

K.S.Ü Bilgisayar Mühendisliği Bölümü

BM 301 Veri Madenciliği

YAZILIM RAPORU

A- YAZILIMIN TANITIMI

Programcının Adı-Soyadı : Hüseyin TAŞ 18110131021

Programcının Adı-Soyadı : Emirhan KOÇAK 17110131049

Geliştirme Ortamları :

Programlama Dili: Python

Programın Adı : Kaggle Notebook

1. Proje Adı : APRİORİ ALGORİTMASI İLE MARKET SEPET ANALİZİ

2. Projenin Amacı :

Yapacağımız bu projenin amacı müşterilerin satın aldıkları ürünler arasındaki ilişkiyi ortaya çıkarmak ve bu ilişkiye göre bir pazarlama stratejisi veya ürünlerin arasındaki raflarda düzenleme yapmaktır.

3. Giriş :

Bu projemizde bir çok kaynaktan ve özellikle Kaggle platformunda yazılmış olan diğer notebooklardan yararlandık. Bu çalışmayı bir marketteki alışverişlerde ürünlerin rastgele mi alınıyor yoksa ürünler arasında bir ilişki var mı ? bu soruyu yanıtlamak için yapıyoruz. Ve bu ürünler arasındaki ilişkilere göre bir pazarlama stratejisi geliştirmek. Bu çalışmamızda diğer çalışmalardan farklı olarak veri setine yeni bir özellik ekleyerek ürünlerin en çok ne zaman satıldığını (haftaiçi veya haftasonu) görselleştirdik. Diğer projelerden kendi projemizde kullanılan Apriori algoritması hakkında birkaç örnek görmek amacıyla diğer örnekleri de araştırdık.

4. Yöntem :

Bu projemizde izlediğimiz yol yöntem şu şekildedir:

1. İlk olarak kütüphaneler yüklendi ve veri setini yükledik.
2. Veri seti hakkında ufak bilgiler ortaya çıkardık.
3. Veri setine yani özellik(attribute) ekledik.
4. Veri seti üzerine görselleştirmeler yaptık.
5. Ve son olarak Apriori algoritmasını uygulayarak veri setindeki ürünler arasındaki ilişkiyi ortaya çıkardık.

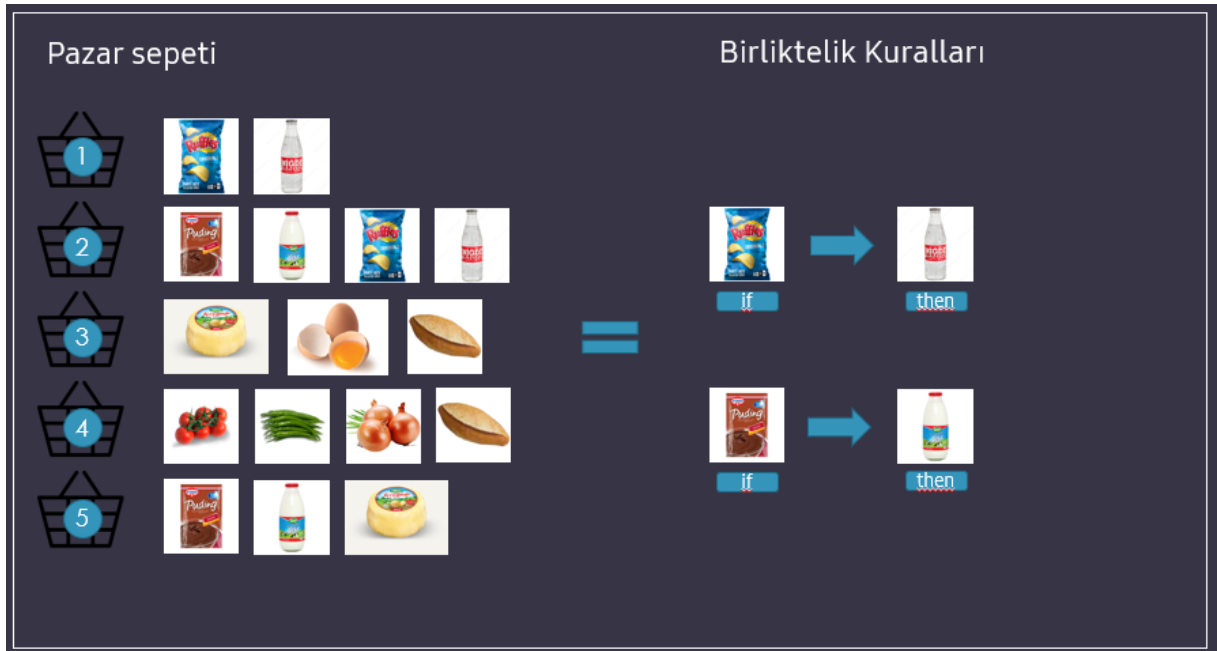
Projede kullanmış olduğumuz Apriori Algoritması hakkında kısa bilgi;

