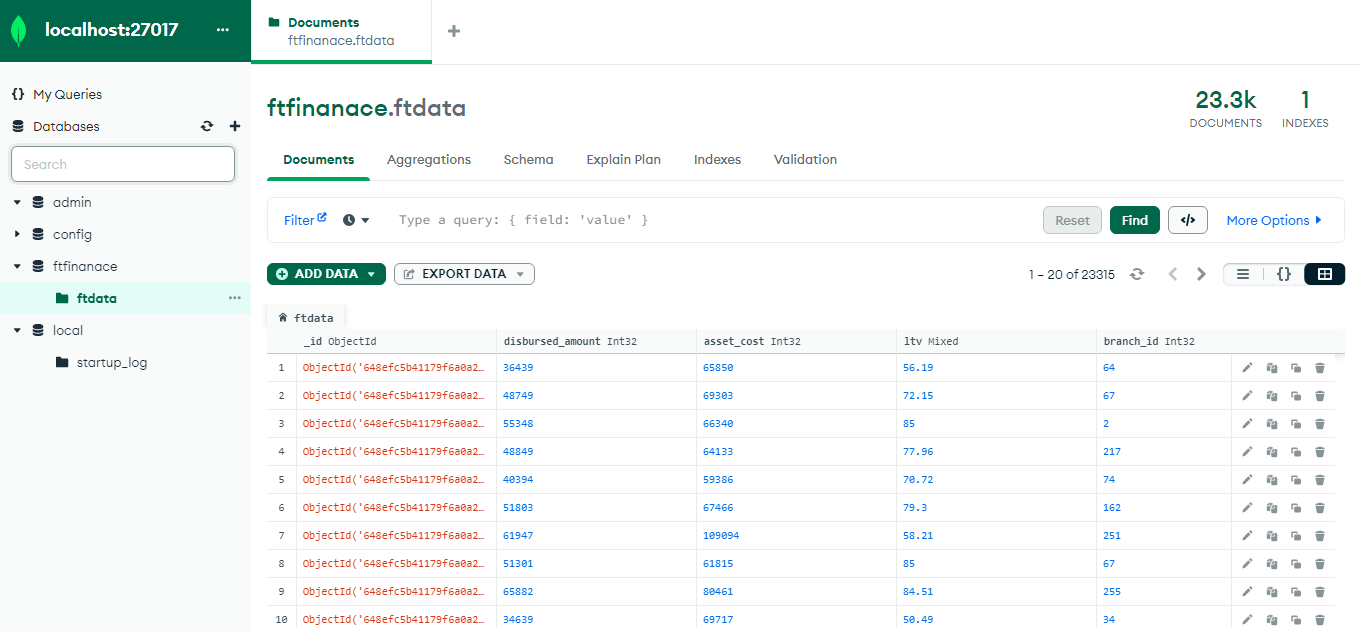
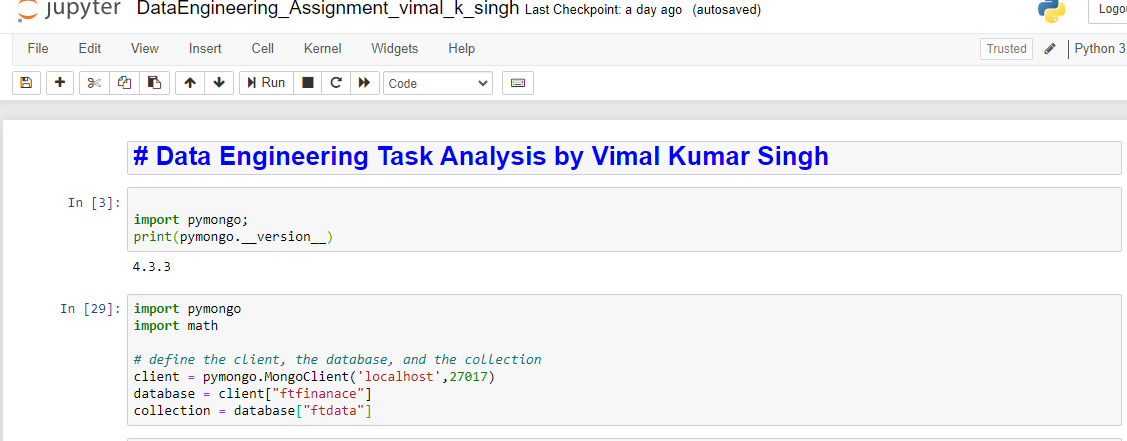
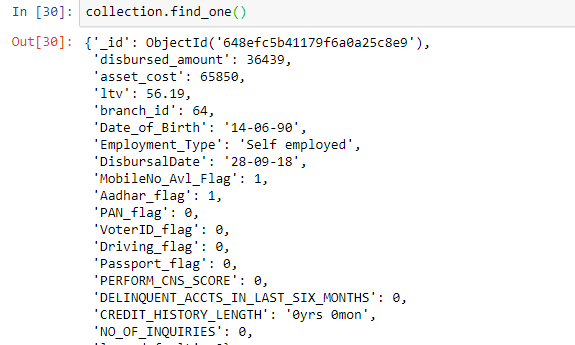
