Introduction to Pandas

CSE 160 Summer 2026

Questions?

Sli.do #cse160

Announcements

- **Homework Assignment 5 due tonight at 11:59PM**
- **Programming Practice 7 due Wednesday, August 19 at 11:59PM**
- **Resubmission Cycle 5 (final resubmission form) available tonight!**
  - Homeworks 1-4 eligible
  - Check Ed post for more information.
- **Final Exam information and Practice Finals** now available on course website!
  - Final Exam on Friday, August 21 at 9:40am in HRC 155

Today's Roadmap

1. Representing Tabular Data
2. Exploring Data
3. Answering Questions with Data

Jupyter Notebook

not this one…

this one :)

pandas

Notebook Philosophy

or, ***How to make the most of your notebook***

Take Small Steps

Avoid large code cells.

- Defining a function? *Give it a cell!*
- Loading a dataset? *Give it a cell!*
- Debugging? *Use a new cell!*

Document!

A good Python notebook combines code and text to **explain each data processing step** and make reproducing results easy!

Data Processing

Often, we need to process (transform) data to be able to derive useful insights.

Address	Type	Datetime	Latitude	Longitude	Report Location	Incident Number
3821 E Newton St	Investigate Out Of Service	2025-05-05 15:22:00	47.636978	-122.282918	POINT (-122.282918 47.636978)	F25005839
3rd Ave / Seneca St	Aid Response	2025-12-10 20:24:00	47.60705	-122.335324	POINT (-122.335324 47.60705)	F25170229
500 5TH AVE	Rescue Elevator	2025-05-04 17:08:00	47.602985	-122.328535	POINT (-122.328535 47.602985)	F25005941
6901 Martin Luther King Jr Way S	Hang-Up- Aid	2025-07-24 01:50:00	47.540684	-122.283048	POINT (-122.283048 47.540684)	F25010016
118 Bell St	Aid Response	2025-07-28 18:21:00	47.613497	-122.346355	POINT (-122.346355 47.613497)	F25010267
3401 W Government Way	Aid Response	2025-09-27 11:58:00	47.659307	-122.400889	POINT (-122.400889 47.659307)	F25013423
121 14TH AVE E	Rubbish Fire	2025-10-23 01:18:00	47.619026	-122.314303	POINT (-122.314303 47.619026)	F25014702
2418 Nw 56th St	Aid Response	2025-09-09 23:55:00	47.669458	-122.387807	POINT (-122.387807 47.669458)	F25012548
2200 Westlake Ave	Aid Response	2025-10-28 16:10:00	47.617752	-122.338344	POINT (-122.338344 47.617752)	F25014972
1715 Nw 59th St	Aid Response Yellow	2025-01-11 09:43:00	47.671633	-122.379034	POINT (-122.379034 47.671633)	F25000476
800 Occidental Ave S	EVENT - Special Event	2025-10-24 17:06:00	47.595692	-122.333273	POINT (-122.333273 47.595692)	F25014778
1736 S Columbian Way	Aid Response Yellow	2025-09-03 12:08:00	47.561081	-122.309252	POINT (-122.309252 47.561081)	F25012227
5270 Rainier Ave S	Aid Response	2025-10-23 01:04:00	47.559937	-122.282393	POINT (-122.282393 47.559937)	F25014701
9134 45th Ave Sw	Aid Response	2025-10-22 05:17:00	47.522014	-122.390207	POINT (-122.390207 47.522014)	F25014652
2318 N 140th St	Aid Response	2025-06-26 23:25:00	47.730467	-122.331075	POINT (-122.331075 47.730467)	F25008558
7927 9th Ave Sw	Aid Response	2025-10-31 01:13:00	47.531689	-122.346777	POINT (-122.346777 47.531689)	F25015086
3920 Stone Way N	Scenes Of Violence Aid	2025-03-04 19:21:00	47.654346	-122.34249	POINT (-122.34249 47.654346)	F25003045
1066 E Blaine St	Aid Response	2025-11-25 11:42:00	47.635027	-122.31869	POINT (-122.31869 47.635027)	F25016326
9419 15th Ave Sw	Aid Response	2025-11-22 21:36:00	47.519017	-122.354003	POINT (-122.354003 47.519017)	F25016223
9200 2nd Ave Sw	Aid Response	2025-09-23 15:38:00	47.521244	-122.336301	POINT (-122.336301 47.521244)	F25013232
916 N 143rd St	Aid Response	2025-08-03 03:13:00	47.732314	-122.347028	POINT (-122.347028 47.732314)	F25010576
6901 SAND POINT WAY NE	Illegal Burn	2025-02-04 21:02:00	47.678774	-122.263605	POINT (-122.263605 47.678774)	F25001650
223 Yesler Way	Rescue Standby	2025-08-28 16:13:00	47.601718	-122.331581	POINT (-122.331581 47.601718)	F25011935
4210 Sw Admiral Way	Aid Response	2025-12-14 19:28:00	47.581187	-122.385299	POINT (-122.385299 47.581187)	F25017211
4301 Woodland Park Ave N	2RED - 1 + 1	2025-12-14 20:38:00	47.659575	-122.344579	POINT (-122.344579 47.659575)	F25017213
6348 24th Ave Sw	Scenes Of Violence Aid	2025-10-21 01:28:00	47.545703	-122.363628	POINT (-122.363628 47.545703)	F25014595
4701 41st Ave Sw	Investigate Out Of Service	2025-05-01 18:59:00	47.561078	-122.384138	POINT (-122.384138 47.561078)	F25005809
8547 29th Ave Nw	Aid Response	2025-12-08 21:47:00	47.690891	-122.394086	POINT (-122.394086 47.690891)	F25016931
1720 E Olive St	Aid Response	2025-12-01 00:54:00	47.616442	-122.309884	POINT (-122.309884 47.616442)	F25016564
400 Pine St	Medic Response	2025-12-13 11:43:00	47.611207	-122.337592	POINT (-122.337592 47.611207)	F25017143
171 12th Ave	Alarm Bell	2025-07-26 14:45:00	47.603031	-122.316782	POINT (-122.316782 47.603031)	F25010154
523 Broadway E	Rescue Elevator	2025-05-24 10:08:00	47.623385	-122.320928	POINT (-122.320928 47.623385)	F25006860
Sb I5 Exp At Ship Canal	MVI - Motor Vehicle Incident	2025-02-27 10:20:00	47.656039	-122.322482	POINT (-122.322482 47.656039)	F25002786
7228 39th Ave Sw	Illegal Burn	2025-03-18 19:21:00	47.537456	-122.381942	POINT (-122.381942 47.537456)	F25003708
2801 Sw Thistle St	Aid Response	2025-09-19 19:03:00	47.528282	-122.368551	POINT (-122.368551 47.528282)	F25013039

