**Initial Analysis**

We are given 2 datasets to perform data analysis, and make a recommendation to maximize the profit. The first dataset “sales.csv” has all the sales information of 3 different burgers “Ham”, “Turkey” and “Veggie”, which were sold between the dates 3/3/2014 to 8/29/2014 (not including the Saturdays and Sundays), and the other dataset “details” has the burgers preparation cost and sale price. There are 130 observations in “sales.csv”.

The sales.csv data set has the following variables:

date – represents the date, on which the burgers were sold.

demand.ham – represents the demand for ham burgers, on the given date.

demand.turkey – represents the demand for turkey burgers, on the given date

demand.veggie – represents the demand for veggie burgers, on the given date

available.ham – represents the number of ham burgers prepared, on the given date

available.turkey – represents the number of turkey burgers prepared, on the given date

available.veggie – represents the number of veggie burgers prepared, on the given date

The other dataset “details.csv” has the following variables

type – represents the type of the burger (can be ham or turkey or veggie)

price – represents the price of selling the burger

cost – represents the cost of preparation of the burger

We have only 3 observations in this data set, and it is given as shown below:

|  |  |  |
| --- | --- | --- |
| type | price | cost |
| ham | 6.5 | 3.5 |
| turkey | 6.5 | 4 |
| veggie | 5 | 2.5 |

**Table 1:** details.csv data

Our main objective is to predict the demand, and increase the profit. Before starting further analysis, we assumed that the sales of the three different burgers are independent of each other, since the coefficient of correlation between the three burgers are close to 0, showing no strong relationship. Here is the R Code used to calculate the coefficient of correlation.

> cor(sales$demand.turkey,sales$demand.veggie)

[1] 0.1070533

> cor(sales$demand.turkey,sales$demand.ham)

[1] 0.0005572513

> cor(sales$demand.veggie,sales$demand.ham)

[1] 0.08680582

We read the sales.csv to a data frame in R.

To track the profits from the sales of 3 burger varieties, we added 3 new variables to the sales data frame to represent the profit made from sales of ham, turkey and veggie respectively.

#Reading the sales.csv to a data frame

sales\_1 <- read.csv(file="sales.csv", header=TRUE, sep=",")

#Adding a variable to represent the ham profits

sales\_1$actual.ham.profit <- (ifelse((sales\_1$demand.ham <= sales\_1$available.ham), ((sales\_1$demand.ham \* 6.5) - (sales\_1$available.ham \* 3.5)),

((sales\_1$available.ham \* 6.5) - (sales\_1$available.ham \* 3.5))))

#Found the actual ham profit as 4837, using the following command

sum(sales\_1$actual.ham.profit)

#Adding a variable to represent the turkey profits

sales\_1$actual.turkey.profit <- (ifelse((sales\_1$demand.turkey <= sales\_1$available.turkey), ((sales\_1$demand.turkey \* 6.5) - (sales\_1$available.turkey \* 4)), ((sales\_1$available.turkey \* 6.5) - (sales\_1$available.turkey \* 4))))

#Found the actual turkey profit as 5106, using the following command

sum(sales\_1$actual.turkey.profit)

#Adding a variable to represent the veggie profits

sales\_1$actual.veggie.profit <- (ifelse((sales\_1$demand.veggie <= sales\_1$available.veggie), ((sales\_1$demand.veggie \* 5) - (sales\_1$available.veggie \* 2.5)), ((sales\_1$available.veggie \* 5) - (sales\_1$available.veggie \* 2.5))))

#Found the actual veggie profit as 2885, using the following command

sum(sales\_1$actual.veggie.profit)

**Approach:**

We started by first computing the profit that could be made, by using the average value of the demand for each burger flavors. Later we analyzed if the demand for the burgers change, depending on the weekdays (Mon, Tue, Wed, Thu, Fri).

We obtained the following results, and recommend the following number of burgers to be prepared on a given week day, to maximize the profits. Please read the remaining document, for a detailed explanation about the analysis we performed and how we arrived at these numbers:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Burger** | **Mon** | **Tues** | **Wed** | **Thurs** | **Fri** |
| **Ham** | 16 | 16 | 15 | 16 | 16 |
| **Turkey** | 24 | 22 | 21 | 22 | 21 |
| **Veggie** | 13 | 13 | 14 | 13 | 12 |

**Table: Recommended number of burgers to be prepared on each week day**

If the burgers are prepared as per the day of the week (as suggested above), then we can get the following profit, when compared to the existing profit.

|  |  |  |  |
| --- | --- | --- | --- |
| **Burger Type** | **Existing profit** | **Potential profit** | **Net profit = (Potential profit – Existing profit)** |
| Ham | 4837$ | 4940$ | 103$ |
| Turkey | 5106$ | 5499$ | 393$ |
| Veggie | 2885$ | 3345$ | 460$ |

**Table: Potential Profit table**

**Detailed Analysis**

The rest of the document describes how we arrived at the “recommended number of burgers to be prepared on each week day, for maximum profit.

