

# Order Forecast in Oasis Store

---

D 4/11/2022

**NOTE:** This project is part of VMware Campus Ambassador Program Virtual Hackathon

## A. Overview

---

Oasis Store is an internal e-commerce site in VMware to order IT Assets. VMware colleagues place order using the store of various IT assets like laptops, monitors etc.

The complete analysis process is divided into several parts

## B. Background

---

As a part of Hackathon, following things are evaluated

1. Ordering pattern basis various KPI's
2. Ordering forecast basis various KPI's

## 1. Data Prepration

---

### Data Source

Data is first party data as it is been generate within the VMware and hence the data is reliable.

## 2. Data Processing

---

Data was checked for the possible errors and processed. Such processing was done on **MS Excel**.

### 2.1 Missing Values

---

Data source was looked for the missing values and were fixed.

- Region column had blank spaces. Those were fixed on the basis of shipping country.
- Data set was searched for the blank row and was removed from analysis.

## Data format

---

For the analysis purpose, the date time has changed to date only.

## Data Transformation

---

- Few item had zero cost. They were changed to the cost respective to the product.
- Data was analysed and found that various product could be covered under one umbrella. Like various laptop can be categorized in Laptop category. Similarly 14 different category has been created. They are **Custom.Accessory, Display.Adapter, Headset, Keyboard, Keyboard.and.Mouse, Laptop, Laptop.Stand, Monitor, Mouse, Multiport.Adapter, Power.Adapter, Speakerphone, Trackpad and Webcam.**
- Using Pivot table, all the same type of order in a single was clubbed together.

## 3. Data Analysis

---

Ordering pattern of the individual product is visualized through graph drawn **MS Excel**. Few of them are shown here

**Conclude:** There is a seasonal relation between the

## Model Formation

---

Linear model is used to predict the future order. **R Programming** language is used for model building and prediction.

Since the data is available for only Sept 2021 to March 2022 so entire data set is used for training and for testing the data of March 2022 is used.

## Import the data

---

```
oasis = read.csv("C:\\Users\\preet\\Desktop\\Project\\NITP Hackathon\\Hackathon Data Oa
# Data manipulation
library(lubridate)
```



```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##    date, intersect, setdiff, union

library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##    filter, lag

## The following objects are masked from 'package:base':
##
##    intersect, setdiff, setequal, union

library(ggplot2)
library(tidyr)
oasis$Date = as.POSIXct(oasis$Date , format = c("%d-%m-%y"))
oasis$month = as.factor(month(oasis$Date)) # Extracting month
oasis$weekday = as.factor(wday(oasis$Date)) # Extracting weekday
oasis$week_num = as.factor(week(oasis$Date)) # extracting week_num
```

## The future order of individual product is predicted.

---

### Custom Accessory

```
# Analysis of the item Custom Accessory
oasis_custom = select(oasis, Date, Custom.Accessory, week_num, weekday, month)

# Drop null values
oasis_custom = drop_na(oasis_custom)

# Testing on entire data set
# Testing model on month of March
oasis_custom_test = subset(oasis_custom, Date > "2022-03-01")

model_custom = lm(Custom.Accessory~month+weekday, data = oasis_custom) # 7.174798
# summary(model_custom)
predict_custom = predict(model_custom, newdata = oasis_custom_test)
```

```
# Warning is given which is due to the reason that testing and training done on same da

result_custom = data.frame(date =oasis_custom_test$Date ,actual = oasis_custom_test$Cus
rsme_custom = sqrt(mean(result_custom$err^2))
```

## Display.Adapter

```
# Analysis of the item Display.Adapter
oasis_disp = select(oasis, Date, Display.Adapter, week_num,weekday,month)

# Drop null values
oasis_disp = drop_na(oasis_disp)

# Testing on entire data set
# Testing model on month of March
oasis_disp_test = subset(oasis_disp, Date > "2022-03-01")

model_disp = lm(Display.Adapter~month+weekday, data = oasis_disp) # 7.174798
# summary(model_disp)
predict_disp = predict(model_disp, newdata = oasis_disp_test)
# Warning is given which is due to the reason that testing and training done on same da

result_disp = data.frame(date =oasis_disp_test$Date, actual = oasis_disp_test$Display.A
rsme_disp = sqrt(mean(result_disp$err^2))
```

## Headset

```
# Analysis of the item Headset
oasis_headset = select(oasis, Date, Headset, week_num,weekday,month)

# Drop null values
oasis_headset = drop_na(oasis_headset)

# Testing on entire data set
# Testing model on month of March
oasis_headset_test = subset(oasis_headset, Date > "2022-03-01")

model_headset = lm(Headset~month+weekday, data = oasis_headset)
# summary(model_headset)
predict_headset = predict(model_headset, newdata = oasis_headset_test)
# Warning is given which is due to the reason that testing and training done on same da

result_headset = data.frame(date =oasis_headset_test$Date, actual = oasis_headset_test$
rsme_headset = sqrt(mean(result_headset$err^2))
```

# Keyboard

```
# Analysis of the item keyboard
oasis_keyboard = select(oasis, Date, Keyboard, week_num, weekday, month)

# Drop null values
oasis_keyboard = drop_na(oasis_keyboard)

# Testing on entire data set
# Testing model on month of March
oasis_keyboard_test = subset(oasis_keyboard, Date > "2022-03-01")

model_keyboard = lm(Keyboard~month+weekday, data = oasis_keyboard)
# summary(model_keyboard)
predict_keyboard = predict(model_keyboard, newdata = oasis_keyboard_test)
# Warning is given which is due to the reason that testing and training done on same da

result_keyboard = data.frame(date =oasis_keyboard_test$Date, actual = oasis_keyboard_te
rsme_keyboard = sqrt(mean(result_keyboard$err^2))
```

# Keyboard.and.Mouse

```
# Analysis of the item keyboard and mouse
oasis_keyboard_mouse = select(oasis, Date, Keyboard.and.Mouse, week_num, weekday, month)

# Drop null values
oasis_keyboard_mouse = drop_na(oasis_keyboard_mouse)

# Testing on entire data set
# Testing model on month of March
oasis_keyboard_mouse_test = subset(oasis_keyboard_mouse, Date > "2022-03-01")

model_keyboard_mouse = lm(Keyboard.and.Mouse~month+weekday, data = oasis_keyboard_mous
# summary(model_keyboard_mouse)
predict_keyboard_mouse = predict(model_keyboard_mouse, newdata = oasis_keyboard_mouse_t
# Warning is given which is due to the reason that testing and training done on same da

result_keyboard_mouse = data.frame(date =oasis_keyboard_mouse_test$Date, actual = oasis
rsme_keyboard_mouse = sqrt(mean(result_keyboard_mouse$err^2))
```

# Laptop

```
# Analysis of the item Laptop
oasis_laptop = select(oasis, Date, Laptop, week_num, weekday, month)
```

```
# Drop null values
oasis_laptop = drop_na(oasis_laptop)

# Testing on entire data set
# Testing model on month of March
oasis_laptop_test = subset(oasis_laptop, Date > "2022-03-01")

model_laptop = lm(Laptop~month+weekday, data = oasis_laptop)
# summary(model_laptop)
predict_laptop = predict(model_laptop, newdata = oasis_laptop_test)
# Warning is given which is due to the reason that testing and training done on same da

result_laptop = data.frame(date =oasis_laptop_test$Date, actual = oasis_laptop_test$Lap
rsme_laptop = sqrt(mean(result_laptop$serr^2))
```

## Laptop.Stand

```
# Analysis of the item Laptop stand
oasis_laptop_stand = select(oasis, Date, Laptop.Stand, week_num,weekday,month)

# Drop null values
oasis_laptop_stand = drop_na(oasis_laptop_stand)

# Testing on entire data set
# Testing model on month of March
oasis_laptop_stand_test = subset(oasis_laptop_stand, Date > "2022-03-01")

model_laptop_stand = lm(Laptop.Stand~month+weekday, data = oasis_laptop_stand)
# summary(model_laptop_stand)
predict_laptop_stand = predict(model_laptop_stand, newdata = oasis_laptop_stand_test)
# Warning is given which is due to the reason that testing and training done on same da

result_laptop_stand = data.frame(date =oasis_laptop_stand_test$Date, actual = oasis_lap
rsme_laptop_stand = sqrt(mean(result_laptop_stand$serr^2))
```

