How visualization works in Python

PYTHON FOR SPREADSHEET USERS

Chris Cardillo
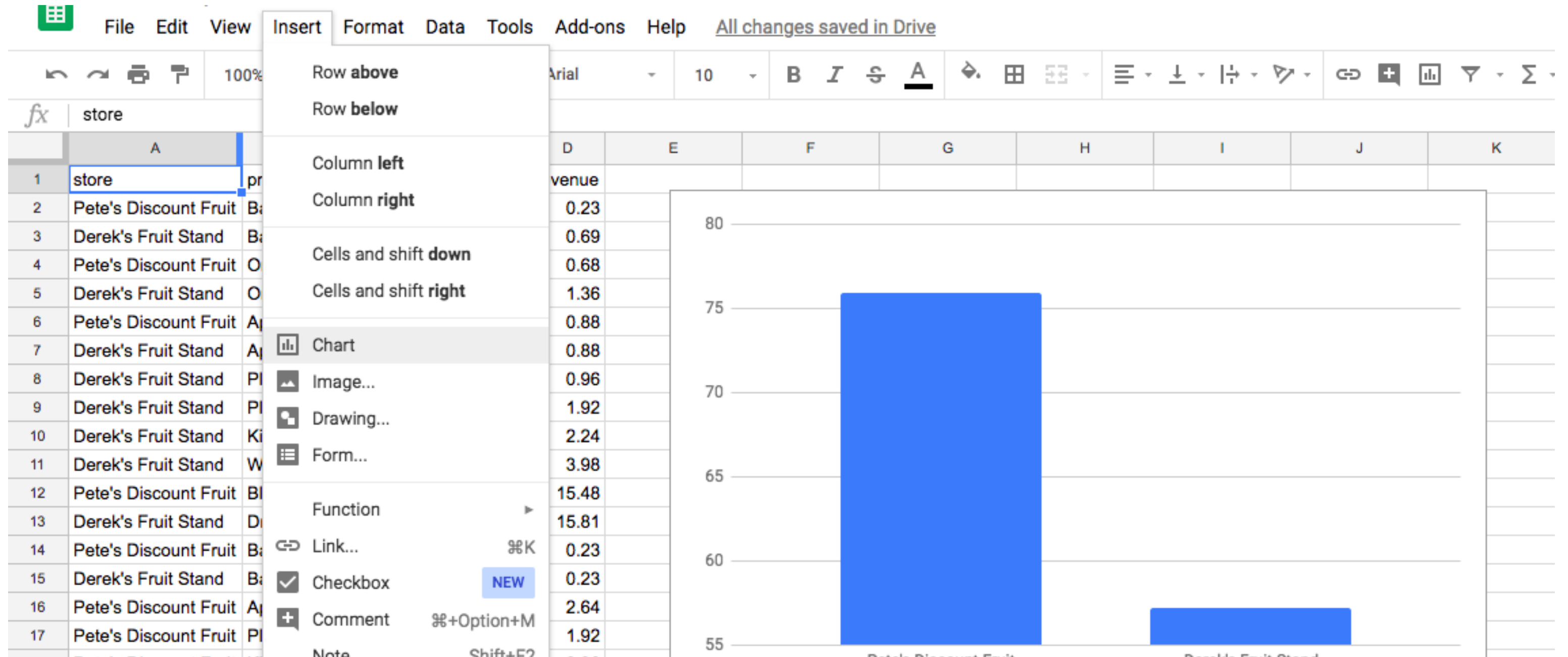Data Scientist at DataCamp

DataCamp

Importing some more packages

```python
import seaborn as sns
import matplotlib.pyplot as plt
```

Plotting functions

- `sns.barplot()` - Creates the bar plot

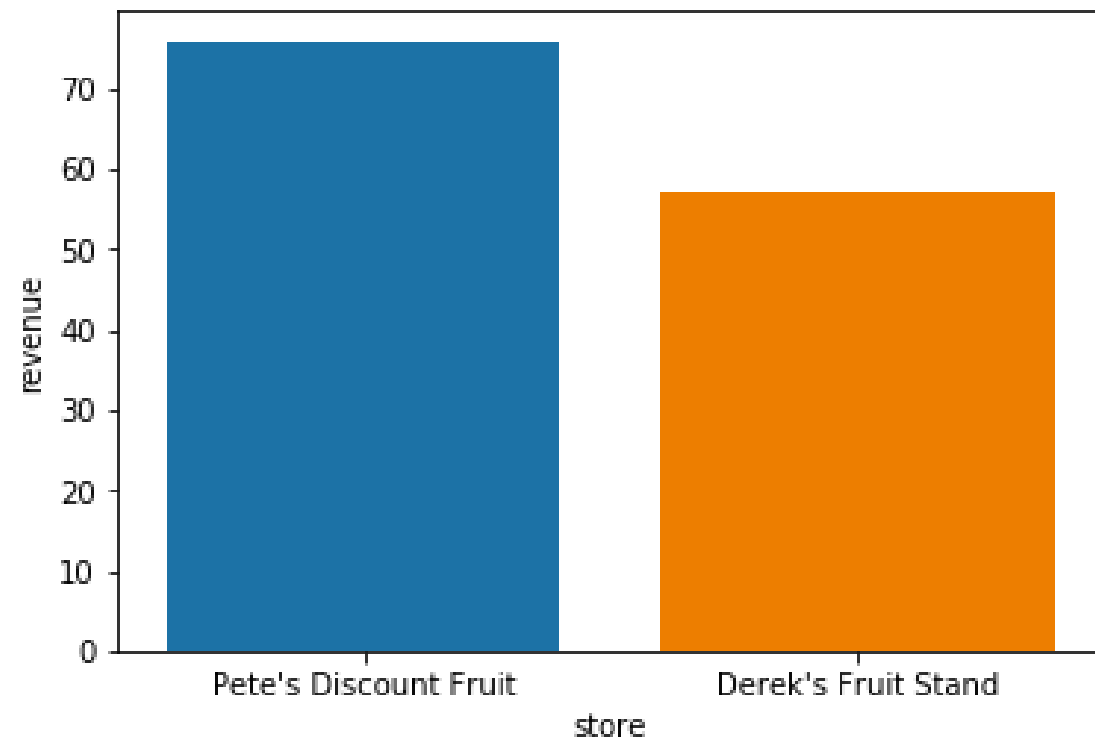- `plt.show()` - Displays the bar plot

Spreadsheet graphs

Python graphs

```
In [3]: totals = fruit_sales.groupby('store', as_index=False).sum().sort_values('revenue', ascending=False).reset_index()
```

```
In [4]: sns.barplot(x='store', y='revenue', data=totals)

plt.show()
```

sns.barplot()

```
sns.barplot(x='store', y='revenue', data=totals)
```

plt.show()

```
sns.barplot(x='store', y='revenue', data=totals)

plt.show()
```

plt.savefig()

```
sns.barplot(x='store', y='revenue', data=totals)

plt.savefig('awesome_plot.png')
```

Your turn!

