**📄 Case Study Demo Notes: TSA Claims Data**

**🏁 1. Setup**

🔹 **Set Working Directory**

setwd("C:/Users/your\_path/ITSE\_2309\_Data\_Files")

🔹 **Load Library**

library(tidyverse)

🔹 **Import Data**

tsa\_claims <- read\_csv("tsa\_claims1.csv")

🔹 **View Data Briefly**

tsa\_claims

📝 About **94,848 rows** loaded.

**🧹 2. Data Cleaning**

🔹 **Remove Spaces in Column Names**

names(tsa\_claims) <- gsub(" ", "", names(tsa\_claims))

tsa\_claims

**📊 3. Data Analysis**

**a) ❓ What is the most common type of insurance claim?**

🔹 **Code**

tsa\_claims %>%

group\_by(ClaimType) %>%

summarize(ClaimCounts = n())

✅ **Answer:** **Passenger Property Loss**

**b) ❓ Which claim site within the airport has the most claims filed?**

🔹 **Code**

tsa\_claims %>%

group\_by(ClaimSite) %>%

summarize(SiteCounts = n())

✅ **Answer:** **Checked Baggage**

**c) ❓ What type of claim is made most at each claim site?**

🔹 **Code**

tsa\_claims %>%

group\_by(ClaimSite, ClaimType) %>%

summarize(Counts = n())

✅ **Answer:**

* **Checked Baggage ➡️ Passenger Property Loss**
* **Checkpoint ➡️ Property Damage**
* **Motor Vehicle ➡️ Passenger Property Loss**
* **Other ➡️ Passenger Property Loss**

**d) ❓ What is the median claim amount?**

🔹 **Code**

summary(select(tsa\_claims, ClaimAmount))

✅ **Answer:** **$184**

📌 *(Max claim observed was $125 million, but median is $184 — huge outlier!)*

**e) ❓ What is the overall claim approval rate for the U.S.?**

🔹 **Step 1: Group by Status**

status <- tsa\_claims %>%

group\_by(Status) %>%

summarize(Count = n())

🔹 **Step 2: Add Total Count Column**

status <- status %>%

mutate(TotalCount = sum(Count))

🔹 **Step 3: Calculate Percentages**

status %>%

mutate(ApprovalRate = Count / TotalCount \* 100)

🔎 **Result Table:**

| **Status** | **Count** | **Total Count** | **% Rate** |
| --- | --- | --- | --- |
| Approved | 23,167 | 94,848 | 24.4% |
| Canceled | 6 | 94,848 | 0.0063% |
| Denied | 53,273 | 94,848 | 56.2% |
| Pending response from claimant | 1 | 94,848 | 0.001% |
| Settled | 18,401 | 94,848 | 19.4% |

✅ **Final Answers:**

* **24.4%** of claims were **Approved**.
* **19.4%** were **Settled**.
* About **43.8%** (24.4% + 19.4%) had some form of **reimbursement**.

**✨ Additional Instructions**

🔹 **Insert a Comment at Top of Script**  
*(State your goal for taking the class)*

Example:

# Goal: I want to become proficient at using R for real-world database programming and data analysis.

**🏆 Final Checklist:**

✅ Working directory set  
✅ Library loaded  
✅ Data imported and cleaned  
✅ Exploratory analysis completed  
✅ Case study questions answered  
✅ Personal comment added