## Monitor

```
# Analysis of the item Monitor
oasis_monitor = select(oasis, Date, Monitor, week_num,weekday,month)

# Drop null values
oasis_monitor = drop_na(oasis_monitor)

# Testing on entire data set
# Testing model on month of March
oasis_monitor_test = subset(oasis_monitor, Date > "2022-03-01")
```

```

model_monitor = lm(Monitor~month+weekday, data = oasis_monitor)
# summary(model_monitor)
predict_monitor = predict(model_monitor, newdata = oasis_monitor_test)
# Warning is given which is due to the reason that testing and training done on same da

result_monitor = data.frame(date =oasis_laptop_stand_test$Date, actual = oasis_monitor_
rsme_monitor = sqrt(mean(result_laptop_stand$err^2))

```

## Mouse

```

# Analysis of the item Laptop stand
oasis_mouse = select(oasis, Date, Mouse, week_num,weekday,month)

# Drop null values
oasis_mouse = drop_na(oasis_mouse)

# Testing on entire data set
# Testing model on month of March
oasis_mouse_test = subset(oasis_mouse, Date > "2022-03-01")

model_mouse = lm(Mouse~month+weekday, data = oasis_mouse)
# summary(model_mouse)
predict_mouse = predict(model_mouse, newdata = oasis_mouse_test)
# Warning is given which is due to the reason that testing and training done on same da

result_mouse = data.frame(date =oasis_mouse_test$Date, actual = oasis_mouse_test$Mouse,
rsme_mouse = sqrt(mean(result_mouse$err^2))

```

## Multiport.Adapter

```

# Analysis of the item Multiport
oasis_multiport = select(oasis, Date, Multiport.Adapter, week_num,weekday,month)

# Drop null values
oasis_multiport = drop_na(oasis_multiport)

# Testing on entire data set
# Testing model on month of March
oasis_multiport_test = subset(oasis_multiport, Date > "2022-03-01")

model_multiport = lm(Multiport.Adapter~month+weekday, data = oasis_multiport)
# summary(model_multiport)
predict_multiport = predict(model_multiport, newdata = oasis_multiport_test)
# Warning is given which is due to the reason that testing and training done on same da

```

```
result_multiport = data.frame(date =oasis_multiport_test$Date, actual = oasis_multiport  
rsme_multiport = sqrt(mean(result_multiport$err^2))
```

## Power.Adapter

```
# Analysis of the item Power Adapter  
oasis_Adapter = select(oasis, Date, Power.Adapter, week_num,weekday,month)  
  
# Drop null values  
oasis_Adapter = drop_na(oasis_Adapter)  
  
# Testing on entire data set  
# Testing model on month of March  
oasis_Adapter_test = subset(oasis_Adapter, Date > "2022-03-01")  
  
model_Adapter = lm(Power.Adapter~month+weekday, data = oasis_Adapter)  
# summary(model_Adapter)  
predict_Adapter = predict(model_Adapter, newdata = oasis_Adapter_test)  
# Warning is given which is due to the reason that testing and training done on same da  
  
result_Adapter = data.frame(date =oasis_Adapter_test$Date, actual = oasis_Adapter_test$  
rsme_Adapter = sqrt(mean(result_Adapter$err^2))
```

