

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ  
УНИВЕРСИТЕТ  
им. Н.Э. Баумана

Факультет «Информатика и системы управления»  
Кафедра «Систем обработки информации и управления»

ОТЧЕТ

**Рубежный контроль №1**  
по дисциплине «Методы машинного обучения в автоматизированных  
системах»

Вариант №6

ИСПОЛНИТЕЛЬ:  
группа ИУ5-22М

\_\_\_\_\_ Калюта Н.И. \_\_\_\_\_  
ФИО

\_\_\_\_\_   
подпись

"30" \_\_\_\_ 05 \_\_\_\_ 2024 г.

ПРЕПОДАВАТЕЛЬ:

\_\_\_\_\_   
ФИО

\_\_\_\_\_   
подпись

" \_\_\_\_ " \_\_\_\_ 2024 г.

Москва - 2024

---

Задача №6.

Для набора данных проведите устранение пропусков для одного (произвольного) числового признака с использованием метода заполнения средним значением.

Задача №26.

Для набора данных для одного (произвольного) числового признака проведите обнаружение и замену (найденными верхними и нижними границами) выбросов на основе правила трех сигм.

Для студентов групп ИУ5-22М, ИУ5И-22М

Для произвольной колонки данных построить гистограмму.

Импортирование необходимых библиотек

```
[1] import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

Подключение гугл диска

```
[3] from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

Загрузка данных из CSV-файла

```
[5] df=pd.read_csv('/content/drive/MyDrive/RK/Electric.csv')
df.head()
```

|   | VIN (1-10) | County | City      | State | Postal Code | Model Year | Make  | Model   | Electric Vehicle Type                  | Alternative Fuel Vehicle (AFV) Eligibility | Electric Range | Base MSRP | Legislative District | DOL Vehicle ID | Vehicle Location                | Electric Utility                             | 2020 Census Tract |
|---|------------|--------|-----------|-------|-------------|------------|-------|---------|--|--|----------------|-----------|----------------------|----------------|---------------------------------|--|-------------------|
| 0 | WBY8P6C58K | King   | Seattle   | WA    | 98115.0     | 2019       | BMW   | I3      | Battery Electric Vehicle (BEV)         | Clean Alternative Fuel Vehicle Eligible    | 153            | 0         | 43.0                 | 259254397      | POINT (-122.3008235 47.6862671) | CITY OF SEATTLE - (WA)CITY OF TACOMA - (WA)  | 5.303300e+10      |
| 1 | 5YJSA1DN4D | Kitsap | Bremerton | WA    | 98312.0     | 2013       | TESLA | MODEL S | Battery Electric Vehicle (BEV)         | Clean Alternative Fuel Vehicle Eligible    | 208            | 69900     | 35.0                 | 127420940      | POINT (-122.6961203 47.5759584) | PUGET SOUND ENERGY INC                       | 5.303508e+10      |
| 2 | 5YJSA1E26J | King   | Kent      | WA    | 98042.0     | 2018       | TESLA | MODEL S | Battery Electric Vehicle (BEV)         | Clean Alternative Fuel Vehicle Eligible    | 249            | 0         | 47.0                 | 170287183      | POINT (-122.1145138 47.3581107) | PUGET SOUND ENERGY INC CITY OF TACOMA - (WA) | 5.303303e+10      |
| 3 | WBY2Z2C54E | King   | Bellevue  | WA    | 98004.0     | 2014       | BMW   | I8      | Plug-in Hybrid Electric Vehicle (PHEV) | Not eligible due to low battery range      | 14             | 0         | 41.0                 | 205545868      | POINT (-122.202397 47.619252)   | PUGET SOUND ENERGY INC CITY OF TACOMA - (WA) | 5.303302e+10      |
| 4 | 5YJXCDE23J | King   | Bellevue  | WA    | 98004.0     | 2018       | TESLA | MODEL X | Battery Electric Vehicle (BEV)         | Clean Alternative Fuel Vehicle Eligible    | 238            | 0         | 41.0                 | 237977386      | POINT (-122.202397 47.619252)   | PUGET SOUND ENERGY INC CITY OF TACOMA - (WA) | 5.303302e+10      |

Считаем количество пропущенных значений (NaN) в каждой колонке

```
df.isna().sum()
```

```
VIN (1-10)      0
County          3
City            3
State           0
Postal Code     3
Model Year      0
Make            0
Model           0
Electric Vehicle Type  0
Clean Alternative Fuel Vehicle (CAFV) Eligibility  0
Electric Range  0
Base MSRP       0
Legislative District  403
DOL Vehicle ID  0
Vehicle Location  8
Electric Utility  3
2020 Census Tract  3
dtype: int64
```

