**МИНОБРНАУКИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ  
<<МИРЭА>> - РОССИЙСКИЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ  
РТУ МИРЭА**  
ИНСТИТУТ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА  
КАФЕДРА БИОКИБЕРНЕТИЧЕСКИХ СИСТИМ И ТЕХНОЛОГИЙ  
Информационные системы в здравоохранении

Задача № 9

* Студент группы

КБМО-01-21

Ваннус. М

* Преподаватель:

Старичкова Ю. В.

**Москва – 2022**

**задание 9: распределение количества студентов по возрасту за анализируемые периоды:**

Основная цель задачи - изучить распределение студентов и сотрудников в течение всего периода исследования, задача будет решена следующим образом, сначала будет объяснена база данных и будет оценена как полезная или непригодная для исследования, затем данные будут продемонстрированы и проиллюстрированы графически, наконец, результаты предыдущего шага будут объяснены и обсуждены, чтобы получить представление и заключительные выводы:

**База данных:**

Приведенные данные представляют собой количество студентов и сотрудников в российских университетах, разделенных на 19 групп по возрасту, поскольку (моложе 15 лет, 15 лет, 16 лет, 17 лет, 18 лет, 19 лет, 20 лет, 21 год, 22 года, 23 года, 24 года, 25 лет, 26 лет, 27 лет, 28 лет, 29 лет, 30–34 года, 35–39 лет, 40 лет и старше) Приведенные данные представляют собой количество студентов и сотрудников в российских университетах, разделенных на 19 групп по возрасту, поскольку (моложе 15 лет, 15 лет, 16 лет,....., 28 лет, 29 лет, от 30 до 34 лет) студенты и сотрудники также демографически распределены по российским города до 84 городов. таким образом, набор данных строится из 1596 строк as (19 групп по годам для каждого города), а количество студентов и сотрудников делится в каждой строке на мужчин и женщин, как показано в таблицах 1 и 2.

Для конкретного пояснения в первой строке таблицы 1 указан возраст группы (младше 15 лет), а город - Алтайский край количество студентов мужского пола в этой группе равно 0, а количество студенток женского пола в группе также равно 0.

**Table 1. первые пять элементов базы данных**

|  |  |  |  |
| --- | --- | --- | --- |
| age | name | man | woman |
| моложе 15 лет | Алтайский край | 0 | 0 |
| моложе 15 лет | Амурская область | 0 | 0 |
| моложе 15 лет | Архангельская область | 0 | 0 |
| моложе 15 лет | Астраханская область | 0 | 0 |
| моложе 15 лет | Белгородская область | 1 | 0 |

**Table 2. последние пять элементов базы данных**

|  |  |  |  |
| --- | --- | --- | --- |
| age | name | man | woman |
| 40 лет и старше | Чеченская Республика | 193 | 130 |
| 40 лет и старше | Чувашская Республика | 637 | 683 |
| 40 лет и старше | Чукотский автономный округ | 4 | 2 |
| 40 лет и старше | Ямало-Ненецкий автономный округ | 15 | 1 |
| 40 лет и старше | Ярославская область | 305 | 438 |

Как упоминалось выше, возрастные группы разделены на 19 групп, но основная проблема заключается в том, что эти группы не разделены поровну, потому что есть 15 групп, которые представляют диапазон всего в один год (от "15 лет" до "29 лет"), с другой стороны, у нас есть две группы, которые представляют определенный диапазон ("моложе 15 лет" и "старше 40 лет"), в дополнение к этим двум у нас есть еще две группы с диапазоном в 5 лет, которые являются ("30-34 года" и "35-39 лет"). В результате эти четыре группы будут исключены из исследования, и будут приняты во внимание только 15 возрастных групп.

**результаты и обсуждение:**

1. Распределение студентов по Российской Федерации:

Чтобы иметь возможность получить общее представление о возрастном распределении студентов и сотрудников в российских вузах, мы рассмотрим общее число без учета демографического распределения.

Распределение студентов и служащих в университетах Российской Федерации показано на графике 1 - а график показывает, что наибольшее количество студентов в России приходится на возрастные группы 20 и 21 год, кривая плотности график 1 - b немного смещена вправо.

график 2 - а и график 2 - b, на которых показаны распределение и кривая плотности учащихся мужского пола в Российской Федерации, и график 2 - c и 3 - d, на которых показано распределение учащихся в Российской Федерации и кривая плотности. Оба распределения показывают сходство с распределением на графике 1, но число студенток превысило число студентов-мужчин в Российской Федерации на 200 тысяч, что считается очень высоким показателем и показывает, что существует огромная разница между числом студентов-женщин и студентов-мужчин.

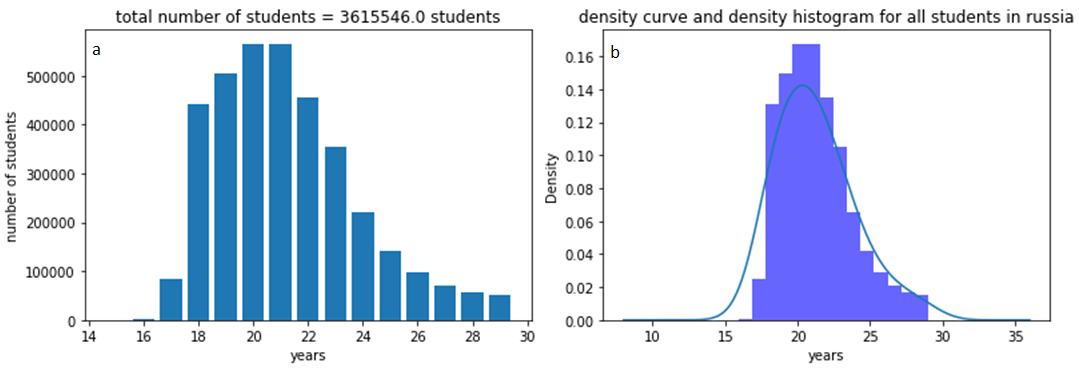


график 1. а - распределение студентов и служащих в Российской Федерации по возрасту, б - кривая плотности распределения по возрасту в Российской Федерации

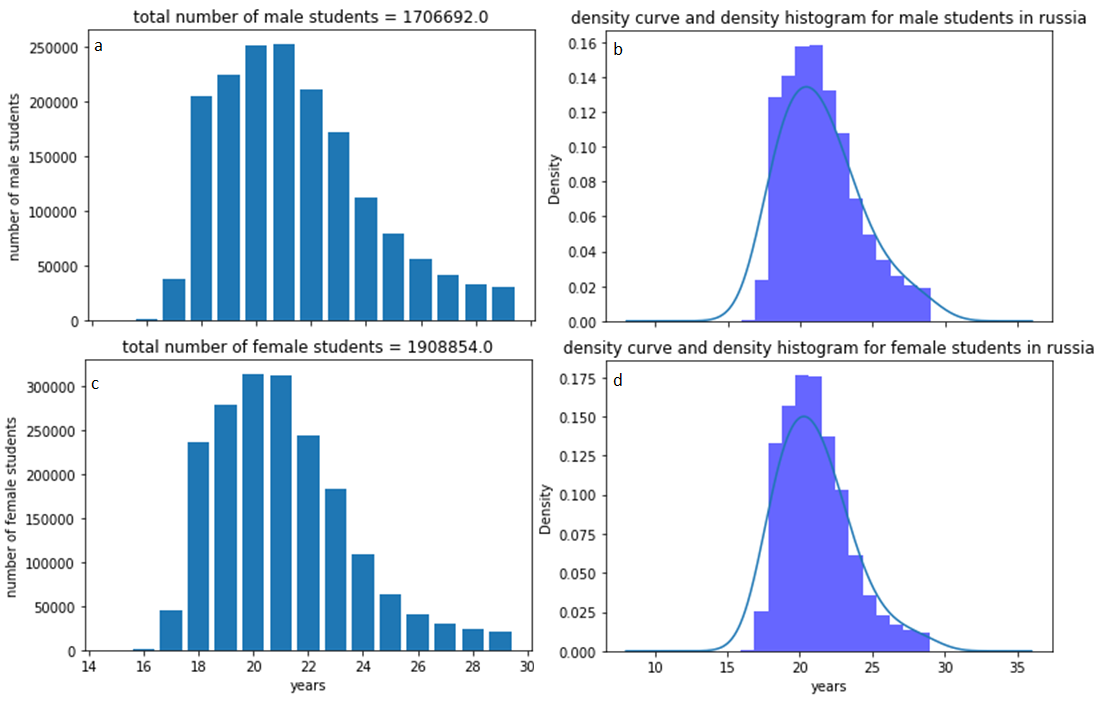


график 2. a и b показывают распределение и кривую плотности возраста студентов и служащих мужского пола в Российской Федерации во время исследования. c и d показывают распределение и кривую плотности возраста студенток и служащих в Российской Федерации во время исследования.

