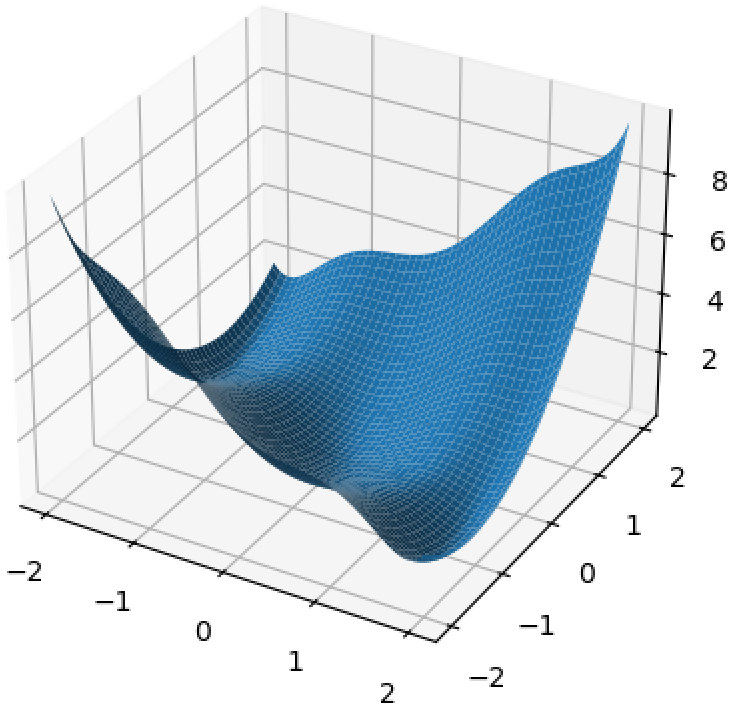
# Tutorial 2

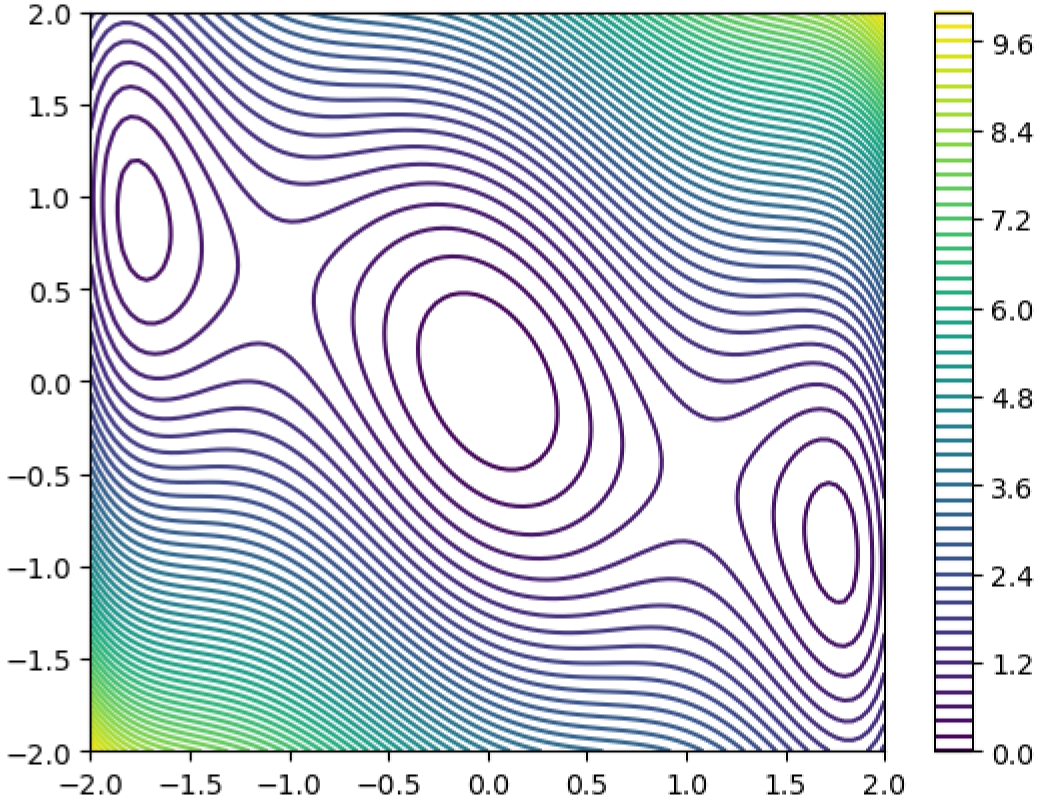
## Outline

- Simulated Annealing (for HW#2)
- Pandas
- Numba

## Simulated Annealing

In [1]:
```
import numpy as np
import matplotlib.pyplot as plt


def Camel(X):
    x, y = X
    return 2 * x**2 - 1.05 * x**4 + x**6 / 6 + x * y + y**2

def plot_surface(func, x_min=-2, x_max=2, y_min=-2, y_max=2):
    a = np.linspace(x_min, x_max, 100)
    b = np.linspace(y_min, y_max, 100)
    x,y = np.meshgrid(a, b)
    z = func((x, y))
    fig = plt.figure()
    ax = fig.add_subplot(projection='3d')
    ax.plot_surface(x, y, z)

plot_surface(Camel)