**Generating the probability distribution and expected values**

Using the R table function, we generated the following probability distributions for the three burgers flavors:

**Table 2:** Ham burger Probability Distribution and expected value

|  |  |  |
| --- | --- | --- |
| **ham** | **probability** | **expected value** |
| 6 | 0.01538462 | 0.09230772 |
| 9 | 0.03076923 | 0.27692307 |
| 10 | 0.03076923 | 0.3076923 |
| 11 | 0.05384615 | 0.59230765 |
| 12 | 0.04615385 | 0.5538462 |
| 13 | 0.07692308 | 1.00000004 |
| 14 | 0.09230769 | 1.29230766 |
| 15 | 0.12307692 | 1.8461538 |
| 16 | 0.1 | 1.6 |
| 17 | 0.12307692 | 2.09230764 |
| 18 | 0.05384615 | 0.9692307 |
| 19 | 0.06923077 | 1.31538463 |
| 20 | 0.06153846 | 1.2307692 |
| 21 | 0.03846154 | 0.80769234 |
| 22 | 0.02307692 | 0.50769224 |
| 23 | 0.03076923 | 0.70769229 |
| 24 | 0.01538462 | 0.36923088 |
| 25 | 0.01538462 | 0.3846155 |
| **Mean demand of ham burgers** | | **15.94615386 =16 (appx)** |

**Table 3:** Turkey burger Probability Distribution and expected value

|  |  |  |
| --- | --- | --- |
| **turkey** | **Probability** | **expected value** |
| 13 | 0.023076923 | 0.299999999 |
| 14 | 0.030769231 | 0.430769234 |
| 15 | 0.030769231 | 0.461538465 |
| 16 | 0.038461538 | 0.615384608 |
| 17 | 0.069230769 | 1.176923073 |
| 18 | 0.061538462 | 1.107692316 |
| 19 | 0.061538462 | 1.169230778 |
| 20 | 0.1 | 2 |
| 21 | 0.092307692 | 1.938461532 |
| 22 | 0.084615385 | 1.86153847 |
| 23 | 0.046153846 | 1.061538458 |
| 24 | 0.053846154 | 1.292307696 |
| 25 | 0.038461538 | 0.96153845 |
| 26 | 0.061538462 | 1.600000012 |
| 27 | 0.069230769 | 1.869230763 |
| 28 | 0.023076923 | 0.646153844 |
| 29 | 0.038461538 | 1.115384602 |
| 30 | 0.023076923 | 0.69230769 |
| 31 | 0.023076923 | 0.715384613 |
| 32 | 0.015384615 | 0.49230768 |
| 34 | 0.007692308 | 0.261538472 |
| 37 | 0.007692308 | 0.284615396 |
| **Mean demand (Turkey)** | | **22.05384615=22(appx)** |

**Table 4:** Veggie burger Probability Distribution and expected value

|  |  |  |
| --- | --- | --- |
| **veggie** | **Probability** | **expected value** |
| 4 | 0.00769231 | 0.030769232 |
| 5 | 0.00769231 | 0.03846154 |
| 6 | 0.01538462 | 0.09230769 |
| 7 | 0.00769231 | 0.053846156 |
| 8 | 0.04615385 | 0.369230768 |
| 9 | 0.08461539 | 0.761538465 |
| 10 | 0.08461539 | 0.84615385 |
| 11 | 0.07692308 | 0.846153847 |
| 12 | 0.09230769 | 1.107692304 |
| 13 | 0.13846154 | 1.799999994 |
| 14 | 0.10769231 | 1.507692312 |
| 15 | 0.1 | 1.5 |
| 16 | 0.09230769 | 1.476923072 |
| 17 | 0.04615385 | 0.784615382 |
| 18 | 0.02307692 | 0.415384614 |
| 19 | 0.02307692 | 0.438461537 |
| 20 | 0.01538462 | 0.3076923 |
| 21 | 0.01538462 | 0.323076915 |
| 23 | 0.00769231 | 0.176923084 |
| 24 | 0.00769231 | 0.184615392 |
| **Mean demand (Veggie)** |  | **13.06153845=13 (appx)** |

Let us use the mean values for ham, turkey and veggie, and calculate the projected demand:

We used the following R Code, and found that the estimated profit would be 4933.5$, if 16 burgers of ham are prepared daily. This profit is 96.5$ more than the current profit for ham (current profit for ham = 4837$)

sales\_1$actual.ham.profit <- (ifelse((sales\_1$demand.ham <= 16), ((sales\_1$demand.ham \* 6.5) - (16 \* 3.5)), ((16 \* 6.5) - (16 \* 3.5))))

Similarly, we found that the profit would be 5486$, if 22 Turkey burgers are prepared daily. This profit is 380$ more than the current profit (current Turkey burgers profit is 5106$)

sum(ifelse((sales\_1$demand.turkey <= 22), ((sales\_1$demand.turkey \* 6.5) - (22 \* 4)), ((22 \* 6.5) - (22 \* 4))))

On the similar lines, if 13 veggie burgers are prepared daily, then we will obtain a profit of 3345$, which is 460$ more than the current veggie burgers profit (current veggie burgers profit is 2885)

sum (ifelse((sales\_1$demand.veggie <= 13), ((sales\_1$demand.veggie \* 5) - (13 \* 2.5)), ((13 \* 5) - (13 \* 2.5))))

If we use the average demand, then we obtain the following profit:

