

FOLIE I – TITELBLATT

TITEL:CAPSTONE-PROJEKT:ANALYSE DER WEINQUALITÄT

NAME:AHMADI,SEYED MOHAMMAD HOSSEIN

MENTOR: MANUEL BORDASCH

DATUM: 15/04/2025



FOLIE 2 – ZIEL DES PROJEKTS

- **Ziel:**

Untersuchung der Faktoren, die die Qualität von Rotwein beeinflussen und Erstellung eines Vorhersagemodells.

Datenquelle: UCI Machine Learning Repository

Anzahl der Einträge: 1143

Anzahl der Merkmale: 13 (z. B. Alkoholgehalt, pH-Wert, Zitronensäure...)

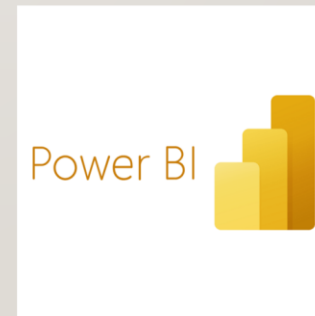
FOLIE 3 – VERWENDETE TOOLS

Haupttool:

- Python (Pandas, Matplotlib, Seaborn, Scikit-Learn, XGBoost)

Optional genutzte Tools:

- Power BI Desktop (für interaktive Visualisierung)
- Excel (für Pivot-Analyse & Charts)



FOLIE 4 – DATENVORBEREITUNG UND EXPLORATION (PYTHON)

