KN31_Nalyvayko_Diana

March 11, 2024

0.0.1

“ ”

:

,

0.0.2

2019.

(data_m1.csv) , , .

Birth
Gender
Region /
TerType
PhysStatus
PhysBall
MathStatus
MathBall

0.0.3

[1]:
```
#
import numpy as np
import pandas as pd

#
from matplotlib import pyplot as plt
import seaborn as sns
sns.set()
```

’ Pandas.DataFrame

[3]:
```
df = pd.read_csv("data_m1.csv", encoding = "cp1251")
df.head()
```

[3]:
```
   Birth    Gender                Region TerType        PhysStatus  PhysBall  \
0   2002                                                  141
1   2001                                                 124
2   2002                                                  0
3   2002                                                  108
4   2001                                                   189

         MathStatus  MathBall
0                    125
1                    119
2                    0
3                    133
4                    191
```

1 (2). ?

[7]:
```
print(len(df))
```

```
19101
```

2 (2). ?

[9]:
```
region = df[df['Region']='          ']

region['Birth'].max() - region['Birth'].min()
```

[9]:
```
34
```

3 (2). ?

[11]:
```
df.groupby('TerType').size()
```

[11]:
```
TerType
      15199
       3902
dtype: int64
```

4 (2). ( )?

[13]:
```
total_passed = df[df['MathStatus'] == '     ']['MathStatus'].count()
total_failed = df[df['MathStatus'] == '         ']['MathStatus'].count()
failed = (total_failed / (total_passed + total_failed)) * 100
print( round(failed))
```

```
11
```

5 (3). (,) ?

[15]:
```
physics = df[df['PhysStatus'] == '      ']
median_m = physics[physics['Gender'] == '     '][ 'PhysBall'].median()
median_f = physics[physics['Gender'] == '    '][ 'PhysBall'].median()
print(median_m - median_f)
```

```
-3.0
```

6 (3). 150 ( )?

[30]:
```
df_rivne = df[(df['Region'] == '          ') & (df['MathBall'] < 150)]
low_math = (len(df_rivne) / len(df[df['Region'] == '          '])) * 100
print(round(low_math))
```

```
50
```

7 (3). ( ) – (TotalBall). TotalBall 8 .

[32]:
```
df['TotalBall'] = df['PhysBall'] + df['MathBall']
sort = df.sort_values(by='TotalBall', ascending=False)
sort.head(8)
```

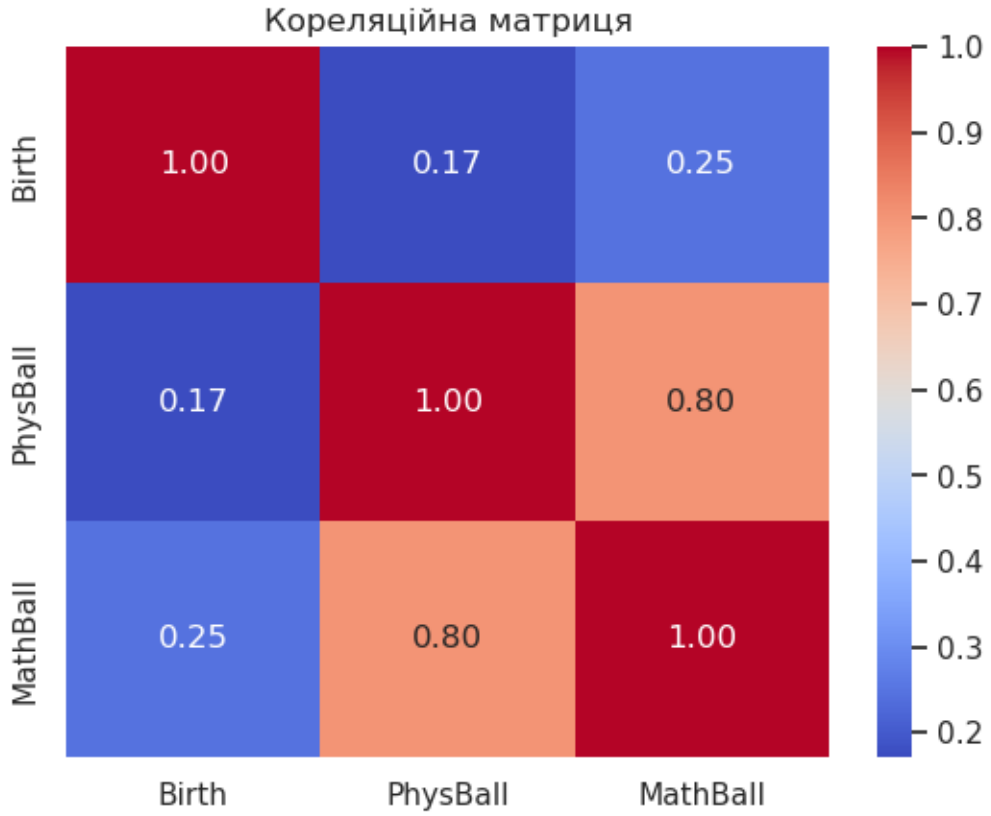
[32]:

	Birth	Gender	Region	TerType	PhysStatus	PhysBall	MathStatus	MathBall	TotalBall
15313	2001					200		200	400
11386	2002		.			199		200	399
13733	2002		-			199		200	399
8887	2002		.			200		199	399
14758	2002					199		200	399
8810	2002		.			198		200	398
12909	2002					198		200	398
1997	2001					199		199	398

8 (3). Birth, PhysBall MathBall, . PhysBall MathBall?

[34]:
```
cor_m = df[['Birth', 'PhysBall', 'MathBall']].corr(method='spearman')
sns.heatmap(cor_m, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('          ')
plt.show()
```

```
print("                PhysBall   MathBall:",cor_m.loc['PhysBall', 'MathBall'])
```



PhysBall MathBall: 0.8048384290601402

9 (2). , . ?

[36]:
```
all = len(df)
failed = df[df['MathStatus'] == '          ']
delete =  len(failed)
print((delete / all) * 100)
```

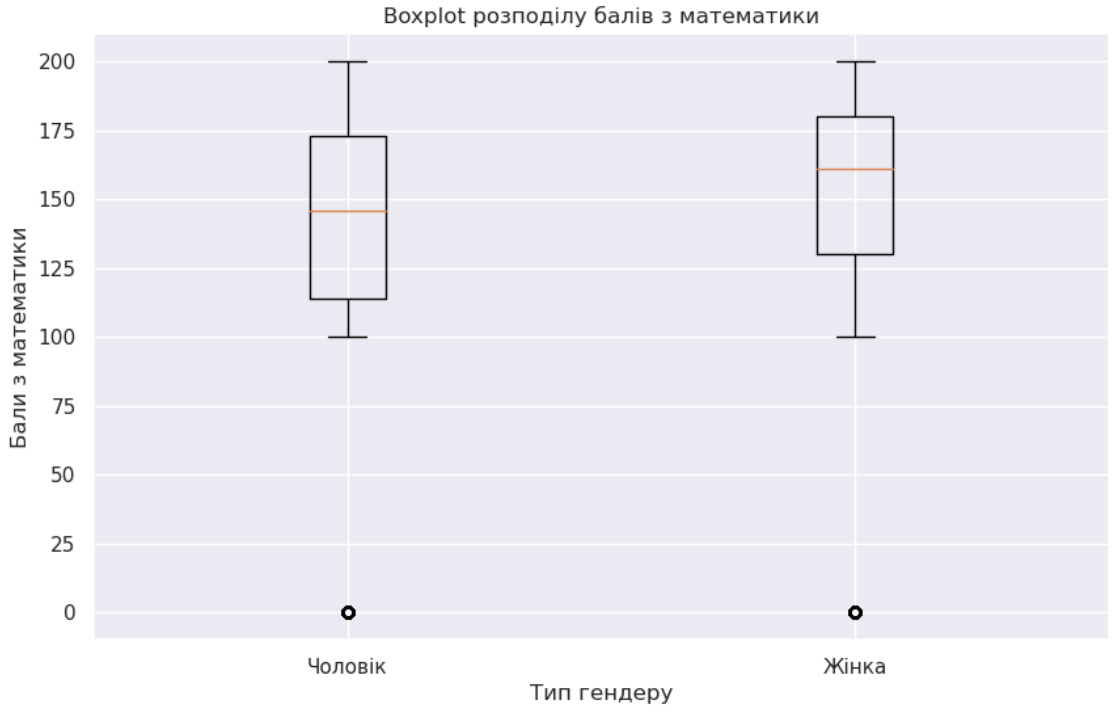
10.711481074289305

10 (2). boxplot- , MathBall :
.

[40]:
```
df_m = df[df['Gender'] == '     ']
df_f = df[df['Gender'] == '    ']
plt.figure(figsize=(10, 6))


plt.boxplot([df_m['MathBall'], df_f['MathBall']], labels=['    ', '   '])
```

```
plt.title('Boxplot                    ')
plt.xlabel('       ')
plt.ylabel('           ')
plt.grid(True)
plt.show()
```



[ ]: