0.1 Question 2e

If we were to drop businesses with MISSING postal code values, what specific types of businesses would we be excluding? In other words, is there a commonality among businesses with missing postal codes?

Hint: You may want to look at the names of the businesses with missing postal codes. Feel free to reuse parts of your code from 2d, but we will not be grading your code.

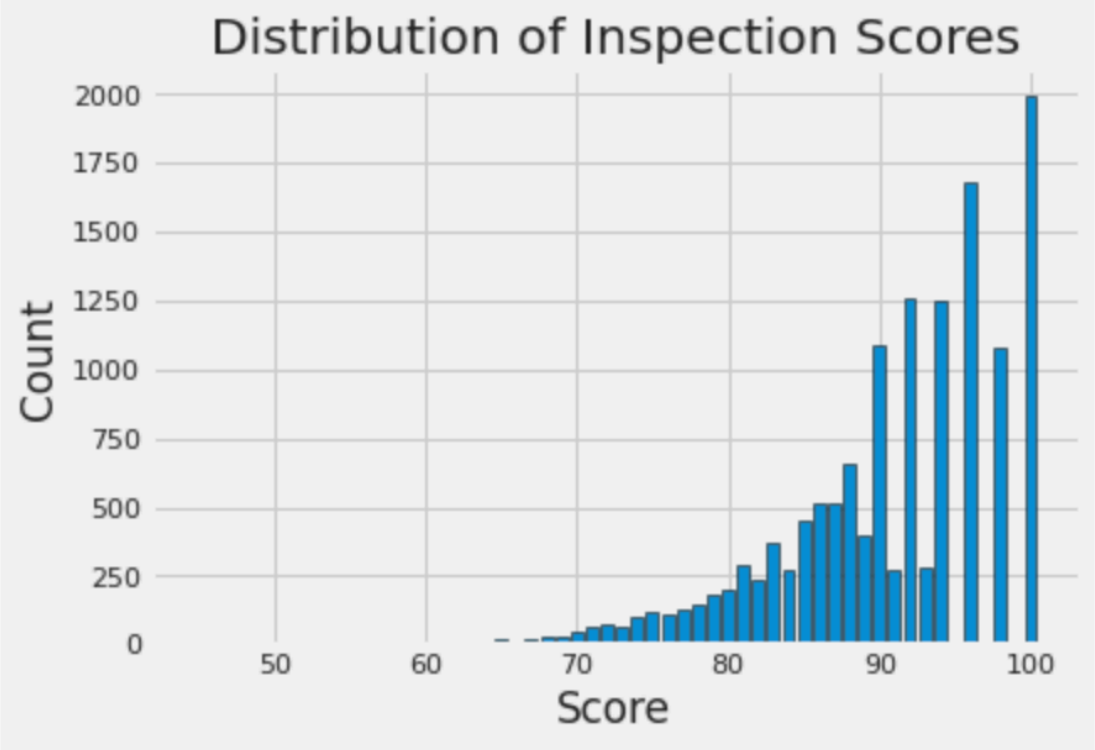
It looks like most of the businesses with missing zip codes are portable businesses, such as food trucks or street snack bars. Meaning that they move around the city, so they don't specifically stay in one zip code like restaurants and other fast food retailers.

0.2 Question 5a

Let's look at the distribution of inspection scores. As we saw before when we called head on this data frame, inspection scores appear to be integer values. The discreteness of this variable means that we can use a bar plot to visualize the distribution of the inspection score. Make a bar plot of the counts of the number of inspections receiving each score.

It should look like the image below. It does not need to look exactly the same (e.g., no grid), but **make sure that all labels and axes are correct**.

You should use the `ins` dataframe, and should ignore any score that is less than 0.



