– Dataset 1—

SELECT CASE WHEN non\_casual\_developers.gender='male' THEN 0 ELSE 1 END as Gender , non\_casual\_developers.is\_gender\_neutral as isGenderNeutral, regression\_dataset.request\_id, regression\_dataset.insertions, regression\_dataset.deletions,

regression\_dataset.total\_churn as patchSize, regression\_dataset.number\_patches as numPatch, regression\_dataset.is\_accepted as isAccepted, regression\_dataset.is\_bugfix as isBugFix,

regression\_dataset.author\_tenure\_months as tenure, regression\_dataset.owner, regression\_dataset.code\_commit\_exp as totalCommit, regression\_dataset.code\_review\_exp as revExp, regression\_dataset.review\_interval\_hours as reviewInterval, regression\_dataset.file\_count as fileCount, regression\_dataset.num\_new\_files as numNewFiles,

regression\_dataset.directory\_count as dirCount, regression\_dataset.avg\_complexity as cyCmplx, regression\_dataset.avg\_comment\_volume as cmtVolume FROM `non\_casual\_developers` INNER JOIN regression\_dataset on non\_casual\_developers.owner=regression\_dataset.owner;

– Dataset 2—

SELECT non\_casual\_developers.owner, CASE WHEN non\_casual\_developers.gender ='male' THEN 0 ELSE 1 END as gender, non\_casual\_developers.is\_gender\_neutral,

CEILING(AVG(code\_commit\_exp)) as commit\_experience, CEILING(AVG(code\_review\_exp)) as review\_experience, MAX(code\_review\_exp) -MIN(code\_review\_exp) as review\_count, author\_tenure\_months

FROM regression\_dataset INNER JOIN

non\_casual\_developers on non\_casual\_developers.owner=regression\_dataset.owner

GROUP BY non\_casual\_developers.owner, non\_casual\_developers.gender, non\_casual\_developers.is\_gender\_neutral, author\_tenure\_months;

SELECT SUM(is\_accepted)/COUNT(request\_id) as acceptance\_rate, non\_casual\_developers.gender

FROM `regression\_dataset` INNER JOIN non\_casual\_developers ON non\_casual\_developers.owner=regression\_dataset.owner

GROUP by gender;

—--------------------------ghtorrent dataset1—---------------

SELECT CASE WHEN user\_resolve.resolvedGender='male' THEN 0 ELSE 1 END as Gender , user\_resolve.is\_gender\_neutral as isGenderNeutral, regression\_dataset.pull\_id, regression\_dataset.additions, regression\_dataset.deletions,

regression\_dataset.total\_churn as patchSize, regression\_dataset.commit\_count as numPatch, regression\_dataset.is\_accepted as isAccepted, regression\_dataset.is\_bugfix as isBugFix,

regression\_dataset.author\_tenure\_months as tenure, regression\_dataset.owner, regression\_dataset.code\_commit\_exp as totalCommit, regression\_dataset.code\_review\_exp as revExp,

regression\_dataset.review\_interval\_hours as reviewInterval, regression\_dataset.num\_changed\_files as fileCount,

regression\_dataset.num\_new\_files as numNewFiles,

regression\_dataset.directory\_count as dirCount, regression\_dataset.avg\_complexity as cyCmplx, regression\_dataset.avg\_comment\_volume cmtVolume FROM

regression\_dataset INNER JOIN user\_resolve on regression\_dataset.owner=user\_resolve.id and user\_resolve.unresolved=0 and regression\_dataset.review\_interval\_hours is not NULL

INNER JOIN projects ON regression\_dataset.project\_id=projects.project\_id and projects.project\_size=1

—--------------------------ghtorrent dataset2—---------------

SELECT user\_resolve.id, CASE WHEN user\_resolve.resolvedGender ='male' THEN 0 ELSE 1 END as gender, user\_resolve.is\_gender\_neutral,

CEILING(AVG(code\_commit\_exp)) as commit\_experience, CEILING(AVG(code\_review\_exp)) as review\_experience, MAX(code\_review\_exp) -MIN(code\_review\_exp) as review\_count, regression\_dataset.author\_tenure\_months

FROM regression\_dataset INNER JOIN

user\_resolve on user\_resolve.id=regression\_dataset.owner and user\_resolve.unresolved=0

INNER JOIN projects ON regression\_dataset.project\_id=projects.project\_id and projects.project\_size=1

GROUP BY user\_resolve.id, gender, user\_resolve.is\_gender\_neutral, regression\_dataset.author\_tenure\_months;