

FOLIE 1 (noch leer / Platzhalter)

FOLIE 2 (noch leer / Platzhalter)

Statistische Kennzahlen für Trainings– und Testdaten

2.1.1 Trainingsdaten Zusammenfassung

ID	Monate Letzte Spende	Anzahl Spenden	Gesamtvolumen	Monate Erste Spende	Spende Maerz 2007	Anzahl Zeilen: 576
Min. : 0.0	Min. : 0.000	Min. : 1.000	Min. : 250	Min. : 2.00	Min. : 0.0000	
1st Qu.:183.8	1st Qu.: 2.000	1st Qu.: 2.000	1st Qu.: 500	1st Qu.:16.00	1st Qu.:0.0000	
Median :375.5	Median : 7.000	Median : 4.000	Median : 1000	Median :28.00	Median :0.0000	
Mean :374.0	Mean : 9.439	Mean : 5.427	Mean : 1357	Mean :34.05	Mean :0.2396	
3rd Qu.:562.5	3rd Qu.:14.000	3rd Qu.: 7.000	3rd Qu.: 1750	3rd Qu.:49.25	3rd Qu.:0.0000	
Max. :747.0	Max. :74.000	Max. :50.000	Max. :12500	Max. :98.00	Max. :1.0000	

Output Variable

Spende März 2007 ja/nein:

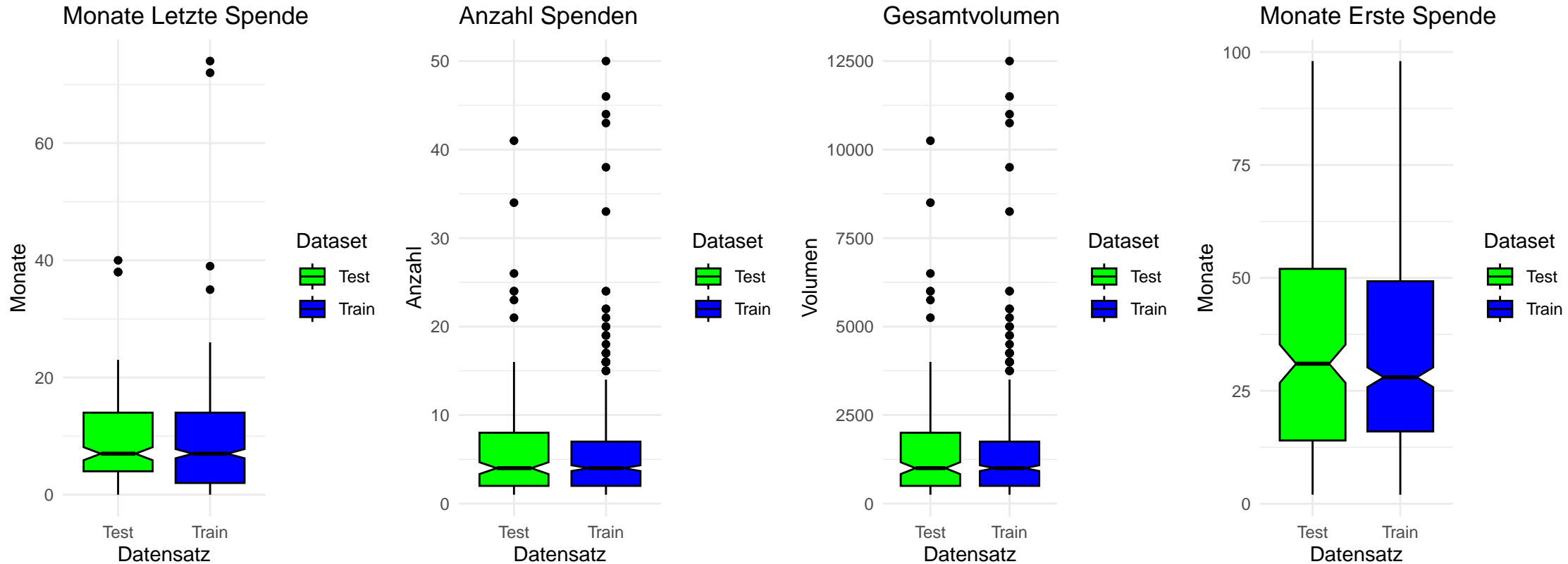
ja: 138

nein: 438

2.1.3 Testdaten Zusammenfassung

ID	Monate Letzte Spende	Anzahl Spenden	Gesamtvolumen	Monate Erste Spende	Anzahl Zeilen: 200
Min. : 1.0	Min. : 0.000	Min. : 1.000	Min. : 250	Min. : 2.00	
1st Qu.:198.2	1st Qu.: 4.000	1st Qu.: 2.000	1st Qu.: 500	1st Qu.:14.00	
Median :377.5	Median : 7.000	Median : 4.000	Median : 1000	Median :31.00	
Mean :374.6	Mean : 9.495	Mean : 5.935	Mean : 1484	Mean :35.48	
3rd Qu.:537.0	3rd Qu.:14.000	3rd Qu.: 8.000	3rd Qu.: 2000	3rd Qu.:52.00	
Max. :745.0	Max. :40.000	Max. :41.000	Max. :10250	Max. :98.00	

Boxplots Train vs. Test

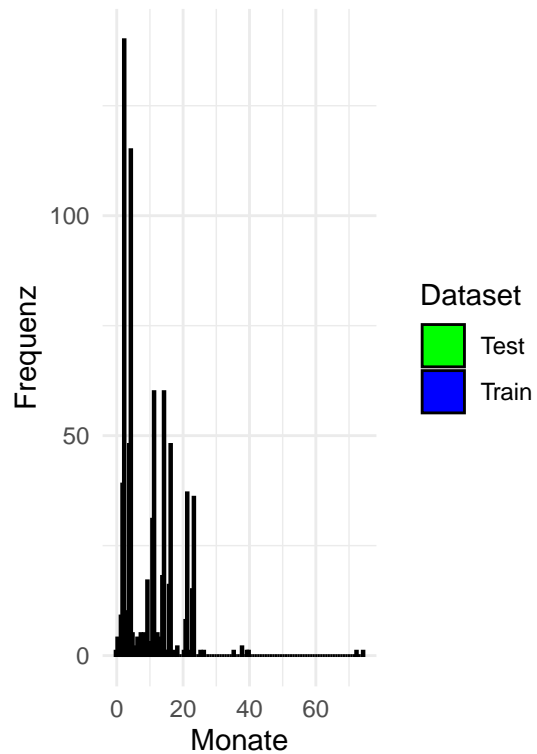


Erkenntnisse:

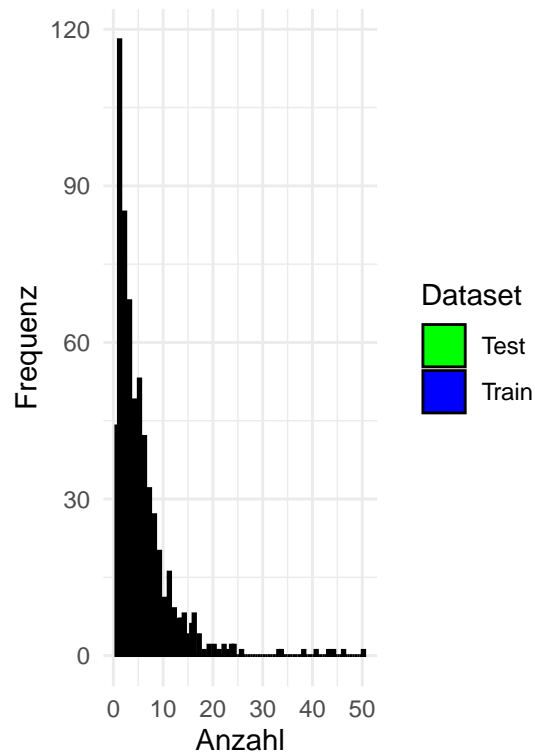
- Verteilungen der Merkmale
- Kein starker Train/Test-Shift

