Assignment 3 Rubric

A screenshot of a black and white screen

Description automatically generated

1.1

# 2 points for using Dataframe

# 4 points for using "unique" or similar to check number of unique 'business id column'

# 4 points for either visually comparing or programatically comparing the unique rows with the size or number of records.

1.2

# 2 points - printing of the top names, printing of the top 3 addresses:

# 2 points for (each) printing the “unique” top business names, and addresses

# 2 points of using numpy array

# 2 points for using a variable named each top\_names, top\_addresses

# 2 point for doing extra data cleaning to find duplicate records that have similar names --

# not all the cleaning needs to be done

1.3

# 5 points for showing a list of address with the count

# 2 points for attempting to fix any of the addresses that seems duplicates

# 3 points for using any extra attempt for finding missing or bad zip codes

1.4

5 points for checking the length of the zip code

2 points for checking that it is a digit.

2 points for setting to None, not the string "None"

1 point for testing for the changes

Question 2.0

10 points for showing just the correct years, 2017 thru 2019.

10 points for setting the figure size by 12 by 8

10 points for generating a box plot.

Question 3.1

# 8 points for creating a database, two tables, and inserting the test data

# extra 1 point for dropping the table if the table already exists.

# extra 1 point for dropping the cats table before the owner

# 2 points for testing that the tables exists by displaying the contents

Question 3.2

There is multiple ways to solve this.

# 5 points for using a valid working SQL

# 3 points for either counting the rows, or using the count() function

# 2 points for showing the correct count

Question 3.3

# there is dataframe or SQL approach

Data frame approach

# 2 points for using pandas and almost no SQL

# 2 points for merging the two dataframes

# 2 points for using some groupby or other dataframe API to find the least cat owner.

# 2 points for showing owner name and ownerID or the calculated least owner

# 2 points for showing progress

SQL approach:

# 4 points for using 1 or more SQL statement without forcing the logic (ie. count < 2)

# 2 points for showing only the 1st record by alphabetical name

# 2 points for showing owner name and ownerID

# 2 points for using ASC technique for the number of cats to get the least

# Connect to CatsDB