T	-0.011744716	-0.007300369		
## Cabin2_	0.014446398	-0.004465816		
## Cabin2_P	-0.010733488	0.004364154		
##	VIP_	VIP_True	Cabin1_	Cabin1_A
## Age	-0.0112543403	0.095456329	4.374327e-05	0.081611947
## RoomService	0.0131770444	0.054169693	2.607187e-02	-0.024665996
## FoodCourt	-0.0004801842	0.137441188	1.591936e-03	0.121736950
## ShoppingMall	-0.0230384257	0.014015755	-8.469795e-03	-0.014962472
## Spa	-0.0059584103	0.064125908	2.091822e-02	0.058471514
## VRDeck	0.0001752033	0.124070392	-8.000954e-03	0.103995999
## Transported	0.0042947626	-0.040820185	3.477566e-03	-0.002506483
## count	-0.0048121834	0.002847841	2.455851e-02	0.032833709
## HomePlanet_	-0.0075820926	0.004259522	9.787134e-03	-0.012525748
## HomePlanet_Europa	-0.0138073251	0.153266608	1.771179e-02	0.302815541
## HomePlanet_Mars	0.0047991920	0.036898412	-6.803652e-03	-0.088751051
## CryoSleep_	0.0022713635	-0.013776752	1.376526e-02	0.001001963
## CryoSleep_True	0.0080904969	-0.078954779	2.463336e-02	-0.033952010
## Destination_	0.0009239070	-0.010638483	-4.711704e-03	-0.004584605
## Destination_55.Cancr.i.e	0.0024204778	0.049181710	1.444640e-02	0.089120629
## Destination_PSO.J318.5.22	0.0066364352	-0.004201270	-4.465816e-03	-0.050893440
## VIP_	1.0000000000	-0.024192056	4.112354e-03	-0.007502264
## VIP_True	-0.0241920558	1.000000000	4.696219e-03	0.135220045
## Cabin1_	0.0041123536	0.004696219	1.000000e+00	-0.027362298
## Cabin1_A	-0.0075022640	0.135220045	-2.736230e-02	1.000000000
## Cabin1_B	-0.0137316255	0.074737366	-4.913527e-02	-0.055888726
## Cabin1_C	0.0042329288	0.064278938	-4.752345e-02	-0.054055369
## Cabin1_D	-0.0110294151	0.070088570	-3.722356e-02	-0.042339795
## Cabin1_E	-0.0101063895	-0.011916690	-5.226345e-02	-0.059446862
## Cabin1_G	0.0113513025	-0.099130371	-9.941512e-02	-0.113079339
## Cabin1_T	-0.0035848760	-0.003544310	-3.554491e-03	-0.004043042
## Cabin2_	0.0041123536	0.004696219	1.000000e+00	-0.027362298
## Cabin2_P	-0.0080501905	0.003907184	-1.495407e-01	-0.013515618
##	Cabin1_B	Cabin1_C	Cabin1_D	Cabin1_E
## Age	0.109880254	0.129878691	0.081083099	0.031203033
## RoomService	-0.067013280	-0.017701241	0.164896444	0.046047738
## FoodCourt	0.161091042	0.263076821	0.023428214	-0.041838651
## ShoppingMall	-0.006406563	0.005529648	0.047195347	0.027179696
## Spa	0.126550367	0.154352399	0.039250469	-0.012847825
## VRDeck	0.126751666	0.203485276	-0.001285289	-0.022900983
## Transported	0.140017430	0.111692402	-0.040070045	-0.099436095
## count	0.080496563	0.102590293	-0.007542590	-0.068702695
## HomePlanet_	-0.013731625	-0.011109022	-0.003454459	0.006935144
## HomePlanet_Europa	0.540672141	0.522166888	0.074514516	-0.077247687

```

## HomePlanet_Mars -0.159372844 -0.154144824 0.236502368 0.143746103
## CryoSleep_ -0.004397072 0.007537120 -0.001685201 -0.004366451
## CryoSleep_True 0.119229742 0.030884353 -0.073772809 -0.108628032
## Destination_ 0.001000367 -0.018954096 0.004952677 0.019851644
## Destination_55.Cancric 0.175685973 0.162716297 0.013789175 -0.066908537
## Destination_PSO.J318.5.22 -0.096187767 -0.078165730 -0.046456554 -0.051857441
## VIP_ -0.013731625 0.004232929 -0.011029415 -0.010106389
## VIP_True 0.074737366 0.064278938 0.070088570 -0.011916690
## Cabin1_ -0.049135273 -0.047523455 -0.037223561 -0.052263453
## Cabin1_A -0.055888726 -0.054055369 -0.042339795 -0.059446862
## Cabin1_B 1.000000000 -0.097068798 -0.076030802 -0.106750459
## Cabin1_C -0.097068798 1.000000000 -0.073536710 -0.103248649
## Cabin1_D -0.076030802 -0.073536710 1.000000000 -0.080871276
## Cabin1_E -0.106750459 -0.103248649 -0.080871276 1.000000000
## Cabin1_G -0.203059857 -0.196398742 -0.153832684 -0.215987592
## Cabin1_T -0.007260208 -0.007022046 -0.005500138 -0.007722426
## Cabin2_ -0.049135273 -0.047523455 -0.037223561 -0.052263453
## Cabin2_P -0.016258396 -0.023506843 0.019756820 0.004056384
## Cabin1_G Cabin1_T Cabin2_ Cabin2_P
## Age -0.221969123 0.012813692 4.374327e-05 -0.009964511
## RoomService -0.144860818 -0.004171434 2.607187e-02 0.011070578
## FoodCourt -0.153955878 0.015407651 1.591936e-03 -0.019061572
## ShoppingMall -0.108004978 -0.006507588 -8.469795e-03 0.021020625
## Spa -0.125929996 0.043450976 2.091822e-02 -0.009267596
## VRDeck -0.129209830 0.009557598 -8.000954e-03 0.004788886
## Transported 0.017553957 -0.011585224 3.477566e-03 -0.107483648
## count 0.037241380 -0.014921495 2.455851e-02 -0.019226335
## HomePlanet_ 0.003784099 0.033938139 9.787134e-03 -0.004609977
## HomePlanet_Europa -0.366959257 0.026742754 1.771179e-02 -0.043971419
## HomePlanet_Mars -0.322458159 -0.011529178 -6.803652e-03 0.025889174
## CryoSleep_ -0.001680573 -0.003684522 1.376526e-02 0.012572762
## CryoSleep_True 0.243068606 -0.016579808 2.463336e-02 -0.025021995
## Destination_ -0.015836300 -0.003366969 -4.711704e-03 -0.008333514
## Destination_55.Cancric -0.102883341 -0.011744716 1.444640e-02 -0.010733488
## Destination_PSO.J318.5.22 0.206325439 -0.007300369 -4.465816e-03 0.004364154
## VIP_ 0.011351303 -0.003584876 4.112354e-03 -0.008050190
## VIP_True -0.099130371 -0.003544310 4.696219e-03 0.003907184
## Cabin1_ -0.099415116 -0.003554491 1.000000e+00 -0.149540727
## Cabin1_A -0.113079339 -0.004043042 -2.736230e-02 -0.013515618
## Cabin1_B -0.203059857 -0.007260208 -4.913527e-02 -0.016258396
## Cabin1_C -0.196398742 -0.007022046 -4.752345e-02 -0.023506843
## Cabin1_D -0.153832684 -0.005500138 -3.722356e-02 0.019756820
## Cabin1_E -0.215987592 -0.007722426 -5.226345e-02 0.004056384
## Cabin1_G 1.000000000 -0.014689536 -9.941512e-02 0.023111205
## Cabin1_T -0.014689536 1.000000000 -3.554491e-03 0.023769383
## Cabin2_ -0.099415116 -0.003554491 1.000000e+00 -0.149540727
## Cabin2_P 0.023111205 0.023769383 -1.495407e-01 1.000000000
##
## [[2]]
## col1 col2 correlation
## 1 Cabin1_ Cabin2_ 1

