

# Project

Tuesday, November 13, 2018 7:34 PM

## Configure and load MS SQL Server

(Instructor, review using SQL Server here.)

1. From the Start menu, start Microsoft SQL Server Management Studio.
2. Right-click on the Databases folder and select *New Database*. Type in *CycleMikes\_Staging* for the name of the database.
3. Import data into the CycleMikes\_Staging database. Open the following zip file and walk through the steps detailed in the Importing a Single Csv section on the csv files contained within:



## Importing a Single Csv

- a. Right-click on the CycleMikes\_Staging database, select Tasks, Import Data...
- b. On the Choose a Data Source dialog, select Flat File Source as a data source. Click browse and select the source csv (you will have to change the extensions to find it). Leave the rest of the selections alone, but do click on the Columns selection on the left-hand side and verify everything looks correct.
- c. On the Choose the Destination Dialog, select Microsoft OLE DB Provider for SQL Server. Choose CycleMikes\_Staging as the database.
- d. Click 'Next' all the way to the end.

## Create the Data Warehouse

1. Create a new database called CycleMikes.
2. Open the following script in Sql Server Management Studio and run it.



fill\_cycle\_...

## Create the Python Script to Generate the Missing Data

With the previous section, you saw several errors while trying to generate the data warehouse tables. The problem is that we're missing the Europe tables in the staging area. Create a Python script to generate three csv files with the European data, retrieving the data from the mongo database. Import those csv files as you did with the headquarters data. Creating the script is the bulk of the project.

The script will need to:

1. **Discard the mongo-generated object id.** To see some object ids, run `db.Sales.find()`; from the mongo shell. The first piece of data is the object id (e.g. `"_id" : ObjectId("5c2e7d924a5e1819e040e97c")`).
2. **Strip commas from data values.**
  - a. You'll see that some of the data items have commas in them. To see one, run this command from within the mongo shell: `db.Sales.find({AddressLine1:"304, rue des Rosiers"})`;
  - b. Since we are creating a .csv, or a comma-separated values file, the commas within data values get confused with the commas delimiting the separate data values.
  - c. The commas within the data need to be removed.
3. **Add in the country codes.** In order for our data warehouse to work correctly, the project lead mentioned that we need to have a column titled 'CountryCode' whose data values contain the two-letter country code. If the country is "Germany", the code is "DE". For France the code is "FR", and for the "United Kingdom" the code is "GB".

## Import the European data

After creating the 3 new csv files by extracting the data from mongo, import them into SQL Server. Then run the fill\_cycle\_mikes.sql script from within SQL Server. If done correctly, you should not see any

errors.