

# CoxAssignment01

**Task: Differentiate between the different types of learning, visualize data, explore it, and understand it.**

**What type of learning and technique is being used in the following examples?**

**Automated sorting of mail by zip code scanning.**

Supervised Classification

**Printing of custom discount coupons at the conclusion of a grocery store checkout based on what you just bought and what others have bought previously.**

Unsupervised Collaborative Filters

**Identifying segments of similar customers.**

Unsupervised Data Reduction

**Discuss the following:**

**Describe the difference between using a training and testing set of data.**

Training data is a subset of the data where testing data is to see whether the model you created is accurate for what you are trying to discover.

**Why are exploratory data analyses performed?**

To help get a better understanding of the data and possibly find any patterns or errors that could be influence the results of what you are trying to discover.

## Data Visualization

**Please use the RidingMowers.csv for this exercise.**

A company that manufactures riding mowers wants to identify the best sales prospects for an intensive sales campaign. In particular, the manufacturer is interested in classifying households as prospective owners or nonowners on the basis of Income (in \$ 1000s) and Lot Size (in 1000 ft<sup>2</sup>). The marketing expert looked at a random sample of 24 households, given in the file RidingMowers.csv.

## Data Dictionary (CREATE THIS ON YOUR OWN)

```
In [1]: # import required packages for this assignment
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
grass = pd.read_csv('RidingMowers.csv')
grass.rename(columns={'Lot_Size': 'Size'})
grass.head(10)
```

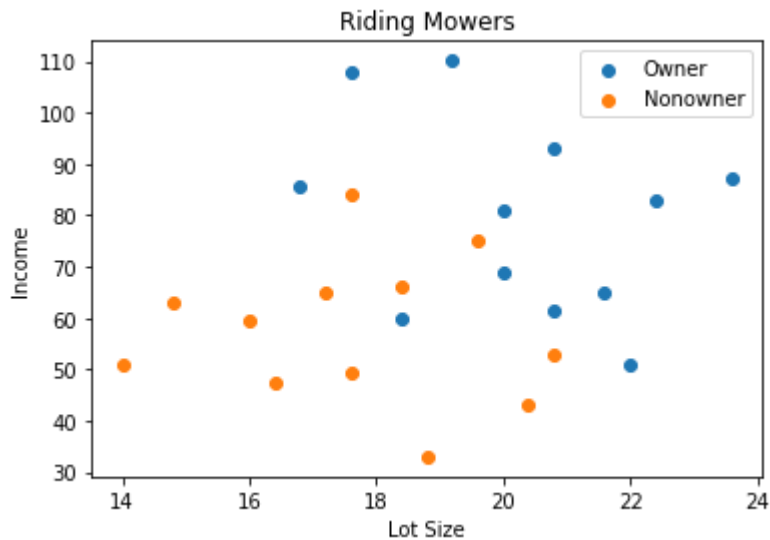
```
Out[1]:
```

	Income	Lot_Size	Ownership
0	60.0	18.4	Owner
1	85.5	16.8	Owner
2	64.8	21.6	Owner
3	61.5	20.8	Owner
4	87.0	23.6	Owner
5	110.1	19.2	Owner
6	108.0	17.6	Owner
7	82.8	22.4	Owner
8	69.0	20.0	Owner
9	93.0	20.8	Owner

Create a scatter plot of Lot Size vs. Income, color-coded by the outcome variable owner/nonowner. Make sure to obtain a well-formatted plot (create legible labels and a legend).

```
In [3]: _, ax = plt.subplots()
for ownership, color in ('Owner', 'C1'), ('Nonowner', 'C0'):
    subset_df = grass[grass.Ownership == ownership]
    ax.scatter(subset_df['Lot_Size'], subset_df['Income'])
ax.set_xlabel('Lot Size')
ax.set_ylabel('Income')
ax.set_title('Riding Mowers')
ax.legend(["Owner", "Nonowner"])
```

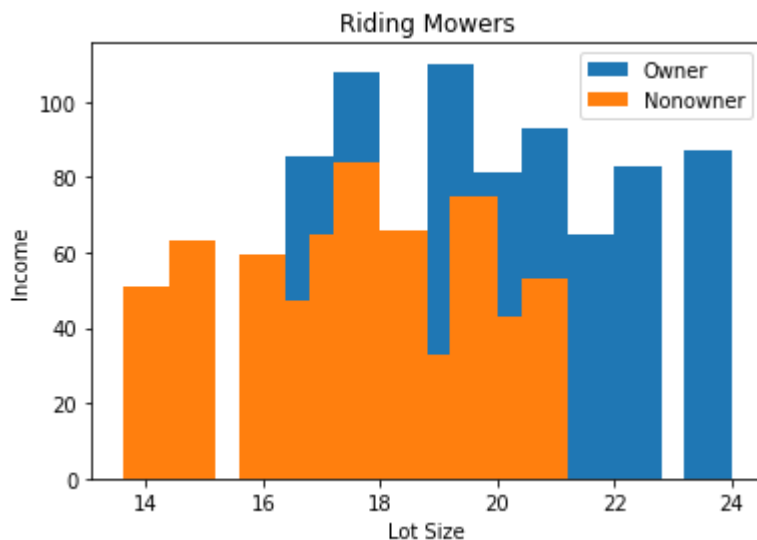
```
Out[3]: <matplotlib.legend.Legend at 0x17ada002610>
```



Create a different chart or plot to visualize the data. Analyze the visualized data.

```
In [4]: _, ax = plt.subplots()
for ownership, color in ('Owner', 'C1'), ('Nonowner', 'C0'):
    subset_df = grass[grass.Ownership == ownership]
    ax.bar(subset_df['Lot_Size'], subset_df['Income'])
ax.set_xlabel('Lot Size')
ax.set_ylabel('Income')
ax.set_title('Riding Mowers')
ax.legend(["Owner", "Nonowner"])
```

Out[4]: <matplotlib.legend.Legend at 0x17adb78a2e0>



From this bar chart You can see that the higher the income does not mean that they have a bigger lot as a large portion of the data is in between 16 and 21 acres. However, it does also show that the more income you have the more likely they are to have a larger lot as owners are the only ones with lots greater than 21 acres and nonowners with less income are the only ones with less than 16 acres.

## Supplemental Resources

### Use the following to help if you get stuck:

1. Google the error you may be receiving. At a higher level, google what you are expected to do. For example: How do I subtract variables in Python?
2. Navigate through the slides and use the find feature to locate keywords.
3. Rewatch a lecture video or two.
4. Check out YouTube. This is a great resource for students.
5. Post in the discussion board. Remember, I receive emails of postings and other students will as well. If you are not receiving emails, subscribe to the assignment questions discussion board.