Welcome to

INTERNSHIP STUDIO

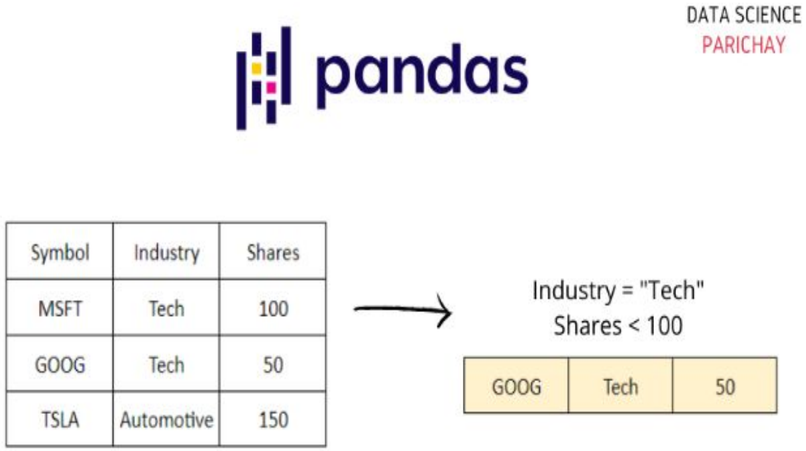
Module 04 | Lesson 05

Data Analytics

Filtering & Aggregation of Data

Data Filtering

- Pandas provides powerful tools for filtering data based on specific conditions.
- The **DataFrame** object in Pandas allows for easy selection of rows and columns based on certain criteria.




```
# Selecting rows where the 'Age' column is greater than 30
filtered_data = df[df['Age'] > 30]

# Display the filtered data
print(filtered_data)
```

Data Grouping

internship studio

- Pandas allows grouping data based on one or more columns.

- The **groupby()** function is used to group data, and then aggregating functions can be applied to the grouped data.

```
# Grouping data by 'City' column and calculating the mean age
grouped_data = df.groupby('City')['Age'].mean()

# Display the grouped data
print(grouped_data)
```

Data Aggregation

internship studio

- Pandas provides various aggregation functions to calculate summary statistics on data such as sum(), max(), min().

- These functions can be applied to individual columns or entire **DataFrame** objects.

```
# Calculating the sum of the 'Sales' column
total_sales = df['Sales'].sum()

# Calculating the maximum value in the 'Revenue' column
max_revenue = df['Revenue'].max()

# Display the aggregated values
print("Total Sales:", total_sales)
print("Max Revenue:", max_revenue)
```

Data Analysis Workflow

1. **Load Data:** Read the data into a **DataFrame** using Pandas' input functions, such as **read_csv()** for CSV files.

2. **Explore Data:** Use descriptive statistics, data visualization, and **DataFrame** methods to gain insights into the data.

3. **Clean and Preprocess Data:** Handle missing values, remove duplicates, and transform data into usable formats.

4. **Perform Analysis:** Apply filtering, grouping, and aggregating techniques to answer specific research questions.

5. **Visualize Results:** Utilize Pandas' integration with visualization libraries like Matplotlib or Seaborn to create meaningful plots.

SUMMARY

You got this

- Pandas offers a range of functionalities for data analysis tasks.
- Filtering allows for selecting data based on specific conditions.
- Grouping enables grouping data based on one or more columns.
- Aggregation functions help in summarizing data for analysis.
- Following a systematic workflow with Pandas can streamline data analysis tasks.

Next session

Data Analytics hands-on coding for a Real Dataset