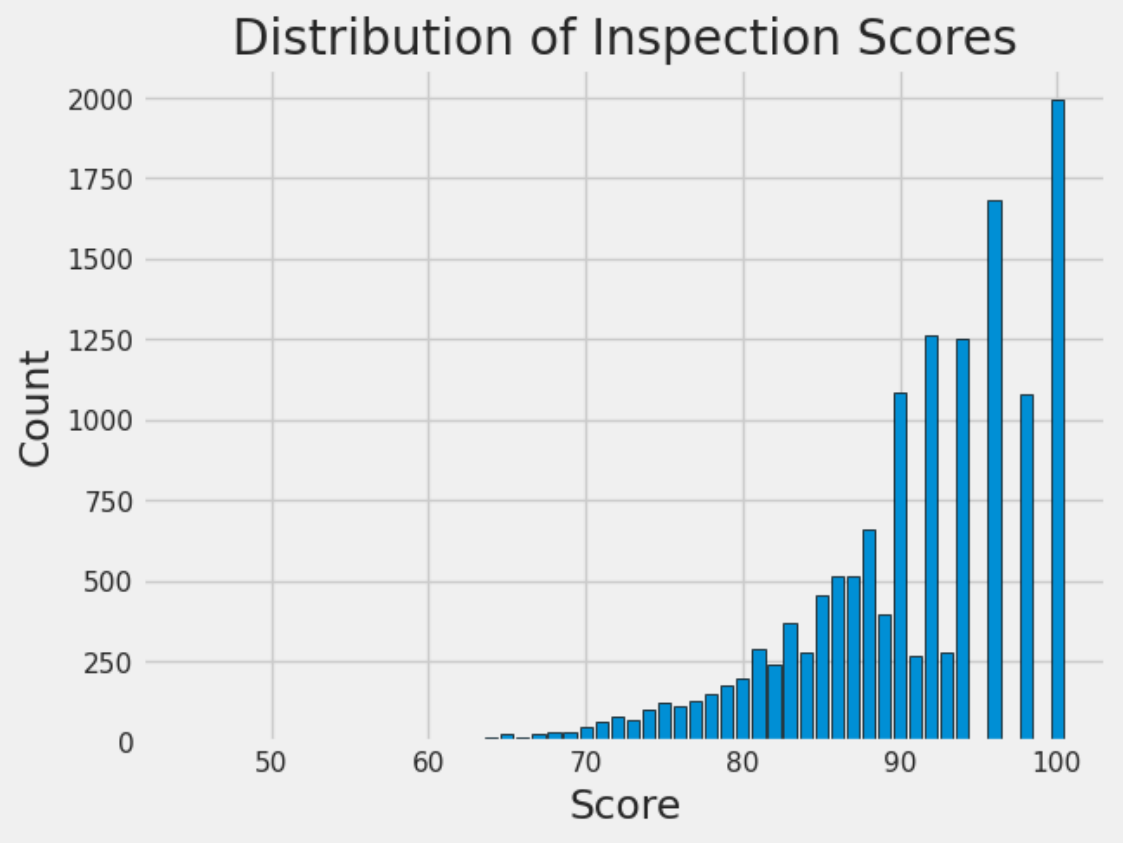
You might find this matplotlib.pyplot tutorial useful. Key syntax that you'll need:

```
plt.bar
plt.xlabel
plt.ylabel
plt.title
```

To set the color of the edges for your bars, include 'edgecolor = 'black'.

```
In [395]: x = ins[ins["score"] > 0]["score"].value_counts().index
          y = ins[ins["score"] > 0]["score"].value_counts()
          plt.bar(x, y, edgecolor = "black")
          plt.xlabel("Score"), plt.ylabel("Count")
          plt.title("Distribution of Inspection Scores")
```

```
Out[395]: Text(0.5, 1.0, 'Distribution of Inspection Scores')
```



0.2.1 Question 5b

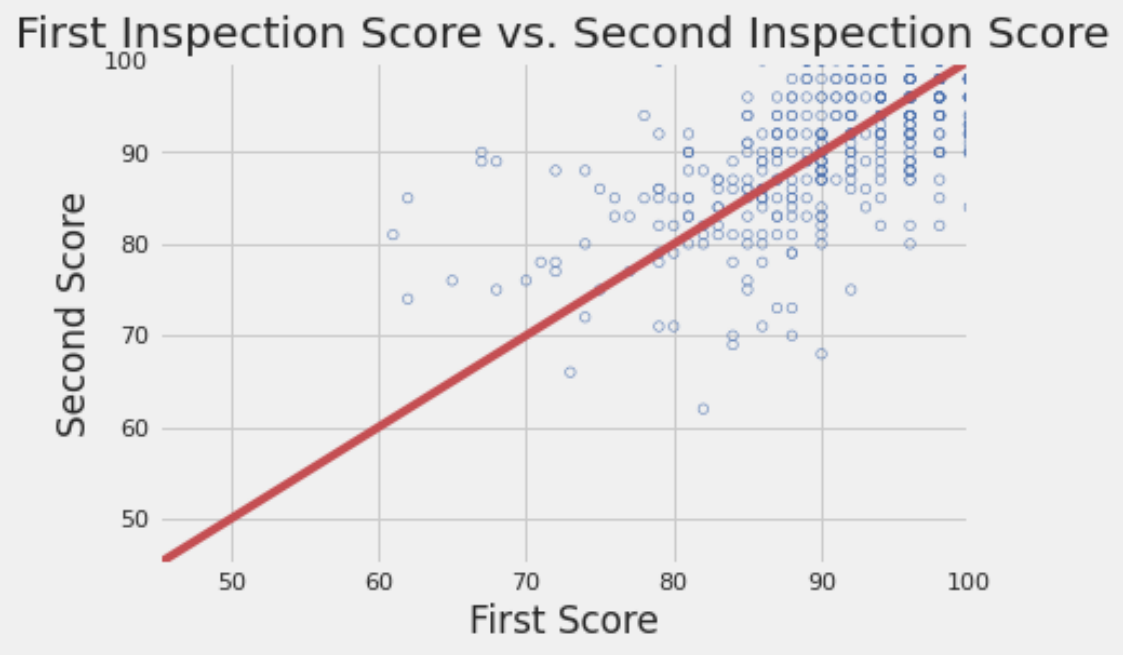
Now let's actually reflect on the histogram that we generated before with a bin size of 1.

