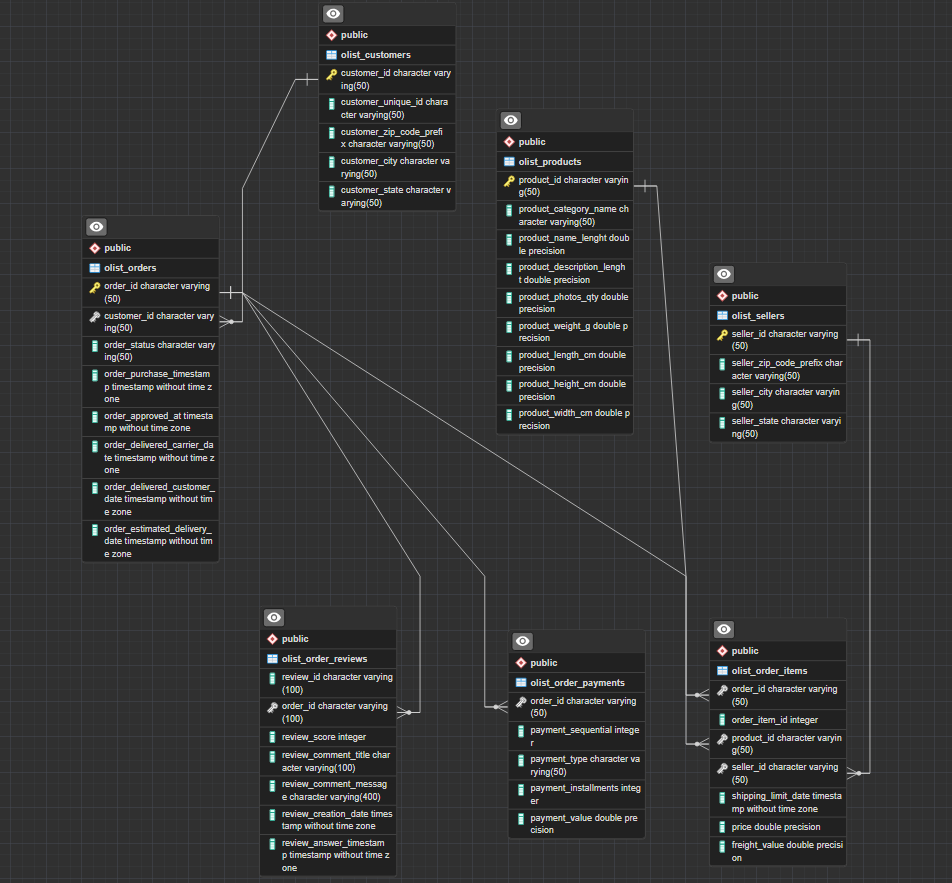
**KODLASAM SQL PROJESİ  
AHMET HAKAN AYDOĞAN**

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

[Brazilian E-Commerce Public Dataset by Olist](https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce) geolocation dataseti olmadan ERD Diagramı



#### **Case 1 : Sipariş Analizi**

##### **Question 1 :**

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

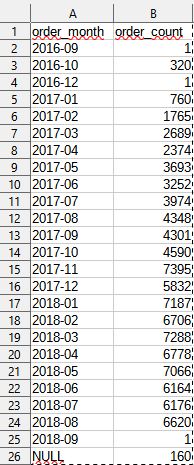
**SQL Query:**

SELECT to\_char(order\_approved\_at,'YYYY-MM') AS order\_month,

count(order\_id) AS order\_count

FROM olist\_orders

GROUP 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.

**SQL Query:**

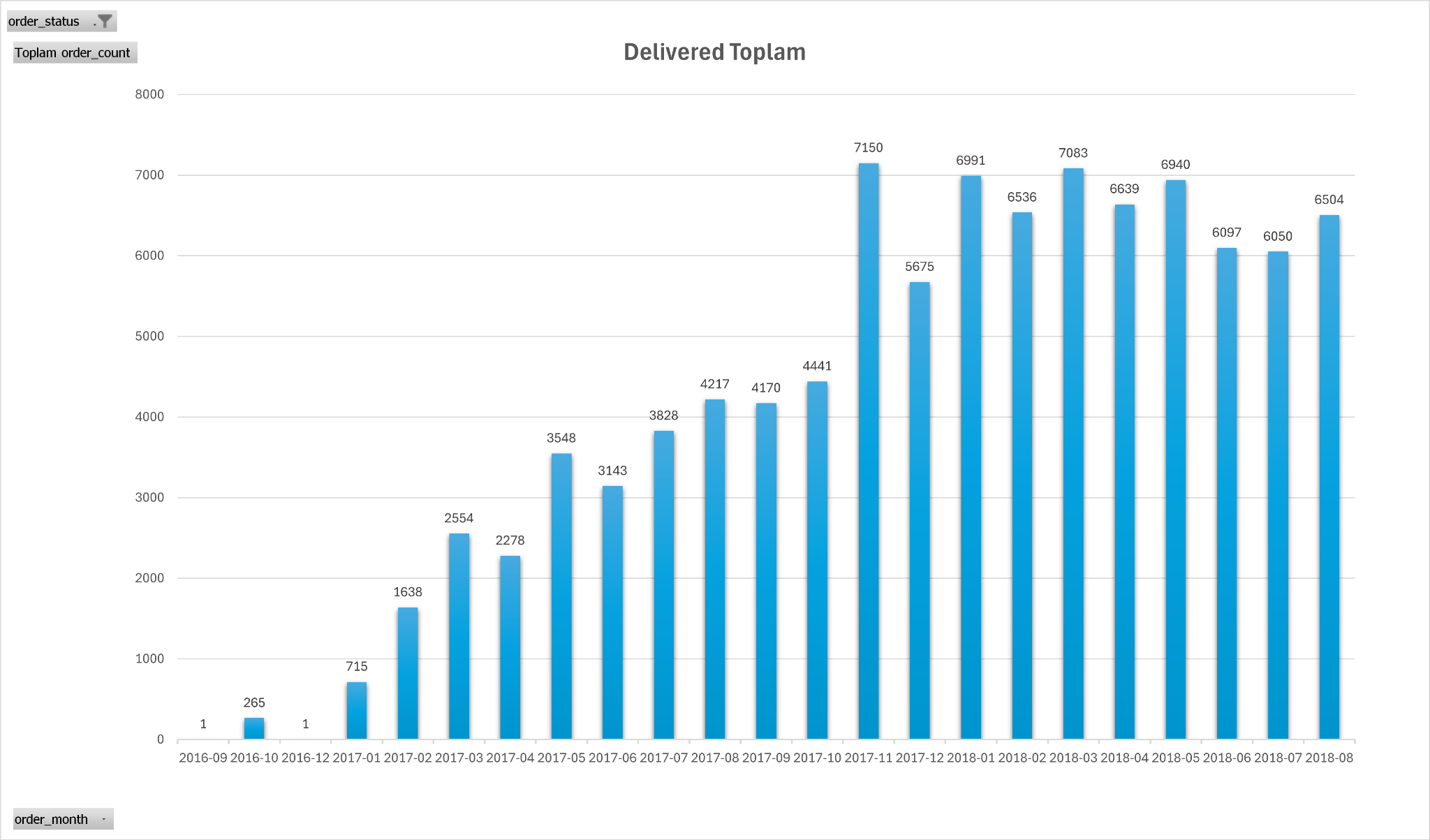
SELECT to\_char(order\_approved\_at,'YYYY-MM') AS order\_month,