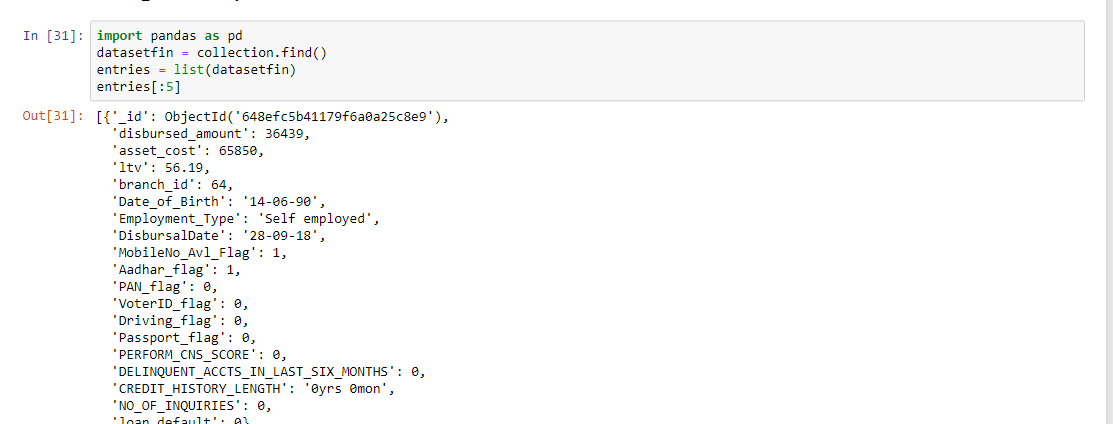
**Data Engineering AWS(Vimal Kumar Singh)**

