Silas_Analysis

Max Kuttner

25 3 2020

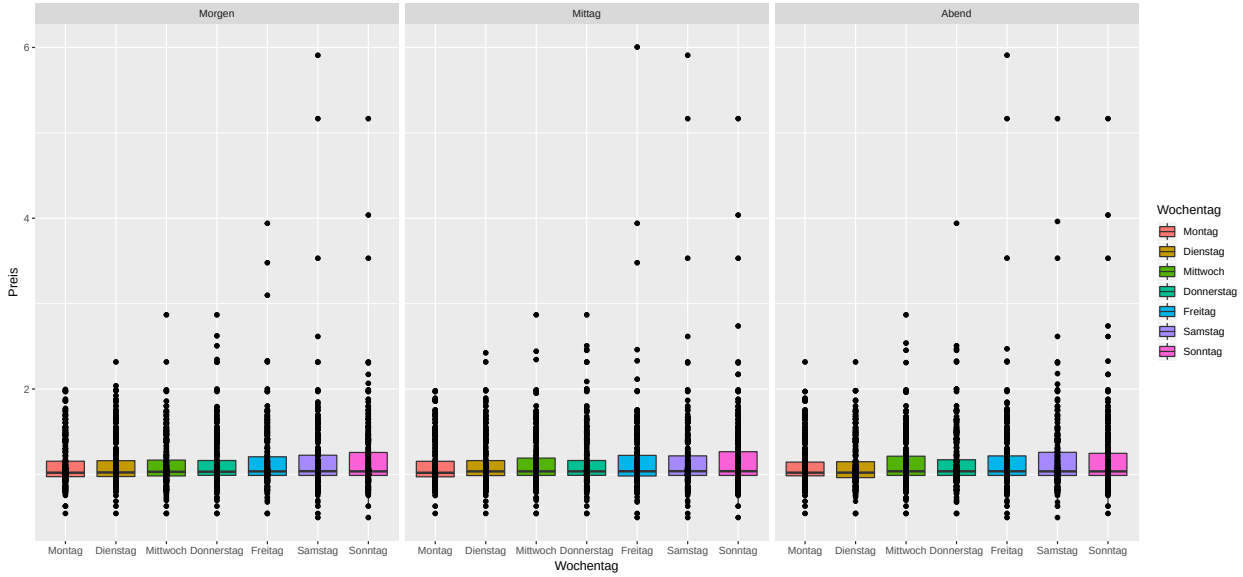
Hypothese 1

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

Deskreptive Statistik:

```
##
##  DESCRIPTIVES
##
##  Descriptives
##  -------------------------------------------------
##                          Preis    Wochentag    Zeit
##  -------------------------------------------------
##    N                      7224         7224    7224
##    Missing                   0            0       0
##    Mean                   1.13
##    Median                 1.04
##    Standard deviation    0.346
##    Minimum               0.495
##    Maximum                6.00
##  -------------------------------------------------
##
##
##  FREQUENCIES
##
##  Frequencies of Wochentag
##  ---------------------------------------------------------
##    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
##  -------------------------------------------------
```

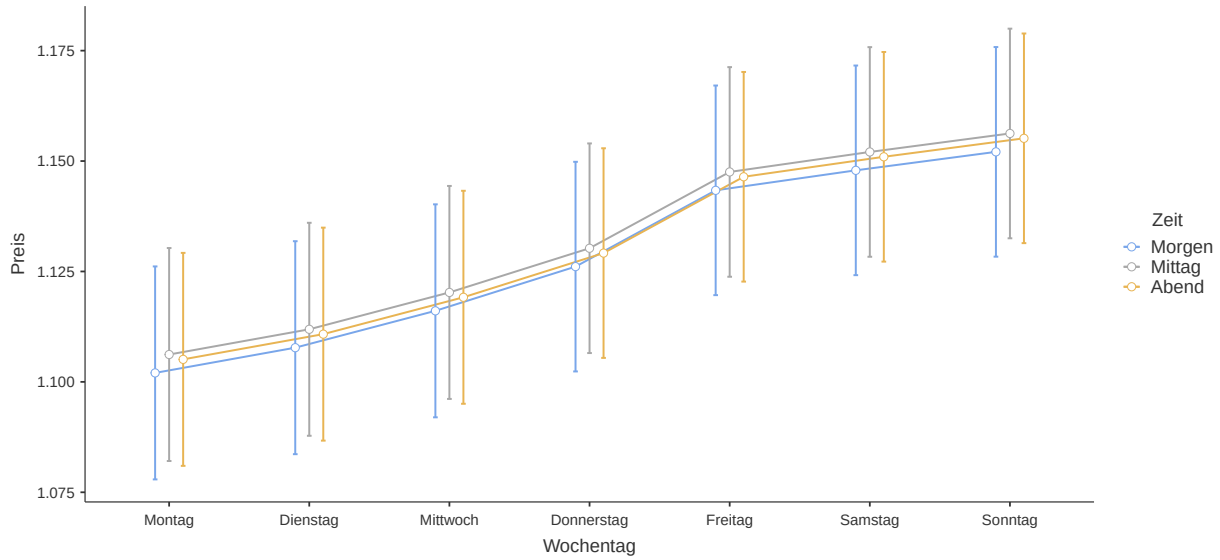
Verteilung - BoxPlot



Mehrfakorielle ANOVA:

```
##
##  ANOVA
##
##  ANOVA
##  ------------------------------------------------------------------------------
##                 Sum of Squares    df      Mean Square    F         p         <U+03B7>²p
##  ------------------------------------------------------------------------------
##    Wochentag            2.5126     6           0.4188    3.5122    0.002     0.003
##    Zeit                 0.0224     2           0.0112    0.0941    0.910     0.000
##    Residuals          860.2688  7215           0.1192
##  ------------------------------------------------------------------------------
##
##
##  ESTIMATED MARGINAL MEANS
##
##  WOCHENTAG:ZEIT
##
##  Estimated Marginal Means - Wochentag:Zeit
##  ----------------------------------------------------------------
##    Zeit      Wochentag     Mean    SE        Lower    Upper
##  ----------------------------------------------------------------
##    Morgen    Montag        1.10    0.0123     1.08     1.13
##              Dienstag      1.11    0.0123     1.08     1.13
##              Mittwoch      1.12    0.0123     1.09     1.14
##              Donnerstag    1.13    0.0121     1.10     1.15
##              Freitag       1.14    0.0121     1.12     1.17
##              Samstag       1.15    0.0121     1.12     1.17
##              Sonntag       1.15    0.0121     1.13     1.18
##    Mittag    Montag        1.11    0.0123     1.08     1.13
##              Dienstag      1.11    0.0123     1.09     1.14
##              Mittwoch      1.12    0.0123     1.10     1.14
```

```
##              Donnerstag    1.13    0.0121     1.11     1.15
##              Freitag       1.15    0.0121     1.12     1.17
##              Samstag       1.15    0.0121     1.13     1.18
##              Sonntag       1.16    0.0121     1.13     1.18
##    Abend     Montag        1.11    0.0123     1.08     1.13
##              Dienstag      1.11    0.0123     1.09     1.13
##              Mittwoch      1.12    0.0123     1.10     1.14
##              Donnerstag    1.13    0.0121     1.11     1.15
##              Freitag       1.15    0.0121     1.12     1.17
##              Samstag       1.15    0.0121     1.13     1.17
##              Sonntag       1.16    0.0121     1.13     1.18
##  ------------------------------------------------------------
```



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   Wochentag   mean     se lower upper
##    <fct>  <fct>      <dbl>  <dbl> <dbl> <dbl>
##  1 Morgen Montag      1.10 0.0123  1.08  1.13
##  2 Abend  Montag      1.11 0.0123  1.08  1.13
##  3 Mittag Montag      1.11 0.0123  1.08  1.13
##  4 Morgen Dienstag    1.11 0.0123  1.08  1.13
##  5 Abend  Dienstag    1.11 0.0123  1.09  1.13
##  6 Mittag Dienstag    1.11 0.0123  1.09  1.14
##  7 Morgen Mittwoch    1.12 0.0123  1.09  1.14
##  8 Abend  Mittwoch    1.12 0.0123  1.10  1.14
##  9 Mittag Mittwoch    1.12 0.0123  1.10  1.14
## 10 Morgen Donnerstag  1.13 0.0121  1.10  1.15
## # ... with 11 more rows
```

Hypothese 2

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

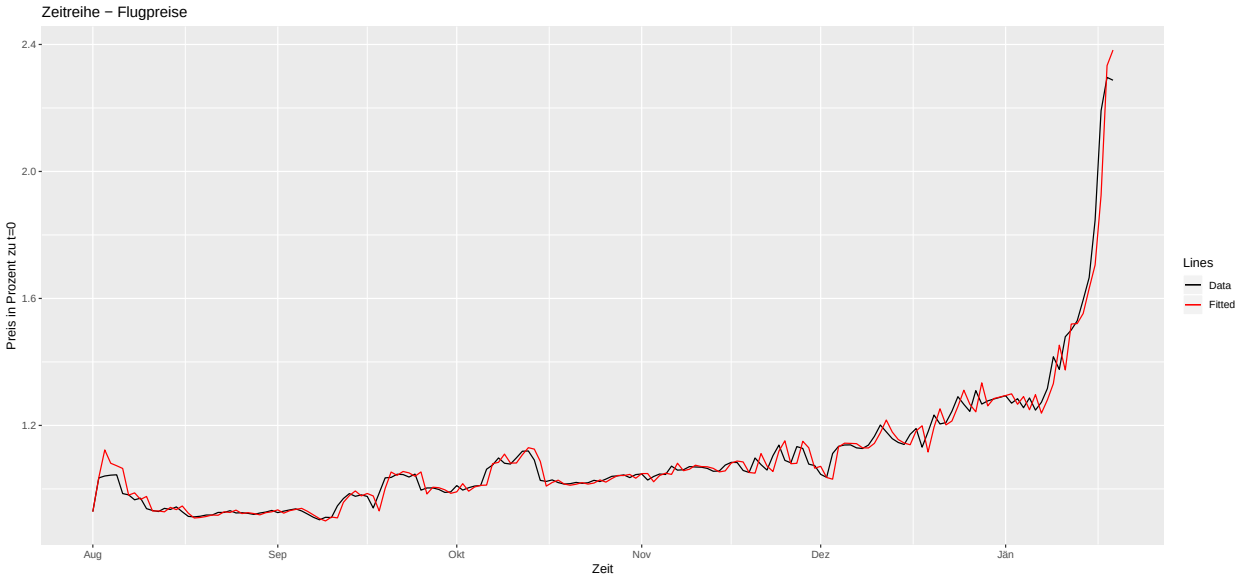
```
## New names:
## * `` -> ...1
## * `` -> ...5
```

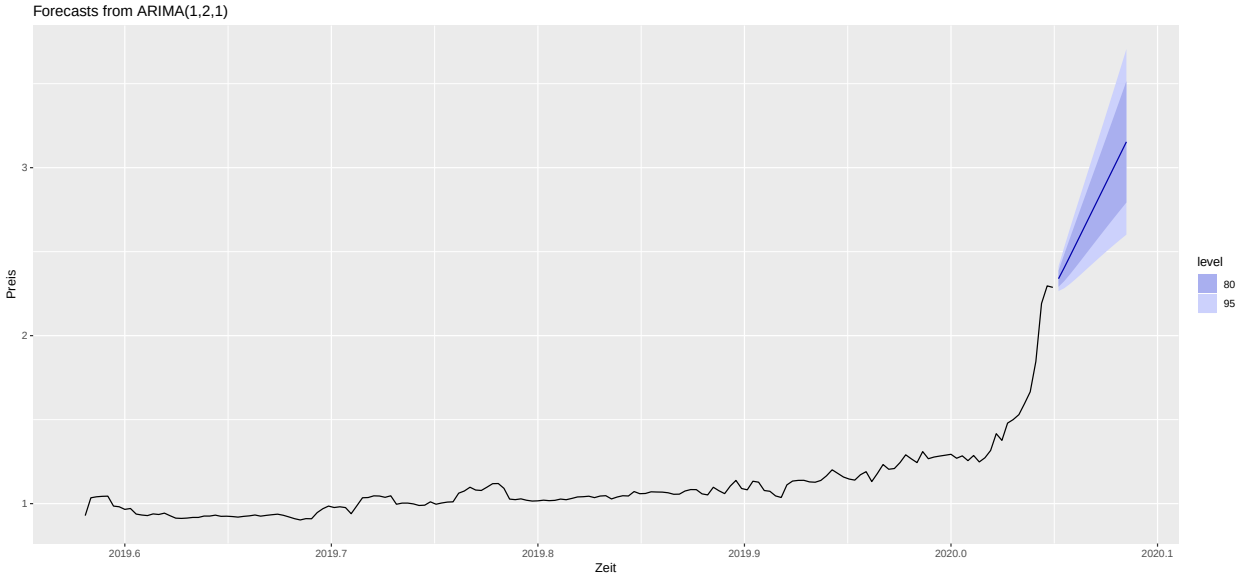
Diese Zeitreihe lässt sich mit folgendem Modell modellieren.

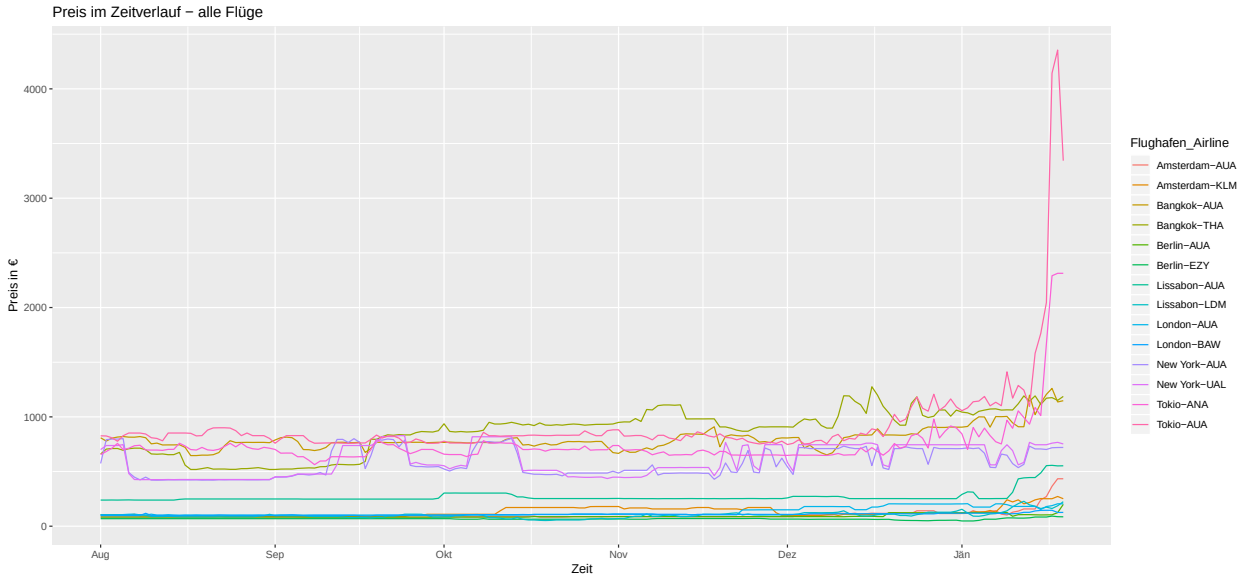
```
## Series: myts
## ARIMA(1,2,1)
##
## Coefficients:
##          ar1      ma1
##       0.2165  -0.8956
## s.e.  0.1100   0.0580
##
## sigma^2 estimated as 0.001404:  log likelihood=317.47
## AIC=-628.94   AICc=-628.79   BIC=-619.53
##
## Training set error measures:
##                       ME       RMSE       MAE         MPE    MAPE MASE
## Training set 0.001348106 0.03703545 0.0216493 0.005775143 1.79834  NaN
##                    ACF1
## Training set 0.0445453
```

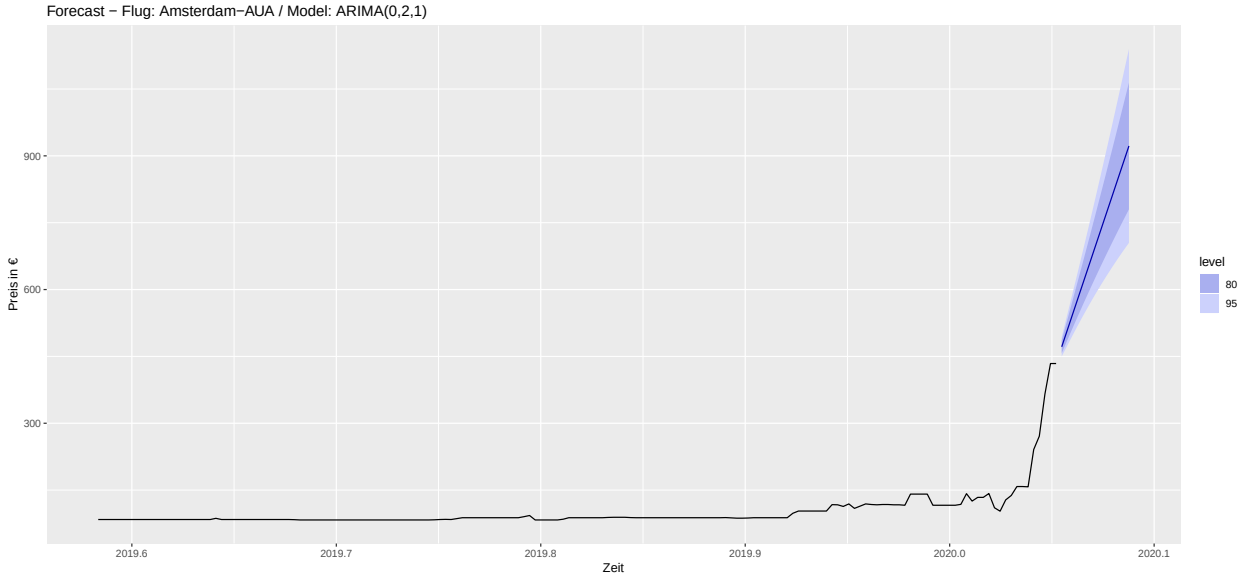
```
##
## z test of coefficients:
##
##      Estimate Std. Error z value Pr(>|z|)
## ar1  0.216510   0.109957   1.969  0.04895 *
## ma1 -0.895598   0.058046 -15.429  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

