Name: Chaoyu Li
Date: 1/27/2022

Explanation of HW1

1. For Chaoyu_Li_load.py:

In this script, I downloaded the dataset “WA_Fn-UseC_-Telco-Customer-Churn.csv” manually, then use Pandas to transfer it to json format. After I get the json dataset, I insert an index for each customer, then load it to Firebase. Therefore, I only sent one “put” request to Firebase.

```
def upload_data(url, data):
    try:
        putResponse = requests.put(url + '.json', data)
        if putResponse.status_code == 200:
            print("Upload data Successfully")
        else:
            print("Upload data failed, Reason: {}".format(putResponse.text))
    except:
        print("Upload data failed")
```

2. For Chaoyu_Li_churn.py:

In this script, I add “`?orderBy="Churn"&equalTo="Yes"&print=pretty`” to the URL of request. It avoids downloading the whole dataset for one query and for each query, it only send one “get” request to Firebase to get all customers information whose “Churn” shows “Yes”. Then I sorted them by “customerID”.

```
baseURL = 'https://dsci551-hw1-89e4f-default-rtdb.firebaseio.com/'

def query_data(path):
    getURL = baseURL + '/' + path + '.json'
    response = requests.get(getURL + '?orderBy="Churn"&equalTo="Yes"&print=pretty')
    return response.json()
```

3. For Chaoyu_Li_tenure.py:

It is very similar with Chaoyu_Li_churn.py and I only change the filter in the URL to "`'?orderBy="tenure"&startAt=' + k + '&print=pretty'`", here "k" stands for the input "k". Therefore, similarly, for each query, it only sends one "get" request to Firebase.

```
def query_data(path, k):
    getURL = baseURL + '/' + path + '.json'
    response = requests.get(getURL + '?orderBy="tenure"&startAt=' + k + '&print=pretty')
    return response.json()
```