Conclusions: === T-Test: Charles River vs MEDV ===

T-statistic: -3.9964, P-value: 0.000074

Conclusion: Significant difference

=== ANOVA: MEDV Across AGE Groups ===

F-statistic: 36.4076, P-value: 0.000000

Conclusion: At least one group differs significantly

=== Pearson Correlation: NOX vs INDUS ===

Correlation (r): 0.7637, P-value: 0.000000

Conclusion: Significant relationship

=== Regression: Distance to Employment vs MEDV ===

Coefficient: 1.0916, Intercept: 18.3901, R²: 0.0625

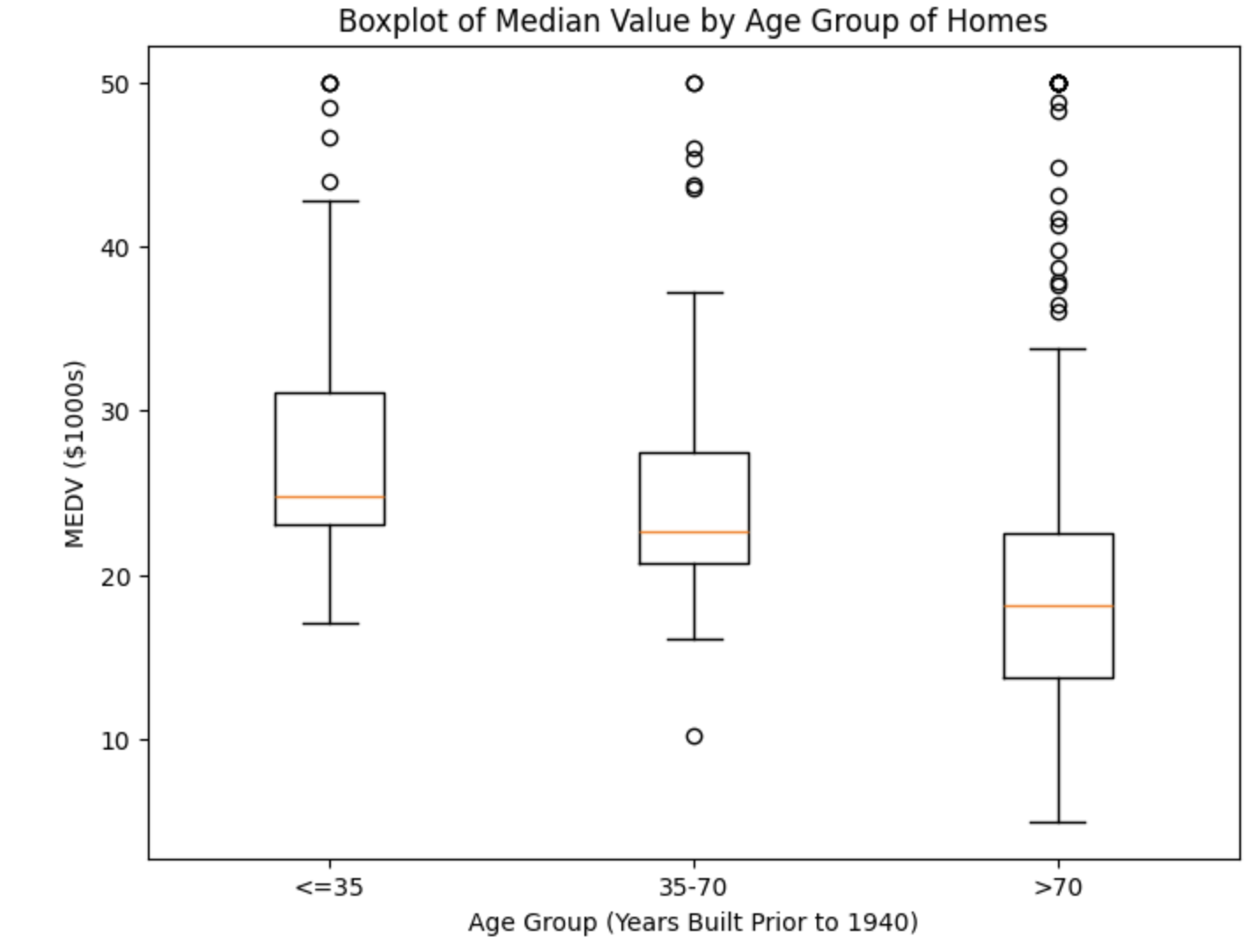
Interpretation: Each additional unit of distance is associated with a 1.09 change in MEDV

