

Silas_Analysis

Max Kuttner

25 3 2020

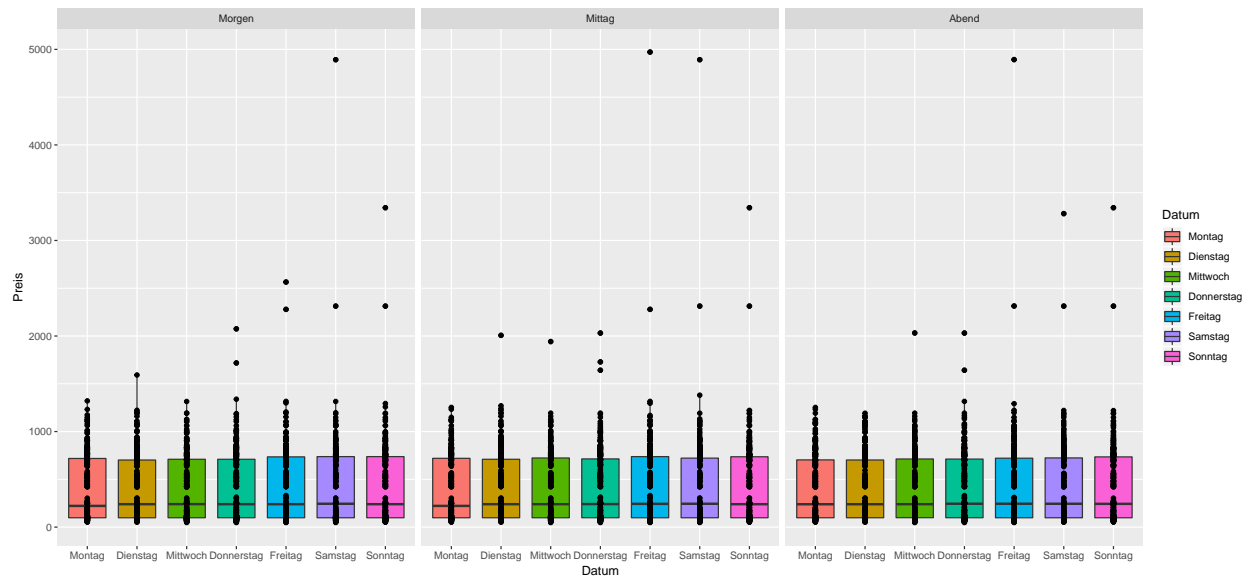
Hypothese 1

[Erfolgt die Buchung von Flugtickets Dienstagabends, kann statistisch gesehen das preiswerteste Offert erzielt werden.]

Deskreptive Statistik:

```
##
## DESCRIPTIVES
##
## Descriptives
## -----
##               Preis      Datum      Zeit
## -----
##      N              7224      7224      7224
##      Missing           0           0           0
##      Mean             395
##      Median            241
##      Standard deviation  364
##      Minimum           48.0
##      Maximum           4972
## -----
##
##
## FREQUENCIES
##
## Frequencies of Datum
## -----
##      Levels      Counts      % of Total      Cumulative %
## -----
##      Montag           1008           14.0           14.0
##      Dienstag          1008           14.0           27.9
##      Mittwoch           1008           14.0           41.9
##      Donnerstag          1050           14.5           56.4
##      Freitag            1050           14.5           70.9
##      Samstag            1050           14.5           85.5
##      Sonntag            1050           14.5          100.0
## -----
##
##
## Frequencies of Zeit
## -----
##      Levels      Counts      % of Total      Cumulative %
## -----
##      Morgen          2408           33.3           33.3
##      Mittag           2408           33.3           66.7
##      Abend            2408           33.3          100.0
## -----
```

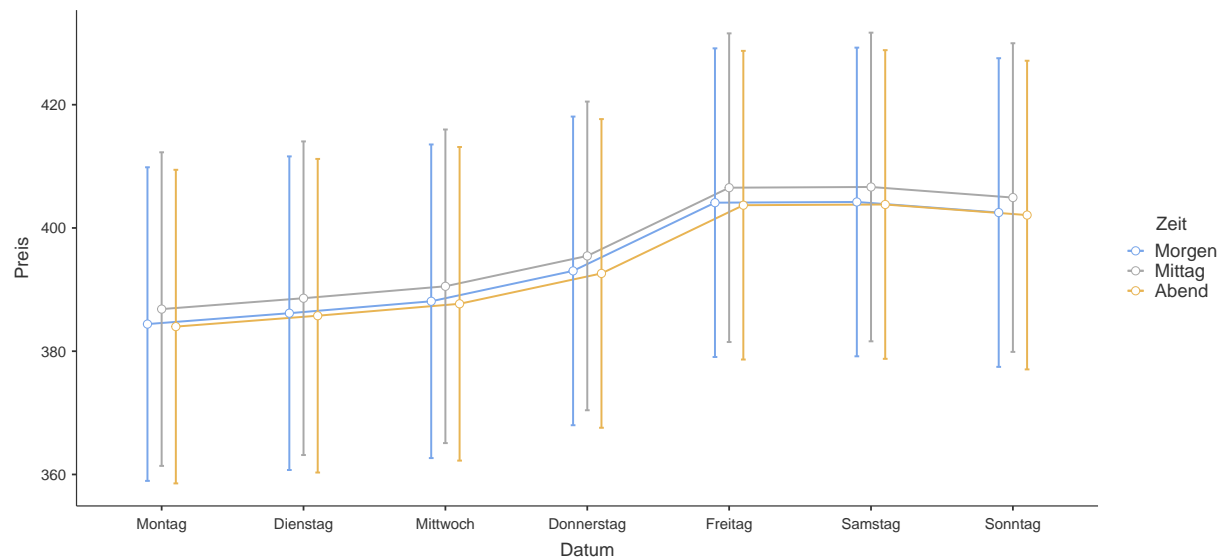
Verteilung - BoxPlot



Mehrfaktorielle ANOVA:

```
##
## ANOVA
##
## ANOVA
## -----
##              Sum of Squares    df      Mean Square    F        p        <U+03B7>²p
## -----
## Datum              478813         6          79802    0.6011    0.730    0.000
## Zeit               11357          2           5678    0.0428    0.958    0.000
## Residuals         9.58e+8       7215        132767
## -----
##
##
## ESTIMATED MARGINAL MEANS
##
## DATUM:ZEIT
##
## Estimated Marginal Means - Datum:Zeit
## -----
##      Zeit      Datum      Mean      SE      Lower      Upper
## -----
## Morgen
##      Montag      384      13.0      359      410
##      Dienstag    386      13.0      361      412
##      Mittwoch     388      13.0      363      414
##      Donnerstag    393      12.8      368      418
##      Freitag      404      12.8      379      429
##      Samstag      404      12.8      379      429
##      Sonntag      403      12.8      377      428
## Mittag
##      Montag      387      13.0      361      412
##      Dienstag    389      13.0      363      414
##      Mittwoch     391      13.0      365      416
```

```
##           Donnerstag    395    12.8    370    421
##           Freitag      407    12.8    381    432
##           Samstag      407    12.8    382    432
##           Sonntag      405    12.8    380    430
##   Abend Montag        384    13.0    359    409
##           Dienstag    386    13.0    360    411
##           Mittwoch     388    13.0    362    413
##           Donnerstag    393    12.8    368    418
##           Freitag      404    12.8    379    429
##           Samstag      404    12.8    379    429
##           Sonntag      402    12.8    377    427
## -----
```



Estimated marginal means korrigiert Missverhältnisse aus unterschiedlich großen Sample-Größen für einzelne Tage. Somit wird jeder/jede Tag/Uhrzeit gleich gewertet. Wie oft jeder einzelne Tag gemessen wurde bzw. im Datensatz vorkommt, ist in der deskriptiven Statistik unter **FREQUENCIES** zu sehen. Für mehr Infos zum EMM: <https://cran.r-project.org/web/packages/emmeans/vignettes/basics.html>

Im folgenden werden Tage und Uhrzeiten nach ihrem mean (also **Preis**) angeordnet.

```
## # A tibble: 21 x 6
##   Zeit Datum    mean    se lower upper
##   <fct> <fct>    <dbl> <dbl> <dbl> <dbl>
## 1 Abend Montag    384.   13.0  359.  409.
## 2 Morgen Montag    384.   13.0  359.  410.
## 3 Abend Dienstag  386.   13.0  360.  411.
## 4 Morgen Dienstag  386.   13.0  361.  412.
## 5 Mittag Montag    387.   13.0  361.  412.
## 6 Abend Mittwoch  388.   13.0  362.  413.
## 7 Morgen Mittwoch  388.   13.0  363.  414.
## 8 Mittag Dienstag  389.   13.0  363.  414.
## 9 Mittag Mittwoch  391.   13.0  365.  416.
## 10 Abend Donnerstag 393.   12.8  368.  418.
## # ... with 11 more rows
```

Hypothese 2

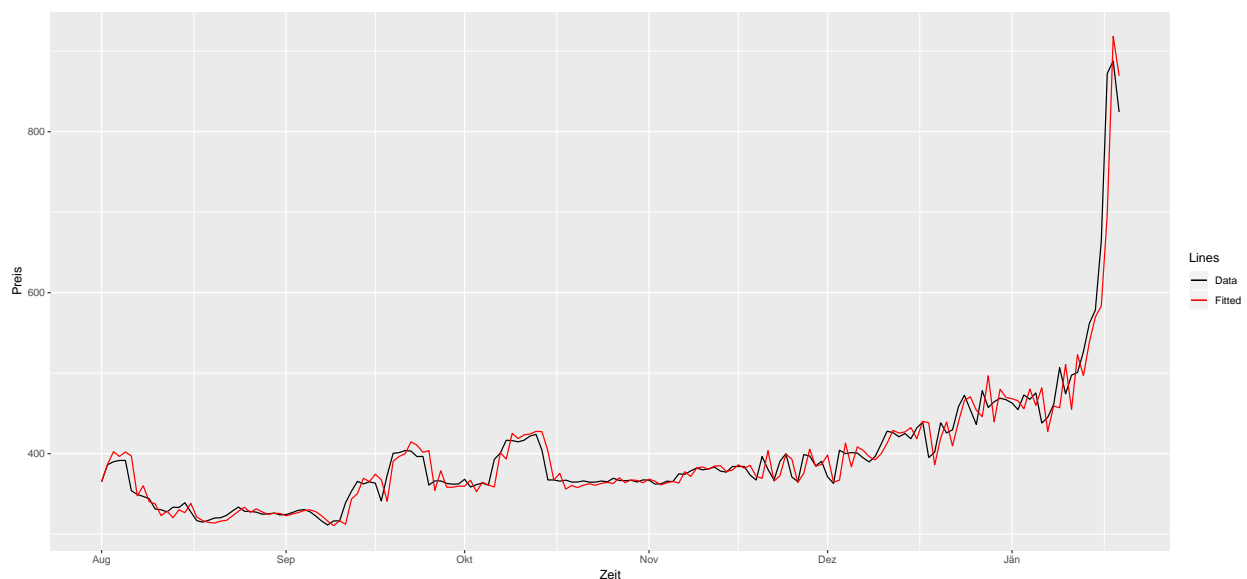
[Je spontaner und kurzfristiger die Kaufentscheidung getroffen wird, desto höher ist der offerierte Preis einer Airline.]

```
## Warning in value[[3L]](cond): The chosen test encountered an error, so no
## seasonal differencing is selected. Check the time series data.
```