```

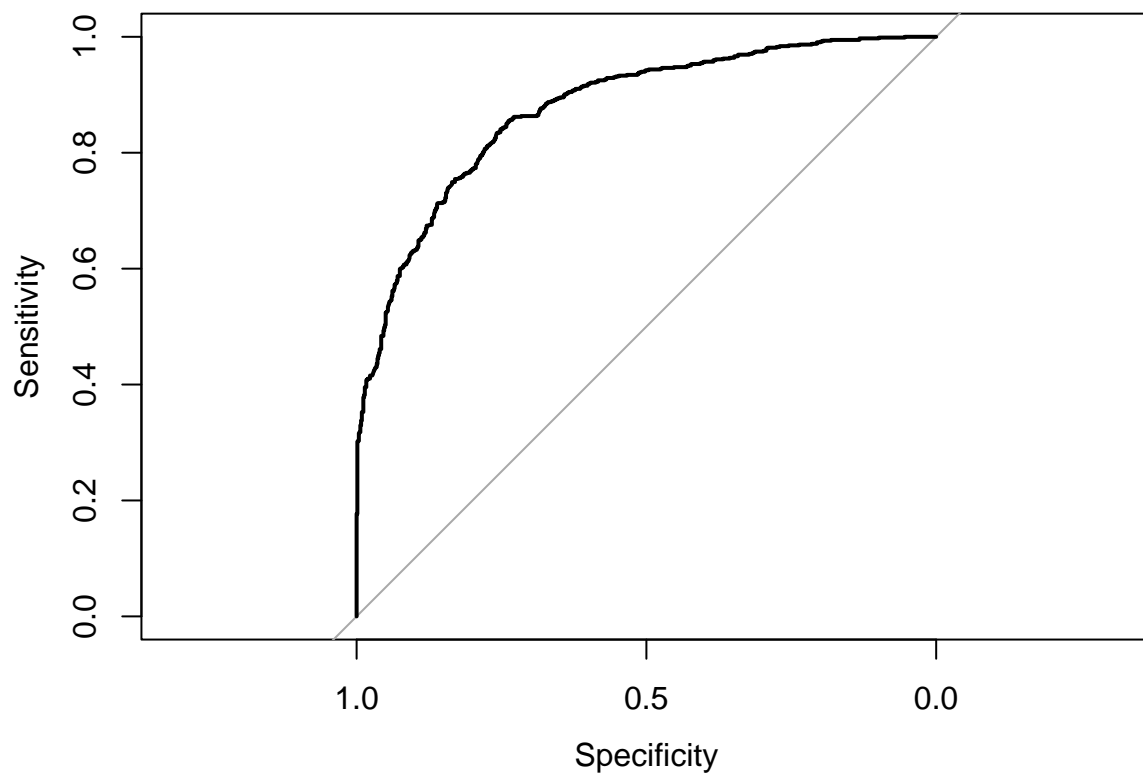
Perform Logistic Regression With Features

```
##
## Call:
## glm(formula = fla, family = "binomial", data = train_reg)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.14364  -0.67105   0.05778   0.69332   3.13064
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      2.518e-01  1.398e-01   1.802  0.07161 .
## Age             -2.751e-03  3.522e-03  -0.781  0.43480
## RoomService     -1.466e-03  1.058e-04 -13.854 < 2e-16 ***
## FoodCourt        5.360e-04  4.870e-05  11.006 < 2e-16 ***
## ShoppingMall     6.428e-04  1.046e-04   6.144 8.04e-10 ***
## Spa             -2.026e-03  1.259e-04 -16.089 < 2e-16 ***
## VRDeck          -1.988e-03  1.241e-04 -16.012 < 2e-16 ***
## HomePlanet_      3.639e-01  2.256e-01   1.613  0.10676
## HomePlanet_Europa 1.595e+00  2.607e-01   6.120 9.37e-10 ***
## HomePlanet_Mars   5.126e-01  1.151e-01   4.454 8.43e-06 ***
## CryoSleep_       2.055e-01  2.060e-01   0.998  0.31832
## CryoSleep_True   1.391e+00  9.793e-02  14.205 < 2e-16 ***
## Destination_     3.262e-01  2.377e-01   1.372  0.16998
## Destination_55.Cancr i.e 5.800e-01  9.808e-02   5.914 3.34e-09 ***
## Destination_PSO.J318.5.22 4.373e-02  1.113e-01   0.393  0.69437
## VIP_            -9.459e-03  2.243e-01  -0.042  0.96637
## VIP_True        -5.168e-01  3.137e-01  -1.647  0.09947 .
## Cabin1_         -7.405e-01  2.422e-01  -3.057  0.00223 **
## Cabin1_A        -9.093e-01  3.507e-01  -2.593  0.00952 **
## Cabin1_B         4.168e-01  3.124e-01   1.334  0.18209
## Cabin1_C         1.954e+00  3.649e-01   5.355 8.53e-08 ***
## Cabin1_D        -1.240e-01  2.004e-01  -0.619  0.53598
## Cabin1_E        -6.826e-01  1.241e-01  -5.502 3.75e-08 ***
## Cabin1_G        -5.193e-01  1.001e-01  -5.187 2.14e-07 ***
## Cabin1_T        -1.149e+01  2.065e+02  -0.056  0.95563
## Cabin2_          NA          NA          NA          NA
## Cabin2_P        -6.473e-01  7.026e-02  -9.213 < 2e-16 ***
## Group           4.174e-01  1.518e-01   2.749  0.00598 **
## Inter_CountShop  -3.551e-04  1.475e-04  -2.407  0.01610 *
## Inter_CountAge   -1.133e-02  4.890e-03  -2.317  0.02052 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 8450.0  on 6095  degrees of freedom
## Residual deviance: 5182.2  on 6067  degrees of freedom
## AIC: 5240.2
##
## Number of Fisher Scoring iterations: 12
##
## Confusion Matrix and Statistics
```

```

##
##           Reference
## Prediction  0   1
##           0 593 132
##           1 185 614
##
##           Accuracy : 0.792
##           95% CI : (0.7707, 0.8121)
##           No Information Rate : 0.5105
##           P-Value [Acc > NIR] : < 2.2e-16
##
##           Kappa : 0.5844
##
## Mcnemar's Test P-Value : 0.003493
##
##           Sensitivity : 0.7622
##           Specificity : 0.8231
##           Pos Pred Value : 0.8179
##           Neg Pred Value : 0.7685
##           Prevalence : 0.5105
##           Detection Rate : 0.3891
##           Detection Prevalence : 0.4757
##           Balanced Accuracy : 0.7926
##
##           'Positive' Class : 0
##

```



```
##  
## Call:  
## roc.default(response = dfPred_raw$class, predictor = dfPred_raw$predict_reg,      plot = TRUE)  
##  
## Data: dfPred_raw$predict_reg in 778 controls (dfPred_raw$class 0) < 746 cases (dfPred_raw$class 1).  
## Area under the curve: 0.8758
```

Perform PCA

Perform Random Forest

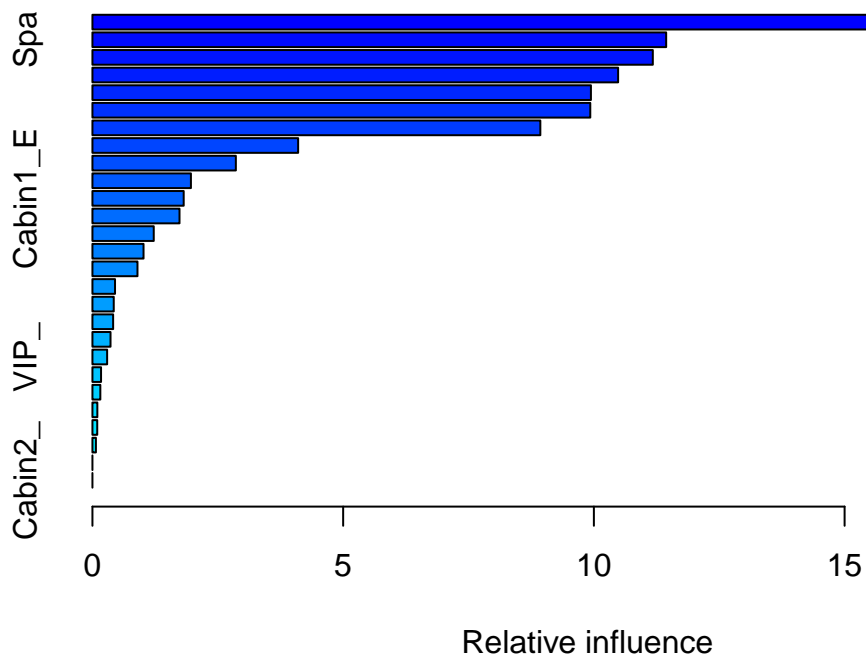
Didn't Do This:

Perfomr K-Means

Create a smaller database

Perform GBM

```
## gbm(formula = Transported ~ ., distribution = "gaussian", data = dfTrain5,
##      n.trees = 10000, interaction.depth = 4, shrinkage = 0.01)
## A gradient boosted model with gaussian loss function.
## 10000 iterations were performed.
## There were 27 predictors of which 25 had non-zero influence.
```



	var	rel.inf
## CryoSleep_True	CryoSleep_True	19.94118588
## Spa	Spa	11.44174754
## VRDeck	VRDeck	11.17560123
## RoomService	RoomService	10.48358046
## FoodCourt	FoodCourt	9.94099348
## Age	Age	9.92830480
## ShoppingMall	ShoppingMall	8.93237193
## Cabin1_G	Cabin1_G	4.10258937
## count	count	2.86023330
## Cabin2_P	Cabin2_P	1.96396781
## Cabin1_E	Cabin1_E	1.81958019
## HomePlanet_Europa	HomePlanet_Europa	1.73656746
## HomePlanet_Mars	HomePlanet_Mars	1.22240228
## Destination_PS0.J318.5.22	Destination_PS0.J318.5.22	1.01886146
## Destination_55.Cancric.e	Destination_55.Cancric.e	0.89843224
## Cabin1_	Cabin1_	0.45049454
## HomePlanet_	HomePlanet_	0.42499150
## Cabin1_C	Cabin1_C	0.41279166
## CryoSleep_	CryoSleep_	0.36047937
## VIP_	VIP_	0.29322517
## Destination_	Destination_	0.17200304
## Cabin1_A	Cabin1_A	0.15741628
## Cabin1_D	Cabin1_D	0.09767297
## VIP_True	VIP_True	0.09555446
## Cabin1_B	Cabin1_B	0.06895158
## Cabin1_T	Cabin1_T	0.00000000
## Cabin2_	Cabin2_	0.00000000

Perform SVM

Perfomr K means

Perform XGBoost - also does some feature engineering

## [1] train-rmse:0.415047+0.000841	train-auc:0.813023+0.001987	test-rmse:0.416344+0.005941	test-auc:0.813023+0.001987
## [2] train-rmse:0.385845+0.001185	train-auc:0.860516+0.002116	test-rmse:0.390597+0.007574	test-auc:0.860516+0.002116
## [3] train-rmse:0.378852+0.001398	train-auc:0.870728+0.001990	test-rmse:0.386041+0.008923	test-auc:0.870728+0.001990
## [4] train-rmse:0.372729+0.001637	train-auc:0.880756+0.002273	test-rmse:0.381464+0.009477	test-auc:0.880756+0.002273
## [5] train-rmse:0.368863+0.001861	train-auc:0.886316+0.002073	test-rmse:0.378730+0.009714	test-auc:0.886316+0.002073
## [6] train-rmse:0.366441+0.001607	train-auc:0.889350+0.001927	test-rmse:0.376971+0.010199	test-auc:0.889350+0.001927
## [7] train-rmse:0.364508+0.001586	train-auc:0.891987+0.001954	test-rmse:0.375765+0.010405	test-auc:0.891987+0.001954
## [8] train-rmse:0.361741+0.002500	train-auc:0.895200+0.002572	test-rmse:0.374605+0.011307	test-auc:0.895200+0.002572
## [9] train-rmse:0.359367+0.002674	train-auc:0.897881+0.002820	test-rmse:0.373190+0.010896	test-auc:0.897881+0.002820
## [10] train-rmse:0.357505+0.002598	train-auc:0.900123+0.002623	test-rmse:0.372752+0.010770	test-auc:0.900123+0.002623
## [11] train-rmse:0.356330+0.002516	train-auc:0.901504+0.002411	test-rmse:0.372415+0.010664	test-auc:0.901504+0.002411
## [12] train-rmse:0.354930+0.002560	train-auc:0.903052+0.002540	test-rmse:0.371868+0.010613	test-auc:0.903052+0.002540
## [13] train-rmse:0.353317+0.002560	train-auc:0.905031+0.002491	test-rmse:0.371075+0.011119	test-auc:0.905031+0.002491
## [14] train-rmse:0.352609+0.002423	train-auc:0.905840+0.002382	test-rmse:0.370518+0.011415	test-auc:0.905840+0.002382
## [15] train-rmse:0.351316+0.002204	train-auc:0.907134+0.002231	test-rmse:0.370852+0.011695	test-auc:0.907134+0.002231
## [16] train-rmse:0.350333+0.002146	train-auc:0.908162+0.002161	test-rmse:0.370726+0.011704	test-auc:0.908162+0.002161
## [17] train-rmse:0.349126+0.002166	train-auc:0.909447+0.002254	test-rmse:0.371293+0.011492	test-auc:0.909447+0.002254
## [18] train-rmse:0.348295+0.002153	train-auc:0.910399+0.002297	test-rmse:0.371874+0.011599	test-auc:0.910399+0.002297
## [19] train-rmse:0.347184+0.002109	train-auc:0.911515+0.002141	test-rmse:0.371745+0.011720	test-auc:0.911515+0.002141

