

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение
высшего образования

«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»

Факультет информационных систем и технологий
Кафедра «Измерительно-вычислительные комплексы»

Отчет по лабораторной работе №3

«Исследование библиотек CSV, pandas, numpy»

по дисциплине «Методы искусственного интеллекта»

Выполнила:
ст. гр. ИСТбд-41 Гильметдинова Е.Д.

Проверил:
к.т.н., доцент Шишкин.В.В.

Ульяновск

2022

Задание:

1. Создать симулированный набор данных и записать его на диск в виде csv файла.

Реализовано в файле `main.py`

Отрывок полученного файла:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Табельный номер,Фамилия И. О.,Пол,Год рождения,Начало работы (год),Подразделение,Должность,Оклад,Количество выполненных проектов															
2	0,А.	Д.	Кокорхин,М,	1985,2005,	Отдел охраны труда,	Инженер-исследователь,	44985,	20								
3	1,П.	П.	Чепукалин,М,	1972,1992,	Аэрометрические системы воздушных сигналов,	Бухгалтер,	58916,	19								
4	2,Я.	Г.	Бугайцов,М,	1970,1990,	Отдел кадров,	Техник,	47665,	2								
5	3,И.	Е.	Путрин,М,	1972,1992,	Системы автоматизированного управления наземной техникой,	Инженер-программист,	52596,	4								
6	4,А.	К.	Евдокимцев,М,	1966,1986,	Аэрометрические системы воздушных сигналов,	Инженер-исследователь,	33215,	18								
7	5,Г.	Ю.	Кусакин,М,	1998,2018,	Информационно-управляющие системы для вертолетов,	Техник,	29582,	19								
8	6,Ф.	Ю.	Сейфудинов,М,	1978,1998,	Отдел охраны труда,	Инженер,	29629,	4								
9	7,Г.	В.	Коломацкий,М,	1985,2005,	Аэрометрические системы воздушных сигналов,	Техник,	35867,	16								
10	8,В.	Г.	Очкин,М,	1978,1998,	Проектирование и верификация программно-математического обеспечения,	Техник,	46960,	19								
11	9,Б.	А.	Толстоухов,М,	1967,1987,	Отдел охраны труда,	Инженер,	33675,	20								
12	10,И.	И.	Чекулдаев,М,	1967,1987,	Аэрометрические системы воздушных сигналов,	Инженер-программист,	48392,	20								
13	11,В.	Д.	Курилев,М,	1986,2006,	Отдел охраны труда,	Начальник ТКБ,	22133,	17								
14	12,В.	Е.	Озерянский,М,	1990,2010,	Отдел кадров,	Бухгалтер,	57399,	12								
15	13,Э.	Г.	Боткин,М,	1993,2013,	Отдел кадров,	Инженер-исследователь,	38767,	1								
16	14,К.	О.	Пращикин,М,	1998,2018,	Системы автоматизированного управления наземной техникой,	Начальник ТКБ,	29674,	3								
17	15,В.	Н.	Берруких,М,	1987,2007,	Отдел кадров,	Техник,	53741,	5								
18	16,Б.	В.	Отморский,М,	1976,1996,	Проектирование и верификация программно-математического обеспечения,	Инженер-исследователь,	53908,	2								
19	17,Ю.	Л.	Цекин,М,	1962,1982,	Проектирование и верификация программно-математического обеспечения,	Бухгалтер,	43176,	9								

2. Прочитать сгенерированный набор данных в виде списков и получить с помощью программирования и методов библиотеки `numpy` для разных по типу признаков столбцов основные статистические характеристики.

Реализовано в файле `lab3.py` в функции `def numpy_analyze()`

Результат:

`Numpy` анализ

Анализ даты рождения:

Минимальная: 1961

Максимальная: 2001

Сумма: 2363788

Среднее: 1981.3813914501256

Дисперсия: 135.8537023388037

Стандартное отклонение: 11.655629641456686

Медиана: 1982.0

Анализ зарплаты:

Минимальная: 20166

Максимальная: 59914

Сумма: 48145010

Среднее: 40356.25314333613

Дисперсия: 130628850.91328642

Стандартное отклонение: 11429.297918651277

Медиана: 40810.0

Анализ выполненных проектов:

Минимальное количество: 1

Максимальное количество: 20
Сумма: 12535
Среднее: 10.507124895222129
Дисперсия: 34.99429123083873
Стандартное отклонение: 5.91559728436941
Медиана: 10.0

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

Реализовано в файле lab3.py в функции def pandas_analyze()

Результат:

Pandas анализ

Анализ даты рождения:

Минимальная: 1961
Максимальная: 2001
Сумма: 2363788
Среднее: 1981.3813914501256
Дисперсия: 135.967673565598
Стандартное отклонение: 11.660517722879975
Медиана: 1982.0

Анализ зарплаты:

Минимальная: 20166
Максимальная: 59914
Сумма: 48145010
Среднее: 40356.25314333613
Дисперсия: 130738438.87546198
Стандартное отклонение: 11434.09108217448
Медиана: 40810.0

