

Bölüm A: SQL

Soru 1: Her şubenin genel memnuniyet oranını hesaplayan bir SQL sorgusu yazın.

```
--Her Şubenin Genel Memnuniyet Oranı
SELECT
  model AS sube, --şube bilgisi
  COUNT(*) AS anket_sayisi, --her şubedeki toplam anket
  SUM(CASE WHEN muster_i_memnun = 'evet' THEN 1 ELSE 0 END) AS memnun_sayisi, --evet=1 topl
  ROUND(100.0 * SUM(CASE WHEN muster_i_memnun = 'evet' THEN 1 ELSE 0 END) / COUNT(*), 2) AS memnuniyet_yuzde -- memnuniyeti yüzdelerik olarak
FROM restoran_memnuniyet --veri kaynağı
GROUP BY model --her şube için ayrı satır
ORDER BY model; --önce Sube_A sonra Sube_B
```

Soru 2: Memnuniyet oranlarını lokasyon tipine göre kırarak gösteren bir SQL sorgusu yazın.

```
-- Her Şubenin Lokasyon Tipine Göre Memnuniyet Oranı
SELECT
  model AS sube, -- şube bilgisi
  lokasyon_tipi, -- AVM veya sanayi
  COUNT(*) AS anket_sayisi, -- şube-lokasyon grubundaki toplam anket
  SUM(CASE WHEN muster_i_memnun = 'evet' THEN 1 ELSE 0 END) AS memnun_sayisi, -- evet=1 topl
  ROUND(100.0 * SUM(CASE WHEN muster_i_memnun = 'evet' THEN 1 ELSE 0 END) / COUNT(*), 2) AS memnuniyet_yuzde -- memnuniyeti yüzdelerik hesapla
FROM restoran_memnuniyet -- veri kaynağı
GROUP BY model, lokasyon_tipi -- şube ve lokasyon kombinasyonlarına göre ayrı satır
ORDER BY model, lokasyon_tipi; -- önce şubeye sonra lokasyona göre sıralar
```

Bölüm B: Python

Soru 3: restaurant_branch_success.csv dosyasını Pandas ile okuyun. Aşağıdakileri hesaplayın.

a) Her şube için genel memnuniyet oranı:

```
# Her Şube İçin Genel Memnuniyet Oranı

# muster_i_memnun evet/hayır --> oran hesabı için 1/0 çevir
df["memnun_bool"] = (df["muster_i_memnun"] == "evet").astype(int)

# şubeye göre grupla: toplam anket, memnun sayısı, oran
sube_memnuniyet = df.groupby("model").agg(
  anket_sayisi = ("ticket_id", "count"),
  memnun_sayisi = ("memnun_bool", "sum"),
  memnuniyet_orani = ("memnun_bool", "mean")
).reset_index()

# oranı yüzdeye çevir
sube_memnuniyet["memnuniyet_yuzde"] = (sube_memnuniyet["memnuniyet_orani"] * 100).round(2)

sube_memnuniyet
```

✓ 0.0s

	model	anket_sayisi	memnun_sayisi	memnuniyet_orani	memnuniyet_yuzde
0	Sube_A	1000	539	0.539	53.9
1	Sube_B	1000	774	0.774	77.4

İlk olarak her şubenin genel memnuniyet oranını çıkardım

muster_i_memnun sütununu 1/0'a çevirerek şubeye göre grupladım. Bu sayede Şube A ve Şube B'nin ham verideki performansını yan yana görebildim.

b) Her Őube iin lokasyon tipine gre memnuniyet oranı:

```
# Her Őube İin Lokasyon Tipine Gre Memnuniyet Oranı

df["memnun_bool"] = (df["musteri_memnun"] == "evet").astype(int)

# Őube ve lokasyon ikilisine gre grupla
sube_lokasyon_memnuniyet = df.groupby(["model", "lokasyon_tipi"]).agg(
    anket_sayisi = ("ticket_id", "count"),
    memnun_sayisi = ("memnun_bool", "sum"),
    memnuniyet_orani = ("memnun_bool", "mean")
).reset_index()

# oranı yzdeye evir (2 ondalık)
sube_lokasyon_memnuniyet["memnuniyet_yuzde"] = (
    sube_lokasyon_memnuniyet["memnuniyet_orani"] * 100
).round(2)

sube_lokasyon_memnuniyet
✓ 0.0s
```

	model	lokasyon_tipi	anket_sayisi	memnun_sayisi	memnuniyet_orani	memnuniyet_yuzde
0	Sube_A	avm	200	185	0.9250	92.50
1	Sube_A	sanayi	800	354	0.4425	44.25
2	Sube_B	avm	800	690	0.8625	86.25
3	Sube_B	sanayi	200	84	0.4200	42.00

