

Eksplorativna analiza kvaliteta sna i njegova korelacija sa zdravljem i načinom života



Odabir skupa podataka

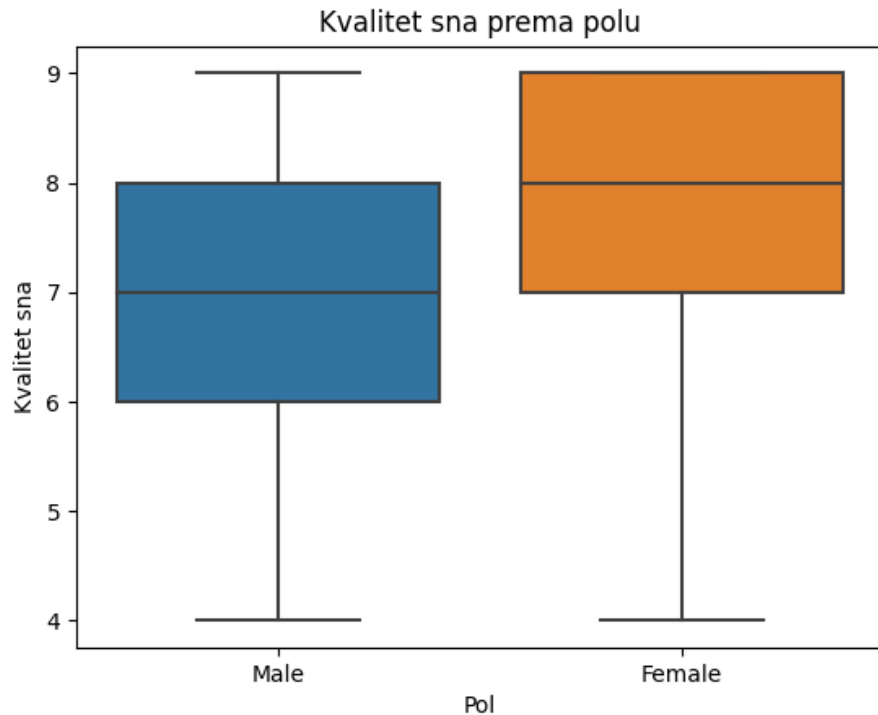
- Sleep Health and Lifestyle Dataset
- Sleep Health:
 - Duration of Sleep (5.8 – 8.5)
 - Quality of Sleep (4 - 9)
 - Sleep Disorder (Insomnia, Sleep apnea)
- Lifestyle:
 - Occupation (Software Engineer, Doctor, Sales Representative, Teacher, Nurse, Engineer, Accountant, Scientist, Lawyer, Salesperson, Manager)
 - Stress Level (3 - 8)
 - Physical Activity Level (30 - 90)
 - BMI Category (Normal, Normal Weight, Overweight, Obese)
 - Daily Steps
- Zdravlje kardiovaskularnog sistema:
 - Heart rate
 - Blood pressure
- Age (27 - 59)
- Gender (Male, Female)



Kvalitet sna u zavisnosti od pola

```
sb.boxplot(x='Gender', y='Quality of Sleep', data=data)
plt.xlabel('Pol')
plt.ylabel('Kvalitet sna')
plt.title('Kvalitet sna prema polu')
plt.show()
```

- Da li je razlika u kvalitetu sna statistički značajna?





ANOVA

• Filtriranje podataka

```
male_data = data[data['Gender'] == 'Male']['Quality of Sleep']
female_data = data[data['Gender'] == 'Female']['Quality of Sleep']

#ANOVA analiza (oneway)
anova, p_value = f_oneway(male_data, female_data)

#ispis rezultata
print('Anova: ', anova)
print('p vrednost: ', p_value)
```

•

Output:

Anova: 34.510311389862665

p vrednost: 9.416446532689026e-09

- Razlika je statistički značajna



Korelacija između godina i kvaliteta sna

dve kontinuirane varijable

```
age = data['Age'] # x-osa
quality_of_sleep = data['Quality of Sleep'] # y-osa
fit = np.polyfit(age, quality_of_sleep, 1)
fit_fn = np.poly1d(fit)

korelacija = np.corrcoef(age, quality_of_sleep)[0, 1]
print("Korelacija godina i kvaliteta sna: ", korelacija)
```

• Best fit line

• Rezultat će biti koeficijent korelacije (r)

Output:

Korelacija godina i kvaliteta sna: 0.4737338761619902

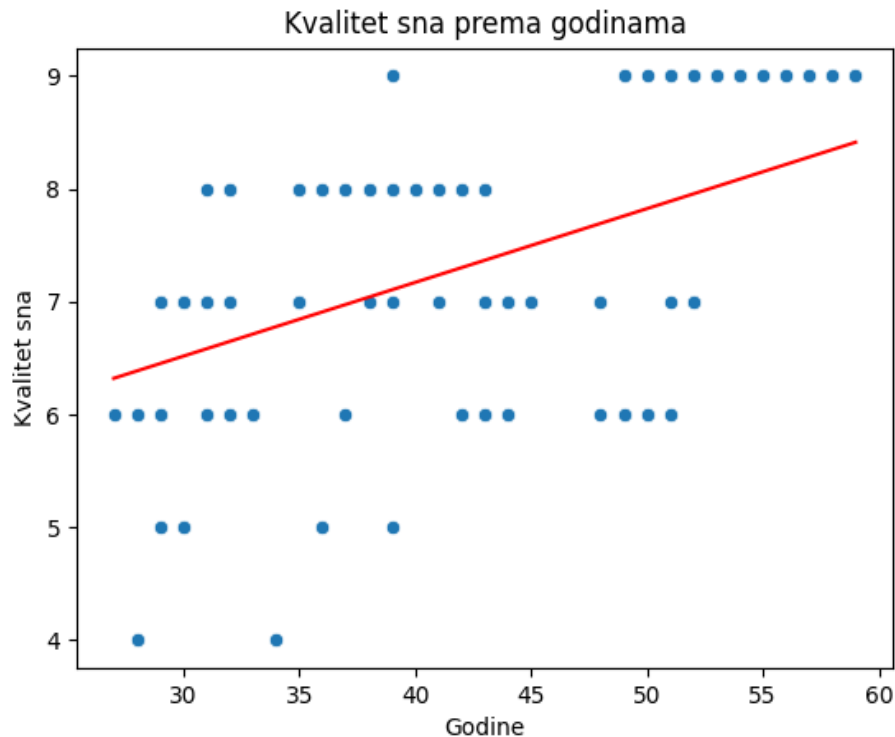
- $0.4 \leq |r| < 0.6$ – umereno jaka korelacija (pozitivna)



Kvalitet sna u zavisnosti od godina

```
sb.scatterplot(x='Age', y='Quality of Sleep', data=data)
plt.plot(age, fit_fn(age), color='red')
plt.xlabel('Godine')
plt.ylabel('Kvalitet sna')
plt.title('Kvalitet sna prema godinama')
plt.show()
```

- Best fit line: trend povecanja kvaliteta sna u odnosu na godine
- $r = 0.4737$
- Sa povecanjem godina raste i kvalitet sna





Korelacija između nivoa stresa i kvaliteta sna

```
stress = data['Stress Level']
quality_of_sleep = data['Quality of Sleep']
fit1 = np.polyfit(stress, quality_of_sleep, 1)
fit_fn1 = np.poly1d(fit1)

korelacija1 = np.corrcoef(stress, quality_of_sleep)[0, 1]
print("Korelacija nivoa stresa i kvaliteta sna: ", korelacija1)
```

• Best fit line

• Rezultat će biti koeficijent korelacije (r)

Output:

Korelacija nivoa stresa i kvaliteta sna: -0.8987520310040424

- $0.8 \leq |r| < 1.0$ – veoma jaka korelacija (negativna)