Diese Zeitreihe lässt sich mit einem ARIMA(2,2,1)-Modell modellieren.

```
## Series: myts
## ARIMA(2,2,1)
##
## Coefficients:
##          ar1      ar2      ma1
##      0.0271  -0.4283  -0.7750
## s.e.  0.1274   0.1189   0.1074
##
## sigma^2 estimated as 467.7:  log likelihood=-763.15
## AIC=1534.3   AICc=1534.54   BIC=1546.84
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE  MASE      ACF1
## Training set 1.289643 21.31025 11.96192 0.140989 2.742449   NaN 0.005852603
##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1  0.02708    0.12737   0.2126 0.8316330
## ar2 -0.42829    0.11888  -3.6027 0.0003149 ***
## ma1 -0.77498    0.10736  -7.2182 5.268e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Mit diesem Modell kann in weiterer Folge die Zeitreihe angenähert werden und bis zum 1. Februar 2020 vorhergesagt werden.



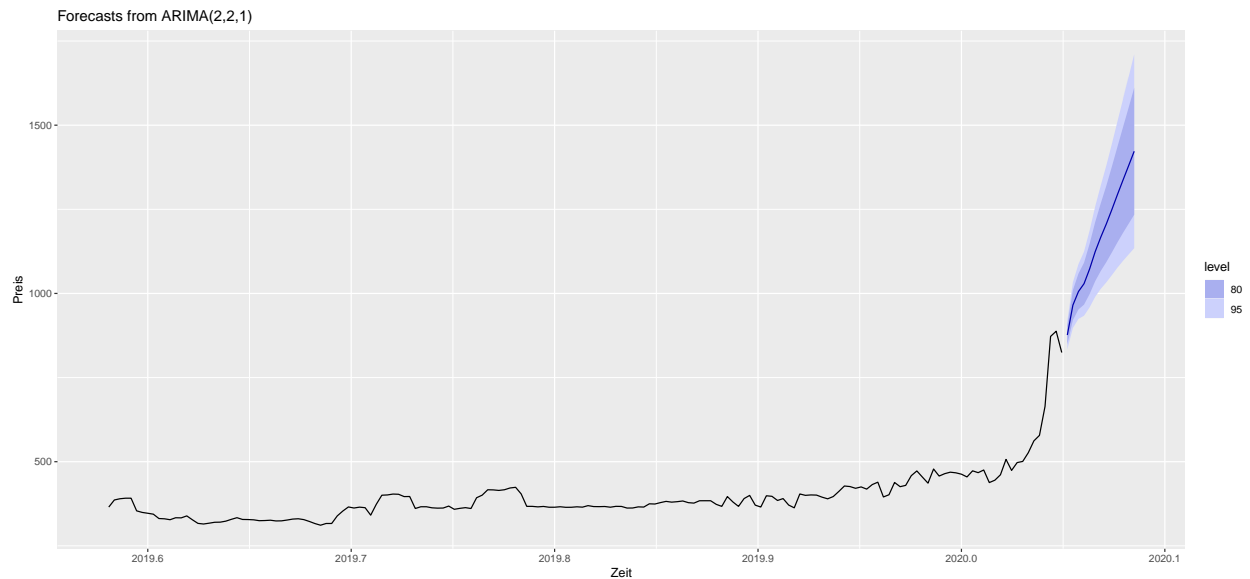
```
## $title
## [1] "Zeitreihe - Flugpreise"
##
## attr("class")
## [1] "labels"

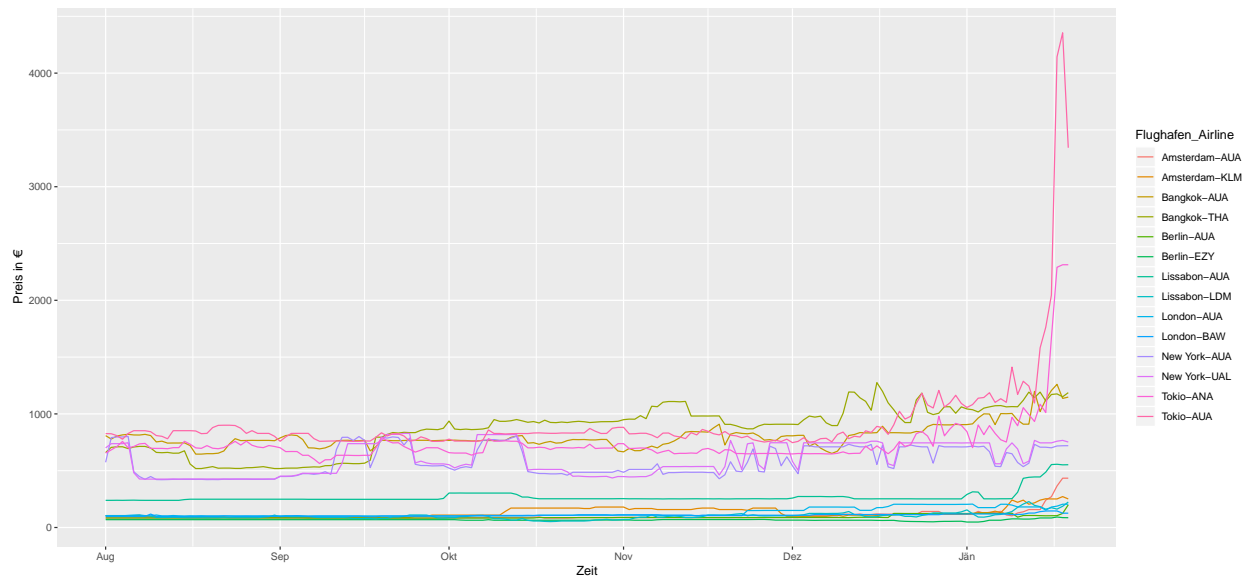
## Warning: package 'lubridate' was built under R version 3.5.3

##
## Attaching package: 'lubridate'

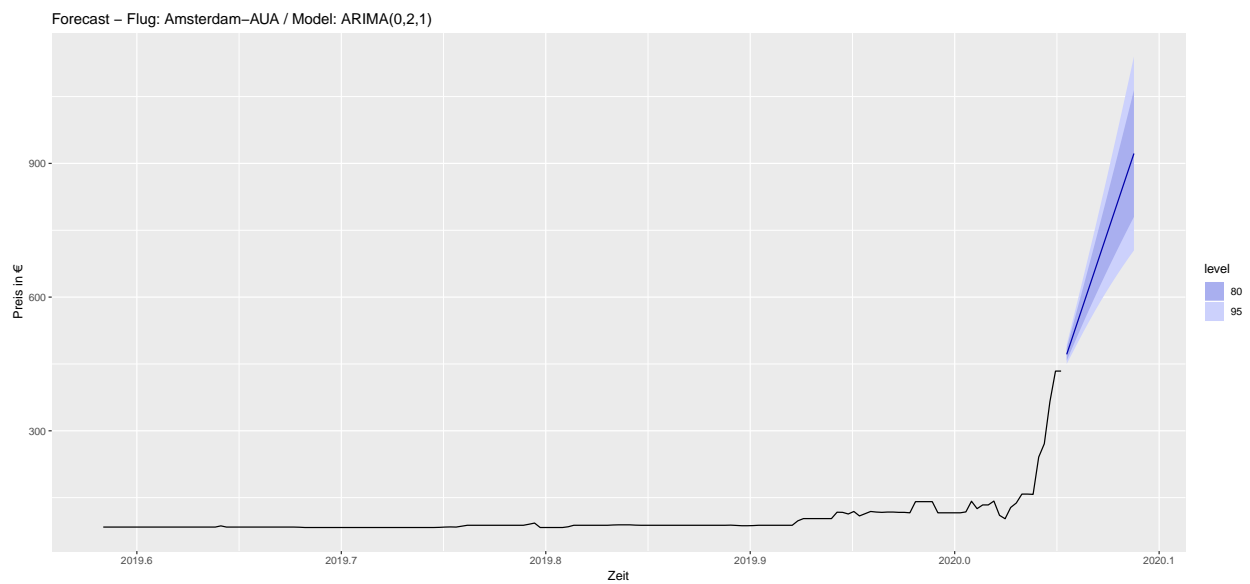
## The following objects are masked from 'package:data.table':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday,
##     week, yday, year