1. Birliktelik Kural Analizi (Association Rule Analysis)

Veri setinin içindeki alışverişler arasındaki ilişkiyi ortaya çıkarmak için kullanılan temel veri madenciliği tekniklerinden biridir.

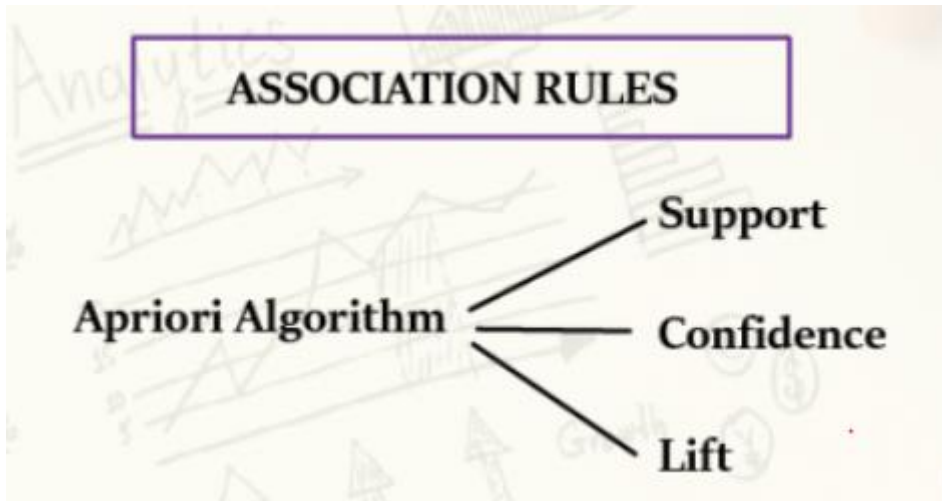
- Hangi ürünler ne sıklıkla satın alınmış?
- Bu ürün satın alındığında diğer ürünün satın alma olasılığı kaçtır?
- Bu ürün alındığında diğer ürünün satın alınma olasılığı % kaç artmıştır ?

Gibi sorulara cevap bulmuş oluyoruz.



2. APRIORİ ALGORİTMASI

Apriori algoritması veri setindeki öge kümelerini bulur ve ögeler arasındaki birliktelik kurallarını tanımlar.



Apriori algoritması 3 bileşeni vardır:

1. Destek (Support)
2. Güven (confidence)
3. Etki (Lift)

- Destek (Support) : Bir ürünün/ilişkinin tüm olaylar içerisinde görülme olasılığıdır.
N: Toplam alışverişlerin sayısı

$$\text{Support}(X \rightarrow Y) = \text{Frequency}(X,Y) / N$$

- Güven Değeri (Confidence) : X ürününün alan bir müşterinin Y ürününü alma olasılığıdır.

$$\text{Confidence}(X \rightarrow Y) = \text{Frequency}(X,Y) / \text{Frequency}(X)$$

- Lift : X olayının gerçekleşmesinin , Y olayının gerçekleşme ihtimalini kaç kat artırdığını gösterir.

$$\text{Lift} = \text{Support}(X \rightarrow Y) / (\text{Support}(X) * \text{Support}(Y))$$

2.1 Algoritmayı Projeye nasıl uyguladık ?

- Her bir ürünün frekansı hesaplanır. Yani ürünlerin alışveriş kayıtlarında görülme sıklığıdır. Daha sonra support değerleri hesaplanır.
- Belirlediğimiz min_support değerine eşit veya üstünde bir support değerine sahip ürünlerimiz ile yeni bir tablo oluşturulur. Min_support değerini kendimiz belirleyebiliriz.
- Daha sonra yeniden oluşturacağımız tabloda elimizdeki kalan tüm ürünlerin kombinasyonlarını barındırır.
- Oluşturduğumuz son tabloda kurallar çıkarımı bu adımda gerçekleştiririz.

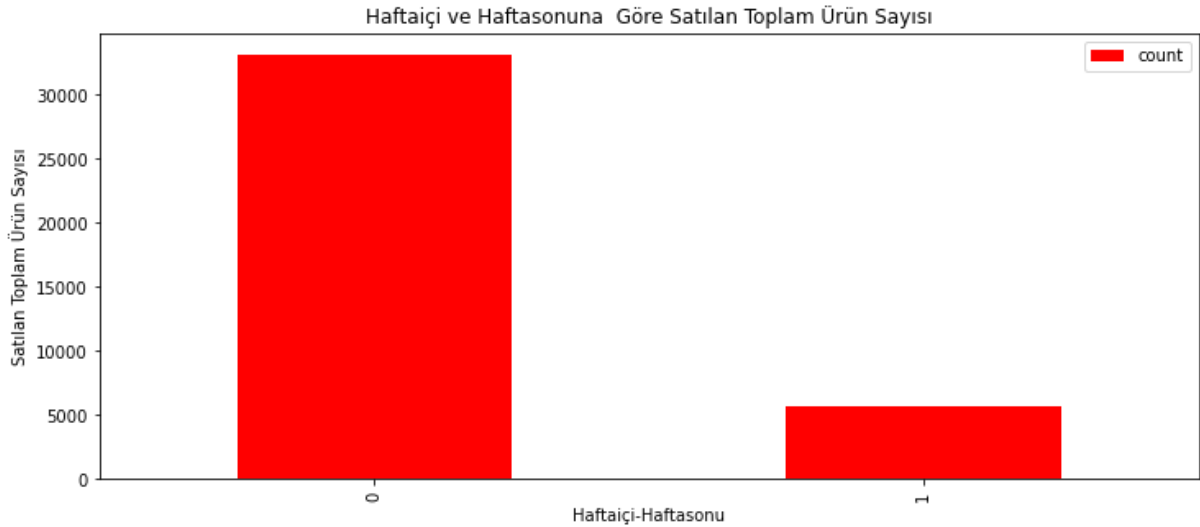
4.2 Verilerin Toplama Aracı:

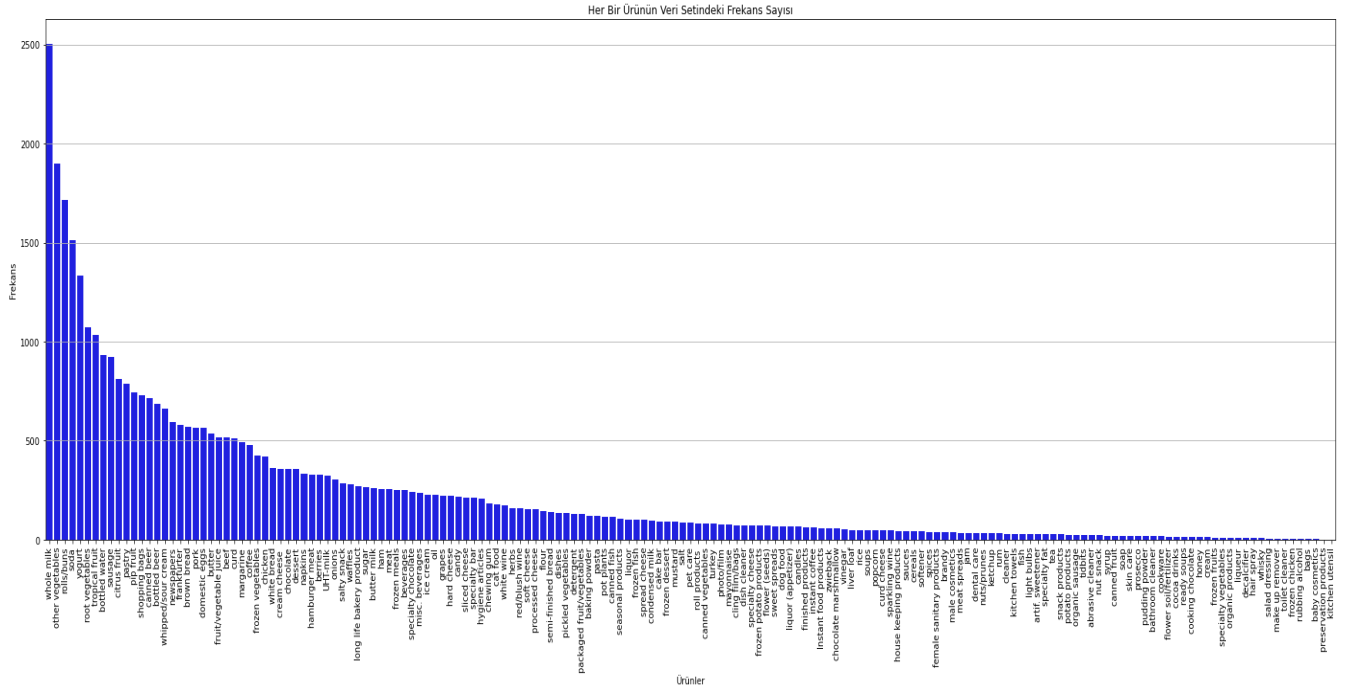
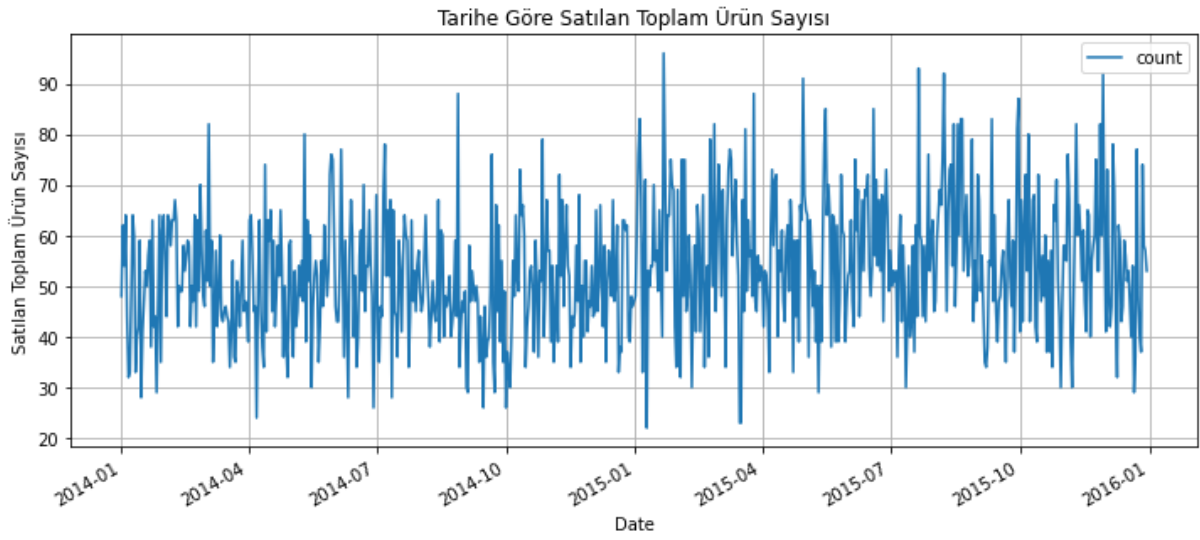
Veri setimizi Kaggle platformundan elde ettik. Veri seti 38765 gözlem ve 3 sütundan oluşmaktaydı.

