# into user\_history\_by\_refactor

# Standard sql

# Large results

Select

refactor.refactor\_commit as refactor\_commit

, refactor.refactor\_repo\_name as refactor\_repo\_name

, refactor.file as refactor\_file

# Typo

, refactor.refctor\_email as refactor\_email

#, refactor.refactor\_email as refactor\_email

, max(refactor.refctor\_date) as refactor\_date

# overall experience

, count(distinct hist.commit) as overall\_commits

, count(distinct if(hist.refactor\_matches > 0, hist.commit, null)) as overall\_refactors

, count(distinct if(hist.bq\_classification , hist.commit, null)) as overall\_correctives

, if( count(distinct hist.commit) > 0, 1.0\*count(distinct if(hist.refactor\_matches > 0, hist.commit, null))/count(distinct hist.commit), 0) as overall\_refactors\_ratio

, count(distinct hist.repo\_name) as overall\_repositories

, count(distinct CONCAT(hist.repo\_name, hist.difference\_old\_path)) as overall\_files

, max(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY)) as overall\_days\_to\_refactor

# repo experience

, count(distinct if(refactor.refactor\_repo\_name = hist.repo\_name, hist.commit, null)) as repo\_commits

, count(distinct if(refactor.refactor\_repo\_name = hist.repo\_name and hist.refactor\_matches > 0, hist.commit, null)) as repo\_refactors

, count(distinct if(refactor.refactor\_repo\_name = hist.repo\_name and hist.bq\_classification , hist.commit, null)) as repo\_correctives

, if( count(distinct if(refactor.refactor\_repo\_name = hist.repo\_name, hist.commit, null)) > 0, 1.0\*count(distinct if(refactor.refactor\_repo\_name = hist.repo\_name and hist.refactor\_matches > 0, hist.commit, null))/ count(distinct if(refactor.refactor\_repo\_name = hist.repo\_name, hist.commit, null)), 0) as repo\_refactors\_ratio

#, count(distinct hist.repo\_name) as overall\_repositories

, count(distinct if(refactor.refactor\_repo\_name = hist.repo\_name, hist.difference\_old\_path, null)) as repo\_files

, max(if(refactor.refactor\_repo\_name = hist.repo\_name,TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY), null)) as repo\_days\_to\_refactor

# file experience

, count(distinct if(refactor.refactor\_repo\_name = hist.repo\_name and refactor.file = hist.difference\_old\_path, hist.commit, null)) as file\_commits

, count(distinct if(refactor.refactor\_repo\_name = hist.repo\_name and refactor.file = hist.difference\_old\_path and hist.refactor\_matches > 0, hist.commit, null)) as file\_refactors

, count(distinct if(refactor.refactor\_repo\_name = hist.repo\_name and refactor.file = hist.difference\_old\_path and hist.bq\_classification , hist.commit, null)) as file\_correctives

, if( count(distinct if(refactor.refactor\_repo\_name = hist.repo\_name and refactor.file = hist.difference\_old\_path, hist.commit, null)) > 0, 1.0\*count(distinct if(refactor.refactor\_repo\_name = hist.repo\_name and refactor.file = hist.difference\_old\_path and hist.refactor\_matches > 0, hist.commit, null))/ count(distinct if(refactor.refactor\_repo\_name = hist.repo\_name and refactor.file = hist.difference\_old\_path, hist.commit, null)), 0) as file\_refactors\_ratio

#, count(distinct hist.repo\_name) as overall\_repositories

#, count(distinct if(refactor.refactor\_repo\_name = hist.repo\_name, hist.difference\_old\_path, null)) as repol\_files

, max(if(refactor.refactor\_repo\_name = hist.repo\_name and refactor.file = hist.difference\_old\_path,TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY), null)) as file\_days\_to\_refactor

#### Recent (3 months experience)

# Recent overall

, count(distinct if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30, hist.commit, null)) as overall\_3m\_commits

, count(distinct if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30 and hist.refactor\_matches > 0, hist.commit, null)) as overall\_3m\_refactors

, count(distinct if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30 and hist.bq\_classification , hist.commit, null)) as overall\_3m\_correctives

, if(count(distinct if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30, hist.commit, null)) > 0, 1.0\*count(distinct if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30 and hist.refactor\_matches > 0, hist.commit, null))/ count(distinct if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30, hist.commit, null)), 0) as overall\_m3\_refactors\_ratio

, count(distinct if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30, hist.repo\_name, null)) as overall\_3m\_repositories

, count(distinct if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30, hist.difference\_old\_path, null)) as overall\_3m\_files

, max(if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30 ,TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY), null)) as overall\_3m\_days\_to\_refactor

#recent repo experience

, count(distinct if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30 and refactor.refactor\_repo\_name = hist.repo\_name, hist.commit, null)) as repo\_3m\_commits

, count(distinct if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30 and refactor.refactor\_repo\_name = hist.repo\_name and hist.refactor\_matches > 0, hist.commit, null)) as repo\_3m\_refactors

, count(distinct if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30 and refactor.refactor\_repo\_name = hist.repo\_name and hist.bq\_classification , hist.commit, null)) as repo\_3m\_correctives

, if(count(distinct if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30 and refactor.refactor\_repo\_name = hist.repo\_name, hist.commit, null)) > 0, 1.0\*count(distinct if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30 and refactor.refactor\_repo\_name = hist.repo\_name and hist.refactor\_matches > 0, hist.commit, null))/ count(distinct if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30 and refactor.refactor\_repo\_name = hist.repo\_name, hist.commit, null)), 0) as repo\_3m\_refactors\_ratio

#, count(distinct hist.repo\_name) as overall\_repositories

, count(distinct if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30 and refactor.refactor\_repo\_name = hist.repo\_name, hist.difference\_old\_path, null)) as repo\_3m\_files

, max(if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30 and refactor.refactor\_repo\_name = hist.repo\_name,TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY), null)) as repo\_3m\_days\_to\_refactor

# recent file experience

, count(distinct if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30 and refactor.refactor\_repo\_name = hist.repo\_name and refactor.file = hist.difference\_old\_path, hist.commit, null)) as file\_3m\_commits

, count(distinct if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30 and refactor.refactor\_repo\_name = hist.repo\_name and refactor.file = hist.difference\_old\_path and hist.refactor\_matches > 0, hist.commit, null)) as file\_3m\_refactors

, count(distinct if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30 and refactor.refactor\_repo\_name = hist.repo\_name and refactor.file = hist.difference\_old\_path and hist.bq\_classification , hist.commit, null)) as file\_3m\_correctives

, if( count(distinct if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30 and refactor.refactor\_repo\_name = hist.repo\_name and refactor.file = hist.difference\_old\_path, hist.commit, null)) > 0, 1.0\*count(distinct if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30 and refactor.refactor\_repo\_name = hist.repo\_name and refactor.file = hist.difference\_old\_path and hist.refactor\_matches > 0, hist.commit, null))/ count(distinct if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30 and refactor.refactor\_repo\_name = hist.repo\_name and refactor.file = hist.difference\_old\_path, hist.commit, null)), 0) as file\_3m\_refactors\_ratio

#, count(distinct hist.repo\_name) as overall\_repositories

#, count(distinct if(refactor.refactor\_repo\_name = hist.repo\_name, hist.difference\_old\_path, null)) as repo\_3m\_files

, max(if(TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY) < 3\*30 and refactor.refactor\_repo\_name = hist.repo\_name and refactor.file = hist.difference\_old\_path ,TIMESTAMP\_DIFF(refactor.refctor\_date, hist.author\_date, DAY), null)) as file\_3m\_days\_to\_refactor

From

`citric-trees-187317.repo\_attributes.refactor\_stats\_before\_2013` as refactor

Join

`citric-trees-187317.repo\_attributes.valid\_commits\_classified` as hist

On

Refactor.refctor\_email = hist.author\_email

And

refactor.refctor\_date > hist.author\_date

#Where

#Refactor.refactor\_matches > 0

Group by

refactor\_commit

, refactor\_repo\_name

, refactor\_file

, refactor\_email