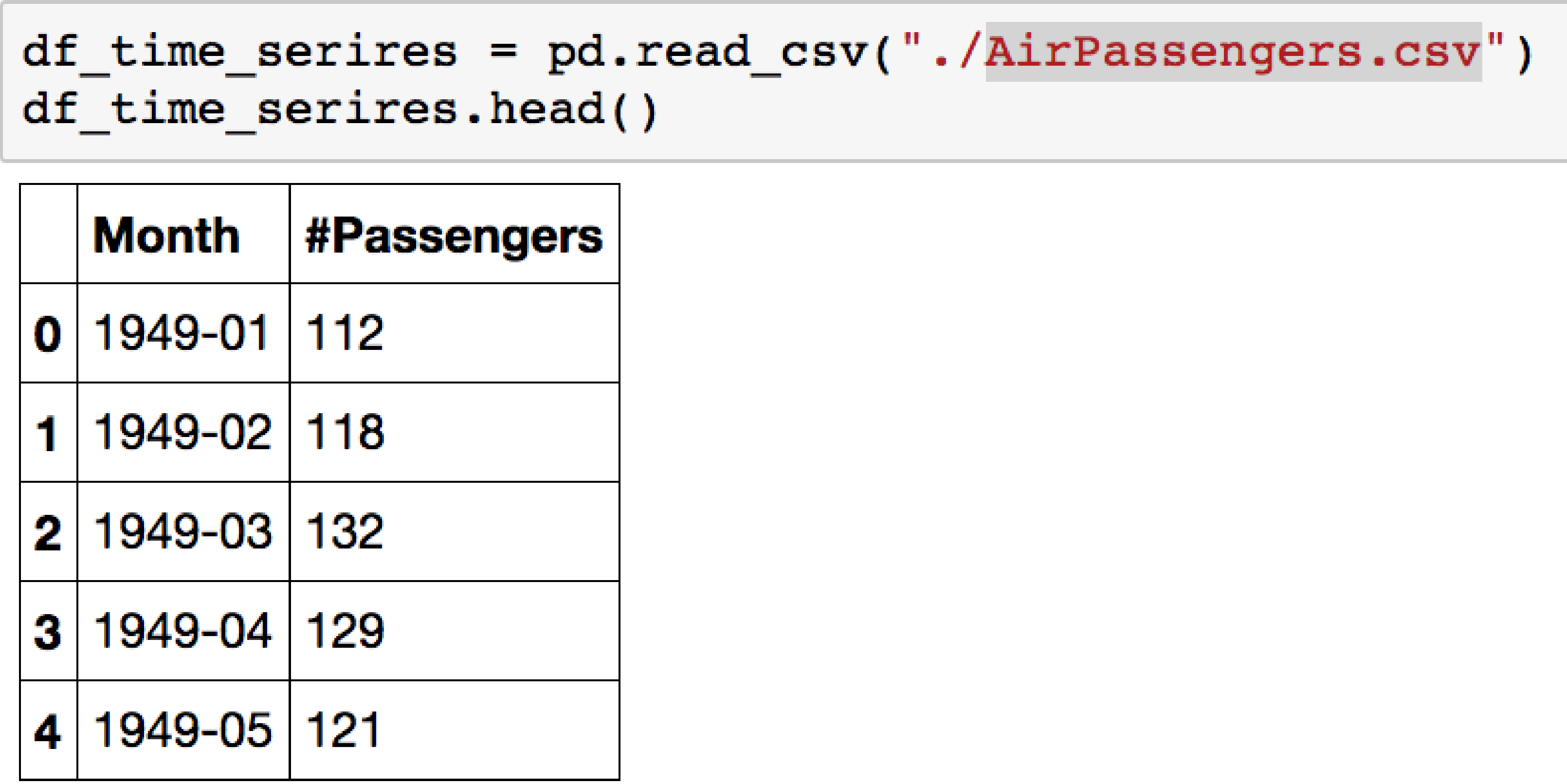
Case study I

Data

- 1940~1960까지 매달 비행기를 이용객 수를 정리한 데이터
- 시계열 toy data

https://www.kaggle.com/rakannimer/air-passengers

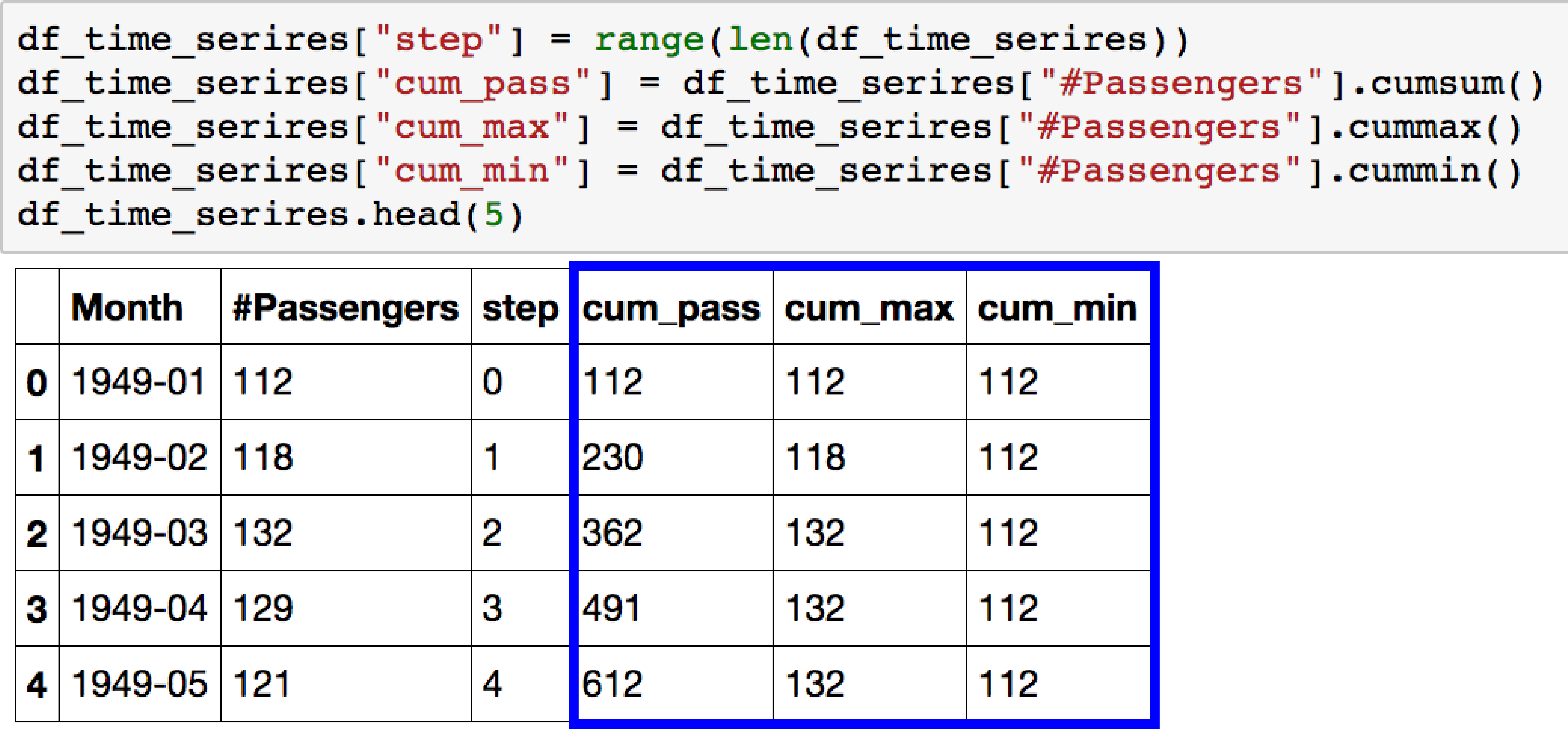


```
df_time_serires = pd.read_csv("./AirPassengers.csv")
df_time_serires.head()
```

	Month	#Passengers
0	1949-01	112
1	1949-02	118
2	1949-03	132
3	1949-04	129
4	1949-05	121

