load-diabetes

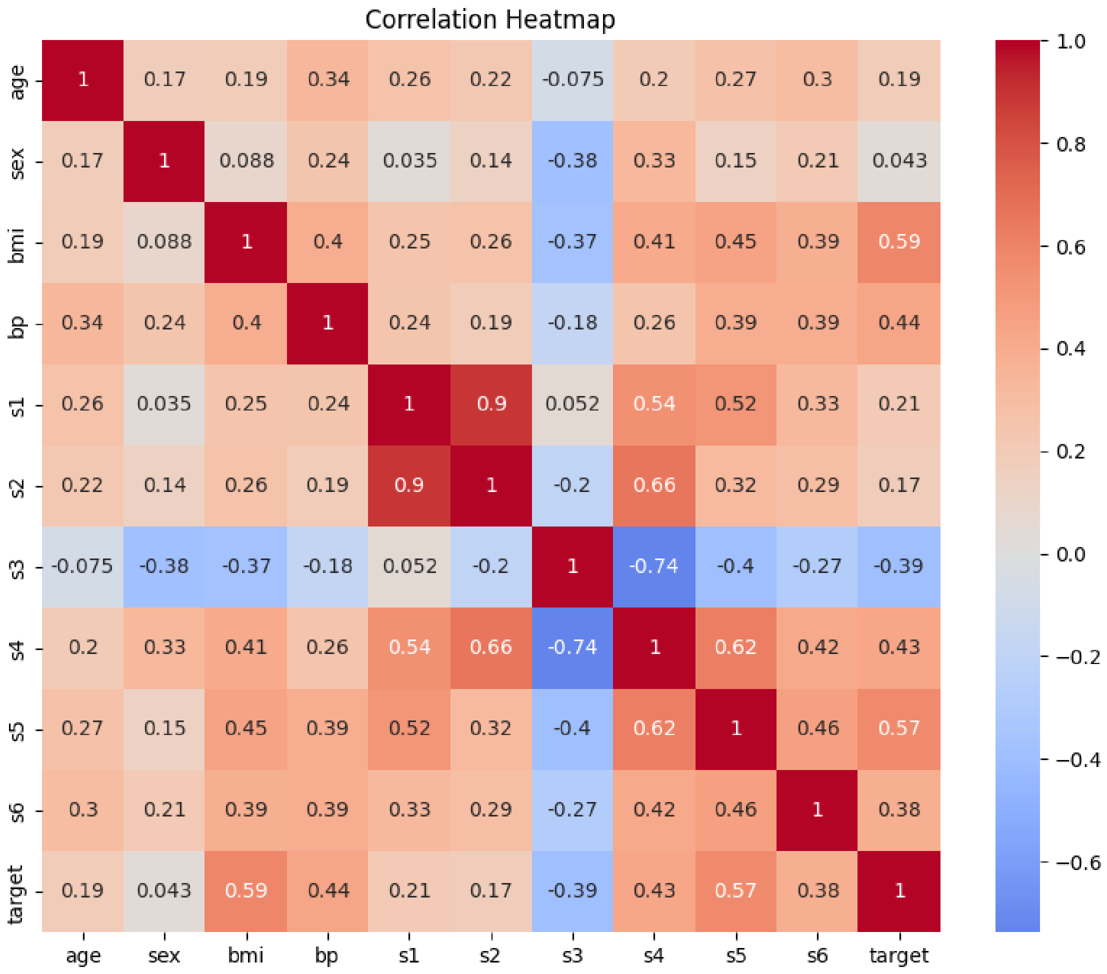
June 20, 2024

[5]:

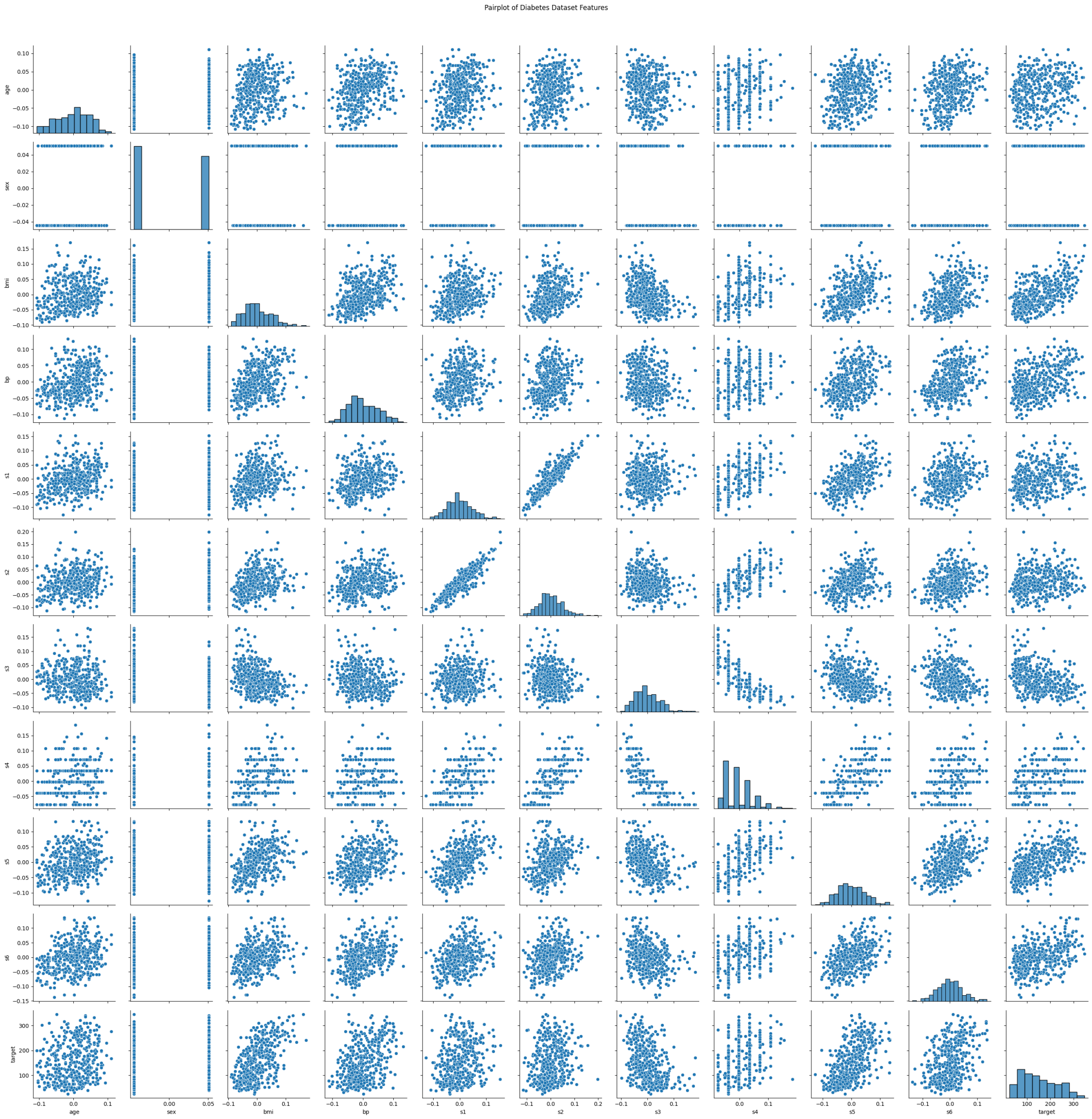
```python
# Import necessary libraries
from sklearn.datasets import load_diabetes
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

# Load dataset
diabetes = load_diabetes()
X, y = diabetes.data, diabetes.target

# Convert data to DataFrame for easier analysis
df = pd.DataFrame(data=X, columns=diabetes.feature_names)
df['target'] = y

# Display basic statistics and information
print(f"Dataset shape: {df.shape}")
print(f"Columns: {df.columns}")
print(f"Target variable summary:\n{df['target'].describe()}")

# Display correlation heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(df.corr(), annot=True, cmap='coolwarm', center=0)
plt.title('Correlation Heatmap')
plt.show()

# Pairplot for visualizing relationships and distributions
sns.pairplot(df, diag_kind='hist')
plt.suptitle('Pairplot of Diabetes Dataset Features', y=1.02)
plt.tight_layout()
plt.show()

# Train-test split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
 ↪random_state=42)
```

```python
# Train Linear Regression model
lr = LinearRegression()
lr.fit(X_train, y_train)

# Predict and evaluate
y_pred = lr.predict(X_test)
print(f"\nMean Squared Error: {mean_squared_error(y_test, y_pred)}")
```

```
Dataset shape: (442, 11)
Columns: Index(['age', 'sex', 'bmi', 'bp', 's1', 's2', 's3', 's4', 's5', 's6',
       'target'],
      dtype='object')
Target variable summary:
count    442.000000
mean     152.133484
std       77.093005
min       25.000000
25%       87.000000
50%      140.500000
75%      211.500000
max      346.000000
Name: target, dtype: float64
```



	age	sex	bmi	bp	s1	s2	s3	s4	s5	s6	target
age	1	0.17	0.19	0.34	0.26	0.22	-0.075	0.2	0.27	0.3	0.19
sex	0.17	1	0.088	0.24	0.035	0.14	-0.38	0.33	0.15	0.21	0.043
bmi	0.19	0.088	1	0.4	0.25	0.26	-0.37	0.41	0.45	0.39	0.59
bp	0.34	0.24	0.4	1	0.24	0.19	-0.18	0.26	0.39	0.39	0.44
s1	0.26	0.035	0.25	0.24	1	0.9	0.052	0.54	0.52	0.33	0.21
s2	0.22	0.14	0.26	0.19	0.9	1	-0.2	0.66	0.32	0.29	0.17
s3	-0.075	-0.38	-0.37	-0.18	0.052	-0.2	1	-0.74	-0.4	-0.27	-0.39
s4	0.2	0.33	0.41	0.26	0.54	0.66	-0.74	1	0.62	0.42	0.43
s5	0.27	0.15	0.45	0.39	0.52	0.32	-0.4	0.62	1	0.46	0.57
s6	0.3	0.21	0.39	0.39	0.33	0.29	-0.27	0.42	0.46	1	0.38
target	0.19	0.043	0.59	0.44	0.21	0.17	-0.39	0.43	0.57	0.38	1



```
Mean Squared Error: 2900.193628493482
```