```
## Import Necessary Packages ##

import pandas as pd
import pickle
import zlib
import json
import redis
from pyspark import SparkConf, SparkContext
from pyspark.sql import SparkSession, SQLContext
spark = SparkSession.builder.master("local[*]").appName("Womes-Shoe").getOrCreate()

sc = spark.sparkContext
sqlContext = SQLContext(spark.sparkContext)

## Read Input file from Hadoop ##

data = 'hdfs:///tmp/7210_1.csv'
data1 = 'hdfs:///tmp/Datafiniti_Womens_Shoes.csv'
data2 = 'hdfs:///tmp/Datafiniti_Womens_Shoes_Jun19.csv'

data_df = spark.read.csv(path=data, header=True, inferSchema='true').cache()
data_df1 = spark.read.csv(path=data1, header=True, inferSchema='true').cache()
data_df2 = spark.read.csv(path=data2, header=True, inferSchema='true').cache()

new_df = data_df.select("id", "dateAdded", "brand", "colors")
new_df1 = data_df1.select("id", "dateAdded", "brand", "colors")
new_df2 = data_df2.select("id", "dateAdded", "brand", "colors")

final_df = new_df.where((new_df.brand != '') | (new_df.colors != '') | (new_df.dateAdded != ''))
final_df1 = new_df1.where((new_df1.brand != '') | (new_df1.colors != '') | (new_df1.dateAdded !=
''))
final_df2 = new_df2.where((new_df2.brand != '') | (new_df2.colors != '') | (new_df2.dateAdded !=
''))

EXPIRATION_SECONDS = 14400
combined = final_df.union(final_df1.union(final_df2))
final_df1 = combined.toPandas()
print(final_df1.count())
f = pickle.dumps(final_df1)

## REDIS ##
```

```
r = redis.StrictRedis(host='localhost', port=6379)
r.setex("WomensShoesList", EXPIRATION_SECONDS,  zlib.compress(f))
output = pickle.loads(zlib.decompress(r.get("WomensShoesList")))

print(output)
```