Bu aŐamada genel oranların yanıtıcı olabileceđini dŐunerek verileri lokasyon tipine gre ayırdım. Gruopby'a lokasyon_tipi stununu ekleyerek her Őubenin AVM ve sanayi lokasyonlarındaki memnuniyet oranını ayrı ayrı hesapladım. Bu segmentasyon sonucunda Őube A'nın her iki lokasyonda da Őube B'den daha yksek memnuniyet rettiđini net bir Őekilde grdm.

c) Her şubeye atanan lokasyon tipi dağılımı:

```
# Her Şubeye Atanan Lokasyon Tipi Dağılımı

# Şube x lokasyon kombinasyonuna göre anket sayısı
sube_lokasyon_dagilim = df.groupby(["model", "lokasyon_tipi"]).size().reset_index(name="anket_sayisi")

# Her şubenin kendi içindeki yüzde payı
sube_lokasyon_dagilim["yuzde"] = (
    sube_lokasyon_dagilim.groupby("model")["anket_sayisi"]
    .transform(lambda x: x / x.sum() * 100)
    .round(2)
)

sube_lokasyon_dagilim
```

✓ 0.0s

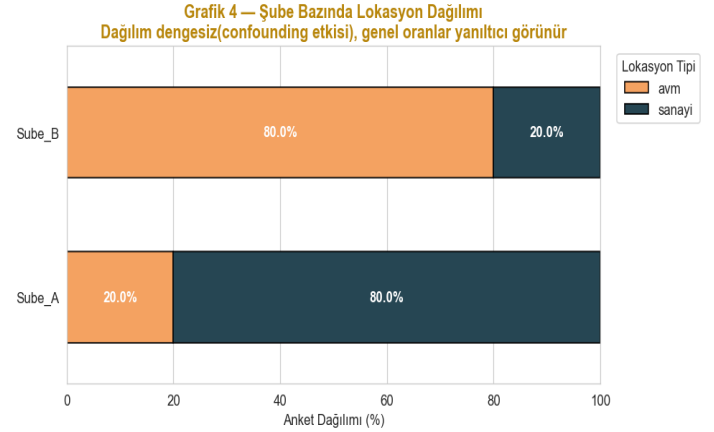
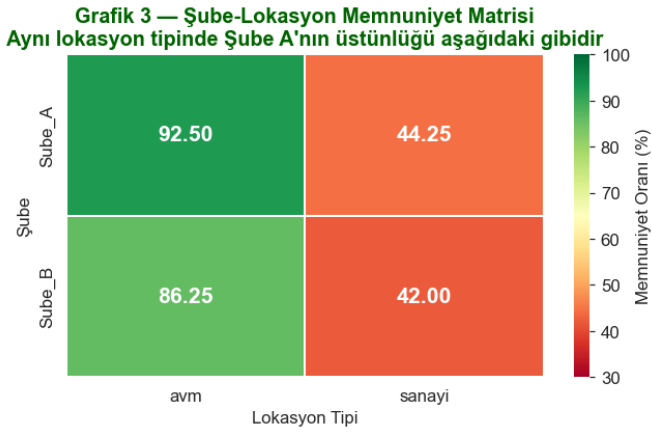
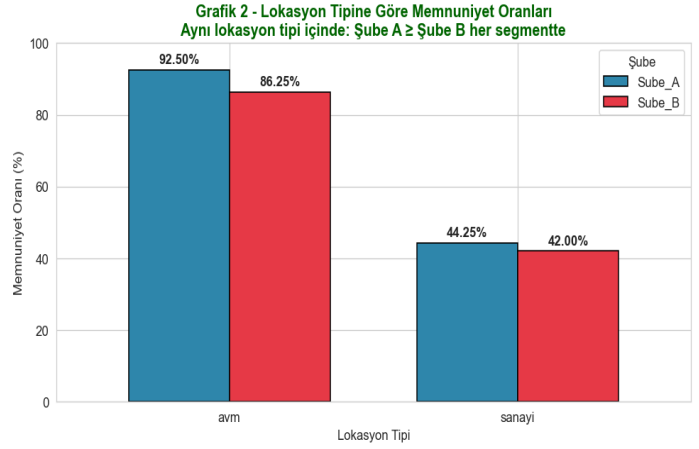
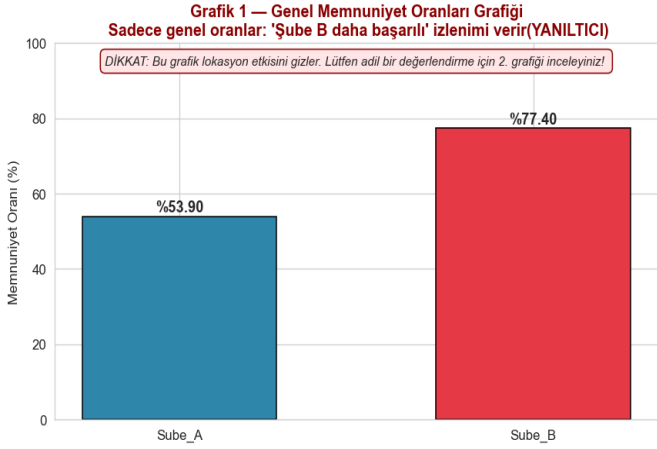
	model	lokasyon_tipi	anket_sayisi	yuzde
0	Sube_A	avm	200	20.0
1	Sube_A	sanayi	800	80.0
2	Sube_B	avm	800	80.0
3	Sube_B	sanayi	200	20.0

Her şubeye hangi tür lokasyondan kaç adet anket geldiğine baktım. Tablo net bir şekilde dengesizliği gösteriyor. Şube A çoğunlukla sanayi, Şube B ise çoğunlukla AVM müşterisiyle değerlendirilmiş olduğu çıkarımını yaptım.

Yaptığım sorgular ve Pandas hesaplarıyla sonuç olarak değerlendirmelerim ve vardığım yargılar;

Şube B'nin ham verideki üstünlüğünün gerçek bir performans farkından değil, büyük ölçüde atanan lokasyon dağılımındaki avantajdan kaynaklandığı görülmektedir. Lokasyon tipi sabit tutulduğunda Şube A her iki segmentte de daha başarılı sonuç verdiği için, adil bir karşılaştırma yapmak adına confounding variable'ın (lokasyon) kontrol altına alınması gerekliliği ortaya çıkmaktadır. Bu bulgular, sonraki bölümlerdeki sergilediğim istatistiksel testlere ve karar yorumlarıma temel oluşturmaktadır.

Soru 4: İnsan kaynakları kararına temel oluşturacak grafik tasarlayın.



Sadece genel oranları gösteren grafik, yöneticileri kesinlikle yanıltır. Genel memnuniyet oranlarının bulunduğu Grafik 1 üzerinden yapılan yorumlar ve değerlendirmeler yanıltıcıdır. 4. Grafikte görüldüğü üzere lokasyon dağılımının şubeler arasında dengesiz olması, 1. Grafik üzerinden alınacak kararın neden yanıltıcı olacağını açıklar niteliktedir. **İnsan Kaynaklarına sunmak için 2. Grafik kullanılmalıdır.** Lokasyon tipinin bir kontrol değişkeni olarak dikkate alınması gerektiğini ve bunun sonucunda genel memnuniyet oranlarından farklı olarak her iki lokasyon tipinde de Şube A, Şube B'ye göre üstün performans sergilemekte olduğunu gösterir. Günün sonunda İnsan Kaynaklarına karar oluşturmak için 2. veya farklı bir gösterimi olarak 3. grafik tercih edilmelidir.

Bölüm C: Karar

Soru 5: "Şube B daha başarılı" sonucu istatistiksel olarak doğru mudur? Teknik terimi açıklayın.

- **"Şube B daha başarılı" sonucu istatistiksel olarak yanlıştır.** Sonucu ham, düzeltilmemiş veri üzerinden incelendiğinde güvenilir gibi gözükse de istatistiksel olarak incelendiğinde bu çıkarımın güvenli olmadığına kanaat getirilir. İki oran için Z Test uygulandığında Şube A (%53.90) ile Şube B (%77.40) arasındaki fark anlamlıdır ($z \approx -11.07$, $p < 0.001$; $\alpha = 0.05$). Oluşturulan hipotez reddedilir ($H_0: p_A = p_B$) ve ham veri üzerinde Şube B lehine anlamlı bir fark olduğu söylenebilir. Fakat lokasyon tipi sabit tutularak yapılan testlerde: AVM'de Şube A (%92.50) ile Şube B (%86.25) farkı Şube A lehine anlamlıdır ($p \approx 0.017$). Sanayi'de Şube A (%44.25) ile Şube B (%42) farkı anlamsızdır ($p \approx 0.566$). Sonuçlarından hareketle her iki lokasyon tipinde de Şube B, Şube A'dan daha başarılı değildir. Ham veriden yapılan sorgu ile gözlenen üstünlük, lokasyon tipi göz önünde bulundurulduğunda yok olmakta, hatta tersine dönmektedir.
- **Simpson Paradoksu**'na göre alt gruplarda izlenen eğilim gruplar birleştirilince tersine dönebilir. Confounding variable (lokasyon tipi) üzerinde dengesiz örneklem dağılımından kaynaklanır. Şube A ağırlıklı olarak sanayi, Şube B ise ağırlıklı olarak AVM lokasyonuna atanmıştır. AVM'de doğal olarak daha yüksek memnuniyet oluştuğundan (AVM: %87.5, sanayi: %43.8), Şube B'nin genel oranı yapay şekilde şişmiş görünmektedir.

Soru 6: Lokasyon dağılımındaki sorunu açıklayın.

- Lokasyon dağılımındaki temel sorun, müşteri anket routing sisteminin eşit dağılım ve rastgele atama yapmamasıdır. Şube A'ya gelen anketlerin %80'i sanayi, Şube B'ye gelen anketlerin %80'i ise AVM lokasyonuna aittir; iki şube tamamen farklı müşteri popülasyonlarıyla değerlendirilmektedir. AVM ve sanayi lokasyonlarının genel memnuniyet düzeyleri arasında yaklaşık 44 puanlık doğal bir fark bulunduğundan (AVM %87.5, sanayi %43.8), bu dengesiz dağılım şubeler arasındaki performans karşılaştırmasını baştan adaletsiz kılmakta ve bir **selection bias durumu** yaratmaktadır. Şubelerin karşılaştırılabilmesi için gereken "eşit koşullar" varsayımı ihlal edildiğinden mevcut yapı kontrollü bir deney değil, dengesiz **quasi-experiment** niteliğindedir.

Soru 7:

a) Hangi Őube yneticisi gerekten daha baŐarılı? Neden?

- Her ne kadar ham veriden ıkan sonu farklı olsa da memnuniyet oranlarını lokasyon tipine gre kırarak elde ettiĐimiz veriden hareketle Őube A yneticisinin Őube B yneticisine kıyasla daha baŐarılı olduĐu anlaŐılmaktadır. Őube A her iki lokasyonda da Őube B'den fazla performans gstermektedir.

AVM lokasyonunda Őube A memnuniyet oranı %92.50 iken Őube B'nin oranı %86.25'tir. Aradaki 6.25 puanlık fark istatistiksel olarak anlamlıdır ($p \approx 0.017$).

Sanayi lokasyonunda Őube A memnuniyet oranı %44.25 iken Őube B'nin oranı %42.00'dır. Aradaki 2.25 puanlık fark istatistiksel olarak anlamlı ıkmasa da yn olarak A'nın lehinedir.

- Őube A'ya ynlendirilen anketlerin %80'i sanayi lokasyonundan gelirken bu lokasyondaki mŐteri memnuniyeti genel olarak dŐktr (%43.8). Yani Őube A ynetici yapısal olarak daha zorlu bir lokasyonda alıŐmasına raĐmen Őube B'ye gre daha iyi sonular retmeyi baŐarmıŐtır. Őube B ise %80 gibi byk bir blm AVM lokasyonundan gelen mŐteri memnuniyeti anketine tabi tutulmuŐtur. Bu lokasyonda genel olarak yksek memnuniyet verisinin (%87.5) olduĐu gz nnde bulundurularak, sonuta ham veride grnen %77.40'lık baŐarı yneticinin becerisinden deĐil, routing sisteminin yarattıĐı avantajlı mŐteri daĐılımından kaynaklandıĐını gsterir.
- Btn bunların sonucunda Őube A yneticisini grevden almak ve Őube B yneticisinin dllendirilmesi kararı, verinin gsterdiĐi gerekle tam olarak zıttır. Uzun yıllardır iŐletmeyi yneten A yneticisi, daha zorlu bir mŐteri segmentine sahip olmasına raĐmen tutarlı bir biimde daha yksek memnuniyet retmekte, dolayısıyla B yneticisine kıyasla daha baŐarılı olduĐu anlaŐılmaktadır.

b) Adil bir performans deęerlendirmesi nasıl tasarlanır?