order\_status,

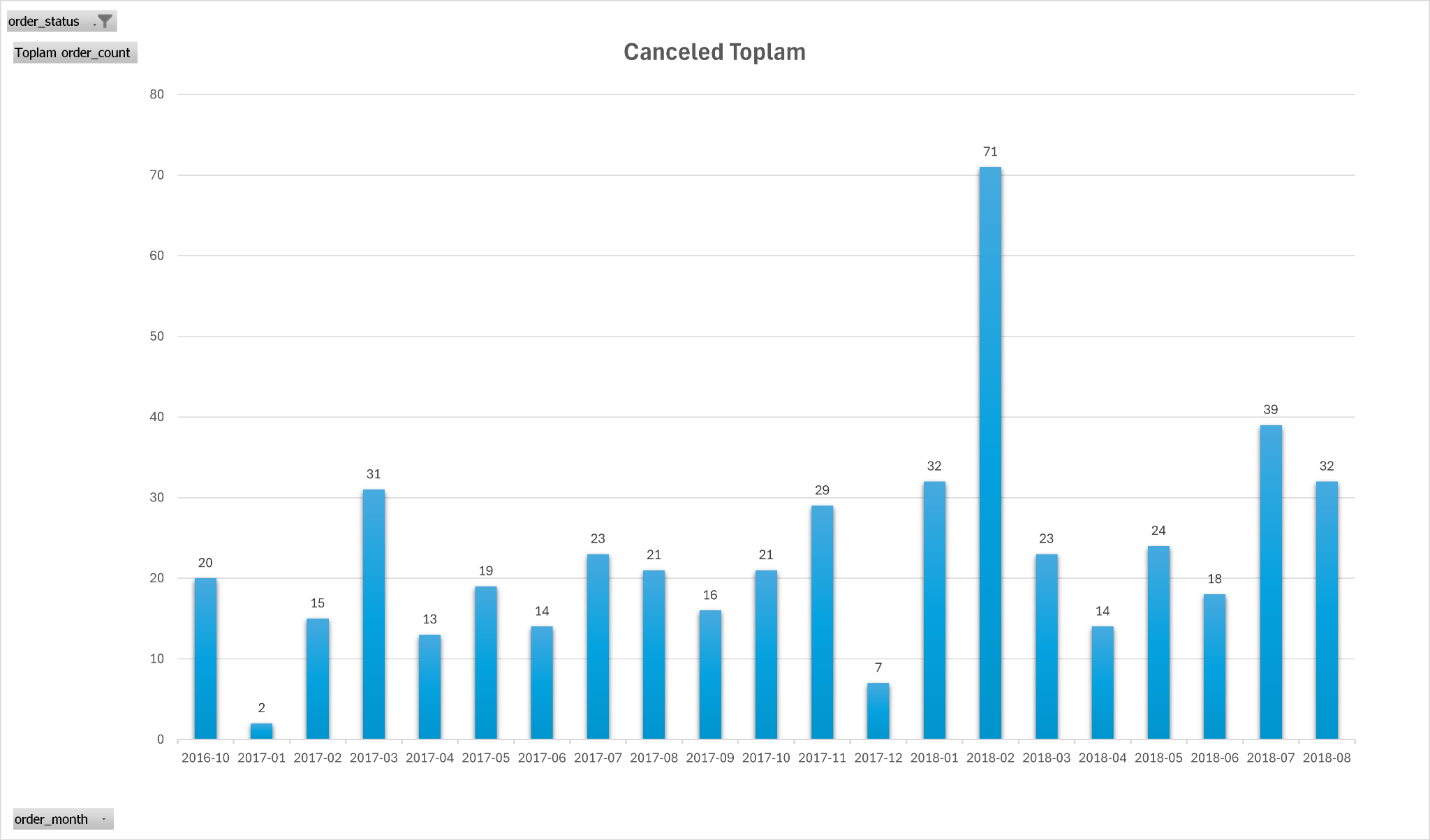
count(order\_id) AS order\_count

FROM olist\_orders

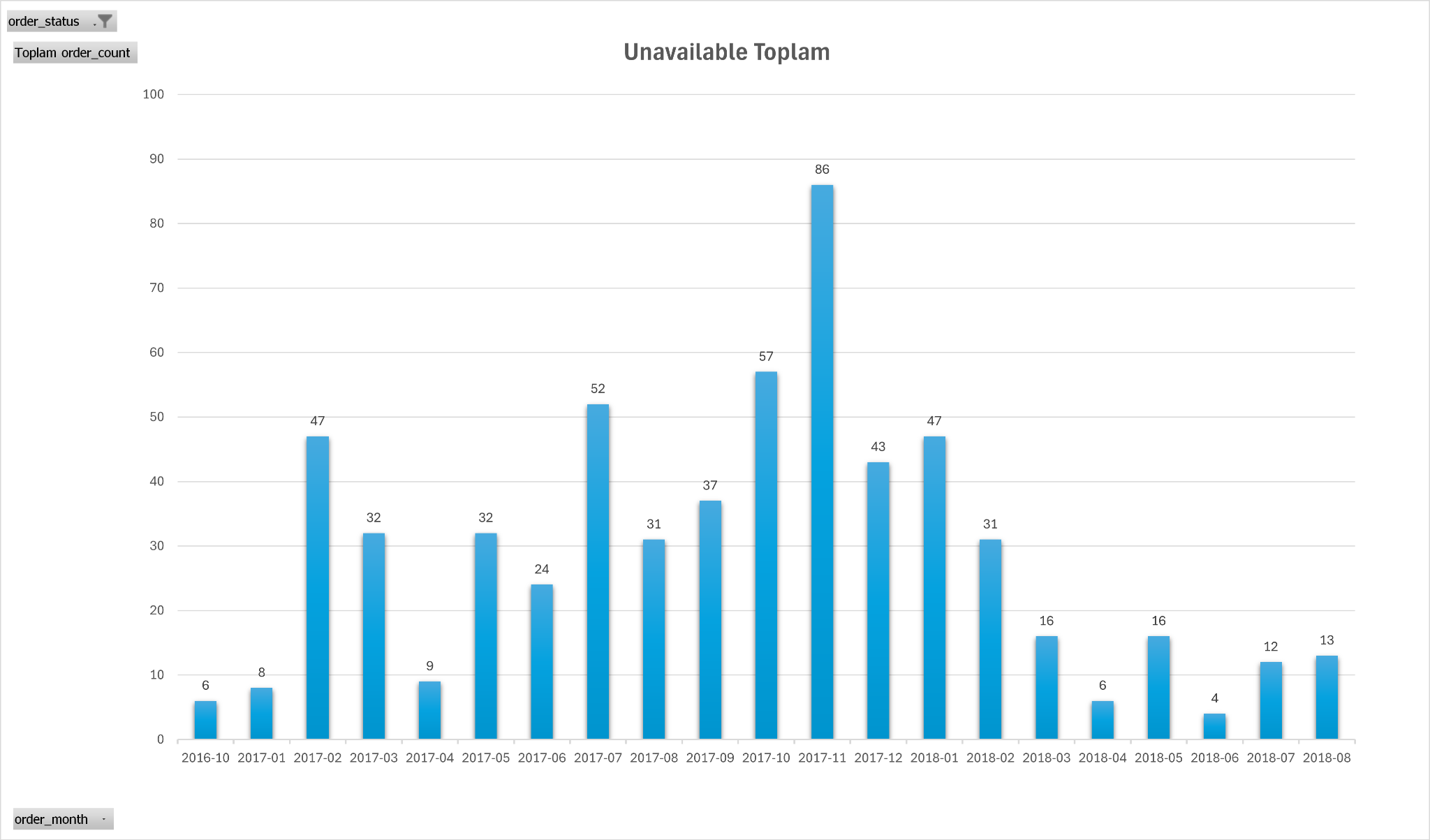
GROUP BY order\_month,order\_status;

Exel Chart:****

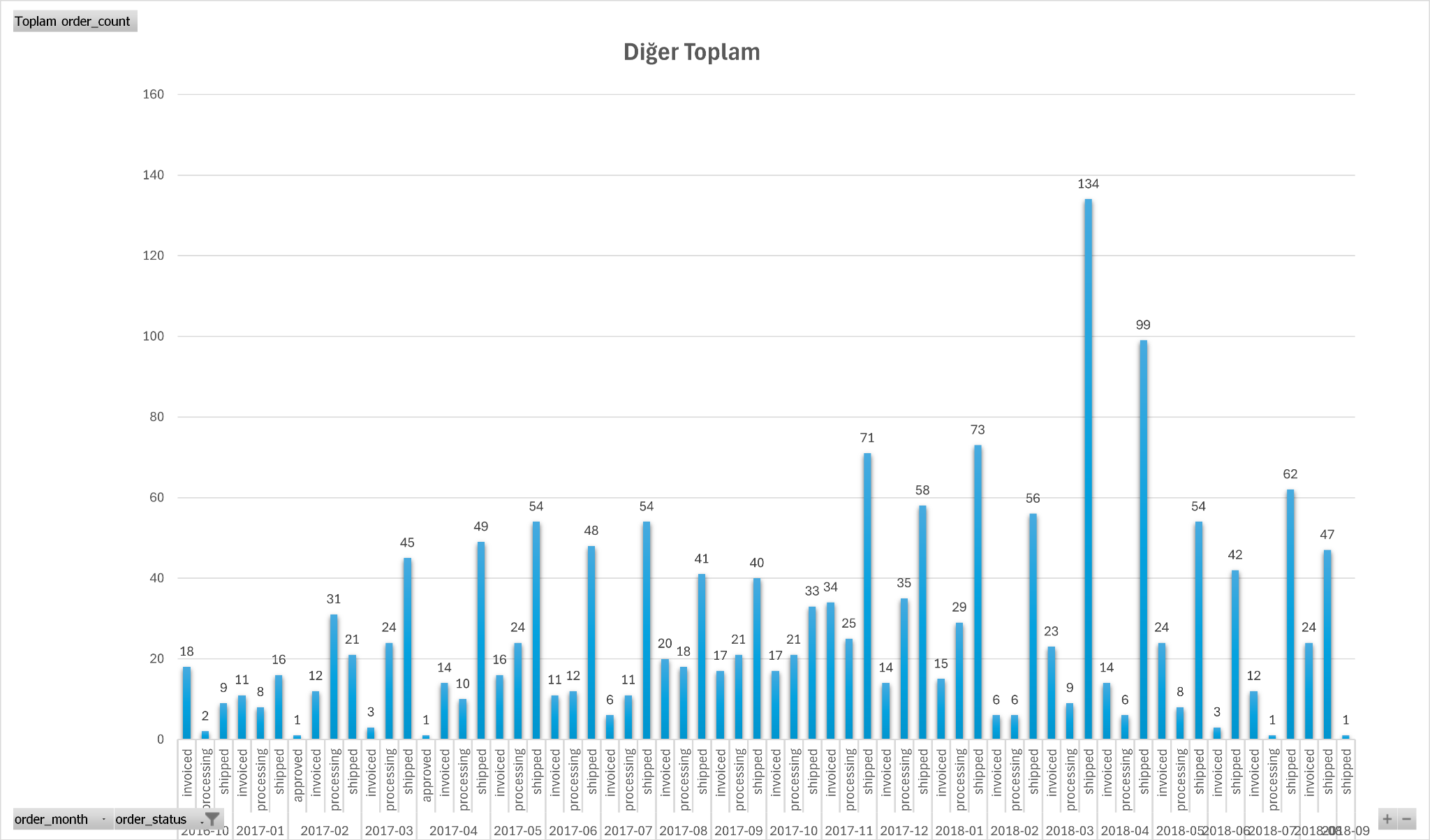
2016 Eylül, Ekim ve Kasım sonrasında 2017 Kasıma kadar genellikle bir artış göstermektedir. Ancak 2017 Aralık ayında yaklaşık olarak %20’lik bir düşüş yaşanmış bunun herhangi bir sebebinin olup olmadığı araştırılabilir ancak 2017 Kasım ayından aralık ayına %61 civarında yaşanan artış da bu azalmayı aşırı fazlaymış gibi gösteriyor olabilir. Aralık ayından sonra teslim sayısı ortalama 6000~7000 arasında stabil olarak kalmıştır.

****

2016 Ekim ayında teslimat sayısına oranla çok fazla iptal edilen ürün varken bu oran ilerleyen tarihlerde ürün teslimatı artmasına rağmen bu sayı 15~33 civarında kalmıştır. Sadece 2018 Şubat ayında anormal miktarda iptal edilen ürün olmuştur. Burada gerekli incelemeler yapılarak durumun olist tarafında mı yada müşteri veya satıcı tarafından teknik bir aksaklığın olup olmadığı tespit edilip tekrardan bir müşteri kaybı yaşanmaması için gerekli aksiyonlar alınabilir.

****

Unavailable grafiğine baktığımızda ise herhangi düzenli bir artış veya azalış yok. Artış yaşanan durumlarda genellikle bir önceki aya göre delivered orderlarda bir artış yaşanmış oluyor. Sadece 2017 Kasım ayında çok yüksek bir miktarda stoksuz kalındığı bir durum yaşanmış ancak teslim edilmiş toplam siparişlere de baktığımızda bir önceki aya göre %61 civarında yaşanan satış artışının da bu duruma etken olmuş olabilir.

****

Diğer chartlara bakmamızın bize herhangi bir katkısı olmuyor çünkü bu chartlar bize sipariş edilen ancak olumlu veya olumsuz herhangi bir teslimat bilgisi ile alakalı geri dönüş yapmayan bilgiler içeriyor.

##### **Question 3 :**

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

**SQL Query:**

SELECT op.product\_category\_name,

count(o.order\_id) AS order\_count

FROM olist\_orders AS o

