

Иванов Кирилл Евгеньевич ИУ5-61Б

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

file_path = "heart.csv"
df = pd.read_csv(file_path)

print("Шаг 1: Данные загружены")
print(f"Форма данных: {df.shape[0]} строк, {df.shape[1]} столбцов")

missing_values = df.isnull().sum()
if missing_values.any():
    print("\n Предупреждение: Обнаружены пропущенные значения:")
    print(missing_values)

    df_clean = df.dropna()
    print(f"\n Пропущенные значения удалены. Новая форма данных: {df_clean.shape}")
else:
    df_clean = df.copy()
    print("\n Пропущенных значений не обнаружено.")

corr_matrix = df_clean.corr(numeric_only=True)

plt.figure(figsize=(12, 8))
sns.heatmap(corr_matrix, annot=True, fmt=".2f", cmap='coolwarm', linewidths=0.5)
plt.title("Корреляционная матрица признаков")
plt.tight_layout()
plt.show()

correlation_with_target = corr_matrix['target'].sort_values(ascending=False)
print("\n Корреляция признаков с целевой переменной 'target':")
print(correlation_with_target)

plt.figure(figsize=(10, 6))
correlation_with_target.drop('target').plot(kind='barh', color='skyblue')
plt.title("Степень корреляции признаков с наличием сердечной болезни")
plt.xlabel("Корреляция")
plt.ylabel("Признаки")
plt.axvline(0, color='black', linestyle='--')
plt.tight_layout()
plt.show()

print("\n Выводы:")
```

```
print("- Наиболее сильно влияют на наличие сердечной болезни:")
print("  - thalach (максимальная ЧСС): отрицательная корреляция ( $\sim -0.4$ ")
print("  - slope (наклон ST-сегмента): положительная ( $\sim +0.35$ ")
print("  - cp (тип боли в груди): положительная ( $\sim +0.24$ )\n")
print("- Менее значимые признаки:")
print("  - trestbps (давление в покое), fbs (уровень сахара натощак)\n")
print("- Модель машинного обучения может быть построена, так как признаки имеют")
print("- Рекомендуется использовать модели классификации: Logistic Regression, R
```