

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  
по дисциплине «Методы искусственного интеллекта»

Выполнила

ст. гр. ИСТбд-41 Карташова М.В.

Проверил:

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.

Результат программы, представленной в Приложении 1:

- 1) Сгенерированный набор данных в файле `example.csv`.

1	Табельный номер,И.О. Фамилия,Пол,Год рождения,Год начала работы в компании,Подразделение,Должность,Оклад,Кол-во выполненных проектов
2	1210,С. Е. Куринская,Женщина,1959,2019,Отдел по работе с клиентами,ГИС- специалист,32467,15
3	223,В. А. Шмонин,Мужчина,1955,1982,Отдел бухгалтерского учета,Бухгалтер,22914,8
4	447,Д. И. Мысотин,Мужчина,1973,2021,Отдел администрирования,Водитель,32271,15
5	1504,А. О. Зыдина,Женщина,1979,2008,Отдел бухгалтерского учета,Старший бухгалтер,43982,17
6	942,К. Ф. Дзодзиева,Женщина,1969,1993,Отдел администрирования,Системный администратор,43100,19
7	21,И. Е. Шамшанов,Мужчина,2000,2018,Отдел бухгалтерского учета,Бухгалтер,47920,16
8	762,К. А. Барбачева,Женщина,2002,2022,Отдел администрирования,Техник,33864,9
9	486,А. Ю. Костометов,Мужчина,1991,2018,Отдел по работе с общественностью,PR-менеджер,34119,8
10	1137,К. Д. Музузина,Женщина,1965,1983,Отдел бухгалтерского учета,Старший бухгалтер,29538,11
11	1533,В. В. Солдаткина,Женщина,1968,2008,Отдел по работе с общественностью,Комьюнити-менеджер,33867,16
12	8,А. Б. Коныгин,Мужчина,2004,2022,Отдел по работе с общественностью,PR-менеджер,23332,24
13	439,Э. Н. Метельский,Мужчина,1974,1994,Отдел бухгалтерского учета,Старший бухгалтер,27481,17
14	194,Р. Л. Горетов,Мужчина,1955,1976,Отдел по работе с общественностью,PR-менеджер,42016,17
15	616,Э. В. Кентова,Женщина,1998,2017,Отдел по работе с клиентами,Специалист по работе с клиентами,47826,15
16	1055,К. М. Лульянова,Женщина,1963,2005,Отдел по работе с общественностью,Комьюнити-менеджер,26738,23
17	1159,С. С. Евшова,Женщина,1994,2017,Отдел коммерции,Специалист по рекламе,49945,8
18	1648,Л. Ф. Заигрина,Женщина,1991,2010,Отдел по работе с клиентами,Специалист по сбору инф-ции,31545,12
19	1435,Ж. А. Ладырева,Женщина,2003,2021,Отдел администрирования,Администратор офиса,38827,23
20	1546,К. А. Малетина,Женщина,1955,2005,Отдел коммерции,Специалист по рекламе,34325,7
21	301,В. Л. Чилиндин,Мужчина,1979,1997,Отдел по работе с клиентами,Специалист по работе с клиентами,33214,5
22	1610,К. Ф. Рагунова,Женщина,1966,2002,Отдел по работе с общественностью,Комьюнити-менеджер,31713,9
23	221,Э. Л. Карিশев,Мужчина,1998,2017,Отдел администрирования,Администратор офиса,36655,5
24	427,В. А. Завернягин,Мужчина,1960,1990,Отдел по работе с клиентами,Специалист по сбору инф-ции,45462,24
25	1208,Д. Ф. Колмычкова,Женщина,1992,2015,Отдел по работе с клиентами,Специалист по сбору инф-ции,47935,22
26	1070,А. Р. Шишкина,Женщина,1963,2005,Отдел по работе с общественностью,PR-менеджер,31115,9
27	969,В. Ф. Перемятова,Женщина,2003,2021,Отдел по работе с клиентами,Специалист по работе с клиентами,25433,8
28	1628,А. С. Ворчалова,Женщина,1986,2016,Отдел коммерции,Дизайнер-верстальщик,39615,23
29	430,Г. А. Подлиткин,Мужчина,2002,2022,Отдел коммерции,Администратор по учету продаж,39496,5
30	262,Е. М. Скazin,Мужчина,1968,2014,Отдел администрирования,Водитель,25112,8

