



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 №4

Студент

ИУ6-32М

(Группа)

(Подпись, дата)

А.М. Панфилкин

(И.О. Фамилия)

Преподаватель

(Подпись, дата)

М.А. Захаров

(И.О. Фамилия)

Задание

1. Построить тепловую карту. Оси - года и штаты. Значения – процент убийств от общего количества убийств в этот год
2. Создать Bar Chart, отражающий процент жертв по полу в каждом штате.
3. Отсортировать штаты в предыдущем пункте по убыванию процентного соотношения мужчин

Выполнение

Листинг 1 – Выполнение задания

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import squarify

# Построить тепловую карту. Оси - года и штаты. Значения - процент
# убийств от общего количества убийств в этот год

homicide_data = pd.read_csv('data/homicide-data.csv', encoding='ISO-8859-1')
homicide_data_b = homicide_data[homicide_data['victim_race'] == 'Black']
homicide_data_b['year'] = homicide_data_b['reported_date'].astype(str).str[:4]
homicides_by_year_state = homicide_data_b.groupby(['year',
'state']).size().unstack(fill_value=0)

total_homicides_per_year = homicides_by_year_state.sum(axis=1)
percentage_homicides = (homicides_by_year_state.T / total_homicides_per_year).T
* 100

plt.figure(figsize=(12, 8))
sns.heatmap(percentage_homicides, annot=True, fmt=".1f", cmap='coolwarm',
cbar_kws={'label': 'Percentage of Homicides'})
plt.xticks(rotation=0)
plt.title('Percentage of Homicides by State and Year')
plt.xlabel('State')
plt.ylabel('Year')
plt.show()

# Процент male female по годам
state_gender_counts = homicide_data.groupby(['state',
'victim_sex']).size().unstack(fill_value=0)
state_gender_percentage =
state_gender_counts.div(state_gender_counts.sum(axis=1), axis=0) * 100
state_gender_percentage =
state_gender_percentage.reset_index().melt(id_vars='state',
var_name='victim_sex', value_name='percentage')

# sort by males
```

```

state_gender_percentage.sort_values()

plt.figure(figsize=(12, 8))
sns.barplot(x='state', y='percentage', hue='victim_sex',
data=state_gender_percentage, palette='coolwarm')
plt.title('Percentage of Men and Women in Each State')
plt.xlabel('State')
plt.ylabel('Percentage')
plt.xticks(rotation=45)
plt.legend(title='Victim Gender')
plt.tight_layout()
plt.show()

grouped = homicide_data.groupby(['state',
'victim_sex']).size().unstack(fill_value=0)
grouped['Total'] = grouped.sum(axis=1)
grouped['Male_Percentage'] = (grouped['Male'] / grouped['Total']) * 100
grouped['Female_Percentage'] = (grouped['Female'] / grouped['Total']) * 100
grouped['Unknown_Percentage'] = (grouped['Unknown'] / grouped['Total']) * 100

# Sorting states by the percentage of men
sorted_states = grouped.sort_values('Male_Percentage', ascending=False).index

# Reordering the DataFrame according to sorted states
grouped = grouped.loc[sorted_states]

# Reshape the data for plotting
stacked_data = grouped[['Male_Percentage', 'Female_Percentage',
'Unknown_Percentage']].stack().reset_index()
stacked_data.columns = ['State', 'Gender', 'Percentage']

