

Homework 1: Firebase

DSCI 551 – Fall 2024

Due: 11:59pm, September 20, 2024, Friday

Points: 100

In this homework, we will use Firebase real-time database to manage a coffee sales dataset (coffee_shop_sales.csv). You can find more information about the data set here:

<https://www.kaggle.com/datasets/ahmedabbas757/coffee-sales?resource=download>

Please use the dataset provided along with this handout, above link is just for reference.

Complete the homework by writing the below methods in the given template.

The homework consists of the following tasks.

- [20 points] Write a Python function "load" that loads the transactions in the csv file to Firebase. **All the transactions can be formatted and can be inserted with a single request.**

Execution format:

```
python3 <yourname>_hw1.py load coffee_shop_sales_2023.csv
```

- [20 points] Write a Python function "search_by_id" that takes an id and returns the details of the transaction with the given id. The output should be in the JSON format.

- **Requirement:** Do not query the entire database. You should only query one requested transaction. Zero points otherwise.

Execution format:

```
python3 <yourname>_hw1.py search_by_id <id>
```

where <id> is the id of a transaction.

- [30 points] Write a Python function "index" that takes the csv file, creates an index on the "product_type" column, and stores the index on your Firebase database. Your index should have an entry for each unique keyword in the column and for each keyword, list ids of transactions that contain the keyword.

Execution format:

```
python3 <yourname>_hw1.py create_index coffee_shop_sales_2023.csv
```

- [30 points] Write a Python function "search_by_keywords" that takes a list of keywords and returns the ids of transactions whose product type contains all the keywords.

- **Requirement:** Do not query the entire index. You should only query one keyword at a time. Zero points otherwise.

Execution format:

```
python3 <yourname>_hw1.py search_by_keywords "herbal tea"
```

Permitted libraries: pandas, json, requests, sys

SUBMISSION INSTRUCTIONS:

1. A single python file with name: **[Student_Name]_hw1.py** [replace Student_Name with your name] Eg. John_Smith_hw1.py
2. Do not copy paste the commands from handout to terminal. Please rewrite the commands in the terminal. PDF format will encode special characters which are different from UTF-8 encoding in the terminal.
3. Do not modify any contents in the template. Just fill the template by reading the comments.
4. The test script will accept the return data same as specified in the template.
5. Testing is done by test script with different test cases. So points will only be awarded if the method returns the expected result.
6. You will get 0 points if the code breaks for any syntax errors or any other problems. Please test the code thoroughly before submitting.
7. **Please delete the data in firebase when submitting the HW. The database should be empty without any data [keys and values].**