```python
In [1]: import numpy as np
        from csv import reader
        the_file = open("taxis.csv", encoding="UTF-8")
        read_file = reader(the_file)
        list_taxi = list(read_file)
        list_taxi_header = list_taxi[0]
        list_taxi = list_taxi[1:]
```

```python
In [2]: list_taxi_header
```

```
Out[2]: ['pickup_year',
         'pickup_month',
         'pickup_day',
         'pickup_dayofweek',
         'pickup_time',
         'pickup_location_code',
         'dropoff_location_code',
         'trip_distance',
         'trip_length',
         'fare_amount',
         'fees_amount',
         'tolls_amount',
         'tip_amount',
         'total_amount',
         'payment_type']
```

```python
In [3]: list_taxi[0][13]
```

```
Out[3]: '69.99'
```

In [4]:
```python
list_taxis = []
for checks in list_taxi:
    list_tax = []
    for check in checks:
        list_tax.append(float(check))
    list_taxis.append(list_tax)
list_taxis
```

Out[4]:
```
[[2016.0,
  1.0,
  1.0,
  5.0,
  0.0,
  2.0,
  4.0,
  21.0,
  2037.0,
  52.0,
  0.8,
  5.54,
  11.65,
  69.99,
  1.0],
 [2016.0,
  1.0,
  1.0,
  5.0,
```

In [6]:
```python
import numpy as np
taxi_ndarray = np.array(list_taxis)
```

In [ ]:
```python
taxi_ndarray[:, 1:3]
```

In [ ]:
```python
this = []
for checks in list_taxi:
    this.append([checks[1], checks[2]])
this
```

In [ ]:
```python
like_nd = []
for checks in list_taxi:
    before_nd = []
    for check in checks[1:3]:
        before_nd.append(check)
    like_nd.append(before_nd)
like_nd
```

In [ ]:
```python
list_taxi[0][5]
```

In [ ]:
```python
a = np.array([2, 7])
b = np.array([2, 7])
a + b
```

In [ ]:
```python
total_amount = taxi_ndarray[:, 9:13]
```

```python
total_amount
```

```python
total_amount.sum(axis=1)
```

```python
a = np.array([[2, 4, 6],
              [1, 2, 3]])
a.sum(axis=1)
```

```python
tax_ndarray[0, 13]
```

```python
list_taxi_header
```

```python
taxi_average = taxi_ndarray[:, 7] / (taxi_ndarray[:, 8] / 3600)
```

```python
taxi_average
```

```python
print(tax_ndarray)
```

```python
list_taxi[0][8]
```

```python
data = np.genfromtxt('taxis.csv', delimiter=',', skip_header=1)
```

```python
data
```

```python
data[0][8]
```

```python
pick = taxi_ndarray[:, 1]
```

```python
pick_bool = pick == 5
```

```python
pick[pick_bool] = 5
```

```python
picker.shape
```

```python
feb = taxi_ndarray[:, 1]
```

```python
feb_bool = feb == 2
```

```python
february = feb[feb_bool]
```

```python
february.shape[0]
```

```python
taxi_average
```

```python
miles_per_hour = taxi_average > 2000
```

```python
needed = taxi_ndarray[miles_per_hour, 5:9]
```

```python
needed.shape[0]
```

```python
needed
```

```python
tip = taxi_ndarray[:, 12]
```

```python
tip_bool = tip > 50
```

```python
tip_boolean = tip[tip_bool]
```

```python
tiping_boolean = tip_boolean.shape[0]
```

```python
taxi_ndarray[tip_bool, 0:13]
```

```python
taxi_modified = taxi_ndarray.copy()
```

```python
taxi_modified
```

```python
taxi_modified[550:552, 7] = means
```

```python
taxi_modified
```

```python
col_7 = taxi_modified[:, 7]
```

```python
taxi_modified
```

```python
means = taxi_modified[:, 7].mean()
```

```python
means = round(means, 2)
```

```python
means
```

```python
total_amount = taxi_modified[:, 8]
```

```python
total_amount_bool = total_amount > 0
```

```python
total_amount[total_amount_bool] = 1
```

```python
total_amount[total_amount < 0, 1] = 0
```

```python
amount = taxi_ndarray[:, 13]
```

```python
amount_bool = amount < 0
```

```python
amount_tot = amount[amount_bool]
```

```python
amount_tot.shape[0]
```

```python
taxi_modified[taxi_modified[total_amount < 0, ]
```

```python
taxi_modified[0, 13]
```

```python
total_amount[total_amount_bool] = 0
```

```python
a2 = np.array([1, 2, 3, 4, 5])

a2_bool = a2 > 2

a2[a2_bool] = 99

print(a2)
```

```python
taxi_modified[0, 1]
```

```python
pickup = taxi_modified[:, 1]
```

```python
pickup_bool = pickup >= 1
```

```python
pickup[pickup_bool] = 0
```

```python
taxi_modified[1000, 1]
```

```python
taxi_modified.shape[0]
```

```python
zeros = np.zeros([taxi_modified.shape[0], 1])
```

```python
the_taxi_modified = np.concatenate([taxi_modified, zeros], axis=1)
```

```python
the_taxi_modified[0, 15]
```

```python
jfk = the_taxi_modified[:, 5]
```

```python
jfk_bool = jfk == 2
```

```python
the_taxi_modified[jfk_bool, 15] = 1
```

```python
heathrow = the_taxi_modified[:, 5]
```

```
In [ ]:  heathrow_bool = heathrow == 4
```

```
In [ ]:  heathrow[heathrow_bool] = 2
```

```
In [7]:  taxi_ndarray
```

```
Out[7]:  array([[2.016e+03, 1.000e+00, 1.000e+00, ..., 1.165e+01, 6.999e+01,
                 1.000e+00],
                [2.016e+03, 1.000e+00, 1.000e+00, ..., 8.000e+00, 5.430e+01,
                 1.000e+00],
                [2.016e+03, 1.000e+00, 1.000e+00, ..., 0.000e+00, 3.780e+01,
                 2.000e+00],
                ...,
                [2.016e+03, 6.000e+00, 3.000e+01, ..., 5.000e+00, 6.334e+01,
                 1.000e+00],
                [2.016e+03, 6.000e+00, 3.000e+01, ..., 8.950e+00, 4.475e+01,
                 1.000e+00],
                [2.016e+03, 6.000e+00, 3.000e+01, ..., 0.000e+00, 5.484e+01,
                 2.000e+00]])
```

```
In [8]:  zeros = np.zeros([taxi_ndarray.shape[0], 1])
```

```
In [9]:  the_taxi_ndarray = np.concatenate([taxi_ndarray, zeros], axis=1)
```

```
In [28]:  the_jfk = the_taxi_ndarray[:, 6]
```

```
In [29]:  the_jfk_bool = the_jfk == 2
```

```
In [30]:  my_jfk = the_jfk[the_jfk_bool]
```

```
In [31]:  jfk_counts = my_jfk.shape[0]
```

```
In [32]:  jfk_counts
```

```
Out[32]:  285
```

```
In [33]:  the_laguacha = the_taxi_ndarray[:, 6]
          the_laguach_bool = the_laguacha == 3
          my_laguacha = the_laguacha[the_laguach_bool]
          laguaja_counts = my_laguacha.shape[0]
          laguaja_counts
```

```
Out[33]:  308
```

```
In [34]:  the_newark = the_taxi_ndarray[:, 6]
          the_newark_bool = the_newark == 5
          my_newark = the_newark[the_newark_bool]
          newark_counts = my_newark.shape[0]
          newark_counts
```

```
Out[34]:  2
```

```
In [35]: taxi_average = the_taxi_ndarray[:, 7] / (the_taxi_ndarray[:, 8] / 3600)
```

```
In [41]: taxi_average.shape[0]
```

Out[41]: 2013

```
In [37]: taxi_average_bool = taxi_average < 100
```

```
In [55]: average_taxi_under_100 = the_taxi_ndarray[taxi_average_bool]
```

```
In [56]: average_taxi_under_100
```

Out[56]: array([[2.016e+03, 1.000e+00, 1.000e+00, ..., 6.999e+01, 1.000e+00,
            0.000e+00],
           [2.016e+03, 1.000e+00, 1.000e+00, ..., 5.430e+01, 1.000e+00,
            0.000e+00],
           [2.016e+03, 1.000e+00, 1.000e+00, ..., 3.780e+01, 2.000e+00,
            0.000e+00],
           ...,
           [2.016e+03, 6.000e+00, 3.000e+01, ..., 6.334e+01, 1.000e+00,
            0.000e+00],
           [2.016e+03, 6.000e+00, 3.000e+01, ..., 4.475e+01, 1.000e+00,
            0.000e+00],
           [2.016e+03, 6.000e+00, 3.000e+01, ..., 5.484e+01, 2.000e+00,
            0.000e+00]])

```
In [42]: meaning = average_taxi_under_100.shape[0]
```

```
In [45]: meaning
```

Out[45]: 2004

```
In [47]: average_taxi_under_100.mean()
```

Out[47]: 24.440989804211227

```
In [54]: trip_mph = taxi[:,7] / (taxi[:,8] / 3600)

         cleaned_taxi = taxi[trip_mph < 100]

         mean_distance = cleaned_taxi[:,7].mean()
         mean_length = cleaned_taxi[:,8].mean()
         mean_total_amount = cleaned_taxi[:,13].mean()
```

```
         ---------------------------------------------------------------------------
         IndexError                                Traceback (most recent call last)
         ~\AppData\Local\Temp\ipykernel_3392\873581025.py in <module>
         ----> 1 distance = average_taxi_under_100[:, 7].mean()

         IndexError: too many indices for array: array is 1-dimensional, but 2 were inde
         xed
```

In [ ]: