VERİ MADENCİLİĞİ PROJESİ

HAZIRLAYAN

Cansu UZUNŞİMŞEK

# KONU: En çok Ramen’i

# Nasıl ve Hangi Ülkede Satabiliriz ?

MAYIS 2018

*İçindekiler Tablosu*

[KONU: En çok Ramen’i1](#_Toc513715053)

[Nasıl ve Hangi Ülkede Satabiliriz ? 1](#_Toc513715054)

[*İçindekiler Tablosu* 2](#_Toc513715055)

[KONUNUN TANITILMASI 2](#_Toc513715056)

[VERİ KÜMESİNİN TANITILMASI 3](#_Toc513715057)

[VERİLERİN TANITILMASI 3](#_Toc513715058)

[VERİLERİN DÖNÜŞTÜRME İŞLEMİ: 3](#_Toc513715059)

[VERİLERİN TEMİZLENMESİ 3](#_Toc513715060)

[VERİLERİN FİLTRELENMESİ 6](#_Toc513715061)

[VERİ MADENCİLİĞİNDE NEDEN KÜMELEME ANALİZİ KULLANDIM? 6](#_Toc513715062)

[K-MEANS KÜMELEME ALGORİTMASI 7](#_Toc513715063)

[TWO-STEP KÜMELEME ALGORİTMASI 7](#_Toc513715064)

[VERİ MADENCİLİĞİ İŞLEMİ 7](#_Toc513715065)

[1. Two-Step Kümeleme Analizi 8](#_Toc513715066)

[TWO-STEP ANALİZİNİN SONUCU 9](#_Toc513715067)

[2. K-Means Kümeleme Analizi 9](#_Toc513715068)

[K-MEANS ANALİZİNİN SONUCU 11](#_Toc513715069)

[VERİ ANALİZİNİN GENEL SONUCU 12](#_Toc513715070)

[12](#_Toc513715071)

## KONUNUN TANITILMASI

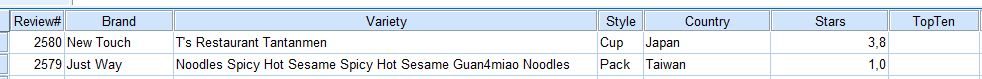
Veri kümesi ,“ Ramen “ adı verilen ve daha çok Uzak Doğu ülkelerinde tüketilen , pirinç unundan yapılan bir çeşit eriştenin tüketimine dair verileri içeri. Her bir veri ayrı bir siparişi temsil eder. Sipariş farklı ülkelerden sipariş verilerini içerir. Bunun yanında sipariş edilen restoranın bilgisi ve sipariş edilen Ramen’in içeriği de veri kümesine dahildir.

## VERİ KÜMESİNİN TANITILMASI

Veriler, uluslararası veri bilimi yarışmalarının düzenlendiği ve içerisinde işletmelere ve çeşitli bilimsel araştırmalara yönelik gerçek veri bulunduran [www.kaggle.com](http://www.kaggle.com) ‘dan alınmıştır. Web sitesinde veri setleri site üyelerince indirilip analiz edilebilmektedir.

Veri Seti; Ramen sipariş etme ve bu siparişin verildiği restoran zinciri markaları ve restoranların bulunduğu ülkelere ilişkindir.

## VERİLERİN TANITILMASI



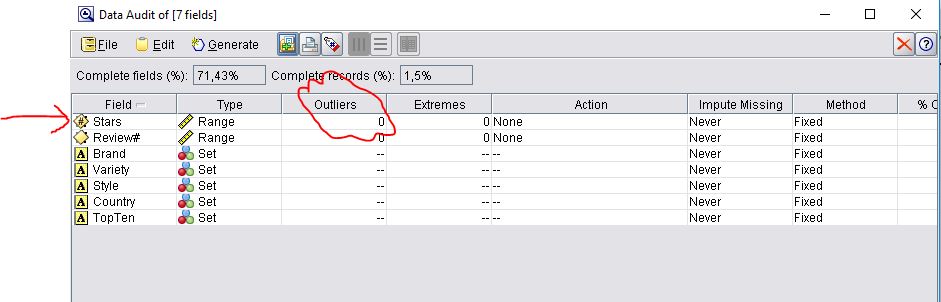
1. Review: Her bir satır için başka bir sıra numarasını ifade eder. Ordinal bir veri sütunudur. Sıralayıcı veya gruplayıcı bir veri çeşidi değildir.
2. Brand: “Ramen” in sipariş edildiği her bir Restoran Markasını ifade eder. String tipinde bir veridir. Nominal bir veridir, sıralayıcı veya ölçülebilen bir veri değildir.
3. Variety : Müşterinin Restorandan sipariş ettiği “Ramen” çeşidini ifade eder.Örneğin; “Soya soslu Ramen” gibi…
4. Style: Müşterinin “Ramen” i tüketme biçimini ifade eder. Örneğin; “ Cup (bardakta) ,Box(kutuda), Tray (tepside)..vb”
5. Country: Müşterinin bulunduğu ve “Ramen” sipariş etiği ülkeyi ifade eder. Örneğin ; “ Japonya, Güney Kore, ABD.. vb”
6. Stars: Müşterinin sipariş ettiği “Ramen”i beğeni puanıdır.0 ile 5 arası tam sayılarla ifade edilir.
7. TopTen: 1 yıl içerisinde en iyi “Ramen” i yapan Restoran seçilen işletme , 1.’den 10.’ya kadar seçilir. Bu sütun sadece sıralamada yer alabilen restoranlar için veri taşır.

## VERİLERİN DÖNÜŞTÜRME İŞLEMİ:

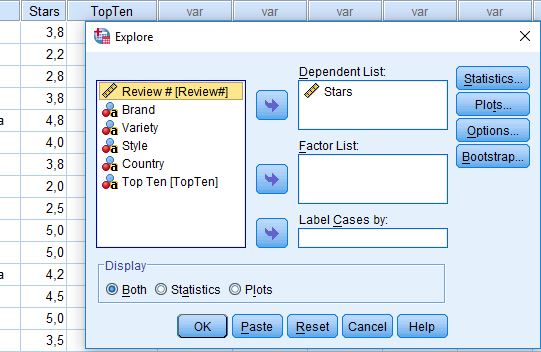
Veriler .csv formatında alınmıştır. Daha sonrasın da Excel programında düzenlenebilmek için ; Not Pad programında “,” seperatörü “;” seperatörüne dönüştürülmüştür. Daha sonra bu dosya; kayıp verileri temizlendikten sonra “ Excel Çalışma Kitabı“ olarak kaydedilmiştir. Oluşan bu yeni dosya IBM SPSS Modeller’da veri türleri belirlendikten sonra “.sav “ dosya uzantısıyla kaydedilmiştir. Kaydedilen dosya son olarak Clementine programında analiz edilmiştir.

## VERİLERİN TEMİZLENMESİ

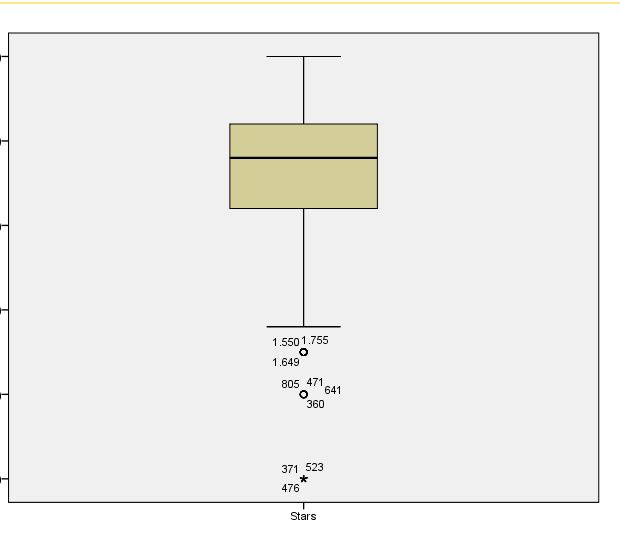
Clementine programının “Data Audit “ özelliğini veri tabanına bağlanarak ve verinin kalitesini kontrol edilmiştir.