data processing

Type	Count
Aid Response	57618
Medic Response	10784
Auto Fire Alarm	7053
Aid Response Yellow	4748
Triaged Incident	3650
...	...
BC Medic Response- 7 per Rule	1
Brush Fire Major	1
Mutual Aid- Hazmat	1
Train Derailment wFireHzmt	1
Law Enforcement Standby	1

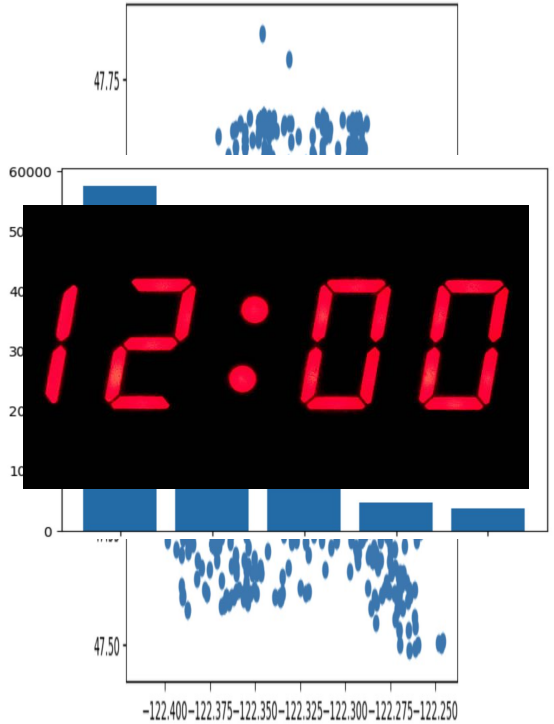
Original Data

Transformed Data

Exploring Our Data

What might we want to do with our data?

1. Map incidents ✅
2. What proportion of incidents took place between 12am and 1am? 🤔
3. What are the most common incident types? 🔁



What is pandas?

pandas is a widely-used Python library for working with **tabular data**.

- In pandas, data is stored as a **DataFrame**, which is a collection of columns called **series**) with associated column names

	Address	Type	Datetime	Latitude	Longitude	Report Location	Incident Number
0	1333 Harbor Ave Sw	Illegal Burn	2025-12-31 23:59:00	47.591918	-122.382656	POINT (-122.382656 47.591918)	F260000002
1	3323 S Graham St	Aid Response	2025-12-31 23:57:00	47.546103	-122.289056	POINT (-122.289056 47.546103)	F250180050
2	81 CLAY ST	Ladder Code Red	2025-12-31 23:56:00	47.615840	-122.352214	POINT (-122.352214 47.61584)	F250180051
3	E Pine St / Boylston Ave	Aid Response	2025-12-31 23:55:00	47.615231	-122.323415	POINT (-122.323415 47.615231)	F250180049
4	1325 E Madison St	Aid Response	2025-12-31 23:52:00	47.613492	-122.315187	POINT (-122.315187 47.613492)	F250180048

*Whereas **CSV DictReader** stores a list of row dictionaries, **pandas** stores a collection of columns.*

Importing Data

We can easily import a CSV file as a pandas **DataFrame** with `pd.read_csv` (**pd** is the standard alias for pandas)