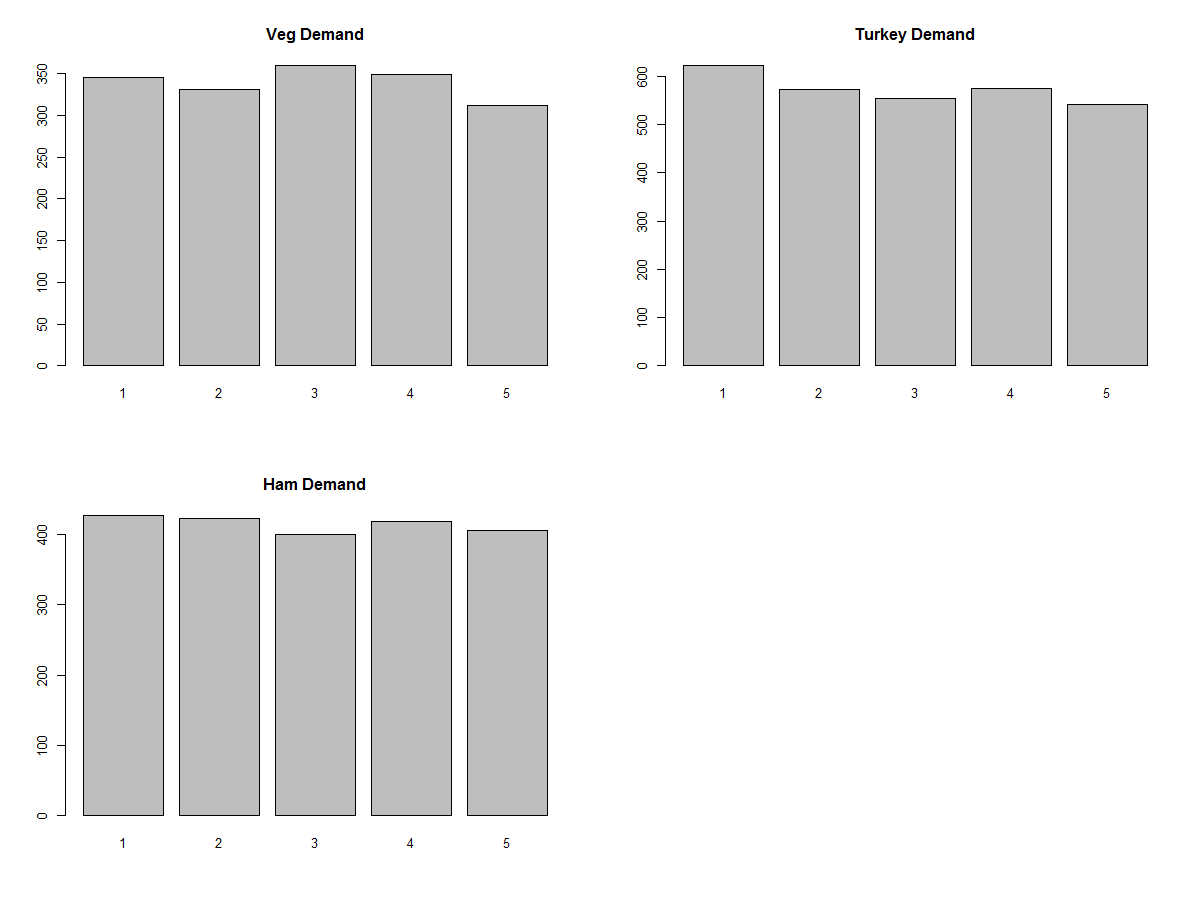
|  |  |  |  |
| --- | --- | --- | --- |
| **Burger Type** | **Existing profit**  **(EP)** | **Potential profit (if average is used)**  **(PP)** | **Profits difference**  **(PP – EP)** |
| Ham | 4837$ | 4933.5$ (if 16 ham burgers are prepared on every day) | 4933.5 – 4837 = 96.5$ |
| Turkey | 5106$ | 5486$ (if 22 Turkey burgers are prepared on every day) | 5486 – 5106 = 380$ |
| Veggie | 2885$ | 3345$ (if 13 Veggie burgers are prepared on every day) | 3345 – 2885 = 460$ |

**Table 5:** Existing and potential profits, if average daily demand is used to prepare the number of burgers per day

**Effect of week day on burgers demand:**

We analyzed if there is any effect on the week day, and the demand of the burgers. We drew the following graphs, which shows that the demand for the burgers vary depending on the week day.

**Figure 1: Week day demand, for each burger type**



Given the shapes of the histograms, we concluded that the demand for burgers change depending on the week day, and there is room to maximize the profits using the conditional probability concept. For instance, veg burgers have the most demand on Wednesday, while Ham has the least demand on Wednesday. So if we design a strategy to prepare more veggie burgers on Wed and least Ham burgers on Wednesday, then we could potentially increase the profit margins.

We know that we have 5 days – Mon, Tue, Wed, Thu, Fri. Since these are equally likely, we can write the following Probability distribution for week days in the sales data:

P(Mon) = P(Tue) = P(Wed) = P(Thu) = P(Fri) = 1/5 = 0.2

To make our analysis easier, we added a new variable “wday” to the sales dataset:

library("lubridate")

sales\_1$wday <- wday(ymd(sales$date),label=T)

Here are the conditional probability distribution for veggie burger, given the week day:

P(Vd = X | Week\_Day = D); where D = { MON, TUES, WED, THURS, FRI}

Vd = Veggie demand = {4 to 24}; if any number is not given then that probability is 0, and is not given in the following table.

