Challenge Summary

This is a short challenge to begin applying what you are learning to the problem at hand. You will go through a series of questions related to the course project goals:

1. Coming up with a new product idea, and
2. Segmenting the customer-base

Objectives

1. Apply `lubridate` and `stringr` functions to answer questions related to the course projects.
2. Gain exposure to `rmarkdown`.

Data

To read the data, make sure that the paths point to the appropriate data sets. Saving the file in the "challenges folder" should enable the paths to be detected correctly.

```
# Load libraries
library(tidyverse)
library(lubridate)
```

```
# Read bike orderlines data
path_bike_orderlines <- "00_data/bike_sales/data_wrangled/bike_orderlines.rds"
bike_orderlines_tbl <- read_rds(path_bike_orderlines) %>%

    # Fix typos found in Feature Engineering
    mutate(model = case_when(
        model == "CAAD Disc Ultegra" ~ "CAAD12 Disc Ultegra",
        model == "Syapse Carbon Tiagra" ~ "Synapse Carbon Tiagra",
        model == "Supersix Evo Hi-Mod Utegra" ~ "Supersix Evo Hi-Mod Ultegra",
        TRUE ~ model
    ))

glimpse(bike_orderlines_tbl)
```

```
## Observations: 15,644
## Variables: 13
## $ order_date     <dttm> 2011-01-07, 2011-01-07, 2011-01-10, 2011-01-10...
## $ order_id       <dbl> 1, 1, 2, 2, 3, 3, 3, 3, 3, 4, 5, 5, 5, 5, 6, 6,...
## $ order_line     <dbl> 1, 2, 1, 2, 1, 2, 3, 4, 5, 1, 1, 2, 3, 4, 1, 2,...
## $ quantity       <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1,...
## $ price          <dbl> 6070, 5970, 2770, 5970, 10660, 3200, 12790, 533...
## $ total_price    <dbl> 6070, 5970, 2770, 5970, 10660, 3200, 12790, 533...
## $ model          <chr> "Jekyll Carbon 2", "Trigger Carbon 2", "Beast o...
## $ category_1     <chr> "Mountain", "Mountain", "Mountain", "Mountain",...
## $ category_2     <chr> "Over Mountain", "Over Mountain", "Trail", "Ove...
## $ frame_material <chr> "Carbon", "Carbon", "Aluminum", "Carbon", "Carb...
## $ bikeshop_name  <chr> "Ithaca Mountain Climbers", "Ithaca Mountain Cl...
## $ city           <chr> "Ithaca", "Ithaca", "Kansas City", "Kansas City...
## $ state          <chr> "NY", "NY", "KS", "KS", "KY", "KY", "KY", "KY",...
```

Questions

lubridate: Which month has the highest bike sales? (Difficulty = Medium)

- Start with `bike_orderlines_tbl`
- Select columns `order_date` and `total_price`
- Add a column called `month`
- Group by, summarize, and ungroup calculating the `sales`
- Arrange the sales values by month (Jan - Dec)
- Format the sales values as `dollar()`
- Adjust column names to title case

What does this tell us about a time of year to focus marketing efforts?

```
bike_orderlines_tbl %>%

    # Select columns
    select(order_date, total_price) %>%

    # Add month column
    mutate(month = month(order_date, label = TRUE, abbr = FALSE)) %>%

    # Group by, summarize, ungroup
    group_by(month) %>%
    summarize(sales = sum(total_price)) %>%
    ungroup() %>%

    # Arrange sales by month
    arrange(month) %>%

    # Format dollar
    mutate(sales = sales %>% scales::dollar()) %>%

    # Format column names
    rename(
        Month = month,
        Sales = sales
    )
```

```
## # A tibble: 12 x 2
##    Month     Sales
##    <ord>     <chr>
##  1 January   $4,089,460
##  2 February  $5,343,295
##  3 March     $7,282,280
##  4 April     $8,386,170
##  5 May       $7,935,055
##  6 June      $7,813,105
##  7 July      $7,602,005
##  8 August    $5,346,125
##  9 September $5,556,055
## 10 October   $4,394,300
## 11 November  $4,169,755
```

```
## 12 December  $3,114,725
```

```
    # Alternate Methods:
    # set_names(names(.) %>% str_to_title())
    # rename_all(~ str_to_title(.))
```

stringr: What is the median orderline sales value by Bike Attribute? (Difficulty = Medium)

