

Exercitii la clasa - 22 oct 2020

- O lista cu titlurile, id-ul si textul a 47 de articole din iarna lui 2018 ordonate crescator dupa scor

```
SELECT id, title, text
FROM `bigquery-public-data.hacker_news.full` as f
WHERE Date(timestamp) between '2018-01-01' and '2018-12-31'
and type="story"
ORDER BY score asc
LIMIT 47
```

- Top 35 sondaje de opinie cu optiuni sortate dupa scor din quarter-ul 3 din 2019

```
SELECT id, title, text
FROM `bigquery-public-data.hacker_news.full` as f
WHERE Date(timestamp) between '2019-07-01' and '2019-09-30'
and type="pollopt"
ORDER BY score asc
LIMIT 35
```

Tema - 27 oct 2020

- Cate articole (story) au fost publicate in Ianuarie 2020?

```
SELECT count(*) as NumberOfStories
FROM `bigquery-public-data.hacker_news.full`
WHERE Date(timestamp) between '2020-01-01' and '2020-01-31'
and type="story"
```

34955

- Cate sondaje de opinie (poll) au fost facute intre Martie si Noiembrie 2016?

```
SELECT count(*) as NumberOfPolls
FROM `bigquery-public-data.hacker_news.full`
WHERE Date(timestamp) between '2016-03-01' and '2016-11-30'
and type="poll"
```

57

- Care e continutul celor mai bune 5 comentarii din perioada 15 Martie 2020 si 16 Aprilie 2020

```
SELECT text as Comments
FROM `bigquery-public-data.hacker_news.full`
WHERE Date(timestamp) between '2020-03-15' and '2020-04-16'
and type = "comment"
ORDER BY ranking DESC
LIMIT 5
```

- Care e continutul celor mai proaste 25 de articole?

```
SELECT text as Stories
FROM `bigquery-public-data.hacker_news.full`
WHERE Date(timestamp) between '2020-03-15' and '2020-04-16'
and type = "story"
ORDER BY score asc
LIMIT 25
```

- Cine este autorul celui mai prost articol din 2020?

```
SELECT f.by
FROM `bigquery-public-data.hacker_news.full` as f
WHERE Date(timestamp) between '2020-01-01' and '2020-12-31'
and type = "story"
ORDER BY score asc
LIMIT 1
```

oeFrha

- Care sunt link-urile si titlurile pentru top 20 articole din vara lui 2017?

```
SELECT f.title, f.url
FROM `bigquery-public-data.hacker_news.full` as f
```

```
WHERE Date(timestamp) between '2017-06-01' and '2017-08-31'
and type = 'story'
ORDER BY score desc
LIMIT 20
```

Exercitii la clasa - 29 oct 2020

- Folosind tabela stories, extrageti o lista cu toti autorii unici al caror articole au un scor intre 100 si 1000 dar diferit de 500, sunt mai noi de 2015, si au un numar de copii diferit de 100,110,120,130

```
SELECT DISTINCT author FROM `bigquery-public-data.hacker_news.stories`
WHERE score BETWEEN 100 and 1000
and score !=500
and date(time_ts)> '2015-01-01'
and descendants not in (100, 110, 120, 130)
```

- Folosind tabela stories, extrageti o lista cu toate titlurile unice de articole din anul 2015, cu mai putin de 50 de copii, care au url-ul pe protocolul http si al caror text se termina in ro

```
SELECT DISTINCT title FROM `bigquery-public-data.hacker_news.stories`
WHERE date(time_ts) BETWEEN '2015-01-01' AND '2015-12-31'
AND descendants < 50
AND url like 'http%'
AND text like '%ro'
```

- Titlul a top 50 cel mai noi articole cele mai noi care sunt dead si autorul lor

```
SELECT title, author FROM `bigquery-public-data.hacker_news.stories`
WHERE dead =true
ORDER BY time_ts desc
LIMIT 50
```

- Textul si data a Top 300 de comentarii dupa ranking din 1 Februarie 2017, scrise de autori al caror nume incep cu litera j, si al caror text contine caracterele a si c despartite de oricare 3 caractere

```
SELECT f.text as Text, date(timestamp) as Date
FROM `bigquery-public-data.hacker_news.full` as f
WHERE f.type = "comment"
AND date(timestamp) = "2017-02-01"
AND f.by LIKE "j%"
AND f.text LIKE "%a__c%"
ORDER BY ranking desc
LIMIT 300
```

- Titlul, autorul si data a 1200 de sondaje de opinie care au un scor mai mare 50, mai mic de 500, dar diferit de 79, 121 sau 417

```
SELECT f.title as Title, f.by as Author, date(timestamp) as Date
FROM `bigquery-public-data.hacker_news.full` as f
WHERE f.type = "poll"
AND f.score BETWEEN 50 AND 500
AND f.score NOT IN (79, 121, 417)
LIMIT 1200
```

- Selectati autorul si id-ul a 500 de comentarii din perioada 1 Martie 2017 - 11 Iulie 2019, dar nu din Februarie 2018 si nici al caror text contine cuvantul apple

```
SELECT f.by as Author, f.id as ID
FROM `bigquery-public-data.hacker_news.full` as f
WHERE f.type = "comment"
AND date(timestamp) BETWEEN "2017-03-01" AND "2019-07-11"
AND date(timestamp) NOT BETWEEN "2018-02-01" AND "2018-02-28"
AND f.text NOT LIKE " %apple% "
LIMIT 500
```

- Vreau toate datele mai putin textul si titlul sondajelor cu optiuni postate in 2018 si 2019 al caror text contine mentiuni despre limba engleza

```
SELECT * EXCEPT (text, title)
FROM `bigquery-public-data.hacker_news.full` as f
WHERE f.type = "pollopt"
AND date(timestamp) BETWEEN "2018-01-01" AND "2019-12-31"
AND f.text LIKE "%english%"
```

- Vreau o lista cu toti autorii unici care au postat job-uri pe site-ul jobs.lever.co inainte de toamna lui 2017

```
SELECT DISTINCT f.by as Author
FROM `bigquery-public-data.hacker_news.full` as f
WHERE f.type = "job"
AND date(timestamp) < "2017-09-01"
AND f.url LIKE "%jobs.lever.co%"
```

- Vreau o lista a titlurilor a top 10 articole dupa scor cu intre 1 si 7 descendenti, dar diferit de 3 sau 5 si al caror text contine ase dar nu contine tan, din trimestrele 2 si 3 din 2019

```
SELECT f.title as Title
FROM `bigquery-public-data.hacker_news.full` as f
WHERE f.type = "story"
AND f.descendants BETWEEN 1 AND 7
AND f.descendants NOT IN (3, 5)
AND f.text LIKE "%ase%"
AND f.text NOT LIKE "%tan%"
AND date(timestamp) BETWEEN "2019-04-01" AND "2019-09-30"
ORDER BY score DESC
LIMIT 10
```

- *De ce nu merge cu f.date(timestamp) ?*

Tema - 3 noiembrie 2020

- Vreau o lista a titlurilor a top 25 articole cu cei mai putini descendenti cu scor mai mare de 1 mai mic de 21 si diferit de 14 si al caror text contine bar dar nu contine foo, din anul american fiscal 2019

```
SELECT f.title as Title
FROM `bigquery-public-data.hacker_news.full` as f
WHERE f.type = "story"
AND date(timestamp) BETWEEN "2018-07-01" AND "2019-06-30"
AND f.score BETWEEN 1 AND 21
AND f.score != 14
AND f.text LIKE "%bar%"
AND f.text NOT LIKE "%foo%"
ORDER BY f.descendants ASC
LIMIT 25
```

- Folosind tabela stories, vreau url-urile tuturor articolelor care au fost publicate de catre altcineva decat autorul lor si care au un scor negativ

```
SELECT s.url as URL
FROM `bigquery-public-data.hacker_news.stories` as s
WHERE s.by != s.author
AND s.score < 0
```

- Vreau lista tuturor ID-urile articolelor care au fost sterse in vara lui 2018, excluzand zilele de sarbatoare legala din Romania.

```
SELECT s.id as ID
FROM `bigquery-public-data.hacker_news.stories` as s
WHERE s.dead = true
AND date(time_ts) BETWEEN "2018-06-01" AND "2018-08-31"
AND date(time_ts) != "2018-06-01"
AND date(time_ts) != "2018-08-15"
```

- Toate coloanele mai putin textul si titlul a Top 850 cele mai vechi postari de job-uri legate de management, dar nu mai vechi de 14 Martie 2012 ora 3:35 dupa-amiaza UTC

```
SELECT * EXCEPT (text, title)
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type = "job"
AND text LIKE "%management%"
AND timestamp < "2012-03-14 15:35:00 UTC"
ORDER BY timestamp ASC
LIMIT 850
```

- Id-ul si titlul articolelor din anul 2016, dar nu din Martie postate pe ycombinator.com sau pe techcrunch.com al caror text contine cuvintele startup,money

```
SELECT s.id as ID, s.title as Title
FROM `bigquery-public-data.hacker_news.stories` as s
WHERE date(time_ts) BETWEEN "2016-01-01" AND "2016-12-31"
AND date(time_ts) NOT BETWEEN "2016-03-01" AND "2016-03-31"
AND s.url LIKE "%ycombinator.com%"
AND s.url LIKE "%techcrunch.com%"
AND text LIKE "%startup%"
AND text LIKE "%money%"
AND type = "story"
```

- Id-ul si tipul de continut al tuturor elementelor ce tin de autori al caror nume contine 3 litere e cu separate de cate un caracter si care au fost publicate intre orele 10:30 dimineata si 4:45 dupamiaza

```
SELECT f.id as ID, f.type as Type
FROM `bigquery-public-data.hacker_news.full` as f
WHERE f.by LIKE "%e_e_e%"
AND time(timestamp) BETWEEN "10:30:00" AND "16:45:00"
```

- Titlul, url-ul si scorul a top 35 job-uri dupa scor postate in sectiuni de careers a site-ului mai noi de 1 Martie 2013, insa care au un scor mai mic de 80

```
SELECT f.title as Title, f.url as URL, f.score as Score
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type = "job"
AND f.url LIKE "%/careers%"
AND date(timestamp) > "2013-03-01"
AND f.score < 80
ORDER BY score DESC
LIMIT 35
```

