



**Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное  
учреждение  
высшего образования  
«Московский государственный технический университет  
имени Н.Э. Баумана  
(национальный исследовательский университет)»  
(МГТУ им. Н.Э. Баумана)**

---

**ФАКУЛЬТЕТ ИНФОРМАТИКА, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И СИСТЕМЫ  
УПРАВЛЕНИЯ**

**КАФЕДРА КОМПЬЮТЕРНЫЕ СИСТЕМЫ И СЕТИ (ИУ6)**

**НАПРАВЛЕНИЕ ПОДГОТОВКИ 09.04.01 Информатика и вычислительная техника**

**МАГИСТЕРСКАЯ ПРОГРАММА 09.04.01/07 Интеллектуальные системы анализа,  
обработки и интерпретации больших данных**

**О Т Ч Е Т**

**по лабораторной работе №1**

Студент

ИУ6-32М

(Группа)

\_\_\_\_\_

(Подпись, дата)

А.М. Панфилкин

(И.О. Фамилия)

Преподаватель

\_\_\_\_\_

(Подпись, дата)

М.А. Захаров

(И.О. Фамилия)

## Задание

1. Считать набор данных 'data/homicide-data.csv'
2. Вывести информацию про первых 5 пассажиров, всех, кроме последних 2х и с 3 по 15 включительно
3. Проанализировать на предмет наличия особенностей, таких как:
  - a. Средний возраст, максимальный и минимальный
  - b. Узнать количество женщин и мужчин
  - c. Количество женщин, которым больше 18 лет
  - d. Количество мужчин, которые "Closed by arrest" (то есть преступник был пойман)
4. Отсортировать всех людей по возрасту по возрастанию и по убыванию и вывести
5. Для каждого штата найти год, где было максимальное число пострадавших. Вывести таблицу

## Выполнение

### Листинг 1 – Выполнение задания

```
# Подключение библиотек
import pandas as pd
from IPython.display import display
# Считать набор данных 'data/homicide-data.csv'
homicide_data = pd.read_csv('data/homicide-data.csv', encoding='ISO-8859-1')
# Вывести информацию про первых 5 пассажиров, всех, кроме последних 2х и с 3 по
15 включительно
display(homicide_data.head())
display(homicide_data[:-2])
display(homicide_data[3:16])
# Проанализировать на предмет наличия особенностей, таких как:
# 1. Средний возраст, максимальный и минимальный
# 2. Узнать количество женщин и мужчин
# 3. Количество женщин, которым больше 18 лет
# 4. Количество мужчин, которые "Closed by arrest" (то есть преступник был
пойман)
# Отсортировать всех людей по возрасту по возрастанию и по убыванию и вывести

# 1. Средний возраст, максимальный и минимальный
# P.S. В колонке victim_age есть значения "Unknowн", которые не надо учитывать
clean_df = homicide_data[homicide_data.victim_age != 'Unknowн']
# В одну строку можно было бы сделать так:

print(f'Средний возраст: {clean_df.victim_age.astype(int).mean()}')
print(f'Максимальный возраст: {clean_df.victim_age.astype(int).max()}')
print(f'Минимальный возраст: {clean_df.victim_age.astype(int).min()}')

# 2. Узнать количество женщин и мужчин (victim_sex = / Female)
```

```

print(f'Количество женщин: {homicide_data[homicide_data.victim_sex ==
"Female"].shape[0]}')
print(f'Количество мужчин: {homicide_data[homicide_data.victim_sex ==
"Male"].shape[0]}')

# 3. Количество женщин, которым больше 18 лет
print(f'Количество женщин, которым больше 18 лет:
{clean_df[(clean_df.victim_sex == "Female") & (clean_df.victim_age.astype(int)
> 18)].shape[0]}')

# 4. Количество мужчин, которые "Closed by arrest" (то есть преступник был
пойман)
print(f'Количество мужчин, которые "Closed by arrest":
{homicide_data[(homicide_data.victim_sex == "Male") &
(homicide_data.disposition == "Closed by arrest")].shape[0]}')

# Отсортировать всех людей по возрасту по возрастанию и по убыванию и вывести
print('Возраст по возрастанию:')
display(homicide_data.sort_values(by=['victim_age']))
print('Возраст по убыванию:')
display(homicide_data.sort_values(by=['victim_age'], ascending=False))
# Таблица - для каждого штата найти год, где было максимальное число
пострадавших
homicide_data['year'] = homicide_data['reported_date'].astype(str).str[:4]

grouped = homicide_data.groupby(['state',
'year']).size().reset_index(name='incidents')
max_incidents = grouped.loc[grouped.groupby('state')['incidents'].idxmax()]

result_df = max_incidents.rename(columns={'incidents':
'max_incidents'})[['state', 'year', 'max_incidents']]

display(result_df)

