In [1]:
```
import pandas as pd
import numpy as np
import statistics
from statistics import mean
import matplotlib.pyplot as plt
```

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

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)
```

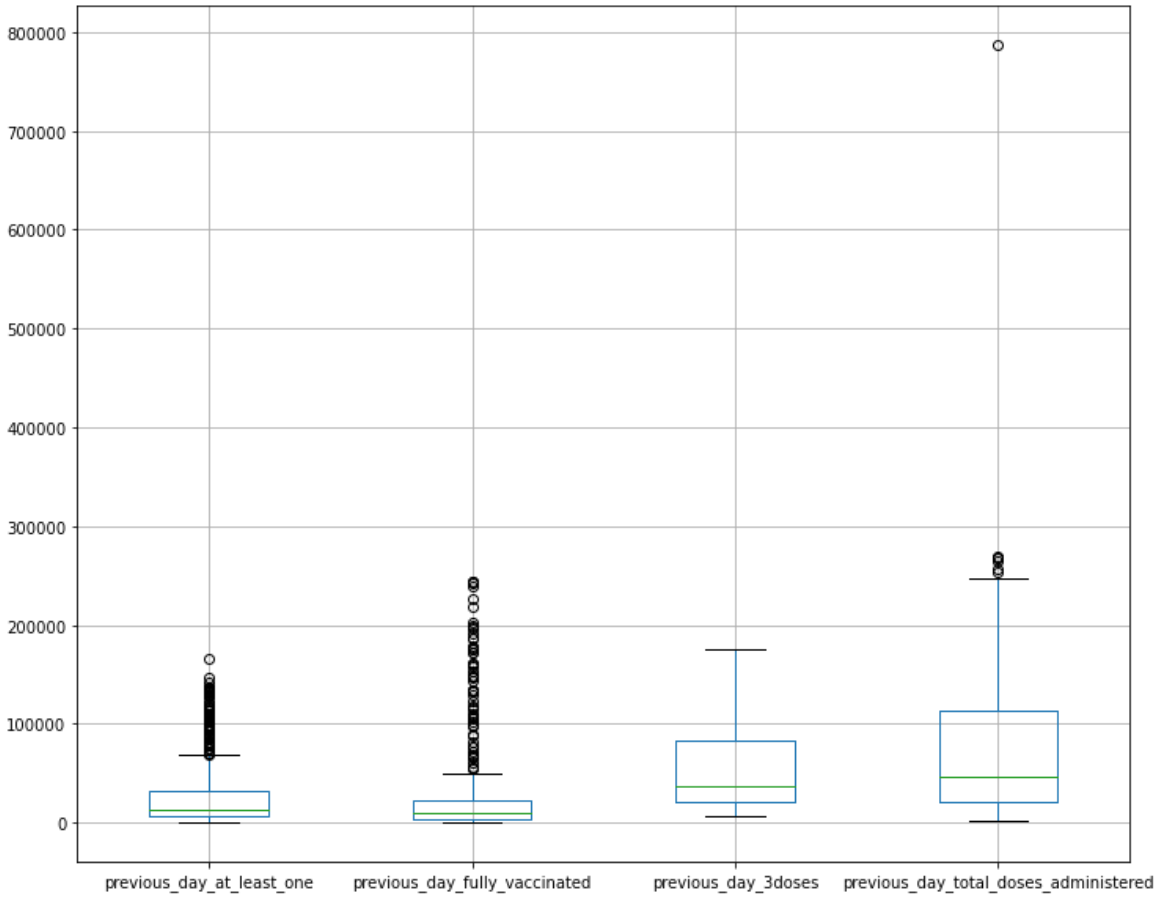
Out[8]:

	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))
```

