ELECTIVE III : Data Analytics - Use of Select Tools and Techniques
CSE 457

CA4 Project

Done by

Gokul P - E0119056
Saai Swetha M K - E0119016

Datasets

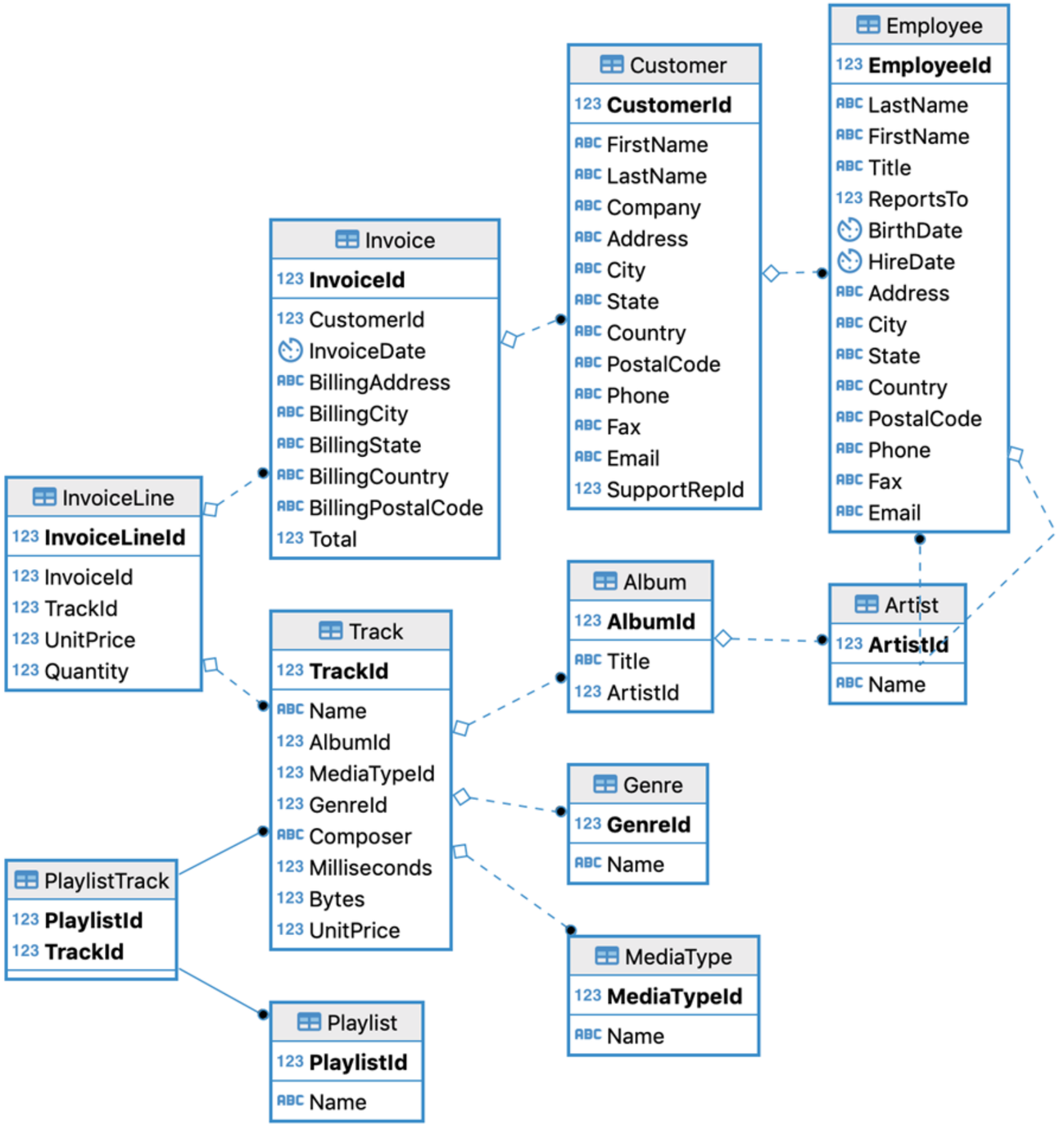
1. Chinook Database
2. Cars India dataset

Chinook Db

data model represents a digital music media store, including tables for artists, albums, media tracks, invoices and customers

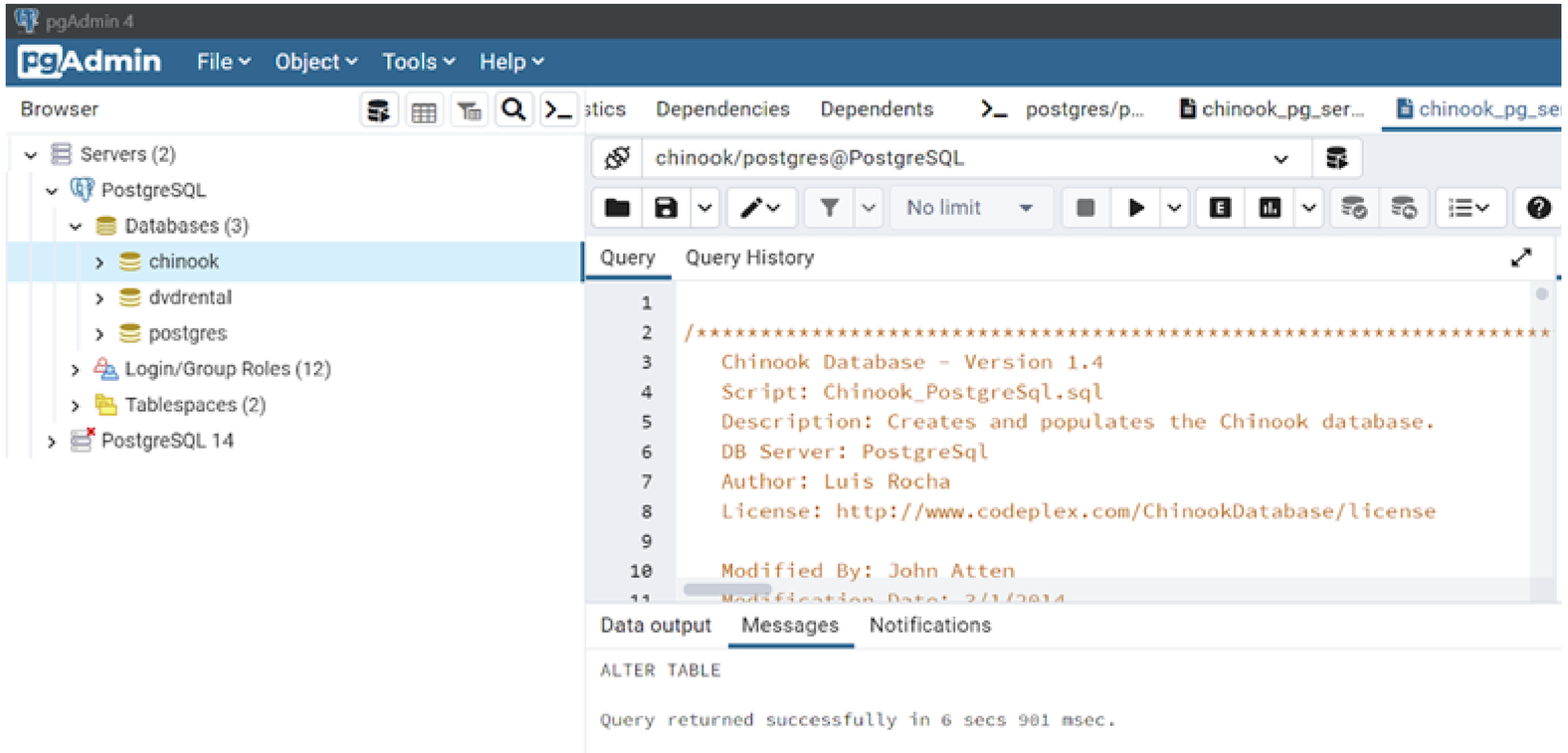
Car India Dataset

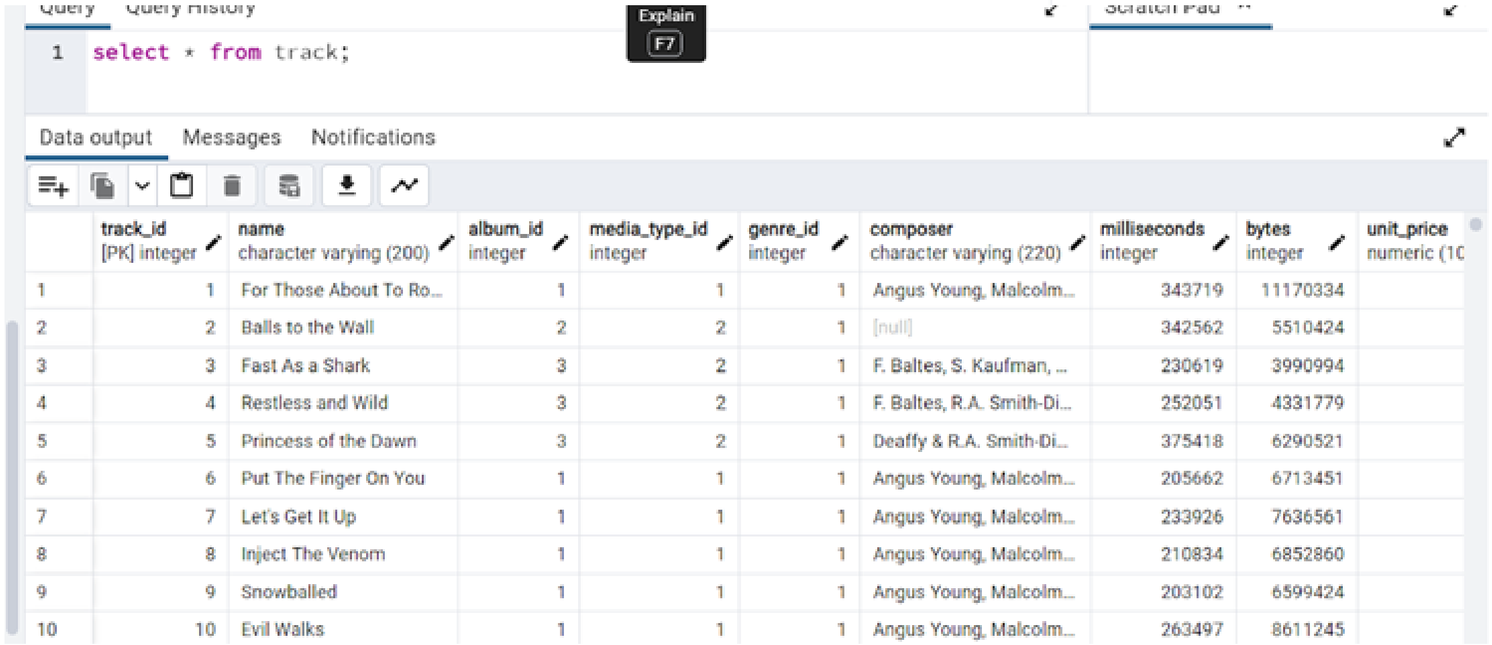
information on each car’s model type, makers, no.of cylinders, height, width etc.


Employee
EmployeeId
LastName
FirstName
Title
ReportsTo
BirthDate
HireDate
Address
City
State
Country
PostalCode
Phone
Fax
Email
Customer
CustomerId
FirstName
LastName
Company
Address
City
State
Country
PostalCode
Phone
Fax
Email
SupportRepId
Invoice
InvoiceId
CustomerId
InvoiceDate
BillingAddress
BillingCity
BillingState
BillingCountry
BillingPostalCode
Total
InvoiceLine
InvoiceLineId
InvoiceId
TrackId
UnitPrice
Quantity
Track
TrackId
Name
AlbumId
MediaTypeId
GenreId
Composer
Milliseconds
Bytes
UnitPrice
Album
AlbumId
Title
ArtistId
Artist
ArtistId
Name
Genre
GenreId
Name
MediaType
MediaTypeId
Name
PlaylistTrack
PlaylistId
TrackId
Playlist
PlaylistId
Name

Postgres

https://raw.githubusercontent.com/xivSolutions/ChinookDb_Pg_Modified/master/chinook_pg_serial_pk_proper_naming.sql





	track_id [PK] integer	name character varying (200)	album_id integer	media_type_id integer	genre_id integer	composer character varying (220)	milliseconds integer	bytes integer	unit_price numeric (10
1	1	For Those About To Ro...	1	1	1	Angus Young, Malcolm...	343719	11170334	
2	2	Balls to the Wall	2	2	1	[null]	342562	5510424	
3	3	Fast As a Shark	3	2	1	F. Baltes, S. Kaufman, ...	230619	3990994	
4	4	Restless and Wild	3	2	1	F. Baltes, R.A. Smith-Di...	252051	4331779	
5	5	Princess of the Dawn	3	2	1	Deaffy & R.A. Smith-Di...	375418	6290521	
6	6	Put The Finger On You	1	1	1	Angus Young, Malcolm...	205662	6713451	
7	7	Let's Get It Up	1	1	1	Angus Young, Malcolm...	233926	7636561	
8	8	Inject The Venom	1	1	1	Angus Young, Malcolm...	210834	6852860	
9	9	Snowballed	1	1	1	Angus Young, Malcolm...	203102	6599424	
10	10	Evil Walks	1	1	1	Angus Young, Malcolm...	263497	8611245	

Postgres

No.of songs whose duration is less than or equal to 2 mins or 120000ms

Query Query History

```
5
6  select count(*) from track where milliseconds<=120000;
```

Data output Messages Notifications

	count bigint
1	94

Postgres

Average price of the music tracks composed by Steven Tyler

Query | Query History

```
7  /*average unit price of the tracks composed by steven tyler*/
8  select avg(unit_price) from track where composer like '%Steven Tyler%';
```

Data output | Messages | Notifications

	avg numeric
1	0.990000000

Postgres

Country wise total sales in the year 2011

```
14  select billing_country,sum(total) as TotalSales from invoice where EXTRACT(year FROM invoice_date)=2011
15  group by billing_country;
16
```

Data output | Messages | Notifications

	billing_country character varying (40)	totalsales numeric
1	Argentina	0.99
2	Australia	1.98
3	Belgium	24.75
4	Brazil	19.80
5	Canada	55.44
6	Chile	5.94
7	Czech Republic	12.87
8	Finland	15.88
9	France	42.61
10	Germany	48.57
11	India	24.75
12	Ireland	32.75

Postgres

Country with highest sales in the year 2009

```
19  select billing_country,sum(total) as TotalSales from invoice
20  where EXTRACT(year FROM invoice_date)=2009
21  group by billing_country
22  order by TotalSales desc limit 1;
23
```

Data output | Messages | Notifications

	billing_country character varying (40)	totalsales numeric
1	USA	103.95

Postgres

Country with most invoices

```
27  SELECT billing_country, COUNT(invoice_id) FROM invoice
28  GROUP BY 1
29  ORDER BY 2 DESC;
30
```

Data output | Messages | Notifications

	billing_country character varying (40)	count bigint
1	USA	91
2	Canada	56
3	France	35
4	Brazil	35
5	Germany	28
6	United Kingdom	21
7	Portugal	14
8	Czech Republic	14
9	India	13
10	Norway	7
11	Sweden	7

Postgres

Country with most invoices

```
27  SELECT billing_country, COUNT(invoice_id) FROM invoice
28  GROUP BY 1
29  ORDER BY 2 DESC;
30
```

Data output | Messages | Notifications

	billing_country character varying (40)	count bigint
1	USA	91
2	Canada	56
3	France	35
4	Brazil	35
5	Germany	28
6	United Kingdom	21
7	Portugal	14
8	Czech Republic	14
9	India	13
10	Norway	7
11	Sweden	7

Postgres

Highest no.of music tracks purchased by a single customer

```
21  SELECT count(*)
22  FROM CUSTOMER C
23  JOIN INVOICE I ON C.customer_id = I.customer_id
24  GROUP BY I.customer_id
25  ORDER BY 1 DESC Limit 1;
26
```

Data output | Messages | Notifications

	count bigint
1	7

Postgres

No.of songs per Genre

Query | Query History

```
28  SELECT G.name, COUNT(T.track_id)
29  FROM TRACK T JOIN GENRE G ON T.genre_id = G.genre_id
30  GROUP BY 1
31  ORDER BY 2 DESC;
32
```

Data output | Messages | Notifications

	name character varying (120)	count bigint
1	Rock	1297
2	Latin	579
3	Metal	374
4	Alternative & Punk	332
5	Jazz	130
6	TV Shows	93
7	Blues	81
8	Classical	74
9	Drama	64
10	R&B/Soul	61
11	Reggae	58

Postgres

List of all music tracks that has a song length greater than avg song length

```
34  SELECT name, milliseconds FROM track
35  WHERE milliseconds > (SELECT AVG(milliseconds) AS avg_track_length FROM track)
36  ORDER BY milliseconds DESC;
37
```

Data output | Messages | Notifications

	name character varying (200)	milliseconds integer
1	Occupation / Precipice	5286953
2	Through a Looking Gla...	5088838
3	Greetings from Earth, ...	2960293
4	The Man With Nine Liv...	2956998
5	Battlestar Galactica, Pt...	2956081
6	Battlestar Galactica, Pt...	2952702
7	Murder On the Rising S...	2935894
8	Battlestar Galactica, Pt...	2927802
9	Take the Celestra	2927677
10	Fire In Space	2926593
11	The Long Patrol	2925008

Postgres

List of all music tracks that has a song length greater than avg song length

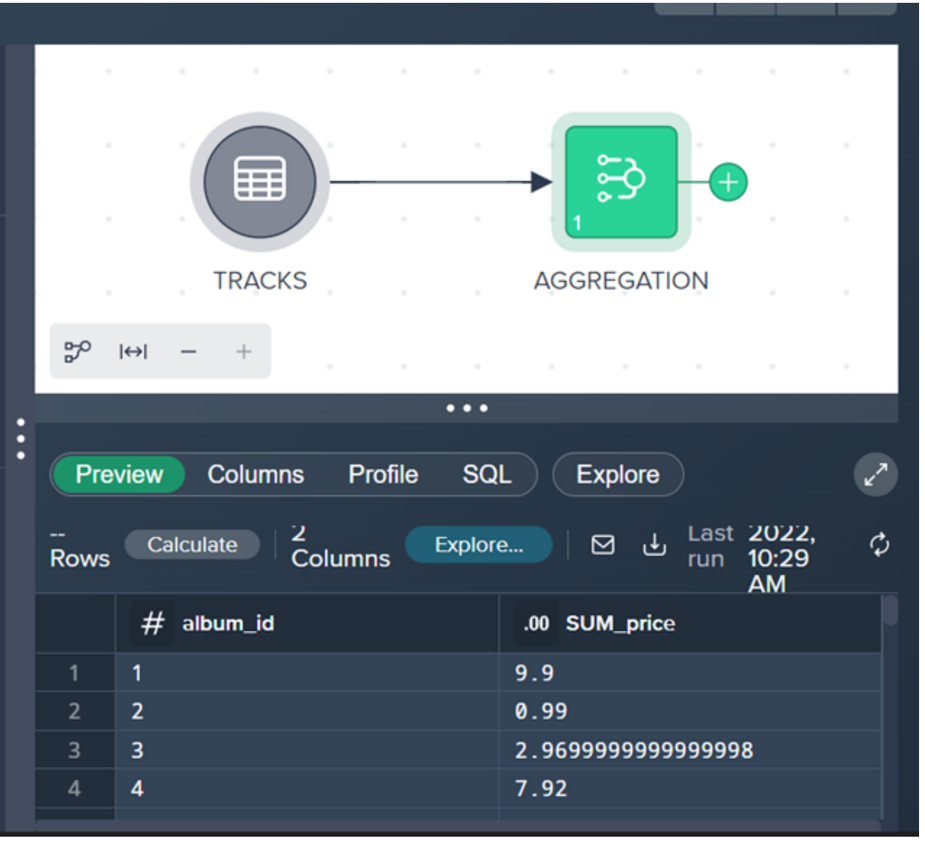
```
34  SELECT name, milliseconds FROM track
35  WHERE milliseconds > (SELECT AVG(milliseconds) AS avg_track_length FROM track)
36  ORDER BY milliseconds DESC;
37
```

Data output | Messages | Notifications

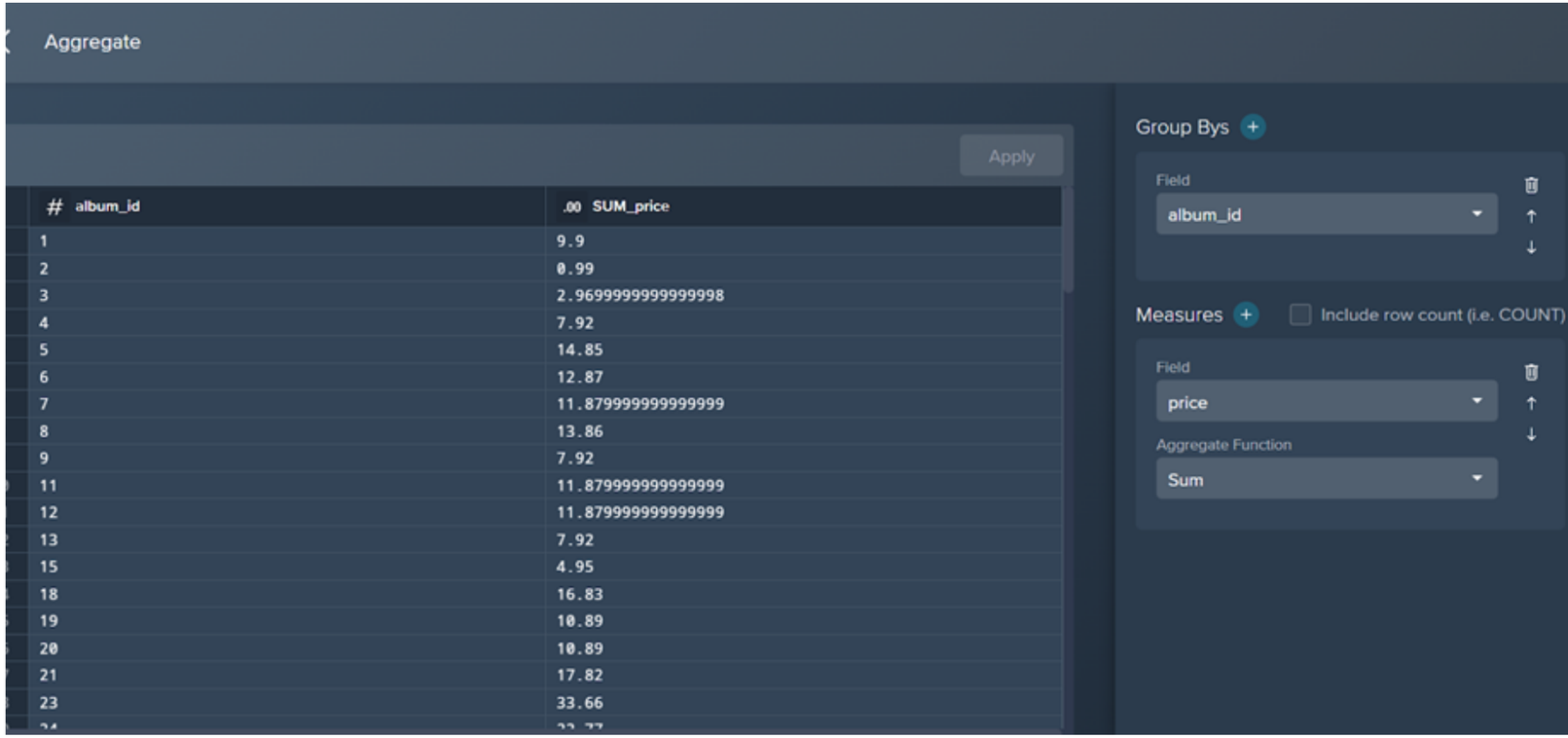
	name character varying (200)	milliseconds integer
1	Occupation / Precipice	5286953
2	Through a Looking Gla…	5088838
3	Greetings from Earth, …	2960293
4	The Man With Nine Liv…	2956998
5	Battlestar Galactica, Pt…	2956081
6	Battlestar Galactica, Pt…	2952702
7	Murder On the Rising S…	2935894
8	Battlestar Galactica, Pt…	2927802
9	Take the Celestra	2927677
10	Fire In Space	2926593
11	The Long Patrol	2925008