- Titlul, autorul si timpul publicarii pentru top 5 sondaje dupa scor legate de Facebook

```
SELECT f.title as Title, f.by as Author, time(timestamp) as Time
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type = "poll"
AND f.title LIKE "%facebook%"
ORDER BY score DESC
LIMIT 5
```

- Id-ul, autorul si tipul pentru elementele ce au fost sterse cu un numar de descendenti diferit de 11,12,13,14 sau 15 si mai mare de 18, excluzand sondajele atat cele cu optiuni cat si cele simple

```
SELECT f.id as ID, f.by as Author, f.type as Type
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type NOT IN ( "poll", "pollopt")
AND f.dead = true
AND f.descendants NOT BETWEEN 11 AND 15
AND f.descendants > 18
```

- Titlul articolelor scrise de autori al caror nume incepe cu J dar nu cu j, din perioada Aprilie 2012 - 17 Mai 2018, insa doar din intervalul orar 8-20, in ordinea crescatoare a scorului

```
SELECT f.title as Title
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type = "story"
AND f.by LIKE "J%"
AND f.by NOT LIKE "j%"
AND date(timestamp) BETWEEN "2012-04-01" AND "2018-05-17"
```

AND time(timestamp) BETWEEN "08:00:00" AND "20:00:00"
ORDER BY score ASC

Tema - 5 noiembrie 2020

- Titlul, autorul si timpul publicarii pentru top 25 sondaje dupa scor legate de retelele de socializare Facebook si Twitter

SELECT f.title as Title, f.by as Author, time(timestamp) as Time
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type = "poll"
AND (title LIKE "%facebook%" OR title LIKE "%twitter%")
AND (text LIKE "%facebook%" OR text LIKE "%twitter%")
ORDER BY score DESC
LIMIT 25

- Numarul unic de utilizatori care au postat job-uri de Python in anii 2015 si 2016 insa doar cele postate in intervalul orar 10 dimineata - 10 seara

SELECT
COUNT (DISTINCT f.by) as Users
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type = "job"
AND title LIKE "python"
AND date(timestamp) BETWEEN "2015-01-01" AND "2016-12-31"
AND time(timestamp) BETWEEN "10:00:00" AND "22:00:00"

Exercitii la clasa - 5 noiembrie 2020

- Vreau scorul mediu si numarul de articole la nivel de zi si autor scrise de unul din autorii abennett sau taylorbuley sau exolymph, cu url pe protocol securizat in ordine calendaristica

SELECT
avg(score) as Average_score,
count(*) as Number_of_stories,
f.by as Author,
date(timestamp) as Date
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type = "story"
AND f.by IN ("abenett", "taylorbuley", "exolymph")

AND url LIKE "https%"
GROUP BY Author, Date
ORDER BY Date

- Top 15 utilizatori al caror nume incepe cu e sau f dupa numarul de comentarii cu mai putin de 175 de descendenti postate in perioada 16 Ianuarie 2016 - 19 August 2019

SELECT
f.by as Users,
count(*) as Number_of_comments
FROM `bigquery-public-data.hacker_news.full` as f
WHERE (f.by LIKE "e%" OR f.by LIKE "f%")
AND descendants < 175
AND date(timestamp) BETWEEN "2016-01-16" AND "2019-08-19"
AND type = "comment"
GROUP BY Users
ORDER BY Number_of_comments DESC
LIMIT 15

- Evolutia numarului de elemente care sunt deleted dupa tipul lor care au fost publicate in perioada 2012 - 2020

SELECT
date(timestamp) as Date,
type as Type,
count(*) as Number_of_elements
FROM `bigquery-public-data.hacker_news.full` as f
WHERE deleted = true
AND date(timestamp) >"2012-01-01"
GROUP BY Date, Type
ORDER BY Date ASC

- Vreau numarul unic de autori la nivel de zi care au scris comentarii in anul 2020 si al caror text contine cuvantul awesome sau al caror titlu contine cuvantul amazing

SELECT
f.by as Author, date(timestamp) as Date,
count(distinct f.by) as NumberOfAuthors
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type ="comment"
AND date(timestamp) >"2020-01-01"
AND (text LIKE "%awesome%" OR title LIKE "amazing")

GROUP BY Date, Author
ORDER BY NumberOfAuthors

- Numarul mediu, minim si maxim de descendenti la nivel de tip de continut pentru job-uri si articole din anii 2016,2017 si 2019 de pe iam-magazine.com

```
SELECT  
descendants as Descendants,  
type as Type,  
AVG(Descendants) as Average_Descendants,  
MIN(Descendants) as Minimum_Descendants,  
MAX(Descendants) as Maximum_Descendants  
FROM `bigquery-public-data.hacker_news.full` as f  
WHERE type IN ("job", "story")  
AND (Date(timestamp) BETWEEN "2016-01-01" AND "2017-12-31" OR  
date(timestamp) BETWEEN "2019-01-01" AND "2019-12-31")  
AND url LIKE "%iam-magazine.com%"  
GROUP BY Descendants, Type
```

- Ranking-ul minim si maxim la nivel de autor pentru comentariile din Martie 2007,2008 si 2011 care nu au fost sterse sau eliminate

```
SELECT  
author as Author,  
MIN (ranking) as MINrank,  
MAX (ranking) as MAXrank,  
Date(time_ts) as Date  
FROM `bigquery-public-data.hacker_news.comments`  
WHERE (date(time_ts) BETWEEN "2007-03-01" AND "2007-03-31" or  
date(time_ts) BETWEEN "2008-03-01" AND "2008-03-31" or date(time_ts)  
BETWEEN "2011-03-01" AND "2011-03-31")  
AND (deleted is null and dead is null)  
GROUP BY Author, Date
```

- Top 35 zile calendaristice dupa numarul de articole care sunt dead din perioada Aprilie - Iunie si Octombrie - Decembrie din anul 2009

```
SELECT  
count(*) as Number_of_stories,  
Date(time_ts) as Date  
FROM `bigquery-public-data.hacker_news.stories`  
WHERE (date(time_ts) BETWEEN "2009-04-01" AND "2009-06-30" or  
date(time_ts) BETWEEN "2009-10-01" AND "2009-12-31")  
AND dead is true
```

GROUP BY Date
ORDER BY Number_of_stories DESC
LIMIT 35

- Scorul mediu, numarul maxim de descendenti si numarul de elemente dupa tipul de continut postate pe site-uri de tip .org, dar nu de tip .com sau .de

SELECT
avg(score) as Average_Score,
MAX(descendants) as Maximum_Descendants,
count(*) as Number_of_Elements,
type as Type
FROM `bigquery-public-data.hacker_news.full`
WHERE url LIKE "%.org%"
AND (url NOT LIKE "%.com%" AND url NOT LIKE "%.de%")
GROUP BY Type

- Data primului si ultimului comentariu pentru fiecare utilizator al carui nume incepe cu oricare din literele a,b,c

SELECT
date(MIN(time_ts)) as First_comment,
date(MAX(time_ts)) as Last_comment,
author as Author
FROM `bigquery-public-data.hacker_news.comments` as c
WHERE (c.by LIKE "%a" or c.by LIKE "%b" or c.by LIKE "%c")
GROUP BY Author

Tema - 10 noiembrie 2020

- Pentru fiecare tip de retea vreau sa vad care a fost viteza medie si numarul total de sateliti in Barcelona in perioada 1 Ianuarie 14 Martie 2015

SELECT
IFNULL(net, "2G") as Network_Type,
avg(speed) as Average_Speed,
sum(satellites) as Number_of_Satellites
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE town_name = "Barcelona"
AND date BETWEEN "2015-01-01" AND "2015-03-14"
GROUP BY Network_Type

- Vreau numarul unic de operatori si numarul mediu de sateliti care au deservit zonele Girona si Lleida in vara si toamna lui 2016

```
SELECT
avg(satellites) as Average_Number_of_Satellites,
count(distinct operator) as Unique_Operators,
town_name as City
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE (postal_code LIKE "17%" OR postal_code LIKE "25%")
AND date BETWEEN "2016-06-01" AND "2016-11-30"
GROUP BY City
```

- Vreau top 10 retele care au servit zona Tarragona in timpul verii, in toti anii si puterea lor medie, ordonate alfabetic descrescator

```
SELECT
avg(signal) as Average_Power,
network as Network
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE postal_code LIKE "43%"
AND (date BETWEEN "2015-06-01" AND "2015-08-31" OR date BETWEEN
"2016-06-01" AND "2016-08-31" OR date BETWEEN "2017-06-01" AND
"2017-08-31")
--AND extract(month from date) IN (6,7,8)
GROUP BY Network
ORDER BY Network DESC
LIMIT 10
```

How to limit for only one group by?

- Vreau numarul de masuratori din luna August din 2017 dupa tipul de provider al pozitiei, ale retelelor Orange si Vodafone si cu o precizie mai mica de 25 din orasele Tarragona si Girona

```
SELECT
count(*) as Number_of_Measurements,
provider as Provider,
network as Network
```

```

FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE town_name IN ("Tarragona", "Girona")
AND date BETWEEN "2017-08-01" AND "2017-08-31"
AND (operator = "Orange" OR operator LIKE "vodafone%")
AND precission < 25
GROUP BY Provider, Network

```

- Vreau sa vad cum au evoluat viteza medie si precizia maxima in zona Barcelona pe cei 3 ani

```

SELECT
extract(year from date) as Year
avg(speed) as Average_Speed,
max(precission) as Maximum_Precission
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE postal_code LIKE "08%"
--AND date BETWEEN "2015-01-01" AND "2017-12-31"
GROUP BY Year
ORDER BY Year ASC

```

- Vreau numarul de masuratori si precizia medie dupa tipul de activitate al utilizatorului din retelele al caror nume se termina in emergencia din zonele Lleida si Girona in trimestrele 2 si 3 din toti anii sortate crescator dupa precizia medie

```

SELECT
activity as Activity,
count(*) as Number_of_Measurements,
avg(precission) as Average_Precission,
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE (postal_code LIKE "25%" OR postal_code LIKE "17%")
AND network LIKE "%emergencia"
AND (extract(quarter from date)=2 or extract(quarter from date)=3)
GROUP BY Activity
ORDER BY Average_Precission ASC

```

- Vreau toate orasele si numarul total de sateliti Orange sau Vodafone care i-au deservit in 2015 sau 2016 ordonate crescator dupa deviatia standard a vitezei

```

SELECT
town_name as City,
sum(satellites) as Number_of_Satellites,
stddev(speed) as Speed_Standard_Deviation,
network as Network
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE (network LIKE "%orange%" or network LIKE "%vodafone%")
AND date BETWEEN "2015-01-01" AND "2016-12-31"
GROUP BY City, Network
ORDER BY Speed_Standard_Deviation ASC

```

- Data primei si ultimei masuratori pentru fiecare operator la nivel de tip de retea pentru masuratori cu o viteza mai mare de 7 sau o precizie intre 2 si 250

```

SELECT
min(date) as First_Measure,
max(date) as Last_Measure,
operator as Operator,
IFNULL(net, "2G") as Network_Type
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE speed > 7
AND precision BETWEEN 2 AND 250
GROUP BY Operator, Network_Type
ORDER BY Operator

```

- Providerul cu cele mai multe masuratori in anii 2015 si 2016 din districtele Lleida si Barcelona

```

SELECT
provider as Provider,
count(*) as Number_of_Measurements
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE date BETWEEN "2015-01-01" AND "2016-12-31"
AND (postal_code LIKE "08%" OR postal_code LIKE "25%")
GROUP BY Provider

```

ORDER BY Number_of_Measurements DESC
LIMIT 1

- Pentru fiecare oras din districtul Tarragona vreau numarul total de sateliti si precizia medie dupa tipul de activitate al utilizatorului

SELECT
town_name as City,
sum(satellites) as Number_of_Satellites,
avg(preciision) as Average_Precision,
activity as User_Activity
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE postal_code LIKE "43%"
GROUP BY City, User_Activity

- Puterea medie a semnalului si perioada pe care acestea au fost facute pentru masuratorile din anul 2016 pe retele mai noi de 3G in functie de starea telefonului. Perioada este formata din data primei masuratori si data ultimei masuratori

SELECT
avg(signal) as Average_Signal,
description as Phone_State,
MIN(date) as First_Measure,
MAX(date) as Last_Measure
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE date BETWEEN "2016-01-01" AND "2016-12-31"
AND net IN ("4G", "5G")
GROUP BY Phone_State

- Viteza medie pentru fiecare oras din districtele Girona, Lleida si Tarragona la nivel de operator

SELECT
avg(speed) as Average_Speed,
town_name as City,
operator as Operator
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE (postal_code LIKE "17%"OR postal_code LIKE "25%"OR postal_code LIKE "43%")

GROUP BY City, Operator

Tema - 17 noiembrie 2020

- Folosind tabela stories, extrageți o listă cu toți autorii unici al căror articol are un scor între 75 și 2500 diferit de 450, 551, sau 783, și nu din intervalul 565-695, sunt mai noi de 2014, și au un număr de copii diferit de mai mare de 3

```
SELECT  
distinct author as Author,  
time_ts as Date,  
descendants as Number_of_descendants,  
score as Score  
FROM `bigquery-public-data.hacker_news.stories`  
WHERE score BETWEEN 75 AND 2500  
AND score NOT IN (450, 551,783)  
AND score NOT BETWEEN 565 AND 695  
AND time_ts > "2014-12-31"  
AND descendants <3
```

- Folosind tabela stories, extrageți o listă cu toate titlurile unice de articole din anul 2017, cu mai puțin de 25 de copii înșă mai mare de 3, care au url-ul pe protocolul https și al căror text conține fast

```
SELECT  
distinct title as Title  
FROM `bigquery-public-data.hacker_news.stories`  
WHERE extract(year from time_ts)=2017  
AND descendants BETWEEN 3 AND 25  
AND url LIKE "https%"  
AND text LIKE "%fast%"
```

- Autorul și Titlul a top 40 cele mai vechi articole care sunt dead

```
SELECT  
author as Author,  
title as Title,  
time_ts as Timestamp  
FROM `bigquery-public-data.hacker_news.stories`  
WHERE dead=true  
ORDER BY Timestamp ASC  
LIMIT 40
```


- Folosind tabela comments extrageti textul si data a Top 215 cele mai proaste comentarii dupa ranking din perioada 14 August 2008 - 15 Martie 2011 , scrise de autori al caror nume se termina ib, si al caror text contine caracterele d si o despartite de oricare 4 caractere

```
SELECT
text as Text,
time_ts as Date,
ranking as Ranking
FROM `bigquery-public-data.hacker_news.comments`
WHERE time_ts BETWEEN "2008-08-14" AND "2011-03-15"
AND author LIKE "%ib"
AND text LIKE "%d____o%"
ORDER BY ranking ASC
LIMIT 215
```

- Url-ul, autorul, data si timpul a 300 de sondaje de opinie care au un scor mai mare 25, mai mic de 600, dar diferit de 65, 145 sau 301

```
SELECT
url as URL,
f.by as Author,
timestamp as Date_Time
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type="poll"
AND score BETWEEN 25 AND 600
AND score NOT IN (65,145,301)
LIMIT 300
```

- Selectati autorul si id-ul a 450 de sondaje de opinie cu optiuni din perioada 13 Ianuarie 2015 - 18 Decembrie 2018, dar nu din Iulie 2016 si nici al caror text nu se termina in pear

```
SELECT
f.by as Author,
id as ID
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type="pollopt"
AND date(timestamp) BETWEEN "2015-01-13" AND "2018-12-18"
AND date(timestamp) NOT BETWEEN "2016-07-01" AND "2016-07-30"
```

**AND text NOT LIKE "%pear"
LIMIT 450**

- Vreau toate datele mai puțin textul, url-ul și autorul sondajelor de opinie postate în 2017 și 2020 al căror text conține mențiuni despre limba franceză

**SELECT * except (text, url, `by`)
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type="poll"
AND (extract(year from timestamp)=2017 OR extract(year from timestamp)=2020)
AND text LIKE "%french%"**

- Vreau o listă cu toți autorii unici care au postat job-uri pe site-ul jobs.lever.co înainte de primăvara lui 2019 dar al căror nume nu începe cu litera a

**SELECT
distinct f.by as Author
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type='job'
AND url LIKE "%jobs.lever.co%"
AND date(timestamp) < "2019-03-01"
AND f.by NOT LIKE "a%"**

- Vreau o listă a url-urilor a top 25 articole după număr de descendenți cu un scor între 4 și 25 descendenți, dar diferit de 7,8,9 sau 10 și al căror text conține mar dar nu conține bon, din trimestrul 4 din 2017 și trimestrul 1 din 2018

**SELECT
url as URL
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type='story'
AND descendants BETWEEN 4 AND 25
AND descendants NOT IN (7,8,9,10)
AND text LIKE "%mar%"
AND text NOT LIKE "%bon%"
AND (extract(year from timestamp)=2017 AND extract(quarter from timestamp)=4) OR (extract(year from timestamp)=2018 AND extract(quarter from timestamp)=1)**

**ORDER BY score DESC
LIMIT 25**

- Vreau o lista a titlurilor si autorilor a top 10 articole cu cel mai mic cu scor cu un numar de descendenti mai mic 21 si diferit de 14 si al caror text contine rab dar nu contine ti, din perioada primului mandat de presedinte al lui Obama

**SELECT
title as Title,
f.by as Author
FROM `bigquery-public-data.hacker_news.full` as f
WHERE descendants <21
AND descendants != 14
AND text LIKE "%rab%"
AND text NOT LIKE "%ti%"
and date(timestamp) BETWEEN "2009-01-20" AND "2012-11-06"
ORDER BY score ASC
LIMIT 10**

- Folosind tabela stories, vreau titlurile tuturor articolelor care au fost publicate de catre altcineva decat autorul lor si care au un scor pozitiv dar mai mic de 1000

**SELECT title as Title
FROM `bigquery-public-data.hacker_news.stories` as s
WHERE s.by!= author
AND score BETWEEN 0 AND 1000**

- Vreau unica (ID) a tuturor autorilor al caror articole au fost sterse insa au fost publicate in primavara lui 2017, excluzand zilele de sarbatoare legala din Romania

**SELECT id as ID
FROM `bigquery-public-data.hacker_news.stories` as s
WHERE deleted = true
AND date(time_ts) between '2017-03-01' AND '2017-05-31'
AND date(time_ts) NOT IN ('2017-04-16','2017-04-17', '2017-05-01')**

- Toate coloanele mai putin textul si titlul a Top 30 cele mai recente postari de job-uri legate de Java, dar nu mai vechi de 25 Septembrie 2020 ora 4:48 dupa-amiaza UTC

```
SELECT * except (text, title)
FROM `bigquery-public-data.hacker_news.full`
WHERE type='job'
AND title LIKE '%Java%'
AND timestamp > '2020-09-15 16:48:00'
ORDER BY timestamp DESC
LIMIT 30
```

- Url-ul si titlul articolelor din perioada 2014 - 2018, dar nu din Aprilie 2015 sau Mai 2016 postate pe ycombinator.com sau pe techcrunch.com al caror text contine cuvintele growth hacking, investment

```
SELECT url as URL,
title as Title
FROM `bigquery-public-data.hacker_news.full`
WHERE type = 'story'
AND timestamp BETWEEN '2014-01-01' AND '2018-12-31'
AND (extract(year from timestamp)!=2015 AND extract(month from timestamp)!=4)
AND (extract(year from timestamp)!=2016 AND extract(month from timestamp)!=5)
AND (url LIKE '%ycombinator.com%' OR url LIKE '%techcrunch.com%')
AND (text LIKE '%growth hacking%' OR text LIKE '%investment%')
```

- Id-ul si tipul de continut al tuturor elementelor ce tin de autori al caror nume contine literele a si e de un caracter si care au fost publicate intre orele 8:45 dimineata si 6:15 dupamiaza

```
SELECT id as ID,
type as Type
FROM `bigquery-public-data.hacker_news.full` as f
WHERE f.by LIKE '%a_e%'
AND time(timestamp) BETWEEN '08:45:00' AND '18:15:00'
```

