

DSBK2023 – Abschlusspräsentation

StonksAnalytics

Tran Anh Hoang

1. *Reminder – Datensätze und Datenarten*
2. *Zusammenfassung Ergebnisse Chapter 2 - Modelling and Evaluation - Analyse*
3. *Zusammenfassung Ergebnisse Chapter 3 - Modelling and Evaluation - Prädiktion*
4. *Bewertung des Ergebnisses des Data Minings in Bezug auf Ziele und Erfolgskriterien*
5. *Letzter Meilenstein*

1. Reminder – Datensätze und Datenarten

2. Reminder – Datensätze und Datenarten

ERR_USD.csv
EUNL_holdings_MSCI_World.csv
ICP_SDW_data_23Apr2023.00.01.42.csv
iShares Core MSCI World UCITS ETF USD EUNL.DE.csv
iShares Core MSCI World UCITS ETF USD (Acc)_Monthly.csv
leitzins Europa.csv
Leitzins_USA_FED.csv
PRIME.csv
snb-data-snbffzisa-de-all-20230421_0900.csv
snb-data-snbffzisa-de-selection-20230421_0900.csv

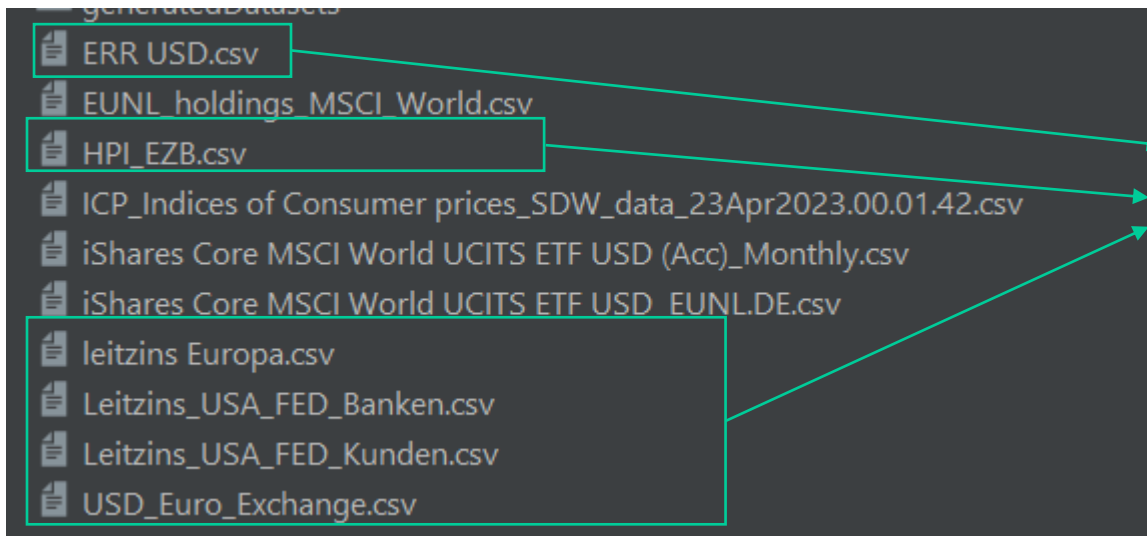
num. kontinuierlich

- historische Datenpunkte werden zu bestimmten Zeitintervallen aufgezeichnet, wie zum Beispiel täglich, wöchentlich oder monatlich
- Jeder Datenpunkt repräsentiert den Wert des ETFs zu einem bestimmten Zeitpunkt

Quelle:
yahoo finance

	Date	Open	High	Low	Close	Adj Close	Volume
1	2009-09-25	17.209999	17.209999	17.209999	17.209999	17.209999	0
2	2009-09-28	17.209999	17.209999	17.209999	17.209999	17.209999	0
3	2009-09-29	17.209999	17.209999	17.209999	17.209999	17.209999	0
4	2009-09-30	17.209999	17.209999	17.209999	17.209999	17.209999	0
5	2009-10-01	17.209999	17.209999	17.209999	17.209999	17.209999	0
6	2009-10-02	17.209999	17.209999	17.209999	17.209999	17.209999	0
7	2009-10-05	17.209999	17.209999	17.209999	17.209999	17.209999	0
8	2009-10-06	17.209999	17.209999	17.209999	17.209999	17.209999	0
9	2009-10-07	17.209999	17.209999	17.209999	17.209999	17.209999	0
10	2009-10-08	17.209999	17.209999	17.209999	17.209999	17.209999	0
11	2009-10-09	17.209999	17.209999	17.209999	17.209999	17.209999	0
12	2009-10-12	17.209999	17.209999	17.209999	17.209999	17.209999	0
13	2009-10-13	17.209999	17.209999	17.209999	17.209999	17.209999	0
14	2009-10-14	17.209999	17.209999	17.209999	17.209999	17.209999	0
15	2009-10-15	17.209999	17.209999	17.209999	17.209999	17.209999	0
16	2009-10-16	17.209999	17.209999	17.209999	17.209999	17.209999	0
17	2009-10-19	17.209999	17.209999	17.209999	17.209999	17.209999	0
18	2009-10-20	17.209999	17.209999	17.209999	17.209999	17.209999	0
19	2009-10-21	17.280001	17.280001	17.200001	17.250000	17.250000	200
20	2009-10-22	17.010000	17.010000	17.010000	17.010000	17.010000	0
21	2009-10-23	17.080000	17.080000	16.940001	16.940001	16.940001	50
22	2009-10-26	17.000000	17.000000	16.840000	16.840000	16.840000	100

2. Reminder – Datensätze und Datenarten



- num./diskret Kontinuierlich
- historische Datenpunkte werden zu bestimmten Zeitintervallen aufgezeichnet, wie zum Beispiel täglich, wöchentlich oder monatlich
 - Jeder Datenpunkt repräsentiert den Wert des Leitzinses, Geldpolitik und Wechselkurse zu einem bestimmten Zeitpunkt

Quelle: EZB, FED

	Date	Leitzins per annum	Value status	Value comment
1	2023-04-22	3.5000	Normal value (A)	
2	2023-04-21	3.5000	Normal value (A)	
3	2023-04-20	3.5000	Normal value (A)	
4	2023-04-19	3.5000	Normal value (A)	
5	2023-04-18	3.5000	Normal value (A)	
6	2023-04-17	3.5000	Normal value (A)	
7	2023-04-16	3.5000	Normal value (A)	
8	2023-04-15	3.5000	Normal value (A)	
9	2023-04-14	3.5000	Normal value (A)	
10	2023-04-13	3.5000	Normal value (A)	
11	2023-04-12	3.5000	Normal value (A)	
12	2023-04-11	3.5000	Normal value (A)	
13	2023-04-10	3.5000	Normal value (A)	
14	2023-04-09	3.5000	Normal value (A)	
15	2023-04-08	3.5000	Normal value (A)	
16	2023-04-07	3.5000	Normal value (A)	
17	2023-04-06	3.5000	Normal value (A)	
18	2023-04-05	3.5000	Normal value (A)	
19	2023-04-04	3.5000	Normal value (A)	

2. Erklärung der Datensätze

ERR_USD.csv
EUNL_holdings_MSCI_World.csv
ICP_SDW_data_23Apr2023.00.01.42.csv
iShares Core MSCI World UCITS ETF USD EUNL.DE.csv
iShares Core MSCI World UCITS ETF USD (Acc)_Monthly.csv
leitzins Europa.csv
Leitzins_USA_FED.csv
PRIME.csv
snb-data-snbffzisa-de-all-20230421_0900.csv
snb-data-snbffzisa-de-selection-20230421_0900.csv

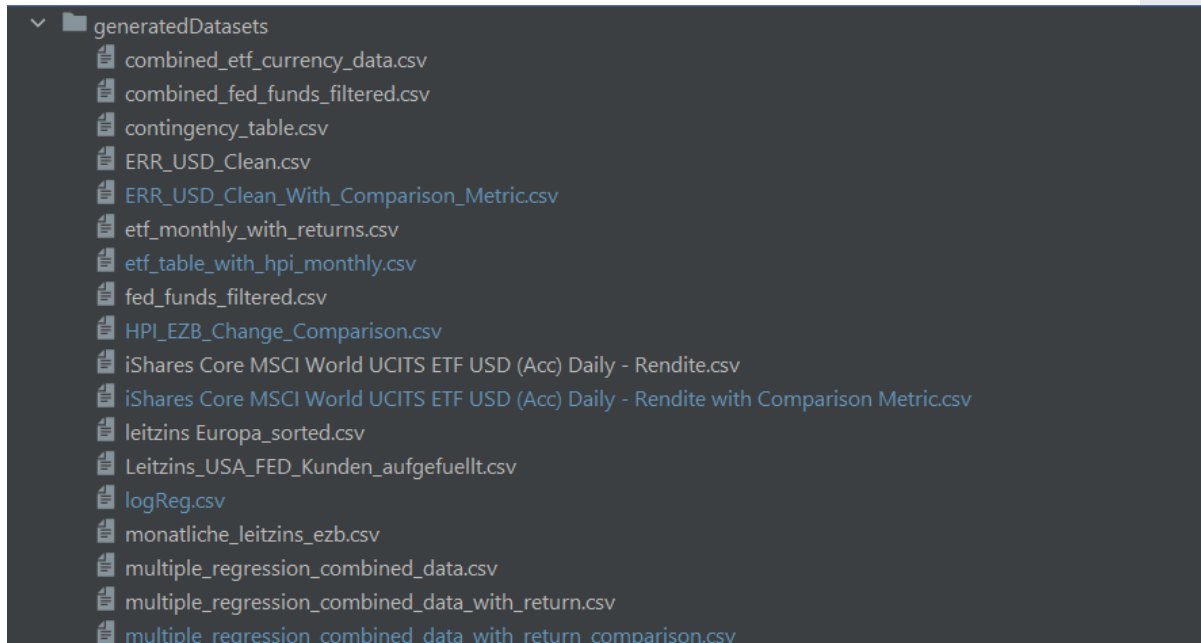
Kategorial

- Listet alle Unternehmen, die in ETF erhalten sind
- Sortierung der Werte nach vordefinierten Kategorien und Faktoren

Quelle: IShares

ondsposition per	31.Mai.2023	<unset>	<unset>	<unset>	<unset>	<unset>	<unset>	<unset>
BSP	<unset>	<unset>	<unset>	<unset>	<unset>	<unset>	<unset>	<unset>
mittententicker	Name	Sektor	Anlageklasse	Marktwert	Gewichtung (%)	Nominalwert	Nominale	K
APL	APPLE INC	IT	Aktien	2.712.235.187,50	5,21	2.712.235.187,50	15.301.750,00	1
SFT	MICROSOFT CORP	IT	Aktien	2.247.309.051,85	4,32	2.247.309.051,85	6.843.415,00	3
MZN	AMAZON COM INC	Zyklische Konsumgüter	NBSP Aktien	1.077.635.276,84	2,07	1.077.635.276,84	8.937.098,00	1
VDA	NVIDIA CORP	IT	Aktien	903.775.186,94	1,74	903.775.186,94	2.388.791,00	3
GOGL	ALPHABET INC CLASS A	Kommunikation	Aktien	706.938.688,50	1,36	706.938.688,50	5.753.550,00	1
GOOG	ALPHABET INC CLASS C	Kommunikation	Aktien	641.623.806,33	1,23	641.623.806,33	5.200.809,00	1
META	META PLATFORMS INC CLASS A	Kommunikation	Aktien	567.698.393,28	1,09	567.698.393,28	2.144.524,00	2
SLA	TESLA INC	Zyklische Konsumgüter	NBSP Aktien	563.093.147,44	1,08	563.093.147,44	2.761.208,00	2
UNH	UNITEDHEALTH GROUP INC	Gesundheitsversorgung	Aktien	440.843.058,24	0,85	440.843.058,24	904.776,00	4
BRKB	BERKSHIRE HATHAWAY INC CLASS B	Financials	Aktien	402.770.779,00	0,77	402.770.779,00	1.254.425,00	3
EXXON	EXXON MOBIL CORP	Energie	Aktien	401.021.452,26	0,77	401.021.452,26	3.924.657,00	1
JOJ	JOHNSON & JOHNSON	Gesundheitsversorgung	Aktien	389.523.124,80	0,75	389.523.124,80	2.512.080,00	1

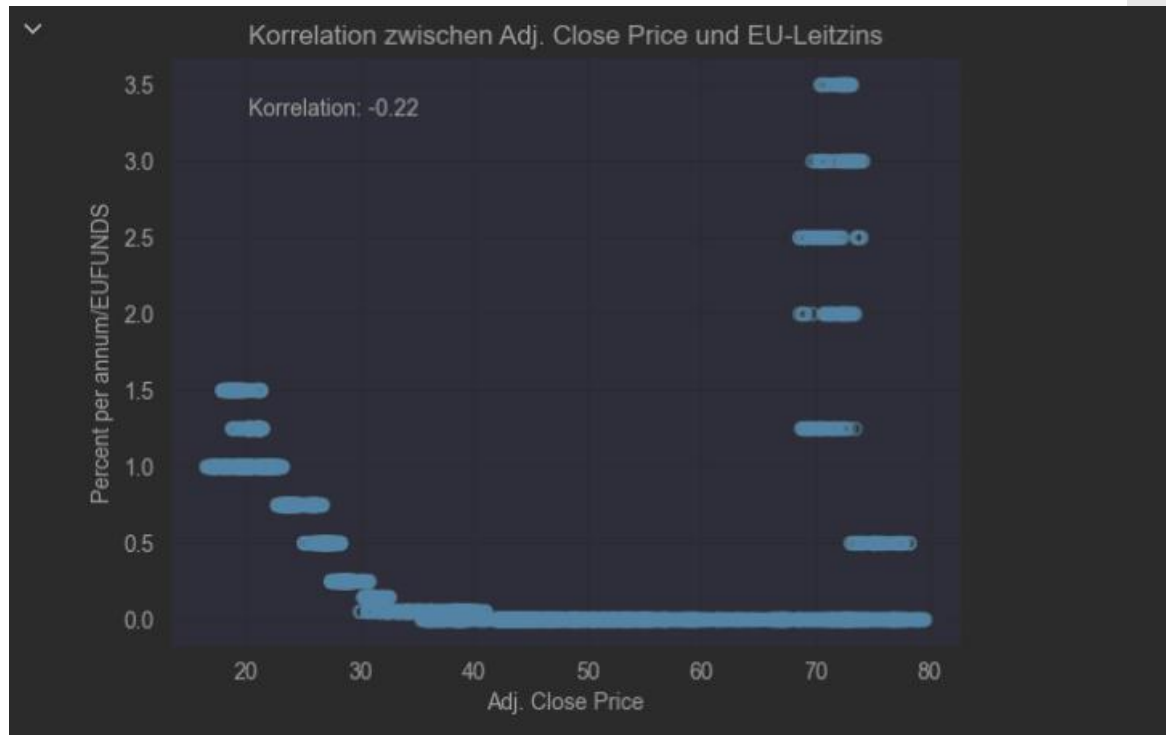
2. Erklärung der Datensätze



Zusammensetzung und Bereinigung von verschiedenen kategorialen als auch num./diskret kontinuierlich Daten für verschiedene Analysen und als CSV-Dateien abgespeichert

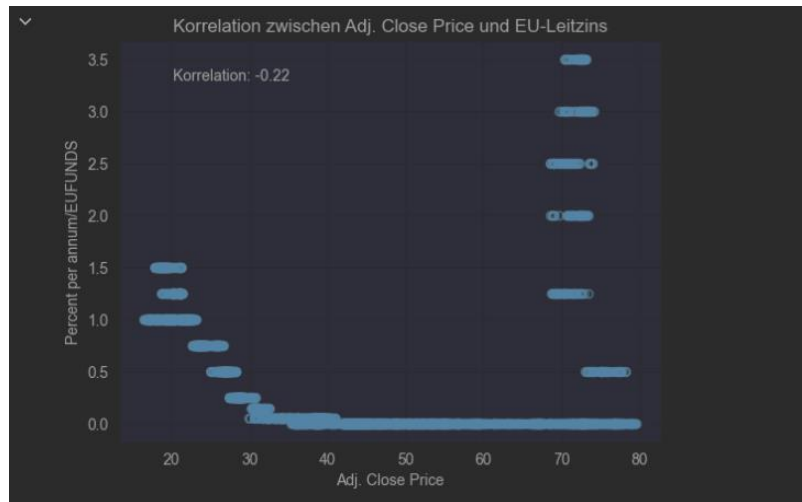
2. Zusammenfassung Ergebnisse Chapter 2 - Modelling and Evaluation - Analyse

2. T-Test



Aus den **Korrelationsdaten** und das **Streudiagramm** konnte man einen **gewissen Zusammenhang** zwischen dem **Leitzins der EZB** und den **Adj. Close Price** des ETFs erkennen.