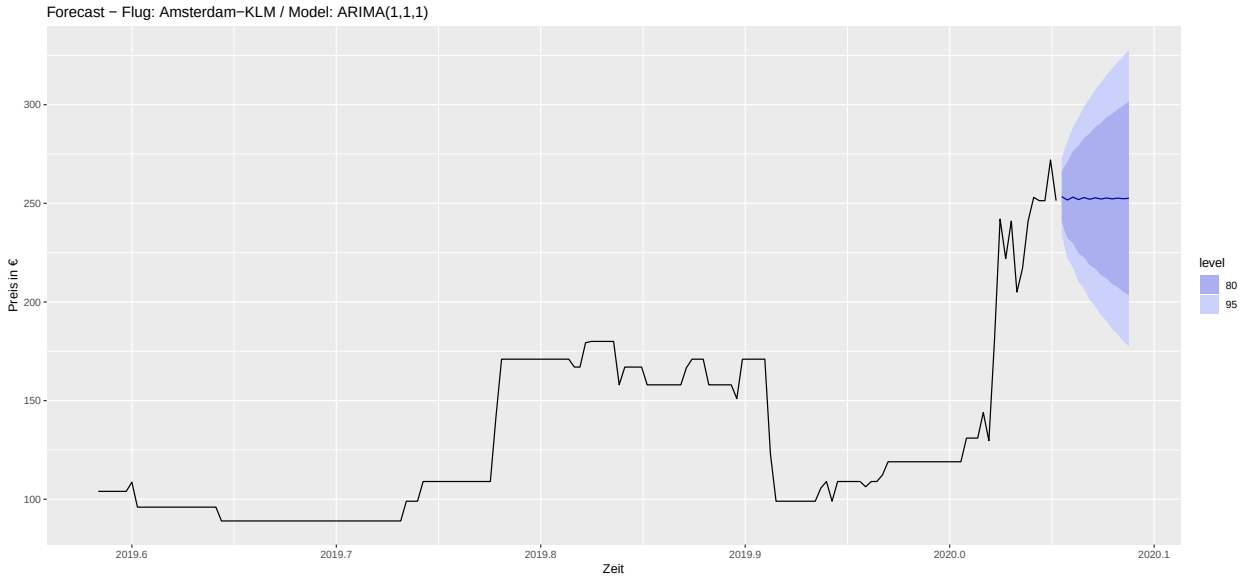
Mit diesem Modell kann in weiterer Folge die Zeitrehe angenähert werden und bis zum 1.Februar 2020 vorhergesagt werden.

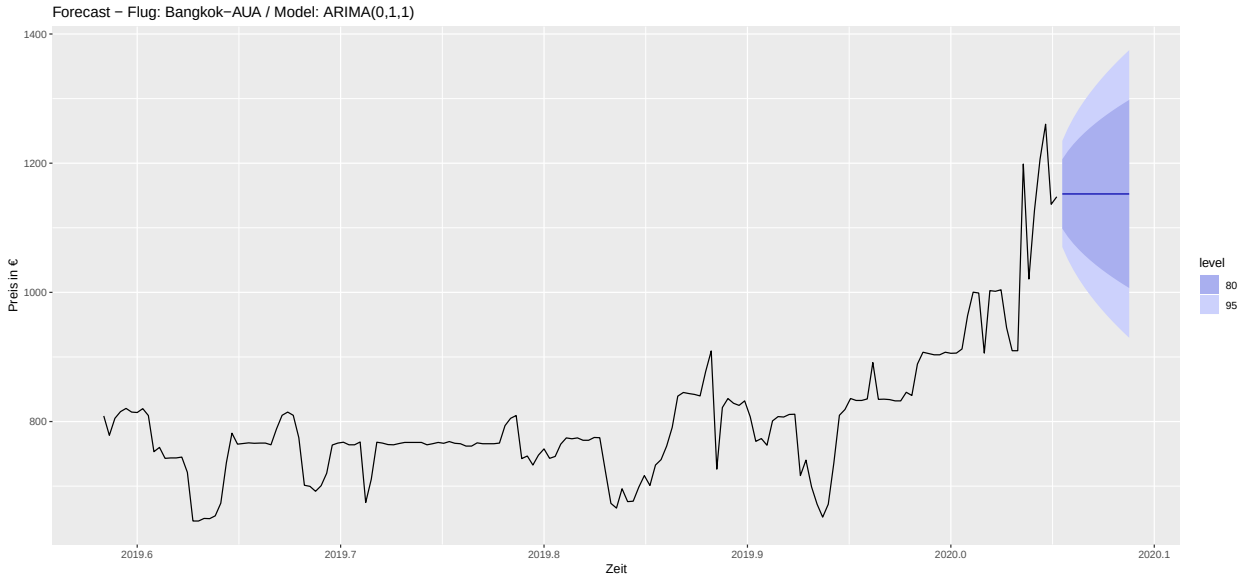

Zeitreihe − Flugpreise
Preis in Prozent zu t=0
2.4
2.0
1.6
1.2
Aug
Sep
Okt
Nov
Dez
Jän
Zeit
Lines
Data
Fitted

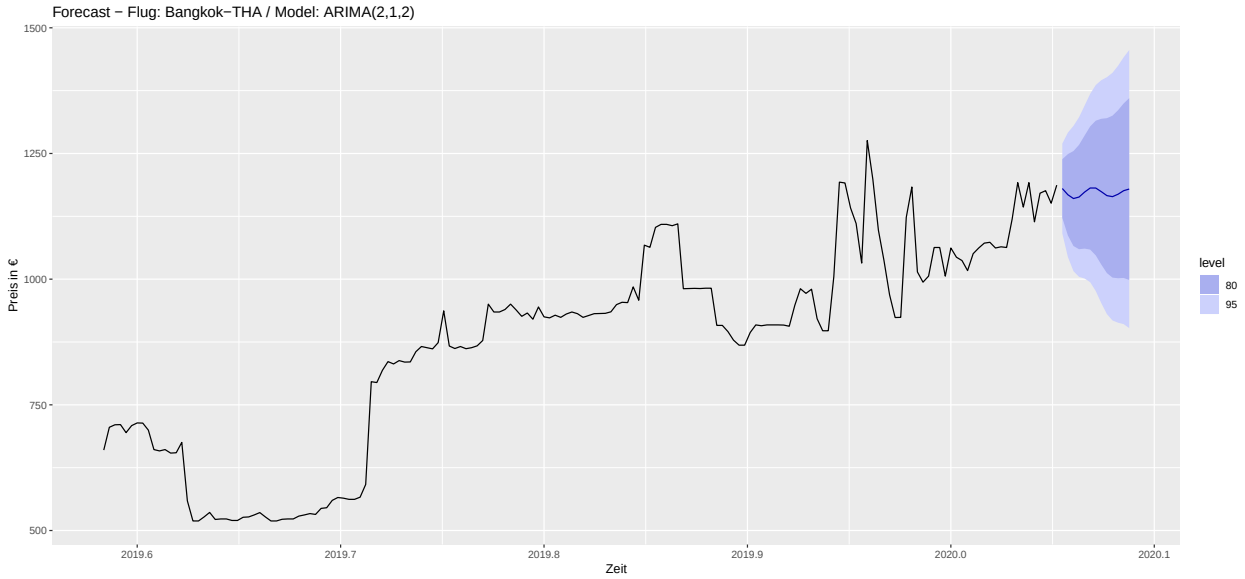

Forecasts from ARIMA(1,2,1)
Preis
3
2
1
2019.6
2019.7
2019.8
2019.9
2020.0
2020.1
Zeit
level
80
95

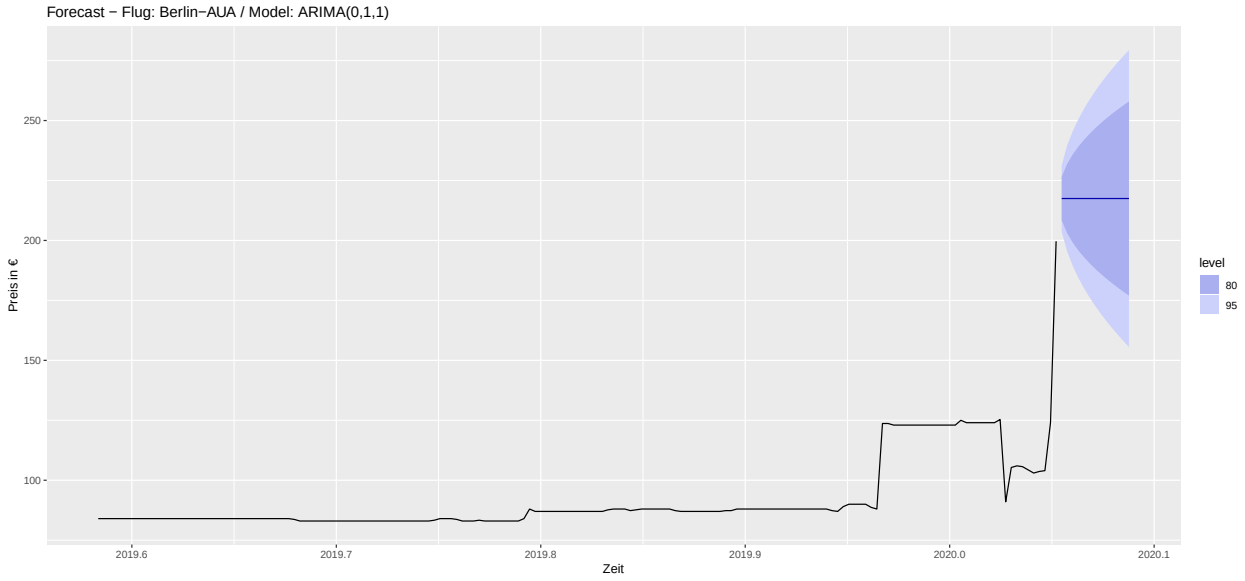

Preis im Zeitverlauf − alle Flüge
Preis in €
4000
3000
2000
1000
0
Aug
Sep
Okt
Nov
Dez
Jän
Zeit
Flughafen_Airline
Amsterdam−AUA
Amsterdam−KLM
Bangkok−AUA
Bangkok−THA
Berlin−AUA
Berlin−EZY
Lissabon−AUA
Lissabon−LDM
London−AUA
London−BAW
New York−AUA
New York−UAL
Tokio−ANA
Tokio−AUA

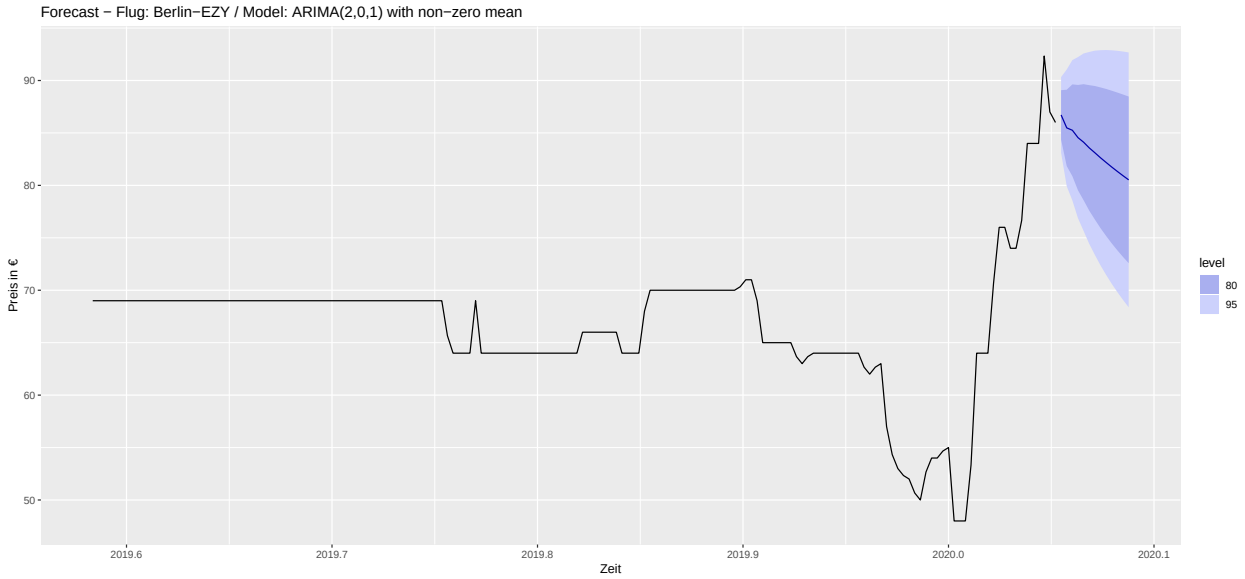

Forecast − Flug: Amsterdam−AUA / Model: ARIMA(0,2,1)
Preis in €
900
600
300
2019.6
2019.7
2019.8
2019.9
2020.0
2020.1
Zeit
level
80
95

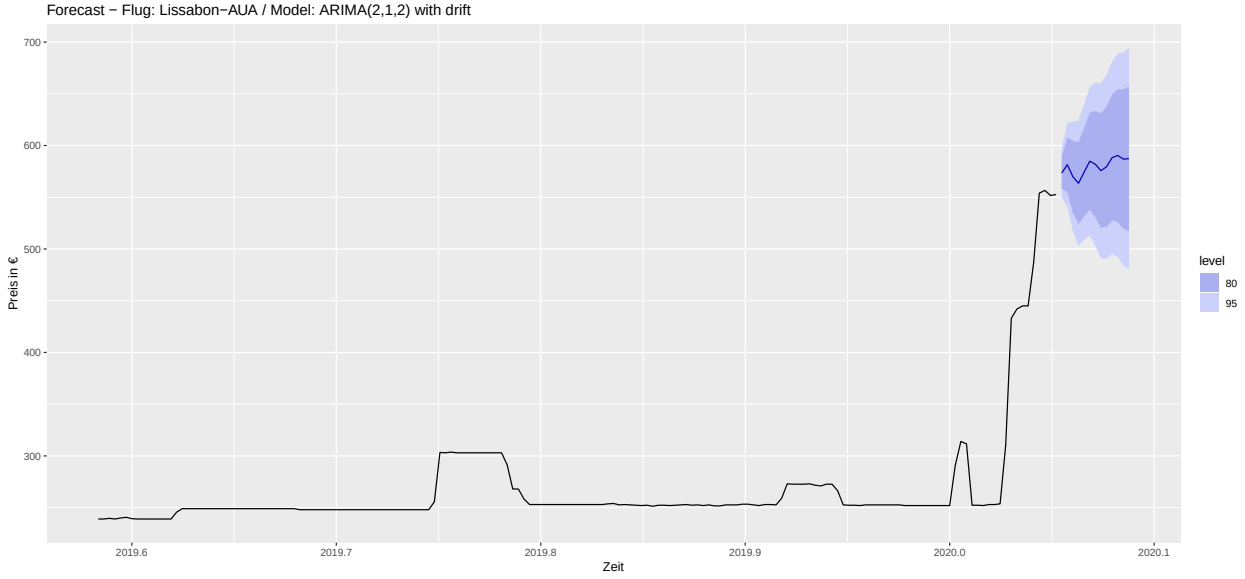

Forecast − Flug: Amsterdam−KLM / Model: ARIMA(1,1,1)
Preis in €
300
250
200
150
100
2019.6
2019.7
2019.8
2019.9
2020.0
2020.1
Zeit
level
80
95