- Titlul, url-ul si scorul a top 35 job-uri dupa scor postate in sectiuni de jobs a site-ului mai noi de 15 August 2015, insa care au un scor mai mare de 27

```

SELECT
title as Title,
url as URL,
score as Score
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type='job'
AND url LIKE '%job%'
AND date(timestamp) > '2015-08-15'
AND score>27
ORDER BY score DESC
LIMIT 35

```

- Titlul, autorul si timpul publicarii pentru top 15 cele mai proaste sondaje dupa scor legate de Twitter

```

SELECT
title as Title,
f.by as Author,
time(timestamp) as Time_of_post
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type="poll"
AND (title LIKE "%Twitter%" or text LIKE "%Twitter%")
ORDER BY score ASC
LIMIT 15

```

- Id-ul, autorul si tipul pentru elementele ce au fost sterse cu un numar de descendenti diferit de 31,32,33,34 sau 65 si mai mare de 9, excluzand sondajele atat cele cu optiuni cat si cele simple

```

SELECT
id as ID,
f.by as Author,
type as Type_of_content
FROM `bigquery-public-data.hacker_news.full` as f
WHERE deleted = true
AND descendants NOT IN (31,32,33,34,65)
AND descendants>9
AND type NOT IN ('poll', 'pollopt')

```

- Folosind tabela comments extrageti Id-ul si parintele comentariilor scrise de autori al caror nume incepe cu F dar nu cu f, din perioada Iulie 2007 - 3 Septembrie 2011, insa doar din intervalul orar 10-16, in ordinea decrescatoare a ranking-ului

```
SELECT
id as ID,
parent as Parent
FROM `bigquery-public-data.hacker_news.comments`
WHERE author LIKE 'F%'
AND author NOT LIKE 'f%'
AND date(time_ts) BETWEEN '2007-07-01' AND '2011-09-03'
AND time(time_ts) BETWEEN '10:00:00' AND '16:00:00'
ORDER BY ranking DESC
```

- Titlul, data si timpul a 45 de job-uri care au un scor mai mare 1, mai mic de 705, dar nu din intervalele 100-151 si nici 201-250

```
SELECT
title as Title,
timestamp as Date_Time
FROM `bigquery-public-data.hacker_news.full`
WHERE type='job'
AND score BETWEEN 1 AND 705
AND score NOT BETWEEN 100 AND 151
AND score NOT BETWEEN 201 AND 250
LIMIT 45
```

- Autorul, ranking-ul, data si timpul tuturor comentariilor din Ianuarie - Martie 2008 si Aprilie-Mai 2009 al caror id este intre 20000 si 30000 sau 40000 si 70000

```
SELECT
author as Author,
ranking as Ranking,
time_ts as Date_Time
FROM `bigquery-public-data.hacker_news.comments`
WHERE date(time_ts) BETWEEN '2008-01-01' AND '2008-03-31'
AND date(time_ts) BETWEEN '2009-04-01' AND '2009-05-31'
AND (id BETWEEN 20000 AND 30000 OR id=40000)
```

- Titlul, autorul si timpul publicarii pentru top 50 cele mai proaste sondaje dupa scor legate de retelele de socializare Facebook, Twitter, Instagram si Snapchat

```
SELECT
title as Title,
f.by as Author,
time(timestamp) as Posting_time
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type='poll'
AND (title LIKE '%Facebook%' OR title LIKE '%Twitter%' OR title LIKE '%Instagram%' OR title LIKE '%Snapchat%')
ORDER BY score ASC
LIMIT 50
```

- Top 55 autori dupa numarul mediu de descendenti al articolelor publicate in perioadele Iunie - August 2018 si Septembrie - Decembrie 2019

```
SELECT
f.by as Author,
avg(descendants) as Average_Nr_of_Descendants
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type='story'
AND date(timestamp) BETWEEN '2018-06-01' AND '2018-08-31'
AND date(timestamp) BETWEEN '2019-09-01' AND '2019-12-31'
GROUP BY Author
ORDER BY Average_Nr_of_Descendants
LIMIT 55
```

- Numarul de comentarii din anul 2015 scrise de utilizatori al caror nume incepe oricare din literele d,f,k

```
SELECT
count(*) as Nr_of_comments
FROM `bigquery-public-data.hacker_news.comments` as c
WHERE extract(year from time_ts)=2015
AND (c.by LIKE 'd%' OR c.by LIKE 'f%' OR c.by LIKE 'k%')
```

- Numarul unic de utilizatori si numarul de job-uri job-uri legate Scala in anii 2014 si 2018 insa doar cele postate in intervalul orar 8 dimineata - 6 seara

```

SELECT
count(distinct 'f.by') as Unique_nr_of_Users,
count (*) as Nr_of_jobs
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type='job'
AND text LIKE '%Scala%'
AND (extract(year from timestamp)=2014 OR extract(year from
timestamp)=2018)
AND time(timestamp) BETWEEN '08:00:00' AND '18:00:00'

```

- Vreau numarul mediu de descendenti si numarul de comentarii la nivel de zi si autor scrise de unul din autorii dragonwriter sau TeMPOraL sau dang sau tptacek, cu url pe protocol nesecurizat in ordine calendaristica

```

SELECT
avg(descendants) as Average_Descendants,
count(*) as Nr_of_comments,
f.by as Author,
timestamp as Date_Time
FROM `bigquery-public-data.hacker_news.full` as f
WHERE (title LIKE '%dragonwriter%' OR title LIKE '%TeMPOraL%' OR title LIKE
'%dang%' OR title LIKE '%tptacek%')
AND url LIKE '%http%'
GROUP BY Author, Date_Time
ORDER BY timestamp ASC

```

- Top 25 utilizatori al caror nume incepe cu g sau i dupa numarul de articole cu un score mai mic 175 postate in perioada 28 Iulie 2016 - 30 Mai 2018

```

SELECT
f.by as Users,
count(*) as Nr_of_Stories
FROM `bigquery-public-data.hacker_news.full` as f
WHERE (f.by LIKE 'g%' OR f.by LIKE 'i%')
AND type='story'
AND score<175
AND date(timestamp) BETWEEN '2016-07-28' AND '2018-05-30'
GROUP BY Users
ORDER BY Nr_of_Stories DESC
LIMIT 25

```


- Evolutia la nivel de zi a numarului de elemente care sunt dead sau deleted dupa tipul lor care au fost publicate in perioada 2015 - 2018 dar nu din anul 2017

```
SELECT
date(timestamp) as Date,
count(*) as Nr_of_Elements,
type as Type
FROM `bigquery-public-data.hacker_news.full`
WHERE (dead=true OR deleted=true)
AND date(timestamp) BETWEEN '2015-01-01' AND '2018-12-31'
AND extract(year from timestamp)!=2017
GROUP BY Type, Date
ORDER BY Date
```

- Vreau numarul unic de autori la nivel de zi care au postat job-uri in anul 2017 si al caror text contine cuvantul Ruby sau al caror titlu contine cuvantul Rails

```
SELECT
count (distinct 'f.by') as Author,
date(timestamp) as Date
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type='job'
AND extract (year from timestamp)=2017
AND (text LIKE '%Ruby%' OR title LIKE '%Rails%')
GROUP BY Date
```

- Scorul mediu, minim si maxim la nivel de tip de continut pentru sondaje de opinie cu optiuni sau fara din anii 2018,2019 si 2013 de pe news.ycombinator.com

```
SELECT
avg(score) as Average_Score,
min(score) as Min_Score,
max(score) as Max_Score,
type as Type
FROM `bigquery-public-data.hacker_news.full`
WHERE (type='pollopt' OR type='poll')
AND extract(year from timestamp) IN (2013,2018,2019)
AND url LIKE '%news.ycombinator.com%'
```

GROUP BY Type

- Numarul de comentarii, ranking-ul mediu si deviatia lui standard la nivel de autor pentru comentariile din Decembrie 2009,2010 si 2012 care nu au fost sterse sau eliminate

```
SELECT
count(*) as Nr_of_Comments,
avg(ranking) as Average_Ranking,
stddev(ranking) as Ranking_Standars_Deviation,
author as Author
FROM `bigquery-public-data.hacker_news.comments`
WHERE extract (year from time_ts) IN (2009, 2010, 2012)
AND extract (month from time_ts)=12
AND (dead!=true or deleted!=true)
GROUP BY Author
```

- 15 zile calendaristice cu cele mai putine articole dead din perioada Ianuarie - Iulie si Septembrie - Decembrie din anul 2011

```
SELECT
count(*) as Nr_of_Stories,
date(timestamp) as Date
FROM `bigquery-public-data.hacker_news.full`
WHERE dead=true
AND type='story'
AND date(timestamp) BETWEEN '2011-01-01'AND '2011-07-30'
AND date(timestamp) BETWEEN '2011-09-01'AND '2011-12-31'
GROUP BY Date
ORDER BY Nr_of_Stories ASC
LIMIT 15
```

- Numarul mediu de descendenti, scorul minim si numarul de elemente dupa tipul de continut postate pe site-uri de tip .net, dar nu de tip .org sau .io

```
SELECT
avg(descendants) as Average_nr_of_Descendants,
min(score) as Min_Score,
count(*) as Nr_of_Elements,
```

```

type as Type
FROM `bigquery-public-data.hacker_news.full`
WHERE url LIKE '%.net%'
AND (url NOT LIKE '%.org%' AND url NOT LIKE '%.io%')
GROUP BY Type

```

- Data primului si ultimului articol publicat pentru fiecare utilizator al carui nume incepe cu oricare din literele d,e,f

```

SELECT
distinct s.by as User,
min(date(time_ts)) as First_Story,
max(date(time_ts)) as Last_Story
FROM `bigquery-public-data.hacker_news.stories` as s
WHERE (s.by LIKE 'd%' OR s.by LIKE 'e%' OR s.by LIKE 'f%')
GROUP BY User
ORDER BY User ASC

