# Bike Price Data Analysis Project

import pandas as pd

import matplotlib.pyplot as plt

from pptx import Presentation

from pptx.util import Inches

from pptx.chart.data import CategoryChartData

from pptx.enum.chart import XL\_CHART\_TYPE

# ------------------------------

# 1. Create Sample Dataset

# ------------------------------

data = {

'Brand': ['Hero', 'Honda', 'Yamaha', 'Royal Enfield', 'Bajaj', 'Suzuki', 'TVS', 'KTM', 'BMW', 'Ducati'],

'Model': ['Splendor', 'Shine', 'FZ', 'Classic 350', 'Pulsar', 'Gixxer', 'Apache', 'Duke 390', 'G310R', 'Monster'],

'Engine\_CC': [97, 125, 150, 346, 220, 155, 160, 373, 313, 937],

'Mileage\_kmpl': [70, 60, 50, 35, 45, 50, 48, 30, 28, 20],

'Fuel\_Type': ['Petrol'] \* 10,

'Price\_in\_Lakhs': [0.8, 0.85, 1.05, 2.2, 1.2, 1.15, 1.1, 3.3, 3.0, 11.5]

}

df = pd.DataFrame(data)

print("Sample Dataset:\n", df)

# ------------------------------

# 2. Data Summary

# ------------------------------

print("\nBasic Information:")

print(df.info())

print("\nDescriptive Statistics:")

print(df.describe())

# ------------------------------

# 3. Visualization

# ------------------------------

plt.figure(figsize=(10,6))

plt.bar(df['Brand'], df['Price\_in\_Lakhs'], color='orange')

plt.title("Bike Price Comparison by Brand")

plt.xlabel("Brand")

plt.ylabel("Price (in Lakhs)")

plt.xticks(rotation=45)

plt.tight\_layout()

plt.savefig("bike\_price\_chart.png")

plt.show()

# ------------------------------

# 4. Correlation

# ------------------------------

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

plt.scatter(df['Engine\_CC'], df['Price\_in\_Lakhs'], color='blue')

plt.title("Engine CC vs Price")

plt.xlabel("Engine CC")

plt.ylabel("Price (in Lakhs)")

plt.grid(True)

plt.tight\_layout()

plt.savefig("engine\_vs\_price.png")

plt.show()

# ------------------------------

# 5. Create PowerPoint

# ------------------------------

prs = Presentation()

# Slide 1 - Title

slide1 = prs.slides.add\_slide(prs.slide\_layouts[0])

slide1.shapes.title.text = "Bike Price Data Analysis Project"

slide1.placeholders[1].text = "Exploring how engine power, mileage, and brand affect bike prices."

# Slide 2 - Dataset Table

slide2 = prs.slides.add\_slide(prs.slide\_layouts[5])

slide2.shapes.title.text = "Dataset Overview"

rows, cols = df.shape

table = slide2.shapes.add\_table(rows+1, cols, Inches(0.5), Inches(1.5), Inches(9), Inches(4)).table

for i, col\_name in enumerate(df.columns):

table.cell(0, i).text = col\_name

for r in range(rows):

for c in range(cols):

table.cell(r+1, c).text = str(df.iloc[r, c])

# Slide 3 - Bar Graph

slide3 = prs.slides.add\_slide(prs.slide\_layouts[5])

slide3.shapes.title.text = "Bike Price Comparison"

slide3.shapes.add\_picture("bike\_price\_chart.png", Inches(1), Inches(1.5), Inches(8), Inches(4.5))

# Slide 4 - Scatter Plot

slide4 = prs.slides.add\_slide(prs.slide\_layouts[5])

slide4.shapes.title.text = "Engine CC vs Price"

slide4.shapes.add\_picture("engine\_vs\_price.png", Inches(1), Inches(1.5), Inches(8), Inches(4.5))

# Slide 5 - Conclusion

slide5 = prs.slides.add\_slide(prs.slide\_layouts[1])

slide5.shapes.title.text = "Conclusion"

slide5.placeholders[1].text = (

"- Higher Engine CC bikes generally have higher prices.\n"

"- Premium brands (BMW, Ducati) are priced much higher.\n"

"- Mileage inversely affects price — higher mileage bikes are cheaper.\n"

"- Ideal balance for Indian consumers: Mileage + Affordable Engine."

)

prs.save("Bike\_Price\_Analysis.pptx")

print("\n✅ PowerPoint 'Bike\_Price\_Analysis.pptx' has been created successfully!")