Apache Hive ile FLO Veri Analizi

1. **Sorgu performansını önceleyecek şekilde bir tablo oluşturunuz.**

   Csv dosyasını inceleyerek oluşturulacak Hive tablosunun şemasını çıkarınız.
   Tablo oluşturunuz.

```
beeline -u jdbc:hive2://localhost:10000
```

```
drop table test1.flo_transactions ;
create table if not exists test1.flo_transactions
(master_id string,
order_channel string,
platform_type string,
last_order_channel string,
first_order_date string,
last_order_date string,
last_order_date_online string,
last_order_date_offline string,
order_num_total_ever_online int,
order_num_total_ever_offline int,
customer_value_total_ever_offline float,
customer_value_total_ever_online float,
interested_in_categories_12 array<string>,
online_product_group_amount_top_name_12 string,
offline_product_group_name_12 string,
last_order_date_new date,
store_type string )
row format delimited
fields terminated by '|'
lines terminated by '\n'
stored as textfile
tblproperties('skip.header.line.count'='1');
```

2. Veri setini oluşturduğunuz tabloya yükleyiniz.
   Csv dosyasını tabloya yükleyiniz.

   Mevcutta olan tablonun üstüne direkt yazar.

   LOCAL hdfs değil, **linux alt dizinindeki veri**

```
LOAD DATA LOCAL INPATH '/home/train/datasets/flo100k.csv' OVERWRITE INTO TABLE test1.flo_transactions;
```

   LOAD DATA INPATH `hdfs` üzerinden veriyi hivea aktarır.

```
LOAD DATA INPATH "/user/train/flo_odev/flo100k.csv" into table test1.flo_transactions;
```

Yükleme sonrası kontrol.

```
SELECT * FROM test1.flo_transactions LIMIT 15;
```

ORC çevirmeden önce kontrol edelim

```
select
cast(concat(substr(first_order_date, 1,10), ' ',
substr(first_order_date,12,8)) as timestamp)
from test1.flo_transactions LIMIT 15;
```

Mevcut tabloyu kullanarak ORC tablosu oluşturalım

```
create table test1.flo_transactions_orc stored as orc
as select
master_id,
order_channel,
platform_type,
last_order_channel,
cast(concat(substr(first_order_date, 1,10), ' ',
substr(first_order_date,12,8)) as timestamp) first_order_date,
cast(concat(substr(last_order_date, 1,10), ' ',
substr(last_order_date,12,8)) as timestamp) last_order_date,
cast(concat(substr(last_order_date_online, 1,10), ' ',
substr(last_order_date_online,12,8)) as timestamp) last_order_date_online,
cast(concat(substr(last_order_date_offline, 1,10), ' ',
substr(last_order_date_offline,12,8)) as timestamp) last_order_date_offline,
order_num_total_ever_online,
order_num_total_ever_offline,
customer_value_total_ever_offline,
customer_value_total_ever_online,
interested_in_categories_12,
online_product_group_amount_top_name_12,
offline_product_group_name_12,
last_order_date_new,
store_type
from test1.flo_transactions ;
```

Kontrol

```
describe test1.flo_transactions_orc;
SELECT * FROM test1.flo_transactions_orc LIMIT 15;
```

3. Mağaza türlerine (store_type) göre işlem sayılarını bulan sorguyu hazırlayınız.

```
SELECT store_type, COUNT(1) as total_count
FROM test1.flo_transactions_orc
GROUP BY store_type
ORDER BY total_count DESC;
```

Result

```
store_type|total_count|
----------+-----------+
A         |      89225|
A,B       |       8497|
B         |       1491|
A,C       |        702|
A,B,C     |         75|
B,C       |         10|
```

4. Sipariş kanallarına (order_channel) göre işlem sayılarını bulan sorguyu hazırlayınız.

```
SELECT order_channel, COUNT(1) as total_count
FROM test1.flo_transactions_orc
GROUP BY order_channel
ORDER BY total_count DESC;
```

```
order_channel|total_count|
-------------+-----------+
Offline      |      70784|
Android App  |      11989|
```

```
Mobile       |       8512|
Desktop      |       4751|
Ios App      |       3964|
```

5. İlk sipariş yılı (first_order_year) baz alınarak yıllara göre işlem sayılarını bulan sorguyu hazırlayınız.

```
set hive.groupby.orderby.position.alias=true;
```

```
SELECT YEAR (first_order_date) AS first_order_year, COUNT(1) as total_count
FROM test1.flo_transactions_orc
GROUP BY 1
ORDER BY total_count DESC;
```

Result

```
first_order_year|total_count|
----------------+-----------+
            2019|      51604|
            2020|      28605|
            2021|      11807|
            2018|       3737|
            2017|       1742|
            2016|       1070|
            2015|        771|
            2014|        445|
            2013|        219|
```

6. Omni-channel satışlarında en çok değer getiren 15 müşteriyi bulan sorguyu hazırlayınız.

```
SELECT master_id, customer_value_total_ever_offline,
customer_value_total_ever_online,
(customer_value_total_ever_offline + customer_value_total_ever_online) as
customer_value
FROM test1.flo_transactions_orc
WHERE customer_value_total_ever_offline > 0.0 AND
customer_value_total_ever_online > 0.0
ORDER BY customer_value DESC
LIMIT 15;
```