Вычисляем процент пропущенных значений (NaN) для каждого столбца

```
[*((c,df[c].isna().sum()/len(df[c])*100) for c in df.columns)]
```

```
[('VIN (1-10)', 0.0),
 ('County', 0.0016053168092723098),
 ('City', 0.0016053168092723098),
 ('State', 0.0),
 ('Postal Code', 0.0016053168092723098),
 ('Model Year', 0.0),
 ('Make', 0.0),
 ('Model', 0.0),
 ('Electric Vehicle Type', 0.0),
 ('Clean Alternative Fuel Vehicle (CAFV) Eligibility', 0.0),
 ('Electric Range', 0.0),
 ('Base MSRP', 0.0),
 ('Legislative District', 0.21564755804558028),
 ('DOL Vehicle ID', 0.0),
 ('Vehicle Location', 0.0042808448247261596),
 ('Electric Utility', 0.0016053168092723098),
 ('2020 Census Tract', 0.0016053168092723098)]
```

Заполнение пропущенных значений в столбце 'Legislative District' в DataFrame df средним значением этого столбца

```
[20] df["Legislative District"].fillna(df["Legislative District"].mean(), inplace=True)
```

df.head()

|   | VIN (1-10) | County | City      | State | Postal Code | Model Year | Make  | Model   | Electric Vehicle Type                  | Clean Alternative Fuel Vehicle (CAFV) Eligibility | Electric Range | Base MSRP | Legislative District | DOL Vehicle ID | Vehicle Location                | Electric Utility                             | 2020 Census Tract |
|---|------------|--------|-----------|-------|-------------|------------|-------|---------|--|---|----------------|-----------|----------------------|----------------|---------------------------------|--|-------------------|
| 0 | WBY8P6C58K | King   | Seattle   | WA    | 98115.0     | 2019       | BMW   | I3      | Battery Electric Vehicle (BEV)         | Clean Alternative Fuel Vehicle Eligible           | 153            | 0         | 43.0                 | 259254397      | POINT (-122.3008235 47.6862671) | CITY OF SEATTLE - (WA) CITY OF TACOMA - (WA) | 5.303300e+10      |
| 1 | 5YJSA1DN4D | Kitsap | Bremerton | WA    | 98312.0     | 2013       | TESLA | MODEL S | Battery Electric Vehicle (BEV)         | Clean Alternative Fuel Vehicle Eligible           | 208            | 69900     | 35.0                 | 127420940      | POINT (-122.6961203 47.5759584) | PUGET SOUND ENERGY INC CITY OF TACOMA - (WA) | 5.303508e+10      |
| 2 | 5YJSA1E26J | King   | Kent      | WA    | 98042.0     | 2018       | TESLA | MODEL S | Battery Electric Vehicle (BEV)         | Clean Alternative Fuel Vehicle Eligible           | 249            | 0         | 47.0                 | 170287183      | POINT (-122.1145138 47.3581107) | PUGET SOUND ENERGY INC CITY OF TACOMA - (WA) | 5.303303e+10      |
| 3 | WBY2Z2C54E | King   | Bellevue  | WA    | 98004.0     | 2014       | BMW   | I8      | Plug-in Hybrid Electric Vehicle (PHEV) | Not eligible due to low battery range             | 14             | 0         | 41.0                 | 205545868      | POINT (-122.202397 47.619252)   | PUGET SOUND ENERGY INC CITY OF TACOMA - (WA) | 5.303302e+10      |
| 4 | 5YJXCDE23J | King   | Bellevue  | WA    | 98004.0     | 2018       | TESLA | MODEL X | Battery Electric Vehicle (BEV)         | Clean Alternative Fuel Vehicle Eligible           | 238            | 0         | 41.0                 | 237977386      | POINT (-122.202397 47.619252)   | PUGET SOUND ENERGY INC CITY OF TACOMA - (WA) | 5.303302e+10      |

Вычисляем процент пропущенных значений (NaN) для каждого столбца

```
[22] [*((c, df[c].isna().sum() / len(df[c]) * 100) for c in df.columns)]
```

```
[('VIN (1-10)', 0.0),
 ('County', 0.0016053168092723098),
 ('City', 0.0016053168092723098),
 ('State', 0.0),
 ('Postal Code', 0.0016053168092723098),
 ('Model Year', 0.0),
 ('Make', 0.0),
 ('Model', 0.0),
 ('Electric Vehicle Type', 0.0),
 ('Clean Alternative Fuel Vehicle (CAFV) Eligibility', 0.0),
 ('Electric Range', 0.0),
 ('Base MSRP', 0.0),
 ('Legislative District', 0.0),
 ('DOL Vehicle ID', 0.0),
 ('Vehicle Location', 0.0042808448247261596),
 ('Electric Utility', 0.0016053168092723098),
 ('2020 Census Tract', 0.0016053168092723098)]
```