Datameer

Album wise total price of music tracks



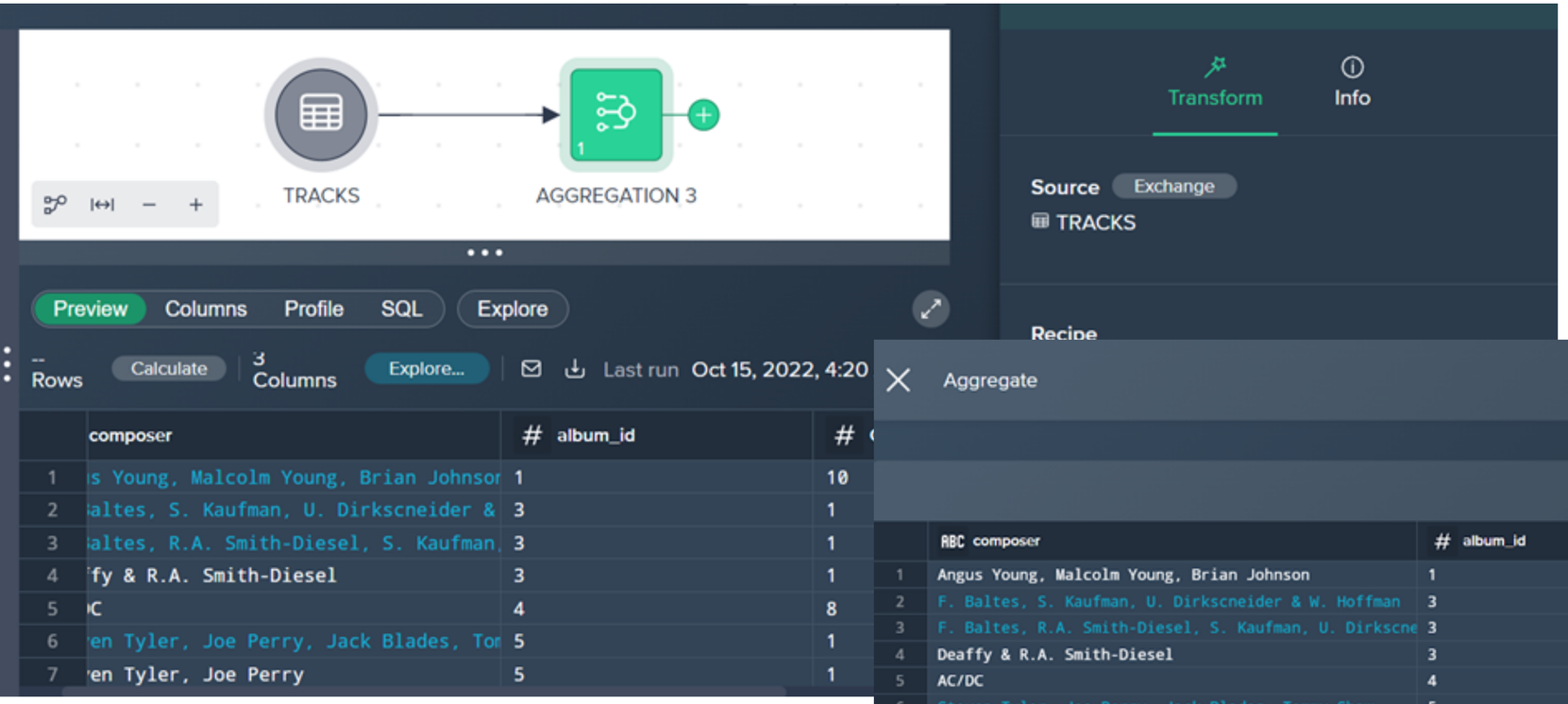
	# album_id	.00 SUM_price
1	1	9.9
2	2	0.99
3	3	2.9699999999999998
4	4	7.92

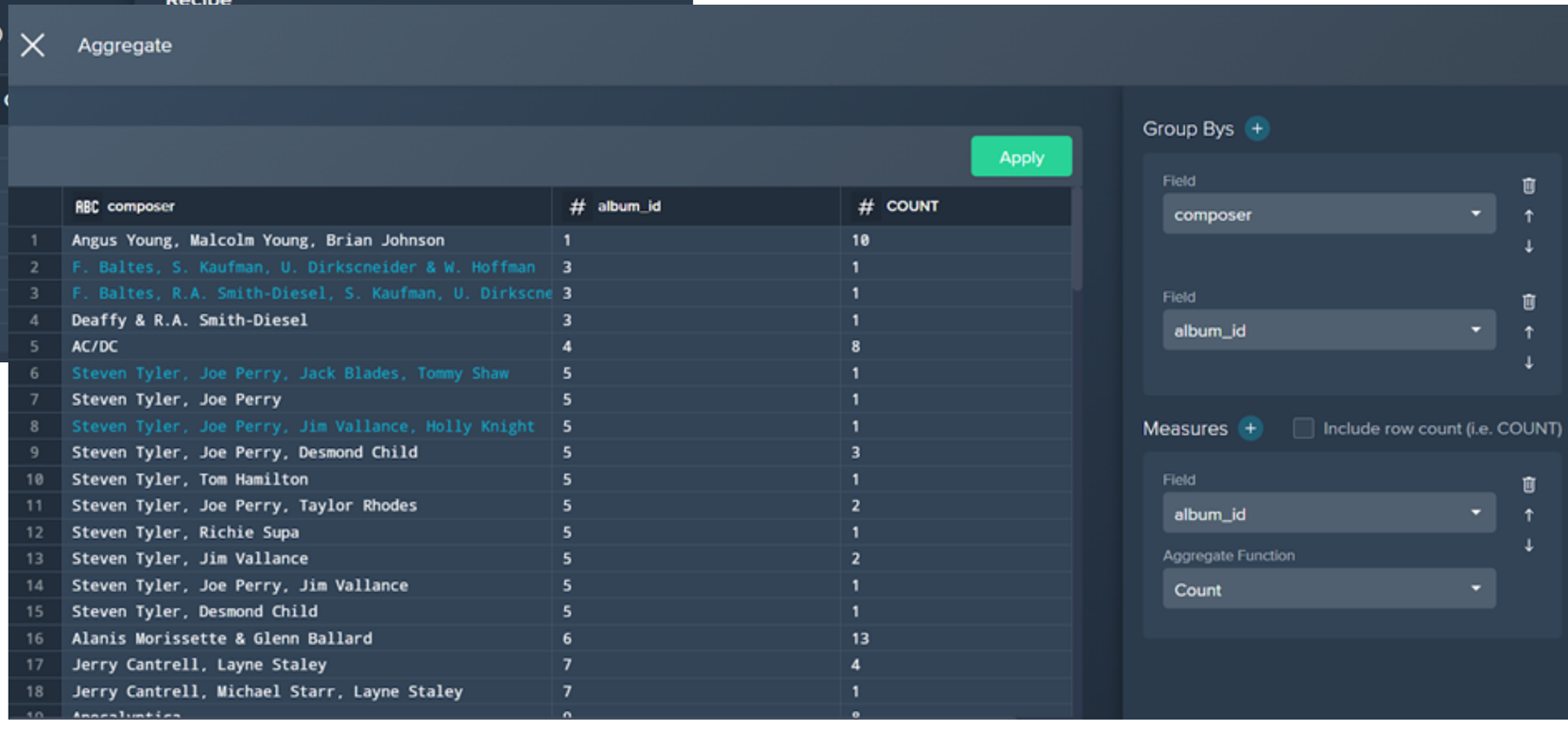


# album_id	.00 SUM_price
1	9.9
2	0.99
3	2.9699999999999998
4	7.92
5	14.85
6	12.87
7	11.879999999999999
8	13.86
9	7.92
11	11.879999999999999
12	11.879999999999999
13	7.92
15	4.95
18	16.83
19	10.89
20	10.89
21	17.82
23	33.66

Datameer

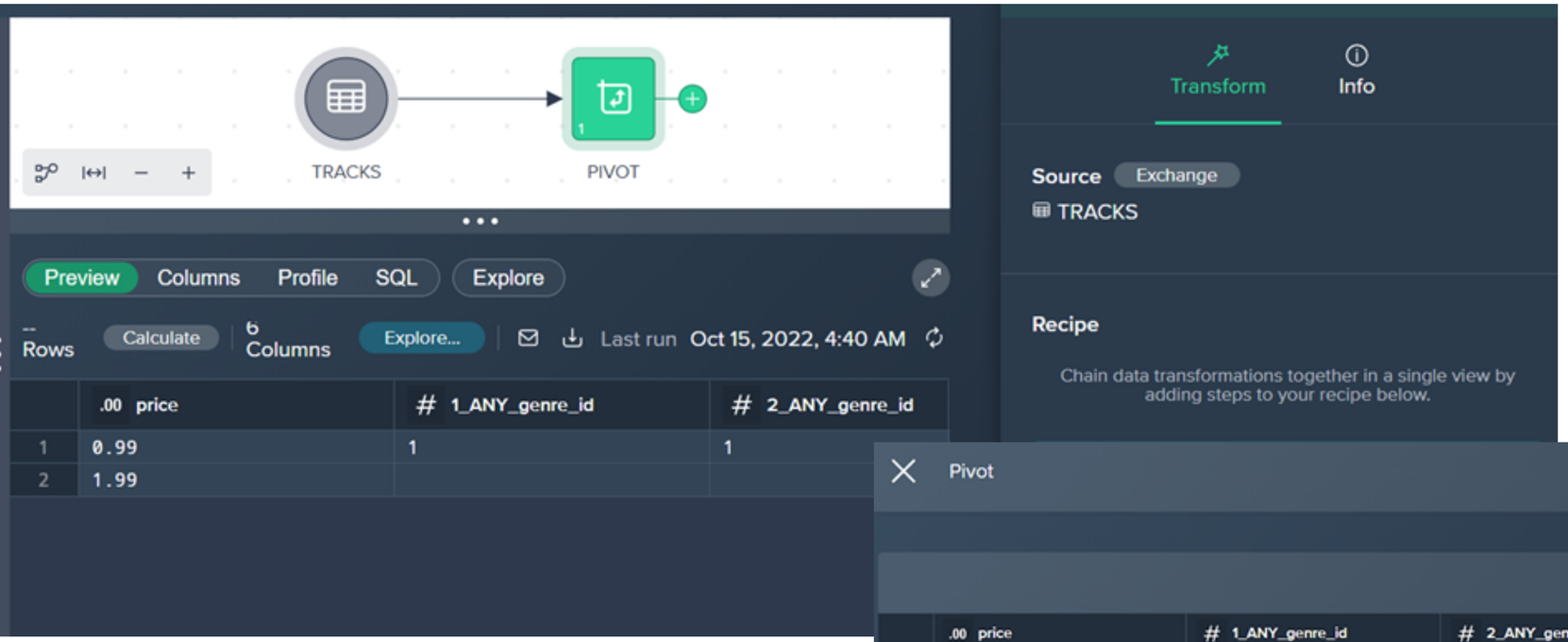
No.of songs composed by each composer album wise

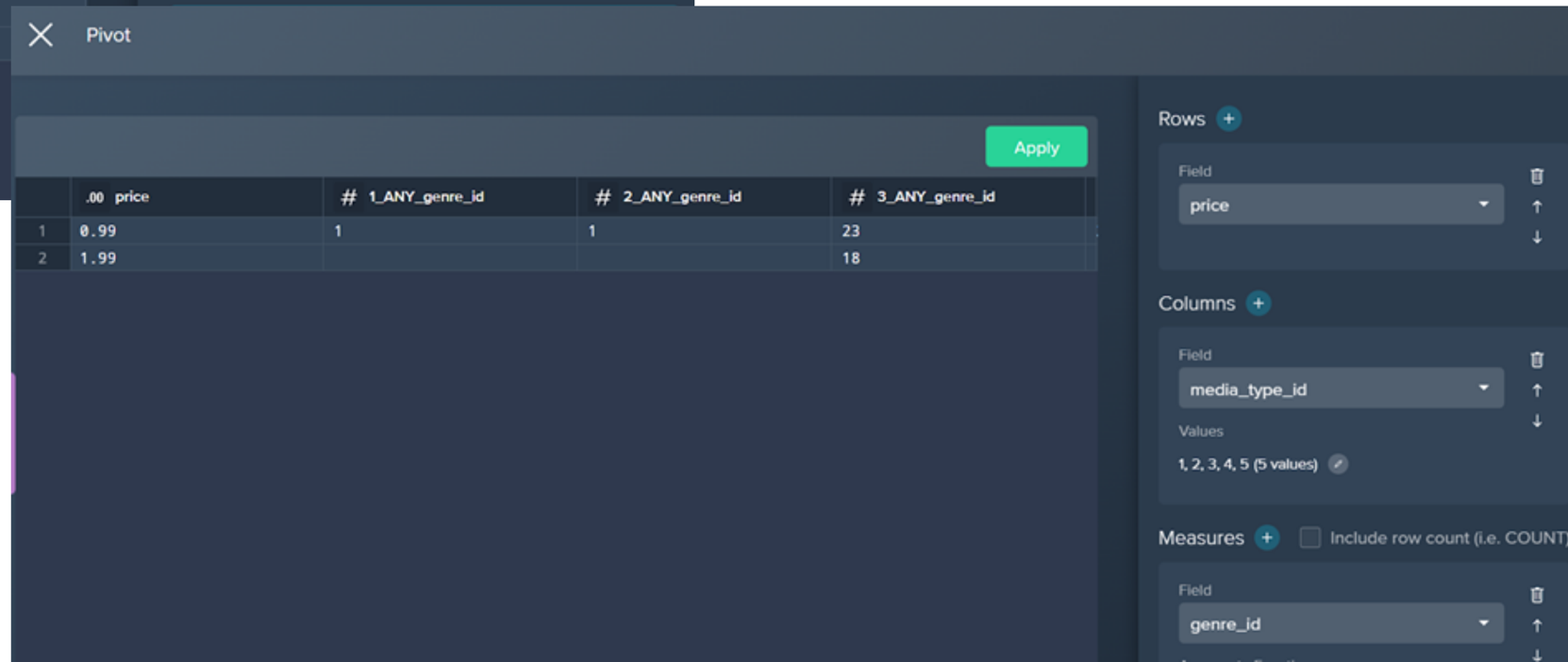




Datameer

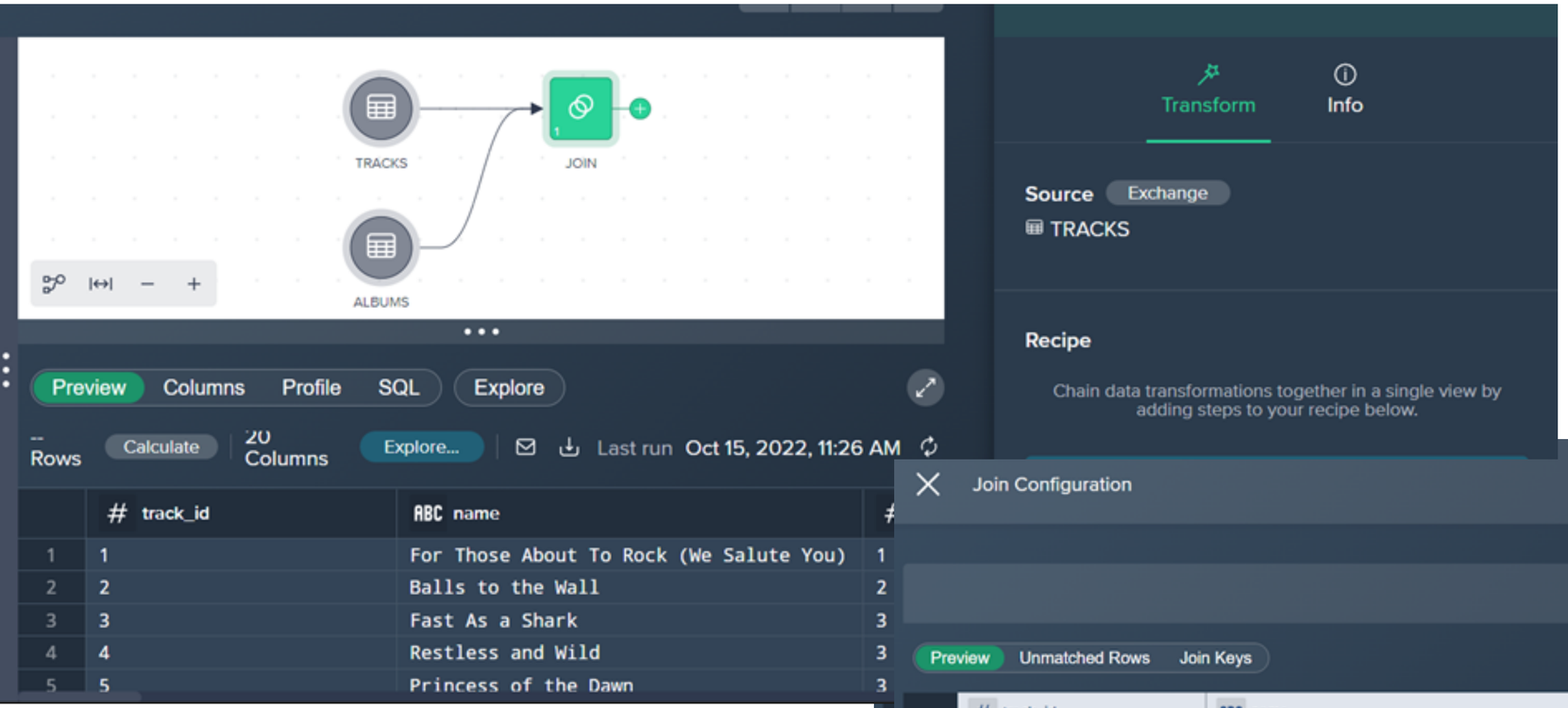
No.of songs genre and price wise - using pivot table

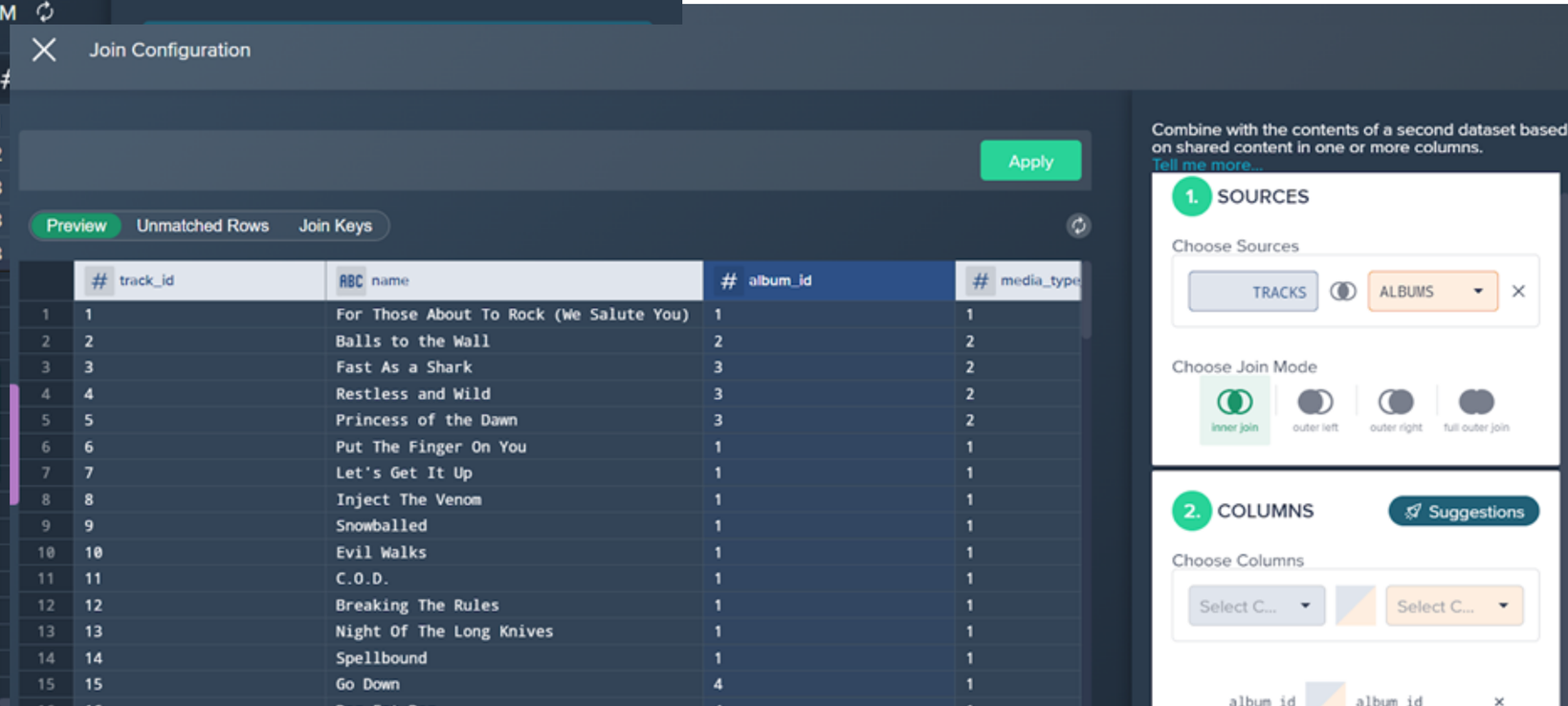




Datameer

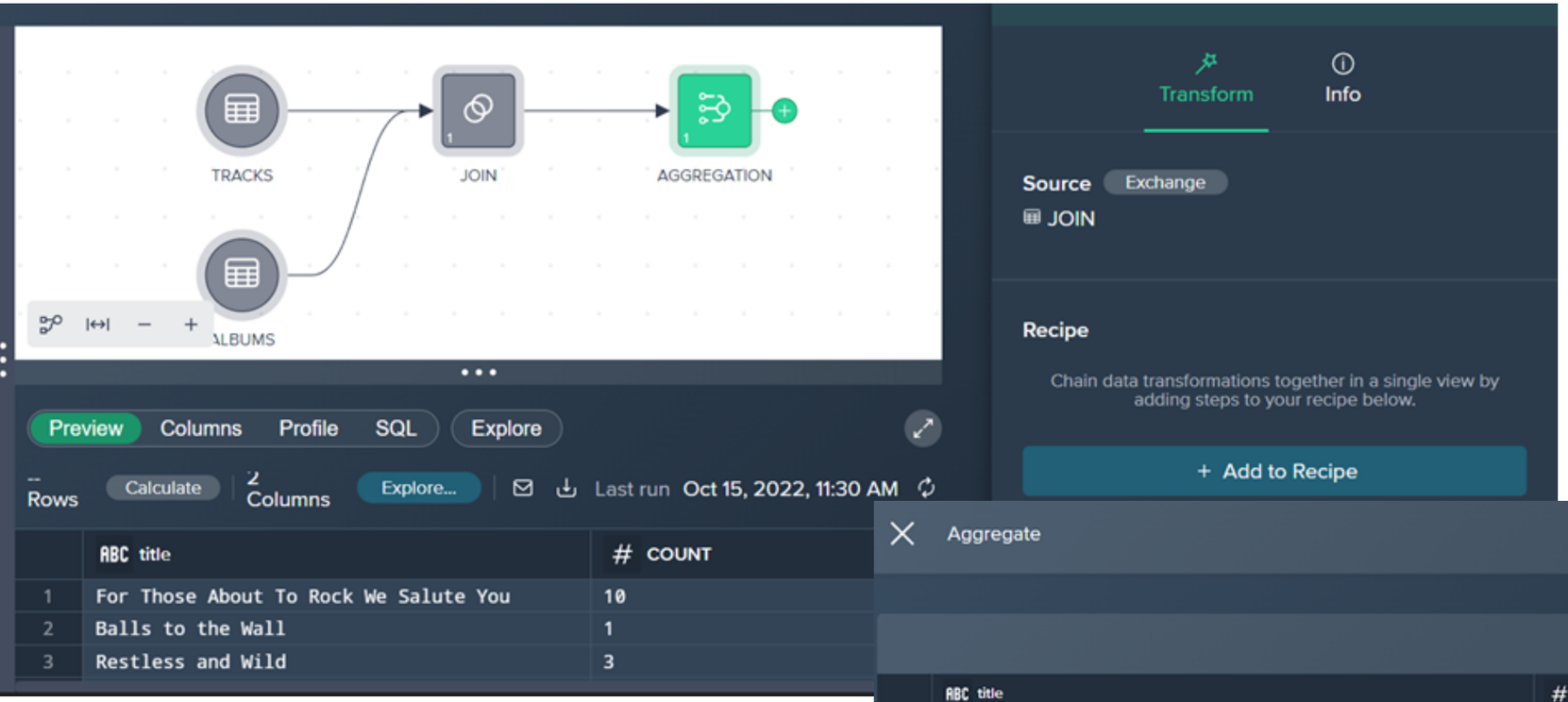
Join tables track and album



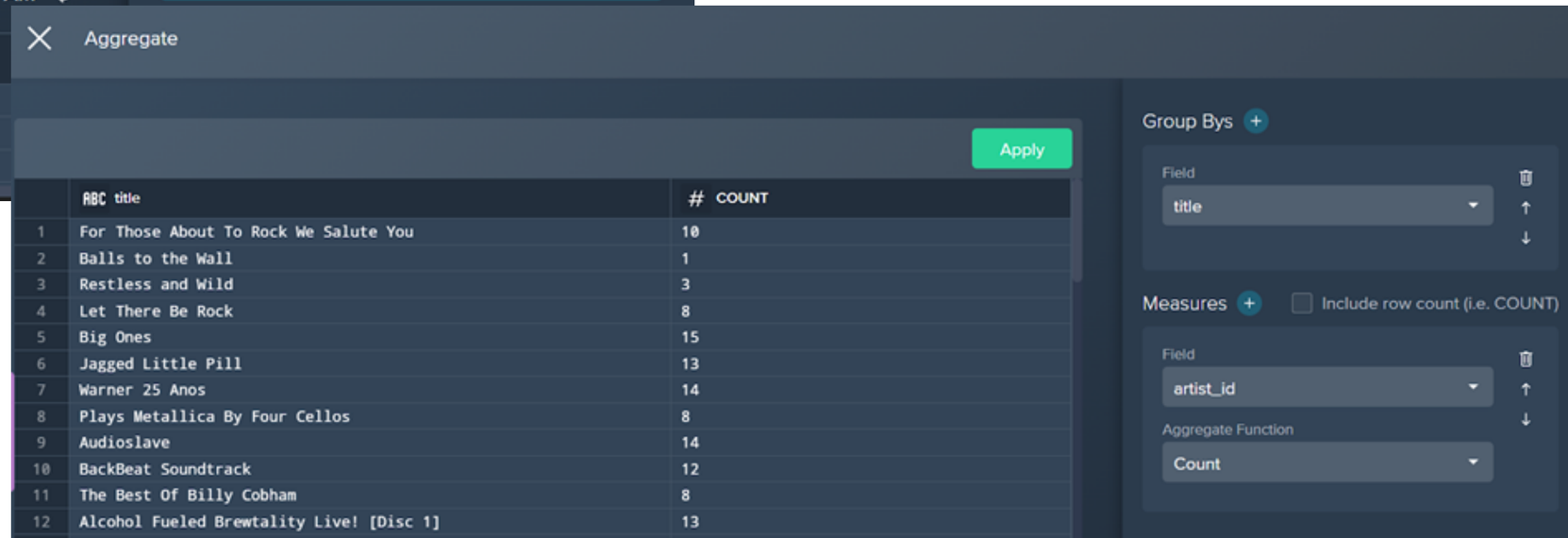


Datameer

No.of artists who sang in each album - using both track and album tables



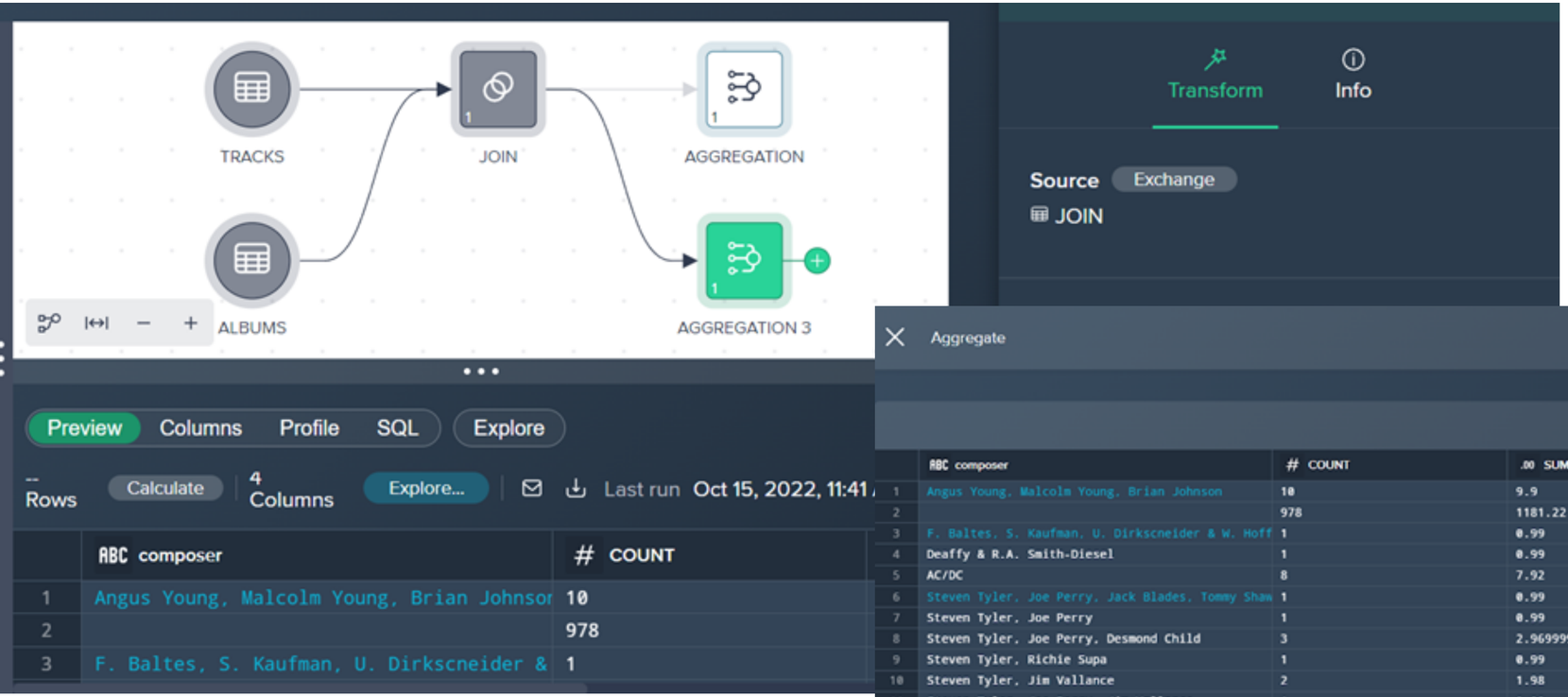
	ABC title	# COUNT
1	For Those About To Rock We Salute You	10
2	Balls to the Wall	1
3	Restless and Wild	3

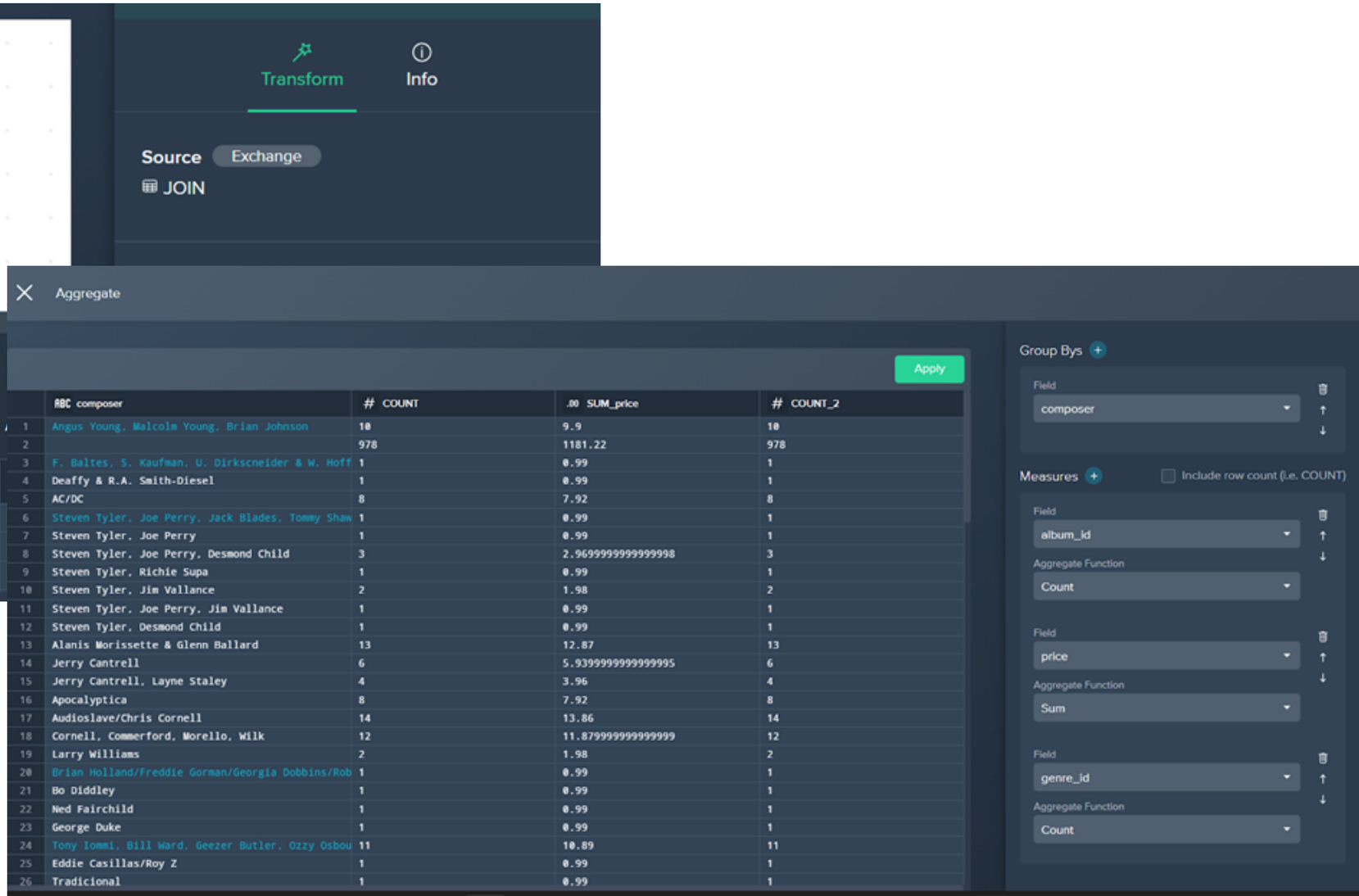


	ABC title	# COUNT
1	For Those About To Rock We Salute You	10
2	Balls to the Wall	1
3	Restless and Wild	3
4	Let There Be Rock	8
5	Big Ones	15
6	Jagged Little Pill	13
7	Warner 25 Anos	14
8	Plays Metallica By Four Cellos	8
9	Audioslave	14
10	BackBeat Soundtrack	12
11	The Best Of Billy Cobham	8
12	Alcohol Fueled Brewtality Live! [Disc 1]	13

Datameer

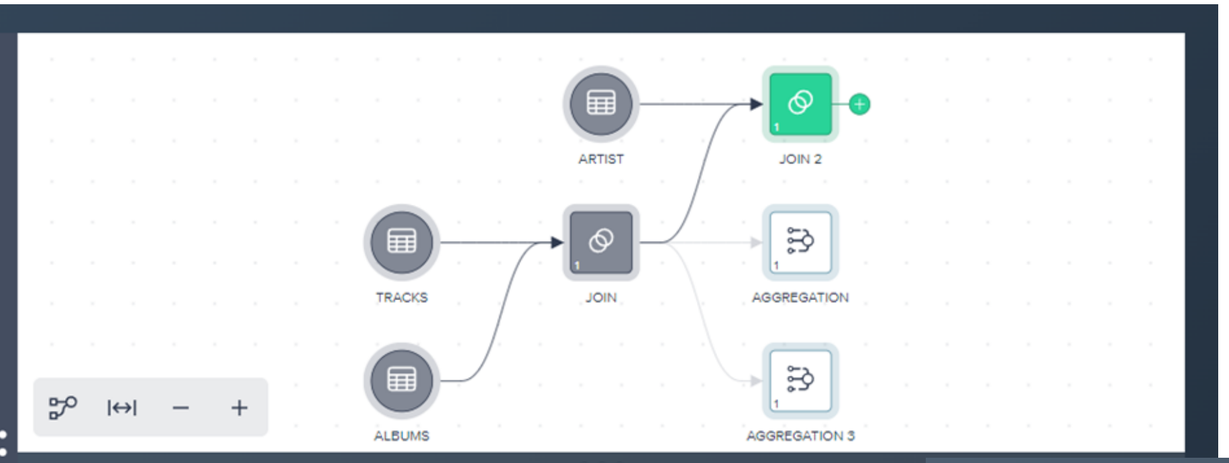
No.of albums composed by each composer along with total price of all the songs and total no.of genres composed





Datameer

Join 3 tables - track, album and artist



Preview | Columns | Profile | SQL | Explore

-- Rows | Calculate | 21 Columns | Explore... | Last run Oct 15

	# track_id	ABC name
1	1	For Those About To Rock (We Sa
2	2	Balls to the Wall
3	3	Fast As a Shark

Join Configuration

Apply

Preview | Unmatched Rows | Join Keys

	# track_id	ABC name	# album_id	# media_type
1	1	For Those About To Rock (We Salute You)	1	1
2	2	Balls to the Wall	2	2
3	3	Fast As a Shark	3	2
4	4	Restless and Wild	3	2
5	5	Princess of the Dawn	3	2
6	6	Put The Finger On You	1	1
7	7	Let's Get It Up	1	1
8	8	Inject The Venom	1	1
9	9	Snowballed	1	1
10	10	Evil Walks	1	1
11	11	C.O.D.	1	1
12	12	Breaking The Rules	1	1
13	13	Night Of The Long Knives	1	1
14	14	Spellbound	1	1
15	15	Go Down	4	1
16	16	Dog Eat Dog	4	1

Combine with the contents of a second dataset based on shared content in one or more columns.
Tell me more...