## Speakerphone

```
# Analysis of the item Speakerphone  
oasis_Speaker = select(oasis, Date, Speakerphone, week_num,weekday,month)  
  
# Drop null values  
oasis_Speaker = drop_na(oasis_Speaker)  
  
# Testing on entire data set  
# Testing model on month of March  
oasis_Speaker_test = subset(oasis_Speaker, Date > "2022-03-01")  
  
model_Speaker = lm(Speakerphone~month+weekday, data = oasis_Speaker)  
#summary(model_Speaker)  
predict_Speaker = predict(model_Speaker, newdata = oasis_Speaker_test)  
# Warning is given which is due to the reason that testing and training done on same da  
  
result_Speaker = data.frame(date =oasis_Speaker_test$Date, actual = oasis_Speaker_test$  
rsme_Speaker = sqrt(mean(result_Speaker$err^2))
```

## Trackpad



```
# Analysis of the item Trackpad
oasis_trackpad = select(oasis, Date, Trackpad, week_num, weekday, month)

# Drop null values
oasis_trackpad = drop_na(oasis_trackpad)

# Testing on entire data set
# Testing model on month of March
oasis_trackpad_test = subset(oasis_trackpad, Date > "2022-03-01")

model_trackpad = lm(Trackpad~month+weekday, data = oasis_trackpad)
# summary(model_trackpad)
predict_trackpad = predict(model_trackpad, newdata = oasis_trackpad_test)
# Warning is given which is due to the reason that testing and training done on same da

result_trackpad = data.frame(date =oasis_trackpad_test$Date, actual = oasis_trackpad_te
rsme_trackpad = sqrt(mean(result_trackpad$err^2))
```

## Webcam

```
# Analysis of the item Webcam
oasis_webcam = select(oasis, Date, Webcam, week_num, weekday, month)

# Drop null values
oasis_webcam = drop_na(oasis_webcam)

# Testing on entire data set
# Testing model on month of March
oasis_webcam_test = subset(oasis_webcam, Date > "2022-03-01")

model_webcam = lm(Webcam~month+weekday, data = oasis_webcam)
# summary(model_webcam)
predict_webcam = predict(model_webcam, newdata = oasis_webcam_test)
# Warning is given which is due to the reason that testing and training done on same da

result_webcam = data.frame(date =oasis_webcam_test$Date, actual = oasis_webcam_test$Web
rsme_webcam = sqrt(mean(result_webcam$err^2))
```

Individual item has been predicted. Now combining the result to get the final prediction.  
Combining the result

## Combining the result

```
a1 = subset(oasis, oasis$Date > "2022-03-01")
predicted_order = data.frame(date = a1$Date)
```

```

predicted_order1 = merge(y = select(result_Adapter, date, predicted_adapter), x = predic
predicted_order1 = merge(x = predicted_order1, y = select(result_custom, date, predicte
predicted_order1 = merge(x = predicted_order1, y = select(result_disp, date, predicted_
predicted_order1 = merge(x = predicted_order1, y = select(result_headset, date, predict
predicted_order1 = merge(x = predicted_order1, y = select(result_keyboard, date, predic
predicted_order1 = merge(x = predicted_order1, y = select(result_keyboard_mouse, date,
predicted_order1 = merge(x = predicted_order1, y = select(result_laptop, date, predicte

```

```

predicted_order1 = merge(x = predicted_order1, y = select(result_laptop_stand, date, pr
predicted_order1 = merge(x = predicted_order1, y = select(result_monitor, date, predict
predicted_order1 = merge(x = predicted_order1, y = select(result_mouse, date, predicted
predicted_order1 = merge(x = predicted_order1, y = select(result_multiport, date, predi
predicted_order1 = merge(x = predicted_order1, y = select(result_Speaker, date, predict
predicted_order1 = merge(x = predicted_order1, y = select(result_trackpad, date, predic
predicted_order1 = merge(x = predicted_order1, y = select(result_webcam, date, predicte

```

```

#predicted_order1$Predicted_total = rowSums(predicted_order1$predicted_adapter, na.rm =
a1 = predicted_order1
a1$date = NULL
a1[is.na(a1)] = 0

```

```

a1 = a1 %>% mutate(total_prediction = rowSums(across(where(is.numeric), na.rm = T)))
a1 = cbind(date = predicted_order1$date, a1)
b = subset(oasis, oasis$Date > "2022-03-01")
a1 = cbind(a1, actual_order = b$Grand.Total)
rmse = sqrt(mean(a1$actual_order - a1$total_prediction)^2)
colnames(a1)