Histogramme Train vs. Test

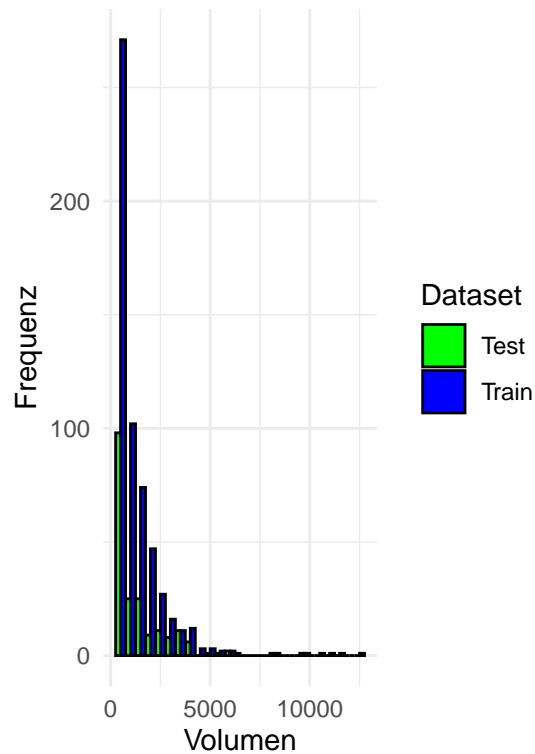
Monate Letzte Spende



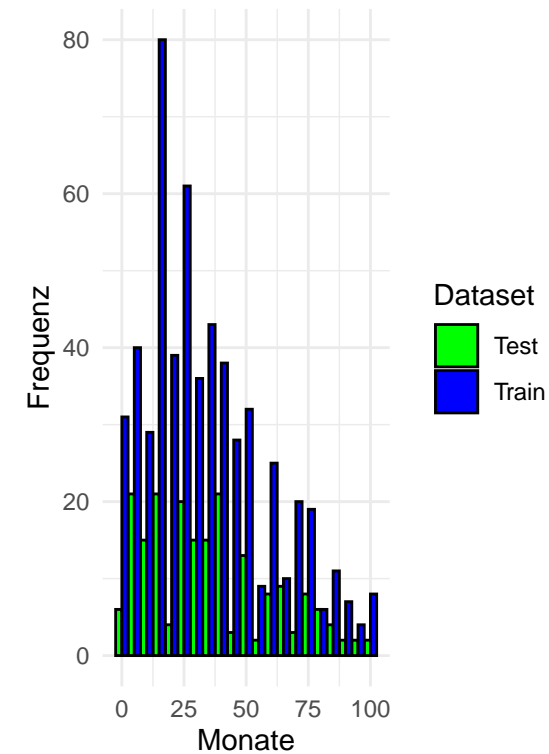
Anzahl Spenden



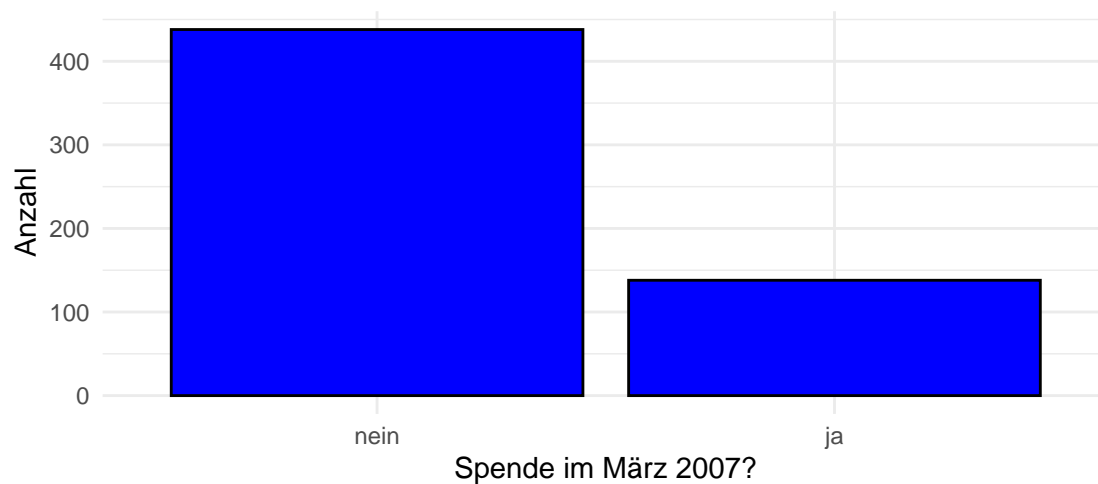
Gesamtvolumen



Monate Erste Spende



Klassenverteilung (Train)

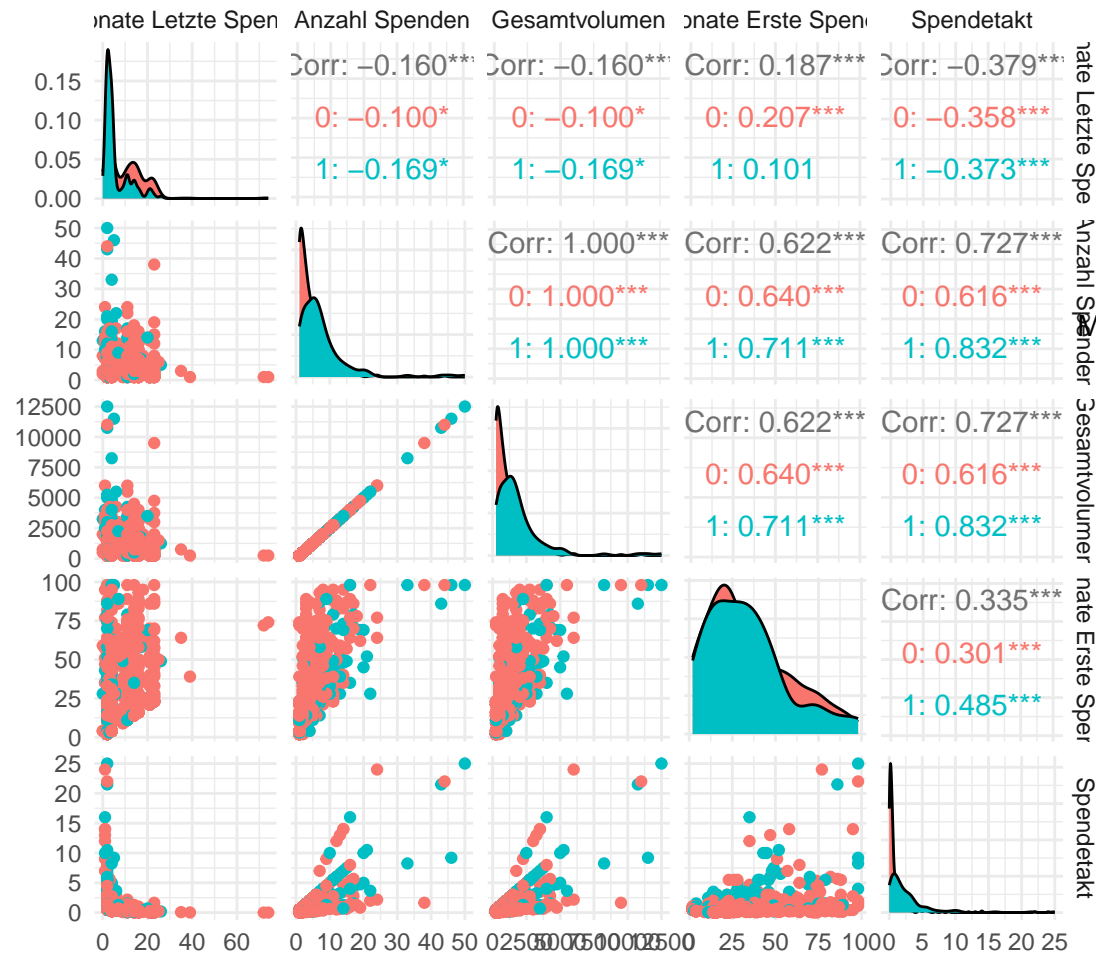


Erkenntnisse:

- Verteilungen ähnlich
- Test hat weniger hohe Spenderzahlen
- ~60% 'ja', 40% 'nein'

