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 №1

Вариант №14

Выполнил:

студент группы ИСТбд-42

Миридонов Павел

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;

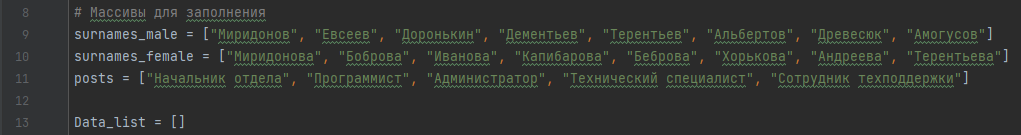
- подразделение;

- должность;

- оклад;

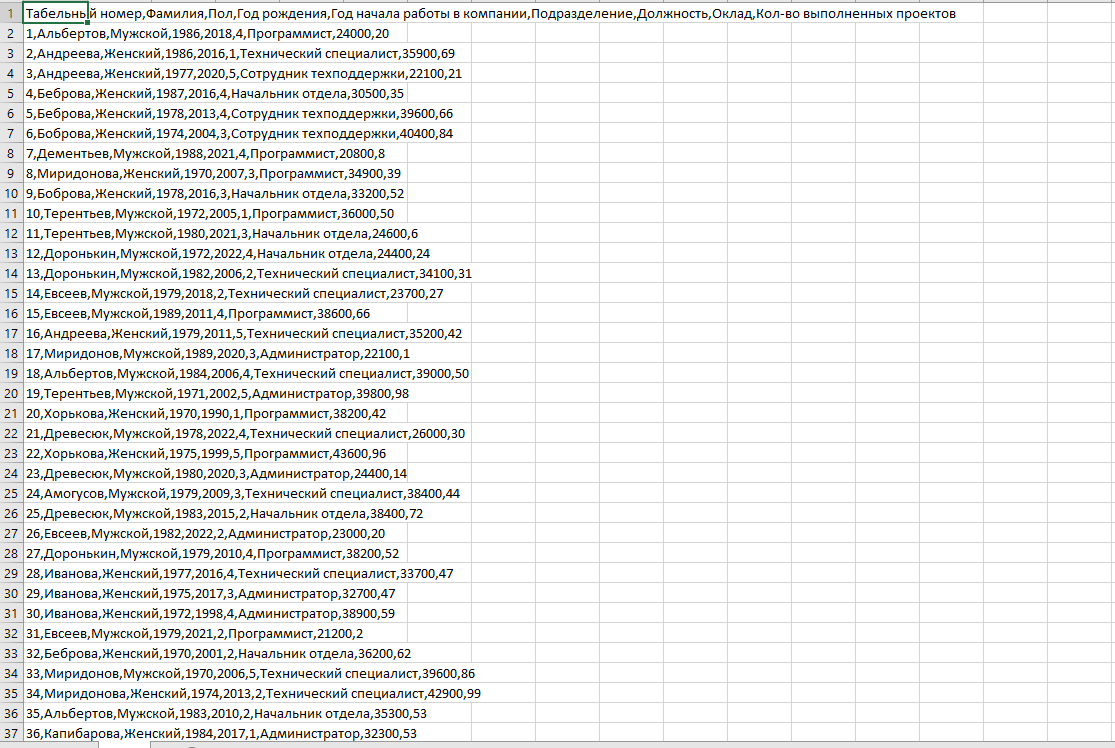
- количество выполненных проектов

Код:



# Заполнение цсв файла  
with open("Data.csv", "w") as f:  
 writer = csv.writer(f, delimiter=",", lineterminator="\r")  
 writer.writerow(  
 ["Табельный номер", "Фамилия", "Пол", "Год рождения", "Год начала работы в компании", "Подразделение",  
 "Должность", "Оклад", "Кол-во выполненных проектов"])  
 for i in range(1, 1100):  
 temp\_sex = random.randint(0, 1)  
 temp\_birthyear = random.randint(1970, 1990)  
 temp\_startwork = random.randint(temp\_birthyear + random.randint(18, 31), 2022)  
 temp\_post = random.choice(posts)  
 temp\_subdiv = random.randint(1, 5)  
 ye\_of\_work = 2022 - temp\_startwork  
 if ye\_of\_work < 5:  
 temp\_projects = random.randint(1, 30)  
 elif ye\_of\_work > 5:  
 temp\_projects = random.randint(30, 100)  
  
 if ye\_of\_work == 1:  
 temp\_salary = random.randrange(20000, 25000, 1000) + temp\_projects \* 100  
 elif ye\_of\_work < 5:  
 temp\_salary = random.randrange(20000, 25000, 1000) + temp\_projects \* 100  
 elif ye\_of\_work < 7:  
 temp\_salary = random.randrange(25000, 30000, 1000) + temp\_projects \* 100  
 elif ye\_of\_work > 7:  
 temp\_salary = random.randrange(30000, 35000, 1000) + temp\_projects \* 100  
 temp\_id = i  
 if temp\_sex == 0:  
 sex\_str = "Мужской"  
 temp\_surname = random.choice(surnames\_male)  
 elif temp\_sex == 1:  
 sex\_str = "Женский"  
 temp\_surname = random.choice(surnames\_female)  
 writer.writerow(  
 [temp\_id, temp\_surname, sex\_str, temp\_birthyear, temp\_startwork, temp\_subdiv,  
 temp\_post, temp\_salary, temp\_projects])

Результат работы кода:



**Задание 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).

Статистический метод с помощью библиотеки numpy:

# Статистические данные с помощью нумпай  
def numpy\_stat(column, csv\_file, headers):  
 header = headers[column]  
 data = []  
  
 for cell in csv\_file:  
 data.append(int(cell[column]))  
  
 stat\_array = numpy.array(data)  
 print("\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_NumPY\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_")  
 print("Статистические данные для столбца \"%s\" " % header)  
 print("Минимальное значение:", numpy.min(stat\_array))  
 print("Максимальное значение:", numpy.max(stat\_array))  
 print("Среднее значение:", numpy.average(stat\_array))  
 print("Дисперсия:", numpy.var(stat\_array))  
 print("Стандартное отклонение:", numpy.std(stat\_array))  
 print("Медиана:", numpy.median(stat\_array))  
 print("Мода:", statistics.mode(stat\_array))