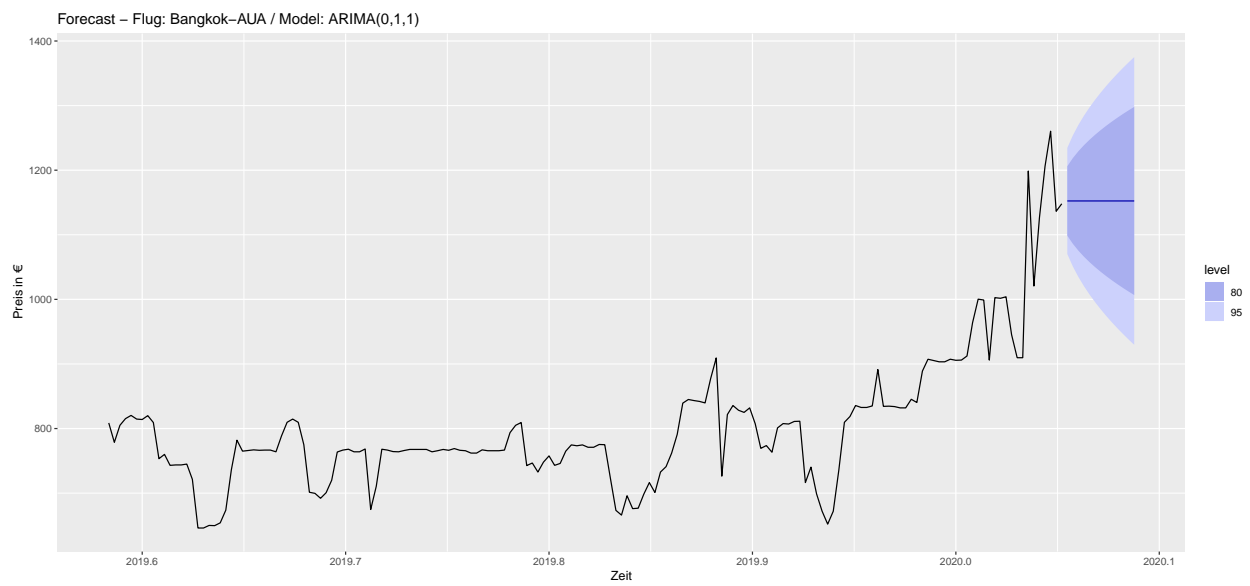
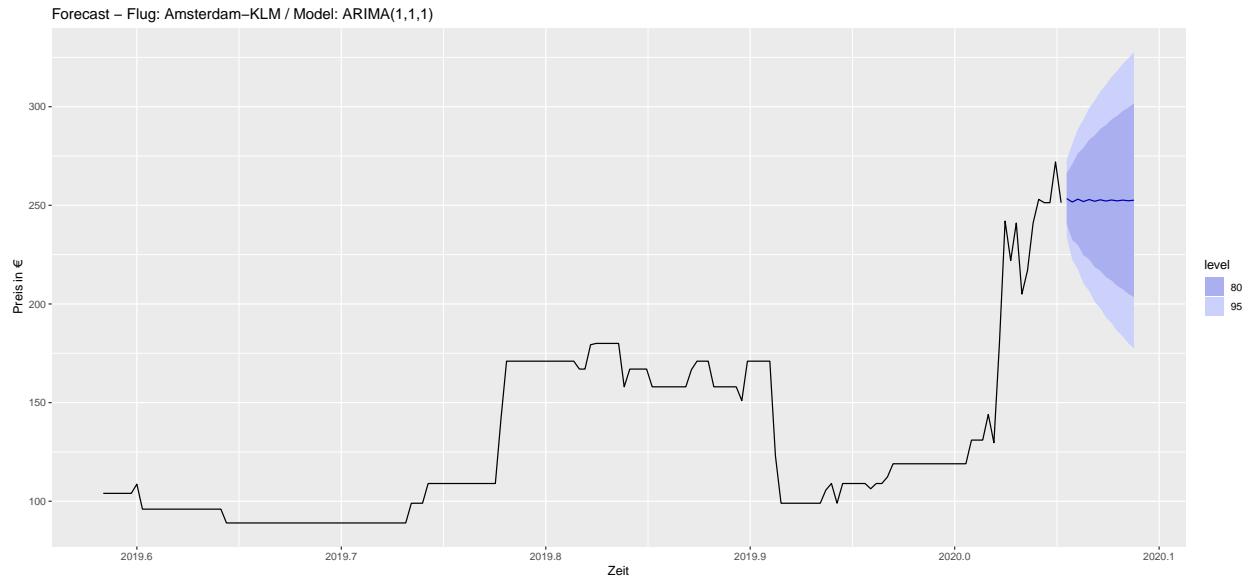
## The following object is masked from 'package:base':
##
##     date
```

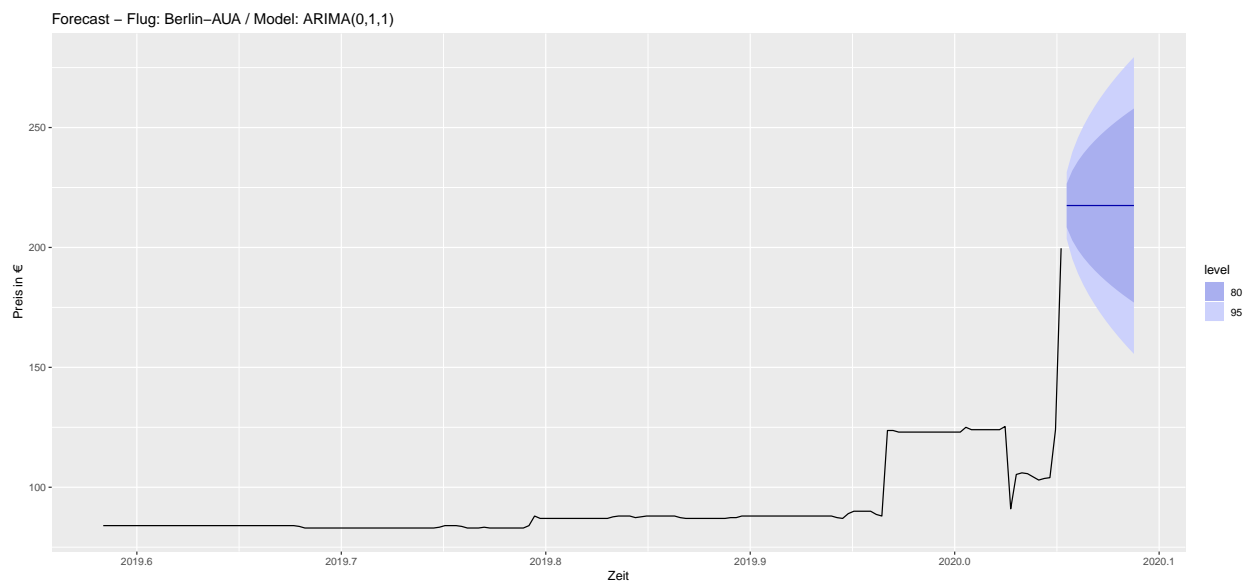
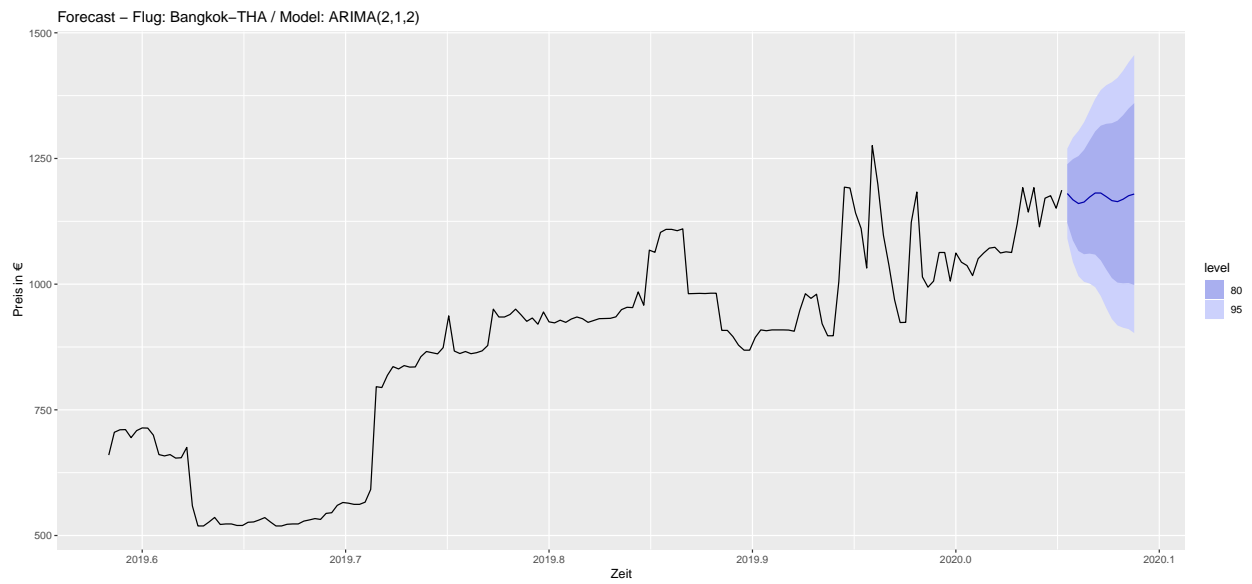


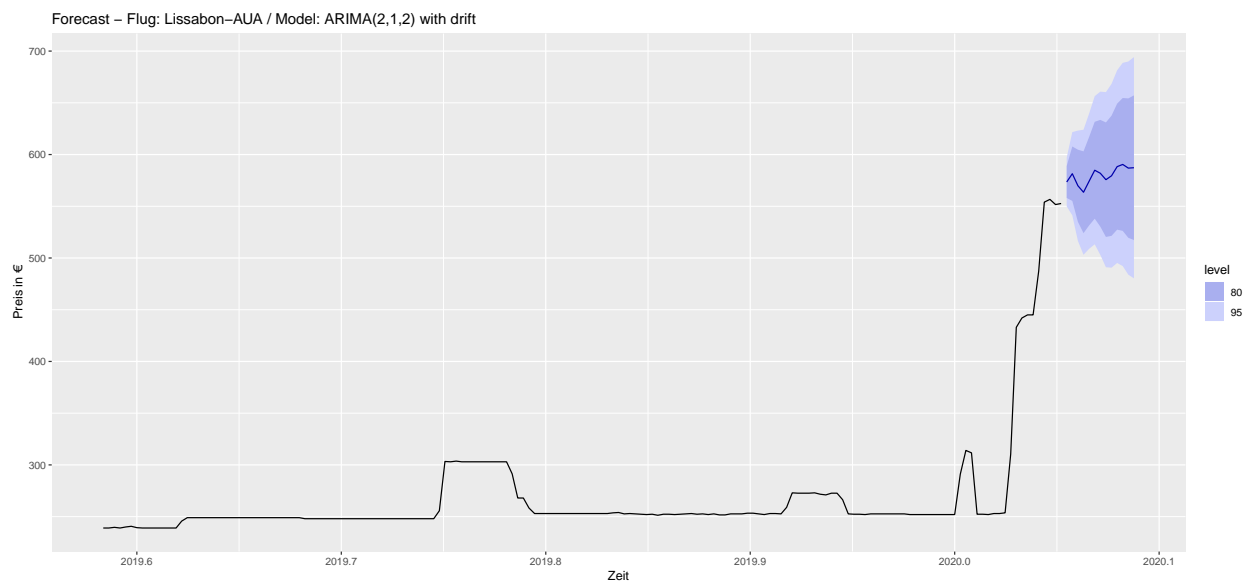
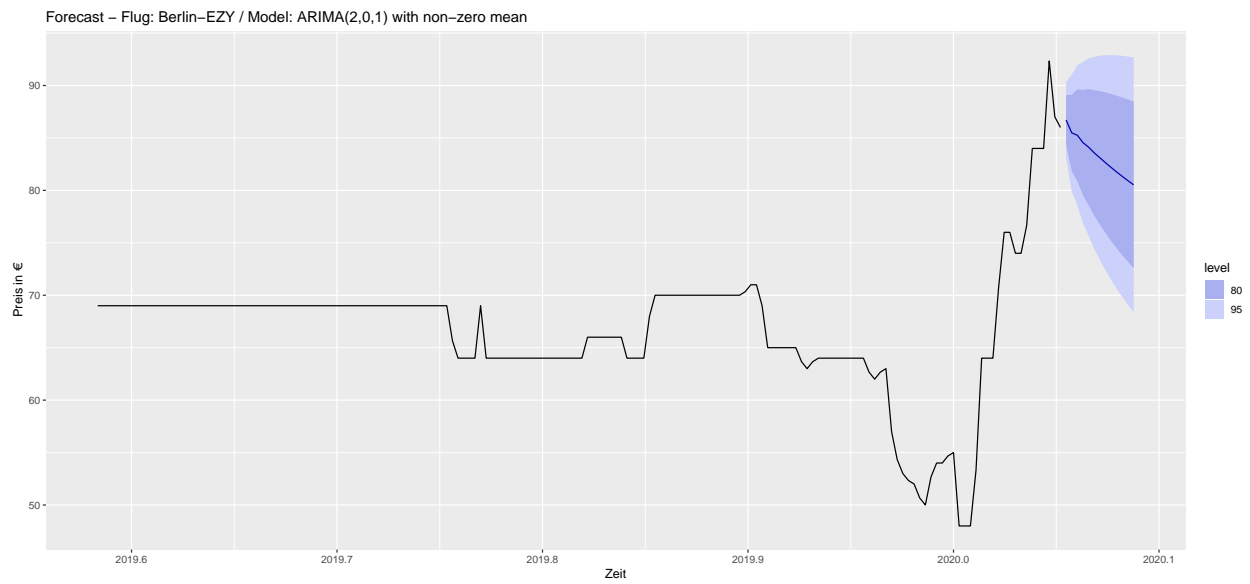


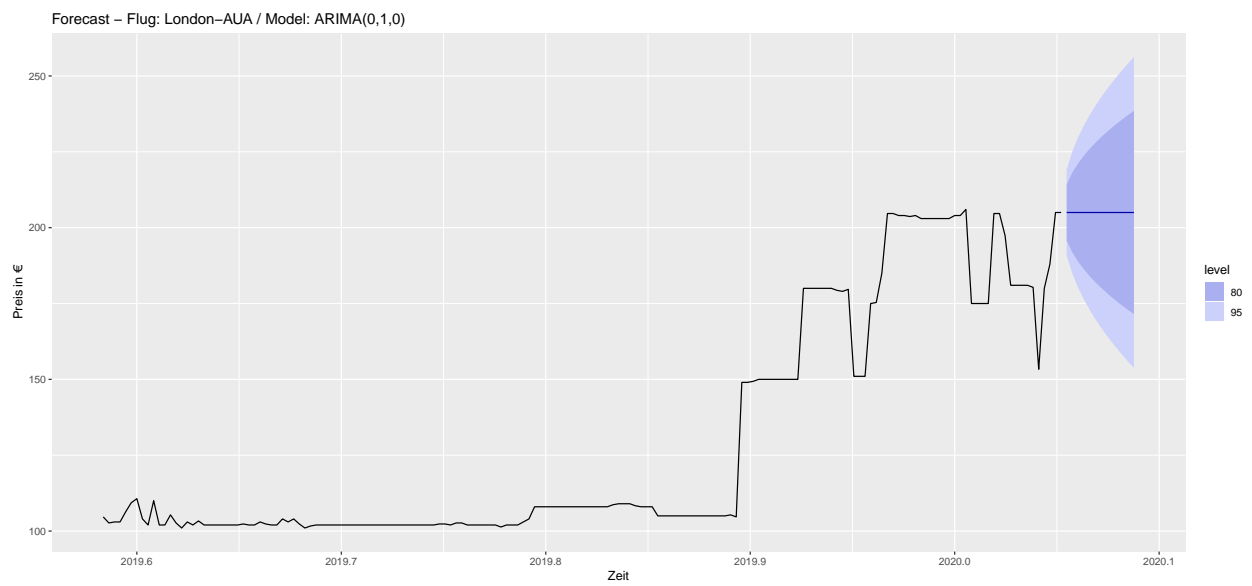
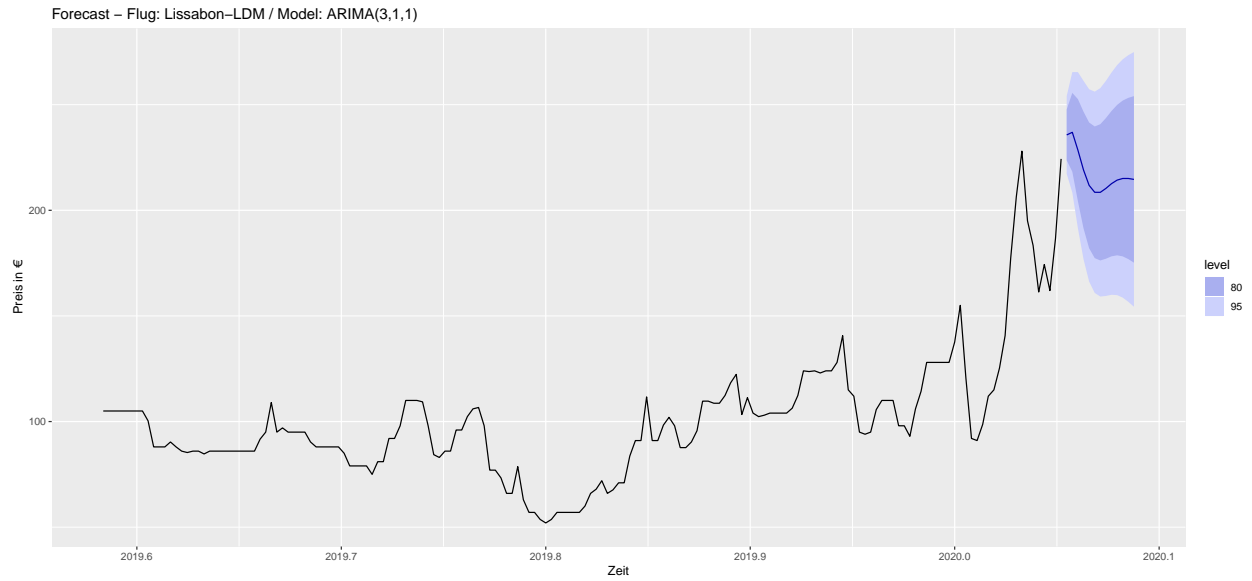
```
