

YZM 511 - VİZE SINAV SORULARI

04.12.2023

Lütfen problemlerden **yalnızca birini** seçerek soruları cevaplandırınız. Aşamaların açıkça gösterilmesi ve teslim edeceğiniz kodların hiçbir müdahale gerektirmeksizin çalıştırılabilir durumda olması beklenmektedir. Lütfen, her bir adım için yapmış olduğunuz yorumları, ilgili kod bloklarının hemen yanında ve başında # işareti kullanarak ekleyiniz. Bu, kodunuzun okunabilirliğini ve anlaşılabilirliğini artıracaktır.

Cevap kağıdınızı Enstitü tarafından ilan edilen yerde ve saatte imza karşılığı teslim ediniz. Sınav duyurusuna [buradan](#) erişebilirsiniz.

Başarılar dilerim.

İŞ PROBLEMİ A:

Bir sağlık ve fitness uygulaması şirketi, geliştirdiği mobil uygulamanın genel sağlık göstergeleri üzerindeki etkisini anlamak istemektedir. Şirket, düzenli uygulama kullanımının kullanıcıların adım sayısını ve genel aktivite seviyelerini artırabileceğine inanmaktadır. Bu nedenle şirket, kullanıcıları aşağıdaki şekilde iki gruba ayırmaktadır.

A Grubu: Uygulamayı haftada en az 5 gün kullananlar.

B Grubu: Uygulamayı haftada 5 günden az kullananlar.

Grup	Yas	Cinsiyet	Ortalama_Kalp_Atis_Hizi	Gunluk_Uyku_Suresi	Haftalik_Ortalama_Adim_Sayisi
A	62	Kadın	73	8	920
A	18	Erkek	65	6	920
A	21	Erkek	60	7	902
A	21	Erkek	68	8	943
A	57	Kadın	75	6	840

Veri setinde yer alan gözlemlerin ilk 5 satırı yukarıdaki gibidir. Eğer düzenli uygulama kullanımının adım sayısını artırdığı istatistiksel olarak anlamlıysa, şirket bu bilgiyi kullanıcıları daha fazla hareket etmeye teşvik etmek için kullanmayı amaçlamaktadır. Bu kapsamda tarafınızdan yapılması beklenenler aşağıda yer almaktadır. Bu iş problemini çözerken [health_app_usage.csv](#) isimli veri setini kullanınız.

SORULAR

- 1) Yalnızca gerekli kütüphaneleri ve testleri çalışma ortamına yükleyiniz.¹ (5 puan)
- 2) Veri setini çalışma ortamına yükleyiniz. (5 puan)
- 3) Veriden edindiğiniz ilk izlenim nedir? (15 puan)
 - a) Veri setinin baştan ve sondan ilk 10 satırını görüntüleyiniz. (5 puan)
 - b) Karşılaştırma yaptığınız gruplar içerisinde, önemli gördüğünüz hususlar (örneğin, grupların ortalama değerleri gibi) hakkında yorum yapınız. (10 puan)
- 4) %1 anlamlılık düzeyi için aşağıdaki hususları içerek şekilde istatistiksel testlerinizi yapınız ve sonuçları yorumlayınız. Kullanacağınız tüm testlerin temel-alternatif hipotezini ve değerlendirme kriterini açıkça belirtiniz. Her bir test için elde edeceğiniz test istatistiği ve p-value'nin yorum olarak ilgili kodun hemen altına yazılması ve yorumlanması beklenmektedir. (45 puan)
 - a) Doğru istatistiksel testin hangisi olduğuna karar verebilmek için sınamanız gereken varsayımlar nelerdir gösteriniz. Gerekli değilse varsayımı sınamayınız. (20 puan)
 - b) Nihai test ve sonuçlarını gösteriniz. (15 puan)
 - c) Nihai test sonuçlarını dikkate alarak iş problemini yorumlayınız. (10 puan)
- 5) %5 anlamlılık düzeyi için 4. soruda elde ettiğiniz sonuçları tekrar yorumlayınız. Yalnızca gerekli olduğunu düşündüğünüz testleri tekrarlamanız beklenmektedir. Diğer türlü 4. soruda elde ettiğiniz test sonuçlarını kopyalayıp altına yorumlarınızı yazabilirsiniz. (30 puan)

BONUS: Seçtiğiniz iş probleminin nedenini kısaca açıklayınız. Kararınızı etkileyen faktörler nelerdir? (10 puan)

¹ Kullanılmadığı halde yüklenmiş kütüphane ve testlerin her biri için 1 puanınız düşürülecektir.

