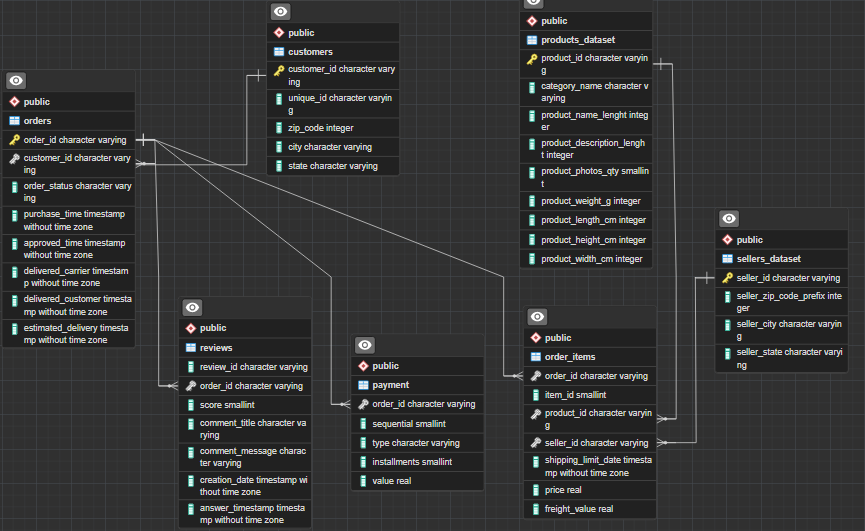
### SQL PROJE

Verilen linkteki veri setinden veri tabanını oluşturarak ERD oluşturunuz ve görsel olarak ekleyiniz. Geolocation tablosu projede kullanılmayacaktır.

Link : [Brazilian E-Commerce Public Dataset by Olist | Kaggle](https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce)

Aşağıda örnek olarak travel veri tabanının ERD görseli eklenmiştir.



#### **Case 1 : Sipariş Analizi**

##### Question 1 :

-Aylık olarak order dağılımını inceleyiniz. Tarih verisi için order\_approved\_at kullanılmalıdır.

SELECT

DATE\_TRUNC('month', order\_approved\_at) AS order\_month,

COUNT(\*) AS total\_orders

FROM

olist\_orders\_dataset

WHERE

order\_approved\_at IS NOT NULL

GROUP BY

order\_month

ORDER BY

order\_month;

##### Question 2 :

-Aylık olarak order status kırılımında order sayılarını inceleyiniz. Sorgu sonucunda çıkan outputu excel ile görselleştiriniz. Dramatik bir düşüşün ya da yükselişin olduğu aylar var mı? Veriyi inceleyerek yorumlayınız.

SELECT

TO\_CHAR(order\_purchase\_timestamp, 'YYYY-MM') AS year\_month,

order\_status,

COUNT(\*) AS order\_count

FROM

olist\_orders\_dataset

GROUP BY

year\_month,

order\_status

ORDER BY

year\_month DESC,

order\_count DESC;

Grafikte, 2016-10 ile başlayan ve 2018-08 ile biten dönemdeki sipariş sayıları gösteriliyor. Genel bir trend olarak, sipariş sayılarının zaman içinde arttığını görebiliyoruz. Ancak bu artışın sürekli ve düzenli olduğunu söylemek zor. Belli başlı aylar var ki, özellikle 2017-12 ve 2018-01 gibi tatil sezonlarında bir artış göze çarpıyor. Bu, sezonluk alışveriş trendlerinin bir göstergesi olabilir.

Grafikteki dalgalanmaların sebebini tam olarak anlamak için daha fazla veriye ve analize ihtiyaç vardır. Örneğin, sipariş sayısındaki artışın belirli ürün kategorileri, promosyonlar veya özel günlerle ilişkili olup olmadığını incelemek yararlı olabilir. Ayrıca, müşteri yorumları ve geri bildirimler gibi diğer tabloların verileri de sipariş trendleri hakkında daha fazla bilgi verebilir.

Özetlemek gerekirse, grafikte bir yükseliş trendi var, ancak bu trendin nedenleri ve dinamikleri hakkında daha derinlemesine bir analiz yapmak için ek verilere ve bağlam bilgisine ihtiyaç duyuluyor. Sipariş sayılarındaki artışın yanı sıra, ay bazında önemli dalgalanmalar da dikkate alınmalı ve bu dalgalanmaların nedenleri araştırılmalıdır.

