BDA

Assignment 1

Reshan Faraz (PhD19006)

Prakrati Gupta(MT20014)

Methodology and Approaches :

We used PostgreSQL,which is an open source relational database management system first we open the dataset and export in excel sheet .After that we create a database name BDA and under that we create table ***pull_request*** with the following attributes *id(Primary Key),name_of_author ,events,counts,date_of_event.*

We used *group by* clause that returned the selected elements into groups.

We also used *where* clauses to filter the query result.

Creating the Table :

```
CREATE TABLE pull_request (
  id SERIAL,
  Name_of_Author VARCHAR(50),
  events VARCHAR(50),
  counts INT,
  Date_of_Event DATE,
  PRIMARY KEY (id)
)
```

AFTER that we import the CSV

```
COPY pull_request(counts, name_of_author, events, date_of_event)
FROM 'C:\Program Files\PostgreSQL\13\pgAdmin 4\a.csv' DELIMITER ',';
```

Table Info :

id	name_of_author	counts	events	Date_of_events
Primary Key	Name of the author	Pull request Id	Type of Events	Date of the Events

Data Analysis :

Following are the Query used for analysis of Dataset :

select * from pull_request;

select Count(*) from pull_request;

select name_of_author from pull_request group by name_of_author order by name_of_author;

SQL query and its corresponding Screenshot of output

ANSWER 1 -

a) **select date_of_event,count(*) as opened_per_day from pull_request where events='opened' group by date_of_event order by date_of_event ;**

```
14  select date_of_event,count(*) as opened_per_day from
15  pull_request where events='opened'
16  group by date_of_event order by date_of_event ;
```

Data Output | Explain | Messages | Notifications

	date_of_event date	opened_per_day bigint
1	2010-09-02	2
2	2010-09-06	1
3	2010-09-08	1
4	2010-09-09	4
5	2010-09-10	3
6	2010-09-11	3
7	2010-09-12	3
8	2010-09-13	3
9	2010-09-15	2
10	2010-09-16	2
11	2010-09-18	6

b) **select date_of_event,count(*) as opened_per_day from pull_request where events='discussed' group by date_of_event order by date_of_event ;**

```
14  select date_of_event,count(*) as opened_per_day from
15  pull_request where events='discussed'
16  group by date_of_event order by date_of_event ;
```

Data Output | Explain | Messages | Notifications

	date_of_event date	opened_per_day bigint
1	2010-09-09	6
2	2010-09-10	10
3	2010-09-11	13
4	2010-09-12	5
5	2010-09-13	7
6	2010-09-14	1
7	2010-09-15	3
8	2010-09-16	2
9	2010-09-17	1
10	2010-09-21	6
11	2010-09-22	3

ANSWER 2

Here we are taking separate for each month for all years and Output the person who has the highest number of comments that month.

```
select Distinct on(mm) name_of_author from(
        select* from (
select name_of_author, max(ee) as ee1,mm from (
select name_of_author, count(events) as ee,
date_trunc('month', date_of_event) as mm
from pull_request where events='discussed' group by mm,name_of_author)
as tempp group by name_of_author,mm order by mm DESC) as tempp3
group by mm,name_of_author,ee1 order by ee1 DESC) as dd order by mm,ee1 DESC
```

```
26  ---------------------------------- 2 ----------------------------------
27  select Distinct on(mm) name_of_author from(
28      select* from (
29  select name_of_author, max(ee) as ee1,mm from (
30  select name_of_author, count(events) as ee,
31  date_trunc('month', date_of_event) as mm
32  from pull_request where events='discussed' group by mm,name_of_author)
33  as tempp group by name_of_author,mm order by mm DESC) as tempp3
34  group by mm,name_of_author,ee1 order by ee1 DESC) as dd order by mm,ee1 DESC
```

Data Output | Explain | Messages | Notifications

	name_of_author character varying (50)
1	mikel
2	josevalim
3	josevalim
4	josevalim
5	jeremy
6	spastorino
7	josevalim

Answer 3)

Here we are taking separate for each week for all months and Output the person who has the highest number of comments that week.

select Distinct on(mm) name_of_author from(

select* from (

select name_of_author, max(ee) as ee1,mm from (

select name_of_author, count(events) as ee,

date_trunc('week', date_of_event) as mm

from pull_request where events='discussed' group by mm,name_of_author)

as tempp group by name_of_author,mm order by mm DESC) as tempp3

group by mm,name_of_author,ee1 order by ee1 DESC) as dd order by mm,ee1 DESC

```
67  select Distinct on(mm) name_of_author from(
68      select* from (
69  select name_of_author, max(ee) as ee1,mm from (
70  select name_of_author, count(events) as ee,
71  date_trunc('week', date_of_event) as mm
72  from pull_request where events='discussed' group by mm,name_of_author)
73  as tempp group by name_of_author,mm order by mm DESC) as tempp3
74  group by mm,name_of_author,ee1 order by ee1 DESC) as dd order by mm,ee1 DESC
75
```

Data Output | Explain | Messages | Notifications

	name_of_author character varying (50)
1	mikel
2	mikel
3	josevalim
4	josevalim
5	josevalim
6	josevalim
7	krekoten
8	fxn

ANSWER - 4)

select count(*),

date_trunc('week', date_of_event)

from pull_request

where events = 'opened'

group by

date_trunc('week', date_of_event)

order by date_trunc desc ;

```
72  select count(*),
73    date_trunc('week', date_of_event)
74  from pull_request
75  where events = 'opened'
76  group by
77    date_trunc('week', date_of_event)
78  order by date_trunc desc ;
79
```

Data Output | Explain | Messages | Notifications

	count bigint	date_trunc timestamp with time zone
1	30	2018-08-27 00:00:00+05:30
2	39	2018-08-20 00:00:00+05:30
3	39	2018-08-13 00:00:00+05:30
4	34	2018-08-06 00:00:00+05:30
5	28	2018-07-30 00:00:00+05:30
6	29	2018-07-23 00:00:00+05:30
7	27	2018-07-16 00:00:00+05:30
8	28	2018-07-09 00:00:00+05:30

ANSWER - 5)

select sum(res),Extract(month from date_of_event) as month_in_2010 from(

select Count(*) as res,date_of_event from pull_request where Extract(year from date_of_event)=2010 and events=

'merged' group by date_of_event) as xx group by Extract(month from date_of_event);

```
12  select sum(res),Extract(month from date_of_event) as month_in_2010 from(
13  select Count(*) as res,date_of_event from pull_request where Extract(year from date_of_event)=2010 and events=
14  'merged' group by date_of_event) as xx group by Extract(month from date_of_event);
15
```

Data Output | Explain | Messages | Notifications

sum numeric	month_in_2010 double precision

Output : There is no record for this query

ANSWER- 6

select count(events),date_of_event from pull_request

group by EXTRACT(day from date_of_event),date_of_event

order by date_of_event;

```
18  select count(events),date_of_event from pull_request
19  group by EXTRACT(day from date_of_event ),date_of_event
20  order by date_of_event;
21
```

Data Output | Explain | Messages | Notifications

	count bigint	date_of_event date
1	2	2010-09-02
2	1	2010-09-06
3	1	2010-09-08
4	10	2010-09-09
5	13	2010-09-10
6	16	2010-09-11
7	8	2010-09-12
8	10	2010-09-13
9	1	2010-09-14
10	5	2010-09-15
11	4	2010-09-16

ANSWER 7)

```
select name_of_author
from(
        select name_of_author,count(events) as
        pull_request_from_author_in_2011
        from pull_request where events='opened'
        and EXTRACT(year from date_of_event )='2011'
        group by name_of_author) as
```

dd order by pull_request_from_author_in_2011 DESC Limit 1;

```
41  select name_of_author
42  from(
43      select name_of_author,count(events) as
44      pull_request_from_author_in_2011
45      from pull_request where events='opened'
46      and EXTRACT(year from date_of_event )='2011'
47      group by name_of_author) as
48      dd order by pull_request_from_author_in_2011 DESC Limit 1;
49
```

Data Output | Explain | Messages | Notifications

	name_of_author character varying (50)
1	arunagw

Learning in doing this assignment :

- How to work with PostgreSql, creating a database in pgAdmin.
- How to write a Query based on months, weeks, days or particular years.
- SQL syntax for different features such as Date_trunc,Extract features.
- Writing SQL inner and outer Queries, joins, Group By and conditions.