1. SOURCES

Choose Sources

JOIN | ARTIST

Choose Join Mode

inner join | outer left | outer right | full outer join

2. COLUMNS | Suggestions

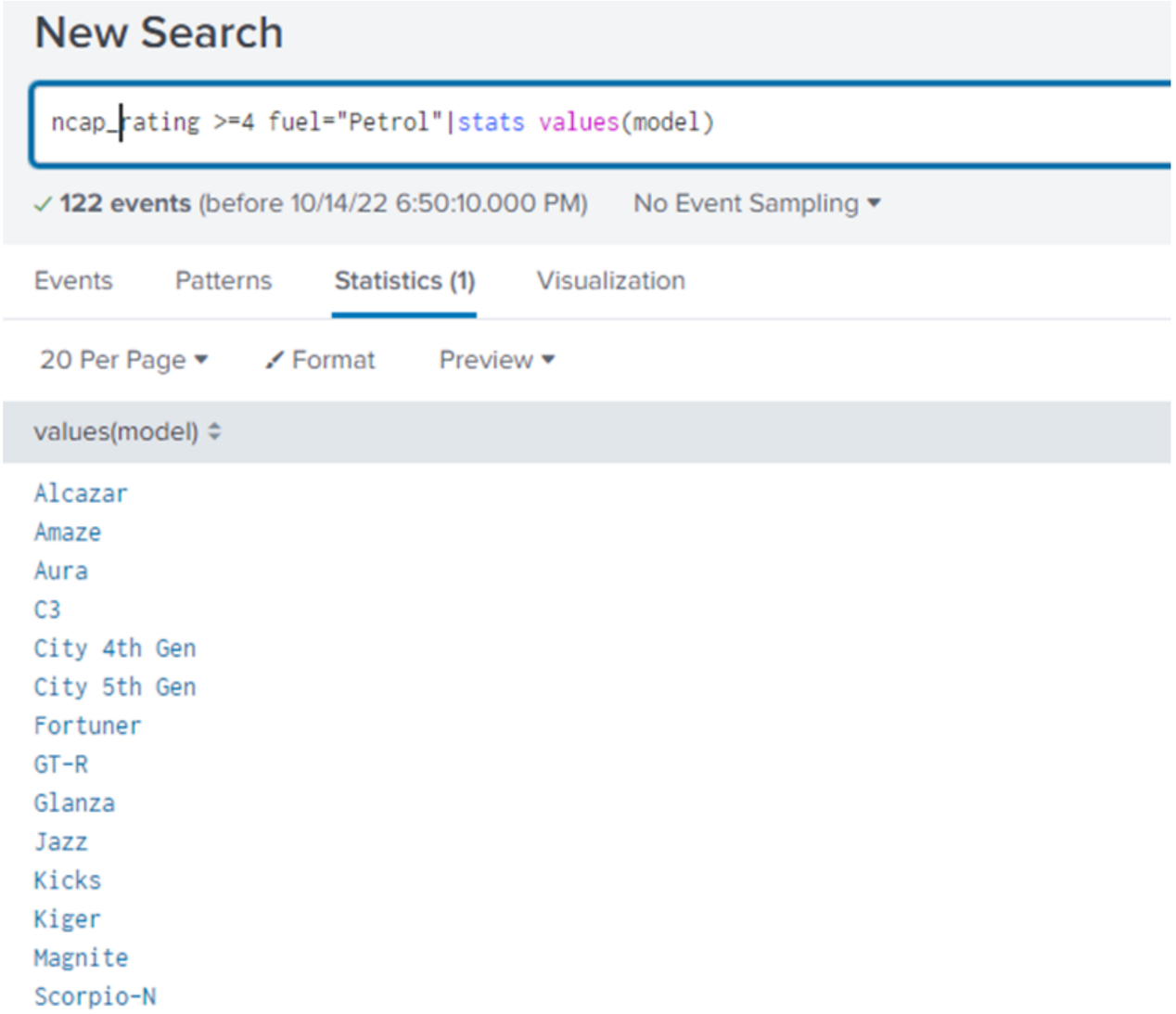
Choose Columns

Select C... | Select C...

artist_id | artist_id

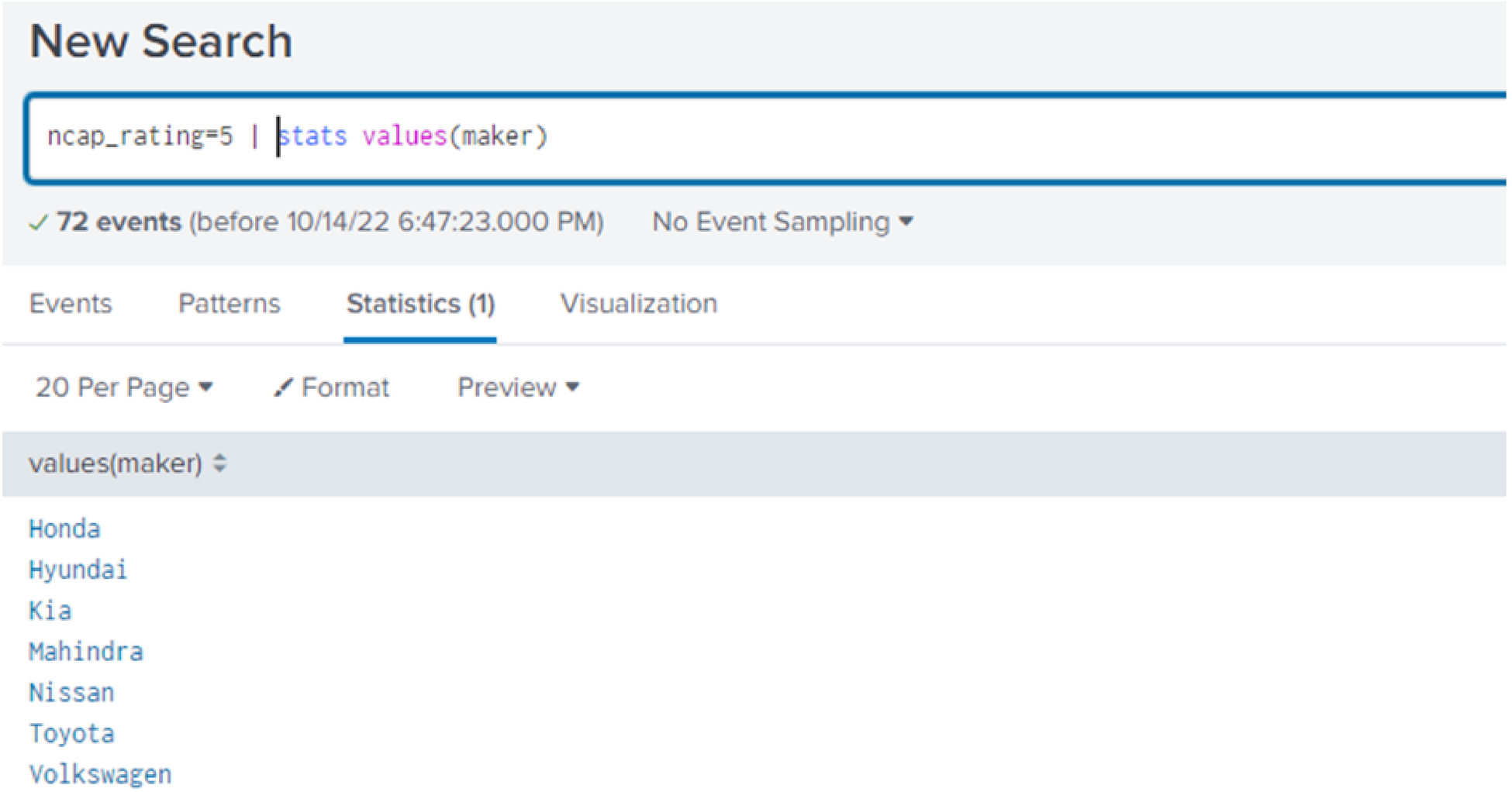
Splunk

List of all petrol car models with high safety measures i.e NCAP Rating greater than or equal to 4



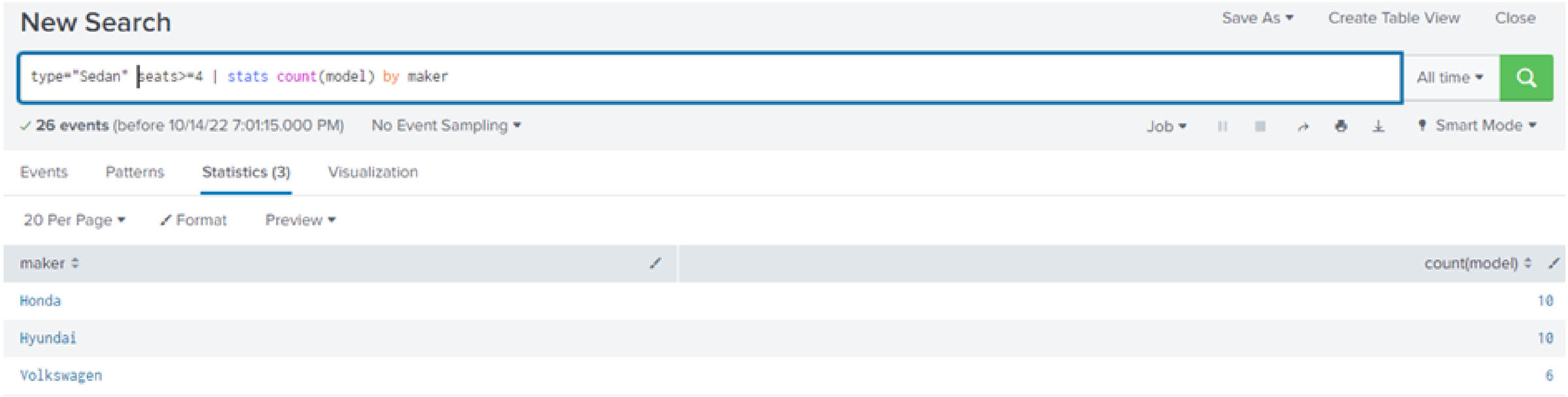
Splunk

List of all car makers who produce cars with high safety measure with an NCAP Rating of 5



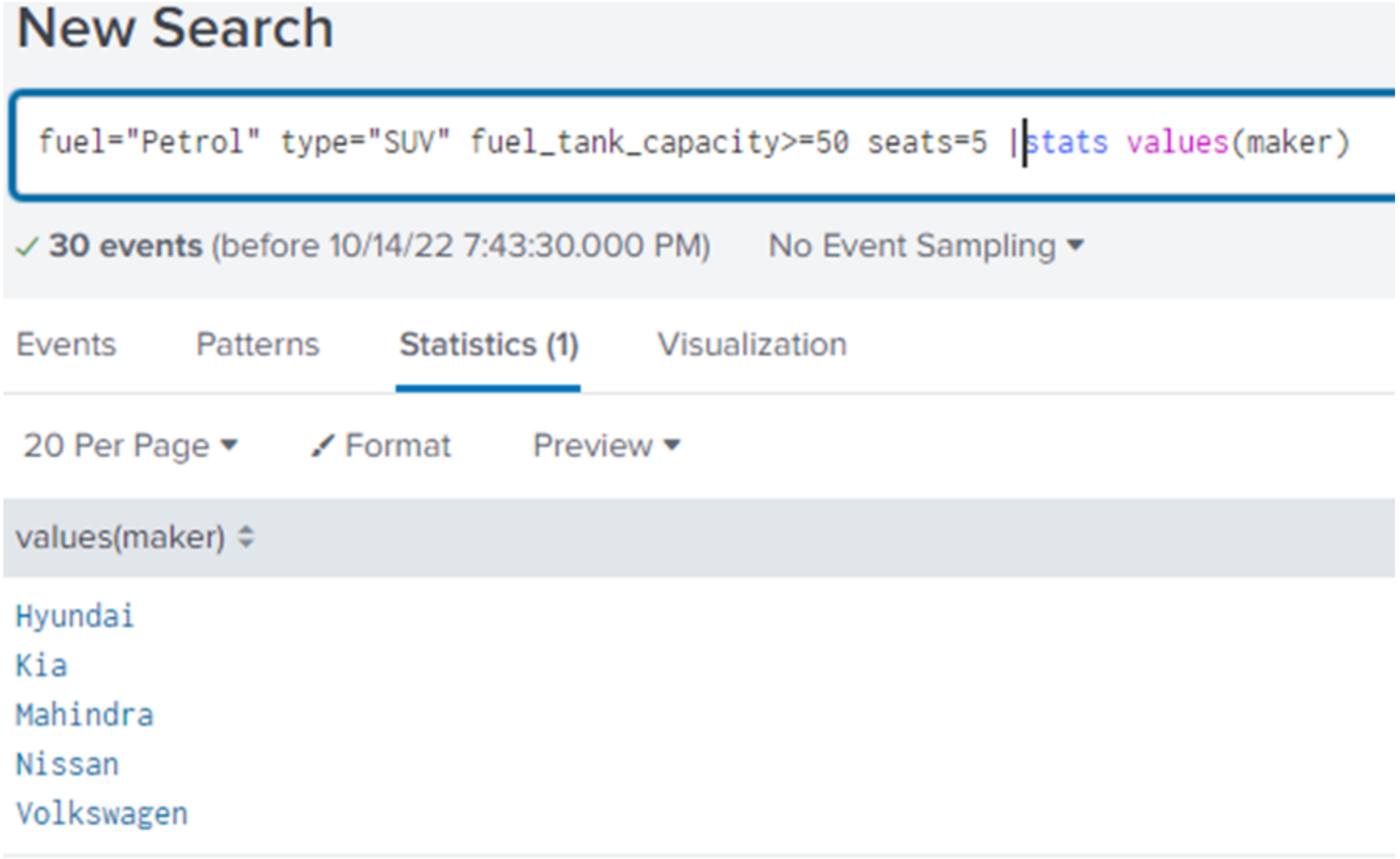
Splunk

No.of different Sedan type car models produced by each maker with more than 4 seats



Splunk

Find list of types of Petrol based SUV cars that have fuel tank capacity >= 50 and with 5 seats



Splunk

Type wise no.of cars produced by each maker

New Search

Save As ▾ Create Table View Close

```
source="cars_india_dataset.csv" | stats count(model) by maker,type
```

All time ▾

✓ **468 events** (before 10/14/22 7:11:22.000 PM) No Event Sampling ▾ Job ▾ Smart Mode ▾

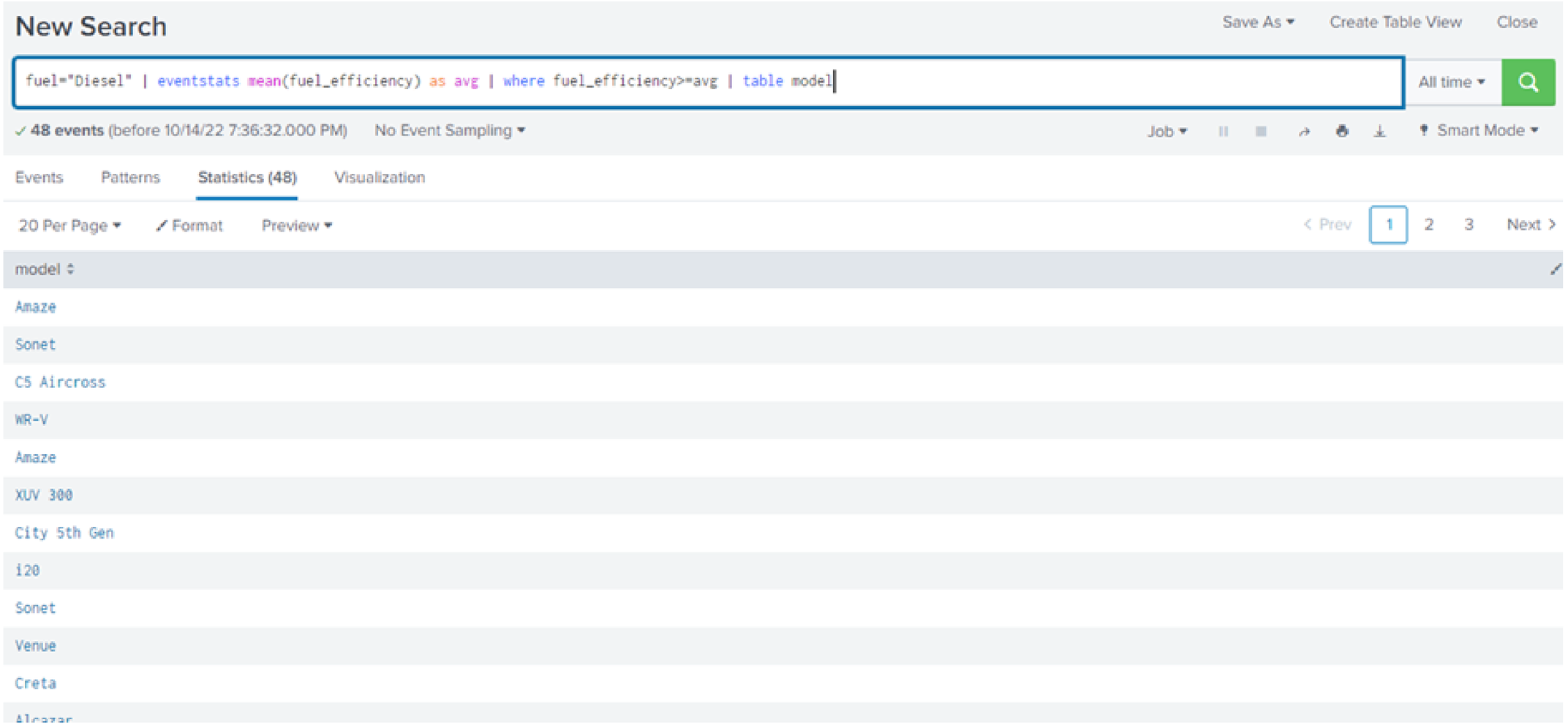
Events Patterns **Statistics (31)** Visualization

20 Per Page ▾ Format Preview ▾ < Prev 1 2 Next >

maker	type	count(model)
Citroen	Hatchback	4
Citroen	SUV	2
Honda	Compact SUV	4
Honda	Compact Sedan	8
Honda	Hatchback	4
Honda	Sedan	10
Hyundai	Compact SUV	10
Hyundai	Compact Sedan	8
Hyundai	Hatchback	22
Hyundai	SUV	34
Hyundai	Sedan	10
Kia	Compact SUV	10

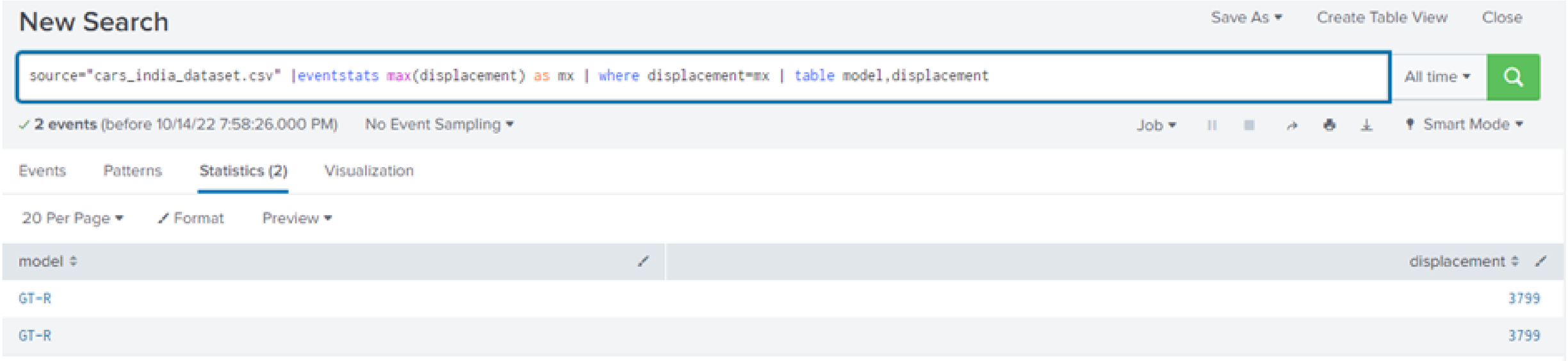
Splunk

List of all Diesel cars that have a fuel efficiency greater than the average fuel efficiency



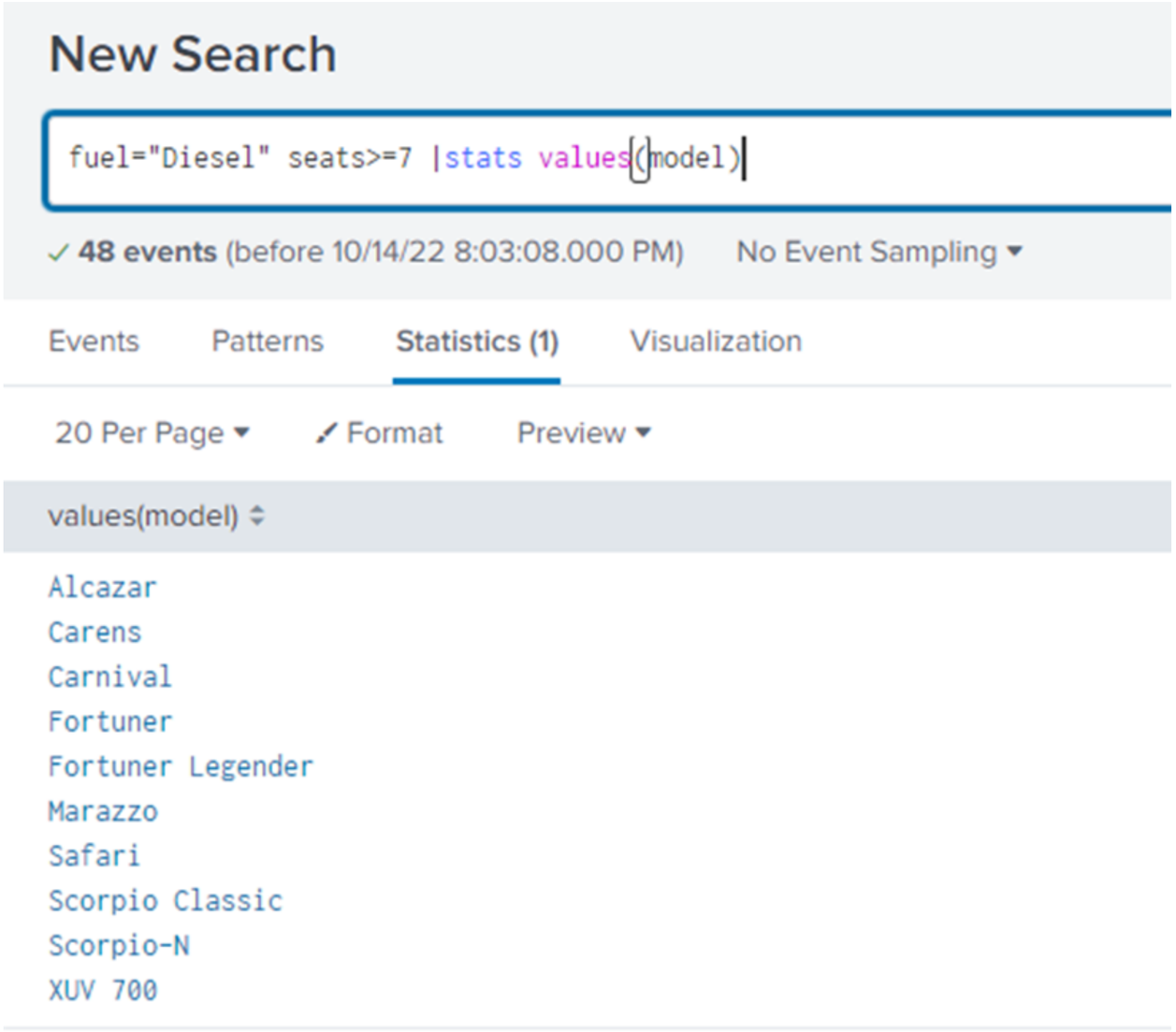
Splunk

Model of the car with highest displacement



Splunk

Distinct car models that has 7 seats and uses diesel consumption



Splunk

Visualizing types of car models via a pie chart

New Search

Save As ▾ Create Table View Close

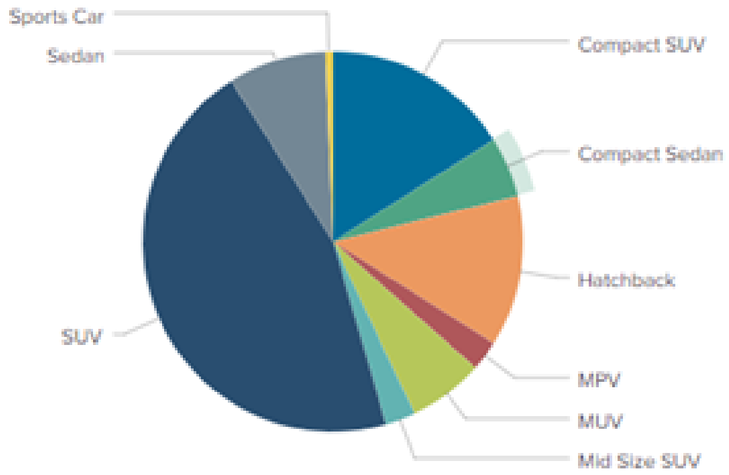
```
source="cars_india_dataset.csv" | stats count(model) by type
```