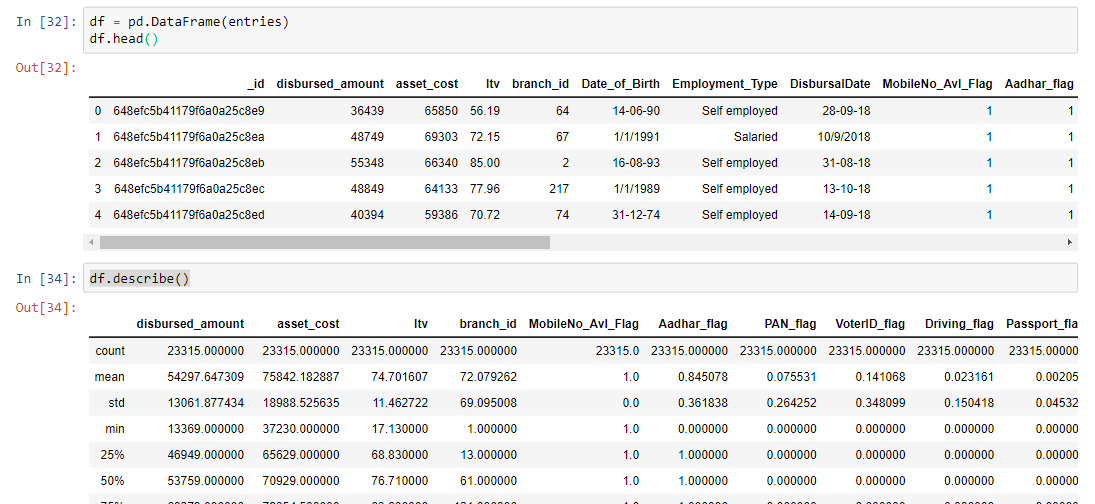
GitHub URL: https://github.com/vimalrns/DataEnginnering\_AWS.git

