1.Discuss your solution's time complexity. What tradeoffs did you make?

Regarding time complexity, I made some tradeoffs in this solution. One of them is the use of pandas for simplicity, although it can be slow with big datasets. To handle larger datasets, it may be necessary to either process the data without using pandas or use a more specialized library such as pyspark. Another tradeoff is that the merge operation using pandas can be slow, and to make it easier to modify in the future, I added indexes to the tables before merging. However, this approach may not be efficient with large datasets.

In general, the time complexity of a merge operation in pandas with an inner join can be expressed as O(m * n), where m is the number of rows in the left dataframe and n is the number of rows in the right dataframe. This is because pandas needs to compare every row in the left dataframe with every row in the right dataframe to identify the matching rows.

However, in practice, pandas may use indexing or other optimization techniques to achieve better performance, resulting in a lower time complexity. For example, if the dataframes being merged are sorted on the join key, pandas can use a more efficient algorithm to perform the merge, resulting in a time complexity of O(m + n) instead of O(m * n). So, in this assignment was done 4 merges in indexes : O(m+n+p+q), where m,n,p and q are the size of each dataframe. The first question add two lambdas with time complexity of O(n) + O(n) and a groupby with time complexity of O(m*n), where m is the number of unique groups and n is the number of rows.
The Second question add a groupby and a map, the map in small dataframe is insignificante and the groupby depends of many variants like the first question.
We can simplify the time complexi of this assignment to O(m+n+p+q).

2.How would you change your solution to account forfuture columns that might be requested, such as "Bill Voted On Date" or"Co-Sponsors"?

To account for future columns that might be requested, such as "Bill Voted On Date" or "Co-Sponsors," I believe the current structure of merging all CSVs into a unified table and working with that table should be sufficient. However, some adjustments may be necessary depending on the specific requirements of the new columns.

3. How would you change your solution if instead ofreceiving CSVs of data, you were given a list of legislators or bills that you should generate a CSV for?

If given a list of legislators or bills to generate a CSV for instead of receiving CSVs of data, some changes would be required to transform the data into a dataframe. However, the overall structure of the solution would remain the same..

4. How long did you spend working on the assignment?

It took me a total of 4.5 hours to complete the assignment.