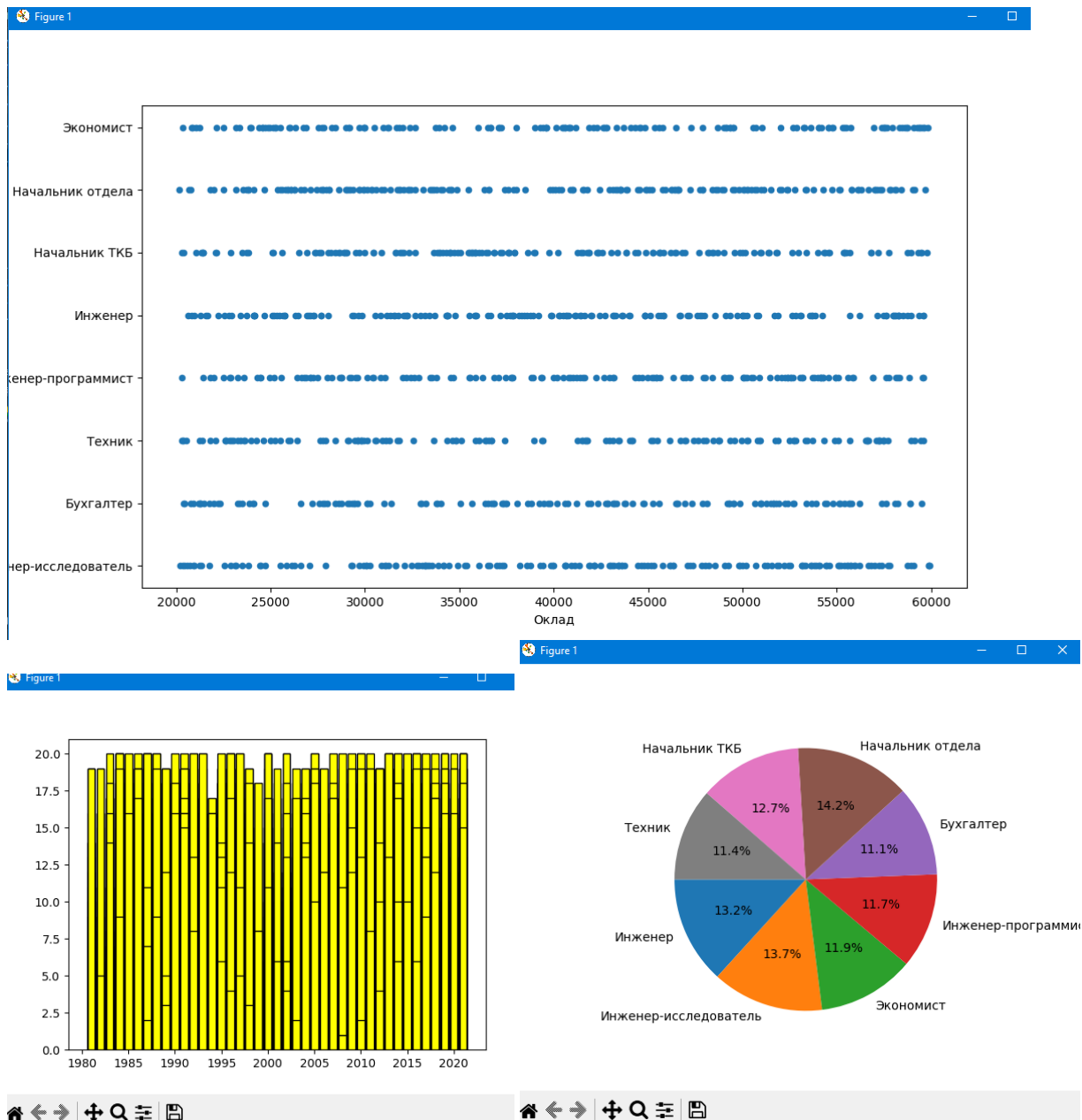
Анализ выполненных проектов:

Минимальное количество: 1
Максимальное количество: 20
Сумма: 12535
Среднее: 10.507124895222129
Дисперсия: 35.02364885771024
Стандартное отклонение: 5.918078138864867
Медиана: 10.0

4. Построить не менее 3 разнотипных графиков.

Реализовано в файле lab3.py в функции def graph()

Получившиеся графики:



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

CSV (comma-separated value) - это формат представления табличных данных (например, это могут быть данные из таблицы или данные из БД).

В этом формате каждая строка файла - это строка таблицы. Несмотря на название формата, разделителем может быть не только запятая.

Программист может осуществлять чтение, запись и анализ csv-файлов.

Существует великое множество сайтов, которые размещают свою информацию в этом формате, также он часто используется в сфере бизнеса.

NumPy — это библиотека Python, которую применяют для математических вычислений: начиная с базовых функций и заканчивая линейной алгеброй.

Используется для работы с многомерными массивами.

Pandas — это библиотека Python для обработки и анализа структурированных данных, её название происходит от «panel data» («панельные данные»). Панельными данными называют информацию, полученную в результате исследований и структурированную в виде таблиц.

Навык работы с этой библиотекой пригодится дата-сайентистам или аналитикам данных. С помощью Pandas эти специалисты могут группировать и визуализировать данные, создавать сводные таблицы и делать выборку по определенным признакам.