

```
In [88]: import json
from pathlib import Path
import os

import pandas as pd
import s3fs

def read_cluster_csv(file_path):
    #file_path = ("C:\Users\spashtunyar\Documents\School\dsc650\data\external\tidynomi
    return pd.read_csv(open(file_path, mode='rb'))
#commenting out this part, the full file path is input and that works
current_dir = Path(os.getcwd()).absolute()
results_dir = current_dir.joinpath('results')
kv_data_dir = results_dir.joinpath('kvdb')
kv_data_dir.mkdir(parents=True, exist_ok=True)

people_json = kv_data_dir.joinpath('people.json')
visited_json = kv_data_dir.joinpath('visited.json')
sites_json = kv_data_dir.joinpath('sites.json')
measurements_json = kv_data_dir.joinpath('measurements.json')
```

```
In [89]: class KVDB(object):
    def __init__(self, db_path):
        self._db_path = Path(db_path)
        self._db = {}
        self._load_db()

    def _load_db(self):
        if self._db_path.exists():
            with open(self._db_path) as f:
                self._db = json.load(f)

    def get_value(self, key):
        return self._db.get(key)

    def set_value(self, key, value):
        self._db[key] = value

    def save(self):
        with open(self._db_path, 'w') as f:
            json.dump(self._db, f, indent=2)
```

```
In [90]: sites_json
```

```
Out[90]: WindowsPath('C:/Users/spashtunyar/Documents/School/dsc650/dsc650/assignments/assignment02/results/kvdb/sites.json')
```

```
In [91]: people_json
```

```
Out[91]: WindowsPath('C:/Users/spashtunyar/Documents/School/dsc650/dsc650/assignments/assignment02/results/kvdb/people.json')
```

```
In [92]: visited_json
```

```
Out[92]: WindowsPath('C:/Users/spashtunyar/Documents/School/dsc650/dsc650/assignments/assignment02/results/kvdb/visited.json')
```

```
In [93]: measurements_json
```

```
Out[93]: WindowsPath('C:/Users/spashtunyar/Documents/School/dsc650/dsc650/assignments/assignment02/results/kvdb/measurements.json')
```

```
In [94]: def create_sites_kvdb():
    db = KVDB(sites_json)
    df = read_cluster_csv(r'C:\Users\spashtunyar\Documents\School\dsc650\data\external')
    for site_id, group_df in df.groupby('site_id'):
        db.set_value(site_id, group_df.to_dict(orient='records')[0])
    db.save()

def create_people_kvdb():
    db = KVDB(people_json)
    df = read_cluster_csv(r'C:\Users\spashtunyar\Documents\School\dsc650\data\external')
    for person_id, group_df in df.groupby('person_id'):
        db.set_value(person_id, group_df.to_dict(orient='records')[0])
    db.save()

def create_visits_kvdb():
    db = KVDB(visited_json)
    df = read_cluster_csv(r'C:\Users\spashtunyar\Documents\School\dsc650\data\external')
    for key, group_df in df.groupby(['visit_id', 'site_id']):
        db.set_value(str(key), group_df.to_dict(orient='records')[0])
    db.save()

def create_measurements_kvdb():
    db = KVDB(measurements_json)
    df = read_cluster_csv(r'C:\Users\spashtunyar\Documents\School\dsc650\data\external')
    for key, group_df in df.groupby(['person_id', 'visit_id', 'quantity']):
        db.set_value(str(key), group_df.to_dict(orient='records')[0])
    db.save()
```

```
In [95]: create_sites_kvdb()
create_people_kvdb()
create_visits_kvdb()
create_measurements_kvdb()
```

```
In [96]: kvdb_path = 'visits.json'
kvdb = KVDB(kvdb_path)
key = (619, 'DR-1')
value = dict(
    visit_id=619,
    site_id='DR-1',
    visit_date='1927-02-08'
)
kvdb.set_value(key, value)
retrieved_value = kvdb.get_value(key)
```

```
In [ ]:
```