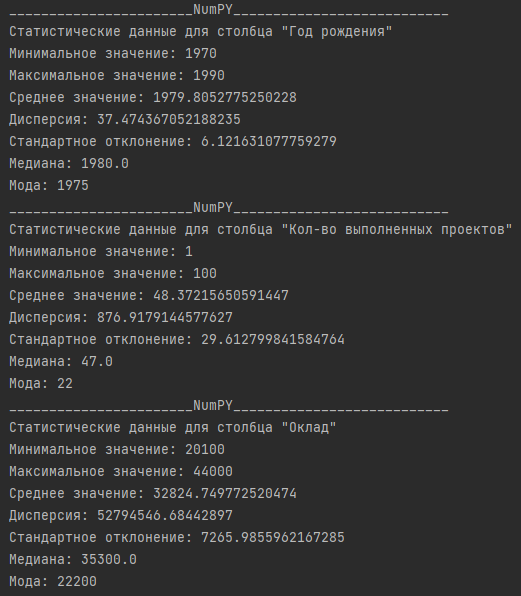
**Задание 3.**

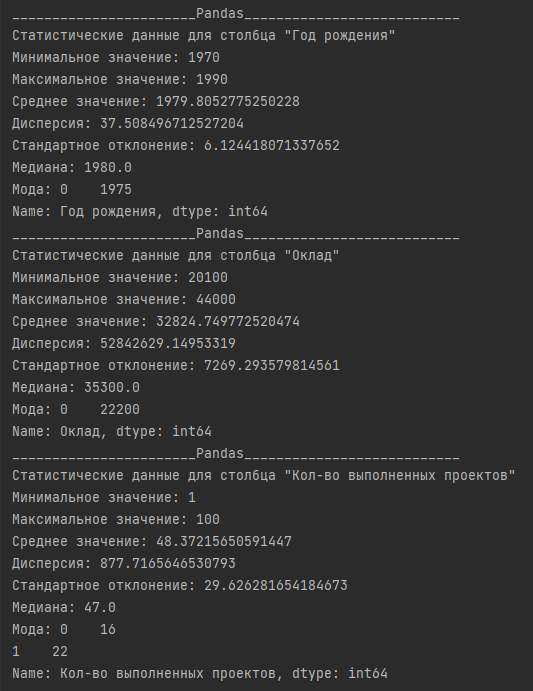
Прочитать сгенерированный набор данных в виде датафрейма и получить с помощью методов библиотеки pandas для тех же столбцов те же статистические характеристики. Продемонстрировать применение не менее 3 методов библиотеки pandas.

Статистический метод с помощью библиотеки pandas:

# Статистические данные с помощью пандас  
def pandas\_stat(column, dataframe):  
 print("\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Pandas\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_")  
 print("Статистические данные для столбца \"%s\" " % column)  
 print("Минимальное значение:", dataframe[column].min())  
 print("Максимальное значение:", dataframe[column].max())  
 print("Среднее значение:", dataframe[column].mean())  
 print("Дисперсия:", dataframe[column].var())  
 print("Стандартное отклонение:", dataframe[column].std())  
 print("Медиана:", dataframe[column].median())  
 print("Мода:", dataframe[column].mode())

Результат работы статистических методов:





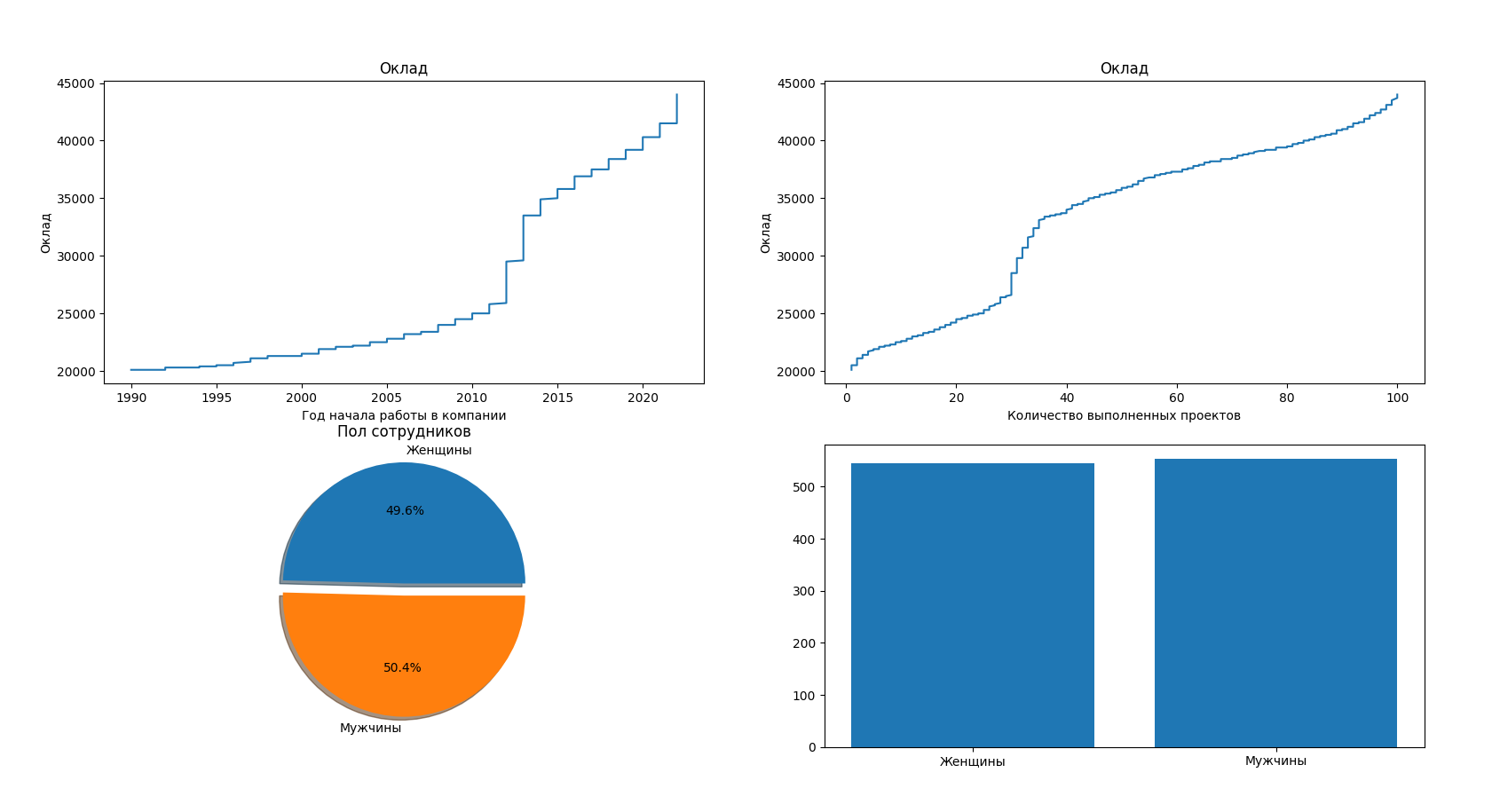
**Задание 4.**

Построить не менее 3 разнотипных графиков.

Код:

# Вывод графика зависимости Оклада от года начала работы в компании  
plt.figure()  
plt.subplot(2, 2, 1)  
plt.title("Оклад")  
plt.xlabel("Год начала работы в компании")  
plt.ylabel("Оклад")  
plt.plot(df\_stat["Год начала работы в компании"].sort\_values(), df\_stat["Оклад"].sort\_values())  
# plt.show()  
# Вывод графика зависимости Оклада от года начала работы в компании  
plt.subplot(2, 2, 2)  
plt.title("Оклад")  
plt.xlabel("Количество выполненных проектов")  
plt.ylabel("Оклад")  
plt.plot(df\_stat["Кол-во выполненных проектов"].sort\_values(), df\_stat["Оклад"].sort\_values())  
# plt.show()  
# Построение круговой диаграммы  
plt.subplot(2, 2, 3)  
plt.title("Пол сотрудников")  
labels = "Женщины", "Мужчины"  
male\_count = 0  
female\_count = 0  
for sex in df\_stat["Пол"]:  
 if sex == "Мужской":  
 male\_count = male\_count + 1  
 elif sex == "Женский":  
 female\_count = female\_count + 1  
sizes = [female\_count, male\_count]  
explode = (0.1, 0)  
plt.pie(sizes, explode=explode, labels=labels, autopct='%1.1f%%', shadow=True)  
  
# Построение графика столбиками  
plt.subplot(2,2,4)  
plt.bar(labels, sizes)  
plt.show()

Вывод графиков:



**Задание 5.**

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

В ходе выполнения работы были использованы различные библиотеки, в том числе csv, numpy и pandas

CSV.

Библиотека csv позволяет работать с csv файлами. Формат CSV является наиболее удобным форматом для работы с таблицами данных. Этот формат универсален и позволяет хранить различные типы данных в удобном формате.

Pandas и NumPy

Эти библиотеки предлагают методы для осуществления различных математических и статистических вычислений. Обе эти библиотеки предлагают широкий спектр функций. Но работа с библиотекой pandas показалась мне немного более удобной, потому что с помощью метода read\_csv(), можно скопировать данные из csv файла в датафрейм и дальше с ним работать. В случае с нумпай приходилось создавать дополнительный массив с использованием дополнительного цикла.