на графике 3 показана лучшая визуализация распределения студентов мужского и женского пола в Российской Федерации

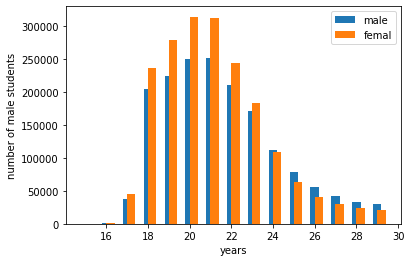


график 3 распределение студентов мужского и женского пола в Российской Федерации

1. Распределение студентов и служащих в Российской Федерации демографически:

Было нанесено распределение студентов и служащих в каждом городе Российской Федерации и получена кривая плотности. Также самые высокие пики были в диапазоне (19-21) лет, аналогичном основному распределению на графике 1. наиболее интересным наблюдением было то, что 20% студентов и служащих, проходящих обучение, были сосредоточены в Москва, в то время как 40% студентов находятся только в 6 регионах из 84 регионов с числом студентов и сотрудников, равным и выше 100 тыс. что приводит нас к одной из самых острых проблем в Российской Федерации - студенческой иммиграции, когда студенты предпочитают покидать свои города и поселки в поисках лучших университетов и возможностей трудоустройства в крупных городах России, что приводит к упадку этих маленьких городов и не оставляет шансов для улучшения и поддержания там численности населения.

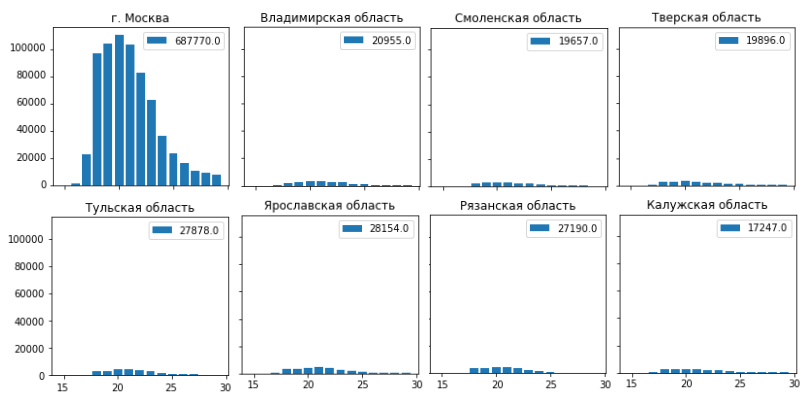


график 4. распределение по возрасту в городе Москве и еще семи регионах, близких к городу Москве

график 4 - показывает распределение по возрасту в городе Москве и еще 7 регионах, которые близки к Москве, график показывает, что в семи регионах очень низкое количество студентов и служащих в университетах по сравнению с Москвой, что показывает, что иммиграция студентов и рабочей силы из малых городов в большие города серьезно воздействуя на эти маленькие города и делая и без того перенаселенную Москву более населенной.

график 5 - показывает кривые плотности первых восьми регионов внутри базы данных, кривые плотности всех регионов показаны в записной книжке внутри папки задачи.

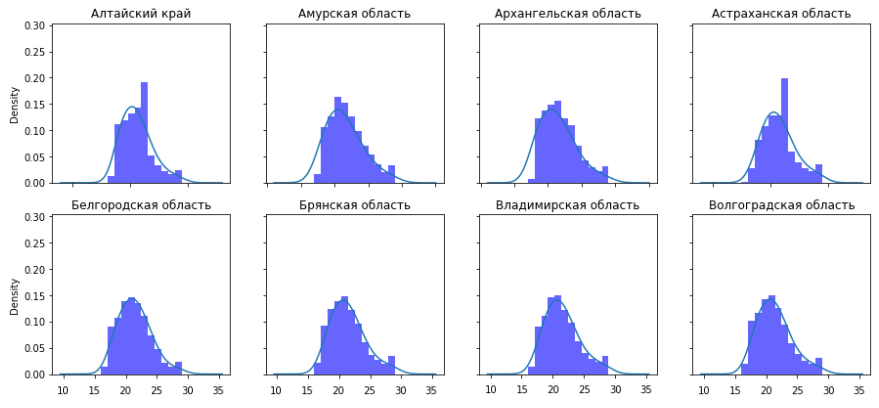


график 5. кривые плотности первых восьми регионов в базе данных

1. сравнение возрастных групп мужчин и женщин:

Распределение и кривые плотности мужской и женской возрастных групп были довольно схожими, с одним лишь основным отличием, которое заключается в том, что количество женских групп превысило количество мужских групп, что в целом обусловлено основной численностью населения в Российской Федерации, поскольку мы уже знаем, что население в Российской Федерации имеет большое количество самок по сравнению с количеством самцов.

на графике 6 показаны распределение и кривые плотности мужских возрастных групп для четырех регионов России, в то время как на графике 7 показаны распределение и кривая плотности тех же регионов, но для женских групп.

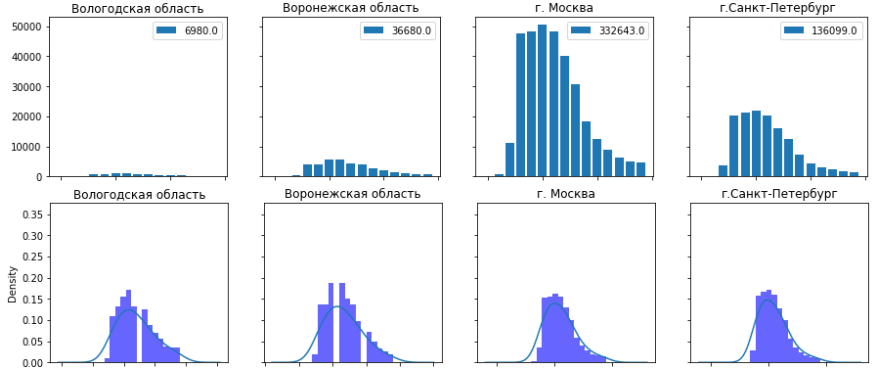


график 6. кривые распределения и плотности мужских возрастных групп

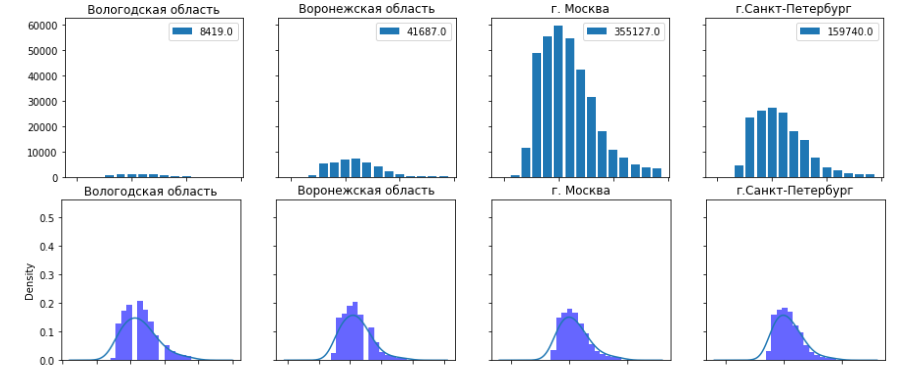


график 7. кривые распределения и плотности женских возрастных групп

график 8 показывает лучшую визуализацию распределения мужчин и женщин в каждом регионе Российской Федерации

