**Case Study – Cyclistic**

* ASK
  + Business Task – “How annual members and causal riders differ, and how do we covert causal riders to members?”
* PREPARE
  + Data Source Description
    - I pulled the last 12 months of rider data of the Cysclistic’s historical trip data.
      * July 2020 – June 2021
    - After downloading the data, I unzipped each file and placed into a CSV folder.
      * My original unedited files
    - I opened each file and saved as an .xls document into a different EXCEL folder.
      * My cleaned and sorted data files.
    - It appears some of the data is missing and I have some NULL values. I will delete and lean all this data up to ensure all accuracy.
* PROCESS
  + Excel
    - Subtract start and end date and times to find total duration of the ride.
      * Casual riders on average have longer rides than members
      * Number of rides starts to decrease in the fall months starting with October.
    - Use WEEKDAY function to identify the day of week rentals were made.
      * Create new tab with a day of the week table.
        + Used VLOOKUP to convert number days into the actual days of the week. (i.e. 1 is converted to Sunday)
    - Deleted NULL values and any blank location-based data.
      * Used filter function to sort locations by BLANK and then removed this data.
      * Created a PIVOT TABLE to find average ride time and number of rides per day.
        + Sorted this data into descending order to find what days are most popular for each type of rider.
        + Then created a bar chart to show visuals of the most popular days of the week for each member type.
        + Snipped the bar chart to place into a Power Point for a month by month analysis.
        + Most popular days for each rider:

Casual – weekends

Member – weekdays

* + - * + Causal riders drop off drastically in the winter months.
      * REPEAT STEPS FOR REMAINING 11 MONTHS, SAVE A SECOND COPY AS A CSV FILE.
  + R
* # Cyclistic Case Study
* # Install the necessary R packages
* install.packages("tidyverse")
* library(tidyverse)
* install.packages("ggplot2")
* library(ggplot2)
* install.packages("janitor")
* library(janitor)
* install.packages("lubridate")
* library(lubridate)
* install.packages("scales")
* library(scales)
* # Rename data sets and label accordingly.
* br2007 <- read.csv("Cyclistic Bike Ride Data/202007\_Rides.csv")
* br2008 <- read.csv("Cyclistic Bike Ride Data/202008\_Rides.csv")
* br2106 <- read.csv("Cyclistic Bike Ride Data/202106\_Rides.csv")
* br2009 <- read.csv("Cyclistic Bike Ride Data/202009\_Rides.csv")
* br2010 <- read.csv("Cyclistic Bike Ride Data/202010\_Rides.csv")
* br2011 <- read.csv("Cyclistic Bike Ride Data/202011\_Rides.csv")
* br2012 <- read.csv("Cyclistic Bike Ride Data/202012\_Rides.csv")
* br2101 <- read.csv("Cyclistic Bike Ride Data/202101\_Rides.csv")
* br2102 <- read.csv("Cyclistic Bike Ride Data/202102\_Rides.csv")
* br2103 <- read.csv("Cyclistic Bike Ride Data/202103\_Rides.csv")
* br2104 <- read.csv("Cyclistic Bike Ride Data/202104\_Rides.csv")
* br2105 <- read.csv("Cyclistic Bike Ride Data/202105\_Rides.csv")
* # Combine data by season
* summer\_rides <- rbind(br2007,br2008,br2106)
* autumn\_rides <- rbind(br2009,br2010,br2011)
* winter\_rides <- rbind(br2012,br2101,br2102)
* spring\_rides <- rbind(br2103,br2104,br2105)
* # Export File to Computer
* write\_csv(summer\_rides,"summer\_rides.csv")
* write\_csv(autumn\_rides,"autumn\_rides.csv")
* write\_csv(winter\_rides,"winter\_rides.csv")
* write\_csv(spring\_rides,"spring\_rides.csv")
  + Excel Again
    - For each season:
      * Created two pivot tables
        + 1. To visualize the number to bike rides per day, and per bike type.
        + 2. To visualize the average ride time per day, and per bike type