<https://medium.com/@krglnahmetcan/rfm-analizi-ile-m%C3%BC%C5%9Fteri-segmentasyonu-b3c7233b4243>

RFM analizi müşteri analitiği alanında sıkça kullanılan bir yöntemdir. Bu analiz türünde müşterileri segmentlere ayırarak her bir segmentteki müşteriye özel, satış ve pazarlama teknikleri geliştirilmektedir.

**RFM Analizi Nedir?**

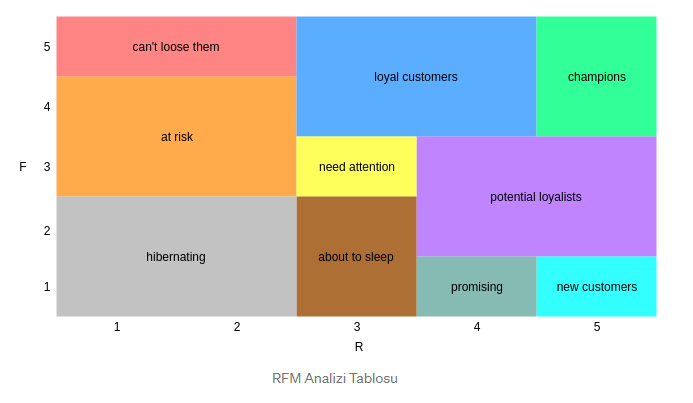
· **Recency —**Yenilik — Müşterinin son satın almasından bugüne kadar geçen süredir. Diğer bir ifade ile “Müşterinin son temasından bugüne kadar geçen süre” dir. **Formül = Bugünün tarihi — Son satın alma tarihi** genel anlamda analizler yapılıyorken “Bugünün tarihi” olarak tanımladığımız kısım analizin yapıldığı gün olarak kabul edilmektedir.

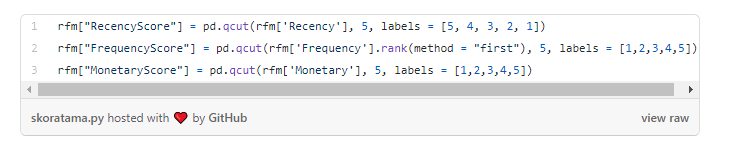
· **Frequency**— Sıklık — Müşterinin toplam satın alma sayısıdır.

· **Monetary —**Parasal Değer — Müşterinin yaptığı toplam harcamadır.

RFM analizinde yaygın olarak kullanılmakta olan bir tablo bulunmaktadır. Genelde bu tabloda 2 parametre kullanılmaktadır. Bunlar **Recency** ve **Frequency** değerleridir. Öncelikli olarak **Recency**, **Frequency** ve **Monetary** değerleri bulunduktan sonra bunları 1 ile 5 arasında değerler verilerek müşterileri segmentlere ayırmaktayız. 5 çok iyi, 1 ise çok kötü anlamına gelmektedir. Tabloda müşteriler segmentlere ayrılmış durumdadır.

Bizlerin burada dikkat etmesi gereken grup aslında herkesin düşündüğü şampiyonlar grubu değil bu grup zaten sürekli olarak bizi ziyaret ediyor ve alışveriş yapıp para bırakıyor. Burada en dikkat gerektiren sınıf “can’t loose them” (onları kaybedemeyiz) sınıfıdır. Çünkü bu kişiler bizi nerdeyse hiç ziyaret etmiyor ve çok fazla alışveriş yapıyor yani frequency değerleri 5 gördüğünüz üzere bu nedenle bu sınıfa odaklanmamız gerekiyor. Çünkü bu grupta yer alan müşteriler bizi bıraktı bırakacak, churn olmak üzereler bu nedenle dikkat gerektirmekte ve bu gruptaki müşterileri gerekli analizler yaparak bize kazandırmalıyız.