## [20]	train-rmse:0.346005+0.001498	train-auc:0.912769+0.001518	test-rmse:0.371314+0.012274	test-auc:0.912769+0.001518
## [21]	train-rmse:0.345138+0.001769	train-auc:0.913644+0.001712	test-rmse:0.371627+0.012131	test-auc:0.913644+0.001712
## [22]	train-rmse:0.344329+0.001754	train-auc:0.914464+0.001689	test-rmse:0.371266+0.012413	test-auc:0.914464+0.001689
## [23]	train-rmse:0.343350+0.001915	train-auc:0.915459+0.001870	test-rmse:0.371731+0.013099	test-auc:0.915459+0.001870
## [24]	train-rmse:0.342475+0.001964	train-auc:0.916324+0.001901	test-rmse:0.371916+0.012878	test-auc:0.916324+0.001901
## [25]	train-rmse:0.341509+0.002072	train-auc:0.917216+0.002013	test-rmse:0.371866+0.013102	test-auc:0.917216+0.002013
## [26]	train-rmse:0.340689+0.001933	train-auc:0.918042+0.001888	test-rmse:0.372126+0.012656	test-auc:0.918042+0.001888
## [27]	train-rmse:0.339792+0.002083	train-auc:0.918946+0.002021	test-rmse:0.372931+0.012650	test-auc:0.918946+0.002021
## [28]	train-rmse:0.339044+0.002041	train-auc:0.919682+0.001946	test-rmse:0.373496+0.012743	test-auc:0.919682+0.001946
## [29]	train-rmse:0.338046+0.001954	train-auc:0.920647+0.001815	test-rmse:0.373787+0.012831	test-auc:0.920647+0.001815
## [30]	train-rmse:0.337128+0.001795	train-auc:0.921598+0.001656	test-rmse:0.373811+0.013434	test-auc:0.921598+0.001656
## [31]	train-rmse:0.336358+0.001755	train-auc:0.922325+0.001597	test-rmse:0.373794+0.013034	test-auc:0.922325+0.001597
## [32]	train-rmse:0.335645+0.001621	train-auc:0.922911+0.001497	test-rmse:0.373752+0.012883	test-auc:0.922911+0.001497
## [33]	train-rmse:0.334922+0.001529	train-auc:0.923596+0.001316	test-rmse:0.374523+0.012649	test-auc:0.923596+0.001316
## [34]	train-rmse:0.333932+0.001624	train-auc:0.924532+0.001408	test-rmse:0.374884+0.013187	test-auc:0.924532+0.001408
## [35]	train-rmse:0.332912+0.001554	train-auc:0.925443+0.001347	test-rmse:0.374749+0.013705	test-auc:0.925443+0.001347
## [36]	train-rmse:0.332262+0.001438	train-auc:0.926035+0.001194	test-rmse:0.375115+0.013390	test-auc:0.926035+0.001194
## [37]	train-rmse:0.331246+0.001364	train-auc:0.926959+0.001162	test-rmse:0.374810+0.013510	test-auc:0.926959+0.001162
## [38]	train-rmse:0.330345+0.001437	train-auc:0.927820+0.001221	test-rmse:0.375184+0.013597	test-auc:0.927820+0.001221
## [39]	train-rmse:0.329495+0.001612	train-auc:0.928612+0.001336	test-rmse:0.375546+0.014105	test-auc:0.928612+0.001336
## [40]	train-rmse:0.328807+0.001581	train-auc:0.929218+0.001320	test-rmse:0.375701+0.013685	test-auc:0.929218+0.001320
## [41]	train-rmse:0.327937+0.001694	train-auc:0.930023+0.001379	test-rmse:0.375881+0.013405	test-auc:0.930023+0.001379
## [42]	train-rmse:0.327184+0.001656	train-auc:0.930684+0.001377	test-rmse:0.375684+0.013582	test-auc:0.930684+0.001377
## [43]	train-rmse:0.326425+0.001521	train-auc:0.931303+0.001260	test-rmse:0.375522+0.013492	test-auc:0.931303+0.001260
## [44]	train-rmse:0.325603+0.001620	train-auc:0.931990+0.001330	test-rmse:0.375850+0.013920	test-auc:0.931990+0.001330
## [45]	train-rmse:0.324805+0.001692	train-auc:0.932658+0.001348	test-rmse:0.376160+0.013887	test-auc:0.932658+0.001348
## [46]	train-rmse:0.324063+0.001575	train-auc:0.933289+0.001216	test-rmse:0.376169+0.013544	test-auc:0.933289+0.001216
## [47]	train-rmse:0.323439+0.001762	train-auc:0.933793+0.001388	test-rmse:0.375983+0.013562	test-auc:0.933793+0.001388
## [48]	train-rmse:0.322854+0.001788	train-auc:0.934262+0.001377	test-rmse:0.376130+0.013322	test-auc:0.934262+0.001377
## [49]	train-rmse:0.322252+0.001745	train-auc:0.934807+0.001355	test-rmse:0.376373+0.013048	test-auc:0.934807+0.001355
## [50]	train-rmse:0.321600+0.001629	train-auc:0.935366+0.001298	test-rmse:0.376584+0.012697	test-auc:0.935366+0.001298
## [51]	train-rmse:0.320875+0.001747	train-auc:0.935970+0.001400	test-rmse:0.376835+0.012753	test-auc:0.935970+0.001400
## [52]	train-rmse:0.320162+0.001893	train-auc:0.936542+0.001521	test-rmse:0.377153+0.012650	test-auc:0.936542+0.001521
## [53]	train-rmse:0.319647+0.001986	train-auc:0.936932+0.001590	test-rmse:0.377232+0.012663	test-auc:0.936932+0.001590
## [54]	train-rmse:0.319180+0.002051	train-auc:0.937360+0.001665	test-rmse:0.377856+0.012505	test-auc:0.937360+0.001665
## [55]	train-rmse:0.318495+0.002023	train-auc:0.937930+0.001619	test-rmse:0.377686+0.012397	test-auc:0.937930+0.001619
## [56]	train-rmse:0.317847+0.001816	train-auc:0.938420+0.001408	test-rmse:0.378082+0.012285	test-auc:0.938420+0.001408
## [57]	train-rmse:0.317274+0.001728	train-auc:0.938898+0.001332	test-rmse:0.378395+0.012456	test-auc:0.938898+0.001332
## [58]	train-rmse:0.316655+0.001743	train-auc:0.939373+0.001312	test-rmse:0.378724+0.012390	test-auc:0.939373+0.001312
## [59]	train-rmse:0.316001+0.001710	train-auc:0.939908+0.001291	test-rmse:0.379343+0.012630	test-auc:0.939908+0.001291
## [60]	train-rmse:0.315394+0.001733	train-auc:0.940395+0.001311	test-rmse:0.379405+0.012469	test-auc:0.940395+0.001311
## [61]	train-rmse:0.314790+0.001697	train-auc:0.940824+0.001289	test-rmse:0.379191+0.012942	test-auc:0.940824+0.001289
## [62]	train-rmse:0.314159+0.001587	train-auc:0.941314+0.001225	test-rmse:0.379223+0.012687	test-auc:0.941314+0.001225
## [63]	train-rmse:0.313643+0.001755	train-auc:0.941697+0.001368	test-rmse:0.378866+0.012774	test-auc:0.941697+0.001368
## [64]	train-rmse:0.313051+0.001694	train-auc:0.942164+0.001256	test-rmse:0.378491+0.012298	test-auc:0.942164+0.001256
## [65]	train-rmse:0.312558+0.001828	train-auc:0.942522+0.001354	test-rmse:0.378371+0.012428	test-auc:0.942522+0.001354
## [66]	train-rmse:0.312013+0.001876	train-auc:0.942916+0.001425	test-rmse:0.378508+0.012472	test-auc:0.942916+0.001425
## [67]	train-rmse:0.311434+0.001698	train-auc:0.943358+0.001266	test-rmse:0.378844+0.012370	test-auc:0.943358+0.001266
## [68]	train-rmse:0.310875+0.001640	train-auc:0.943777+0.001186	test-rmse:0.378949+0.012547	test-auc:0.943777+0.001186
## [69]	train-rmse:0.310278+0.001617	train-auc:0.944223+0.001195	test-rmse:0.379260+0.012677	test-auc:0.944223+0.001195
## [70]	train-rmse:0.309696+0.001594	train-auc:0.944673+0.001148	test-rmse:0.379544+0.012526	test-auc:0.944673+0.001148
## [71]	train-rmse:0.309181+0.001532	train-auc:0.945069+0.001093	test-rmse:0.379735+0.012658	test-auc:0.945069+0.001093
## [72]	train-rmse:0.308746+0.001576	train-auc:0.945353+0.001151	test-rmse:0.379650+0.012406	test-auc:0.945353+0.001151
## [73]	train-rmse:0.308254+0.001624	train-auc:0.945715+0.001192	test-rmse:0.379444+0.012359	test-auc:0.945715+0.001192


```

## [74] train-rmse:0.307611+0.001518 train-auc:0.946229+0.001126 test-rmse:0.379600+0.012350 test-auc:0.946229+0.001126
## [75] train-rmse:0.307122+0.001547 train-auc:0.946569+0.001098 test-rmse:0.380029+0.012011 test-auc:0.946569+0.001098
## [76] train-rmse:0.306571+0.001594 train-auc:0.946961+0.001127 test-rmse:0.380702+0.012300 test-auc:0.946961+0.001127
## [77] train-rmse:0.305946+0.001564 train-auc:0.947416+0.001071 test-rmse:0.380684+0.012528 test-auc:0.947416+0.001071
## [78] train-rmse:0.305473+0.001577 train-auc:0.947754+0.001094 test-rmse:0.380686+0.012480 test-auc:0.947754+0.001094
## [79] train-rmse:0.305042+0.001672 train-auc:0.948028+0.001165 test-rmse:0.380843+0.012558 test-auc:0.948028+0.001165
## [80] train-rmse:0.304559+0.001754 train-auc:0.948360+0.001245 test-rmse:0.380739+0.012845 test-auc:0.948360+0.001245
## [81] train-rmse:0.304144+0.001801 train-auc:0.948627+0.001221 test-rmse:0.381011+0.013132 test-auc:0.948627+0.001221
## [82] train-rmse:0.303739+0.001734 train-auc:0.948914+0.001159 test-rmse:0.381287+0.012875 test-auc:0.948914+0.001159
## [83] train-rmse:0.303312+0.001882 train-auc:0.949232+0.001253 test-rmse:0.381426+0.013058 test-auc:0.949232+0.001253
## [84] train-rmse:0.302778+0.001926 train-auc:0.949588+0.001289 test-rmse:0.381812+0.012985 test-auc:0.949588+0.001289
## [85] train-rmse:0.302109+0.001988 train-auc:0.950044+0.001358 test-rmse:0.381994+0.012993 test-auc:0.950044+0.001358
## [86] train-rmse:0.301626+0.002012 train-auc:0.950331+0.001356 test-rmse:0.382231+0.012890 test-auc:0.950331+0.001356
## [87] train-rmse:0.300992+0.001877 train-auc:0.950830+0.001264 test-rmse:0.383184+0.013100 test-auc:0.950830+0.001264
## [88] train-rmse:0.300468+0.001871 train-auc:0.951155+0.001258 test-rmse:0.383213+0.013012 test-auc:0.951155+0.001258
## [89] train-rmse:0.299921+0.001991 train-auc:0.951500+0.001340 test-rmse:0.383235+0.012890 test-auc:0.951500+0.001340
## [90] train-rmse:0.299382+0.002110 train-auc:0.951856+0.001388 test-rmse:0.383879+0.012765 test-auc:0.951856+0.001388
## [91] train-rmse:0.298813+0.002120 train-auc:0.952226+0.001396 test-rmse:0.383953+0.012664 test-auc:0.952226+0.001396
## [92] train-rmse:0.298246+0.002226 train-auc:0.952622+0.001474 test-rmse:0.383782+0.012788 test-auc:0.952622+0.001474
## [93] train-rmse:0.297793+0.002160 train-auc:0.952924+0.001428 test-rmse:0.383734+0.013002 test-auc:0.952924+0.001428
## [94] train-rmse:0.297216+0.002172 train-auc:0.953288+0.001439 test-rmse:0.383941+0.012873 test-auc:0.953288+0.001439
## [95] train-rmse:0.296801+0.002170 train-auc:0.953534+0.001385 test-rmse:0.384043+0.012999 test-auc:0.953534+0.001385
## [96] train-rmse:0.296367+0.002323 train-auc:0.953836+0.001481 test-rmse:0.384360+0.012980 test-auc:0.953836+0.001481
## [97] train-rmse:0.295864+0.002369 train-auc:0.954187+0.001540 test-rmse:0.384128+0.012834 test-auc:0.954187+0.001540
## [98] train-rmse:0.295462+0.002359 train-auc:0.954423+0.001510 test-rmse:0.384236+0.012831 test-auc:0.954423+0.001510
## [99] train-rmse:0.295035+0.002460 train-auc:0.954663+0.001593 test-rmse:0.384399+0.012750 test-auc:0.954663+0.001593
## [100] train-rmse:0.294489+0.002418 train-auc:0.955032+0.001500 test-rmse:0.384445+0.012599 test-auc:0.955032+0.001500

```

xgb.cv 10-folds

```

## iter train_rmse_mean train_rmse_std train_auc_mean train_auc_std
## 1 0.4150472 0.0008407872 0.8130229 0.001987129
## 2 0.3858453 0.0011852734 0.8605165 0.002116012
## 3 0.3788522 0.0013976987 0.8707282 0.001989507
## 4 0.3727295 0.0016374695 0.8807559 0.002273087
## 5 0.3688626 0.0018613585 0.8863161 0.002073216
## 6 0.3664408 0.0016074771 0.8893501 0.001927122
## 7 0.3645081 0.0015860282 0.8919871 0.001953559
## 8 0.3617411 0.0025001798 0.8952004 0.002572208
## 9 0.3593669 0.0026737110 0.8978812 0.002819932
## 10 0.3575046 0.0025977840 0.9001226 0.002623137
## 11 0.3563304 0.0025156403 0.9015038 0.002411376
## 12 0.3549304 0.0025602207 0.9030519 0.002540477
## 13 0.3533172 0.0025595232 0.9050305 0.002490587
## 14 0.3526090 0.0024227740 0.9058402 0.002381847
## 15 0.3513156 0.0022040046 0.9071336 0.002231236
## 16 0.3503330 0.0021458483 0.9081623 0.002161142
## 17 0.3491261 0.0021664241 0.9094473 0.002254360
## 18 0.3482954 0.0021525416 0.9103991 0.002296694
## 19 0.3471842 0.0021090532 0.9115153 0.002141252
## 20 0.3460052 0.0014983255 0.9127690 0.001518343
## 21 0.3451379 0.0017689474 0.9136437 0.001711919
## 22 0.3443292 0.0017538449 0.9144641 0.001688819
## 23 0.3433504 0.0019147572 0.9154587 0.001869824
## 24 0.3424754 0.0019642783 0.9163236 0.001900750

```

##	25	0.3415089	0.0020722321	0.9172161	0.002012797
##	26	0.3406887	0.0019334262	0.9180421	0.001887678
##	27	0.3397922	0.0020830397	0.9189460	0.002021011
##	28	0.3390438	0.0020406827	0.9196823	0.001946301
##	29	0.3380463	0.0019536782	0.9206469	0.001815359
##	30	0.3371280	0.0017953080	0.9215984	0.001656312
##	31	0.3363576	0.0017554653	0.9223249	0.001597113
##	32	0.3356455	0.0016210755	0.9229110	0.001497178
##	33	0.3349221	0.0015292546	0.9235958	0.001315839
##	34	0.3339316	0.0016236697	0.9245321	0.001407679
##	35	0.3329116	0.0015539443	0.9254432	0.001346647
##	36	0.3322624	0.0014381169	0.9260354	0.001193753
##	37	0.3312456	0.0013643091	0.9269594	0.001161588
##	38	0.3303449	0.0014368349	0.9278201	0.001220801
##	39	0.3294946	0.0016117879	0.9286122	0.001335569
##	40	0.3288075	0.0015811839	0.9292182	0.001319932
##	41	0.3279367	0.0016937697	0.9300226	0.001378730
##	42	0.3271845	0.0016557488	0.9306839	0.001376820
##	43	0.3264252	0.0015210003	0.9313029	0.001260430
##	44	0.3256029	0.0016199637	0.9319899	0.001330492
##	45	0.3248051	0.0016917590	0.9326583	0.001347714
##	46	0.3240628	0.0015749109	0.9332886	0.001216220
##	47	0.3234387	0.0017624129	0.9337930	0.001388292
##	48	0.3228536	0.0017878012	0.9342619	0.001376842
##	49	0.3222521	0.0017451265	0.9348072	0.001354840
##	50	0.3216003	0.0016293940	0.9353660	0.001298031
##	51	0.3208746	0.0017466966	0.9359704	0.001399812
##	52	0.3201616	0.0018929619	0.9365422	0.001520536
##	53	0.3196474	0.0019863154	0.9369315	0.001590095
##	54	0.3191798	0.0020510904	0.9373596	0.001664705
##	55	0.3184953	0.0020233171	0.9379301	0.001619470
##	56	0.3178470	0.0018159202	0.9384201	0.001408072
##	57	0.3172745	0.0017284655	0.9388979	0.001331981
##	58	0.3166554	0.0017430991	0.9393732	0.001312450
##	59	0.3160015	0.0017098937	0.9399080	0.001291092
##	60	0.3153941	0.0017327089	0.9403952	0.001311240
##	61	0.3147899	0.0016971185	0.9408238	0.001289445
##	62	0.3141589	0.0015871036	0.9413142	0.001224871
##	63	0.3136426	0.0017545167	0.9416971	0.001367983
##	64	0.3130513	0.0016935131	0.9421638	0.001256322
##	65	0.3125581	0.0018283325	0.9425217	0.001354058
##	66	0.3120128	0.0018758543	0.9429164	0.001425453
##	67	0.3114345	0.0016978604	0.9433579	0.001266203
##	68	0.3108745	0.0016403151	0.9437774	0.001186087
##	69	0.3102778	0.0016174699	0.9442233	0.001194706
##	70	0.3096960	0.0015938708	0.9446733	0.001147959
##	71	0.3091811	0.0015323005	0.9450690	0.001093086
##	72	0.3087464	0.0015762548	0.9453532	0.001151400
##	73	0.3082539	0.0016238172	0.9457153	0.001191572
##	74	0.3076112	0.0015179670	0.9462288	0.001126127
##	75	0.3071218	0.0015471330	0.9465694	0.001098388
##	76	0.3065711	0.0015937457	0.9469606	0.001127085
##	77	0.3059460	0.0015644019	0.9474155	0.001071474
##	78	0.3054734	0.0015765148	0.9477536	0.001094148

##	79	0.3050415	0.0016721615	0.9480278	0.001164874
##	80	0.3045590	0.0017535599	0.9483599	0.001245139
##	81	0.3041436	0.0018011682	0.9486268	0.001220946
##	82	0.3037391	0.0017341519	0.9489142	0.001158834
##	83	0.3033118	0.0018817398	0.9492318	0.001253110
##	84	0.3027782	0.0019258601	0.9495880	0.001288597
##	85	0.3021095	0.0019879170	0.9500436	0.001358454
##	86	0.3016264	0.0020116503	0.9503314	0.001355558
##	87	0.3009923	0.0018768301	0.9508298	0.001264026
##	88	0.3004677	0.0018706476	0.9511545	0.001258138
##	89	0.2999206	0.0019907126	0.9515000	0.001339754
##	90	0.2993822	0.0021101515	0.9518561	0.001388026
##	91	0.2988134	0.0021202069	0.9522256	0.001395612
##	92	0.2982457	0.0022264526	0.9526222	0.001474256
##	93	0.2977932	0.0021597408	0.9529243	0.001427907
##	94	0.2972163	0.0021721367	0.9532876	0.001439034
##	95	0.2968009	0.0021696078	0.9535342	0.001384682
##	96	0.2963672	0.0023232846	0.9538355	0.001481315
##	97	0.2958643	0.0023686242	0.9541874	0.001539564
##	98	0.2954618	0.0023590862	0.9544232	0.001509885
##	99	0.2950347	0.0024596957	0.9546631	0.001592549
##	100	0.2944892	0.0024184142	0.9550320	0.001500150
##	iter train_rmse_mean train_rmse_std train_auc_mean train_auc_std				
##	test_rmse_mean test_rmse_std test_auc_mean test_auc_std				
##	0.4163444	0.005940971	0.8103009	0.01108728	
##	0.3905971	0.007574149	0.8523477	0.01083094	
##	0.3860413	0.008923041	0.8609303	0.01309413	
##	0.3814645	0.009476891	0.8682670	0.01300392	
##	0.3787298	0.009713631	0.8723888	0.01294677	
##	0.3769712	0.010198804	0.8756100	0.01269970	
##	0.3757651	0.010404697	0.8769801	0.01277184	
##	0.3746049	0.011306811	0.8789893	0.01406098	
##	0.3731903	0.010896145	0.8807695	0.01321164	
##	0.3727517	0.010769920	0.8816784	0.01274347	
##	0.3724146	0.010664375	0.8823298	0.01248041	
##	0.3718681	0.010613391	0.8831247	0.01235360	
##	0.3710748	0.011118640	0.8842339	0.01307136	
##	0.3705184	0.011414561	0.8849245	0.01331109	
##	0.3708519	0.011694643	0.8841869	0.01342856	
##	0.3707264	0.011704054	0.8845269	0.01362023	
##	0.3712928	0.011491559	0.8838648	0.01332994	
##	0.3718740	0.011599427	0.8833306	0.01341941	
##	0.3717453	0.011720242	0.8834212	0.01330782	
##	0.3713145	0.012274063	0.8843503	0.01402412	
##	0.3716268	0.012131086	0.8838853	0.01388941	
##	0.3712662	0.012412699	0.8844709	0.01416349	
##	0.3717313	0.013099009	0.8839264	0.01474153	
##	0.3719164	0.012877542	0.8839594	0.01453009	
##	0.3718655	0.013102286	0.8840172	0.01482702	
##	0.3721261	0.012656256	0.8836352	0.01439909	
##	0.3729312	0.012649515	0.8829951	0.01430383	
##	0.3734957	0.012742949	0.8824653	0.01434710	
##	0.3737875	0.012831275	0.8821490	0.01457266	
##	0.3738110	0.013433636	0.8820928	0.01534312	

