

ECON 0150 | Economic Data Analysis

The economist's data analysis pipeline.

Part 1.4 | Numerical Variables by Category

Example 1.4 | Coffee Shop Transactions

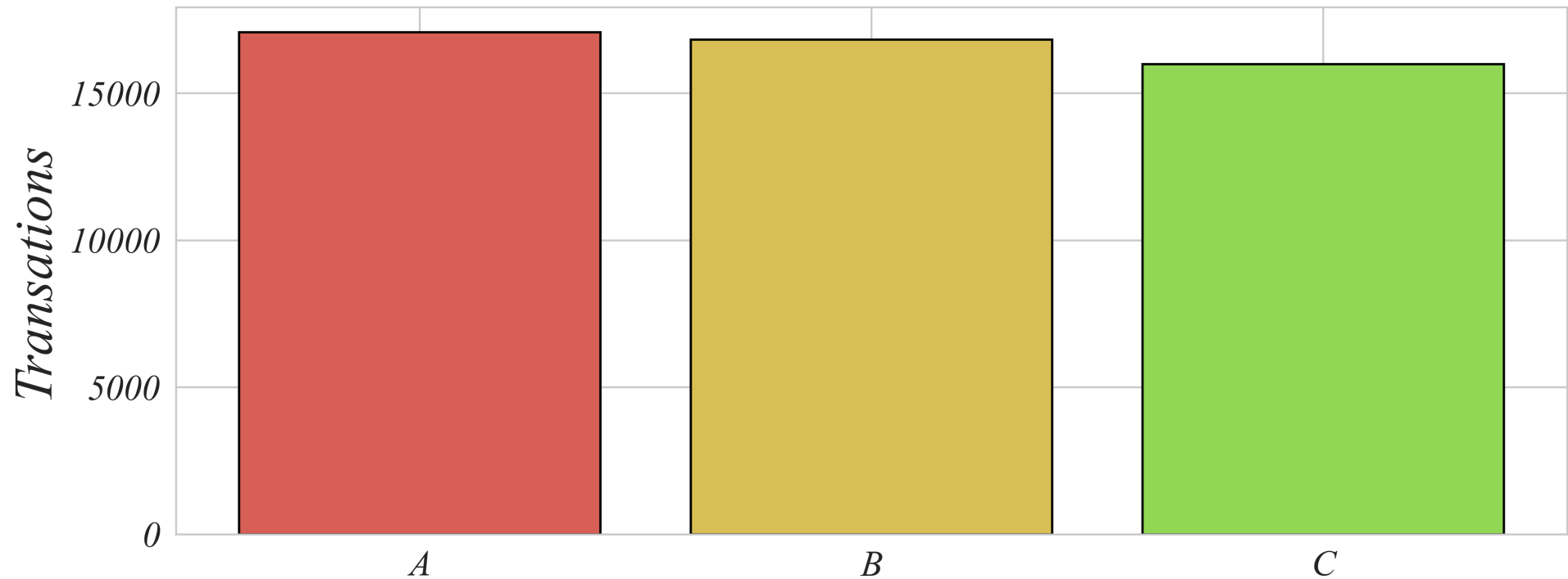
Use Coffee_Sales_Receipts.csv to help inform where to hire a barista.

```
1 # Load the data
2 sales = pd.read_csv(file_path + file_name)
```

Hiring a New Barista

Q. Which coffee shop is the busiest?

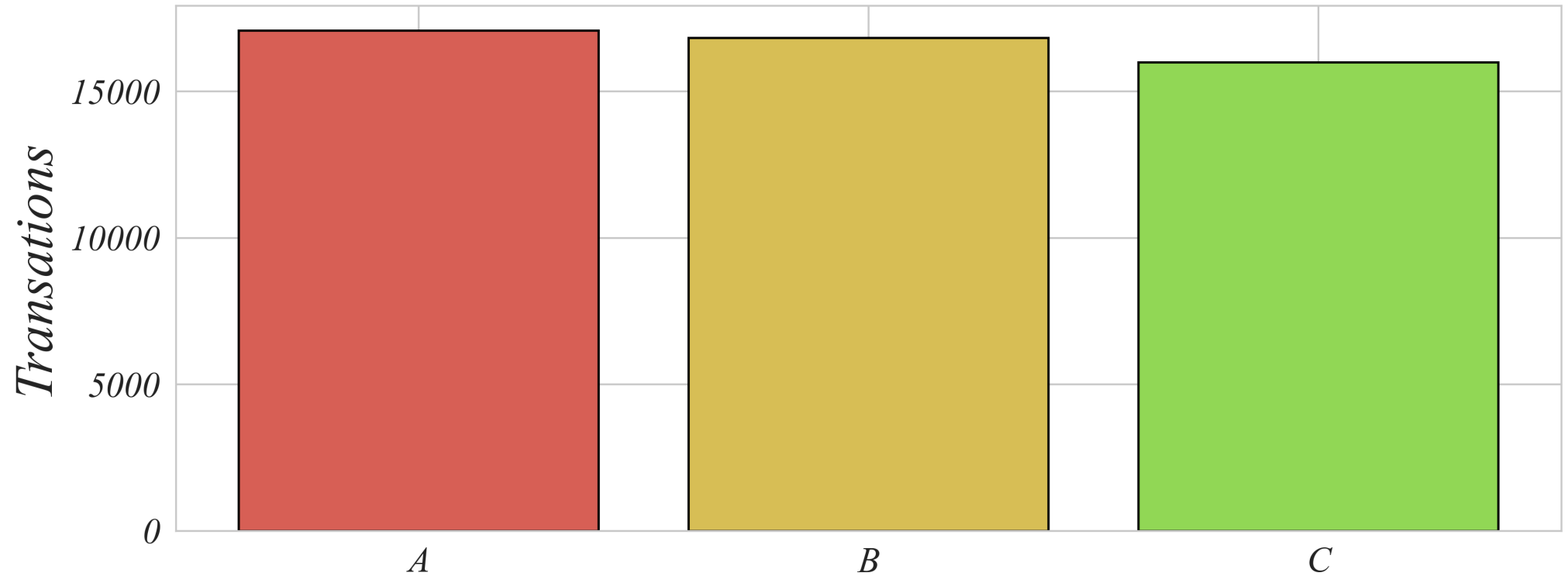
```
1 # Count by category
2 sales_counts = sales['sales_outlet_id'].value_counts()
3
4 # Bar graph
5 plt.bar(['A','B','C'], sales_counts.values)
```



Hiring a New Barista

Q. Which coffee shop is the busiest?

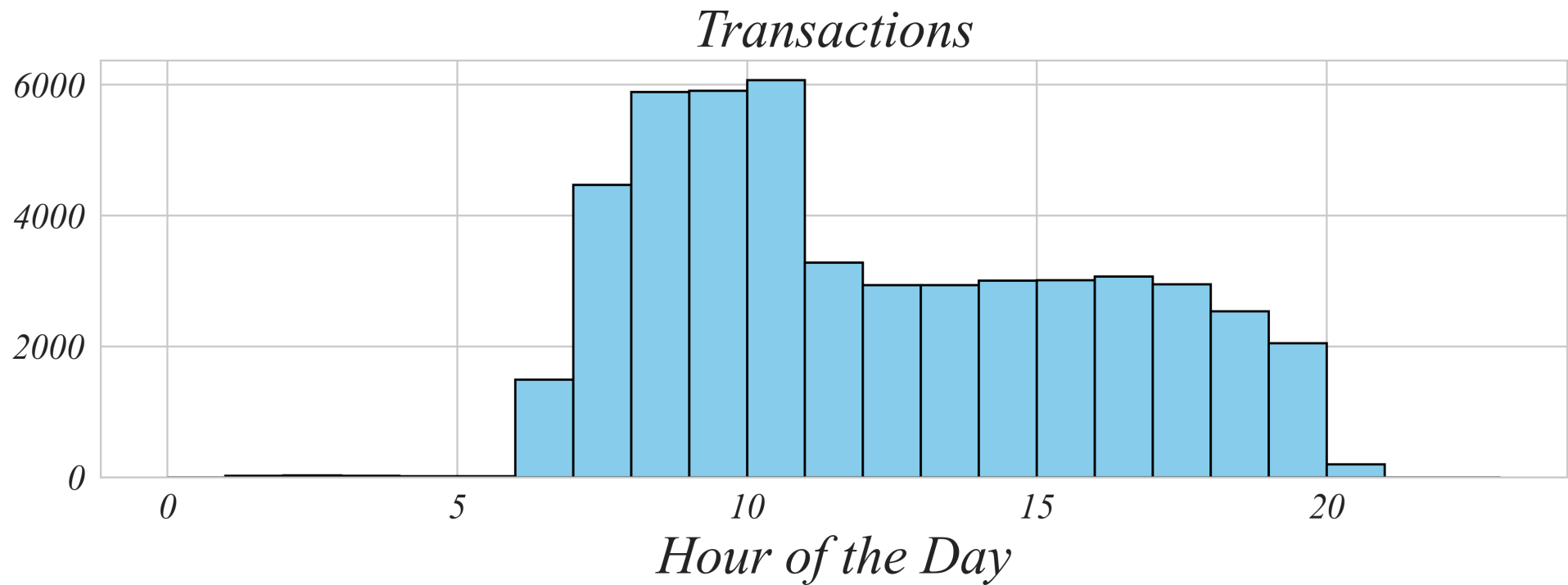
> a bar chart makes it easy to compare between categories



Hiring a New Barista

Q. What time of day is the busiest?

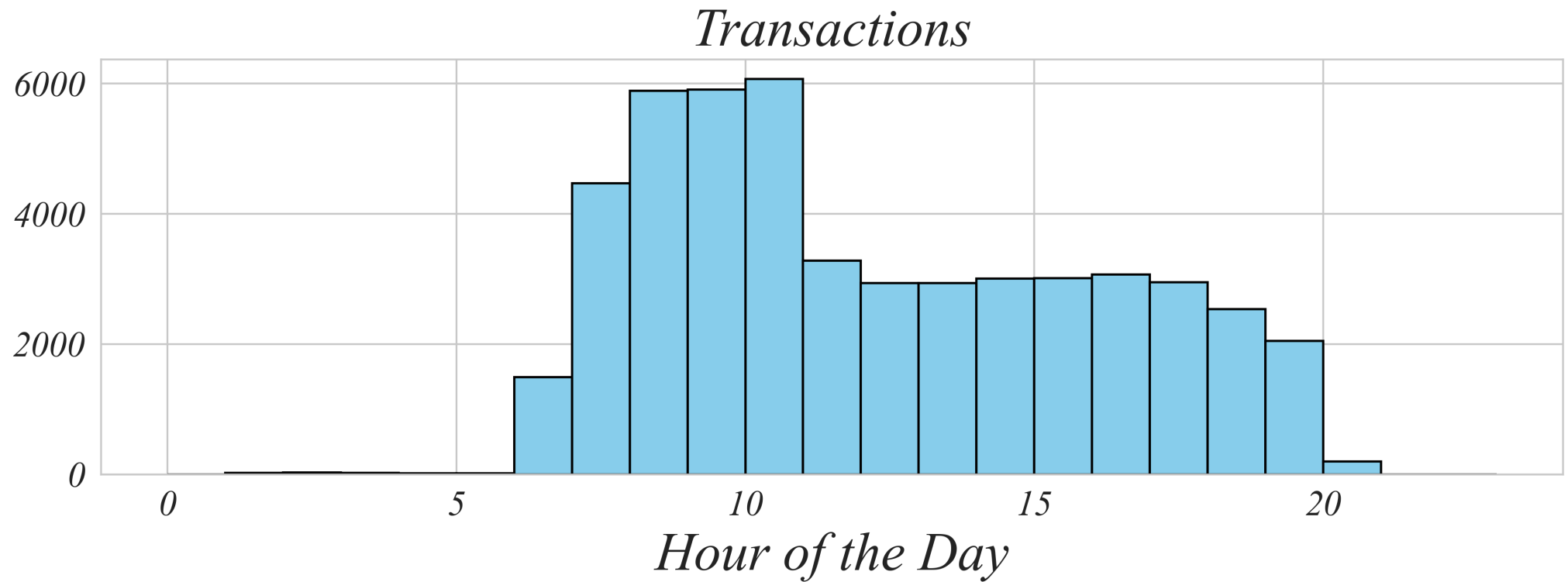
```
1 # Create bins from 0 to 24
2 bins = range(0, 24, 1)
3
4 # Create a histogram
5 plt.hist(sales['Hours'], bins=bins)
```



Hiring a New Barista

Q. What time of day is the busiest?

> *a histogram makes it easy to compare time of day*

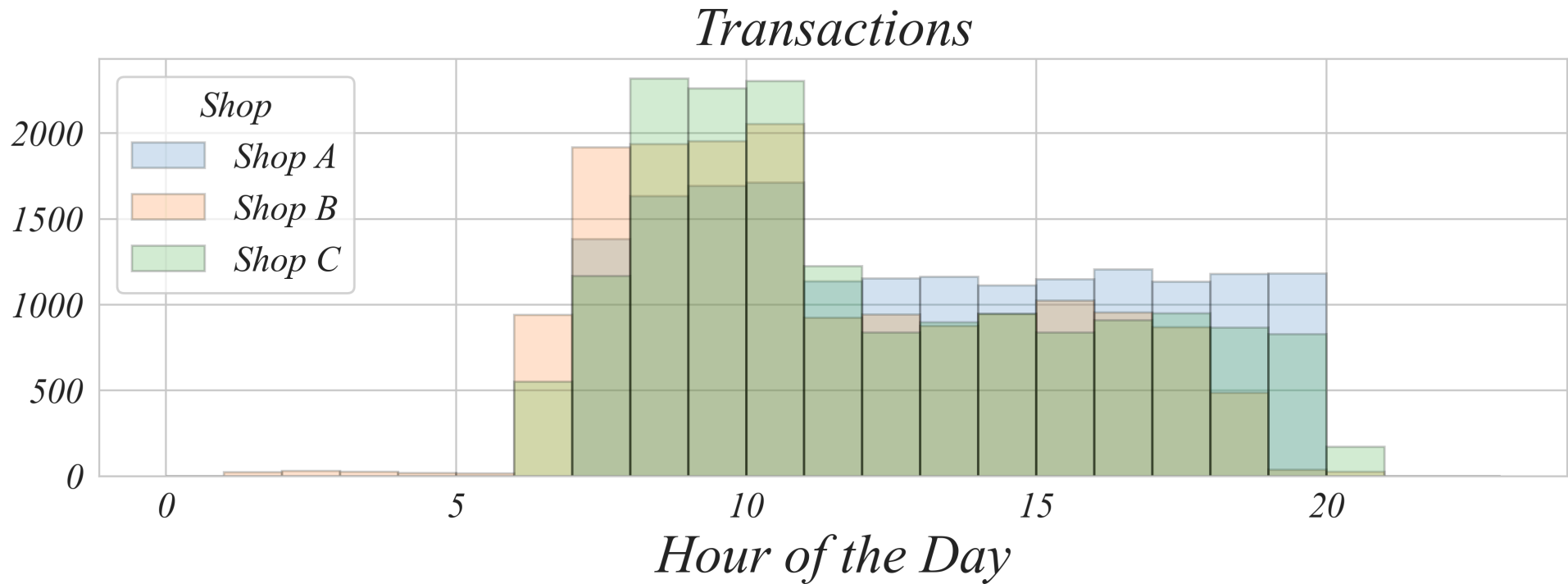


> *does this mean the morning shift at Shop A is the busiest?*

Hiring a New Barista

Q. Which shift is the busiest?

> *an overlaid histogram can show all three groups*



> *does this show the data clearly?*

Hiring a New Barista

Q. Which shift is the busiest?

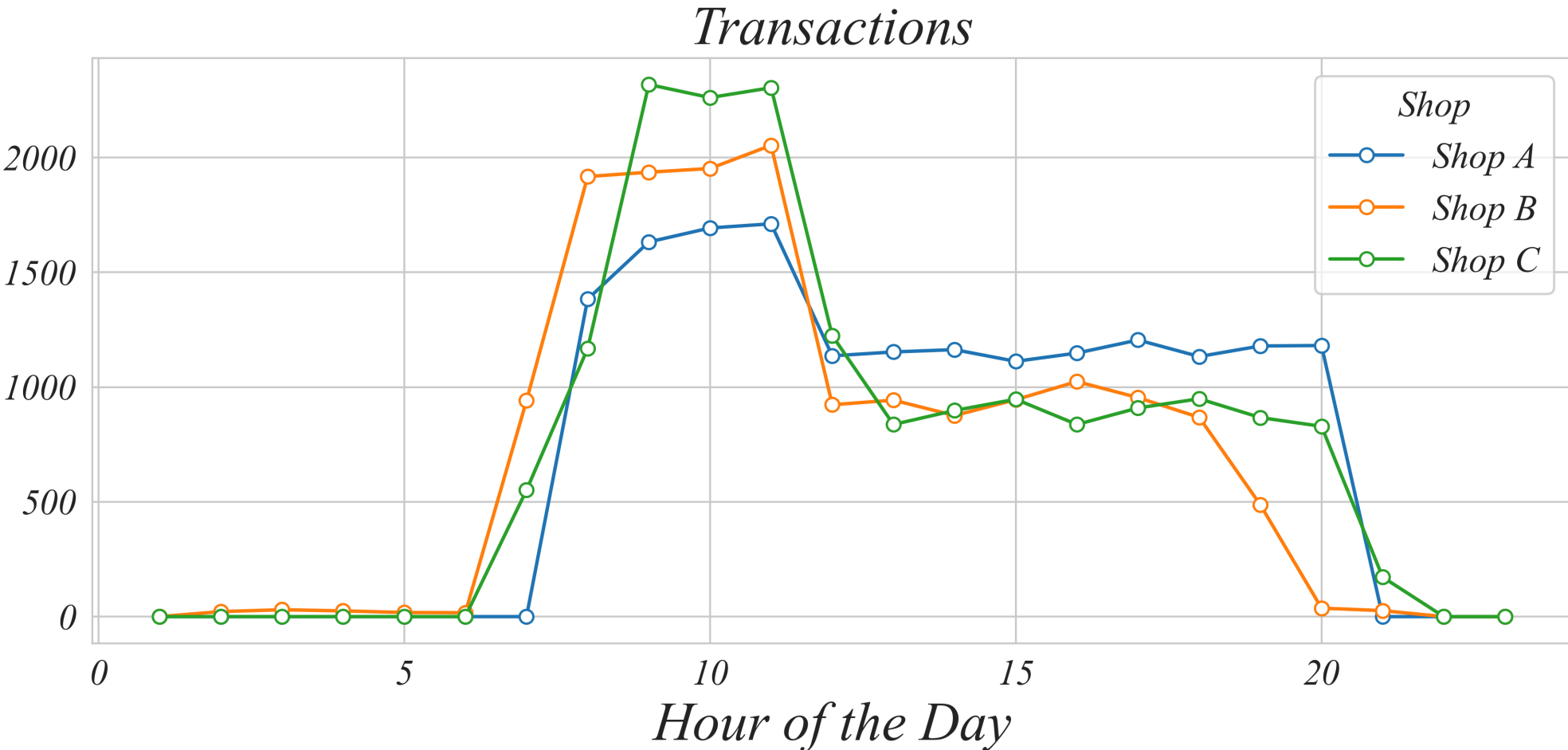
> instead, lets use a line graph

```
1 # Select Shop A data, summarize, and sort
2 shop_A = sales[sales.Shop == 'A'].Hours.value_counts().sort_index()
3
4 # Plot Shop A
5 plt.plot(shop_A, label='Shop A')
6
7 # Shop B
8 shop_B = sales[sales.Shop == 'B'].Hours.value_counts().sort_index()
9 plt.plot(shop_B, label='Shop B')
10
11 # Shop C
12 shop_C = sales[sales.Shop == 'C'].Hours.value_counts().sort_index()
13 plt.plot(shop_C, label='Shop C')
```


Hiring a New Barista

Q. Which shift is the busiest?

> *instead, lets use a line graph*



Part 1.4 | Numerical Variables by Category

Summary

- *Categorical variables and continuous variables can give us different views of the same data.*
- *We can visualize both views on the same graph.*
- *Line graphs help simplify the visualization of multiple categories.*