P(Vd = X | Week\_Day = D) = P(Vd =X and Week\_Day = D) / P(Week\_Day = D)

We used the R Command: prop.table(table(sales\_1$demand.veggie,sales\_1$wday))/0.2

to generate the following table (this has conditional probability of demand of Veggie burger, on a given day):

**Demand Mon Tues Wed Thurs Fri**

**4**  0.00000000 0.00000000 0.03846154 0.00000000 0.00000000

**5**  0.00000000 0.00000000 0.00000000 0.00000000 0.03846154

**6**  0.00000000 0.03846154 0.03846154 0.00000000 0.00000000

**7**  0.03846154 0.00000000 0.00000000 0.00000000 0.00000000

**8**  0.03846154 0.07692308 0.00000000 0.03846154 0.07692308

**9**  0.15384615 0.03846154 0.03846154 0.15384615 0.03846154

**10** 0.07692308 0.07692308 0.03846154 0.07692308 0.15384615

**11** 0.03846154 0.15384615 0.07692308 0.03846154 0.07692308

**12** 0.07692308 0.15384615 0.11538462 0.07692308 0.03846154

**13** 0.03846154 0.07692308 0.07692308 0.07692308 0.42307692

**14** 0.07692308 0.11538462 0.11538462 0.19230769 0.03846154

**15** 0.15384615 0.11538462 0.11538462 0.11538462 0.00000000

**16** 0.15384615 0.03846154 0.15384615 0.07692308 0.03846154

**17** 0.03846154 0.03846154 0.07692308 0.03846154 0.03846154

**18** 0.03846154 0.00000000 0.03846154 0.00000000 0.03846154

**19** 0.03846154 0.03846154 0.00000000 0.03846154 0.00000000

**20** 0.00000000 0.00000000 0.03846154 0.03846154 0.00000000

**21** 0.03846154 0.00000000 0.00000000 0.03846154 0.00000000

**23** 0.00000000 0.03846154 0.00000000 0.00000000 0.00000000

**24** 0.00000000 0.00000000 0.03846154 0.00000000 0.00000000

P(Hd = X | Week\_Day = D); where D = { MON, TUES, WED, THURS, FRI}

Hd = Ham demand = X = {6 to 25}; if any number is not given then that probability is 0, and is not given in the following table (this has conditional probability of demand of Ham burger, on a given day):

P(Hd = X | Week\_Day = D) = P(Hd =X and Week\_Day = D) / P(Week\_Day = D)

We used the R Command: prop.table(table(sales\_1$demand.ham,sales\_1$wday))/0.2

to generate the following table:

**Demand Mon Tues Wed Thurs Fri**

**6**  0.00000000 0.03846154 0.03846154 0.00000000 0.00000000

**9**  0.00000000 0.00000000 0.07692308 0.00000000 0.07692308

**10**  0.07692308 0.03846154 0.00000000 0.03846154 0.00000000

**11**  0.03846154 0.00000000 0.07692308 0.07692308 0.07692308

**12**  0.03846154 0.03846154 0.03846154 0.03846154 0.07692308

**13**  0.07692308 0.03846154 0.11538462 0.07692308 0.07692308

**14**  0.07692308 0.07692308 0.07692308 0.07692308 0.15384615

**15**  0.11538462 0.19230769 0.07692308 0.15384615 0.07692308

**16**  0.07692308 0.15384615 0.03846154 0.19230769 0.03846154

**17**  0.15384615 0.07692308 0.19230769 0.07692308 0.11538462

**18**  0.11538462 0.03846154 0.03846154 0.00000000 0.07692308

**19**  0.03846154 0.11538462 0.03846154 0.03846154 0.11538462

**20**  0.07692308 0.11538462 0.03846154 0.07692308 0.00000000

**21**  0.00000000 0.03846154 0.07692308 0.07692308 0.00000000

**22**  0.00000000 0.00000000 0.03846154 0.03846154 0.03846154

**23**  0.03846154 0.00000000 0.03846154 0.03846154 0.03846154

**24**  0.03846154 0.00000000 0.00000000 0.00000000 0.03846154

**25**  0.03846154 0.03846154 0.00000000 0.00000000 0.00000000

P(Td = X | Week\_Day = D); where D = { MON, TUES, WED, THURS, FRI}

Td = Turkey demand = X = {13 to 37}; if any number is not given then that probability is 0, and is not given in the following table (this has conditional probability of demand of Turkey burger, on a given day):

P(Td = X | Week\_Day = D) = P(Td =X and Week\_Day = D) / P(Week\_Day = D)

We used the R Command: prop.table(table(sales\_1$demand.turkey,sales\_1$wday))/0.2

to generate the following table:

**Demand Mon Tues Wed Thurs Fri**

**13**  0.00000000 0.00000000 0.07692308 0.00000000 0.03846154

**14**  0.03846154 0.07692308 0.00000000 0.03846154 0.00000000

**15**  0.03846154 0.00000000 0.03846154 0.03846154 0.03846154

**16**  0.00000000 0.03846154 0.07692308 0.03846154 0.03846154

**17**  0.07692308 0.11538462 0.03846154 0.03846154 0.07692308

**18**  0.03846154 0.07692308 0.00000000 0.00000000 0.19230769

**19**  0.11538462 0.00000000 0.03846154 0.11538462 0.03846154

**20**  0.11538462 0.03846154 0.15384615 0.15384615 0.03846154

**21**  0.03846154 0.11538462 0.11538462 0.07692308 0.11538462

**22**  0.03846154 0.11538462 0.07692308 0.03846154 0.15384615

**23**  0.03846154 0.03846154 0.03846154 0.03846154 0.07692308

**24**  0.00000000 0.03846154 0.11538462 0.07692308 0.03846154

**25**  0.00000000 0.11538462 0.00000000 0.07692308 0.00000000

**26**  0.03846154 0.03846154 0.07692308 0.11538462 0.03846154

**27**  0.11538462 0.03846154 0.07692308 0.07692308 0.03846154

**28**  0.00000000 0.03846154 0.03846154 0.03846154 0.00000000

**29**  0.07692308 0.07692308 0.00000000 0.00000000 0.03846154

**30**  0.07692308 0.00000000 0.03846154 0.00000000 0.00000000

**31**  0.03846154 0.03846154 0.00000000 0.03846154 0.00000000

**32**  0.03846154 0.00000000 0.00000000 0.00000000 0.03846154

**34**  0.03846154 0.00000000 0.00000000 0.00000000 0.00000000

**37**  0.03846154 0.00000000 0.00000000 0.00000000 0.00000000

Now let us find the average values of burgers, on all the week days. Using the above conditional probability distributions above, we obtained the following expected values for each burger on a given week day (The complete R Code is given at the end):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Burger | Mon | Tues | Wed | Thurs | Fri |
| Ham | 16.423 | 16.2693 | 15.384 | 16.07692 | 15.57692 |
| Turkey | 23.961 | 22 | 21.34615 | 22.11358 | 20.846 |
| Veggie | 13.30769 | 12.73077 | 13.8461 | 13.42308 | 12 |

Rounding the above number of burgers, we get the following table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Burger** | **Mon** | **Tues** | **Wed** | **Thurs** | **Fri** |
| **Ham** | 16 | 16 | 15 | 16 | 16 |
| **Turkey** | 24 | 22 | 21 | 22 | 21 |
| **Veggie** | 13 | 13 | 14 | 13 | 12 |

Using the above number of burgers on a given day, we could obtain the following profits:

|  |  |
| --- | --- |
| **Burger** | **Profit** |
| Ham | 4940$ |
| Turkey | 5499$ |
| Veggie | 3345$ |

The above profits (obtained using conditional probability) for Ham and Turkey are more than the profits obtained using non-conditional probability. The following table shows all the possible profits, when compared to the current profits:

|  |  |  |  |
| --- | --- | --- | --- |
| **Burger Type** | **Existing profit**  **(EP)** | **Potential profit (non-conditional probability)**  **(PP)** | **Potential profit, if conditional (on week day) is used** |
| Ham | 4837$ | 4933.5$ | 4940$ |
| Turkey | 5106$ | 5486$ | 5499$ |
| Veggie | 2885$ | 3345$ | 3345$ |

Since the profit for conditional probability model has the maximum value, we recommend to prepare the burgers as per the given table (shown below):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Burger** | **Mon** | **Tues** | **Wed** | **Thurs** | **Fri** |
| **Ham** | 16 | 16 | 15 | 16 | 16 |
| **Turkey** | 24 | 22 | 21 | 22 | 21 |
| **Veggie** | 13 | 13 | 14 | 13 | 12 |

**R- Code used for analysis**

**#606 PROJECT 1 – Ham data analysis**

sales\_1 <- read.csv(file="sales.csv", header=TRUE, sep=",")

library("lubridate")

sales\_1$wday <- wday(ymd(sales$date),label=T)

sales\_1

#To get probabilities, irrespective of the week day:

data.frame(prop.table(table(sales\_1$demand.ham)))

prop.table(table(sales\_1$demand.ham,sales\_1$wday))

hist(prop.table(table(sales\_1$demand.ham,sales\_1$wday))[,2])

cumsum(prop.table(table(sales\_1$demand.ham,sales\_1$wday))[,2])

summary(prop.table(table(sales\_1$demand.ham,sales\_1$wday))[,2])

#The probability of week day can be found like this. This is needed to find the cond prob.

sum(prop.table(table(sales\_1$demand.ham,sales\_1$wday))[,2])

sum(prop.table(table(sales\_1$demand.ham,sales\_1$wday))[,3])

sum(prop.table(table(sales\_1$demand.ham,sales\_1$wday))[,4])

sum(prop.table(table(sales\_1$demand.ham,sales\_1$wday))[,5])

sum(prop.table(table(sales\_1$demand.ham,sales\_1$wday))[,6])

#To get the conditional probability by day (the following command will get the probability of ham demand, given the day as friday):

#P(ham.demand = x | day = FRIDAY) = P(DAY=FRIDAY AND ham.demand = x) / P(DAY = FRIDAY)

#where X is the demand number. For instance, if X = 11, then P(DAY = FRIDAY AND X=11) = 0.015384615

#P(FRIDAY) = 0.2. Therefore P(demand.ham = 11 | FRIDAY) = 0.015384615/0.2

prop.table(table(sales\_1$demand.ham,sales\_1$wday))[,6]/0.2

#To find all the conditional probabilities, cond on week day:

prop.table(table(sales\_1$demand.ham,sales\_1$wday))/0.2

data.frame((prop.table(table(sales\_1$demand.ham,sales\_1$wday))/0.2))

#To find marginal probabilities, use the following command. This will find the marginal probabilities horizintally for

