

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение

высшего образования

«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»

Кафедра «Измерительно-вычислительные комплексы»

«Методы искусственного интеллекта»

Отчёт по лабораторной работе №3

Вариант №15

Выполнила:

студентка группы ИСТбд-42

Саушкина Наталья

Проверил:

доцент кафедры ИВК, к.т.н.

Шишкин В.В.

Ульяновск
2022

Задание:

1. Создать симулированный набор данных и записать его на диск в виде csv файла со следующими параметрами:

- количество строк не менее 1000 (задается случайным образом);

- структура набора:

- табельный номер;
- Фамилия И.О.;
- пол;
- год рождения;
- год начала работы в компании;
- подразделение;
- должность;
- оклад;
- количество выполненных проектов

2. Прочитать сгенерированный набор данных в виде списков и получить с помощью программирования и методов библиотеки numpy для разных по типу признаков столбцов (не менее 3) основные статистические характеристики (например для порядкового типа: минимум, максимум, среднее, дисперсия, стандартное отклонение, медиана, мода).

3. Прочитать сгенерированный набор данных в виде датафрейма и получить с помощью методов библиотеки pandas для тех же столбцов те же статистические характеристики. Продемонстрировать применение не менее 3 методов библиотеки pandas.

4. Построить не менее 3 разнотипных графиков.

5. Оценить возможности библиотек csv, numpy, pandas в форме отчета по лабораторной работе.

1. Листинг программы:

```
# data generation
```

```
class Employee:
```

```
    def __init__(self, number, name, gender, birth_year, start_work, department, post, salary, number_projects):
        self.number = number
        self.name = name
        self.gender = gender
        self.birth_year = birth_year
        self.start_work = start_work
        self.department = department
        self.post = post
        self.salary = salary
        self.number_projects = number_projects
```

```
def get_str(self):  
    return [self.number, self.name, self.gender, self.birth_year,  
            self.start_work, self.department, self.post, self.salary,  
            self.number_projects]
```

```
full_name = ('Сиборов У. А.', 'Мишин В. Ф.', 'Николаев Д. В.', 'Похрутин С. К.',  
             'Иванов Б. Д.',  
             'Ермилов В. Ч.', 'Никишкин Ш. Л.', 'Емельянов Д. К.', 'Связов К. В.',  
             'Кандалинцев А. Ц.',  
             'Макаров В. Р.', 'Малов У. Ш.', 'Буров Р. Д.', 'Гордеев С. Т.', 'Митин С.  
             Л.',  
             'Ганин В. С.', 'Горинович Ц. Т.', 'Синотов У. И.', 'Кузмячук В. Н.',  
             'Демянчук В. Л.',  
             'Гранин З. Л.', 'Бельский З. Д.', 'Плеханов З. В.', 'Желябов Х. В.', 'Богров З.  
             Н.',  
             'Семенова Е. А.', 'Коннова Г. С.', 'Койнова В. З.', 'Завьялова у. З.',  
             'Шевцова Г. Л.',  
             'Петрова Г. В.', 'Селезнева Ш. Л.', 'Рубина К. Т.', 'Князева О. Л.',  
             'Стрелецкая Ж. Г.',  
             'Плетнева Л. В.', 'Сибирцева Л. В.', 'Кумова Д. Г.', 'Шитова Л. А.',  
             'Максимова Г. З.',  
             'Тульцева К. Н.', 'Зиновьева Г. В.', 'Каменева Л. В.', 'Тахтилова В. К.',  
             'Железнова Ш. Л.')
```

```
department = ('отделение по работе с общественностью', 'секретариат',  
              'техническое отделение',  
              'отделение построения', 'служба упревления персоналом',  
              'отдел организации труда', 'бухгалтерия', 'служба управления  
              персоналом',  
              'финансовое подразделение', 'отдел внешнеэкономических связей',  
              'склады готовой продукции и материалов', 'планово-экономический  
              отдел',  
              'служба стандартизации', 'юридическая служба', 'отдел кадров',  
              'служба безопасности', 'вычислительный центр')
```

```
rank = ("руководитель отдела информационной безопасности", "начинающий  
        специалист",  
        "руководитель отдела разработки ПО", "senior разработчик", "middle  
        разработчик",  
        "руководитель отдела контроля качества и процессов", "специалист",  
        "работник",  
        "руководитель отдела развития ИТ", "специалист",  
        "руководитель отдела поддержки ИТ", 'отдел кадров',  
        'служба безопасности', 'вычислительный центр')
```

```
gender = ('муж', 'жен')
```

```

count_str = random.randint(1000, 1200)

workers = list()

for i in range(count_str):
    number = random.randint(1, 1500)
    name = random.choice(full_name)
    gen = random.choice(gender)
    birth_year = random.randint(1970, 2000)
    start_work = random.randint(2005, 2022)
    dep = random.choice(department)
    pos = random.choice(rank)
    salar = random.randint(15000, 70000)
    number_projects = random.randint(1, 15)
    person = Employee(number, name, gen, birth_year, start_work, dep, pos, salar,
                      number_projects)
    workers.append(person)

with open("data.csv", mode="w", encoding='utf-8') as w_file:
    file_writer = csv.writer(w_file, delimiter=",", lineterminator="\r")
    file_writer.writerow(["number", "name", "gender", "birth_year", "start_work",
                          "department", "rank", "salary",
                          "number_projects"])
    for per in workers:
        file_writer.writerow(per.get_str())

def get_info(max_res,min_res,average,disp,std,median):
    print('Max:',max_res,'\nMin:',min_res,'\nAverage:',average,
          '\nDisp:',disp,'\nStd:',std,'\nMedian:',median)

```

Содержимое data.csv после отработки кода:

```

1 number,name,gender,birth_year,start_work,department,rank,salary,number_projects
2 1002,Макаров В. Р.,жен,1996,2020,склады готовой продукции и материалов,главный инженер,69215,14
3 135,Малов У. Ш.,муж,1978,2005,техническое отделение,менеджер,15206,8
4 790,Шитова Л. А.,жен,1975,2015,отдел кадров,секретарь,64517,3
5 1384,Макимова Г. З.,жен,1997,2007,отдел кадров,главный инженер,67357,3
6 411,Никишкин Ш. Л.,жен,1973,2009,служба безопасности,секретарь,43336,13
7 1204,Ганин В. С.,муж,1997,2010,склады готовой продукции и материалов,руководитель,32024,2
8 467,Тулъцева К. Н.,муж,1998,2021,секретариат,заместитель директора,36323,3
9 1129,Железнова Ш. Л.,жен,1996,2013,отдел организации труда,главный инженер,34515,1
10 1005,Гранин З. Л.,муж,1977,2020,отдел кадров,заместитель директора,66639,1
11 1308,Митин С. Л.,жен,1995,2008,склады готовой продукции и материалов,главный инженер,33226,3
12 283,Мишин В. Ф.,жен,1990,2006,вычислительный центр,офис-менеджер,40133,8
13 451,Зиновьева Г. В.,жен,1981,2007,служба безопасности,руководитель,61211,2
14 843,Петрова Г. В.,муж,1994,2008,служба управления персоналом,заместитель директора,36101,4
15 1022,Селезнева Ш. Л.,жен,1995,2017,бухгалтерия,директор,24342,8
16 1345,Николаев Д. В.,жен,1971,2015,техническое отделение,инженер,50477,4
17 660,Койнова В. З.,муж,2000,2020,секретариат,начальник,37300,15
18 563,Петрова Г. В.,жен,1980,2015,отделение по работе с общественностью,начальник,50350,10
19 173,Тулъцева К. Н.,муж,1978,2015,бухгалтерия,директор,66496,7
20 460,Никишкин Ш. Л.,муж,1991,2006,служба стандартизации,офис-менеджер,57732,3
21 148,Князева О. Л.,жен,1983,2009,отдел кадров,офис-менеджер,65579,3
22 1301,Гранин З. Л.,жен,1978,2021,юридическая служба,начальник,59174,6
23 127,Ермилов В. Ч.,муж,1977,2014,отдел организации труда,инженер,66019,1
24 959,Гранин З. Л.,муж,1997,2007,склады готовой продукции и материалов,инженер,34354,10
25 1423,Никишкин Ш. Л.,муж,1980,2015,финансовое подразделение,инженер,51519,15
26 1267,Завьялова У. З.,жен,2000,2007,отдел кадров,начальник,49397,8
27 524,Макаров В. Р.,жен,1995,2019,финансовое подразделение,главный инженер,55038,9
28 1379,Завьялова У. З.,жен,1982,2015,отделение построения,офис-менеджер,38647,6
29 1245,Гранин З. Л.,муж,1976,2005,бухгалтерия,начальник,18345,11
30 331,Шитова Л. А.,муж,1988,2017,финансовое подразделение,директор,34009,1
31 309,Малов У. Ш.,муж,1975,2022,служба управления персоналом,менеджер,29669,2
32 447,Зиновьева Г. В.,жен,1973,2017,секретариат,заместитель директора,43191,3
33 157,Богров З. Н.,жен,1984,2011,склады готовой продукции и материалов,начальник,45147,12
34 514,Малов У. Ш.,жен,1984,2022,бухгалтерия,инженер,34832,10
35 22,Князева О. Л.,муж,1989,2017,отдел кадров,офис-менеджер,23784,12
36 884,Желябов Х. В.,жен,1999,2009,отдел организации труда,секретарь,18170,10
37 49,Коннова Г. С.,жен,1985,2019,бухгалтерия,менеджер,40830,1
38 656,Кузмячук В. Н.,жен,1970,2010,секретариат,заместитель директора,51480,3
39 1426,Тахтилова В. К.,жен,1974,2007,служба стандартизации,офис-менеджер,58983,14
40 191,Шитова Л. А.,жен,1983,2009,служба стандартизации,начальник,19465,5
41 584,Малов У. Ш.,жен,1997,2017,служба безопасности,руководитель,41465,12

```

2. Листинг программы:

working with a file

with open('data.csv') as scvfile:

```

    tab_num = []
    year_birth = []
    salary=[]

    reader = csv.reader(scvfile)
    for row in reader:
        tab_num.append(row[0])
        year_birth.append(row[3])
        salary.append(row[7])

    tab_num.pop(0)
    tab_num_int = list(map(int,tab_num))

    print('Numpy:\n')

    print('Information on the service number:\n')

```

```

get_info(np.max(tab_num_int),np.min(tab_num_int),

np.mean(tab_num_int),np.var(tab_num_int),np.std(tab_num_int),np.median(tab_num_int))

print('\nInformation about the year of birth:\n')

year_birth.pop(0)
year_birth_int = list(map(int,year_birth))

get_info(np.max(year_birth_int),np.min(year_birth_int),

np.mean(year_birth_int),np.var(year_birth_int),np.std(year_birth_int),np.median(year_birth_int))

print('\nSalary information:\n')

salary.pop(0)
salary_int = list(map(int,salary))

get_info(np.max(salary_int),np.min(salary_int),

np.mean(salary_int),np.var(salary_int),np.std(salary_int),np.median(salary_int))

```

Вывод программы:

Numpy:

Information on the service number:

```

Max: 1500
Min: 6
Average: 757.4977856510186
Disp: 186862.18976480435
Std: 432.2755946902443
Median: 746.0

```

Information about the year of birth:

```

Max: 2000
Min: 1970
Average: 1985.319751992914
Disp: 79.91990215284147
Std: 8.939793182889717
Median: 1986.0

```

Salary information:

Max: 69788
Min: 15019
Average: 41980.19574844996
Disp: 260014443.4382105
Std: 16124.963362383509
Median: 41334.0

3. Листинг программы:

```
# pandas
```

```
data = pd.read_csv('data.csv', delimiter=',', encoding="utf-8")  
print('*****')  
print('Pandas:\n')
```

```
print('Information on the service number:\n')
```

```
data_num = data['number']
```

```
get_info(data_num.max(),data_num.min(),  
          data_num.mean(),data_num.var(),data_num.std(),data_num.median())
```

```
print("\nInformation about the year of birth:\n')
```

```
data_birth = data['birth_year']
```

```
get_info(data_birth.max(),data_birth.min(),  
          data_birth.mean(),data_birth.var(),data_birth.std(),data_birth.median())
```

```
print("\nSalary information:\n')
```

```
data_sal = data['salary']
```

```
get_info(data_sal.max(),data_sal.min(),  
          data_sal.mean(),data_sal.var(),data_sal.std(),data_sal.median())
```