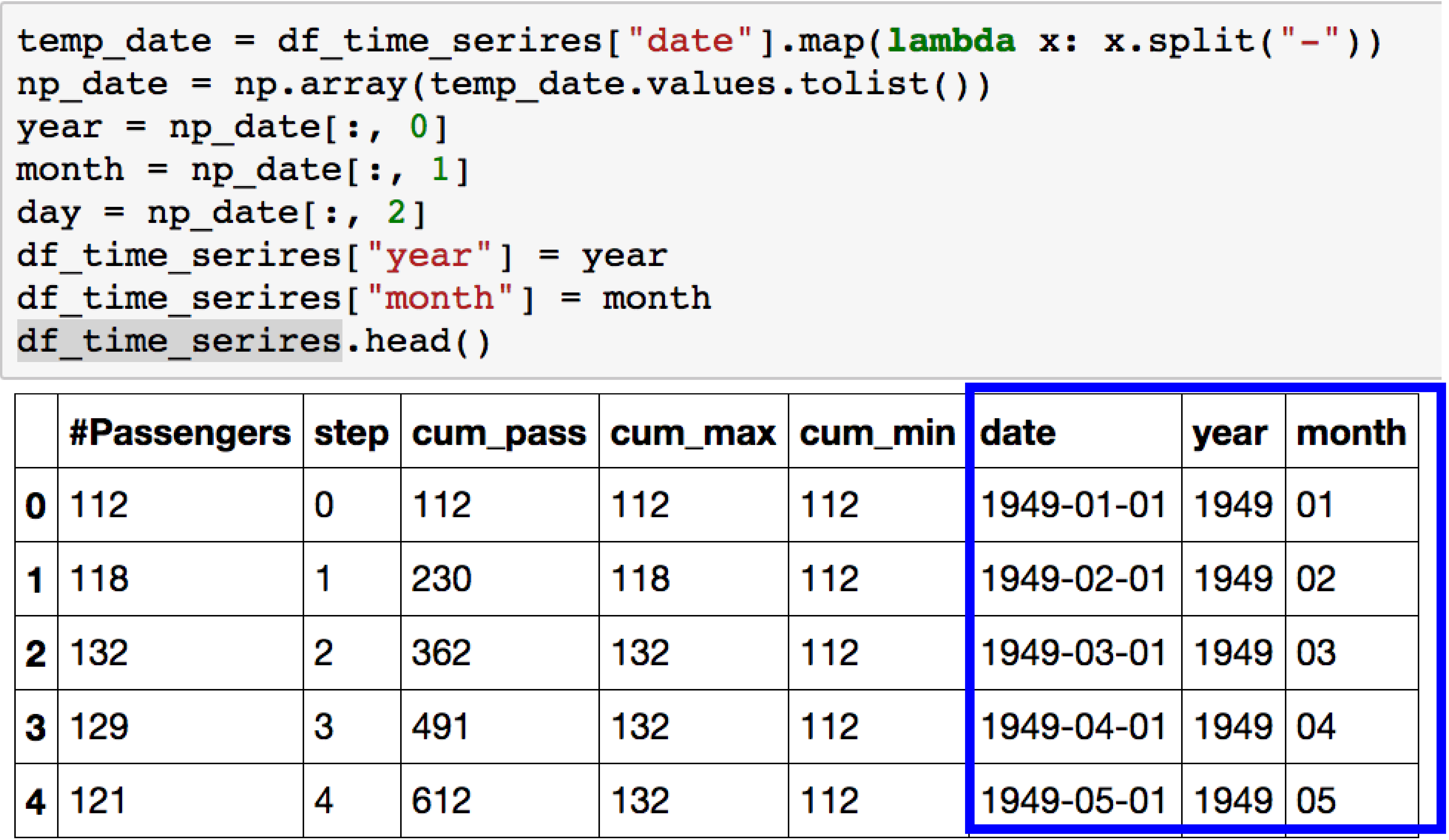
Data

- 월별, 년도별 승객은 얼마나 될까?
- 이전달과 이번달의 승객의 차이는? 상승율은?
- 누적 승객수, 최고, 최소 승객수?



