

Лабораторная работа №1

1. Текстовое описание датасета "Iris"

Датасет "Iris" содержит информацию о 150 образцах ирисов, разделенных на три вида: Iris-setosa, Iris-versicolor и Iris-virginica. Для каждого образца измерены четыре признака: длина и ширина чашелистника (sepal length и sepal width), а также длина и ширина лепестка (petal length и petal width). Целью работы с этим датасетом обычно является классификация образцов по видам на основе измеренных признаков.

2. Основные характеристики датасета

Количество образцов: 150

Количество признаков: 4 (sepal length, sepal width, petal length, petal width)

Количество классов: 3 (Iris-setosa, Iris-versicolor, Iris-virginica)

Пропуски в данных: отсутствуют

3. Визуальное исследование датасета

```
import pandas as pd
from sklearn.datasets import load_iris

# Загрузка датасета
iris = load_iris()
df = pd.DataFrame(iris.data, columns=iris.feature_names)
df['species'] = iris.target
df['species'] = df['species'].map({0: 'setosa', 1: 'versicolor', 2:
'virginica'})

# Основные характеристики датасета
print("Основные характеристики датасета:")
print(f"Количество образцов: {df.shape[0]}")
print(f"Количество признаков: {df.shape[1] - 1}") # исключаем столбец
'species'
print(f"Количество классов: {df['species'].nunique()}")
print("\nПервые 5 строк датасета:")
print(df.head())
```

Основные характеристики датасета:

Количество образцов: 150

Количество признаков: 4

Количество классов: 3

Первые 5 строк датасета:

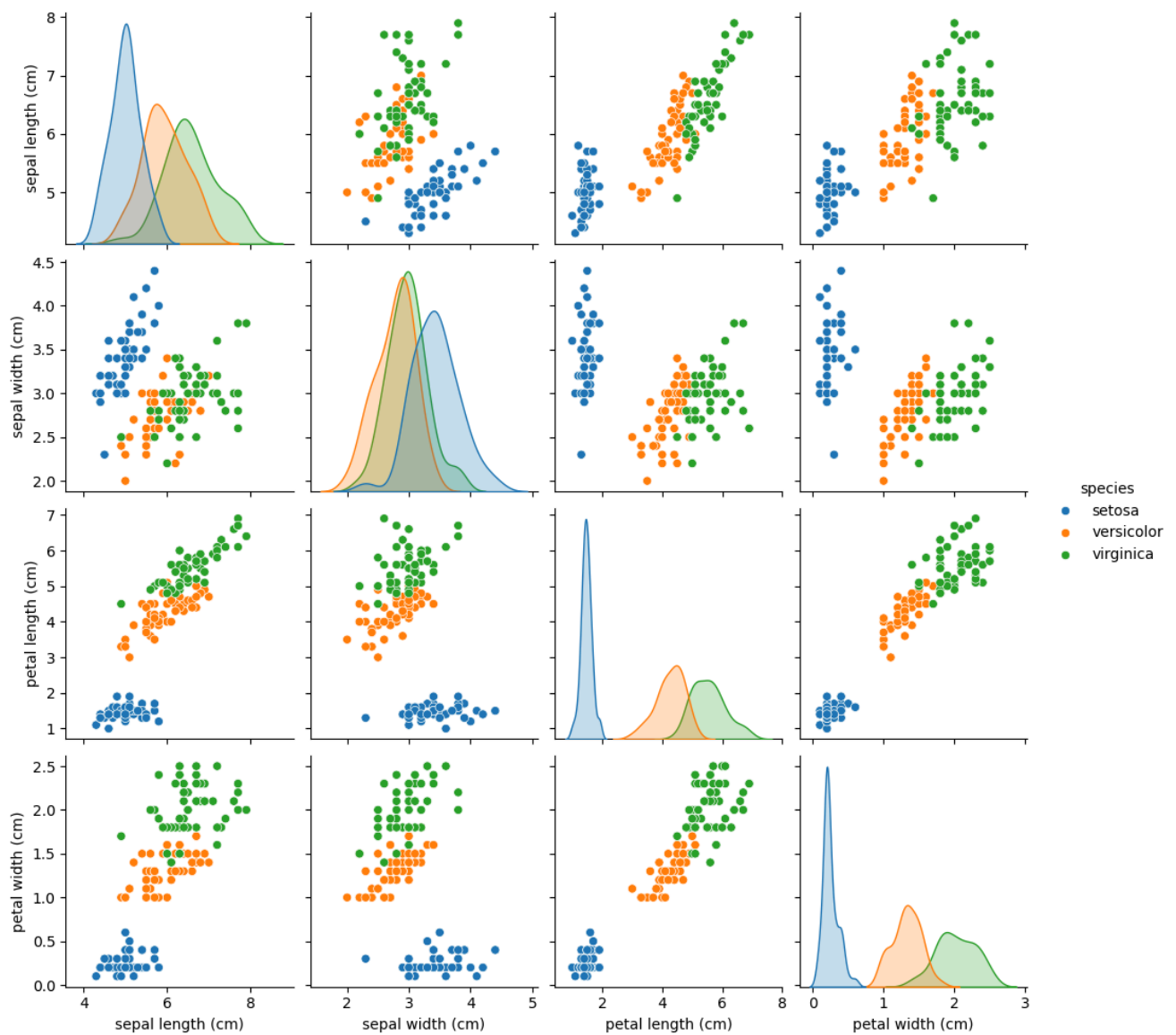
sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
-------------------	------------------	-------------------	------------------

```
\
0      5.1      3.5      1.4      0.2
1      4.9      3.0      1.4      0.2
2      4.7      3.2      1.3      0.2
3      4.6      3.1      1.5      0.2
4      5.0      3.6      1.4      0.2

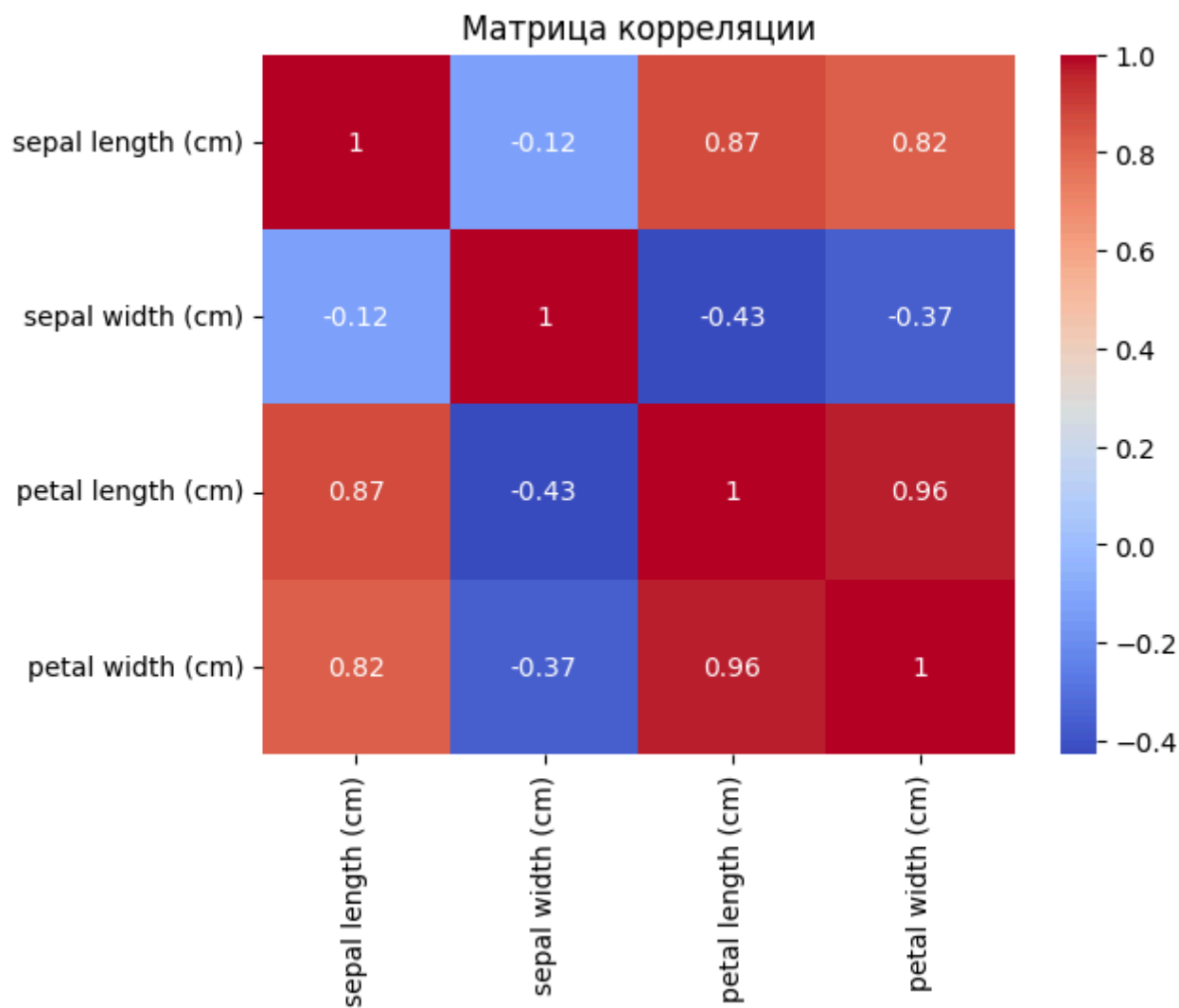
species
0  setosa
1  setosa
2  setosa
3  setosa
4  setosa
```

```
import seaborn as sns
import matplotlib.pyplot as plt

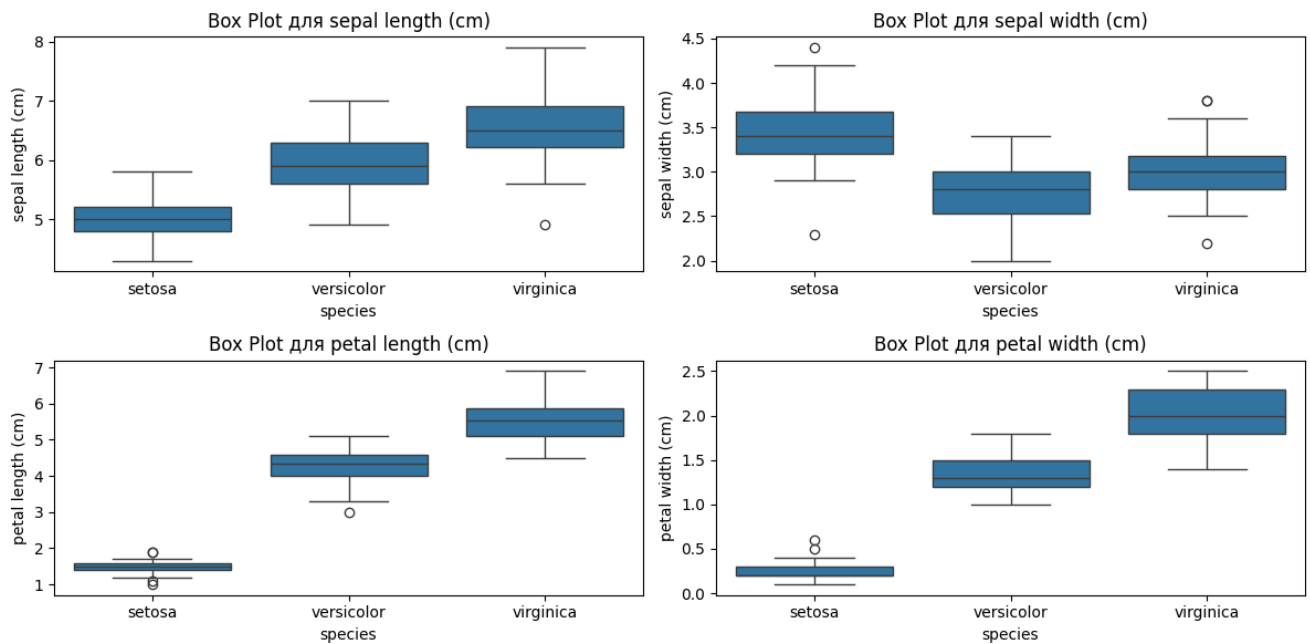
# Визуализация данных с помощью Pair Plot
sns.pairplot(df, hue='species')
plt.show()
```



```
# Информация о корреляции признаков
corr = df.iloc[:, :-1].corr()
sns.heatmap(corr, annot=True, cmap='coolwarm')
plt.title('Матрица корреляции')
plt.show()
```



```
# Box Plot для анализа распределения признаков по видам ирисов
plt.figure(figsize=(12, 6))
for i, feature in enumerate(iris.feature_names):
    plt.subplot(2, 2, i + 1)
    sns.boxplot(x='species', y=feature, data=df)
    plt.title(f'Box Plot для {feature}')
plt.tight_layout()
plt.show()
```



```
# Violin Plot для анализа распределения и плотности данных
plt.figure(figsize=(12, 6))
for i, feature in enumerate(iris.feature_names):
    plt.subplot(2, 2, i + 1)
    sns.violinplot(x='species', y=feature, data=df, inner='quartile')
    plt.title(f'Violin Plot для {feature}')
plt.tight_layout()
plt.show()
```

