

# Pandas Cheatsheet

Onur Yilmaz

June 1, 2023

## Erste Schritte mit pandas

Importieren der Pandas-Bibliothek:

```
1 import pandas as pd
```

Erstellen einer Series:

```
1 s = pd.Series([1, 3, 5, np.nan, 6, 8])
```

Erstellen eines DataFrames:

```
1 data = {'Name': ['John', 'Emily', 'Ryan'],  
2         'Alter': [25, 30, 35]}  
3 df = pd.DataFrame(data)
```

Erstellen eines Indexobjekts:

```
1 index = pd.Index(['a', 'b', 'c'])
```

Neuindizierung eines DataFrames:

```
1 df_reindexed = df.reindex(...)
```

## Deskriptive Statistik

Korrelation, Korrelationskoeffizient und Korrelationsmatrix:

```
1 corr_matrix = df.corr()  
2 corr_coeff = df['A'].corr(df['B'])
```

Kovarianz, Kovarianzmatrix und lineare Diskriminanzanalyse nach Fisher:

```
1 cov_matrix = df.cov()
```

## Laden und Speichern von Daten

Stückweises Lesen und Schreiben von Daten in Textformaten:

```
1 df = pd.read_csv('data.csv', chunksize=1000)
2 df.to_csv('output.csv')
```

Web-Scraping und Lesen von MS Excel-Dateien:

```
1 df = pd.read_html('https://www.example.com')
2 df = pd.read_excel('data.xlsx')
```

## Daten bereinigen und vorbereiten

Fehlende Daten herausfiltern und einsetzen:

```
1 df.dropna()
2 df.fillna(0)
```

Duplikate entfernen, Daten transformieren und Erkennen von Ausreißern:

```
1 df.drop_duplicates()
2 df['column'].apply(lambda x: x*2)
3 df['column'].quantile(0.95)
```

Methoden von String-Objekten und Reguläre Ausdrücke:

```
1 df['column'].str.upper()
2 df['column'].str.contains('pattern')
```

## Datenaufbereitung: Verknüpfen, Kombinieren und Umformen

Ebenen neu anordnen und sortieren:

```
1 df.sort_index()
2 df.sort_values(by='column')
```

Tabellen permutieren und transponieren:

```
1 df.pivot(index='A', columns='B', values='C')
2 df.transpose()
```

Daten über Index verknüpfen:

```
1 df1.join(df2)
2 df.merge(df2, on='key')
```

Hierarchische Indizierung:

```
1 df.set_index(['A', 'B'])
2 df.reset_index()
```

## Aggregation von Daten und Gruppenoperationen

Gruppenaggregationen:

```
1 df.groupby('column').mean()
2 df.groupby(['col1', 'col2']).sum()
```

Gruppieren mit Dictionaries, Series und Funktionen:

```
1 group_dict = {'col1': 'group1', 'col2': 'group2'}
2 df.groupby(group_dict).max()
3 df.groupby(df['column']).apply(lambda x: x*2).min()
```

Spaltenweise und mehrfache Anwendung:

```
1 df['column'].apply(func)
2 df.apply(func, axis=1)
```