**Instructions:**

The following questions are structured to be completed within 1 hour or less. Please do your best to answer as many questions as possible. Please submit the answers to this question as a `.txt` file.

Important tips:

1. You can assume you are using any version of Python 3.0 or later
2. For Question 2, you can assume the Pandas library is available to use
3. For Questions 1 & 2 make sure you submit complete code that is runnable and not pseudo-code

*Please do not repost these questions in any public forums*

**Question 1:** (Python Skills)

Implement a function in Python, that gives a list of tuples, eg [(1 “a”), (2, “b”), (3, “c”), (4, “a”), ((100, “e”)], the function returns a dictionary, where the keys in the dictionary are the first element in each tuple. Note, that the first element in each tuple could repeat in the list.

Input:

[(1, “a”), (2, “b”), (1, “c”), (4, “a”), ((100, “e”)]

Expected Output:

{1: [“a”, “c”],

2: [“b”],

4: [“a”],

100: [“e”]

}

Provide the Big-O cost for the function you implement.

**Question 2:** (Data Manipulation and Intuition for data)

Given a CSV file that has the following data sample data:

| currency | transaction\_amount | legal |
| --- | --- | --- |
| btc | 1000.105001 | True |
| BTC | 2000.111199 | True |
| btc | 5000.112111 | False |
| eth | 100.011111 | True |
| eth | nan | True |
| eth | 200 | False |

Write a Python function, that when provided a CSV file with the data above, returns the SUM of all legal transactions by currency. Document any assumptions you are making with the data.

(You are free to use pandas or an equivalent library for this question if you choose to)

**Bonus**: Give the sum of all legal transactions. Document any assumptions you are making while answering that question.

**Question 3:**

Briefly answer the following. Please make sure you provide concrete examples for the answers.

1. Provide an example of a time you had performed a feature engineering task and how it improved the model's performance.
2. List 2 to 3 examples of metrics you have monitored in your previous ML projects while in production and an example of how you went about improving the metrics.