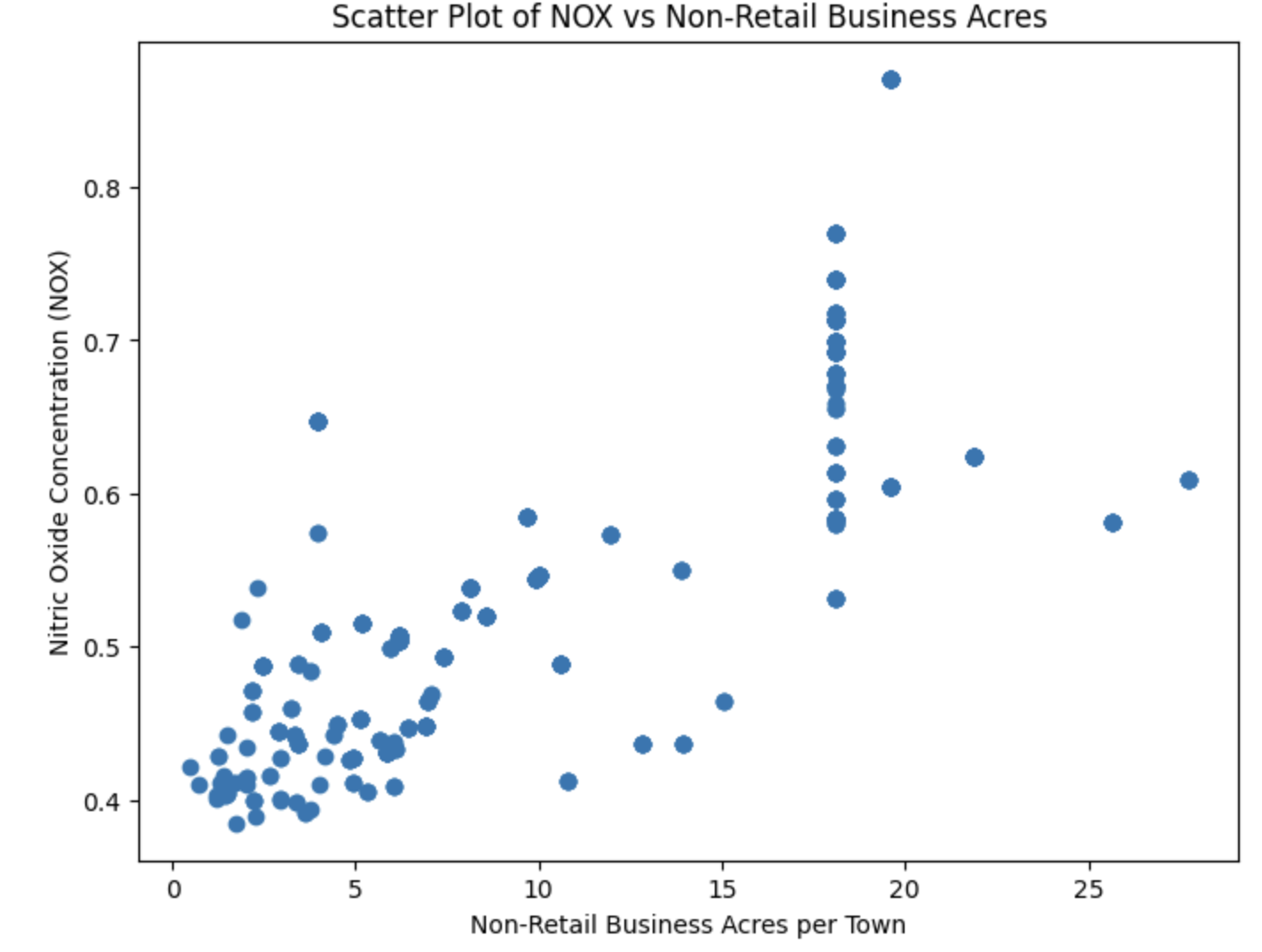
A diagram of a box with a square and a line