LEFT JOIN olist\_order\_items AS oi

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

LEFT JOIN olist\_products AS op

ON oi.product\_id=op.product\_id

GROUP BY product\_category\_name

ORDER BY 2 DESC;

**SQL Query 2:**

SELECT to\_char(order\_approved\_at,'YYYY-MM') AS order\_month,

op.product\_category\_name,

count(o.order\_id) AS order\_count

FROM olist\_orders AS o

LEFT JOIN olist\_order\_items AS oi

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

LEFT JOIN olist\_products AS op

ON oi.product\_id=op.product\_id

GROUP BY order\_month, product\_category\_name

ORDER BY 3 DESC;

Elimizdeki veriye baktığımızda en çok toplam order’ın olduğu 8 kategori ve order sayıları şu şekilde: cama\_mesa\_banho 11115 adet order, beleza\_saude 9670 adet order, esporte\_lazer 8641 adet order, moveis\_decoracao 8334 adet order, informatica\_acessorios 7827 adet order, utilidades\_domesticas 6964 adet order, relogios\_presentes 5991 adet order, telefonia 4545 adet order.

Verileri incelediğimizde tek 1 ayda satan en çok ürün olan bilgisayar aksesuarları (informatica\_acessorios) Şubat 2018’de 991 adet satarak bu satışını gerçekleştirmiştir. Bu kategorinin diğer aylardaki satışına bakacak olursak bu miktarın çok altında kalıyor. Bu sebeple şubat ayında uygulanan bir indirim veya sevgililer gününün bu ayda oluyor olması bu duruma etken olmuş olabilir. Bunun haricinde tek 1 ayda en çok satış yapan 2. kategori olan yatak masa banyo (cama\_mesa\_banho) kategorisi Kasım 2017’de 956 adet order yapmıştır. Aynı şekilde dekorasyon mobilyası (moveis\_decoracao) kategorisi ve bahçe aletleri (ferramentas\_jardim) kategorisi de kendi en yüksek orderını Kasım 2017’de 776 ve 548 adet order ile yapmıştır. Bu durum göz önünde bulundurulduğunda, 2017 Kasım ayında olist sitesinde bir indirim uygulanmış olabilir.

##### **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.

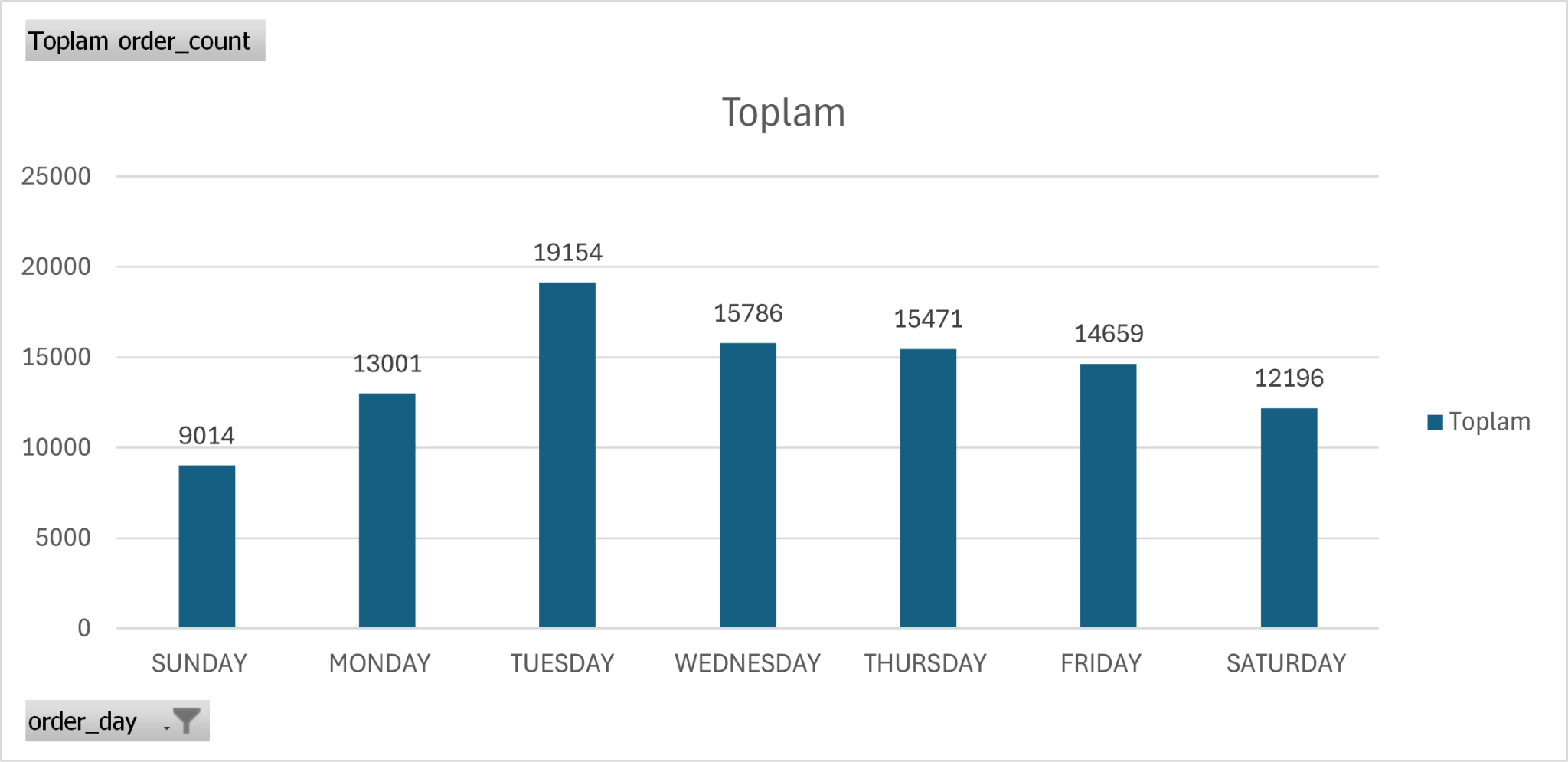
**SQL Query:**

SELECT to\_char(order\_approved\_at,'DAY') as order\_day,

count(order\_id) as order\_count

FROM olist\_orders

GROUP BY order\_day

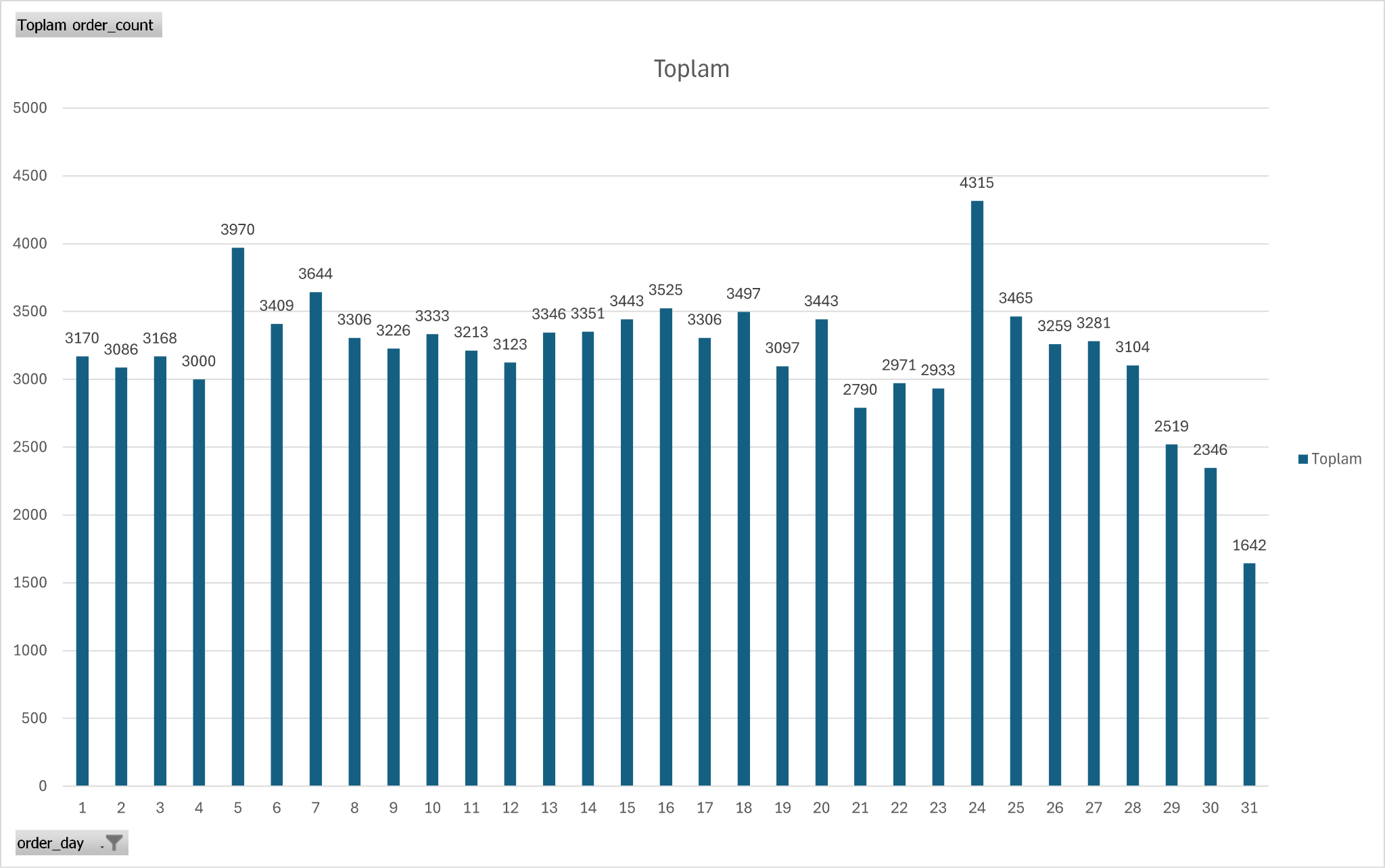


**SQL Query 2:**

SELECT extract (day from order\_approved\_at) as order\_day,

count(order\_id) as order\_count

FROM olist\_orders

GROUP BY order\_day

Yukarıdaki verileri incelediğimizde müşteriler en çok salı günleri, en az pazar günleri alışveriş yapıyorlar. Haftaiçi - haftasonu karşılaştırması yapacak olursak müşteriler çoğunlukla hafta içi alışveriş yapmaya meyilliler. Hafta sonu yapılan alışveriş hafta içine kıyasla daha düşük oluyor.

Günlere bakacak olursak ayın 24. gününde en çok alışveriş yapılmış en az ise 31. gününde yapılmış. Sipariş ayın 5. gününden sonra çok daha yüksek bu sebeple brezilyada maaş günü ayın ilk haftasında olabilir. 24. günde yaşanan bu en çok satışın sebebi noel tatilinin ve kutsal cuma günlerinin ayın 25. gününde olmasından kaynaklı oluyor olabilir.

#### **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.

**SQL Querry:**

SELECT cc.customer\_city,

count(o.order\_id) AS order\_count

FROM olist\_orders AS o

LEFT JOIN( SELECT oc.customer\_id,

oc.customer\_city

FROM( SELECT o.customer\_id,

c.customer\_city,

count(o.order\_id) AS order\_count

FROM olist\_orders AS o

LEFT JOIN olist\_customers AS c ON c.customer\_id = o.customer\_id

GROUP BY o.customer\_id, c.customer\_city

) AS oc

JOIN( SELECT o.customer\_id,

ROW\_NUMBER() OVER (

PARTITION BY o.customer\_id ORDER BY count(o.order\_id) DESC

) AS rn

FROM olist\_orders AS o

LEFT JOIN olist\_customers AS c ON c.customer\_id = o.customer\_id

GROUP BY o.customer\_id

) AS rn ON oc.customer\_id = rn.customer\_id AND oc.order\_count = rn.rn

) AS cc ON o.customer\_id = cc.customer\_id

GROUP BY cc.customer\_city

ORDER BY order\_count DESC;

Elimize geçen veri sonucunda en çok sipariş veren 5 şehir aşağıdaki gibidir:

1. sao paulo
2. rio de janeiro
3. belo horizonte
4. brasilia
5. curitiba

bu şehirlerde yaşayan kişiler daha çok sipariş veriyor. Verilerde başta gelen ilk 39 şehirden verilen siparişlerin toplamı geriye kalan 4080 şehirden verilen siparişten daha fazla.

#### **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.

**SQL Query:**

SELECT top5.seller\_id,

count(DISTINCT oi.order\_id) AS order\_count,

AVG(r.review\_score) AS avg\_review\_score,

count(DISTINCT r.review\_comment\_message) AS total\_review\_message

FROM (

SELECT seller\_id

FROM (

SELECT oi.seller\_id,

EXTRACT(DAY FROM order\_delivered\_customer\_date - order\_purchase\_timestamp) AS order\_delivery\_time

FROM olist\_orders AS o

LEFT JOIN olist\_order\_items AS oi ON o.order\_id = oi.order\_id

) AS order\_delivery\_date

GROUP BY seller\_id

ORDER BY AVG(order\_delivery\_time)

LIMIT 5

) AS top5

LEFT JOIN olist\_order\_items AS oi ON oi.seller\_id = top5.seller\_id

LEFT JOIN olist\_order\_reviews AS r ON r.order\_id = oi.order\_id

GROUP BY top5.seller\_id;

Eline ürünü en hızlı ulaşan 5 kişiden 4’ü 5 puan vermiş, 1 kişi ise 1 puan vermiştir ve 5 kişi de herhangi bir review mesajı yayınlamamış ve eline en hızlı ürünü ulaşan 5 kişide sadece 1 kez order vermiştir. 1 puan veren kişinin siparişi ile ilgili inceleme yapılıp sorunun ne olduğu çözülmeye çalışılabilir.

##### **Question 2 :**

-Hangi satıcılar daha fazla kategoriye ait ürün satışı yapmaktadır?

Fazla kategoriye sahip satıcıların order sayıları da fazla mı?

**SQL Query:**

SELECT oi.seller\_id,

count(DISTINCT p.product\_category\_name) AS category\_count,

count(DISTINCT oi.order\_id) AS order\_count

FROM olist\_order\_items AS oi

LEFT JOIN olist\_products AS p ON oi.product\_id=p.product\_id

GROUP BY 1

ORDER BY 2 DESC

Elimizdeki veriye baktığımızda ve order sıralamasına order\_count’a göre baktığımızda ilk 20 sırada olan satıcılardan hepsinin birden fazla kategoride satış yaptığını görüyoruz ancak en çok kategoride satış yapan kişilere baktığımızda en çok satış yapan kişilerle ilk 20 bazında baktığımızda sadece 2 kişinin en çok satış yapanlar kategorisinde de listede olduğunu görüyoruz. Category\_count’a göre baktığımızda ilk 20 de olan kişilerin en azının 59 orderı olduğunu görüyoruz. Yani kısacası çok fazla kategoride satış yapıyor olmak order sayılarında en üst sıralarda olmak için çok avantaj sağlamıyor gibi gözüküyor.

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

##### **Question 1 :**

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

**SQL Query:**

SELECT payment\_installments,

c.customer\_city,

count(DISTINCT o.customer\_id) AS customer\_count

FROM olist\_order\_payments AS p

LEFT JOIN olist\_orders AS o ON p.order\_id=o.order\_id

LEFT JOIN olist\_customers AS c ON c.customer\_id=o.customer\_id

GROUP BY 1, 2

ORDER BY 1 DESC, 3 DESC

Elimizdeki veriye baktığımızda müşterilerimiz en fazla 24 taksit yapmıştır. 24 taksit ile alışveriş yapan her biri farklı şehirlerde olacak şekilde 18 kişi bulunmaktadır. 10 taksite kadar elimizdeki veriye göre her taksit miktarında farklı şehirlerde 1~15 arasında kullanıcı bulunmaktadır. 10 taksitte ise bu sayı yükselerek farklı şehirlerde 600’e yakın kullanıcının bu ödeme şeklini tercih ettiği görülebiliyor. "sao paulo", "rio de janeiro", "belo horizonte" adlı şehirler yüksek taksitlerlde en çok müşteri bulunan şehirler. Buralarda vade farkında oynamalar yapılarak elde edilen kar arttırılmaya çalışılabilir yada daha az vade farkı uygulanılarak müşteri sayısı arttırılmaya çalışılabilir.

##### **Question 2 :**

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

**SQL Query:**

SELECT p.payment\_type,

SUM (CASE WHEN o.order\_status NOT IN('cancelled','unavailable') THEN p.payment\_value END) AS succ\_payment\_value,

count(DISTINCT CASE WHEN o.order\_status NOT IN ('cancelled', 'unavailable') THEN o.order\_id END) AS order\_count

FROM olist\_order\_payments AS p

LEFT JOIN olist\_orders AS o ON p.order\_id=o.order\_id

GROUP BY 1

Elimizdeki veriye göre orderlarda en çok kullanılan yöntemler sırasıyla "credit\_card", "boleto", "voucher", "debit\_card".

##### **Question 3 :**

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

**SQL Query 1:**

SELECT p.payment\_installments,

product\_category\_name,

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

FROM olist\_orders AS o

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

LEFT JOIN olist\_order\_items AS oi ON oi.order\_id= o.order\_id

LEFT JOIN olist\_products AS pro ON pro.product\_id= oi.product\_id

WHERE p.payment\_installments=1

GROUP BY 1,2

ORDER BY 3 DESC, 1 DESC

**SQL Query 2:**

SELECT p.payment\_installments,

product\_category\_name,

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

FROM olist\_orders AS o

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

LEFT JOIN olist\_order\_items AS oi ON oi.order\_id= o.order\_id

LEFT JOIN olist\_products AS pro ON pro.product\_id= oi.product\_id

WHERE p.payment\_installments>1

GROUP BY 1,2

ORDER BY 3 DESC, 1 DESC

Elimizdeki verilere baktığımızda tek çekim olarak en çok order yapılan kategori 4299 order ile spor eğlence (esporte\_lazer) kategorisi. Bu kategoriyi daha sonrasında 4177 ve 3880 order ile bilgisayar aksesuarları (informatica\_acessorios) ve güzellik (beleza\_saude) kategorisi takip etmektedir. En çok taksitle ödeme seçeneği kullanılan kategori 5969 taksitli order ile yatak masa banyo (cama\_mesa\_banho) kategorisidir. Bu kategoriyi daha sonrasında 5012 taksitli order ile güzellik (beleza\_saude) ve 3482 taksitli order ile spor eğlence (esporte\_lazer) kategorileri takip ediyor. Bu 3 kategoride en çok order yapılan 3 kategoriden biri. Yani order sayısı çok olan kategorinin aynı zamanda en çok taksit ile ödeme yapılan kategoriler olduğu söylenebilir.