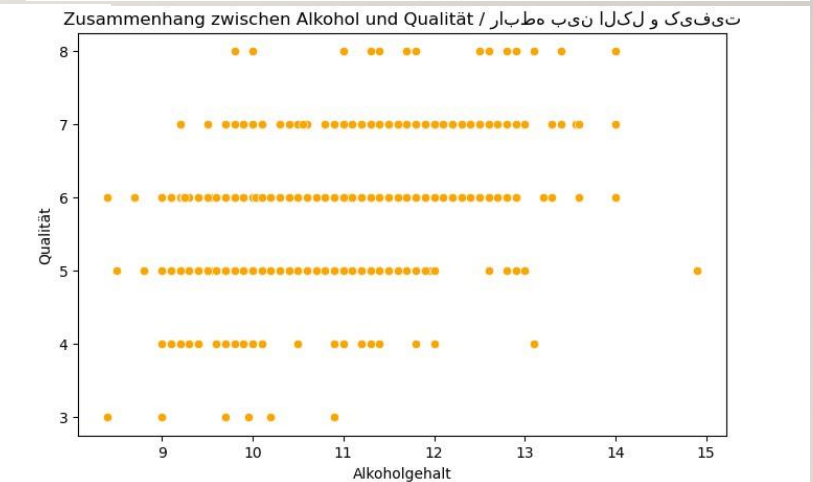
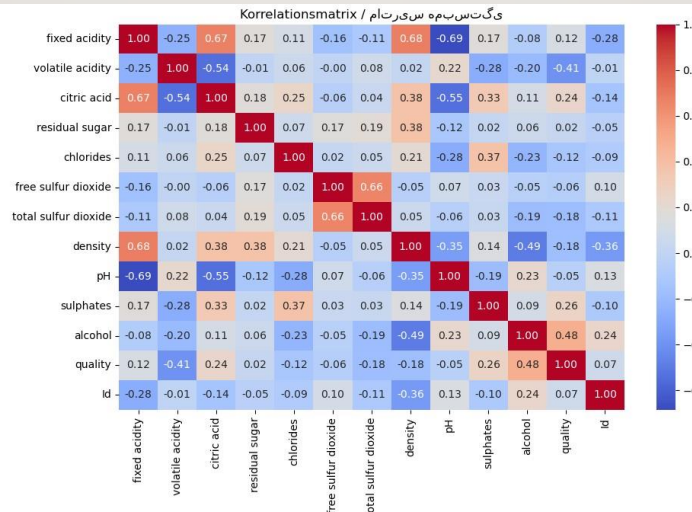
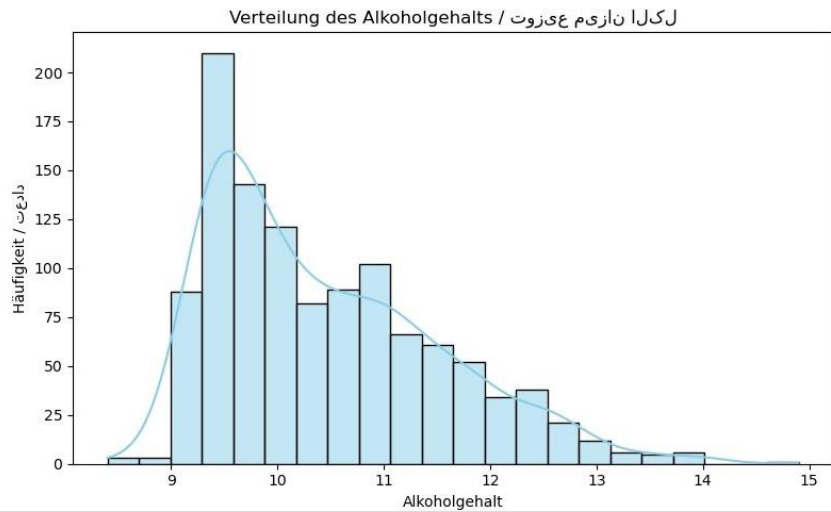
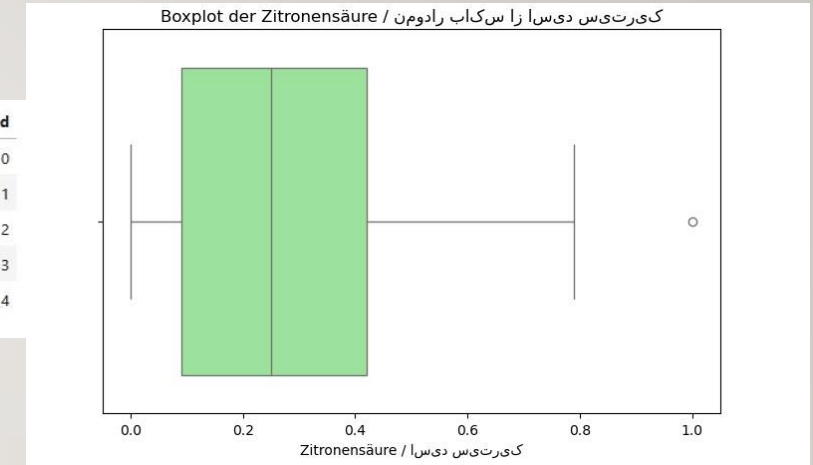
- **Einlesen und erste Inspektion der Daten**
- **Statistische Zusammenfassung (`describe()`)**
- **Überprüfung auf fehlende Werte**
- **Erkennung und Behandlung von Ausreißern**
- **Standardisierung der numerischen Merkmale**
- **Aufteilung in Trainings- und Testdaten (80/20)**

FOLIE 5 – VISUALISIERUNG IN PYTHON

🖼️ Platz für Bild:

- Histogramme
- Boxplots
- Korrelationsmatrix
- Scatterplots für wichtige Features

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality	Id
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5	0
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9.8	5	1
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9.8	5	2
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9.8	6	3
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5	4



FOLIE 6 – MODELLVERGLEICH: LINEARES MODELL, RANDOM FOREST, XGBOOST

Ziel: Vorhersage der Weinqualität (Zahl von 0 bis 10)

Trainierte Modelle:

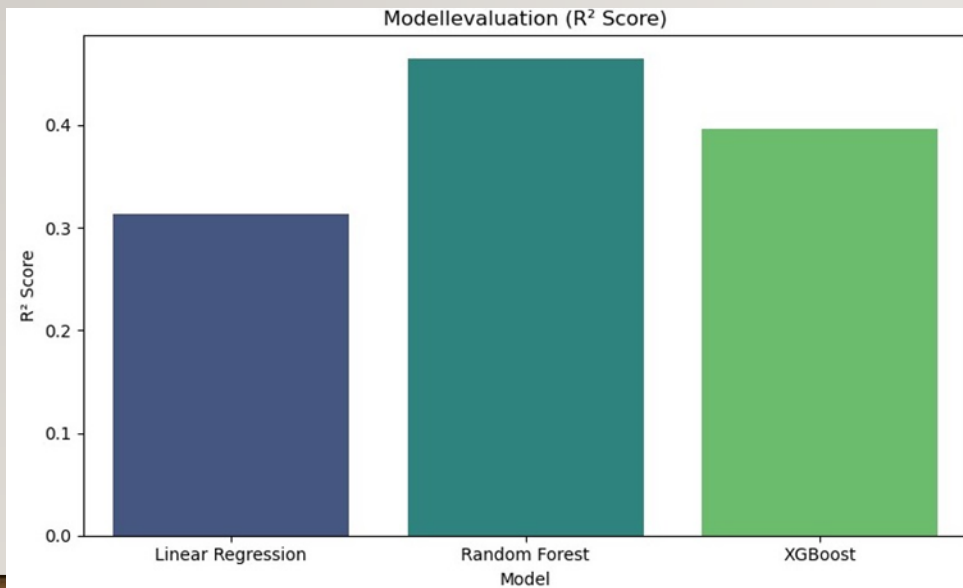
- Lineare Regression
- Random Forest Regressor
- XGBoost Regressor

Metriken:

- Mean Squared Error (MSE)
- R^2 Score

FOLIE 7 – MODELLVERGLEICH: ERGEBNISSE

- 📷 **Platz für Bild:**
Vergleich der Modelle (Tabelle oder Bar Plot)



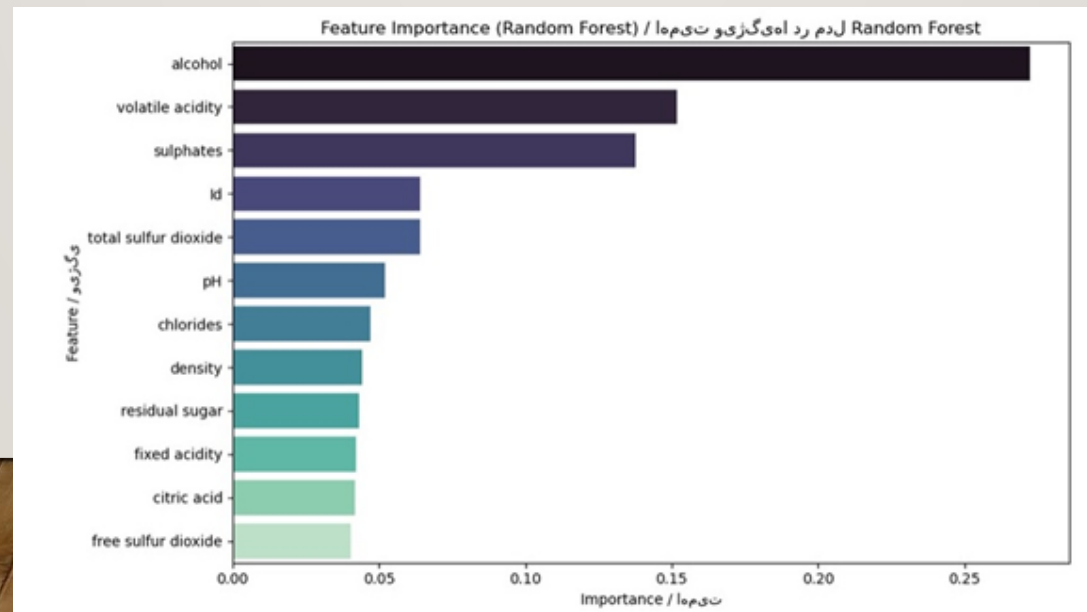
R ² -Score	MSE	Modell
0.317	0.3800	Lineare Regression
0.462	0.2989	Random Forest
0.343	0.3655	XGBoost

