Expectations - What I expect from you

- No Late Assignments

- No Cheating

- Be a good classmate

- Don't waste your time

- Show up to class

Agenda

On the right is what we will cover today.

Jupyter & Pandas: Interactive Data Analysis
Why Jupyter and Pandas?
Data is everywhere that you can use!
Getting Started: Setup
Launching Jupyter Notebook
First Steps in the Notebook
Loading Data with Pandas
Exploring the Data
Unique Classes and Filtering
Sorting and Top Performers
Grouping and Aggregation
Visualizing Results
Exercises
Summary
Handy Jupyter Shortcuts
Next Steps

WE ARE ESSENTIAL TO ALBERTA | NAIT

Jupyter & Pandas: Interactive Data Analysis

- Jupyter Notebooks + Pandas = powerful, interactive data exploration

- Great for learning, prototyping, and sharing insights

Why Jupyter and Pandas?

- Jupyter lets you run code, see results, and document your work in one place

- Pandas makes data cleaning, analysis, and visualization easy

- Used in data science, research, and business analytics

Data is everywhere that you can use!

There's a ton of open datasets available for practice or for your own projects:

- General dataset searches

  - Kaggle Datasets
  - Google Dataset Search

- Local government open data portals

  - edmonton open data portal
  - alberta open data portal
  - open data canada
  - so many more!

Getting Started: Setup

- Create a virtual environment and install the tools:

```
python -m venv venv
venv\Scripts\activate
pip install jupyter pandas
pip freeze > requirements.txt
```

- This keeps your project isolated and reproducible

Launching Jupyter Notebook

- Start the notebook server:

```
jupyter notebook
```

- Opens in your browser; create a new Python 3 notebook

First Steps in the Notebook

- Rename your notebook to `student_grades_analysis`

- In the first cell, test your setup:

```python
print("Ready to analyze student grades!")
```

Loading Data with Pandas

- Imagine you have a CSV file: `data/student_grades.csv`

- Example contents:

```
Name,Class,Grade
Alice,Math,88
Bob,Math,92
Charlie,Science,85
Dana,Math,76
Eli,Science,90
```

- Load the data:

```python
import pandas as pd
grades = pd.read_csv('data/student_grades.csv')
grades.head()
```

Exploring the Data

- Get a summary:

```
grades.info()
```

- See the shape (rows, columns):

```
grades.shape
```

- Basic statistics:

```
grades.describe()
```

Unique Classes and Filtering

- List all classes:

```
grades['Class'].unique()
```

- Filter for Math students:

```
math_grades = grades[grades['Class'] == 'Math']
math_grades
```

Sorting and Top Performers

- Sort Math grades, highest first:

```
top_math = math_grades.sort_values(by='Grade', ascending=False)
top_math
```

- Show the top student in Math:

```
top_math.iloc[0]
```

Grouping and Aggregation

- Average grade per class:

```python
avg_by_class = grades.groupby('Class')['Grade'].mean()
print(avg_by_class)
```

- Count of students per class:

```python
count_by_class = grades['Class'].value_counts()
print(count_by_class)
```

Visualizing Results

- Plot average grades by class:

```python
import matplotlib.pyplot as plt

avg_by_class.plot(kind='bar', title='Average Grade by Class')
plt.ylabel('Average Grade')
plt.show()
```

Exercises

- Try these in new cells:

  1. What is the highest grade in Science?

  2. How many students scored above 85 in Math?

  3. List all students with grades below 80.

Summary

- You set up Jupyter and Pandas in a virtual environment

- Loaded and explored a dataset

- Filtered, sorted, grouped, and visualized data

- These skills are foundational for data analysis in Python

Handy Jupyter Shortcuts

Shortcut	Action
Shift+Enter	Run cell, go to next
Ctrl+Enter	Run cell, stay
A/B (cmd mode)	Add cell above/below
M/Y	Change cell to Markdown/Code
D,D	Delete cell
Ctrl+S	Save notebook
Tab/Shift+Tab	Autocomplete / Show doc

Next Steps

- Try with your own dataset (e.g., sports scores, survey results)

- Explore more Pandas features: merging, pivot tables, missing data handling

- Share your notebook with classmates or on GitHub