

VERİ MADENCİLİĞİ PROJESİ

MARKET SEPETİ ANALİZİ

HAZIRLAYANLAR:

ŞEVKET KARABULUT
(02205076038)

MEHMET MERT MENEVŞE
(02205076048)

REŞAT ERAY ÖZTÜRK
(02195076043)

ALPEREN GÜL (02205076025)



MARKET SEPETİ ANALİZİ UYGULAMASINDA KULLANILAN ARAÇLAR

- colab.Google (Python kodlamak için)
- Excel (Veri seti için)

MARKET SEPETİ ANALİZİ NEDİR:

- Market sepeti analizi, müşterilerin alışveriş alışkanlıklarını analiz etmek için kullanılan bir veri madenciliği tekniğidir. Bu analiz, müşterilerin satın alma davranışlarını anlamak ve bu bilgileri iş stratejilerini geliştirmek için kullanılır.
- Market sepeti analizi, müşterilerden veya POS (satış noktası) sistemlerinden toplanan veri kümeleri kullanılarak yapılır. Bu veri kümeleri, müşterilerin satın aldıkları ürünleri, satın alma tarihlerini ve satın alma miktarlarını içerir.
- Market sepeti analizi, birliktelik kuralları adı verilen bir teknik kullanarak gerçekleştirilir. Birliktelik kuralları, iki veya daha fazla ürünün birlikte satın alınma olasılığını ölçen kurallardır. Örneğin, "Bir müşteri süt alıyorsa, aynı zamanda ekmek de alma olasılığı %80'dir." gibi bir kural, bir birliktelik kuralıdır.

MARKET SEPETİ ANALİZİ AMACI:

- Market sepeti analizi aşağıdaki amaçlar için kullanılabilir:
- Müşteri satın alma davranışlarını anlamak
- Yeni ürün önerileri yapmak
- İndirim kampanyaları tasarlamak
- Mağaza raf düzenlemelerini optimize etmek

MARKET SEPETİ ANALİZİ UYGULAMALARI

- Bir süpermarket, birliktelik kurallarını kullanarak, süt ve ekmek gibi birlikte sıklıkla satın alınan ürünlerin raflarını birbirine yakın konumlandırabilir. Bu, müşterilere bu ürünleri daha kolay görmelerini ve satın almalarını sağlayabilir.
- Bir çevrimiçi perakendeci, birliktelik kurallarını kullanarak, müşterilerin geçmiş satın alımlarına dayalı olarak kişiselleştirilmiş ürün önerileri sunabilir. Bu, müşterilerin ilgi alanlarına daha uygun ürünler bulmalarına yardımcı olabilir.
- Bir giyim mağazası, birliktelik kurallarını kullanarak, belirli bir ürün satın alan müşterilerin aynı zamanda diğer hangi ürünleri satın alma olasılığının yüksek olduğunu belirleyebilir. Bu bilgiler, mağazanın hangi ürünlerini daha fazla vurgulaması gerektiğini belirlemesine yardımcı olabilir.

BİRLİKTELİK KURALLARI

- Birliktelik kuralları, aşağıdaki iki temel bileşenden oluşur:
- Destek(Support) : Birliktelik kuralının geçerli olduğu öğelerin oranı. Örneğin, destek; süt ve ekmek satın alan müşterilerin toplam müşteri sayısına oranıdır.
- Güven(Confidence): Birliktelik kuralının geçerli olduğu öğelerin, kuralın öncül öğelerini içeren öğelerin oranı. Örneğin, güven; süt satın alan müşterilerin aynı zamanda ekmek de alma olasılığıdır.
- Müşteri satın alma davranışlarını anlamak: Birliktelik kuralları, müşterilerin hangi ürünleri birlikte satın aldıklarını anlamaya yardımcı olabilir. Bu bilgiler, işletmelerin ürün önerileri, indirim kampanyaları ve mağaza raf düzenlemeleri gibi stratejilerini geliştirmelerine yardımcı olabilir.
- Yeni ürün önerileri yapmak: Birliktelik kuralları, müşterilerin geçmiş satın alımlarına dayalı olarak yeni ürün önerileri yapmak için kullanılabilir. Bu bilgiler, müşterilerin ilgi alanlarına daha uygun ürünler bulmalarına yardımcı olabilir.
- İndirim kampanyaları tasarlamak: Birliktelik kuralları, belirli ürünler satın alan müşterileri hedeflemek için indirim kampanyaları tasarlamak için kullanılabilir. Bu bilgiler, işletmelerin kampanyalardan en iyi şekilde yararlanmalarına yardımcı olabilir.
- Mağaza raf düzenlemelerini optimize etmek: Birliktelik kuralları, birlikte sıklıkla satın alınan ürünlerin raflarını birbirine yakın konumlandırmak için kullanılabilir. Bu, müşterilere bu ürünleri daha kolay görmelerini ve satın almalarını sağlayabilir.

APRİORİ ALGORİTMASI

- Apriori algoritması, birliktelik kuralları üretmek için kullanılan bir veri madenciliği algoritmasıdır. Bu algoritma, market sepeti analizi gibi uygulamalarda yaygın olarak kullanılır.
- Apriori algoritması, aşağıdaki adımlardan oluşur:
- Destek gereksinimi belirlenir. Birliktelik kuralının geçerli olması için, kuralın öncül öğelerinin veri kümesinde belirli bir yüzdeye karşılık gelmesi gerekir. Bu yüzde, destek gereksinimi olarak adlandırılır.
- Tüm olası öge kümeleri oluşturulur. Bu kümeler, veri kümesinde bulunan tüm öğeleri içerir.
- Destek gereksinimini karşılayan öge kümeleri belirlenir. Bu kümeler, veri kümesinde belirli bir yüzdeye karşılık gelir.
- Güven gereksinimi belirlenir. Birliktelik kuralının güvenilir olması için, kuralın öncül öğelerinin kuralın sonucunu içeren öğelerin belirli bir yüzdesine karşılık gelmesi gerekir. Bu yüzde, güven gereksinimi olarak adlandırılır.
- Destek ve güven gereksinimlerini karşılayan birliktelik kuralları belirlenir.
- Apriori algoritması, aşağıdaki iki temel varsayıma dayanır:
 - Geniş bir öge kümesinin herhangi bir alt kümesi de geniş olacaktır. Örneğin, eğer {süt, ekmek} kümesi geniş ise, {süt} kümesi de geniş olacaktır.
 - Bir öge kümesi geniş ise, bu kümenin destek değeri de yüksek olacaktır.

APRORI UYGULAMALARI

- Müşteri satın alma davranışlarını anlamak: Apriori algoritması, müşterilerin hangi ürünleri birlikte satın aldıklarını anlamaya yardımcı olabilir. Bu bilgiler, işletmelerin ürün önerileri, indirim kampanyaları ve mağaza raf düzenlemeleri gibi stratejilerini geliştirmelerine yardımcı olabilir.
- Yeni ürün önerileri yapmak: Apriori algoritması, müşterilerin geçmiş satın alımlarına dayalı olarak yeni ürün önerileri yapmak için kullanılabilir. Bu bilgiler, müşterilerin ilgi alanlarına daha uygun ürünler bulmalarına yardımcı olabilir.
- İndirim kampanyaları tasarlamak: Apriori algoritması, belirli ürünler satın alan müşterileri hedeflemek için indirim kampanyaları tasarlamak için kullanılabilir. Bu bilgiler, işletmelerin kampanyalardan en iyi şekilde yararlanmalarına yardımcı olabilir.
- Mağaza raf düzenlemelerini optimize etmek: Apriori algoritması, birlikte sıklıkla satın alınan ürünlerin raflarını birbirine yakın konumlandırmak için kullanılabilir. Bu, müşterilere bu ürünleri daha kolay görmelerini ve satın almalarını sağlayabilir.

BİRLİKTELİK KURALLARI FORMÜLLERİ

- Confidence(Güven):Birliktelik kuralının güvenilirliğini ölçen bir değerdir. Confidence, kuralın öncül öğelerini içeren öğelerin, kuralın sonucunu içeren öğelerin oranı olarak hesaplanır.
- Lift(İlginçlik):Birliktelik kuralının kuralın öncül öğelerinin bağımsız olarak satın alma olasılığına göre ne kadar güçlü olduğunu ölçen bir değerdir. Lift, kuralın güven değerinin, kuralın öncül öğelerinin bağımsız olarak satın alma olasılığına olan oran olarak hesaplanır.
- Conviction:Birliktelik kuralının kuralın öncül öğelerini içeren öğelerin, kuralın sonucunu içermeyen öğelerin oranı olarak hesaplanır.
- Leverage:Birliktelik kuralının kuralın öncül öğelerinin bağımsız olarak satın alma olasılığını ne kadar artırdığını ölçen bir değerdir.
- Coverage:Birliktelik kuralının veri kümesi içindeki öğelerin ne kadarını kapsadığını ölçen bir değerdir. Coverage, kuralın öncül öğelerini içeren öğelerin, veri kümesi içindeki öğelerin oranı olarak hesaplanır.

