

DSCI 551 – Spring 2022

Homework #1: Firebase, JSON, and Data Modeling

Deadline: February 4, Friday (100 points)

Consider managing customer churn data in Firebase realtime database. The data are stored in a CSV file, with 7044 rows and 21 columns. You can find the details about the data set at Kaggle web site: <https://www.kaggle.com/blastchar/telco-customer-churn>. You can also download the data set from the web site (archive.zip containing WA_Fn-UseC_-Telco-Customer-Churn.csv).

To reduce the amount of data to be handled, this homework will only consider customers who are senior citizens (1142 of them).

Tasks:

1. [40 points] Write a python script “load.py” which load the rows for the above senior customers to your database.

Execution format:

```
python3 load.py
```

You can assume the WA_Fn-UseC_-Telco-Customer-Churn.csv file is stored at the same directory where you execute your script.

2. [30 points] Write a Python script “churn.py” to find the first k (senior) customers who has churned. Only need to return IDs of first k customers (ordered by their IDs).

Execution format:

```
python3 churn.py <k>
```

For example:

```
python3 churn.py 10
```

will return IDs of first 10 customers who have churned.

3. [30 points] Write a Python script “tenure.py” to find out how many customers who have used the service for at least k months.

Execution format:

```
python3 tenure.py <k>
```

For example,

```
python3 tenure.py 10
```

Requirements:

- For each query in both patterns, only one round trip (send request and receive response) is permitted to the Firebase server.
- You should not download entire database to answer the query.
- You should create indexes in Firebase console that allow the above programs to execute without errors.

Permitted libraries: pandas, requests, json, and other common Python libraries (e.g., sys). Do not use firebase-admin, firebase python libraries.

Submissions:

- Above 3 scripts.
- Prepend your full name to the script name, e.g., John_Smith_load.py, so on.
- A document (word/pdf) explaining why your program sends only one request to Firebase for each query.
- A JSON dump of your Firebase database for this app.
- A screenshot of your Firebase, showing the structure of your database.
- Please make sure your codes can run on EC2 using python3 (3.7 on EC2), your codes will be tested on EC2.
- You can submit multiple times but only the latest one within the deadline will be graded.
- Submit online. See syllabus for late penalty!