

Veri Ayırıştırması ile Kampanya Verimliliği Ölçme

(Measuring Campaign Efficiency with Data Disaggregation)

Selim SÜLEYMANOĞLU¹

1. OptiWisdom Inc. ve Antalya Bilim Üniversitesi, Elektrik ve Bilgisayar Mühendisliği

Özet

Veri birleştirme ve veri ayırıştırma şirketlerin pazarlama ve satışla ilgili sorularını cevaplayabilmek için sıklıkla kullanılan yöntemlerdir. Bu çalışmada fiziksel bir işyerinden (bina, dükkan vb.) mal satan hayali bir şirketin satışlarıyla ilgili bir veri seti üzerinde Python programlama dili kullanarak veri işlemesi yapıp şirketin kampanyalarının verimliliği ölçülmeye çalışılmıştır. Veri setinde satış detayları ve kuponlar aracılığıyla bazı müşterilere belli ürünlerde indirim hakkı sağlayan kampanya bilgileri yer alır. İlk adımda veriyi incelemek, örüntüleri ve anomalileri tespit etmek için keşifçi veri çözümlemesi yapılmıştır. Daha sonra verinin temsil ettiği bilgiyi daha iyi anlamak için farklı regresyon yöntemleri (doğrusal regresyon, polinomsal regresyon, destek vektör regresyonu vb.) karşılaştırmalı olarak denenmiştir. Bu deneysel yöntem sonucunda doğrusal regresyon tercih edilmiştir. Son aşamada ise sonuçlar uygun Python kütüphaneleri kullanılarak grafiğe dökülmüş ve okurun daha rahat anlayabilmesi için yorumlanmıştır.

Anahtar Kavramlar: Veri Birleştirme, Veri Ayırıştırma, Veri Bilimi, Kampanya Verimliliği, Doğrusal Regresyon

Abstract

Data aggregation and data disaggregation are widely used in data science to solve common marketing and sales problems. In this study, we used Python programming language to apply data processing on a given dataset about the sales of an imaginary Brick & Mortar retailer for measuring the efficiency of its campaigns. The dataset includes transaction details and information about the campaigns, each of which gives some certain customers the opportunity to get a discount on certain items via coupons. First, exploratory data analysis is used to investigate the data, discover patterns and spot anomalies. Then, in order to get a meaningful reading on what the data represents we compared some regression methods such as linear regression, polynomial regression, and support vector regression. Linear regression is the approach we chose. Lastly, we plotted the results with the help of appropriate Python libraries and commented on them to make it more understandable for the reader.

Keywords: Data Aggregation, Data Disaggregation, Data Science, Campaign Efficiency, Linear Regression

1. Giriş

Bu bölümde probleme ait en son çözümleri ve algoritmaları içeren (state-of-art) literatür bilgisi paylaşılacak ve problem tanımlanacaktır.

Zamansal sinyaller üzerinde ayırıştırma (disaggregation) kavramı istatistiksel bir verinin değerinin yardımcı bilgileri aracılığıyla kaba ve geniş bir uzaysal seviyeden daha düşük bir uzaysal seviyeye naklidir [1].

Garrett [9] ekonomi alanındaki çalışmasında bireysel davranışın açıklanmasında toparlanmış (aggregated) verinin kullanılmasının bireylerin ekonomik davranışlarıyla ilgili yanlış yönlendirici çıkarımlarla sonuçlanabileceğini iddia etmiştir.

Verma ve Anvar [10], kesintisiz elektrik yükü analizi – Nonintrusive load monitoring (NILM) ile kullanarak bir evdeki elektrikli eşyaların hangilerinin daha çok yük çektiğini tespit etmek için veri ayırıştırması yapmışlardır.

Almutairi ve arkadaşları [11], çalışmalarında birden çok geniş ve kaba bakış açısından ve farklı boyutlar üzerinden toparlanmış ince ölçekli veriyi yeniden inşa edecek bir yöntem geliştirmeyi amaçlamışlardır.

Schirmer ve arkadaşları [13], enerji veri ayırıştırması konulu çalışmalarında 5 farklı algoritma (K-en yakın komşu, destek vektör makineleri, derin sinir ağları ve rastgele orman) denemiş ve rastgele ormanı en başarılı bulmuşlardır.

Arambepola ve arkadaşları [14], farklı ortamlarda simüle edilmiş veriler üzerinde ayırıştırma için regresyon uygulamışlar ve elde ettikleri ince ölçekli tahminleri simüle edilmiş kesin referanslarla karşılaştırarak boyutsal hastalık haritalandırması yapmışlardır. Toparlanmış veride çapraz geçerlilik korelasyonunun ince ölçekli tahmin performansı için kabul edilebilir olacağını bulmuşlardır.

Chamberlin [15], ekonomi istatistiğinde zaman serileri üzerinde veri ayırıştırmasının kullanım alanlarını örneklendirdiği çalışmasında başarıyı ölçmek için Fernandez Rastgele Yürüyüş Modeli, Litterman Rastgele Yürüyüş Markov Modeli, AR(1) Modeli gibi seçenekler sunmuştur. Regresyon yaklaşımının temel avantajlarından birinin bazı indikatör serilerinin ayırıştırılmış zaman serisinde kısa dönem hareketlerin tespitinde kullanılabilmesi olduğunu belirtmiştir.

Bütün bu kaynaklardan anlaşılmıştır ki çok boyutlu veri yaygınlaşmıştır ve sıklıkla karşılaşılan bir fenomen haline gelmiştir. Veri toplaması zaman, konum vb. farklı öznitelikler üzerinden yapılabilir [11]. Mesela ürün satışları zaman boyutu üzerinden ve mağazaların konumu üzerinden toparlanmış olabilir. Ancak veri madenciliği ve makine öğrenmesi modelleri tahlil ve tahmin için kişiselleştirilmiş yahut özelleştirilmiş detaylı veriden büyük fayda görür. Bu yüzden veri ayırıştırma algoritmalarının önemi kamu yönetimi [3], eğitim [4], tarım [5], beşerî coğrafya [6], enerji araştırmaları [7] gibi değişik bilimsel çalışma alanlarında artmaktadır.

Dolgun ve Ersel [16], bankacılık sektöründe doğrudan pazarlama kampanyalarının nasıl yönlendirilebileceği üzerinde durmuşlar, “dengesizlik” problemini incelemiştir. Çalışmanın sonucunda, genel başarı ölçütüne göre sırasıyla, SVM linear, lojistik regresyon ve SVM RBF, F ölçütüne göre sırasıyla, Lojistik regresyon, SVM RBF ve CHAID ve matthews korelasyon katsayısına göre sırasıyla, SVM linear, lojistik regresyon ve CHAID yöntemleri en başarılı yöntemler olarak tespit edilmiştir.

Bu çalışma kapsamında, hayali ABC firmasının belli tarihlerdeki kampanyalarının satışa etkisi araştırılıp kaggle.com’da bulunan Predicting Coupon Redemption_PCA veri kümesindeki [2] train.csv dosyasında yer alan 18 indirim kampanyasının toplam gelire etkisinin birbiriyle kıyaslamasının nasıl yapıldığı anlatılacaktır. Kampanyaların her biri kapsamında müşterilere belli ürünlerde indirim almalarını sağlayan kuponlar verilmiştir. Her bir kuponun indirim sağladığı ürünler listesi birbirinden farklıdır.

Müşteriler alışveriş için bu kuponları kullanmak zorunda değildir, customer_transaction_data.csv dosyasında verileri bulunan her alışverişte ABC firması bir gelir elde etmiştir ve bu gelir hesaplanırken kampanyalarda tanınan indirimlerden faydalanılmış ya da faydalanılmamış tüm alışverişlerin gelirleri her bir gün toplanarak o günkü toplam gelir bulunmuştur. Veri bilimi araçları kullanılarak her bir kampanyanın aktif olduğu zaman dilimi boyuca gelire etkisi skorlanmıştır.

Probleminiz kaggle.com’da veri kümesini paylaşan vasudeva mahlaslı kullanıcının probleminden farklıdır. O problemde geçmişteki kampanya bilgilerinden gelecekteki 10 kampanyayla ilgili tahmin yapılması istenmiştir. Bizim

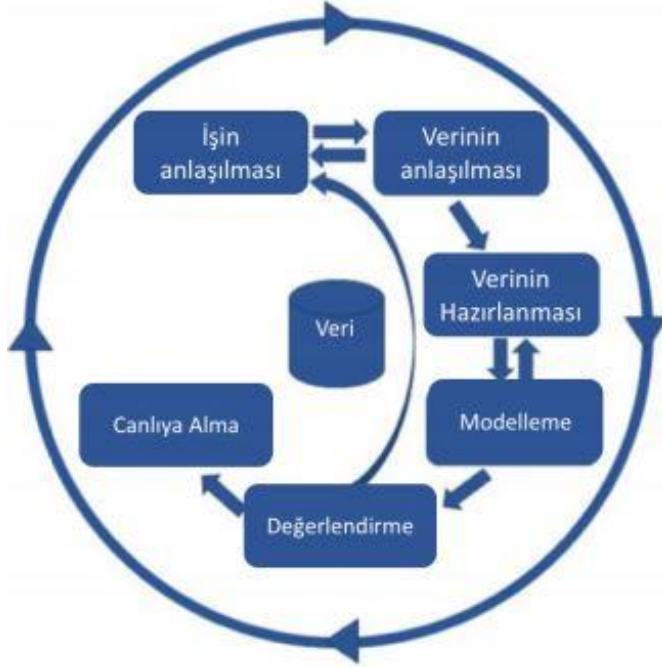
problemimiz ise geçmiş kampanyaların performanslarının birbirleriyle karşılaştırılmasıdır ve literatür taramasında problemimizin daha önce çözülmeye çalışıldığı görülmemiştir.

2018 yılında tüm dünyada reklamlara yapılan toplam harcama yaklaşık 617.32 milyar dolardır [19]. 2019 yılında şirketlerin tüm dünyada reklama yaptıkları bütünlük harcamalara bakıldığında ilk sıradaki Procter & Gamble'in 12.2 milyar dolar, ikinci sıradaki Amazon'un ve üçüncü sıradaki L'Oréal'in 6.7'şer milyar dolar harcadıkları görülebilir [18]. Purba'ya [20] göre reklamın amacı insanları ürünü satın almaya ikna etmektir. Bu reklam maliyetleri düşürülürse şirketler artan bütçeyi daha etkin ve etkili kullanabileceklerdir. Böylece toplumların ve uygarlıkların medeniyet seviyelerini artırmak için değerlendirilebilecek fırsatlar artacak, çevre kirliliği azalacak, indirim kampanyalarını reklamlarla duyurmak için harcanan vaktin düşmesiyle çalışanlar daha işe yarar uğraşlar (örneğin dünya çapında kadın haklarının erkek haklarıyla aynı seviyeye getirilmesi) için efor sarf edebileceklerdir.

Problemimizin çözümünün geçmiş kampanyalarının performanslarını kıyaslamak isteyen şirketlere kılavuz olmasını umuyoruz. Bu sayede şirketler indirim kampanyaları için daha az ve öz harcayabilecek, maliyetlerini düşürebilecek, zaman ve enerjiden tasarruf sağlayarak çevreye de katkıda bulunabileceklerdir. Buldukları toplumlarda genç nesillerin eğitime daha çok zaman ayırabileceklerdir.

2. Metodoloji

Uygulamada izlenen adımların daha sade ve anlaşılır olabilmesi için CRISP-DM: Endüstriler Arası Standart İşleme – Veri Madenciliği için (Cross Industry Standard Processing – Data Mining) metodolojisine uyulmuştur. CRISP-DM 6 aşamadan oluşan bir süreçtir [8]. Bu aşamalar Şekil 1'de özetlenmiştir.



Şekil 1: CRISP-DM Adımları ve Akışı

- İş süreçlerinin anlaşılması: Bu aşamada problemin tanımı yapılır. Problemimizin tanımı 3. kısımda yapılmıştır.
- Verinin anlaşılması: Bu aşamada probleme uygun olarak veri toplanır veya mevcut verinin üzerinden geçilir. Çalışmamızın bu aşaması 4. kısımda anlatılmıştır.

- iii. Veri Ön İşleme aşaması: Bu aşamada veri üzerinde yapılacak işlemlere ve bu işlemlerin hangi yöntemlerle yapılacağına karar verilir. Çalışmamızın bu aşaması 4. kısımda anlatılmıştır
- iv. Model aşaması: Bu aşamada tanımlanan problem ve veri kaynakları üzerinde bir makine öğrenmesi veya istatistiksel model geliştirilir. Çalışmamızın bu aşaması 4. kısımda anlatılmıştır.
- v. Değerlendirme aşaması: Bu aşamada, şimdiye kadar olan adımların genel bir değerlendirmesi yapılır ve aslında ilk adımda (problem tanımı aşaması) konulan başarı kriterlerini ne ölçüde sağladığı test edilir. Çalışmamızın bu aşaması 5. kısımda anlatılmıştır. Problemimizde kampanyaları aralarında kıyaslamak için hangi regresyon modelinin seçileceğine karar verilirken Şekil 1'deki gibi İşin Anlaşılması aşamasına tekrar dönülmesi birkaç kez uygulanmıştır.
- vi. Ürün aşaması: Çalışmamızın sonunda yazmış olduğumuz Python programı bu gibi problemlere çözüm sunacak bir ürün olarak düşünülebilir.

3. Veri Kümesinin Tanımı ve Özellikleri (İş Sürecinin Anlaşılması)

Veri kümemiz kaggle.com'da bulunan Predicting Coupon Redemption_PCA veri kümesinde yer alan iki dosyadan oluşur.

Veri kümesinin içeriğini veren iki dosyadan ilki olan customer_transaction_data.csv dosyasındaki her bir satırda bir ürünün satışıyla ilgili bilgiler yer alır. date kolonu altında satışın yapıldığı tarih, customer_id kolonu altında ürünü alan müşterinin ID numarası, item_id kolonu altında ürünün ID numarası, quantity kolonu altında üründen kaç kalem alındığı, selling_price kolonu altında ürün için ödenen miktar, other_discount kolonu altında incelenen kampanyalardan kaynaklanmayan indirimler, coupon_discount kolonu altında ise kampanyalardan kaynaklanan indirimler yer alır.

Veri kümesinin içeriğini veren iki dosyadan ikincisi olan campaign_data.csv dosyasındaki her bir satırda birer kampanyayla ilgili bilgiler yer alır. campaign_id kolonu kampanyaya ait ID numarasını, campaign_type kolonu kampanyanın türünü start_date ve end_date kolonları ise sırasıyla kampanyanın başlangıç ve bitiş tarihlerini gösterir.

Problemimiz bu veriyi kullanarak 18 kampanyanın hangilerinin gelire en olumlu, hangilerinin en olumsuz etki yaptığı sorusudur.

4. Uygulama

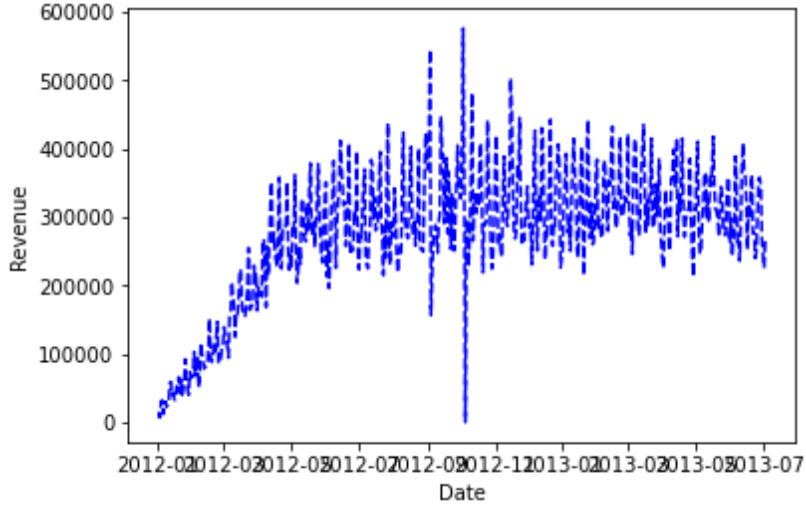
Bu bölümde problemi çözmek için takip edilen adımlar anlatılacaktır. Kod için Python 3.8.5 kullanılmıştır. Faydalanılan kütüphaneler şunlardır:

- i. Datetime: Tarih bilgilerini işleyebilmek için kullanılmıştır.
- ii. Numpy: Verilerimizi dizin (array) ve liste şeklinde toparlayıp ihtiyaca göre manipüle etmemizi sağlamıştır.
- iii. Pandas: Veri çerçeveleri için kullanılmıştır.
- iv. Matplotlib: Grafiklerin çizimi ve manipülasyonu için kullanılmıştır.
- v. Sklearn: Regresyon modelini çalıştırabilmek için kullanılmıştır

4.1 Keşifçi Veri Analizi EDA (Exploratory Data Analysis) ve Veriyi Toparlama (Aggregation)

Veri, 1.324.566 satırdan oluşmaktadır ve her satırda bir satışın (transaction) bilgileri yer alır. Verinin kolay işlenmesi amacıyla benzer özellikte olan satırların bir araya getirilmesi planlanmıştır. Böylece verilerin gruplar halinde birlikte işlenebilmesi hedeflenmiştir.

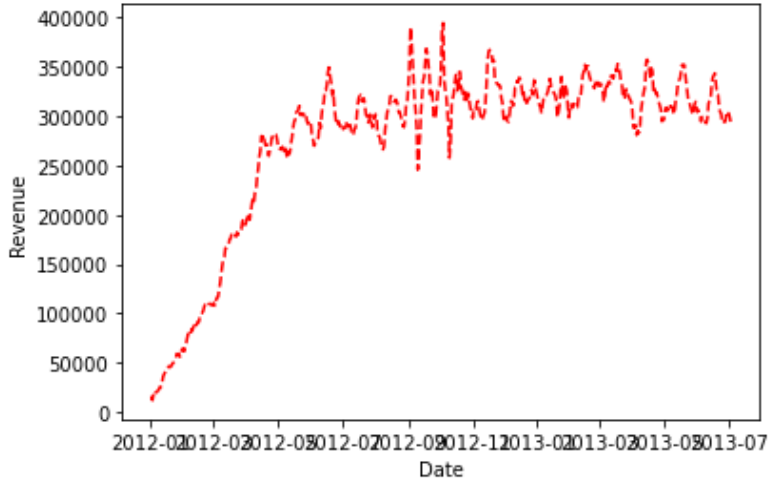
Her bir satışın günü ve gelirini elde etmek için `customer_transaction_data.csv` dosyasının `date` ve `selling_price` kolonları kullanılmıştır. Aynı gün yapılan satışların toplamı `.groupby()` ve `.sum()` yöntemleri ile bir araya getirilmiş, böylece 1.324.566 satırlık veri 549 satıra indirgenmiştir. Bu 549 günün gelirleri Şekil 2’de gösterilmiştir. 5 Ekim 2012 801,11 ile en düşük gelirin elde edildiği gündür. 3 Ekim 2012 ise 575.967 ile en yüksek gelirin elde edildiği gündür. Veri kümesinde kayıp veri yoktur.



Şekil 2: 549 Günlük Gelir Bilgisi.

4.2 Veri Ön İşleme ve Tüm Dönemlerdeki Gelirin Hareketli Ortalaması

Hareketli ortalama, bir zaman serisinden belirli bir sayıdaki ardışık değerlerin ortalamasını alarak oluşturulan yeni bir zaman serisidir [12]. Varyansı yüksek verileri daha düz (smooth) hale getirmeye yarar. Şekil 2’de gösterilen verilerin daha düz hale getirilmesi için 7 günlük hareketli ortalaması alınıp sonucu Şekil 3’te gösterilmiştir.



Şekil 3: 549 Günlük Gelir Bilgisinin Hareketli Ortalaması (window=7, min_periods=1).

İnceleyeceğimiz 18 kampanya başlamadan önceki dönemde günlük ortalama gelir 106.975,88, kampanyaların olduğu dönemde ise günlük ortalama gelir 319.955,73 bulunmuştur. Buna göre kampanyalar boyunca günlük ortalama gelir kampanyasız dönemdekine göre yaklaşık yüzde 50 yüksektir.

4.3 Her Bir Kampanyanın Aktif Olduğu Günlerin Ayrıştırılması

Bu kampanyaların gelire etkilerinin kendi aralarında mukayesesi için sadece kampanyaların aktif olduğu dönemin incelenmesine karar verilmiştir. Bunun için 549 günlük veriden kampanyaların bulunduğu 326 günlük veri çekilmiştir. Hangi gün hangi kampanyaların aktif olduğu, her bir günkü gelirin kampanyasız dönemin ortalamasından farkı ve o gün için hareketli ortalama ile elde edilen gelir bilgileri, pandas ve numpy kütüphaneleri kullanılarak `campaigns_by_day_id_binary` isimli tek bir veri çerçevesinde toplanmıştır. Tablo 1, bu veri çerçevesinin ilk 20 satırını yani kampanyaların aktif olduğu ilk 20 günü gösterir. 1, kampanyanın o gün aktif olduğunu, 0 (sıfır) ise kampanyanın o gün aktif olmadığını belirtir. Bunu yapmamızın amacı bir sonraki adımda doğrusal regresyon uygulandığında çıkacak katsayıların aktif olmayan kampanyalar için 0 gelmesini ve aktif olan kampanyaların katkılarının hesaplanmasını sağlamaktır.

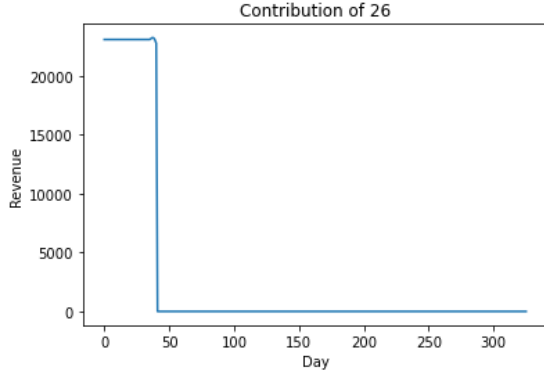
Tablo 1: İlk 20 gün için kampanyalarla ilgili aktiflik ve gelir bilgisi.

Index	date	13	11	12	10	9	8	7	6	3	5	4	1	2	30	29	28	27	26	extra_revenue_that_day	revenue_rolling_avg
0	2012-08-12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	69737.4	303175
1	2012-08-13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	66900.1	307713
2	2012-08-14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	56130.7	309572
3	2012-08-15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	104211	319913
4	2012-08-16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	149961	320204
5	2012-08-17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	189752	317214
6	2012-08-18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	80735.6	315469
7	2012-08-19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	60162	314102
8	2012-08-20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	85290.5	316729
9	2012-08-21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	45410.2	315197
10	2012-08-22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	74824.9	310999
11	2012-08-23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	99289.8	303761
12	2012-08-24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	188587	303594
13	2012-08-25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	50320.4	299249
14	2012-08-26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	39458	296291
15	2012-08-27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	40472.8	289889
16	2012-08-28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	36903.7	288674
17	2012-08-29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	131491	296769
18	2012-08-30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	210298	312627
19	2012-08-31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	198714	314074
20	2012-09-01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	157211	329344

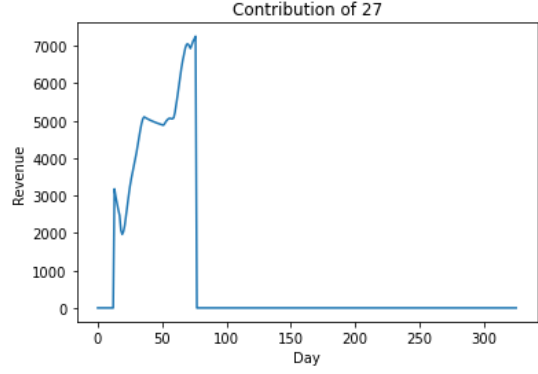
4.4 Kampanyaların Verimliliğin Ölçülmesi ve Grafikleri

326 satırlık bu veri çerçevesinden 60'ar satırlık, başlangıç günleri birer gün arayla artan yeni veri çerçeveleri çıkarılmıştır. Öyle ki, yeni oluşan veri çerçevelerinden ilki `campaigns_by_day_id_binary`'nin ilk 60 satırını (`index=0`'dan `index=59`'a kadar), ikincisi 2. satırdan 61. satıra kadarını (`index=1`'den `index=60`'a kadar), 267. yani sonuncusu ise `campaigns_by_day_id_binary`'nin son 60 satırını içermiştir. Yeni oluşan 267 veri çerçevesinin her biri için 18 kampanya girdi, hareketli ortalama ile elde edilen gelir ise çıktı olacak şekilde doğrusal regresyon uygulanmıştır. Böylece her doğrusal regresyonun sonucunda 18'er katsayı elde edilmiştir. Bu katsayılardan her biri sırasıyla denk geldiği kampanyanın kolonu (içeriği 1 ve 0'lardan oluşan) ile çarpılarak bu içerikler güncellenmiştir.

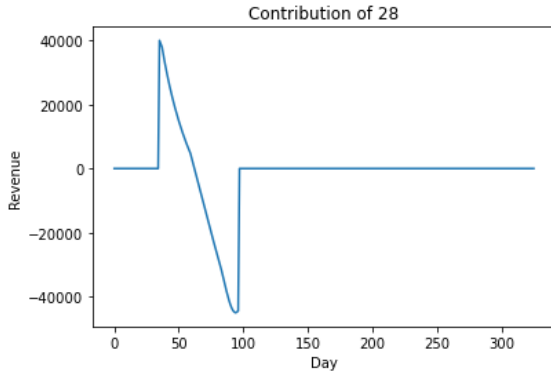
Elde edilen 60x18'lik güncel veri çerçeveleri art arda eklenerek 16020x18'lik bir veri çerçevesi oluşturulmuştur. Bu yeni veri çerçevesi içinden index'i aynı olan satırlar ortalamaları alınıp tek satır haline getirilmiştir. Böylece her bir kampanyanın 326 günün her birindeki katkısının görüldüğü 326x18'lik bir veri çerçevesi elde edilmiştir. Şekil 4'ten Şekil 21'e kadarki grafikler her bir kampanyanın etkisini göstermektedir.