Aynı zamanda bir de “need attention” grubu bulunmakta **Recency** değeri 3, **Frequency** değeri 3 olan gruba ve uyku halinde olan gruba özel mailler atarak, promosyonlar yaparak canlandırıp tabloda sağa veya yukarıya doğru çıkarmamız gerekmektedir. Kısacası bu tabloyu bu şekilde açıklayabiliriz. RFM analizinin temel mantığının anlatıldığı tablo bu tablodur. Elde ettiğimiz parametreleri skorlara ayırmamız gerekmektedir. Az önce yukarıda tabloyu anlatırken değindiğim değerlerden bahsediyorum. Aşağıdaki kod parçasını incelediğimizde burada dikkatinizi çeken nokta **Recency** değeri olacaktır. Çünkü tersten bir skorlama yapılmaktadır. Bunun sebebi ise örneğin siz analiz yapılan günden 1 gün önce bizleri ziyaret ederek alışveriş yaptığınızı, diğer müşterimiz ise 100 gün önce bizleri ziyaret ederek alışveriş yaptığını düşünelim. Bu durumda recency değeri **Analiz yapılan günün (Bugünün) tarihi — Son satın alma tarihi** formülü ile hesaplandığında burada siz bizi 1 gün önce ziyaret edip alışveriş yaptığınız hesaplanacaktır. Yani yakın zamanda alışveriş yaptığınız için **recency puanınız: 5** olacaktır. Diğer müşterimiz ise analiz yapılan günden 100 gün önce alışveriş yaptığı için **recency puanı: 1 olacaktır.** Bu nedenle recency değeri düşük olana 5 puan, recency değeri yüksek olana 1 puan verilmektedir. Diğer iki parametre zaten satış arttıkça ve ödenen toplam para arttıkça skor değerleri de artacaktır.



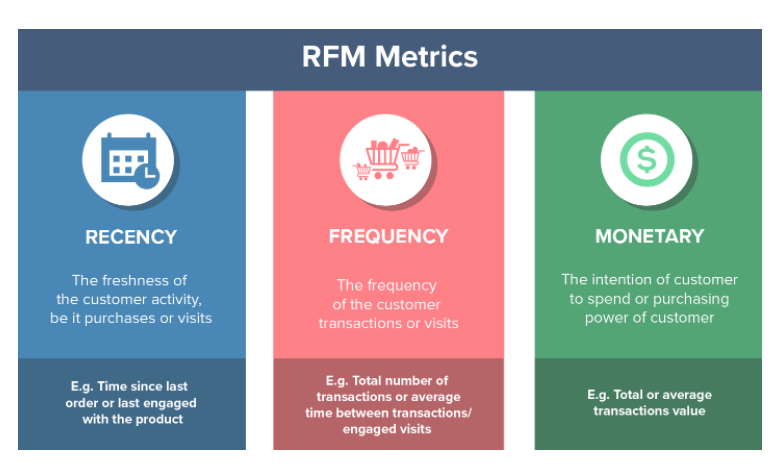
|  |  |  |
| --- | --- | --- |
| rfm["RecencyScore"] = pd.qcut(rfm['Recency'], 5, labels = [5, 4, 3, 2, 1]) | | |
|  | rfm["FrequencyScore"] = pd.qcut(rfm['Frequency'].rank(method = "first"), 5, labels = [1,2,3,4,5]) |
|  | rfm["MonetaryScore"] = pd.qcut(rfm['Monetary'], 5, labels = [1,2,3,4,5]) |

Yapmış olduğum analizi Kaggle platformunda paylaştım. Dilerseniz aşağıdaki linkten notebooka erişim sağlayabilirsiniz.

Kaggle Notebook Link: <https://www.kaggle.com/ahmetcankaraolan/customer-segmentation>

<https://www.linkedin.com/pulse/rfm-recency-frequency-monetary-analizi-nedir-neden-yilmaz-msc/?originalSubdomain=tr>

Bu yöntemlerden biri de RFM Analizi yöntemidir. Genel olarak Online alışveriş sitelerinde alışveriş yapan müşteriler üzerinde analizler yapılırken kullanımının yaygın olduğunu söyleyebiliriz



RFM algoritmalarının kullanım amacı aslında bir müşteri segmentasyonu yaparak müşteri bazında yapılan alışveriş ve ödenen ücret üzerinden bir skor hesaplanmasıdır. Skor doğrultusunda müşterinin RFM skoru bulunmuş olmaktadır.

RFM analizi ile birlikte bazı durumlarda gruplama analizi de faydalı olabilmektedir. Grup analizi yapılmasının sebebi aslında aynı davranışları gösteren müşterileri bir arada görmektir. Ayrıca grup türlerine göre alınacak aksiyonun veya aksiyonların kararlaştırılmasıdır. Örnek vermek gerekir ise belirli bir ürün veya zaman içerisinde ürün alan veya müşteriyi baz aldığımızda ilgili ürün veya zamanda tetiklenebilecek müşteri kitlesi olduğunu düşünülmesi gerektiğini düşünüyorum.

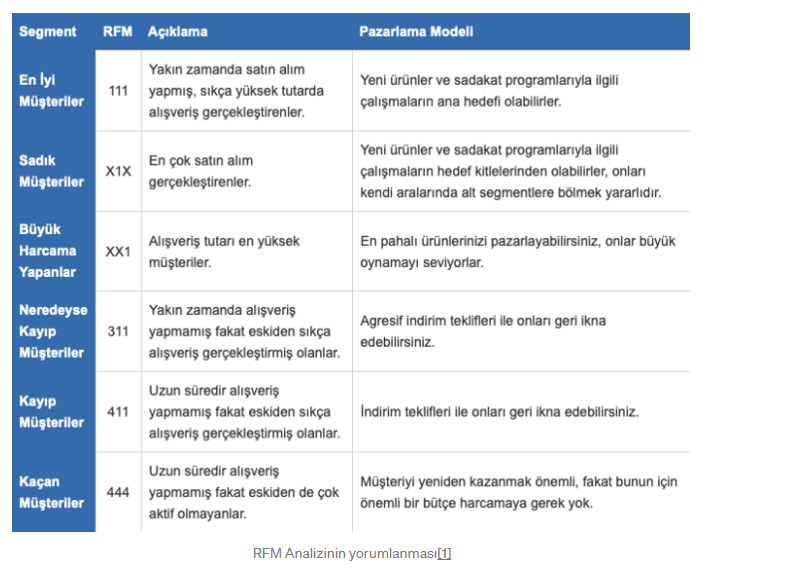
<https://medium.com/kaveai/rfm-analizi-ve-kmeans-algoritmas%C4%B1yla-m%C3%BC%C5%9Fteri-segmentasyonu-fad0842945b1>

bu yazımızda sizlerle RFM analizinin ne olduğunu, nasıl yorumlandığını, örnek bir veride Python kullanarak nasıl uygulandığını ve Kmeans algoritmasını da kullanarak müşteri segmentasyonun nasıl yapıldığından bahsedeceğim.

**RFM analizi** müşterinin geçmiş satın alım davranışlarını inceleyip, onların yenilik, sıklık ve parasallık özelliklerini hesaplayarak müşterileri bölütlemeye (segmente etmeye) yarayan bir analiz tekniğidir.

*Recency — Yenilik (R) — Son satım alımdan bugüne kadar geçen süre  
Frequency — Sıklık (F) — Toplam satın alım sayısı  
Monetary — Parasallık (M) — Tüm satın alımların parasal toplamı*

Her müşteriye ayrı ayrı **yenilik**(R),**sıklık**(F) ve **parasallık**(M) skoru atanır. Bu skor 1'den 5'e kadar skor alır. Biz yapacağımız örnekte müşterileri 1–4 arası skorlayacağız. 1 en iyi, 4 en kötü olarak tanımlanabilir. Yani her skor için 4 ihtimalimiz olduğundan, 4x4x4 = 64 farklı kombinasyon oluşturabiliriz. Böylece detaylı bir bilgi edinmek için yeterli kombinasyon elde etmiş oluruz.

Elde ettiğimiz yenilik(R), sıklık(F), parasallık(M) skorlarını yan yana tek bir skor olarak birleştirip RFM skorumuzu elde ederiz. RFM skorlarının genel yorumlanması yukarıdaki tablodaki gibidir. İşletmelerin türüne göre final RFM skoruna ulaşırken R,F ve M skorlarında göreceli olarak anlamını arttırabilir veya azaltabilirsiniz. Örneğin, sizin işletmeniz beyaz eşya satan bir işletmeyse müşterilerinizin yüksek fiyatlı alışverişler yapmanızı beklersiniz fakat satın alım sıklığı ve yenilik zamanı düşük olacaktır. Bu yüzden bu şartlar altında ayrı ayrı skorları yorumlamak daha doğru sonuçlar verir.

# **KMeans Algoritması ve Müşterilerimizin kümelenmesi**

K merkezli öbekleme (KMeans), iteratif bir algoritmadır. Önceden herhangi bir etiketlenmiş veri verilmediğinden gözetimsiz öğrenme algoritmasıdır. Elinizdeki veri setini birbiriyle kesişmeyecek şekilde K adet öbeğe bölmektedir. Veri setindeki her nokta bu öbeklerden birine ait olur. Veri noktalarını öbeklere mesafesinin karesinin toplamını minimum yapacak şekilde atanır.

Adımları sıralarsak;

*1- Öbek sayınızı (K) seçtikten sonra rastgele K adet merkez seçilir.*

*2- Her veri noktasıyla merkez arasındaki uzaklık hesaplandıktan sonra en yakın öbeğe atanır.*

*3- Daha sonra öbeklerde bulunan verilerin ortalamasına göre yeni merkezler belirlenir ve tekrardan noktalar yakın olduğu öbeklere atanır.*

*4- Bu işlem öbek merkezlerinde değişiklik olmayana kadar devam eder.*

**Dikkat:** 1. adımda K adet merkezin rastgele atandığını söylemiştik. KMeans iteratif bir algoritma olduğu için farklı atanacak merkezler farklı kümelere sebep olmaktadır çünkü algoritmamız local optimuma ulaşırsa orada kalıcak ve global optimuma yakınsayamayacaktır. Bu yüzden algoritmayı merkezleri farklı atayarak birkaç kez çalıştırıp mesafe kareler toplamını sonuç olarak en küçük döndüren sonucu kullanmak başarınızı arttıracaktır.

Standardize etmemiz verimizde eşit varyans elde etmemizi sağlar. Böylece noktalarımızı eşit olarak ağırlandırabiliriz. Amaç fonksiyonuna dikkat ederseniz algoritmamız her küme için varyans tahmini yapmakta ve bunu minimize etmeye çalışmaktadır. Çünkü öbek içinde elde edilen verilen birbirine yakın yani benzer olmasını isteriz.

Verilen veriyi incelediğimizde standartlaştırmadan önce logaritma dönüşümü yapacağız. Bunu yapmamızın sebebi recency,frequency,monetary değerlerine baktığımızda çarpıklık(skewness) söz konusu olması. Yukarıdaki ilk fotoğrafta dağılıma göre çarpıklık çeşitlerini, ikinci fotoğrafta logaritma dönüşümü uygulanmadan önceki recency dağılımını görüyorsunuz.(Koda benzer olarak isterseniz diğer değişkenlerin dağılımınıda inceleyebilirsiniz). Bizim dağılımımızda negatif çarpıklık söz konusu. Algoritmamızda uzaklık metriğimizi öklid kullandığımız için bu çarpıklık sonucumuzu kötü etkileyebilir. Bu yüzden önce logaritma dönüşümü yaparak çarpıklaştırmayı bir miktar düzeltip daha sonra standartlaştırma işlemi uygulayacağız.

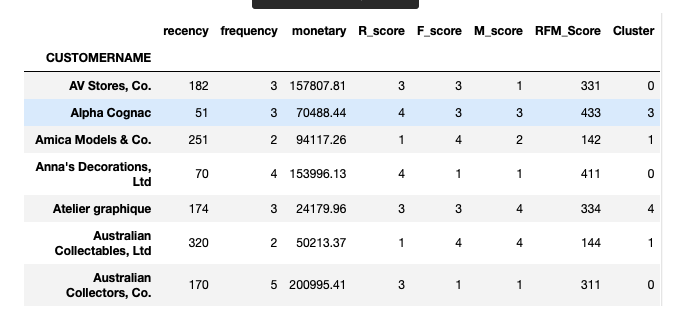
rfm\_vis = np.log(rfm\_vis)

Standartlaştırma işlemimizide aşağıdaki kodla uyguluyoruz.

scaler = StandardScaler()scaler.fit(rfm\_vis)  
data\_normalized = scaler.transform(kmd\_log)

Daha sonra K merkezli öbekleme metodu için öbek sayımızı seçmek için dirsek yöntemini kullanıcaz. Dirsek metodu, x ekseninde öbek sayısı, y ekseninde hatanın kareler toplamını grafik haline bakarak kırılımın azaldığı noktaları tespit etmemizi sağlar. Bizim hedefimiz olabildiğince öbek sayısını az seçerek en küçük hata kare toplamı elde etmektir.

Algoritmamızı 6 öbekle çalıştırıp müşterilerimifzi kümeledikten sonra aynı tablo içerisinde ‘Cluster’ özniteliği içerisinde kaydediyoruz. Böylece tablomuzun son çıktısı şu hali alıyor:



<https://www.kaggle.com/emresirma/k-means-algoritmas-ile-m-teri-segmentasyonu>

# **K-Means Segmentasyonu**

Gözetimsiz öğrenme algoritmalarından biridir. Yapılan işlem, verileri kendine özgü özelliklerine göre ayırmak ve karakteristik özelliklerini ortaya çıkarmaktır. Amaç, noktalar arasındaki toplam mesafeyi en aza indirmek ve kümeler arasındaki mesafeyi en üst düzeye çıkarmaktır. Bu süreç kümelemedir.

Burada iki metrik kullanılabilir.

* Inertia: bir kümedeki noktaların ne kadar uzak olduğunu gösterir. Bu nedenle, küçük bir Inertia hedeflenir. Inertia değeri sıfırdan başlar ve artar. (min SSE)
* Silhoutte Puanı: Bir kümedeki veri noktalarının başka bir kümedeki veri noktalarından ne kadar uzakta olduğunu gösterir. Puan aralığı -1 ile 1 arasındadır ve 1'e yaklaşması beklenmektedir.

Bu çalışmada Inertia kullanılmıştır.

Bu çalışmada, daha önce belirlediğimiz rfm metriklerini k-means ile segmentlere ayıracağız. Veri setimize bu haliyle cluster işlemini uygulayamıyoruz. Bu işlemi yapabilmemiz için sayısal değişkenler lazım. Bu değişkenler şu anda elimizde yok. Biz cluster işlemini rfm metrik değerleri üzerinden k-means ile uygulayacağız.

linkcode

Recency, Frequency and Monetary Değerlerinin Görselleştirilmesi

Burada önemli olan doğru küme sayısını belirlemektir. Bunun için Elbow Method'unu uygularız.

Burada küme sayısı arttıkça, dirsek dediğimiz kısmı kullanıyoruz ve belli bir noktadan sonra değerin çok fazla düşmediğini görüyoruz. Bu nedenle 3 veya 4 kullanmak bizim için daha değerli olabilir.

# **Sonuç**

Bu çalışmada RFM metrikleri üzerinden 3 değişkene dayalı olarak Customer ID değerleri tekilleştirilmiştir.

Bu veri seti, k-means yöntemi uygulanarak sayısal değişkenler temelinde 4 kümeye indirilmiştir.

<https://www.youtube.com/watch?v=fdUofaT8gUw>

# **Customer Segmentation using RFM K-Means & Python | Who are your Loyal Customers ?**

<https://www.youtube.com/watch?v=DowbzgtH0-0>

# **Müşteri Analitiğinde Veri Bilimi**

<https://www.kaggle.com/fszlnwr/customer-segmentation-rfm-cohort-analysis>

Customer Segmentation (RFM & Cohort Analysis)

<https://campus.datacamp.com/courses/customer-segmentation-in-python/cohort-analysis?ex=3>

\*\*\* <https://www.kaggle.com/mahmoudelfahl/cohort-analysis-customer-segmentation-with-rfm>

**\*\*\* Cohort Analysis & Customer Segmentation with RFM**

# Kohort Analizi Nedir? Kohort Analizi Nasıl Yapılır?

<https://www.cancankiran.com/kohort-analizi-nedir-kohort-analizi-nasil-yapilir/>

<https://www.karabulut.co/dijital-pazarlamada-kohort-analizi-cohort/#more-1985>

# Dijital Pazarlamada Kohort Analizi (Cohort)

<https://ichi.pro/tr/python-ve-pandalar-kullanarak-kohort-analizi-231610663111245>

# **Python ve Pandalar Kullanarak Kohort Analizi**

<https://yigitsener.medium.com/cohort-kohort-ile-m%C3%BC%C5%9Fteri-hareket-analizi-ve-python-uygulamas%C4%B1-276cb8882bee>

# Cohort (Kohort) ile Müşteri Hareket Analizi ve Python Uygulaması

<https://www.analyticsvidhya.com/blog/2021/06/cohort-analysis-using-python-for-beginners-a-hands-on-tutorial/>

<https://towardsdatascience.com/a-step-by-step-introduction-to-cohort-analysis-in-python-a2cbbd8460ea>

# A step-by-step introduction to Cohort Analysis in Python