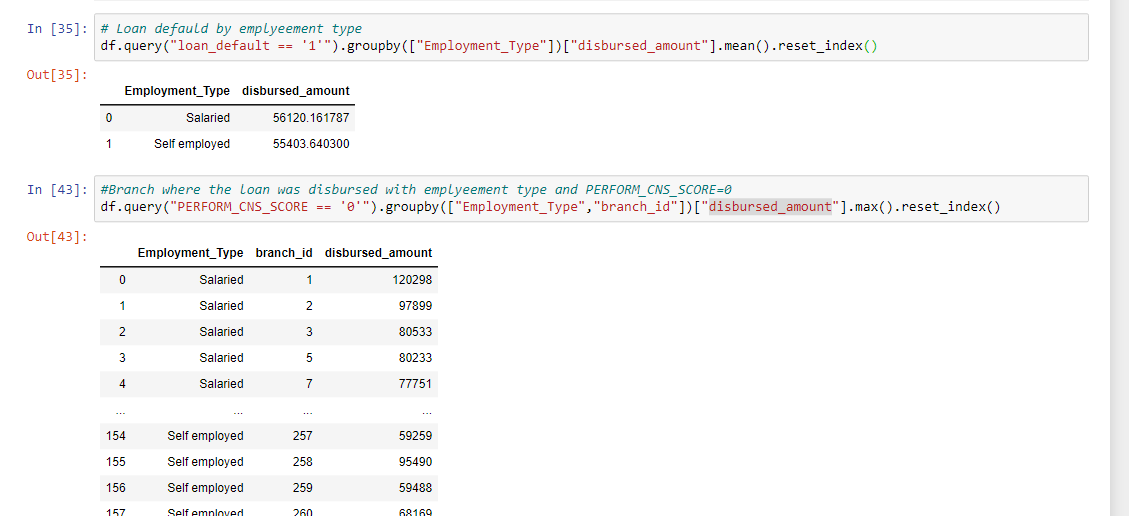


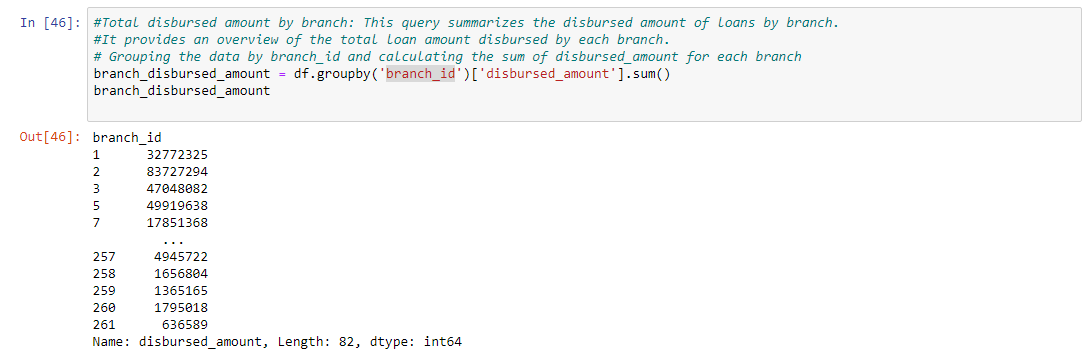


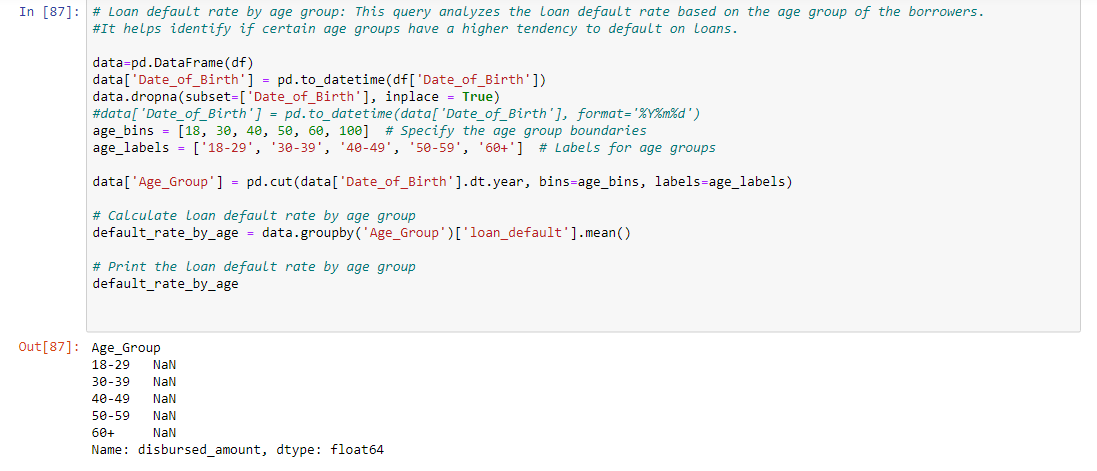


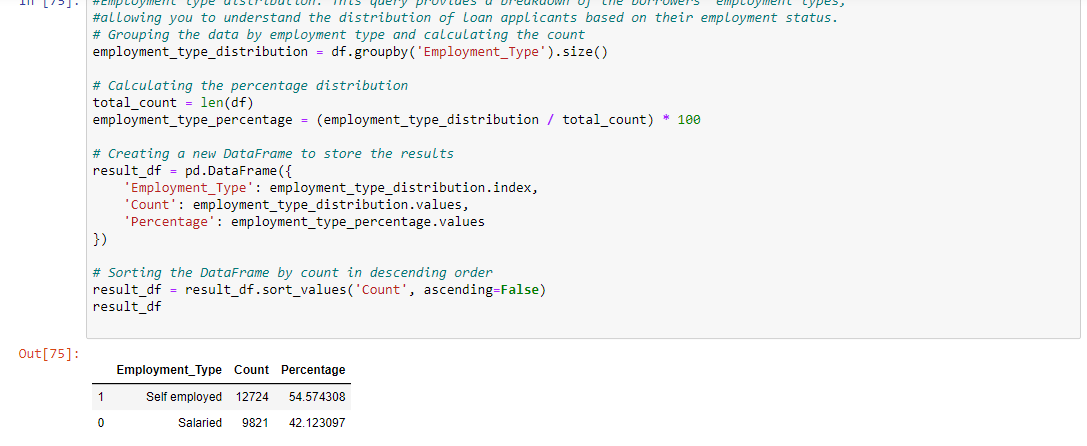


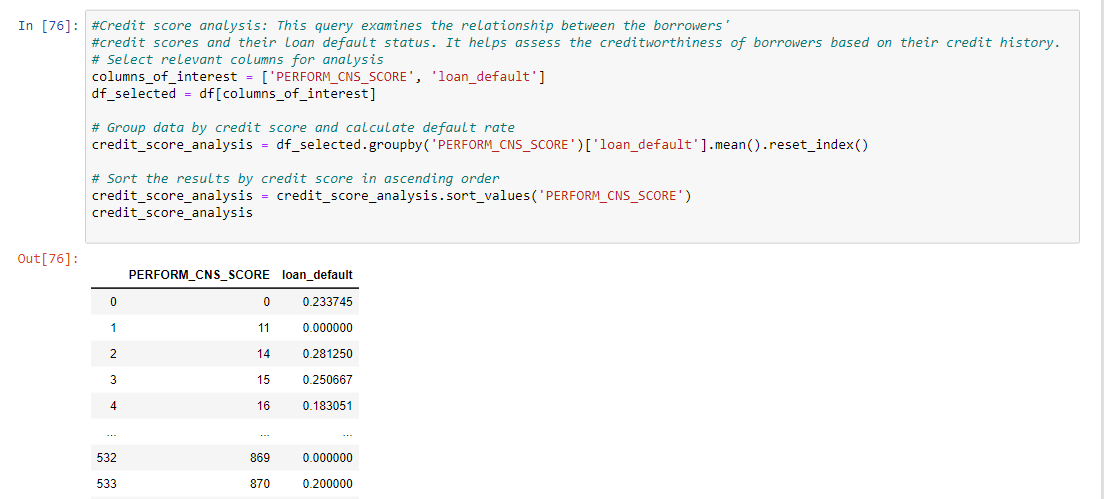


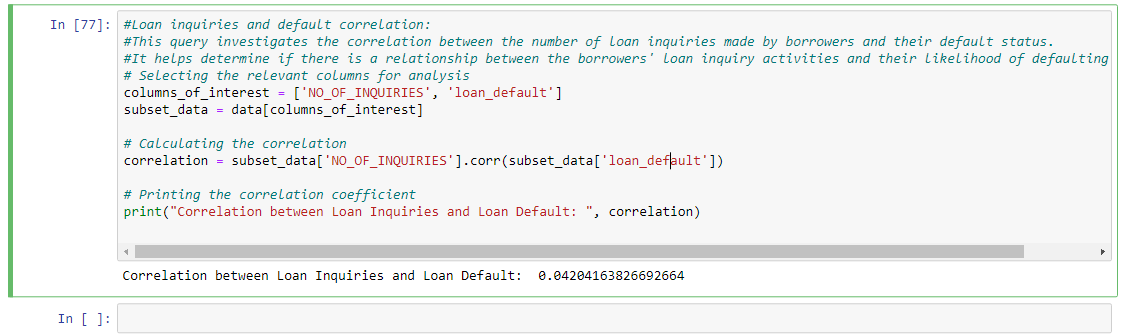






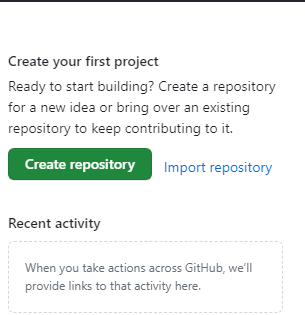


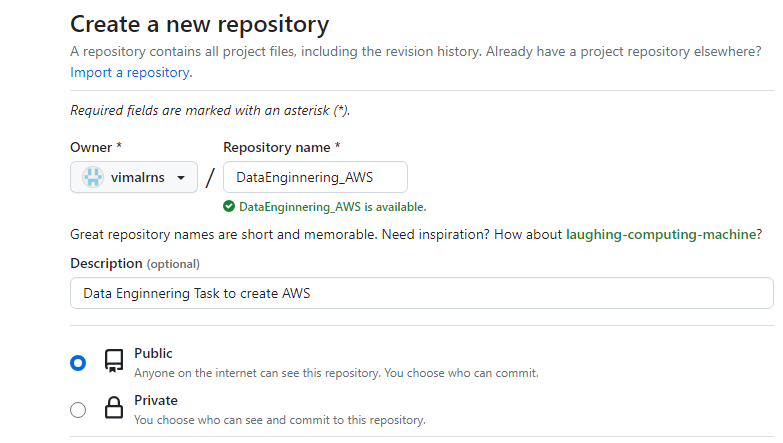


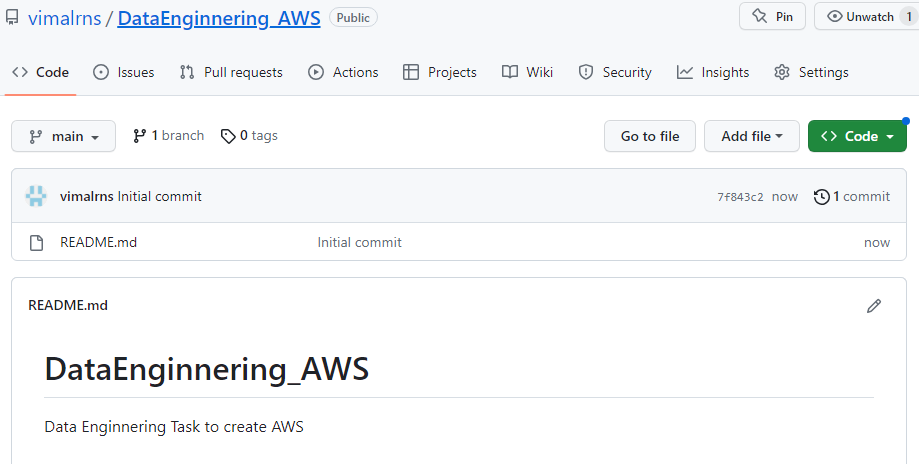


**Task 1 Create a repository in github and upload the dataset**

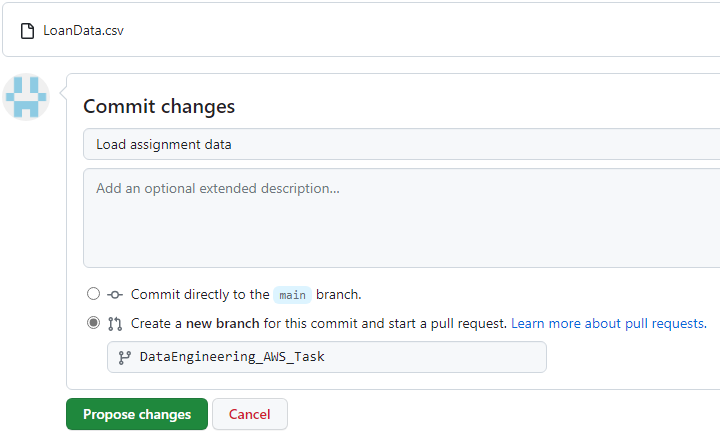
* 1. Create a repository in github.com

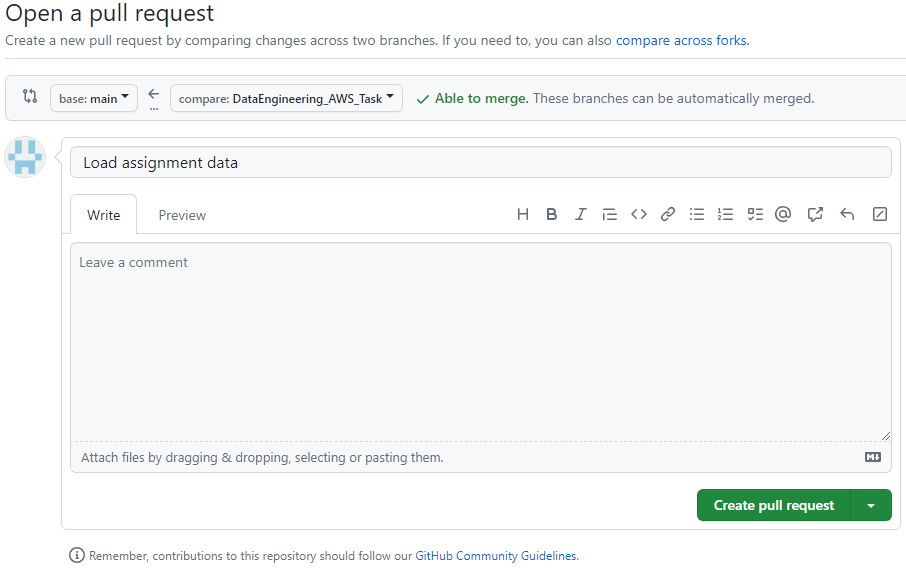