```

- Pentru fiecare operator vreau sa vad care a fost viteza medie si numarul total de sateliti in orasele Lleida, Badalona si Sabadell in perioada 5 Martie 2015 - 17 Iulie 2016

```

SELECT
IFNULL(operator, 'Not found') as Operator,
avg(speed) as Average_Speed,
count(*) as Nr_of_satellites
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE town_name IN ('Lleida', 'Badalona', 'Sabadell')
AND date BETWEEN '2015-03-05' AND '2016-07-17'
GROUP BY Operator

```

- Vreau numarul unic de orase si numarul mediu de sateliti care au deservit zonele Tarragona si Barcelona in primavara si iarna lui 2017

```

SELECT
count(distinct town_name) as Unique_nr_of_Cities,
avg(satellites) as Average_nr_of_Satellites,
town_name as City
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`

```

WHERE (postal_code LIKE '43%' OR postal_code LIKE '08%')
AND extract(year from date)=2017
AND extract(month from date) IN (3,4,5,12,1,2)
GROUP BY City

- Vreau top 25 operatori care au servit zona Girona in timpul verii, in toti anii si puterea lor medie, ordonati alfabetic

SELECT
IFNULL(operator, 'Not found') as Operator,
avg(signal) as Average_Power
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE postal_code LIKE '17%'
AND extract(month from date) IN (6,7,8)
GROUP BY Operator
ORDER BY Operator ASC, Average_Power DESC
LIMIT 25

- Vreau numarul de masuratori din luna Mai din 2016 dupa tipul de provider al pozitiei, ale retelelor T-Mobile si Movistar si cu o precizie mai mare de 5 din orasele Barcelona si Lleida

SELECT
count(*) as Nr_of_Measurements,
provider as Provider
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE extract(month from date)=5
AND extract(year from date)=2016
AND (operator='T-Mobile' OR operator='Movistar')
AND precission>5
AND town_name IN ('Barcelona', 'Lleida')
GROUP BY Provider

- Vreau sa vad cum au evoluat numarul mediu de sateliti si precizia minima la nivel de zi in zona Tarragona pe cei 3 ani

SELECT
date as Date,
avg(satellites) as Average_nr_of_Satellites,
min(precission) as Min_Precision
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE postal_code LIKE '43%'

GROUP BY Date
ORDER BY Date ASC

- Vreau numarul de masuratori si puterea maxima a semnalului dupa tipul de activitate al utilizatorului din retelele al caror nume se termina in emergencia din zonele Barcelona si Girona in trimestrele 1 si 4 din toti anii sortate crescator dupa precizia puterea maxima a semnalului

```
SELECT  
count(*) as Nr_of_Measurements,  
max(signal) as Max_Power,  
activity as User_Activity  
FROM  
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`  
WHERE network LIKE '%emergencia'  
AND (postal_code LIKE '08%' OR postal_code LIKE '17%')  
AND extract(quarter from date) IN (1,4)  
GROUP BY User_Activity  
ORDER BY Max_Power ASC
```

- Vreau top 20 orase si numarul total de sateliti Telekom care i-au deservit in Martie, Mai, August si Decembrie 2017 dupa viteza medie

```
SELECT  
town_name as City,  
sum(satellites) as Total_nr_of_Satellites,  
avg(speed) as Average_Speed  
FROM  
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`  
WHERE network LIKE '%Telekom%'  
AND extract(month from date) IN (3,5,8,12)  
AND extract(year from date)=2017  
GROUP BY City  
ORDER BY Average_Speed DESC  
LIMIT 20
```

- Data primei si ultimei masuratori pentru fiecare operator la nivel de tip de retea pentru masuratori cu o putere mai mare de 5 sau o precizie intre 10 si 112

```

SELECT
min(date) as First_Measurement,
max(date) as Last_Measurement,
IFNULL(net, '2G') as Type_of_Net,
operator as Operator
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE (signal>5 OR precission BETWEEN 10 AND 112)
GROUP BY Type_of_Net, Operator

```

- Providerul cu cele mai multe masuratori in anii 2016 si 2017 din toate districtele in afara de Barcelona

```

SELECT
provider as Provider,
count(*) as Nr_of_Measurements
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE extract(year from date) IN (2016,2017)
AND postal_code NOT LIKE '08%'
GROUP BY Provider
ORDER BY Nr_of_Measurements DESC
LIMIT 1

```

- Pentru fiecare operator cu masuratori in districtul Barcelona vreau numarul total de sateliti si precizia medie dupa tipul de activitate al utilizatorului

```

SELECT
IFNULL(operator, 'Not found') as Operator,
sum(satellites) as Total_nr_of_Satellites,
count(*) as Nr_of_Measurements,
avg(precission) as Average_Precision,
activity as User_Activity
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE postal_code LIKE '08%'
GROUP BY Operator, User_Activity
ORDER BY Nr_of_Measurements ASC

```

- Precizia maxima a masuratorilor si perioada pe care acestea au fost facute din anii 2015 si 2017 pe retele mai vechi de 4G in functie de starea telefonului. Perioada este formata din data primei masuratori si data ultimei masuratori

```
SELECT
max(precision) as Max_Precision,
min(date) as First_Measure,
max(date) as Last_Measure,
description as Phone_Status
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE net IN ('2G', '3G')
AND extract(year from date) IN (2015, 2017)
GROUP BY Phone_Status
```

- Puterea media a semnalului pentru fiecare oras din districtele Barcelona la nivel de retea

```
SELECT
avg(signal) as Average_Power,
town_name as City,
IFNULL(net, '2G') as Type_of_Net
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE postal_code LIKE '08%'
GROUP BY City, Type_of_Net
ORDER BY City ASC
```

- Comparatie intre viteza medie in cele 4 districte din Catalonia pe toti cei 3 ani

```
SELECT
CASE
WHEN postal_code LIKE '08%' THEN 'Barcelona'
WHEN postal_code LIKE '17%' THEN 'Girona'
WHEN postal_code LIKE '25%' THEN 'Lleida'
WHEN postal_code LIKE '43%' THEN 'Tarragona'
ELSE 'This should not be here'
END as Catalonia_Districts,
avg(speed) as Average_Speed
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
```

GROUP BY Catalonia_Districts

Exercitii la curs - 19 noiembrie 2020

- Vreau o comparatie intre operatorii Vodafone, Orange, Telekom dupa numarul de masuratori si viteza medie

```
SELECT
CASE
    WHEN (operator LIKE "%Orange%" OR operator LIKE "%orange%") THEN
'Orange'
    WHEN (operator LIKE "%Vodafone%" OR operator LIKE "%vodafone%")
THEN 'Vodafone'
    WHEN (operator LIKE "%Telekom%" OR operator LIKE "%telekom%") THEN
'Telekom'
END as Operator,
count(*) as Nr_of_Measure,
avg(speed) as Average_Speed
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE(operator LIKE "%Vodafone%" OR operator LIKE "%vodafone%" OR
operator LIKE "%Orange%" OR operator LIKE "%orange%" OR operator LIKE
"%Telekom%" OR operator LIKE "%telekom%")
GROUP BY Operator
ORDER BY Average_Speed DESC
```

- Pentru districtele Barcelona si Lleida vreau sa vad numarul de masuratori cu o putere mai mica de 100 pe intervale ce tin de numarul de sateliti. Intervalele de numar de sateliti sunt: intre 1 si 5, intre 6 si 10, intre 10 si 20, mai mult de 20

```
SELECT
CASE
    WHEN satellites BETWEEN 1 AND 5 THEN 'I1'
    WHEN satellites BETWEEN 6 AND 10 THEN 'I2'
    WHEN satellites BETWEEN 11 AND 20 THEN 'I3'
    ELSE 'I4'
END as Satellites_Interval,
count(*) as Nr_of_measures
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE (postal_code LIKE '08%' OR postal_code LIKE '25%')
AND signal <100
```


GROUP BY Satellites_Interval

- Vreau sa vad evolutia la nivel de zi a comparatiei intre masuratorile pentru telefoanele care erau disponibile vs cele indisponibile dupa numarul de masuratori si putere maxima in anul 2017. Telefoanele sunt disponibile doar pentru STATE_IN_SERVICE

```
SELECT  
Date,  
IF(description='STATE_IN_SERVICE', ' Available' , ' Offline' ) as State_Service,  
count(*) as Nr_of_Measurements,  
max(signal) as Max_Signal  
FROM  
bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017  
WHERE extract(year from date) =2017  
GROUP BY Date, State_Service  
ORDER BY Date ASC
```

- Vreau sa vad orasele care in anul 2016 au fost deservite de cel putin trei retele diferite si care a fost viteza medie in aceste orase

```
SELECT  
town_name as City,  
avg(speed) as Average_Speed  
FROM  
bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017  
WHERE extract(year from date)=2016  
GROUP BY City  
HAVING count(distinct network)>=3
```

- Pentru fiecare district din Catalonia vreau sa vad numarul unic de oras din care avem masuratori la nivel de tip de retea

```
SELECT  
CASE  
  WHEN postal_code LIKE '08%' THEN 'Barcelona'  
  WHEN postal_code LIKE '17%' THEN 'Girona'  
  WHEN postal_code LIKE '25%' THEN 'Lleida'  
  WHEN postal_code LIKE '43%' THEN 'Tarragona'  
  ELSE 'Not in Catalonia'  
END as Catalonia_District,  
count(distinct town_name) as Unique_nr_of_Cities,  
IFNULL(net, '2G') as Net  
FROM
```

```
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`  
GROUP BY Catalonia_District, Net  
ORDER BY Catalonia_District ASC, Net ASC
```

- Vreau o comparatie a numarului de masuratori la nivel de anotimp

```
SELECT Season, Nr_of_Measurements  
FROM  
(  
  SELECT  
  CASE  
    WHEN extract(month from date) IN (1,2,12) THEN 'Winter'  
    WHEN extract(month from date) IN (3,4,5) THEN 'Spring'  
    WHEN extract(month from date) IN (6,7,8) THEN 'Summer'  
    WHEN extract(month from date) IN (9,10,11) THEN 'Autumn'  
  END as Season,  
  CASE  
    WHEN extract(month from date) IN (1,2,12) THEN 1  
    WHEN extract(month from date) IN (3,4,5) THEN 2  
    WHEN extract(month from date) IN (6,7,8) THEN 3  
    WHEN extract(month from date) IN (9,10,11) THEN 4  
  END as Season_order,  
  count(*) as Nr_of_Measurements  
FROM  
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`  
GROUP BY Season, Season_order  
)  
ORDER BY Season_order
```