#prop.table(table(sales\_1$demand.ham,sales\_1$wday))

margin.table(prop.table(table(sales\_1$demand.ham,sales\_1$wday)),1)

#The following command will find the marginal probabilities vertically

margin.table(prop.table(table(sales\_1$demand.ham,sales\_1$wday)),2)

#Let us convert the following command's o/p to dataframe:

#prop.table(table(sales\_1$demand.ham,sales\_1$wday))/0.2

#NOTE That the prob column in the following column, is the conditional probability of demand, given a week day.

week\_day\_prob <- data.frame((prop.table(table(sales\_1$demand.ham,sales\_1$wday))/0.2),stringsAsFactors=F)

#Rename the column names

names(week\_day\_prob) <- c("demand","week\_day","prob")

#To verify, we have conditional probabilities based on a week day

sum(week\_day\_prob[week\_day\_prob$week\_day == "Mon",]$prob)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Tues",]$prob)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Wed",]$prob)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Thurs",]$prob)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Fri",]$prob)

#Let us add another variable expected\_val

# problem here...but resolved by using "levels" in as.numeric

week\_day\_prob$expected\_val <- ( as.numeric(levels(week\_day\_prob$demand)) \* (week\_day\_prob$prob))

#Number of ham burgers needed on a week day

sum(week\_day\_prob[week\_day\_prob$week\_day == "Mon",]$expected\_val)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Tues",]$expected\_val)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Wed",]$expected\_val)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Thurs",]$expected\_val)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Fri",]$expected\_val)

#Round

round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Mon",]$expected\_val))

round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Tues",]$expected\_val))

round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Wed",]$expected\_val))

round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Thurs",]$expected\_val))

round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Fri",]$expected\_val))

#######################################################

##Now finding the probability by ignoring the week day#

#######################################################

data.frame(table(sales\_1$demand.ham))

all\_prob <- data.frame(prop.table(table(as.numeric(levels(sales\_1$demand.ham)))))

names(all\_prob) <- c("demand","freq")

sum(all\_prob$demand \* all\_prob$freq)

#######################################################

#Let us evaluate the profit

#######################################################

#Adding a new variable

#new\_avail\_cond\_by\_wday

sales\_1$new\_avail\_cond\_by\_wday <- 0

sales\_1$new\_avail\_cond\_by\_wday[sales\_1$wday=="Mon"] <- round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Mon",]$expected\_val))

sales\_1$new\_avail\_cond\_by\_wday[sales\_1$wday=="Tues"] <- round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Tues",]$expected\_val))

sales\_1$new\_avail\_cond\_by\_wday[sales\_1$wday=="Wed"] <- round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Wed",]$expected\_val))

sales\_1$new\_avail\_cond\_by\_wday[sales\_1$wday=="Thurs"] <- round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Thurs",]$expected\_val))

sales\_1$new\_avail\_cond\_by\_wday[sales\_1$wday=="Fri"] <- round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Fri",]$expected\_val))

sales\_1$cond\_prob\_profit <- (ifelse((sales\_1$demand.ham <= sales\_1$new\_avail\_cond\_by\_wday), ((sales\_1$demand.ham \* 6.50) - (sales\_1$new\_avail\_cond\_by\_wday \* 3.5)),

((sales\_1$new\_avail\_cond\_by\_wday \* 6.50) - (sales\_1$new\_avail\_cond\_by\_wday \* 3.5)))

)

print(sum(sales\_1$cond\_prob\_profit))

**#606 PROJECT 1 Turkey data analysis**

sales\_1 <- read.csv(file="sales.csv", header=TRUE, sep=",")

library("lubridate")

sales\_1$wday <- wday(ymd(sales$date),label=T)

sales\_1

#To get probabilities, irrespective of the week day:

data.frame(prop.table(table(sales\_1$demand.turkey)))

prop.table(table(sales\_1$demand.turkey,sales\_1$wday))

hist(prop.table(table(sales\_1$demand.turkey,sales\_1$wday))[,2])

cumsum(prop.table(table(sales\_1$demand.turkey,sales\_1$wday))[,2])

summary(prop.table(table(sales\_1$demand.turkey,sales\_1$wday))[,2])

#The probability of week day can be found like this. This is needed to find the cond prob.

sum(prop.table(table(sales\_1$demand.turkey,sales\_1$wday))[,2])

sum(prop.table(table(sales\_1$demand.turkey,sales\_1$wday))[,3])

sum(prop.table(table(sales\_1$demand.turkey,sales\_1$wday))[,4])

sum(prop.table(table(sales\_1$demand.turkey,sales\_1$wday))[,5])

sum(prop.table(table(sales\_1$demand.turkey,sales\_1$wday))[,6])

#To get the conditional probability by day (the following command will get the probability of turkey demand, given the day as friday):

#P(turkey.demand = x | day = FRIDAY) = P(DAY=FRIDAY AND turkey.demand = x) / P(DAY = FRIDAY)

#where X is the demand number. For instance, if X = 17, then P(DAY = FRIDAY AND X=17) = 0.015384615

#P(FRIDAY) = 0.2. Therefore P(demand.ham = 17 | FRIDAY) = 0.015384615/0.2

prop.table(table(sales\_1$demand.turkey,sales\_1$wday))[,6]/0.2

#To find all the conditional probabilities, cond on week day:

prop.table(table(sales\_1$demand.turkey,sales\_1$wday))/0.2

data.frame((prop.table(table(sales\_1$demand.turkey,sales\_1$wday))/0.2))

