מטלה 6

יבוא נתונים וסידור שמות עמודות ואינדקס

In [1]:

```python
import pandas as pd
import numpy as np
import re

df = pd.read_csv('LP_destinations.csv')
df = df.rename(columns = {'Unnamed: 0':'index'})
df.set_index('index', inplace=True)
df.head(5)
```

Out[1]:

index	city	description	countrys
0	Antalya	Once seen simply as the gateway to the Turkish...	Turkey
1	Athens	With equal measures of grunge and grace, Athen...	Greece
2	Ayia-napa	Endless blue-sky days and a coastline riddled ...	Cyprus
3	Bangkok	Same same, but different. This Thailish T-shir...	Thailand
4	Barcelona	NaN	Catalonia

ניקוי נתונים

In [2]:

```python
df = df[df['description'].notnull()]
df.head(5)
```

Out[2]:

index	city	description	countrys
0	Antalya	Once seen simply as the gateway to the Turkish...	Turkey
1	Athens	With equal measures of grunge and grace, Athen...	Greece
2	Ayia-napa	Endless blue-sky days and a coastline riddled ...	Cyprus
3	Bangkok	Same same, but different. This Thailish T-shir...	Thailand
6	Batumi	With a backdrop of mist-wrapped hills and soar...	Georgia

שאלה 1 ספירת מילים

In [3]:

```python
def count_words(string):
    return len(string.split())

df['num_of_words'] = df.apply(lambda row:count_words(row['description']), axis=1)
df.head(3)
```

Out[3]:

index	city	description	countrys	num_of_words
0	Antalya	Once seen simply as the gateway to the Turkish...	Turkey	112
1	Athens	With equal measures of grunge and grace, Athen...	Greece	391
2	Ayia-napa	Endless blue-sky days and a coastline riddled ...	Cyprus	148

שאלה 2 הסרת תווים

In [4]:

```python
def clear(words):
    return re.sub(r'[^\w\s]', '', words).strip()

df['No_punct_description'] = df.apply(lambda row:clear(row['description']), axis=1)
df.head(3)
```

Out[4]:

index	city	description	countrys	num_of_words	No_punct_description
0	Antalya	Once seen simply as the gateway to the Turkish...	Turkey	112	Once seen simply as the gateway to the Turkish...
1	Athens	With equal measures of grunge and grace, Athen...	Greece	391	With equal measures of grunge and grace Athens...
2	Ayia-napa	Endless blue-sky days and a coastline riddled ...	Cyprus	148	Endless bluesky days and a coastline riddled w...

שאלה 3 ספירת מילה

In [5]:

```python
def has_restaurants(words):
    return words.count('restaurant')

df['has_restaurants'] = df.apply(lambda row: has_restaurants(row['No_punct_description']),
df.head(3)
```

Out[5]:

index	city	description	countrys	num_of_words	No_punct_description	has_restaurants
0	Antalya	Once seen simply as the gateway to the Turkish...	Turkey	112	Once seen simply as the gateway to the Turkish...	0
1	Athens	With equal measures of grunge and grace, Athen...	Greece	391	With equal measures of grunge and grace Athens...	1
2	Ayia-napa	Endless blue-sky days and a coastline riddled ...	Cyprus	148	Endless bluesky days and a coastline riddled w...	0

שאלה 4 ספירת מילה

In [6]:

```python
def has_museums(words):
    return words.count('museum')

df['has_museums'] = df.apply(lambda row:has_museums(row['No_punct_description']), axis=1)
df.head(3)
```

Out[6]:

index	city	description	countrys	num_of_words	No_punct_description	has_restaurants	ha
0	Antalya	Once seen simply as the gateway to the Turkish...	Turkey	112	Once seen simply as the gateway to the Turkish...	0	
1	Athens	With equal measures of grunge and grace, Athen...	Greece	391	With equal measures of grunge and grace Athens...	1	
2	Ayia-napa	Endless blue-sky days and a coastline riddled ...	Cyprus	148	Endless bluesky days and a coastline riddled w...	0	

◄ ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬   ▶

שאלה 5 ספירת מילים

In [7]:

```python
def has_beaches(words):
    return (words.count('beach') + words.count('ocean') + words.count('sea'))

df['has_beaches'] = df.apply(lambda row: has_beaches(row['No_punct_description']), axis=1)
df.head(3)
```

Out[7]:

index	city	description	countrys	num_of_words	No_punct_description	has_restaurants	ha
0	Antalya	Once seen simply as the gateway to the Turkish...	Turkey	112	Once seen simply as the gateway to the Turkish...	0	
1	Athens	With equal measures of grunge and grace, Athen...	Greece	391	With equal measures of grunge and grace Athens...	1	
2	Ayia-napa	Endless blue-sky days and a coastline riddled ...	Cyprus	148	Endless bluesky days and a coastline riddled w...	0	

שאלה 6 ממוצע מילים

In [8]:

```python
df['num_of_words'].mean()
```

Out[8]:

250.30769230769232

שאלה 7 העיר עם התיאור הארוך ביותר

In [9]:

```python
max__num = df['num_of_words'].max()
line_max = df[df['num_of_words'] == max__num].reset_index()
line_max.loc[0,'city']
```
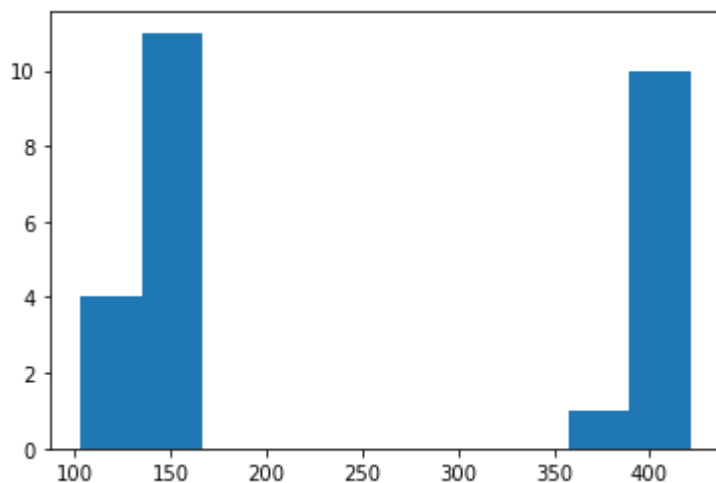
Out[9]:

'Lisbon'

שאלה 8 גרף

In [11]:

```python
import matplotlib.pyplot as plt

plt.hist(df['num_of_words'])
```

Out[11]:

```
(array([ 4., 11.,  0.,  0.,  0.,  0.,  0.,  0.,  1., 10.]),
 array([103. , 134.8, 166.6, 198.4, 230.2, 262. , 293.8, 325.6, 357.4,
        389.2, 421. ]),
 <BarContainer object of 10 artists>)
```
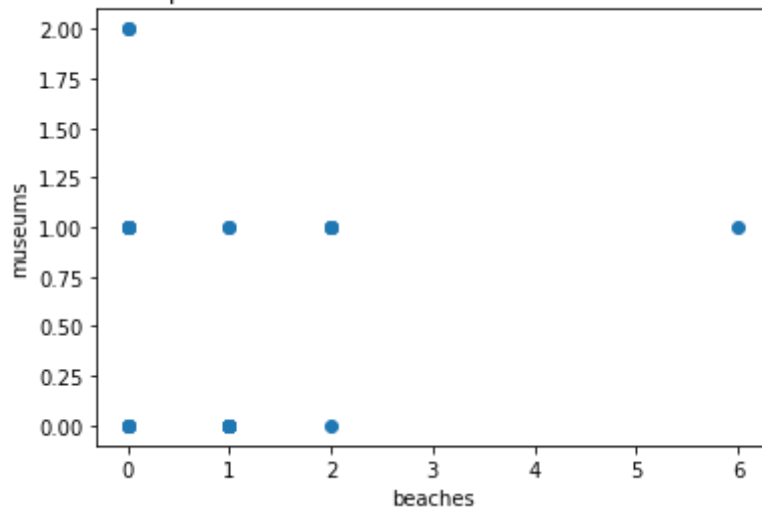

שאלה 9 יחסים

In [13]:

```python
x = df['has_beaches']
y = df['has_restaurants']
plt.scatter(x, y)

plt.xlabel("beaches")
plt.ylabel("museums")
plt.title("the relationship between number of beaches and number of restaurants")

plt.show()
```


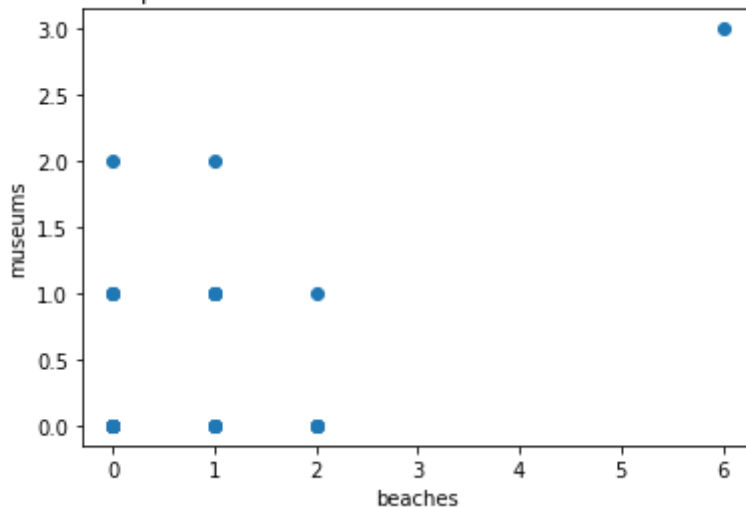the relationship between number of beaches and number of restaurants

שאלה 10 יחסים

In [15]:

```python
x = df['has_beaches']
y = df['has_museums']
plt.scatter(x, y)

plt.xlabel("beaches")
plt.ylabel("museums")
plt.title("the relationship between number of museums and the number of beaches")

plt.show()
```

the relationship between number of museums and the number of beaches

In [ ]: