Adversarial AI. Project Step 1 - Ivan Miller

a) The number of rows in the sample data set

643,214

In [3]:
```
row_num = df.shape[0]
print(f'The number of rows is {row_num:,}')
```

```
The number of rows is 643,214
```

b) The average value of each numerical column

NV_rand column average: 1,734,836.188
VOLUME_rand column average: 32,226.431

b) The average value of each numerical column (there are two of them)

In [4]:
```
# Option1.
# Checking descriptive statistics of the numerical columns which will return the average (mean)
df.describe()
```

Out[4]:

	NV_rand	VOLUME_rand
count	6.432140e+05	6.432140e+05
mean	1.734836e+06	3.222643e+04
std	2.031613e+07	1.849409e+05
min	6.630502e-02	1.000000e+00
25%	6.475344e+03	4.460000e+02
50%	5.821660e+04	3.108000e+03
75%	4.158061e+05	1.445500e+04
max	6.053256e+09	1.736854e+07

In [5]:
```
# Option2 checking the averages separately for both columns
NV_rand_mean = df.NV_rand.mean()
VOLUME_rand_mean = df.VOLUME_rand.mean()

print(f'NV_rand column average: {NV_rand_mean:,.3f}')
print(f'VOLUME_rand column average: {VOLUME_rand_mean:,.3f}')
```

```
NV_rand column average: 1,734,836.188
VOLUME_rand column average: 32,226.431
```

c) The min and max values of a time-based column

Min value of the time-based column("Time") is 2021-09-20 09:30:00
Max value of the time-based column("Time") is 2021-09-21 16:00:00

In [6]:
```
first_record = df.Time.min()
last_record = df.Time.max()

print(f'Min value of the time-based column("Time") is {first_record}')
print(f'Max value of the time-based column("Time") is {last_record}')
```

```
Min value of the time-based column("Time") is 2021-09-20 09:30:00
Max value of the time-based column("Time") is 2021-09-21 16:00:00
```