PYTHON FOR SPREADSHEET USERS
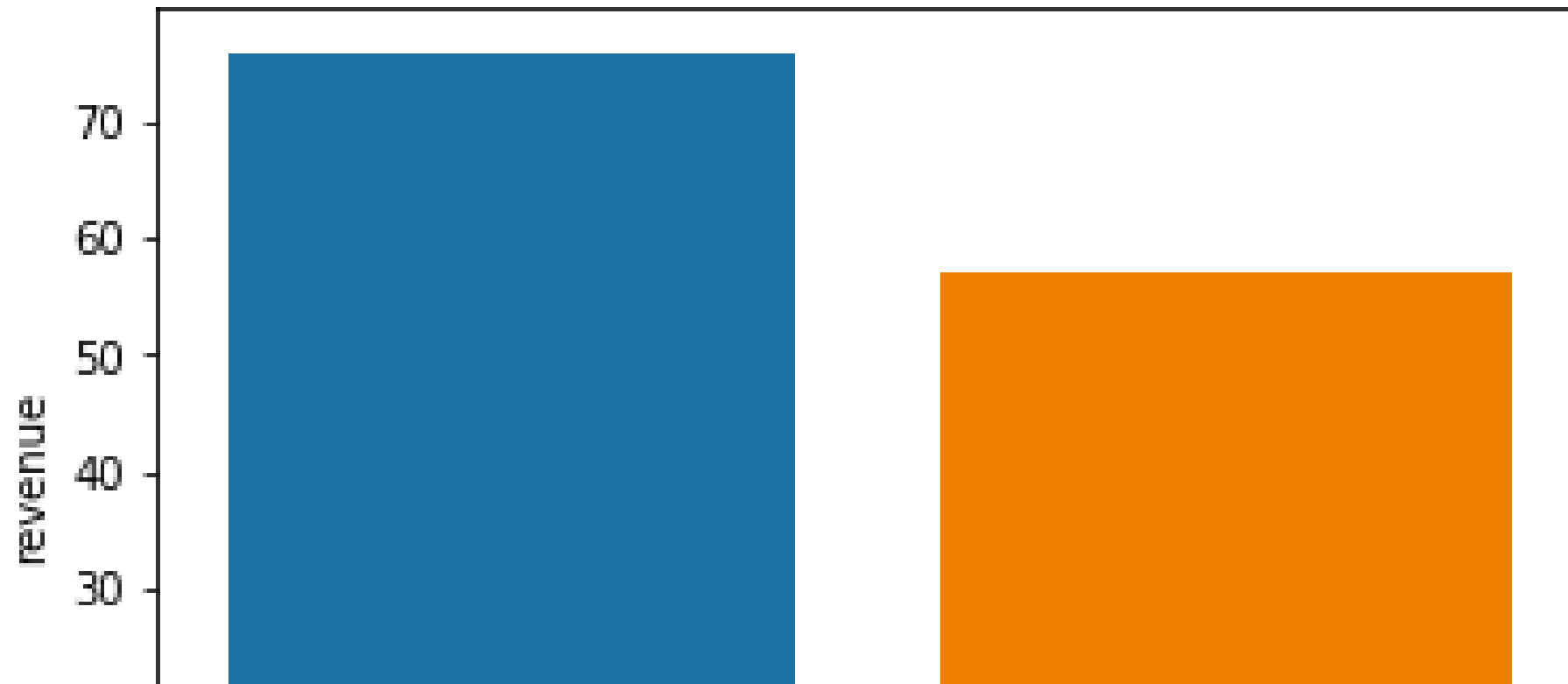
Building up the barplot

PYTHON FOR SPREADSHEET USERS

Chris Cardillo
Data Scientist at DataCamp

Previous barplot
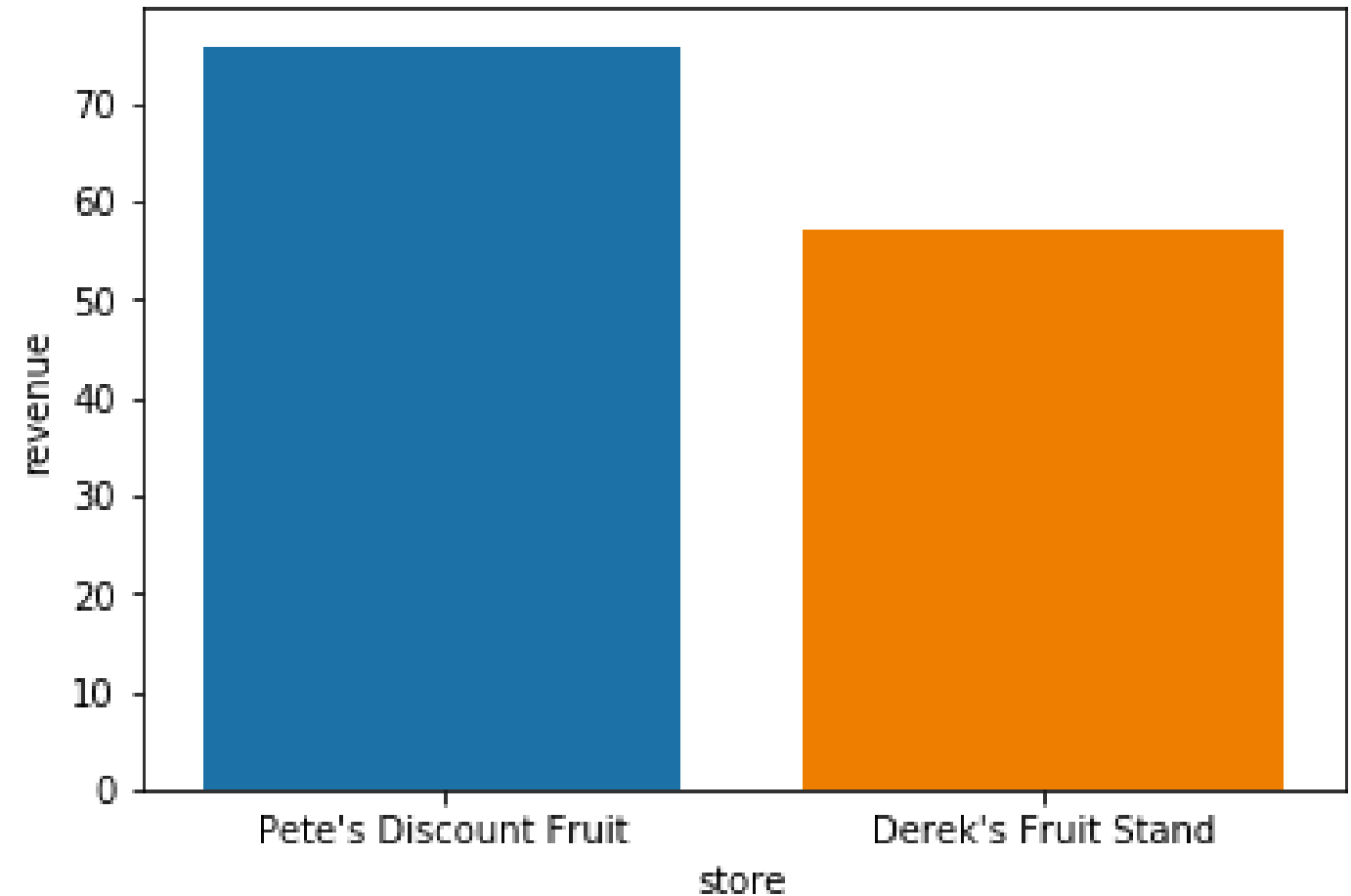
```
sns.barplot(x='store', y='revenue', data=totals)

plt.show()
```
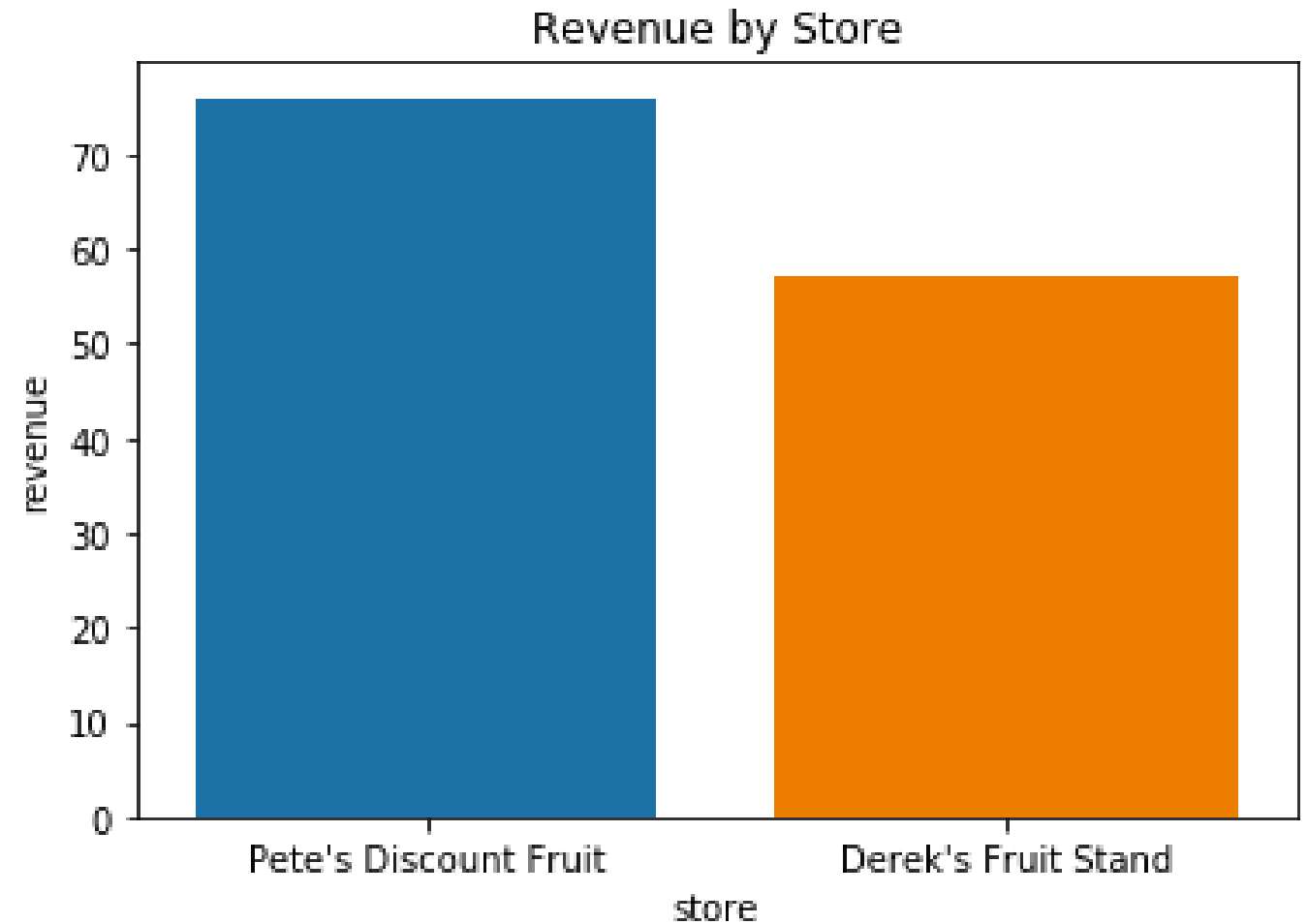
Adding labels

```python
sns.barplot(x='store',
            y='revenue',
            data=totals)

# add a title

# add x-axis label

# add y-axis label

plt.show()
```

Adding a title with plt.title()

```
sns.barplot(x='store',
            y='revenue',
            data=totals)

plt.title('Revenue by Store')

# add x-axis label

# add y-axis label

plt.show()
```

Adding an x-axis label with plt.xlabel()

```python
sns.barplot(x='store',
            y='revenue',
            data=totals)

plt.title('Revenue by Store')

plt.xlabel('Store')

# add y-axis label

plt.show()
```
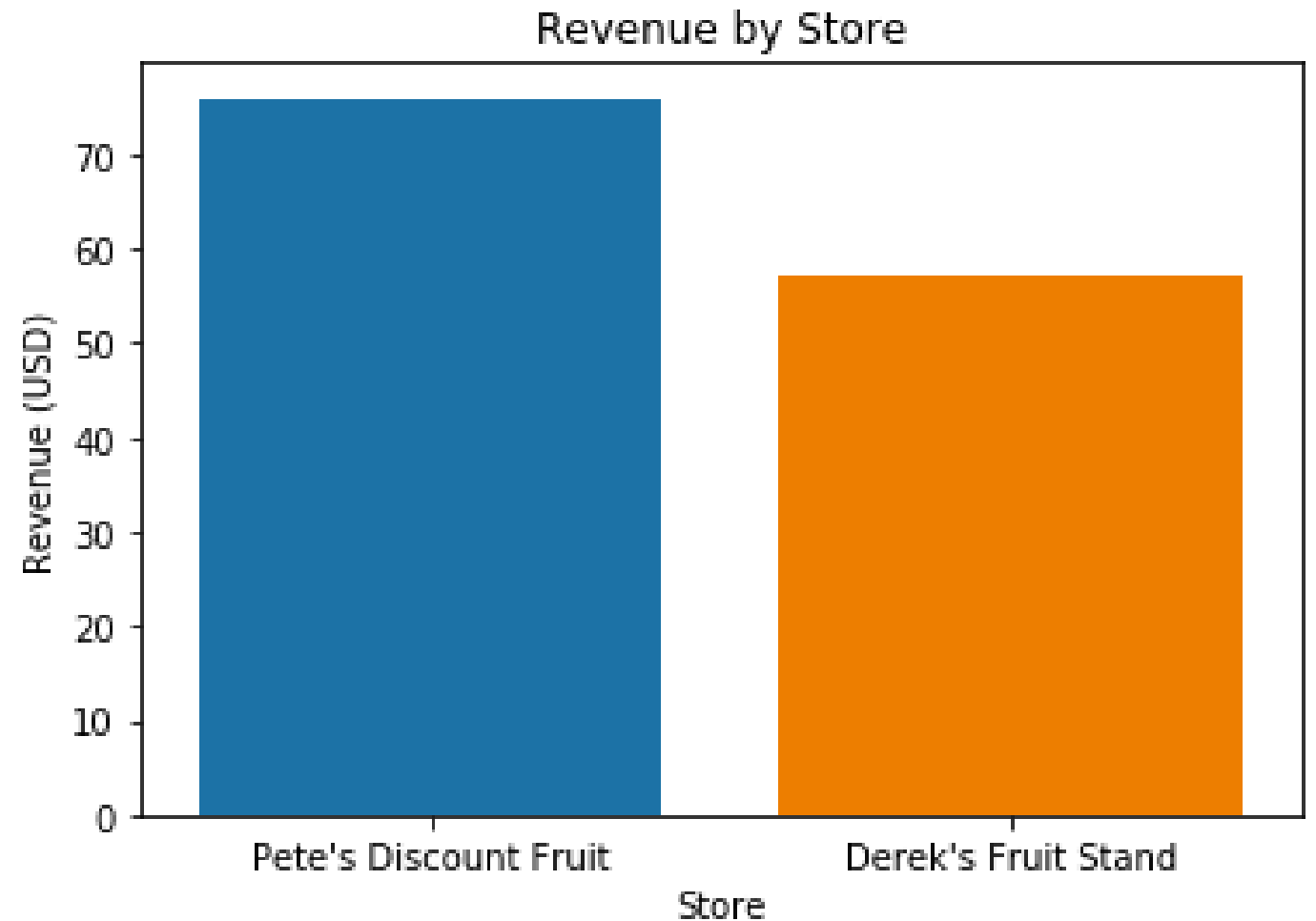
Adding an y-axis label with plt.ylabel()

```python
sns.barplot(x='store',
            y='revenue',
            data=totals)

plt.title('Revenue by Store')

plt.xlabel('Store')

plt.ylabel('Revenue (USD)')

plt.show()
```