Определение границ выбросов в DataFrame с использованием метода, основанного на стандартном отклонении.

```
[1] from enum import Enum

class OutlierBoundaryType(Enum):
    SIGMA = 1
    QUANTILE = 2
    IRQ = 3

def get_outlier_boundaries(df, col):
    K1 = 3
    lower_boundary = df[col].mean() - (K1 * df[col].std())
    upper_boundary = df[col].mean() + (K1 * df[col].std())

    return lower_boundary, upper_boundary
```

## Удаление выбросов из столбца "Legislative District"

```
[1] lower_boundary, upper_boundary = get_outlier_boundaries(df, "Legislative District")
outliers_temp = np.where(
    df["Legislative District"] > upper_boundary, True, np.where(df["Legislative District"] < lower_boundary, True, False)
)
data_trimmed = df.loc[~(outliers_temp),]
```

data\_trimmed

|        | VIN (1-10) | County    | City      | State | Postal Code | Model Year | Make   | Model      | Electric Vehicle Type                  | Clean Alternative Fuel Vehicle (CAFEV) Eligibility | Electric Range | Base MSRP | Legislative District | DOL Vehicle ID | Vehicle Location                | Electric Utility                                 | 2020 Ce T |
|--------|------------|-----------|-----------|-------|-------------|------------|--------|------------|--|--|----------------|-----------|----------------------|----------------|---------------------------------|--|-----------|
| 0      | WBY8P6C58K | King      | Seattle   | WA    | 98115.0     | 2019       | BMW    | I3         | Battery Electric Vehicle (BEV)         | Clean Alternative Fuel Vehicle Eligible            | 153            | 0         | 43.0                 | 259254397      | POINT (-122.3008235 47.6862671) | CITY OF SEATTLE - (WA) CITY OF TACOMA - (WA)     | 5.303300  |
| 1      | 5YJSA1DN4D | Kitsap    | Bremerton | WA    | 98312.0     | 2013       | TESLA  | MODEL S    | Battery Electric Vehicle (BEV)         | Clean Alternative Fuel Vehicle Eligible            | 208            | 69900     | 35.0                 | 127420940      | POINT (-122.6961203 47.5759584) | PUGET SOUND ENERGY INC CITY OF TACOMA - (WA)     | 5.303508  |
| 2      | 5YJSA1E26J | King      | Kent      | WA    | 98042.0     | 2018       | TESLA  | MODEL S    | Battery Electric Vehicle (BEV)         | Clean Alternative Fuel Vehicle Eligible            | 249            | 0         | 47.0                 | 170287183      | POINT (-122.1145138 47.3581107) | PUGET SOUND ENERGY INC CITY OF TACOMA - (WA)     | 5.303303  |
| 3      | WBY222C54E | King      | Bellevue  | WA    | 98004.0     | 2014       | BMW    | I8         | Plug-in Hybrid Electric Vehicle (PHEV) | Not eligible due to low battery range              | 14             | 0         | 41.0                 | 205545868      | POINT (-122.202397 47.619252)   | PUGET SOUND ENERGY INC CITY OF TACOMA - (WA)     | 5.303302  |
| 4      | 5YJXCDE23J | King      | Bellevue  | WA    | 98004.0     | 2018       | TESLA  | MODEL X    | Battery Electric Vehicle (BEV)         | Clean Alternative Fuel Vehicle Eligible            | 238            | 0         | 41.0                 | 237977386      | POINT (-122.202397 47.619252)   | PUGET SOUND ENERGY INC CITY OF TACOMA - (WA)     | 5.303302  |
| ...    | ...        | ...       | ...       | ...   | ...         | ...        | ...    | ...        | ...                                    | ...  | ...            | ...       | ...                  | ...            | ...                             | ...  | ...       |
| 186874 | JTMEB3FVXM | Snohomish | Arlington | WA    | 98223.0     | 2021       | TOYOTA | RAV4 PRIME | Plug-in Hybrid Electric Vehicle (PHEV) | Clean Alternative Fuel Vehicle Eligible            | 42             | 0         | 10.0                 | 167257387      | POINT (-122.11597 48.194109)    | BONNEVILLE POWER ADMINISTRATION PUD 1 OF SNOH... | 5.306105  |

## Дополнительное задание

```
[23] plt.figure(figsize=(10, 6))
sns.histplot(df['Legislative District'], kde=False)
plt.title('Гистограмма распределения законодательных округов')
plt.xlabel('Законодательный округ')
plt.ylabel('Количество')
plt.show()
```



Гистограмма распределения законодательных округов