Kvalitet sna u zavisnosti od nivoa stresa

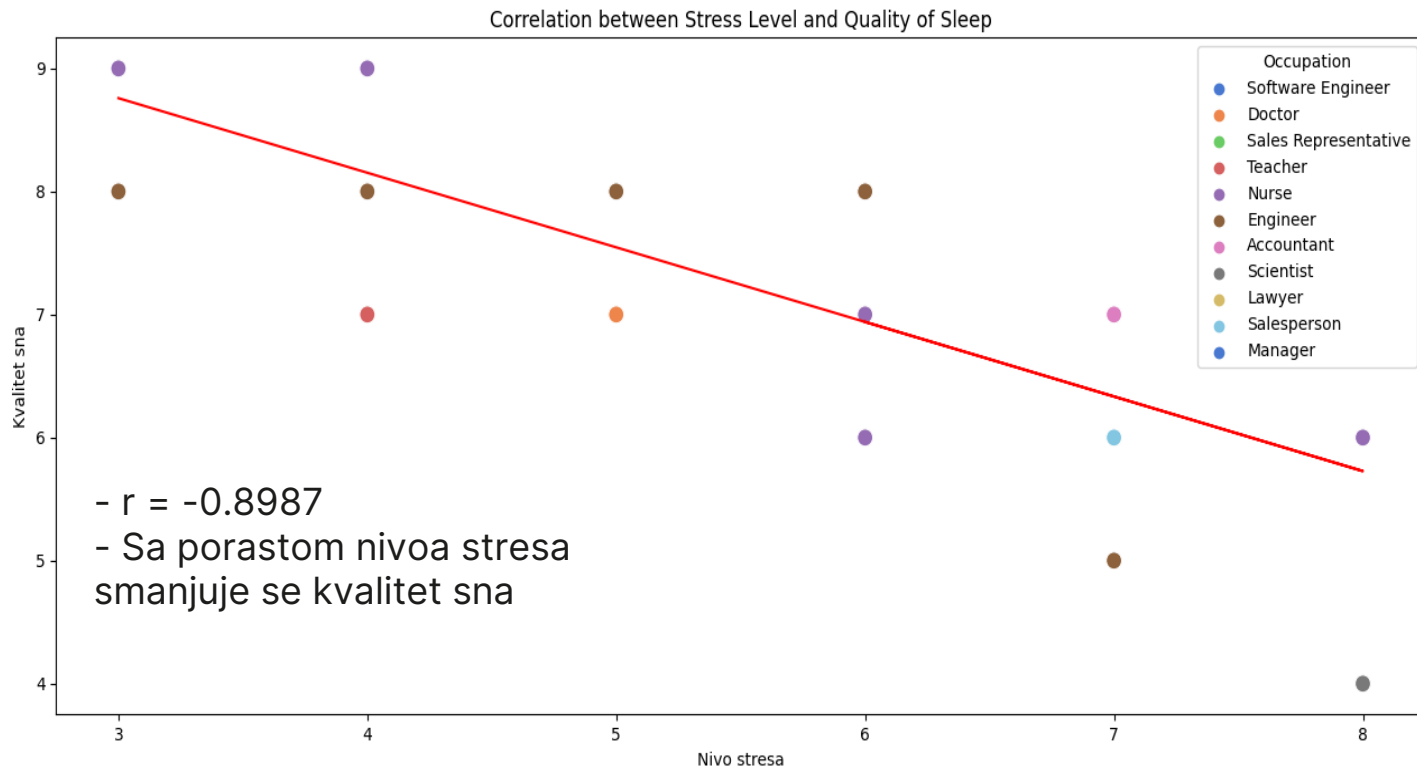
```
color = sb.color_palette('muted', len(data['Occupation'].unique()))  
  
sb.scatterplot(data=data, x='Stress Level', y='Quality of Sleep', hue='Occupation', palette=color, s=100)  
plt.plot(stress, fit_fn1(stress), color='red')  
plt.xlabel('Nivo stresa')  
plt.ylabel('Kvalitet sna')  
plt.title('Correlation between Stress Level and Quality of Sleep')  
plt.legend(title='Occupation')  
plt.tight_layout()  
plt.show()
```

Generiše paletu boja koja će se koristiti na grafikonu

Različita zanimanja će biti prikazana različitom bojom



Kvalitet sna u zavisnosti od nivoa stresa

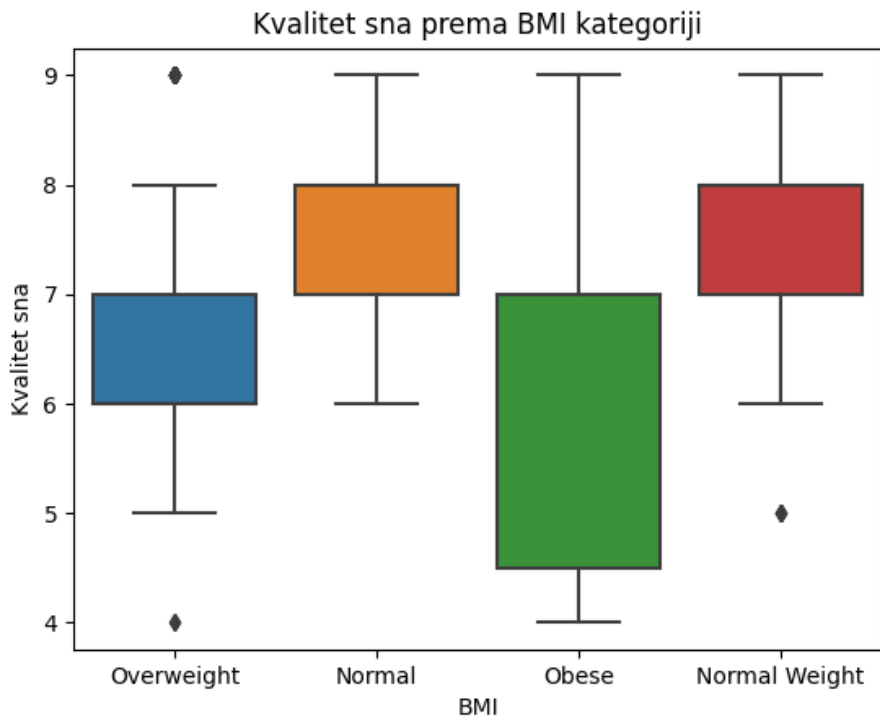




Kvalitet sna u zavisnosti od BMI kategorije

```
sb.boxplot(x='BMI Category', y='Quality of Sleep', data=data)
plt.xlabel('BMI')
plt.ylabel('Kvalitet sna')
plt.title('Kvalitet sna prema BMI kategoriji')
plt.show()
```

- Da li je razlika u kvalitetu sna statistički značajna?





ANOVA između svake BMI kategorije

```
bmi_categories = ['Normal', 'Normal Weight', 'Overweight', 'Obese']
results = pd.DataFrame(index=bmi_categories, columns=bmi_categories, dtype=str)

for category1 in bmi_categories:
    for category2 in bmi_categories:
        if category1 == category2:
            results.loc[category1, category2] = '-'
        else:
            data_category1 = data[data['BMI Category'] == category1]['Quality of Sleep']
            data_category2 = data[data['BMI Category'] == category2]['Quality of Sleep']

            anova_result = f_oneway(data_category1, data_category2)

            nivo_znacajnosti = 0.05
            if anova_result.pvalue < nivo_znacajnosti:
                results.loc[category1, category2] = '✓'
            else:
                results.loc[category1, category2] = 'X'

results = results.rename_axis("Statisticki znacajna razlika kvaliteta sna", axis="columns")
print(results)
```



ANOVA između svake BMI kategorije - rezultat

Statistički značajna razlika kvaliteta sna	Normal	Normal Weight	Overweight	Obese
Normal	-	X	✓	✓
Normal Weight	X	-	X	X
Overweight	✓	X	-	X
Obese	✓	X	X	-

Rezultujuća tabela koja prikazuje da li postoji statistički značajna razlika između svake dve BMI kategorije



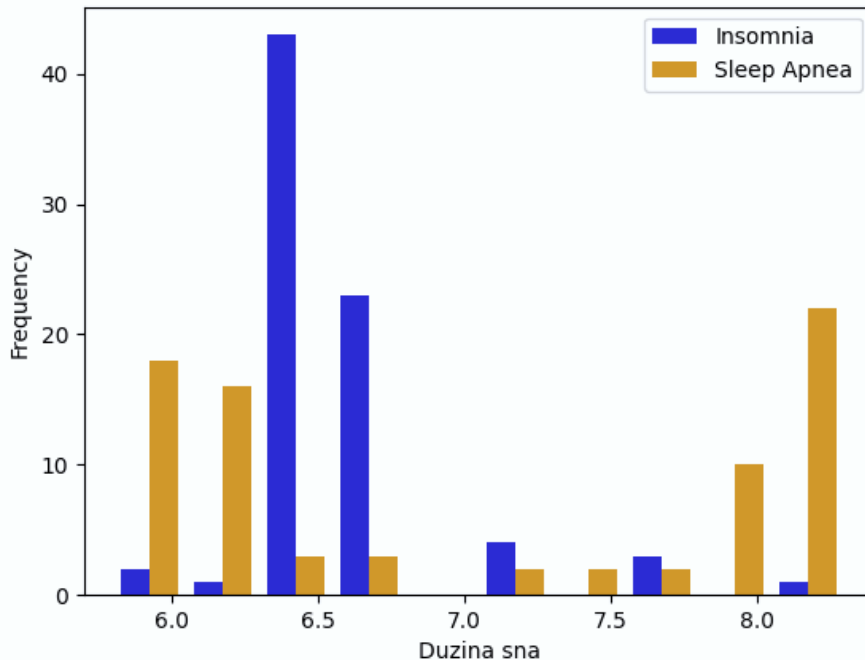
Dužina sna u zavisnosti od prisutnosti poremećaja sna

```
insomnia_data = data[data['Sleep Disorder'] == 'Insomnia']['Sleep Duration']  
sleep_apnea_data = data[data['Sleep Disorder'] == 'Sleep Apnea']['Sleep Duration']
```

```
plt.hist([insomnia_data, sleep_apnea_data], bins=  
plt.xlabel('Duzina sna')  
plt.ylabel('Frequency')  
plt.title('Histogram koji prikazuje duzinu sna sa  
plt.legend()  
plt.show()
```

