Лабораторная работа №3

Задание

- 1. Создать симулированный набор данных и записать его на диск в виде сsv 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 питру 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.

Оценка возможностей библиотек csv, numpy, pandas.

CSV

CSV - формат представления табличных данных, где каждая строка файла - строка таблицы, разделитель может быть любым. Разработан для работы с CSV файлами, созданными в Excel, легко настраивается для работы с любыми разновидностями этого формата. CSV содержит объекты и методы

для чтения файлов CSV, обработки данных в этом формате, а также записи данных в файл CSV.

Numpy

Библиотека numpy предоставляет множество удобных математических и статистических методов, которые гораздо менее требовательны к типу получаемых методом данных, чем такие же методы библиотек statistics и math, и присутствуют в большем количестве, чем в этих библиотеках. Она также предоставляет методы генерации случайных чисел, в целом похожие на встроенные из библиотеки random. Реализует инструменты для обработки больших одномерных и многомерных массивов и матриц.

Pandas

Pandas - пакет, который делает Python мощным инструментом для анализа данных.

Пакет дает возможность строить сводные таблицы, выполнять группировки, предоставляет удобный доступ к табличным данным, а при наличии пакета matplotlib дает возможность рисовать графики на полученных наборах данных. Позволяет использовать объект DataFrame для манипулирования двумерными массивами данных.

Порядок выполнения работы

1. Сгенерированный набор данных в файле data.csv.

	ABCDEFGHIJKLMNO
1	Табельный номер, Фамилия И.О., Пол, Год рождения, Год начала работы в компании, Подразделение, Должность, Оклад, Количество выполненных проекто
2	1,Степанова Ева Валерьевна,Женский,1965,1983,Отдел развития ИТ,Программист РНР,64557,45
3	2,Макарова Варвара Валентиновна,Женский,1960,1981,Отдел поддержки ИТ,Программист JavaScript,23015,23
4	3,Михайлова Антонина Анатольевна,Женский,1969,1989,Отдел поддержки ИТ,Web-дизайнер,99706,43
5	4,Новикова Вера Георгиевна,Женский,2000,2020,Отдел контроля и качества ИТ,Программист Swift,41113,30
6	5,Афанасьев Роман Борисович,Мужской,1980,1998,Отдел разработки ПО,Тестировщик,36543,41
7	6,Карташов Артём Иванович,Мужской,1979,2001,Отдел контроля и качества ИТ,Программист JavaScript,83132,32
8	7,Волкова Валентина Васильевна,Женский,1987,2005,Отдел разработки ПО,Web-дизайнер,67987,25
9	8,Миничкин Илья Викторович,Мужской,1973,1992,Отдел информационной безопасности,Программист Android,27190,40
10	9,Михайлова Белла Романовна,Женский,1969,1988,Отдел разработки ПО,Web-дизайнер,76748,31
11	10,Попова Елизавета Борисовна,Женский,1999,2019,Отдел разработки ПО,Программист Swift,39863,41
12	11,Никитина Варвара Антоновна,Женский,1990,2010,Отдел развития ИТ,DevOps,82477,17
13	12,Михайлова Ирина Денисовна,Женский,1974,1992,Отдел разработки ПО,Сетевой инженер,31050,22
14	13,Грязнов Даниил Викторович,Мужской,1961,1980,Отдел информационной безопасности,Программист РНР,60178,41
15	14,Макарова Анна Васильевна,Женский,1998,2019,Отдел контроля и качества ИТ,Программист JavaScript,66158,47
16	15,Цыпарев Даниил Вадимович,Мужской,1977,1998,Отдел информационной безопасности,DevOps,71100,48
17	16,Глухов Олег Моисеевич,Мужской,1989,2007,Отдел информационной безопасности,Программист Android,51472,37
18	17,Иванова Жанна Михайловна,Женский,1997,2015,Отдел поддержки ИТ,Системный аналитик,58886,48
19	18,Михайлова Дарья Витальевна,Женский,1984,2004,Отдел контроля и качества ИТ,Программист Python,59209,32
20	19,Степанова Галина Борисовна,Женский,1972,1993,Отдел развития ИТ,Программист JavaScript,40391,26
21	20,3ахаров Всеволод Леонидович,Мужской,1964,1985,Отдел контроля и качества ИТ,Web-дизайнер,53845,30
22	21,Александрова Варвара Васильевна,Женский,1981,1999,Отдел информационной безопасности,DevOps,27779,39
23	22,Винаев Евгений Леонидович,Мужской,1967,1986,Отдел поддержки ИТ,Программист Swift,33331,44
24	23,Александрова Диана Васильевна,Женский,1984,2005,Отдел развития ИТ,Web-дизайнер,61126,40
25	24,Волкова Виктория Богдановна,Женский,1981,2003,Отдел разработки ПО,Тестировщик,16315,28
26	25,Николаева Дарья Валентиновна,Женский,1999,2019,Отдел поддержки ИТ,Тестировщик,16139,27
27	26,Николаева Антонина Михайловна,Женский,1994,2014,Отдел поддержки ИТ,Тестировщик,54117,50
28	27,Фокин Игорь Глебович,Мужской,1973,1991,Отдел контроля и качества ИТ,Программист Swift,62775,35
29	28,Попова Варвара Денисовна,Женский,1999,2017,Отдел разработки ПО,Программист Swift,62231,10
30	29.Булидин Рафаэль Захарович.Мужской.1998.2019.Отлел разработки ПО.Программист РНР.57613.15

2. Вывод статистических данных по прочитанным спискам при помощи питру

3. Вывод статистических данных по прочитанным датафреймам при помощи pandas

```
Количество сотрудников: 1203 чел.
Количество мужчин: 587 чел.
Количество мужчин: 616 чел.

Год рождения самого возрастного сотрудника: 2000 год
Год рождения самого молодого сотрудника: 1960 год
Количество должностей: 11

Среднее арифметическое значение оклада: 57798.26 руб.
Размах оклада: 84853 руб.
Медианное значение оклада: 57008 руб.
Стандартное отклонение оклада: 24240.94 руб.
Дисперсия оклада: 587623061.98 руб.

Суммарное выполненных количество проектов: 29789 шт.
Среднее количество выполненных проектов на одного сотрудника: 25 шт.
Мода количества выполненных проектов: 5 шт.
```

4. Построенные разнотипные графики

График с использованием функции plot()

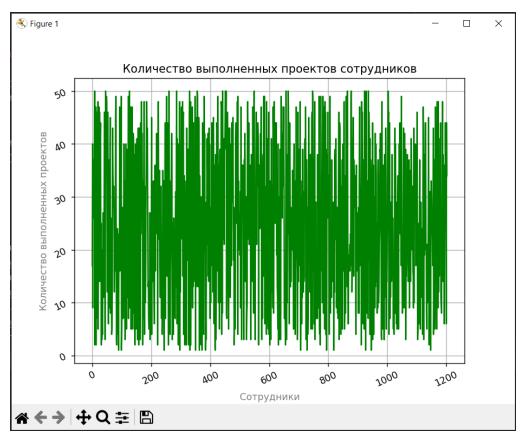


График с использованием функции hist()

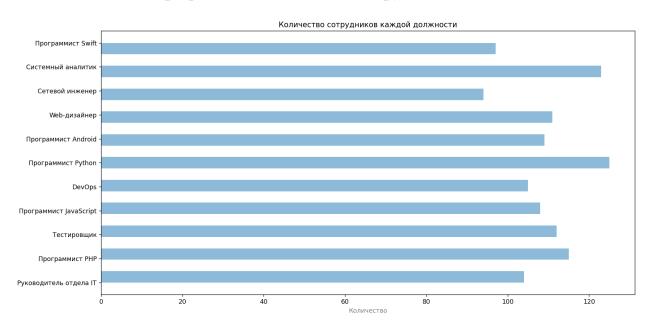


График с использованием функции barh()