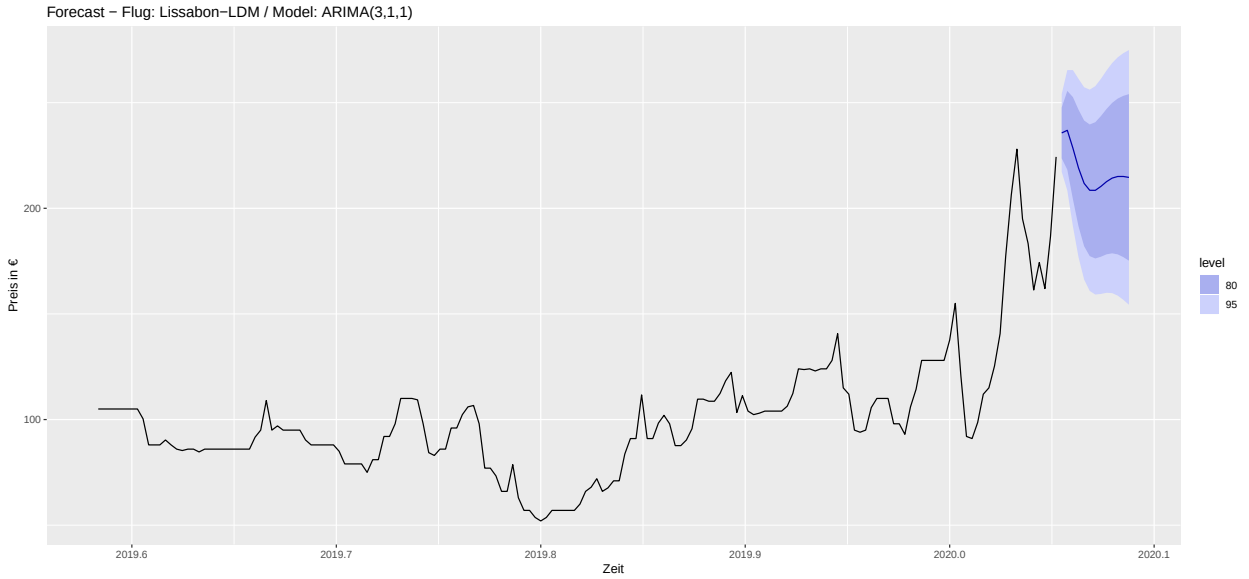

Forecast − Flug: Bangkok−AUA / Model: ARIMA(0,1,1)
Preis in €
1400
1200
1000
800
2019.6
2019.7
2019.8
2019.9
2020.0
2020.1
Zeit
level
80
95

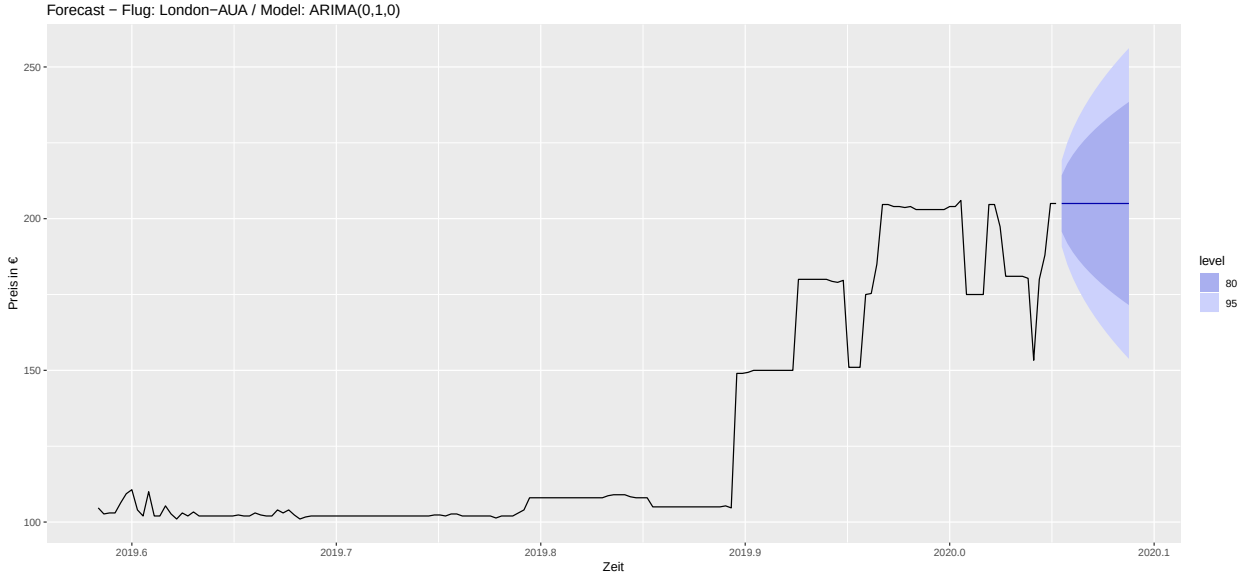

Forecast − Flug: Bangkok−THA / Model: ARIMA(2,1,2)
Preis in €
1500
1250
1000
750
500
2019.6
2019.7
2019.8
2019.9
2020.0
2020.1
Zeit
level
80
95

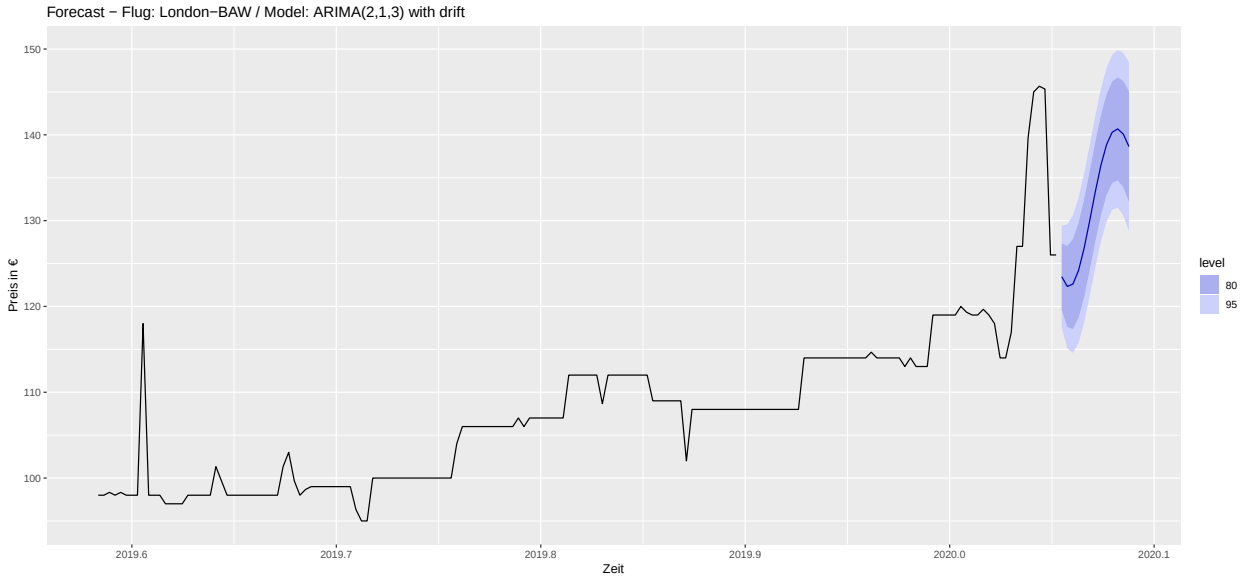

Forecast − Flug: Berlin−AUA / Model: ARIMA(0,1,1)
Preis in €
250
200
150
100
2019.6
2019.7
2019.8
2019.9
2020.0
2020.1
Zeit
level
80
95

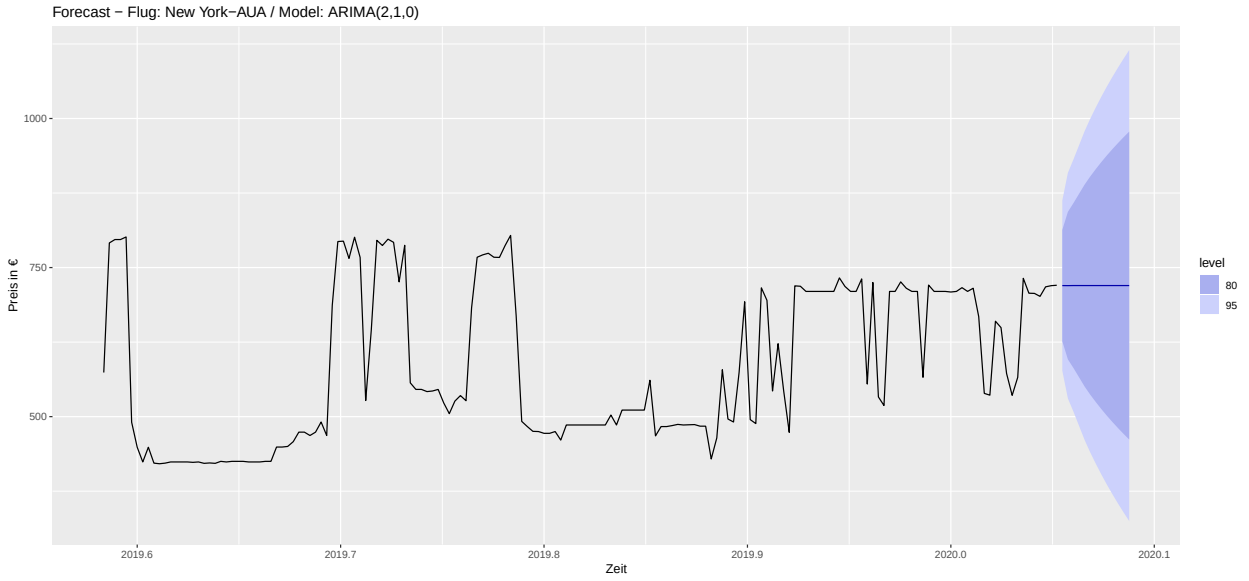

Forecast − Flug: Berlin−EZY / Model: ARIMA(2,0,1) with non−zero mean
Preis in €
90
80
70
60
50
2019.6
2019.7
2019.8
2019.9
2020.0
2020.1
Zeit
level
80
95

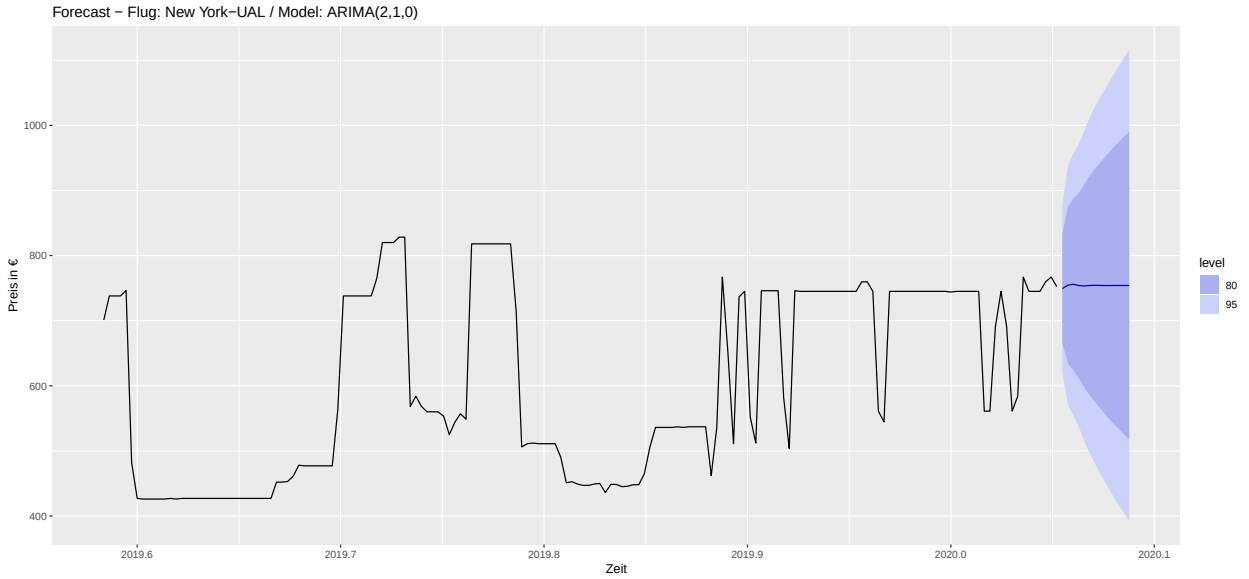

Forecast − Flug: Lissabon−AUA / Model: ARIMA(2,1,2) with drift
Preis in €
700
600
500
400
300
2019.6
2019.7
2019.8
2019.9
2020.0
2020.1
Zeit
level
80
95

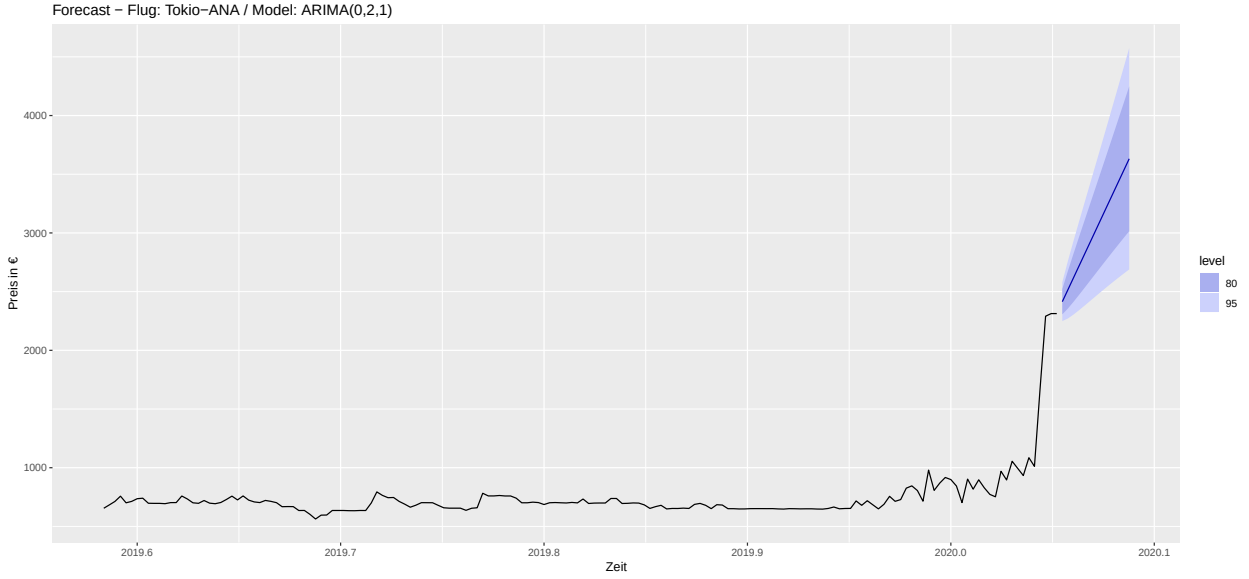

Forecast − Flug: Lissabon−LDM / Model: ARIMA(3,1,1)
Preis in €
200
100
2019.6
2019.7
2019.8
2019.9
2020.0
2020.1
Zeit
level
80
95

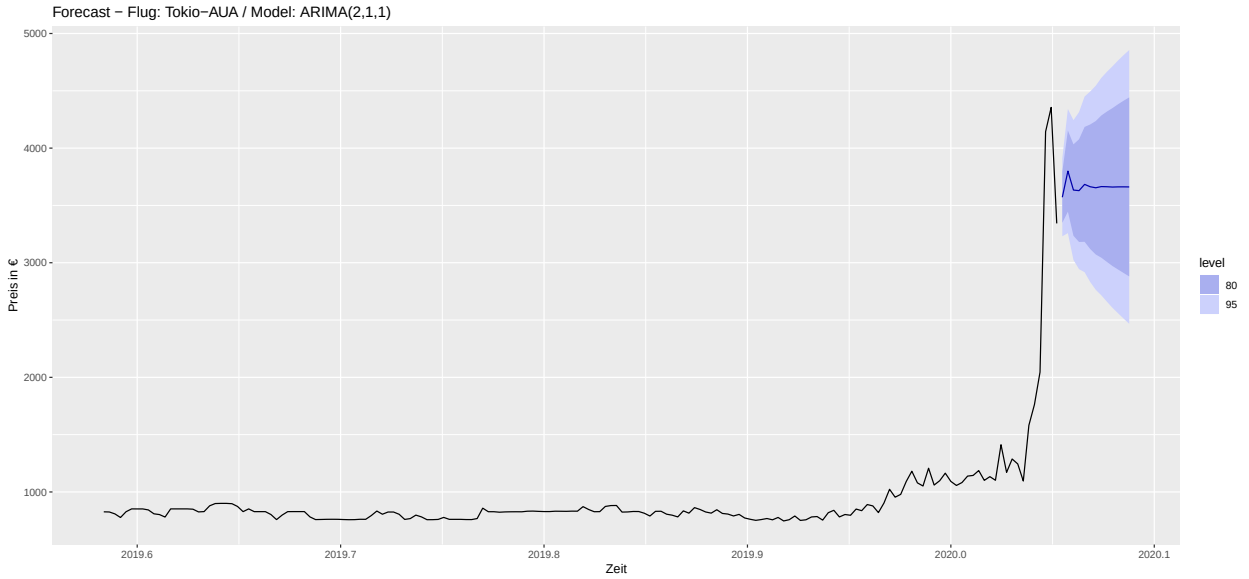

Forecast − Flug: London−AUA / Model: ARIMA(0,1,0)
Preis in €
250
200
150
100
2019.6
2019.7
2019.8
2019.9
2020.0
2020.1
Zeit
level
80
95

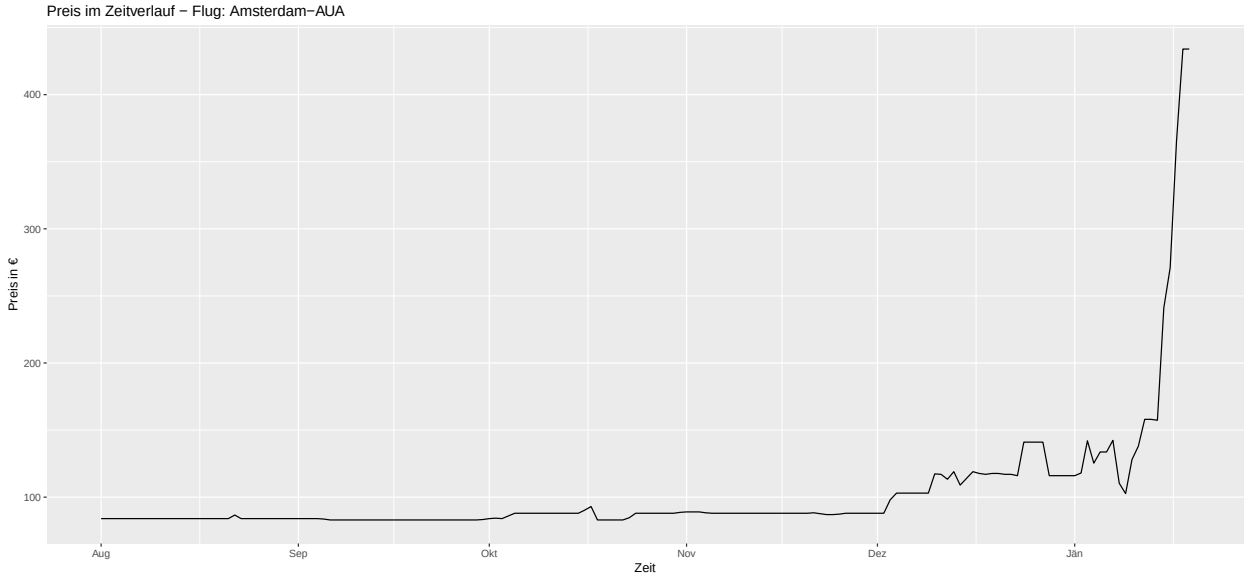

Forecast − Flug: London−BAW / Model: ARIMA(2,1,3) with drift
Preis in €
150
140
130
120
110
100
2019.6
2019.7
2019.8
2019.9
2020.0
2020.1
Zeit
level
80
95

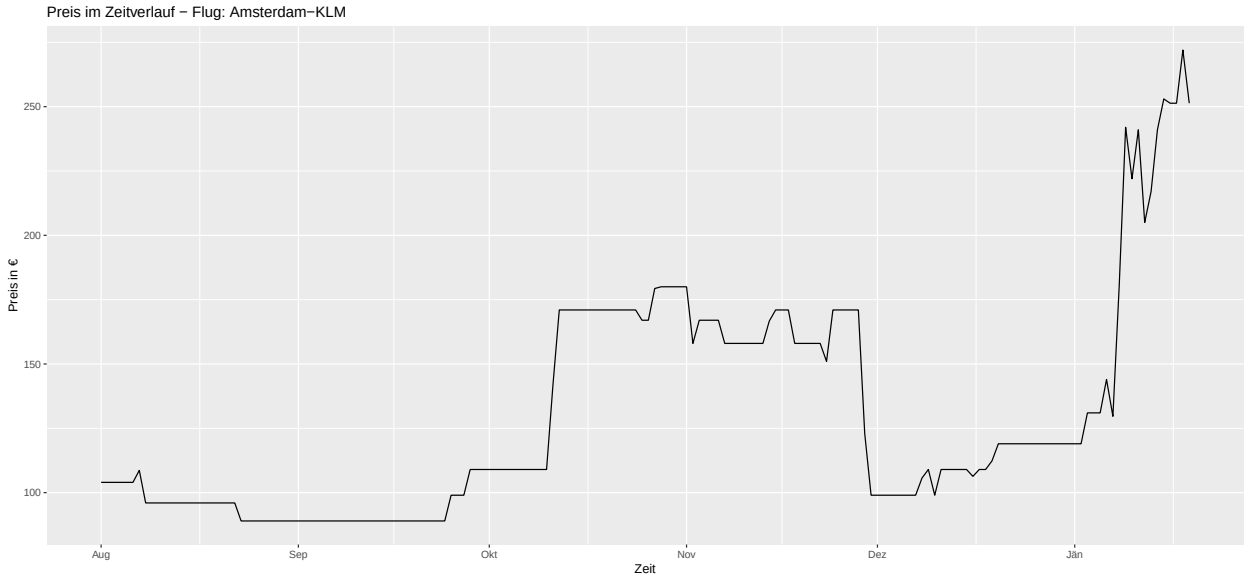

Forecast − Flug: New York−AUA / Model: ARIMA(2,1,0)
Preis in €
1000
750
500
2019.6
2019.7
2019.8
2019.9
2020.0
2020.1
Zeit
level
80
95

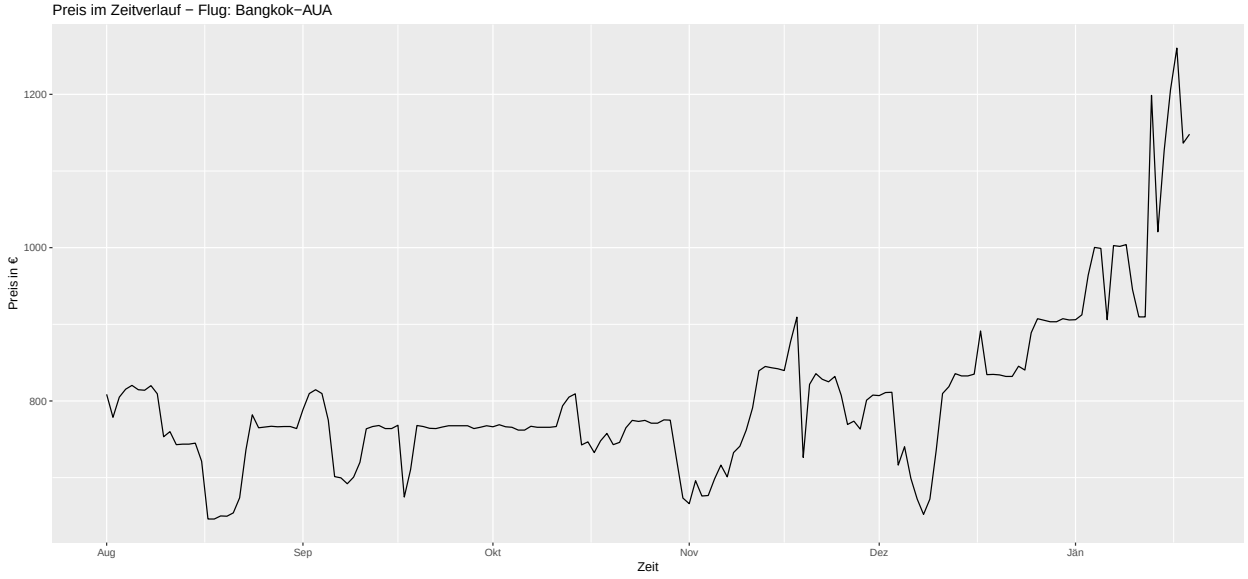

Forecast − Flug: New York−UAL / Model: ARIMA(2,1,0)
Preis in €
1000
800
600
400
2019.6
2019.7
2019.8
2019.9
2020.0
2020.1
Zeit
level
80
95

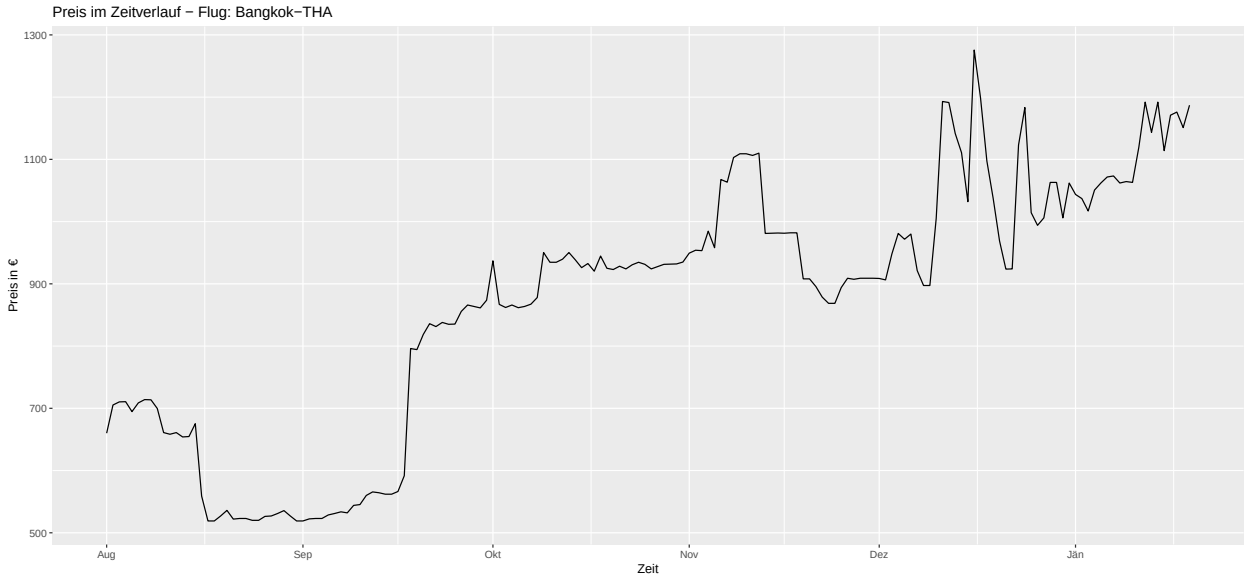

Forecast − Flug: Tokio−ANA / Model: ARIMA(0,2,1)
Preis in €
4000
3000
2000
1000
2019.6
2019.7
2019.8
2019.9
2020.0
2020.1
Zeit
level
80
95

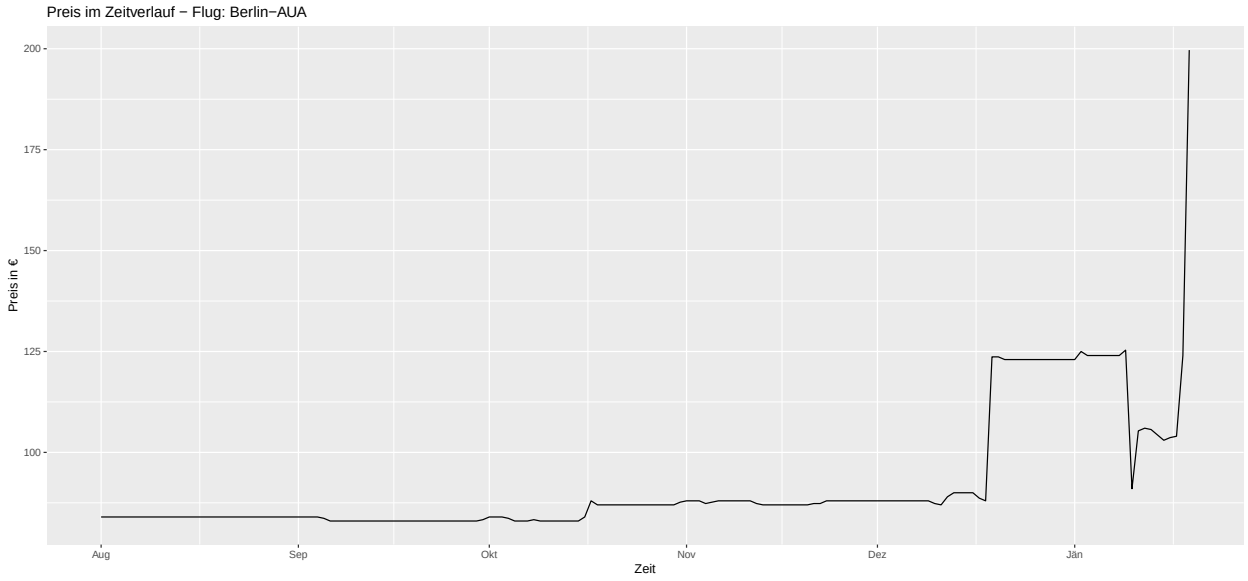

Forecast − Flug: Tokio−AUA / Model: ARIMA(2,1,1)
Preis in €
5000
4000
3000
2000
1000
2019.6
2019.7
2019.8
2019.9
2020.0
2020.1
Zeit
level
80
95

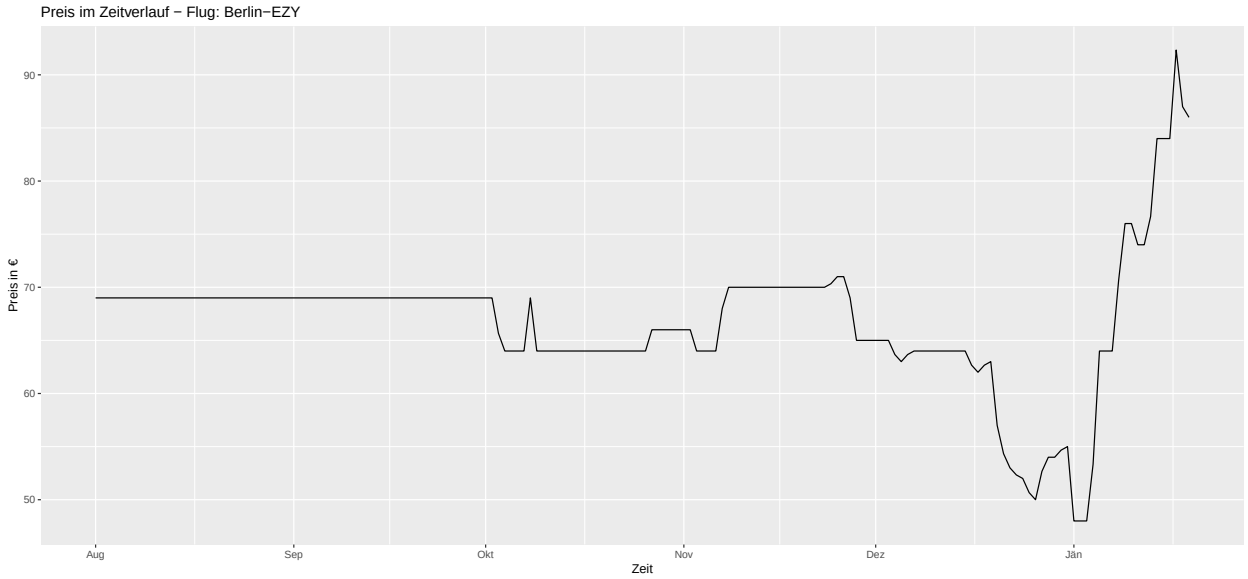

Preis im Zeitverlauf − Flug: Amsterdam−AUA
Preis in €
400
300
200
100
Aug
Sep
Okt
Nov
Dez
Jän
Zeit

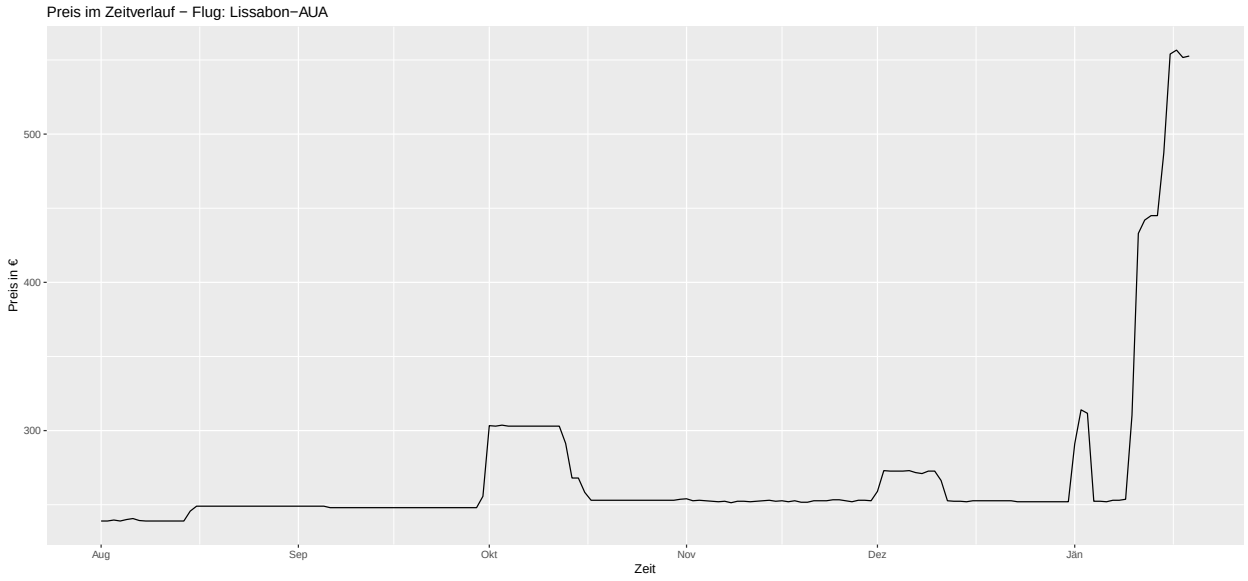

Preis im Zeitverlauf − Flug: Amsterdam−KLM
250
200
150
100
Preis in €
Aug
Sep
Okt
Nov
Dez
Jän
Zeit

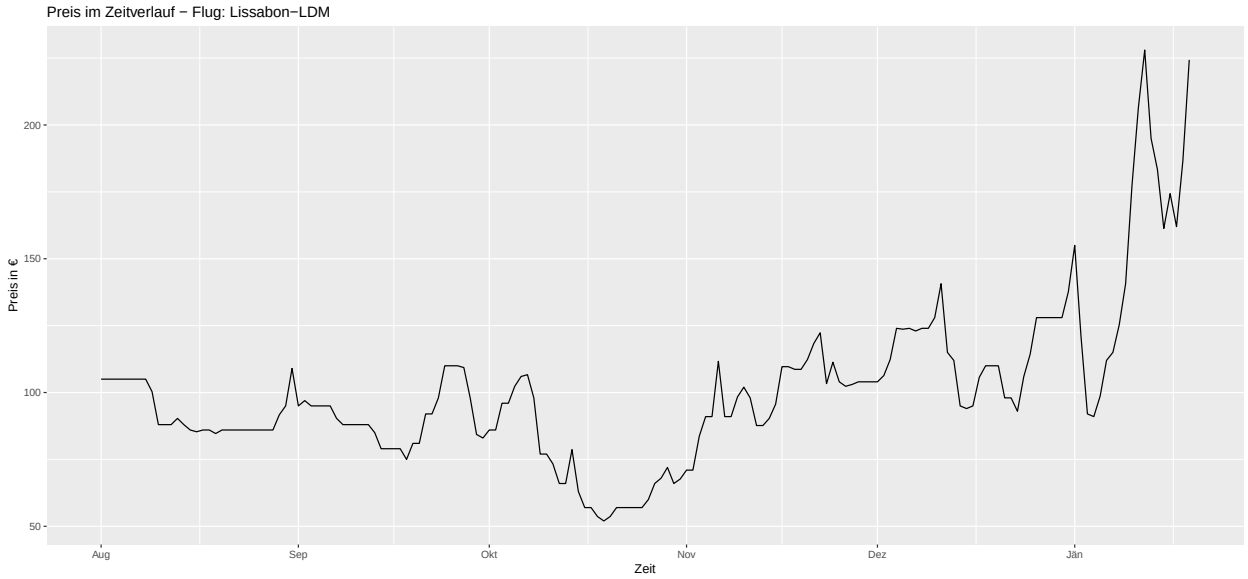

Preis im Zeitverlauf − Flug: Bangkok−AUA
1200
1000
800
Preis in €
Aug
Sep
Okt
Nov
Dez
Jän
Zeit

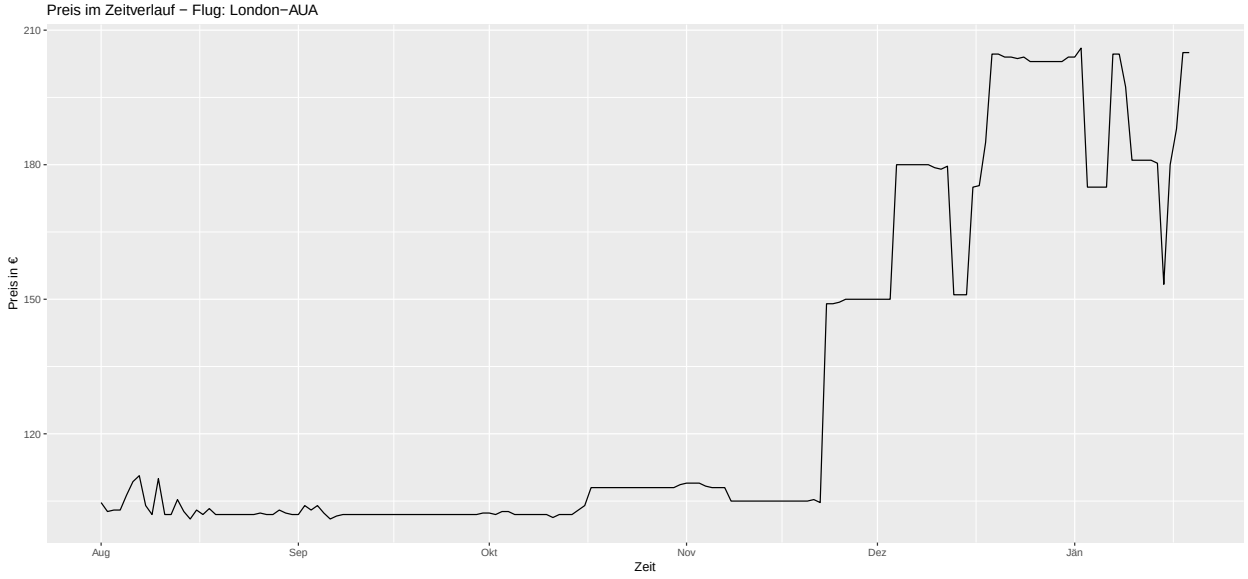

Preis im Zeitverlauf − Flug: Bangkok−THA
1300
1100
900
700
500
Preis in €
Aug
Sep
Okt
Nov
Dez
Jän
Zeit

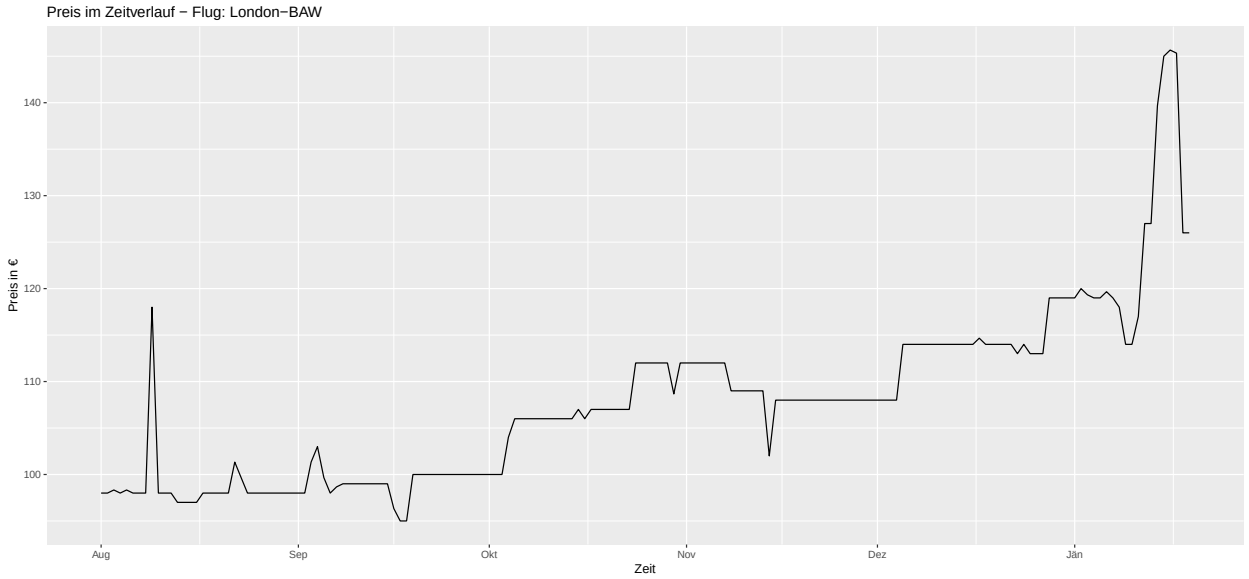

Preis im Zeitverlauf − Flug: Berlin−AUA
200
175
150
125
100
Preis in €
Aug
Sep
Okt
Nov
Dez
Jän
Zeit

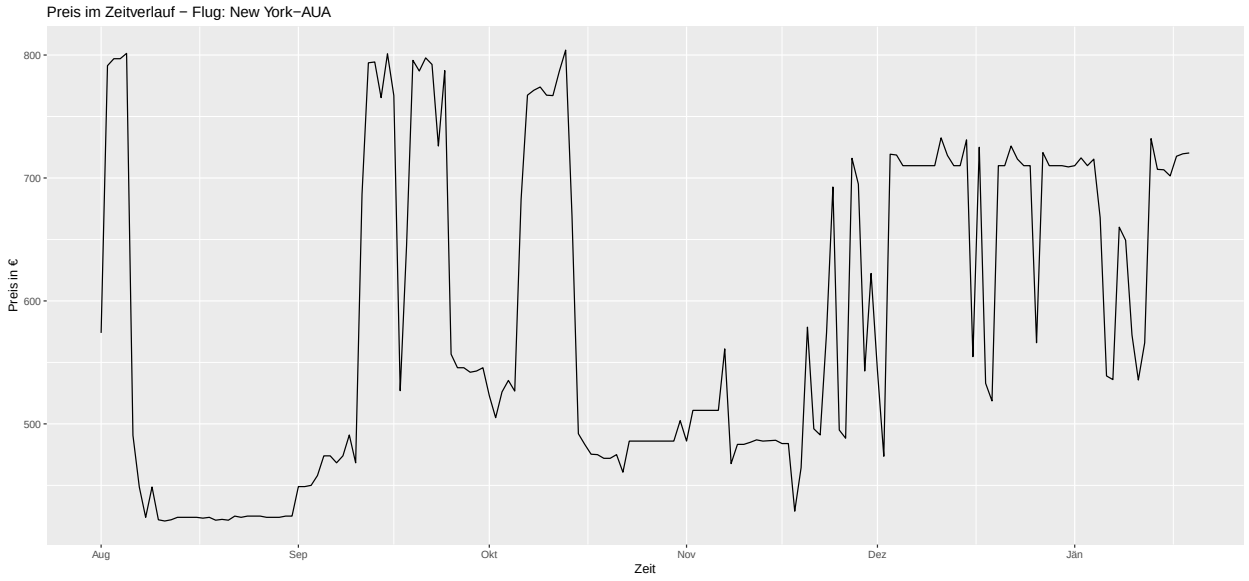

Preis im Zeitverlauf − Flug: Berlin−EZY
Preis in €
90
80
70
60
50
Aug
Sep
Okt
Nov
Dez
Jän
Zeit

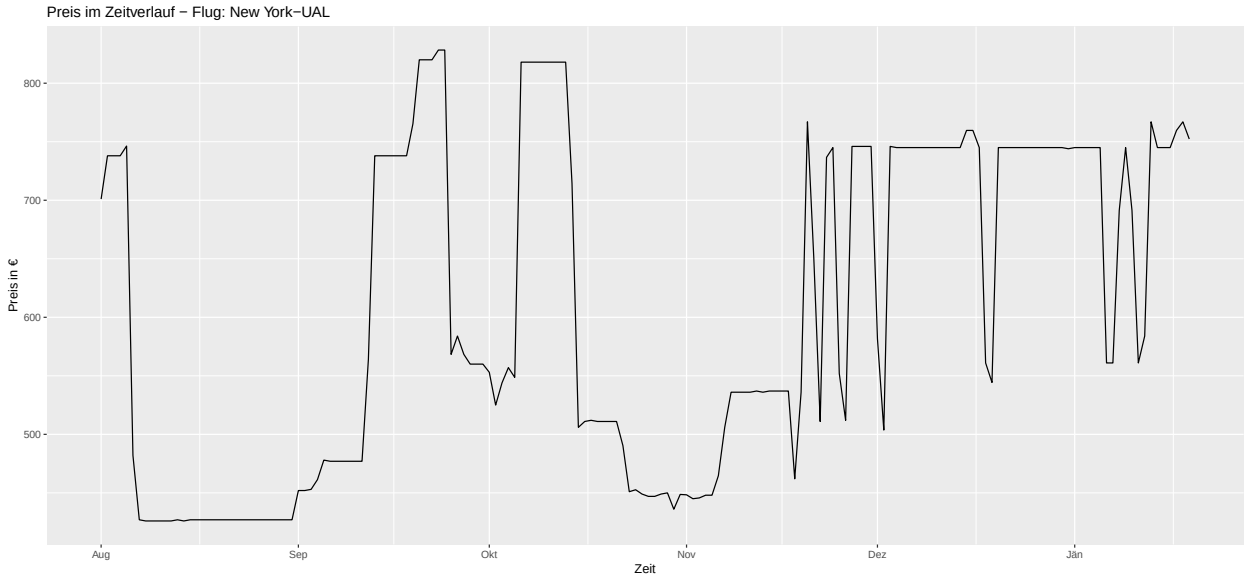

Preis im Zeitverlauf − Flug: Lissabon−AUA
Preis in €
500
400
300
Aug
Sep
Okt
Nov
Dez
Jän
Zeit

