



Hafta 3 (CRM-RFM-CLTV-Buy Till You Die-NG/NBD-Gamma Gamma)

[@mebaysan](#)

12/09/2021

Bu haftanın veri seti: Dataset

İlgili Okuma Listesi

- A Complete Data Analytics Project with Python
- RFM Analysis with Python
- Pandas DataFrame: rank(.) function
- Python | Pandas Dataframe.rank().
- Pandas Rank – Rank Your Data – pd.df.rank().
- All You Need to Know About Pandas Cut and Qcut Functions
- Predicting Customer Lifetime Value with “Buy ‘Til You Die” probabilistic models in Python
- How to calculate CLV using BG/NBD and Gamma-Gamma

Benim Yazdığım Yazılar:

- Rule-Based Customer Segmentation Without ML
- What is RFM Analysis? An Applied Example in Python
- Customer Life Time Value Prediction by Using BG-NBD & Gamma-Gamma Models and Applied Example in Python

CRM Analitiği

Customer Relationship Management → Müşteri İlişkileri Yönetimi

CRM Çalışma Alanları

- Müşteri yaşam döngüsü optimizasyonları yapılır

- Customer lifecycle
- Customer journey
- Customer funnel
- İletişim (dil, renk, görseller, kampanyalar)
- Müşteri edinme/bulma çalışmaları
- Müşteri elde tutma (terk) çalışmaları
- Müşteri segmentasyonları
- Çapraz satış (cross-sell), üst-satış(up-sell)
 - Up-Sell → bir üst modeli satmak
 - Cross-Sell → yan ürün satmak



KPI

Key Performance Indicators → Temel Performans Göstergeleri

Şirket, departman ya da çalışanların performanslarını değerlendirmek için kullanılan matematiksel göstergelerdir.

KPI Örnekleri

- Customer Acquisition Rate (Müşteri Kazanma Oranı)
- Customer Retention Rate (Müşteri Elde Tutma Oranı)
- Customer Churn Rate (Müşteri Terk Oranı)
- Conversion Rate (Dönüşüm Oranı)
- Growth Rate (Büyüme Oranı)

Cohort Analizi

*Bunların yeri Python değil, Data Studio, Tableau vb. araçlar olmalıdır
(- Vahit Keskin)

Cohort: Ortak özelliklere sahip bir grup insan

Ortak özelliklere sahip gruplara yapılan analiz → Cohort Analizi

RFM Nedir?

RFM metriklerini kullanarak, müşterileri segmentlere ayırma işlemidir. Metrikler standartlaştırılarak **RFM skorları** elde edilir. Skorlara göre müşteriler segmentlere ayrılır ve aksiyon alınır.

Satış verileri için süper çalışır.

RFM analizi farklı zaman dilimlerinde çalıştırılarak elde edilen çıktılar karşılaştırılabilir. Bu karşılaştırmalar neticesinde:

- Alınan aksiyon sonuçları
- Yapılan harcamalar
- Maliyetler

Ölçülebilir ve değerlendirilebilir.

RFM Metrikleri

Recency (Yenilik)

En son alış-veriş yapma uzaklığı (kaç gün önce?)

En son kaç gün önce satın alım yaptı

Frequency (Sıklık)

Toplamda kaç kere alış-veriş yapmış

Toplam kaç satın alım yaptı

Monetary (Parasal Değer)

Toplam alış-verişlerinde ne kadar para bırakmış

Satın alımları neticesinde ne kadar para bıraktı

Basit, etkili bir yöntemdir.

Satın alma senaryolarında, hızlı satış kanallarına iyi giden bir yöntem.

Çeşitli zaman periyotlarında çalıştırılıp segment geçişlerinin takip edilmesi faydalı olacaktır.

RFM analizi frekansın olduğu satış verilerinin bulunduğu senaryolar için iyi çalışır.

