Cheat Sheet
Data Cleaning

Comprehensive Python Techniques

Kostya Numan
AI & AGI Researcher

A guide to essential python functions for data cleaning, ensuring data consistency, and preparing datasets for analysis.

Kostya Numan

AI & AGI Researcher

Exploring Data Structure

Display Data Types

```
df.dtypes
```

Check for Missing Values

```
df.isnull().sum()
```

Get Unique Values in Column

```
df['column'].unique()
```

Kostya Numan
AI & AGI Researcher

Handling Missing Values

Fill Missing with Mean

```
df['age'] = df['age'].fillna(df['age'].m-
ean())
```

Fill Missing with Mode

```
df['x'] = df['x'].fillna(df['x'].mode()[-
0])
```

Drop Rows with Any NaNs

```
df = df.dropna(how='any')
```

Kostya Numan
AI & AGI Researcher

Removing Duplicates

Find Duplicate Rows

```
df[df.duplicated()]
```

Drop Duplicates Based on Column

```
df = df.drop_duplicates(subset=['column'])
```

Kostya Numan

AI & AGI Researcher

Transforming Data

Apply Function to Column

```
df['x'] = df['x'].apply(lambda x: x * 2)
```

Convert Column to String

```
df['column'] = df['column'].astype(str)
```

Replace Values in Column

```
df['column'] = df['column'].replace({
    'old_value': 'new_value'
})
```

Kostya Numan
AI & AGI Researcher

Filtering & Selecting Data

Filter Rows Based on Condition

```
df[df['column'] > 10]
```

Select Specific Columns

```
df[['col1', 'col2']]
```

Select Rows by Index

```
df.iloc[0:5]
```

Kostya Numan
AI & AGI Researcher

Combining & Merging Data

Merge DataFrames on Column

```
df1.merge(df2, on='column')
```

Concatenate DataFrames Vertically

```
pd.concat([df1, df2])
```

Join DataFrames on Index

```
df1.join(df2)
```

Kostya Numan
AI & AGI Researcher

→

Handling Outliers

Using Z-score Filtering

```python
from scipy import stats
df = df[(stats.zscore(df['column']) < 3)]
```

Using IQR Filtering

```python
Q1 = df['column'].quantile(0.25)
Q3 = df['column'].quantile(0.75)
IQR = Q3 - Q1
df = df[
  (df['column'] ≥ (Q1 - 1.5 * IQR)) &
  (df['column'] ≤ (Q3 + 1.5 * IQR))
]
```

Kostya Numan

AI & AGI Researcher

Subscribe

to Numan Substack
for more insights

Kostya Numan
AI & AGI Researcher