2. T-Test

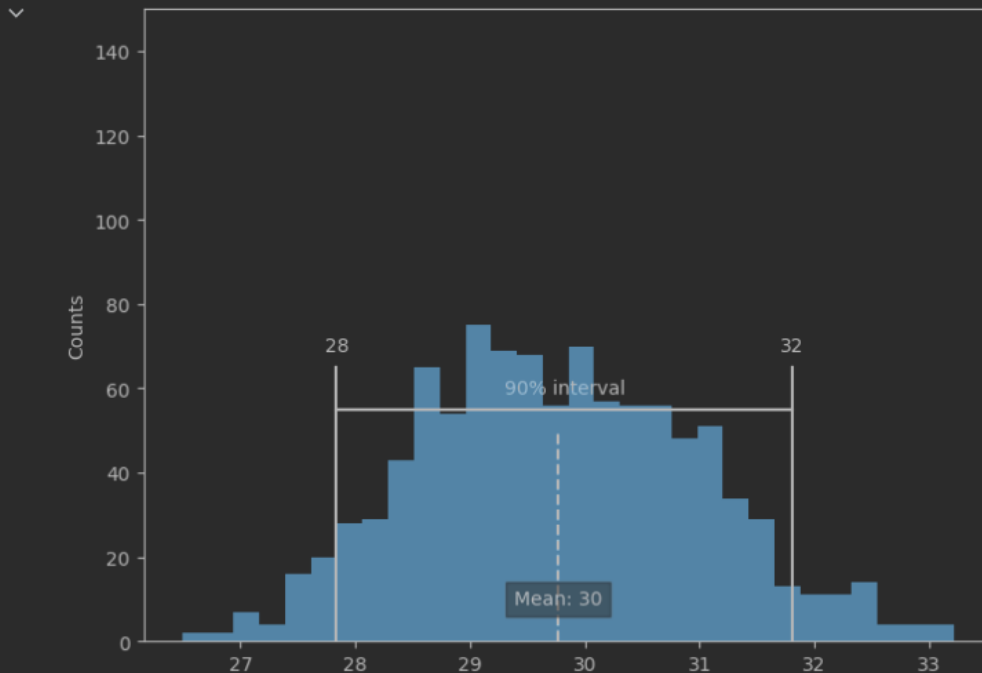


- Nullhypothese (H_0): Es besteht **keine statistisch signifikante Beziehung** zwischen den **Adj. Closed Price des MSCI World und dem Leitzins der EU-Zentralbank**. Der Leitzins der EU-Zentralbank hat **keinen signifikanten Einfluss** auf den Adj. Closed Price des Aktienmarktes
- Alternativhypothese (H_1): Es besteht **eine statistisch signifikante Beziehung** zwischen **Adj. Closed Price des MSCI World und dem Leitzins der EU-Zentralbank**. Der Leitzins der EU-Zentralbank **hat einen signifikanten Einfluss** auf den Adj. Closed Price des Aktienmarktes

2. T-Test

✓ t-Statistic: 29.60855086116252
p-value: 1.0600119261545373e-67

Durchschnittliche Werte:
Durchschnittliche t-Statistik: 29.76566133619707
Durchschnittlicher p-Wert: 1.1371456639268555e-63
Standardabweichung von p_value: 1.7331534604061102e-62

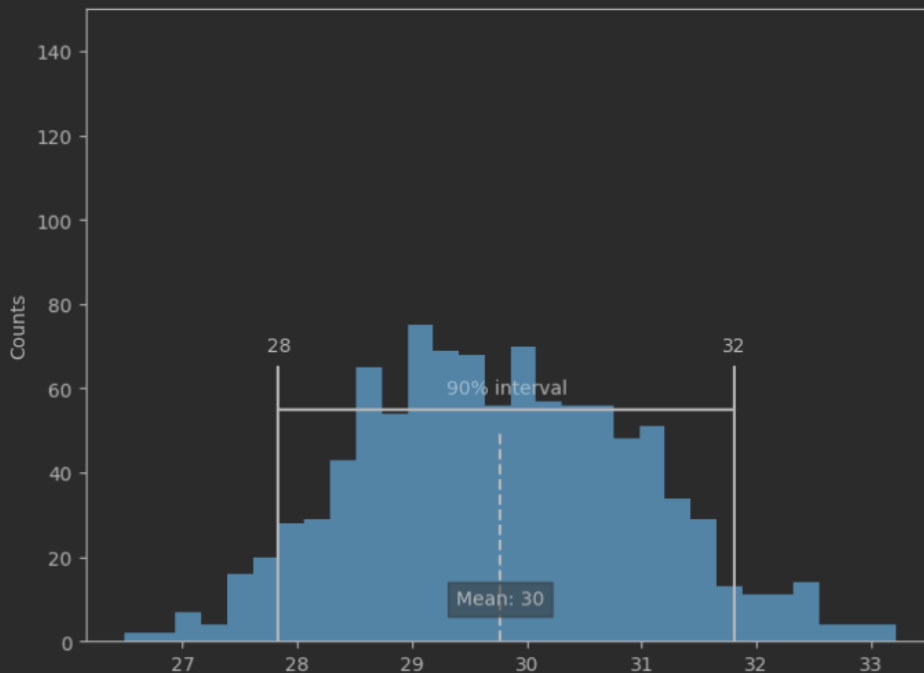


- durchschnittliche Wert der t-Statistik von zeigt an, wie stark der beobachtete Wert vom erwarteten Wert abweicht.
- die t-Statistik liegt mit 29.76566133619707 in einem etwas höheren Bereich:
 - **statistisch signifikante Beziehung zwischen den Variablen gibt**

2. T-Test

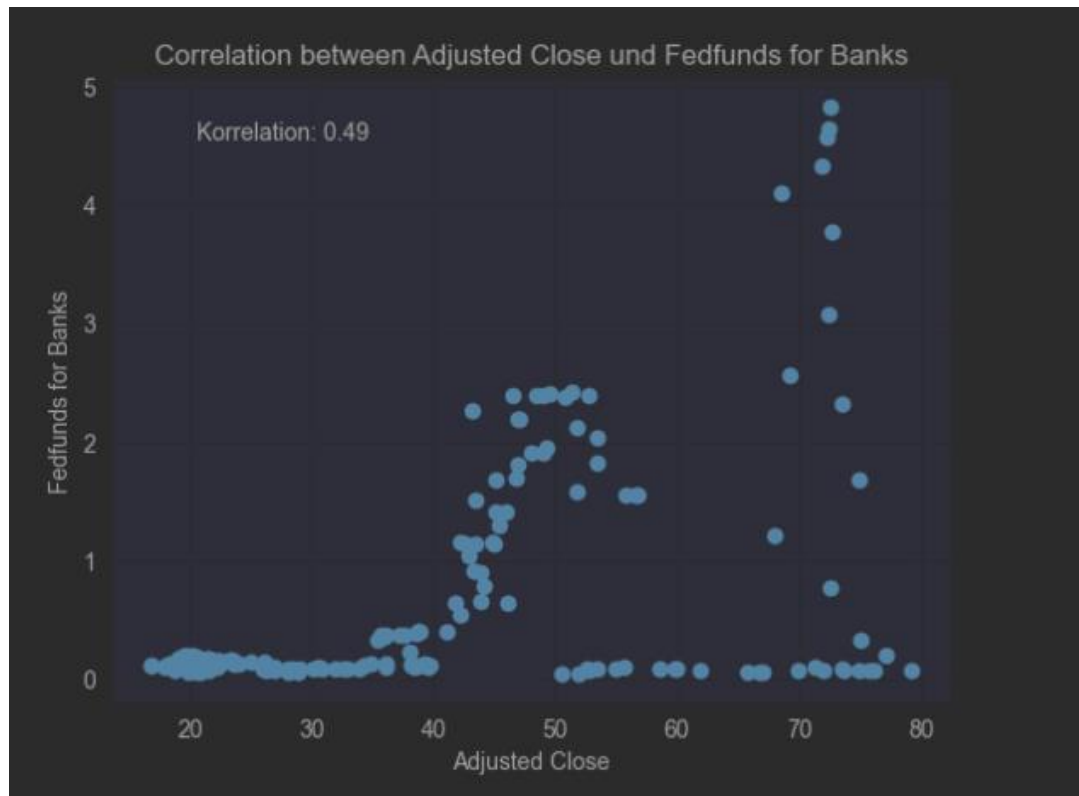
✓ t-Statistic: 29.60855086116252
p-value: 1.0600119261545373e-67

Durchschnittliche Werte:
Durchschnittliche t-Statistik: 29.76566133619707
Durchschnittlicher p-Wert: 1.1371456639268555e-63
Standardabweichung von p_value: 1.7331534604061102e-62



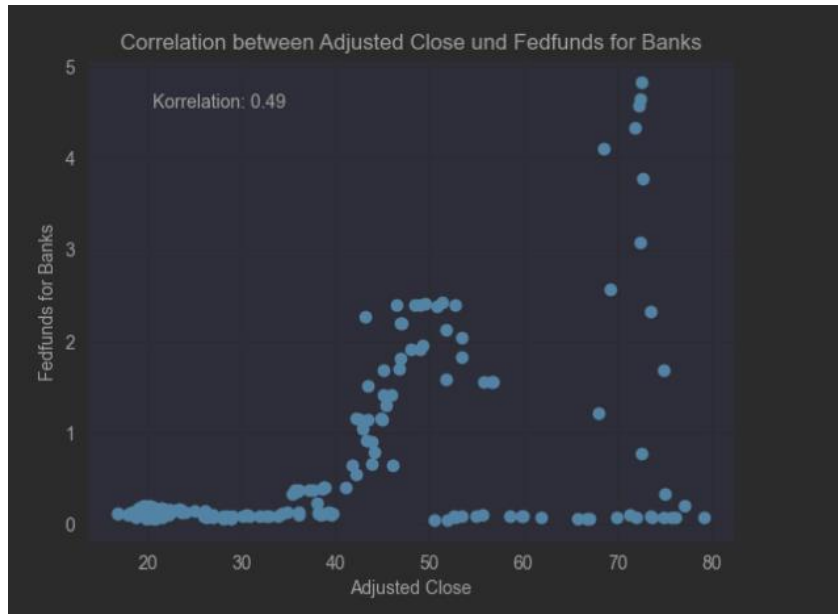
- durchschnittliche p-Wert liegt $1.2835205117325122e-62 \rightarrow$ Unterschiede zwischen den Adj. Close-Werten und den Leitzins-Werten höchstwahrscheinlich nicht zufällig und statistisch signifikant sind
- p-Wert liegt unter den Signifikanzniveau von 0,05
 - Ablehnung der Nullhypothese.
 - Akzeptanz der Alternativhypothese
- Deutung: Leitzins der EU-Zentralbank hat einen signifikanten Einfluss auf den Adj. Closed Price des MSCI World
- Standardabweichung der p-Werte von $1.7331534604061102e-62$
 - Standardabweichung sehr klein kann man sehen, dass die p-Werte in den Bootstrap-Samples relativ eng um den Durchschnitt gruppiert sind
 - die Bootstrap-Samples konsistente und stabile Ergebnisse in Bezug auf die p-Werte liefern == stärkt die Vertrauenswürdigkeit der statistischen Tests

2. Anova - Test



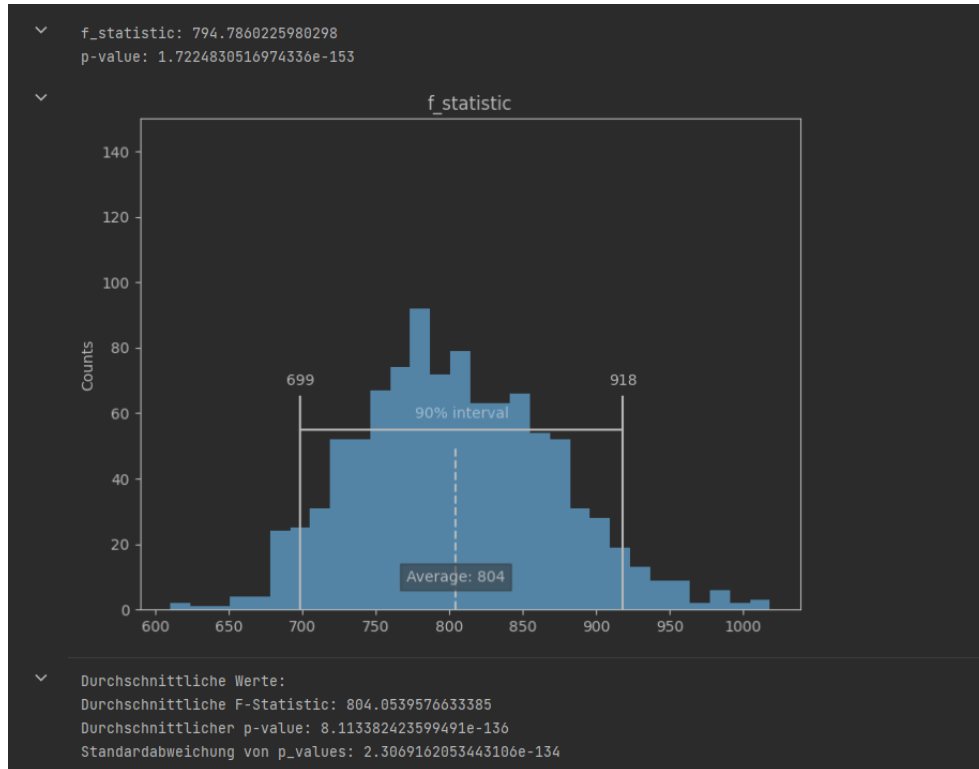
Aus den **Korrelationsdaten** und das **Streudiagramm** konnte man einen **gewissen Zusammenhang** zwischen dem **Leitzins der der FED für Banken** und den **Adj. Close Price** des ETFs erkennen.