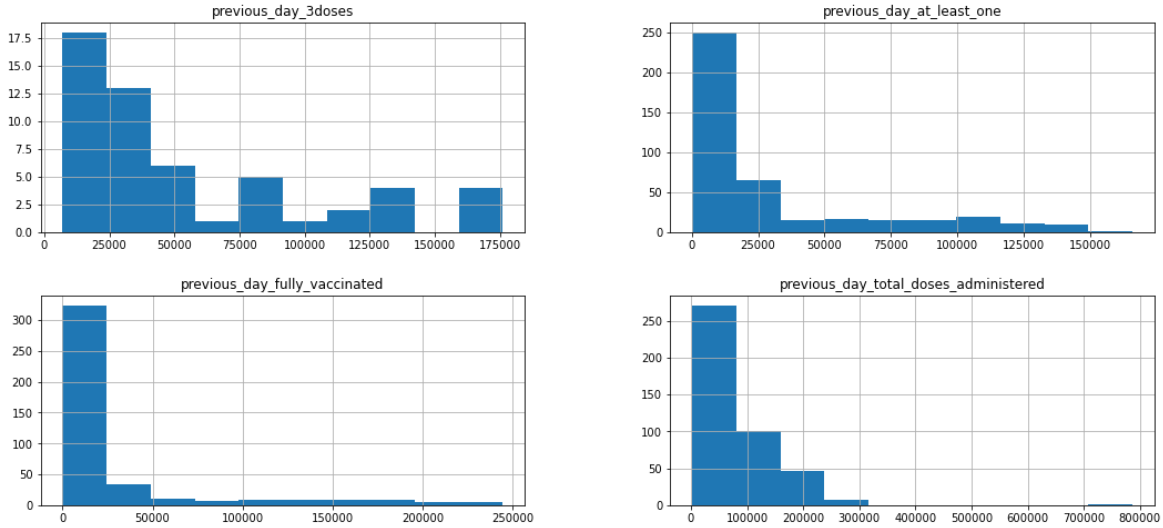
Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x2187092c340>



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]:

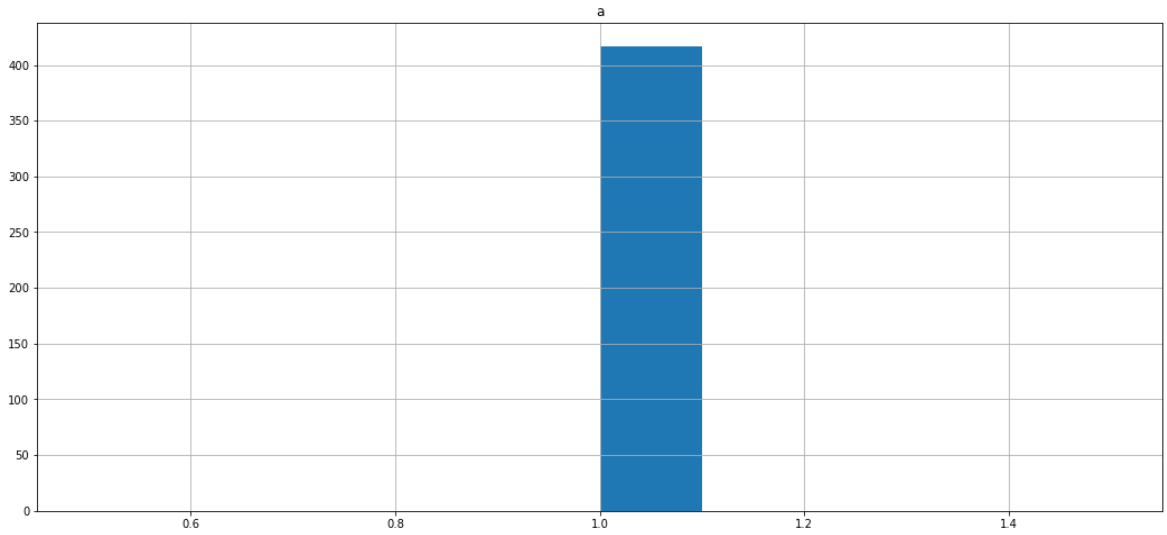
	previous_day_at_least_one	previous_day_fully_vaccinate
previous_day_at_least_one	1.000000	-0.03609
previous_day_fully_vaccinated	-0.036094	1.00000
previous_day_3doses	0.877246	0.01329
previous_day_total_doses_administered	0.376057	0.56899