- Begin with `bike_orderlines`
- Select `model` and `total_price`
- Detect if string is present (e.g. "black inc")
- Groupby, summarize, and ungroup calculating the `median()` orderline
- Format numeric price as `dollar()` (Hint: investigate `largest_with_cents` argument)
- Rename column to evaluation string (e.g. "Black Inc")

Evaluate "Black Inc". *What does this tell us about the "Black Inc" feature?*

```
bike_orderlines_tbl %>%

    # Select columns
    select(model, total_price) %>%

    # Detect string present
    mutate(option_detected = model %>% str_to_lower() %>% str_detect("black inc")) %>%

    # Group by, summarize, ungroup
    group_by(option_detected) %>%
    summarise(
        median_price = median(total_price)
        ) %>%
    ungroup() %>%

    # Format dollar
    mutate(median_price = median_price %>% scales::dollar(largest_with_cents = 10)) %>%

    # Rename columns
    rename(
        `Black Inc` = option_detected,
        `Median Orderline` = median_price
    )
```

```
## # A tibble: 2 x 2
##   `Black Inc` `Median Orderline`
##   <lgl>       <chr>
## 1 FALSE       $2,880
## 2 TRUE        $12,250
```

Evaluate "Ultegra". *What does this tell us about the "Ultegra" feature?*

```
bike_orderlines_tbl %>%

    # Select columns
    select(model, total_price) %>%
```

```
    # Detect string present
    mutate(option_detected = model %>% str_to_lower() %>% str_detect("ultegra")) %>%

    # Group by, summarize, ungroup
    group_by(option_detected) %>%
    summarise(
        median_price = median(total_price)
        ) %>%
    ungroup() %>%

    # Format dollar
    mutate(median_price = median_price %>% scales::dollar(largest_with_cents = 10)) %>%

    # Rename columns
    rename(
        Ultegra = option_detected,
        `Median Orderline` = median_price
    )
```

```
## # A tibble: 2 x 2
##   Ultegra `Median Orderline`
##   <lgl>   <chr>
## 1 FALSE   $3,200
## 2 TRUE    $3,200
```

Evaluate "Disc" option. *What does this tell us about the "Disc" feature?*

```
bike_orderlines_tbl %>%

    # Select columns
    select(model, total_price) %>%

    # Detect string present
    mutate(option_detected = model %>% str_to_lower() %>% str_detect("disc")) %>%

    # Group by, summarize, ungroup
    group_by(option_detected) %>%
    summarise(
        median_price = median(total_price)
        ) %>%
    ungroup() %>%

    # Format dollar
    mutate(median_price = median_price %>% scales::dollar(largest_with_cents = 10)) %>%

    # Rename columns
    rename(
        Disc = option_detected,
        `Median Orderline` = median_price
    )
```

```
## # A tibble: 2 x 2
##   Disc  `Median Orderline`
##   <lgl> <chr>
## 1 FALSE $3,200
```

```
## 2 TRUE  $2,660
```

stringr: What are the average, min, and max prices by Base Model? (Difficulty = High)

- Start with `bike_orderlines_tbl`
- Select distinct primary category, secondary category, model, and price (unit price, not total price)
- Create the base feature, `model_base` (Hint: Use the Feature Engineering code)
  - separate the models
  - Create a base feature that combines the appopriate parts (e.g. "Beast of the East")
- Remove any unnecessary columns (Hint: Deselect any columns matching `"model_[0-9]"`)
- Group by, summarize, and ungroup (Hint: use `mean()`, `min()`, and `max()`)
- Arrange descending by average price
- Format any numeric columns as `dollar()` (Hint: Check out `largest_with_cents`)
- Adjust the column names to title case

