**Python Question Bank**

**Q.1:**

There is a hotel chain of various branches at different places in the country. The guests are as per the subscription policy.(eg: Mahindra holiday homes).

Guests have been having chats with the hotel hosts. Need to provide rank to the guests based on the number of messages they've exchanged with the hosts. If the total no of messages of the guests are same, they would have the same rank.

Pandas Dataframe to be taken as input and as output expected.

**csv:** with data of guest id,date of interaction and host id of hotel.

**Output:** The output should contain rank, guest id, and number of total messages they've sent.

**Case 1:** Do not skip rankings if the preceding rankings are identical.

**Q.2:**

For each postal code, find the different number of streets names. The first part of the street name should be counted for a longer/multiple street name. The result should display the postal code and the count of number of streets in descending(or ascending) order.

(Complexity can be increased to make the result with descending count of streets and ascending postal code.)

Pandas Dataframe to be taken as input and as output expected.

**Csv**: The dataset has to be provided for the given area.

**Case 1:** Old Mahaballipuram road and mahaballiburam road should be considered as old Mahaballipuram road.

**Case 2:** case sensitivity should be checked. TAZHAMBUR ROAD and tazhambur road should be considered the same.

**Q.3:**

From a pdf or long text, extract the dates. The dates can be in different formats (dd-mm-yyyy,dd-mmm-yyyy,dd-mm-yy – different combinations can be provided) .

Extract the dates along with the page numbers.

**Input:** A pdf/Strings/word document

**Output:** Pandas Dataframe

**Case 1:** Format the dates in a particular format (dd-mm-yyyy) and display in ascending order along with the page numbers it occurs.

**Case 2:** Also replace the original text with the corrected date format.

**Q.4:**

There are sighting of tiger in the national parks recorded for over the period of 100 years in a country. The dataset is vast. Can be around 70k-80k

**Input:** CSV/SQL TABLE

**Output:** Pandas Dataframe

**Case 1:** Calculate todays date using pandas and from todays date calculate the period of the sighting days. Find out the oldest date of the sighting date. Calculate the dates which are older than 50 years.

**Case 2:** Calculate the sightings based on the two specific dates.

**Case 3:** Draw a scatter plot of the sighting/national park between the two specific dates.

**Q.5**

Based on the language, find the number of users with Apple and Samsung.

Apple products can be set of iPad,MACBook and phone.

Samsung products can be the phones

**Input:** CSV/SQL TABLE

**Output:** Pandas Dataframe

**Case 1:** Output the data along with the number of users based on the two specific dates or between today and a specific data. The date has to be grouped by the language.

**Q.6:**

The dataset contains the rainfall of the districts in a state alongwith the humidity,day and night temperature. The rainfall can be negative too.

**Input:** CSV/SQL TABLE

**Output:** Pandas Dataframe

**Case 1:** Check for the missing data and fill with an avg of min and max rainfall.

**Case 2:** Get the flattened array of the rainfall. Generate a random number and partition the array based on the random number position. Move all the smaller values to the left of the partition and the remaining values to the right.

**Q.7:**

Data contains the opening price, low price, high price, trading volume of a ABC Ltd. company .

**Input:** CSV/SQL TABLE

**Case 1:** Write the code to draw a bar plot of trading volume.

**Case 2:** Write the code to create histograms plot of opening,closing,high,low stocks prices of ABC Ltd. Between two specific dates.