İŞ PROBLEMİ B:

Bir e-ticaret şirketi, müşterilerine yönelik iki farklı e-posta pazarlama kampanyasını test etmek istemektedir. Bunları "Kampanya A" ve "Kampanya B" olarak adlandırabiliriz. Kampanya A, ürünlerin tanıtıldığı geleneksel bir e-posta içeriği sunarken, Kampanya B interaktif ve kişiselleştirilmiş içerikler içermektedir. Bu deney kapsamında müşteriler rastgele olarak kampanyalardan birine atanmış ve her birine e-posta gönderilmiştir. Bu işlemler sonucunda üretilen veriler şirketin bilgi işlem birimi tarafından [email_marketing.csv](#) isimli bir dosya ile size sunulmaktadır. Verinin ilk 5 satırı aşağıdaki gibidir.

Musteri_ID	Kampanya	Eposta_Acilma_Orani	Tıklama_Orani	Satin_Alma_Miktari	Musteri_Segmenti	Bölge
1	Kampanya A	90.83	77.71	13	Düşük Harcama	Bölge 4
2	Kampanya B	23.96	27.86	21	Yeni Müşteri	Bölge 2
3	Kampanya A	14.49	55.84	16	Yeni Müşteri	Bölge 3
4	Kampanya A	48.95	42.42	22	Sık Müşteri	Bölge 1
5	Kampanya A	98.57	90.64	1	Düşük Harcama	Bölge 2

Şirket, hangi kampanyanın satın alma miktarlarını artırdığını belirlemek istiyor. Bu kapsamda tarafınızdan yapılması beklenenler aşağıda belirtilmektedir.

SORULAR

- 1) Yalnızca gerekli kütüphaneleri ve testleri çalışma ortamına yükleyiniz². (5 puan)
- 2) Veri setini çalışma ortamına yükleyiniz. (5 puan)
- 3) Veriden edindiğiniz ilk izlenim nedir? (15 puan)
 - a) Veri setinin baştan ve sondan ilk 10 satırını görüntüleyiniz. (5 puan)
 - b) Karşılaştırma yaptığınız gruplar içerisinde, önemli gördüğünüz hususlar (örneğin, grupların ortalama değerleri gibi) hakkında yorum yapınız. (10 puan)
- 4) %1 anlamlılık düzeyi için aşağıdaki hususları içerek şekilde istatistiksel testlerinizi yapınız ve sonuçları yorumlayınız. Kullanacağınız tüm testlerin temel-alternatif hipotezini ve değerlendirme kriterini açıkça belirtiniz. Her bir test için elde edeceğiniz test istatistiği ve p-value'nin yorum olarak ilgili kodun hemen altına yazılması ve yorumlanması beklenmektedir. (45 puan)
 - a) Doğru istatistiksel testin hangisi olduğuna karar verebilmek için sınamanız gereken varsayımlar nelerdir gösteriniz. Gerekli değilse varsayımı sınamayınız. (20 puan)
 - b) Nihai test ve sonuçlarını gösteriniz. (15 puan)
 - c) Nihai test sonuçlarını dikkate alarak iş problemini yorumlayınız. (10 puan)
- 5) %5 anlamlılık düzeyi için 4. soruda elde ettiğiniz sonuçları tekrar yorumlayınız. Yalnızca gerekli olduğunu düşündüğünüz testleri tekrarlamamanız beklenmektedir. Diğer türlü 4. soruda elde ettiğiniz test sonuçlarını kopyalayıp altına yorumlarınızı yazabilirsiniz. (30 puan)

BONUS: Seçtiğiniz iş probleminin nedenini kısaca açıklayınız. Kararınızı etkileyen faktörler nelerdir? (10 puan)

² Kullanılmadığı halde yüklenmiş kütüphane ve testlerin her biri için 1 puanınız düşürülecektir.