What does this tell us about how bikes are priced?

```
bike_orderlines_tbl %>%

    # Select distinct category_1, category_2, model, and price
    distinct(category_1, category_2, model, price) %>%

    # Separate models
    separate(col     = model,
             into    = str_c("model_", 1:7),
             sep     = " ",
             remove  = FALSE,
             fill    = "right") %>%

    # creating a "base" feature
    mutate(model_base = case_when(

        # Fix Supersix Evo
        str_detect(str_to_lower(model_1), "supersix") ~ str_c(model_1, model_2, sep = " "),

        # Fix Fat CAAD bikes
        str_detect(str_to_lower(model_1), "fat") ~ str_c(model_1, model_2, sep = " "),

        # Fix Beast of the East
        str_detect(str_to_lower(model_1), "beast") ~ str_c(model_1, model_2, model_3, model_4, sep = "

        # Fix Bad Habit
        str_detect(str_to_lower(model_1), "bad") ~ str_c(model_1, model_2, sep = " "),

        # Fix Scalpel 29
        str_detect(str_to_lower(model_2), "29") ~ str_c(model_1, model_2, sep = " "),

        # catch all
        TRUE ~ model_1)
        ) %>%

    # Remove unnecessary columns
```

```
    select(-matches("model_[0-9]")) %>%

    # Group by, summarize, ungroup
    group_by(category_1, category_2, model_base) %>%
    summarize(
        mean_price = mean(price),
        min_price  = min(price),
        max_price  = max(price)
        ) %>%
    ungroup() %>%

    # Arrange descending by mean price
    arrange(desc(mean_price)) %>%

    # Format dollar
    # Alternate method
    # mutate_if(is.numeric, ~ scales::dollar(., largest_with_cents = 10)) %>%
    mutate(
        mean_price = mean_price %>% scales::dollar(largest_with_cents = 10),
        min_price  = min_price  %>% scales::dollar(largest_with_cents = 10),
        max_price  = max_price  %>% scales::dollar(largest_with_cents = 10)
    ) %>%


    # Adjust Column Names
    # Alternate method
    # rename_all(~ str_replace(., "_", " ") %>% str_to_title())
    set_names(names(.) %>% str_replace("_", " ") %>% str_to_title())
```

```
## # A tibble: 18 x 6
##    `Category 1` `Category 2` `Model Base` `Mean Price` `Min Price`
##    <chr>        <chr>        <chr>        <chr>        <chr>
##  1 Mountain     Cross Count~ Scalpel-Si   $6,927       $3,200
##  2 Road         Elite Road   Supersix Evo $5,491       $1,840
##  3 Mountain     Over Mounta~ Jekyll       $5,275       $3,200
##  4 Mountain     Over Mounta~ Trigger      $5,275       $3,200
##  5 Mountain     Cross Count~ F-Si         $5,070       $1,840
##  6 Mountain     Trail        Habit        $5,052       $1,950
##  7 Mountain     Cross Count~ Scalpel 29   $4,795       $3,200
##  8 Road         Triathalon   Slice        $3,870       $1,950
##  9 Mountain     Fat Bike     Fat CAAD1    $3,730       $3,730
## 10 Road         Endurance R~ Synapse      $3,514       $870
## 11 Road         Elite Road   CAAD12       $3,121       $1,680
## 12 Mountain     Trail        Bad Habit    $2,930       $2,660
## 13 Road         Cyclocross   SuperX       $2,415       $1,750
## 14 Mountain     Trail        Beast of th~ $2,173       $1,620
## 15 Mountain     Fat Bike     Fat CAAD2    $2,130       $2,130
## 16 Mountain     Sport        Trail        $1,149       $815
## 17 Road         Elite Road   CAAD8        $1,126       $815
## 18 Mountain     Sport        Catalyst     $546         $415
## # ... with 1 more variable: `Max Price` <chr>
```