RFM Uygulama Adımları

- Metrikler 1-5 arasında standartlaştırılır
- **Standartlaştırılan metrikler neticesinde skorlar elde edilir**
- Skorlar birleştirilir ÖR: (352)
 - 1. hane R (3)
 - 2. hane F (5)
 - 3. hane M (2)
- Skorlara göre segmentler oluşturulur
- Segmentlere göre aksiyon alınır

	R	F	M	RFM
Müşteri 1	1	4	5	145
Müşteri 2	4	5	4	454
Müşteri 3	5	1	3	513
.	.	.	.	
.	.	.	.	
.	.	.	.	
Müşteri 1000	2	4	4	244

Klasik RFM Tablosu



Skorlar Birleştirildikten Segmentlerin Oluşturulmasına Örnek

	R	F	M	RFM	Segmentler
<u>Müşteri 1</u>	1	4	5	145	→ At Risk
<u>Müşteri 2</u>	4	5	4	454	→ Loyal Customers
<u>Müşteri 3</u>	5	1	3	513	→ New Customers
.
.
.
Müşteri 1000	2	4	4	244	→ At risk

Skor Açıklamaları

- **R** değeri arttıkça müşterinin daha yeni satın alım yaptığını düşünebiliriz
- **F** değeri arttıkça sık satın alım yaptığını düşünebiliriz
- **M** değeri arttıkça müşterinin bize daha fazla kazandırdığını düşünebiliriz
- Unutmayalım ki bu skorlar hesaplanırken metrikler, bir müşteri başına tüm müşterilerin aksiyonlarına göre hesaplanıyor.
 - Yani M değerinin 5 çıkması, bu müşterinin diğer müşterilere göre daha fazla para bıraktığını düşünmemizi sağlar.

Python Pandas Rank Metodu İle İlgili Trickler

```
rfm.loc[:, 'frequency_score'] = pd.qcut(rfm["frequency"].rank(method="first"), 5, labels=[1, 2, 3, 4, 5])
```

Rank hesaplama: `rank` metodu aksi belirtilmediği sürece verileri küçükten büyüğe (ascending) sıralar. Bu örneğimizde `size` değişkenine göre rank atadık. Atama metodumuz `first` olduğu için, `size` değişkeni içerisindeki her bir unique değer için ilk görüldüğü (`metod=first`) index numarasını rank olarak atadı. Metodumuz `average` olsaydı, unique değerlerin görüldüğü index numaralarını alır ve bu index numaralarını toplayıp ortalamasını rank olarak atardı.

Yani, min ile 20 çeyreklik arasını 1. segmente atayacak (çeyreklik değeri 1), 21 ve 40 çeyreklik arasını 2'ye atayacak. Fakat biz `method=first` demeseydik 21 çeyreklikteki 1 değerini de 1 sınıfına atayacaktı (çünkü min-20 çeyreklik ile 21 çeyrekliği aynı segmente atayacaktı, çeyreklik değerleri ikisinde de 1 olduğu için). Bunun önüne geçmek için `method=first` kullandık.

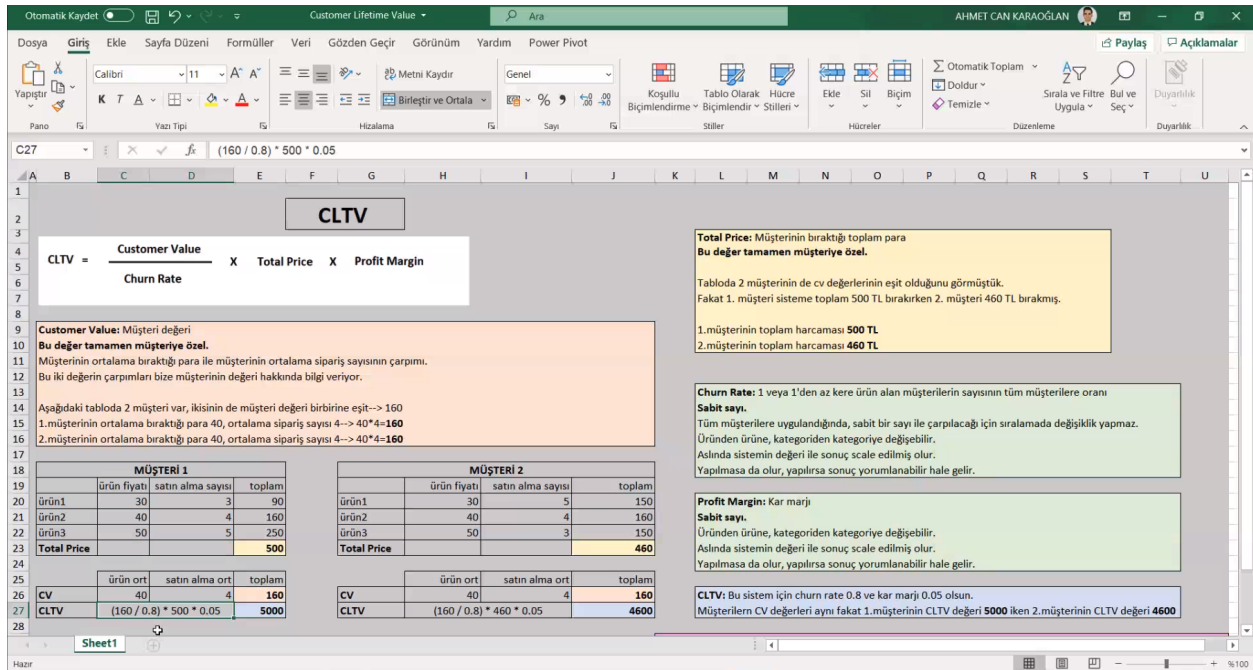
Frequency Çeyreklik Değerleri		Segmentasyon
min	1.00000	1
20%	1.00000	
21%	1.00000	2
40%	2.00000	
41%	2.00000	3
60%	3.00000	
61%	3.00000	4
80%	6.00000	
81%	6.00000	5
100%	209.00000	

```
df = sns.load_dataset('tips')

df['rank'] = df['size'].rank(method='first')

df.sort_values('size')['size']
```

Customer Life Time Value



Otomatik Kaydet

Arayış

AHMET CAN KARAOĞAN

DosyaGirişEkleSayfa DüzeniFormüllerVeriGözden GeçirGörünümYardımPower Pivot

Yapıştır

Kalibri11A⁺

KT

Yazı Tipi

Hızalama

Sayın

Stiller

Hücreler

Düzenleme

Duyarlılık

Genel

Kopullu Biçimlendirme

Tablo Olarak Biçimlendir

Hücre Stilleri

Ekle

Sil

Biçim

Otomatik Toplam

Doldur

Temizle

Sırala ve Filtre Uygula

Bul ve Seç

W28

$\text{satin alma sayısı} * \text{satin alma başına ortalama kazanç}$

$\text{Customer Value} = \text{Purchase Frequency} * \text{Average Order Value}$

CLTV Prediction yapmak istediğimiz zaman ise aşağıdaki formülü kullanırız.

$\text{CLTV} = (\text{Conditional}) \text{ Expected Number of Transaction} * (\text{Conditional}) \text{ Expected Average Profit}$

Yani olasılıksal bir yaklaşım getiriyoruz (expected). Formülü şu şekle sokabiliriz.

$\text{CLTV} = \text{BG/NBD Model} * \text{Gamma Gamma Model}$

- BG/NBD Model, tüm müşterilerin satın alma davranışlarını modelleyecek ve genel kitle patternini kişiye indirgeyerek beklenen satış sayısını tahmin edecek
- Gamma Gamma Model, bütün kitle üzerinden expected average profit dağılımı modelleyecek ve kişiye indirgeyerek beklenen ortalama kar değerini tahmin edecek

BG-NBD Model (Buy Till You Die)

Beta Geometric / Negative Binomial Distribution

Bize koşullu beklenen işlem sayısını verir (expected number of transactions)

Bu model, 1 hafta sonra, 3 ay sonra tahmini ne kadar satış olacak? vb sorulara cevap verir

Bu model, **Expected Number Of Transactions** için iki süreci olasılıksal olarak modeller:

- Transaction Process (Buy)
 - Satın alma işlemidir
 - Müşteri alive olduğu sürece, belirli bir zaman periyodunda, bir müşteri tarafından gerçekleştirilecek işlem sayısı transaction rate parametresi ile poisson dağılır
 - Bir müşteri alive olduğu sürece kendi transaction rate'i etrafında rastgele satın alma yapmaya devam edecektir
 - Transaction rate'ler her bir müşteriye göre değişir ve tüm kitle için **gamma dağılır (r, α)**
- Dropout Process (Till You Die)
 - Satın almayı bırakmak demektir

- Her bir müşterinin p olasılığı ile dropout rate (dropout probability)'i vardır
- Bir müşteri satın alım yaptıktan sonra belirli bir olasılıkla drop olur
- Dropout rate'ler her bir müşteriye göre değişir ve tüm kitle için **beta dağılır (a,b)**

BG-NBD Model Formülasyonu

$$E(Y(t)|X = x, t_x, T, r, \alpha, a, b) = \frac{a+b+x-1}{a-1} \times \frac{\left[1 - \left(\frac{\alpha+T}{\alpha+T+t} \right)^{r+x} {}_2F_1(r+x, b+x; a+b+x-1; \frac{t}{\alpha+T+t}) \right]}{1 + \delta_{(x>0)} \frac{a}{b+x-1} \left(\frac{\alpha+T}{\alpha+t_x} \right)^{r+x}}$$

- **E** beklenen değer ifadesini temsil etmektedir
- **I** koşullu olasılığı ifade ediyor (conditional expected number of transaction)
- **x** müşteri için frequency'i ifade ediyor. Yani, en az 2 kere satış yapmış kullanıcılar için tekrar eden işlem (satış) sayısı
- **tx** müşteri için recency'i ifade ediyor. Buradaki recency günlük olarak değil haftalık olarak kullanılır. Müşterinin ilk satın alma ile son satın alma arasındaki geçen süre (hafta)
 - RFM'den farklı olarak; burada müşterinin **son satın alım tarihi - ilk satın alım tarihi** hesaplanır. RFM'de ise **analiz tarihi - son satın alım tarihi** hesaplanıyordu.
- **T** müşterinin ilk satın alımı üzerinden geçen zamanı ifade eder. Müşterinin yaşı diyebiliriz. Analiz tarihin - ilk satın alım tarihi. Haftalık cinsten.
- **r, α** gamma dağılımından gelen parametreler (buy process). **Tüm kitlenin transaction rate dağılımı.**

- a, b beta dağılımından gelen parametreler (till you die process). **Tüm kitlenin dropout rate dağılımı.**
- $\gamma(t)$ müşteri için beklenen işlem sayısını ifade eder. Expected number of transactions.

Gamma Gamma Model

Bir müşterinin işlem başına ortalama ne kadar kar getirebileceğini tahmin etmek için kullanılır.

Bütün kitle için average profiti modelledikten sonra bir müşteri özelinde expected average profiti verir.

- Bir müşterinin işlemlerinin parasal değeri (monetary) transaction value'larının ortalaması etrafında rastgele dağılır.
- Ortalama transaction value, zaman içinde kullanıcılar arasında değişebilir fakat tek bir kullanıcı için değişmez.
- Ortalama transaction value tüm müşteriler arasında gamma dağılır

Gamma Gamma Model Formülasyon

$$E(M|p, q, \gamma, m_x, x) = \frac{(\gamma + m_x x)p}{px + q - 1} = \left(\frac{q - 1}{px + q - 1}\right) \frac{\gamma p}{q - 1} + \left(\frac{px}{px + q - 1}\right) m_x$$

- E expected değeri ifade eder
- x frequency'i ifade eder
- m_x gözlemlenen transaction value değeri (monetary)
- M beklenen transaction value değeri (hesaplamaya çalıştığımız expected average profit)
- p, q, γ gamma dağılımdan gelen parametreler (bütün kitlenin average profiti modellendikten sonra)

Bu modelleri ayrı ayrı da kullanabiliriz. Sadece BG-NBD Modeli ile satış tahmini de gerçekleştirebiliriz.

