



**FLO**

Apache Hive ile FLO Veri Analizi



**miuul**



# İş Problemi

Online ayakkabı mağazası olan FLO'nun verisi üzerinde sorgu performansı öncelikli olacak şekilde Apache Hive'da tablo oluşturulup, istenen sorguların yapılması gerekmektedir.

Veri seti Flo'dan son alışverişlerini 2020 - 2021 yıllarında alışveriş yapan müşterilerin geçmiş alışveriş davranışlarından elde edilen bilgilerden oluşmaktadır.

16 Değişken	100K Gözlem	20MB
<b>master_id</b>	Eşsiz müşteri numarası	
<b>order_channel</b>	Alışveriş yapılan platforma ait hangi kanalın kullanıldığı (Android, ios, Desktop, Mobile)	
<b>platform_type</b>	Alışveriş yapılan platform tipi. (Online, Offline, Omnichannel) <i>Omnichannel: Hem online hem offline alışveriş.</i>	
<b>last_order_channel</b>	En son alışverişin yapıldığı kanal.	
<b>first_order_date</b>	Müşterinin yaptığı ilk alışveriş tarihi	
<b>last_order_date</b>	Müşterinin yaptığı son alışveriş tarihi	
<b>last_order_date_online</b>	Müşterinin online platformda yaptığı son alışveriş tarihi	
<b>last_order_date_offline</b>	Müşterinin offline platformda yaptığı son alışveriş tarihi	
<b>order_num_total_ever_online</b>	Müşterinin online platformda yaptığı toplam alışveriş sayısı	
<b>order_num_total_ever_offline</b>	Müşterinin offline'da yaptığı toplam alışveriş sayısı	
<b>customer_value_total_ever_offline</b>	Müşterinin offline alışverişlerinde ödediği toplam ücret	
<b>customer_value_total_ever_online</b>	Müşterinin online alışverişlerinde ödediği toplam ücret	
<b>interested_in_categories_12</b>	Müşterinin son 12 ayda alışveriş yaptığı kategorilerin listesi	
<b>online_product_group_amount_top_name_12</b>	Müşterinin son 12 ayda online da en çok para harcadığı ürün grubu	
<b>offline_product_group_name_12</b>	Müşterinin offline da son 12 ayda en çok para harcadığı ürün grubu	
<b>last_order_date_new</b>	Last order date'in data hazırlarken değiştirilmiş hali (aynı alanı ifade eder.)	
<b>store_type</b>	3 farklı company'i ifade eder. A company'sinden alışveriş yapan kişi B'dende yaptı ise A,B şeklinde yazılmıştır.	



# Proje Görevleri

---

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

**Adım 1:** Csv dosyasını inceleyerek oluşturulacak Hive tablosunun şemasını çıkarınız.

**Adım 2:** Bu şemaya göre CREATE TABLE komutunu kullanarak tabloyu oluşturunuz.

**Görev 2:** Veri setini oluşturduğunuz tabloya yükleyiniz.

---

**Adım 1:** Csv dosyasını, Görev 1’de oluşturduğunuz tabloya yükleyiniz.

**Adım 2: SELECT** komutu ile yükleme sonrası kontrol ediniz.

**Adım 3:** Mevcut tabloyu kullanarak yeni bir **ORC** formatlı tablo oluşturunuz.

**Not:** ORC formatına çevirmeden önce kontrol ediniz.

## Görev 3: Hive Sorguları ile veriyi analiz etme

**Adım 1:** Mağaza türlerine (**store\_type**) göre işlem sayılarını bulan sorguyu hazırlayınız.

**Adım 2 :** Sipariş kanallarına (**order\_channel**) göre işlem sayılarını bulan sorguyu hazırlayınız.

**Adım 3 :** İlk sipariş yılı (**first\_order\_year**) baz alınarak yıllara göre işlem sayılarını bulan sorguyu hazırlayınız.

**Adım 4 : Omni-channel** satışlarında en çok değer getiren 15 müşteriyi bulan sorguyu hazırlayınız.

**miuul**

miuul.com