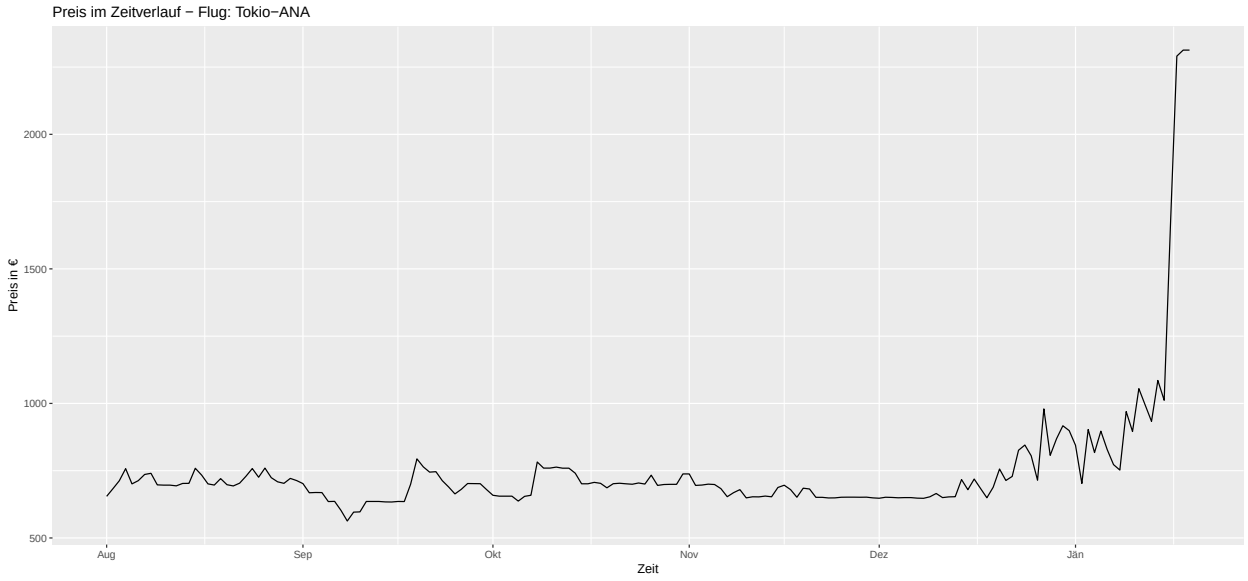

Preis im Zeitverlauf − Flug: Lissabon−LDM
Preis in €
200
150
100
50
Aug
Sep
Okt
Nov
Dez
Jän
Zeit

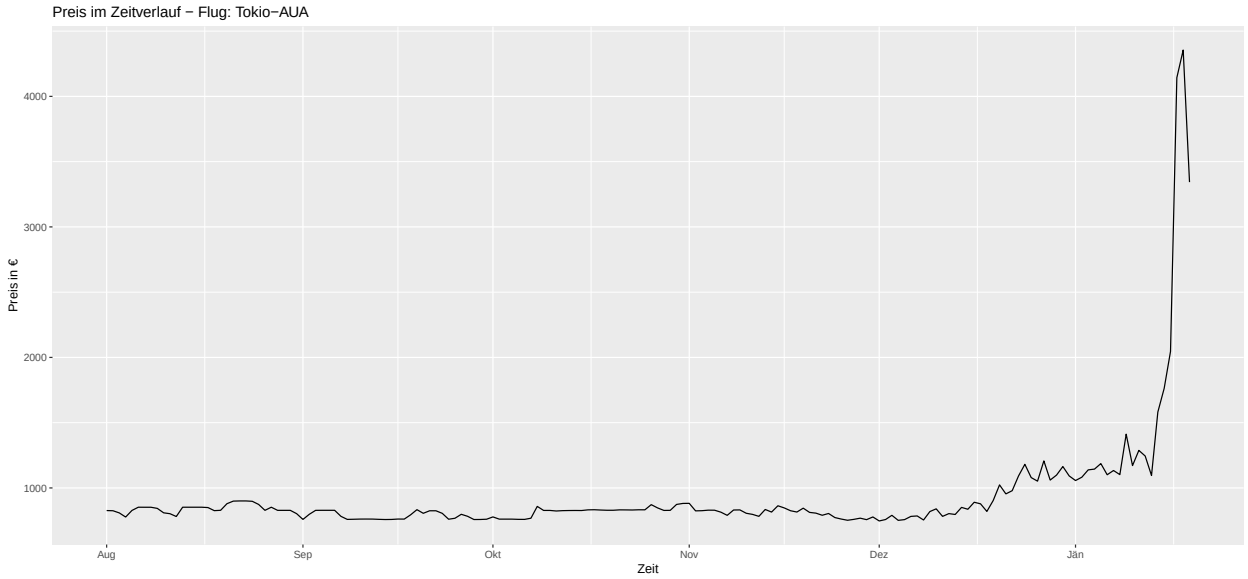

Preis im Zeitverlauf − Flug: London−AUA
Preis in €
210
180
150
120
Aug
Sep
Okt
Nov
Dez
Jän
Zeit


Preis im Zeitverlauf − Flug: London−BAW
Preis in €
140
130
120
110
100
Aug
Sep
Okt
Nov
Dez
Jän
Zeit

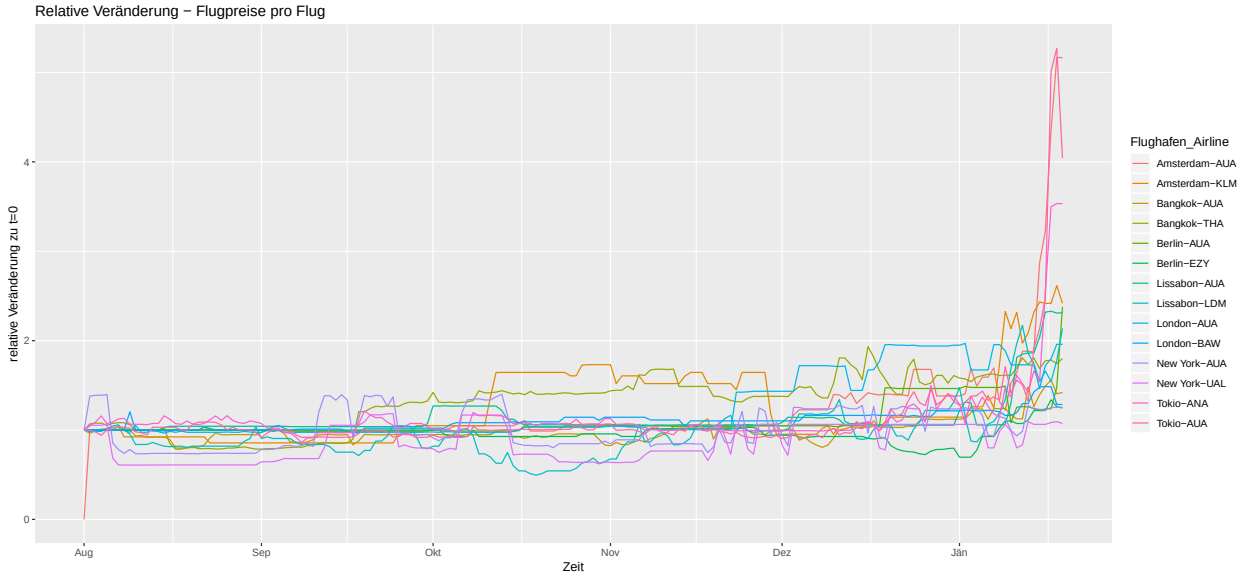

Preis im Zeitverlauf − Flug: New York−AUA
Preis in €
800
700
600
500
Aug
Sep
Okt
Nov
Dez
Jän
Zeit

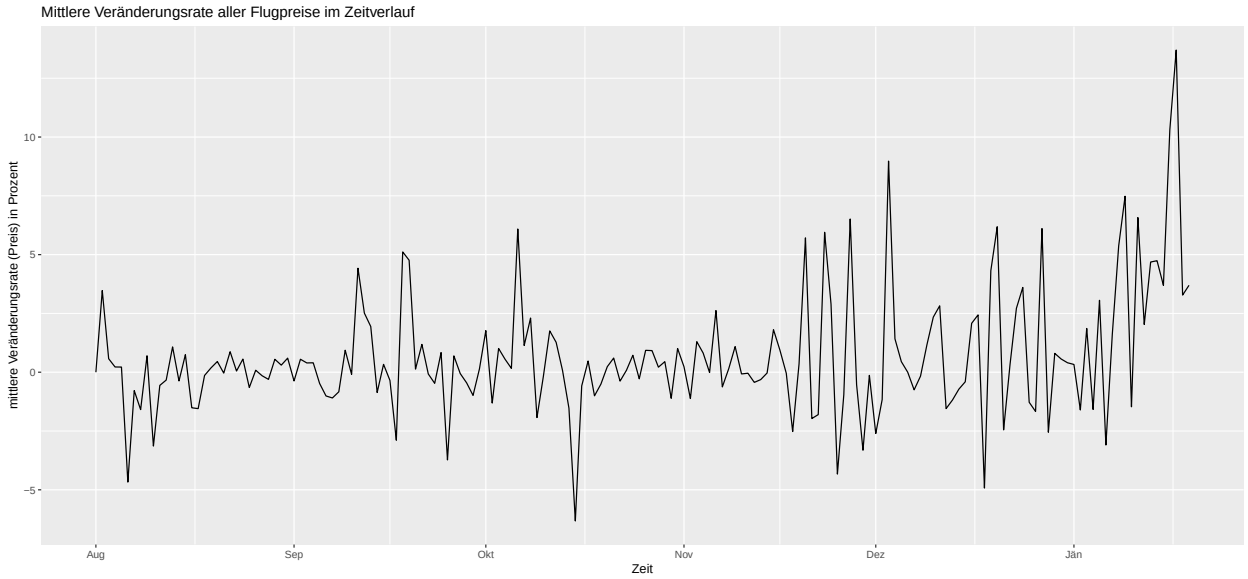

Preis im Zeitverlauf − Flug: New York−UAL
800
700
600
500
Preis in €
Aug
Sep
Okt
Nov
Dez
Jän
Zeit


Preis im Zeitverlauf − Flug: Tokio−ANA
2000
1500
1000
500
Preis in €
Aug
Sep
Okt
Nov
Dez
Jän
Zeit

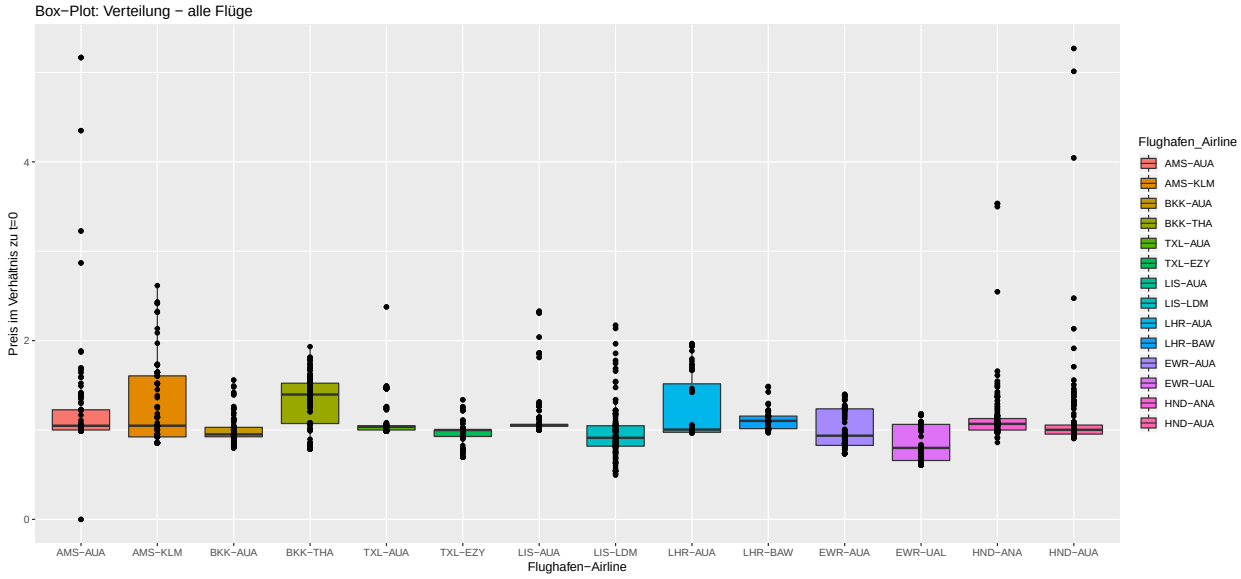


Relative Veränderung zu t=0





BoxPlot Vergleich aller Flüge



Hypothese 3

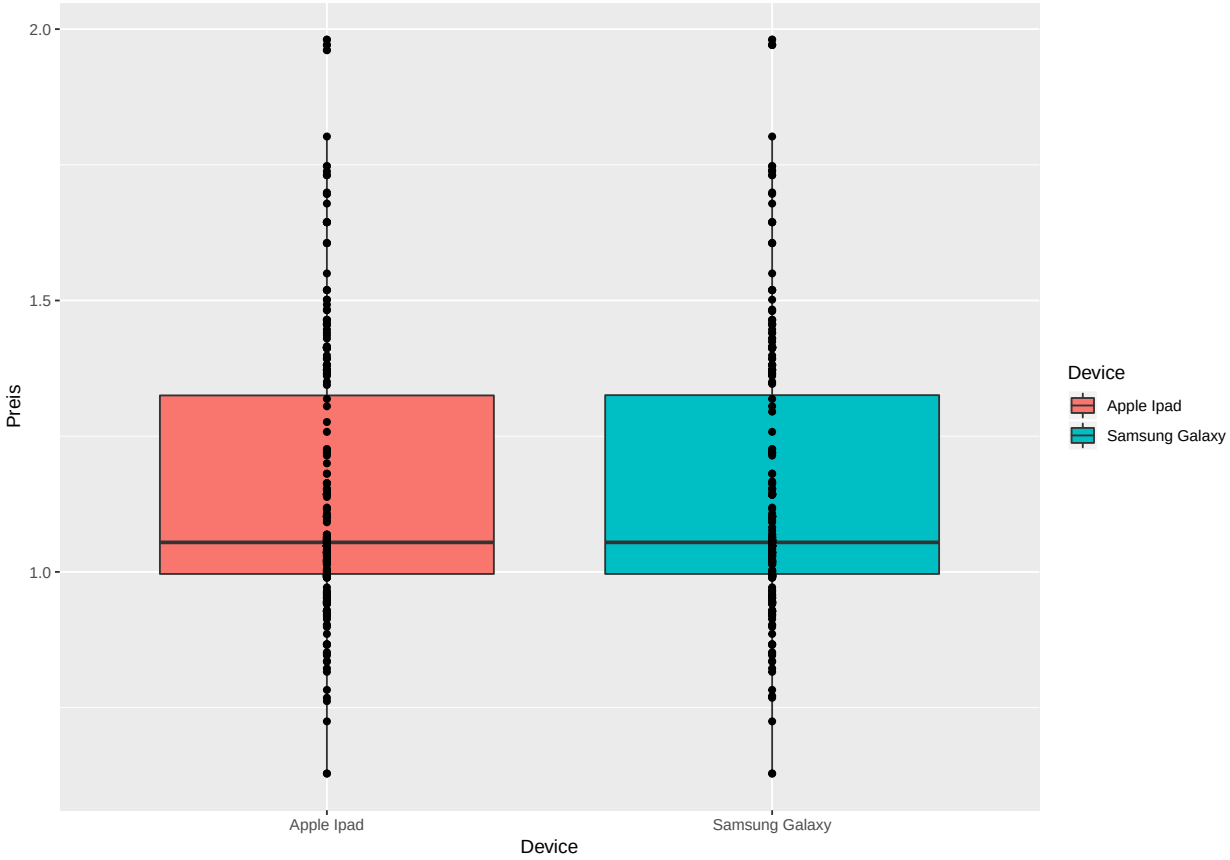
[Die Wahl des Betriebssystems respektive die Marke des Nutzerendgeräts mit dem die Reise-Website abgerufen wird, hat eine Auswirkung auf den offerierten Preis einer Airline.]