- Sleep Apnea: uglavnom nema intermedijarne vrednosti
- Veći deo ispitanika sa insomnijom ima dužinu sna oko 6.5 sati

Histogram koji prikazuje duzinu sna sa obeleženim poremećajima sna





Korelacija između kvaliteta sna i otkucaja srca

```
heart_rate = data['Heart Rate'] # x-osa
sleep_quality = data['Quality of Sleep'] # y-osa
fit4 = np.polyfit(heart_rate, quality_of_sleep, 1)
fit_fn4 = np.poly1d(fit4)

korelacija = np.corrcoef(heart_rate, sleep_quality)[0, 1]
print("Korelacija kvaliteta sna i otkucaja srca: ", korelacija)
```

• Best fit line

• Rezultat će biti koeficijent korelacije (r)

Output:

Korelacija kvaliteta sna i otkucaja srca: -0.6598647331480089

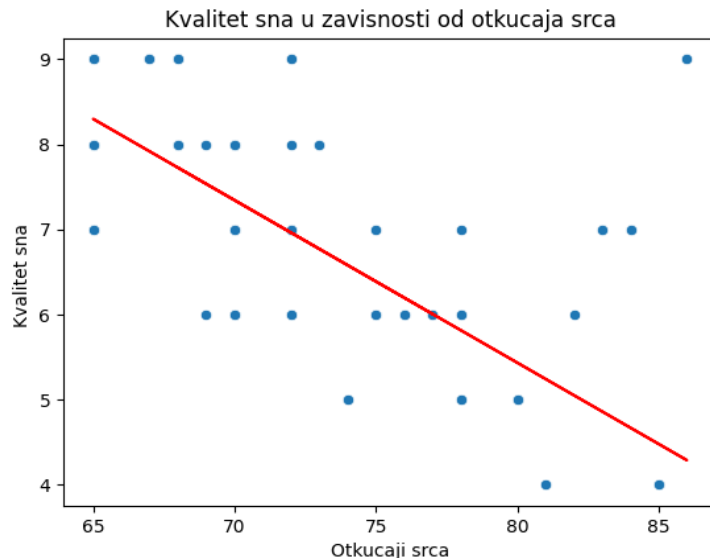
- $0.6 \leq |r| < 0.8$ – jaka korelacija (negativna)



Kvalitet sna u zavisnosti od otkucaja srca

```
sb.scatterplot(x='Heart Rate', y='Quality of Sleep', data=data)
plt.plot(heart_rate, fit_fn4(heart_rate), color='red')
plt.xlabel('Otkucaji srca')
plt.ylabel('Kvalitet sna ')
plt.title('Kvalitet sna u zavisnosti od otkucaja srca')
plt.show()
```

- Best fit line: trend smanjenja kvaliteta sna u odnosu na otkucaje srca
- $r = -0.6598$
- Sa povećanjem otkucaja srca smanjuje se kvalitet sna





Korelacija između kvaliteta sna i fizičke aktivnosti

```
physical_activity = data['Physical Activity Level'] # x-osa
quality_of_sleep = data['Quality of Sleep'] # y-osa
fit2 = np.polyfit(physical_activity, quality_of_sleep, 1)
fit_fn2 = np.poly1d(fit2)

korelacija = np.corrcoef(physical_activity, quality_of_sleep)[0, 1]
print("Korelacija fizičke aktivnosti i kvaliteta sna: ", korelacija)
```

• Best fit line

• Rezultat će biti koeficijent korelacije (r)

Output:

Korelacija fizičke aktivnosti i kvaliteta sna: 0.1928964549397526

- $0.0 < |r| < 0.2$ – veoma slaba korelacija (pozitivna)



Korelacija između dužine sna i fizičke aktivnosti

```
physical_activity = data['Physical Activity Level'] # x-osa
sleep_duration = data['Sleep Duration'] # y-osa
fit3 = np.polyfit(physical_activity, sleep_duration, 1)
fit_fn3 = np.poly1d(fit3)

korelacija = np.corrcoef(physical_activity, sleep_duration)[0, 1]
print("Korelacija fizičke aktivnosti i dužine sna: ", korelacija)
```

• Best fit line

• Rezultat će biti koeficijent korelacije (r)

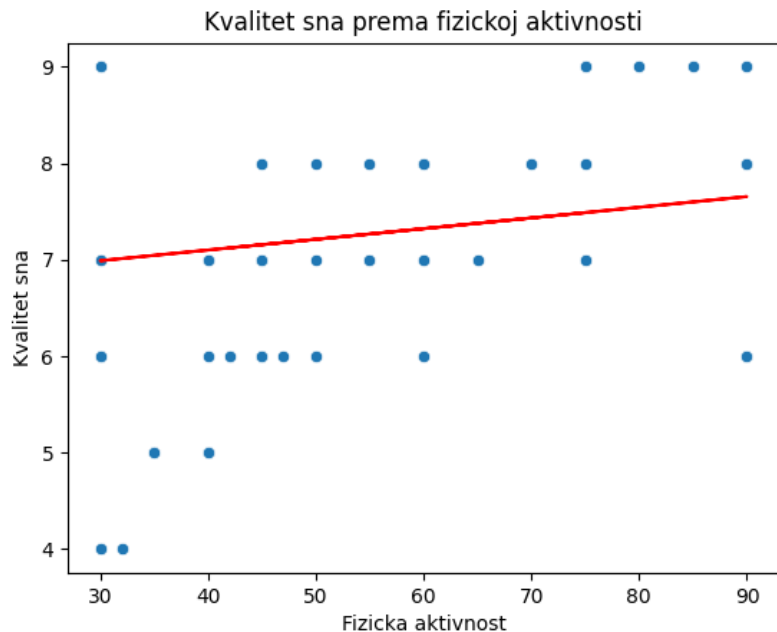
Output:

Korelacija fizičke aktivnosti i dužine sna: 0.2123603147257588

- $0.2 \leq |r| < 0.4$ - slaba korelacija (pozitivna)



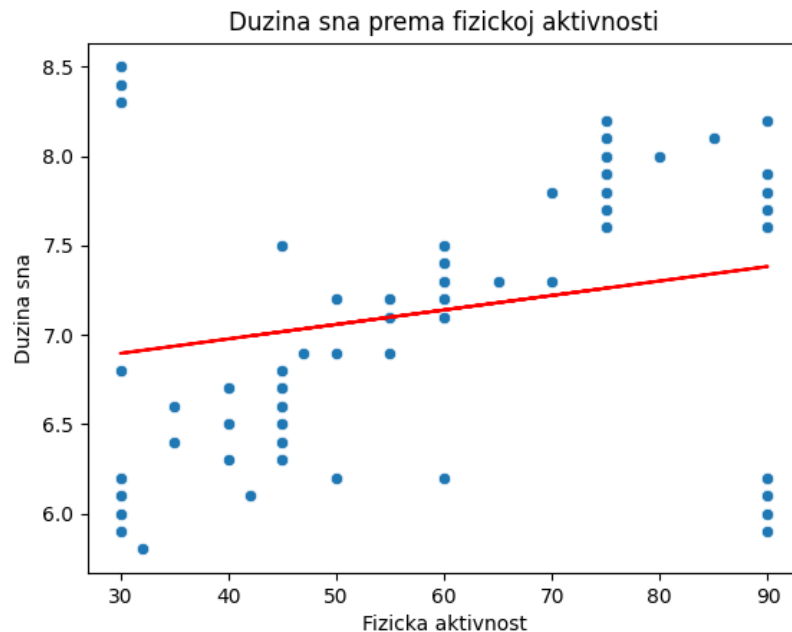
Kvalitet sna u zavisnosti od fizičke aktivnosti



- $r = 0.1928$



Dužina sna u zavisnosti od fizičke aktivnosti



- $r = 0.2123$

Hvala na pažnji!