Describe the qualities of the distribution of the inspections scores based on your bar plot. Consider the mode(s), symmetry, tails, gaps, and anomalous values. Are there any unusual features of this distribution? What do your observations imply about the scores?

The distribution is skewed to the left and there are gaps between bars starting from 100 up to 80, perhaps, the inspection violations is deducted by even numbers rather than odd numbers, like 2, 4, 6 etc. Also, the distribution has a peak of 100 which is called a unimodel distribution in statistics.

Now let's make a scatter plot to display these pairs of scores. Include on the plot a reference line with slope 1 and y-intercept 0. Since restaurant scores bottom out at 45 points, we'll only focus on ratings between 45 and 100. Thus your reference line should start at `[45, 45]` and go up to `[100, 100]`.

Create your scatter plot in the cell below. It does not need to look exactly the same (e.g., no grid) as the sample below, but make sure that all labels, axes and data itself are correct.



Key pieces of syntax you'll need:

`plt.scatter` plots a set of points. Use `facecolors='none'` and `edgecolors='b'` to make circle markers with blue borders.

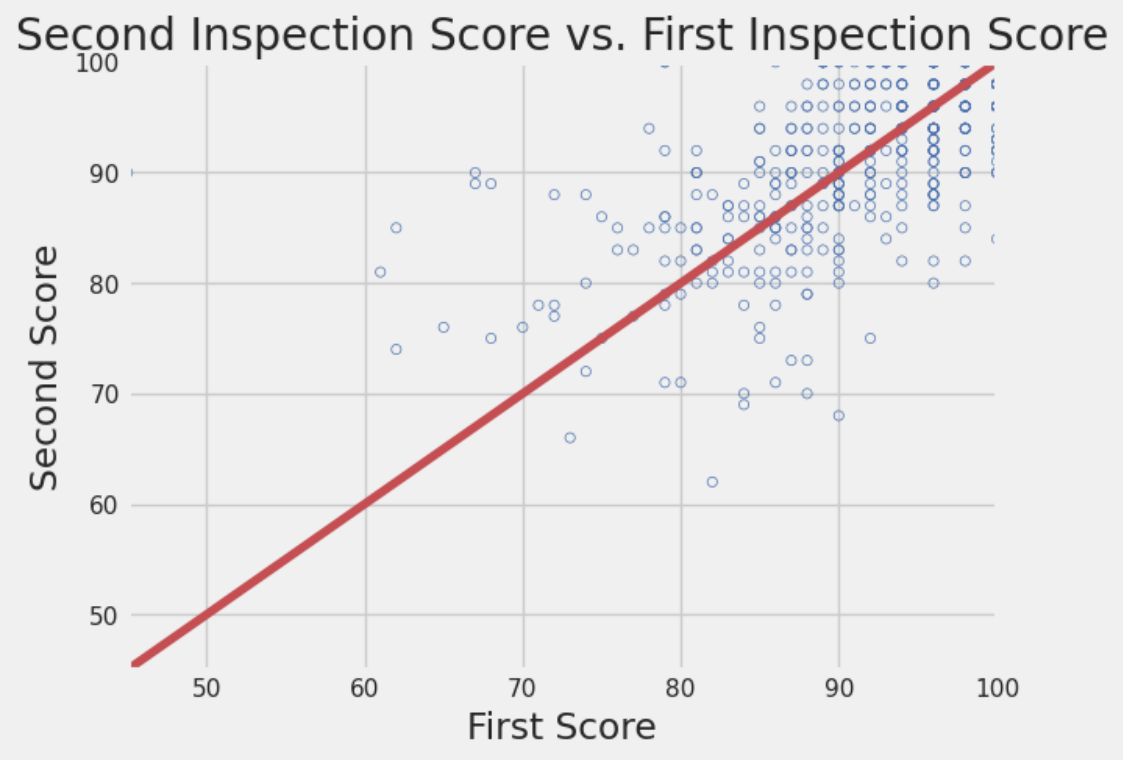
`plt.plot` for the reference line. Using the argument `r` will make the line red.

`plt.xlabel`, `plt.ylabel`, `plt.axis`, and `plt.title`.

```
In [399]: a = scores["first score"]
          b = scores["second score"]
          plt.axis([45,100,45,100])
          plt.plot([45,100],[45,100], 'r')

          plt.scatter(a,b,s = 20, facecolors='none',edgecolors='b')
          plt.title("Second Inspection Score vs. First Inspection Score");
          plt.xlabel('First Score')
          plt.ylabel('Second Score')
```

Out[399]: Text(0, 0.5, 'Second Score')



0.2.2 Question 6c

If restaurants' scores tend to improve from the first to the second inspection, what do you expect to see in the scatter plot that you made in question 6b? What do you oberve from the plot? Are your observations consistent with your expectations?

Hint: What does the slope represent?

If the restaurants scores improved compatered to thier first inspection the points would moved above the line and vise versa. The points are surounded mostly on the top right of the scatter plot which means the most of the restaurants have good score compared to the few whose second scores are below 80. Overall, the scores seem good which means most of the restaurants care about food safety.