# Create the bar plot using Seaborn
plt.figure(figsize=(12, 8))
sns.barplot(x='State', y='Percentage', hue='Gender', data=stacked_data,
palette='muted')
plt.title('Percentage of Men, Women, and Unknown Genders in Each State')
plt.xlabel('State')
plt.ylabel('Percentage')
plt.xticks(rotation=90)
plt.legend(title='Gender')
plt.tight_layout()
plt.show()

```

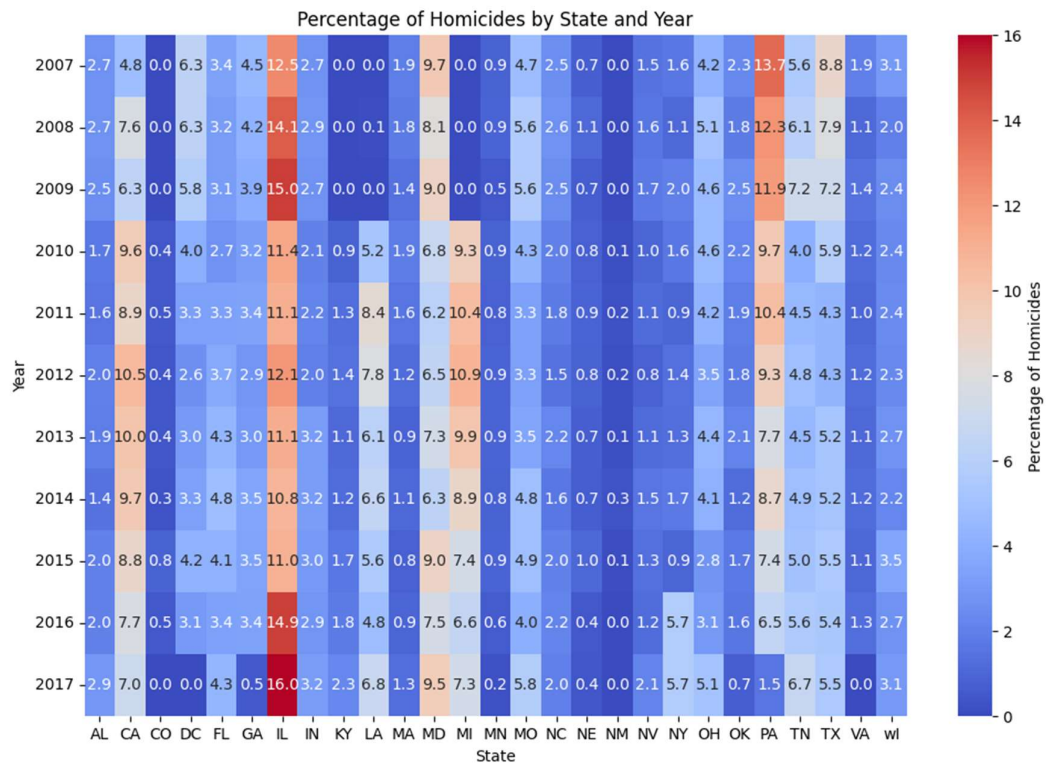


Рисунок 1 – Результат выполнение пункта 1 задания

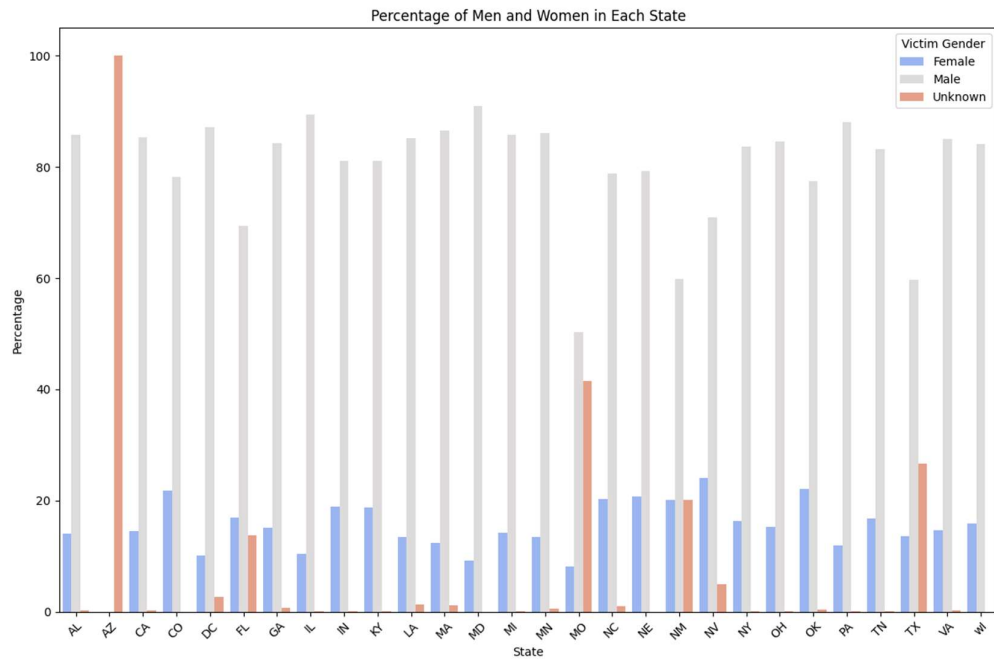


Рисунок 2 – Результат выполнения пункта 2 задания