Removing unwanted borders with sns.despine()

```python
sns.barplot(x='store',
            y='revenue',
            data=totals)

plt.title('Revenue by Store')

plt.xlabel('Store')

plt.ylabel('Revenue (USD)')

sns.despine()

plt.show()
```


Revenue by Store

Adding style with sns.set_style()

```python
sns.set_style('whitegrid')

sns.barplot(x='store',
            y='revenue',
            data=totals)

plt.title('Revenue by Store')

plt.xlabel('Store')

plt.ylabel('Revenue (USD)')

sns.despine()

plt.show()
```
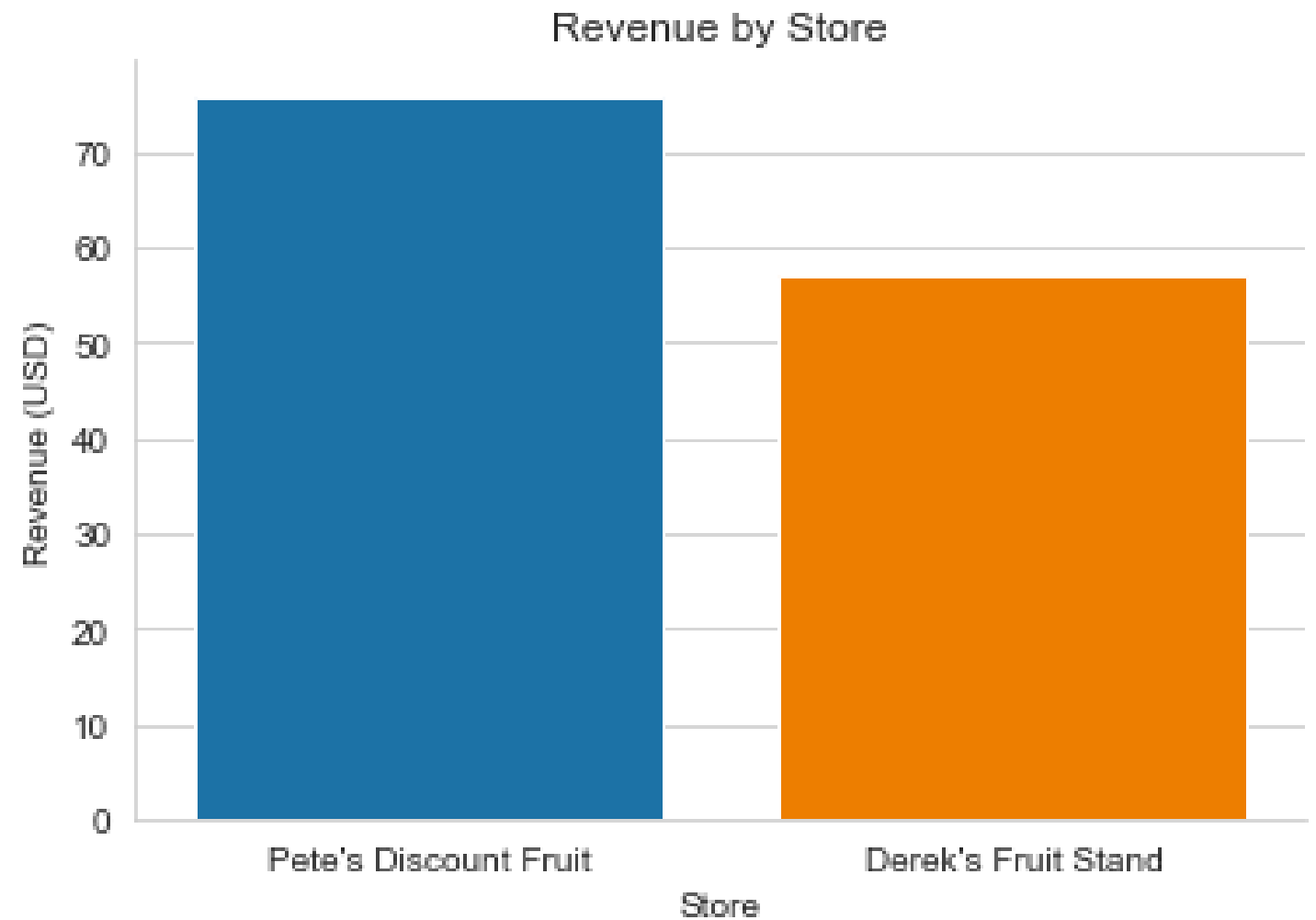
Side by side

