Data Wrangling

Blake Pappas

September 12, 2022

Load Necessary Packages

```r
library(tidyverse)
```

Import and Tidy Data

```r
# Import AcctAddresses.txt
AcctAddresses <- read_tsv("AcctAddress.txt",
                          col_names = c("account_number", "shipping_address"))
```

```r
AcctAddresses <- AcctAddresses %>%
# Filter for Only Those Rows That End with a Number (Zip Codes)
  filter(str_detect(shipping_address, "[0123456789]$")) %>%
# Pull Only the Last 3 Digits of the Zip Code
  mutate(ZipCode = str_sub(shipping_address, -5, -3))
```

```r
# Import USPS_Zone.csv
USPS_Zone <- read_csv("USPS_Zone.csv")
```

```r
# Create Tibble for Cost per Zone
Zone_Costs <- tribble(
                      ~Zone, ~cost,
                      1, 8.50,
                      2, 8.50,
                      3, 9.90,
                      4, 11.15,
                      5, 14.15,
                      6, 18.70,
                      7, 20.65,
                      8, 23.15,
                      9, 30.25
)
```

Join Data

```
# Join All 3 Tibbles
AddressZoneCost <- AcctAddresses %>%
  select(account_number, ZipCode) %>%
  left_join(USPS_Zone, by = "ZipCode") %>%
  select(account_number, ZipCode, Zone) %>%
  left_join(Zone_Costs, by = "Zone")
```

Strategy 1

Code:

```
# Total Number of Customers Multiplied by the Flat Rate Box Cost
(Question_1 <- nrow(AddressZoneCost) * 13.65)
```

```
## [1] 13650
```

Answer: Shipping all 1,000 gifts in a priority mail flat rate box that cost $13.65 each which would cost $13,650.00.

Strategy 2

Code:

```
# Sum All Values in the `cost` Column
(Question_2 <- sum(AddressZoneCost$cost, na.rm = TRUE))
```

```
## [1] 12343.05
```

Answer: Shipping all 1,000 gifts by priority mail standard box rated by shipping zone would cost $12,343.05.

Strategy 3

Code:

```
# Sum All Values in the `cost` Column for Zones 1-4
zone1234_cost <- AddressZoneCost %>%
  filter(Zone %in% c(1, 2, 3, 4)) %>%
  summarize(
    zone1234_cost = sum(cost, na.rm = TRUE)
  )

# Total Number of Customers in Zones 5-9 Multiplied by the Flat Rate Box Cost
cost2 <- AddressZoneCost %>%
  filter(!Zone %in% c(1, 2, 3, 4))
zone56789_cost <- nrow(cost2) * 13.65

# Sum of Total Standard Rates for Zones 1-4 and Total Flat Rates for Zones 5-9
(Question_3 <- zone1234_cost + zone56789_cost)
```

```
##   zone1234_cost
## 1     11604.8
```

Answer: Shipping all gifts being mailed through Zones 1,2,3, & 4 (standard rates which are less than flat rate) via standard boxes and all gifts being mailed to Zones 5, 6, 7, 8, and 9 via flat rate boxes would cost $11,604.80.