

Adı Soyadı:		Telefon Numarası:	
E-Mail Adresi:		Tarihi:	

*Aşağıda bulunan görevlerin çözüm senaryolarını oluşturarak yazmanızı ve mümkünse kodlamasını yapmanızı rica ederiz.

*Yaparken tercih ettiğiniz herhangi bir dili ve aracı kullanabilirsiniz.

*Zorlandığınız noktaları, yeni öğrendiğiniz veya önceden tecrübe ettiğiniz için hızlı yaptığınız yerleri anlatmanız artı olacaktır.

*Kafanıza takılan kısımlar olursa anladığınız kadarıyla yapıp yorum yazabilirsiniz.

*Araştırma yapmaktan çekinmeyin ve neyi nasıl araştırıp yaptığınızı belirtin.

*Aşağıdaki verilen sorulardan en az 1 tanesinin "**SAS Data Step**" programlama dili ile yapılması beklenmektedir.

Bu amaçla kullanılabilecek "**SAS University Edition**" içinde gelen web tabanlı "**SAS Studio**" aracı aşağıdaki adresten ücretsiz olarak indirilebilir.

- https://www.sas.com/en_us/software/university-edition/download-software.html

* SAS Studio aracının öğrenimi ve kullanımı ile ilgili ücretsiz yardımcı kaynaklara aşağıdaki linklerden erişilebilir:

- <https://video.sas.com/category/videos/sas-studio>
- <https://support.sas.com/edu/elearning.html?ctry=us&productType=library>
- <https://documentation.sas.com/?docsetId=basess&docsetTarget=n053a58fwk57v7n14h8x7y7u34y4.htm&docsetVersion=9.4&locale=en>

Vakit ayırdığınız için şimdiden teşekkür ederiz.

- 1.) Bir banka müşterilerine ait 3 milyon telefon numarası verisini bir text dosyasında veriyor ve bizden telefonların sabit hat mı, cep telefonu hattı mı olduğunu, sabit hat ise hangi şehre ait olduğunu bulmamızı istiyor. Görevin yerine getirilmesi için nasıl bir yol izlersiniz?

MUSTERI_NO	TELEFON
1	05333371000
2	5354321938
3	3840812
3	545 12 15
4	0212 512 12 78
5	0312 244 12 45
6	0216 441 45 12
7	0273 574 19 1
7	53534156 43
8	5422388080
9	1111111
10	3121111111
11	+90543 337 91 31
12	272 415 56 58
13	0312 244 12 47
14	0312 244 12 48
15	0315 233 12 33
15	5444321939
16	02164177175

- 2.) Elimizde çok sayıda e-mail adresi var. Bu adreslerin sahiplerinin hangilerinin aynı firmada çalıştığını bulmak için nasıl bir yol izlersiniz?

ID	EMAIL
1	deniz@sadeyazilim.com
2	kadir@SADEYAZILIM.COM
3	kadir@afirma.com
4	Ceyda@gmail.com
5	ali@gmail.com
6	tamer@sadeyazilim.com
7	müge@testbank.com.tr

- 3.) Bankada veri tabanında müşterilerin adresleri, iş adresi ve ev adresi olarak tutulmaktadır. Bazı müşteriler yazım farkları olsa da iş ve ev adresi olarak aynı adresi vermişlerdir. İş ve ev adresi aynı olan müşterileri nasıl bulursunuz?

İS_ADRESİ	EV_ADRESİ
Tugay Yolu Cad. No:60 Maltepe	Tugay Yolu Caddesi No60 MALTEPE
İnkılap Mah. Küçüksu Cd. No 6 Ümraniye	İnkılap Mahallesi Küçüksu Caddesi No 6 Ümraniye
Soğanlık Mah. Muş Sk. No 63 Kartal	Soğanlık Mah. Muş Sk. No 61 Kartal

- 4.) Elimizde bazı okulların isimleri ve bulundukları mahallelerin olduğu bir liste mevcut. Başka bir listede ise il, bu illerdeki ilçeler ve bu ilçelerdeki mahalleleri bulunduran referans tablomuz var. Mahalle bilgisini ve bu referans tabloları kullanarak okulların hangi il ve ilçede olduklarını nasıl bulursunuz?