```

	uid	reported_date	victim_last	victim_first	victim_race	victim_age	victim_sex	city	state	lat	lon	disposition
0	Alb-000001	20100504	GARCIA	JUAN	Hispanic	78	Male	Albuquerque	NM	35.095788	-106.538555	Closed without arrest
1	Alb-000002	20100216	MONTOYA	CAMERON	Hispanic	17	Male	Albuquerque	NM	35.056810	-106.715321	Closed by arrest
2	Alb-000003	20100601	SATTERFIELD	VIVIANA	White	15	Female	Albuquerque	NM	35.086092	-106.695568	Closed without arrest
3	Alb-000004	20100101	MENDIOLA	CARLOS	Hispanic	32	Male	Albuquerque	NM	35.078493	-106.556094	Closed by arrest
4	Alb-000005	20100102	MULA	VIVIAN	White	72	Female	Albuquerque	NM	35.130357	-106.580986	Closed without arrest

  

	uid	reported_date	victim_last	victim_first	victim_race	victim_age	victim_sex	city	state	lat	lon	disposition
0	Alb-000001	20100504	GARCIA	JUAN	Hispanic	78	Male	Albuquerque	NM	35.095788	-106.538555	Closed without arrest
1	Alb-000002	20100216	MONTOYA	CAMERON	Hispanic	17	Male	Albuquerque	NM	35.056810	-106.715321	Closed by arrest
2	Alb-000003	20100601	SATTERFIELD	VIVIANA	White	15	Female	Albuquerque	NM	35.086092	-106.695568	Closed without arrest
3	Alb-000004	20100101	MENDIOLA	CARLOS	Hispanic	32	Male	Albuquerque	NM	35.078493	-106.556094	Closed by arrest
4	Alb-000005	20100102	MULA	VIVIAN	White	72	Female	Albuquerque	NM	35.130357	-106.580986	Closed without arrest
...	...	...	...	...	...	...	...	...	...	...	...	...
52172	Was-001378	20160415	IVEY	PAUL	Black	37	Male	Washington	DC	38.826458	-77.003590	Closed by arrest
52173	Was-001379	20160715	HARRIS	SHAROD	Black	20	Male	Washington	DC	38.827266	-77.001572	Open/No arrest
52174	Was-001380	20160908	WILLIAMS	EVAN	Black	29	Male	Washington	DC	38.828704	-77.002075	Closed by arrest
52175	Was-001381	20160913	SMITH	DEON	Black	19	Male	Washington	DC	38.822852	-77.001725	Open/No arrest
52176	Was-001382	20161114	WASHINGTON	WILLIE	Black	23	Male	Washington	DC	38.828025	-77.002511	Open/No arrest

52177 rows × 12 columns

  

	uid	reported_date	victim_last	victim_first	victim_race	victim_age	victim_sex	city	state	lat	lon	disposition
3	Alb-000004	20100101	MENDIOLA	CARLOS	Hispanic	32	Male	Albuquerque	NM	35.078493	-106.556094	Closed by arrest
4	Alb-000005	20100102	MULA	VIVIAN	White	72	Female	Albuquerque	NM	35.130357	-106.580986	Closed without arrest
5	Alb-000006	20100126	BOOK	GERALDINE	White	91	Female	Albuquerque	NM	35.151110	-106.537797	Open/No arrest
6	Alb-000007	20100127	MALDONADO	DAVID	Hispanic	52	Male	Albuquerque	NM	35.111785	-106.712614	Closed by arrest
7	Alb-000008	20100127	MALDONADO	CONNIE	Hispanic	52	Female	Albuquerque	NM	35.111785	-106.712614	Closed by arrest
8	Alb-000009	20100130	MARTIN-LEYVA	GUSTAVO	White	56	Male	Albuquerque	NM	35.075380	-106.553458	Open/No arrest
9	Alb-000010	20100210	HERRERA	ISRAEL	Hispanic	43	Male	Albuquerque	NM	35.065930	-106.572288	Open/No arrest
10	Alb-000011	20100212	BARRIUS-CAMPANIONI	HECTOR	Hispanic	20	Male	Albuquerque	NM	35.077375	-106.560569	Closed by arrest
11	Alb-000012	20100218	LUJAN	KEVIN	White	Unknown	Male	Albuquerque	NM	35.077011	-106.564910	Closed without arrest
12	Alb-000013	20100222	COLLAMORE	JOHN	Hispanic	46	Male	Albuquerque	NM	35.064076	-106.608281	Open/No arrest
13	Alb-000014	20100306	CHIQUEITO	CORIN	Other	16	Male	Albuquerque	NM	35.041491	-106.742147	Closed by arrest
14	Alb-000015	20100308	TORRES	HECTOR	Hispanic	54	Male	Albuquerque	NM	35.070656	-106.615845	Closed by arrest
15	Alb-000016	20100308	GRAY	STEFANIA	White	43	Female	Albuquerque	NM	35.070656	-106.615845	Closed by arrest

Рисунок 1 – Результат выполнение пункта 2 задания

```

Средний возраст: 31.801220008133388
Максимальный возраст: 102
Минимальный возраст: 0
Количество женщин: 7209
Количество мужчин: 40739
Количество женщин, которым больше 18 лет: 5977
Количество мужчин, которые "Closed by arrest": 19092