BEFORE
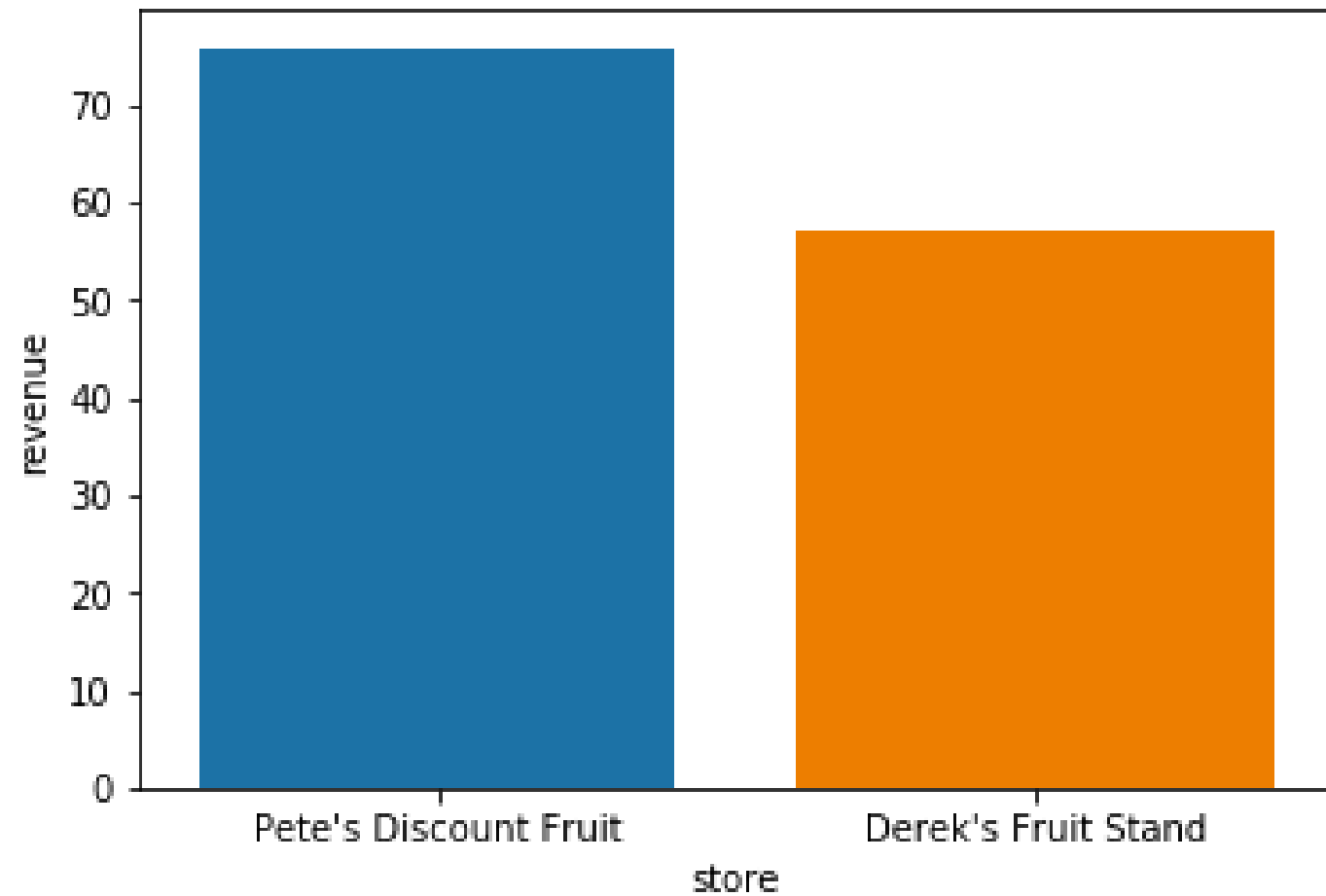
AFTER

Your turn!

PYTHON FOR SPREADSHEET USERS

The power of hue

PYTHON FOR SPREADSHEET USERS

Chris Cardillo
Data Scientist at DataCamp

Our nice barplot

THE PLOT

Revenue by Store
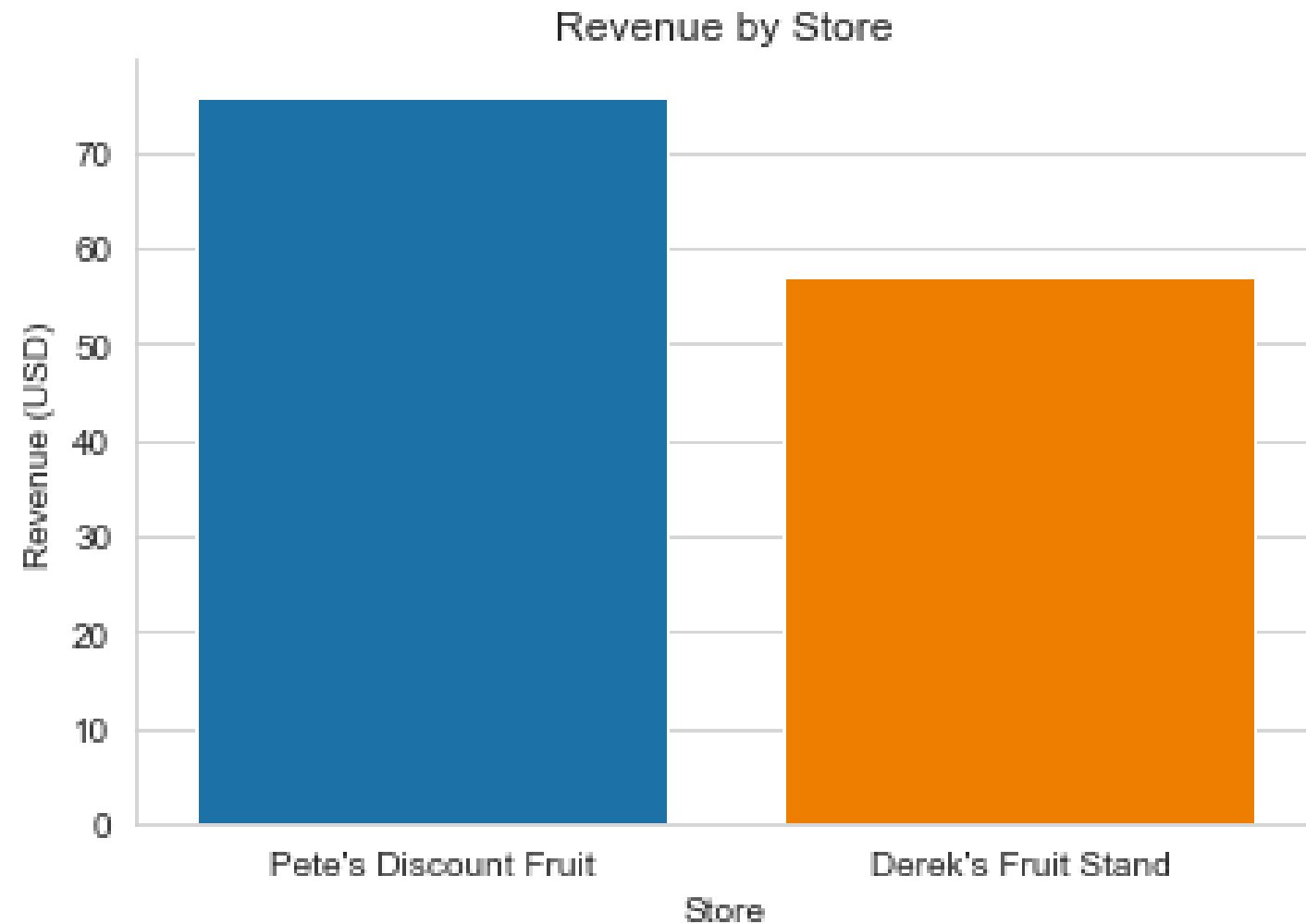
THE DATA BEHIND THE PLOT

```
                  store  quantity_purchased  revenue
0  Pete's Discount Fruit                  30    75.90
1     Derek's Fruit Stand                 27    57.22
```

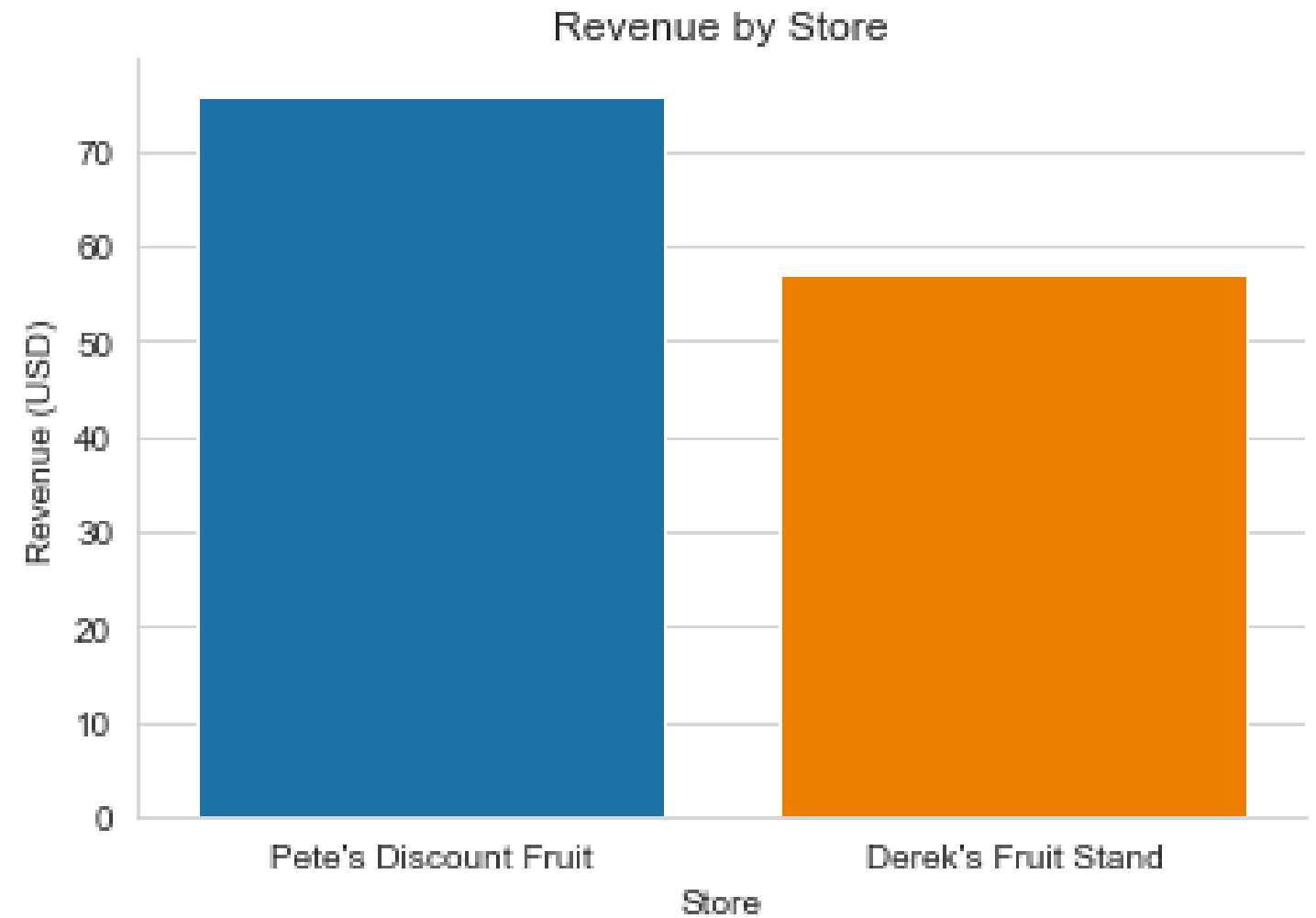# A more granular barplot

THE DATA BEHIND THE NEW PLOT

```
                   store product_name  quantity_purchased  revenue
0     Derek's Fruit Stand        Apple                   2     1.76
1     Derek's Fruit Stand       Banana                   4     0.92
2     Derek's Fruit Stand  Blueberries                   2    10.32
3     Derek's Fruit Stand  Dragonfruit                   4    21.08
4     Derek's Fruit Stand         Kiwi                   5     5.60
5     Derek's Fruit Stand       Orange                   4     2.72
6     Derek's Fruit Stand         Plum                   3     2.88
7     Derek's Fruit Stand   Watermelon                   3    11.94
8    Pete's Discount Fruit        Apple                   4     3.52
9    Pete's Discount Fruit       Banana                   2     0.46
10   Pete's Discount Fruit  Blueberries                   6    30.96
11   Pete's Discount Fruit  Dragonfruit                   4    21.08
12   Pete's Discount Fruit         Kiwi                   6     6.72
13   Pete's Discount Fruit       Orange                   2     1.36
14   Pete's Discount Fruit         Plum                   4     3.84
15   Pete's Discount Fruit   Watermelon                   2     7.96
```