Beziehung der Variablen

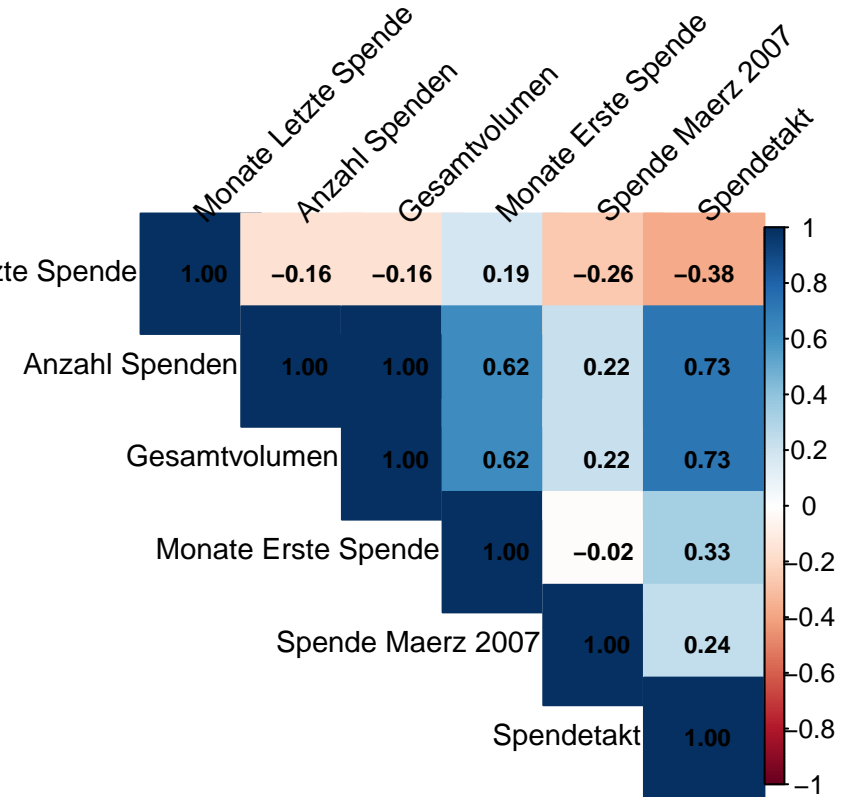
Pairs-Plot Trainingsdaten



Erkenntnisse:

- Hohe Korrelation von Spenden & Volumen
- Spendetakt korreliert ebenso

Korrelationsmatrix



Fazit:

- 'Gesamtvolumen' redundant
- 'Spendetakt' als Feature nutzen

Modellerstellung & Variable Importance

```
Call:
glm(formula = `Spende Maerz 2007` ~ `Anzahl Spenden` + `Monate Letzte Spende` +
  `Spenden Rate` + `Monate Erste Spende`, family = binomial(link = "logit"),
  data = glm_train)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.509543	0.213421	-2.388	0.01696	*
`Anzahl Spenden`	0.140069	0.034712	4.035	5.46e-05	***
`Monate Letzte Spende`	-0.100581	0.022230	-4.525	6.05e-06	***
`Spenden Rate`	-0.041674	0.060164	-0.693	0.48851	
`Monate Erste Spende`	-0.018455	0.006589	-2.801	0.00509	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

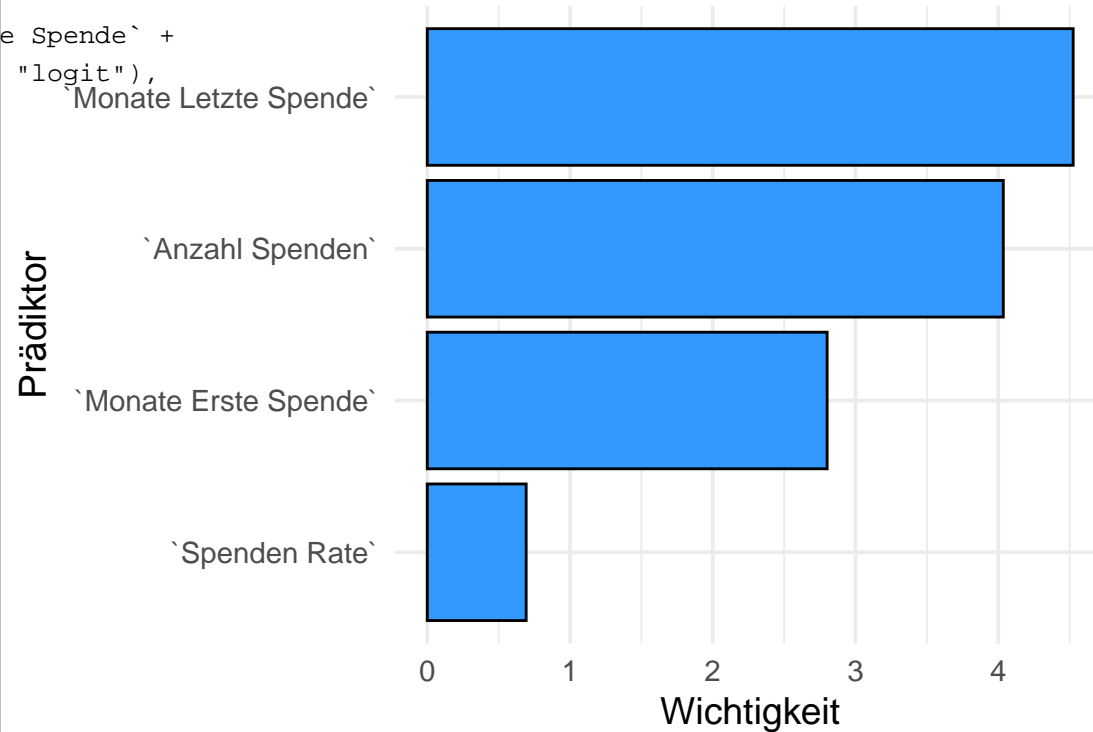
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 629.79 on 571 degrees of freedom
Residual deviance: 552.22 on 567 degrees of freedom
AIC: 562.22

Erkenntnisse:
Number of Fisher Scoring iterations: 5

- Höchste Bedeutung: Monate Letzte Spende
- 'Spenden Rate' liefert Mehrwert

Variable Importance (Logit)



Metriken und Konfusionsmatrix

Confusion Matrix and Statistics

	Reference	
Prediction	0	1
0	425	124
1	10	13

Accuracy : 0.7657

95% CI : (0.7288, 0.7999)

No Information Rate : 0.7605

P-Value [Acc > NIR] : 0.4064

Kappa : 0.1006

McNemar's Test P-Value : <2e-16

Sensitivity : 0.97701

Specificity : 0.09489

Pos Pred Value : 0.77413

Neg Pred Value : 0.56522

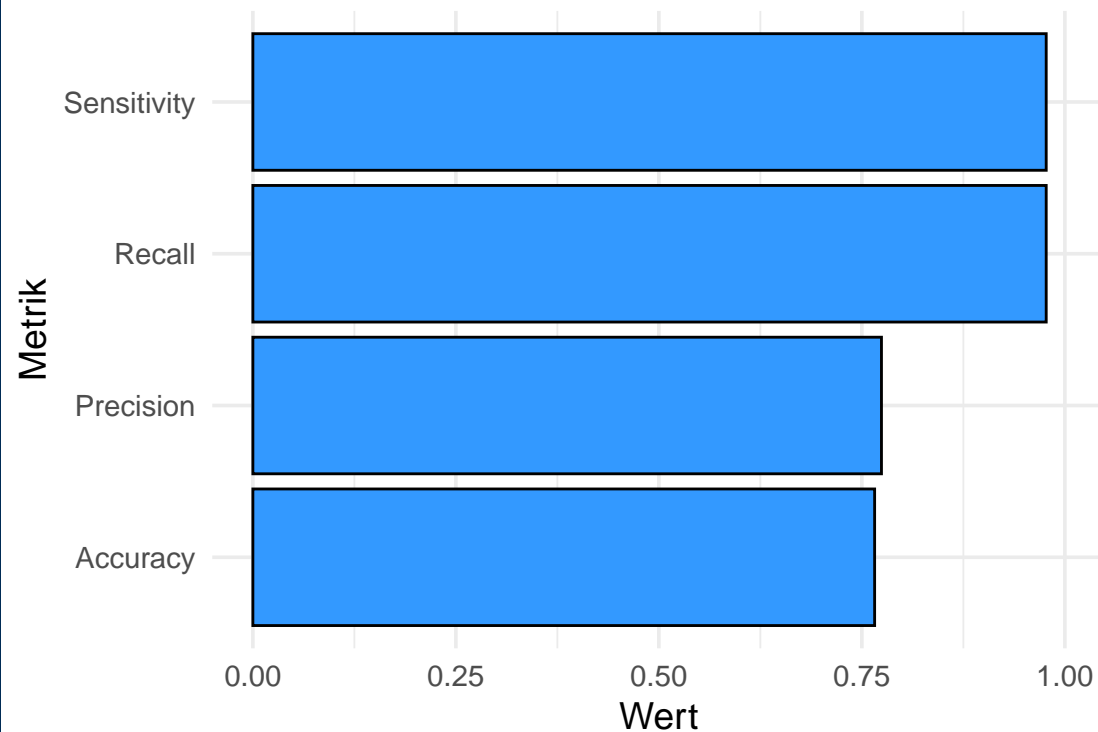
Prevalence : 0.76049

Erkenntnisse:

- Modellwerte solide
- Sensitivität = Recall

Feintuning möglich

Kennzahlen (Train)



Modellleistung & ROC-Kurve

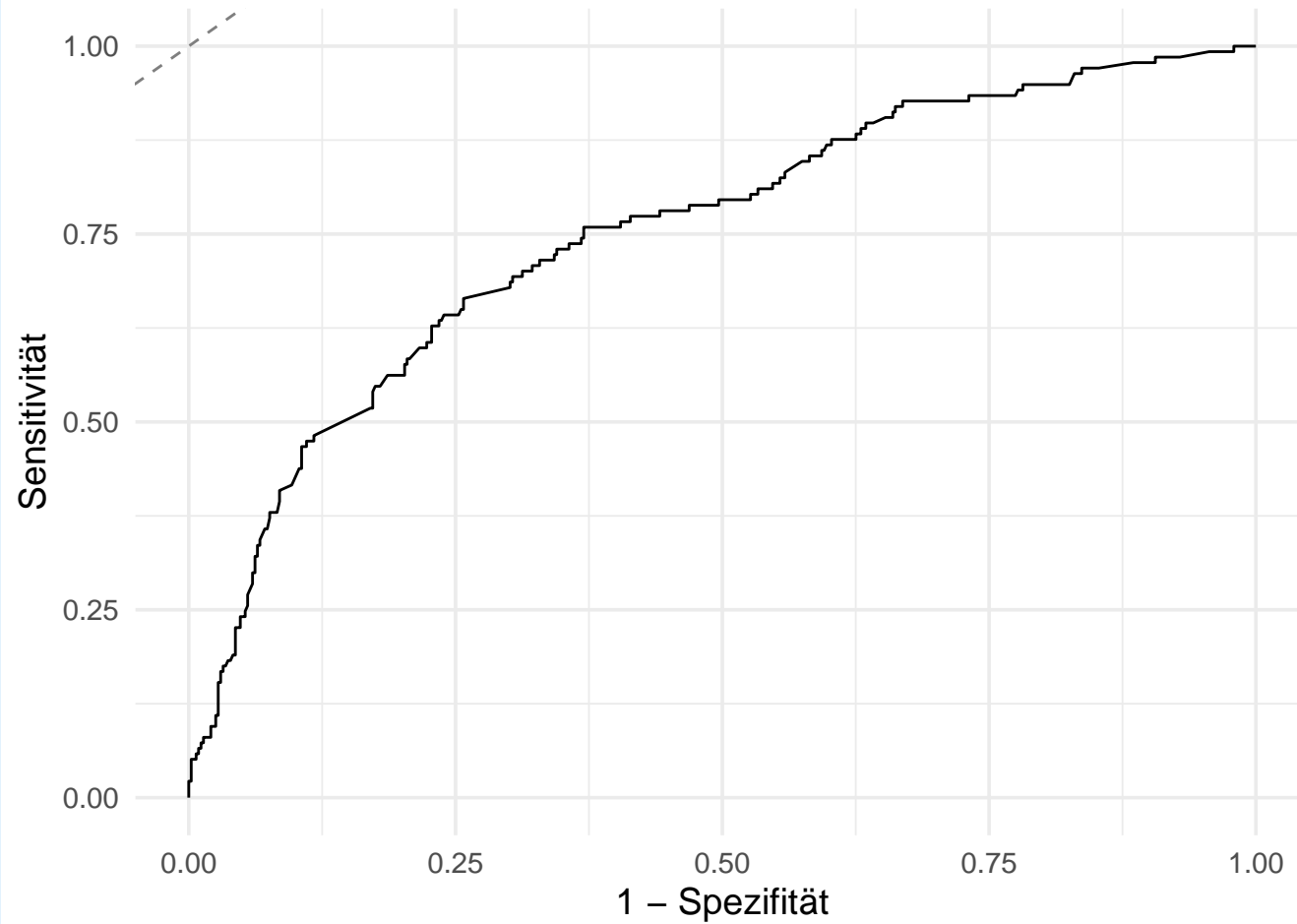
Modell: Logistische Regression
Metrik: ROC/AUC zur Trennschärfe

AUC = 0.751

Interpretation:

- 1 = perfekte Trennung, 0.5 = Zufall

ROC-Kurve (AUC = 0.751)



Erkenntnisse:

- AUC > 0.8 ... gute Trennschärfe
- Nächster Schritt: Testdatenprädiktion

FOLIE 10 (noch leer / Platzhalter)