def draw_path(func, path=None, x_min=-2, x_max=2, y_min=-2, y_max=2):
    a = np.linspace(x_min, x_max, 100)
    b = np.linspace(y_min, y_max, 100)
    x, y = np.meshgrid(a, b)
    z = func((x, y))
    fig, ax = plt.subplots()
    contour = ax.contour(x, y, z, 50)
    plt.colorbar(contour)
    if path:
        ax.plot(path[:, 0], path[:, 1], color='red')
        print(len(path))

draw_path(Camel)
```





No description has been provided for this image

- Random Displacement:

$$X_{i+1} = X_i + \Delta * (2 * \mathrm{URN} - 1)$$

- Metropolis Rule:

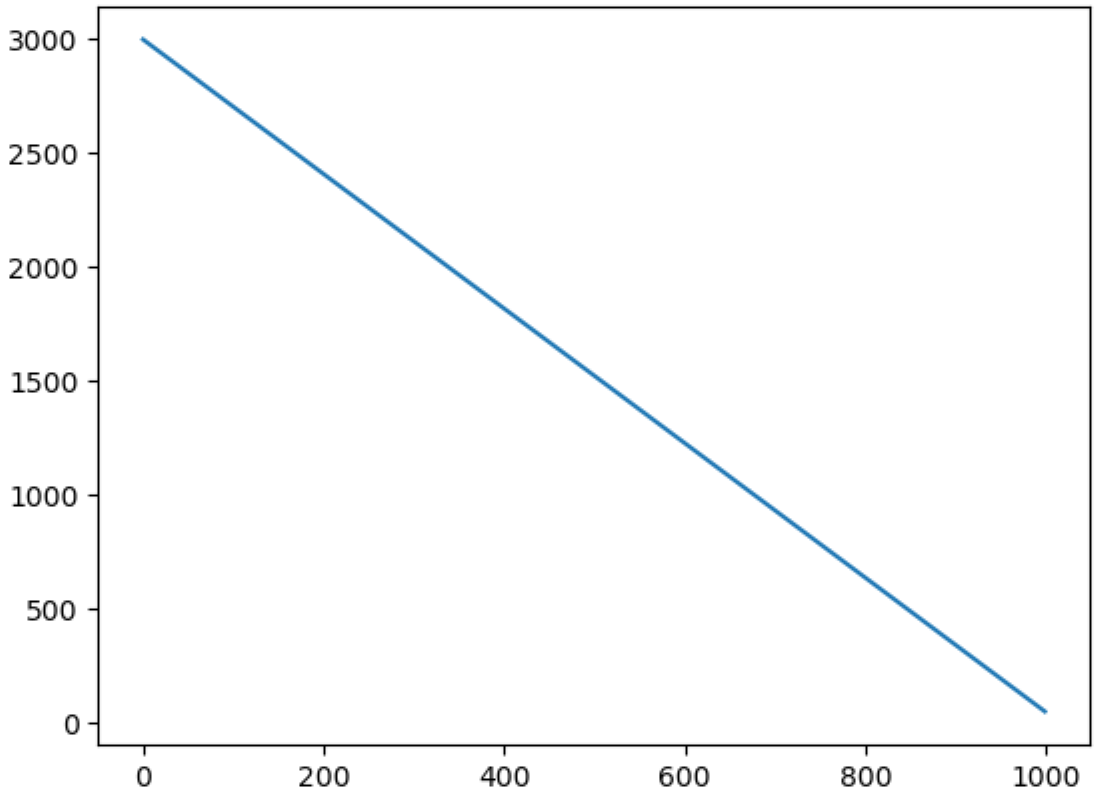
$$P(\text{accept}) = \exp\left(-\frac{\Delta E}{T}\right)$$

