**1. Load the Data**

First, load the CSV file into R:

# Load required libraries

library(ggplot2)

library(dplyr)

# Load the dataset

file\_path <- "C:/Users/umars/Desktop/team/Group---A147/combined\_flight\_data.csv"

flight\_data <- read.csv(file\_path)

# View the first few rows and column names

head(flight\_data)

colnames(flight\_data)

**2. Explore the Data**

Verify that FlightName and Price exist in the dataset:

# Check for missing values in FlightName and Price

sum(is.na(flight\_data$FlightName))

sum(is.na(flight\_data$Price))

# Check the structure of the columns

str(flight\_data)

**3. Generate Visualizations**

**a. Bar Chart**

Compare the average price for each flight name:

ggplot(flight\_data, aes(x = FlightName, y = Price)) +

geom\_bar(stat = "summary", fun = "mean", fill = "skyblue") +

labs(title = "Average Price by Flight Name", x = "Flight Name", y = "Average Price") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))

**b. Box Plot**

Show the distribution of prices for each flight name:

ggplot(flight\_data, aes(x = FlightName, y = Price)) +

geom\_boxplot(fill = "orange") +

labs(title = "Price Distribution by Flight Name", x = "Flight Name", y = "Price") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))

**c. Violin Plot**

Visualize the density distribution of prices for each flight name:

ggplot(flight\_data, aes(x = FlightName, y = Price)) + geom\_violin(fill = "purple", alpha = 0.7) + labs(title = "Price Density by Flight Name", x = "Flight Name", y = "Price") + theme\_minimal() + theme(axis.text.x = element\_text(angle = 45, hjust = 1))

**d. Scatter Plot**

If there’s an additional numeric variable (e.g., flight duration or distance), include it for a scatter plot:

ggplot(flight\_data, aes(x = FlightName, y = Price, color = FlightName)) +

geom\_point(size = 3, alpha = 0.6) +

labs(title = "Scatter Plot of Price by Flight Name", x = "Flight Name", y = "Price") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))

**e. Histogram**

Show the distribution of prices across all flights:

ggplot(flight\_data, aes(x = Price)) +

geom\_histogram(binwidth = 50, fill = "green", color = "black") +

labs(title = "Histogram of Prices", x = "Price", y = "Frequency") +

theme\_minimal()

**f. Density Plot**

Visualize the density of prices across all flights:

ggplot(flight\_data, aes(x = Price, fill = FlightName)) +

geom\_density(alpha = 0.5) +

labs(title = "Density Plot of Prices", x = "Price", y = "Density") +

theme\_minimal()

**4. Save Visualizations**

You can save any plot using ggsave:

ggsave("bar\_chart\_flight\_prices.png", width = 10, height = 6)

ggsave("box\_plot\_flight\_prices.png", width = 10, height = 6)

**5. Additional Notes**

* Replace FlightName and Price with the exact column names from the dataset (use colnames(flight\_data) to verify).
* Ensure Price is numeric; if not, convert it using:

flight\_data$Price <- as.numeric(as.character(flight\_data$Price))