Lab I assignment Report:

**Lab 1 Assignment**

Name: Nagulapati Rohith Kumar, Class ID: 16

Name: Nageswara Rao Nandigam, Class ID: 17

**Question 1**

1.Spark ProgrammingDownload the movielens dataset from the following link -

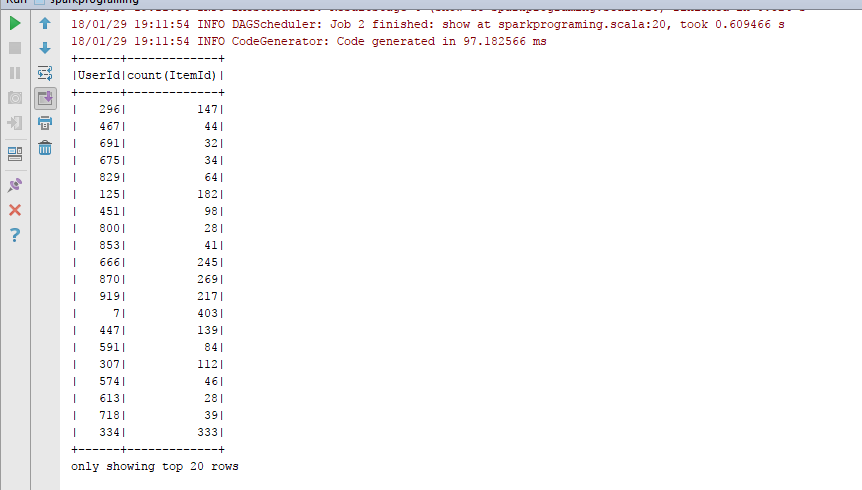
<https://grouplens.org/datasets/movielens/100k/>

Please use the u.data file in the dataset as your input file.The input file contains 100000 movie ratings by 943 users on 1682 items. The data is randomly ordered. This is a tab separated list of "user id | item id | rating | timestamp". Write Spark Transformations and Actions to find the users who have rated more than 25 items.Input: u.data file from the movielens dataset.Output: List of (userid of the user who has rated more than 25 items, Number of items that he has rated)

Source code snippet:



Output :

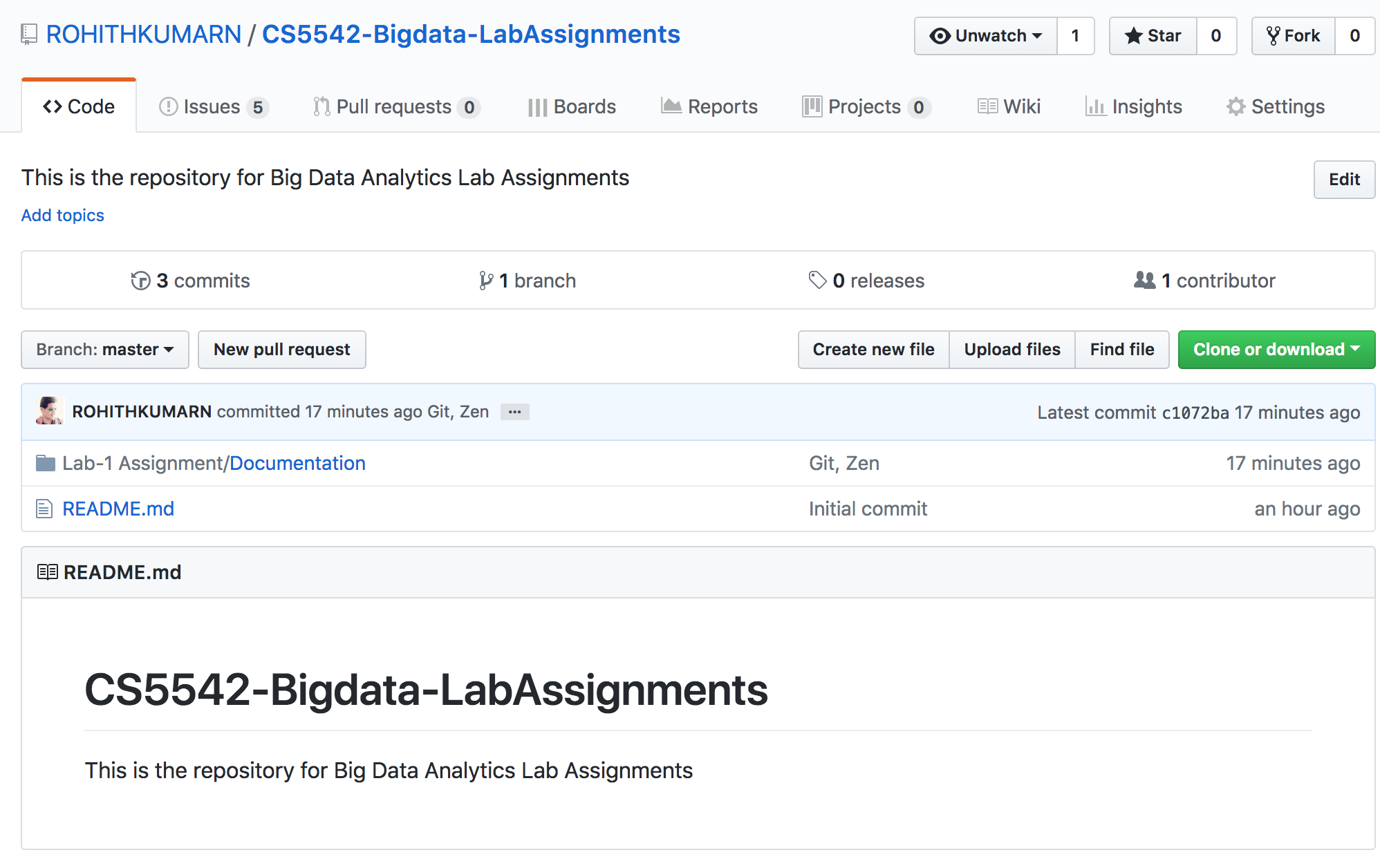


## Question 2 and 3:

* Create GitHub Account. Create a repository in remote Github. Clone it to the local machine.

### Steps:

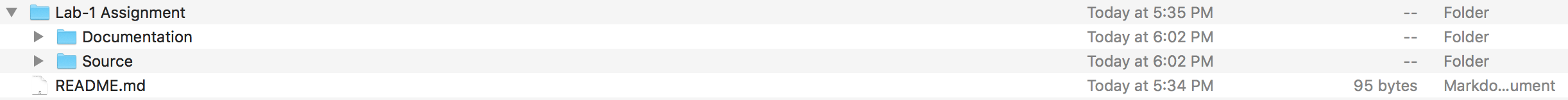
1. created the github account
2. created a repository with name "CS5542-Bigdata-LabAssignments" and cloned to local machine



* Create 2 (Source and Documentation) directories in local GitHub.

### Steps:

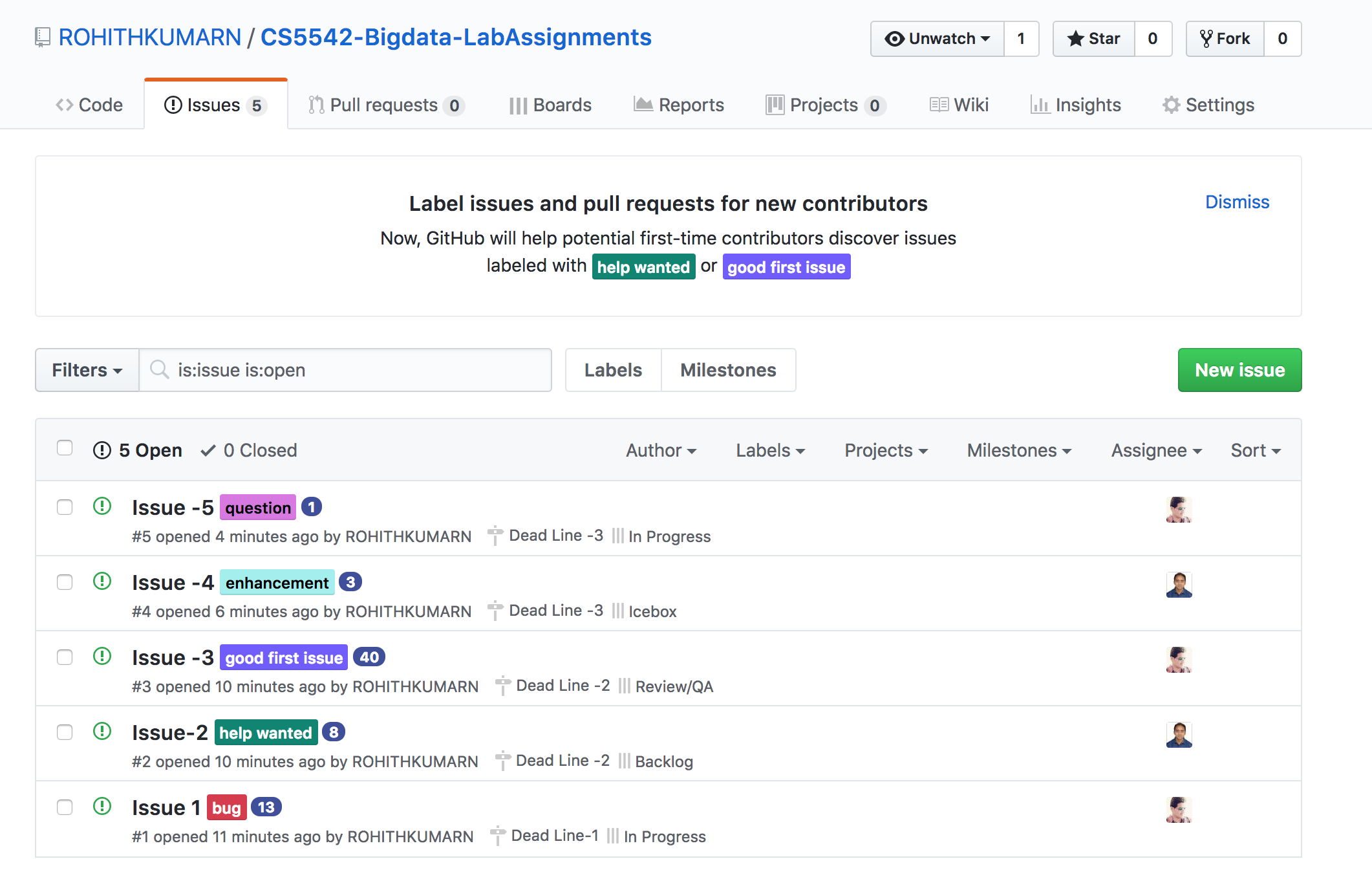
1. Created two folders source and documentation
2. Pushed it to Remote Repository



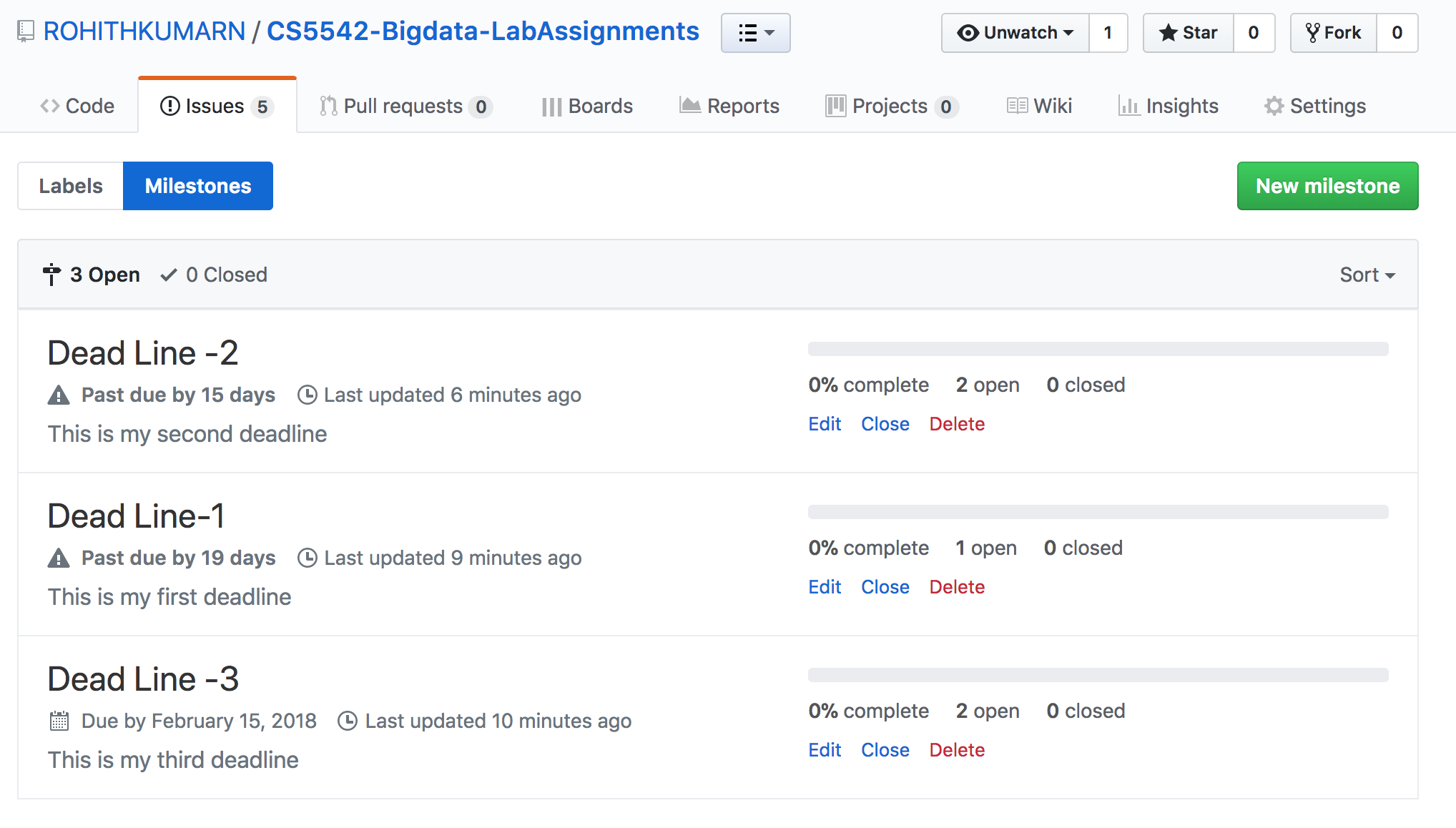
Create ZenHub Tool Account. Create a board, 3 milestones, at least 5 issues and show the analytics graph.

### Steps:

1. created five issues



1. created 3 milestones



1. Assigned five issues to 3 milestones
2. Draw burndown chart