FOLIE 8 – WICHTIGSTE EINFLUSSFAKTOREN (FEATURE IMPORTANCE)

- **Alkoholgehalt, pH-Wert und Zitronensäure** hatten den größten Einfluss auf die Weinqualität.
- Diese Merkmale zeigen hohe Korrelation mit der Zielvariable.
- **Random Forest und XGBoost** bestätigten diese Features.

📷 Platz für Bild:

- Bar Plot der Feature Importance in Random Forest/XGBoost

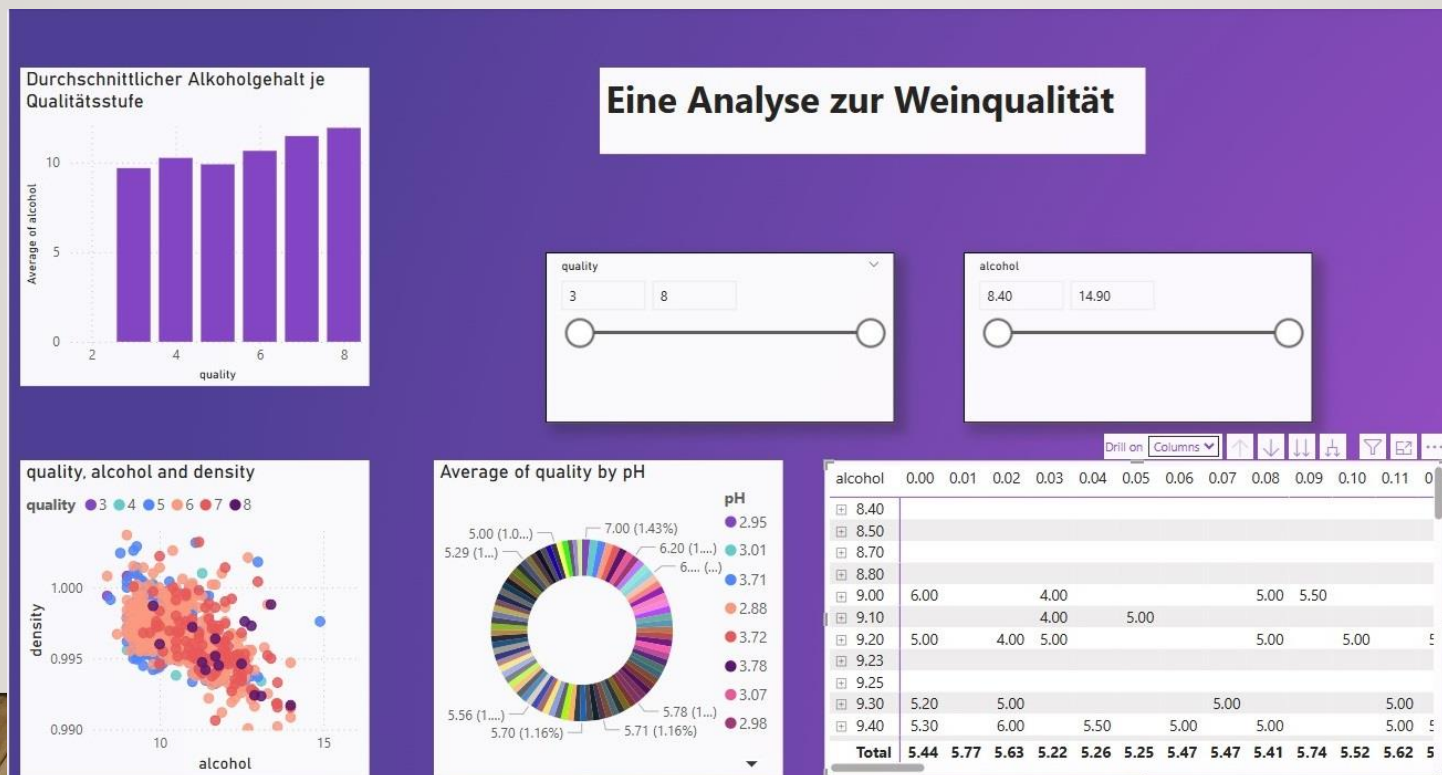


FOLIE 9 – ERGÄNZENDE ANALYSE MIT POWER BI

- Interaktive Visualisierung der Merkmale
- Darstellung von Mittelwerten, Korrelationen und Verteilungen

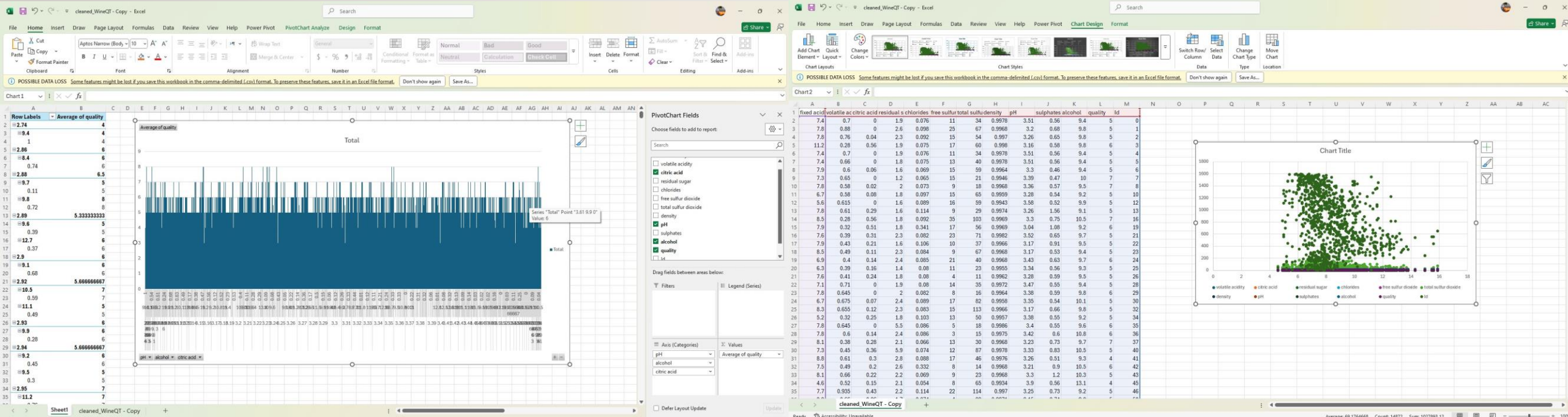
📷 Platz für Bild:

- Power BI Dashboard mit Diagrammen (z.B. Streudiagramm, Balken)



FOLIE 10 – ERGÄNZENDE ANALYSE MIT EXCEL

- Erstellung einer Pivot-Tabelle
 - Berechnung von Mittelwerten nach Qualitätsstufen
 - Diagramme zur Veranschaulichung (z.B. Streudiagramm, Balkendiagramm)
- 📷 Platz für Bild:
- Pivot-Tabelle und Diagramme aus Excel



FOLIE 11 – FAZIT

- Random Forest erzielte die beste Vorhersagequalität (höchster R^2 Score).
- Alkoholgehalt war das wichtigste Merkmal für die Weinqualität.
- **Python** ermöglichte eine tiefgehende Datenanalyse und Modellierung.
- **Power BI** und **Excel** waren nützliche Ergänzungen zur Visualisierung und für die Analyse.

Vielen Dank
für Ihre
Aufmerksamkeit!