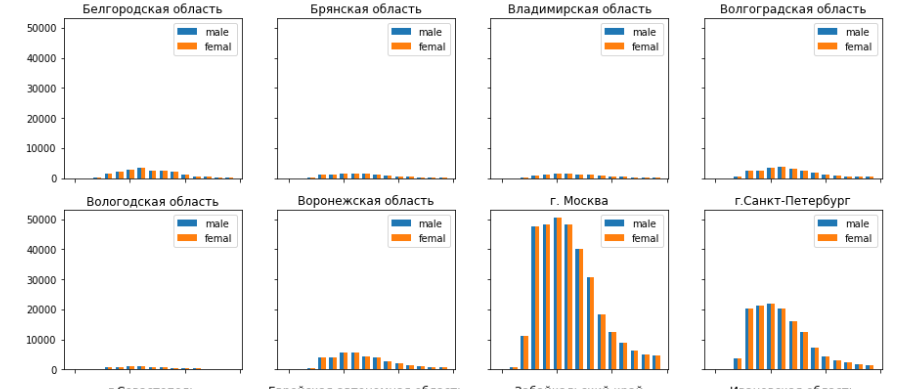


график 8 распределение мужчин и женщин в каждом регионе Российской Федерации

**Выводы:**

1. распределение возрастных групп студентов в диапазоне (15-29) лет смещено вправо с наибольшей концентрацией в диапазоне (19-21) лет.
2. Изучая демографическое распределение студентов, мы обнаружили высокую численность студентов в небольшом количестве крупных городов Российской Федерации, что обусловлено главным образом иммиграцией в большие города в поисках лучших университетов и лучших возможностей трудоустройства.
3. Общая форма распределения и кривые плотности студентов схожи по форме в разных регионах, но различаются с точки зрения численности населения.

**код:  
распространение по всей Российской Федерации**

# import the libraries

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# import the dataset

df = pd.read\_csv("students by age.csv",delimiter=';',encoding='cp1251')

df

# defining two arrays with the name of the unique age groups and the unique regions

ages = df.age.unique()

regions= df.name.unique()

regions.shape

# taking columns number [1,2,3,4] of the dataset into a numpy array

X = df.iloc[:,[0,1,2,3]].values

# defining a dictionary with the key "age" which contains an array of the required age groups for

# this study (15-29) and the key "val" which represents the number of students of an age group in whole Russia

d ={"age": np.linspace(15,29,15),"val":np.zeros(15)}

for i in range(1596):

ID=int(np.where(ages== X[i,0])[0])

if ID>0 and ID<16:

d["val"][ID-1]=(float(X[i,2]+X[i,3]))+d["val"][ID-1]

d = pd.DataFrame(d)

d

# distribution plot

leg = d["val"].sum()

plt.bar(d.age, d.val)

plt.title("total number of students = {} students".format(leg))

plt.xlabel("years")

plt.ylabel("number of students")

plt.show()

#density curve

a = pd.Series(d.age.repeat(d.val))

a.plot.kde(bw\_method=0.5)

plt.hist(a,bins=15, density=True, alpha=0.6, color='b')

plt.title("density curve and density histogram for all students in russia")

plt.xlabel("years")

plt.show

# create a data frame for male and female students

X = df.iloc[:,[0,1,2,3]].values

d ={"age": np.linspace(15,29,15),"men":np.zeros(15),"women":np.zeros(15)}

for i in range(1596):

ID=int(np.where(ages== X[i,0])[0])

if ID>0 and ID<16:

d["men"][ID-1]=(float(X[i,2]))+d["men"][ID-1]

d["women"][ID-1]=(float(X[i,3]))+d["women"][ID-1]

d = pd.DataFrame(d)

# plot male and female distribution

plt.bar(d.age, d.men,align="center",width=0.4,label="male")

plt.bar(d.age,d.women,align="edge",width=0.4,label="femal")

plt.xlabel("years")

plt.ylabel("number of male students")

plt.legend()

plt.show()

# density curve for male students

a = pd.Series(d.age.repeat(d.male))

a.plot.kde(bw\_method=0.5)

plt.hist(a,bins=15, density=True, alpha=0.6, color='b')

plt.title("density curve and density histogram for male students in russia")

plt.xlabel("years")

plt.show

# density curve for female students

a = pd.Series(d.age.repeat(d.female))

a.plot.kde(bw\_method=0.5)

plt.hist(a,bins=15, density=True, alpha=0.6, color='b')

plt.title("density curve and density histogram for male students in russia")

plt.xlabel("years")

