

Explorer

Open Editors

- main.ipynb portal code\day2\conce... U
- question.md portal code\day2\co... 2, U
- image.png portal code\day2\conce... U
- ML470\_S2\_Diabetes\_Data\_Concep... U

advanced-ml-techniques

- .venv
- cheatsheets
- notes
- day1
  - resources
    - session1.md U
- day2
- portal code
  - day1
  - day2
    - concept
      - 1. ML470\_Sprint02\_Support\_Ve... U
      - 2. ML470\_Sprint02\_Support\_Ve... U
        - image.png U
        - main.ipynb U
        - ML470\_S2\_Diabetes\_Data\_Co... U
        - question.md 2, U
      - 3. ML470\_Sprint02\_Support\_Vector\_... U
  - practice
  - syllabus
  - .gitignore
  - readme.md M

portal code > day2 > concept > 2. ML470\_Sprint02\_Support\_Vector\_Machines\_Concept\_Q2 > main.ipynb > assess\_outliers{

+ Code + Markdown | Run All Restart Clear All Outputs Jupyter Variables Outline ...

.venv (Python 3.12.10)

```
def assess_outliers(data, title):
    features = ['Glucose', 'BMI', 'Age', 'Insulin']

    plt.figure(figsize=(12, 5))
    sns.boxplot(data=data[features])
    plt.title(title, fontsize=14)
    plt.xticks(rotation=45)
    plt.tight_layout()
    plt.show()
```

[14] ✓ 0.0s Python

```
print("Visualization 1: Before Outlier Treatment")
assess_outliers(data, "Before Outlier Treatment")
```

[15] ✓ 0.2s Python

Visualization 1: Before Outlier Treatment

Before Outlier Treatment

800

600

400

200

0

Glucose BMI Age Insulin

W

FileEditSelectionViewGoRunTerminalHelp

←→

advanced-ml-techniques - main.ipynb

W

View Changelog X

—

X

main.ipynb U

question.md 2, U

image.png U

ML470\_S2\_Diabetes\_Data\_Concept.xlsx U

portal code > day2 > concept > 2. ML470\_Sprint02\_Support\_Vector\_Machines\_Concept\_Q2 > main.ipynb > assess\_outliers()

+ Code + Markdown ▶ Run All ⌂ Restart 🗑 Clear All Outputs 📄 Jupyter Variables ☰ Outline ...

main.ipynb portal code\day2\conce... U

question.md portal code\day2\co... 2, U

image.png portal code\day2\conce... U

ML470\_S2\_Diabetes\_Data\_Concep... U

advanced-ml-techniques

> .env

> cheatsheets

> notes

> day1

> resources

> session1.md U

> day2

> portal code

> day1

> day2

> concept

> 1. ML470\_Sprint02\_Support\_Ve...

> 2. ML470\_Sprint02\_Support\_Ve...

> image.png U

> main.ipynb U

> ML470\_S2\_Diabetes\_Data\_Co... U

> question.md 2, U

> 3. ML470\_Sprint02\_Support\_Vec...

> practice

> syllabus

> .gitignore

> readme.md M

def treat\_outliers(data):

df = data.copy()

features = ['Glucose', 'BMI', 'Age', 'Insulin']

for col in features:

Q1 = df[col].quantile(0.25)

Q3 = df[col].quantile(0.75)

IQR = Q3 - Q1

lower = Q1 - 1.5 \* IQR

upper = Q3 + 1.5 \* IQR

df[col] = df[col].clip(lower, upper)

return df

[17] ✓ 0.0s

Python

treated\_data = treat\_outliers(data)

[18] ✓ 0.0s

Python

assess\_outliers(

treated\_data,

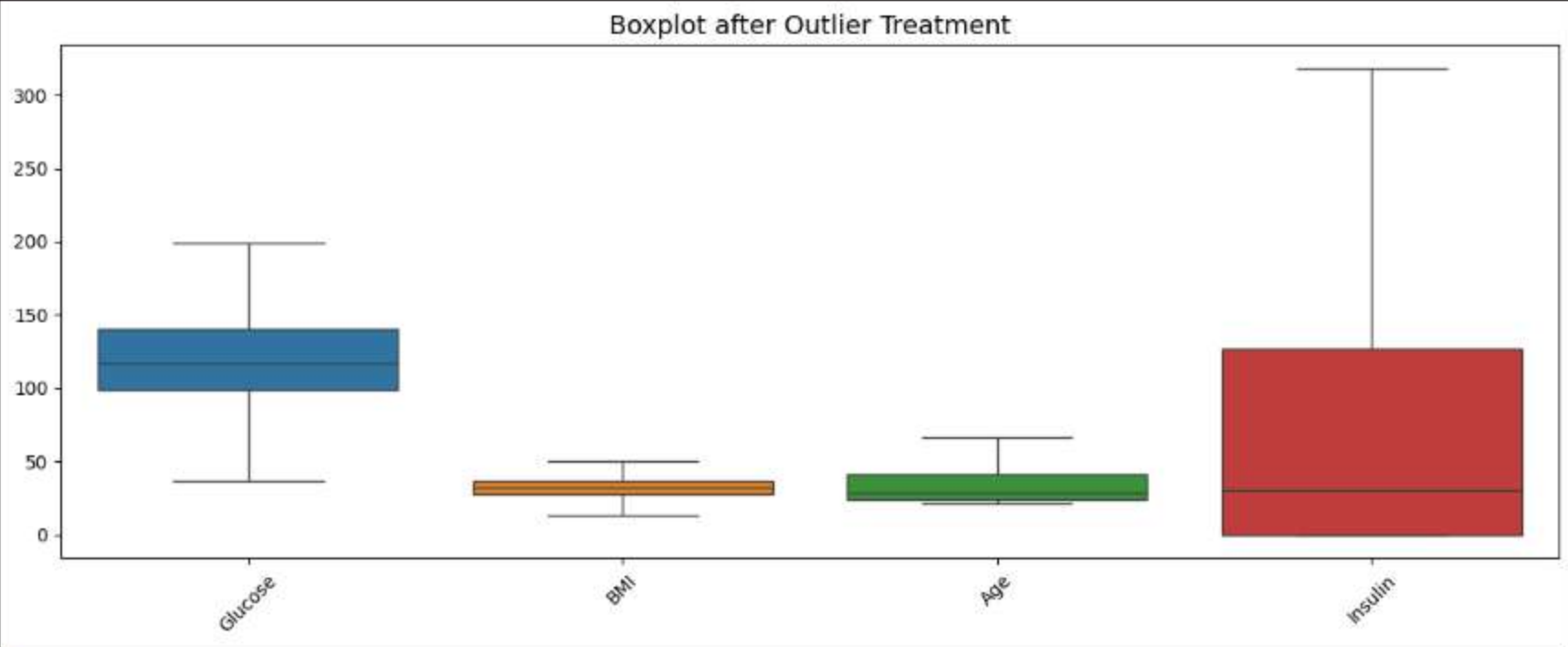
title="Boxplot after Outlier Treatment"

)

[19] ✓ 0.3s

Python

Boxplot after Outlier Treatment



A boxplot titled "Boxplot after Outlier Treatment" showing the distribution of four variables: Glucose, BMI, Age, and Insulin. The y-axis represents values from 0 to 300. Glucose (blue box) has a median around 115, with whiskers extending from approximately 35 to 200. BMI (orange box) has a median around 30, with whiskers from approximately 15 to 50. Age (green box) has a median around 30, with whiskers from approximately 20 to 65. Insulin (red box) has a median around 30, with whiskers from approximately 0 to 320. The boxes represent the interquartile range (IQR), and the horizontal line inside each box is the median.

Variable	Min	Q1	Median	Q3	Max
Glucose	35	100	115	140	200
BMI	15	25	30	35	50
Age	20	25	30	40	65
Insulin	0	0	30	125	320

> Outline

> Timeline

master\*

Launchpad

0 2

Git Graph

Spaces: 4

()

Cell 7 of 7

Go Live

Free - Upgrade Now

Windsurf - Settings

Prettier