All time ▾

✓ **468 events** (before 10/14/22 8:10:13.000 PM) No Event Sampling ▾ Job ▾ Smart Mode ▾

Events Patterns Statistics (31) **Visualization**

Pie Chart Format Trellis



Splunk

Visualizing different car makers via a pie chart

New Search

Save As ▾ Create Table View Close

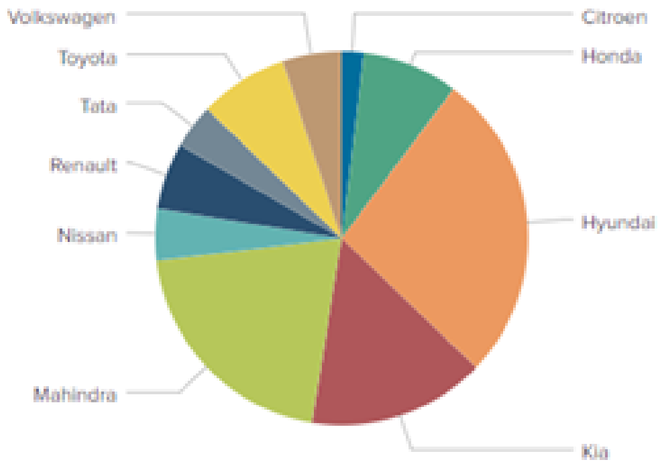
```
source="cars_india_dataset.csv" | stats count(model) by maker
```

All time ▾

✓ **468 events** (before 10/14/22 8:11:11.000 PM) No Event Sampling ▾ Job ▾ Smart Mode ▾

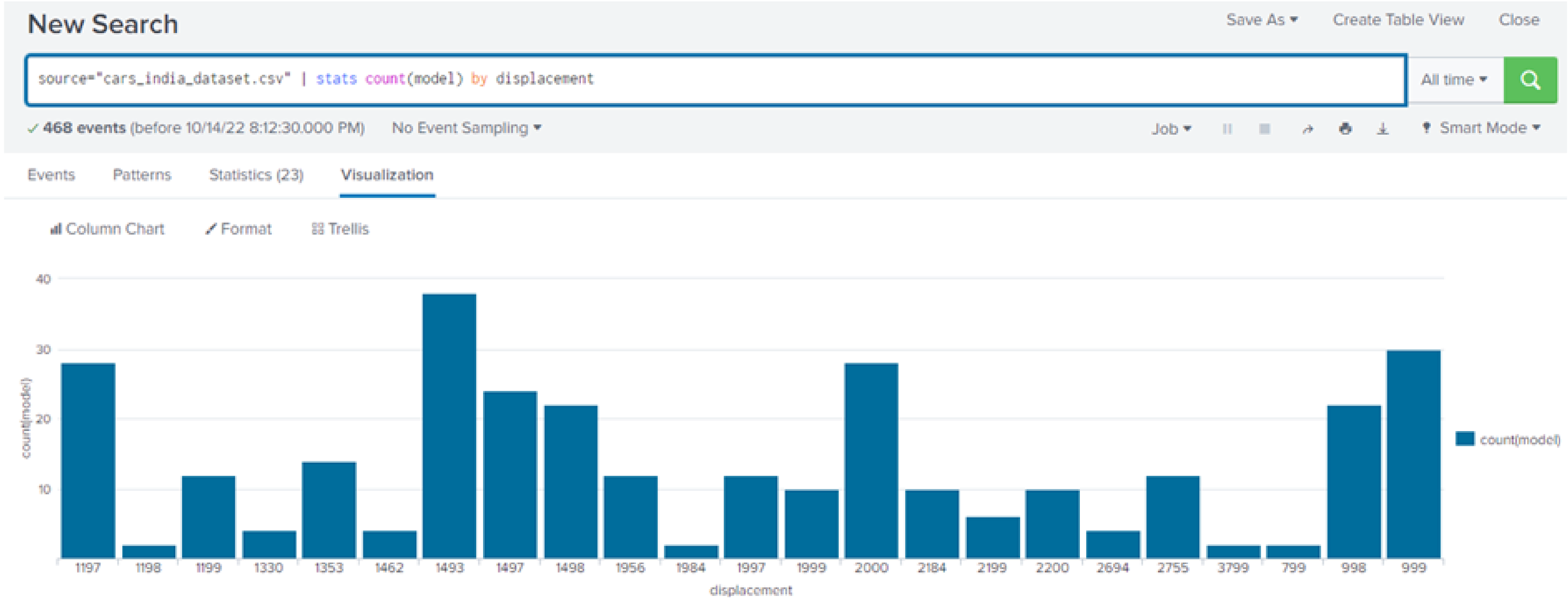
Events Patterns Statistics (10) **Visualization**

Pie Chart Format Trellis



Splunk

Visualizing no.of cars under each displacement via a bar chart



MongoDB

Updating the categorical values such as “Not Rated”, “Not Tested” as 0 in NCAP Rating column

```
> db.car.updateMany({"NCAP Rating": {$in:["Not Tested","Not Rated"]}},{$set:{"NCAP Rating":0}})
{ "acknowledged" : true, "matchedCount" : 33, "modifiedCount" : 33 }
```

MongoDB

Display all car models whose transmission is 6AT or 6MT

```
> db.car.distinct("Model",{"Transmission":{$in:["6 AT","6 MT"]}})
[
        "Alcazar",
        "C3",
        "Carens",
        "City 5th Gen",
        "Creta",
        "Fortuner",
        "Fortuner Legender",
        "Harrier",
        "Kicks",
        "Marazzo",
        "Safari",
        "Scorpio Classic",
        "Scorpio-N",
        "Seltos",
        "Sonet",
        "Thar AX OPT",
        "Thar LX",
        "Tucson",
        "Venue",
        "Verna",
        "WR-V",
        "XUV 300",
        "XUV 700",
        "i20"
]
```

MongoDB

Find all car makers who produce cars with NCAP Rating between 3 and 4. And fuel efficiency is less than 10

```
> db.car.find({"Fuel Efficiency":{$lte:10},"NCAP Rating":{$lte:5,$gte:3}}).pretty()
{
        "_id" : ObjectId("634ae70ae6b4582cfee10143"),
        "Model" : "Thar LX",
        "Maker" : "Mahindra",
        "Type" : "SUV",
        "Seats" : 4,
        "Displacement" : 2184,
        "Length" : 3985,
        "Width" : 1855,
        "Height" : 1844,
        "Wheelbase" : 2450,
        "No_of_Cylinders" : 4,
        "Fuel" : "Diesel",
        "Engine Type" : "mHawk 130",
        "Transmission" : "6 TC Automatic",
        "Front Brake" : "Disc",
        "Rear Brake" : "Drum",
        "Drive" : "4WD",
        "Turning Radius" : 5.25,
        "Fuel Tank Capacity" : 57,
        "Boot Space" : "",
        "Fuel Efficiency" : 9,
        "Emission Type" : "BS VI",
        "Tyre Size" : "255/65 R18",
        "Variants" : 1,
        "NCAP Rating" : 4
}
```

MongoDB

No.of different types of Petrol cars manufactured by Mahindra car maker

```
> db.car.aggregate([{$match:{"Fuel":{$ne:"Diesel"},"Maker":"Mahindra"}},{$count:"carTypes"}])
{ "carTypes" : 13 }
```

MongoDB

Average fuel efficiency Model wise

```
> db.car.aggregate([{$group:{_id:"$Model",avgEfficiency:{$avg:"$Fuel Efficiency"}}}])
{ "_id" : "Kona Electric", "avgEfficiency" : null }
{ "_id" : "Aura", "avgEfficiency" : 22.5 }
{ "_id" : "Amaze", "avgEfficiency" : 20.65 }
{ "_id" : "Triber", "avgEfficiency" : 18.6 }
{ "_id" : "Creta", "avgEfficiency" : 18.018 }
{ "_id" : "i20 N-Line", "avgEfficiency" : 20.25 }
{ "_id" : "Thar LX", "avgEfficiency" : 13.649999999999999 }
{ "_id" : "C5 Aircross", "avgEfficiency" : 18.6 }
{ "_id" : "Fortuner Legender", "avgEfficiency" : 14.3 }
{ "_id" : "Scorpio-N", "avgEfficiency" : null }
{ "_id" : "Scorpio Classic", "avgEfficiency" : null }
{ "_id" : "Jazz", "avgEfficiency" : 16.85 }
{ "_id" : "Kicks", "avgEfficiency" : 15.299999999999999 }
{ "_id" : "Marazzo", "avgEfficiency" : 17 }
{ "_id" : "Venue N-Line", "avgEfficiency" : 18.1 }
{ "_id" : "Urban Cruiser", "avgEfficiency" : 17.895000000000003 }
{ "_id" : "GT-R", "avgEfficiency" : 10.16 }
{ "_id" : "Kwid", "avgEfficiency" : 21.46666666666667 }
{ "_id" : "Verna", "avgEfficiency" : 20.32 }
{ "_id" : "i20", "avgEfficiency" : 21.14 }
Type "it" for more
```

MongoDB

Model and type of car wise avg mileage - group by model and type of car

```
> db.car.aggregate([{$group:{_id:{"model":"$Model","type":"$Type"},avgMileage:{$avg:"$Displacement"}}}])
{ "_id" : { "model" : "XUV 300", "type" : "Compact SUV" }, "avgMileage" : 1347 }
{ "_id" : { "model" : "City 5th Gen", "type" : "Sedan" }, "avgMileage" : 1498 }
{ "_id" : { "model" : "Urban Cruiser", "type" : "Compact SUV" }, "avgMileage" : 1462 }
{ "_id" : { "model" : "Amaze", "type" : "Compact Sedan" }, "avgMileage" : 1348.5 }
{ "_id" : { "model" : "XUV 700", "type" : "SUV" }, "avgMileage" : 2111.1111111111113 }
{ "_id" : { "model" : "Tiguan", "type" : "SUV" }, "avgMileage" : 1984 }
{ "_id" : { "model" : "Safari", "type" : "SUV" }, "avgMileage" : 1956 }
{ "_id" : { "model" : "Tucson", "type" : "SUV" }, "avgMileage" : 1997.6666666666667 }
{ "_id" : { "model" : "Kiger", "type" : "Compact SUV" }, "avgMileage" : 999 }
{ "_id" : { "model" : "Kwid", "type" : "Hatchback" }, "avgMileage" : 932.3333333333334 }
{ "_id" : { "model" : "C5 Aircross", "type" : "SUV" }, "avgMileage" : 1997 }
{ "_id" : { "model" : "Taigun", "type" : "Mid Size SUV" }, "avgMileage" : 1248.5 }
{ "_id" : { "model" : "GT-R", "type" : "Sports Car" }, "avgMileage" : 3799 }
{ "_id" : { "model" : "Kicks", "type" : "SUV" }, "avgMileage" : 1386 }
{ "_id" : { "model" : "Scorpio-N", "type" : "SUV" }, "avgMileage" : 2000 }
{ "_id" : { "model" : "Thar LX", "type" : "SUV" }, "avgMileage" : 2090.5 }
{ "_id" : { "model" : "Thar AX OPT", "type" : "SUV" }, "avgMileage" : 2090.5 }
{ "_id" : { "model" : "Marazzo", "type" : "MPV" }, "avgMileage" : 1497 }
{ "_id" : { "model" : "Scorpio Classic", "type" : "SUV" }, "avgMileage" : 2184 }
{ "_id" : { "model" : "Glanza", "type" : "Hatchback" }, "avgMileage" : 1197 }
Type "it" for more
```

MongoDB

Max NCAP Rating according to car maker wise, among all those 5 seater Petrol cars

```
> db.car.aggregate([{$match:{"Seats":5,"Fuel":"Petrol"}},{$group:{_id:"$Maker",maxRating:{$max:"$NCAP Rating"}}},{$sort:{maxRating:-1}}])
{ "_id" : "Volkswagen", "maxRating" : 5 }
{ "_id" : "Honda", "maxRating" : 5 }
{ "_id" : "Nissan", "maxRating" : 5 }
{ "_id" : "Mahindra", "maxRating" : 5 }
{ "_id" : "Toyota", "maxRating" : 4 }
{ "_id" : "Hyundai", "maxRating" : 4 }
{ "_id" : "Renault", "maxRating" : 4 }
{ "_id" : "Kia", "maxRating" : 3 }
{ "_id" : "Citroen", "maxRating" : 0 }
>
```

Thank You !