ECON 0150 | Economic Data Analysis

The economist's data analysis pipeline.

Part 1.7 | Transforming Data

Example 1.7 | Starbucks Location Hours

How many shops are open at once?

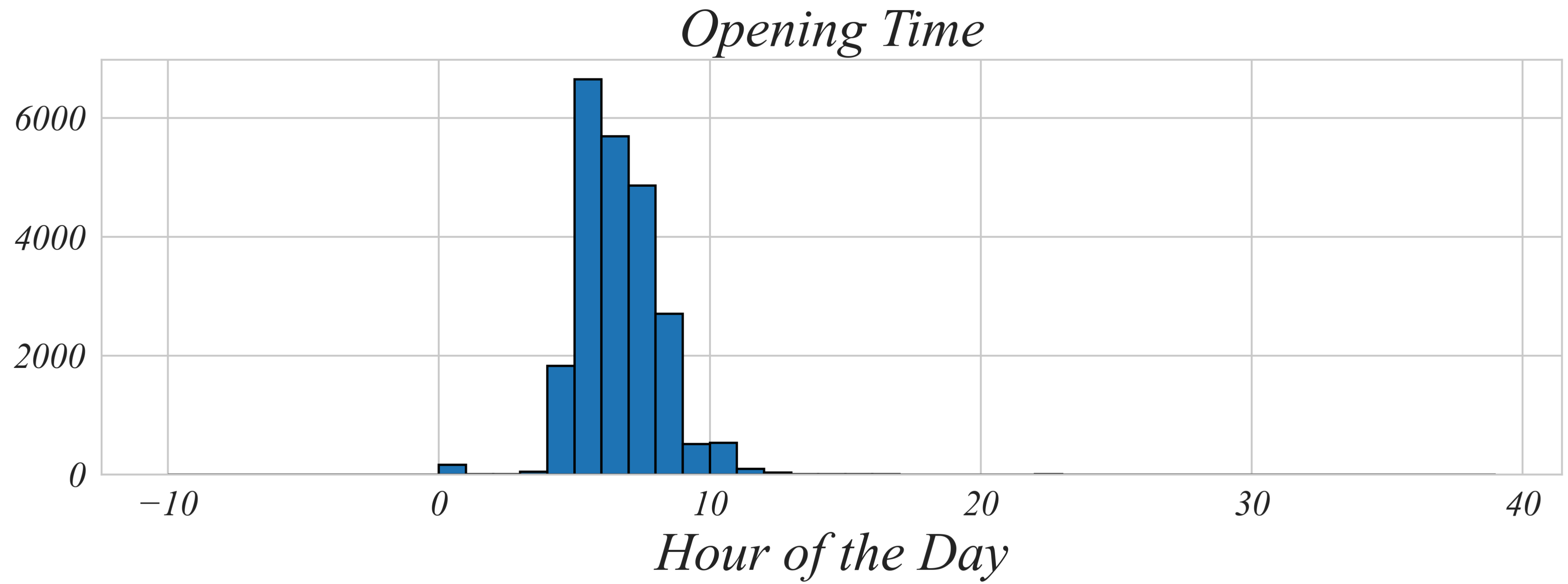
```python
# Import packages
import pandas as pd

# Load data
hours = pd.read_csv("Starbucks_Location_Hours.csv")
```

>as is common, it’s difficult to understand the raw data on its own

Location Hours

What times to shops open?

```
# Histogram of opening times
plt.hist(hours.open)
```

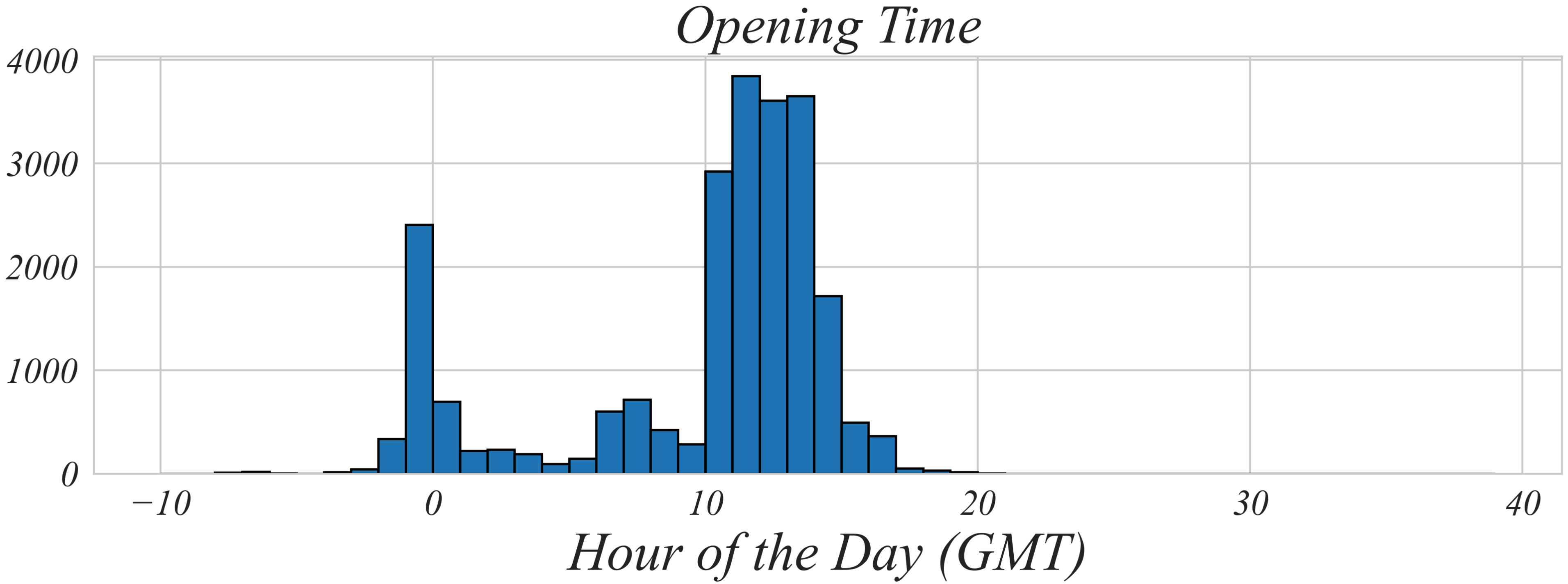


> but does this tell us how many shops are open at one time?

Location Hours

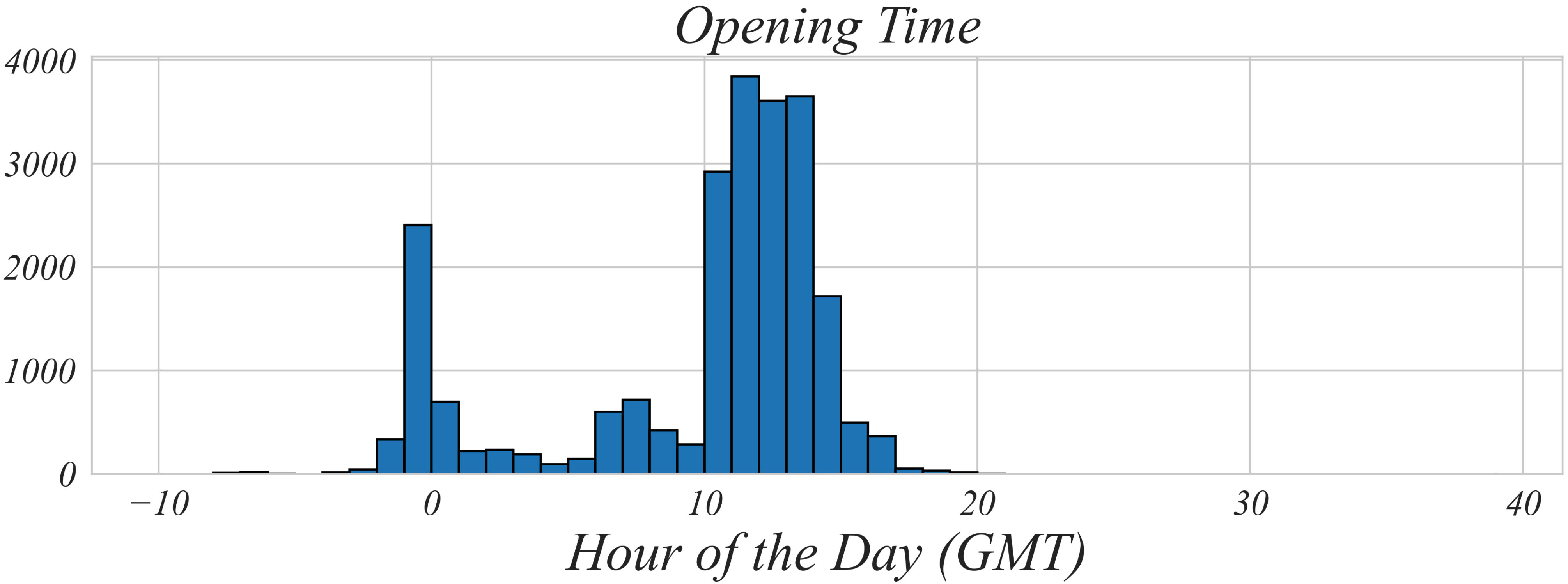
What times do shops open (GMT)?

```
# Normalize to GMT
hours['open_GMT'] = hours['open'] - hours['timezone']

# Histogram of opening times (GMT)
plt.hist(hours.open_GMT)
```



Location Hours

What times do shops open (GMT)?

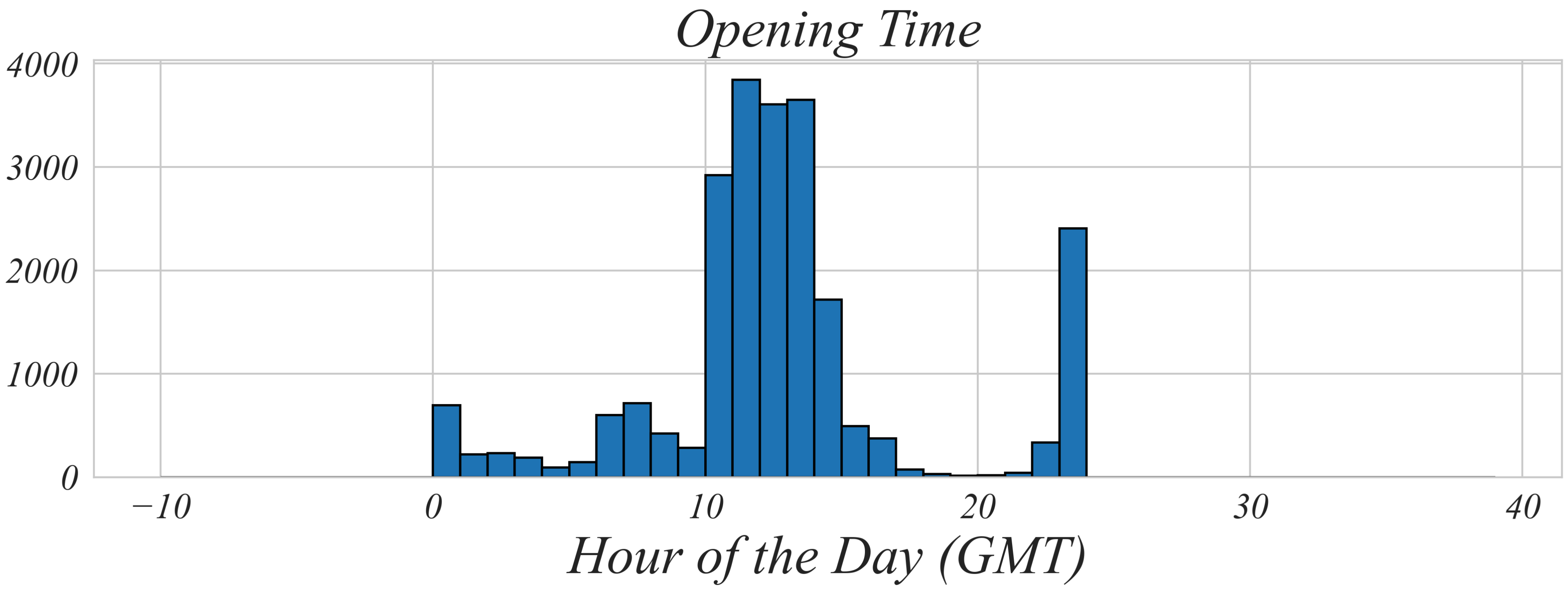


> what do the negative values mean?

Location Hours

Normalize the negative values 24 hours.

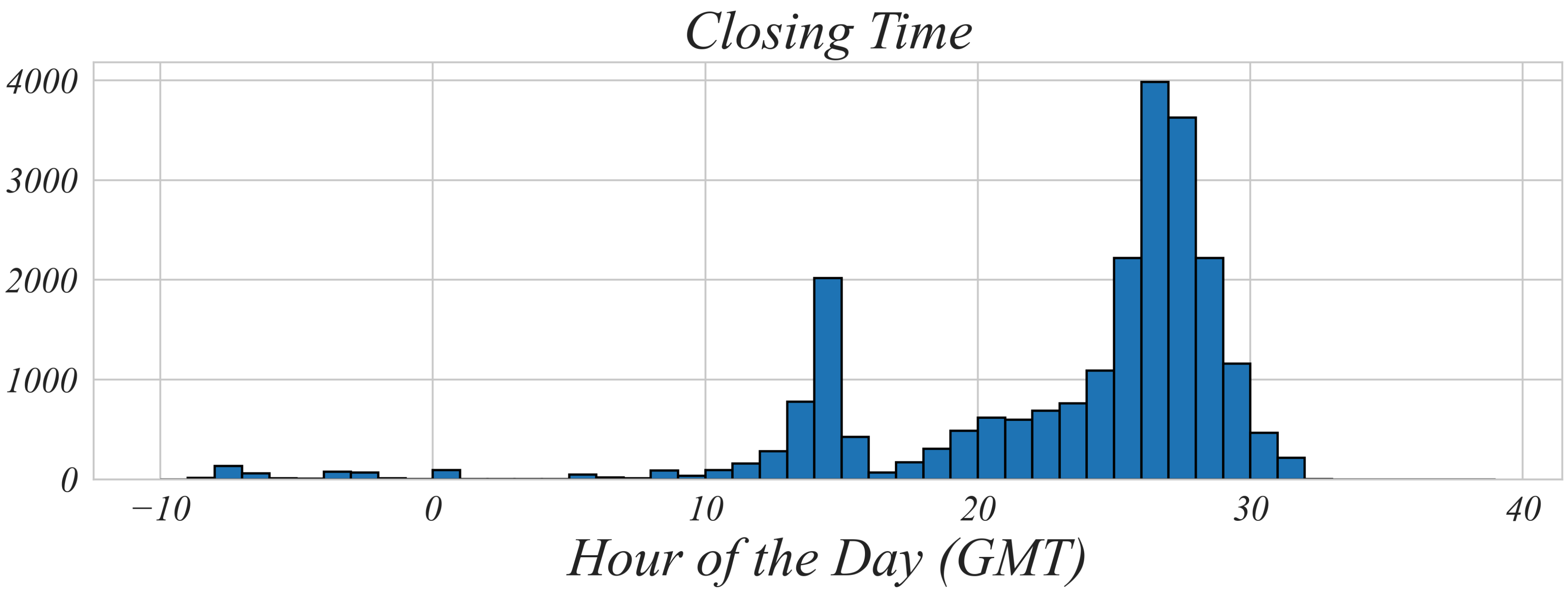
```python
# Normalize to 24 hours
hours['open_GMT'] = hours['open_GMT'].mod(24)
```



Location Hours

Closing times have the same issue.

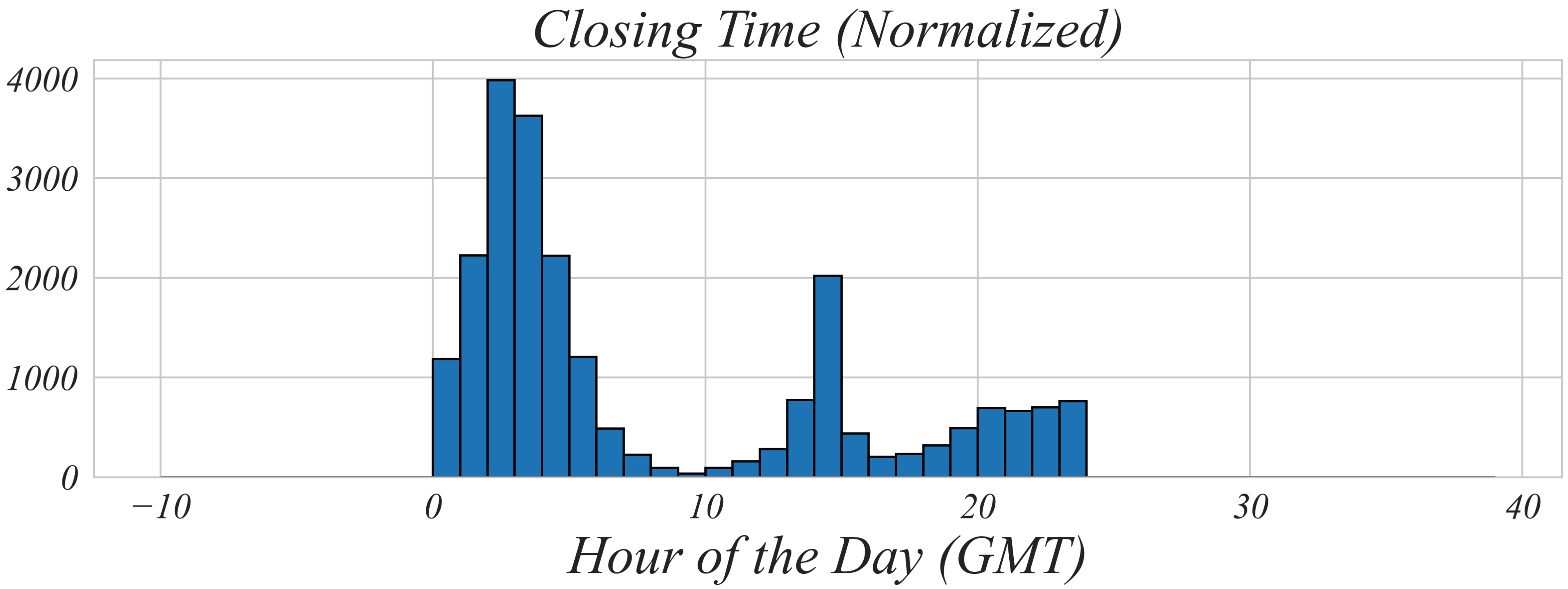
```python
# Normalize to GMT
hours['close_GMT'] = hours['close'] - hours['timezone']

# Histogram of opening times (GMT)
plt.hist(hours.close_GMT)
```



Location Hours

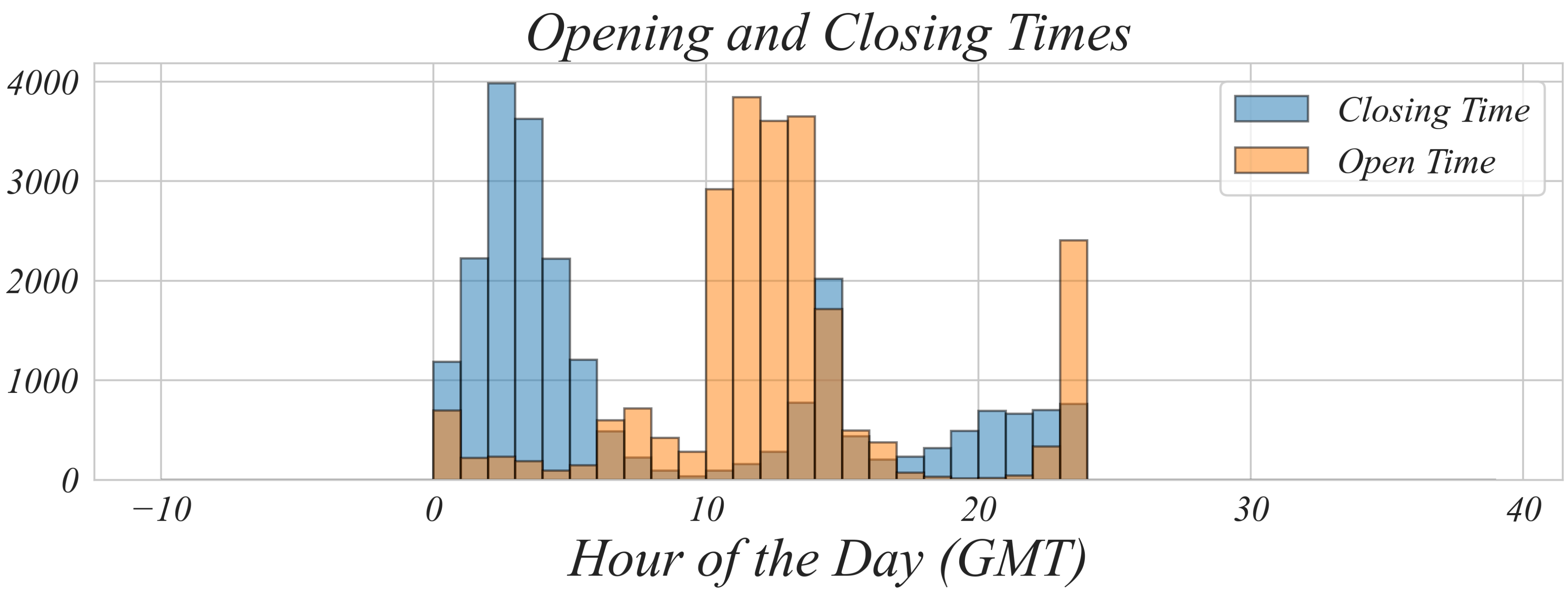
Normalize the positive values to 24 hours.

```python
# Normalize to 24 hours
hours['close_GMT'] = hours['close_GMT'].mod(24)
```



Location Hours

So, how many locations are open at each hour of the day?



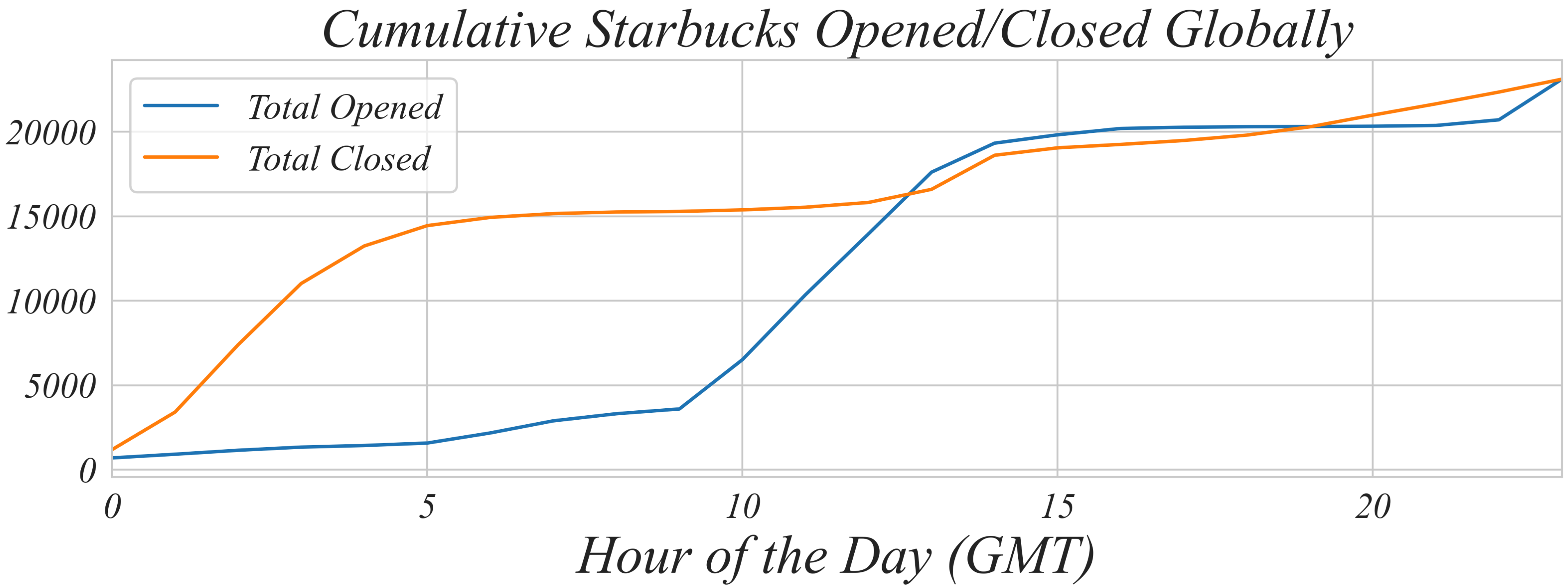
> this only tells us openings and closings at each hour, not total open

*> instead, lets sum up all the shops that have opened **that day***

Location Hours

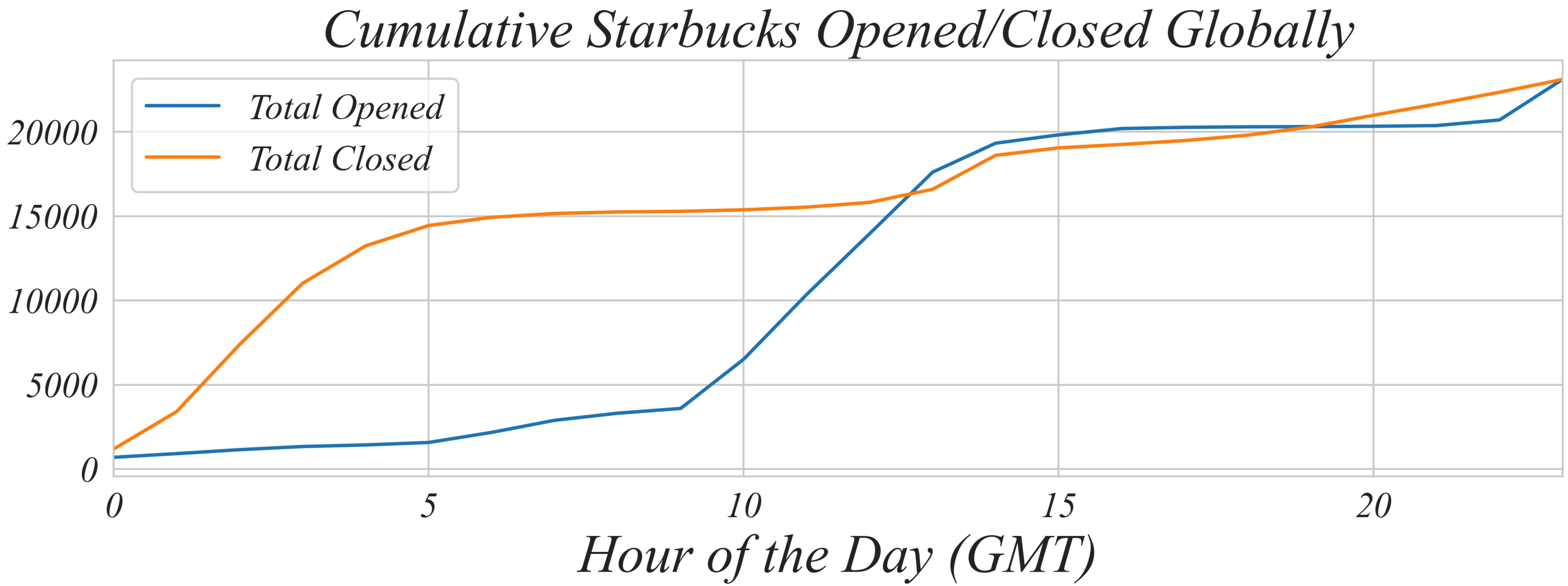
So, how many locations are open at each hour of the day?

```python
# Construct values by bin
opened_values = hours['open_GMT'].value_counts(bins=24, sort=False)

# Cumulative sum
total_opened = opened_values.cumsum()
```



Location Hours

So, how many locations are open at each hour of the day?

