

pandas

April 14, 2023

1 Pandas Cheat Sheet

1.1 Importing Pandas

```
import pandas as pd
```

1.2 Reading Data

Read data from CSV, Excel, and JSON files.

```
data = pd.read_csv("filename.csv")
data = pd.read_excel("filename.xlsx")
data = pd.read_json("filename.json")
```

1.3 Writing Data

Write data to CSV, Excel, and JSON files.

```
data.to_csv("filename.csv", index=False)
data.to_excel("filename.xlsx", index=False)
data.to_json("filename.json")
```

1.4 Data Exploration

Explore data: display first/last n rows, shape, information, data types, and descriptive statistics.

```
data.head(n)
data.tail(n)
data.shape
data.info()
data.dtypes
data.describe()
```

1.5 Data Selection

Select data: single/multiple columns, row by index, rows based on a condition, or specific rows and columns.

```
data["column_name"]
data[["column1", "column2"]]
data.loc[index]
```

```
data[data["column"] > value]
data.loc[data["column"] > value, ["column1", "column2"]]
```

1.6 Data Manipulation

Manipulate data: rename columns, drop columns, change data types, set/reset index, sort values, apply a function to a column, group data, aggregation functions, pivot tables, merge and concatenate DataFrames.

```
data.rename(columns={"old_name": "new_name"}, inplace=True)
data.drop("column_name", axis=1, inplace=True)
data["column"] = data["column"].astype("new_data_type")
data.set_index("column_name", inplace=True)
data.reset_index(inplace=True)
data.sort_values(by="column_name", ascending=True, inplace=True)
data["new_column"] = data["column"].apply(lambda x: function(x))
grouped = data.groupby("column_name")
grouped["column"].agg(["count", "mean", "min", "max"])
pd.pivot_table(data, values="column1", index="column2", columns="column3", aggfunc=np.mean)
merged = pd.merge(data1, data2, on="column_name", how="inner")
concatenated = pd.concat([data1, data2], axis=0)
```

1.7 Handling Missing Data

Handle missing data: drop missing values, fill missing values, interpolate missing values, and detect missing values.

```
data.dropna(subset=["column_name"], inplace=True)
data["column_name"].fillna(value, inplace=True)
data["column_name"].interpolate(method="linear", inplace=True)
data.isna().sum()
```