##### Question 3 :

-Ürün kategorisi kırılımında sipariş sayılarını inceleyiniz. Özel günlerde öne çıkan kategoriler nelerdir? Örneğin yılbaşı, sevgililer günü…

SELECT

TO\_CHAR(o.order\_purchase\_timestamp, 'YYYY-MM-DD') AS order\_month,

pc.product\_category\_name\_english AS category,

COUNT(DISTINCT o.order\_id) AS order\_count

FROM

olist\_orders\_dataset o

JOIN

order\_items oi ON o.order\_id = oi.order\_id

JOIN

olist\_products\_dataset p ON oi.product\_id = p.product\_id

JOIN

product\_category\_name\_translation pc ON p.product\_category\_name = pc.product\_category\_name

GROUP BY

order\_month,

category

ORDER BY

order\_month, order\_count DESC;

##### Question 4 :

-Haftanın günleri(pazartesi, perşembe, ….) ve ay günleri (ayın 1’i,2’si gibi) bazında order sayılarını inceleyiniz. Yazdığınız sorgunun outputu ile excel’de bir görsel oluşturup yorumlayınız.

SELECT

TO\_CHAR(order\_purchase\_timestamp, 'Month') AS month\_name,

EXTRACT(DAY FROM order\_purchase\_timestamp) AS day\_of\_month,

TO\_CHAR(order\_purchase\_timestamp, 'Day') AS weekday,

COUNT(\*) AS order\_count

FROM

olist\_orders\_dataset

GROUP BY

month\_name,

day\_of\_month,

weekday

ORDER BY

MIN(order\_purchase\_timestamp),

day\_of\_month;

Grafikteki verilere bakıldığında, Cuma günlerinde en yüksek sipariş hacmi olduğunu, Pazartesi günlerinde ise en düşük sipariş hacmi olduğunu görebiliyoruz. Hafta içi günlerde (Pazartesi’den Perşembeye kadar) sipariş sayıları nispeten benzer ve yüksek düzeyde seyretmekte. Bu durum, müşterilerin hafta sonuna girdikçe ve hafta sonu geçirdikçe alışveriş yapma eğiliminde olduğunu gösteriyor olabilir. Cuma günlerinin yüksek olması, hafta sonu için son dakika alışverişlerinin yapılmasından kaynaklanıyor olabilir.

Pazar günlerindeki düşük sipariş hacmi, bu günün genellikle dinlenme günü olarak kabul edilmesi ve bazı perakende ve lojistik hizmetlerinin daha az çalıştığı bir gün olması nedeniyle olabilir. Ayrıca, e-ticaret platformları bazen hafta sonlarında sipariş işleme ve gönderimde yavaşlayabilir.

Özetle, haftanın günlerine göre sipariş hacminde belirgin bir farklılık olduğunu ve özellikle Cuma günlerinin diğer günlerden daha yoğun olduğunu söyleyebiliriz. Bu bulgular, satış ve pazarlama stratejilerini planlamak için kullanılabilir.

#### **Case 2 : Müşteri Analizi**

##### Question 1 :

-Hangi şehirlerdeki müşteriler daha çok alışveriş yapıyor? Müşterinin şehrini en çok sipariş verdiği şehir olarak belirleyip analizi ona göre yapınız.

WITH MusteriSiparisleri AS (

SELECT

c.customer\_unique\_id,

c.customer\_city,

COUNT(\*) as order\_count

FROM

olist\_customers\_dataset c

INNER JOIN

olist\_orders\_dataset o ON c.customer\_id = o.customer\_id

GROUP BY

c.customer\_unique\_id, c.customer\_city

),

SehirSıralaması AS (

SELECT

customer\_unique\_id,

customer\_city,

order\_count,

RANK() OVER (PARTITION BY customer\_unique\_id ORDER BY order\_count DESC) as city\_rank

FROM

MusteriSiparisleri

)

SELECT

customer\_city,

SUM(order\_count) as total\_orders

FROM

SehirSıralaması

WHERE

city\_rank = 1

GROUP BY

customer\_city

ORDER BY

total\_orders DESC;

#### **Case 3: Satıcı Analizi**

##### Question 1 :

-Siparişleri en hızlı şekilde müşterilere ulaştıran satıcılar kimlerdir? Top 5 getiriniz. Bu satıcıların order sayıları ile ürünlerindeki yorumlar ve puanlamaları inceleyiniz ve yorumlayınız.

WITH SatıcıTeslimatSüresi AS (

SELECT

oi.seller\_id,

o.order\_id,

EXTRACT(EPOCH FROM (o.order\_delivered\_customer\_date - o.order\_purchase\_timestamp)) / 3600 AS teslimat\_saati

FROM order\_items oi

INNER JOIN olist\_orders\_dataset o ON oi.order\_id = o.order\_id

WHERE o.order\_delivered\_customer\_date IS NOT NULL

),

OrtalamaTeslimat AS (

SELECT

seller\_id,

AVG(teslimat\_saati) AS ortalama\_teslimat\_zamani

FROM SatıcıTeslimatSüresi

GROUP BY seller\_id

ORDER BY ortalama\_teslimat\_zamani

LIMIT 5

),

SiparişVeYorumlar AS (

SELECT

oi.seller\_id,

COUNT(DISTINCT oi.order\_id) AS sipariş\_sayısı,

COUNT(DISTINCT r.review\_id) AS yorum\_sayısı,

AVG(r.review\_score) AS ortalama\_puan

FROM order\_items oi

INNER JOIN olist\_order\_reviews\_dataset r ON oi.order\_id = r.order\_id

GROUP BY oi.seller\_id

)

SELECT

ot.seller\_id,

ROUND(ot.ortalama\_teslimat\_zamani,2),

sy.sipariş\_sayısı,

sy.yorum\_sayısı,

ROUND(sy.ortalama\_puan,2),

r.review\_comment\_title,

r.review\_comment\_message

FROM OrtalamaTeslimat ot

INNER JOIN SiparişVeYorumlar sy ON ot.seller\_id = sy.seller\_id

INNER JOIN order\_items oi ON ot.seller\_id = oi.seller\_id

INNER JOIN olist\_order\_reviews\_dataset r ON oi.order\_id = r.order\_id

WHERE ot.seller\_id IN (SELECT seller\_id FROM OrtalamaTeslimat)

ORDER BY ot.ortalama\_teslimat\_zamani, r.review\_creation\_date DESC;

Bu sorgunun sonucu, satıcıların ne kadar hızlı ve etkili hizmet verdiklerini anlamamıza yardımcı olur. Top 5 listesi, müşterilere en hızlı ulaşan ve en yüksek müşteri memnuniyetini sağlayan satıcıları gösterir. E-ticaret platformları için bu tür analizler, satıcı performansını izlemek ve geliştirmek, ayrıca müşteri memnuniyetini artırmak için önemlidir.

Bu bilgiler, satıcıların gönderim süreçlerini iyileştirmelerine, müşteri ilişkilerini güçlendirmelerine ve genel olarak hizmet kalitelerini artırmalarına yardımcı olacak şekilde kullanılabilir. Aynı zamanda, platform üzerindeki satıcıların sıralamasını belirlemede veya öne çıkarmada da etkili bir faktör olabilir.

##### Question 2 :

-Hangi satıcılar daha fazla kategoriye ait ürün satışı yapmaktadır?

Fazla kategoriye sahip satıcıların order sayıları da fazla mı?

WITH SatıcıKategorileri AS (

SELECT

oi.seller\_id,

COUNT(DISTINCT p.product\_category\_name) AS kategori\_sayısı

FROM order\_items oi

JOIN olist\_products\_dataset p ON oi.product\_id = p.product\_id

GROUP BY oi.seller\_id

),

SatıcıSiparişSayıları AS (

SELECT

seller\_id,

COUNT(order\_id) AS sipariş\_sayısı

FROM order\_items

GROUP BY seller\_id

)

SELECT

sk.seller\_id,

sk.kategori\_sayısı,

ss.sipariş\_sayısı

FROM SatıcıKategorileri sk

JOIN SatıcıSiparişSayıları ss ON sk.seller\_id = ss.seller\_id

ORDER BY sk.kategori\_sayısı DESC, ss.sipariş\_sayısı DESC;

#### **Case 4 : Payment Analizi**

##### Question 1 :

-Ödeme yaparken taksit sayısı fazla olan kullanıcılar en çok hangi bölgede yaşamaktadır? Bu çıktıyı yorumlayınız.