```

```

## [1] "date" "predicted_adapter"
## [3] "predicted_custom" "predicted_disp"
## [5] "predicted_headset" "predicted_keyboard"
## [7] "predicted_keyboard_mouse" "predicted_laptop"
## [9] "predicted_laptop_stand" "predicted_monitor"
## [11] "predicted_mouse" "predicted_multiport"
## [13] "predicted_speaker" "predicted_trackpad"
## [15] "predicted_webcam" "total_prediction"
## [17] "actual_order"

```

```

ggplot(a1) + geom_bar(aes(x = date, y = actual_order, fill="Actual Order"), stat = "iden

```

## Actual and Predicted Order

Comparison of actual and predicted order of March 2022

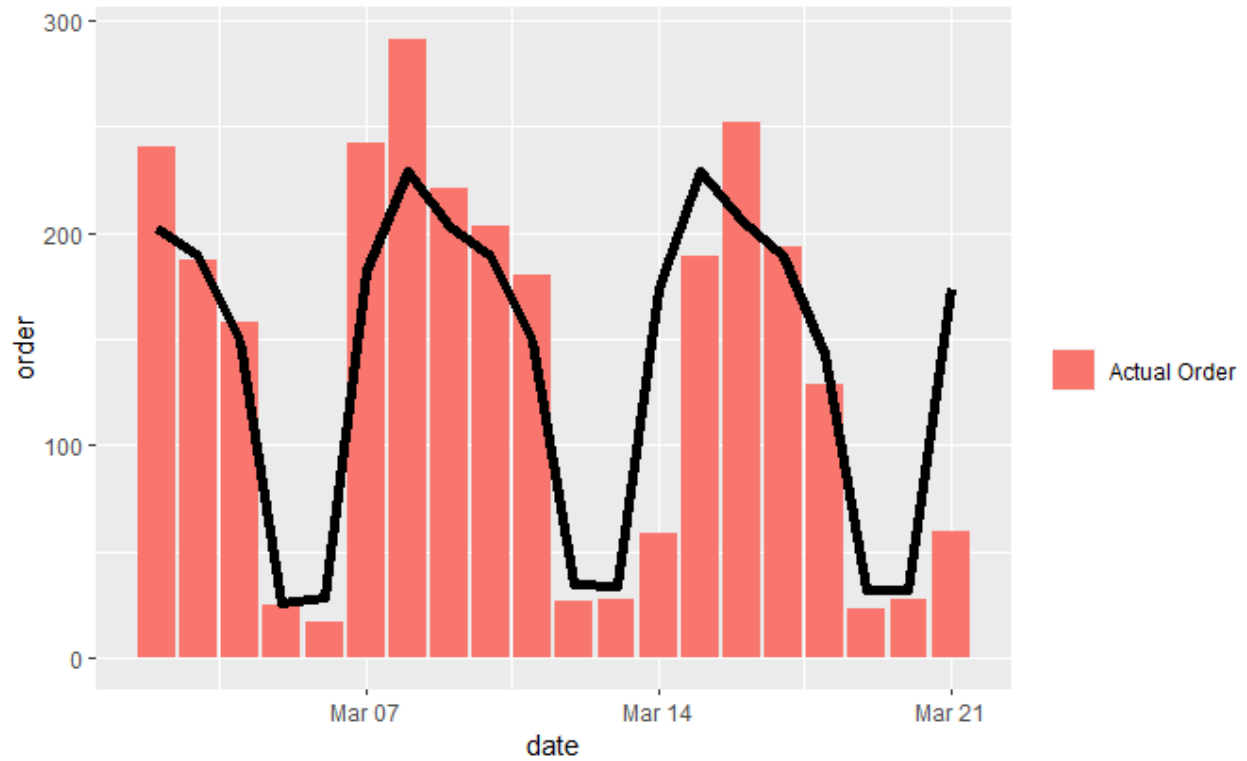


Figure 1: Date wise actual and predicted order in month of March 2022.