INFOSYS 750- Final Project

Evolution of Open Source Projects

Open source projects are software artefacts, which are developed and maintained by software developers and volunteers. Generally, the source code of these projects are available online and end-users can freely use them under the constraints defined on project license type. GitHub is the largest open source repository with millions of projects. You have been provided with a panel dataset of projects hosted on GitHub. It contains two years (8 quarters) of data for each project. You can how projects change over time by analysing these datasets.

Academic Report:

In recent years, software companies have started collaborating with and/or sponsoring open source software projects. Before forging any partnership, software companies try to understand OSS projects' progress and how they change over time.

Mr. Richard Thompson, a software architect at Mega Software who often collaborates with or sponsors OSS projects, needs your help in understanding evolution in some of the OSS projects. Mr. Thompson, would like any insights about these projects (see the dataset).

By applying longitudinal data models, you need to provide an insight about the projects evolution. Clearly define your research questions and test your hypotheses using this dataset. The final report should be have following structure-

- Introduction
- Research questions,
- Definition of main variables, Visual Exploration
- Data cleaning and preparation (any modification, transformation and sampling techniques that you have applied)
- Your multi-level longitudinal model,
- Results an discussions
- References

Also, you need to provide your R Scripts in a separate file along with the report in (Doc, Docx, or PDF format)

Project Grading:

Project Report	10%	8th November 2019, 23:59 PM (submit on Canvas)
R Script File	15%	8th November 2019, 23:59 PM (submit on Canvas)
Total	25%	

Everyone needs to submit a project report. All group members are expected to contribute equally. For each group, R script file would be same for every group member, however, report should be written individually.

Length of the report (excluding script file): 2000 words (max.)

If there are any problems in your group, please notify the instructor immediately. You need to get familiar with Git terminology to understand the dataset better. Bring your questions to lectures and tutorials.

Dataset details:

Variable	**Definition**
PrjID	A unique id number for each project
Period	Represents the current record contains data for which period of year
Time	A sequence for time of observations
SatrtDate	Beginning of observation
EndDate	End of observation
Forks	Number of times a project is forked
Members	Number of members
Commits	Number of coding activities
Issues	Number of problem/bugs raised or requests for new features
Watchers	Number of people interested in project
PullReq	Number of code changes request for review.
CommitCmnt	Number of discussion on commits