## $title
## [1] "Preis im Zeitverlauf - alle Flüge"
##
## attr("class")
## [1] "labels"
```

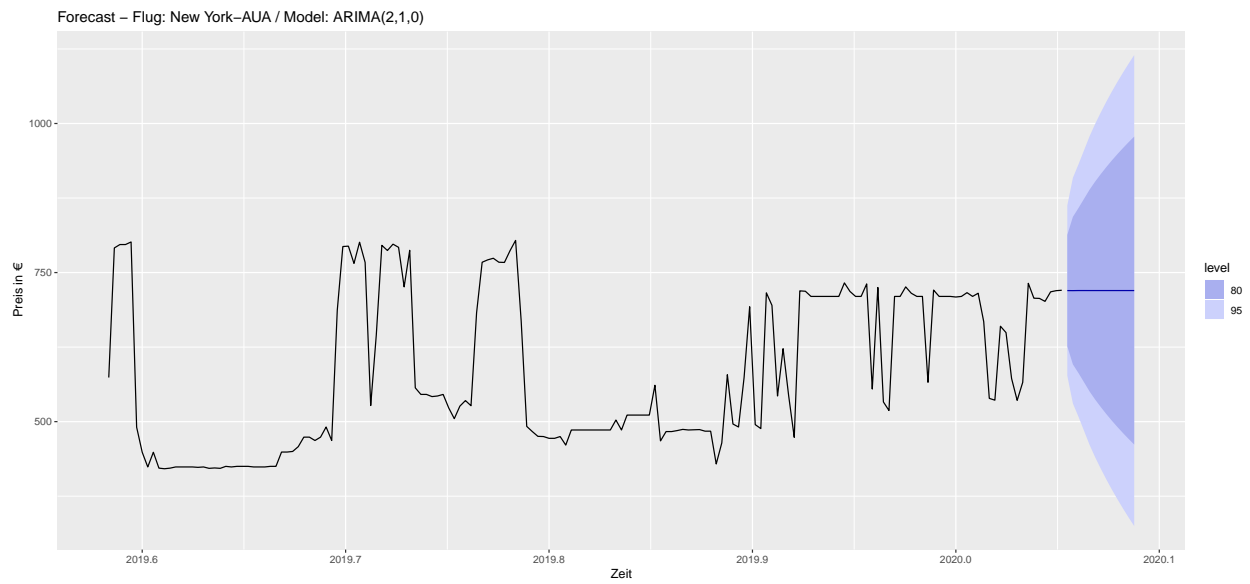
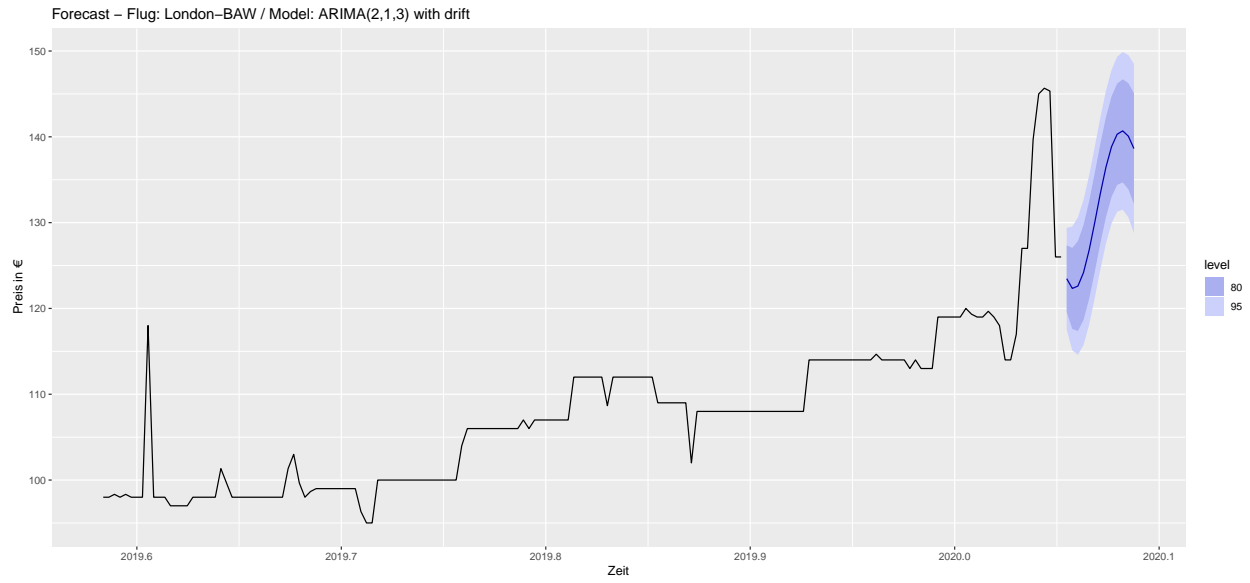


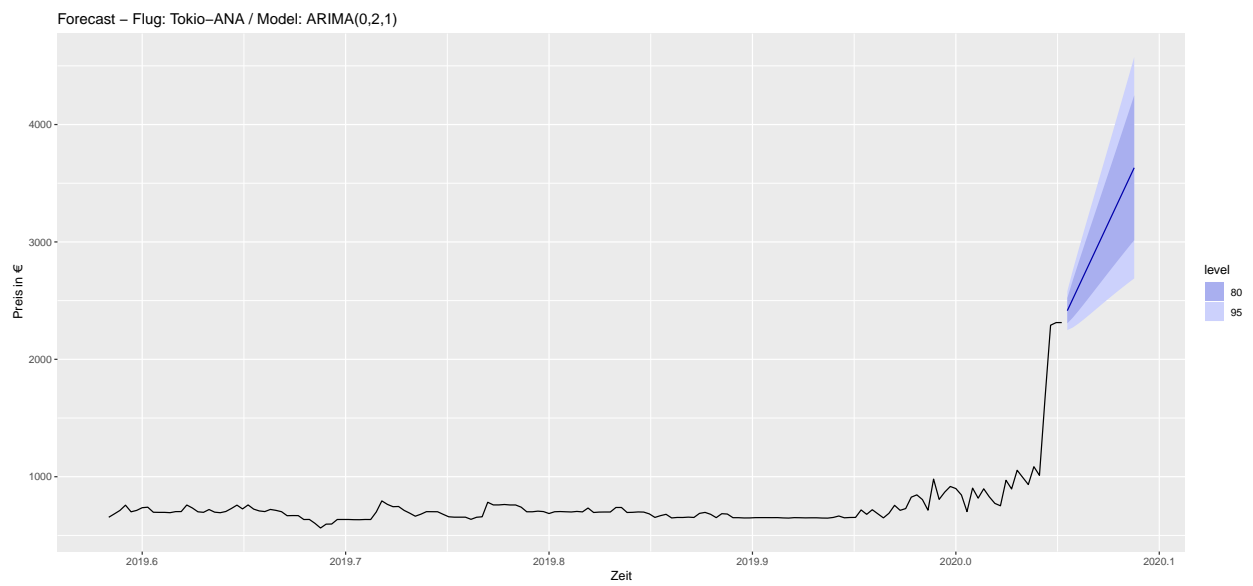
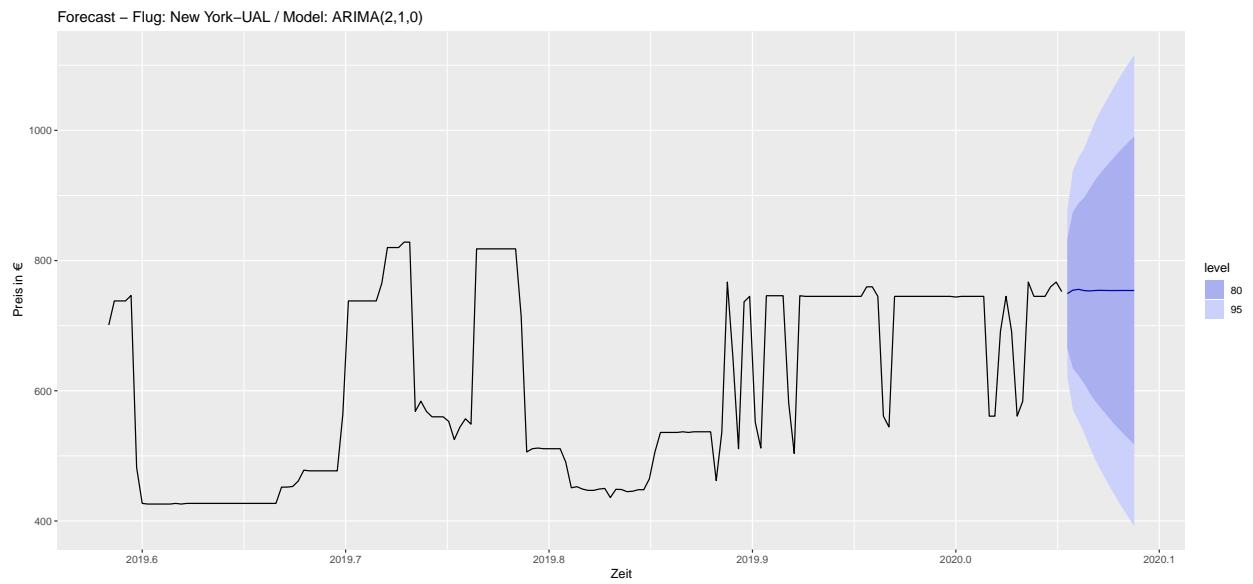


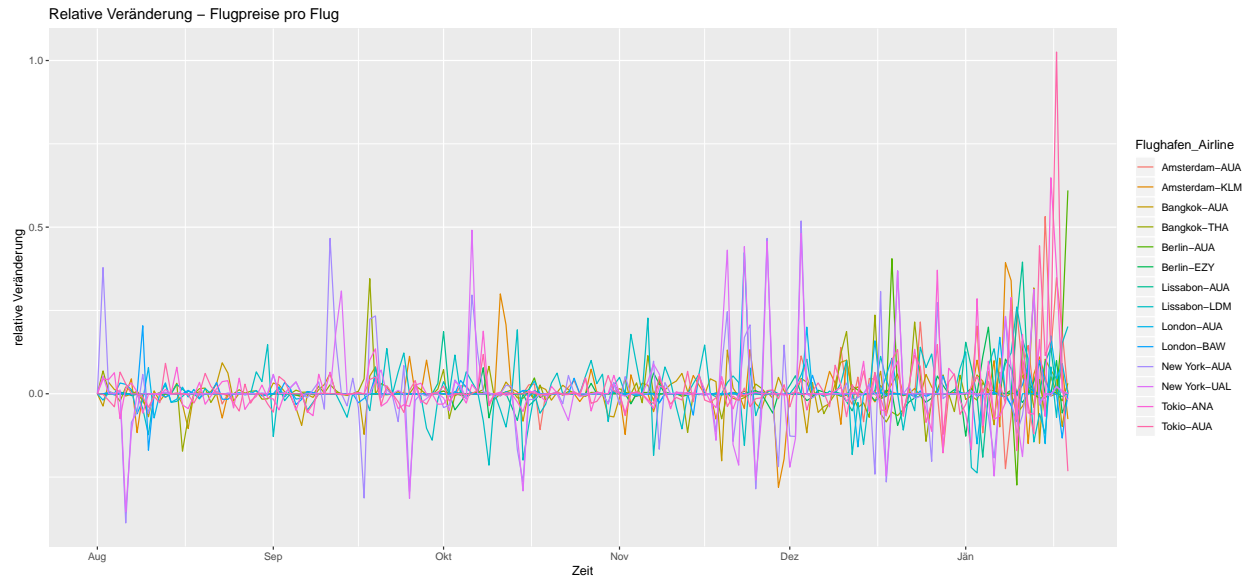
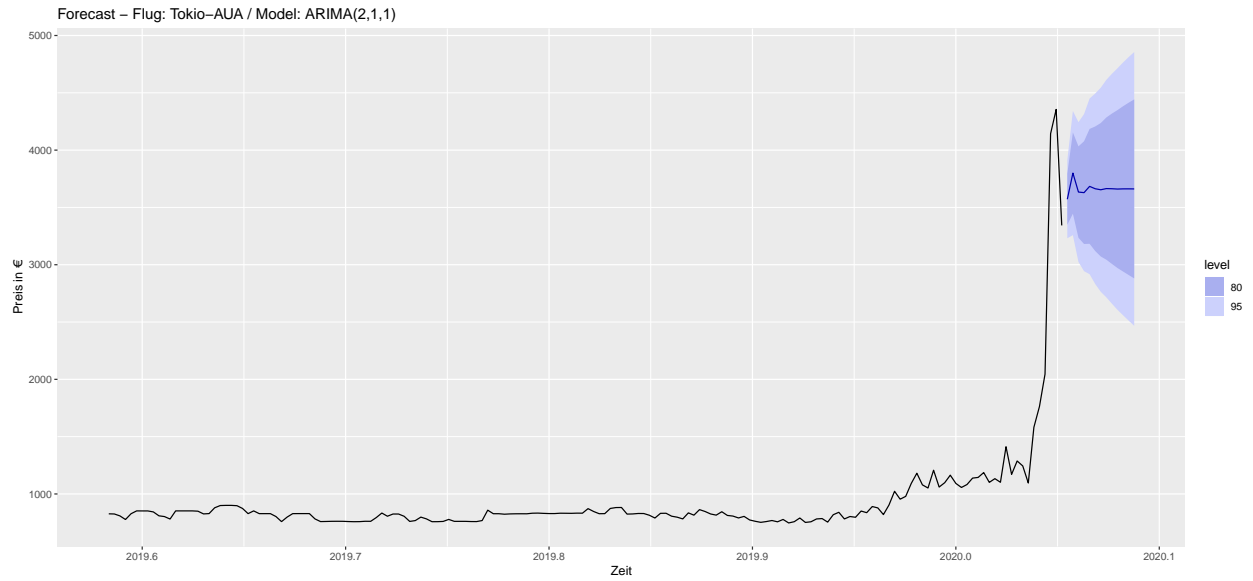


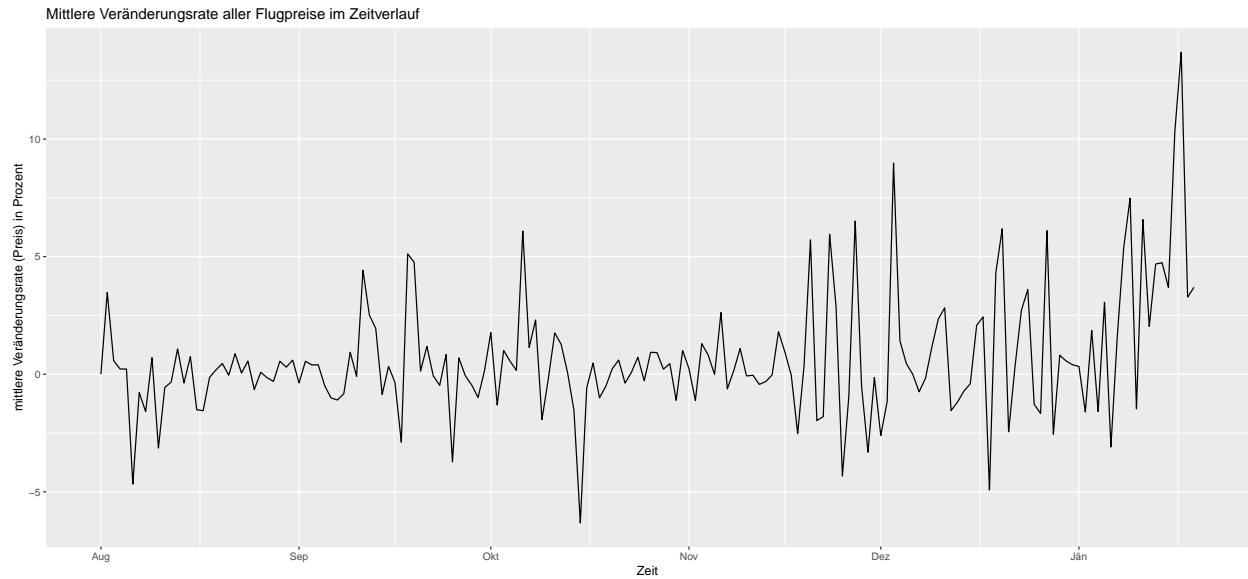




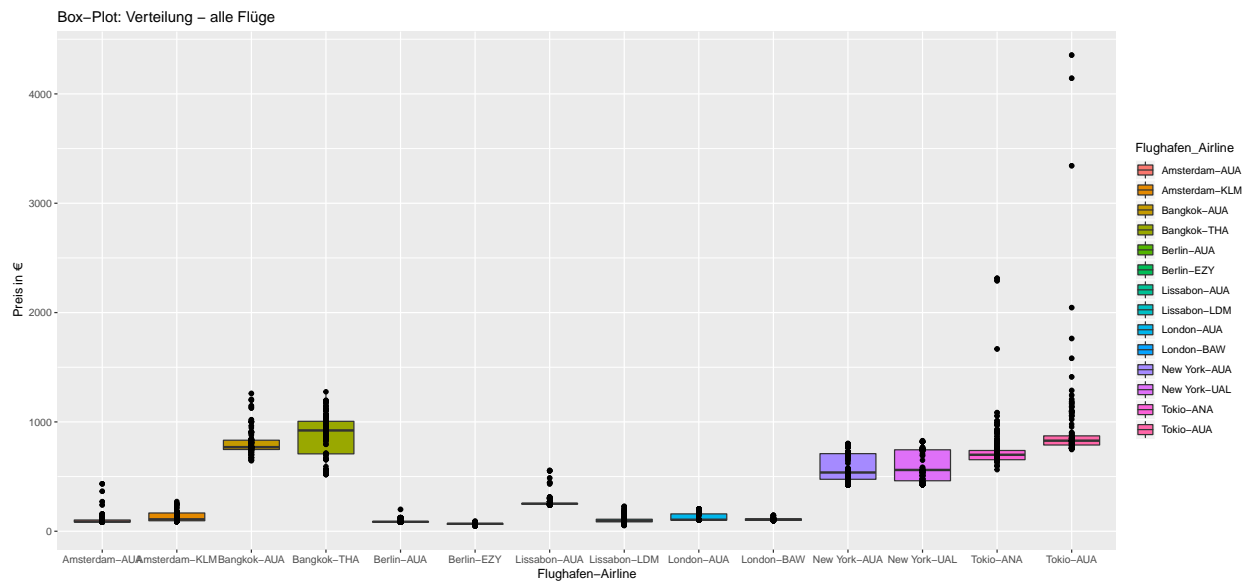