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

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

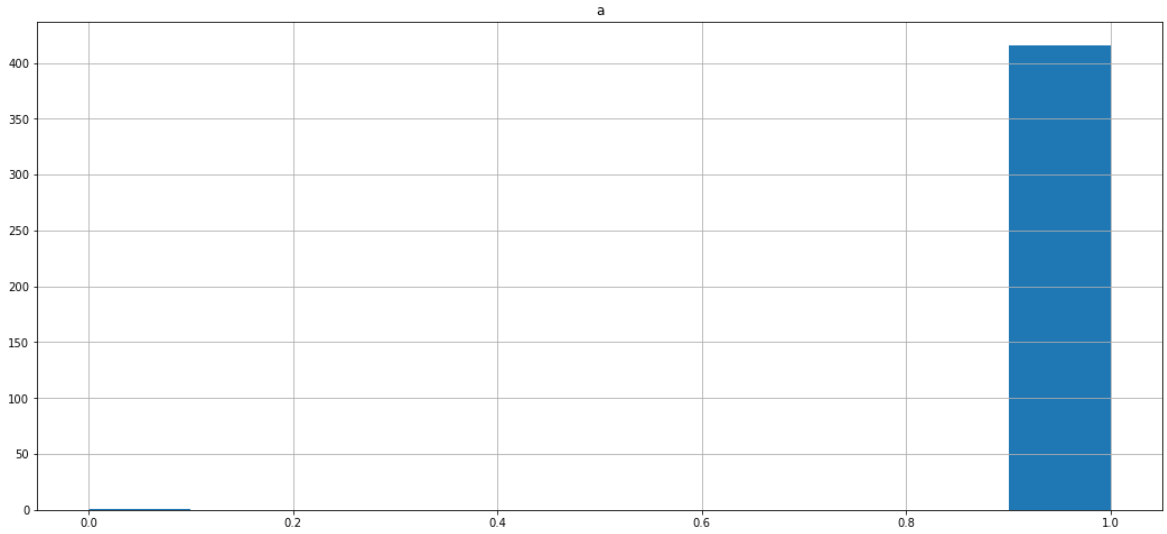
In [ ]:
```
dft1=dft[['previous_day_at_least_one']]
dft1=dft1.dropna()
d=preprocessing.normalize(dft1)
norm_df=pd.DataFrame(d,columns=['a'])
norm_df.hist(figsize=(18,8))
```

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



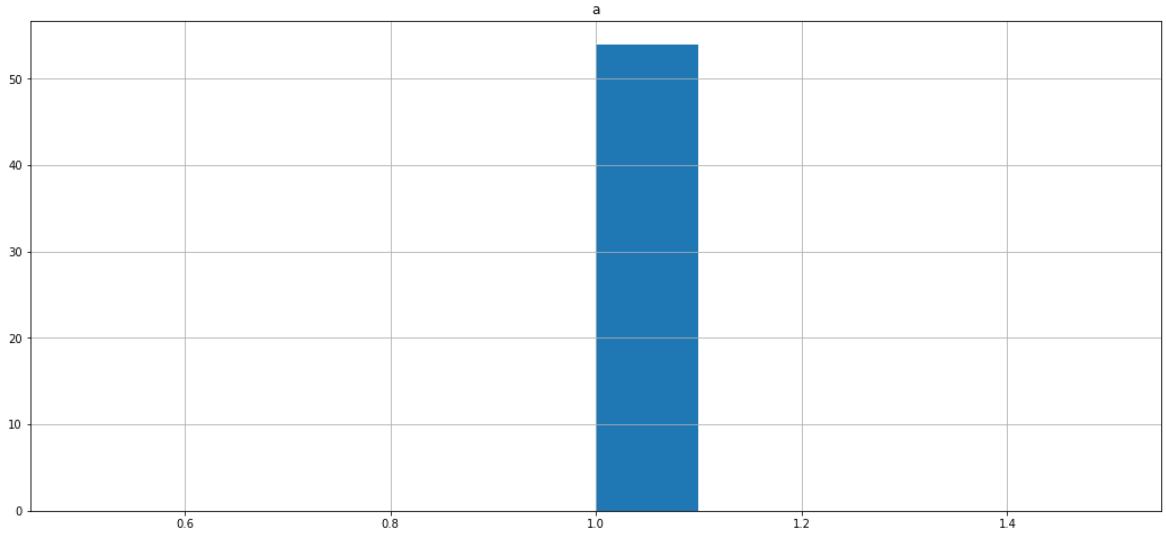
In [ ]:
```
dft2=dft[['previous_day_fully_vaccinated']]
dft2=dft2.dropna()
d=preprocessing.normalize(dft2)
norm_df=pd.DataFrame(d,columns=['a'])
norm_df.hist(figsize=(18,8))
```

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



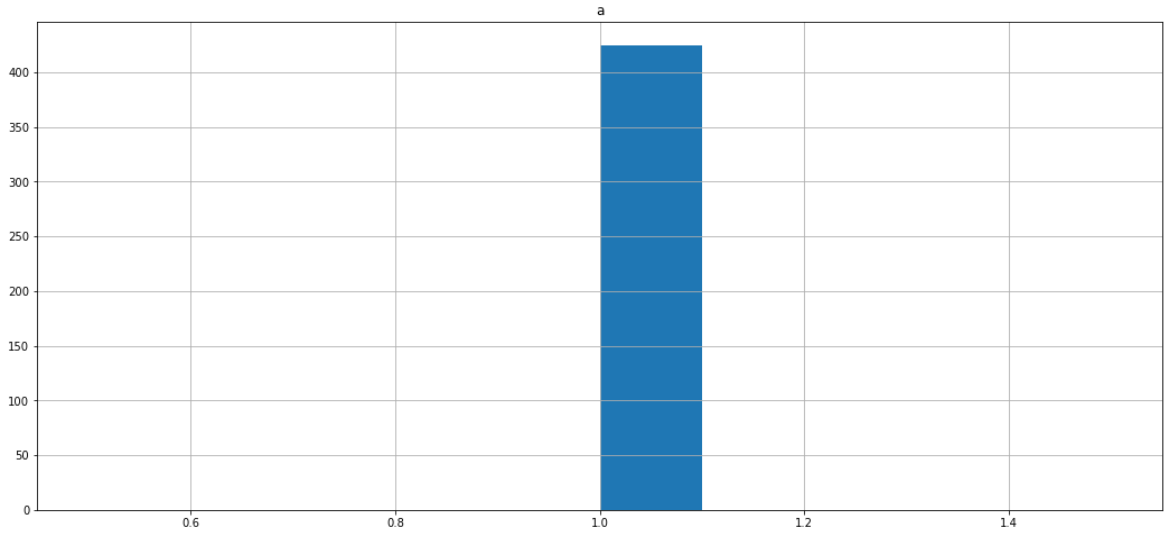
In [ ]:
```
dft3=dft[['previous_day_3doses']]
dft3=dft3.dropna()
d=preprocessing.normalize(dft3)
norm_df=pd.DataFrame(d,columns=['a'])
norm_df.hist(figsize=(18,8))
```