In [2]:
```
def SA(solution, func, schedule, delta, boundary, n_iter, report_interval=None):
    """
    Simulated Annealing for minimization

    Parameters
    ----------
    solution: np.ndarray
        Initial guess
    func: Callable
        Function to minimize
    schedule: np.ndarray
        An array of temperatures for simulated annealing
    delta: float
        Magnitude of random displacement
    boundary: tuple
        Boundary of the variables to minimize. (lowerbound,upperbound)
    n_iter: int
        Number of random displacement move in each temperature
    report_interavl: int
        Number of temperature steps to report result

    Returns
    -------
    res: dict
        Minimized point and its evaulation value
    """

    best_solution = solution.copy()
    lowest_eval = func(best_solution)

    for idx, temp in enumerate(schedule):
        if report_interval is not None and ((idx + 1) % report_interval == 0 or idx
            msg = (
                f"{idx + 1}/{len(schedule)}, Temp: {temp:.2f}, "
                f"Best solution: {best_solution}, Value: {lowest_eval:.7f}"
            )
            print(msg)

        for n in range(n_iter):
            trial = solution.copy()
            trial += delta * (2 * np.random.random(trial.shape) - 1)
            if np.all(trial >= boundary[0]) and np.all(trial <= boundary[1]):
                # fill in acceptance criterion
                if np.exp(-(func(trial) - func(solution)) / temp) > np.random.rando
                    solution = trial
                    if func(solution) < lowest_eval:
                        # update solution here
                        best_solution = solution.copy()
                        lowest_eval = func(best_solution)
    return {"solution":best_solution, "evaluation":lowest_eval}
```

Try linear cooling

In [3]:
```
linear_cooling = np.linspace(3000, 50, 1000)
plt.plot(linear_cooling)
```

Out[3]:
```
[<matplotlib.lines.Line2D at 0x1a54d604710>]
```



In [4]:
```
# Start from a point which is close to local minimum
starting_point = np.array([-1.7, 0.7])
SA(starting_point, Camel, linear_cooling, 0.1, (-2, 2), 5, 100)
```

```
1/1000, Temp: 3000.00, Best solution: [-1.7  0.7], Value: 0.3332232
100/1000, Temp: 2707.66, Best solution: [ 0.02742354 -0.38066902], Value: 0.1359731
200/1000, Temp: 2412.36, Best solution: [0.08827732 0.04828203], Value: 0.0221154
300/1000, Temp: 2117.07, Best solution: [ 0.01358763 -0.03255679], Value: 0.0009868
400/1000, Temp: 1821.77, Best solution: [ 0.01358763 -0.03255679], Value: 0.0009868
500/1000, Temp: 1526.48, Best solution: [ 0.01358763 -0.03255679], Value: 0.0009868
600/1000, Temp: 1231.18, Best solution: [ 0.01358763 -0.03255679], Value: 0.0009868
700/1000, Temp: 935.89, Best solution: [ 0.01358763 -0.03255679], Value: 0.0009868
800/1000, Temp: 640.59, Best solution: [ 0.01358763 -0.03255679], Value: 0.0009868
900/1000, Temp: 345.30, Best solution: [ 0.01358763 -0.03255679], Value: 0.0009868
1000/1000, Temp: 50.00, Best solution: [ 0.01358763 -0.03255679], Value: 0.0009868
```

Out[4]:
```
{'solution': array([ 0.01358763, -0.03255679]),
 'evaluation': 0.000986786501804797}
```

In [5]:
```
# Local minimization methods fail to find the global minimum
from scipy.optimize import minimize
```

```
minimize(Camel, starting_point, method="BFGS")
```

Out[5]:
```
  message: Optimization terminated successfully.
  success: True
   status: 0
      fun: 0.298638442236861
        x: [-1.748e+00  8.738e-01]
      nit: 5
      jac: [ 3.129e-07 -7.451e-09]
 hess_inv: [[ 8.564e-02 -4.261e-02]
            [-4.261e-02  5.250e-01]]
     nfev: 21
     njev: 7
```