BoxPlot Vergleich aller Flüge



Hypothese 3

[Die Wahl des Betriebssystems respektive die Marke des Nutzerendgeräts mit dem die Reise-Website abgerufen wird, hat eine Auswirkung auf den offerierten Preis einer Airline.]

Deskreptive Statistik:

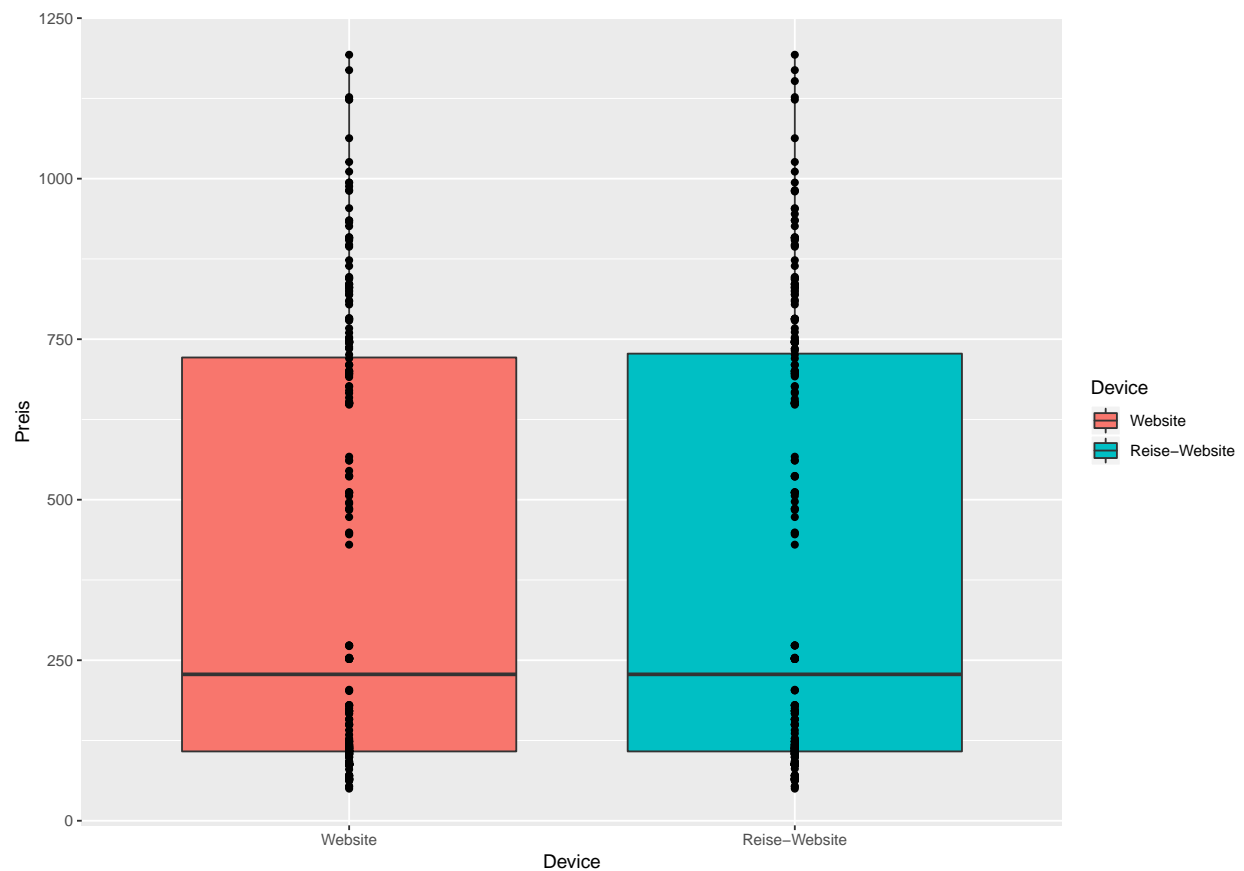
```
##
## DESCRIPTIVES
##
## Descriptives
## -----
##               Preis      Device
## -----
## N                560      560
```