2. Anova - Test



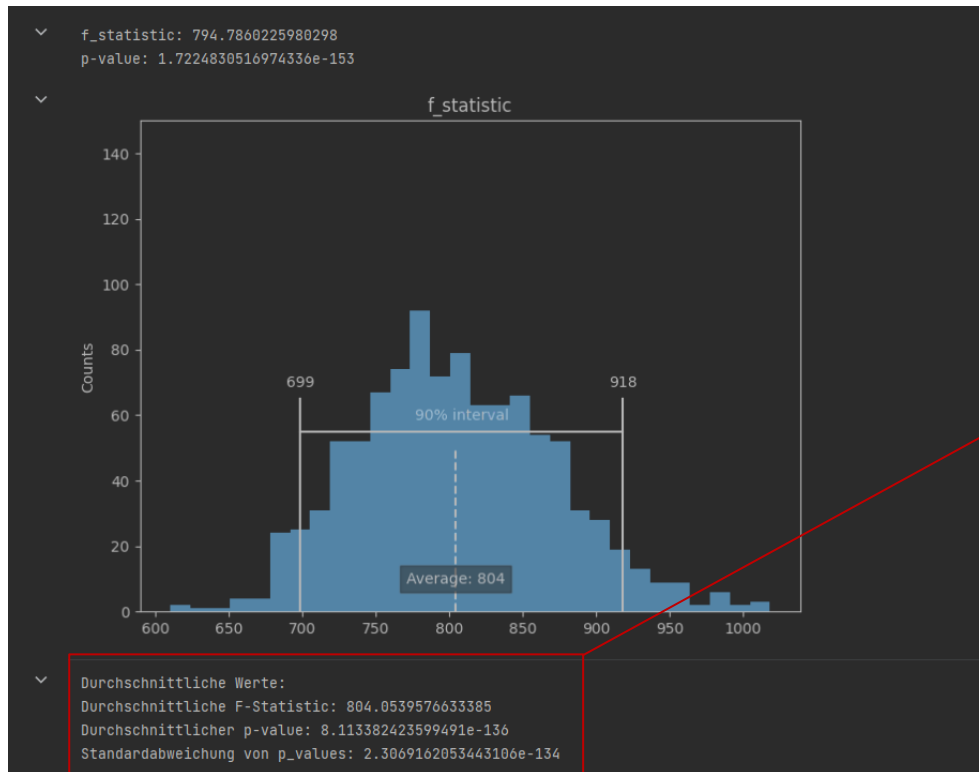
- Nullhypothese (H_0): Es besteht keine statistisch signifikante Beziehung zwischen den Closed Adj.Value des MSCI World und dem Leitzins der FED. Der Leitzins der FED hat keinen signifikanten Einfluss auf den Adj. Closed Price des Aktienmarktes
- Alternativhypothese (H_1): Es besteht eine statistisch signifikante Beziehung zwischen Closed Adj.Value des MSCI World und dem Leitzins der FED. Der Leitzins der FED hat einen signifikanten Einfluss auf den Adj. Closed Price des Aktienmarktes

2. Anova - Test



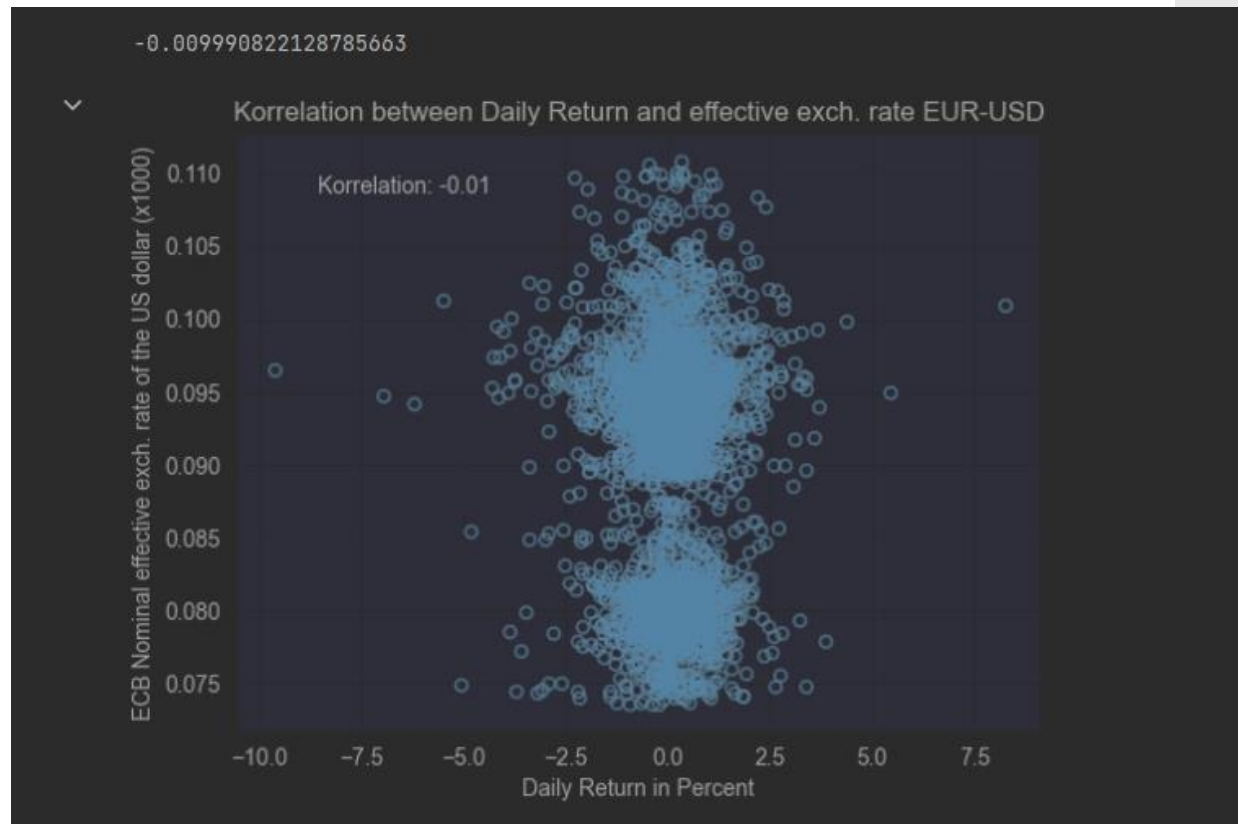
- hohe F-Statistik deutet darauf hin, dass die Varianz zwischen den Gruppen im Vergleich zur Varianz innerhalb der Gruppen signifikant ist
- durchschnittliche F-Statistik von 804.0539576633385 sehr hoch, was darauf hindeutet:
 - **Beziehung zwischen den Variablen besteht.**

2. Anova - Test



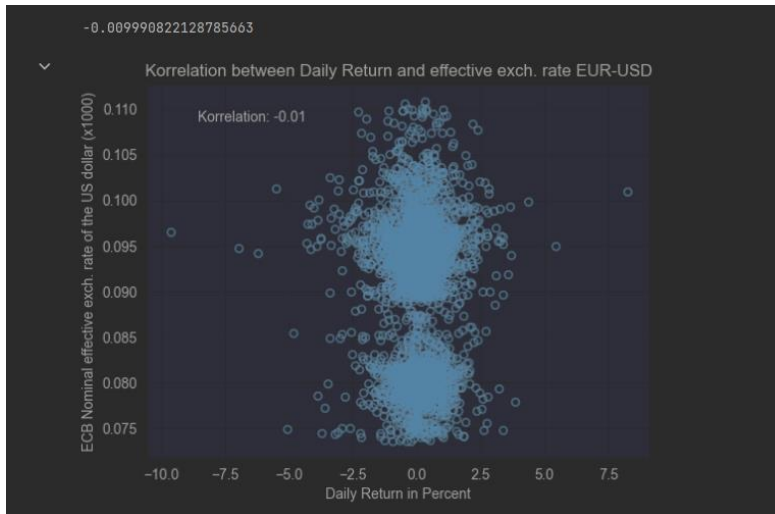
- durchschnittlicher p-value von $8.113382423599491e-136$
 - sehr klein ist
 - **Ablehnung der Nullhypothese (H_0)**
 - **Akzeptanz der Alternativhypothese (H_1)**
- Standardabweichung der p-Werte von $2.3069162053443106e-134$ zeigt die Streuung der p-Werte um den Durchschnitt:
 - Standardabweichung sehr klein ist → dass die p-Werte in den Bootstrap-Samples relativ eng um den Durchschnitt gruppiert
 - konsistente und stabile Ergebnisse in Bezug auf die p-Werte
 - stärkt die **Vertrauenswürdigkeit der statistischen Tests** bzw. die Hypothesenüberprüfungen.

2. Chi-Square Test



Aus den **Korrelationsdaten** und das **Streudiagramm** konnte man einen **gewissen Zusammenhang** zwischen dem **Währungstausch EUR-USD** und den **Adj. Close Price** des ETFs erkennen.

2. Chi-Square Test



- Nullhypothese (H_0): Es besteht kein Zusammenhang zwischen der Rendite des ETFs und den Währungsumtauschdaten.
- Alternativhypothese (H_A): Es besteht ein Zusammenhang zwischen der Rendite des ETFs und den Währungsumtauschdaten.

FH **FACHHOCHSCHULE**
ERFURT UNIVERSITY
OF APPLIED SCIENCES
Angewandte
Informatik

Daily Return ↕	Cumulative Return ↕		Change ↕	Comparison ↕
<null>	<null>		0.0	stabil
0.0	0.0		0.0	stabil
0.0	0.0		0.0	stabil
0.0	0.0		0.0	stabil
0.0	0.0		0.0	stabil
0.0	0.0		0.0	stabil
0.0	0.0		0.0	stabil
0.0	0.0		0.0	stabil
0.0	0.0		0.0	stabil
0.0	0.0		0.0	stabil
0.0	0.0		0.0	stabil
0.0	0.0		0.0	stabil
0.0	0.0		0.0	stabil
0.0	0.0		0.0	stabil
0.0	0.0		0.0	stabil
0.0	0.0		0.0	stabil
0.0	0.0		0.0	stabil
0.0	0.0		0.0	stabil
0.0	0.0		0.0	stabil
0.2324288339586781	0.2324288339586866	0.2324288339586781	positive	
-1.3913043478260778	-1.1588755138673967	-1.623733181784756	negative	
-0.4115167548501047	-1.570392268717502	0.9797875929759732	positive	
-0.5903246404766972	-2.1607169091942047	-0.1788078856265925	negative	
0.2969061757719696	-1.863810733422244	0.8872308162486668	positive	
-1.657773928820239	-3.521584662242483	-1.9546801045922086	negative	
0.9632690569976432	-2.558315605244843	2.6210429858178825	positive	
-0.1192605843768609	-2.6775761896217	-1.082529641374504	negative	
0.0	-2.6775761896217	0.1192605843768609	positive	
-1.432835820895513	-4.1104120105172175	-1.432835820895513	negative	
0.0	-4.1104120105172175	1.432835820895513	positive	
1.3930950938824764	-2.717316916634738	1.3930950938824764	positive	
0.2389545997610524	-2.478362316873695	-1.154140494121424	negative	
0.0	-2.478362316873695	-0.2389545997610524	negative	

19

2. Chi-Square Test

```
✓ Zusammenfassung Vergleich der Renditen:  
Status  negative  positive  stabil  
Anzahl    1794    1635     25  
  
Zusammenfassung Vergleich der Exchange Rate USD-EUR:  
Status  empty  negative  positive  
Anzahl  1479   3222     256
```

Zusammenfassung der Spalten „Comparison“ in Klassen empty, negative positive

2. Chi-Square Test

```
Comparison negative positive stabil
Comparison
empty          548      477      8
negative       1116     1045      4
positive       130      113     13
```

Chi-Quadrat-Test Ergebnis:

Chi2-Statistik: 77.40614180275973

p-Wert: 6.170125165820085e-16

Durchschnittliche Werte:

Mittelwert der Chi-Quadrat-Statistiken: 77.40614180275972

Mittelwert der p-Werte: 6.170125165820084e-16

Standardabweichung von p_values: 9.860761315262648e-32

- Chi-Quadrat-Statistik von 77.40614180275973:
 - relativ hoher Chi-Quadrat-Statistik Wert
 - beobachteten Häufigkeiten deutlich von den erwarteten Häufigkeiten abweichen und somit auf einen Zusammenhang hindeuten.
- extrem niedrigen p-Wert von 6.170125165820085e-16:
 - **sehr kleinen** p-Wert, geringer als Signifikanzniveau von 0,05
 - **Verwerfung Nullhypothese**
 - **Akzeptanz Alternativhypothese**
 - Die, was bedeutet, dass die Rendite des ETFs tatsächlich von den Währungsumtauschdaten beeinflusst wird

2. Chi-Square Test

```
Comparison negative positive stabil
Comparison
empty          548      477      8
negative       1116     1045      4
positive       130      113     13
```

Chi-Quadrat-Test Ergebnis:

Chi2-Statistik: 77.40614180275973

p-Wert: 6.170125165820085e-16

Durchschnittliche Werte:

Mittelwert der Chi-Quadrat-Statistiken: 77.40614180275972

Mittelwert der p-Werte: 6.170125165820084e-16

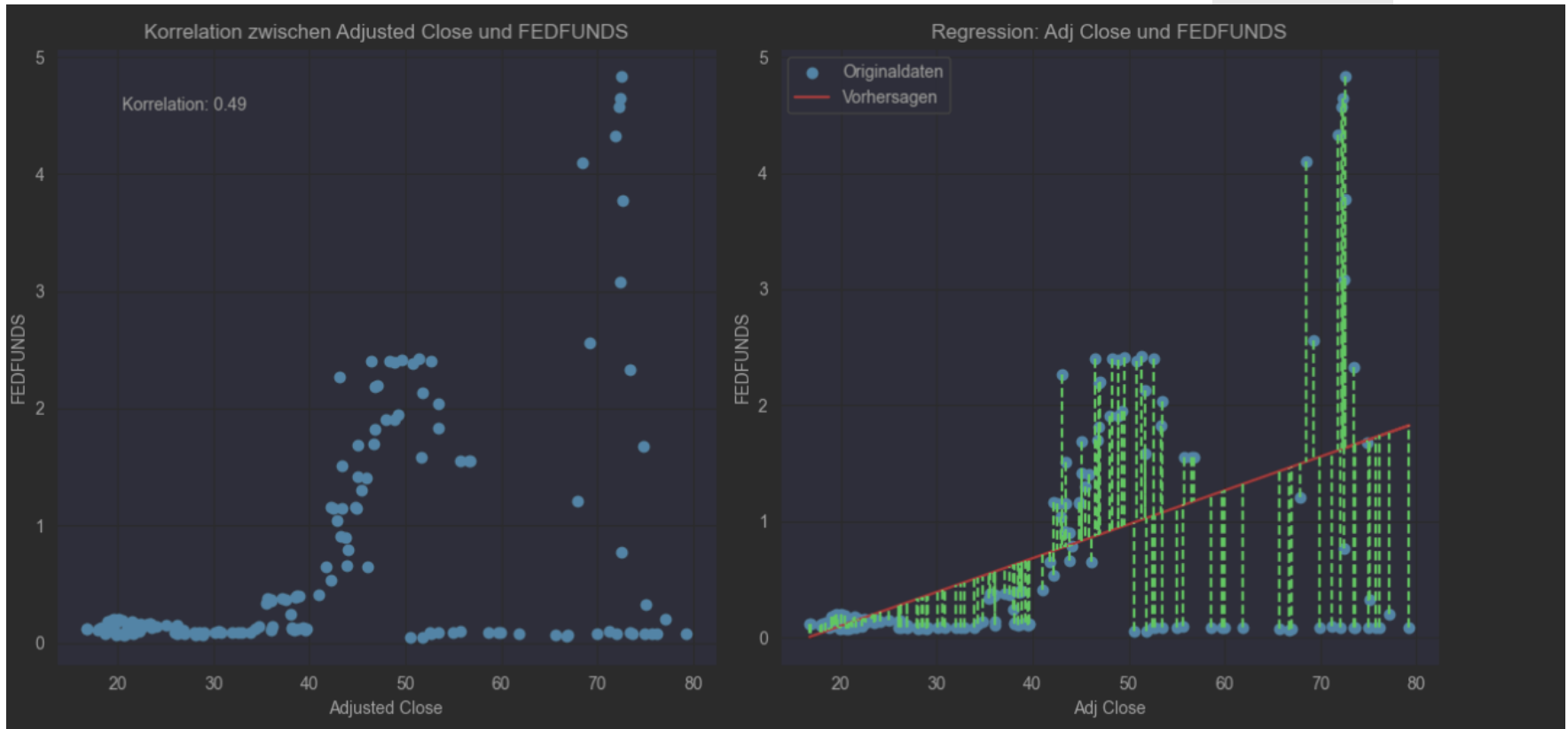
Standardabweichung von p_values: 9.860761315262648e-32

- Standardabweichung der p-Werte von 9.860761315262648e-32:
 - Standardabweichung **sehr klein** ist
 - die p-Werte in den Bootstrap-Samples relativ eng um den Durchschnitt gruppiert
 - Bootstrap-Samples liefert konsistente und stabile Ergebnisse
 - stärkt die **Vertrauenswürdigkeit der statistischen Tests bzw. die Hypothesenüberprüfungen**

3. Zusammenfassung Ergebnisse Chapter 3 - Modelling and Evaluation - Prädiktion

3. Einfache Lineare Regression

```
[163 rows x 2 columns]  
Intercept: -0.485  
Coefficient Exposure: 0.029
```



3. Einfache Lineare Regression

```
[163 rows x 2 columns]  
Intercept: -0.485  
Coefficient Exposure: 0.029
```

eine positive Beziehung
zwischen der Einflussgröße
und der Zielgröße

Steigt die Einflussgröße,
steigt auch die Zielgröße.

