

```

---
title: "Untitled"
output: pdf_document
date: "2024-07-06"
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
install.packages("tidyverse")
library(tidyverse)
library(tidyr)
library(readr)
library(dplyr)
library(tibble)
library(purrr)
install.packages("here")
library(here)
install.packages("skimr")
library(skimr)
install.packages("janitor")
library(janitor)
```

```

## ## Data Science Assignment

### \*Question 1

1. Implement Programming Techniques in Processing the Dataset.

- Use the OrderData.csv provided with this assignment to do the following:

i. Store a column in a vector

```

```{r cars}
#Reading our dataset into R enviroment
Order_data <- read_csv("OrderData_converted.csv")

#Cleaning our dataset
Tidying_orderdata <- clean_names(Order_data)

#Vector
OrderData_vector <-c(Tidying_orderdata$units)
print(OrderData_vector)
```

```

## ## Steps

I started with reading my dataset into Rstudio enviroment. I proceeded to clean my column header using the clean\_names function. Then I stored the units column in a vector named OrderData\_vector using the subsetting function(\$).

ii. Create a list and assign multiple columns to it.

```

```{r pressure, echo=FALSE}
#List
OrderData_list <- list(Tidying_orderdata$order_date, Tidying_orderdata$region,
Tidying_orderdata$rep, Tidying_orderdata$item)
print(OrderData_list)
```

```

To create a list I used the list function. Then I used the subsetting function to extract four columns from the dataset

iii. Create a Data Frame consisting first 4 columns and 30 rows from the entire dataset.

```

```{r}
#Data frame
OrderData_dataframe <- Tidying_orderdata[1:30, 1:4]
print(OrderData_dataframe, n=30)
```

```

Determining the attributes of the data structures:

The str() function was used to determine the type of elements the data structure contains. While the length function was used to determine the length of the elements in the data structure.

i. Vectors: The vector contained numbers with a length of 43.

```

```{r}
#Determining the vector's attributes
str(OrderData_vector)
length(OrderData_vector)
```

```

ii. List: The list contained 4 columns

```

```{r}
str(OrderData_list)
length(OrderData_list)
```

```

iii. Data frame: The data frame used is a tibble containing 4 columns and a length of 30 rows.

```

```{r}
str(OrderData_dataframe)
length(OrderData_dataframe)
```

```

The difference between the data structures (Vector, List, data frame):

A vector is a group of data elements that must be of the same data type. A list is a group of data elements that can consist of different data types while a data frame is collection of columnnd similar to a spreadsheet table.