Deskreptive Statistik:

```
##
##  DESCRIPTIVES
##
##  Descriptives
##  ---------------------------------------
##                          Preis    Device
##  ---------------------------------------
```

```
##    N                      560       560
##    Missing                  0         0
##    Mean                  1.15
##    Median                1.05
##    Standard deviation   0.243
##    Minimum              0.629
##    Maximum               1.98
##  ----------------------------------------
##
##
##  FREQUENCIES
##
##  Frequencies of Device
##  ------------------------------------------------------------
##    Levels            Counts    % of Total    Cumulative %
##  ------------------------------------------------------------
##    Apple Ipad           280          50.0            50.0
##    Samsung Galaxy       280          50.0           100.0
##  ------------------------------------------------------------
```

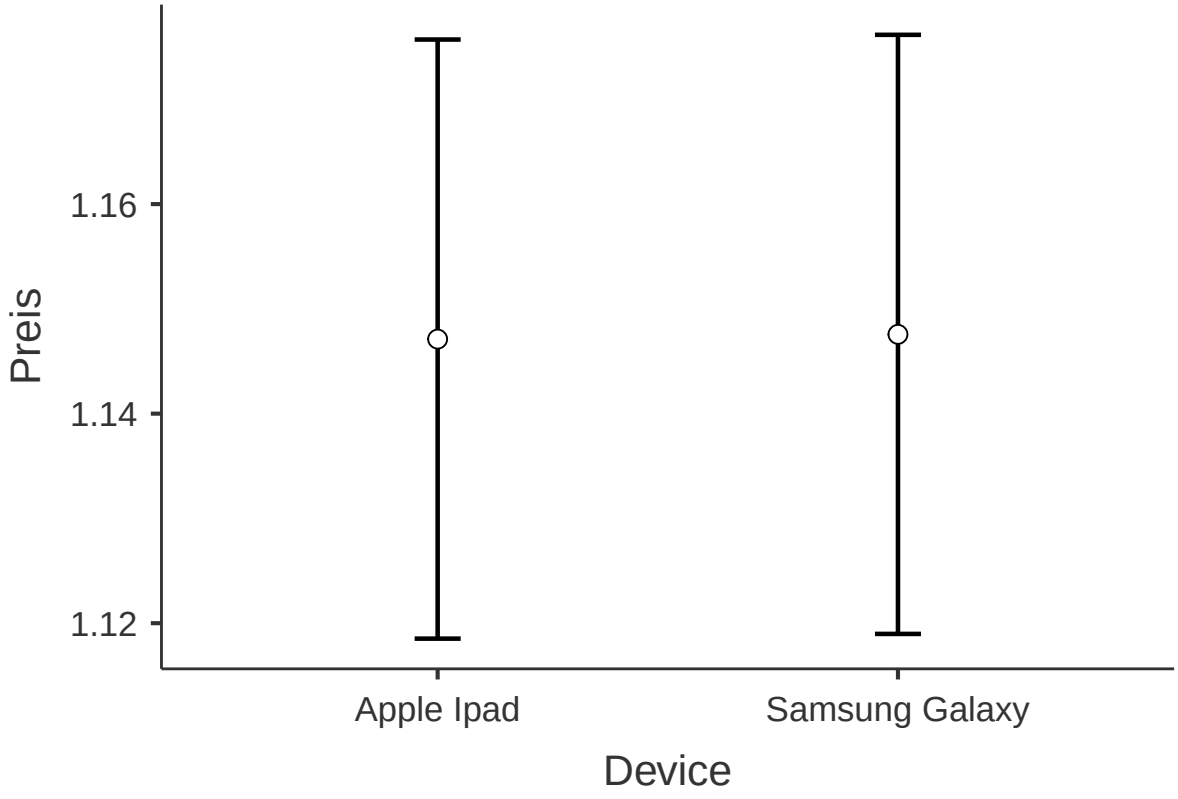
Verteilung - BoxPlot



Einfaktorielle ANOVA:

```
##
##  ANOVA
```

```
##
##  ANOVA
##  ------------------------------------------------------------------------------
##                  Sum of Squares    df     Mean Square    F          p
##  ------------------------------------------------------------------------------
##    Device              2.87e-5      1        2.87e-5    4.83e-4    0.982
##    Residuals              33.1    558         0.0593
##  ------------------------------------------------------------------------------
##
##
##  ESTIMATED MARGINAL MEANS
##
##  DEVICE
##
##  Estimated Marginal Means - Device
##  ----------------------------------------------------
##    Device             Mean    SE        Lower    Upper
##  ----------------------------------------------------
##    Apple Ipad         1.15    0.0146     1.12     1.18
##    Samsung Galaxy     1.15    0.0146     1.12     1.18
##  ----------------------------------------------------
```



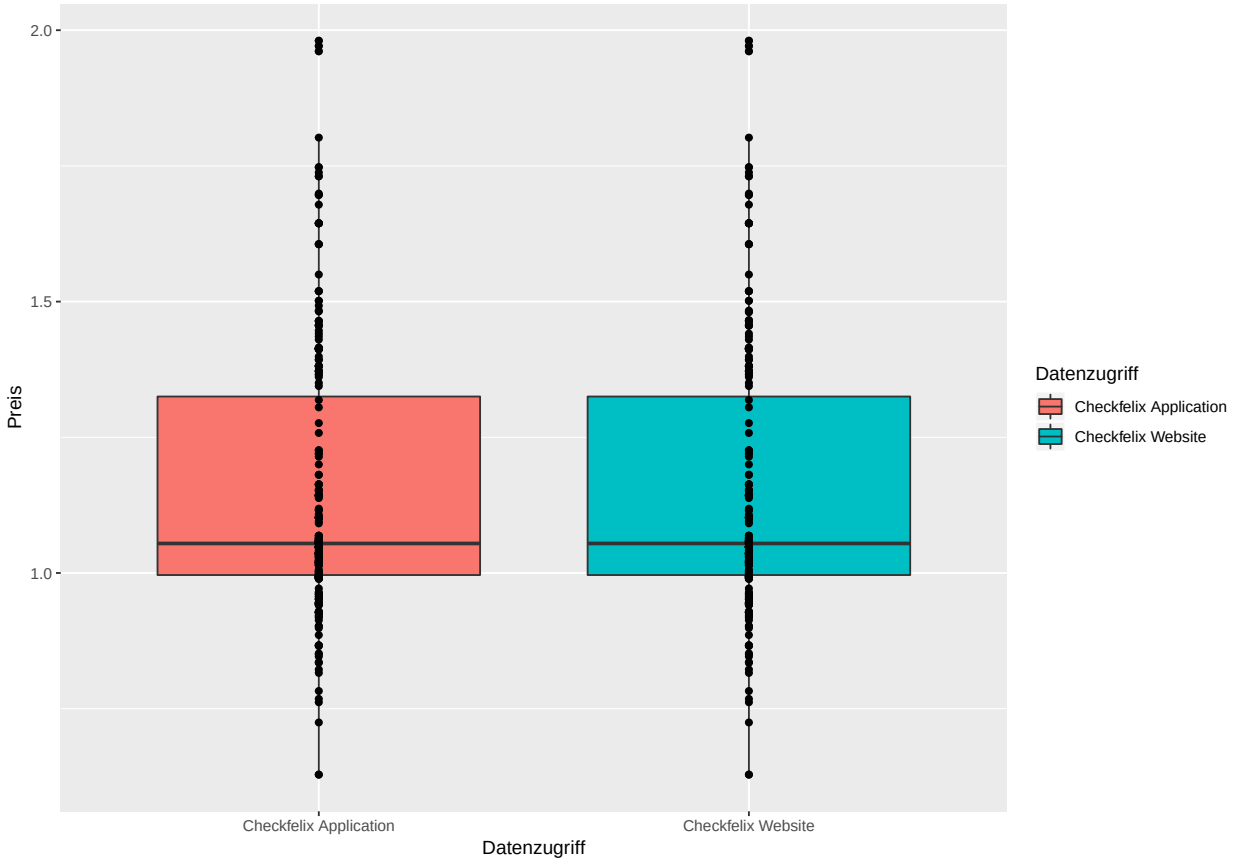
Hypothese 4

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

Deskreptive Statistik:

```
##
##  DESCRIPTIVES
##
##  Descriptives
##  ---------------------------------------------
##                           Preis    Datenzugriff
##  ---------------------------------------------
##    N                        560             560
##    Missing                    0               0
##    Mean                    1.15
##    Median                  1.05
##    Standard deviation     0.244
##    Minimum                0.629
##    Maximum                 1.98
##  ---------------------------------------------
##
##
##  FREQUENCIES
##
##  Frequencies of Datenzugriff
##  ------------------------------------------------------------------
##    Levels                     Counts    % of Total    Cumulative %
##  ------------------------------------------------------------------
##    Checkfelix Application        280          50.0            50.0
##    Checkfelix Website            280          50.0           100.0
##  ------------------------------------------------------------------
```

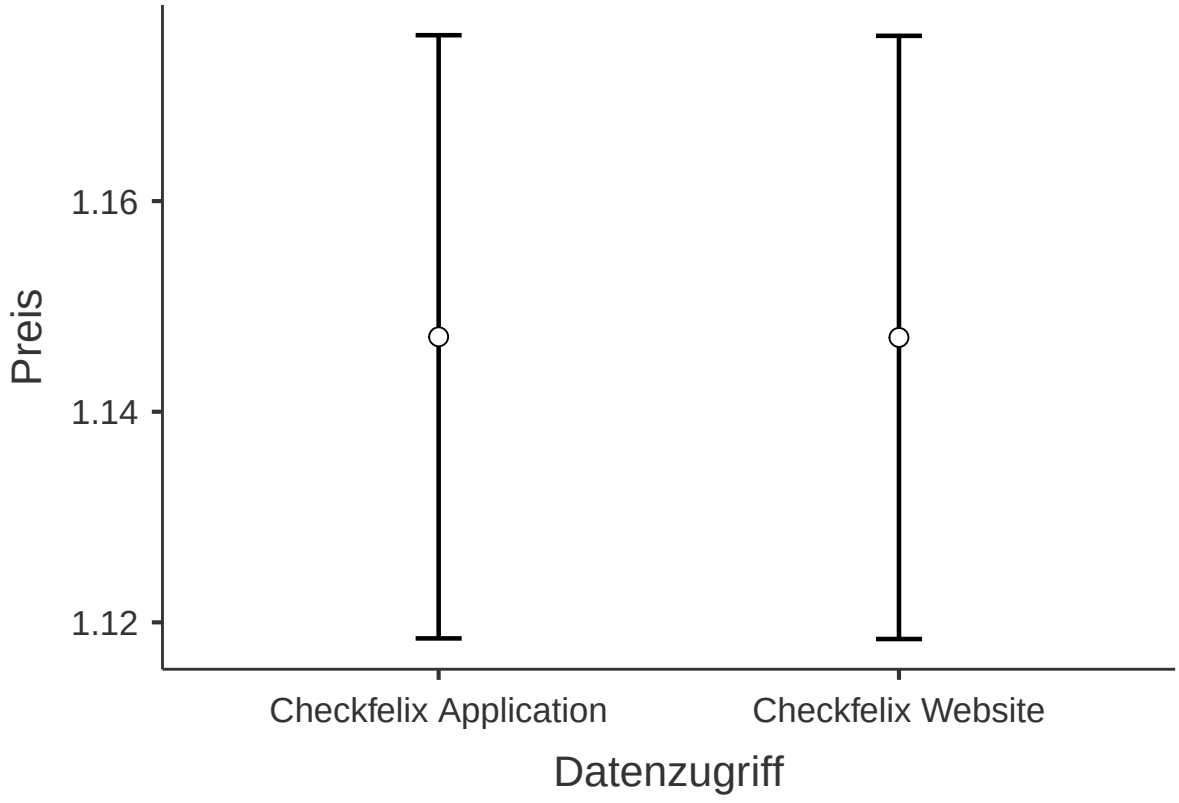
Verteilung - BoxPlot



Einfaktorielle ANOVA:

```
##
##  ANOVA
##
##  ANOVA
##  ------------------------------------------------------------------------------------
##                    Sum of Squares    df     Mean Square    F          p        <U+03B7>²p
##  ------------------------------------------------------------------------------------
##    Datenzugriff           4.85e-7     1        4.85e-7    8.14e-6    0.998    0.000
##    Residuals                 33.2   558         0.0595
##  ------------------------------------------------------------------------------------
##
##
##  ESTIMATED MARGINAL MEANS
##
##  DATENZUGRIFF
##
##  Estimated Marginal Means - Datenzugriff
##  ------------------------------------------------------------
##    Datenzugriff              Mean    SE        Lower    Upper
##  ------------------------------------------------------------
##    Checkfelix Application    1.15    0.0146     1.12     1.18
##    Checkfelix Website        1.15    0.0146     1.12     1.18
```

```
##  ------------------------------------------------------------
```



Hypothese 6

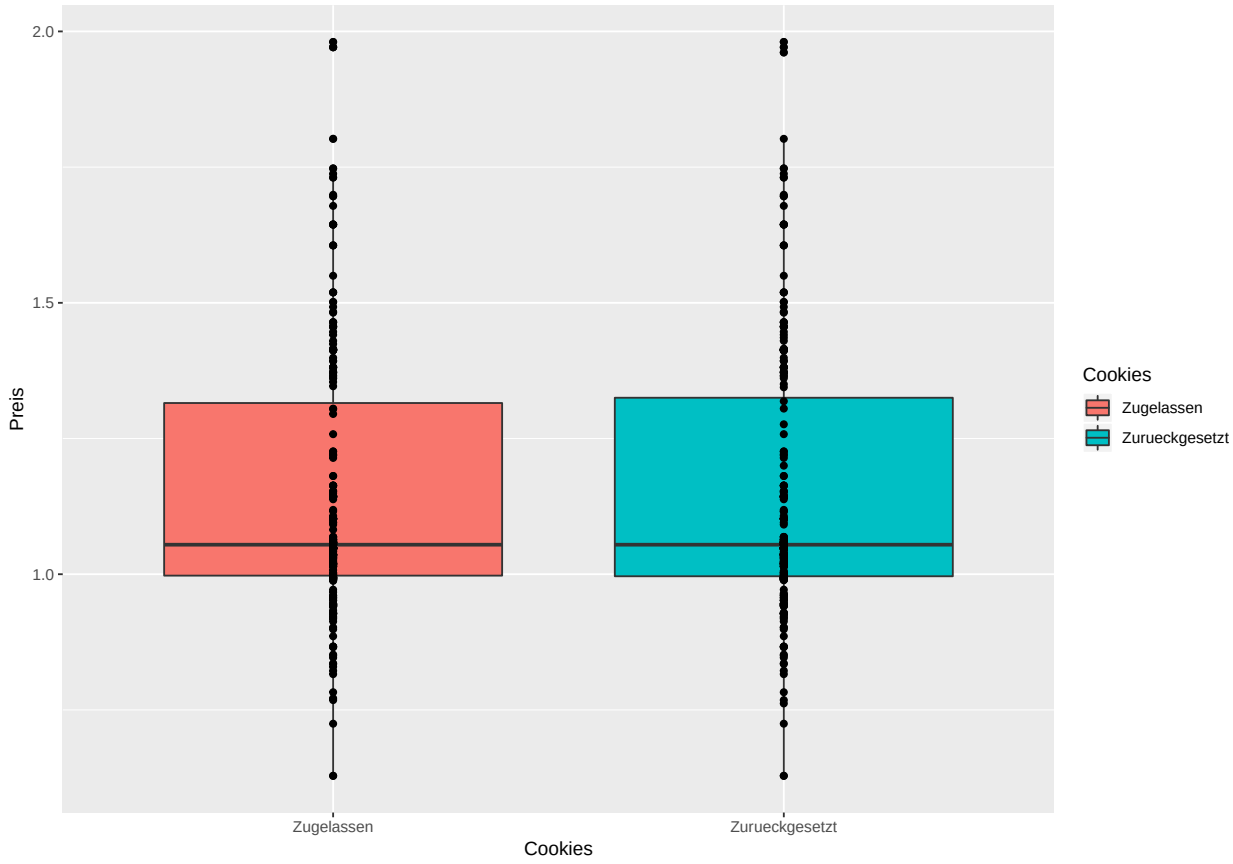
[Das Zurücksetzen von Cookies respektive dem Browserverlauf erwirkt ein Sinken des offerierten Preises einer Airline.]