3. Multiple Linear Regression

```
multiple_regression_combined_data = pd.read_csv('data/generatedDatasets/multiple_regression_combined_data.csv')
multiple_regression_combined_data = multiple_regression_combined_data.dropna(axis=0)

predictors = ['ezb_funds', 'fed_funds_for_banks', 'fed_funds_for_customer']
outcome = 'Adj Close'

adjClose_etf_lm = LinearRegression()
adjClose_etf_lm.fit(multiple_regression_combined_data[predictors], multiple_regression_combined_data[outcome])

print(f'Intercept: {adjClose_etf_lm.intercept_:.3f}')
print('Coefficients:')
for name, coef in zip(predictors, adjClose_etf_lm.coef_):
    print(f' {name}: {coef}')
```

```
✓ Intercept: 28.944
Coefficients:
ezb_funds: -12.76264906264126
fed_funds_for_banks: 7.228215416972676
fed_funds_for_customer: 3.248414184528484
```

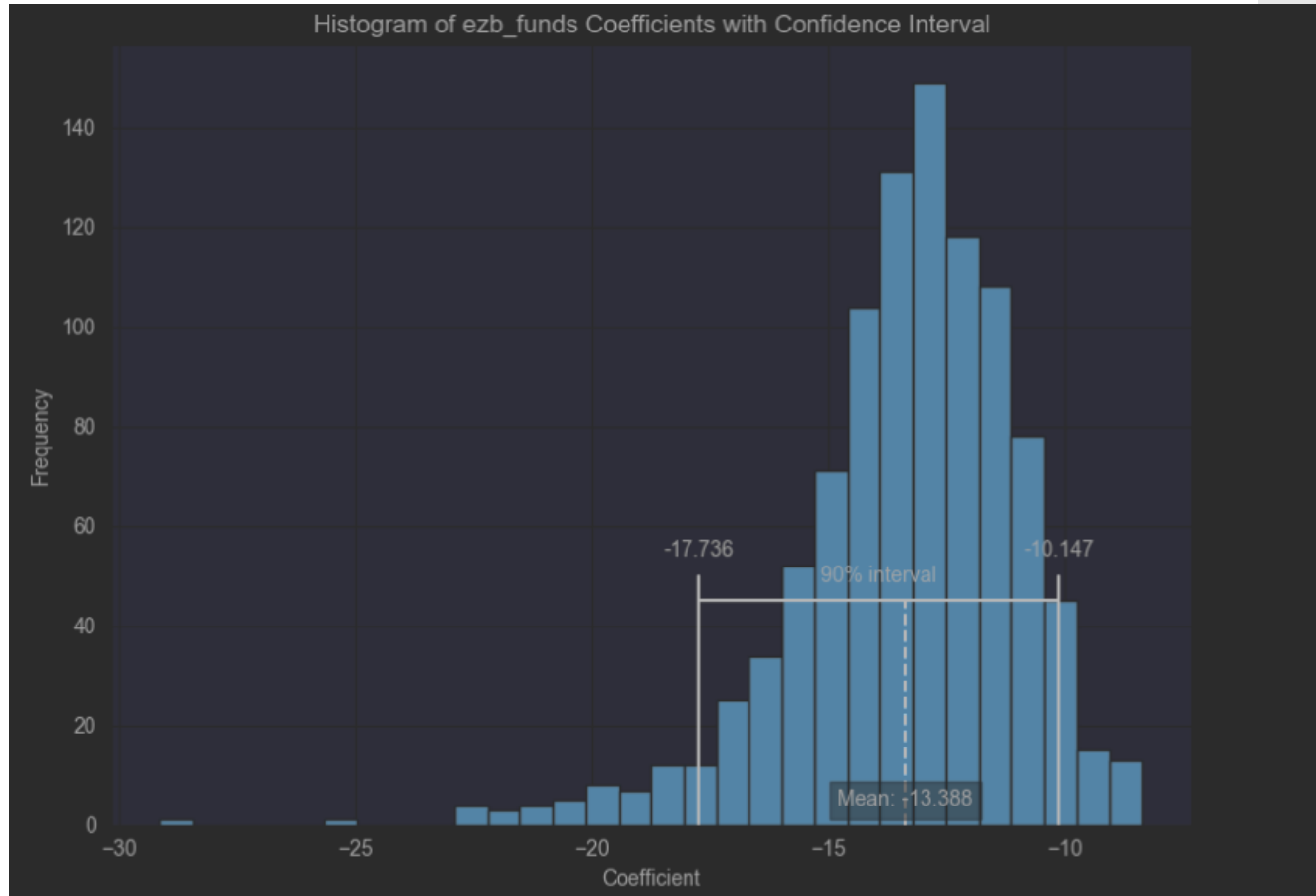
Steigt (negativ) die Einflussgröße, **sinkt** auch die Zielgröße

3. Multiple Linear Regression

```
✓ Intercept: 28.944  
Coefficients:  
  ezb_funds: -12.76264906264126  
  fed_funds_for_banks: 7.228215416972676  
  fed_funds_for_customer: 3.248414184528484
```

Koeffizient	Deutung
Leitzins EZB (Consumer)	negative Beziehung zwischen der Leitzins EZB und Adj. Close Price
Leitzins FED für Banken	positive Beziehung zwischen der Leitzins FED und Adj. Close Price der Banken
Leitzins FED für Consumer	positive Beziehung zwischen der Leitzins FED für Consumer und Adj. Close Price

3. Konfidenzintervalle der MLR-Koeffizienten

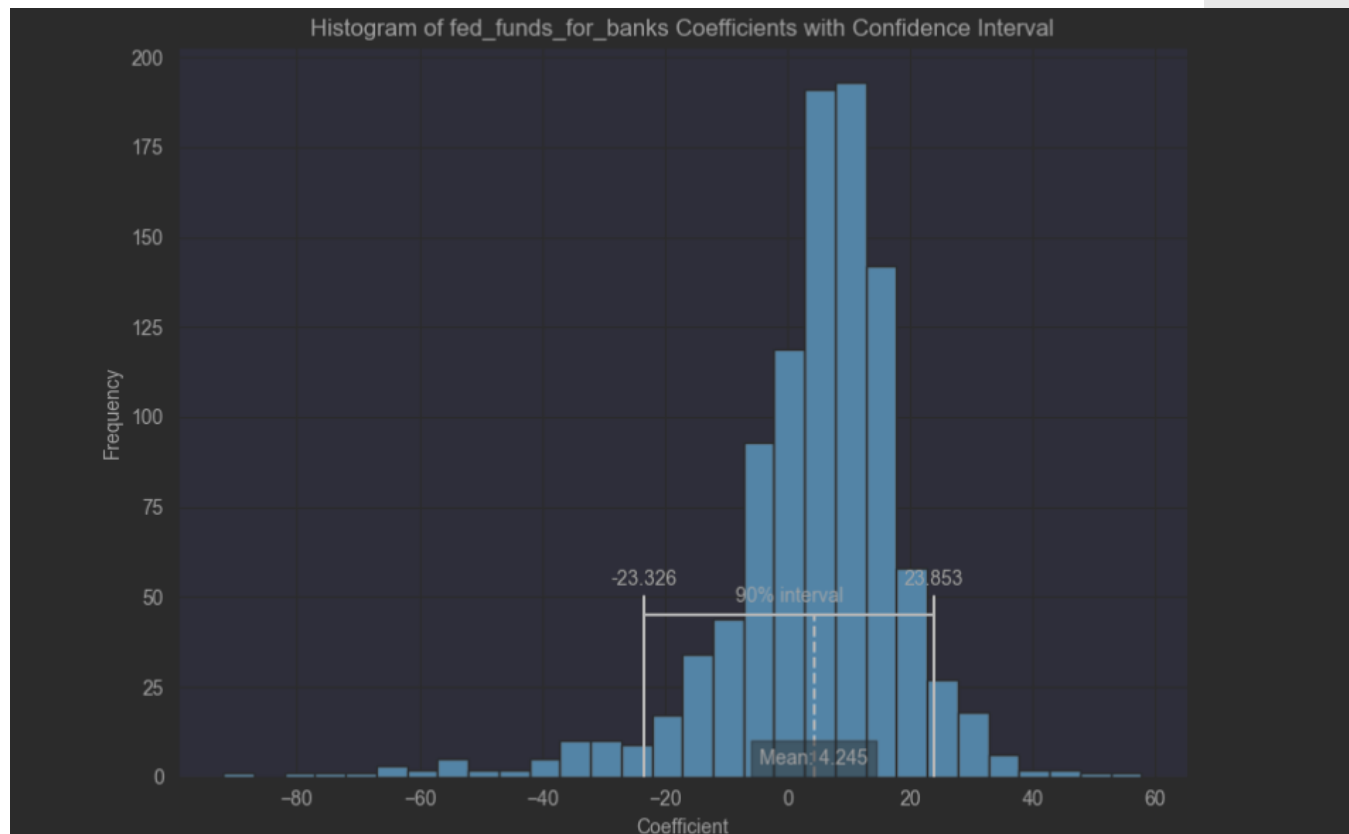


starke Abweichung des Konfidenzintervalle für Koeffizient des EZB-Leitzins:

→ Schätzung des **Koeffizienten weniger präzise**

→ **Unsicherheit** des wahren des Koeffizient Wert ist **größer**

3. Konfidenzintervalle der MLR-Koeffizienten

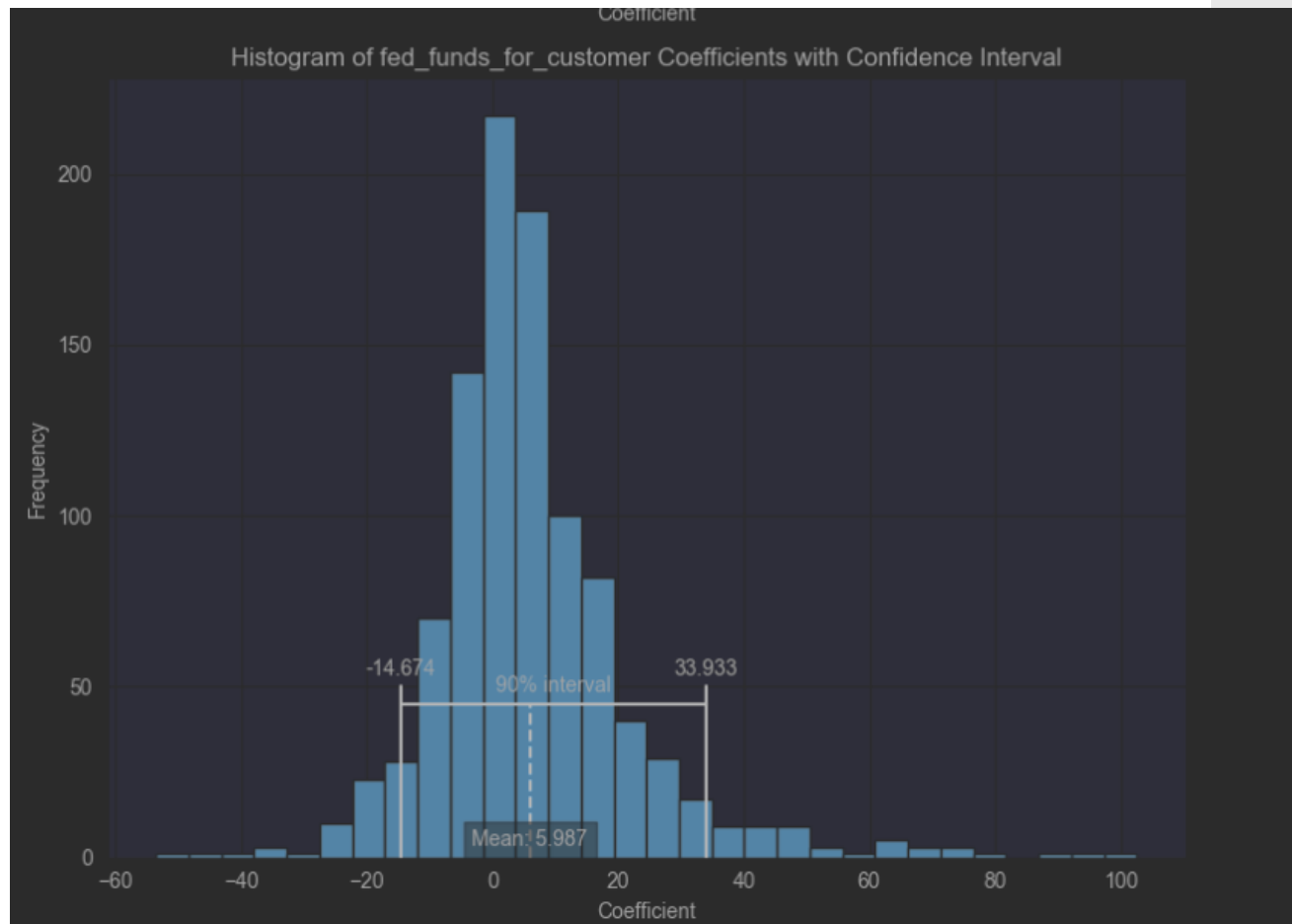


starke Abweichung des Konfidenzintervalle für Koeffizient des FED-Leitzins für Banken:

→ Schätzung des **Koeffizienten weniger präzise**

→ **Unsicherheit** des wahren des Koeffizient Wert ist **größer**

3. Konfidenzintervalle der MLR-Koeffizienten



starke Abweichung des Konfidenzintervalle für Koeffizient des FED-Leitzins für Business/Consumer:

→ Schätzung des **Koeffizienten weniger präzise**

→ **Unsicherheit** des wahren des Koeffizient Wert ist **größer**

3. Gewichtete Multiple Linear Regression

```
# Definiere die Prädiktoren und das Ziel
predictors = ['ezb_funds', 'fed_funds_for_banks', 'fed_funds_for_customer']
outcome = 'Adj Close'
```

```
# Gewichtung basierend auf dem Volumen
weights = multiple_regression_combined_data['Volume']

# Initialisiere das lineare Regressionsmodell mit gewichteten Daten
adjClose_etf_wt = LinearRegression()
adjClose_etf_wt.fit(multiple_regression_combined_data[predictors], multiple_regression_combined_data[outcome], sample_weight=weights)

# Zeige die Ergebnisse an
print('Berechnung basierend auf der Gewichtung des Volumens')
print(f'Intercept: {adjClose_etf_wt.intercept_.3f}')
print('Coefficients:')
for name, coef in zip(predictors, adjClose_etf_wt.coef_):
    print(f'{name}: {coef}')
```

Gewichtung nach
Volumen des ETFS

```
▼ Berechnung basierend auf der Gewichtung des Volumens
Intercept: 87.835
Coefficients:
ezb_funds: 6.512985853309622
fed_funds_for_banks: 8.114393693831994
fed_funds_for_customer: -9.556114176848865
```

Steigt die Einflussgröße,
steigt auch die Zielgröße

Steigt (negativ) die
Einflussgröße, **sinkt** auch
die Zielgröße

3. Gewichtete Multiple Linear Regression

```
# Definiere die Prädiktoren und das Ziel
predictors = ['ezb_funds', 'fed_funds_for_banks', 'fed_funds_for_customer']
outcome = 'Monthly Return'
```

Gewichtung nach
Zeitliche Auflösung
des ETFs

```
train_data = multiple_regression_combined_data[multiple_regression_combined_data['Date'] <= multiple_regression_combined_data['Date'].max()]

# Modell initialisieren und trainieren
# train_data[predictors] sind die unabhängigen Variablen (Prädiktoren)
# train_data[outcome] ist die abhängige Variable (Ziel).
monthly_return_wl = LinearRegression()
monthly_return_wl.fit(train_data[predictors], train_data[outcome])

# Vorhersage für den 1. August 2023 treffen
# [pd.to_datetime('2023-08-01'), 0, 0, 0]], columns=['Date'] + predictors ist notwendig um die Datenstruktur für das trainierten Modell zu verwenden
# Wert 0 wird für alle Prädiktoren (ezb_funds, fed_funds_for_banks, fed_funds_for_customer) verwendet, da es sich um eine Vorhersage handelt und keine ta
prediction_data = pd.DataFrame([[pd.to_datetime('2023-08-01'), 0, 0, 0]], columns=['Date'] + predictors)
prediction = monthly_return_wl.predict(prediction_data[predictors])
```