SELECT c.customer\_state, AVG(p.payment\_installments) AS average\_installments

FROM olist\_customers\_dataset c

JOIN olist\_orders\_dataset o ON c.customer\_id = o.customer\_id

JOIN olist\_order\_payments\_dataset p ON o.order\_id = p.order\_id

GROUP BY c.customer\_state

ORDER BY average\_installments DESC

LIMIT 1;

Bu sorgunun çıktısı, taksitli ödemeyi tercih eden kullanıcıların hangi bölgede yoğunlaştığını gösterir. Bu veri, belirli bir bölgedeki tüketicilerin satın alma gücü, ödeme tercihleri ve taksitli alışveriş eğilimleri hakkında fikir verebilir. Yüksek ortalama taksit sayısı, bölgenin ekonomik durumu, kredi erişimi veya tüketici davranışları gibi çeşitli faktörlerle ilişkilendirilebilir.

Bu tür bir analiz, bölgesel pazarlama stratejileri ve kredi teklifleri için değerli içgörüler sunar. Örneğin, taksitli ödemeyi daha fazla tercih eden bölgelere yönelik özel finansman seçenekleri veya promosyonlar geliştirebilir. Aynı zamanda, taksit seçeneklerinin müşteri tercihleri üzerindeki etkisini anlamak için de kullanılabilir.

##### Question 2 :

-Ödeme tipine göre başarılı order sayısı ve toplam başarılı ödeme tutarını hesaplayınız. En çok kullanılan ödeme tipinden en az olana göre sıralayınız.

**SELECT**

**op.payment\_type,**

**COUNT(DISTINCT o.order\_id) AS successful\_order\_count,**

**SUM(op.payment\_value) AS total\_successful\_payment\_amount**

**FROM**

**olist\_order\_payments\_dataset op**

**INNER JOIN**

**olist\_orders\_dataset o ON op.order\_id = o.order\_id**

**WHERE**

**o.order\_status = 'delivered'**

**GROUP BY**

**op.payment\_type**

**ORDER BY**

**successful\_order\_count DESC,**

**total\_successful\_payment\_amount DESC;**

##### Question 3 :

-Tek çekimde ve taksitle ödenen siparişlerin kategori bazlı analizini yapınız. En çok hangi kategorilerde taksitle ödeme kullanılmaktadır?