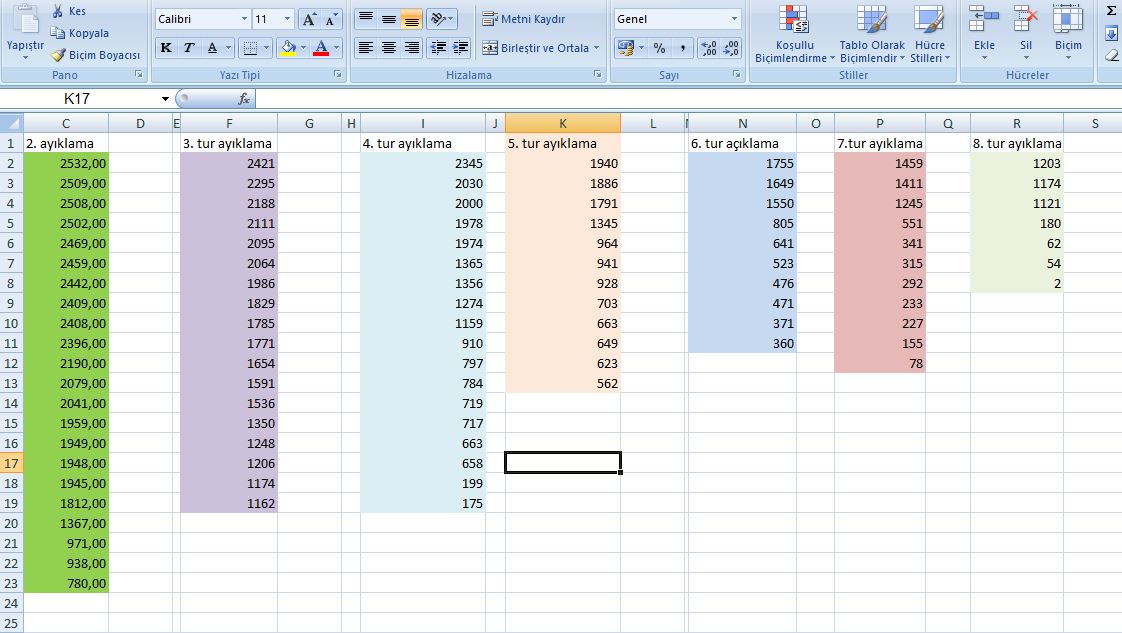
“Outlier” yani aşırı uç değerler 49 idi. Daha sonra verileri aşağıdaki şekilde aşırı uç değerlerden temizlenmiştir.

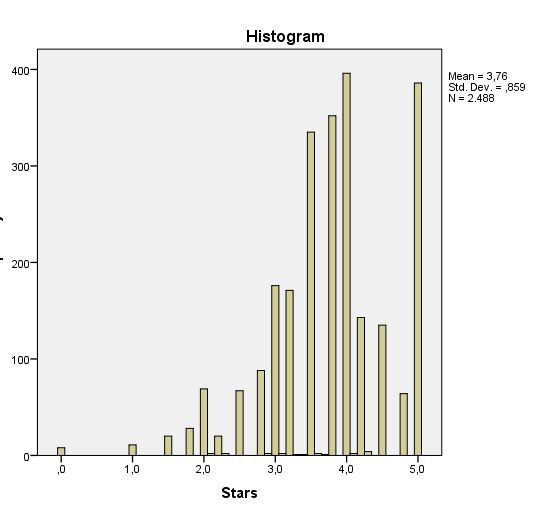


IBM SPSS kullanarak “Anayze” sekmesinden “ Descriptive Statistics” sekmesinden “explore” bölüne tıklayıp yukarıda görülen ekrandan “Stars” sekmesinin seçilip ve verinin normal dağılıp dağılmadığına dair histogram grafiği elde edilmiştir.

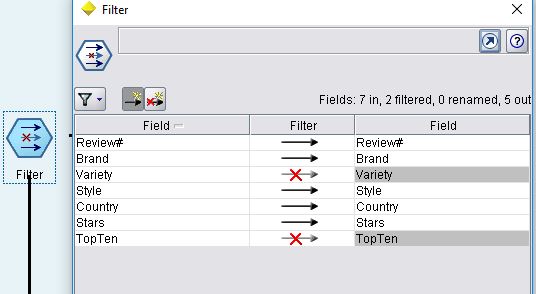


Bu grafikdeki veriler aşağıdaki Excel dosyasına aktarılarak büyükten küçüğe sıralandıktan sonra IBM SPSS kullanılarak tek tek temizlenmiştir. Toplamda ayıklama işlemi bir kez tekrarlandıktan sonra Data Audit’de tekrar kontrol edilmiştir. Tekrar çıkan uç değerler yine aynı yöntemle temizlenmiş ve bu işlem tam 8 kez devam edilmiştir.





## VERİLERİN FİLTRELENMESİ

****

Analizde Clementine programını kullanarak” Variety” ve “TopTen” kısımlarını yani Ramen çeşidini ve en iyi seçilen restoranlar dahil edilmemiştir. İki veri de “String “ değişken tipinde olduğu ve nominal olmadığı için kümeleme analizinde değer oluşturamamaktadır.

## VERİ MADENCİLİĞİNDE NEDEN KÜMELEME ANALİZİ KULLANDIM?

Kümeleme analizi nominal verileri analiz etmede kullanılır. Sınıflayıcı analiz bu veri setini analiz etmede uygun değildir.Çünkü Kümeleme bir çeşit denetimsiz öğrenmedir. Kümeleme benzer eğilim ve modelleri paylaşan verileri gruplandırma metodudur. Veri kümeleme, büyük veri setleri içindeki benzer küçük setleri gruplandıran bir yöntemdir.

## K-MEANS KÜMELEME ALGORİTMASI

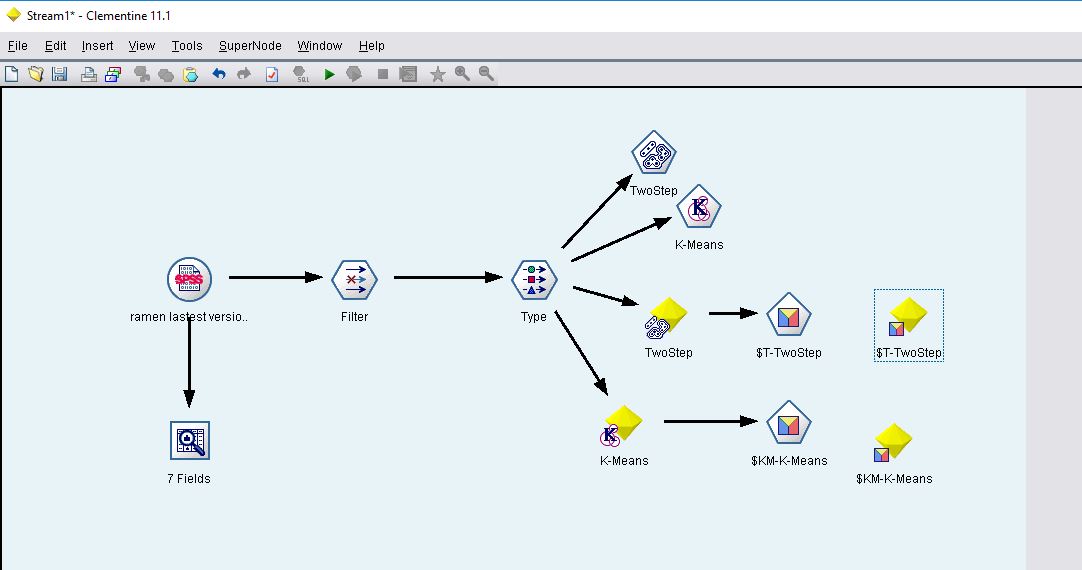
K-Means (K-Ortalama) algoritmasında ; “K” verilerin ayrılacağı küme sayısını belirler. K sayısına yani verilerin ayrılacağı kümelerin sayısına analizi yapacak kişi karar verir. K-Means algoritması aykırı uç değerlere karşı çok duyarlıdır. Bu nedenle verilerin aykırı uç değerlerden temizlenmesi önemlidir.

## TWO-STEP KÜMELEME ALGORİTMASI

Two-Step algoritması ; yaygın olarak veri madenciliği uygulamalarında kullanılan bir algoritmadır. Bu algoritmada veri analizini yapan kişi oluşturulacak maksimum ve minimum küme sayısını belirler. Belirleme işleminden sonra uygun küme sayısına algoritma kendisi karar verir. Algoritma en uygun küme sayısını belirlemek için BIC ( Schwarz’s Beyesian In formatin Criterion) veya AIC ( Akaike’s Information Criterion) yöntemlerini kullanır.İki aşamadan kümeleme gerçekleştirir ve bu aşamalarda hiyerarşik kümeleme yapar.

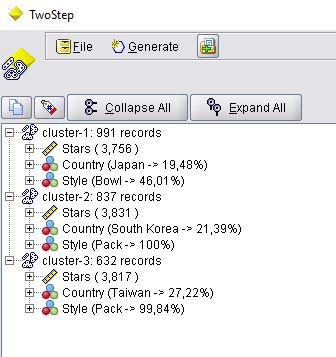
## VERİ MADENCİLİĞİ İŞLEMİ

Temizlenen ve kayıp değerleri düzeltilen veri setine 2 farklı kümeleme analizi uygulanmıştır. Bunlar Two-Step ve K-Means kümeleme algoritmalarıdır.



Veri temizleme işlemi yaptığımız SPSS dosyası “ Data Audit “ yöntemiyle tekrar kontrol edilmiştir. Verinin aykırı uç değerlerden tamamen temizlendiğine karar verildiğinde “Filter” kısmından analize dahil edilmeyecek veri sütunları elenmiştir.Daha sonra “Type” yöntemiyle verilerin tipi belirlenmiş ve en sonunda da kümeleme algoritmalarıyla işleme konulmuştur. En sonunda da “Discriminant” sekmesi ile kümelerin kalitesini ölçmek için diskiriminant analizi yapılmıştır.

## Two-Step Kümeleme Analizi

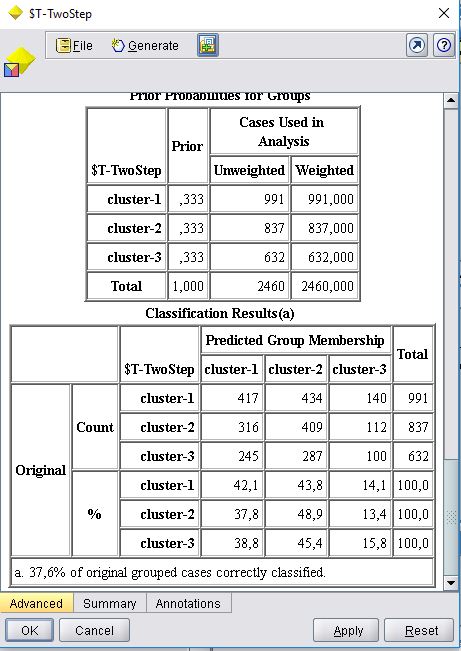


3 kümeleme oluşturulmuştur. Bu üç küme en çok beğenilen restoran markalarının bulunduğu kümeleri ifade etmektedir.Bunlar :

* En çok oyu alarak Japonya , Ramen sipariş memnuniyeti en yüksek olan ülke konumundadır. %46,01oranı ile en çok sevilen sipariş biçimi , kase ile sipariş edilen Ramen’dir. Kase biçiminde siparişi en çok sevilen ülkeler; 2.olarak %14,93 oranıyla Malezya ve 3. olarak %12,31 oranla Endonezya’dır.
* %21,39 oranında “ tepsi” biçiminde sipariş ile Ramen sipariş memnuniyeti en yüksek ülke Güney Kore’dir.Bunu %14,93 beğenilme oranıyla Malezya izlemektedir.
* %99,84 oranıyla “ paketleme” biçimi sipariş ile en çok sevilen ülke Tayland’dır.Bu oranı %23, 89 oranıyla Japonya izlemektedir. %14,56 oranıyla Çin tepsi ile servis tercihlerinde 3. Olmuştur.

Oluşturulan küme sayısının kalitesi diskriminant analizi ile test edilmiştir. %36,7 ile en anlamlı küme sayısı 3

olduğu gözlenmiştir.



## TWO-STEP ANALİZİNİN SONUCU

Ramen siparişi en çok beğenilen ülkeler sırasıyla şöyledir;

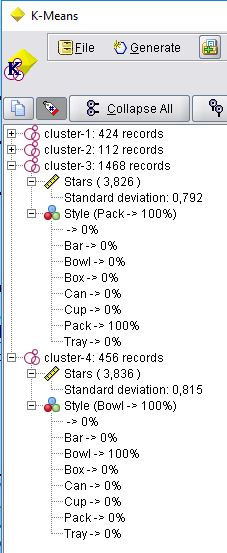
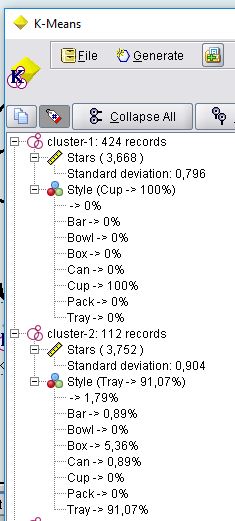
1.Japonya

2.Güney Kore

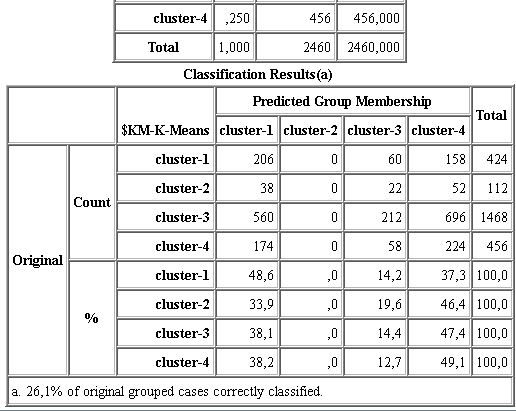
3. Tayland

Bu ilkeleri Çin ve Endonezya izlemektedir. Genel olarak Uzak Doğu ilkelerindeki beğenilme oranları çok daha yüksektir.