## Pandas

- Documentation

In [6]:
```
import pandas as pd
```

```
C:\Users\artui\AppData\Local\Temp\ipykernel_12292\4080736814.py:1: DeprecationWarnin
g:
Pyarrow will become a required dependency of pandas in the next major release of pan
das (pandas 3.0),
(to allow more performant data types, such as the Arrow string type, and better inte
roperability with other libraries)
but was not found to be installed on your system.
If this would cause problems for you,
please provide us feedback at https://github.com/pandas-dev/pandas/issues/54466

  import pandas as pd
```

### Read CSV file

In [7]:
```
df = pd.read_csv("Datasets/titanic.csv")
print(type(df))
df
```

```
<class 'pandas.core.frame.DataFrame'>
```

Out[7]:

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | F |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 0 | 3 | Braund, Mr. Owen Harris | male | 22.0 | 1 | 0 | A/5 21171 | 7.2 |
| **1** | 2 | 1 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1 | 0 | PC 17599 | 71.2 |
| **2** | 3 | 1 | 3 | Heikkinen, Miss. Laina | female | 26.0 | 0 | 0 | STON/O2. 3101282 | 7.9 |
| **3** | 4 | 1 | 1 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | female | 35.0 | 1 | 0 | 113803 | 53.1 |
| **4** | 5 | 0 | 3 | Allen, Mr. William Henry | male | 35.0 | 0 | 0 | 373450 | 8.0 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **886** | 887 | 0 | 2 | Montvila, Rev. Juozas | male | 27.0 | 0 | 0 | 211536 | 13.0 |
| **887** | 888 | 1 | 1 | Graham, Miss. Margaret Edith | female | 19.0 | 0 | 0 | 112053 | 30.0 |
| **888** | 889 | 0 | 3 | Johnston, Miss. Catherine Helen "Carrie" | female | NaN | 1 | 2 | W./C. 6607 | 23.4 |
| **889** | 890 | 1 | 1 | Behr, Mr. Karl Howell | male | 26.0 | 0 | 0 | 111369 | 30.0 |
| **890** | 891 | 0 | 3 | Dooley, Mr. Patrick | male | 32.0 | 0 | 0 | 370376 | 7.7 |

891 rows × 12 columns

In [8]:
```
df.head(40)
```

Out[8]:

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 0 | 3 | Braund, Mr. Owen Harris | male | 22.0 | 1 | 0 | A/5 21171 | 7 |
| **1** | 2 | 1 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1 | 0 | PC 17599 | 71 |
| **2** | 3 | 1 | 3 | Heikkinen, Miss. Laina | female | 26.0 | 0 | 0 | STON/O2. 3101282 | 7 |
| **3** | 4 | 1 | 1 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | female | 35.0 | 1 | 0 | 113803 | 53 |
| **4** | 5 | 0 | 3 | Allen, Mr. William Henry | male | 35.0 | 0 | 0 | 373450 | 8 |
| **5** | 6 | 0 | 3 | Moran, Mr. James | male | NaN | 0 | 0 | 330877 | 8 |
| **6** | 7 | 0 | 1 | McCarthy, Mr. Timothy J | male | 54.0 | 0 | 0 | 17463 | 51 |
| **7** | 8 | 0 | 3 | Palsson, Master. Gosta Leonard | male | 2.0 | 3 | 1 | 349909 | 21 |
| **8** | 9 | 1 | 3 | Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg) | female | 27.0 | 0 | 2 | 347742 | 11 |
| **9** | 10 | 1 | 2 | Nasser, Mrs. Nicholas (Adele Achem) | female | 14.0 | 1 | 0 | 237736 | 30 |
| **10** | 11 | 1 | 3 | Sandstrom, Miss. Marguerite Rut | female | 4.0 | 1 | 1 | PP 9549 | 16 |
| **11** | 12 | 1 | 1 | Bonnell, Miss. Elizabeth | female | 58.0 | 0 | 0 | 113783 | 26 |
| **12** | 13 | 0 | 3 | Saundercock, Mr. William Henry | male | 20.0 | 0 | 0 | A/5. 2151 | 8 |

