Write-Up

1. List of team members: Michal Golovanevsky, Brooke Hanna, Anish Yakkala.
2. Initial decisions: Python, Jupyter Notebook and terminal (for testing purposes).
3. Internal architecture: Pandas Dataframe (two instead of one) for easier access to subsets based on the filtering and or grouping specifications and also to assist us with clear table visualization of the data.
4. Task log: we completed this task during the given lab periods on 04/08 and 04/10 working in a team around one computer and rotating who is typing.
5. Testing: we tested our functions in the Jupyter notebook after writing each function to ensure proper functionality, especially with the new functionalities that differed from part 1. We found minor bugs that we fixed while we were writing the functions. Jupyter notebook made it easy to do step by step testing without needing to compile everything at once. Then we used the command line to test that our input/output looks correct to the user and performs the correct computations.
6. Final Notes: We would recommend doing this task using pandas, as it made it easier to complete and made our work more organized and straight forward. Note that we used the functions that are written in jupyter and are using pandas to see what our output should be when testing functions such as the average.
7. Modifying Part I: We used two data frames instead of one data frame in our main when we read the files. Then we merged the two data frames so we could use the same code.
8. We used groupby() and pivot tables in NR5 and in other places we use basic slicing and subsetting of data frames. For example, for NR5 the user can input any valid categorical

column and then the output will be a pivot table aggregated by the specified column showing the mean in variance. Then for NR1 through NR3 we subsetted the dataframe by the given classroom number or grade and outputted the desired information. Finally, for NR4 we looped through the classrooms and outputted the number students.