Veri setinde üye numarası, tarih ve ürün açıklaması olmak üzere 3 değişkene sahip. Veri seti birkaç yıl içerisinde insanların yapmış olduğu alışverişlerinde almış olduğu ürünleri tutan bir veri seti (veri tabanı) diyebiliriz.

4.3 Verilerin Analizi :

Bu kısımda ise veri seti üzerinde analizler ve görselleştirmeler yaptık.





5. Sonuç ,Bulgular ve Yorumlar

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
1236	(sausage)	(whole milk, yogurt)	0.060349	0.011161	0.001470	0.024363	2.182917	0.000797	1.013532
1233	(whole milk, yogurt)	(sausage)	0.011161	0.060349	0.001470	0.131737	2.182917	0.000797	1.082219
1232	(whole milk, sausage)	(yogurt)	0.008955	0.085879	0.001470	0.164179	1.911760	0.000701	1.093681
1237	(yogurt)	(whole milk, sausage)	0.085879	0.008955	0.001470	0.017121	1.911760	0.000701	1.008307
474	(citrus fruit)	(specialty chocolate)	0.053131	0.015973	0.001403	0.026415	1.653762	0.000555	1.010726
475	(specialty chocolate)	(citrus fruit)	0.015973	0.053131	0.001403	0.087866	1.653762	0.000555	1.038081
1234	(sausage, yogurt)	(whole milk)	0.005748	0.157923	0.001470	0.255814	1.619866	0.000563	1.131541
1235	(whole milk)	(sausage, yogurt)	0.157923	0.005748	0.001470	0.009310	1.619866	0.000563	1.003596
618	(tropical fruit)	(flour)	0.067767	0.009757	0.001069	0.015779	1.617141	0.000408	1.006118
619	(flour)	(tropical fruit)	0.009757	0.067767	0.001069	0.109589	1.617141	0.000408	1.046969
90	(beverages)	(sausage)	0.016574	0.060349	0.001537	0.092742	1.536764	0.000537	1.035704
91	(sausage)	(beverages)	0.060349	0.016574	0.001537	0.025471	1.536764	0.000537	1.009129
1230	(sausage)	(whole milk, soda)	0.060349	0.011629	0.001069	0.017719	1.523708	0.000368	1.006200
1227	(whole milk, soda)	(sausage)	0.011629	0.060349	0.001069	0.091954	1.523708	0.000368	1.034806
846	(pastry)	(napkins)	0.051728	0.022121	0.001738	0.033592	1.518529	0.000593	1.011869
847	(napkins)	(pastry)	0.022121	0.051728	0.001738	0.078550	1.518529	0.000593	1.029109
1024	(root vegetables)	(processed cheese)	0.069572	0.010158	0.001069	0.015370	1.513019	0.000363	1.005293
1025	(processed cheese)	(root vegetables)	0.010158	0.069572	0.001069	0.105263	1.513019	0.000363	1.039891
750	(pip fruit)	(hard cheese)	0.049054	0.014703	0.001069	0.021798	1.482586	0.000348	1.007254
751	(hard cheese)	(pip fruit)	0.014703	0.049054	0.001069	0.072727	1.482586	0.000348	1.025530