Berechnung basierend auf der Gewichtung des historischen Zeitrahmens:

Intercept: 5.7913023987362715

Coefficients:

ezb_funds: -0.104806680694586

fed_funds_for_banks: 1.0925324558159264

fed_funds_for_customer: -1.4527754089689695

Vorhersage für den 01.08.2023:

Monatliche Rendite: 5.79

Steigt die Einflussgröße,
steigt auch die Zielgröße

Steigt (negativ) die
Einflussgröße, **sinkt** auch
die Zielgröße

3. Gewichtete Multiple Lineare Regression

```
Berechnung basierend auf der Gewichtung des historischen Zeitrahmens:  
Intercept: 5.7913023987362715  
Coefficients:  
  ezb_funds: -0.104806680694586  
  fed_funds_for_banks: 1.0925324558159264  
  fed_funds_for_customer: -1.4527754089689695
```

```
Vorhersage für den 01.08.2023:  
Monatliche Rendite: 5.79
```

Prognose mittels
historischer Daten und
Einflussgrößen

3. Lineare Diskriminanzanalyse

```
multiple_regression_combined_data_without_stabil = multiple_regression_combined_data[multiple_regression_combined_data['Comparison'] != 'stabil']
predictors = ['ezb_funds', 'fed_funds_for_banks']
outcome = 'Comparison'

# Entferne alle Zeilen mit Null- oder NaN-Werten
multiple_regression_combined_data.dropna(inplace=True)

# Daten für die Diskriminanzanalyse auswählen
X = multiple_regression_combined_data_without_stabil[predictors]
y = multiple_regression_combined_data_without_stabil[outcome]

etf_lda = LinearDiscriminantAnalysis()
etf_lda.fit(X, y)

# je größer der Score desto besser ist der Aussagekraft
# je kleiner der Score desto schlechter ist der Aussagekraft

print(pd.DataFrame(etf_lda.scalings_, index=X.columns))
```

0

ezb_funds	1.460757
fed_funds_for_banks	-0.742099

3. Lineare Diskriminanzanalyse

```
                                0  
ezb_funds          1.460757  
fed_funds_for_banks -0.742099
```

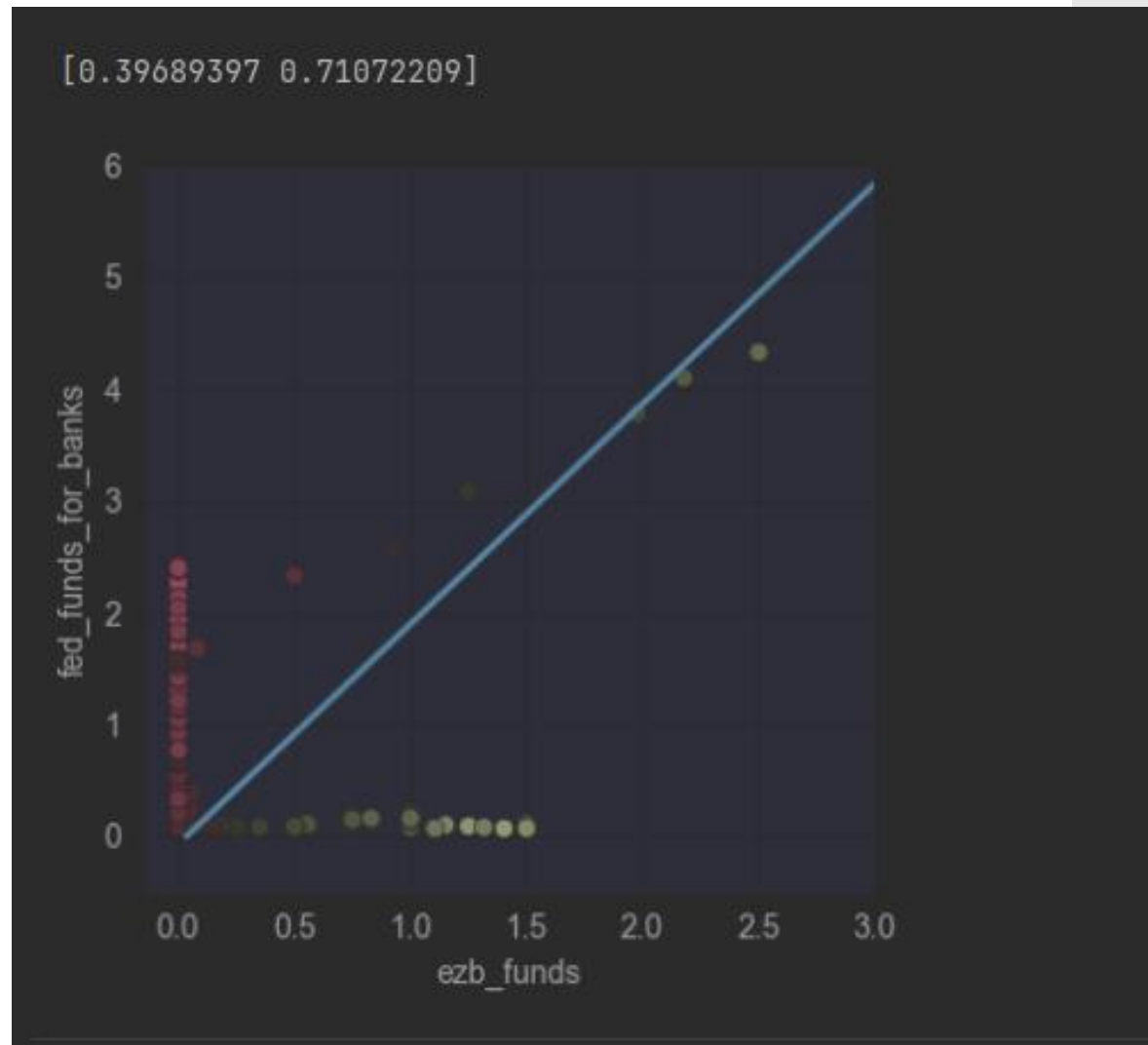
- **positiver** Koeffizient ezb_funds
 - Erhöhung des Wertes von ezb_funds führt zu einer Erhöhung der Diskriminanzfunktion
 - **höhere Ausprägung** von ezb_funds führt zu einer **höheren Wahrscheinlichkeit**, dass das Ergebnis der Variable „Comparison“ in die Klasse **positiv reinfällt**

3. Lineare Diskriminanzanalyse

```
                                0  
ezb_funds                    1.460757  
fed_funds_for_banks -0.742099
```

- **negativer** Koeffizient fed_funds_for_banks:
 - Erhöhung des Wertes von ezb_funds führt zu einer Abnahme der Diskriminanzfunktion
 - **höhere Ausprägung** von ezb_funds führt zu einer **höheren Wahrscheinlichkeit**, dass das Ergebnis der Variable „Comparison“ in die Klasse **negativ reinfällt**

3. Lineare Diskriminanzanalyse



3. Lineare Diskriminanzanalyse

Confusion Matrix (Accuracy 0.6687)

	Prediction	
Actual	negative	positive
negative	0	53
positive	0	107

Konfusionsmatrix, LDA - Modell sagt mit einer 66,87% der Fälle korrekt vor

3. Lineare Diskriminanzanalyse

	Labels	Precision	Recall	F-Score	Support
0	negative	0.00000	0.0	0.000000	53
1	positive	0.66875	1.0	0.801498	107

- Precision gibt an, wie viele der als positiv vorhergesagten Fälle tatsächlich positiv sind:
 - die negative Klasse beträgt die Precision 0.0, negativ vorhergesagten Fällen kein einziger korrekt ist.
 - die positive Klasse beträgt die Precision $0.66875 = 66,875\%$ der als positiv vorhergesagten Fälle tatsächlich positiv sind
- Recall (Sensitivität) gibt an, wie viele der tatsächlich positiven Fälle vom Modell korrekt erkannt wurde:
 - Recall für die negative Klasse 0.0 kein einziger negativer Fall richtig ist
 - die positive Klasse beträgt der Recall 1.0 alle positiven Fälle korrekt erkannt wurden

3. Naiver Bayes

Prädikate

Outcome

ezb_funds	fed_funds_for_banks	fed_funds_for_customer	Monthly Return	Comparison
1.0	0.12	3.25	0.0	stabil
1.0	0.12	3.25	0.6268656716417936	positive
1.0	0.12	3.25	6.8525660041530765	positive
1.0	0.11	3.25	-0.3886674069961288	negative
1.0	0.13	3.25	2.3968783502297453	positive
1.0	0.16	3.25	7.076749750857409	positive
1.0	0.2	3.25	2.135231316725969	positive
1.0	0.2	3.25	-2.638133399701337	negative
1.0	0.18	3.25	-2.760731224986267	negative
1.0	0.18	3.25	0.9989432176656164	positive
1.0	0.19	3.25	0.9890734507586324	positive
1.0	0.19	3.25	0.2061907216494854	positive
1.0	0.19	3.25	0.6172736307986944	positive
1.0	0.19	3.25	4.754606582546335	positive
1.0	0.18	3.25	4.929233772571994	positive
1.0	0.17	3.25	-0.7441860465116286	negative
1.0	0.16	3.25	2.5304545454545533	positive
1.0	0.14	3.25	-3.839113520983256	negative
1.15	0.1	3.25	-0.3327091096621087	negative
1.25	0.09	3.25	1.049127374779547	positive
1.25	0.09	3.25	-2.170839916430388	negative
1.403225806451613	0.07	3.25	-0.6753497346840356	negative
1.5	0.1	3.25	-7.285089849441477	negative
1.5	0.08	3.25	-1.990576217915139	negative
1.5	0.07	3.25	6.306788151084347	positive
1.3166666666666669	0.08	3.25	0.5027702615771821	positive
1.1048387096774193	0.07	3.25	4.05202101050526	positive
1.0	0.08	3.25	3.7500049879810025	positive
1.0	0.1	3.25	2.919365152919377	positive
1.0	0.13	3.25	1.4408015056641776	positive
1.0	0.14	3.25	-0.621402546764197	negative
1.0	0.16	3.25	-2.456449417438561	negative

Zusammensetzung der Daten zu einer neuen Tabelle für Naiver Bayes Modell

3. Naiver Bayes

```
predictors = ['ezb_funds', 'fed_funds_for_banks']  
outcome = 'Comparison'
```

```
# Aufteilung in Trainings- und Testdaten  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)  
  
# Naive Bayes Klassifikator erstellen und trainieren  
naive_model = MultinomialNB(alpha=1e-10, fit_prior=False)  
naive_model.fit(X_train, y_train)  
  
# Vorhersagen auf Testdaten machen  
y_pred = naive_model.predict(X_test)  
  
# Genauigkeit der Vorhersagen auswerten  
accuracy = accuracy_score(y_test, y_pred)
```

3. Naiver Bayes

```
predicted class:  negative
predicted probabilities
      negative  positive
0  0.501568  0.498432
```

*Klassifikationsvorhersage über das Naiver
Bayer Modell*

3. Naiver Bayes

```
Confusion Matrix (Accuracy 0.5125)
```

```
          Prediction
Actual    negative positive
negative     36       17
positive     61       46
```

Konfusionsmatrix, Naiven-Bayes-Modell sagt mit einer 66,87% der Fälle korrekt vor

3. Naiver Bayes

	Labels	Precision	Recall	F-Score	Support
0	negative	0.371134	0.679245	0.480000	53
1	positive	0.730159	0.429907	0.541176	107

- Precision (Relevanz) gibt an, wie viele der als positiv vorhergesagten Fälle tatsächlich positiv sind:
 - negative Klasse beträgt die Precision 37,1134% der als negativ vorhergesagten Fälle tatsächlich negativ sind
 - die positive Klasse beträgt die Precision 73,0159% der als positiv vorhergesagten Fälle tatsächlich positiv sind
- Recall (Sensitivität) gibt an, wie viele der tatsächlich positiven Fälle vom Modell korrekt erkannt wurden:
 - negative Klasse 67,9245% der tatsächlich negativen Fälle richtig erkannt wurden
 - positive Klasse beträgt der Recall 42,9907% der tatsächlich positiven Fälle richtig erkannt wurden

3. Logistische Regression

Date	hpi_percentage_change_monthly	Adj Close	Volume	ezb_funds	fed_funds_for_banks	fed_funds_for_customer	Monthly Return	Monthly Return Comparison
2023-03-01	6.9	72.377998	10135602.0	3.161290322580645	4.65	7.75	0.1854756027292656	positive
2023-02-01	8.5	72.244003	5824871.0	2.875	4.57	7.5	0.5931688352061082	positive
2023-01-01	8.6	71.818001	6731261.0	2.5	4.33	7.5	4.877484953305502	positive
2022-12-01	9.2	68.477997	6728267.0	2.177419354838709	4.1	7.0	-5.703670929344351	negative
2022-11-01	10.1	72.620003	8034631.0	1.975	3.78	6.25	0.3011007983204828	positive
2022-10-01	10.6	72.402	11679350.0	1.25	3.08	6.25	4.536526002410455	positive
2022-09-01	9.9	69.260002	11304564.0	0.925	2.56	5.5	-5.676307325866048	negative
2022-08-01	9.1	73.428001	8035747.0	0.5	2.33	5.5	-1.9024155093684445	negative
2022-07-01	8.9	74.851997	9847150.0	0.0806451612903225	1.68	4.75	10.261312781389	positive
2022-06-01	8.6	67.886002	10987360.0	0.0	1.21	4.0	-6.353797923376277	negative
2022-05-01	8.1	72.491997	10734418.0	0.0	0.77	3.5	-3.4598470347501435	negative
2022-04-01	7.4	75.089996	10364008.0	0.0	0.33	3.5	-2.6044826065526987	negative
2022-03-01	7.4	77.098	16536926.0	0.0	0.2	3.25	4.678759572041489	positive
2022-02-01	5.9	73.652	21606509.0	0.0	0.08	3.25	-1.7108362481873731	negative
2022-01-01	5.1	74.933998	18793358.0	0.0	0.08	3.25	-5.441289012914496	negative
2021-12-01	5.0	79.246002	11783850.0	0.0	0.08	3.25	4.002838784225408	positive
2021-11-01	4.9	76.195999	9895153.0	0.0	0.08	3.25	0.6445804934749327	positive
2021-10-01	4.1	75.708	8102372.0	0.0	0.08	3.25	5.129558141472495	positive
2021-09-01	3.4	72.014	8076837.0	0.0	0.08	3.25	-1.9337065101644155	negative
2021-08-01	3.0	73.433998	7276449.0	0.0	0.09	3.25	3.114461506636369	positive
2021-07-01	2.2	71.216003	8410984.0	0.0	0.1	3.25	1.833164178601656	positive
2021-06-01	1.9	69.933998	6275746.0	0.0	0.08	3.25	4.738656795206983	positive
2021-05-01	2.0	66.769997	7636691.0	0.0	0.06	3.25	-0.2986471509833078	negative
2021-04-01	1.6	66.970001	7249398.0	0.0	0.07	3.25	1.942340244162319	positive
2021-03-01	1.3	65.694	8767454.0	0.0	0.07	3.25	6.1498190279214136	positive
2021-02-01	0.9	61.888	6523875.0	0.0	0.08	3.25	3.205149584764696	positive
2021-01-01	0.9	59.966	9271385.0	0.0	0.09	3.25	0.4253717703184744	positive
2020-12-01	-0.3	59.712002	5950417.0	0.0	0.09	3.25	1.918483274416548	positive
2020-11-01	-0.3	58.588001	9148926.0	0.0	0.09	3.25	9.592222401650242	positive
2020-10-01	-0.3	53.459999	5141356.0	0.0	0.09	3.25	-2.743414400076662	negative
2020-09-01	-0.3	54.967999	6675954.0	0.0	0.09	3.25	-1.250362869172843	negative
2020-08-01	-0.2	55.664001	4518274.0	0.0	0.1	3.25	5.962080142349799	positive

Zusammensetzung der ETF, FED-Funds, EZB-Funds und EZB-HPI (Inflationsdaten) zu einer neuen Tabelle

3. Logistische Regression

```
predictors = ['ezb_funds', 'fed_funds_for_banks', 'fed_funds_for_customer', 'hpi_percentage_change_monthly']
outcome = 'Monthly Return Comparison'
```

Date	hpi_percentage_change_monthly	Adj Close	Volume	ezb_funds	fed_funds_for_banks	fed_funds_for_customer	Monthly Return	Monthly Return Comparison
2023-03-01	6.9	72.377998	10135602.0	3.161290322580645	4.65	7.75	0.1854756027292656	positive
2023-02-01	8.5	72.244003	5824871.0	2.875	4.57	7.5	0.5931688352061082	positive
2023-01-01	8.6	71.818001	6731261.0	2.5	4.33	7.5	4.877484953305502	positive
2022-12-01	9.2	68.477997	6728267.0	2.177419354838709	4.1	7.0	-5.703670929344351	negative
2022-11-01	10.1	72.620003	8034631.0	1.975	3.78	6.25	0.3011007983204828	positive
2022-10-01	10.6	72.402	11679350.0	1.25	3.08	6.25	4.536526002410455	positive
2022-09-01	9.9	69.260002	11304564.0	0.925	2.56	5.5	-5.676307325866048	negative
2022-08-01	9.1	73.428001	8035747.0	0.5	2.33	5.5	-1.9024155093684445	negative
2022-07-01	8.9	74.851997	9847150.0	0.0806451612903225	1.68	4.75	10.261312781389	positive
2022-06-01	8.6	67.886002	10987360.0	0.0	1.21	4.0	-6.353797923376277	negative
2022-05-01	8.1	72.491997	10734418.0	0.0	0.77	3.5	-3.4598470347501435	negative
2022-04-01	7.4	75.089996	10364008.0	0.0	0.33	3.5	-2.6044826065526987	negative
2022-03-01	7.4	77.098	16536926.0	0.0	0.2	3.25	4.678759572041489	positive
2022-02-01	5.9	73.652	21606509.0	0.0	0.08	3.25	-1.7108362481873731	negative
2022-01-01	5.1	74.933998	18793358.0	0.0	0.08	3.25	-5.441289012914496	negative
2021-12-01	5.0	79.246002	11783850.0	0.0	0.08	3.25	4.002838784225408	positive
2021-11-01	4.9	76.195999	9895153.0	0.0	0.08	3.25	0.6445804934749327	positive
2021-10-01	4.1	75.708	8102372.0	0.0	0.08	3.25	5.129558141472495	positive
2021-09-01	3.4	72.014	8076837.0	0.0	0.08	3.25	-1.9337065101644155	negative
2021-08-01	3.0	73.433998	7276449.0	0.0	0.09	3.25	3.114461506636369	positive
2021-07-01	2.2	71.216003	8410984.0	0.0	0.1	3.25	1.833164178601656	positive
2021-06-01	1.9	69.933998	6275746.0	0.0	0.08	3.25	4.738656795206983	positive
2021-05-01	2.0	66.769997	7636691.0	0.0	0.06	3.25	-0.2986471509833078	negative
2021-04-01	1.6	66.970001	7249398.0	0.0	0.07	3.25	1.942340244162319	positive
2021-03-01	1.3	65.694	8767454.0	0.0	0.07	3.25	6.1498190279214136	positive
2021-02-01	0.9	61.888	6523875.0	0.0	0.08	3.25	3.205149584764696	positive
2021-01-01	0.9	59.966	9271385.0	0.0	0.09	3.25	0.4253717703184744	positive
2020-12-01	-0.3	59.712002	5950417.0	0.0	0.09	3.25	1.918483274416548	positive
2020-11-01	-0.3	58.588001	9148926.0	0.0	0.09	3.25	9.592222401650242	positive
2020-10-01	-0.3	53.459999	5141356.0	0.0	0.09	3.25	-2.743414400076662	negative
2020-09-01	-0.3	54.967999	6675954.0	0.0	0.09	3.25	-1.250362869172843	negative
2020-08-01	-0.2	55.664001	4518274.0	0.0	0.1	3.25	5.962080142349799	positive

3. Logistische Regression

```
X = pd.get_dummies(etf_table_with_hpi_monthly_data_without_stabil[predictors], prefix='', prefix_sep='',
                    drop_first=True)
y = etf_table_with_hpi_monthly_data_without_stabil[outcome]

logit_reg = LogisticRegression(penalty='l2', C=1e42, solver='liblinear')
logit_reg.fit(X, y)

print('intercept ', logit_reg.intercept_[0])
print('classes', logit_reg.classes_)
pd.DataFrame({'coeff': logit_reg.coef_[0]},
              index=X.columns)
```

```
intercept 3.3491330671307993
classes ['negative' 'positive']
```

|< < 4 rows > >| 4 rows × 1 columns [pd.DataFrame](#)

	coeff
ezb_funds	0.336359
fed_funds_for_banks	0.908847
fed_funds_for_customer	-0.792358
hpi_percentage_change_monthly	-0.190698

3. Logistische Regression

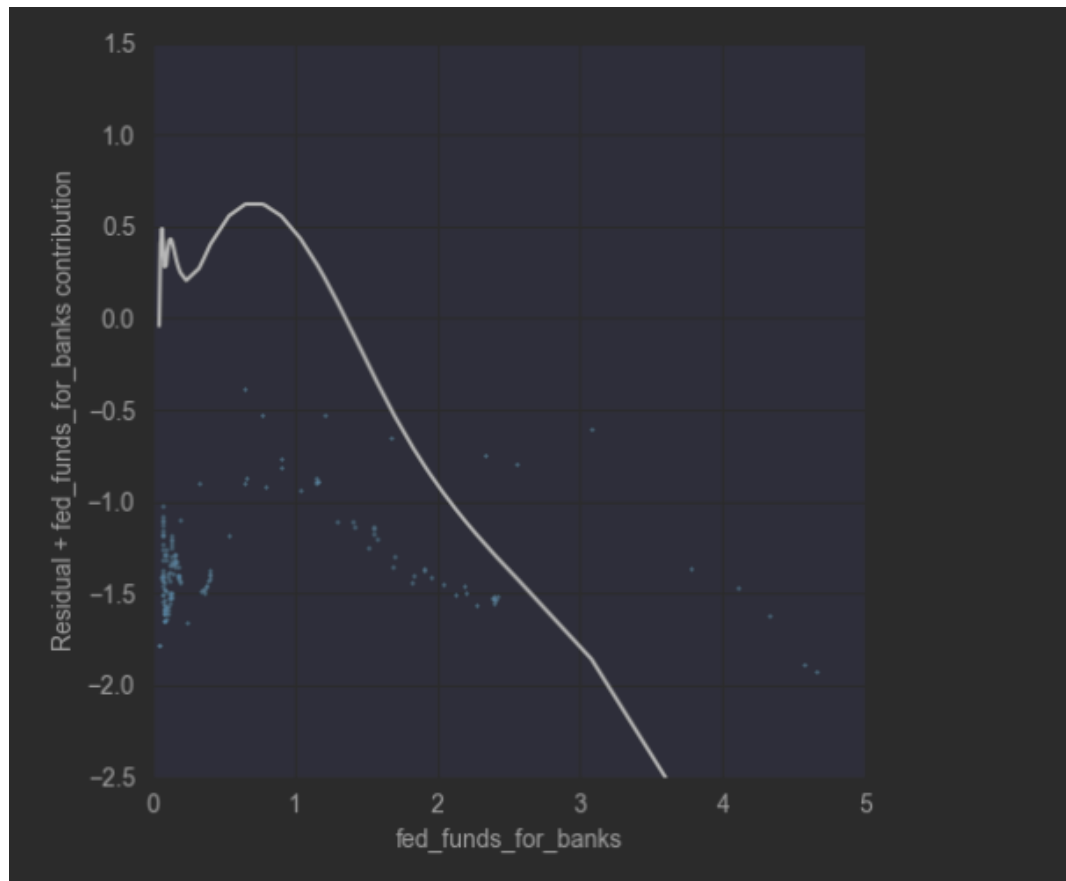
```
Predicted Class  
positive      148  
negative       12  
Name: count, dtype: int64
```

*Aus 160 Beobachtungen wurde
mittels Logistische Regression
148 Werte in den positiven Klasse
und 12 Werte in der negativen
Klasse vorhergesagt*

```
          negative    positive  
count    160.000000    160.000000  
mean     -1.126833    -0.411497  
std       0.196588     0.144656  
min      -1.449482    -1.010442  
25%      -1.242367    -0.396978  
50%      -1.172407    -0.370514  
75%      -1.115805    -0.340661  
max       -0.452648    -0.267477
```

*statistische Zusammenfassungen
der Ergebnisse der logistischen
Regression für die negative und
positive Klassen*

3. Logistische Regression



*partiellen Residuen Diagramm zwischen FED-Leitzins
für Banken und partiellen Residuen*

4. Bewertung der Ergebnisse des Data Minings in Bezug auf Ziele und Erfolgskriterien

4. Bewertung der Ergebnisse des Data Minings in Bezug auf Ziele und Erfolgskriterien

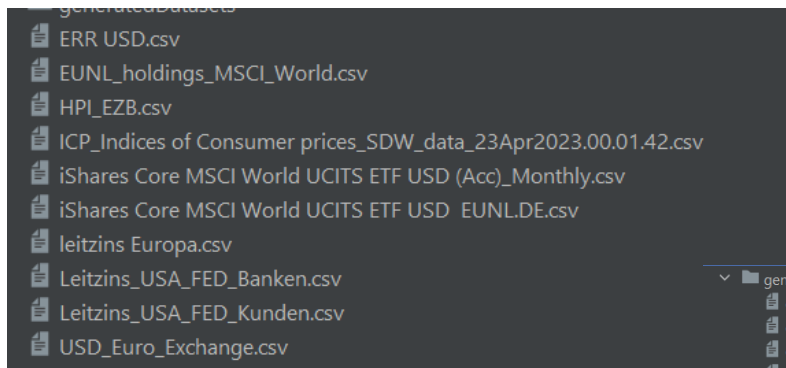
Ziele:

- geeignet Datensätze zu den Zielgrößen und Einflussgrößen finden und diese ggf. anpassen
- Modellierungsverfahren erarbeiten, um Korrelationen und Muster zu erkennen, um thesenspezifische Prognosen durchzuführen
- Interpretation der Ergebnisse und Schlussfolgerungen daraus ziehen

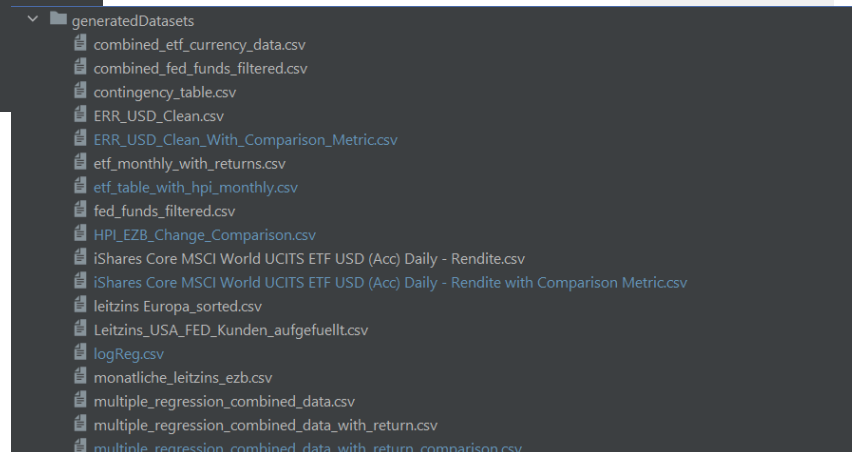
4. Bewertung der Ergebnisse des Data Minings in Bezug auf Ziele und Erfolgskriterien

Ziele:

- geeignet Datensätze zu den Zielgrößen und Einflussgrößen finden und diese ggf. anpassen



Ziel Datensätze



*Verschiedene Datensatzzusammensetzung für
unters. Analysen und Modelle*

4. Bewertung der Ergebnisse des Data Minings in Bezug auf Ziele und Erfolgskriterien

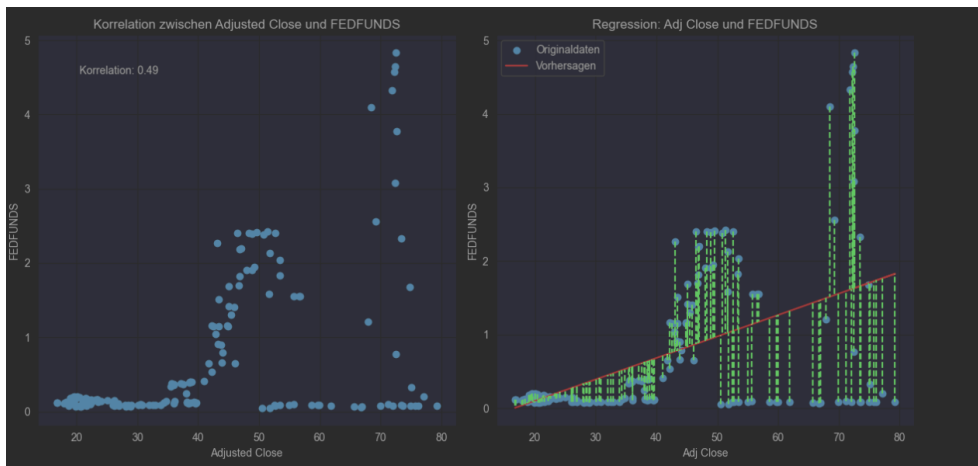
Ziele:

- geeignet Datensätze zu den Zielgrößen und Einflussgrößen finden und diese ggf. anpassen ✓
- Modellierungsverfahren erarbeiten, um Korrelationen und Muster zu erkennen, um thesenspezifische Prognosen durchzuführen
- Interpretation der Ergebnisse und Schlussfolgerungen daraus ziehen

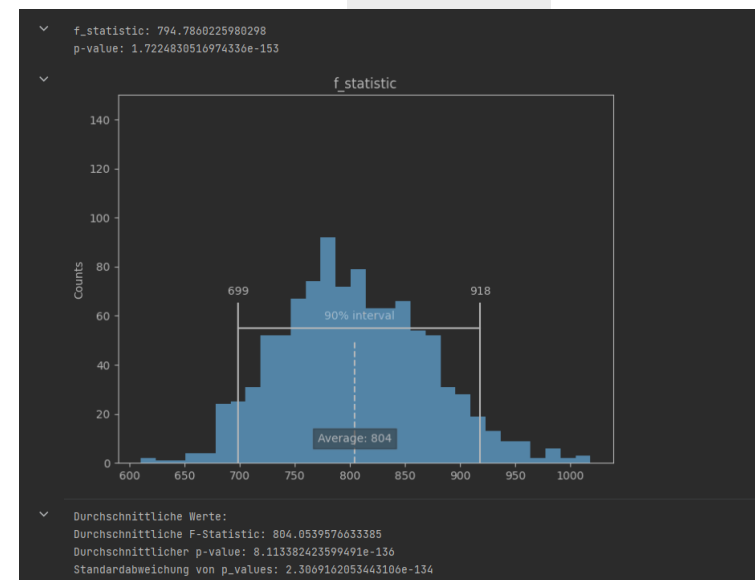
4. Bewertung der Ergebnisse des Data Minings in Bezug auf Ziele und Erfolgskriterien

Ziele:

- Modellierungsverfahren erarbeiten, um Korrelationen und Muster zu erkennen, um thesenspezifische Prognosen durchzuführen