$$\text{Confidence}(A \rightarrow B) = P(A|B) = \frac{\text{support}(A,B)}{\text{support}(A)}$$

Confidence(Güven) değerinin formülü

$$\text{Lift}(A \rightarrow B) = \frac{P(B|A)}{P(B)} = \frac{\text{support}(A,B)}{\text{support}(A) * \text{support}(B)}$$

Lift(İlginçlik) değerinin formülü

$$\text{Conviction}(A \rightarrow B) = \frac{(1 - \text{support}(B))}{(1 - \text{Confidence}(A \rightarrow B))}$$

Conviction değerinin formülü

$$\text{Leverage}(A \rightarrow B) =$$

$$\text{support}(A,B) - (\text{support}(A) * \text{support}(B))$$

Leverage değerinin formülü

$$\text{Coverage}(A \rightarrow B) = \text{support}(A)$$

Coverage değerinin formülü

VERİ SETİ

- Veri setinden alınan ilk 5 satır yazdırılır.

	0	1	2	3	4	5	6
0	Ekmek	Kola	Yumurta	Kiyma	Peynir	Kalem	Bebek Bezi
1	Ekmek	Peynir	Kiyma	Bebek Bezi	Kola	Sut	Kalem
2	Peynir	Kiyma	Yumurta	Sut	Kola	NaN	NaN
3	Peynir	Kiyma	Yumurta	Sut	Kola	NaN	NaN
4	Kiyma	Kalem	Kola	NaN	NaN	NaN	NaN

VERİLERİN ETİKETLENMESİNİN KOD BLOĞU

- `encoded_vals = []` #Boş bir liste oluşturulur. Bu liste, her satır için oluşturulan etiketleme sözlüklerini içerecek.
- `for index, row in hamveri.iterrows():` #hamveri veri seti üzerinde bir döngü başlatılır. Bu döngü, veri seti'nin her satırını temsil eder.
- `labels = {}` #Her satır için boş bir etiketleme sözlüğü oluşturulur. Bu sözlük, sütun isimleriyle ilişkilendirilmiş değerlere sahip olacak.
- `uncommons = list(set(column_names) - set(row))` #uncommons, o satırdaki benzersiz sütun isimleri ile `column_names` içindeki sütun isimleri arasındaki farkı bulur. Yani, bu satırda bulunmayan sütun isimlerini belirler.
- `commons = list(set(column_names).intersection(row))` #commons, o satırdaki benzersiz sütun isimleri ile `column_names` içindeki sütun isimleri arasındaki kesişimi bulur. Yani, bu satırda bulunan sütun isimlerini belirler.
- `for uc in uncommons:` #Her bir eksik sütun ismi için, etiketleme sözlüğüne bu sütun ismini ve karşılık gelen değeri (0) ekler. Yani, bu sütunlar bu satırda bulunmuyorsa, değer 0 olarak atanır.
- `labels[uc] = 0`
- `for com in commons:` # Her bir ortak sütun ismi için, etiketleme sözlüğüne bu sütun ismini ve karşılık gelen değeri (1) ekler. Yani, bu sütunlar bu satırda bulunuyorsa, değer 1 olarak atanır.
- `labels[com] = 1`
- `encoded_vals.append(labels)` #Her satırın etiketleme sözlüğü, `encoded_vals` listesine eklenir. Bu, her bir satır için oluşturulan etiketleme sözlüklerini içeren bir liste oluşturur.