#To find marginal probabilities, use the following command. This will find the marginal probabilities horizintally for

#prop.table(table(sales\_1$demand.turkey,sales\_1$wday))

margin.table(prop.table(table(sales\_1$demand.turkey,sales\_1$wday)),1)

#The following command will find the marginal probabilities vertically

margin.table(prop.table(table(sales\_1$demand.turkey,sales\_1$wday)),2)

#Let us convert the following command's o/p to dataframe:

#prop.table(table(sales\_1$demand.turkey,sales\_1$wday))/0.2

#NOTE That the prob column in the following column, is the conditional probability of demand, given a week day.

week\_day\_prob <- data.frame((prop.table(table(sales\_1$demand.turkey,sales\_1$wday))/0.2),stringsAsFactors=F)

#Rename the column names

names(week\_day\_prob) <- c("demand","week\_day","prob")

#To verify, we have conditional probabilities based on a week day

sum(week\_day\_prob[week\_day\_prob$week\_day == "Mon",]$prob)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Tues",]$prob)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Wed",]$prob)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Thurs",]$prob)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Fri",]$prob)

#Let us add another variable expected\_val

# problem here...but resolved by using "levels" in as.numeric

week\_day\_prob$expected\_val <- ( as.numeric(levels(week\_day\_prob$demand)) \* (week\_day\_prob$prob))

#Number of ham burgers needed on a week day

sum(week\_day\_prob[week\_day\_prob$week\_day == "Mon",]$expected\_val)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Tues",]$expected\_val)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Wed",]$expected\_val)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Thurs",]$expected\_val)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Fri",]$expected\_val)

#Round

round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Mon",]$expected\_val))

round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Tues",]$expected\_val))

round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Wed",]$expected\_val))

round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Thurs",]$expected\_val))

round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Fri",]$expected\_val))

#######################################################

##Now finding the probability by ignoring the week day#

#######################################################

data.frame(table(sales\_1$demand.turkey))

all\_prob <- data.frame(prop.table(table(sales\_1$demand.turkey)))

names(all\_prob) <- c("demand","freq")

sum(as.numeric(levels(all\_prob$demand)) \* all\_prob$freq)

#######################################################

#Let us evaluate the profit

#######################################################

#Adding a new variable

#new\_avail\_cond\_by\_wday

sales\_1$turkey.new\_avail.cond\_prb <- 0

sales\_1$turkey.new\_avail.cond\_prb[sales\_1$wday=="Mon"] <- round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Mon",]$expected\_val))

sales\_1$turkey.new\_avail.cond\_prb[sales\_1$wday=="Tues"] <- round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Tues",]$expected\_val))

sales\_1$turkey.new\_avail.cond\_prb[sales\_1$wday=="Wed"] <- round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Wed",]$expected\_val))

sales\_1$turkey.new\_avail.cond\_prb[sales\_1$wday=="Thurs"] <- round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Thurs",]$expected\_val))

sales\_1$turkey.new\_avail.cond\_prb[sales\_1$wday=="Fri"] <- round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Fri",]$expected\_val))

sales\_1$turkey.cond\_prob\_profit <- (ifelse((sales\_1$demand.turkey <= sales\_1$turkey.new\_avail.cond\_prb), ((sales\_1$demand.turkey \* 6.50) - (sales\_1$turkey.new\_avail.cond\_prb \* 4)),

((sales\_1$turkey.new\_avail.cond\_prb \* 6.50) - (sales\_1$turkey.new\_avail.cond\_prb \* 4)))

)

print(sum(sales\_1$turkey.cond\_prob\_profit))

sales\_1$turkey.actual <- (ifelse((sales\_1$demand.turkey <= sales\_1$available.turkey), ((sales\_1$demand.turkey \* 6.50) - (sales\_1$available.turkey \* 4)),

((sales\_1$available.turkey \* 6.50) - (sales\_1$available.turkey \* 4)))

)

print(sum(sales\_1$turkey.actual))

#Just considering the 22

sum(ifelse((sales\_1$demand.turkey <= 22), ((sales\_1$demand.turkey \* 6.50) - (22 \* 4)),

((22 \* 6.50) - (22 \* 4)))

)

**#606 PROJECT 1 – Veggie data analysis**

sales\_1 <- read.csv(file="sales.csv", header=TRUE, sep=",")

library("lubridate")

sales\_1$wday <- wday(ymd(sales$date),label=T)

sales\_1

#To get probabilities, irrespective of the week day:

data.frame(prop.table(table(sales\_1$demand.veggie)))

prop.table(table(sales\_1$demand.veggie,sales\_1$wday))

hist(prop.table(table(sales\_1$demand.veggie,sales\_1$wday))[,2])

cumsum(prop.table(table(sales\_1$demand.veggie,sales\_1$wday))[,2])

summary(prop.table(table(sales\_1$demand.veggie,sales\_1$wday))[,2])

#The probability of week day can be found like this. This is needed to find the cond prob.

sum(prop.table(table(sales\_1$demand.veggie,sales\_1$wday))[,2])

sum(prop.table(table(sales\_1$demand.veggie,sales\_1$wday))[,3])

sum(prop.table(table(sales\_1$demand.veggie,sales\_1$wday))[,4])

sum(prop.table(table(sales\_1$demand.veggie,sales\_1$wday))[,5])

sum(prop.table(table(sales\_1$demand.veggie,sales\_1$wday))[,6])