Deskreptive Statistik:

```
##
##  DESCRIPTIVES
##
##  Descriptives
##  ----------------------------------------
##                          Preis    Cookies
##  ----------------------------------------
##    N                       560        560
##    Missing                   0          0
##    Mean                   1.15
##    Median                 1.05
##    Standard deviation    0.244
##    Minimum               0.629
##    Maximum                1.98
##  ----------------------------------------
##
##
##  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             4.64e-5     1         4.64e-5    7.81e-4    0.978    0.000
##    Residuals              33.2   558          0.0595
##  ------------------------------------------------------------------------------
##
##
##  ESTIMATED MARGINAL MEANS
```

```
##
##  COOKIES
##
##  Estimated Marginal Means - Cookies
##  ---------------------------------------------------
##    Cookies           Mean    SE        Lower    Upper
##  ---------------------------------------------------
##    Zugelassen        1.15    0.0146     1.12     1.18
##    Zurueckgesetzt    1.15    0.0146     1.12     1.18
##  ---------------------------------------------------
```

Hypothese 7

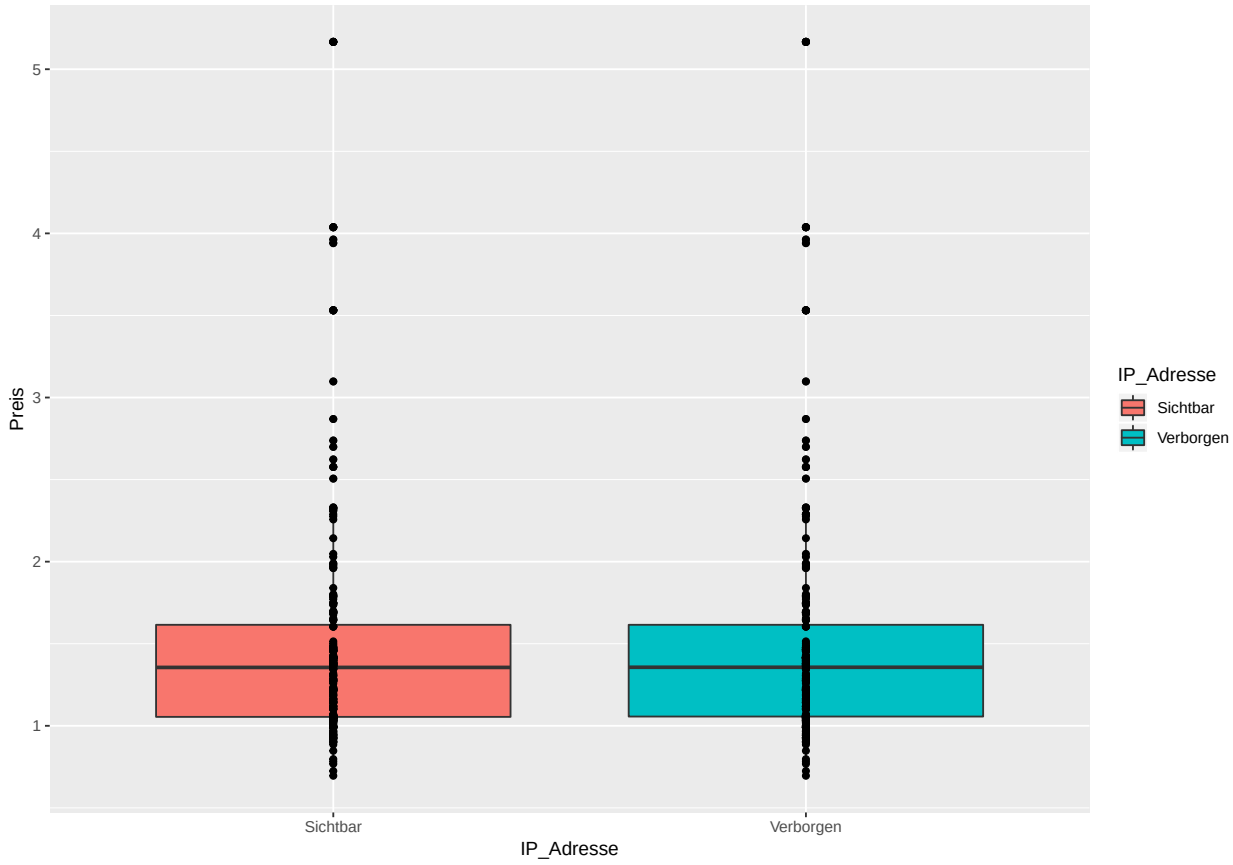
[Das Verbergen der Internetprotokoll-Adresse und folglich der ortsspezifischen Parameter mittels Virtual Private Network verursacht eine Differenz im offerierten Preis einer Airline.]

Deskreptive Statistik:

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

```
##    Mean                   1.50
##    Median                 1.36
##    Standard deviation    0.739
##    Minimum               0.696
##    Maximum                5.17
##  ---------------------------------------------
##
##
##  FREQUENCIES
##
##  Frequencies of IP_Adresse
##  ---------------------------------------------------
##    Levels       Counts    % of Total    Cumulative %
##  ---------------------------------------------------
##    Sichtbar        280          50.0            50.0
##    Verborgen       280          50.0           100.0
##  ---------------------------------------------------
```

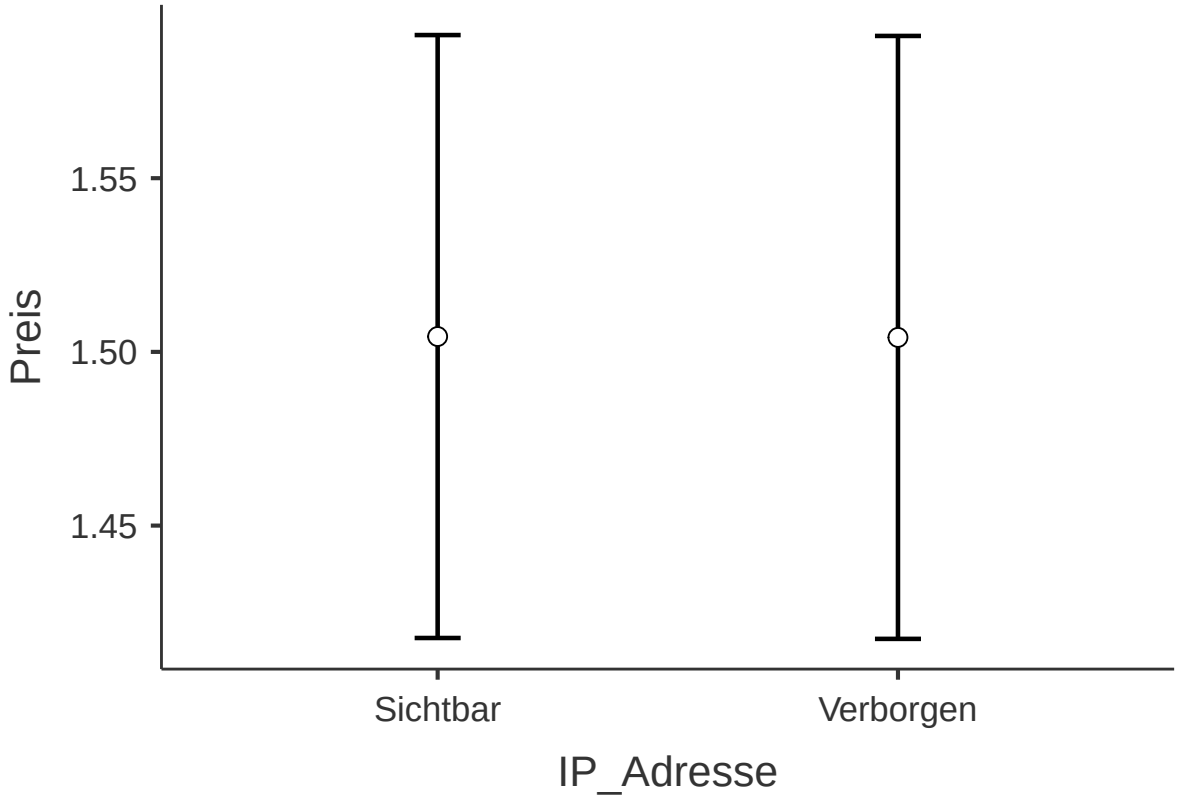
Verteilung - BoxPlot



Einfaktorielle ANOVA:

```
##
##  ANOVA
##
##  ANOVA
```

```
##  ------------------------------------------------------------------------------
##                  Sum of Squares    df     Mean Square    F          p        <U+03B7>²p
##  ------------------------------------------------------------------------------
##    IP_Adresse           9.15e-6      1        9.15e-6    1.67e-5    0.997    0.000
##    Residuals                305    558          0.547
##  ------------------------------------------------------------------------------
##
##
##  ESTIMATED MARGINAL MEANS
##
##  IP_ADRESSE
##
##  Estimated Marginal Means - IP_Adresse
##  -------------------------------------------------
##    IP_Adresse    Mean    SE        Lower    Upper
##  -------------------------------------------------
##    Sichtbar      1.50    0.0442     1.42     1.59
##    Verborgen     1.50    0.0442     1.42     1.59
##  -------------------------------------------------
```



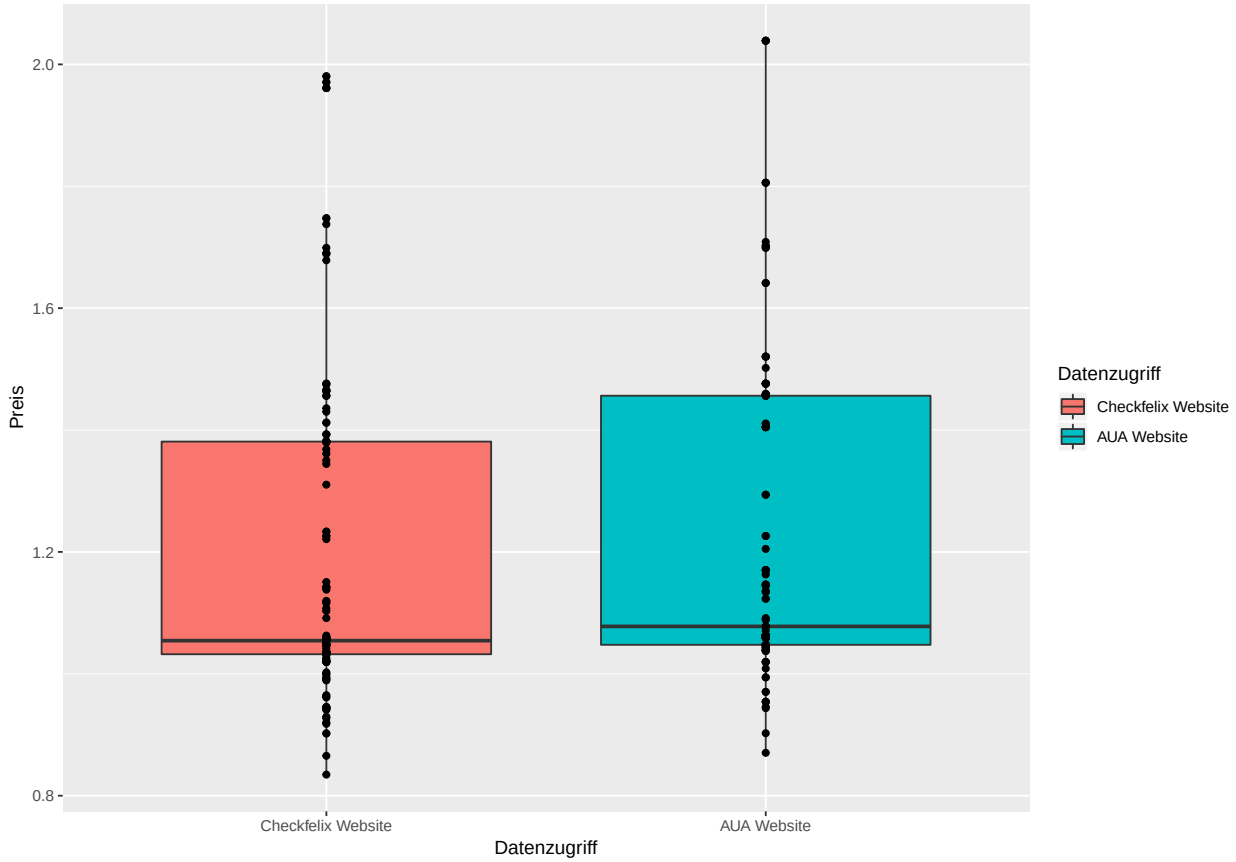
Hypothese 8

[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                        280             280
##    Missing                    0               0
##    Mean                    1.21
##    Median                  1.06
##    Standard deviation     0.260
##    Minimum                0.835
##    Maximum                 2.04
##  ---------------------------------------------
##
##
##  FREQUENCIES
##
##  Frequencies of Datenzugriff
##  ------------------------------------------------------------------
##    Levels                 Counts    % of Total    Cumulative %
##  ------------------------------------------------------------------
##    Checkfelix Website        140          50.0            50.0
##    AUA Website               140          50.0           100.0
##  ------------------------------------------------------------------
```

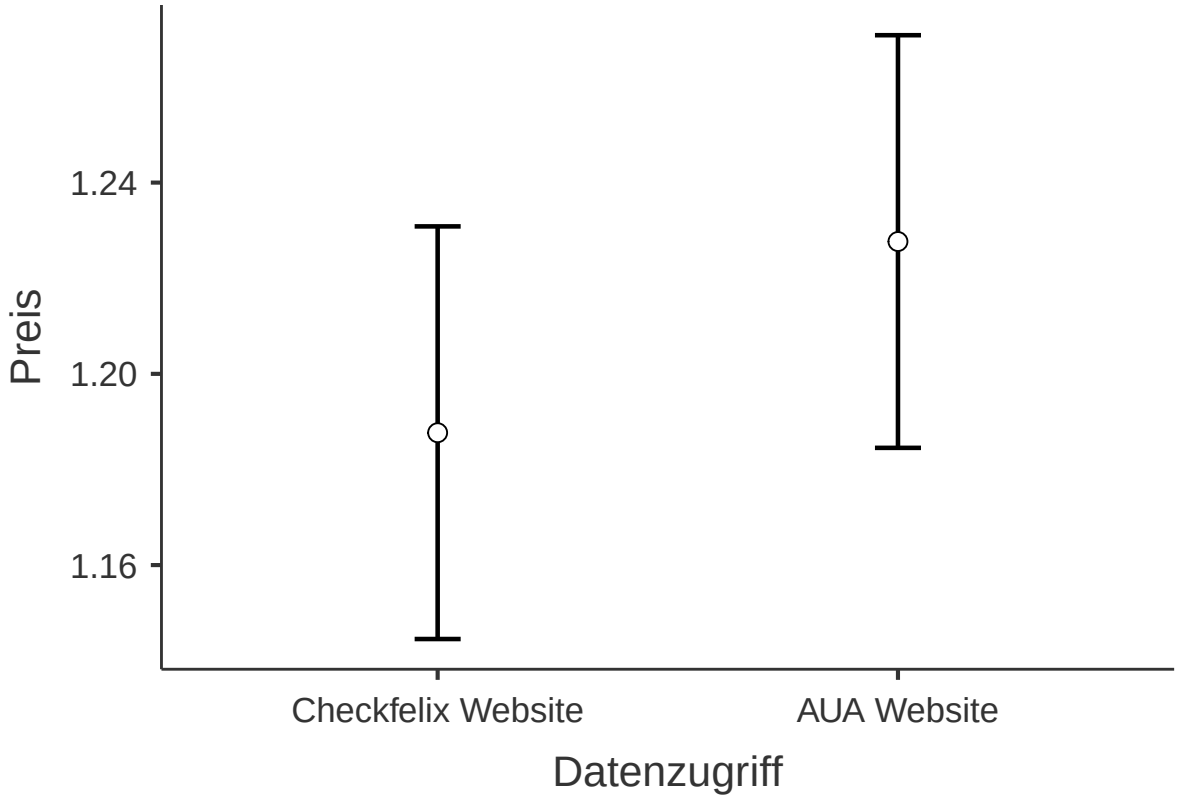
Verteilung - BoxPlot



Einfaktorielle ANOVA:

```
##
##  ANOVA
##
##  ANOVA
##  ---------------------------------------------------------------------------------
##                     Sum of Squares    df     Mean Square    F       p        <U+03B7>²p
##  ---------------------------------------------------------------------------------
##    Datenzugriff              0.112     1          0.1119    1.66    0.198    0.006
##    Residuals                18.704   278          0.0673
##  ---------------------------------------------------------------------------------
##
##
##  ESTIMATED MARGINAL MEANS
##
##  DATENZUGRIFF
##
##  Estimated Marginal Means - Datenzugriff
##  ----------------------------------------------------------
##    Datenzugriff          Mean    SE        Lower    Upper
##  ----------------------------------------------------------
##    Checkfelix Website    1.19    0.0219     1.14     1.23
##    AUA Website           1.23    0.0219     1.18     1.27
```

```
## ------------------------------------------------------------
```


1.24
1.20
1.16
Preis
Checkfelix Website
AUA Website
Datenzugriff