**Ctk**

* **Downline\_Orders\_\_c** (%95.6) ve **Order\_Campaign\_\_c** (%91.8) sütunlarında çok yüksek oranda eksik veri var.İçerik olarak anlamlandırılamıyor. Bu yüzden bu sütunları silerim.
* Downline\_orders\_c kolonunun herhangi bir kolon ile ilişkisini göremedim ve verinin 12123 adet %75-80 kadarlık bir kısmı boş. Bu yüzden bu kolonu komple silebiliriz.
* Order\_campaign\_c kolonu için 11635 adetlik büyük bi boş veri var bu kolonunda diğer veriler ile ilişkisini fazla kuramadım ve bu kolonu da veriden silebiliriz.

KARAR:Silmek

* **Facility\_Type\_\_c (%1),** Sadece %1 lik bir eksiklik olduğu için en yüksek değer olan obgyn clinic yada other seçeneği ile doldurulabilir.

Yada Facility\_Type\_\_c ve Shipping\_State\_\_c arasındaki ilişkiye bakılarak konum bazlı Mod ile de doldurulabilir.(o siparişin bulunduğu Shipping\_State\_\_c için en sık görünen tesis tipi (mod) ile doldurulabilir)

* Yöntem 2 ;Facility\_type\_c kolonu için ilk olarak her kullanıcı için kaç tip yerden alışveriş yaptığına bakarız.
  + Eğer kullanıcı sadece 1 yerden sipariş yapmış ise bu kullanıcın boş veri kısmına diğer facility\_type değerini yazabiliriz.
  + Tarihe ve ardından bölge olarak gruplama yapıp o gün içierisinde o bölgede en çok satılan facility\_type ı diğer boş veriler için kullanabiliriz.

KARAR:Yöntem 2 uygulanacak vakit kalmazsa others ile dolduracağız !!

* **Shipping\_City\_\_c** (%0.05) ve **Shipping\_Zip\_Code\_\_c** (%0.1) sütunlarındaki eksik verileri Customer\_Account\_\_c, Customer\_Name\_\_c sütunlarındaki değerlerle eşleştirerek doldurulabilir(Zipcode için 5 kayıt bulundu.City için de 1 kayıt)

Shipping\_city\_c kolonu için boş olan kısımları zip\_code kolonundan bularak doldurabiliriz. Örneğini yaptım.

KARAR: Shipping\_city\_c kolonu için boş olan kısımları zip\_code un en çok tekrar eden değeri ile dolduracağız

Shipping\_zip için de aynı şekilde.

CTK TABLOSUNUN MAVİ KISMINI AYRI BİR TABLO OLARAK DEĞERLENDİRECEĞİZ!!!!!

**Oppurtunity**

* **Contactid** (%5.8) contactid kolonu doldurulamaz account name oppid ile bağlantılarına baktım ama eşleşmedi.Satırları silmek yada bilinmeyen olarak doldurmak en doğru yaklaşım gibi duruyor
* contact\_id kolonunda 931 boş veri var ve bunu doldurmamız sakıncalı olur bu yüzden doldurmayalım

KARAR:Silinecek

* **Account.name** (%0.025 - çok az) silmek daha mantıklı duruyor.
* account.name kolonunda 4 adet boş veri var ve bunu da doldurmamız anlamsız olabilir.

KARAR:Silinecek

* **Amount** (%2.4)
* Yüksek Çarpıklık (Skewness: 64.26): Bu değerin 1'den çok çok yüksek olması, amount dağılımının aşırı derecede sağa çarpık olduğunu matematiksel olarak doğruluyor. Yani, verilerin büyük çoğunluğu düşük tutarlarda yığılmışken, çok yüksek değerli birkaç aykırı değer ortalamayı aşırı derecede yukarı çekiyor. Bu durum, eksik değerleri doldururken medyan kullanmanın ortalamadan çok daha doğru ve güvenilir olacağını kesinleştiriyor.
* Amount kolonunda 382 adet boş veri var bu verileri Platform -> recordType ve belki stagename kolununu da dahil edip ona göre gruplayıp ortalama, medyan veya mod değerlerinden herhangi biri ile boş verileri doldurabiliriz

KARAR: Verileri Platform -> recordType a göre gruplayıpayrı ayrı medyanına göre doldurulacak (Anlamlı ilişki varmı kontrol edelim) yoksa genel medyanla doldurulacak

* **Platform\_\_c (**%13.4) Koşullu mod ile doldurma Platform\_c campaign name ilişkisi üzerinden bir eşleştirme listesi yapılır.(bazı kampanyaların neredeyse her zaman (%90-95 gibi yüksek oranlarda) belirli bir platform üzerinden geliyor )

örneğin

* '2023 EKC' kampanyası neredeyse hep 'PledgeIt' platformundan geliyordu.
* 'Lydie's Loop 2022' kampanyası neredeyse hep 'Qgiv' platformundan geliyordu.
* '2022 Giving Tuesday...' kampanyası neredeyse hep 'Facebook' platformundan geliyordu.

**Doldurma Döngüsü:** platform\_\_c'nin eksik olduğu her satır için:

1. O satırdaki campaign.name (kampanya adı) değerine bakar.
2. Bu kampanya adının bizim oluşturduğumuz "eşleştirme listesinde" (haritada) olup olmadığını kontrol eder.
3. **Eğer** kampanya adı listede **varsa**, liste ona karşılık gelen platformu (örn: 'PledgeIt' veya 'Qgiv') söyler ve kod eksik platformu bu bilgiyle doldurur.
4. **Eğer** kampanya adı listede **yoksa** (çünkü o kampanya ile belirli bir platform arasında o kadar güçlü bir bağ bulamadık), kod bu listeyi kullanmaz ve diğer kurallara geçer (kampanya adı eksik mi diye bakar veya genel modu kullanır).  
   Platform\_c kolonundaki boşlukları bağışın türü olan recordType.name alanına göre doldurabiliriz.
   * Mesela en fazla donation facebook platformundan atılmışsa boş alanı facebook ile doldurabiliriz.

KARAR:KOŞULLU MOD KULLANILACAK

* **Donation\_date\_\_c** (%52.8 - çok yüksek) close date ve stage name ile arasında anlamlı bir ilişki bulunamadı en mantıklısı kolonu silmek
* Donationdate kolonunu silebiliriz. Çok manalı gelmedi bana ve veri setinin yarısı kadar boş veriye sahip

KARAR:Silmek

* **Campaign.name** (%21.9)
* Kampanya yok olarak doldurulmalı,kampanyasız gelen kişiler olabilir.Diğer kolonlarla anlamlı ilişkisi yok.
* Campaign name alanında ise RecordType\_Name,Platform\_c,Campaign\_Name, CloseDate, count( Campaign\_Name) countOfCampaignName sırasında gruplama yapıldıktan sonra platformlara göre en fazla kullanılan campaign name i boş verilere ekleyebiliriz.
* KARAR: Kampanya yok olarak doldurulmalı,kampanyasız gelen kişiler olabilir.Diğer kolonlarla anlamlı ilişkisi yok.