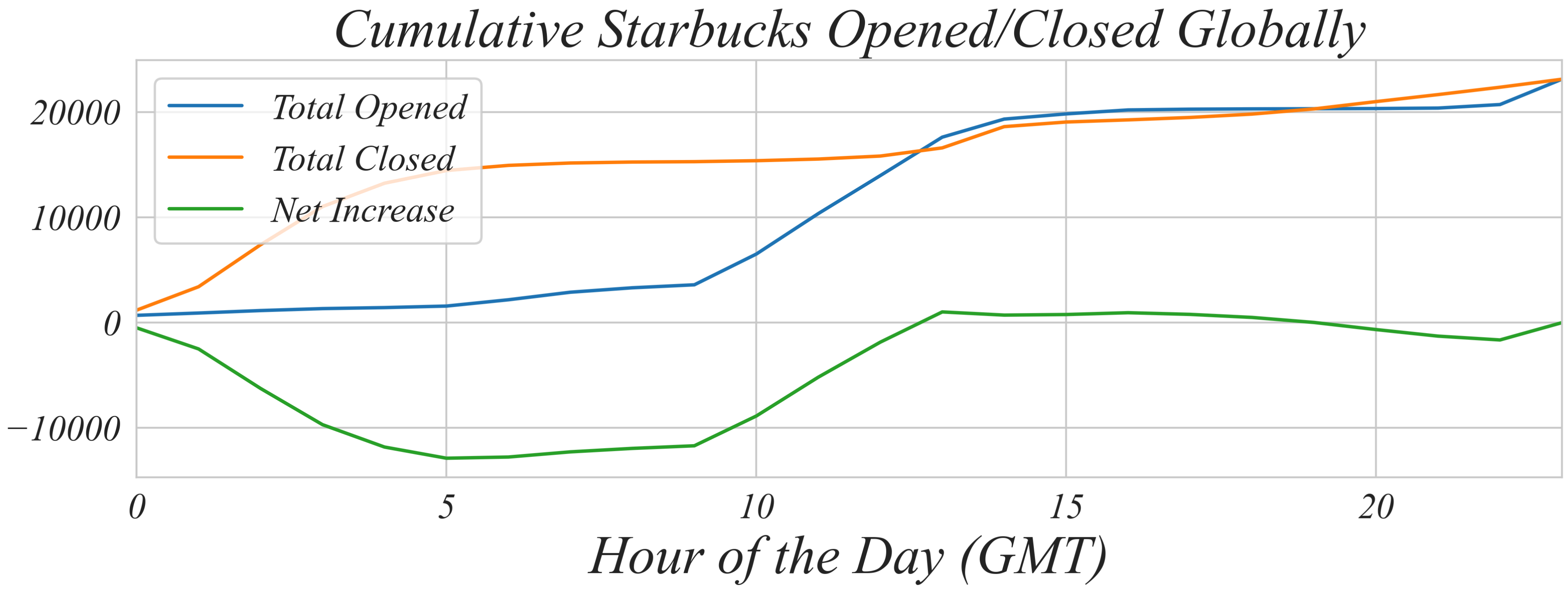


> from here, to find the total that have opened/closed, we take the difference

Location Hours

So, how many locations are open at each hour of the day?

```
# Take the difference
net_increase = total_opened - total_closed
```



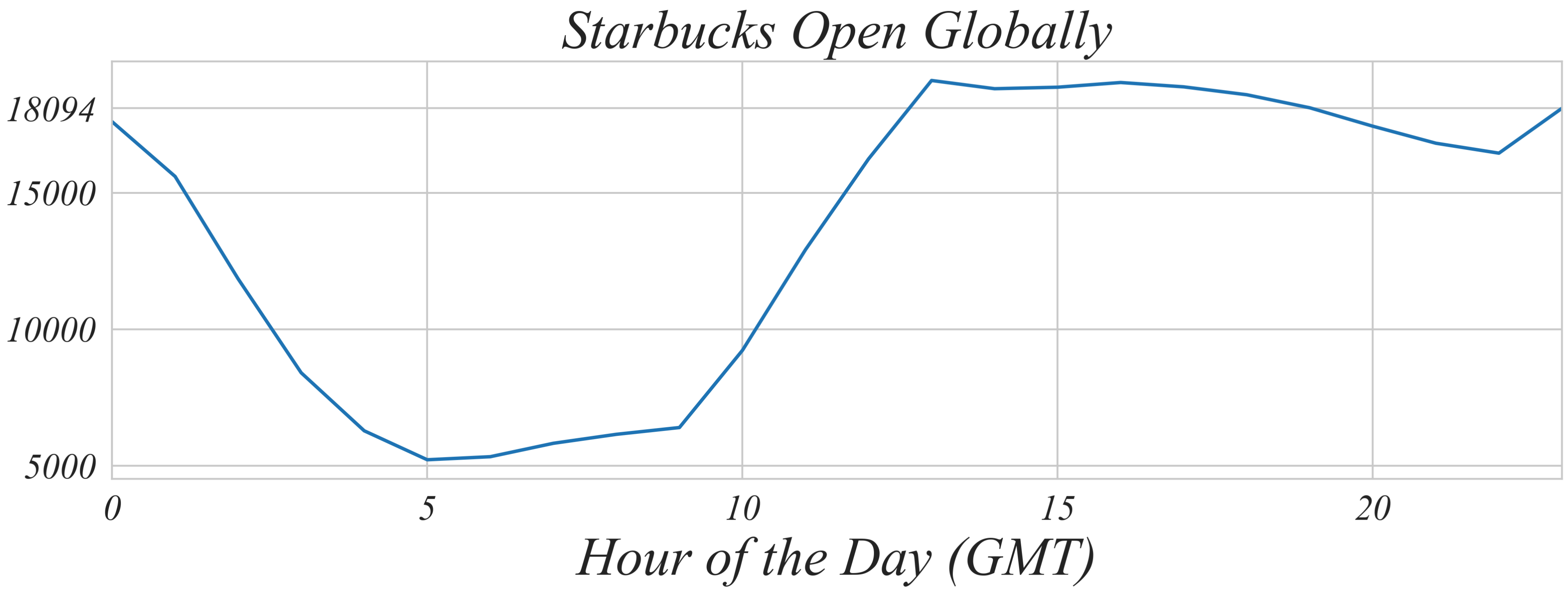
> *why is the green line negative?*

> *lets add the number open at midnight (GMT).*

Location Hours

So, how many locations are open at each hour of the day?

```python
# Add those open at midnight
count_open_after_close = len(hours[hours['open_GMT'] >= hours['close_GMT']])
cumulative_open = net_increase + count_open_after_close
```


Starbucks Open Globally
Hour of the Day (GMT)