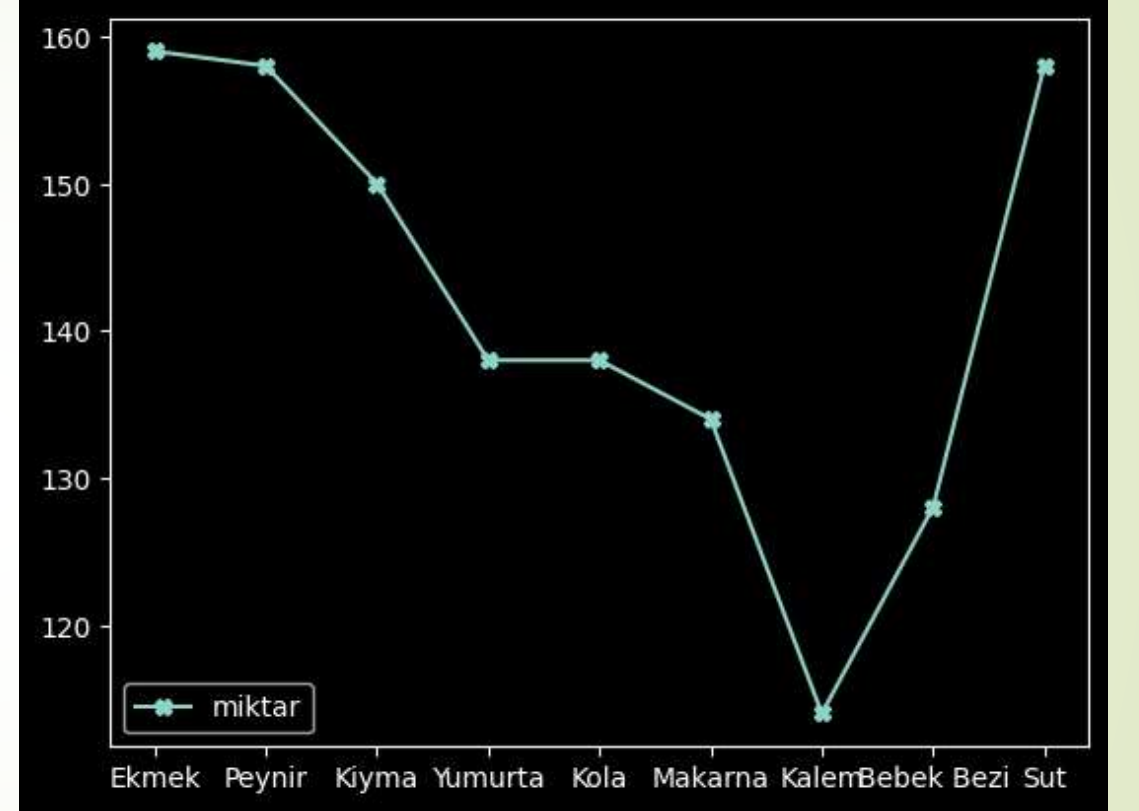
VERİ SETİ MİKTAR ANALİZİ

- Veri setindeki ürünlerin toplam satış miktar analizi yapılır.

	Miktar
Ekmek	159
Peynir	158
Kiyma	150
Yumurta	138
Kola	138
Makarna	134
Kalem	114
Bebek Bezi	128
Sut	158

VERİ SETİ FREKANS GRAFİĞİ

- Veri seti analiz edilerek ürünlerin toplam satış miktarı analizi ile frekans değeri(Tekrar sayısının) grafiğı oluşturulur.



