

Pandas Tutorial — Lesson Notes

What is Pandas?

Pandas is a Python library for data manipulation and analysis. It provides DataFrames — powerful table-like structures for working with structured data.

Core Concepts

Creating DataFrames

```
import pandas as pd

# From dictionary
df = pd.DataFrame({
    "name": ["Alice", "Bob", "Charlie"],
    "age": [25, 30, 35],
    "score": [92, 85, 78]
})

# From CSV file
df = pd.read_csv("students.csv")
```

Exploring Data

```
df.head()           # First 5 rows
df.info()           # Column types and null counts
df.describe()       # Statistical summary
df.shape            # (rows, columns)
df.columns          # Column names
```

Selecting Data

```
df["name"]          # Single column (Series)
df[["name", "age"]] # Multiple columns (DataFrame)
df.loc[0]           # Row by label
df.iloc[0:3]        # Rows by position
df[df["score"] > 80] # Filter rows
```

Common Operations

```
df["grade"] = df["score"].apply(lambda x: "A" if x >= 90 else "B")
df.sort_values("score", ascending=False)
df.groupby("grade")["score"].mean()
df.dropna()                                # Remove rows with NaN
df.fillna(0)                              # Replace NaN with 0
```

Merging Data

```
merged = pd.merge(students, grades, on="student_id", how="left")
```

Key Takeaways

1. Pandas is built on top of NumPy — it's fast for structured data.
2. Use `.loc` for label-based indexing, `.iloc` for position-based.
3. `groupby()` is essential for aggregation and summary statistics.
4. Always check for missing data with `df.isnull().sum()`.