OKUL_ADI	MAHALLE
Kavaklıdere İÖO	Kavaklıdere Mah.
Etimesgut Anadolu Lisesi	Eryaman Mah.
Cumhuriyet Lisesi	Bozkurt Mah.
Zafer İlkokulu	Cumhuriyet Mah.

İL
İSTANBUL
ANKARA
İZMİR
DENİZLİ

İL	İLCE
ANKARA	ÇANKAYA
ANKARA	ALTINDAĞ
ANKARA	ETİMESGUT
İSTANBUL	ŞİŞLİ

İLCE	MAHALLE
ÇANKAYA	KAVAKLIDERE
ÇANKAYA	ESAT
ÇANKAYA	AYRANCI
ÇANKAYA	CUMHURİYET
ŞİŞLİ	BOZKURT
ŞİŞLİ	ESKİŞEHİR
ŞİŞLİ	CUMHURİYET

- 5.) Bir bankanın veri tabanı tablosunda (örneğin Oracle) müşteri isimleri ve müşterilerin TC kimlik numaraları var. Fakat bazı TC kimlik numaraları aynı olmasına rağmen müşteri isimleri farklı. Bu durumdaki kayıtları nasıl bir sorgu ile bulursunuz?

TCKN	AD
71824829102	Ali
12381949284	Ali
71824829102	Ali
59382838321	Ahmet
71824829102	Mehmet

6.) Bir adres verisi içinde, farklı yerlerde TC kimlik numarası bilgisi de bulunmaktadır. Bu bilgiyi adres bilgisinden nasıl ayırıp ikisini de ayrı yazarsınız?

Tugay Yolu Cad. No:60 12537489122 Maltepe
İnkılap Mah. 88372637771 Küçüksu Cd. No 6 Ümraniye
11293888122 Soğanlık Mah. Muş Sk. No 63 Kartal
Bilezikçi sk 147 Şişli İstanbul 12229386

7.) Bir müşterimiz enerji üretimini planlamak üzere, enerji tüketimi üzerinden bir forecasting çalışması yapmak istediğini belirterek bize danıştı.

Veri: Coğrafi bir bölgenin, MWh bazında günlük elektrik tüketimi verisi ekteedir.
Yukarıda paylaşılan veriyi kullanarak,

1. Data Quality: Veri’de gözlemlendiğiniz (missing, duplicates, out of range, vb.) durumlar var mıdır?
2. Descriptive Statistics: Veriyi bir zaman serisi olarak plot ederek gözlemlerinizi paylaşınız (saatlik, aylık, sezonluk, haftanın günleri).
3. Forecasting: İstediğiniz bir forecasting modeli ile verinizi “initialization” set, ve “holdout” set olarak ikiye bölün. Forecasting modelinizi “initialization” set üzerinde geliştirin Sonrasında da modelinizin performansını holdout set üzerinden değerlendirin. Değerlendirmede model parametrelerinizin ne anlama geldiğini ve parametrelerle ilgili verdiğiniz kararları açıklayın.
4. Evaluation: Bu projede performans indikatörü olarak MAPE kullanılacaktır. Bu kavramı içgörüş/intuition kullanarak kısaca açıkladıktan sonra, kullandığınız ya da seçtiğiniz modeldeki sonuçları açıklayınız. Ek olarak tahmin, gerçekleşen ve fark olmak üzere bir plot yaratın ve yorumlayın.
5. Proje kapsamında başka hangi verilere ihtiyaç olabilir? Ya da dikkat edilmesi dış faktörler neler olabilir?