Food Delivery Company Menu Consolidation Case Study Script

This case study is being done by using a data set from Kaggle that was originally published by Edwin U Kannanaikkal.
(https://www.kaggle.com/datasets/kannanaikkal/food-demand-forecasting)
This case study is evaluating the popularity of cuisines and meals purchased from a food delivery company. The purpose of this script is to get the data into a single data frame and be able to provide the brand team with information so that they can determine what items they should remove from their menu and what cuisines to select for the brand identity.

```
# loading all necessary packages to clean and evaluate data
> install.packages("here")
> library(here)
> install.packages("skimr")
> library("skimr")
> install.packages("janitor")
> library(janitor)
> install.packages("dplyr")
> library(dplyr)
> install.packages('tidyverse')
> library(tidyverse)
> library(readr)
> install.packages("ggplot2")
> library(ggplot)

# loading data into R
> train <- read_csv("train.csv")
> View(train)
> meal_info <- read_csv("meal_info.csv")
> View(meal_info)
> fulfilment_center_info <- read_csv("fulfilment_center_info.csv")
> View(fulfilment_center_info)

# starting to clean the data and arrange it to make it more clear
> summary(train)
> head(train)
> summary(meal_info)
> meal_info %>% arrange(cuisine)
> meal_info2 <- meal_info %>% arrange(cuisine)
> View(meal_info2)

# removing unnecessary columns from datasets
> train2 <- train[, -c(1,3,5,6,7,8)]
> View(train2)

# checking the data for errors
# check if data frame is NULL
> is.null(train2)
```

```
[1] FALSE
> is.null(meal_info2)
[1] FALSE

# find the mean number of orders for each meal id
> data <- read.csv("train.csv")
> data <- data[c('week', 'meal_id', 'num_orders')]
> mean_orders <- aggregate(num_orders ~ meal_id, data, mean)
> mean_orders

# merge the mealinfo2 data and mean_orders data so we have one data set to
work with
> complete_meal_info <- merge(x = mean_orders, y = mealinfo2, by =
"meal_id")
> View(complete_meal_info)

# inspect the new data table
> colnames(complete_meal_info)
> nrow(complete_meal_info)
> dim(complete_meal_info)
> head(complete_meal_info)
> str(complete_meal_info)
> summary(complete_meal_info)

# Need to fix column name to average number of orders
> colnames(complete_meal_info_2)[2] ='avg_num_orders'


# conduct descriptive analysis
> max(complete_meal_info$avg_num_orders) #meal with highest number of
orders
> min(complete_meal_info$avg_num_orders) #meal with lowest number of
orders
> arrange(complete_meal_info, avg_num_orders, .by_group = FALSE) #creating
a table that shows the data in ascending order to see the max and min more
easily

# make new table for complete meal info that is sorted by least number of
order to most number of orders
> complete_meal_info_2 <- complete_meal_info %>% arrange(avg_num_orders)

# create visualization of the data
> ggplot(data = complete_meal_info_2, aes(x = meal_id, y =
avg_num_orders))
+    geom_bar(stat = 'identity')

> ggboxplot(complete_meal_info_2, x = "cuisine", y = "avg_num_orders",
color = "meal_id", ylab = "Avg Num Orders", xlab = "Cuisine Type")
```

```
#export cleaned data sets and new data set
> write.csv(mealinfo2, file = "meal_info_clean.csv")
> write.csv(train2, file = "train_cleaned.csv")
> write.csv(mean_orders, file = "mean_orders.csv")
> write.csv(complete_meal_info_2, file = "complete_meal_info.csv")
```