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)
```



In [2]:
```
df1=pd.read_excel('C:\\Users\\eric.park\\Downloads\\vaccines_by_age.xlsx')
```

In [4]:
```
dft_age=df1[['Date','At least one dose_cumulative','Second_dose_cumulative','
```

In [5]:
```
dft_age.isna().sum()
```

Out[5]:
```
Date                                0
At least one dose_cumulative        0
Second_dose_cumulative            897
third_dose_cumulative            4305
dtype: int64
```

In [7]:
```
dft_age.dtypes
```

Out[7]:
```
Date                            datetime64[ns]
At least one dose_cumulative             int64
Second_dose_cumulative                 float64
third_dose_cumulative                  float64
dtype: object
```

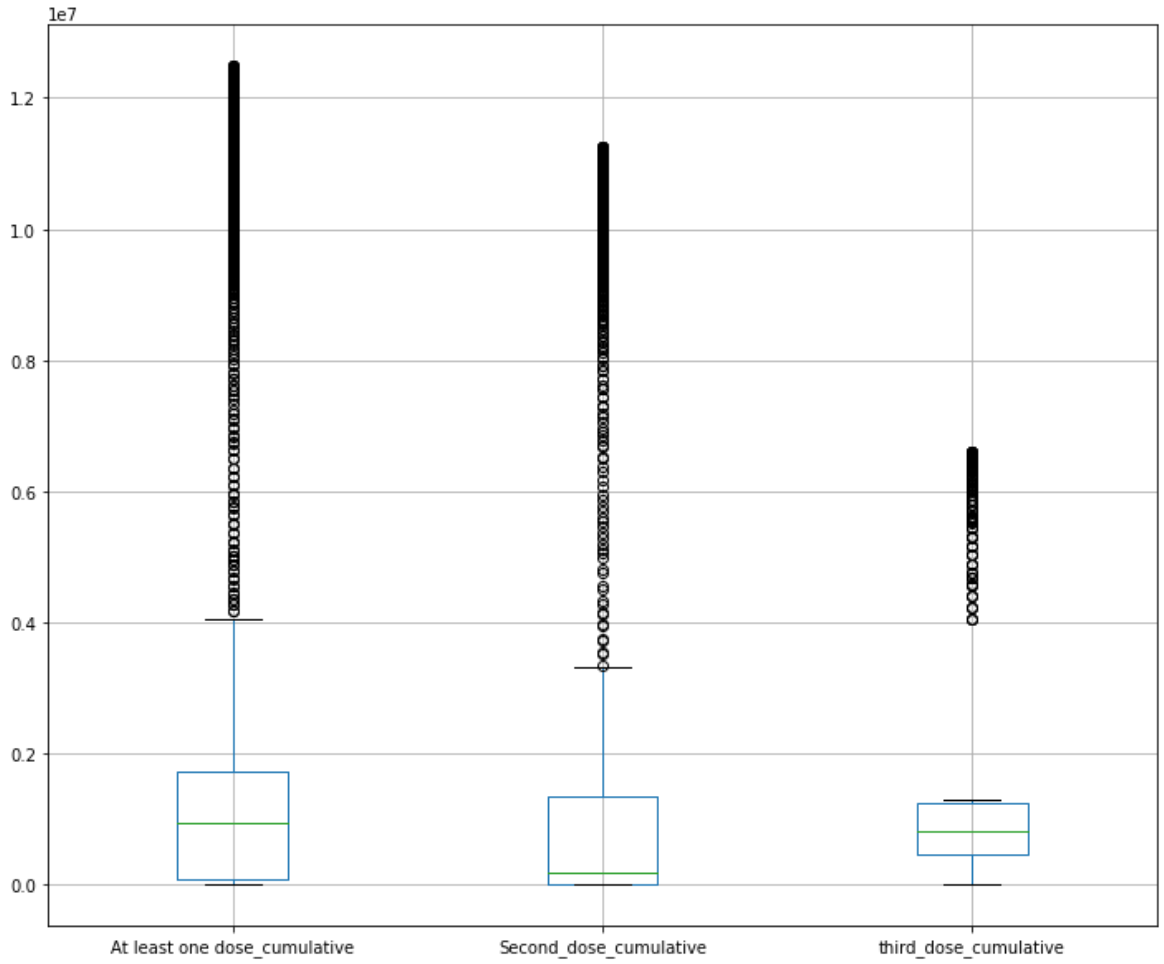
In [8]:
```
dft_age.describe().applymap('{:,.2f}'.format)
```

Out[8]:

	At least one dose_cumulative	Second_dose_cumulative	third_dose_cumulative
count	4,747.00	3,850.00	442.00
mean	2,118,820.53	1,231,565.42	1,755,697.18
std	3,387,902.81	2,550,252.97	2,225,109.43
min	0.00	0.00	0.00
25%	80,860.50	13,507.50	468,789.75
50%	943,317.00	193,616.50	816,695.50
75%	1,720,422.00	1,343,163.75	1,254,583.75
max	12,492,618.00	11,257,423.00	6,604,028.00

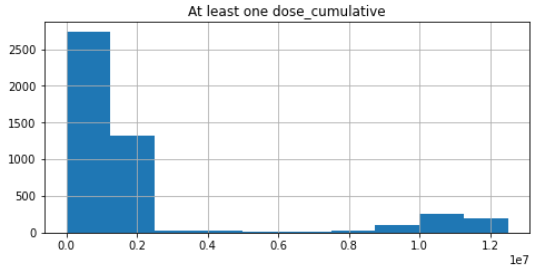
In [10]:
```
dft_age.boxplot(figsize=(12,10))
```

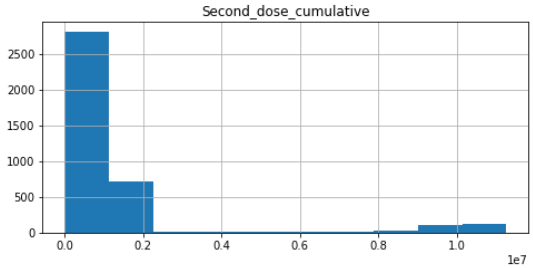
Out[10]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x20181b29f70>
```