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | |
|---|---|---|---|---|---|---|---|---|---|---|
| **13** | 14 | 0 | 3 | Andersson, Mr. Anders Johan | male | 39.0 | 1 | 5 | 347082 | 31 |
| **14** | 15 | 0 | 3 | Vestrom, Miss. Hulda Amanda Adolfina | female | 14.0 | 0 | 0 | 350406 | 7 |
| **15** | 16 | 1 | 2 | Hewlett, Mrs. (Mary D Kingcome) | female | 55.0 | 0 | 0 | 248706 | 16 |
| **16** | 17 | 0 | 3 | Rice, Master. Eugene | male | 2.0 | 4 | 1 | 382652 | 29 |
| **17** | 18 | 1 | 2 | Williams, Mr. Charles Eugene | male | NaN | 0 | 0 | 244373 | 13 |
| **18** | 19 | 0 | 3 | Vander Planke, Mrs. Julius (Emelia Maria Vande... | female | 31.0 | 1 | 0 | 345763 | 18 |
| **19** | 20 | 1 | 3 | Masselmani, Mrs. Fatima | female | NaN | 0 | 0 | 2649 | 7 |
| **20** | 21 | 0 | 2 | Fynney, Mr. Joseph J | male | 35.0 | 0 | 0 | 239865 | 26 |
| **21** | 22 | 1 | 2 | Beesley, Mr. Lawrence | male | 34.0 | 0 | 0 | 248698 | 13 |
| **22** | 23 | 1 | 3 | McGowan, Miss. Anna "Annie" | female | 15.0 | 0 | 0 | 330923 | 8 |
| **23** | 24 | 1 | 1 | Sloper, Mr. William Thompson | male | 28.0 | 0 | 0 | 113788 | 35 |
| **24** | 25 | 0 | 3 | Palsson, Miss. Torborg Danira | female | 8.0 | 3 | 1 | 349909 | 21 |
| **25** | 26 | 1 | 3 | Asplund, Mrs. Carl Oscar (Selma Augusta Emilia... | female | 38.0 | 1 | 5 | 347077 | 31 |
| **26** | 27 | 0 | 3 | Emir, Mr. Farred | male | NaN | 0 | 0 | 2631 | 7 |

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Chehab | | | | | | |
| **27** | 28 | 0 | 1 | Fortune, Mr. Charles Alexander | male | 19.0 | 3 | 2 | 19950 | 263 |
| **28** | 29 | 1 | 3 | O'Dwyer, Miss. Ellen "Nellie" | female | NaN | 0 | 0 | 330959 | 7 |
| **29** | 30 | 0 | 3 | Todoroff, Mr. Lalio | male | NaN | 0 | 0 | 349216 | 7 |
| **30** | 31 | 0 | 1 | Uruchurtu, Don. Manuel E | male | 40.0 | 0 | 0 | PC 17601 | 27 |
| **31** | 32 | 1 | 1 | Spencer, Mrs. William Augustus (Marie Eugenie) | female | NaN | 1 | 0 | PC 17569 | 146 |
| **32** | 33 | 1 | 3 | Glynn, Miss. Mary Agatha | female | NaN | 0 | 0 | 335677 | 7 |
| **33** | 34 | 0 | 2 | Wheadon, Mr. Edward H | male | 66.0 | 0 | 0 | C.A. 24579 | 10 |
| **34** | 35 | 0 | 1 | Meyer, Mr. Edgar Joseph | male | 28.0 | 1 | 0 | PC 17604 | 82 |
| **35** | 36 | 0 | 1 | Holverson, Mr. Alexander Oskar | male | 42.0 | 1 | 0 | 113789 | 52 |
| **36** | 37 | 1 | 3 | Mamee, Mr. Hanna | male | NaN | 0 | 0 | 2677 | 7 |
| **37** | 38 | 0 | 3 | Cann, Mr. Ernest Charles | male | 21.0 | 0 | 0 | A./5. 2152 | 8 |
| **38** | 39 | 0 | 3 | Vander Planke, Miss. Augusta Maria | female | 18.0 | 2 | 0 | 345764 | 18 |
| **39** | 40 | 1 | 3 | Nicola-Yarred, Miss. Jamila | female | 14.0 | 1 | 0 | 2651 | 11 |

In [9]:
```
df.tail(10)
```

Out[9]:

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | |
|---|---|---|---|---|---|---|---|---|---|---|
| **881** | 882 | 0 | 3 | Markun, Mr. Johann | male | 33.0 | 0 | 0 | 349257 | 7. |
| **882** | 883 | 0 | 3 | Dahlberg, Miss. Gerda Ulrika | female | 22.0 | 0 | 0 | 7552 | 10. |
| **883** | 884 | 0 | 2 | Banfield, Mr. Frederick James | male | 28.0 | 0 | 0 | C.A./SOTON 34068 | 10. |
| **884** | 885 | 0 | 3 | Sutehall, Mr. Henry Jr | male | 25.0 | 0 | 0 | SOTON/OQ 392076 | 7. |
| **885** | 886 | 0 | 3 | Rice, Mrs. William (Margaret Norton) | female | 39.0 | 0 | 5 | 382652 | 29. |
| **886** | 887 | 0 | 2 | Montvila, Rev. Juozas | male | 27.0 | 0 | 0 | 211536 | 13. |
| **887** | 888 | 1 | 1 | Graham, Miss. Margaret Edith | female | 19.0 | 0 | 0 | 112053 | 30. |
| **888** | 889 | 0 | 3 | Johnston, Miss. Catherine Helen "Carrie" | female | NaN | 1 | 2 | W./C. 6607 | 23. |
| **889** | 890 | 1 | 1 | Behr, Mr. Karl Howell | male | 26.0 | 0 | 0 | 111369 | 30. |
| **890** | 891 | 0 | 3 | Dooley, Mr. Patrick | male | 32.0 | 0 | 0 | 370376 | 7. |

## Drop columns

In [10]:
```
df2 = df.drop(['Cabin'], axis=1, inplace=True)
```

## Drop NaN values

Drop the rows where at least one element is missing

```
In [11]: df.dropna()
```

Out[11]:

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fa |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 0 | 3 | Braund, Mr. Owen Harris | male | 22.0 | 1 | 0 | A/5 21171 | 7.25 |
| **1** | 2 | 1 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1 | 0 | PC 17599 | 71.28 |
| **2** | 3 | 1 | 3 | Heikkinen, Miss. Laina | female | 26.0 | 0 | 0 | STON/O2. 3101282 | 7.92 |
| **3** | 4 | 1 | 1 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | female | 35.0 | 1 | 0 | 113803 | 53.10 |
| **4** | 5 | 0 | 3 | Allen, Mr. William Henry | male | 35.0 | 0 | 0 | 373450 | 8.05 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **885** | 886 | 0 | 3 | Rice, Mrs. William (Margaret Norton) | female | 39.0 | 0 | 5 | 382652 | 29.12 |
| **886** | 887 | 0 | 2 | Montvila, Rev. Juozas | male | 27.0 | 0 | 0 | 211536 | 13.00 |
| **887** | 888 | 1 | 1 | Graham, Miss. Margaret Edith | female | 19.0 | 0 | 0 | 112053 | 30.00 |
| **889** | 890 | 1 | 1 | Behr, Mr. Karl Howell | male | 26.0 | 0 | 0 | 111369 | 30.00 |
| **890** | 891 | 0 | 3 | Dooley, Mr. Patrick | male | 32.0 | 0 | 0 | 370376 | 7.75 |

712 rows × 11 columns

Drop the columns where at least one element is missing.

```
In [12]: df.dropna(axis=1) # or axis='columns'
```

Out[12]:

| | PassengerId | Survived | Pclass | Name | Sex | SibSp | Parch | Ticket | Fare |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 0 | 3 | Braund, Mr. Owen Harris | male | 1 | 0 | A/5 21171 | 7.2500 |
| **1** | 2 | 1 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 1 | 0 | PC 17599 | 71.2833 |
| **2** | 3 | 1 | 3 | Heikkinen, Miss. Laina | female | 0 | 0 | STON/O2. 3101282 | 7.9250 |
| **3** | 4 | 1 | 1 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | female | 1 | 0 | 113803 | 53.1000 |
| **4** | 5 | 0 | 3 | Allen, Mr. William Henry | male | 0 | 0 | 373450 | 8.0500 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **886** | 887 | 0 | 2 | Montvila, Rev. Juozas | male | 0 | 0 | 211536 | 13.0000 |
| **887** | 888 | 1 | 1 | Graham, Miss. Margaret Edith | female | 0 | 0 | 112053 | 30.0000 |
| **888** | 889 | 0 | 3 | Johnston, Miss. Catherine Helen "Carrie" | female | 1 | 2 | W./C. 6607 | 23.4500 |
| **889** | 890 | 1 | 1 | Behr, Mr. Karl Howell | male | 0 | 0 | 111369 | 30.0000 |
| **890** | 891 | 0 | 3 | Dooley, Mr. Patrick | male | 0 | 0 | 370376 | 7.7500 |

891 rows × 9 columns

Define in which columns to look for missing values.

```
In [13]: df.dropna(subset=['Age'])
```

Out[13]:

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fa |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 0 | 3 | Braund, Mr. Owen Harris | male | 22.0 | 1 | 0 | A/5 21171 | 7.25 |
| **1** | 2 | 1 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1 | 0 | PC 17599 | 71.28 |
| **2** | 3 | 1 | 3 | Heikkinen, Miss. Laina | female | 26.0 | 0 | 0 | STON/O2. 3101282 | 7.92 |
| **3** | 4 | 1 | 1 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | female | 35.0 | 1 | 0 | 113803 | 53.10 |
| **4** | 5 | 0 | 3 | Allen, Mr. William Henry | male | 35.0 | 0 | 0 | 373450 | 8.05 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **885** | 886 | 0 | 3 | Rice, Mrs. William (Margaret Norton) | female | 39.0 | 0 | 5 | 382652 | 29.12 |
| **886** | 887 | 0 | 2 | Montvila, Rev. Juozas | male | 27.0 | 0 | 0 | 211536 | 13.00 |
| **887** | 888 | 1 | 1 | Graham, Miss. Margaret Edith | female | 19.0 | 0 | 0 | 112053 | 30.00 |
| **889** | 890 | 1 | 1 | Behr, Mr. Karl Howell | male | 26.0 | 0 | 0 | 111369 | 30.00 |
| **890** | 891 | 0 | 3 | Dooley, Mr. Patrick | male | 32.0 | 0 | 0 | 370376 | 7.75 |

714 rows × 11 columns

## Indexing

In [14]:
```
df.loc[0, "Pclass"]
```

Out[14]:
```
3
```

In [15]:
```
df.iloc[0, 2]
```

Out[15]:
```
3
```

In [16]:
```
df[['Pclass', 'Survived']]
```

Out[16]:

| | Pclass | Survived |
|---|---|---|
| **0** | 3 | 0 |
| **1** | 1 | 1 |
| **2** | 3 | 1 |
| **3** | 1 | 1 |
| **4** | 3 | 0 |
| **...** | ... | ... |
| **886** | 2 | 0 |
| **887** | 1 | 1 |
| **888** | 3 | 0 |
| **889** | 1 | 1 |
| **890** | 3 | 0 |

891 rows × 2 columns

In [17]:
```
df[df['Survived'] == 1]
```

Out[17]:

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fa |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 2 | 1 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1 | 0 | PC 17599 | 71.28 |
| **2** | 3 | 1 | 3 | Heikkinen, Miss. Laina | female | 26.0 | 0 | 0 | STON/O2. 3101282 | 7.92 |
| **3** | 4 | 1 | 1 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | female | 35.0 | 1 | 0 | 113803 | 53.1 |
| **8** | 9 | 1 | 3 | Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg) | female | 27.0 | 0 | 2 | 347742 | 11.13 |
| **9** | 10 | 1 | 2 | Nasser, Mrs. Nicholas (Adele Achem) | female | 14.0 | 1 | 0 | 237736 | 30.07 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **875** | 876 | 1 | 3 | Najib, Miss. Adele Kiamie "Jane" | female | 15.0 | 0 | 0 | 2667 | 7.22 |
| **879** | 880 | 1 | 1 | Potter, Mrs. Thomas Jr (Lily Alexenia Wilson) | female | 56.0 | 0 | 1 | 11767 | 83.15 |
| **880** | 881 | 1 | 2 | Shelley, Mrs. William (Imanita Parrish Hall) | female | 25.0 | 0 | 1 | 230433 | 26.0 |
| **887** | 888 | 1 | 1 | Graham, Miss. | female | 19.0 | 0 | 0 | 112053 | 30.0 |

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fa |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Margaret Edith | | | | | | |
| **889** | 890 | 1 | 1 | Behr, Mr. Karl Howell | male | 26.0 | 0 | 0 | 111369 | 30.0( |

342 rows × 11 columns

## Other useful methods

Use `.values` attribute to get values in numpy.ndarray

In [18]:
```
df.values
```

Out[18]:
```
array([[1, 0, 3, ..., 'A/5 21171', 7.25, 'S'],
       [2, 1, 1, ..., 'PC 17599', 71.2833, 'C'],
       [3, 1, 3, ..., 'STON/O2. 3101282', 7.925, 'S'],
       ...,
       [889, 0, 3, ..., 'W./C. 6607', 23.45, 'S'],
       [890, 1, 1, ..., '111369', 30.0, 'C'],
       [891, 0, 3, ..., '370376', 7.75, 'Q']], dtype=object)
```

