



Zesty.ai Analyst Challenge

Part 1: Data Quality

As a property analytics company, property information is the basis of every problem here at Zesty.ai. Valid addresses are the first step to analyzing properties with our machine learning models. These addresses may have a number of issues. For example, an address may have missing or incorrect information, incorrect formatting, or extra information not relevant to that property.

Before any of our models can be run, we need to maximize the **coverage rate**. To maximize the coverage rate, we need to check property addresses are correct and formatted.

Deliverable

The objective is to improve the address coverage rate of test.csv. Addresses may have missing or incorrect information, wrong formats, or extra information.

Write a Python function (or functions) that will parse, clean, and format the addresses to maximize coverage rate. Match test.csv against all_addresses.csv to check the coverage rate. Please comment on what your code is doing.

An example of a correctly formatted address is: *street address, city, state zip (e.g. 710 Florida St, San Francisco, CA 94110)*. For this exercise, a correctly formatted address only has the information above. ANY other address information is not needed to maximize the coverage rate.

Create no more than 3-4 slides describing what techniques you used, what % of the total addresses matched, and how many percentage points you've increased the coverage rate by. Add any additional information you think may be relevant.

Please submit your code and document your sources. Zip all relevant files and upload them to the link you received name "Part 1".

Definitions

coverage rate - the percentage of properties in a file that has correct addresses and matches up with a standard address database

Files

[all_addresses.csv](#) - List of all addresses to be matched against

[test.csv](#) - List of addresses to be cleaned and formatted

Part 2: Z-FLOOD and Risk Analysis

Over the last year, Zesty.ai developed a flood product designed to help insurance companies (our clients) understand the risk of flood for each property they consider insuring. The flood product, named Z-FLOOD, has shown great predictive power and is capable of establishing the following:

1. Flood occurrence score: the risk that a property could be involved in a flood
 - The score of 1 (very low risk - e.g., top of a Mountain) to 10 (very high risk - e.g., right next to a river)
2. Flood damage score: the risk that a home may be damaged if a flood occurs
 - The score of 1 (very low risk - e.g., the house is on pillars) to 10 (very high risk - e.g., deep basement)

Grizzly Insurance is one of Zesty.ai's prospective customers, and they shared their portfolio (collection of policies they insure) to test the product. Zesty.ai scored all of their insurance policies with Z-FLOOD ([LINK](#)) as an output. Grizzly is looking to understand if they should purchase a license for Z-FLOOD. They have asked for you to analyze and present the value of Z-FLOOD.

Deliverable: In a vertical document (Google Sheet or Word) and in a spreadsheet (Excel or Google Sheet), describe the value of Z-FLOOD for Grizzly Insurance. Below are some elements we would like to discuss:

- What does their portfolio look like, and how would it change after implementing Z-FLOOD?
- How can Grizzly Insurance benefit from using Z-FLOOD?
- Explain and quantify the value that Z-FLOOD can have on Grizzly's profitability?
- *What else can Zesty.ai do to convince Grizzly Insurance to adopt Z-FLOOD?*

Please submit your doc and excel and zip all relevant files and upload them to the link you received name "Part 2".