In [ ]:
```
import pandas as pd
import numpy as np
import statistics
from statistics import mean
import matplotlib.pyplot as plt
```

In [ ]:
```
df=pd.read_csv('C:\\Users\\eric.park\\Downloads\\vaccine_doses.csv')
```

In [ ]:
```
dft=df[['report_date','previous_day_at_least_one','previous_day_fully_vaccina
```

In [ ]:
```
#count NAs
```

In [ ]:
```
dft.isna().sum()
```

Out[6]:
```
report_date                             0
previous_day_at_least_one               9
previous_day_fully_vaccinated           9
previous_day_3doses                   372
previous_day_total_doses_administered   1
dtype: int64
```

In [ ]:
```
#Data Type
```

In [ ]:
```
dft.dtypes
```

Out[7]:
```
report_date                            object
previous_day_at_least_one             float64
previous_day_fully_vaccinated         float64
previous_day_3doses                   float64
previous_day_total_doses_administered float64
dtype: object
```

In [ ]:
```
dft.describe().applymap('{:,.2f}'.format)
```

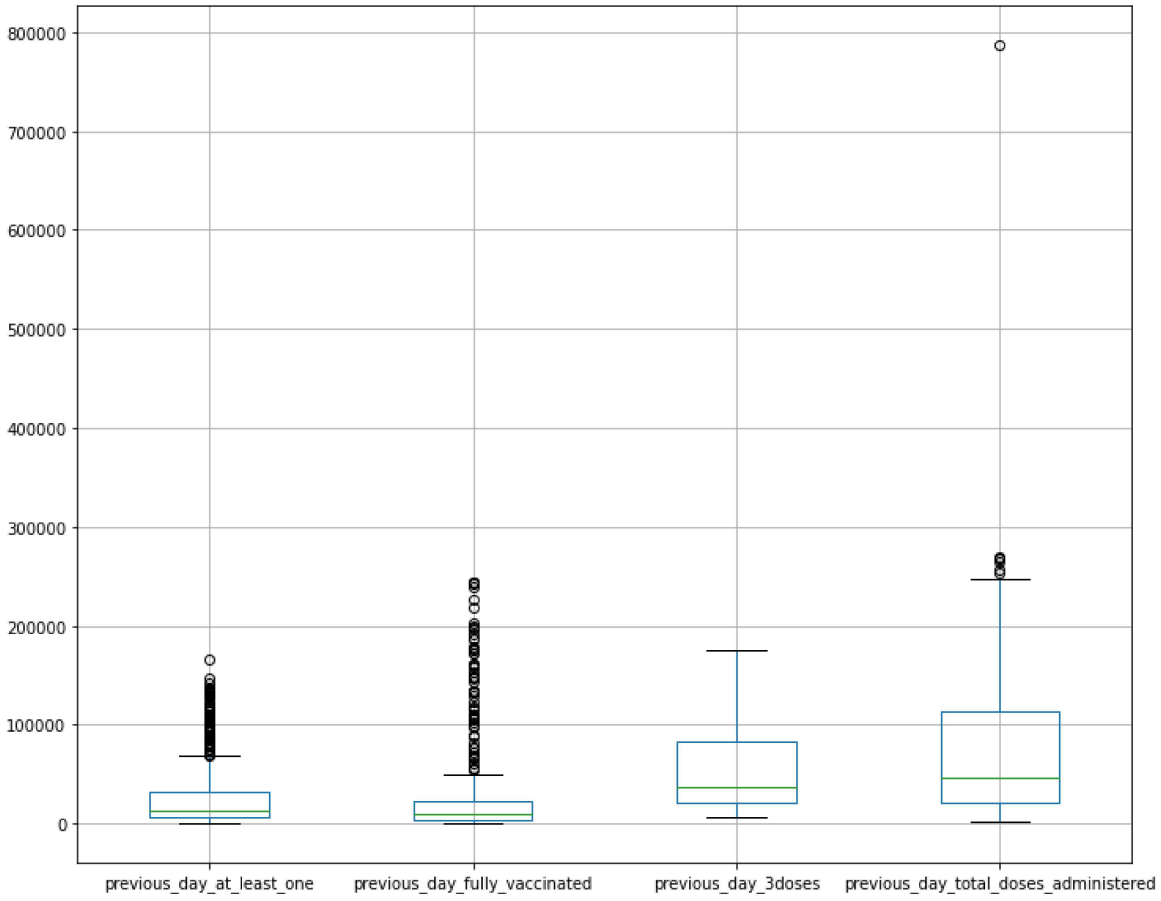
Out[8]:

| | previous_day_at_least_one | previous_day_fully_vaccinated | previous_day_3doses | previou |
|---|---|---|---|---|
| **count** | 417.00 | 417.00 | 54.00 | |
| **mean** | 29,988.06 | 28,789.57 | 57,052.72 | |
| **std** | 37,344.18 | 49,956.21 | 48,873.07 | |
| **min** | 204.00 | 0.00 | 7,021.00 | |
| **25%** | 6,635.00 | 3,854.00 | 20,645.75 | |
| **50%** | 13,171.00 | 9,226.00 | 36,666.00 | |
| **75%** | 31,530.00 | 22,811.00 | 82,862.50 | |
| **max** | 165,905.00 | 244,701.00 | 176,118.00 | |

In [ ]:
```
#box plot
```

In [ ]:
```
dft.boxplot(figsize=(12,10))
```

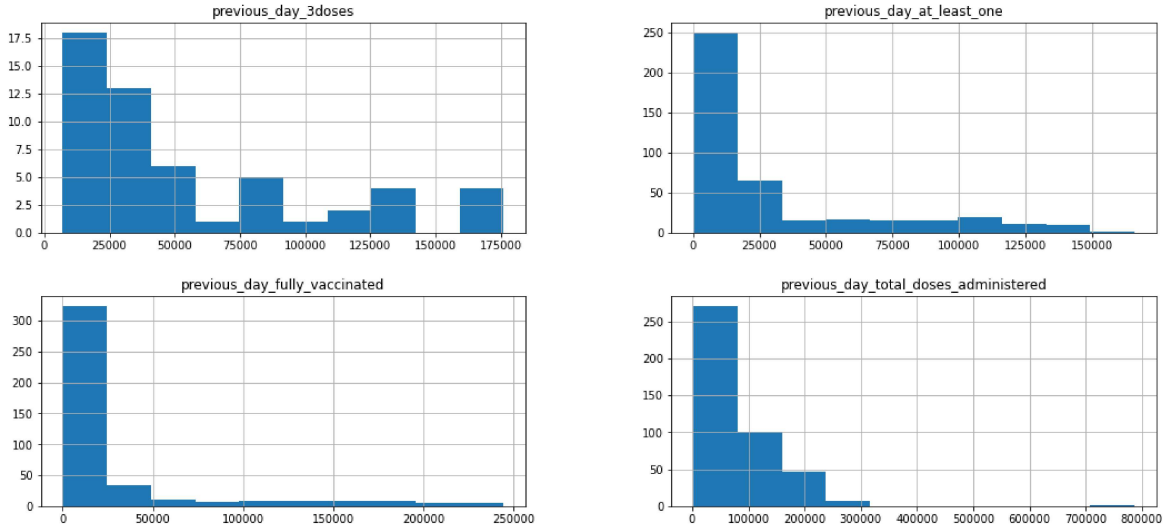
Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x2187092c340>




In [ ]:
```
# histograph
```

In [ ]:
```
dft.hist(figsize=(18,8))
```

Out[10]:
```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x00000218710B1EB0
>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x0000021871372280
>],
       [<matplotlib.axes._subplots.AxesSubplot object at 0x00000218710ED670
>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x0000021871119AF0
>]],
      dtype=object)
```




In [ ]:
```
#correlation
```

In [ ]:
```
dft.corr()
```

Out[11]:

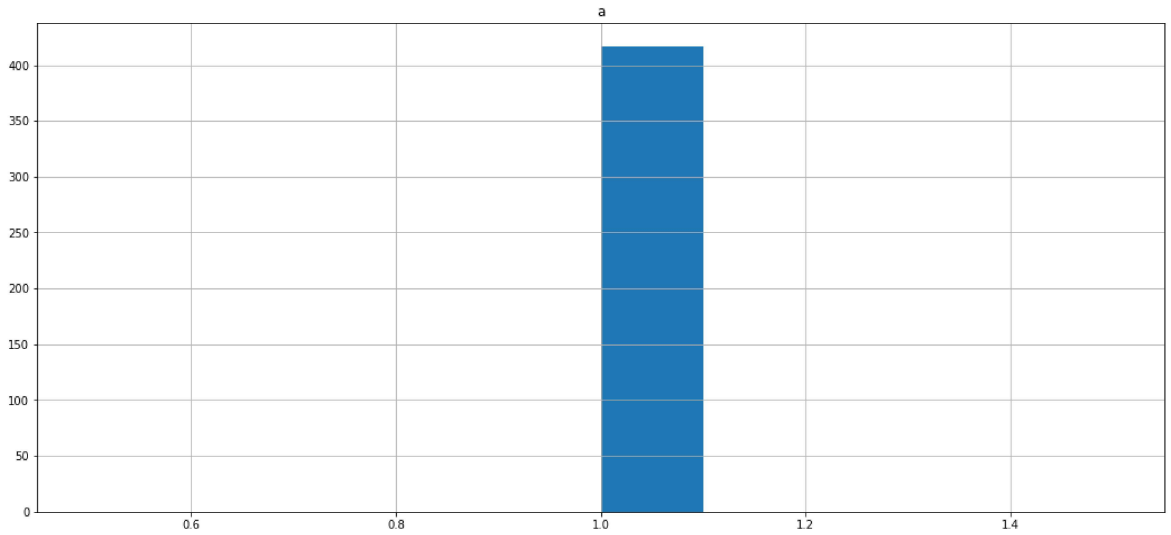
| | previous_day_at_least_one | previous_day_fully_vaccinate |
|---|---|---|
| **previous_day_at_least_one** | 1.000000 | -0.03609 |
| **previous_day_fully_vaccinated** | -0.036094 | 1.00000 |
| **previous_day_3doses** | 0.877246 | 0.01329 |
| **previous_day_total_doses_administered** | 0.376057 | 0.56899 |

In [ ]:
```
#normalize the data
```

In [ ]:
```
from sklearn import preprocessing
```

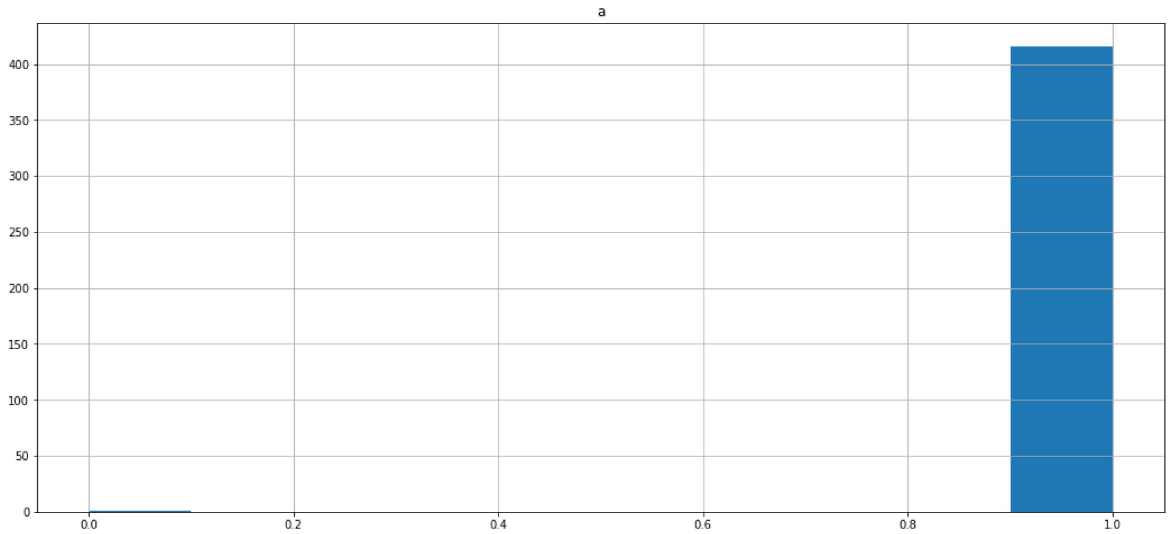
In [ ]:
```
dft1=dft[['previous_day_at_least_one']]
dft1=dft1.dropna()
d=preprocessing.normalize(dft1)
norm_df=pd.DataFrame(d,columns=['a'])
norm_df.hist(figsize=(18,8))
```