##	0.3737938	0.013033989	0.8822264	0.01483236
##	0.3737518	0.012882601	0.8821574	0.01454883
##	0.3745227	0.012648728	0.8813211	0.01419541
##	0.3748835	0.013186861	0.8810441	0.01491990
##	0.3747488	0.013704860	0.8812463	0.01542082
##	0.3751152	0.013389884	0.8811790	0.01514652
##	0.3748097	0.013509896	0.8814768	0.01519824
##	0.3751836	0.013597112	0.8809343	0.01535407
##	0.3755461	0.014105492	0.8804968	0.01564721
##	0.3757008	0.013685368	0.8804653	0.01510704
##	0.3758811	0.013404811	0.8801763	0.01473146
##	0.3756843	0.013582188	0.8803998	0.01490411
##	0.3755217	0.013491695	0.8807061	0.01467348
##	0.3758504	0.013920480	0.8806825	0.01507226
##	0.3761600	0.013886949	0.8804458	0.01502624
##	0.3761690	0.013544243	0.8805973	0.01477085
##	0.3759833	0.013561906	0.8808755	0.01478820
##	0.3761295	0.013322055	0.8807837	0.01463435
##	0.3763730	0.013047615	0.8805153	0.01456083
##	0.3765837	0.012697463	0.8802232	0.01438638
##	0.3768346	0.012753244	0.8800933	0.01436449
##	0.3771535	0.012650149	0.8797425	0.01412330
##	0.3772315	0.012662881	0.8796460	0.01413643
##	0.3778558	0.012504540	0.8790490	0.01408184
##	0.3776862	0.012396650	0.8791791	0.01404165
##	0.3780824	0.012285204	0.8788216	0.01387336
##	0.3783945	0.012455634	0.8784282	0.01426612
##	0.3787235	0.012390253	0.8783465	0.01405308
##	0.3793426	0.012629892	0.8775908	0.01445949
##	0.3794049	0.012468829	0.8776270	0.01435124
##	0.3791911	0.012942039	0.8779519	0.01472845
##	0.3792233	0.012687439	0.8778298	0.01449095
##	0.3788662	0.012774251	0.8783715	0.01473005
##	0.3784912	0.012298131	0.8788565	0.01396577
##	0.3783710	0.012427593	0.8790959	0.01437535
##	0.3785079	0.012471644	0.8789987	0.01418317
##	0.3788439	0.012370071	0.8786583	0.01413923
##	0.3789487	0.012547271	0.8785655	0.01430384
##	0.3792602	0.012676776	0.8782160	0.01456424
##	0.3795440	0.012525504	0.8778857	0.01457869
##	0.3797347	0.012657868	0.8776412	0.01484526
##	0.3796498	0.012405659	0.8777717	0.01448711
##	0.3794444	0.012358788	0.8779069	0.01428116
##	0.3796003	0.012349827	0.8778526	0.01435274
##	0.3800292	0.012010576	0.8774506	0.01404611
##	0.3807024	0.012300158	0.8767426	0.01462188
##	0.3806841	0.012528177	0.8767919	0.01483964
##	0.3806865	0.012479562	0.8767785	0.01477349
##	0.3808431	0.012557640	0.8766432	0.01495237
##	0.3807394	0.012845058	0.8767672	0.01530188
##	0.3810106	0.013132345	0.8765135	0.01549512
##	0.3812867	0.012874867	0.8763985	0.01525579
##	0.3814256	0.013057811	0.8762678	0.01553935
##	0.3818120	0.012985211	0.8759474	0.01547304

```

##      0.3819941    0.012992894    0.8758744    0.01544001
##      0.3822309    0.012889672    0.8759527    0.01527018
##      0.3831844    0.013099716    0.8749906    0.01538256
##      0.3832133    0.013012175    0.8751051    0.01529104
##      0.3832353    0.012890424    0.8750778    0.01501692
##      0.3838789    0.012765165    0.8746764    0.01476730
##      0.3839534    0.012664215    0.8746476    0.01482685
##      0.3837816    0.012788417    0.8748136    0.01505861
##      0.3837339    0.013001662    0.8748475    0.01525643
##      0.3839414    0.012872987    0.8745956    0.01495951
##      0.3840430    0.012998917    0.8745925    0.01518626
##      0.3843602    0.012980077    0.8742902    0.01510705
##      0.3841278    0.012834125    0.8746246    0.01493328
##      0.3842362    0.012830571    0.8746688    0.01479555
##      0.3843985    0.012750189    0.8746084    0.01490047
##      0.3844450    0.012598636    0.8744459    0.01492134
## test_rmse_mean test_rmse_std test_auc_mean test_auc_std

## ##### xgb.cv 10-folds
## call:
##   xgb.cv(data = xgb_train, nrounds = 100, nfold = 10, metrics = list("rmse",
##   "auc"), nthread = 2, max_depth = 3, eta = 1, objective = "binary:logistic")
## params (as set within xgb.cv):
##   nthread = "2", max_depth = "3", eta = "1", objective = "binary:logistic", eval_metric = "rmse", ev
## callbacks:
##   cb.print.evaluation(period = print_every_n, showsd = showsd)
##   cb.evaluation.log()
## niter: 100
## evaluation_log:
##   iter train_rmse_mean train_rmse_std train_auc_mean train_auc_std
##     1      0.4150472    0.0008407872    0.8130229    0.001987129
##     2      0.3858453    0.0011852734    0.8605165    0.002116012
##     3      0.3788522    0.0013976987    0.8707282    0.001989507
##     4      0.3727295    0.0016374695    0.8807559    0.002273087
##     5      0.3688626    0.0018613585    0.8863161    0.002073216
##     6      0.3664408    0.0016074771    0.8893501    0.001927122
##     7      0.3645081    0.0015860282    0.8919871    0.001953559
##     8      0.3617411    0.0025001798    0.8952004    0.002572208
##     9      0.3593669    0.0026737110    0.8978812    0.002819932
##    10      0.3575046    0.0025977840    0.9001226    0.002623137
##    11      0.3563304    0.0025156403    0.9015038    0.002411376
##    12      0.3549304    0.0025602207    0.9030519    0.002540477
##    13      0.3533172    0.0025595232    0.9050305    0.002490587
##    14      0.3526090    0.0024227740    0.9058402    0.002381847
##    15      0.3513156    0.0022040046    0.9071336    0.002231236
##    16      0.3503330    0.0021458483    0.9081623    0.002161142
##    17      0.3491261    0.0021664241    0.9094473    0.002254360
##    18      0.3482954    0.0021525416    0.9103991    0.002296694
##    19      0.3471842    0.0021090532    0.9115153    0.002141252
##    20      0.3460052    0.0014983255    0.9127690    0.001518343
##    21      0.3451379    0.0017689474    0.9136437    0.001711919
##    22      0.3443292    0.0017538449    0.9144641    0.001688819
##    23      0.3433504    0.0019147572    0.9154587    0.001869824
##    24      0.3424754    0.0019642783    0.9163236    0.001900750

```

##	25	0.3415089	0.0020722321	0.9172161	0.002012797
##	26	0.3406887	0.0019334262	0.9180421	0.001887678
##	27	0.3397922	0.0020830397	0.9189460	0.002021011
##	28	0.3390438	0.0020406827	0.9196823	0.001946301
##	29	0.3380463	0.0019536782	0.9206469	0.001815359
##	30	0.3371280	0.0017953080	0.9215984	0.001656312
##	31	0.3363576	0.0017554653	0.9223249	0.001597113
##	32	0.3356455	0.0016210755	0.9229110	0.001497178
##	33	0.3349221	0.0015292546	0.9235958	0.001315839
##	34	0.3339316	0.0016236697	0.9245321	0.001407679
##	35	0.3329116	0.0015539443	0.9254432	0.001346647
##	36	0.3322624	0.0014381169	0.9260354	0.001193753
##	37	0.3312456	0.0013643091	0.9269594	0.001161588
##	38	0.3303449	0.0014368349	0.9278201	0.001220801
##	39	0.3294946	0.0016117879	0.9286122	0.001335569
##	40	0.3288075	0.0015811839	0.9292182	0.001319932
##	41	0.3279367	0.0016937697	0.9300226	0.001378730
##	42	0.3271845	0.0016557488	0.9306839	0.001376820
##	43	0.3264252	0.0015210003	0.9313029	0.001260430
##	44	0.3256029	0.0016199637	0.9319899	0.001330492
##	45	0.3248051	0.0016917590	0.9326583	0.001347714
##	46	0.3240628	0.0015749109	0.9332886	0.001216220
##	47	0.3234387	0.0017624129	0.9337930	0.001388292
##	48	0.3228536	0.0017878012	0.9342619	0.001376842
##	49	0.3222521	0.0017451265	0.9348072	0.001354840
##	50	0.3216003	0.0016293940	0.9353660	0.001298031
##	51	0.3208746	0.0017466966	0.9359704	0.001399812
##	52	0.3201616	0.0018929619	0.9365422	0.001520536
##	53	0.3196474	0.0019863154	0.9369315	0.001590095
##	54	0.3191798	0.0020510904	0.9373596	0.001664705
##	55	0.3184953	0.0020233171	0.9379301	0.001619470
##	56	0.3178470	0.0018159202	0.9384201	0.001408072
##	57	0.3172745	0.0017284655	0.9388979	0.001331981
##	58	0.3166554	0.0017430991	0.9393732	0.001312450
##	59	0.3160015	0.0017098937	0.9399080	0.001291092
##	60	0.3153941	0.0017327089	0.9403952	0.001311240
##	61	0.3147899	0.0016971185	0.9408238	0.001289445
##	62	0.3141589	0.0015871036	0.9413142	0.001224871
##	63	0.3136426	0.0017545167	0.9416971	0.001367983
##	64	0.3130513	0.0016935131	0.9421638	0.001256322
##	65	0.3125581	0.0018283325	0.9425217	0.001354058
##	66	0.3120128	0.0018758543	0.9429164	0.001425453
##	67	0.3114345	0.0016978604	0.9433579	0.001266203
##	68	0.3108745	0.0016403151	0.9437774	0.001186087
##	69	0.3102778	0.0016174699	0.9442233	0.001194706
##	70	0.3096960	0.0015938708	0.9446733	0.001147959
##	71	0.3091811	0.0015323005	0.9450690	0.001093086
##	72	0.3087464	0.0015762548	0.9453532	0.001151400
##	73	0.3082539	0.0016238172	0.9457153	0.001191572
##	74	0.3076112	0.0015179670	0.9462288	0.001126127
##	75	0.3071218	0.0015471330	0.9465694	0.001098388
##	76	0.3065711	0.0015937457	0.9469606	0.001127085
##	77	0.3059460	0.0015644019	0.9474155	0.001071474
##	78	0.3054734	0.0015765148	0.9477536	0.001094148

##	79	0.3050415	0.0016721615	0.9480278	0.001164874
##	80	0.3045590	0.0017535599	0.9483599	0.001245139
##	81	0.3041436	0.0018011682	0.9486268	0.001220946
##	82	0.3037391	0.0017341519	0.9489142	0.001158834
##	83	0.3033118	0.0018817398	0.9492318	0.001253110
##	84	0.3027782	0.0019258601	0.9495880	0.001288597
##	85	0.3021095	0.0019879170	0.9500436	0.001358454
##	86	0.3016264	0.0020116503	0.9503314	0.001355558
##	87	0.3009923	0.0018768301	0.9508298	0.001264026
##	88	0.3004677	0.0018706476	0.9511545	0.001258138
##	89	0.2999206	0.0019907126	0.9515000	0.001339754
##	90	0.2993822	0.0021101515	0.9518561	0.001388026
##	91	0.2988134	0.0021202069	0.9522256	0.001395612
##	92	0.2982457	0.0022264526	0.9526222	0.001474256
##	93	0.2977932	0.0021597408	0.9529243	0.001427907
##	94	0.2972163	0.0021721367	0.9532876	0.001439034
##	95	0.2968009	0.0021696078	0.9535342	0.001384682
##	96	0.2963672	0.0023232846	0.9538355	0.001481315
##	97	0.2958643	0.0023686242	0.9541874	0.001539564
##	98	0.2954618	0.0023590862	0.9544232	0.001509885
##	99	0.2950347	0.0024596957	0.9546631	0.001592549
##	100	0.2944892	0.0024184142	0.9550320	0.001500150
##	iter train_rmse_mean train_rmse_std train_auc_mean train_auc_std				
##	test_rmse_mean test_rmse_std test_auc_mean test_auc_std				
##	0.4163444	0.005940971	0.8103009	0.01108728	
##	0.3905971	0.007574149	0.8523477	0.01083094	
##	0.3860413	0.008923041	0.8609303	0.01309413	
##	0.3814645	0.009476891	0.8682670	0.01300392	
##	0.3787298	0.009713631	0.8723888	0.01294677	
##	0.3769712	0.010198804	0.8756100	0.01269970	
##	0.3757651	0.010404697	0.8769801	0.01277184	
##	0.3746049	0.011306811	0.8789893	0.01406098	
##	0.3731903	0.010896145	0.8807695	0.01321164	
##	0.3727517	0.010769920	0.8816784	0.01274347	
##	0.3724146	0.010664375	0.8823298	0.01248041	
##	0.3718681	0.010613391	0.8831247	0.01235360	
##	0.3710748	0.011118640	0.8842339	0.01307136	
##	0.3705184	0.011414561	0.8849245	0.01331109	
##	0.3708519	0.011694643	0.8841869	0.01342856	
##	0.3707264	0.011704054	0.8845269	0.01362023	
##	0.3712928	0.011491559	0.8838648	0.01332994	
##	0.3718740	0.011599427	0.8833306	0.01341941	
##	0.3717453	0.011720242	0.8834212	0.01330782	
##	0.3713145	0.012274063	0.8843503	0.01402412	
##	0.3716268	0.012131086	0.8838853	0.01388941	
##	0.3712662	0.012412699	0.8844709	0.01416349	
##	0.3717313	0.013099009	0.8839264	0.01474153	
##	0.3719164	0.012877542	0.8839594	0.01453009	
##	0.3718655	0.013102286	0.8840172	0.01482702	
##	0.3721261	0.012656256	0.8836352	0.01439909	
##	0.3729312	0.012649515	0.8829951	0.01430383	
##	0.3734957	0.012742949	0.8824653	0.01434710	
##	0.3737875	0.012831275	0.8821490	0.01457266	
##	0.3738110	0.013433636	0.8820928	0.01534312	