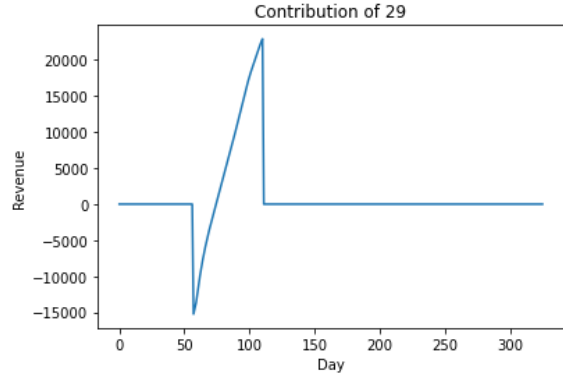
Şekil 4: 26 ID'li kampanya'nın etkisi



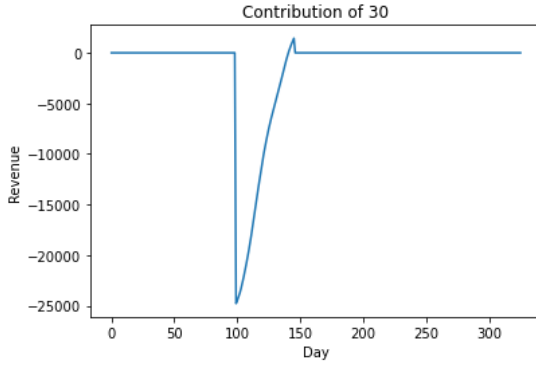
Şekil 5: 27 ID'li kampanya'nın etkisi



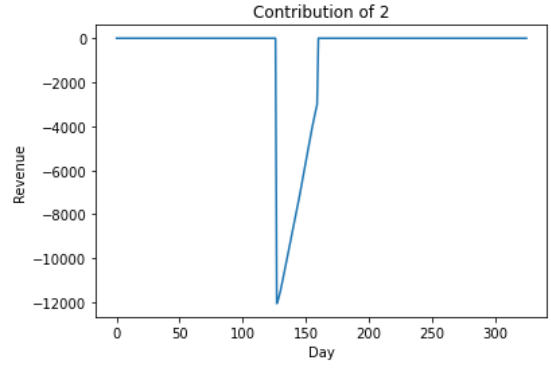
Şekil 6: 28 ID'li kampanya'nın etkisi



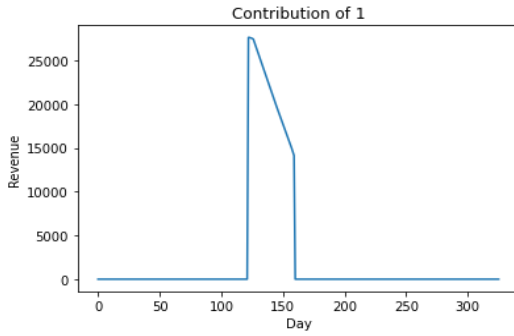
Şekil 7: 29 ID'li kampanya'nın etkisi



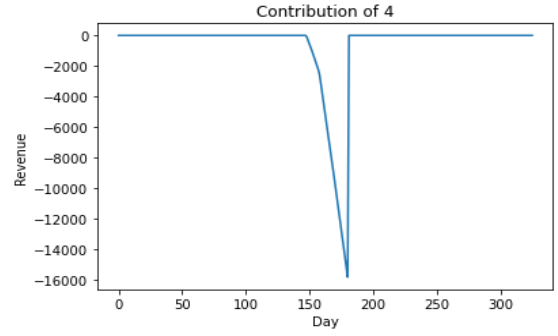
Şekil 8: 30 ID'li kampanya'nın etkisi



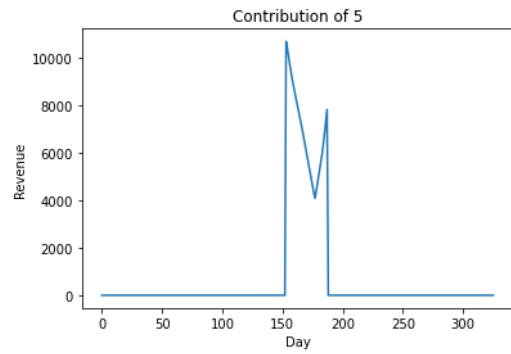
Şekil 9: 2 ID'li kampanya'nın etkisi



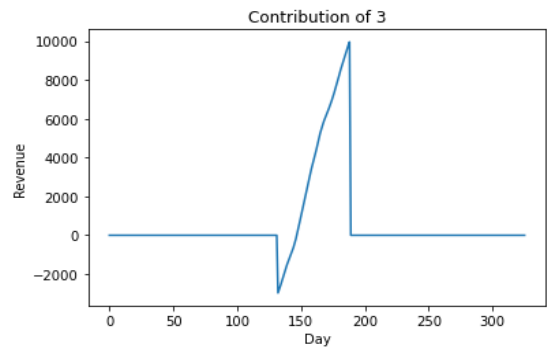
Şekil 10: 1 ID'li kampanya'nın etkisi



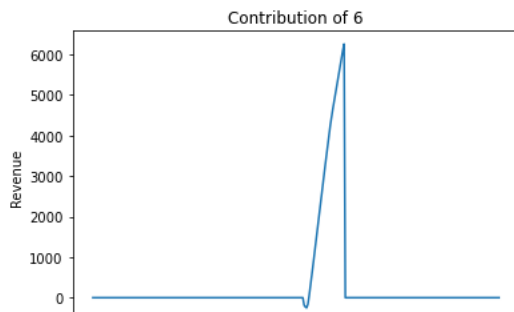
Şekil 11: 4 ID'li kampanya'nın etkisi



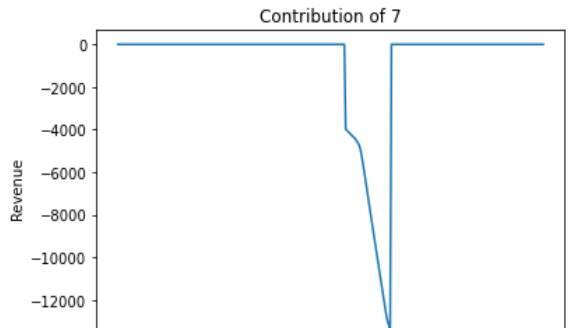
Şekil 12: 5 ID'li kampanya'nın etkisi



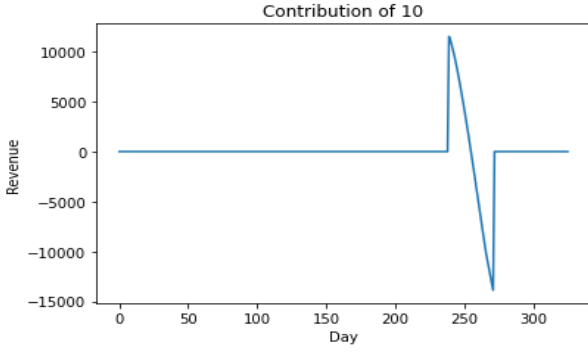
Şekil 13: 3 ID'li kampanya'nın etkisi



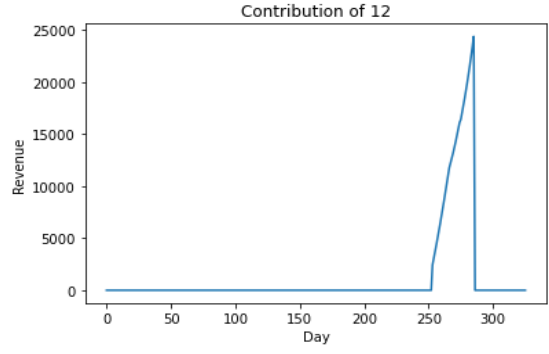
Şekil 14: 6 ID'li kampanya'nın etkisi



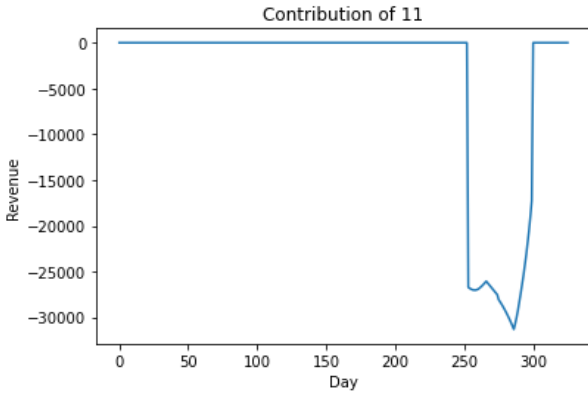
Şekil 15: 7 ID'li kampanya'nın etkisi



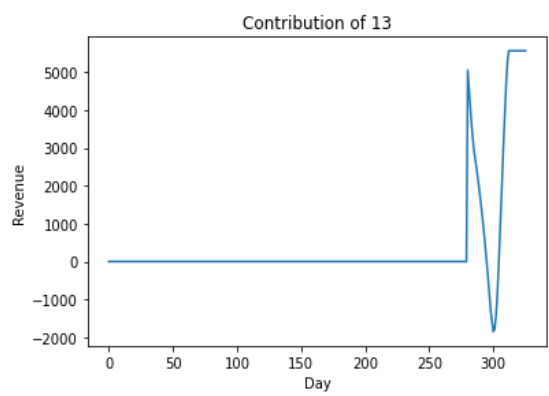
Şekil 18: 10 ID'li kampanya'nın etkisi