## K-Means Kümeleme Analizi



4 kümeleme oluşturulmuştur. Bu 4 küme sırasıyla; bardak,tepsi , paket, kase şeklinde sipariş edildiğinde beğenilme puanlarını büyükten küçüğe doğru sıralanmış biçimini ifade eder.( En çok kullanılan sunum biçimi bu 4 küme olduğu için , küme sayısı 4 olarak belirlenmiştir.) Ayrıca diskrimant analiziyle de kümelerin ayrılmasında en doğru sayı kontrol edilmiştir. %26,1 ile en anlamlı küme sayısının 4 olduğu gözlemlenmiştir.



* Birinci kümede görüldüğü üzere ; 3,668 oy ortalaması ile en çok sevilen 4. Ramen sipariş biçimi bardaktır.
* İkinci kümede görüldüğü üsere; 3,753 oy ortalamasıyla en çok sevilen edilen Ramen sipariş biçimi Tepsi (Tray) dir.Bunu çok küçük oranlarla; %5,36 ile Karton Kutu (Box) ve %0,89 oranıyla Teneke kutu (Can) izler. Çok küçük oranlarda oldukları için diğer siparil biçimleri Tepsi ( Tray) ile birleştirilmiştir. Diğerlerinin ayrı küme olarak değerlendirilmesi diskriminant analizinde küme kalitesini düşürdüğü gözlemlendiği için küme sayısı 4 olarak belirlenmiş ve Box ile Can veri kümesi Tray ile birleştirilmiştir. En anlamlı sonuca böyle ulaşılmıştır. Dolayısıyla “en çok sevilen 3. Sipariş biçimi Tepsi (Tray ) dir” sonucunu çıkartabilmelteyiz.
* Üçüncü kümede görüldüğü üzere; 3,668 oy ortalamasıyla en çok sevilen 2. Ramen sipariş biçimi pakettir.
* Dördüncü kümede görüldüğü üzere ; 3,836 oy oranı ile en çok kullanıcı oy ortalamasına sahip olan ve en çok sevilen Ramen sipariş biçimi olan Kase (Bowl) dur.

## K-MEANS ANALİZİNİN SONUCU

Bowl yani Kase de servis edilme biçimi en çok sevilen Ramen türüdür. En çok kullanıcı tarafından beğenilen ve oy verilen Sipariş türüdür. Sevilme oranlarına göre Ramen servis edilme biçimleri şöyledir:

1.Kase (Bowl)

2.Paket (Pack)

3.Tepsi (Tray)

4.Bardak (Cup)

5.Box (Kutu)

6. Can (Teneke Kutu)

## VERİ ANALİZİNİN GENEL SONUCU

## Analizin sonucunda en çok “Ramen “ satışı yapabilmek için kasede servis yapılacağına karar verilebilmektedir.Ayrıca en çok Güney Kore, Japonya ve Tayland Ülkelerinde satış yapılıcağına karar verilebilmektedir. Bunun için yerel reklam kampanyaları düzenlenebilmektedir. Örneğin :

## Sadece Japonya ,Güney Kore ve Tayland’da reklam yaparak reklam gelirlerini azaltılabilir. Japonya’da daha çok Ramen beğenildiği için Japon müşterilere özel kampanyalar oluşturabilmektedir. Eğer Ramen içeriği için malzeme üreten bir tedarikçi firma varsa Japon Restoranlarıyla uzun süreli tedarik anlaşmaları yapılabilmektedirler.

## Bunun yanında eğer Japonya’da alım gücünden dolayı bir Pazar küçülmesi yaşanırsa yani insanlar Japonya’da daha az Ramen almaya başlarsa ve Ramen satışları düşerse; Pazar payı diğer satış oranı yüksek olan ülkelere kaydırılabilmektedir. Örneğin Japonya’da Ramen satışı olmazsa Güney Kore’ye satış kampanyaları veya Tayland ‘da yatırım yapılabilmektedir.