Out[47]:
```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x00000218737C8E20
>]],
      dtype=object)
```



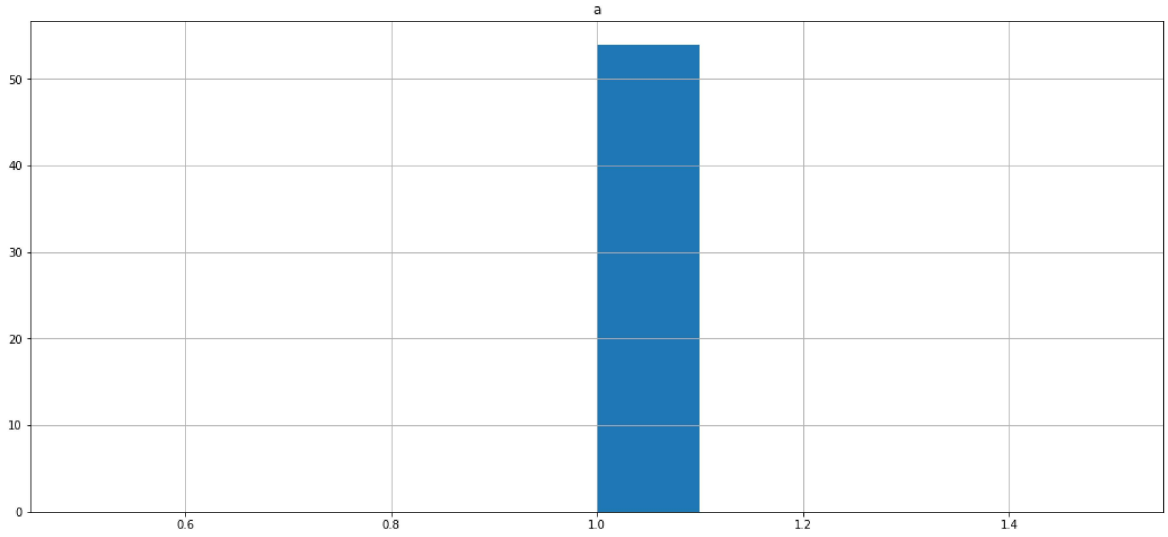

In [ ]:
```
dft2=dft[['previous_day_fully_vaccinated']]
dft2=dft2.dropna()
d=preprocessing.normalize(dft2)
norm_df=pd.DataFrame(d,columns=['a'])
norm_df.hist(figsize=(18,8))
```

Out[48]:
```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x0000021873838A90
>]],
      dtype=object)
```



In [ ]:
```
dft3=dft[['previous_day_3doses']]
dft3=dft3.dropna()
d=preprocessing.normalize(dft3)
norm_df=pd.DataFrame(d,columns=['a'])
norm_df.hist(figsize=(18,8))
```

Out[49]:
```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x000002187383DD90
>]],
      dtype=object)
```




In [ ]:
```
dft4=dft[['previous_day_total_doses_administered']]
dft4=dft4.dropna()
d=preprocessing.normalize(dft4)
norm_df=pd.DataFrame(d,columns=['a'])
norm_df.hist(figsize=(18,8))
```

Out[50]:
```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x0000021873AD3100
>]],
      dtype=object)
```