МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение

высшего образования

''УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ''

Кафедра «Информационно-вычислительные комплексы»

«Методы искусственного интеллекта»

Отчет по лабораторной работе №3

**Выполнила**:

студентка гр. ИСТбд-41

Кашицына Н. А.

**Проверил**:

преподаватель

Шишкин В. В

Ульяновск, 2022

**Исследование библиотек CSV, pandas**

**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).

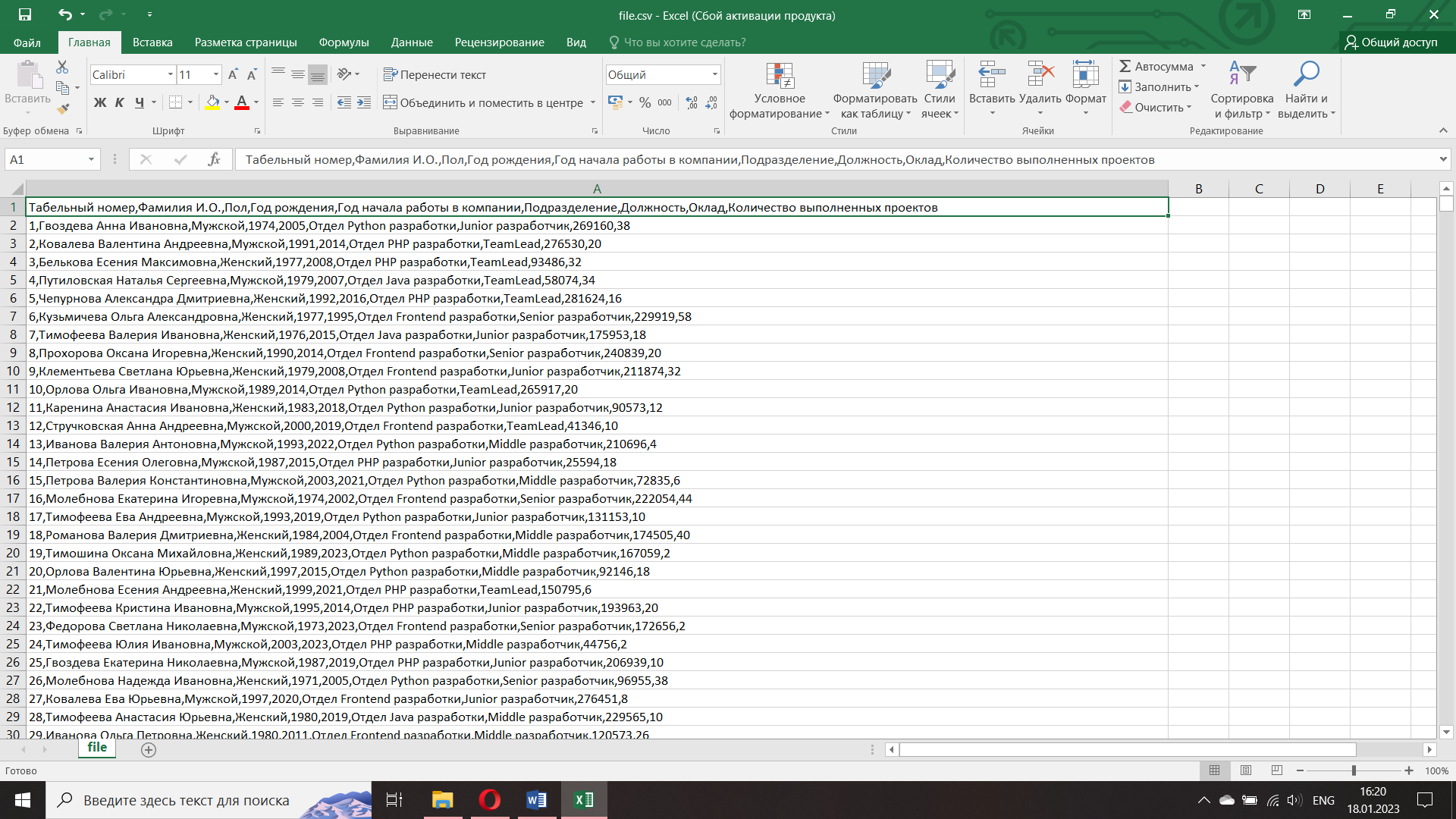
3. Прочитать сгенерированный набор данных в виде датафрейма и получить с помощью методов библиотеки pandas для тех же столбцов те же статистические характеристики. Продемонстрировать применение не менее 3 методов библиотеки pandas.

4. Построить не менее 3 разнотипных графиков.

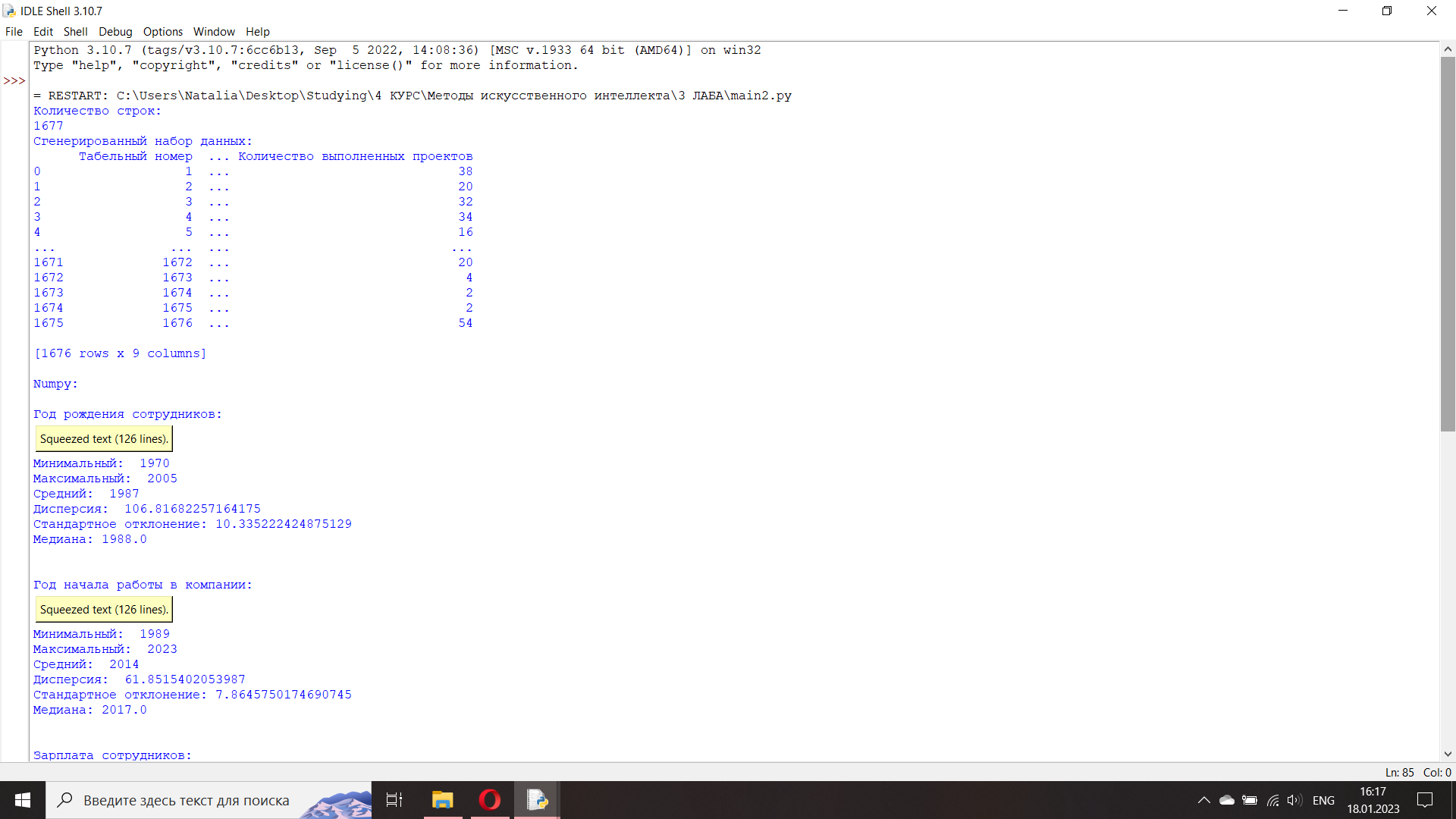
5. Оценить возможности библиотек csv, numpy, pandas в форме отчета по лабораторной работе.

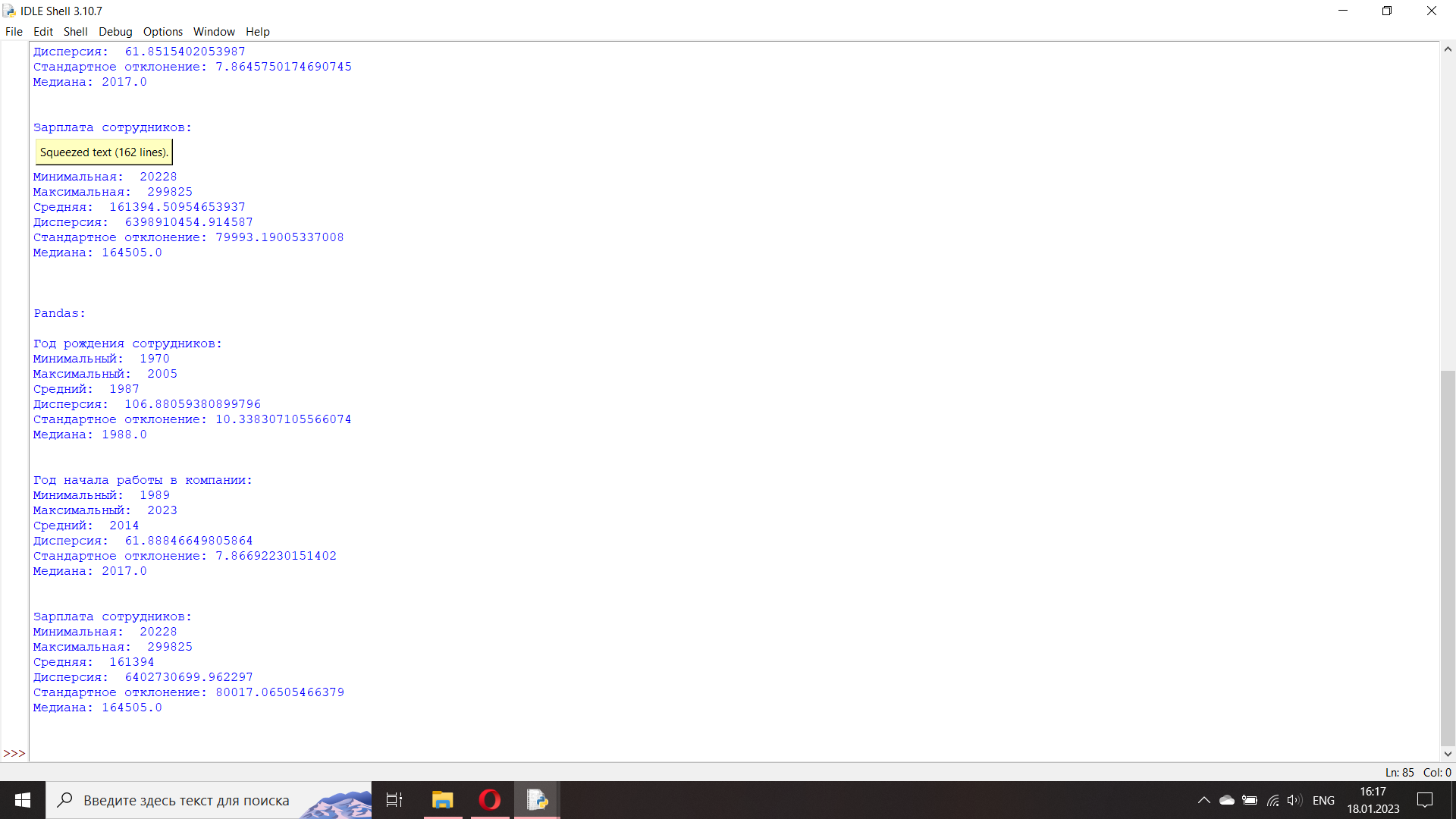
**Результат:**

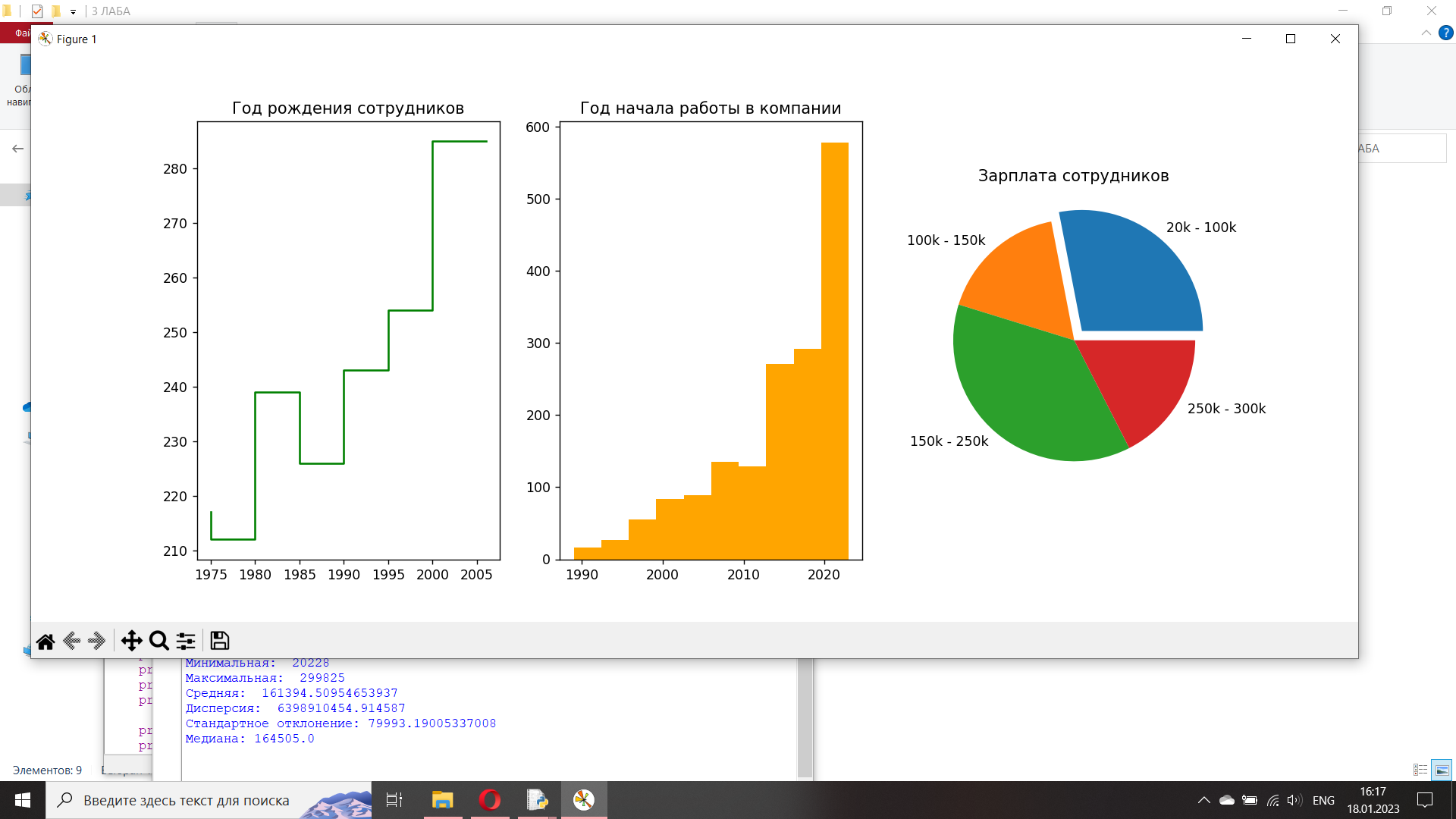
1) Набор данных в файле csv:



2) Результат работы программы:







3) Вывод:

Формат **CSV** является одним из самых распространенных форматов импорта и экспорта электронных таблиц и баз данных. Файл CSV – это особый вид файла, который позволяет структурировать большие объемы данных. По сути, он является обычным текстовым файлом, однако каждый новый элемент отделен от предыдущего запятой или другим разделителем. Обычно каждая запись начинается с новой строки. Данные CSV можно легко экспортировать в электронные таблицы или базы данных. Программист может расширять CSV файл, добавляя новые строки. С помощью csv происходит чтение из файлов / парсинг (csv.reader), запись в файлы (csv.writer), используются дополнительные параметры объекта DictReader и DictWriter (csv.DictReader и csv.DictWriter соответственно).

**NumPy** – это библиотека с открытым исходным кодом, некогда отделившаяся от проекта SciPy. NumPy является наследником Numeric и NumArray. В силу того, что NumPy базируется на Fortran это быстрая библиотека. А в силу того, что поддерживает векторные операции с многомерными массивами — крайне удобная. С помощью данной библиотеки можно создавать массивы, получать доступ к элементам, задавать форму массива и изменять её и т.д.

**Pandas** – это библиотека с открытым исходным кодом, она позволяет обрабатывать, очищать и манипулировать данными. Pandas имеет расширенный и гибкий набор структур данных разных типов. Преимущества pandas в простоте установки и использования, удобстве работы с данными и их отображении. С помощью нее можно читать csv файлы, excel, создавать Series и т.д.