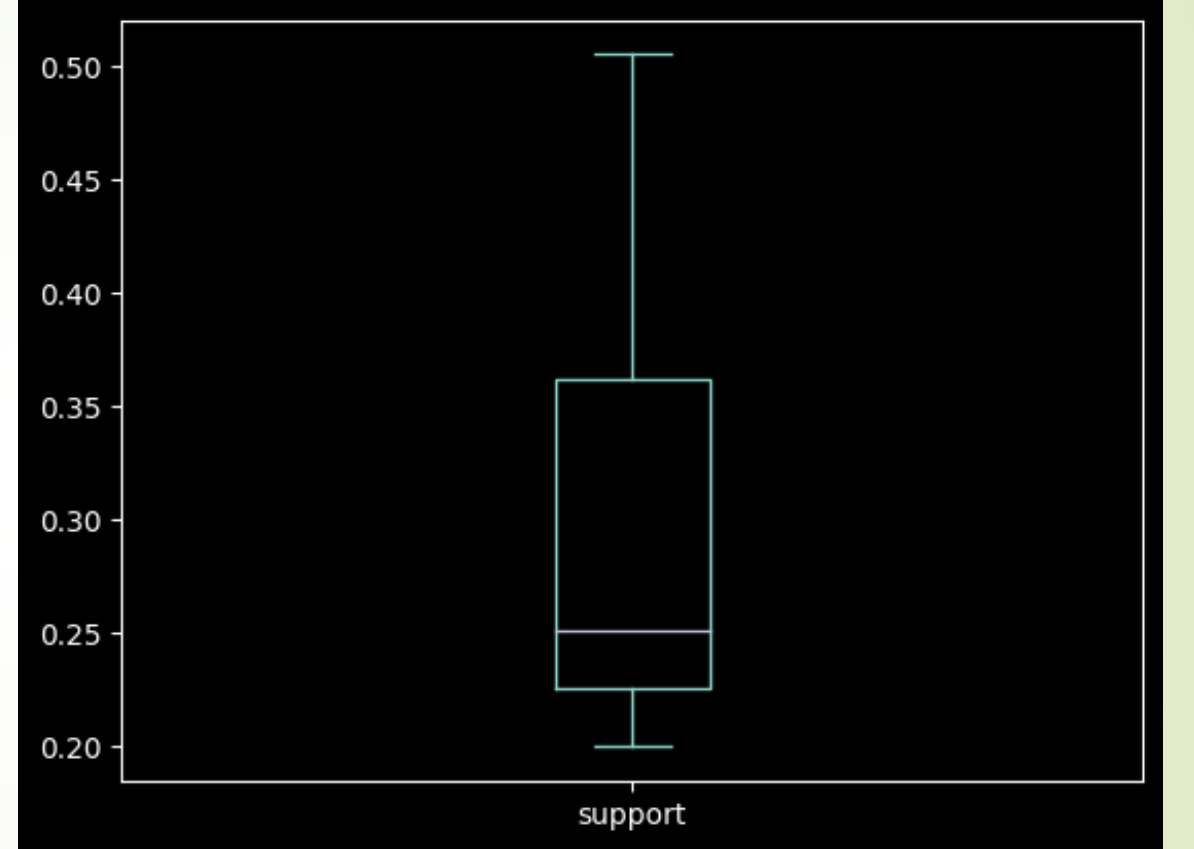
VERİ SETİ SUPPORT DEĞERLERİ

- Veri setindeki verilerin satın alınma oranları.

	support	itemsets
0	0.501587	(Sut)
1	0.425397	(Makarna)
2	0.406349	(Bebek Bezi)
3	0.504762	(Ekmek)
4	0.438095	(Kola)
5	0.501587	(Peynir)
6	0.361905	(Kalem)
7	0.476190	(Kiyıma)
8	0.438095	(Yumurta)

SUPPORT TABLOSU

- Tek veya birlikte ürün kombinasyonları ortalama 0.25 destek değerine sahip olanların plot.box grafiđi çizilir.



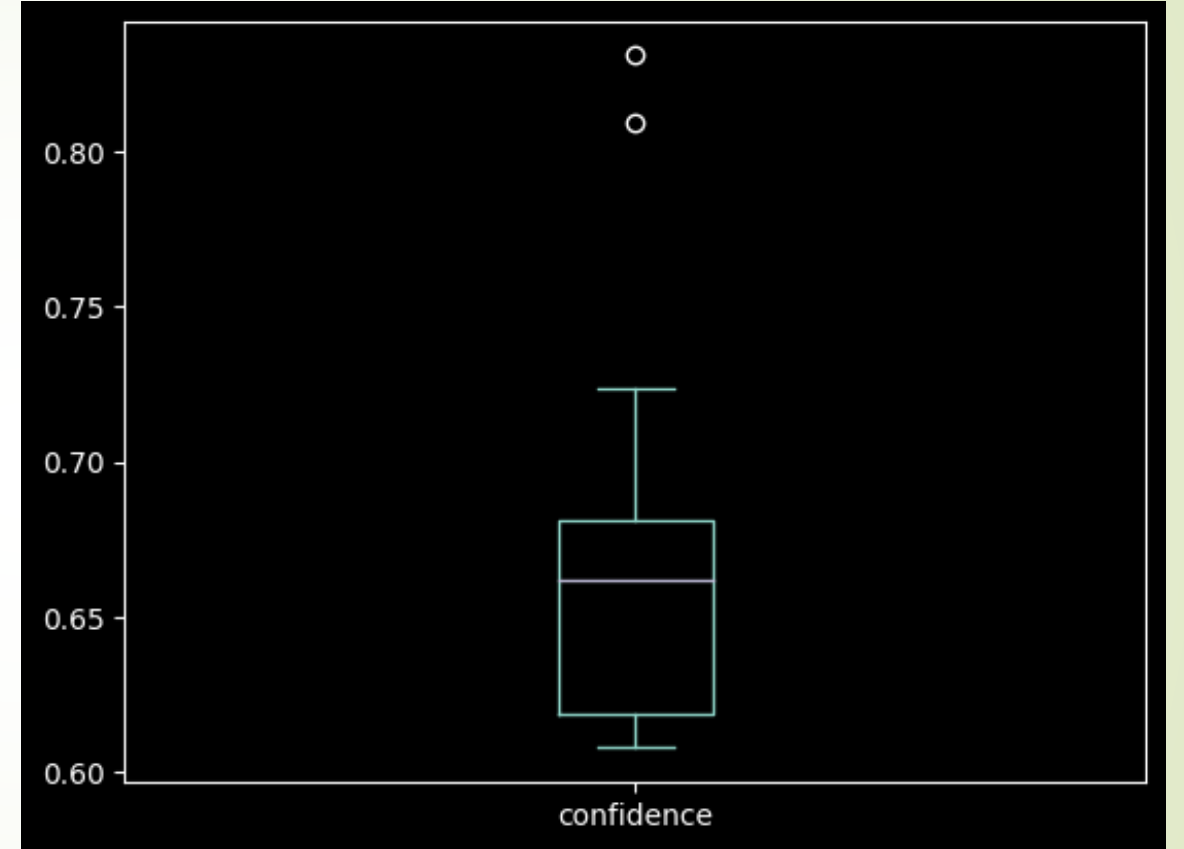
SONUÇ TABLOSU

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(Sut)	(Peynir)	0.501587	0.501587	0.304762	0.607595	1.211344	0.053172	1.270148	0.350053
1	(Peynir)	(Sut)	0.501587	0.501587	0.304762	0.607595	1.211344	0.053172	1.270148	0.350053
2	(Makarna)	(Ekmek)	0.425397	0.504762	0.279365	0.656716	1.301042	0.064641	1.442650	0.402687
3	(Kola)	(Peynir)	0.438095	0.501587	0.269841	0.615942	1.227986	0.050098	1.297754	0.330409
4	(Peynir)	(Kiyma)	0.501587	0.476190	0.323810	0.645570	1.355696	0.084958	1.477891	0.526414
5	(Kiyma)	(Peynir)	0.476190	0.501587	0.323810	0.680000	1.355696	0.084958	1.557540	0.500891
6	(Yumurta)	(Peynir)	0.438095	0.501587	0.298413	0.681159	1.358008	0.078670	1.563203	0.469167
7	(Yumurta)	(Kiyma)	0.438095	0.476190	0.266667	0.608696	1.278261	0.058050	1.338624	0.387409
8	(Sut, Peynir)	(Kiyma)	0.304762	0.476190	0.203175	0.666667	1.400000	0.058050	1.571429	0.410959
9	(Sut, Kiyma)	(Peynir)	0.244444	0.501587	0.203175	0.831169	1.657077	0.080564	2.952137	0.524816
10	(Peynir, Kiyma)	(Sut)	0.323810	0.501587	0.203175	0.627451	1.250931	0.040756	1.337845	0.296655
11	(Peynir, Yumurta)	(Kiyma)	0.298413	0.476190	0.215873	0.723404	1.519149	0.073772	1.893773	0.487091
12	(Peynir, Kiyma)	(Yumurta)	0.323810	0.438095	0.215873	0.666667	1.521739	0.074014	1.685714	0.507042
13	(Yumurta, Kiyma)	(Peynir)	0.266667	0.501587	0.215873	0.809524	1.613924	0.082116	2.616667	0.518717

Minimum confidence(güven) değerimizi belirliyoruz. (%60'tan fazla olanları alıyoruz.)
Confidence değerlerimiz bizim sistemimizin tahmin değerlerini oluşturuyor.

CONFİDENCE GRAFİĞİ

- Tablomuzda oluşan confidence(güven) değerlerinin plot.box grafiğini çizdirelim.



ÇIKTI TABLOSU

Bu 2 değeri listelemek için bir sorgu gerçekleştiriyoruz. Confidence(güven) değeri 0.8'den yukarıda olan yani %80den yukarıda olan değerler.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
9	(Sut, Kiyma)	(Peynir)	0.244444	0.501587	0.203175	0.831169	1.657077	0.080564	2.952137	0.524816
13	(Yumurta, Kiyma)	(Peynir)	0.266667	0.501587	0.215873	0.809524	1.613924	0.082116	2.616667	0.518717

Çıkan sonuca göre

- (Sut, Kiyma) alan müşterilerin (Peynir) alma değeri 0.831169'dir. Bu da yaklaşık olarak %83'e takabül etmektedir.
- (Yumurta, Kiyma) alan müşterilerin (Peynir) alma değeri 0.809524'dir. Bu da yaklaşık olarak %81'e takabül etmektedir.



DİNLEDİĞİNİZ İÇİN TEŞEKKÜRLER