A more granular barplot

THE DATA BEHIND THE NEW PLOT

```
            store product_name  quantity_purchased  revenue
0   Derek's Fruit Stand          Apple                   2     1.76
1   Derek's Fruit Stand         Banana                   4     0.92
2   Derek's Fruit Stand    Blueberries                   2    10.32
3   Derek's Fruit Stand    Dragonfruit                   4    21.08
4   Derek's Fruit Stand           Kiwi                   5     5.60
5   Derek's Fruit Stand         Orange                   4     2.72
6   Derek's Fruit Stand           Plum                   3     2.88
7   Derek's Fruit Stand     Watermelon                   3    11.94
8   Pete's Discount Fruit         Apple                   4     3.52
9   Pete's Discount Fruit        Banana                   2     0.46
10  Pete's Discount Fruit   Blueberries                   6    30.96
11  Pete's Discount Fruit   Dragonfruit                   4    21.08
12  Pete's Discount Fruit          Kiwi                   6     6.72
13  Pete's Discount Fruit        Orange                   2     1.36
14  Pete's Discount Fruit          Plum                   4     3.84
15  Pete's Discount Fruit    Watermelon                   2     7.96
```
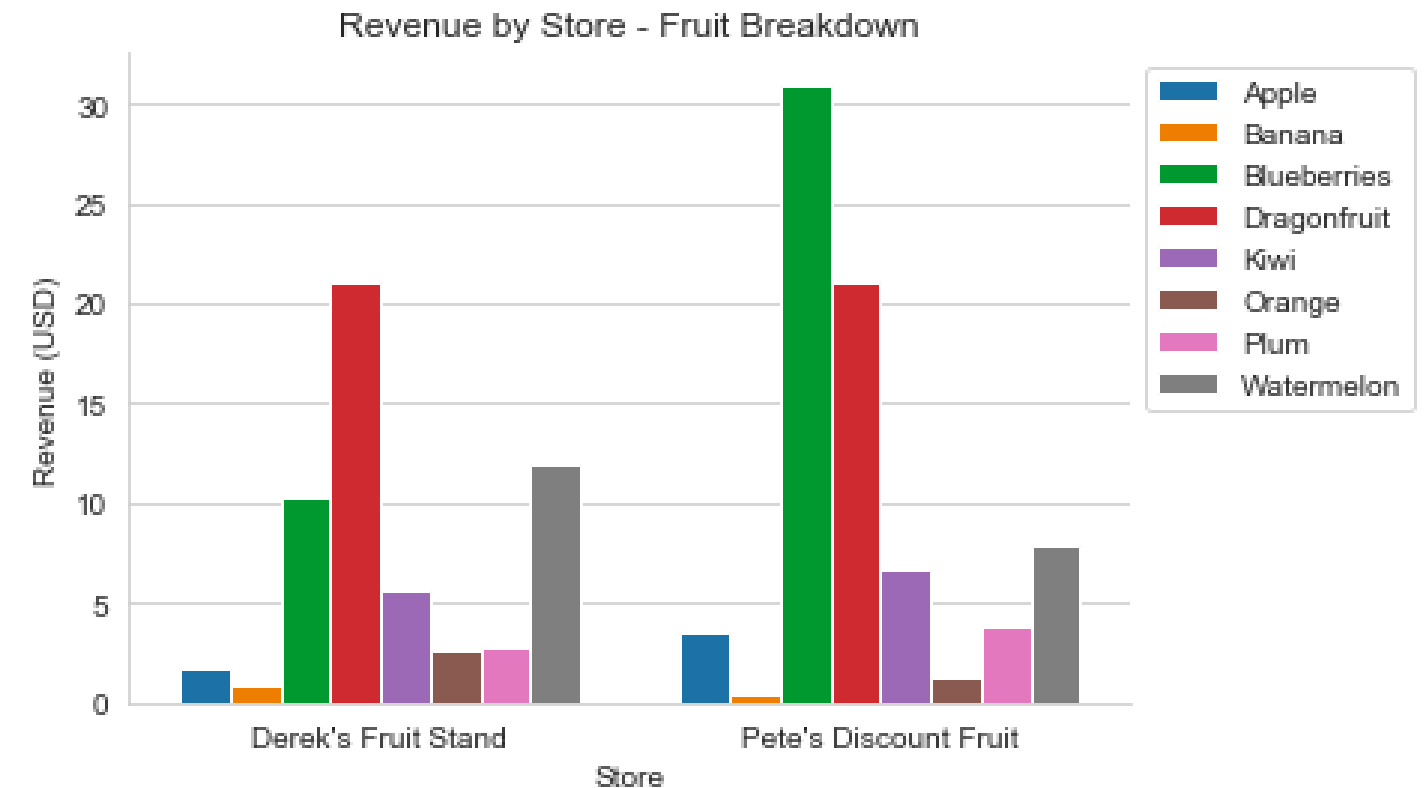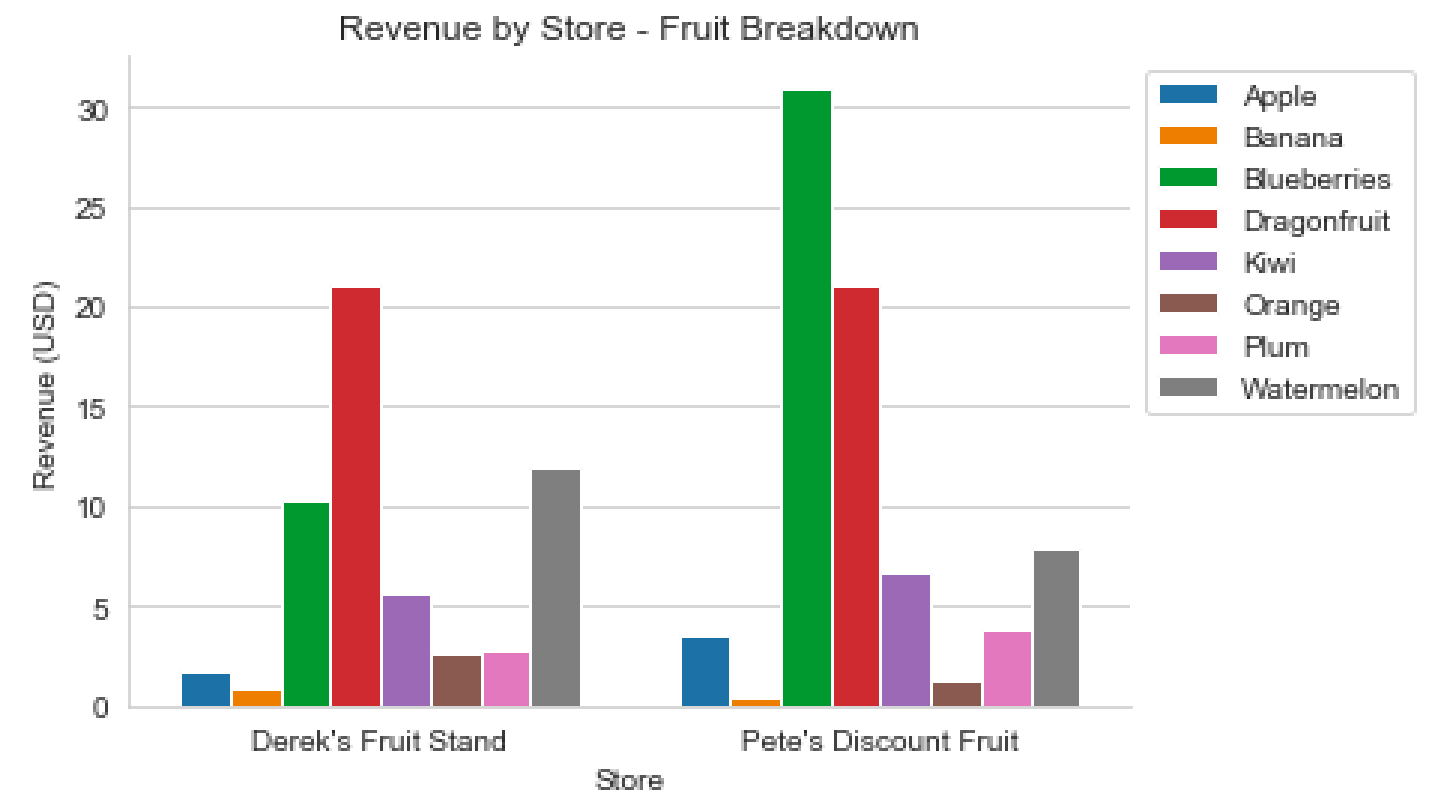
THE NEW PLOT

Hue

```
sns.barplot(x='store',
            y='revenue',
            data=fruit_store_summary,
            hue='product_name')
```

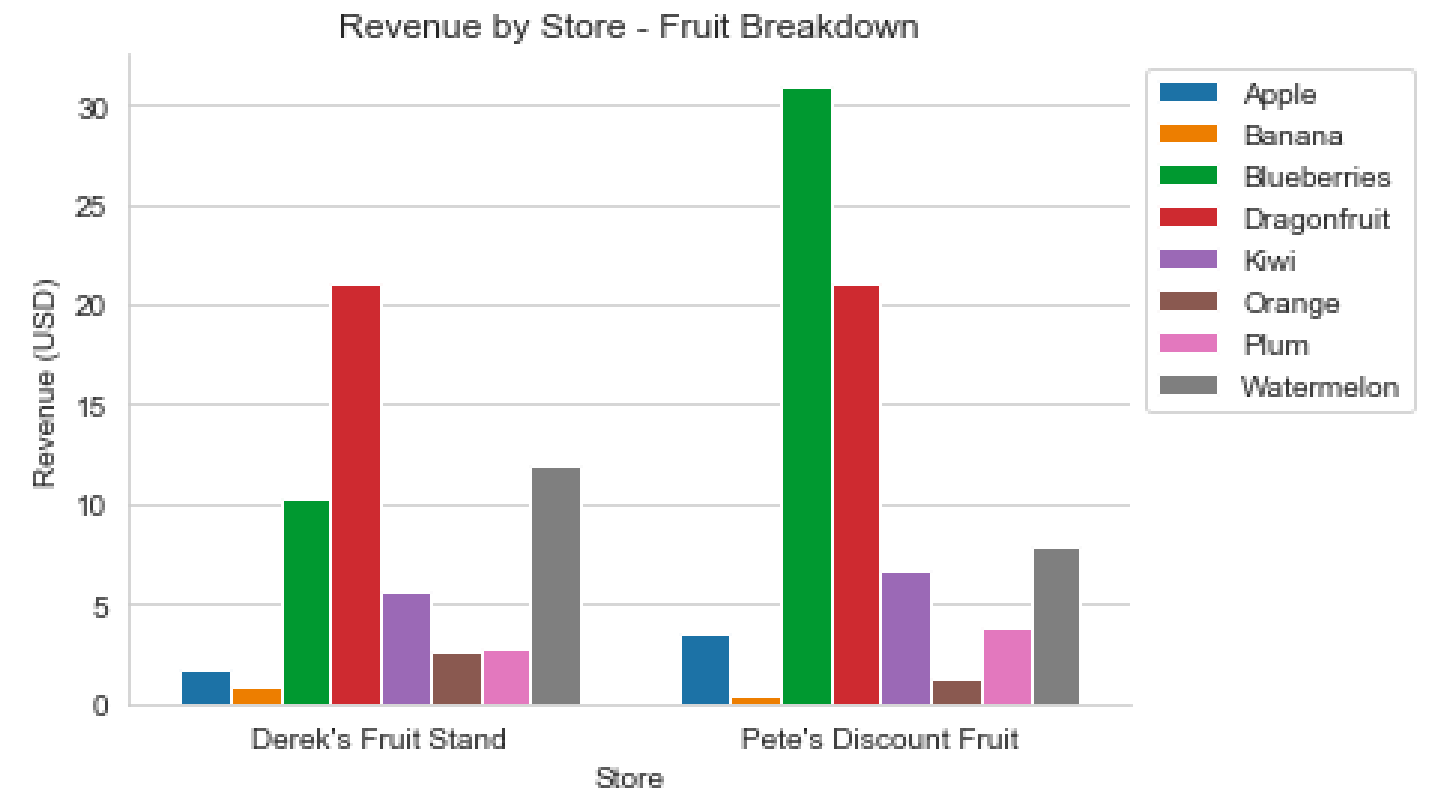# Hue

```
sns.barplot(x='store',
            y='revenue',
            data=fruit_store_summary,
            hue='product_name')
```