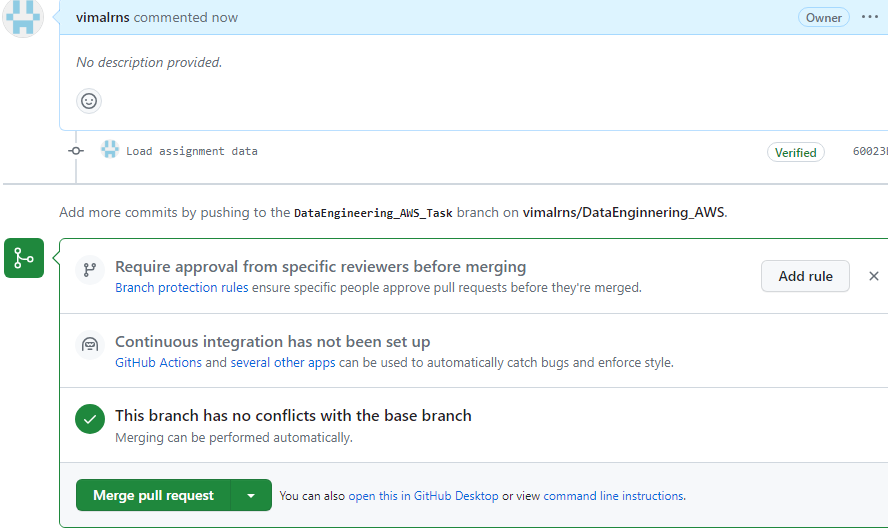


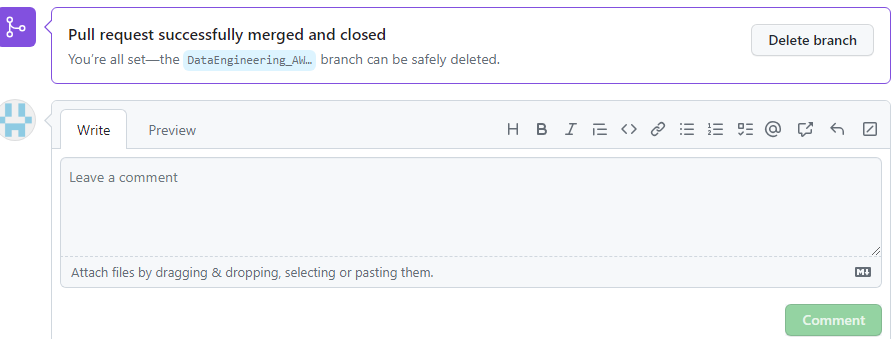


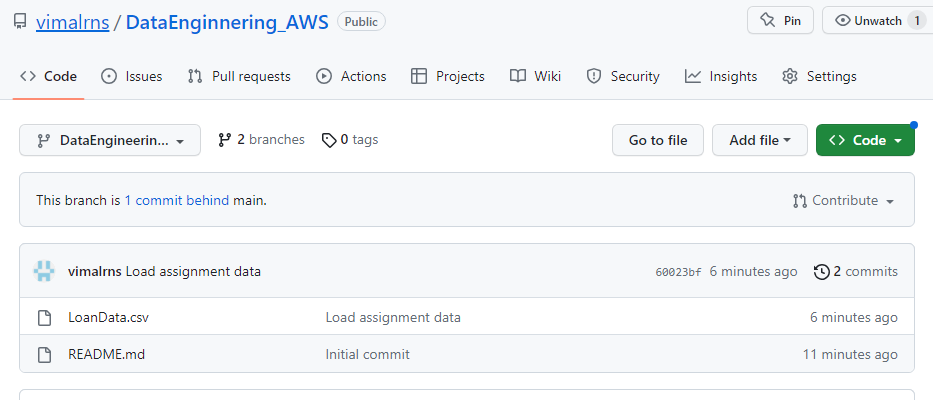
Upload the dataset to github.com through browser