- Sausage olan alışverişlerde {whole milk ,yogurt} satışını 2.18 kat artırmaktadır. Aynı şekilde {whole milk, yogurt} olan alışverişlerde sausage satışını 2.18 kat artırmaktadır.
- Citrus fruit olan alışverişlerde speciality chocolate satışını 1.65 kat artırmaktadır.
- Beverages olan alışverişlerde sausage satışını 1.53 kat artırmaktadır
- Pip fruit olan alışverişlerde hard cheese satışı 1.48 kat artırmaktadır.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
1234	(sausage, yogurt)	(whole milk)	0.005748	0.157923	0.001470	0.255814	1.619866	0.000563	1.131541
1210	(sausage, rolls/buns)	(whole milk)	0.005347	0.157923	0.001136	0.212500	1.345594	0.000292	1.069304
1228	(sausage, soda)	(whole milk)	0.005948	0.157923	0.001069	0.179775	1.138374	0.000130	1.026642
1109	(semi-finished bread)	(whole milk)	0.009490	0.157923	0.001671	0.176056	1.114825	0.000172	1.022008
1222	(yogurt, rolls/buns)	(whole milk)	0.007819	0.157923	0.001337	0.170940	1.082428	0.000102	1.015701
1232	(whole milk, sausage)	(yogurt)	0.008955	0.085879	0.001470	0.164179	1.911760	0.000701	1.093681
585	(detergent)	(whole milk)	0.008621	0.157923	0.001403	0.162791	1.030824	0.000042	1.005814
729	(ham)	(whole milk)	0.017109	0.157923	0.002740	0.160156	1.014142	0.000038	1.002659
149	(bottled beer)	(whole milk)	0.045312	0.157923	0.007151	0.157817	0.999330	-0.000005	0.999874
655	(frozen fish)	(whole milk)	0.006817	0.157923	0.001069	0.156863	0.993287	-0.000007	0.998743
315	(candy)	(whole milk)	0.014369	0.157923	0.002139	0.148837	0.942468	-0.000131	0.989326
1105	(sausage)	(whole milk)	0.060349	0.157923	0.008955	0.148394	0.939663	-0.000575	0.988811
907	(onions)	(whole milk)	0.020250	0.157923	0.002941	0.145215	0.919528	-0.000257	0.985133
1022	(processed cheese)	(rolls/buns)	0.010158	0.110005	0.001470	0.144737	1.315734	0.000353	1.040610
1027	(processed cheese)	(whole milk)	0.010158	0.157923	0.001470	0.144737	0.916503	-0.000134	0.984582
881	(newspapers)	(whole milk)	0.038896	0.157923	0.005614	0.144330	0.913926	-0.000529	0.984114
615	(domestic eggs)	(whole milk)	0.037091	0.157923	0.005280	0.142342	0.901341	-0.000578	0.981834
956	(packaged fruit/vegetables)	(rolls/buns)	0.008488	0.110005	0.001203	0.141732	1.288421	0.000269	1.036967
1036	(seasonal products)	(rolls/buns)	0.007084	0.110005	0.001002	0.141509	1.286395	0.000223	1.036698
375	(cat food)	(whole milk)	0.011829	0.157923	0.001671	0.141243	0.894379	-0.000197	0.980577

- {sausage,yogurt} alan müşterilerin %25'i whole milk'te almaktadır.
- {sausage,roll/buns} alan müşterilerin %21'i whole milk'te almaktadır.
- {semi-finished bread} alan müşterilerin %17'si whole milk'te almaktadır.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
952	(whole milk)	(other vegetables)	0.157923	0.122101	0.014837	0.093948	0.769430	-0.004446	0.968928
953	(other vegetables)	(whole milk)	0.122101	0.157923	0.014837	0.121511	0.769430	-0.004446	0.958551
1056	(whole milk)	(rolls/buns)	0.157923	0.110005	0.013968	0.088447	0.804028	-0.003404	0.976350
1057	(rolls/buns)	(whole milk)	0.110005	0.157923	0.013968	0.126974	0.804028	-0.003404	0.964550
1137	(soda)	(whole milk)	0.097106	0.157923	0.011629	0.119752	0.758296	-0.003707	0.956636
1136	(whole milk)	(soda)	0.157923	0.097106	0.011629	0.073635	0.758296	-0.003707	0.974663
1183	(yogurt)	(whole milk)	0.085879	0.157923	0.011161	0.129961	0.822940	-0.002401	0.967861
1182	(whole milk)	(yogurt)	0.157923	0.085879	0.011161	0.070673	0.822940	-0.002401	0.983638
921	(rolls/buns)	(other vegetables)	0.110005	0.122101	0.010559	0.095990	0.786154	-0.002872	0.971117
920	(other vegetables)	(rolls/buns)	0.122101	0.110005	0.010559	0.086481	0.786154	-0.002872	0.974249

- Whole milk ve {other vegetables} alışveriş sepetlerinin %0.014'inde beraber gözlenmiştir.
- {Rool/buns} ve {other vegetables} alışveriş sepetlerinin %0.010'unda beraber gözlenmiştir.

Apriori algoritması sayesinde veriler üzerinde istatistikler yöntemler ile veriler arasındaki ilişkiyi ortaya koyabilir ve bu çıkarımlar sonucunda ileriye dönük tahminler ve kararlar alabiliriz.

KAYNAKLAR

- <https://www.kaggle.com/heeraldedhia/groceries-dataset>
- <https://medium.com/@cotur/apriori-algoritmas%C4%B1-a90a38856373>
- <https://medium.com/deep-learning-turkiye/https-medium-com-deep-learning-turkiye-apriori-algoritmasi-veri-madenciligi-dae70b0bc1>
- <https://www.veribilimiokulu.com/blog/associationrulesanalysis/>
- https://www.btkakademi.gov.tr/portal/course/player/python-ile-makine-ogrenmesi-11800?language=tr&token=eyJhbGciOiJIUzUxMiJ9.eyJzdWIiOiI0ODQ1MzEiLCJkb21haW4iOiJCVESiLCJpc3MiOiJLQilIm9pZCI6NTEsImV4cCI6MTYwOTUyNTQwNiwiZXVpZCI6IjU4OGFiYWE2LTUyMmEtNDQ5ZS1iMzdhLTFlkNWU0ODcxNmU5MSIsImV4cCI6MTYwOTQzOTAwNiwiianRpIjoibjOTBjOTIzMjEtMTg5MS00Njg2LTJhN2QtOWU1OTY3OTcyYmY1In0.ctsd06dd_zqJFr2AXQIQ3eo27Si-e2ThAXq3F4E2ZL5jflWzHiTTb1y_JHfcEa3wnWFNL-XEmwenZrSHx9LwwGPwnxjgMQWFEQRA67hlyXdqT3SqAufrbItYwv-dqk26WEtlErHnInYwOQZ8gOplhyHMC4X47p0wA1QIWdL1OPWcQvfq_0LZJPxXgBWB2xAzJ6HNZu4oWGAUjCcluETC4I0er4adM9hXRqphPX4HcU2jf1HOgdGY_nP0cuz5zXmdhehiryBwYAlQRmIaoKB9EEqVfX5lGweQ8g4GvT1rNeAFKEGPdMkgaSV5C9Npq010wDgoYiXZXRP0lZfcU3aw&examMode=false
- <https://medium.com/@hacakirdogan/apri%CC%87ori%CC%87-algori%CC%87tmasi-i%CC%87le-pazar-sepeti%CC%87-anali%CC%87zi%CC%87-9cdb8479bf96>