5_data_selection

May 16, 2019

```
In [ ]: import pandas as pd
```

Data loading

- xlrd , conda install xlrd

```
In [ ]: !conda install --y xlrd
```

```
In [ ]: import numpy as np
        df = pd.read_excel("./data/excel-comp-data.xlsx")
        df.head()
```

data selction with index number and column names

```
In [ ]: df["account"].head(2)
```

```
In [ ]: df[["account", "street", "state"]].head(3)
```

```
In [ ]: df[:3]
```

```
In [ ]: df["name"][:3]
```

```
In [ ]: account_serires = df["account"]
        account_serires[:3]
```

```
In [ ]: account_serires[[1,5,2]]
```

```
In [ ]: account_serires<250000
```

```
In [ ]: account_serires[account_serires<250000]
```

```
In [ ]: df.index = df["account"]
```

```
In [ ]: df.head()
```

```
In [ ]: del df["account"]
        df.head()
```

```
In [ ]: df[["name","street"]][:2]
```

```
In [ ]: df.loc[[211829,320563],["name","street"]]
```

```
In [ ]: df[["name", "street"]].iloc[:10]
```

0.0.1 reindex

```
In [ ]: df.index = list(range(0,15))
        df.head()
```

0.0.2 data drop

```
In [ ]: df

In [ ]: df.drop(1)

In [ ]: df

In [ ]: df.drop([0,1, 2,3], inplace=True)

In [ ]: df

In [ ]: matrix = df.as_matrix()
        matrix[:3]

In [ ]: matrix[:,-3:]

In [ ]: matrix[:,-3:].sum(axis=1)

In [ ]: df.drop("city",axis=1)
        df.drop(["city", "state"],axis=1)

In [ ]: df

In [ ]: matrix = df.as_matrix()
        matrix

In [ ]:

In [ ]:
```