1630	1157,Е. Д. Бафаева,Женщина,1967,1999,Отдел по работе с общественностью,PR-менеджер,36534,7	
1631	719,А. А. Оглодкова,Женщина,1997,2022,Отдел бухгалтерского учета,Старший бухгалтер,48379,10	
1632	918,А. Ф. Турмышева,Женщина,1981,2001,Отдел по работе с клиентами,Специалист по сбору инф-ции,46556,21	
1633	352,В. В. Заппаров,Мужчина,1970,1993,Отдел бухгалтерского учета,Бухгалтер,44270,22	
1634	849,И. Н. Семянкова,Женщина,1976,2015,Отдел коммерции,Дизайнер-верстальщик,39853,13	
1635	1524,С. О. Куркова,Женщина,2001,2022,Отдел по работе с клиентами,ГИС- специалист,40265,8	
1636	398,Я. И. Крылецкий,Мужчина,2002,2021,Отдел по работе с клиентами,Специалист по сбору инф-ции,31849,13	
1637	455,Г. О. Червяцов,Мужчина,1971,2006,Отдел по работе с клиентами,Специалист по сбору инф-ции,32352,9	
1638	96,А. А. Завираев,Мужчина,1962,1997,Отдел бухгалтерского учета,Бухгалтер,38081,12	
1639	824,В. Л. Черверкина,Женщина,1982,2009,Отдел администрирования,Системный администратор,34584,10	
1640	1538,Ю. Л. Шанеева,Женщина,1972,2001,Отдел коммерции,Сервис-менеджер,45763,5	
1641	958,Л. Е. Андрюшкова,Женщина,1970,2002,Отдел по работе с клиентами,ГИС- специалист,44632,6	
1642	109,Р. Б. Персидский,Мужчина,1975,1994,Отдел по работе с общественностью,Коммьюнити-менеджер,39286,12	
1643	377,В. В. Обашкин,Мужчина,1973,2015,Отдел по работе с общественностью,Коммьюнити-менеджер,34642,10	
1644	1095,А. А. Смалина,Женщина,1977,2015,Отдел бухгалтерского учета,Бухгалтер,36049,15	
1645	522,Я. В. Будорецкий,Мужчина,1971,2002,Отдел по работе с общественностью,Коммьюнити-менеджер,35832,23	
1646	1295,К. Б. Египкина,Женщина,1975,1993,Отдел администрирования,Водитель,29115,15	
1647	652,К. В. Суших,Женщина,1989,2018,Отдел по работе с клиентами,ГИС- специалист,46439,10	
1648	1079,И. А. Брылсева,Женщина,1960,1999,Отдел коммерции,Сервис-менеджер,44621,24	
1649	1187,Я. К. Моржева,Женщина,1967,2015,Отдел по работе с общественностью,PR-менеджер,49374,14	
1650	1196,Л. Э. Явчунская,Женщина,1960,2000,Отдел по работе с общественностью,PR-менеджер,32304,18	
1651	412,А. А. Коваленков,Мужчина,1964,2022,Отдел администрирования,Системный администратор,48409,13	
1652	1021,И. В. Гардышева,Женщина,1989,2007,Отдел коммерции,Администратор по учету продаж,37798,21	
1653	1279,В. М. Задина,Женщина,1959,2015,Отдел по работе с общественностью,Коммьюнити-менеджер,22537,19	
1654	1654,Е. О. Шингалеева,Женщина,1964,2020,Отдел администрирования,Техник,34151,24	
1655	752,А. О. Лябина,Женщина,1959,2001,Отдел по работе с общественностью,Коммьюнити-менеджер,32973,19	
1656	802,В. Э. Платонова,Женщина,1982,2003,Отдел по работе с клиентами,Специалист по сбору инф-ции,23825,19	
1657	1011,П. Е. Удереvская,Женщина,1964,2013,Отдел коммерции,Дизайнер-верстальщик,24659,8	
1658		
1659		

## 2) Вывод статистических данных по прочитанным спискам при помощи numpy

```
Статистические данные по компании:  
Кол-во сторудников в компании - 1817  
Кол-во мужчин - 1241  
Кол-во женщин - 576  
Год рождения самых старших сотрудников - 1955  
Год рождения самых младших сотрудников - 2004  
Средняя зарплата в компании - 35116.43037974684  
Общая сумма зарплаты в компании - 63806554  
Медиана кол-ва выполненных проектов по сотрудникам - 15.0  
Стандартное отклонение кол-ва выполненных проектов - 5.966417745826937  
Больше всего людей работает в подразделении - Отдел по работе с общественностью  
Одна из популярных должностей - Коммьюнити-менеджер
```

Рис.3. Первая статистика

## 3) Вывод статистических данных по прочитанным датафреймам при помощи pandas

```
Статистические данные по компании через dataframe:  
Кол-во сторудников в компании - 1817  
Кол-во мужчин - 1241  
Кол-во женщин - 576  
Год рождения самых старших сотрудников - 1955  
Год рождения самых младших сотрудников - 2004  
Средняя зарплата в компании - 35116.43037974684  
Общая сумма зарплаты в компании - 63806554  
Медиана кол-ва выполненных проектов по сотрудникам - 15.0  
Стандартное отклонение кол-ва выполненных проектов - 5.96806025590508  
Больше всего людей работает в подразделениях - Отдел по работе с общественностью  
Популярные должности - Коммьюнити-менеджер
```

Рис.4. Вторая статистика

## 4) Построенные разнотипные графики

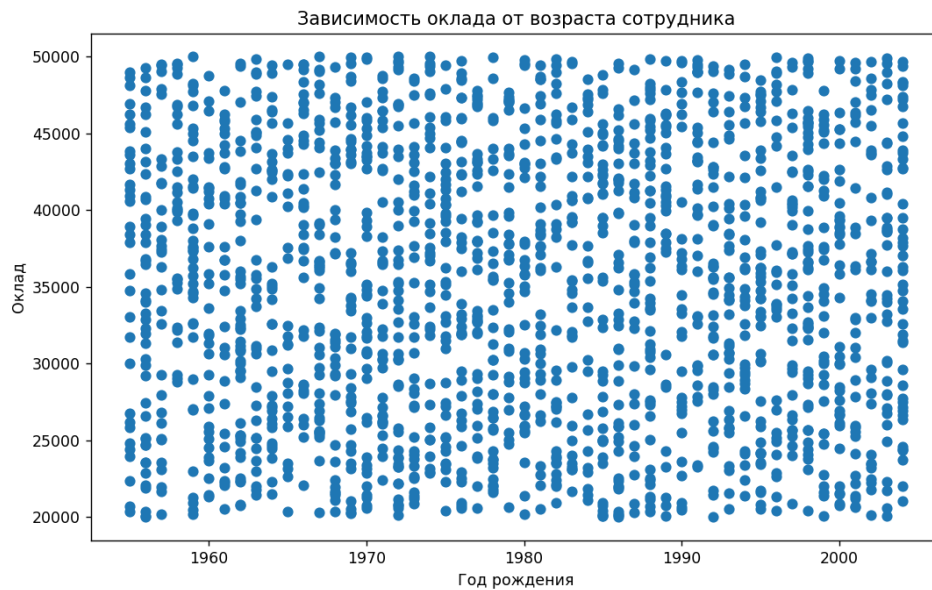


Рис.5. График рассеивания

Доли от фонда заработной платы по подразделениям в %

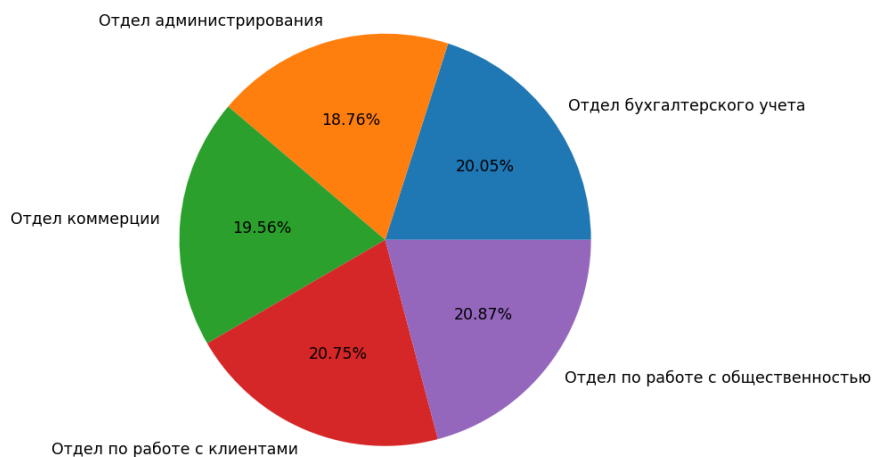


Рис.6. Круговая диаграмма

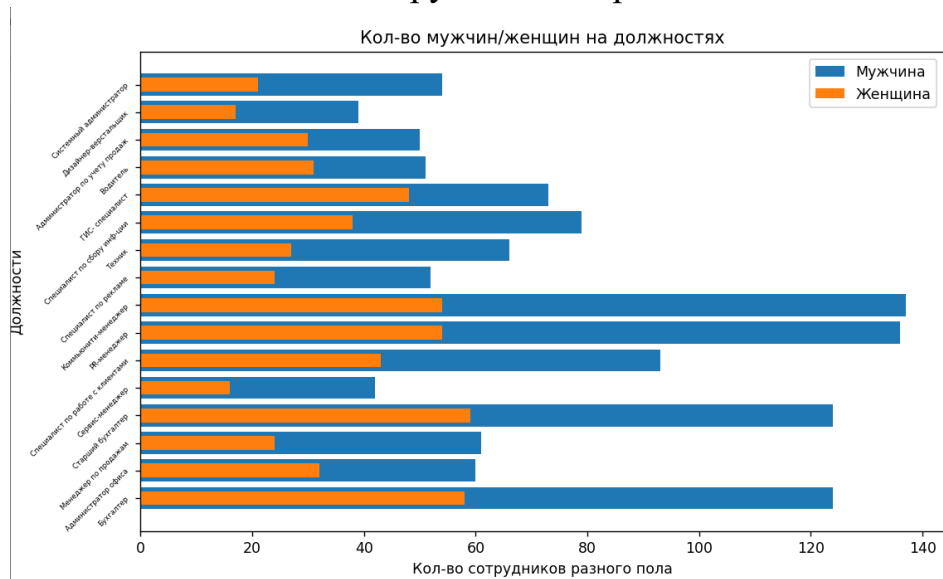


Рис.7. Столбчатая диаграмма



Рис.8. Линейный график

Оценка возможностей библиотек `csv`, `numpy`, `pandas`:

Модуль `csv` обеспечивает стандартный функционал для чтения и записи данных в файлы CSV. Разработанный для работы с CSV файлами созданными в Excel, он 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` — это библиотека для эффективной работы с массивами любого размера. `NumPy` чаще всего используют в анализе данных и обучении нейронных сетей. Массивы в `NumPy` отличаются от обычных списков и кортежей в `Python` тем, что они должны состоять только из элементов одного типа. Такое ограничение позволяет увеличить скорость вычислений в 50 раз, а также избежать ненужных ошибок с приведением и обработкой типов.

`Pandas` — это библиотека `Python` для анализа и обработки табличных данных. `Pandas` позволяет группировать данные, генерировать сводные таблицы, вычислять статистики и рисовать графики. Есть два главных класса библиотеки: `Series` и `DataFrame`. `Series` — одномерный индексированный массив, который способен хранить данные любого типа. Номер записи в массиве называется индексом. `DataFrame` — таблица с разными типами столбцов, то есть двумерная структура данных.

## Листинг программы:

```

import random as rand
from russian_names import RussianNames
import csv

count_str = rand.randint(1000, 2000)
pers_number = list(range(1, count_str + 1))

count_men = int(rand.randint(1, count_str))
count_women = int(count_str - count_men)

name_m = RussianNames(count=count_men, gender=1.0, patronymic=True,
name_reduction=True, patronymic_reduction=True)
name_w = RussianNames(count=count_women, gender=0, patronymic=True,
name_reduction=True, patronymic_reduction=True)
names = name_m.get_batch() + name_w.get_batch()

sex = ['Мужчина'] * count_men + ['Женщина'] * count_women

year_of_birth = list(rand.randint(1955, 2004) for _ in range(count_str))

start_year = list(rand.randint(year + 18, 2022) for year in year_of_birth)

structure_of_company = {
    'Отдел администрирования': ['Администратор офиса', 'Системный
администратор', 'Водитель', 'Техник'],
    'Отдел коммерции': ['Менеджер по продажам', 'Администратор по учету
продаж', 'Специалист по рекламе',
                        'Дизайнер-верстальщик', 'Сервис-менеджер'],
    'Отдел по работе с клиентами': ['Специалист по работе с клиентами',
'Специалист по сбору инф-ции',
                                'ГИС- специалист'],
    'Отдел бухгалтерского учета': ['Бухгалтер', 'Старший бухгалтер'],
    'Отдел по работе с общественностью': ['PR-менеджер', 'Коммьюнити-
менеджер']}
department = list(rand.choice(list(structure_of_company.keys())) for _ in
range(count_str))
post = list(rand.choice(structure_of_company.get(department_item)) for
department_item in department)

salary = list(rand.randint(20000, 50000) for _ in range(count_str))

count_completed_projects = list(rand.randint(5, 25) for _ in
range(count_str))

field_names = ['Табельный номер', 'И.О. Фамилия', 'Пол', 'Год рождения', 'Год
начала работы в компании',
               'Подразделение', 'Должность', 'Оклад', 'Кол-во выполненных
проектов']

data = list(zip(pers_number, names, sex, year_of_birth, start_year,
department, post, salary, count_completed_projects))
rand.shuffle(data)

file = 'example.csv'
with open(file, 'w', newline='', encoding='utf-16') as w_f:
    writer = csv.writer(w_f)
    writer.writerow(field_names)
    writer.writerows(data)

