Anoushka Gurung

Communicate Application Work

```
format:
  html:
    toc: true
    toc-depth: 2
    code-fold: true
    embed-resources: true
  pdf: default
  docx: default
execute:
  echo: false
  warning: false
  message: false
knitr:
  opts_chunk:
    comment: "#>"
    collapse: true
```

1. Specify

We investigated the following six questions:

1. Are shorter delivery times actually more expensive?
2. How many shipments are completed by each carrier?
3. What is the average cost per shipment from each carrier?
4. What is the cost per KG-KM for all carriers?
5. What is the rank for each carrier by cost per KG-KM?
6. How does the longest diagonal affect cost?

---

2. Obtain

```{r}
#| label: setup
#| include: false
```

```
library(tidyverse)
library(readxl)
library(knitr)
library(ggplot2)

data <- read_excel("Whole Game Raw Data.xlsx")

# Rename for clarity
data <- data %>%
  rename(
    shipment_number = number,
    length_mm = L,
    width_mm = W,
    height_mm = H,
    weight_kg = `W.1`,
    carrier = carier,
    distance_km = `dist.`,
    cost_usd = `($)`,
    transit_days = days,
    longest_diagonal_mm = `longest diagonal`,
    volume_m3 = volume,
    density_kg_per_m3 = `KG/m^3`,
    kg_km = `KG-KM`
  ) %>%
  filter(!is.na(carrier), !is.na(cost_usd), !is.na(kg_km)) %>%
  mutate(
    cost_per_kg_km = cost_usd / kg_km,
    cost_per_shipment = cost_usd / weight_kg
  )

# I cleaned column names, converted data types, handled missing values,
# and created two metrics: cost per KG-KM and cost per shipment.
# The dataset is now ready for analysis.



# 3 Analyze

#| label: fig-cost-vs-days
#| fig-cap: "Average Cost by Delivery Days"
#| out-width: "70%"
#| fig-align: center

delivery_cost_analysis <- data %>%
```

```
  group_by(transit_days) %>%
  summarise(avg_cost = mean(cost_usd, na.rm = TRUE)) %>%
  arrange(transit_days)

ggplot(delivery_cost_analysis, aes(x = transit_days, y = avg_cost)) +
  geom_line(color = "blue") +
  geom_point(color = "red") +
  labs(x = "Delivery Days", y = "Average Cost (USD)") +
  theme_minimal()

# Shorter delivery times do not lead to higher costs.
# Average cost fluctuates slightly, but weight and distance remain the key cost drivers.


#4 Shipments per Carrier

#| label: fig-carrier-shipments
#| fig-cap: "Total Shipments per Carrier"
#| out-width: "70%"
#| fig-align: center

carrier_shipments <- data %>%
  group_by(carrier) %>%
  summarise(total_shipments = n())

ggplot(carrier_shipments, aes(x = reorder(carrier, total_shipments), y = total_shipments, fill =
carrier)) +
  geom_col() +
  coord_flip() +
  labs(x = "Carrier", y = "Total Shipments") +
  theme_minimal() +
  theme(legend.position = "none")

#5 Average Cost per Shipment

#| label: tbl-carrier-cost
#| tbl-cap: "Average Cost per Shipment by Carrier"

carrier_avg_cost <- data %>%
  group_by(carrier) %>%
  summarise(avg_cost = mean(cost_usd, na.rm = TRUE)) %>%
  arrange(desc(avg_cost))

kable(carrier_avg_cost, digits = 2)
```

```
#6 Cost per KG-KM

#| label: tbl-kg-km
#| tbl-cap: "Cost per KG-KM by Carrier"

carrier_kg_km_cost <- data %>%
  group_by(carrier) %>%
  summarise(avg_cost_per_kg_km = mean(cost_per_kg_km, na.rm = TRUE))

kable(carrier_kg_km_cost, digits = 4)

#7 Carrier Ranking by Cost per KG-KM

#| label: fig-carrier-rank
#| fig-cap: "Carrier Ranking by Cost Efficiency"
#| out-width: "70%"
#| fig-align: center

carrier_rank <- carrier_kg_km_cost %>%
  arrange(avg_cost_per_kg_km) %>%
  mutate(rank = row_number())

ggplot(carrier_rank, aes(x = reorder(carrier, avg_cost_per_kg_km), y = avg_cost_per_kg_km, fill
= carrier)) +
  geom_col() +
  coord_flip() +
  labs(x = "Carrier", y = "Cost per KG-KM") +
  theme_minimal() +
  theme(legend.position = "none")


#8 Effect of Longest Diagonal on Cost

#| label: fig-diagonal-cost
#| fig-cap: "Does Package Size Affect Cost?"
#| out-width: "70%"
#| fig-align: center

ggplot(data, aes(x = longest_diagonal_mm, y = cost_usd)) +
  geom_point(alpha = 0.5, color = "steelblue") +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(x = "Longest Diagonal (mm)", y = "Cost (USD)") +
```

```
  theme_minimal()
```

REPORT

This analysis answers the six business questions through a combination of statistical summaries and visualizations. Highlights include:

- Delivery time is not a cost driver.
- All carriers charge the same rate per KG-KM, but vary in average cost per shipment.
- Package size has minimal impact on cost.