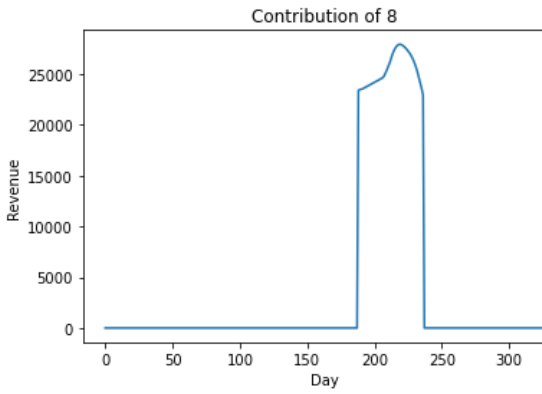
Şekil 19: 12 ID'li kampanya'nın etkisi



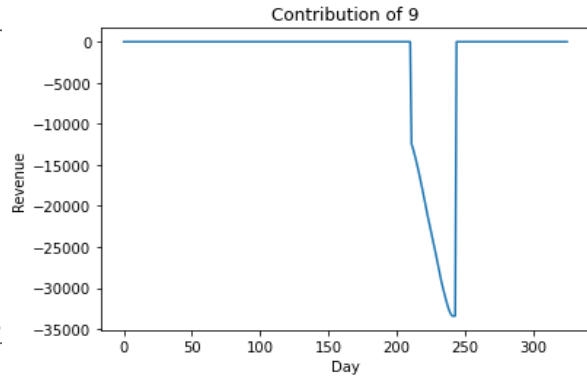
Şekil 20: 11 ID'li kampanya'nın etkisi



Şekil 21: 13 ID'li kampanya'nın etkisi



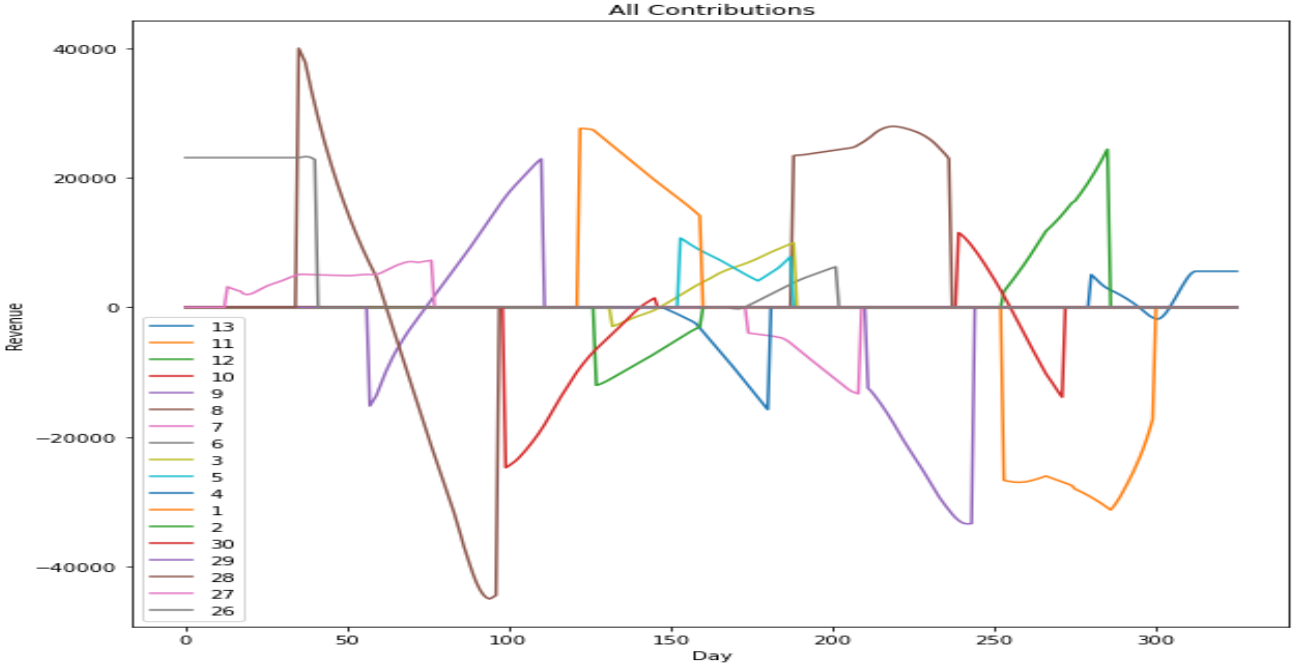
Şekil 16: 8 ID'li kampanya'nın etkisi



Şekil 17: 9 ID'li kampanya'nın etkisi

4.5 Tüm Kampanyaların Verimliliklerinin Birlikte Gösterimi

Bu aşamaya kadar yapılan bütün kampanya analizlerinin tek bir grafik üzerinde görselleştirilmesi mümkündür. Bu grafik Şekil 22'de çizilmiştir.



Şekil 22: Tüm kampanyaların etkilerinin birlikte gösterimi

Şekil 22'de düşeydeki 0 değeri bir kampanyanın aktif olduğu zamanda gelire etkisinin olmadığını, pozitif değerler kampanyanın gelire pozitif etkisi olduğunu, negatif değerler kampanyanın gelire negatif etkisi olduğunu gösterir. Ayrıca, Şekil 22'de görülmeyen bir taban (base) satış değeri de bulunmaktadır. Bu taban satış değeri hiçbir kampanya olmasaydı elde edilecek olan satış değeridir ve Şekil 22'de bu taban değeri 0 kabul edilmiştir.