#To get the conditional probability by day (the following command will get the probability of turkey demand, given the day as friday):

#P(turkey.demand = x | day = FRIDAY) = P(DAY=FRIDAY AND turkey.demand = x) / P(DAY = FRIDAY)

#where X is the demand number. For instance, if X = 17, then P(DAY = FRIDAY AND X=17) = 0.015384615

#P(FRIDAY) = 0.2. Therefore P(demand.ham = 17 | FRIDAY) = 0.015384615/0.2

prop.table(table(sales\_1$demand.veggie,sales\_1$wday))[,6]/0.2

#To find all the conditional probabilities, cond on week day:

prop.table(table(sales\_1$demand.veggie,sales\_1$wday))/0.2

data.frame((prop.table(table(sales\_1$demand.veggie,sales\_1$wday))/0.2))

#To find marginal probabilities, use the following command. This will find the marginal probabilities horizintally for

#prop.table(table(sales\_1$demand.turkey,sales\_1$wday))

margin.table(prop.table(table(sales\_1$demand.veggie,sales\_1$wday)),1)

#The following command will find the marginal probabilities vertically

margin.table(prop.table(table(sales\_1$demand.veggie,sales\_1$wday)),2)

#Let us convert the following command's o/p to dataframe:

#prop.table(table(sales\_1$demand.turkey,sales\_1$wday))/0.2

#NOTE That the prob column in the following column, is the conditional probability of demand, given a week day.

week\_day\_prob <- data.frame((prop.table(table(sales\_1$demand.veggie,sales\_1$wday))/0.2),stringsAsFactors=F)

#Rename the column names

names(week\_day\_prob) <- c("demand","week\_day","prob")

#To verify, we have conditional probabilities based on a week day

sum(week\_day\_prob[week\_day\_prob$week\_day == "Mon",]$prob)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Tues",]$prob)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Wed",]$prob)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Thurs",]$prob)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Fri",]$prob)

#Let us add another variable expected\_val

# problem here...but resolved by using "levels" in as.numeric

week\_day\_prob$expected\_val <- ( as.numeric(levels(week\_day\_prob$demand)) \* (week\_day\_prob$prob))

#Number of ham burgers needed on a week day

sum(week\_day\_prob[week\_day\_prob$week\_day == "Mon",]$expected\_val)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Tues",]$expected\_val)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Wed",]$expected\_val)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Thurs",]$expected\_val)

sum(week\_day\_prob[week\_day\_prob$week\_day == "Fri",]$expected\_val)

#Round

round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Mon",]$expected\_val))

round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Tues",]$expected\_val))

round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Wed",]$expected\_val))

round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Thurs",]$expected\_val))

round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Fri",]$expected\_val))

#######################################################

##Now finding the probability by ignoring the week day#

#######################################################

data.frame(table(sales\_1$demand.veggie))

all\_prob <- data.frame(prop.table(table(sales\_1$demand.veggie)))

names(all\_prob) <- c("demand","freq")

sum(as.numeric(levels(all\_prob$demand)) \* all\_prob$freq)

#######################################################

#Let us evaluate the profit

#######################################################

#Adding a new variable

#new\_avail\_cond\_by\_wday

sales\_1$veggie.new\_avail.cond\_prb <- 0

sales\_1$veggie.new\_avail.cond\_prb[sales\_1$wday=="Mon"] <- round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Mon",]$expected\_val))

sales\_1$veggie.new\_avail.cond\_prb[sales\_1$wday=="Tues"] <- round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Tues",]$expected\_val))

sales\_1$veggie.new\_avail.cond\_prb[sales\_1$wday=="Wed"] <- round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Wed",]$expected\_val))

sales\_1$veggie.new\_avail.cond\_prb[sales\_1$wday=="Thurs"] <- round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Thurs",]$expected\_val))

sales\_1$veggie.new\_avail.cond\_prb[sales\_1$wday=="Fri"] <- round(sum(week\_day\_prob[week\_day\_prob$week\_day == "Fri",]$expected\_val))

sales\_1$veggie.cond\_prob\_profit <- (ifelse((sales\_1$demand.veggie <= sales\_1$veggie.new\_avail.cond\_prb), ((sales\_1$demand.veggie \* 5) -

(sales\_1$veggie.new\_avail.cond\_prb \* 2.5)),

((sales\_1$veggie.new\_avail.cond\_prb \* 5) - (sales\_1$veggie.new\_avail.cond\_prb \* 2.5)))

)

print(sum(sales\_1$veggie.cond\_prob\_profit))

sales\_1$veggie.actual <- (ifelse((sales\_1$demand.veggie <= sales\_1$available.veggie), ((sales\_1$demand.veggie \* 5) - (sales\_1$available.veggie \* 2.5)),

((sales\_1$available.veggie \* 5) - (sales\_1$available.veggie \* 2.5)))

)

print(sum(sales\_1$veggie.actual))

#Making all the days as 13, and seeing the profit (non-cond prob)

sum(ifelse((sales\_1$demand.veggie <= 13), ((sales\_1$demand.veggie \* 5) - (13 \* 2.5)), ((13 \* 5) - (13 \* 2.5)))

)