Pandas

Simple plan

1. Quickly introduce pandas DataFrame object
2. Have a go with pandas in the console
    a. Open up **jupyter notebook** and **tap along** with the slides
3. Practical **Jupyter Notebook** session

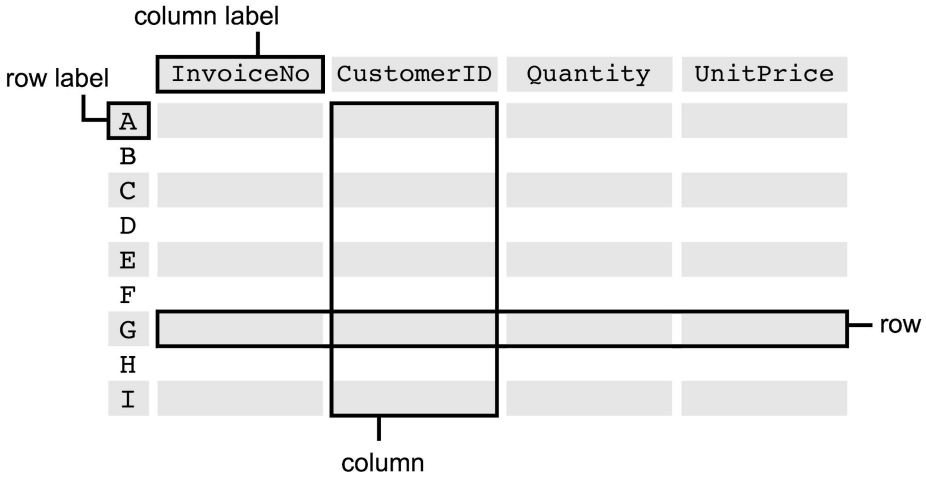
CAMBRIDGE SPARK

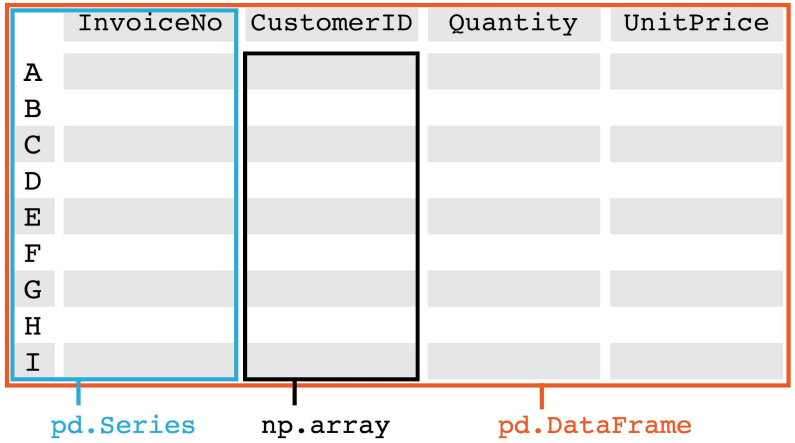
What is Pandas?

- Data **structures** and data **analysis tools**:
  - The ‘excel of python’
- Base data objects are numpy arrays
- Note: **huge** userbase (as with numpy) - your question is on **StackOverflow**!
- Pandas **documentation** is superb
  - https://pandas.pydata.org/pandas-docs/stable/10min.html

CAMBRIDGE SPARK

Anatomy of a pandas DataFrame object





CAMBRIDGE SPARK

Practical introduction

```
Open up a jupyter notebook...
$  import pandas as pd
$  data = {'name': ['alice', 'bob'], 'age': [28, 25]}
$  df = pd.DataFrame(data)
$  df
  >      name  age
     0  alice   28
     1    bob   25
```

CAMBRIDGE SPARK

What's a DataFrame?

```
$  df.values
    >  array([['alice', 28],
              ['bob', 25]], dtype=object)
$  df.dtypes
$  df.columns
$  df.index
$  df.index = ['first', 'second']
$  df
    >            name  age
       first    alice   28
       second     bob   25
```

CAMBRIDGE SPARK

Indexing and selection

```
$  df['name']  # or df.name
$  df.loc['second', 'age']
$  df.iloc[1, 1]
$  df.query('age > 25')
  >          name  age
     first  alice   28
```

Read from CSV

```
$  loc = 'data/airfoil.csv'
$  pd.read_csv(loc)
Can read from websites!
Other arguments, e.g. header, allow for customising the import
such as naming columns, and specifying a column to use for index
$  url = 'https://goo.gl/XE5CrW'
$  pd.read_csv(url, header=None)
Read the documentation!
```

CAMBRIDGE SPARK

Useful functions, methods, and attributes

```
$  pd.isnull(df)
$  df.fillna(value=0)
$  df.describe()
$  df.plot()
$  df.reset_index()
$  df.set_index('name')
$  df.index
$  df.values  # Simba, remember who you are...
$  df.col.unique()  # and maths like df.col.max()
$  df.groupby(...)
```

Hands-on session

01-pandas-skeleton.ipynb
20 mins

CAMBRIDGE SPARK