- İlk olarak yapılması gereken veri toplama seviyesinde routing sisteminin yeniden düzenlenmesidir. İdeal olan anketlerin şubelere rastgele atanmasıdır. Mümkün deęilse stratified sampling gibi örnekleme yöntemleri kullanılabilir. Her şubeye her lokasyon tipinden dengeli sayıda anket yönlendirilmelidir. Bu sayede lokasyon türünün yarattığı yapısal farklar dengelenmiş olur ve confounding etkisi büyük ölçüde kontrol altına alınmış olur.
- Analiz seviyesinde ise tek metrik zayıf kalarak yanıltıcı sonuçlar doğurabilir bizim durumumuzdaki gibi. Bu sebeple memnuniyet oranı önemli bir göstergedir ancak ek olarak operasyonel verimlilik, sipariş süresi, tekrar eden müşteri oranı, masa devir süresi gibi ekstra metriklerle tekrar ele alınmalıdır. Ve sonuç her ne olursa olsun tüm karşılaştırmalar lokasyon tipi gibi kontrol deęişkenleri göz önünde bulundurularak yapılmalıdır. Yani Şube A'nın genel memnuniyet oranı yerine Şube A'nın AVM ve Sanayi lokasyonlarındaki ayrı ayrı performansları incelenmelidir.
- Önemli noktalardan bir dięeri zamansal perspektifin dahil edilmesidir. Tek dönemlik veya kısa vade verilere bakarak karar vermek yerine uzun vadeli trendler izlenmeli, dönem şartları göz önünde bulundurulmalı (şanslı, zorlu, kriz dönemleri) ve etkileri dikkate alınmalı. Son olarak ise İK karar alıcılarının da ekiple birlikte interaktif yaklaşımı ve konulara hakimliği ile yanlış kararların önüne geçilebilir ve adil bir yaklaşım sunulabilir.

c) Yanlış İK kararının (Şube A yöneticisini işten çıkarmak) çalışan motivasyonuna ve kurumsal hafızaya etkisi nedir?

- Yanlış İK kararı ile özellikle köklü ve objektif olarak daha başarılı bir yöneticinin görevden alınması yalnızca o kişiyi deęil tüm organizasyonu olumsuz etkileyen sonuçlar doğurur. Çalışan motivasyonu üzerinde bir kırılma meydana gelir. Başarılı bir yöneticinin veriye dayanmayan (veya yanıltıcı grafik, veri üzerinden) yanlış bir kararla çıkartılmasından sonra dięer çalışanlarda da "ben de haksız yere çıkartılabilirim" hissi yaratır. Kendi durumlarını ve geleceklerini sorgular, kaygı içine ve motivasyon eksikliğine girerler. Psikolojik güvenlik azalır, aidiyet duygusu azalır, risk alma isteğinde düşüş görülür, inisiyatif alma gibi unsurlar yok olur bu da şube performansının düşmesine yol açar. Motivasyonlu ekip yerini belirsizlik duygusu hisseden, savunmacı (arka planda olmak isteyen), karamsar bir çalışan tabanına döndürür.

- Kurumsal hafıza tarafında ise doğrudan anlaşılmayan ama büyük, etkili bir kayıp oluşturur. Deneyimli yöneticinin yıllar içinde biriktirdiği prosedürlere yazılmayan zor müşteri profilleri, sanayi lokasyonunun dinamikleri işleyişi, kriz anı refleksleri gibi kritik bilgiler kişi yönetici ile birlikte gitmiş olur. Bu bilgi transferi yapılmadan verilen karar şubenin veya organizasyonların tekrar tekrar aynı hataları yapmasına yol açar.
- Sonuç olarak veriyi doğru okumayan bir İK kararı kısa vadede motivasyon, güven, istek gibi değerlere zarar verirken; uzun vadede kurumun öğrenme kapasitesini ve sürdürülebilirlik gibi parametreleri olumsuz etkiler. Confounding variable göz önünde bulundurmeyen bir analiz hem istatistik hem de insanlar üzerindeki etkileri olarak yanlış bir karara dönüşür bu sebeple doğru analiz her açıdan önemlidir.