Python'da Hesaplamak

`lifetimes` paketini kullanırız.

```
from lifetimes import BetaGeoFitter
from lifetimes import GammaGammaFitter
```

1-) Sadece belirli bir ürünü (x) filtreledikten sonra bg-nbd modelini kurarsak, ilgili zaman diliminde x ürün tipinden beklenen satış hesaplayabilir miyiz?

Evet bir ürün bazında o ürünü alan müşteriler için belirli bir zamanda örneğin 12 aylık tahmin yaptığımızda o üründen 12 ay içerisinde beklenen ortalama satış tahminleyebiliriz

2-) CLTV prediction sms satışları için çalışır diyebilir miyiz? Çünkü sms satışları belirli periyotlarla açılan, müşterilere belirli zaman dilimlerinde "1234'e X yazın gönderin" gibi atılan smsler olduğundan sürekli bir satış kanalı diyemeyiz gibi düşündüm. Bu tip satışlar için bu modeller çalışır mı?

Bu konuya şöyle yaklaştım şimdi benim bir şirketim var ve ben veri odaklı bir ciro artırmayı hedefliyorum. İlk önce SMS atmadan önce bir CLTV yaparak müşterilerimin bana ortalama ne kadar bırakacağını ve beklenen ortalama satış inceleyip sonrasında birde beklenen ortalama satış üzerinden ben bir segmentasyon yaptığımı düşünelim. Gözlemledim ki C segmentindeki müşterilerime bir şeyler yapmam gerekiyor ve aksiyon almalıyım, aksiyon kararı olarak bu müşterilerime SMS atmayı tercih ediyorum ve SMS'leri atıyorum.

Bu SMS'leri attığım zamandan 1 ay sonraki hem satışları hem gelirleri gözlemliyorum ki C segmentinde müşterilerde hareketlenmeler olmuş ve bana para bırakmaya başlamışlar bir daha cltv yaptığımda görüyorum ki C deki müşteriler B ye geçiş yapmış işte bunları gözlemledikten sonra belirli tarih aralıklarında düzenli olarak SMS atıldığında şirketin cirolarındaki artışı gözlemleyebiliyor olacağımıdır. CRM departmanlarında yoğun olarak SMS çalışmaları da yapılır ve ciroya

oldukça etkisi yüksek olabilmektedir. Bunu bizzat gözlemlediğim için söylüyorum

3-) Elimizde 2000-2021 veri seti olmasına rağmen 2017-2021 filtrelersek bu modelleri kurduğumuzda yanılmış olur muyuz? Çünkü BG-NBD için müşterinin ilk satın alım yaptığı tarihi değiştirmiş olacağız.

Eğer ödev için soruyorsan bunu ödevdeki veri setimiz 2009 -2011 arasındaki veriler. Ama eğer elimizde 2000-2021 arası veri varsa bu durumda 2017-2021 filtresi atmak yanıltıcı olmayacağını düşünüyorum. Çünkü bu analizler belirli periyotlarla yapılması daha sağlıklı karar aldıracaktır yani veri aylık veya haftalık bile olabilecek şekilde periyotlara bölünüp bunun üzerinden analizler ve tahminlemeler yapılır. Birlikte ele alalım 2017'den itibaren veriyi inceliyoruz. İlk alışveriş 2017 1 ocak olacaktır bazı kişiler için. Bazıları için ise daha ilerisi. Ama sonuçta recency ve Tenure yine kişi özelinde son 4 yıl için hesaplanacak. Ayrıca ben şirkete 2021 yılı için beklenen ortalama satış tahmini yapacağım için son yılların etkisi daha da önemli olacaktır benim için bundan dolayı bir sorun olmayacaktır. Keza müşteri alışveriş patternini da son 4 yıllık dataya da yansıtacaktır mutlaka bu durumda yanıltıcı olmayacaktır diyebiliriz