df\_pd = df.toPandas()

df\_list = df\_pd['categorias'].tolist()

l = list(map(lambda x: x.split(", "),df\_list))

cat\_list = sorted(list(dict.fromkeys(list(reduce(lambda x,y: x+y,l)))),key=len)

<https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/5991500465045868/3662336169610342/727247365595281/latest.html>

https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/763780990988853/900404702235493/2620394790310521/latest.html

Seguem as informações do grupo:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Nome | E-mail | Time | Usuario AWS | Senha | link conta AWS | account |
| Pedro |  | los bankeros | pedro | Labdatafiahacka2020 | <https://139538972606.signin.aws.amazon.com/console> | aws account6 |
| Washington | washington.gbjr@gmail.com | los bankeros | washington | Labdatafiahacka2020 | <https://139538972606.signin.aws.amazon.com/console> | aws account6 |
| Jaime |  | los bankeros | jaime | Labdatafiahacka2020 | <https://139538972606.signin.aws.amazon.com/console> | aws account6 |
| Ariel |  | los bankeros | ariel | Labdatafiahacka2020 | <https://139538972606.signin.aws.amazon.com/console> | aws account6 |

1) acesse o link conta AWS informa na tabela acima

2) use o usuario AWS e senha informados na tabela acima para acessar a conta

<https://spark.apache.org/docs/latest/ml-features#onehotencoder-deprecated-since-230>

https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/5991500465045868/3662336169610342/727247365595281/latest.html

# Ecommerce

## 1.1

select touchproduct as produto, count(touchproduct) as qtde from "sessions\_ecommerce2"."rawdata"

where touchproduct <> 0

group by touchproduct

order by qtde DESC

limit 10

|  |  |  |
| --- | --- | --- |
|  | **produto** | **qtde** |
| 1 | 121 | 324 |
| 2 | 149 | 321 |
| 3 | 17 | 319 |
| 4 | 54 | 318 |
| 5 | 56 | 314 |
| 6 | 7 | 314 |
| 7 | 189 | 313 |
| 8 | 75 | 312 |
| 9 | 236 | 311 |
| 10 | 161 | 310 |

3.1

select device, count(device) as qtde from "sessions\_ecommerce2"."aggregated"

group by device

order by qtde DESC

|  |  |  |
| --- | --- | --- |
|  | **device** | **qtde** |
| 1 | app\_tablet | 11217 |
| 2 | browser | 11103 |
| 3 | app\_mobile | 11031 |