Ques 1 - Basic Cryptanalysis

Algorithm:

Please note that although the question pdf uses small letters for both plaintext and ciphertext, I have used capital letters for plaintext and small letters for ciphertext, so that it's easier to understand what the respective mappings will be.

1. First, for a given cipher text, I calculated the following frequencies:
    a. I know that in english, the most frequent letters are: E, T, A, O, I. For cipher text 1, I obtained the following frequencies (reporting a max of 10 here): [[61, 'q'], [35, 'y'], [34, 'n'], [31, 'v'], [30, '@'], [27, 'r'], [25, '1'], [21, 't'], [21, '$'], [19, 'p']]
    b. Most frequent 2 letter words are: OF, TO, IN, IT, IS. For cipher text 1, I obtained the following frequencies (reporting a max of 10 here): [[9, 'nq'], [2, '@$'], [1, 'y$'], [1, 'r@'], [1, '$t']]
    c. Most frequent 3 letter words are: THE, AND, FOR, ARE, BUT. For cipher text 1, I obtained the following frequencies (reporting a max of 10 here): [[6, '51v'], [3, 'nrv'], [2, 'y$@'], [2, '7#@'], [2, '144'], [1, 'zrz'], [1, 'w$t'], [1, 'vqq'], [1, '@nq'], [1, '7qz']]
    d. Most frequent bigrams are: TH, ER, ON, AN, RE. For cipher text 1, I obtained the following frequencies (reporting a max of 10 here): [[16, 'nq'], [16, '@n'], [15, 'ry'], [14, 'yp'], [9, 'tq'], [8, 'qt'], [8, 'nr'], [7, 'qz'], [7, '1v'], [6, 'vq']]
    e. Most frequent trigrams are: THE, AND, THA, ENT, ION. [[14, 'ryp'], [8, '@nq'], [6, '51v'], [4, 'qtq'], [4, 'q51'], [4, 'nry'], [4, 'nqt'], [4, 'nq5'], [4, '@nr'], [3, 'ypn']]
    f. Most frequent doubles are: SS, EE, TT, FF, LL, MM, OO. For cipher text 1, I obtained the following frequencies (reporting a max of 10 here): [[6, 'qq'], [3, '44'], [2, '@@'], [1, 'zz'], [1, 'yy'], [1, 'vv'], [1, 'pp'], [1, '$$']]
    g. All these statistics for cipher text #2 are present in the screenshot.
2. If we go by only the statistics as I have mentioned above, we will get a wrong result, and that is mainly because:
    a. These are just statistics, and there is no guarantee that they will always hold true.
    b. The ciphertext is pretty small, and so the reported statistics for this particular cipher text are not a good representation of the actual world.
3. So, after displaying the statistics, the user can input which cipher character should be matched to which plain character.
4. It may happen that the ciphertext set has some unused characters, which may remain unused. And if there are more than 1 of these, it will be impossible to determine which plain text character they map to.

Answers:

1. **Ciphertext 1:**
    a. **Mapping**: {'A': 'v', 'B': '1', 'C': 't', 'D': '8', 'E': '9', 'F': 'p', 'G': '7', 'H': 'q', 'I': '5', 'J': None, 'K': 'w', 'L': 'u', 'M': '0', 'N': '@', 'O': '$', 'P': '3', 'Q': '4', 'R': '2', 'S': 'y', 'T': 'o',

'U': 's', 'V': '6', 'W': '#', 'X': 'r', 'Y': 'x', 'Z': None}
Note that there is no usage of n and z. So, we can't find the mapping from {J, Z} to {n, z}.

b. **Plaintext**: A DISADVANTAGE OF THE GENERAL MONOALPHABETIC CIPHER IS THAT BOTH SENDER AND RECEIVER MUST COMMIT THE PERMUTED CIPHER SEQUENCE TO MEMORY. A COMMON TECHNIQUE FOR AVOIDING THIS IS TO USE A KEYWORD FROM WHICH THE CIPHER SEQUENCE CAN BE GENERATED. FOR EXAMPLE, USING THE KEYWORD CIPHER, WRITE OUT THE KEYWORD FOLLOWED BY UNUSED LETTERS IN NORMAL ORDER AND MATCH THIS AGAINST THE PLAINTEXT LETTERS. MAKE REASONABLE ASSUMPTIONS ABOUT HOW TO TREAT REDUNDANT LETTERS AND EXCESS LETTERS IN THE MEMORY WORDS AND HOW TO TREAT SPACES AND PUNCTUATION. INDICATE WHAT YOUR ASSUMPTIONS ARE. NOTE, THE MESSAGE IS FROM THE SHERLOCK HOLMES NOVEL, THE SIGN OF FOUR.

2. Ciphertext 2:

a. **Mapping**: {'A': '1', 'B': '7', 'C': '3', 'D': 'z', 'E': 'q', 'F': 'w', 'G': 'p', 'H': 'n', 'I': 'r', 'J': 'x', 'K': '6', 'L': '4', 'M': 's', 'N': 'y', 'O': '$', 'P': '8', 'Q': None, 'R': 't', 'S': 'v', 'T': '@', 'U': '#', 'V': '9', 'W': '5', 'X': None, 'Y': '2', 'Z': None}
Note that there is no usage of 0, o and u. So, we can't find the mapping from {Q, X, Z} to {0, o, u}

b. **Plaintext**: DEFEATED AND LEAVING HIS DINNER UNTOUCHED, HE WENT TO BED. THAT NIGHT HE DID NOT SLEEP WELL, HAVING FEVERISH DREAMS, HAVING NO REST. HE WAS UNSURE WHETHER HE WAS ASLEEP OR DREAMING. CONSCIOUS, UNCONSCIOUS, ALL WAS A BLUR. HE REMEMBERED CRYING, WISHING, HOPING, BEGGING, EVEN LAUGHING. HE FLOATED THROUGH THE UNIVERSE, SEEING STARS, PLANETS, SEEING EARTH, ALL BUT HIMSELF. WHEN HE LOOKED DOWN, TRYING TO SEE HIS BODY, THERE WAS NOTHING. IT WAS JUST THAT HE WAS THERE, BUT HE COULD NOT FEEL ANYTHING FOR JUST HIS PRESENCE.