Data Wrangling NA

with Pandas
Cheat Sheet

df

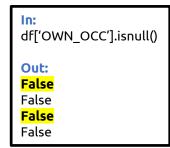| ST_NUM	OWN_OCC
104	--
197	Y
NA	n/a
201	N

df2

ST_NUM	OWN_OCC
104	NA
197	Y
NA	NA
201	N

Detecting NA — Missing values that Pandas recognizes

```
In:
df['ST_NUM'].isnull()

Out:
False
False
True
False
```

```
In:
df['OWN_OCC'].isnull()

Out:
False
False
False
False
```

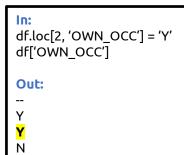Pandas won't detect non-standard missing values such as "**n/a**" and "**—**"

Replacing NA — Replacing different types of missing values

Replace on import

```
In:
missing = ["—","n/a"]
pd.read_csv("data.csv", na_values=missing)
df['OWN_OCC']

Out:
NaN
Y
NaN
N
```

Replace by location

```
In:
df.loc[2, 'OWN_OCC'] = 'Y'
df['OWN_OCC']

Out:
--
Y
Y
N
```

Fill all missing values

```
In:
df['ST_NUM'].fillna(125, inplace=True)
df['ST_NUM']

Out:
104
197
125
201
```

Summarizing — Summarizing missing values

Any missing values?

```
In:
df2.isnull().values.any()

Out:
True
```

Sum by column

```
In:
df2.isnull().sum()

Out:
ST_NUM      1
OWN_OCC     2
```

Total missing values

```
In:
df2.isnull().sum().sum()

Out:
3
```