```

##      Missing              0          0
##      Mean                396
##      Median              228
##      Standard deviation   333
##      Minimum             50.0
##      Maximum             1193
## -----
##
##
##
## FREQUENCIES
##
## Frequencies of Device
## -----
##      Levels      Counts    % of Total    Cumulative %
## -----
##      Website      280       50.0         50.0
##      Reise-Website 280       50.0        100.0
## -----

```

Verteilung - BoxPlot



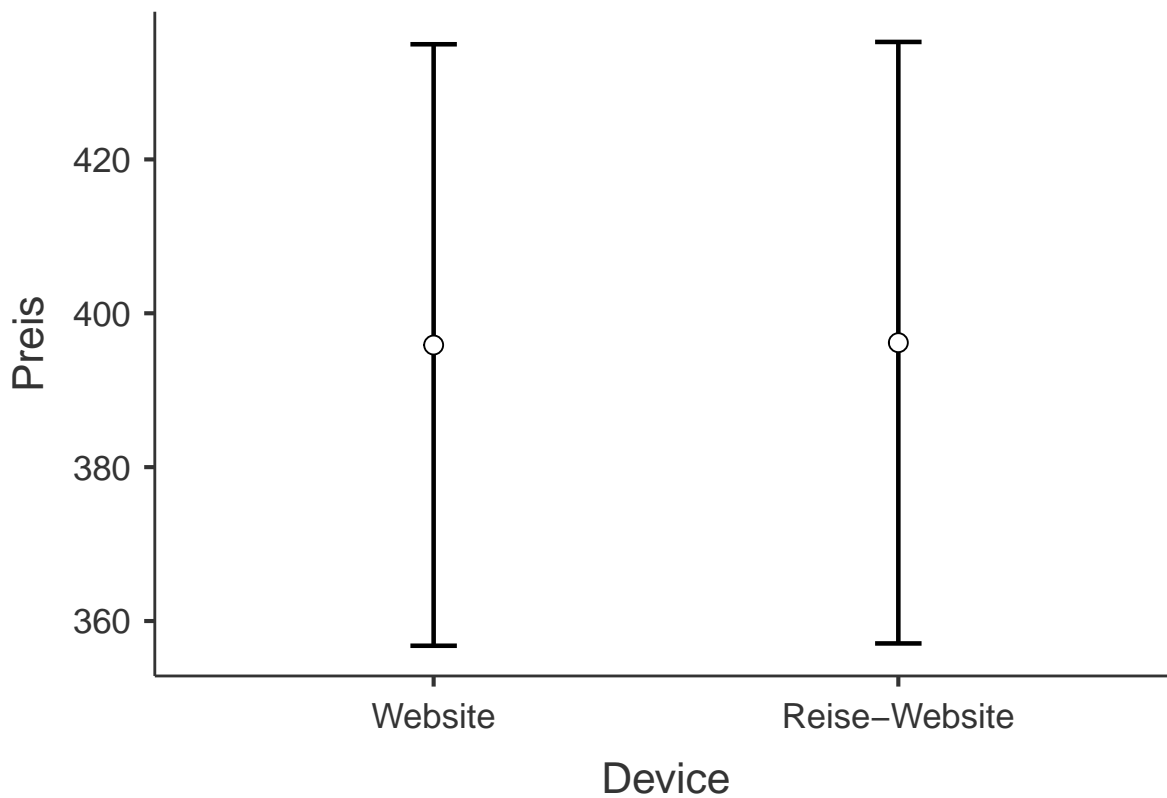
Einfaktorielle ANOVA:

```

##
## ANOVA
##

```

```
## ANOVA
## -----
##              Sum of Squares    df    Mean Square    F        p
## -----
## Device              12.9         1         12.9    1.16e-4    0.991
## Residuals        6.19e+7       558       110897.6
## -----
##
## ESTIMATED MARGINAL MEANS
##
## DEVICE
## Estimated Marginal Means - Device
## -----
## Device      Mean    SE    Lower    Upper
## -----
## Website      396    19.9    357     435
## Reise-Website 396    19.9    357     435
## -----
```



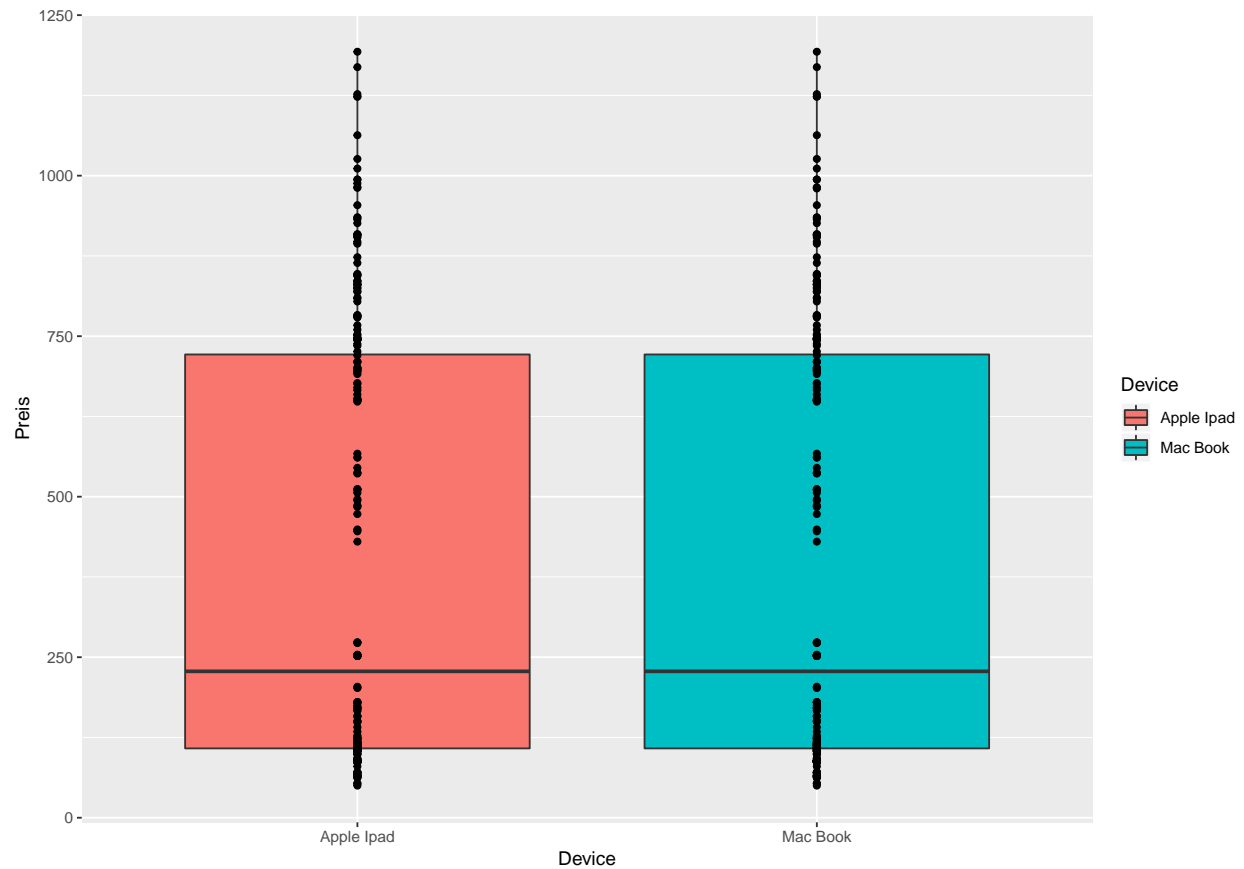
Hypothese 4

[Hypothese 4: Das Abrufen einer Reise-Website mittels Applikation und Website erwirkt einen Unterschied des offerierten Preises einer Airline.]

Deskreptive Statistik:

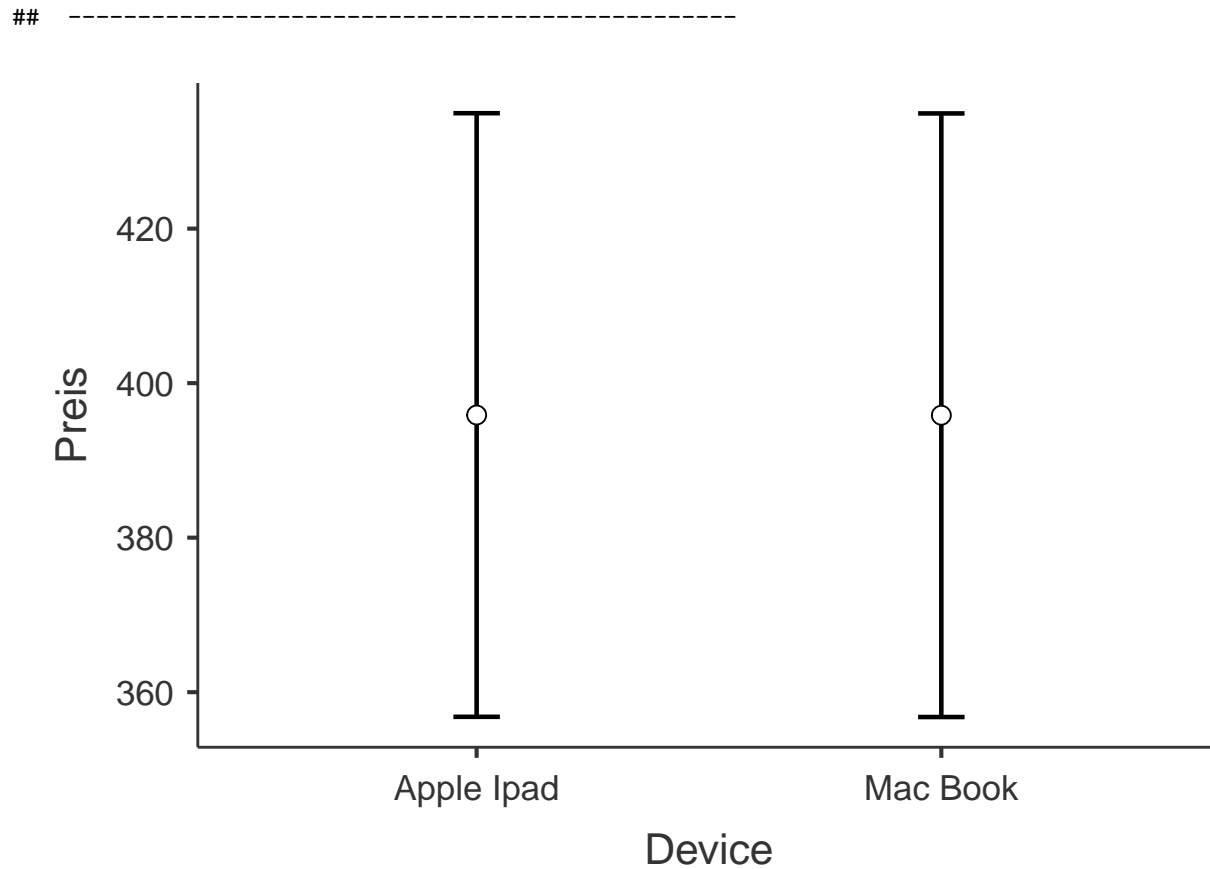
```
##
## DESCRIPTIVES
##
## Descriptives
## -----
##               Preis      Device
## -----
##      N              560      560
##      Missing          0        0
##      Mean            396
##      Median           228
##      Standard deviation 332
##      Minimum          50.0
##      Maximum         1193
## -----
##
##
## FREQUENCIES
##
## Frequencies of Device
## -----
##      Levels      Counts      % of Total      Cumulative %
## -----
##      Apple Ipad      280          50.0          50.0
##      Mac Book        280          50.0          100.0
## -----
```

Verteilung - BoxPlot



Einfaktorielle ANOVA:

```
##
## ANOVA
##
## ANOVA
## -----
##              Sum of Squares    df    Mean Square    F        p        <U+03B7>2p
## -----
## Device              0.114         1         0.114    1.03e-6    0.999    0.000
## Residuals        6.18e+7       558    110753.443
## -----
##
##
## ESTIMATED MARGINAL MEANS
##
## DEVICE
##
## Estimated Marginal Means - Device
## -----
## Device      Mean    SE    Lower    Upper
## -----
## Apple Ipad   396    19.9    357     435
## Mac Book     396    19.9    357     435
```



Hypothese 6

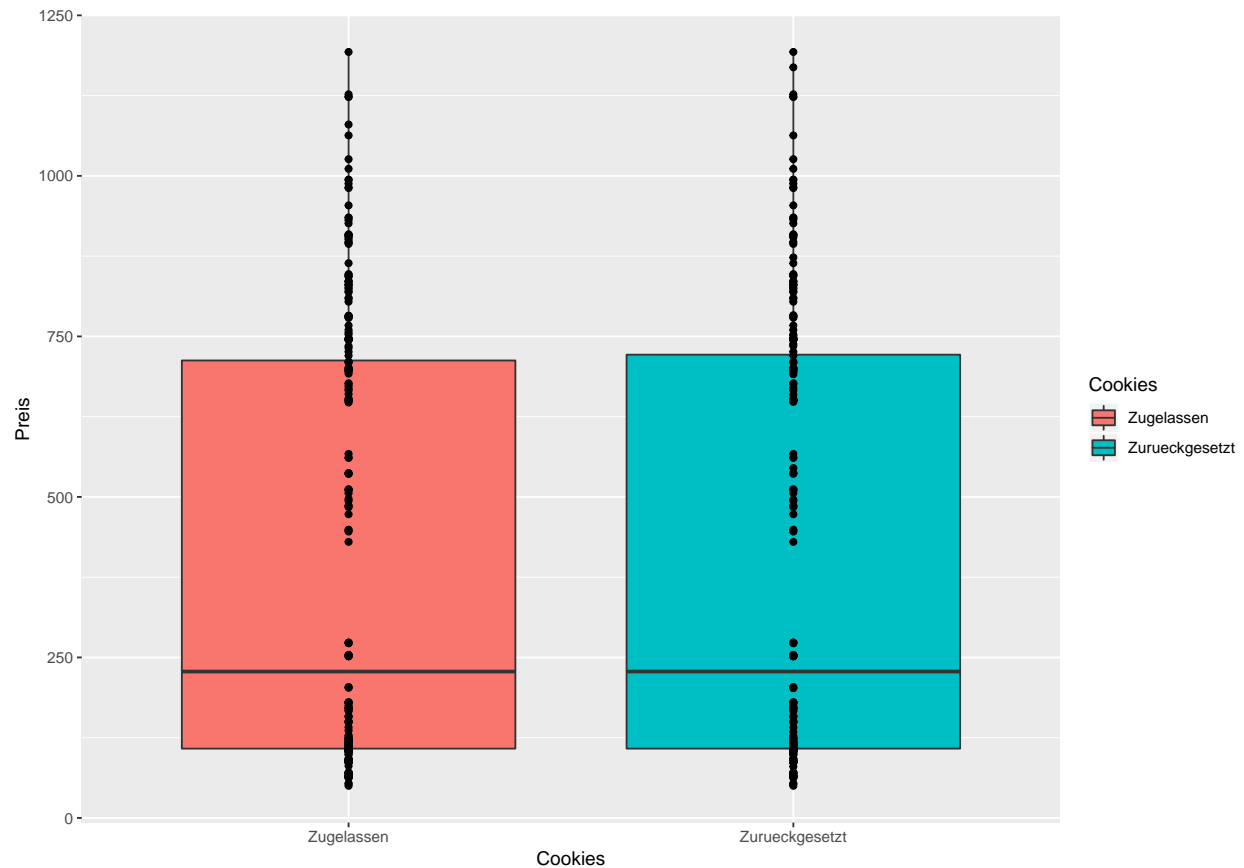
[Das Zurücksetzen von Cookies respektive dem Browserverlauf erwirkt ein Sinken des offerierten Preises einer Airline.]

Deskreptive Statistik:

```
##
## DESCRIPTIVES
##
## Descriptives
## -----
##               Preis      Cookies
## -----
## N                560        560
## Missing           0          0
## Mean             396
## Median           228
## Standard deviation 332
## Minimum          50.0
## Maximum          1193
## -----
##
##
## FREQUENCIES
```

```
##
## Frequencies of Cookies
## -----
##      Levels           Counts    % of Total    Cumulative %
## -----
##      Zugelassen           280         50.0         50.0
##      Zurueckgesetzt       280         50.0        100.0
## -----
```

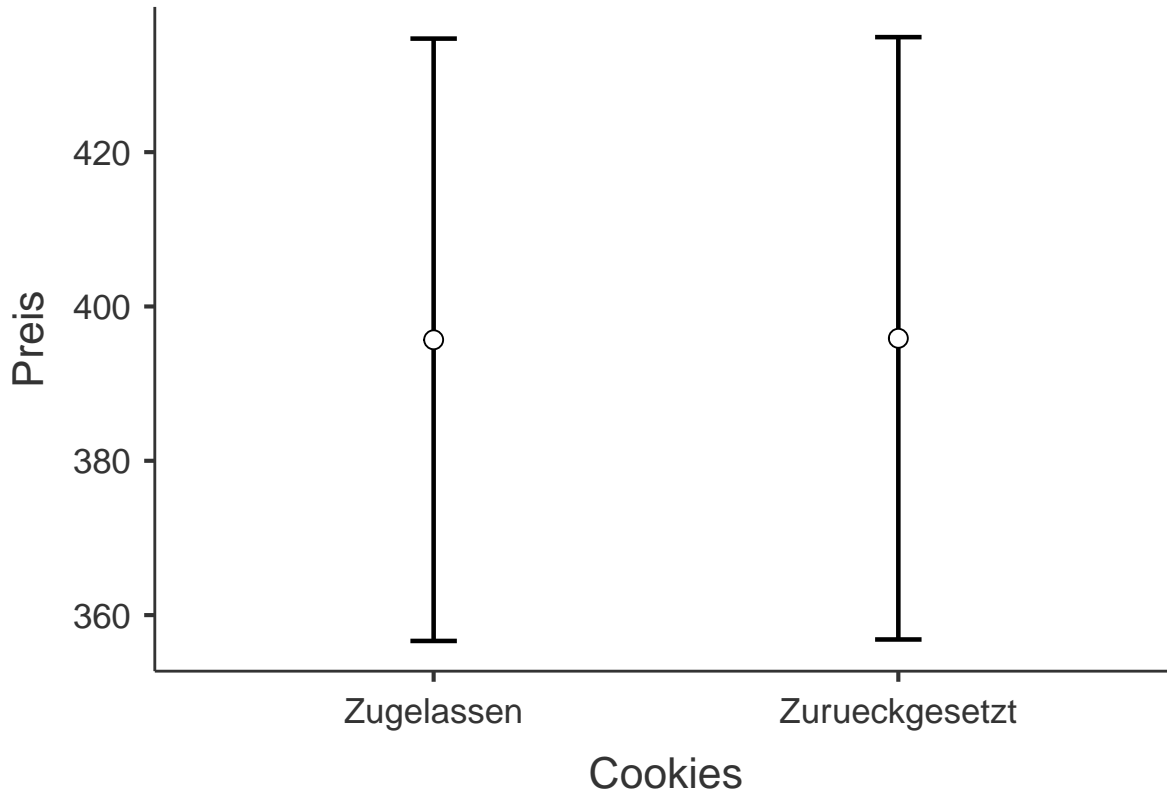
Verteilung - BoxPlot



Einfaktorielle ANOVA:

```
##
## ANOVA
##
## ANOVA
## -----
##      Sum of Squares    df    Mean Square    F        p      <U+03B7>2p
## -----
##      Cookies             5.21      1          5.21    4.71e-5    0.995    0.000
##      Residuals        6.17e+7    558        110573.63
## -----
##
##
## ESTIMATED MARGINAL MEANS
```