```

```
#####
import numpy as np

with open(file, 'r', encoding='utf-16') as r_f:
    reader = csv.reader(r_f, delimiter=",")
    full_data_in_file = list(reader)

    headers = full_data_in_file[0]
    number_of_column = list(range(0, len(headers)))
    dict_header = dict.fromkeys(headers) | dict(zip(headers,
number_of_column))

    data_of_personal = full_data_in_file[1:]

    count_employees = len(data_of_personal)

    def getting_column(name, types):
        return list(types(item[dict_header.get(name)]) for item in
data_of_personal)

    count_m = np.sum(np.array(getting_column('Пол', str)) == 'Мужчина')
    count_w = np.sum(np.array(getting_column('Пол', str)) == 'Женщина')

    birth_oldest_employee = np.min(getting_column('Год рождения', int))
    birth_younger_employee = np.max(getting_column('Год рождения', int))

    mean_salary = np.mean(getting_column('Оклад', int))
    all_salary = np.sum(getting_column('Оклад', int))

    median_of_count_proj = np.median(getting_column('Кол-во выполненных
проектов', int))
    std_of_count_proj = np.std(getting_column('Кол-во выполненных проектов',
int))

    departments = getting_column('Подразделение', str)
    popular_department = max(set(departments), key=lambda x:
departments.count(x))
    posts = getting_column('Должность', str)
    popular_posts = max(set(posts), key=lambda x: posts.count(x))

    print("Статистические данные по компании: ")
    print("Кол-во сотрудников в компании - " + str(count_employees))
    print("Кол-во мужчин - " + str(count_m))
    print("Кол-во женщин - " + str(count_w))
    print("Год рождения самых старших сотрудников - " +
str(birth_oldest_employee))
    print("Год рождения самых младших сотрудников - " +
str(birth_younger_employee))
    print("Средняя зарплата в компании - " + str(mean_salary))
    print("Общая сумма зарплаты в компании - " + str(all_salary))
    print("Медиана кол-ва выполненных проектов по сотрудникам - " +
str(median_of_count_proj))
    print("Стандартное отклонение кол-ва выполненных проектов - " +
str(std_of_count_proj))
    print("Больше всего людей работает в подразделении - " +
popular_department)
    print("Одна из популярных должностей - " + popular_posts + '\n')

#####
import pandas as pd

pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
pd.set_option('display.max_colwidth', None)
```



```

data_df = pd.read_csv(file, sep=',', encoding='utf-16 ')

count_employees_df = len(data_df.index)

count_m_df = len(data_df[data_df['Пол'] == 'Мужчина'])
count_w_df = len(data_df[data_df['Пол'] == 'Женщина'])

birth_oldest_employee_df = data_df['Год рождения'].min()
birth_younger_employee_df = data_df['Год рождения'].max()

mean_salary_df = data_df['Оклад'].mean()
all_salary_df = data_df['Оклад'].sum()

median_of_count_proj_df = data_df['Кол-во выполненных проектов'].median()
std_of_count_proj_df = data_df['Кол-во выполненных проектов'].std()

popular_department_df =
list(data_df['Подразделение'].value_counts().loc[data_df['Подразделение'].value_counts() == data_df['Подразделение'].value_counts().max()].index)
popular_posts_df =
list(data_df['Должность'].value_counts().loc[data_df['Должность'].value_counts() == data_df['Должность'].value_counts().max()].index)

print("Статистические данные по компании через dataframe: ")
print("Кол-во сотрудников в компании - " + str(count_employees_df))
print("Кол-во мужчин - " + str(count_m_df))
print("Кол-во женщин - " + str(count_w_df))
print("Год рождения самых старших сотрудников - " + str(birth_oldest_employee_df))
print("Год рождения самых младших сотрудников - " + str(birth_younger_employee_df))
print("Средняя зарплата в компании - " + str(mean_salary_df))
print("Общая сумма зарплаты в компании - " + str(all_salary_df))
print("Медиана кол-ва выполненных проектов по сотрудникам - " + str(median_of_count_proj_df))
print("Стандартное отклонение кол-ва выполненных проектов - " + str(std_of_count_proj_df))
print("Больше всего людей работает в подразделениях - " + ", ".join(popular_department_df))
print("Популярные должности - " + ", ".join(popular_posts_df))

#####

import matplotlib.pyplot as plt

fig1, ax1 = plt.subplots(figsize=(10, 6))
ax1.scatter(x=data_df['Год рождения'], y=data_df['Оклад'])
plt.xlabel("Год рождения")
plt.ylabel("Оклад")
plt.title("Зависимость оклада от возраста сотрудника")
plt.show()

salary_department_df = {}
departments_df = data_df['Подразделение'].unique()
for item in departments_df:
    department = data_df[data_df['Подразделение'] == item]
    salary_department_df[item] = department['Оклад'].sum()
value_pie = list(salary_department_df[item] for item in departments_df)
fig2, ax2 = plt.subplots(figsize=(10, 6))
ax2.pie(value_pie, labels=departments_df, autopct='%.2f%%')
plt.title("Доли от фонда заработной платы по подразделениям в %")
plt.show()

```

```

count_women_in_post = {}
count_men_in_post = {}
post_df = data_df['Должность'].unique()
for item in post_df:
    post = data_df[data_df['Должность'] == item]
    count_women_in_post[item] = sum(post['Пол'] == 'Женщина')
    count_men_in_post[item] = sum(post['Пол'] == 'Мужчина')
value_men = list(count_men_in_post[item] for item in post_df)
value_women = list(count_women_in_post[item] for item in post_df)
fig3, ax3 = plt.subplots(figsize=(10, 6))
ax3.barh(post_df, value_men)
ax3.barh(post_df, value_women, height=0.5)
plt.yticks(rotation=45, fontsize=5)
plt.legend(data_df['Пол'].unique(), loc=1)
plt.xlabel('Кол-во сотрудников разного пола')
plt.ylabel('Должности')
plt.title("Кол-во мужчин/женщин на должностях")
plt.show()

count_employees_in_year = data_df['Год начала работы в
компании'].value_counts().sort_index()
plt.plot(count_employees_in_year.index, count_employees_in_year)
plt.xlabel('Года')
plt.ylabel('Кол-во людей, взятых на работу')
plt.title("Кол-во персонала, принятого в определенный год")
plt.show()

```