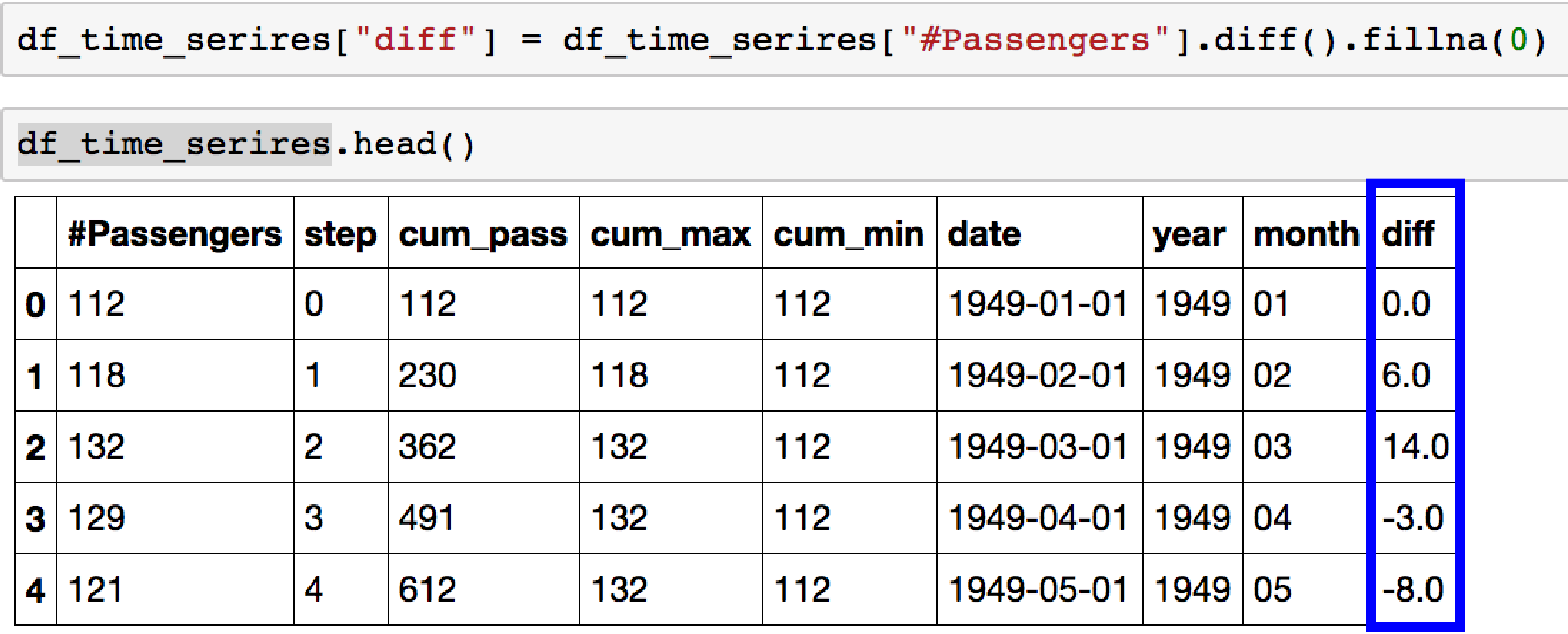
```
df_time_serires["step"] = range(len(df_time_serires))
df_time_serires["cum_pass"] = df_time_serires["#Passengers"].cumsum()
df_time_serires["cum_max"] = df_time_serires["#Passengers"].cummax()
df_time_serires["cum_min"] = df_time_serires["#Passengers"].cummin()
df_time_serires.head(5)
```

	Month	#Passengers	step	cum_pass	cum_max	cum_min
0	1949-01	112	0	112	112	112
1	1949-02	118	1	230	118	112
2	1949-03	132	2	362	132	112
3	1949-04	129	3	491	132	112
4	1949-05	121	4	612	132	112



```
temp_date = df_time_serires["date"].map(lambda x: x.split("-"))
np_date = np.array(temp_date.values.tolist())
year = np_date[:, 0]
month = np_date[:, 1]
day = np_date[:, 2]
df_time_serires["year"] = year
df_time_serires["month"] = month
df_time_serires.head()
```