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

**SQL Query:**

WITH rfm\_metrics AS (

SELECT

customer\_id,

MAX(invoicedate) AS latest\_purchase\_date,

COUNT(DISTINCT invoiceno) AS frequency,

SUM(quantity \* unitprice) AS monetary\_value

FROM rfm

GROUP BY customer\_id

),

rfm\_scores AS (

SELECT

customer\_id,

EXTRACT(DAY FROM '2011-12-09' - latest\_purchase\_date) AS recency,

frequency,

monetary\_value,

NTILE(5) OVER (ORDER BY EXTRACT(DAY FROM '2011-12-09' - latest\_purchase\_date)) AS r\_score,

NTILE(5) OVER (ORDER BY frequency ) AS f\_score,

NTILE(5) OVER (ORDER BY monetary\_value ) AS m\_score

FROM rfm\_metrics

),

merge\_mon\_fre\_score AS (

SELECT

customer\_id,

r\_score,

f\_score + m\_score AS fremon\_score

FROM rfm\_scores

),

final\_rfm\_score AS(

SELECT

customer\_id,

r\_score,

NTILE(5) OVER (ORDER BY fremon\_score ) AS fm\_score

FROM merge\_mon\_fre\_score

)

SELECT customer\_id,

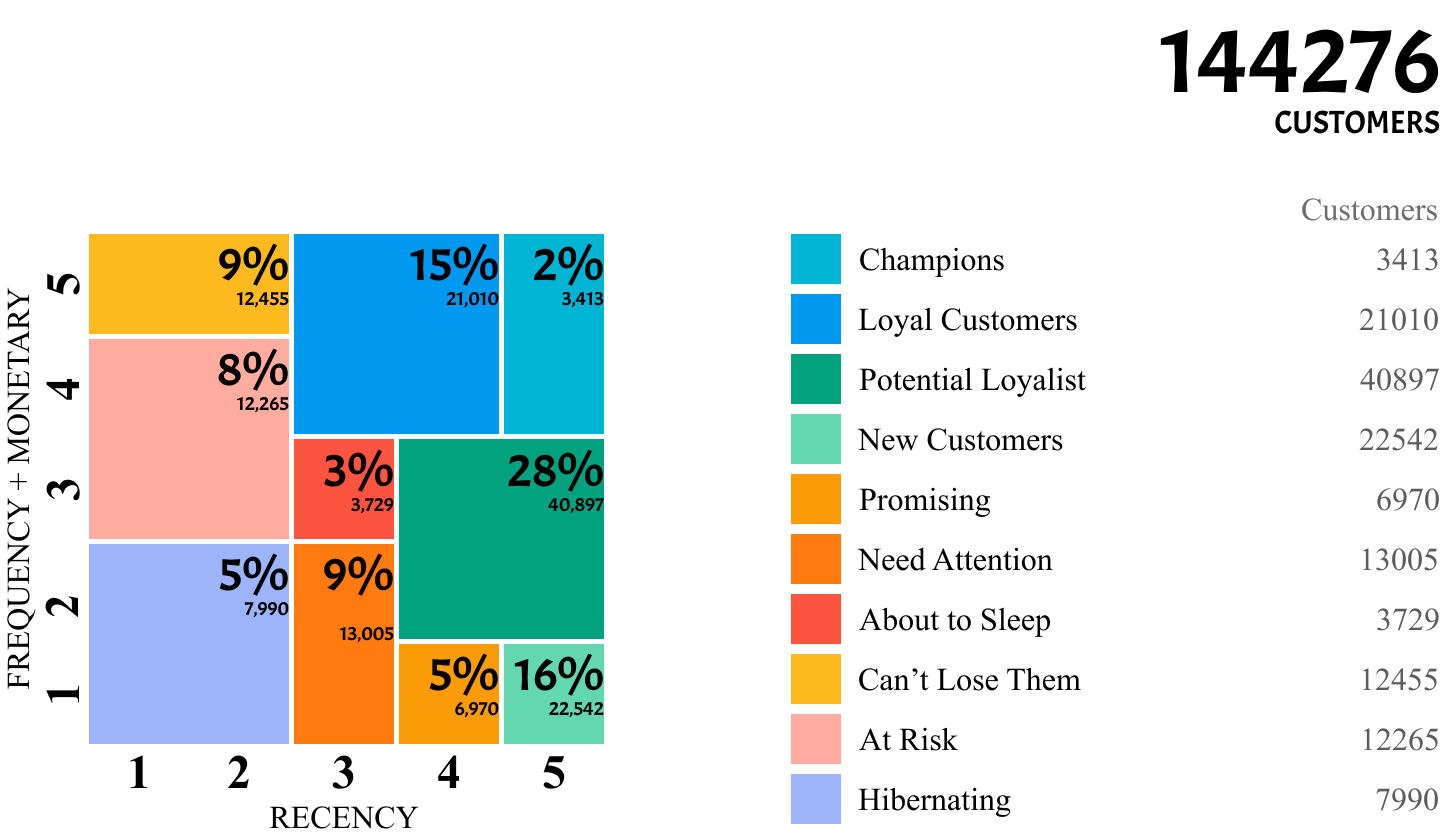
r\_score,

fm\_score,

CONCAT(r\_score,fm\_score) AS rfm\_segment

FROM final\_rfm\_score

**RFM Heatmap:**



Heatmap’i incelediğimizde kullanıcılarımızın büyük bir kısmının potansiyel sadık müşteri olduğunu görüyoruz. Bu kısımda bulunan kişilere ekstra indirimler tanımlanarak üstünde bulunan diğer 2 kategoriye geçiş yapmaları sağlanabilir. Yine aynı şekilde yeni müşterilere de gerekli uygulamalar yapılarak uygulamayı şirket için daha yüksek verimli kullanmaları sağlanarak potansiyel sadık veya sadık kategorilerine çekilmeye çalışılabilir. Geriye kalan yeni müşteri kategorisi altında kalan %39’luk kısım için gerekli analizler yapılarak müşteriler geri uygulamaya çekilmeye çalışılmalı.