

### read\_excel

```
pd.read_excel(path, sheet_name='Sheet2', encoding='utf-16')
```

### Read multiples sheets

```
df_excel = pd.ExcelFile(path)
sheets = df_excel.sheet_names
df_aba = df_excel.parse(<nome da aba>, skiprows=[1,2], header=None)
```

### Read csv

```
pd.read_csv(path, sep=',', header=None)
```

### Create a dataframe

```
my_dict = {'Computer':1500,'Monitor':300}
df = pd.DataFrame(list(my_dict.items()),columns = ['Products','Prices'])
```

### Iterrows

```
For i, row in df.iterrows():
    print(row)
```

### Replace nan to None

```
row = row.replace({np.nan: None})
```

### Find nan

```
pd[isna()]
```

### Select observations between two datetimes

```
dt_inicial = pd.Timestamp(2020, 1, 30)
dt_final = pd.Timestamp(2020, 1, 31)
df.loc[str(dt_inicial):str(dt_final)]
```

OR

```
df.loc['2020-1-1 01:00:00':'2020-1-1 04:00:00']
```

### Diff between 2 df

```
diff = df1[~df1.astype(str).apply(tuple, 1).isin(df2.astype(str).apply(-tuple, 1))]
```

### Append new line

```
df.append(pd.Series(name='new row'))
```

### New column

```
df['new column'] = np.nan
```

### Substitute values with another value

```
df[2].map({'yes':1, 'no':0})
```

### Column to datetime

```
pd.to_datetime(df[3], format="%Y%m%d%H")
```

### Convert decimal with comman to float

```
df.iloc[:,4].str.replace('.', '').str.replace(',', '.').astype(float, inplace=True)
```

### Datetime64[ns] to date

```
df[3].dt.date
```

### Convert a column to int

```
df.astype(int)
```

### Drop rows

```
f.drop([0, 1, 5], inplace=True)
```

### Drop columns

```
df.drop([2, 5], axis=1)
```

### df to dictionary

```
pd.Series(df[1].values, index=df[0]).to_dict()
```

### Save dataframe

```
df.to_csv('file_out.csv', sep='\t', index=False, encoding='utf-8-sig')
```