plt.show

**распределение по регионам**

**# taking columns number [1,2,3,4] of the dataset into a numpy array**

**X = df.iloc[:,[0,1,2,3]].values**

**# defining a dictionary with the key "age" which contains an array of the required age groups for**

**# this study (15-29)**

**d ={"age": np.linspace(15,29,15)}**

**# adding empty keys to the dictionary each key represnts a region in Russia.**

**for i in regions:**

**d[i] = []**

**# filling the list of each region with number of students from the dataset**

**for i in range(1596):**

**ID=int(np.where(ages== X[i,0])[0])**

**if ID>0 and ID<16:**

**ID\_2 = X[i,1]**

**d[ID\_2].append(float(X[i,2]+X[i,3]))**

**d = pd.DataFrame(d)**

# plotting the distribution of the male and female students in each region

figure, axis = plt.subplots(21, 4,sharex=False, sharey=True,figsize=(15,75))

for i in range (1,85):

l=int((i-1)/4)

j=(i-1)%4

#axis[l, j].bar(d\_men.age[1:15], d\_men.iloc[1:15,i])

leg=d[regions[i-1]].sum()

axis[l, j].bar(d.age, d.iloc[:,i],label=leg)

axis[l, j].legend(loc ="upper right")

i=0

for ax in axis.flat:

ax.label\_outer()

ax.set\_title(regions[i])

i+=1

plt.show()

#plotting the density curve of male and female students in each region

figure, axis = plt.subplots(21, 4,sharex=False, sharey=True,figsize=(15,75))

for i in range (1,85):

l=int((i-1)/4)

j=(i-1)%4

a = pd.Series(d.age.repeat(d.iloc[:,i]))

a.plot.kde(ax=axis[l, j],bw\_method=0.5)

axis[l, j].hist(a,bins=15, density=True, alpha=0.6, color='b',rwidth=10)

i=0

for ax in axis.flat:

#ax.label\_outer()

ax.set\_title(regions[i])

i+=1

plt.show()

**нанесите распределение учащихся мужского и женского пола в каждом регионе на один и тот же график**

# create a data frame for male students in each region

# taking the first three columns of the dataset into a numpy array

X\_men = df.iloc[:,[0,1,2]].values

# defining a dictionary with the key "age" which contains an array of the required age groups for

# this study (15-29)

d\_men ={"age": np.linspace(15,29,15)}

# adding empty keys to the dictionary each key represnts a region in Russia.

for i in regions:

d\_men[i] = []

# filling the list of each region with number of students from the dataset

for i in range(1596):

ID=int(np.where(ages== X\_men[i,0])[0])

if ID>0 and ID<16:

ID\_2 = X\_men[i,1]

d\_men[ID\_2].append(float(X\_men[i,2]))

# conveting the dictionary to dataframe

d\_men = pd.DataFrame(d\_men)

# create a data frame for female students in each region:

# taking columns number [1,2,4] of the dataset into a numpy array

X\_women = df.iloc[:,[0,1,3]].values

# defining a dictionary with the key "age" which contains an array of the required age groups for

# this study (15-29)

d\_women ={"age": np.linspace(15,29,15)}

# adding empty keys to the dictionary each key represnts a region in Russia.

for i in regions:

d\_women[i] = []

# filling the list of each region with number of students from the dataset

for i in range(1596):

ID=int(np.where(ages== X\_women[i,0])[0])

if ID>0 and ID<16:

ID\_2 = X\_women[i,1]

d\_women[ID\_2].append(float(X\_women[i,2]))

d\_women = pd.DataFrame(d\_women)

# plotting the distribution of the male and female students in each region

figure, axis = plt.subplots(21, 4,sharex=False, sharey=True,figsize=(15,75))

for i in range (1,85):

l=int((i-1)/4)

j=(i-1)%4

#axis[l, j].bar(d\_men.age[1:15], d\_men.iloc[1:15,i])

axis[l, j].bar(d\_men.age ,d\_men.iloc[:,i],align="center",width=0.6,label="male")

axis[l, j].bar(d\_men.age,d\_men.iloc[:,i],align="edge" ,width=0.4,label="femal")

axis[l, j].legend(loc ="upper right")

i=0

for ax in axis.flat:

ax.label\_outer()

ax.set\_title(regions[i])

i+=1

plt.show()