

SÜLEYMAN DEMİREL ÜNİVERSİTESİ
MÜHENDİSLİK FAKÜLTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ
VERİ MADENCİLİĞİ VİZE SINAV SORULARI

ADI SOYADI:.....NO:.....

Sınav Yönergesi:

1. Sınav Süresi 30 dakikadır
2. Soruların puanları üzerinde belirtildiği gibidir.

1.(15p) Yapılan bir araştırmada, Isparta'daki konut su kaynaklarından birinde KSKD2 adlı moleküler bir bileşik keşfedilmiştir. Su kalitesi araştırmacıları, bileşiği yutmanın tamamen zararsız olmasına rağmen, bileşiğin vücuttaki varlığının aslında CNAM-10 DNA dizisine sahip kişilerin bağışıklık sistemlerini etkileyebileceğini belirlemişlerdir. Araştırma ekibinin bir parçası olarak, KSKD2 moleküler bileşiği içeren suyun düzenli olarak alınmasının CNAM-10 DNA dizisine sahip Isparta sakinlerini nasıl etkilediğini araştırmak için bir araştırma çalışması yürütmek istiyorsunuz.

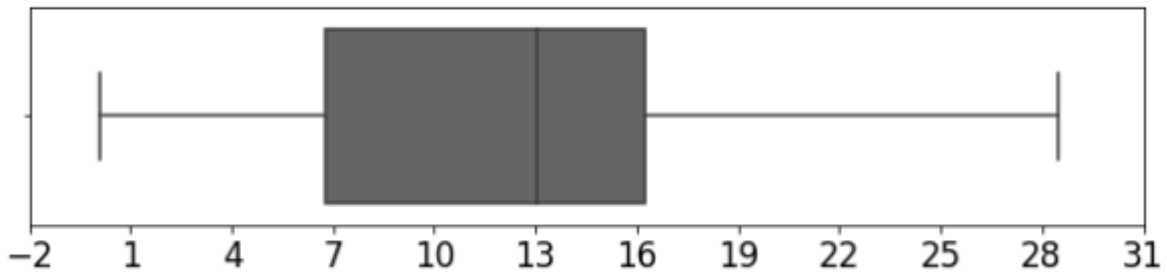
Bu durumda çalışacağınız popülasyonu yazınız?

CNAM-10 DNA dizisine sahip olan ve düzenli olarak KSKD2 molekülü su kaynağından su içen tüm Isparta sakinleri.

2. SDÜ öğrencilerinin bir paket yemek uygulaması olan Yemek Sepeti ile nasıl etkileşimde bulunduklarını inceleyen bir araştırma ekibindesiniz. Ekibiniz, ilk beş satırı aşağıda gösterilen YemekSepeti DataFrame'inde 1000 YemekSepeti verisi kaydı topladı ve temizledi.

	Order ID	Restaurant	Amount Paid	Wait Time	Calories
0	255334	Thai Basil	15.89	12.76	565.39
1	736425	Rice and Bones	5.00	2.34	475.13
2	853635	La Burrita	13.29	7.59	373.89
3	134546	Stuffed Inn	12.50	20.55	360.23
4	953624	Yifang	4.89	0.64	332.21

Wait Time alanı için box-plot grafiği aşağıdaki gibi çizilmiştir. Bu grafiğe göre:



a)(5p) Medyan değeri kaçtır?

13

b) (10p) Ortalama değeri kaçtır?

Belirlenemez

c) (5p) Interquartile Range değerini hesaplayarak bulunuz?

$IQR = 16 - 7 = 9$

3. Dogs dataframe’i, evcil hayvanların bir veterinerin ofisine yaptığı ziyaretlerle ilgili bilgileri içermektedir. Dataframe’in ilk 6 satırı aşağıda gösterilmiştir.

	age	color	fur	name
id				
123	4	brown	shaggy	odie
456	3	grey	short	gabe
821	6	golden	curly	samosa
198	4	grey	shaggy	gabe
3	2	black	curly	bob barker
42	5	brown	shaggy	odie

Aşağıdaki sorguları dogs dataframe’i için yazınız.

a) (15p)select name from dogs where age in (select max(age) from dogs)

dogs[dogs.age == dogs.age.max()].name

b) (20p) select sum(age) from dogs a right join (select fur, count(*) from dogs group by fur having count(*) > 2) b on a.fur=b.fur;

dogs[dogs.fur == dogs.fur.value_counts().index[0]].age.sum()

4.(30p)

	name	age	score	age_missing_data	income	gender		name	age	score	age_missing_data	income	gender	bonus
0	James	30	90%	30.0	100000	Male	0	James	30	90%	30.0	100000	Male	833.333333
1	Jane	40	95%	40.0	80000	Female	1	Jane	40	95%	40.0	80000	Female	600.000000
2	Melissa	32	100%	32.0	55000	Female	2	Melissa	32	100%	32.0	55000	Female	515.625000
3	Ed	67	82%	67.0	62000	Male	3	Ed	67	82%	67.0	62000	Male	231.343284
4	Neil	43	87%	NaN	120000	Male	4	Neil	43	87%	NaN	120000	Male	697.674419

Şekil 1.

Şekil 2.

employee DataFrame’i Şekil 1’den Şekil 2’deki duruma getirilmiştir. Bu sonucu elde etmek için gerekli kodu yazınız. Son sütun için hesaplama şu şekilde yapılmıştır. Erkekler için : $\text{income} / \text{age} * 0.25$ Kadınlar için : $\text{income} / \text{age} * 0.30$. B

```
def bonus(row):
    if row.iloc[5] == "Male":
        return row.iloc[4] / row.iloc[1] * 0.25
    else:
        return row.iloc[4] / row.iloc[1] * 0.30

employee["bonus"] = employee.apply(bonus, axis=1)
```