```
##
## COOKIES
##
## Estimated Marginal Means - Cookies
## -----
##      Cookies      Mean    SE      Lower    Upper
## -----
##      Zugelassen      396   19.9      357      435
##      Zurueckgesetzt  396   19.9      357      435
## -----
```



Hypothese 7

[Das Verbergen der Internetprotokoll-Adresse und folglich der ortsspezifischen Parameter mittels Virtual Private Network verursacht eine Differenz im offerierten Preis einer Airline.]

Deskreptive Statistik:

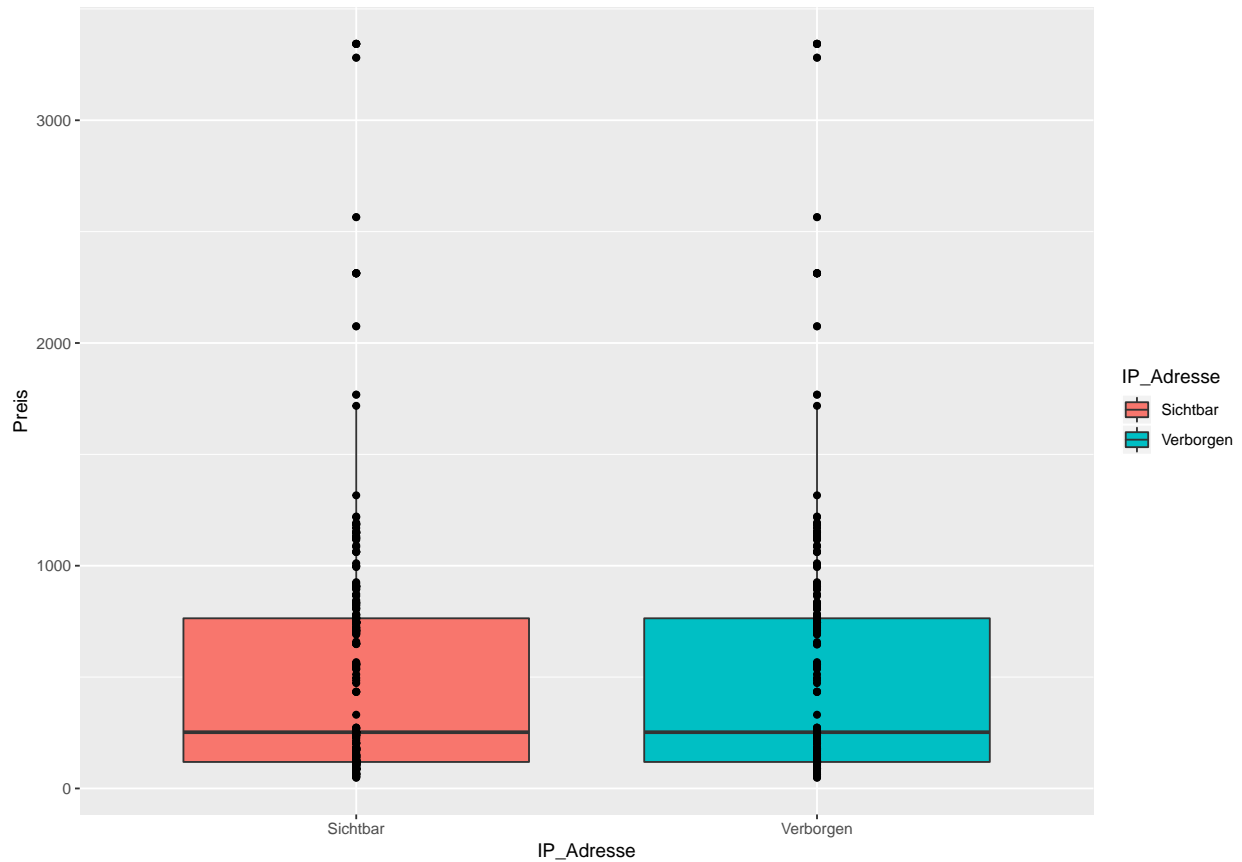
```
##
## DESCRIPTIVES
##
## Descriptives
## -----
##      Preis      IP_Adresse
## -----
##      N          560          560
##      Missing      0           0
```

```

##      Mean                532
##      Median              252
##      Standard deviation   573
##      Minimum             48.0
##      Maximum             3343
##      -----
##
##
##
## FREQUENCIES
##
## Frequencies of IP_Adresse
## -----
##      Levels      Counts    % of Total    Cumulative %
## -----
##      Sichtbar      280        50.0         50.0
##      Verborgen     280        50.0        100.0
## -----

```

Verteilung - BoxPlot



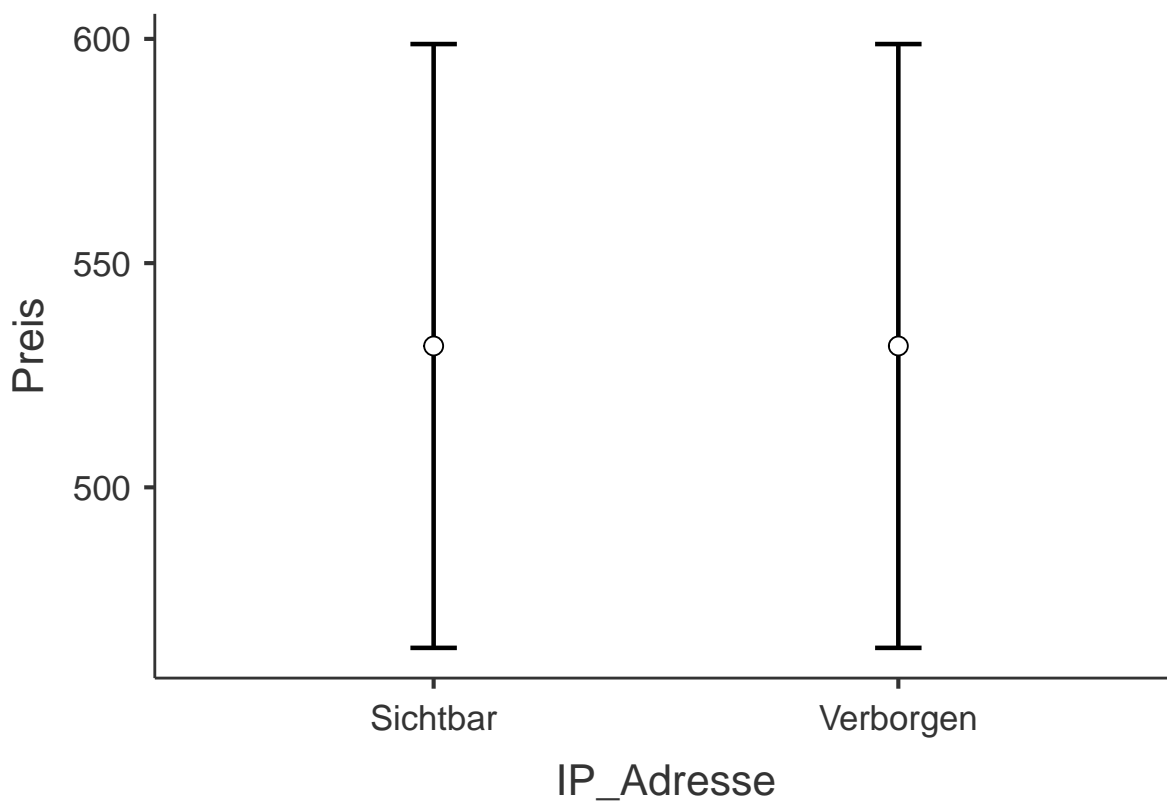
Einfaktorielle ANOVA:

```

##
## ANOVA
##
## ANOVA

```

```
## -----
##               Sum of Squares    df    Mean Square    F          p          <U+03B7>²p
## -----
##   IP_Adresse           0.00        1           0.00    2.21e-30    1.000    0.000
##   Residuals          1.84e+8      558      328963.14
## -----
##
##
## ESTIMATED MARGINAL MEANS
##
## IP_ADRESSE
##
## Estimated Marginal Means - IP_Adresse
## -----
##   IP_Adresse    Mean    SE      Lower    Upper
## -----
##   Sichtbar      532     34.3     464      599
##   Verborgen     532     34.3     464      599
## -----
```



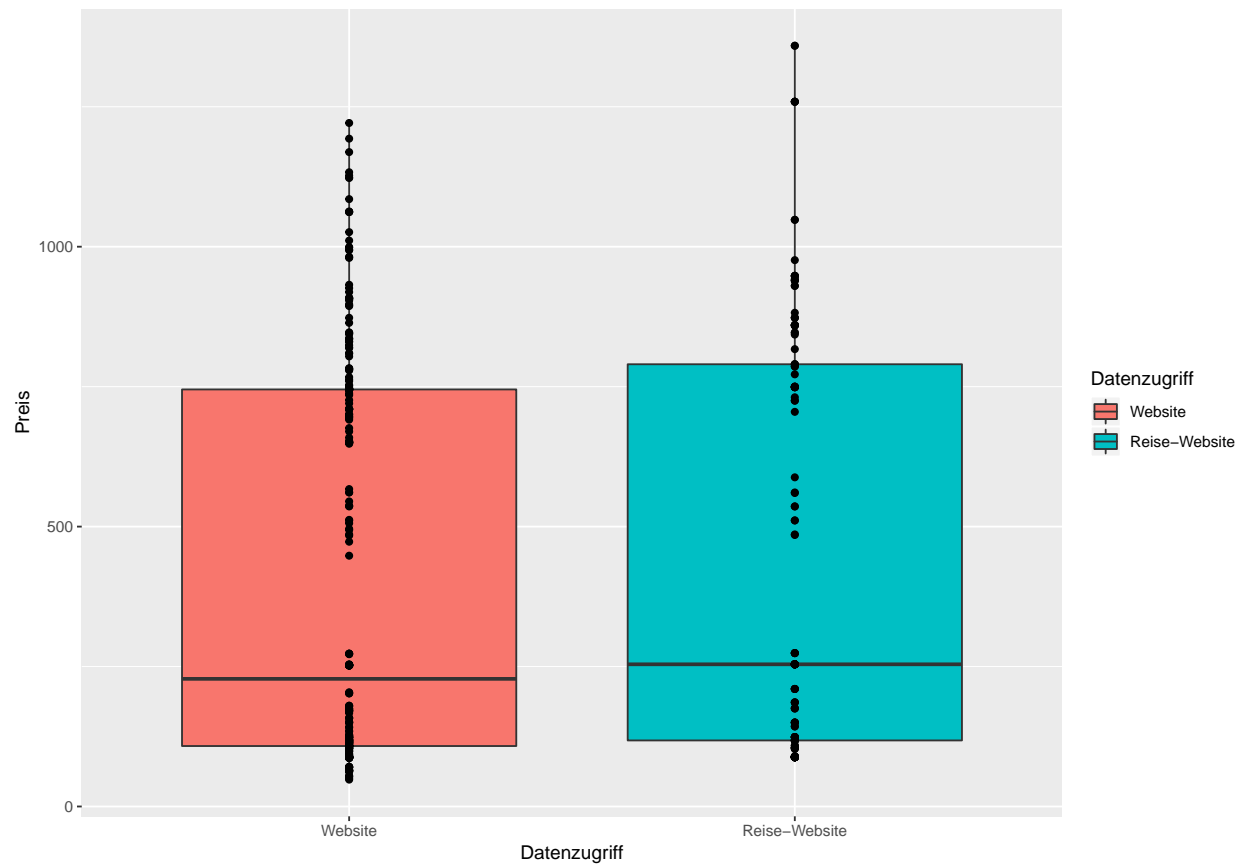
Hypothese 8

[Das Abrufen eines Flugpreises via Reise-Website führt, verglichen mit der Website der Airline selbst, zu einem höheren offerierten Preis.]

Deskreptive Statistik:

```
##
## DESCRIPTIVES
##
## Descriptives
## -----
##               Preis      Datenzugriff
## -----
##      N              420              420
##      Missing          0              0
##      Mean            425
##      Median          252
##      Standard deviation 356
##      Minimum         48.0
##      Maximum        1359
## -----
##
##
## FREQUENCIES
##
## Frequencies of Datenzugriff
## -----
##      Levels      Counts      % of Total      Cumulative %
## -----
##      Website      280          66.7          66.7
##      Reise-Website 140          33.3          100.0
## -----
```


Verteilung - BoxPlot



Einfaktorielle ANOVA:

```
##
## ANOVA
##
## ANOVA
## -----
##              Sum of Squares    df    Mean Square    F    p    <U+03B7>2p
## -----
##   Datenzugriff           135027     1       135027    1.06  0.303  0.003
##   Residuals           5.30e+7   418       126793
## -----
##
##
## ESTIMATED MARGINAL MEANS
##
## DATENZUGRIFF
##
## Estimated Marginal Means - Datenzugriff
## -----
##   Datenzugriff    Mean    SE    Lower    Upper
## -----
##   Website         412    21.3    370    454
##   Reise-Website   450    30.1    391    509
```

##