DataFrame object →

```
df = pd.read_csv('filename.csv')
```

	Address	Type	Datetime	Latitude	Longitude	Report Location	Incident Number
0	1333 Harbor Ave Sw	Illegal Burn	2025-12-31 23:59:00	47.591918	-122.382656	POINT (-122.382656 47.591918)	F260000002
1	3323 S Graham St	Aid Response	2025-12-31 23:57:00	47.546103	-122.289056	POINT (-122.289056 47.546103)	F250180050
2	81 CLAY ST	Ladder Code Red	2025-12-31 23:56:00	47.615840	-122.352214	POINT (-122.352214 47.61584)	F250180051
3	E Pine St / Boylston Ave	Aid Response	2025-12-31 23:55:00	47.615231	-122.323415	POINT (-122.323415 47.615231)	F250180049
4	1325 E Madison St	Aid Response	2025-12-31 23:52:00	47.613492	-122.315187	POINT (-122.315187 47.613492)	F250180048

Use **`df`** or **`df.head()`** at the end of a code cell to display the DataFrame

Accessing Columns

pandas enables the selecting and modifying of columns using the **`df[col_name]`** syntax.

	Selecting a Column	Changing a Column
	`df['column1']`	`df['column1'] = vals`
Example	`calls_df['Address']`	`calls_df['New Column'] = 1`

Questions?

Live and on Sli.do

Sli.do #cse160

Filtering

A common data processing task is to **filter** the rows in a table to only those matching a certain **condition**.

`condition` is a **Series** of boolean values

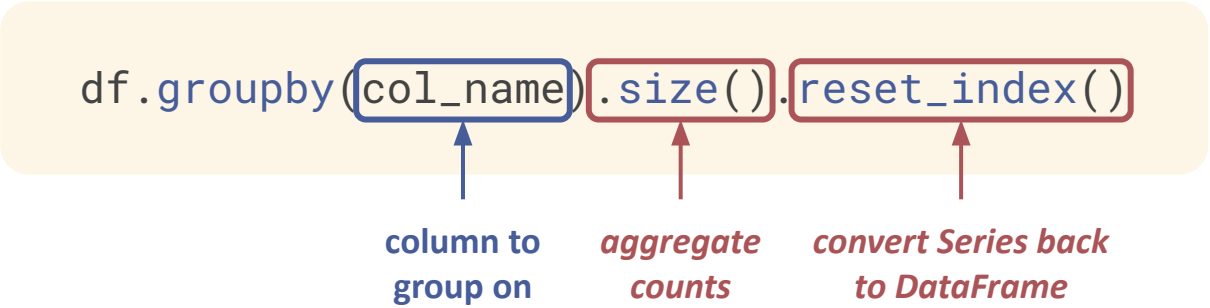
```
condition = df[col_name] == 'target value'

df[condition]
```

select only the rows where the corresponding value in `condition` is **`True`**

Grouping

Grouping involves **combining all of the same values in a given column** into one row and *aggregating* the remaining column values.

To find the frequencies (counts) of each unique value in a column:



Sorting

Sort the rows of a pandas DataFrame by the values in a given column with **`sort_values`**.

```
df.sort_values(col_name, ascending = True)
```

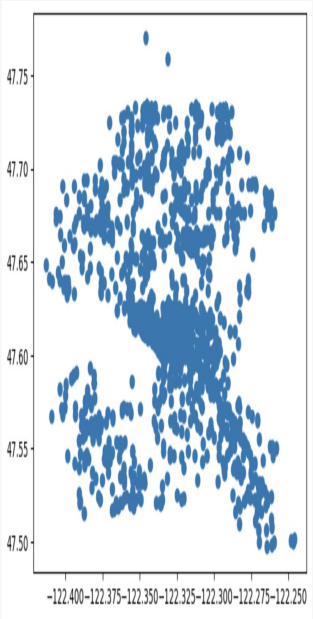
Questions?

Live and on Sli.do

Sli.do #cse160

Think Pair Share

Write the pandas code necessary to determine **which address** in **`calls_df`** had the most SFD 911 calls in 2025.



	Address	Type	Datetime	Latitude	Longitude	Report Location	Incident Number
0	1333 Harbor Ave Sw	Illegal Burn	2025-12-31 23:59:00	47.591918	-122.382656	POINT (-122.382656 47.591918)	F260000002
1	3323 S Graham St	Aid Response	2025-12-31 23:57:00	47.546103	-122.289056	POINT (-122.289056 47.546103)	F250180050
2	81 CLAY ST	Ladder Code Red	2025-12-31 23:56:00	47.615840	-122.352214	POINT (-122.352214 47.61584)	F250180051
3	E Pine St / Boylston Ave	Aid Response	2025-12-31 23:55:00	47.615231	-122.323415	POINT (-122.323415 47.615231)	F250180049
4	1325 E Madison St	Aid Response	2025-12-31 23:52:00	47.613492	-122.315187	POINT (-122.315187 47.613492)	F250180048

“Wow, pandas makes a lot of things much easier!”

Why did I have to take CSE 160???

A few reasons:

1. You need to know **Python syntax** to use pandas

2. You’ll still need to **write and use functions** (or classes or data pipelines) to work with your specific data

3. CSE 160 helps you understand **how things work**, allowing you to know *which* shortcuts to take

Interested in learning more about pandas and data processing?

Take CSE 163!