Midterm Exam Dataframe Reference Sheet - DSC 80, Spring 2024

This page contains information about the dataframes that you will use on the exam. **Only the first few rows are shown for each table.**

The `h` table (left) records addresses within San Diego. Only 50 addresses are recorded. The index of the dataframe contains the numbers 1-50 as unique integers. The data dictionary (right) describes each column.

hid	number	street
1	7370	Torrey Pines Rd
2	960	Mission Blvd
3	5490	La Jolla Village Dr
4	5291	Gilman Dr
5	5834	Torrey Pines Rd

`number` Street address number (`int`).
`street` Street name (`str`).

The `o` table (left) records information on power outages within San Diego in April 2024. Only 100 outages are recorded. The index of the dataframe contains the numbers 1-100 as unique integers. The data dictionary (right) describes each column.

oid	time	duration
1	2024-04-07 03:21:00	3
2	2024-04-20 16:35:00	70
3	2024-04-29 04:29:00	72
4	2024-04-15 13:35:00	27
5	2024-04-11 21:34:00	9

`time` When the outage began (`pd.Timestamp`).
`duration` How long the outage lasted in minutes (`int`).

The `j` table (left) is a table that links outages to addresses. The data dictionary (right) describes each column. Each entry in the `j` table contains the `hid` of the affected address and the `oid` of the outage. For example, the first row of the table records that the outage with an `oid` of 1 caused the power to go out at addresses with `hid` 61 and 88. A single outage can affect multiple addresses. There are no missing values and no duplicated rows in this table, and all values are positive integers. This table records all the addresses affected by all of the outages in 2024 so far.

	hid	oid
0	61	1
1	88	1
2	22	2
3	71	2
4	75	2

`hid` The `hid` of the affected address (`int`).
`oid` The `oid` of the outage (`int`).