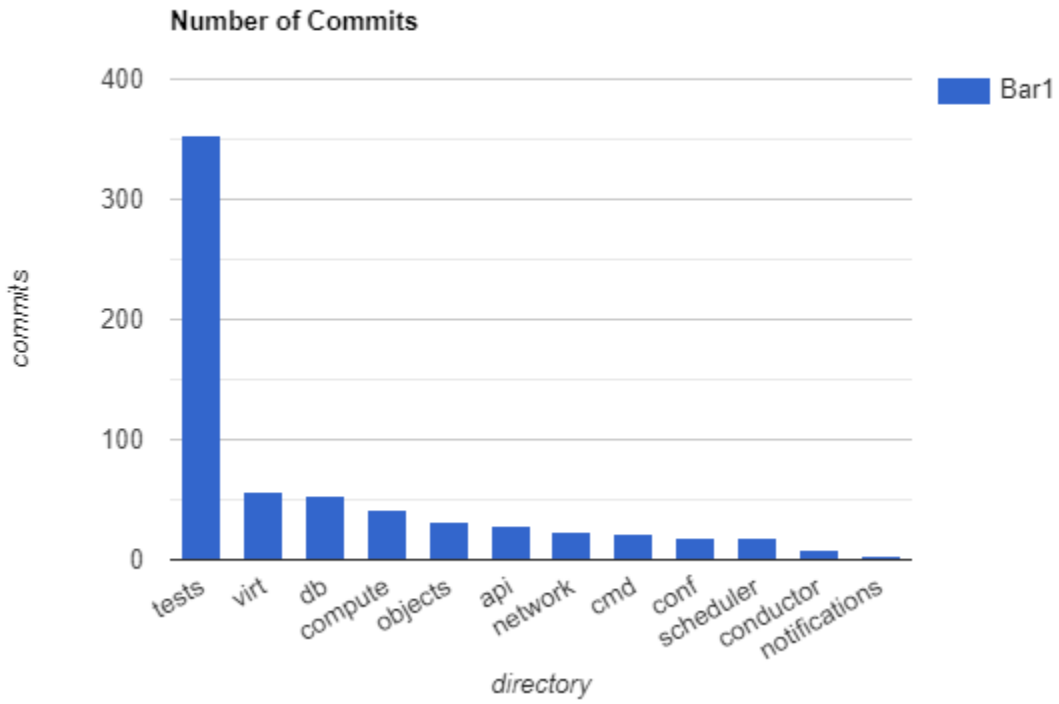
To analyse the activity in the open-source Nova project(https://github.com/openstack/nova), I had used GitPython. The task was to identify the twelve most active subdirectories within the 'nova/' subdirectory. We have analysed the activity in two ways. They are:

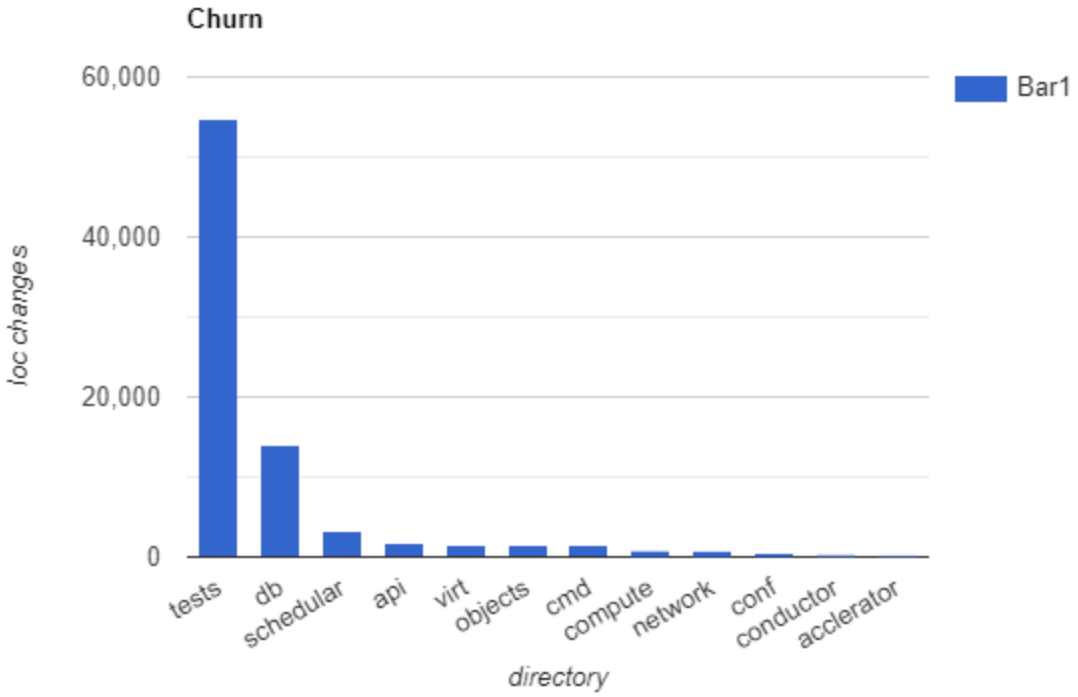
1. Number of commits
2. Code churn

To get details of the subdirectories with the highest number of commits, I had used an internal method, 'iter_commits', to get the commits in the last six months. The results of the activity are depicted in the graph below.



We can see that the most number of commits were made to make changes in the files within the 'nova/tests' subdirectory. The number of commits changing the code within the files of other subdirectories can also be seen from the graph.

To get the code churn that resulted from the commits in the last six months, I used the property of the 'commit' object, namely, 'stats'. The object returned contains the number of lines of code affected by each commit. I wrote code to extract the churn corresponding to one directory at a time. The results of the operation are shown in the graph below.



It can be seen that most churn occurred within only three subdirectories, namely, ‘nova/tests’, ‘nova/db’ and ‘nova/scheduler.