```

Рисунок 2 – Результат выполнения пункта 3 задания

Возраст по возрастанию:												
	uid	reported_date	victim_last	victim_first	victim_race	victim_age	victim_sex	city	state	lat	lon	disposition
31975	Mem-000515	20111209	JONES	ANAYA	Black	0	Female	Memphis	TN	35.096701	-89.925107	Closed by arrest
18937	Dur-000191	20070321	Unknown	Unknown	Black	0	Male	Durham	NC	35.984703	-78.883550	Closed by arrest
34965	Min-000132	20080610	UNKNOWN	UNKNOWN	White	0	Male	Minneapolis	MN	44.939720	-93.261770	Open/No arrest
27614	Las-000746	20131214	MOOSA	DAMIAN	Unknown	0	Unknown	Las Vegas	NV	36.207640	-115.096709	Closed by arrest
16178	Den-000075	20110920	LEATHERS	ALDRICH	White	0	Male	Denver	CO	39.763461	-104.954206	Closed by arrest
...	...	...	...	...	...	...	...	...	...	...	...	...
15720	Dal-002237	20140929	Unknown	Unknown	Unknown	Unknown	Unknown	Dallas	TX	32.747001	-96.648989	Closed by arrest
15721	Dal-002238	20141006	Unknown	Unknown	Unknown	Unknown	Unknown	Dallas	TX	32.793653	-96.809078	Closed by arrest
15722	Dal-002239	20141010	Unknown	Unknown	Unknown	Unknown	Unknown	Dallas	TX	32.757150	-96.749622	Open/No arrest
15704	Dal-002221	20140818	Unknown	Unknown	Unknown	Unknown	Unknown	Dallas	TX	32.696194	-96.768783	Closed by arrest
15460	Dal-001976	20121029	Unknown	Unknown	Unknown	Unknown	Unknown	Dallas	TX	32.773738	-96.769213	Open/No arrest
52179 rows × 12 columns												
Возраст по убыванию:												
	uid	reported_date	victim_last	victim_first	victim_race	victim_age	victim_sex	city	state	lat	lon	disposition
15859	Dal-002376	20151006	Unknown	Unknown	Unknown	Unknown	Unknown	Dallas	TX	32.855187	-96.649427	Closed by arrest
43143	Pho-000051	20110531	BAUTISTA-MADRIGAL	REYNEL	Unknown	Unknown	Unknown	Phoenix	AZ	33.431022	-112.063636	Open/No arrest
43133	Pho-000041	20110513	CARAVEO-GONZALEZ	JESUS	Unknown	Unknown	Unknown	Phoenix	AZ	33.458010	-112.144704	Closed by arrest
43134	Pho-000042	20110516	MARTYNOV	NINA	Unknown	Unknown	Unknown	Phoenix	AZ	33.558672	-112.125859	Closed without arrest
43135	Pho-000043	20110517	HERNANDEZ	MANUEL	Unknown	Unknown	Unknown	Phoenix	AZ	33.480483	-112.190724	Closed by arrest
...	...	...	...	...	...	...	...	...	...	...	...	...
37429	NY-000029	20160210	Unknown	Unknown	Black	0	Female	New York	NY	40.608392	-74.146612	Closed by arrest
44012	Pit-000010	20141209	NEWTON	KAMERO	White	0	Female	Pittsburgh	PA	40.413134	-79.995830	Closed by arrest
50423	Tul-000350	20090927	REDACTED	REDACTED	Black	0	Male	Tulsa	OK	36.062144	-95.969026	Closed by arrest
50427	Tul-000356	20091105	REDACTED	REDACTED	Black	0	Male	Tulsa	OK	36.192440	-95.974440	Closed by arrest
19577	For-000418	20140117	Unknown	Unknown	Hispanic	0	Female	Fort Worth	TX	32.734000	-97.442400	Open/No arrest

Рисунок 3 – Результат выполнения пункта 4 задания

	state	year	max_incidents
10	AL	2017	98
17	AZ	2017	161
23	CA	2012	774
35	CO	2016	57
38	DC	2008	186
57	FL	2017	231
67	GA	2016	148
78	IL	2016	765
90	IN	2017	155
97	KY	2016	117
102	LA	2011	275
112	MA	2010	74
128	MD	2015	342
133	MI	2012	386
147	MN	2015	48
160	MO	2017	352
170	NC	2016	108
180	NE	2015	50
190	NM	2017	75
201	NV	2017	217
211	NY	2016	380
223	OH	2017	211
233	OK	2016	140
235	PA	2007	453
256	TN	2017	290
257	TX	2007	729
277	VA	2016	60
286	wl	2015	146

Рисунок 4 – Результат выполнения пункта 5 задания

## Вывод

В ходе лабораторной работы мы освоили навыки работы с анализом данных на примере набора 'data/homicide-data.csv'. Полученные навыки включают в себя умение извлекать и выводить информацию о подмножестве данных, проводить базовый статистический анализ (в данном случае, возраст

и пол), а также умение сортировать данные и выделять максимальные значения в разрезе групп (например, максимальное число пострадавших в каждом штате).