4.6 Değerlendirme

Şekil 22'de tüm kampanyaların etkileri bir arada gösterilmektedir. Buna göre ID'si 26, 1 ve 8 olan kampanyalar diğer kampanyalara kıyasla en başarılı, ID'si 9 ve 11 olan kampanyalar ise diğer kampanyalara kıyasla en başarısız kampanyalardır.

Literatürde bu tip kampanya başarısı ölçmek için SVM linear, SVM RBF, lojistik regresyon, CHAID [16], eş-zamanlı denklemler regresyon modeli [17] gibi yöntemlerden faydalanılmıştır. Bir yöntemin kesinlikle diğerlerinden başarılı olacağı söylenemez. Çözülecek probleme en uygun yöntemin ve modelin bulunmasında fayda vardır.

5. Sonuç

Bu makalede, hayali ABC şirketinin ilk 549 günlük satış verileri incelenmiştir. 18 kampanyanın aktif olduğu 326 günlük dönemde aktif olmadıkları 223 günlük döneme göre gelirin yaklaşık yüzde 50 arttığı tespit edilmiştir. Kampanyaların aktif olduğu 326 günlük dönemde her bir kampanyanın diğerlerine kıyasla ne derece etkili olduğu anlaşılmasına çalışılmıştır.

Bu amaçla doğrusal regresyon yöntemi kullanılarak veri, kampanya etkilerine ayrııştırılmıştır. Her bir gün için her kampanyanın gelire etkisi hesaplanmıştır. Bunun için 60 günlük pencere büyüklüğü olacak şekilde başlangıç tarihi

1'er gün kaydırılarak 326 günlük kampanya performanslarının hareketli ortalaması alınmıştır. Bu hareketli ortalamalar grafiğe dökülmüş ve böylece her bir kampanyanın diğer kampanyalara kıyasla verimliliği gözlemlenebilmiştir.

Veri ayırıştırma yöntemleri günümüzde yapay zeka ile müşteri-satıcı ilişkilerini, şirketlerin belli kriterlere göre performanslarını ölçmede sıklıkla kullanılmaktadır. Alandaki gelişmeler bu ölçümlerin büyük verilerde daha hızlı ve daha hızlı yapılabilmesi için yolun açık olduğuna işaret etmektedir.

Kaynakça

- [1] Merz, B., Kreibich, H., Schwarze, R., & Thielen, A. (2010). Review article "Assessment of economic flood damage". *Natural Hazards and Earth System Sciences*, 10(8), 1697-1724.
- [2] İnternet kaynağı: <https://www.kaggle.com/vasudeva009/predicting-coupon-redemption-pca>
- [3] Joulfaian, D., & Marlow, M. L. (1990). Government size and decentralization: Evidence from disaggregated data. *Southern Economic Journal*, 1094-1102.
- [4] Cole, D. (2010). The Effects of Student-Faculty Interactions on Minority Students' College Grades: Differences between Aggregated and Disaggregated Data. *Journal of the Professoriate*, 3(2).
- [5] You, L., Wood, S., & Wood-Sichra, U. (2009). Generating plausible crop distribution maps for Sub-Saharan Africa using a spatially disaggregated data fusion and optimization approach. *Agricultural Systems*, 99(2-3), 126-140.
- [6] Skinner, G. W. (1986). Sichuan's population in the nineteenth century: Lessons from disaggregated data. *Late Imperial China*, 7(2), 1-79.
- [7] Odhiambo, N. M. (2021). Energy consumption and economic growth in Botswana: Empirical evidence from a disaggregated data. *International Review of Applied Economics*, 35(1), 3-24.
- [8] Şeker, Ş. E. CRISP-DM: Endüstriler Arası Standart İşleme-Veri Madenciliği için (Cross Industry Standard Processing-Data Mining).
- [9] Garrett, T. A. (2003). Aggregated versus disaggregated data in regression analysis: implications for inference. *Economics Letters*, 81(1), 61-65.
- [10] Verma, A., & Anwar, A. (2021). A Comprehensive Review on the NILM Algorithms for Energy Disaggregation. arXiv preprint arXiv:2102.12578.
- [11] Almutairi, F. M., Kanatsoulis, C. I., & Sidiropoulos, N. D. (2021). PREMA: Principled tensor data recovery from multiple aggregated views. *IEEE Journal of Selected Topics in Signal Processing*, 15(3), 535-549.
- [12] Hyndman, R. J. (2011). Moving Averages.
- [13] Schirmer, P. A., Mporas, I., & Paraskevas, M. (2019, July). Evaluation of regression algorithms and features on the energy disaggregation task. In 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA) (pp. 1-4). IEEE.
- [14] Arambepola, R., Lucas, T. C., Nandi, A. K., Gething, P. W., & Cameron, E. (2020). A simulation study of disaggregation regression for spatial disease mapping. arXiv preprint arXiv:2005.03604.

- [15] Chamberlin, G. (2010). Methods Explained: Temporal Disaggregation. *Economic & Labour Market Review*, 4(11), 106-121.
- [16] Dolgun, M. Ö., & Ersel, D. (2014). Doğrudan pazarlama stratejilerinin belirlenmesinde veri madenciliği yöntemlerinin kullanımı. *İstatistikçiler Dergisi: İstatistik ve Aktüerya*, 7(1), 1-13.
- [17] Bass, F. M. (1969). A simultaneous equation regression study of advertising and sales of cigarettes. *Journal of Marketing Research*, 6(3), 291-300.
- [18] İnternet kaynağı: <https://digiday.com/marketing/how-the-worlds-biggest-advertisers-are-spending-or-not-as-industries-adapt-to-the-coronavirus-pandemic/>
- [19] İnternet kaynağı: <https://www.softwaresuggest.com/blog/global-digital-ad-spending-research/#>
- [20] Purba, R. (2020). MULTIMODAL ANALYSIS ON ERTIGA CAR ADVERTISEMENT. *Wiralodra English Journal*, 4(1), 21-32.