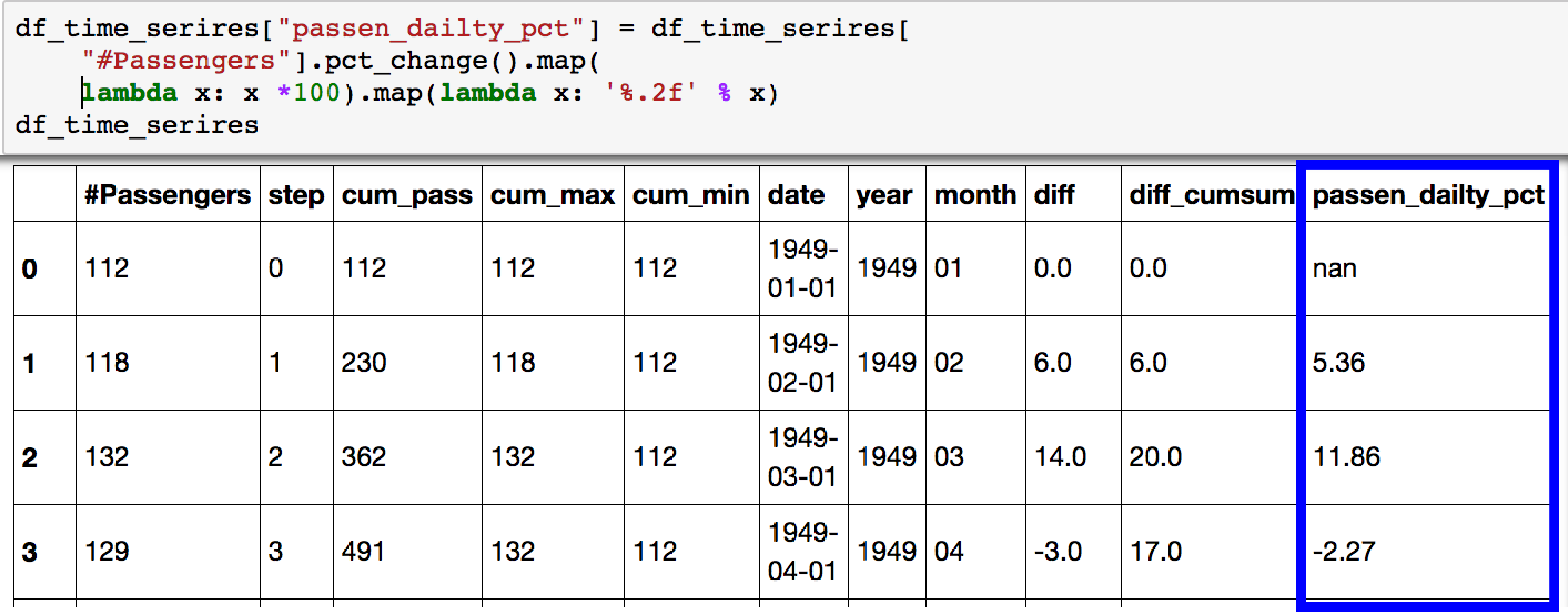
	#Passengers	step	cum_pass	cum_max	cum_min	date	year	month
0	112	0	112	112	112	1949-01-01	1949	01
1	118	1	230	118	112	1949-02-01	1949	02
2	132	2	362	132	112	1949-03-01	1949	03
3	129	3	491	132	112	1949-04-01	1949	04
4	121	4	612	132	112	1949-05-01	1949	05



```
df_time_serires["diff"] = df_time_serires["#Passengers"].diff().fillna(0)
```

