EECS16ML: Quiz Solutions

1. C) Pandas Aggregation methods such as .count() and .sum() are much more efficient, as well as readable and easier to document.

2.

- Setting input1 type as float, since all inputs must be integer. Should be np.int.
- Setting return type as None, since the function returns integer. Should be np.int.
- Setting unrelated parameter n, since it is not part of the inputs. Should delete it.

3. D) In terms of efficient documentation of helper functions, function names should be lowercase and separated with underscores for good readability.

4. To complete this task, we can import pdpipe library and use pdpipe.OneHotEncode() on the column specified. This creates a reusable block that can be applied on other categorical variable as well.

5. A) and D) These options would lead to a better representation of the correlation more aligned with the unity line.

6. When working with documentation for methods, it should emphasize why the function is there, not how it works. We know how it works by quickly examining the code, but explaining why the function exists and where it could be applied is the most important aspect of documenting code.

7. A) Bar plots are appropriate to visualize the amounts in each category.

8. B) Box plots give very useful info about the numerical distribution of variables, including lower and upper quartiles as well as outliers.

9. C) df.dropna(axis=1) specifies that any null values on the axis 1, which is the column axis, should be dropped.

10. B) The TokenizeWords() method tokenizes a sentence into a list of tokens by whitespaces, allowing easy extraction of zip codes.

11. In terms of efficient documentation of class, class names should start with a capital letter without separation of words with underscores.

12. C) A heatmap allows us to visualize a colorized map of intensities based on the magnitude of the numerical values.

13. C) The documentation is written in Numpy/Scipy docstrings.

14.

```
>>> !pip install pycodestyle
>>> pycodestyle module.py
```