Pandas Roadmap: From Basic to Advanced

1. Introduction to Pandas and DataFrames

Pandas is a library for data manipulation and analysis. It introduces two main structures: Series (1D) and DataFrame (2D).

```
import pandas as pd

# Create a Series
data = [10, 20, 30, 40]
s = pd.Series(data)
print(s)

# Create a DataFrame
data = {'Name': ['Alice', 'Bob', 'Charlie'],
        'Age': [25, 30, 35],
        'Salary': [50000, 60000, 70000]}
df = pd.DataFrame(data)
print(df)
```

2. Reading and Writing Data

Pandas allows reading and writing data from various file formats like CSV, Excel, and SQL.

```
import pandas as pd

# Read from a CSV file
df = pd.read_csv('data.csv')
print(df.head())

# Write to a CSV file
df.to_csv('output.csv', index=False)
```

3. Indexing and Selecting Data

You can select rows, columns, or specific elements in a DataFrame using indexing methods like iloc, loc, or direct column access.

```
import pandas as pd

# Create a DataFrame
data = {'A': [1, 2, 3], 'B': [4, 5, 6]}
df = pd.DataFrame(data)

# Access columns
print(df['A'])

# Access rows
print(df.loc[0])   # Using label-based indexing
print(df.iloc[1])  # Using position-based indexing
```

4. Data Cleaning and Transformation

Pandas provides tools for handling missing values, duplicates, and transforming data formats.

```
import pandas as pd

# Create a DataFrame with missing values
data = {'A': [1, None, 3], 'B': [4, 5, None]}
df = pd.DataFrame(data)

# Drop missing values
print(df.dropna())

# Fill missing values
print(df.fillna(0))
```

5. Grouping and Aggregation

You can group data based on certain criteria and perform aggregations like sum, mean, and count.

```
import pandas as pd


# Create a DataFrame
data = {'Category': ['A', 'A', 'B', 'B'],
        'Values': [10, 20, 30, 40]}
df = pd.DataFrame(data)

# Group by 'Category' and aggregate
result = df.groupby('Category').sum()
print(result)
```