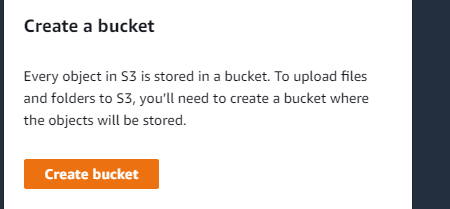


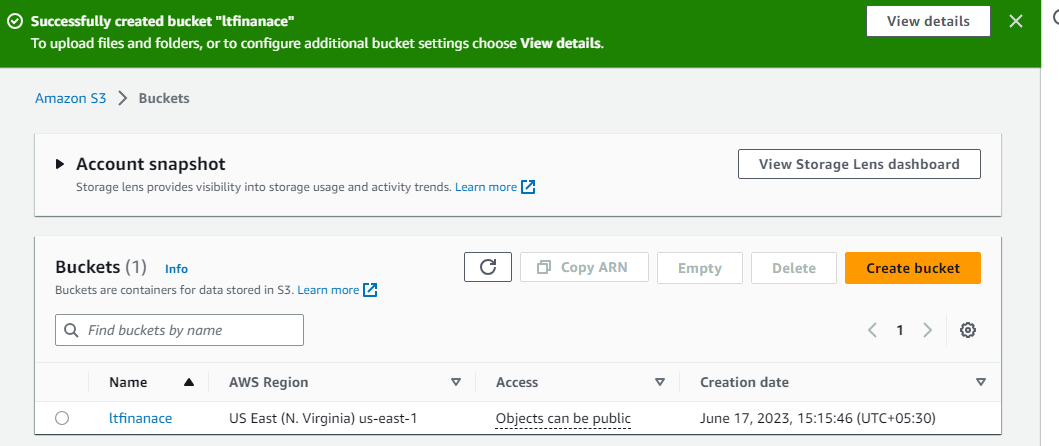


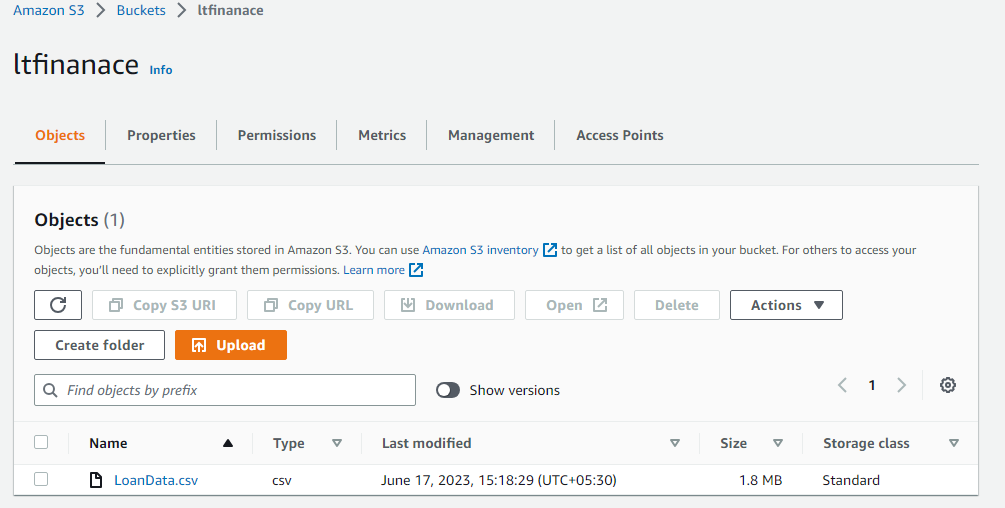


**Cloud Data Engineering Tasks :**

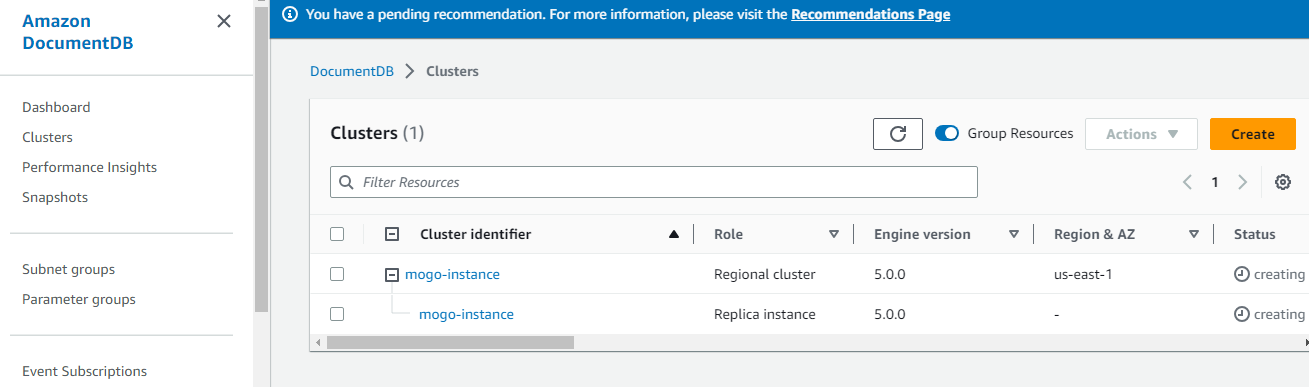
1. Upload the dataset to aws s3







1. Create aws managed mongo db instance and create five reports



1. Create aws code commit repos and perform the same steps as mentioned in Data Versioning Task
2. Create aws emr instance and perform three data engineering tasks such cleaning etc . Example : Removing column values with Unknown , Imputing the column values