**--TEK ÇEKİM ve TAKSİT—**

**SELECT**

**pc.product\_category\_name\_english AS category,**

**SUM(CASE WHEN op.payment\_installments = 1 THEN 1 ELSE 0 END) AS one\_time\_payment\_orders,**

**SUM(CASE WHEN op.payment\_installments > 1 THEN 1 ELSE 0 END) AS installment\_payment\_orders,**

**SUM(CASE WHEN op.payment\_installments = 1 THEN oi.price ELSE 0 END) AS one\_time\_payment\_total,**

**SUM(CASE WHEN op.payment\_installments > 1 THEN oi.price ELSE 0 END) AS installment\_payment\_total**

**FROM**

**olist\_order\_payments\_dataset op**

**INNER JOIN**

**olist\_orders\_dataset o ON op.order\_id = o.order\_id**

**INNER JOIN**

**order\_items oi ON o.order\_id = oi.order\_id**

**INNER JOIN**

**olist\_products\_dataset p ON oi.product\_id = p.product\_id**

**INNER JOIN**

**product\_category\_name\_translation pc ON p.product\_category\_name = pc.product\_category\_name**

**GROUP BY**

**category**

**ORDER BY**

**installment\_payment\_orders DESC, one\_time\_payment\_orders DESC;**

**--EN ÇOK HANGİ KATEGORİLERDE TAKSİTLİ ÖDEME—**

**SELECT**

**pc.product\_category\_name\_english AS category,**

**COUNT(\*) AS number\_of\_installment\_orders**

**FROM**

**olist\_order\_payments\_dataset op**

**INNER JOIN**

**olist\_orders\_dataset o ON op.order\_id = o.order\_id**

**INNER JOIN**

**order\_items oi ON o.order\_id = oi.order\_id**

**INNER JOIN**

**olist\_products\_dataset p ON oi.product\_id = p.product\_id**

**INNER JOIN**

**product\_category\_name\_translation pc ON p.product\_category\_name = pc.product\_category\_name**

**WHERE**

**op.payment\_installments > 1 -- Taksit sayısı 1'den büyük olan ödemeler**

**GROUP BY**

**category**

**ORDER BY**

**number\_of\_installment\_orders DESC;**

#### **Case 5 : RFM Analizi**

**SELECT**

**CustomerID,**

**CURRENT\_DATE - MAX(InvoiceDate) AS Recency,-- Son satın alma tarihinden bugüne kadar geçen gün sayısı**

**COUNT(DISTINCT InvoiceNo) AS Frequency, -- Müşteri başına toplam fatura sayısı**

**SUM(Quantity \* UnitPrice) AS Monetary, -- Müşteri başına toplam harcama**

**CASE**

**WHEN CURRENT\_DATE - MAX(InvoiceDate) > 365 THEN 1**

**WHEN CURRENT\_DATE - MAX(InvoiceDate) BETWEEN 270 AND 365 THEN 2**

**WHEN CURRENT\_DATE - MAX(InvoiceDate) BETWEEN 180 AND 270 THEN 3**

**WHEN CURRENT\_DATE - MAX(InvoiceDate) BETWEEN 90 AND 180 THEN 4**

**ELSE 5**

**END AS RecencyScore,**

**CASE**

**WHEN COUNT(DISTINCT InvoiceNo) = 1 THEN 1**

**WHEN COUNT(DISTINCT InvoiceNo) = 2 THEN 2**

**WHEN COUNT(DISTINCT InvoiceNo) BETWEEN 3 AND 5 THEN 3**

**WHEN COUNT(DISTINCT InvoiceNo) BETWEEN 6 AND 8 THEN 4**

**ELSE 5**

**END AS FrequencyScore,**

**CASE**

**WHEN SUM(Quantity \* UnitPrice) < 100 THEN 1**

**WHEN SUM(Quantity \* UnitPrice) BETWEEN 100 AND 200 THEN 2**

**WHEN SUM(Quantity \* UnitPrice) BETWEEN 200 AND 500 THEN 3**

**WHEN SUM(Quantity \* UnitPrice) BETWEEN 500 AND 1000 THEN 4**

**ELSE 5**

**END AS MonetaryScore,**

**ROUND(**

**(CASE**

**WHEN CURRENT\_DATE - MAX(InvoiceDate) > 365 THEN 1**

**WHEN CURRENT\_DATE - MAX(InvoiceDate) BETWEEN 270 AND 365 THEN 2**

**WHEN CURRENT\_DATE - MAX(InvoiceDate) BETWEEN 180 AND 270 THEN 3**

**WHEN CURRENT\_DATE - MAX(InvoiceDate) BETWEEN 90 AND 180 THEN 4**

**ELSE 5**

**END +**

**CASE**

**WHEN COUNT(DISTINCT InvoiceNo) = 1 THEN 1**

**WHEN COUNT(DISTINCT InvoiceNo) = 2 THEN 2**

**WHEN COUNT(DISTINCT InvoiceNo) BETWEEN 3 AND 5 THEN 3**

**WHEN COUNT(DISTINCT InvoiceNo) BETWEEN 6 AND 8 THEN 4**

**ELSE 5**

**END +**

**CASE**

**WHEN SUM(Quantity \* UnitPrice) < 100 THEN 1**

**WHEN SUM(Quantity \* UnitPrice) BETWEEN 100 AND 200 THEN 2**

**WHEN SUM(Quantity \* UnitPrice) BETWEEN 200 AND 500 THEN 3**

**WHEN SUM(Quantity \* UnitPrice) BETWEEN 500 AND 1000 THEN 4**

**ELSE 5**

**END) / 3.0, 2) AS RFMScore FROM rfm\_analizi**

**GROUP BY CustomerID;**