Einfache lineare Regression



Ergebnis Anova-Test

4. Bewertung der Ergebnisse des Data Minings in Bezug auf Ziele und Erfolgskriterien

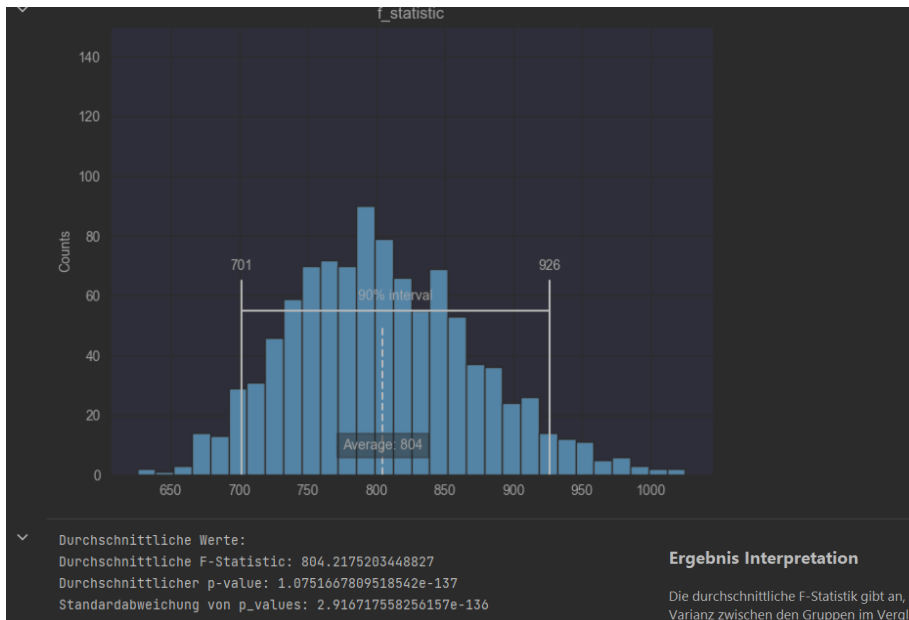
Ziele:

- geeignet Datensätze zu den Zielgrößen und Einflussgrößen finden und diese ggf. anpassen ✓
- Modellierungsverfahren erarbeiten, um Korrelationen und Muster zu erkennen, um thesenspezifische Prognosen durchzuführen ✓
- Interpretation der Ergebnisse und Schlussfolgerungen daraus ziehen

4. Bewertung der Ergebnisse des Data Minings in Bezug auf Ziele und Erfolgskriterien

Ziele:

- Interpretation der Ergebnisse und Schlussfolgerungen daraus ziehen



T-Test Ergebnis

Ergebnis Interpretation

Die durchschnittliche F-Statistik gibt an, wie stark die Varianz der Gruppenmittelwerte im Vergleich zur Varianz innerhalb der Gruppen ist. Eine hohe F-Statistik deutet darauf hin, dass die Varianz zwischen den Gruppen im Vergleich zur Varianz innerhalb der Gruppen signifikant ist. In diesem Fall ist die F-Statistik mit einer durchschnittlichen F-Statistik von 804.0539576633385 sehr hoch, was darauf hindeutet, dass es eine signifikante Beziehung zwischen den Variablen besteht.

Aus den Ergebnis durchschnittliche des p-value Werts von 8.113382423599491e-136, welches sehr klein ist, kann nun interpretiert werden, dass man die Nullhypothese (H0) ablehnt und die Alternativhypothese (H1) akzeptiert.

Die Standardabweichung der p-Werte von 2.3069162053443106e-134 zeigt die Streuung der p-Werte um den Durchschnitt. Da die Standardabweichung sehr klein ist, kann man sehen, dass die p-Werte in den Bootstrap-Samples relativ eng um den Durchschnitt gruppiert sind. Dies bedeutet, dass die Bootstrap-Samples konsistente und stabile Ergebnisse in Bezug auf die p-Werte liefern. Das stärkt die Vertrauenswürdigkeit der statistischen Tests bzw. die Hypothesenüberprüfungen.

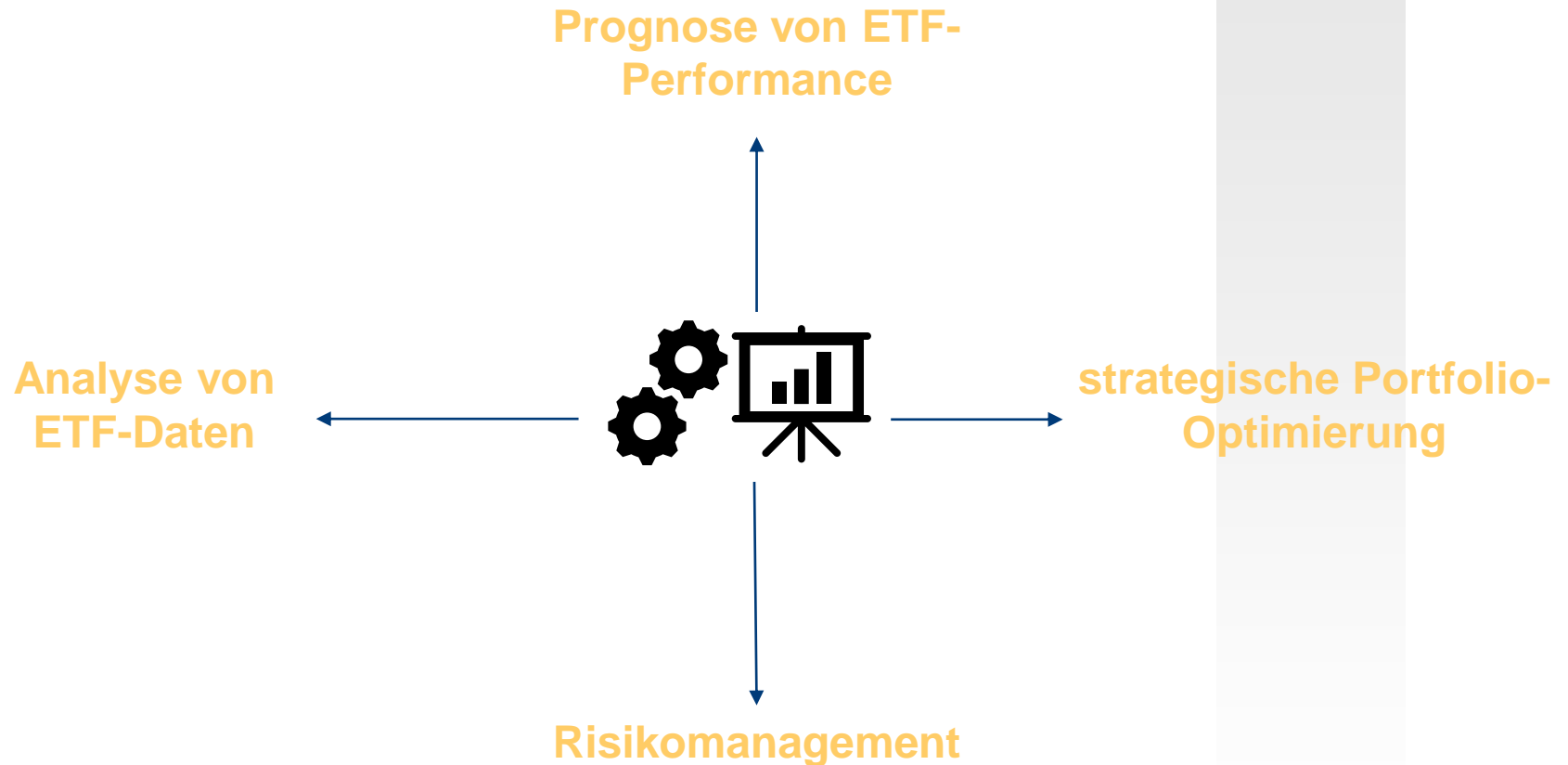
T-Test Ergebnis Interpretation

4. Bewertung der Ergebnisse des Data Minings in Bezug auf Ziele und Erfolgskriterien

Ziele:

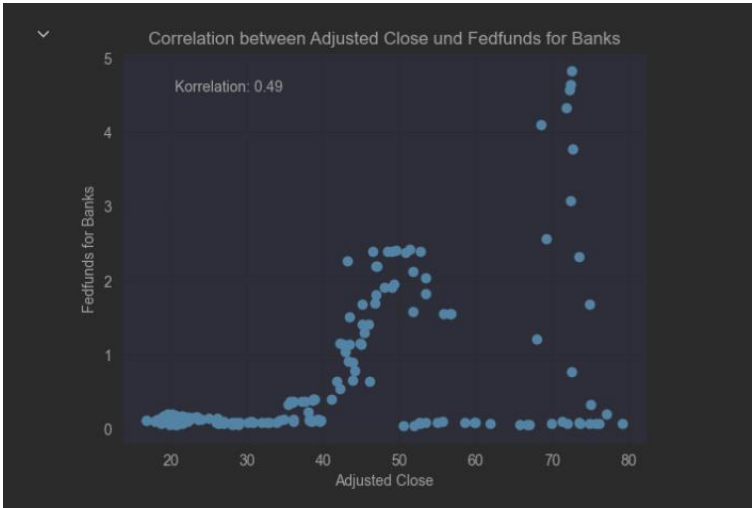
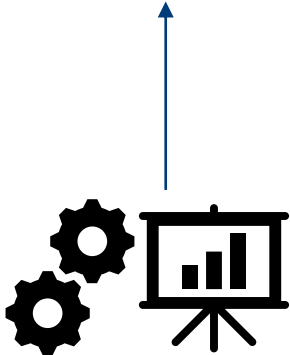
- geeignet Datensätze zu den Zielgrößen und Einflussgrößen finden und diese ggf. anpassen ✓
- Modellierungsverfahren erarbeiten, um Korrelationen und Muster zu erkennen, um thesenspezifische Prognosen durchzuführen ✓
- Interpretation der Ergebnisse und Schlussfolgerungen daraus ziehen ✓

4. Business Objective & Business Success Criteria

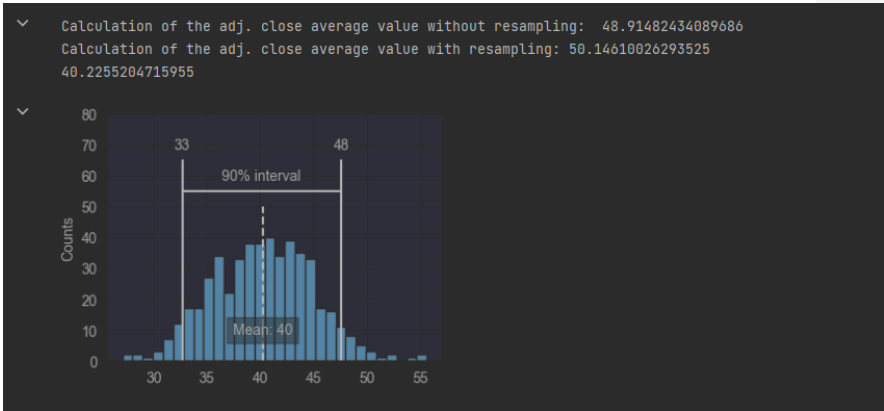


4. Business Objective & Business Success Criteria

Analyse von ETF-Daten

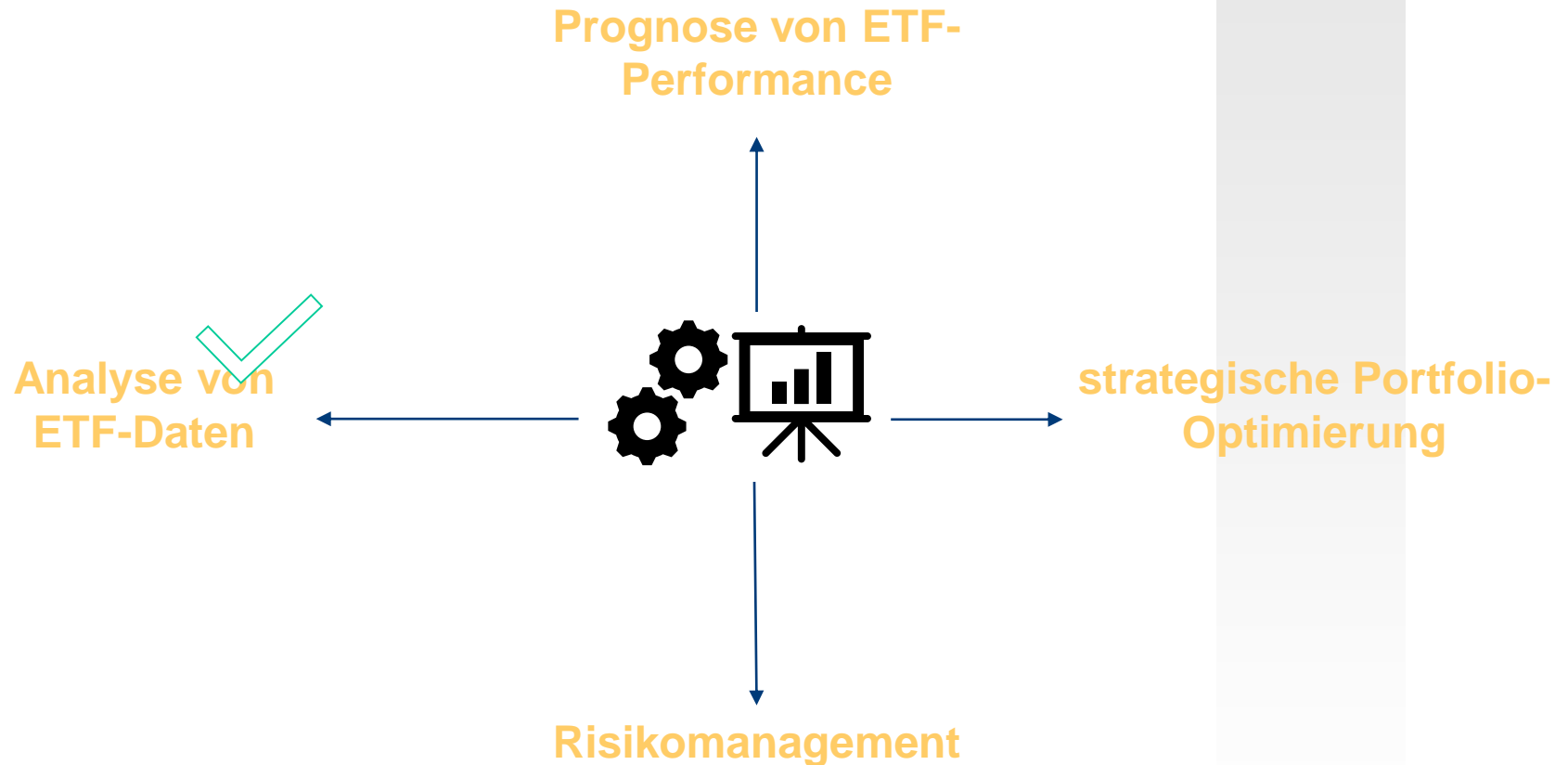


Korrelation & Streudiagramm Adj. Close Price und FED-Funds



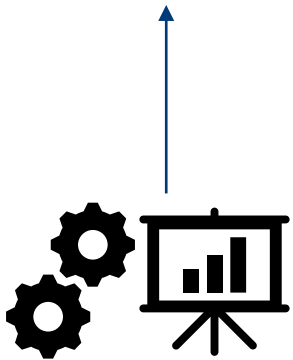
Auswertung Mittelwert Adj. Close Price des ETFs