- Comparatie a puterii medii si vitezei maxime la nivel de perioada a zilei in anii 2016 si 2017 pentru orasele din regiunile Girona si Lleida. Avem 5 perioade ale zilei: dimineata intre orele 7 si 12, pranz intre 12 si 16, dupa-amiaza intre 16 si 19, seara intre 19 si 23 si noapte intre 23 si 7 dimineata.

```
SELECT  
IFNULL (CASE  
  WHEN hour BETWEEN "07:00:00" and "11:59:59" THEN "Morning"  
  WHEN hour BETWEEN "12:00:00" and "15:59:59" THEN "Noon"  
  WHEN hour BETWEEN "16:00:00" and "18:59:59" THEN "Afternoon"  
  WHEN hour BETWEEN "19:00:00" and "22:59:59" THEN "Evening"  
  WHEN hour BETWEEN "23:00:00" and "06:59:59" THEN "Night"
```

```

    END, 'IDK') as Period_of_the_Day,
    avg(signal) as Average_Signal,
    max(speed) as Max_Speed
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE extract(year from date) IN (2016,2017)
AND (postal_code LIKE "17%" OR postal_code LIKE "25%")
GROUP BY Period_of_the_Day

```

Tema - 24 noiembrie 2020

- Pentru lichiorurile de cafea si baturile triplu sec vreau numarul de vanzari, valoarea totala de retail a vanzarii, valoarea totala de en-gros a vanzarii si volumul mediu al sticlei.

```

SELECT
CASE
    WHEN (category_name = 'Coffee Liqueurs' OR category_name = 'COFFEE LIQUEURS') THEN 'Coffee Liqueurs'
    WHEN (category_name = 'Triple Sec' OR category_name = 'TRIPLE SEC') THEN 'Triple Sec'
END as Category_Type,
count(*) as Nr_of_Sales,
round(sum(state_bottle_cost*bottles_sold)/100000, 4) as
EnGros_Value_Million_Dollars,
round(sum(sale_dollars)/1000000, 4) as Retail_Value_Million_Dollars,
round(avg(bottle_volume_ml), 2) as Average_volume_of_bottle_ml
FROM `bigquery-public-data.iowa_liquor_sales.sales`
WHERE (category_name = 'Coffee Liqueurs' OR category_name = 'COFFEE LIQUEURS' OR category_name = 'TRIPLE SEC' OR category_name = 'Triple Sec')
GROUP BY Category_Type

```

- Top 15 judete cu vanzari totale de peste 1000\$ in anul 2019 dupa volumul de bauturi vandut

```

SELECT
upper(county) as County,
sum(sale_dollars) as Total_Sales,
sum(volume_sold_liters) as Volume_Sold_Liters

```

```

FROM `bigquery-public-data.iowa_liquor_sales.sales`
WHERE extract(year from date)=2019
GROUP BY County
HAVING sum(sale_dollars)>1000
ORDER BY Volume_Sold_Liters DESC
LIMIT 15

```

- O comparatie dupa numarul de vanzari si valoarea vanzarilor pentru brandurile de lichior de cafea. Brandurile de interes sunt Kahlua, Patron Xo, Iowa Coffee Company, celelalte produse intra la categoria Other Brands

```

SELECT
CASE
  WHEN item_description LIKE '%ahlua%' THEN 'Kahlua'
  WHEN item_description LIKE '%atron Xo%' THEN 'Patron Xo'
  WHEN item_description LIKE '%owa Coffee Company%' THEN 'Iowa Coffee Company'
  ELSE 'Other Brands'
END as Brand,
count(*) as Nr_of_Sales,
sum(sale_dollars) as Sales_Value_Dollars
FROM `bigquery-public-data.iowa_liquor_sales.sales`
WHERE upper(category_name) = 'COFFEE LIQUEURS'
GROUP BY Brand
ORDER BY Nr_of_Sales DESC

```

- Vendorii care au avut un cost total in anii 2018 si 2019 mai mic de 10000\$ si vanzari totale mai mari de 12000\$ si numarul unic de categorii de produse pe care le-au vandut

```

SELECT
vendor_name as Vendor,
count(distinct category) as Unique_category
FROM `bigquery-public-data.iowa_liquor_sales.sales`
WHERE extract(year from date) IN (2018, 2019)
GROUP BY Vendor
HAVING sum(state_bottle_cost)<10000
AND sum(sale_dollars)>12000
ORDER BY sum(sale_dollars) DESC

```

- Vreau o lista unica a magazinelor care au vandut in cel putin 2 comitate diferite si a numarului lor de vanzari din perioada 1 Aprilie 2013 si 8 August 2017

```
SELECT
store_name as Store_Name,
count(*) as Nr_of_Sales
FROM `bigquery-public-data.iowa_liquor_sales.sales`
WHERE date BETWEEN '2013-04-01' AND '2017-08-08'
GROUP BY Store_Name
HAVING count(county_number)>=2
ORDER BY Nr_of_Sales DESC
```

- In fiecare an vreau sa vad numarul de vanzari per categorie de produse insa doar pentru acele categorii cu peste 1000 de vanzari intr-un an

```
SELECT
extract(year from date) as Year,
category_name as Category_Name,
count(*) as Nr_of_Sales
FROM `bigquery-public-data.iowa_liquor_sales.sales`
GROUP BY Year, Category_Name
HAVING count(category_name)>1000
ORDER BY Year ASC, Category_Name ASC
```

- Care au fost top 5 vendori cu cea mai buna profitabilitate intre anii 2013-2015?

```
SELECT
sum(CASE
  WHEN ceil(bottles_sold/pack) BETWEEN 1 AND 3 THEN
    (state_bottle_retail-state_bottle_cost)*bottles_sold - 0.3*ceil(bottles_sold/pack)
  WHEN ceil(bottles_sold/pack) BETWEEN 4 AND 7 THEN
    (state_bottle_retail-state_bottle_cost)*bottles_sold-0.25*ceil(bottles_sold/pack)
  ELSE
    (state_bottle_retail-state_bottle_cost)*bottles_sold-0.2*ceil(bottles_sold/pack)
  END as Profitability,
vendor_name as Vendor,
FROM bigquery-public-data.iowa_liquor_sales.sales
WHERE date BETWEEN '2013-01-01'AND '2015-12-31'
GROUP BY Vendor
ORDER BY Profitability DESC
LIMIT 5
```

Exercitii la curs - 26 noiembrie 2020

- Functii de procesare STRING: Extrage substring-ul ce apare dupa caracterul ; din string-ul “;;;. bim bam bum bem bem ; chem vrem gem cui **_” eliminand caracterele suplimentare

```
SELECT ';;;. bim bam bum bem bem ; chem vrem gem cui **_'
, TRIM(';;;. bim bam bum bem bem ; chem vrem gem cui **_', ';;. * _')
, STRPOS(TRIM(';;;. bim bam bum bem bem ; chem vrem gem cui **_', ';;. * _')
, ';')
, LENGTH(TRIM(';;;. bim bam bum bem bem ; chem vrem gem cui **_', ';;. * _'))
, SUBSTR(TRIM(';;;. bim bam bum bem bem ; chem vrem gem cui **_', ';;. * _')
, 21+1, 18)
, SUBSTR(TRIM(';;;. bim bam bum bem bem ; chem vrem gem cui **_', ';;. * _')
, STRPOS(TRIM(';;;. bim bam bum bem bem ; chem vrem gem cui **_', ';;. * _')
, ';') +1, LENGTH(TRIM(';;;. bim bam bum bem bem ; chem vrem gem cui **_', ';;. * _')
- STRPOS(TRIM(';;;. bim bam bum bem bem ; chem vrem gem cui **_', ';;. * _')
, ';'))
```

- Functii de procesare DATE: Determina DATE-urile care compun intervalul “de acum 5 saptamani pana ieri”

```
SELECT
DATE_SUB(CURRENT_DATE(), INTERVAL 1 DAY),
DATE_SUB(DATE_SUB(CURRENT_DATE(), INTERVAL 1 DAY), INTERVAL 5 WEEK)
```

- Functii de procesare TIMESTAMP: STRING-ul ‘09/12/22T4:57:03 PM’ reprezinta un TIMESTAMP pe timezone-ul Pacific Daylight Time. Converteste-l intr-un TIMESTAMP pe UTC

```
SELECT PARSE_date('%m/%d/%y', '09/12/22')
, parse_timestamp('%m/%d/%yT%i:%M:%S %p', '09/12/22T4:57:03 PM')
, concat ('09/12/22T4:57:03 PM', ' America/Los_Angeles')
, parse_timestamp('%m/%d/%yT%i:%M:%S %p %Z', concat ('09/12/22T4:57:03 PM', ' America/Los_Angeles'))
```

- Functi Matematice: rotunjeste in sus partea pozitiva a impartirii dintre 53*89 si - 71*101

```
SELECT ceil(abs(div(53*89, -71*101)))
```

- Vreau top 5 autori dupa scorul articolelor publicate in saptamana 47 din anii 2015-2018

```
SELECT
```

```
f.by as Author,
avg(score) as Average_score
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type='story'
AND extract(week(monday) from timestamp)=47
AND extract(year from timestamp) IN (2015,2016,2017,2018)
GROUP BY Author
ORDER BY Average_score DESC
LIMIT 5
```

- Toti operatorii al caror nume incepe cu V, atat cu litera mare cat si cu litera mica, care au avut o viteza medie mai mare de 1.2 pentru masuratori pe bicicleta pe retele mai noi de 3G in Aprilie 2016

```
SELECT
operator as Operator
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE STARTS_WITH(INITCAP(operator), 'V')=TRUE
AND activity='ON_BICYCLE'
AND net IN ('4G', '5G')
AND extract(month from date)=4
AND extract (year from date)=2016
GROUP BY operator
HAVING avg(speed)>1.2
```