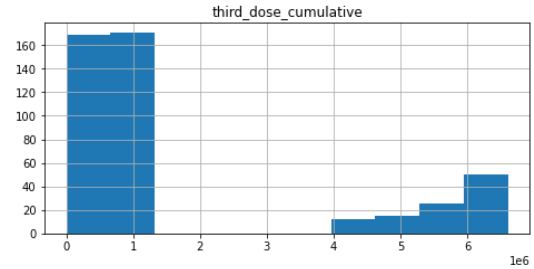


In [11]:
```
dft_age.hist(figsize=(18,8))
```

Out[11]:
```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x0000020182571A00
>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x0000020182599D60
>],
       [<matplotlib.axes._subplots.AxesSubplot object at 0x000002018276B0D0
>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x00000201827964C0
>]],
      dtype=object)
```







In [12]:
```
dft_age.corr()
```

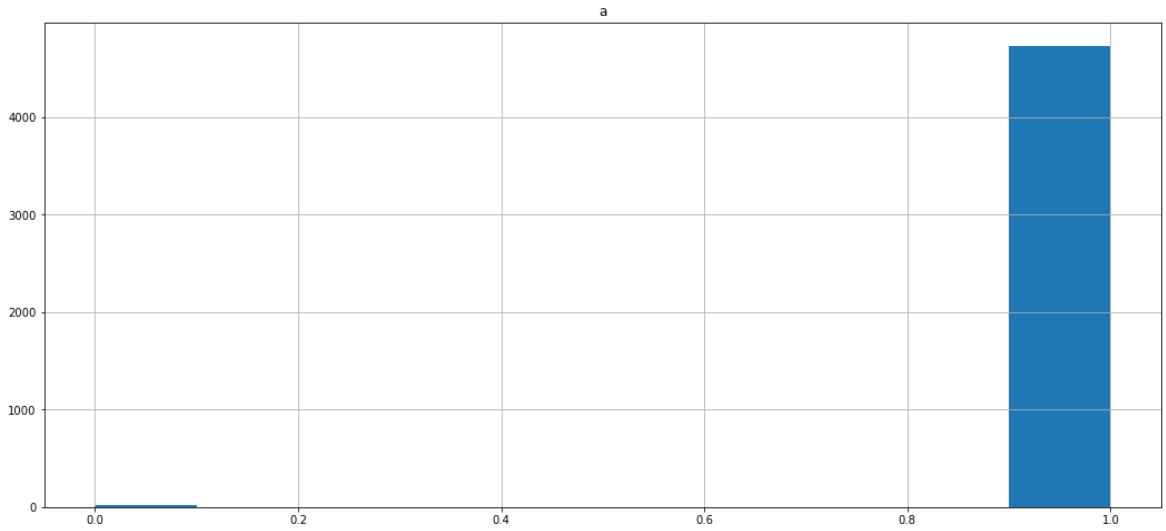
Out[12]:

	At least one dose_cumulative	Second_dose_cumulative	third_dose_cumulative
At least one dose_cumulative	1.000000	0.911421	0.980579
Second_dose_cumulative	0.911421	1.000000	NaN
third_dose_cumulative	0.980579	NaN	1.000000

In [13]:
```
from sklearn import preprocessing
```

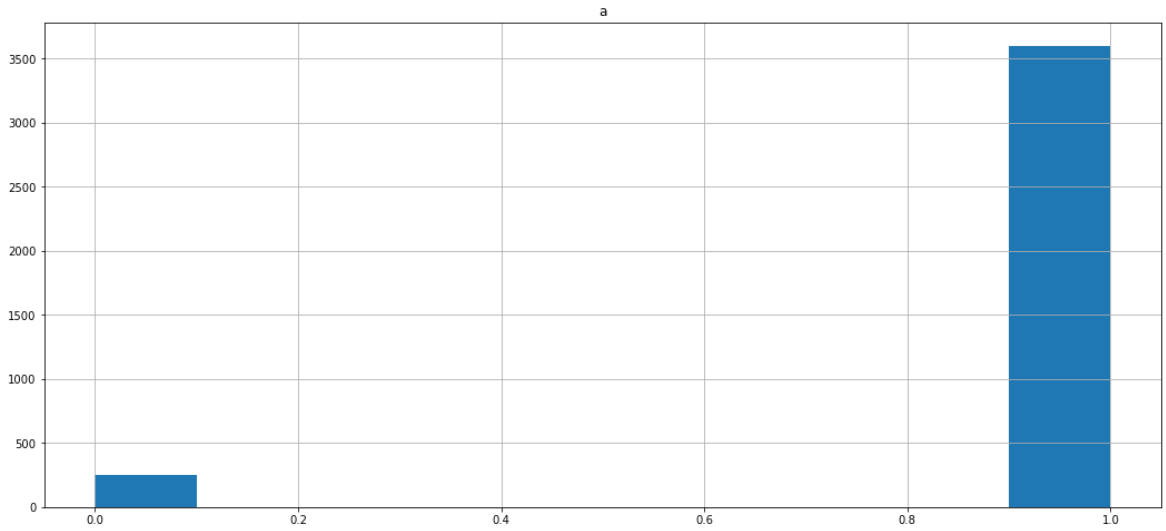
In [14]:
```
dft_age=dft1[['At least one dose_cumulative']]
dft1_age=dft_age.dropna()
d=preprocessing.normalize(dft1_age)
norm_df=pd.DataFrame(d,columns=['a'])
norm_df.hist(figsize=(18,8))
```

Out[14]:
```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x0000020181A2DA30
>]],
      dtype=object)
```



In [15]:
```
dft_age=dft1[['Second_dose_cumulative']]
dft2_age=dft_age.dropna()
d=preprocessing.normalize(dft2_age)
norm_df=pd.DataFrame(d,columns=['a'])
norm_df.hist(figsize=(18,8))
```

Out[15]:
```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x00000201819DED90
>]],
      dtype=object)
```



In [16]:
```
dft_age=dft1[['third_dose_cumulative']]
dft3_age=dft_age.dropna()
d=preprocessing.normalize(dft3_age)
norm_df=pd.DataFrame(d,columns=['a'])
norm_df.hist(figsize=(18,8))
```

Out[16]:
```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x00000201819C2F10
>]],
      dtype=object)
```