

## Вариант 6

### Задача №6.

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

Датасет: Бар

```
import matplotlib
from mpl_toolkits import mplot3d
import pandas as pd
from matplotlib import pyplot as plt
import numpy as np
import seaborn as sns
import os
```

#Загрузка данных из файла

```
path=os.environ["userprofile"]+"\\\\"+".atom"+"\\\\"+"Transactions.csv"
print(path)
```

```
data = pd.read_csv(path)
print(data)
```

#Замена пропущенных значений средним значением

```
mean_value = data['Percentage_markup'].mean()
data['Percentage_markup'].fillna(mean_value, inplace=True)
```

#Построение парных диаграмм

```
sns.pairplot(data, vars=['Percentage_markup', 'Profit'], diag_kind='kde')
```

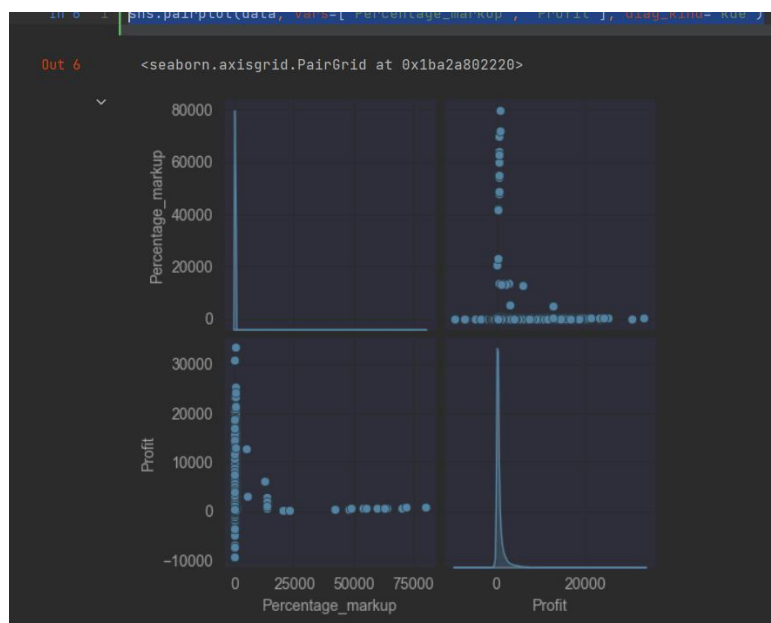


Рисунок 1. Парная диаграмма (Процентная отметка и Прибыль)

#Сохранение изменённых данных в файл

```
data.to_csv('file_with_filled_values.csv', index=False)
```

### **Вывод:**

В этом примере мы загружаем данные из файла 'Transactions.csv', заменяем пропущенные значения в столбце 'Percentage\_markup' средним значением, строим парные диаграммы для столбцов 'Percentage\_markup' и 'Profit' с помощью метода pairplot из библиотеки seaborn и сохраняем измененные данные в файл 'file\_with\_filled\_values.csv'.

Я использовала параметр diag\_kind='kde' для построения графиков плотности распределения на диагонали парных диаграмм.

### **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.

Датасет: Об убийствах

Вот пример кода на языке Python, который демонстрирует, как это можно сделать:

```
import matplotlib
from mpl_toolkits import mplot3d
import pandas as pd
from matplotlib import pyplot as plt
import numpy as np
import seaborn as sns
import os
```

#Загрузка данных из файла

```
path=os.environ["userprofile"]+"\\\\"+".atom"+"\\\\"+"database.csv"
print(path)
```

```
data = pd.read_csv(path)
print(data)
```

#Вычисление среднего и стандартного отклонения

```
mean_value = data['Year']. mean ()
std_value = data['Year']. std()
```

#Вычисление верхней и нижней границы

```
lower_bound = mean_value - 3 * std_value
upper_bound = mean_value + 3 * std_value
```

#Замена выбросов на границы

```
data.loc[data['Year'] < lower_bound, 'Year'] = lower_bound
data.loc[data['Year'] > upper_bound, 'Year'] = upper_bound
```

#Построение парных диаграмм

```
sns.pairplot(data, vars=['Year', 'Incident'][:500], diag_kind='kde')
```

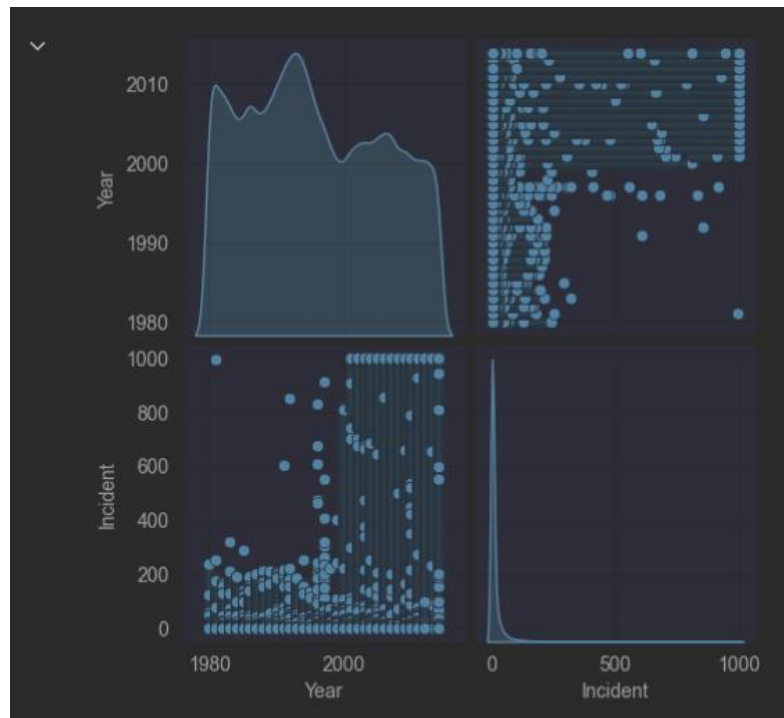


Рисунок 2. Парная диаграмма (Год и количество инцидентов)

### **Вывод:**

В этом примере мы загружаем данные из файла 'database.csv', вычисляем среднее значение и стандартное отклонение для столбца 'Year', вычисляем верхнюю и нижнюю границы на основе правила трех сигм, заменяем выбросы на границы, строим парные диаграммы для столбцов 'Year' и 'Incident' с помощью метода pairplot из библиотеки seaborn и сохраняем измененные данные в файл 'file\_with\_replaced\_outliers.csv'.

Я использовала параметр `diag_kind='kde'` для построения графиков плотности распределения на диагонали парных диаграмм.