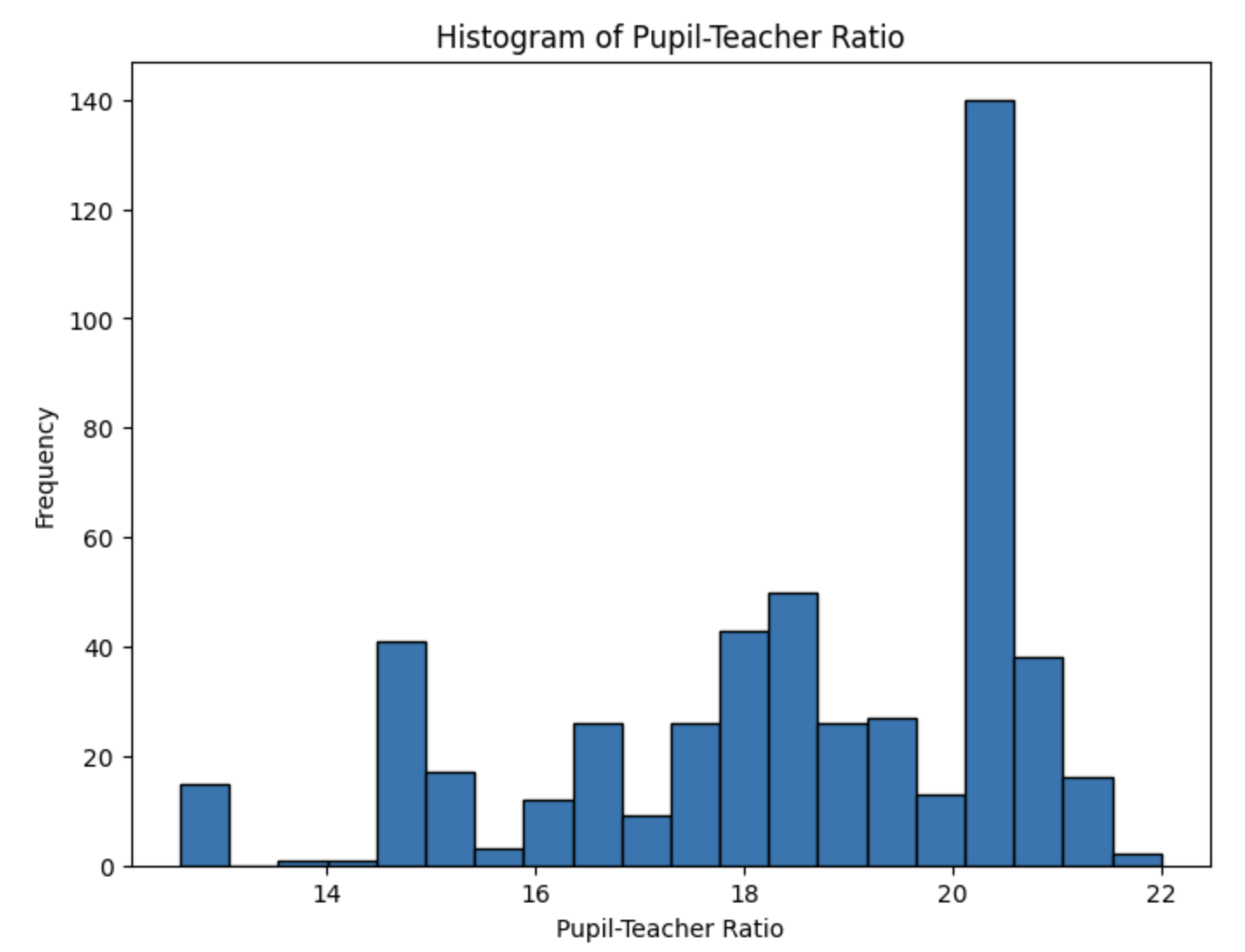
AI-generated content may be incorrect.

A bar graph with blue rectangles

AI-generated content may be incorrect.







from js import fetch

import io

import pandas as pd

URL = 'https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-ST0151EN-SkillsNetwork/labs/boston\_housing.csv'

resp = await fetch(URL)

boston\_url = io.BytesIO((await resp.arrayBuffer()).to\_py())

df = pd.read\_csv(boston\_url)

df.rename(columns={

'CHAS': 'Charles\_River',

'NOX': 'Nitric\_Oxide',

'INDUS': 'Non\_Retail\_Business\_Acres',

'AGE': 'Pre\_1940\_Owner\_Occupied',

'DIS': 'Distance\_to\_Employment\_Centres',

'PTRATIO': 'Pupil\_Teacher\_Ratio',

'MEDV': 'MEDV'

}, inplace=True)

df['AGE\_Group'] = pd.cut(df['Pre\_1940\_Owner\_Occupied'], bins=[0, 35, 70, 100],

labels=['<=35', '35-70', '>70'])

import matplotlib.pyplot as plt

from scipy import stats

from sklearn.linear\_model import LinearRegression

plt.figure(figsize=(8, 6))

plt.boxplot(df['MEDV'])

plt.title('Boxplot of Median Value of Owner-Occupied Homes')

plt.ylabel('MEDV ($1000s)')

plt.show()

plt.figure(figsize=(8, 6))

plt.hist(df['Charles\_River'], bins=2, rwidth=0.8)

plt.title('Bar Plot of Charles River Variable')

plt.xlabel('Bounds Charles River (1 = Yes, 0 = No)')

plt.ylabel('Count of Houses')

plt.xticks([0, 1])

plt.show()

plt.figure(figsize=(8, 6))

data = [

df[df['AGE\_Group'] == '<=35']['MEDV'],

df[df['AGE\_Group'] == '35-70']['MEDV'],

df[df['AGE\_Group'] == '>70']['MEDV']

]

plt.boxplot(data, labels=['<=35', '35-70', '>70'])

plt.title('Boxplot of Median Value by Age Group of Homes')

plt.xlabel('Age Group (Years Built Prior to 1940)')

plt.ylabel('MEDV ($1000s)')

plt.show()

plt.figure(figsize=(8, 6))

plt.scatter(df['Non\_Retail\_Business\_Acres'], df['Nitric\_Oxide'])

plt.title('Scatter Plot of NOX vs Non-Retail Business Acres')

plt.xlabel('Non-Retail Business Acres per Town')

plt.ylabel('Nitric Oxide Concentration (NOX)')

plt.show()

plt.figure(figsize=(8, 6))

plt.hist(df['Pupil\_Teacher\_Ratio'], bins=20, edgecolor='black')

plt.title('Histogram of Pupil-Teacher Ratio')

plt.xlabel('Pupil-Teacher Ratio')

plt.ylabel('Frequency')

plt.show()

# T-Test: Charles River vs MEDV

chas\_0 = df[df['Charles\_River'] == 0]['MEDV']

chas\_1 = df[df['Charles\_River'] == 1]['MEDV']

t\_stat, p\_val\_ttest = stats.ttest\_ind(chas\_0, chas\_1)

# ANOVA: MEDV across AGE groups

anova\_result = stats.f\_oneway(

df[df['AGE\_Group'] == '<=35']['MEDV'],

df[df['AGE\_Group'] == '35-70']['MEDV'],

df[df['AGE\_Group'] == '>70']['MEDV']

)

# Pearson Correlation: NOX vs INDUS

r\_val, p\_val\_corr = stats.pearsonr(df['Nitric\_Oxide'], df['Non\_Retail\_Business\_Acres'])

# Regression: DIS vs MEDV

X = df[['Distance\_to\_Employment\_Centres']]

y = df['MEDV']

model = LinearRegression().fit(X, y)

regression\_coef = model.coef\_[0]

intercept = model.intercept\_

r\_squared = model.score(X, y)

print("=== T-Test: Charles River vs MEDV ===")

print(f"T-statistic: {t\_stat:.4f}, P-value: {p\_val\_ttest:.6f}")

print("Conclusion:", "Significant difference" if p\_val\_ttest < 0.05 else "No significant difference")

print("\n=== ANOVA: MEDV Across AGE Groups ===")

print(f"F-statistic: {anova\_result.statistic:.4f}, P-value: {anova\_result.pvalue:.6f}")

print("Conclusion:", "At least one group differs significantly" if anova\_result.pvalue < 0.05 else "No significant difference among groups")

print("\n=== Pearson Correlation: NOX vs INDUS ===")

print(f"Correlation (r): {r\_val:.4f}, P-value: {p\_val\_corr:.6f}")

print("Conclusion:", "Significant relationship" if p\_val\_corr < 0.05 else "No significant relationship")

print("\n=== Regression: Distance to Employment vs MEDV ===")

print(f"Coefficient: {regression\_coef:.4f}, Intercept: {intercept:.4f}, R²: {r\_squared:.4f}")

print(f"Interpretation: Each additional unit of distance is associated with a {regression\_coef:.2f} change in MEDV")