4. Bewertung des Ergebnisses des Data Minings in Bezug Business Success Criteria



4. Business Objective & Business Success Criteria

Prognose von ETF- Performance

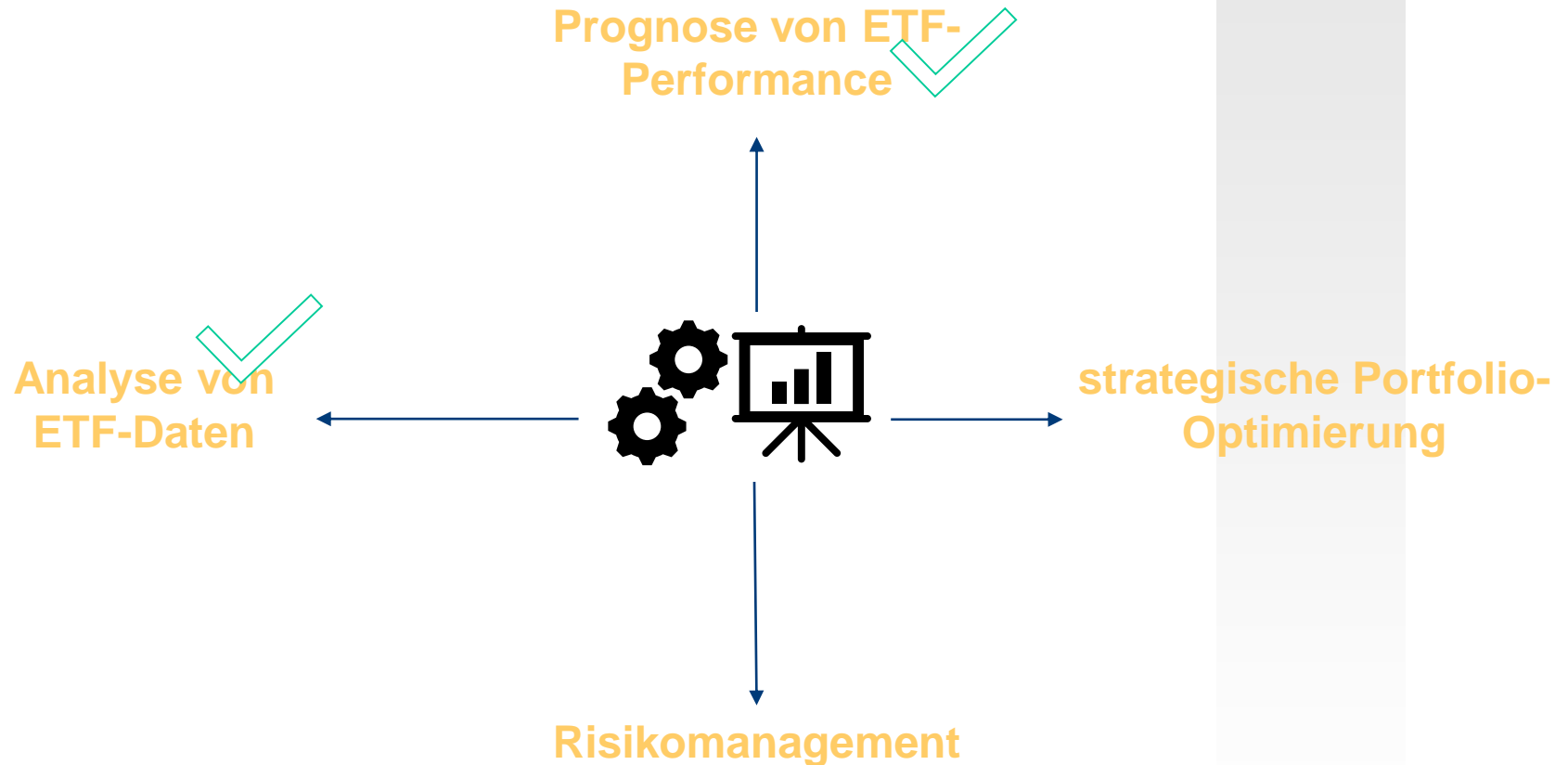


```
Berechnung basierend auf der Gewichtung des historischen Zeitrahmens:  
Intercept: 5.7913023987362715  
Coefficients:  
  ezb_funds: -0.104806680694586  
  fed_funds_for_banks: 1.0925324558159264  
  fed_funds_for_customer: -1.4527754089689695
```

```
Vorhersage für den 01.08.2023:  
Monatliche Rendite: 5.79
```

Prognose mittels
historischer Daten und
Einflussgrößen

4. Business Objective & Business Success Criteria



4. Business Objective & Business Success Criteria

strategische Portfolio-
Optimierung

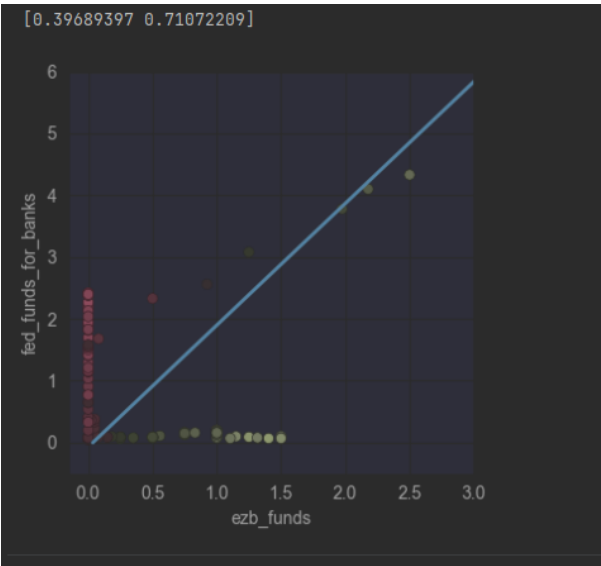
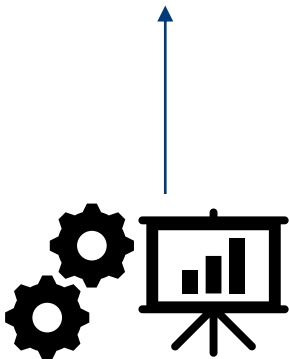
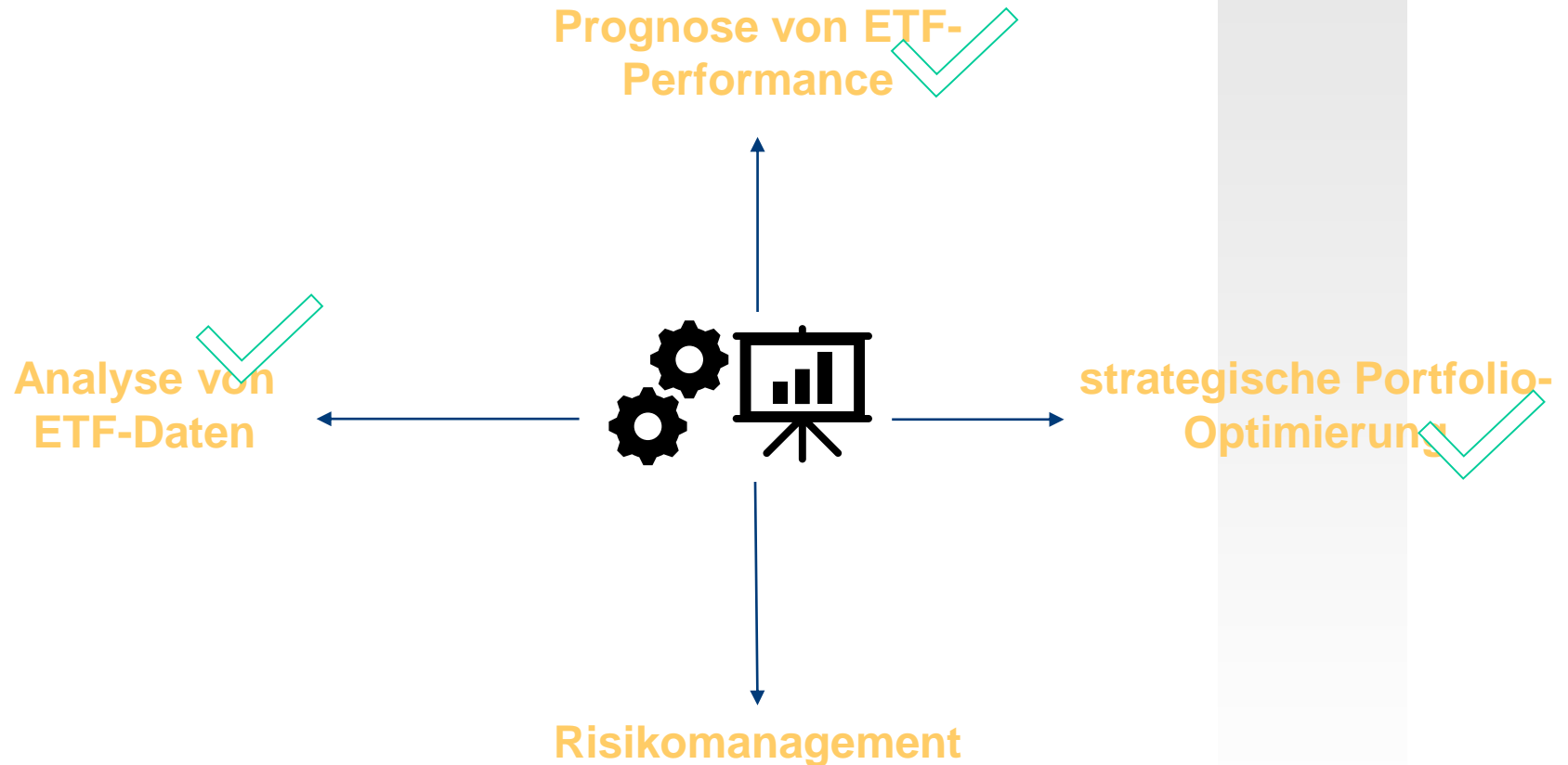


Diagramm LDA

```
predicted class: negative
predicted probabilities
      negative    positive
0  0.501568  0.498432
```

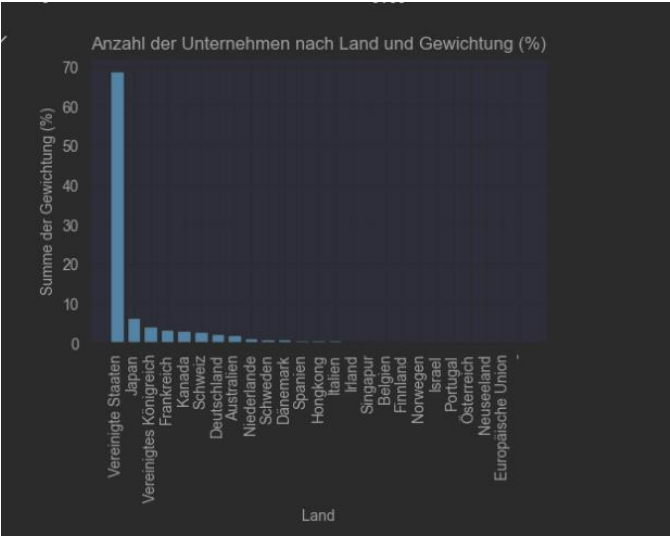
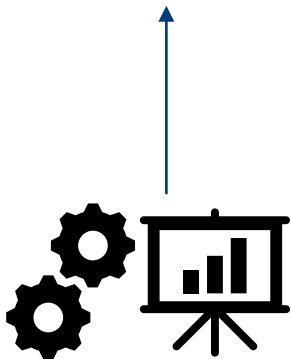
Klassifikationsvorhersage über das Naiver
Bayer Modell

4. Business Objective & Business Success Criteria



4. Business Objective & Business Success Criteria

Risikomanagement



Portfolie Aufteilung

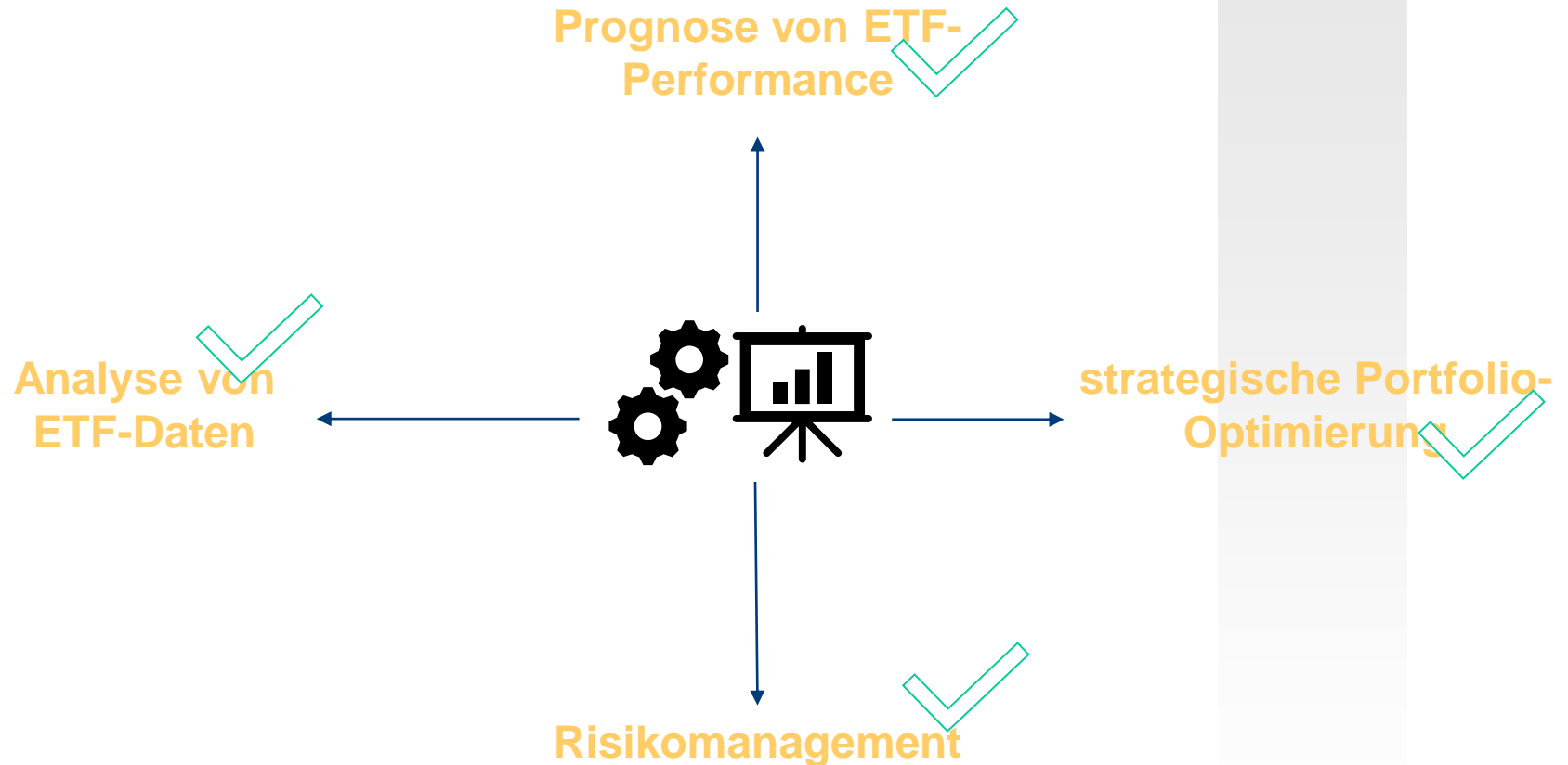
Comparison	negative	positive	stabil
Comparison			
empty	548	477	8
negative	1116	1045	4
positive	130	113	13

Chi-Quadrat-Test Ergebnis:
Chi2-Statistik: 77.40614180275973
p-Wert: 6.170125165820085e-16

Durchschnittliche Werte:
Mittelwert der Chi-Quadrat-Statistiken: 77.40614180275972
Mittelwert der p-Werte: 6.170125165820084e-16
Standardabweichung von p_values: 9.860761315262648e-32

Hypothesen Analyse

4. Business Objective & Business Success Criteria



5. Letzter Meilenstein

5. Letzte Meilensteine



MEILENSTEIN 1

- geeignet Datensätze für Zielgrößen und Einflussgrößen erkunden und Muster, Trends und Beziehungen erkennen

MEILENSTEIN 2

- statistische Analysen und Algorithmen, um die Faktoren und Muster zu identifizieren

MEILENSTEIN 3

- Prognosen für die Zukunft basierend aus den analysierten Daten

MEILENSTEIN 4

- Review und allg. Dokumentation

***Vielen Dank für eure
Aufmerksamkeit***