Use `describe()` method to get statistics

In [19]:
```
df.describe()
```

Out[19]:

| | PassengerId | Survived | Pclass | Age | SibSp | Parch | Fare |
|---|---|---|---|---|---|---|---|
| **count** | 891.000000 | 891.000000 | 891.000000 | 714.000000 | 891.000000 | 891.000000 | 891.00000( |
| **mean** | 446.000000 | 0.383838 | 2.308642 | 29.699118 | 0.523008 | 0.381594 | 32.20420 |
| **std** | 257.353842 | 0.486592 | 0.836071 | 14.526497 | 1.102743 | 0.806057 | 49.69342 |
| **min** | 1.000000 | 0.000000 | 1.000000 | 0.420000 | 0.000000 | 0.000000 | 0.00000( |
| **25%** | 223.500000 | 0.000000 | 2.000000 | 20.125000 | 0.000000 | 0.000000 | 7.91040( |
| **50%** | 446.000000 | 0.000000 | 3.000000 | 28.000000 | 0.000000 | 0.000000 | 14.45420( |
| **75%** | 668.500000 | 1.000000 | 3.000000 | 38.000000 | 1.000000 | 0.000000 | 31.00000( |
| **max** | 891.000000 | 1.000000 | 3.000000 | 80.000000 | 8.000000 | 6.000000 | 512.32920( |

Use `.index` or `.columns` get index/columns

In [20]:
```
for index in df.index:
    print(index)
```

```
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
```

```
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
```

```
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
```

```
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
```

```
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
```

```
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
```

```
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
```

```
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
```

```
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
```

```
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
```

```
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
```

```
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
```

```
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
```

In [21]:
```
for col in df.columns:
    print(col)
```

```
PassengerId
Survived
Pclass
Name
Sex
Age
SibSp
Parch
Ticket
Fare
Embarked
```

Use `to_csv()` to export DataFrame

In [22]:
```
df.to_csv("test_export.csv")
```

Use `sort_values()` method to sort the DataFrame according to values in one column.

In [23]:
```
df.sort_values(by=["Age"], ascending=False)
```

Out[23]:

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | F |
|---|---|---|---|---|---|---|---|---|---|---|
| **630** | 631 | 1 | 1 | Barkworth, Mr. Algernon Henry Wilson | male | 80.0 | 0 | 0 | 27042 | 30.0 |
| **851** | 852 | 0 | 3 | Svensson, Mr. Johan | male | 74.0 | 0 | 0 | 347060 | 7.7 |
| **493** | 494 | 0 | 1 | Artagaveytia, Mr. Ramon | male | 71.0 | 0 | 0 | PC 17609 | 49.5 |
| **96** | 97 | 0 | 1 | Goldschmidt, Mr. George B | male | 71.0 | 0 | 0 | PC 17754 | 34.6 |
| **116** | 117 | 0 | 3 | Connors, Mr. Patrick | male | 70.5 | 0 | 0 | 370369 | 7.7 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **859** | 860 | 0 | 3 | Razi, Mr. Raihed | male | NaN | 0 | 0 | 2629 | 7.2 |
| **863** | 864 | 0 | 3 | Sage, Miss. Dorothy Edith "Dolly" | female | NaN | 8 | 2 | CA. 2343 | 69.5 |
| **868** | 869 | 0 | 3 | van Melkebeke, Mr. Philemon | male | NaN | 0 | 0 | 345777 | 9.5 |
| **878** | 879 | 0 | 3 | Laleff, Mr. Kristo | male | NaN | 0 | 0 | 349217 | 7.8 |
| **888** | 889 | 0 | 3 | Johnston, Miss. Catherine Helen "Carrie" | female | NaN | 1 | 2 | W./C. 6607 | 23.4 |

891 rows × 11 columns

## Numba

Use `pip install numba` or `conda install numba -c conda-forge` to install numba package.

Numba is a package that help users to accelerate the code.

In [24]:
```python
import numba
```

In [25]:
```python
def test():
    i = 0
    for a in range(100000):
        i += a
    return i

%timeit test()
```

```
4.98 ms ± 219 µs per loop (mean ± std. dev. of 7 runs, 100 loops each)
```

In [26]:
```python
@numba.njit()
def test():
    i = 0
    for a in range(100000):
        i += a
    return i

%timeit test()
```

```
The slowest run took 65.00 times longer than the fastest. This could mean that an in
termediate result is being cached.
1.1 µs ± 2.21 µs per loop (mean ± std. dev. of 7 runs, 1 loop each)
```