##	0.3737938	0.013033989	0.8822264	0.01483236
##	0.3737518	0.012882601	0.8821574	0.01454883
##	0.3745227	0.012648728	0.8813211	0.01419541
##	0.3748835	0.013186861	0.8810441	0.01491990
##	0.3747488	0.013704860	0.8812463	0.01542082
##	0.3751152	0.013389884	0.8811790	0.01514652
##	0.3748097	0.013509896	0.8814768	0.01519824
##	0.3751836	0.013597112	0.8809343	0.01535407
##	0.3755461	0.014105492	0.8804968	0.01564721
##	0.3757008	0.013685368	0.8804653	0.01510704
##	0.3758811	0.013404811	0.8801763	0.01473146
##	0.3756843	0.013582188	0.8803998	0.01490411
##	0.3755217	0.013491695	0.8807061	0.01467348
##	0.3758504	0.013920480	0.8806825	0.01507226
##	0.3761600	0.013886949	0.8804458	0.01502624
##	0.3761690	0.013544243	0.8805973	0.01477085
##	0.3759833	0.013561906	0.8808755	0.01478820
##	0.3761295	0.013322055	0.8807837	0.01463435
##	0.3763730	0.013047615	0.8805153	0.01456083
##	0.3765837	0.012697463	0.8802232	0.01438638
##	0.3768346	0.012753244	0.8800933	0.01436449
##	0.3771535	0.012650149	0.8797425	0.01412330
##	0.3772315	0.012662881	0.8796460	0.01413643
##	0.3778558	0.012504540	0.8790490	0.01408184
##	0.3776862	0.012396650	0.8791791	0.01404165
##	0.3780824	0.012285204	0.8788216	0.01387336
##	0.3783945	0.012455634	0.8784282	0.01426612
##	0.3787235	0.012390253	0.8783465	0.01405308
##	0.3793426	0.012629892	0.8775908	0.01445949
##	0.3794049	0.012468829	0.8776270	0.01435124
##	0.3791911	0.012942039	0.8779519	0.01472845
##	0.3792233	0.012687439	0.8778298	0.01449095
##	0.3788662	0.012774251	0.8783715	0.01473005
##	0.3784912	0.012298131	0.8788565	0.01396577
##	0.3783710	0.012427593	0.8790959	0.01437535
##	0.3785079	0.012471644	0.8789987	0.01418317
##	0.3788439	0.012370071	0.8786583	0.01413923
##	0.3789487	0.012547271	0.8785655	0.01430384
##	0.3792602	0.012676776	0.8782160	0.01456424
##	0.3795440	0.012525504	0.8778857	0.01457869
##	0.3797347	0.012657868	0.8776412	0.01484526
##	0.3796498	0.012405659	0.8777717	0.01448711
##	0.3794444	0.012358788	0.8779069	0.01428116
##	0.3796003	0.012349827	0.8778526	0.01435274
##	0.3800292	0.012010576	0.8774506	0.01404611
##	0.3807024	0.012300158	0.8767426	0.01462188
##	0.3806841	0.012528177	0.8767919	0.01483964
##	0.3806865	0.012479562	0.8767785	0.01477349
##	0.3808431	0.012557640	0.8766432	0.01495237
##	0.3807394	0.012845058	0.8767672	0.01530188
##	0.3810106	0.013132345	0.8765135	0.01549512
##	0.3812867	0.012874867	0.8763985	0.01525579
##	0.3814256	0.013057811	0.8762678	0.01553935
##	0.3818120	0.012985211	0.8759474	0.01547304


```
##      0.3819941    0.012992894    0.8758744    0.01544001
##      0.3822309    0.012889672    0.8759527    0.01527018
##      0.3831844    0.013099716    0.8749906    0.01538256
##      0.3832133    0.013012175    0.8751051    0.01529104
##      0.3832353    0.012890424    0.8750778    0.01501692
##      0.3838789    0.012765165    0.8746764    0.01476730
##      0.3839534    0.012664215    0.8746476    0.01482685
##      0.3837816    0.012788417    0.8748136    0.01505861
##      0.3837339    0.013001662    0.8748475    0.01525643
##      0.3839414    0.012872987    0.8745956    0.01495951
##      0.3840430    0.012998917    0.8745925    0.01518626
##      0.3843602    0.012980077    0.8742902    0.01510705
##      0.3841278    0.012834125    0.8746246    0.01493328
##      0.3842362    0.012830571    0.8746688    0.01479555
##      0.3843985    0.012750189    0.8746084    0.01490047
##      0.3844450    0.012598636    0.8744459    0.01492134
## test_rmse_mean test_rmse_std test_auc_mean test_auc_std
```

```
## NULL
```

```
## NULL
```

```
## [1] train-logloss:0.518785 eval-logloss:0.820399
## [2] train-logloss:0.460336 eval-logloss:0.881454
## [3] train-logloss:0.443383 eval-logloss:0.952815
## [4] train-logloss:0.429115 eval-logloss:1.006953
## [5] train-logloss:0.421462 eval-logloss:1.068660
## [6] train-logloss:0.417349 eval-logloss:1.082765
## [7] train-logloss:0.404430 eval-logloss:1.115529
## [8] train-logloss:0.401043 eval-logloss:1.130244
## [9] train-logloss:0.396576 eval-logloss:1.161850
## [10] train-logloss:0.393187 eval-logloss:1.185680
## [11] train-logloss:0.390200 eval-logloss:1.205786
## [12] train-logloss:0.388254 eval-logloss:1.226213
## [13] train-logloss:0.386829 eval-logloss:1.243061
## [14] train-logloss:0.383553 eval-logloss:1.241412
```

```
## [1] 0.850011170 0.199663937 0.003278726 0.004425377 0.187479198 0.122859113
```

Neural network

```
##      Age      RoomService      FoodCourt      ShoppingMall
## Min.   : 0.00   Min.   : 0.0   Min.   : 0   Min.   : 0.0
## 1st Qu.:20.00   1st Qu.: 0.0   1st Qu.: 0   1st Qu.: 0.0
## Median :27.00   Median : 0.0   Median : 0   Median : 0.0
## Mean   :28.89   Mean   : 243.3   Mean   : 476   Mean   : 169.2
## 3rd Qu.:38.00   3rd Qu.: 74.0   3rd Qu.: 79   3rd Qu.: 30.0
## Max.   :78.00   Max.   :14327.0   Max.   :29813   Max.   :9058.0
##      Spa      VRDeck      Transported      count
## Min.   : 0.0   Min.   : 0.0   Min.   :0.0000   Min.   :1.000
## 1st Qu.: 0.0   1st Qu.: 0.0   1st Qu.:0.0000   1st Qu.:1.000
## Median : 0.0   Median : 0.0   Median :1.0000   Median :1.000
```

```

## Mean      : 307.8      Mean      : 308.4      Mean      :0.5037      Mean      :2.019
## 3rd Qu.:   68.0      3rd Qu.:   49.0      3rd Qu.:1.0000      3rd Qu.:3.000
## Max.      :16139.0    Max.      :20336.0    Max.      :1.0000      Max.      :8.000
## HomePlanet_      HomePlanet_Europa      HomePlanet_Mars      CryoSleep_
## Min.      :0.00000      Min.      :0.0000      Min.      :0.0000      Min.      :0.00000
## 1st Qu.:0.00000      1st Qu.:0.0000      1st Qu.:0.0000      1st Qu.:0.00000
## Median :0.00000      Median :0.0000      Median :0.0000      Median :0.00000
## Mean      :0.01942      Mean      :0.2459      Mean      :0.2037      Mean      :0.02283
## 3rd Qu.:0.00000      3rd Qu.:0.0000      3rd Qu.:0.0000      3rd Qu.:0.00000
## Max.      :1.00000      Max.      :1.0000      Max.      :1.0000      Max.      :1.00000
## CryoSleep_True      Destination_      Destination_55.Cancric
## Min.      :0.000      Min.      :0.00000      Min.      :0.000
## 1st Qu.:0.000      1st Qu.:0.00000      1st Qu.:0.000
## Median :0.000      Median :0.00000      Median :0.000
## Mean      :0.342      Mean      :0.02047      Mean      :0.205
## 3rd Qu.:1.000      3rd Qu.:0.00000      3rd Qu.:0.000
## Max.      :1.000      Max.      :1.00000      Max.      :1.000
## Destination_PSO.J318.5.22      VIP_      VIP_True
## Min.      :0.0000      Min.      :0.00000      Min.      :0.00000
## 1st Qu.:0.0000      1st Qu.:0.00000      1st Qu.:0.00000
## Median :0.0000      Median :0.00000      Median :0.00000
## Mean      :0.0895      Mean      :0.02388      Mean      :0.02415
## 3rd Qu.:0.0000      3rd Qu.:0.00000      3rd Qu.:0.00000
## Max.      :1.0000      Max.      :1.00000      Max.      :1.00000
## Cabin1_      Cabin1_A      Cabin1_B      Cabin1_C
## Min.      :0.00000      Min.      :0.00000      Min.      :0.00000      Min.      :0.00000
## 1st Qu.:0.00000      1st Qu.:0.00000      1st Qu.:0.00000      1st Qu.:0.00000
## Median :0.00000      Median :0.00000      Median :0.00000      Median :0.00000
## Mean      :0.02178      Mean      :0.02992      Mean      :0.08819      Mean      :0.08609
## 3rd Qu.:0.00000      3rd Qu.:0.00000      3rd Qu.:0.00000      3rd Qu.:0.00000
## Max.      :1.00000      Max.      :1.00000      Max.      :1.00000      Max.      :1.00000
## Cabin1_D      Cabin1_E      Cabin1_G      Cabin2_
## Min.      :0.00000      Min.      :0.0000      Min.      :0.0000      Min.      :0.00000
## 1st Qu.:0.00000      1st Qu.:0.0000      1st Qu.:0.0000      1st Qu.:0.00000
## Median :0.00000      Median :0.0000      Median :0.0000      Median :0.00000
## Mean      :0.05906      Mean      :0.1005      Mean      :0.2874      Mean      :0.02178
## 3rd Qu.:0.00000      3rd Qu.:0.0000      3rd Qu.:1.0000      3rd Qu.:0.00000
## Max.      :1.00000      Max.      :1.0000      Max.      :1.0000      Max.      :1.00000
## Cabin2_P
## Min.      :0.0000
## 1st Qu.:0.0000
## Median :0.0000
## Mean      :0.4871
## 3rd Qu.:1.0000
## Max.      :1.0000

## user system elapsed
##      0      0      0

```