plt.legend()

```
sns.barplot(x='store',
            y='revenue',
            data=fruit_store_summary,
            hue='product_name')

plt.legend(bbox_to_anchor=(1, 1))
```
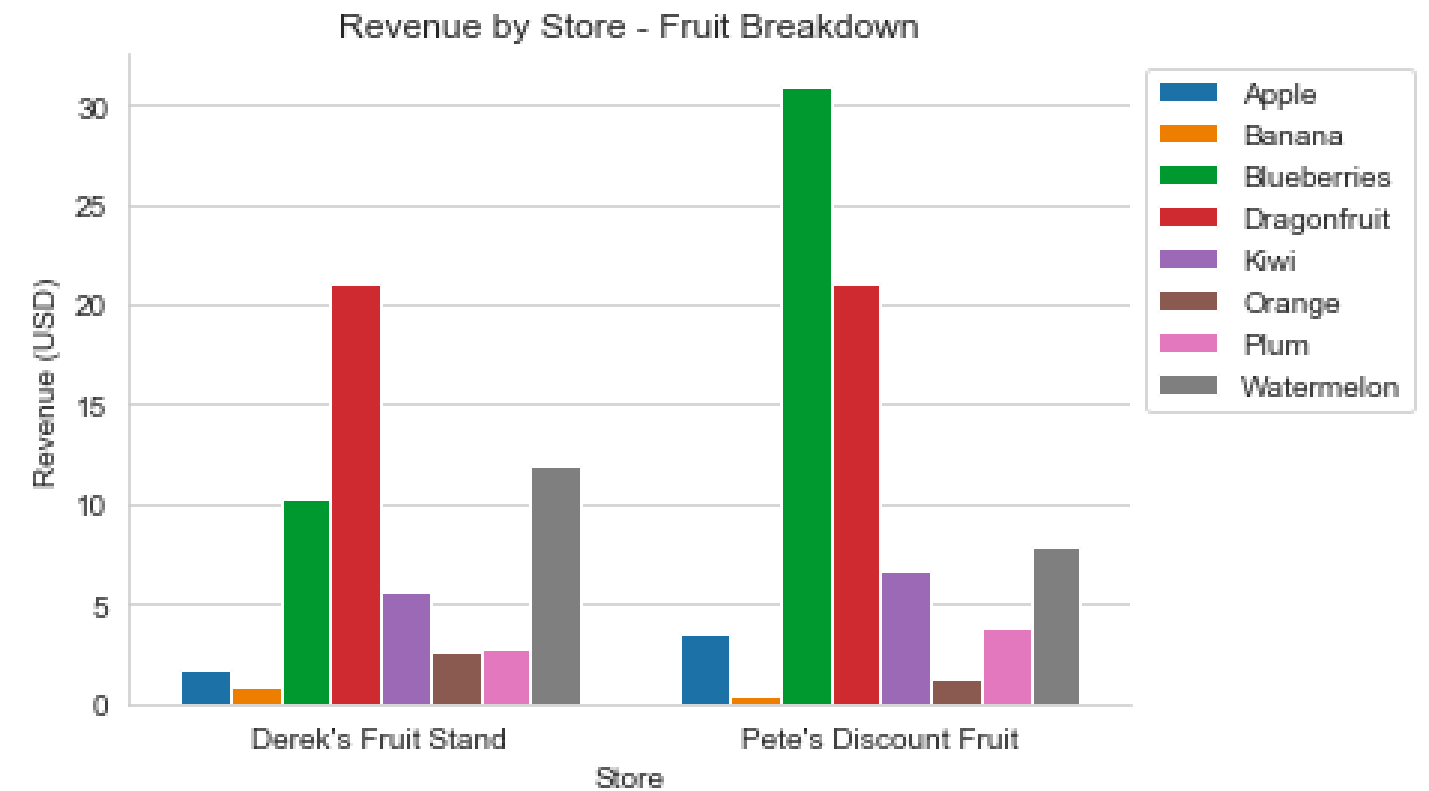

Revenue by Store - Fruit Breakdown

All together

```python
sns.set_style('whitegrid')

sns.barplot(x='store',
            y='revenue',
            data=fruit_store_summary,
            hue='product_name')

plt.legend(bbox_to_anchor=(1, 1))

plt.title('Revenue by Store - Fruit Breakdown')
plt.xlabel('Store')
plt.ylabel('Revenue (USD)')
sns.despine()
plt.show()
```

Your turn!

PYTHON FOR SPREADSHEET USERS

Wrapping up

PYTHON FOR SPREADSHEET USERS

Chris Cardillo
Data Scientist at DataCamp

Chapter 1

- Importing with `pd.read_excel()`

- Understanding with `.head()`, `.describe()`, and `.info()`

- Filtering with `df[df[column] == 'value']`

- Creating with `df[new_column] = df[old_column] + 1`

Chapter 2

- Pivoting with `df.group_by().sum()`

Chapter 3

- Multiple sheets with `pd.ExcelFile()` and `.parse()`

- `.merge()` and left joins

- Functions, methods, & attributes

Chapter 4

- `sns.barplot()` and `plt.show()`

- `plt.title()` , `plt.xlabel()` , and `plt.ylabel()`

- `sns.set_style()` and `sns.despine()`

- `hue`

Next Steps

- **Course Suggestion: pandas Foundations**

- **Career Track Suggestion: Data Analyst with Python**

- **DataCamp Community Resource: Cheat Sheets**

Good luck!

PYTHON FOR SPREADSHEET USERS