Dear Sir/Madam

I enjoyed doing this exercise. The program fetches csv files from s3 and transforms each table.

It joins the relevant tables together based on their unique keys

The result of the above is a one big table – this is normally large in a production environment.

I dealt with some data type conversions and new column creations like “year” and “week” which are not included in the specification but relevant when doing the analyses.

I figured out the relationships between the data and was able to connect them together.

At the end of the program, it sends the cleansed data to AWS bucket as shown below.

I have got the same skills in Both Spark/Pyhton

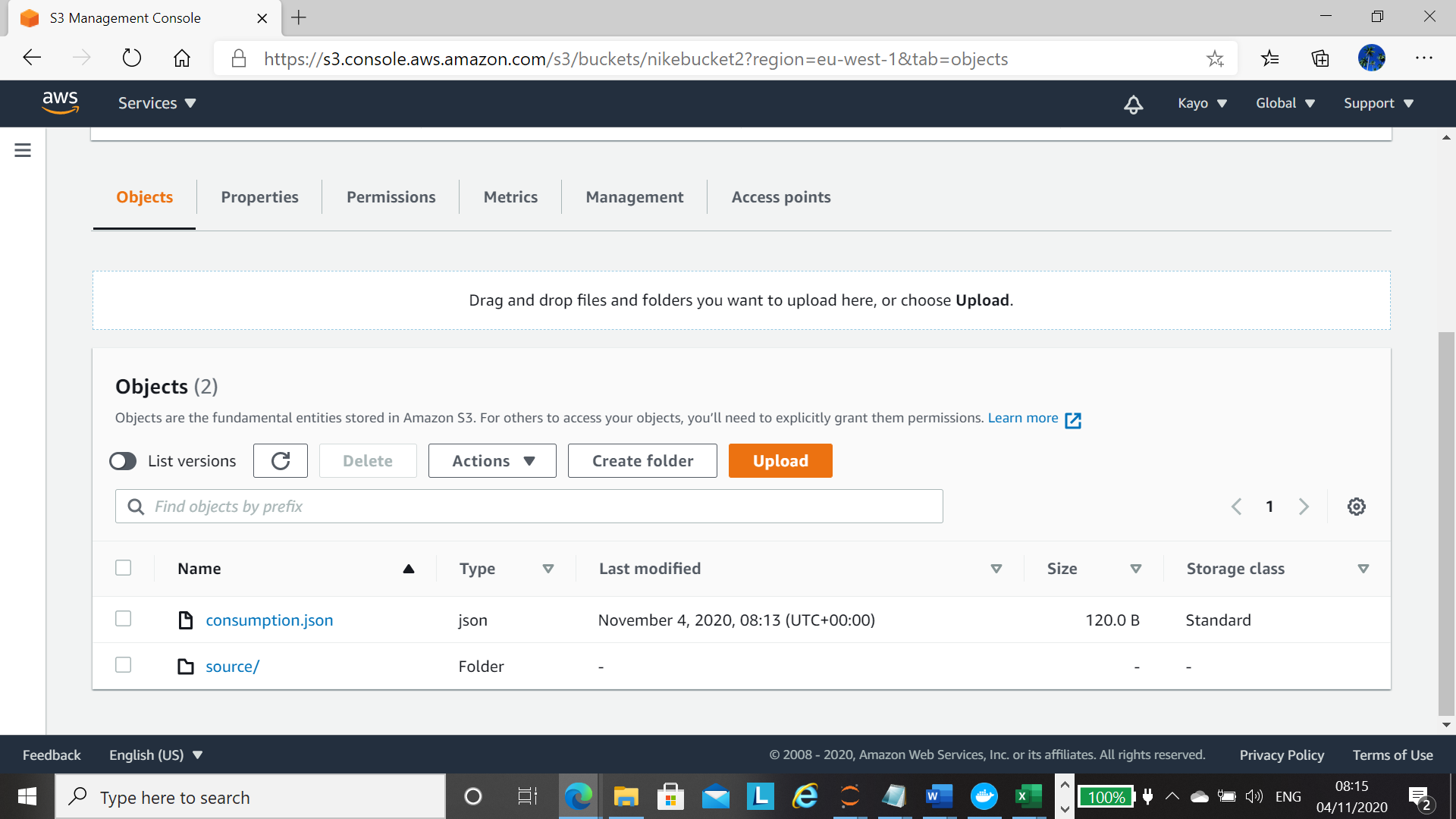
**Improvement:**

Engineer never get tired of looking for new ways to enhance existing application

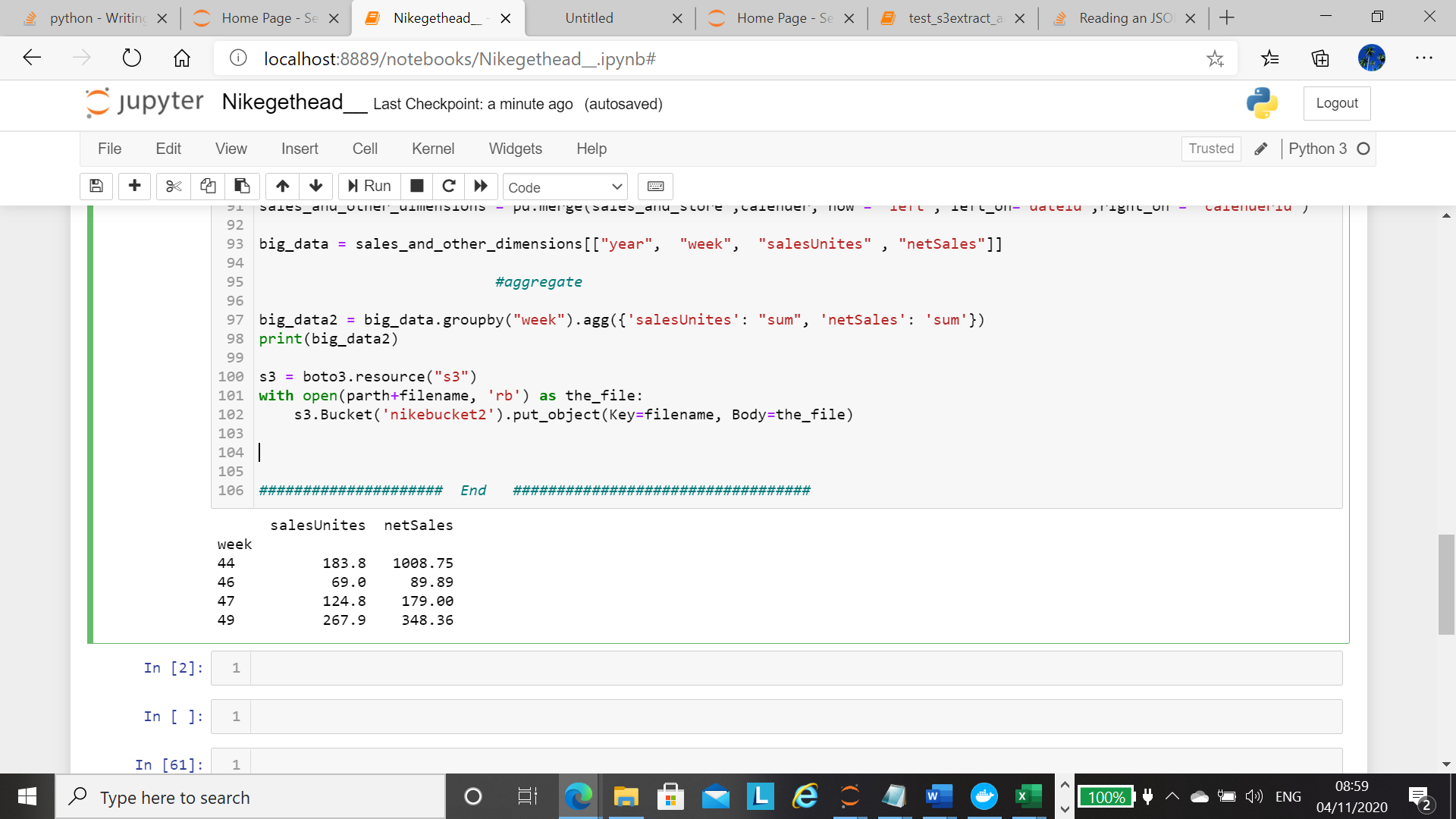
As we find new this to on the program

I have not really looked in looping, functions, or class to perform this job.

I could place some of the repeated lines and make them dynamic depending on the values supply to the method/function



Output from my script.



Hi Re

**Conceptual data model of the database**

Store

Product

Sales

Calender