Вывод программы:

```
*****
```

Pandas:

Information on the service number:

```
Max: 1500
Min: 6
Average: 757.4977856510186
Disp: 187027.84773445403
Std: 432.4671637644343
Median: 746.0
```

Information about the year of birth:

Information about the year of birth:

```
Max: 2000
Min: 1970
Average: 1985.319751992914
Disp: 79.99075312992731
Std: 8.943754979309714
Median: 1986.0
```

Salary information:

```
Max: 69788
Min: 15019
Average: 41980.19574844996
Disp: 260244952.69657767
Std: 16132.109369099184
Median: 41334.0
```

4. Листинг программы:

```
# plotting
plt.figure(figsize=(14, 10), dpi=80)
plt.hlines(y=data['number_projects'], xmin=0, xmax=data['salary'], alpha=0.4,
linewidth=5)
plt.gca().set(ylabel='number of projects', xlabel='salary')
plt.title('dependence of the number of projects on salary')
plt.show()
```

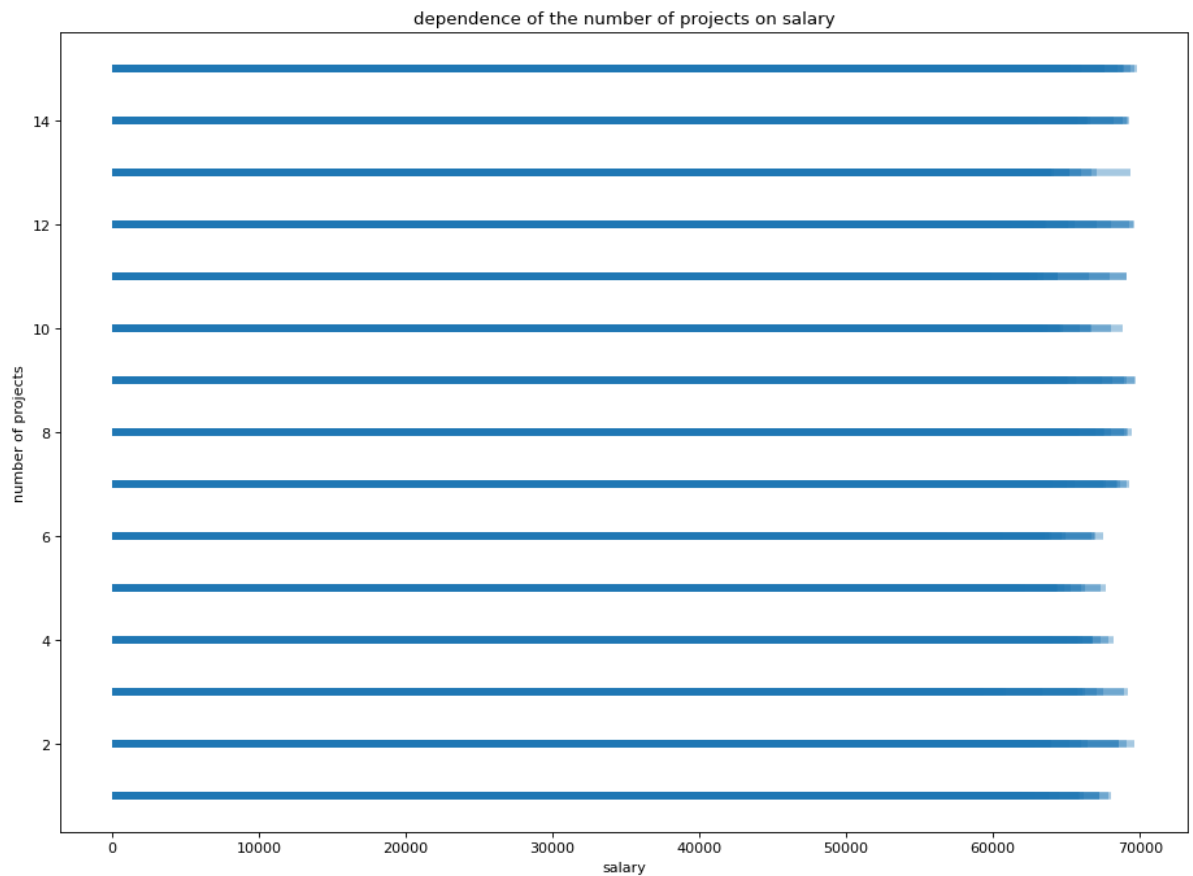
```
graf1 = data['department'].hist()
plt.xlabel('department')
plt.ylabel('number of employees')
plt.xticks(rotation=90)
plt.title('distribution of employees by departments')
plt.show()
```

```
plt.figure(figsize=(16, 10), dpi=80)
```

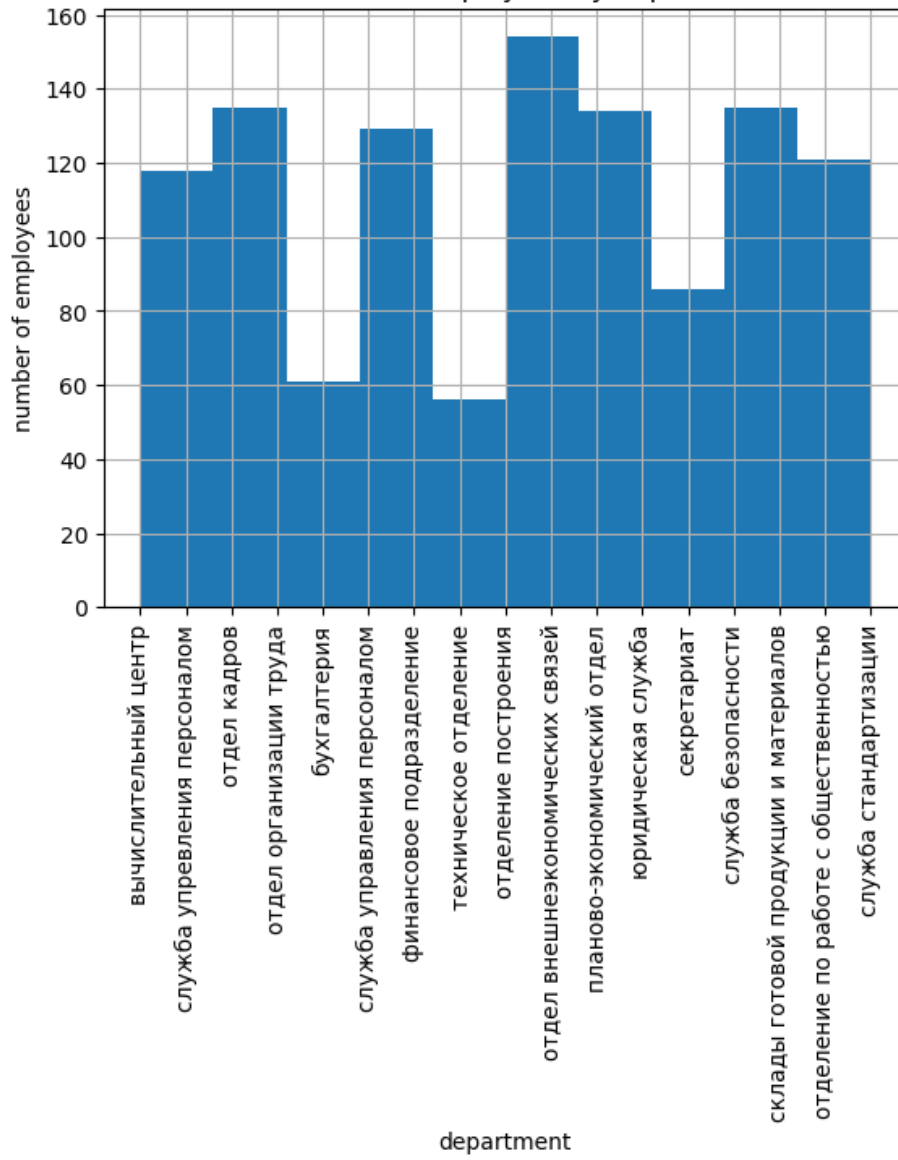


```
plt.plot_date(data["start_work"],data["salary"])
plt.gca().xaxis.set_major_locator(mdates.AutoDateLocator())
plt.ylabel('salary')
plt.xlabel('dates')
plt.title("salary change for the period")
plt.show()
```

Вывод программы:

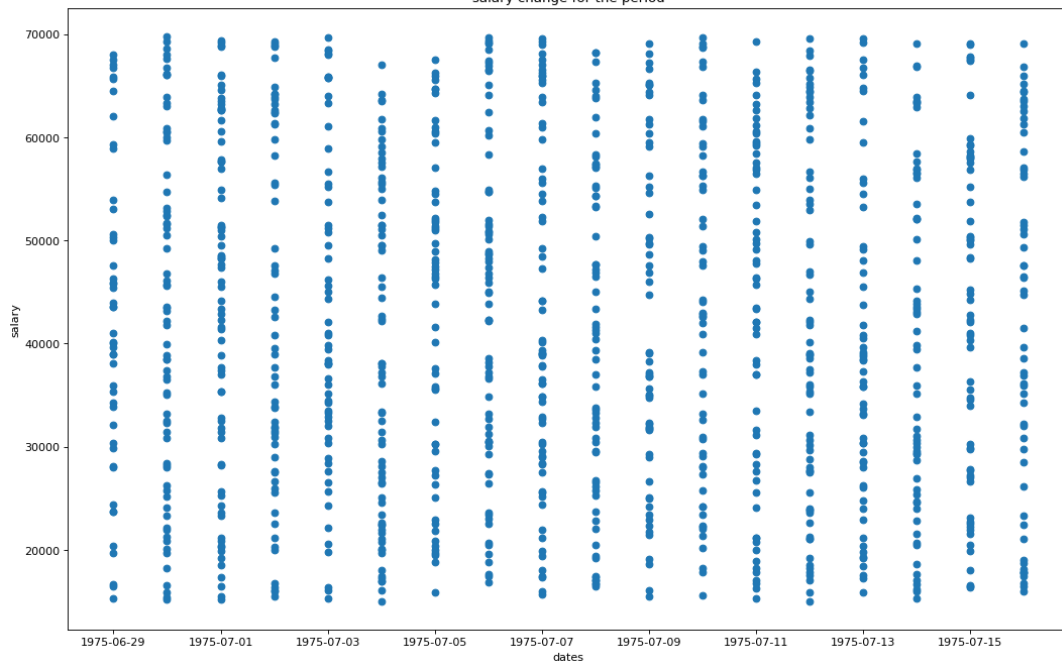


distribution of employees by departments



department

salary change for the period



5. Формат CSV (Comma Separated Values) является наиболее распространенным и удобным форматом для работы с данными в виде таблиц. CSV означает файлы, разделенные запятыми, т.е. данные разделяются запятыми друг от друга. Файлы CSV создаются программой, которая обрабатывает большое количество данных. Данные из файлов CSV можно легко экспортировать в виде электронной таблицы и базы данных, а также импортировать для использования другими программами.

Библиотека NumPy - это популярная библиотека Python, используемая для приложений научных вычислений, и является аббревиатурой от «Numerical Python».

NumPy имеет несколько преимуществ перед использованием основных математических функций Python, некоторые из которых описаны здесь: NumPy чрезвычайно быстр по сравнению с ядром Python благодаря интенсивному использованию расширений C.

Многие продвинутые библиотеки Python, такие как Scikit-Learn, Scipy и Keras, широко используют библиотеку NumPy. Поэтому, если вы планируете продолжить карьеру в области науки о данных или машинного обучения, NumPy - очень хороший инструмент для освоения.

NumPy поставляется с множеством встроенных функций, которые в ядре Python потребуют значительного количества настраиваемого кода.

Pandas это высокоуровневая Python библиотека для анализа данных. Почему я её называю высокоуровневой, потому что построена она поверх более низкоуровневой библиотеки NumPy (написана на Си), что является большим плюсом в производительности. В экосистеме Python, pandas является наиболее продвинутой и быстроразвивающейся библиотекой для обработки и анализа данных.

