

Python/Data Exam

Instructions

1. Create a folder for each question. Specify the folder name for the question number.
2. Put all folders in a directory. You should compress the directory and send it to us. We expect to receive a zip file from you.
3. Put any additional description in the form of comments in the source code.
4. You should follow the rules for clean coding.

Questions

1. **[General]** In python code, write a function for encoding a given input JSON. The encoder should flatten all the objects to a single key-value dictionary.

Note: Do not use the lib that actually performs this function.

Note: There is no need to consider exceptional cases.

Encode Example:

```
{
  "a": {
    "b": "c",
    "d": "e"
  }
}
```

becomes

```
{
  "a_b": "c",
  "a_d": "e"
}
```

Answer:

```
def encode(in_json):
    ...
    return encoded_json
```

2. **[Algorithm]** You have n characters. Return *the number of possible non-empty sequences of letters* you can make using the characters.

Note: We may have duplicate characters in our input characters.

Note: $1 \leq n \leq 6$

Note: Input characters just consist of uppercase English letters.

Example 1:

- **Input:** "MNM"
- **Output:** 8
- **Sequences:** "M", "N", "MM", "MN", "NM", "MMN", "MNM", "NMM"

Example 2:

- **Input:** "ABC"
- **Output:** 15
- **Sequences:** "A", "B", "C", "AB", "AC", "BC", "BA", "CA", "CB", "ABC", "ACB", "BAC", "BCA", "CAB", "CBA"

Example 3:

- **Input:** "XXYYZ"
- **Output:** 188

3. [Data - Timeseries Analysis]

Input Data: There are two attached files in the "Time series" folder, "electricity_data" and "data documentation."

Electricity_data file contains almost one year of historical electricity price, load, and some other data presented through multiple features from column B:L.

There is a Word document where you can find some information about the above data.

Different features are given to you, and some of them are given by an external source, and some are artificially calculated by our data processing module.

The goal is to analyze electricity data and:

1. Use visualization techniques to find statistical patterns inside the data set.
2. Do bivariate analysis over different features and see if any correlation exists between different features.
3. Design a statistical model to predict the electricity price of a given day if all of the other features are given.
4. Design a statistical model to predict the electricity Load of a given day if all of the other features are given.

Feel free to:

1. Create new features and add them to the data set.

What you need to deliver:

- a. Submit one or more .py/.m or notebook file(s) containing your code for steps 1 to 4.
-
4. **[Web Scraping][Bonus/Optional]** To get the price of electricity in CalifornialSO (CAISO), you can visit [this web page](#). The data on this page is represented as a table that is updated every five minutes. Write code using web scripting to extract table data on this page and store it in the database every five minutes. Please use SQLite as a database. The database must be created in the first run of your program if not exist.