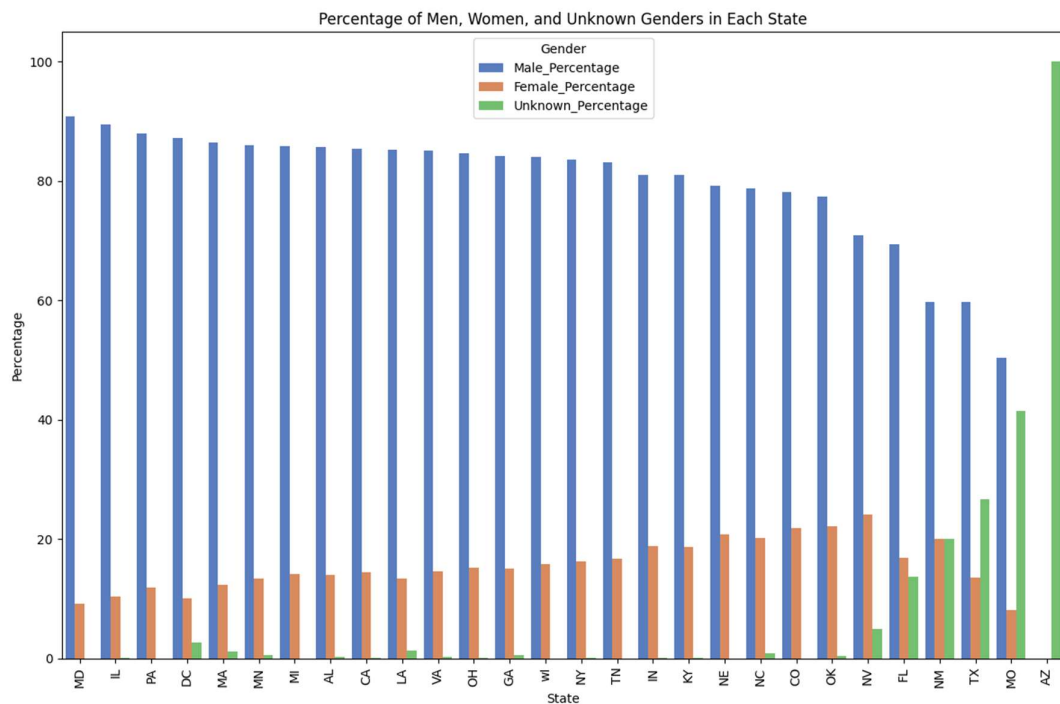


Рисунок 3 – Результат выполнения пункта 3 задания

Вывод

В ходе лабораторной работы мы углубились в работу с библиотекой `seaborn`, освоив методы построения тепловых карт и Bar Chart. Новые навыки включают в себя способность визуализации процентных соотношений убийств по штатам и годам, а также представление данных о проценте жертв по полу в каждом штате с использованием Bar Chart. Кроме того, мы научились эффективно сортировать данные визуализаций для более четкого восприятия информации.