

Bu case çalışması 3 ana soru grubundan oluşmaktadır ve her soru gruplarının içerisinde alt sorular bulunmaktadır. Case çalışması içerisindeki sorulara **Python** ve **SQL** kullanılarak cevap verilmesi beklenmektedir. Tüm sorularda temel programlama becerilerinin yanında analitik bakış açısı, probleme yaklaşım ve yorumlama becerileri gibi faktörler değerlendirme sürecinde önemli bir etken olacaktır. Cevaplardan beklenen format her soruda detaylı bir şekilde ifade edilmiş olup beklenenden farklı **formatta** gönderilmiş olan cevaplar **değerlendirmeye alınmayacaktır**.

Çıktı Beklentisi : Bir pdf dosyası içerisinde aşağıdaki soruların tamamının yazılı olması ve soruların altlarına cevapların ayrıntılı olarak yazılması gerekmektedir. 1. Soru için Python scripti (tercihen Jupyter notebook dosyası) oluşturulmalı ve içerisinde yazılan kodların yorumlarıyla birlikte tüm kodların çalışıyor olması önemlidir. Çalışmayan veya yorumda olan kodlar değerlendirmeye alınmayacaktır. Pdf dosyasının ve Python script dosyasının isimleri, gönderim tarihi ve ad soyad bilgilerini içerecek şekilde olmalıdır. Örneğin 20230807_Hakan_Aras.pdf

İletilen dosyalar örnek amaçlı olup sadece bu çalışma için kullanılacaktır ve gerçek verileri yansıtmamaktadır. İletilen dosyalar yalnızca adaylara özel olup üçüncü kişilerle ve farklı platformlarda kesinlikle paylaşılmamalıdır.

Sorular :

1. A isimli firma yapmış olduğu satışlar üzerinden çeşitli analizler yapmak istemektedir. Yaptığı bu analizler doğrultusunda hem satış performansını hem müşteri bazlı olarak kazanç hesaplamaları yaparak şirketin verimliliği konusunda yorumlar yapmak istemektedir.

Yapılacak olan hesaplamalar için veri seti **first_question_data.feather** olarak paylaşılmıştır. Paylaşılan veri setini açıklamak için aşağıda örnek bir veri seti paylaşılmıştır. (Aşağıdaki tabloda bulunan veri paylaşılan veri seti içerisinde alınmamıştır.)

id	customer_id	sales_date_id	amount
23405	241	20220101	1025
23406	4035	20220203	2306
23407	2030	20220501	2405
23408	241	20220108	78
23409	99	20220203	50

Bu sorunun cevaplarının **Python dosyası** üzerinden paylaşılması istenmektedir. Notebook dosyasında bulunan tüm kodların çalışıyor olması ve yorumların eksiksiz bir şekilde yazıyor olması son derece önemlidir. Sorular içerisinde **SQL** üzerinden sorgu yazmanız istenecek olup **SQL** sorgularını ilgili sorunun altına **SQL** kod parçacığı olarak **pdf** dosyasına eklemeniz gerekmektedir. Belirli sorularda hem yorumlamanız hem çıktı paylaşmanız istenmektedir, o sorulara verdiğiniz cevaplar pdf dosyası içerisinde olmalıdır.

Ad	Açıklama	Veri Tipi
id	Tablodaki satışa ait ID bilgisi	int
customer_id	Satışı gerçekleştiren müşteriye ait ID bilgisi	int
sales_date_id	Satışın gerçekleştiği tarih (YYYYMMDD)	int
amount	Satışın gerçekleştiği tutar bilgisi	float

- a) Gerçekleşmiş olan satışların gerçekleştiği tarihlere göre günlük olarak incelenmek istenmektedir. Her bir günde gerçekleşen **toplam satış adetleri**, **toplam tutar** ve **toplam satış gerçekleştiren müşteri adetlerinin** trendlerini açıklayan nasıl bir grafik çizilebilir? Görselleştirme ve hesaplamalar python üzerinde detaylı olarak gösterilmeli ve pdf dosyası içerisinde yorumlanmalıdır.
- b) Aylar içerisinde satın alım gerçekleştiren müşterilerin yapmış olduğu satın alımları dönemsel olarak karşılaştırmak istenmektedir. Her bir dönemde (örneğin 2023 - Ağustos) **müşteri başına düşen satış adet** ve **toplam tutarların** dağılımını açıklayan nasıl bir grafik çizilebilir? Görselleştirme ve hesaplamalar python üzerinde detaylı olarak gösterilmeli ve pdf dosyası içerisinde yorumlanmalıdır.
- c) İlgili veri seti içerisinde geniş bir tarih aralığındaki satış verileri verilmiştir. Bu verilen veriler içerisinde bir veri manipülasyonu işlemi yapılmak istenmektedir. X tarihinde yapılan satışlar satış yapıldığı tarihten 20 gün önceki satışlarla birlikte değerlendirilerek son 20 gündeki tüm satışların **max**, **ortalama** ve **medyan değerleri** günlük olarak hesaplanmak istenmektedir. Örneğin 2023-06-20 tarihi için hesaplama yaparken 2023-06-01 ile 2023-06-20 tarihleri arasındaki tüm satışlar göz önünde bulundurularak max, medyan ve ortalama değer hesaplaması yapılacaktır. Ulaşılmaması gereken çıktıya ait bir kesit olarak aşağıda paylaşılmıştır. Python notebook üzerinde detaylı kod yazımı ve output çıktısı aşağıda belirtildiği gibi olması beklenmektedir.

sales_date	max	medyan	average
1.10.2022	6.345,68	3.722,01	3.362,84
10.12.2022	13.696,24	3.884,46	3.941,89
25.02.2023	17.394,23	4.466,21	4.648,65
30.03.2023	26.361,89	4.874,42	5.122,07

d) Bu partta olan tüm sorular **SQL** sorusu olup verilecek olan cevaplar pdf dosyası içerisinde ilgili soruların altına yazılmalıdır. Verilmiş olan verinin veri tabanında kayıtlı **customer.sales_detail** tablosunda kayıtlı olduğunu düşünelim.

- I. Aylık olarak müşterilerin satış adetleri ve toplam tutarlarını hesaplayan örnek bir SQL sorgusu hazırlayın.
 - II. Aylık bazda 100'den fazla satış gerçekleştiren customer_id'leri output olarak veren örnek bir SQL sorgusu hazırlayın.
 - III. Aylık bazda 100'den fazla satış gerçekleştirmiş ve gerçekleşen satışlarının toplam tutarı 20000 liradan fazla olan customer_id'leri output olarak veren örnek bir SQL sorgusu hazırlayın.
2. B isimli firma internet sitesi üzerinden müşterilerine hizmet vermektedir. Bu firma internet sitesinde gerçekleşen satış adetlerini çok yakından takip etmekte olup ani artış ve azalışlarda alarma geçmektedir. Mevcut yapılarında bunu tamamen bir grup insan tarafından izlenerek kendilerine ait oluşturdukları kural setleriyle olağan dışı bir durum olup olmadığını manuel olarak tespit etmektedir. Analitik ekip olarak buradaki insan gücünü tamamen bir istatistiksel bir modele bırakacak olsaydık burada nasıl bir modelleme yapısı kurmamız gerekirdi? Bu modelleme hangi problemlere benzetilebilir? Hangi veriler ile çalışılabilir? Bu modeli canlıya aldığımızda başarı kriterleri bizim için ne olurdu? Canlıda böyle bir modeli nasıl çalıştırabilirdik? Hangi modelleme teknikleri burada başarılı olurdu? Sorular örnek vermek amacıyla yazılmış olup, genel olarak bu konuya olan yaklaşımınız değerlendirilecektir. Bu sorular dışında yazacağınız tüm cevaplar değerlendirilecek olup tüm cevapların pdf dosyasının içerisinde yorumlanarak cevaplanması beklenmektedir. Bu soru özelinde kod yazımı beklentisi bulunmamaktadır.
3. C isimli bir firma **1.000'** den fazla kategoride **10.000.000'** dan fazla ürün satışı yapmaktadır. E ticaret firması çalıştığı firmalar ile yapılan satışlar üzerinden **ürün, dönem ve kategori** bazında farklı oranlarda olmak üzere **satış tutarı** üzerinden **komisyon** anlaşmaları yapmaktadır.

Firma, 2022 yılı sonu geldiğinde yapılan tüm satışlardan elde edilen komisyon gelirinin hesaplamasını yapmak istemektedir. Bu hesaplamayı yapabilmesi için aşağıda belirtildiği gibi bir veri setine sahiptir.

Aşağıdaki tablo veri tabanı üzerinde tutulmaktadır. Veri tabanı star schema mantığı ile kurulmuş olup her bir id'nin bir karşılığı farklı tablolarda bulunmaktadır.

sales_id	category_id	product_id	sales_datetime	sales_amount
23405	2030	12349	20.02.2022 23:45	1025
23406	4035	23941	21.08.2022 13:15	2306
23407	2030	53869	2.12.2022 07:45	2405
23408	241	438593	10.05.2022 12:15	78
23409	99	20394	23.06.2022 23:10	50

Komisyon tutarı = sales_amount * komisyon oranı şeklinde hesaplanacaktır.

- a) Star schema mantığına uygun bir şekilde bir veya birden fazla tablo yaratarak komisyon oranları veri tabanı tablolarında hangi sütunlar açılarak kolayca komisyon oranı hesaplaması yapılabilir?
- b) Bu tablo veya tablolar yaratıldıktan sonra nasıl bir SQL sorgusu ile kolayca komisyon tutarı hesaplaması yapılabilir?

Ürünlerin veya kategorilerin sahip oldukları komisyon oranları rastgele olarak belirlenebilir, önemli olan kurulacak olan tablo/tabloların hangi sütunlardan oluşacağını belirlenmesidir. Yukarıda verilmiş olan tablo örnek bir tablodur bu tablodan yola çıkarak sorulara cevap vermeniz beklenmektedir.

Beklenen çıktı : Oluşturulacak olan tablonun sütun adları, sütun veri tipleri ve sütun açıklamalarının pdf dosyasına ilgili sorunun altına eklenmesi ve oluşturulacak olan yeni tablo/tablolar ile komisyon tutarının SQL sorgusu ile hesaplanacak kodun pdf dosyasına yazılması.