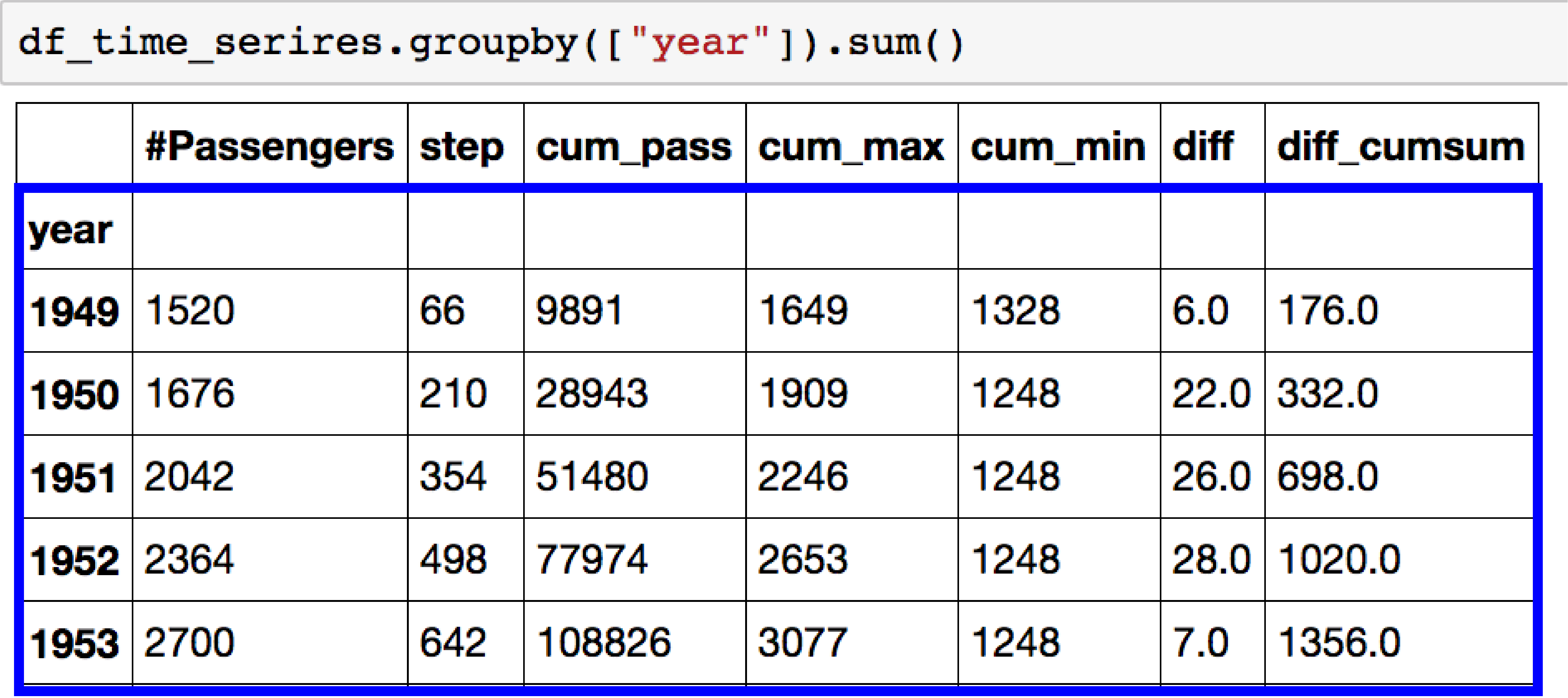
```
df_time_serires.head()
```

	#Passengers	step	cum_pass	cum_max	cum_min	date	year	month	diff
0	112	0	112	112	112	1949-01-01	1949	01	0.0
1	118	1	230	118	112	1949-02-01	1949	02	6.0
2	132	2	362	132	112	1949-03-01	1949	03	14.0
3	129	3	491	132	112	1949-04-01	1949	04	-3.0
4	121	4	612	132	112	1949-05-01	1949	05	-8.0



```
df_time_serires["passen_dailty_pct"] = df_time_serires[
    "#Passengers"].pct_change().map(
    lambda x: x *100).map(lambda x: '%.2f' % x)
df_time_serires
```

	#Passengers	step	cum_pass	cum_max	cum_min	date	year	month	diff	diff_cumsum	passen_dailty_pct
0	112	0	112	112	112	1949-01-01	1949	01	0.0	0.0	nan
1	118	1	230	118	112	1949-02-01	1949	02	6.0	6.0	5.36
2	132	2	362	132	112	1949-03-01	1949	03	14.0	20.0	11.86
3	129	3	491	132	112	1949-04-01	1949	04	-3.0	17.0	-2.27



```
df_time_serires.groupby(["year"]).sum()
```

	#Passengers	step	cum_pass	cum_max	cum_min	diff	diff_cumsum
year							
1949	1520	66	9891	1649	1328	6.0	176.0
1950	1676	210	28943	1909	1248	22.0	332.0
1951	2042	354	51480	2246	1248	26.0	698.0
1952	2364	498	77974	2653	1248	28.0	1020.0
1953	2700	642	108826	3077	1248	7.0	1356.0