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 `dplyr` and `tidyr` 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 main directory should enable the paths to be detected correctly.

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

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

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",...
```

```
# Read bikes data
path_bikes <- "00_data/bike_sales//data_raw/bikes.xlsx"
bikes_tbl <- readxl::read_excel(path_bikes)

glimpse(bikes_tbl)
```

```
## Observations: 97
## Variables: 4
## $ bike.id     <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,...
```

```
## $ model       <chr> "Supersix Evo Black Inc.", "Supersix Evo Hi-Mod Te...
## $ description <chr> "Road - Elite Road - Carbon", "Road - Elite Road -...
## $ price       <dbl> 12790, 10660, 7990, 5330, 4260, 3940, 3200, 2660, ...
```

Questions

1. What are the unique categories of products? (Difficulty = Low)

- Begin with `bike_orderlines_tbl`
- Use `distinct()` to evaluate

Review Primary Product Category (`category_1`).

```
bike_orderlines_tbl %>%
    distinct(category_1)
```

```
## # A tibble: 2 x 1
##   category_1
##   <chr>
## 1 Mountain
## 2 Road
```

Review Secondary Product Category (`category_2`).

```
bike_orderlines_tbl %>%
    distinct(category_2)
```

```
## # A tibble: 9 x 1
##   category_2
##   <chr>
## 1 Over Mountain
## 2 Trail
## 3 Elite Road
## 4 Endurance Road
## 5 Sport
## 6 Cross Country Race
## 7 Cyclocross
## 8 Triathalon
## 9 Fat Bike
```

Review Frame Material (`frame_material`).

```
bike_orderlines_tbl %>%
    distinct(frame_material)
```

```
## # A tibble: 2 x 1
##   frame_material
##   <chr>
## 1 Carbon
## 2 Aluminum
```

2. Which product categories have the most sales? (Difficulty = Medium)

- Select appropriate columns from `bike_orderlines_tbl`
- Group and summarize the data calling the new column `Sales`. Make sure to ungroup.

- Arrange descending by `Sales`
- Rename column names to `Primary Category`, `Secondary Category`, or `Frame Material` (as appropriate).
- Format the Sales as `dollar()`

Review Primary Product Category (`category_1`).

```
bike_orderlines_tbl %>%

    # Select columns
    select(category_1, total_price) %>%

    # Group and summarize
    group_by(category_1) %>%
    summarize(sales = sum(total_price))  %>%
    ungroup() %>%

    # Arrange descending
    arrange(desc(sales)) %>%

    # Rename columns
    rename(
        `Primary Category` = category_1,
        Sales = sales
    ) %>%

    # Format dollar
    mutate(Sales = Sales %>% scales::dollar())
```

```
## # A tibble: 2 x 2
##   `Primary Category` Sales
##   <chr>              <chr>
## 1 Mountain           $39,154,735
## 2 Road               $31,877,595
```

Review Secondary Product Category (`category_2`).

```
bike_orderlines_tbl %>%

    # Select columns
    select(category_2, total_price) %>%

    # Group and summarize
    group_by(category_2) %>%
    summarize(sales = sum(total_price))  %>%
    ungroup() %>%

    # Arrange descending
    arrange(desc(sales)) %>%

    # Rename columns
    rename(
        `Secondary Category` = category_2,
        Sales = sales
    ) %>%
```

```
    # Format dollar
    mutate(Sales = Sales %>% scales::dollar())
```

```
## # A tibble: 9 x 2
##   `Secondary Category` Sales
##   <chr>                <chr>
## 1 Cross Country Race   $19,224,630
## 2 Elite Road           $15,334,665
## 3 Endurance Road       $10,381,060
## 4 Trail                $9,373,460
## 5 Over Mountain        $7,571,270
## 6 Triathalon           $4,053,750
## 7 Cyclocross           $2,108,120
## 8 Sport                $1,932,755
## 9 Fat Bike             $1,052,620
```

Review Frame Material (`frame_material`).

```
bike_orderlines_tbl %>%

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

    # Group and summarize
    group_by(frame_material) %>%
    summarize(sales = sum(total_price))  %>%
    ungroup() %>%

    # Arrange descending
    arrange(desc(sales)) %>%

    # Rename columns
    # Rename columns
    rename(
        `Frame Material` = frame_material,
        Sales = sales
    ) %>%

    # Format dollar
    mutate(Sales = Sales %>% scales::dollar())
```

```
## # A tibble: 2 x 2
##   `Frame Material` Sales
##   <chr>            <chr>
## 1 Carbon           $52,940,540
## 2 Aluminum         $18,091,790
```

3. Do all combinations primary and secondary bike category contain both Aluminum and Carbon frame materials? (Difficulty = High)

Hint - Use summarized sales values and `spread()` to identify gaps in frame materials.

- Select `category_1`, `category_2`, `frame_material`, and `total_price`
- Summarize the data using group by, summarize and ungroup.
- Pivot the frame material and sales column into Alumninum and Carbon

- Fill `NA` values with zeros
- Add a `total_sales` column
- Arrange descending by `total_sales`
- Format all numbers as `dollar()`
- Rename all Columns: Primary Category, Secondary Category, Aluminum, Carbon, Total Sales

```
bike_orderlines_tbl %>%

    # Select columns
    select(category_1, category_2, frame_material, total_price) %>%

    # group_by, summarize, ungroup
    group_by(category_1, category_2, frame_material) %>%
    summarize(sales = sum(total_price)) %>%
    ungroup() %>%

    # spread
    spread(key = frame_material, value = sales) %>%

    # replace NA
    replace_na(list(Aluminum = 0, Carbon = 0)) %>%

    # Add Total Sales column
    mutate(total_sales = Aluminum + Carbon) %>%

    # Arrange descending
    arrange(desc(total_sales)) %>%

    # Format dollar
    mutate(
        Aluminum = scales::dollar(Aluminum),
        Carbon   = scales::dollar(Carbon),
        total_sales = scales::dollar(total_sales)
    ) %>%

    # Rename columns
    rename(
        `Primary Category`   = category_1,
        `Secondary Category` = category_2,
        `Total Sales`        = total_sales
    )
```

```
## # A tibble: 9 x 5
##   `Primary Category` `Secondary Category` Aluminum  Carbon    `Total Sales`
##   <chr>              <chr>                <chr>     <chr>     <chr>
## 1 Mountain           Cross Country Race   $3,318,5~ $15,906,~ $19,224,630
## 2 Road               Elite Road           $5,637,7~ $9,696,8~ $15,334,665
## 3 Road               Endurance Road       $1,612,4~ $8,768,6~ $10,381,060
## 4 Mountain           Trail                $4,537,6~ $4,835,8~ $9,373,460
## 5 Mountain           Over Mountain        $0        $7,571,2~ $7,571,270
## 6 Road               Triathalon           $0        $4,053,7~ $4,053,750
## 7 Road               Cyclocross           $0        $2,108,1~ $2,108,120
## 8 Mountain           Sport                $1,932,7~ $0        $1,932,755
## 9 Mountain           Fat Bike             $1,052,6~ $0        $1,052,620
```