- Functii de procesare STRING -> Extrage ultimele 4 litere din textul
“&& blah bee bun bim vrum ramraNgum **”

```
SELECT “&& blah bee bun bim vrum ramraNgum **”
,TRIM (“&& blah bee bun bim vrum ramraNgum **”, “*& “)
,SUBSTR(TRIM (“&& blah bee bun bim vrum ramraNgum **”, “*& “), -4, 4)
```

- Functii de procesare DATE -> Determina ultima zi a lunii pentru date-ul ‘2019-03-25’

```
SELECT ‘2019-03-25’
, DATE_TRUNC(‘2019-03-25’, month)
, DATE_ADD(DATE_TRUNC(‘2019-03-25’, month), interval 1 month)
```

, DATE_SUB(DATE_ADD(DATE_TRUNC('2019-03-25', month), interval 1 month), interval 1 day)

- Functii de procesare TIMESTAMP -> Timestamp-ul '2020-03-15 18:48:23.987 UTC' este marcat in mod eronat ca fiind pe Timezone UTC. El este de fapt in timezone-ul New Zealand Time. Converteste-l pentru a fi in mod corect pe UTC

```
SELECT "2020-03-15 18:48:23.987 UTC"  
, REPLACE("2020-03-15 18:48:23.987 UTC", "UTC", "Pacific/Auckland")  
, TIMESTAMP(REPLACE("2020-03-15 18:48:23.987 UTC", "UTC",  
"Pacific/Auckland"))
```

- Functii Matematice: Calculeaza restul impartirii dintre $2+9*5$ si $(4+3)*17$

```
SELECT MOD((2+9*5), (4+3)*17)
```

- Vreau o lista cu titlurile, url-ul si autorul tuturor job-urilor remote sau de Python si care nu sunt de management, postate in luna Martie in oricare din ani

```
SELECT  
  CASE  
    WHEN text LIKE "%remote%" THEN "Remote"  
    WHEN text LIKE "%Python%" THEN "Python"  
  END as Job_Type,  
  title as Title,  
  url as URL,  
  f.by as Author  
FROM bigquery-public-data.hacker_news.full as f  
WHERE type="job"  
AND text LIKE "%remote%" OR text LIKE "%Python%"  
AND text NOT LIKE "%management%"  
AND extract (month from timestamp)=3
```

- Viteza medie si precizia maxima la nivel de provider al pozitie in quarterul 3 din 2015 si quarterul 2 din 2017 in districtul Barcelona pentru retelele al caror nume are mai mult de 5 caractere

```
SELECT
```



```

provider as Provider,
avg(speed) as Average_Speed,
max(preciision) as Max_Preciission
FROM
bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017
WHERE ((extract(quarter from date)=3 AND extract(year from date)=2015) OR
(extract(quarter from date)=2 AND extract(year from date)=2017))
AND postal code LIKE '08%'
AND length(network)>5
GROUP BY Provider

```

Tema - 1 decembrie 2020

- Vreau top 5 autori dupa lungimea medie in numar de caractere a titlurilor articolelor din ultima luna calendaristica completa

```

SELECT
f.by as Author,
avg(length(title)) as Average_Title_Lenght,
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type='story'
AND date(timestamp) BETWEEN DATE_SUB(DATE_TRUNC(CURRENT_DATE(),
interval 1 day) AND DATE_SUB(DATE_TRUNC(CURRENT_DATE(), month)
interval 1 month)
GROUP BY Author
ORDER BY Average_Title_Lenght DESC
LIMIT 5

```

- Top 100 cele mai vechi comentarii cu vechimea in numar de zile

```

SELECT
id as ID,
timestamp as Timestamp,
TIMESTAMP_DIFF(CURRENT_TIMESTAMP(), timestamp, DAY)
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type='comment'
ORDER BY Timestamp ASC
LIMIT 100

```

- Vreau titlurile tuturor articolele din ultimele 3 quartere ce contin in titlu caracterul : si sunt pe protocol securizat. Inlocuiti : cu - si folositi functia strpos pentru a verifica daca acest caracter exista in titlu si functia startswith pentru a determina daca protocolul este securizat.

```
SELECT
title as Title
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type='story'
AND date(timestamp) >= DATE_ADD(DATE_SUB(CURRENT_DATE, INTERVAL
3 QUARTER), INTERVAL 1 month)
AND STARTS_WITH(url, 'https')= TRUE
AND STRPOS(title, ':') != 0
```

- Vreau top 20 autori si numarul de articole dupa vechimea medie si scorul mediu. Articolele sa fie dintr-o perioada de 182 de zile incepand cu data de 14 Iunie 2019

```
SELECT
f.by as Author,
count(*) as Nr_of_Stories,
avg(score) as Average_Score,
avg(DATE_DIFF(date(timestamp), '2019-06-14', DAY)) as
Average_Posting_Period_Day
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type='story'
AND date(timestamp) BETWEEN '2019-06-14' AND DATE_ADD('2019-06-14',
interval 182 DAY)
GROUP BY Author
ORDER BY Average_Score DESC, Average_Posting_Period_Day DESC
LIMIT 20
```

- Vreau titlurile tuturor sondajelor de opinie din toata perioada disponibila eliminand cuvantul "Poll" si caracterul "?" si care incep cu litera D

```
SELECT
title as Title_no_Filter
,TRIM(title, 'Poll?: ') as Title
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type='poll'
```

AND STARTS_WITH(TRIM(title, 'Poll?: '), 'D')=TRUE

- Vreau evolutia la nivel de an si luna a numarul de comentarii mai lungi de 50 de caractere din zilele Luni, Miercuri si Sambata din perioada celui de-al doilea mandat al lui Obama.

SELECT

**extract(year from timestamp) as Year,
extract(month from timestamp) as Month,
date_trunc(date(timestamp), month) as Year_Month,
count(*) as Nr_of_comments
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type='comment'
AND length(text)>50
AND date(timestamp) BETWEEN '2013-01-21' AND '2017-01-20'
AND extract(dayofweek from date(timestamp)) IN (1,3,6)
GROUP BY Year, Month, Year_Month
ORDER BY Year ASC, Month ASC**

- Vreau toate URL-urile mai scurte de 60 de caractere ale articolelor scrise de autori al caror nume incepe cu a si se termina cu f din ultima saptamana incepand de ieri ordonate dupa scor crescator

SELECT

**url as URL,
DATE_SUB(DATE_SUB(CURRENT_DATE, INTERVAL 1 DAY), INTERVAL 1 WEEK) as First_Day,
DATE_SUB(CURRENT_DATE, INTERVAL 1 DAY) as Last_Day
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type='story'
AND length(url)<60
AND f.by LIKE 'a%f'
AND date(timestamp) BETWEEN (DATE_SUB(DATE_SUB(CURRENT_DATE, INTERVAL 1 DAY), INTERVAL 1 WEEK)) AND (DATE_SUB(CURRENT_DATE, INTERVAL 1 DAY))
ORDER BY score ASC**

- Vreau numarul de comentarii la nivel de luna ordonate calendaristic. Numele lunilor sa fie complet si literal (Ex: April, September etc.)

```

SELECT
FORMAT_TIMESTAMP('%b', timestamp) as Month_Name,
count(*) as Nr_of_comments,
extract(month from timestamp) as Month
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type='comment'
GROUP BY Month, Month_Name
ORDER BY Month ASC

```

- O comparatie dupa scorul mediu la nivel de prima litera a numelui autorilor in ultimele 40 de saptamani incepand de acum 3 zile definit pe time zone-ul Bucurestiului insa doar pentru cele cu mai mult de 50 articole

```

SELECT
SUBSTR(f.by, 1,1) as Author_First_Letter,
avg(score) as Average_Score,
DATE_SUB(DATE_SUB(CURRENT_DATE('Europe/Bucharest'), INTERVAL 3 DAY), INTERVAL 40 WEEK) as First_Post,
DATE_SUB(CURRENT_DATE('Europe/Bucharest'), INTERVAL 3 DAY) as Last_Post
FROM `bigquery-public-data.hacker_news.full` as f
WHERE date(timestamp) BETWEEN
DATE_SUB(DATE_SUB(CURRENT_DATE('Europe/Bucharest'), INTERVAL 3 DAY), INTERVAL 40 WEEK) AND
DATE_SUB(CURRENT_DATE('Europe/Bucharest'), INTERVAL 3 DAY)
AND type='story'
AND f.by is NOT NULL
GROUP BY Author_First_Letter
HAVING count(*)>50
ORDER BY Author_First_Letter ASC

```

Exercitii la curs - 3 decembrie 2020

- Ce proportie de articole au mai mult de 25 de descendentii sau un scor mai mic de 5?

```
SELECT
count((descendants>25 OR score<5) OR NULL)/count(*)
,count(descendants>25 OR score<5 OR NULL)/count(*)
,sum(CASE
  WHEN (descendants>25 OR score<5)=true THEN 1
  WHEN (descendants>25 OR score<5)=false THEN 0
  END)/count(*)
, avg(CASE
  WHEN (descendants>25 OR score<5)=true THEN 1
  WHEN (descendants>25 OR score<5)=false THEN 0
  END)
FROM `bigquery-public-data.hacker_news.full`
WHERE type='story'
```

- Ce proportie din sondajele de opinie cu optiuni sau simple au fost publicate de autori al caror username incepe cu litera d in timpul orelor de lucru anume intre 9 dimineata si 8 seara?

```
SELECT
sum(CASE
  WHEN STARTS_WITH(INITCAP(f.by), 'D') AND time(timestamp) BETWEEN
'09:00:00' AND '20:00:00' THEN 1
  ELSE 0
  END)/count(*)
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type IN ('poll', 'pollopt')
```

Tema - 8 decembrie 2020

- Ce proportie din vanzarile de lichior au fost de un volum mai mare de 100 de litri si care a fost valoarea acestora?

```

SELECT
(count(volume_sold_liters > 100 OR NULL)/count(*))*100 as
Liquor_Sale_over_100_liters_Percentage,
round(sum(if(volume_sold_liters>100, sale_dollars,NULL))/1000, 4) as
Total_Value_Million_Dollars
FROM `bigquery-public-data.iowa_liquor_sales.sales`

```

- Care este raportul intre volumul cumparat de franciza Hy Vee si Sam's Club in anul 2019?

```

SELECT
sum(IF (store_name LIKE "%Hy-Vee%",bottle_volume_ml, NULL)
/
IF (store_name LIKE "%Sam's Club%",bottle_volume_ml, NULL))
as HyVee_SamsClub_Volume_Ratio
FROM `bigquery-public-data.iowa_liquor_sales.sales`
WHERE extract(year from date)=2019

```

- Care este raportul intre numarul unic de magazine care au cumparat bauturi in districtul Des Moines vs Dubuque?

```

SELECT
count (distinct
CASE
WHEN upper(county) LIKE "%DUBUQUE%" THEN store_name
END)
/
count (distinct
CASE
WHEN upper(county) LIKE "%DES MOINES%" THEN store_name
END)
FROM `bigquery-public-data.iowa_liquor_sales.sales`

```

- Care este raportul intre valoarea vanzarilor de produse ale Kahlua vs Templeton in sa doar pentru vanzari de maxim 1 bax de sticle?

```

SELECT
( sum(CASE
WHEN upper(item_description) LIKE '%KAHLUA%' THEN sale_dollars
ELSE 0

```

```

END)
/
sum(CASE
WHEN upper(item_description )LIKE '%TEMPLETON%' THEN sale_dollars
ELSE 0
END)
) as Kahlua_Templeton_Ratio
FROM `bigquery-public-data.iowa_liquor_sales.sales`
WHERE ceil(bottles_sold/pack)<=1

```

Evaluare - 8 decembrie 2020

- Care este raportul intre valoarea vanzarilor de produse ale Iowa Coffee Company vs Patron Xo insa doar pentru vanzari de minim 10 baxuri de sticle pe o perioada de 25 de zile dupa 3 Ianuarie 2019?

```

SELECT
SUM(CASE
WHEN item_description LIKE '%Iowa Coffee Company%' THEN sale_dollars
ELSE 0
END)
/
SUM(CASE
WHEN item_description LIKE '%Patron Xo%' THEN sale_dollars
ELSE 0
END)
FROM `bigquery-public-data.iowa_liquor_sales.sales`
WHERE ceil(bottles_sold/pack)>=10
AND date BETWEEN '2019-01-03' AND DATE_ADD('2019-01-03', INTERVAL 25
DAY)

```

- Titlul articolelor scrise de autori al caror nume incepe cu B dar nu cu b, din perioada 3 Iunie 2015 - 17 Ianuarie 2019, insa doar din intervalul orar 10-16, in ordinea descrescatoare a scorului

```

SELECT
title as Title,

```

```

score as Score
FROM `bigquery-public-data.hacker_news.full` as f
WHERE f.by LIKE 'B%'
AND f.by NOT LIKE 'b%'
AND date(timestamp) BETWEEN '2015-06-03' AND '2019-01-17'
AND time(timestamp) BETWEEN '10:00:00' AND '16:00:00'
Order by score DESC

```

- Top 25 judete cu vanzari totale de peste 15.000\$ in anul 2017
dupa volumul de bauturi vandut

```

SELECT
upper(county) as County_Name
FROM `bigquery-public-data.iowa_liquor_sales.sales`
WHERE extract(year from date)=2017
GROUP BY County_Name
HAVING sum(sale_dollars)>15000
ORDER BY sum(volume_sold_liters) DESC
LIMIT 25

```

- Pentru fiecare oras din districtul Girona vreau precizia medie si
numarul total de sateliti dupa tipul de activitate al utilizatorului

```

SELECT
town_name as City,
AVG(precision) as AVg_Precision,
sum(satellites) as Nr_of_Satellites,
activity as User_Activity
FROM
`bigquery-public-data.catalonian_mobile_coverage.mobile_data_2015_2017`
WHERE postal_code LIKE '17%'
GROUP BY City, User_Activity
ORDER BY City ASC, User_Activity ASC

```

- Pentru Gin si Vodka vreau numarul de vanzari, valoarea totala de
retail a vanzarii, valoarea totala de en-gros a vanzarii si volumul
mediu al sticlei.

```

SELECT

```



```

CASE
  WHEN category_name LIKE '%Gin%' THEN 'Gin'
  WHEN category_name LIKE '%Vodka%' THEN 'Vodka'
  ELSE 'Another'
END as Category_Type,
count(*) as Nr_of_Sales,
round(sum(state_bottle_cost*bottles_sold)/100000,4) as
En_Gros_Value_Million_Dollars,
round(sum(sale_dollars)/100000,4) as Retail_Value_Million_Dollars,
round(avg(bottle_volume_ml),2) as Average_Bottle_Volume_ml
FROM `bigquery-public-data.iowa_liquor_sales.sales`
WHERE category_name LIKE '%Gin%' OR category_name LIKE '%Vodka%'
GROUP BY Category_Type

```

Exercitii la curs - 10 decembrie 2020

- Care a fost impozitul pe volum in fiecare an? Impozitul se calculeaza in functie de volumul vandut per comanda exprimat in litri si categoria de produse astfel: Vodka impozitul este de 0.1\$ per litru, Whisky impozitul este de 0.15\$ per litru, Rom, impozitul este de 0.13\$ per litru, Tequila impozitul este de 0.175\$ dolari per litru, Lichior impozitul este de 0.1375\$ per litru, Scotch impozitul este de 0.2\$ per litru, Schnapps impozitul este de 0.115\$ per litru, Gin impozitul este de 0.16\$ per litru, Brandy impozitul este de 0.189\$ per litru, iar pentru alte tipuri de bauturi impozitul este de 0.35\$ per litru.

```

SELECT
extract(year from date) as Year,
round(SUM(ceil(volume_sold_liters *
(CASE
  WHEN TRIM(initcap(category_name)) LIKE "%Vodka%" THEN 0.1
  WHEN TRIM(initcap(category_name)) LIKE "%Whisky%" THEN 0.15
  WHEN TRIM(initcap(category_name)) LIKE "%Rum%" THEN 0.13
  WHEN TRIM(initcap(category_name)) LIKE "%Tequila%" THEN 0.175
  WHEN TRIM(initcap(category_name)) LIKE "%Liqueur%" THEN 0.1375
  WHEN TRIM(initcap(category_name)) LIKE "%Scotch%" THEN 0.2
  WHEN TRIM(initcap(category_name)) LIKE "%Schnapps%" THEN 0.115
  WHEN TRIM(initcap(category_name)) LIKE "%Gin%" THEN 0.16
  WHEN TRIM(initcap(category_name)) LIKE "%Brandy%" THEN 0.189
  ELSE 0.35

```

```

END)))) as Tax
FROM `bigquery-public-data.iowa_liquor_sales.sales`
WHERE date IS NOT NULL
GROUP BY Year
ORDER BY Year ASC

```

- Care a fost impozitul pe vanzarea de produse alcoolice per an pentru fiecare vendor in perioada 2016-2019? Impozitul pe vanzarea de produse alcoolice se calculeaza dupa valoarea marfurilor vandute astfel: pentru un cost total anual sub 3000\$ impozitul este 0, pentru un cost total anual mai mare de 3.000\$ dar mai mic de 10.000\$ impozitul este de 500\$, pentru un cost total anual mai mare de 10.000\$ dar mai mic de 50.000\$ impozitul este de 5\$ per sticla vanduta, pentru un cost total anual intre 50.000\$ si 500.000\$ impozitul este de 3.8\$ per sticla vanduta, pentru un cost total anual mai mare de 500.000\$ dar mai mic de 2.000.000\$ impozitul este de 2.5\$ per sticla vanduta, pentru un cost total anual intre 2.000.000\$ si 10.000.000\$ impozitul este de 1.7\$ per sticla vanduta, iar pentru un cost total anual peste 10.000.000\$ impozitul este de 1\$ per sticla vanduta. Costul total se calculeaza ca produsul dintre costul per sticla si numarul de sticle. Se aplica si un discount de 2% la acest cost pentru expunere la risc.

```

SELECT
extract(year from date) as Year,
vendor_name as Vendor_Name,
round(0.98* CASE
  WHEN sum(state_bottle_cost*bottles_sold)<3000 THEN 0
  WHEN sum(state_bottle_cost*bottles_sold)<10000 THEN 500
  WHEN sum(state_bottle_cost*bottles_sold)<50000 THEN
5*SUM(bottles_sold)
  WHEN sum(state_bottle_cost*bottles_sold)<500000 THEN
3.8*SUM(bottles_sold)
  WHEN sum(state_bottle_cost*bottles_sold)<2000000 THEN
2.5*SUM(bottles_sold)
  WHEN sum(state_bottle_cost*bottles_sold)<10000000 THEN
1.7*SUM(bottles_sold)
  WHEN sum(state_bottle_cost*bottles_sold)>10000000 THEN
1*SUM(bottles_sold)

```

```

END,2) as Tax
FROM `bigquery-public-data.iowa_liquor_sales.sales`
WHERE date BETWEEN '2016-01-01' AND '2019-12-31'
GROUP BY Vendor_Name, Year
ORDER BY Year ASC

```

- Titlul, autorul si timpul publicarii pentru top 25 sondaje dupa scor legate de retelele de socializare Facebook si Twitter

```

SELECT
f.by as Author,
title as Title,
time(timestamp) as Time
FROM `bigquery-public-data.hacker_news.full` as f
WHERE (text LIKE '%Facebook%' OR text LIKE '%Twitter%' OR title LIKE
'%Facebook%' OR title LIKE '%Twitter%')
AND type='poll'
ORDER BY score DESC
LIMIT 25

```

- Numarul unic de utilizatori care au postat job-uri de Python in anii 2015 si 2016 insa doar cele postate in intervalul orar 10 dimineata - 10 seara

```

SELECT
count (distinct f.by)
FROM `bigquery-public-data.hacker_news.full` as f
WHERE type='job'
AND title LIKE '%Python%'
AND extract(year from timestamp) IN (2015,2016)
AND time(timestamp) BETWEEN '10:00:00' AND '22:00:00'

```