USE CASE STUDY REPORT

Group No.: Group 07

Student Names: Rohan Singh

Predicting Import Expenditure using R

Executive Summary:

U.S. consumers demand variety, quality, and convenience in the foods they consume. As Americans have become wealthier and more ethnically diverse, there is an increase in the import of tropical products, spices, and gourmet products. Seasonal and climatic factors drive U.S. imports of popular types of fruits and vegetables and tropical products, such as cocoa and coffee. In addition, a growing share of U.S. imports can be attributed to intra-industry trade, whereby agricultural-processing industries based in the United States carry out certain processing steps offshore and import products at different levels of processing from their subsidiaries in foreign markets.

The population of United States is expected to grow to over 438 million by 2050. The rapid growth of U.S. agricultural imports relative to exports in recent years may come as a surprise to many because the U.S. is still the world's leading exporter of farm products. In fact, U.S. agricultural exports grew by almost $3 billion in 2003. And, higher commodity prices point to export gains in 2004. But the U.S. is also the world's largest agricultural importer. Over the last 7 years, U.S. agricultural imports have increased by more than $13 billion, from $32 billion in 1996 to $46 billion in 2003. If these trends continue, the current agricultural trade surplus will turn into a deficit toward the end of the decade.

The data which we are working on has been obtained from The United States Department of Agriculture. Our dataset contains about 36 variables which mainly include the various commodities which are being imported like processed foods, wine & beer, fresh vegetables etc. We'll be predicting the trends for 6 randomly selected food items.

As the data has been acquired from an official government website its reliable, it did contain missing values which were replaced, also as the data needed for processing was purely numerical so we didn't have to deal with any categorical values. After cleaning the dataset and making it complete and free from any redundancies it helped us get optimal results and accurate predictions.

The data mining techniques which were used included linear regression and K nearest neighbor, both the techniques were used to process the data and the results obtained were optimal and the predictions were accurate.

Data Description:
Variables Included:

Agricultural Products	Wine and Beer	Tobacco	Live Animals	Rice	Spices
Red Meat	Fresh Vegetables	Misc. Consumer Products	Fresh Fruit	Snack Foods	Processed Fruits and Vegetables
Nursery Products	Bananas and Plantains	Dairy Products	Tree Nuts	Cheese	Fruit/Vegetable Juices
Roasted/Instant Coffee	Vegetable Oils	Tropical Oils	Cocoa Paste/Butter	Planting Seeds	Essential Oils
Feed and Fodder	Sugar/Sweeteners	Hides/Skins	Coffee(Unroasted)	Rubber Products	Raw Beet/Cane Sugar
Cocoa Beans	Wheat	Tea(Herbal	Coarse Grains	Red Meat(Processed)	

I. Background and Introduction

Importing helps national economies grow and flourish. The United States imports about $147 billion worth of food, feeds and beverages. For consumers, the chief advantages of the import boom are the increased availability and variety of fresh produce, particularly in winter. The United States imports more than it exports as a result it as accumulated a trade deficit of about $621 Billion. As the United States population is on the rise and the world getting more and more connected via trading relations, the import/export business is flourishing.

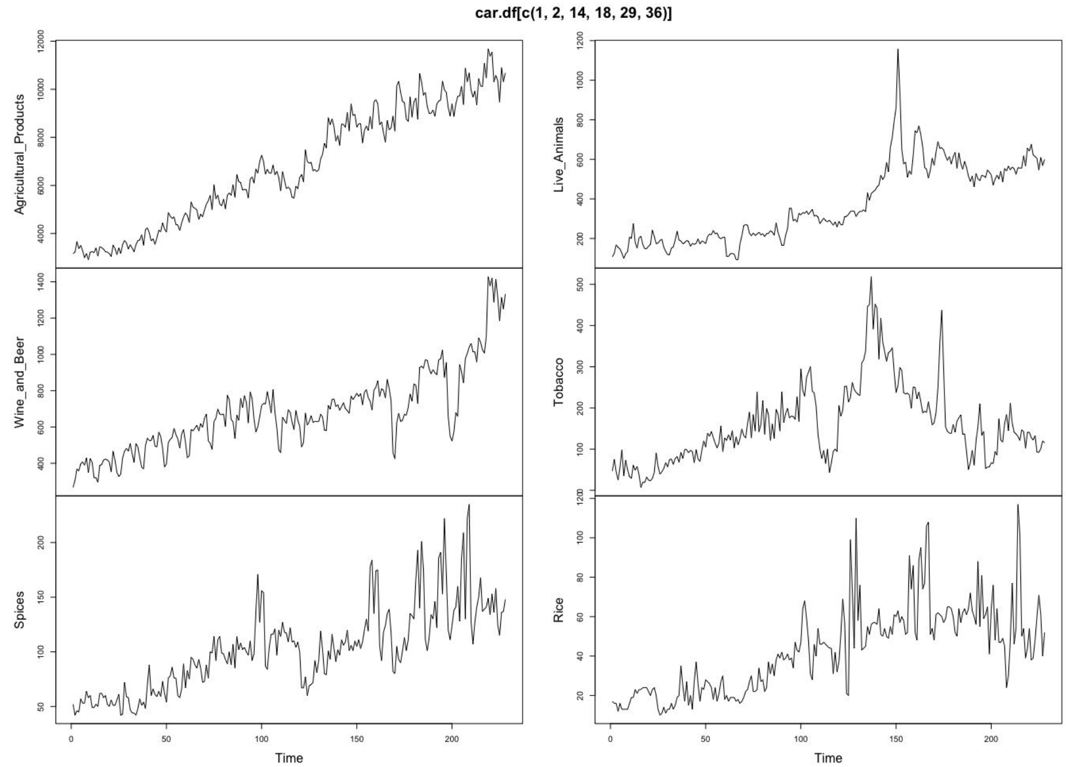
Through this case study we'll be studying the steady increase in the imports of food items and its effects on the United States economy. This case study will provide information regarding possible future expenditure on food imports which can help to make necessary changes in the annual budget of the United States Government relating to imports for a particular year.

Our goal is to predict the total expenditure on food imports by the United States in the next 1 year and map the increase or decrease in the demands of certain food items being currently imported by the government.

We first organize the data into an understandable form which can be worked on and which can be modified. We then use data mining techniques to further make sense of the available data and to understand the relation between the factors. We then apply 2 data mining models to predict the expenditure on imports as well as know increase or decrease in the demands of food items being imported.

II. Data Exploration and Visualization

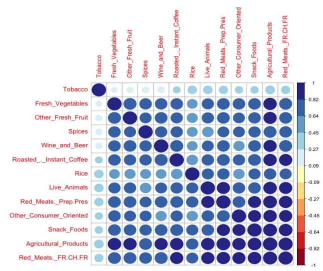
Time Series Analysis of the Dataset(No predictions):



Time series analysis is performed to extract meaningful statistics and other characteristics of the data. Some distinguishable patterns appear when we plot the data. We performed time series analysis on the 6 variables which were chosen and observed a steady increase in the demands of a few while an irregular pattern in others. From the plot we can see that since the year 2000 there has been an increase in the expenditure on imports with a low observed for a few years.

Correlation Analysis:

Correlation analysis is used to study the strength of two or more numerically measured continuous variables. We performed correlation analysis on a few variables and found out the relation between them with the increase in expenditure on one commodity there is an increase/decrease in the expenditure on another commodity. Few variables are highly correlated.



III. Data Preparation and Preprocessing

The data was obtained from an official government website which makes it reliable. The data was neatly categorized with respect to the type of food item along with the month it was imported, and the total cost of that food item. The dataset contains data of food items imported from the year 2000 till the year 2018 and top 36 food categories that are imported by the Unites States.

The data had to be organized and compiled also it had a few redundancies and missing values. To correct the data, we identified special characters and removed them. The missing values were replaced with respective column mean. Redundant rows and columns were identified and removed.

The dataset has about 36 variables out of which we chose 6 of the variables to work on.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1			Agricultural_Products	Wine_and_Beer	Spices	Live_Animals	Tobacco	Rice	Red_Meats,_FR/CH/FR	sh_Vegeta	onsumer_(	er_Fresh_F	nack_Food	Fruit_&_\	as_and_Pl	sery_Prod	Dairy_Prc
2	January		3171	267	52	107	47	17	246	241	199	207	169	191	91	103	75
3	February		3229	307	42	125	75	16	236	223	207	232	169	190	80	124	71
4	March		3667	368	46	168	45	16	299	234	219	257	191	210	91	83	70
5	April		3372	359	45	155	25	12	276	230	197	210	160	187	97	120	62
6	May		3504	397	57	147	56	16	304	184	220	197	187	204	109	132	64
7	June		3283	407	53	125	98	13	311	147	215	135	187	199	100	68	86
8	July		2992	390	53	99	35	13	296	128	211	87	199	197	102	66	76
9	August		3170	430	64	123	73	13	310	117	220	85	230	189	86	97	80
10	September		2913	348	58	134	50	13	254	113	205	74	254	186	81	85	60
11	October		3224	426	60	207	33	16	255	135	227	79	273	210	98	90	68
12	November		3247	405	49	200	29	19	271	180	243	143	250	209	84	109	77
13	December		3203	320	49	275	61	19	253	224	218	216	197	196	83	82	78
14	January		3403	321	52	180	49	23	218	194	320	188	246	307	91	101	82
15	February		3060	295	50	151	58	21	201	180	267	171	214	269	85	126	65
16	March		3452	388	62	202	38	23	237	204	316	191	219	309	95	86	58
17	April		3413	393	62	212	7	23	236	191	302	186	234	261	104	124	56
18	May		3348	418	58	169	20	24	243	200	302	202	170	215	105	136	85
19	June		3244	423	60	148	19	24	236	191	348	199	220	161	96	63	101
20	July		3224	418	52	149	32	24	240	201	358	220	123	145	101	57	76
21	August		3168	407	50	161	25	22	248	204	344	248	93	139	87	92	90
22	September		3042	352	56	170	23	20	260	212	308	287	67	122	94	86	76
23	October		3516	467	51	242	29	23	261	241	343	322	93	147	96	91	79
24	November		3374	417	51	211	43	24	255	224	298	267	135	180	91	108	77
25	December		3122	350	56	174	91	20	230	213	211	209	188	186	99	78	75
26	January		3404	327	61	177	58	13	244	321	269	200	277	214	88	95	82
27	February		3166	340	42	190	39	10	231	243	258	191	268	205	87	113	55
28	March		3540	429	43	196	45	11	254	316	269	201	286	224	103	102	70
29	April		3706	466	72	156	53	14	275	350	257	214	255	228	103	107	74

III. Data Mining Techniques and Implementation:

Techniques Used: K-Nearest Neighbor and Multiple Linear Regression

Multiple Linear Regression: We used multiple linear regression as MLR allows us to make predictions about one variable based on the information known about another variable. As we are predicting the import trends on food items based on the history of

imports by the United States Government multiple linear regression will give the optimal results.
K-Nearest Neighbor: KNN is a non-parametric, Lazy learning algorithm. KNN is a non-parametric, lazy learning algorithm. Its purpose is to use a database in which the data points are separated into several classes to predict the classification of a new sample point

IV. Performance Evaluation

#Data output comparing Test and Train

```
> data.frame("predicted" = car.lm.pred[1:12], "actual" = valid.df$Wine_and_Beer[1:12])
    predicted actual
31        519    509
34        506    509
35        482    490
48        468    483
66        600    610
69        593    597
77        719    699
80        675    671
83        759    750
98        603    601
104       731    741
111       681    654
> data.frame("predicted" = caragri.lm.pred[1:12], "actual" =
valid.df$Agricultural_Products[1:12])
    predicted actual
31       3522   3534
34       3536   3534
35       3710   3703
48       4462   4444
66       4988   4979
69       4697   4693
77       5581   5600
80       5419   5426
83       5673   5687
98       6520   6516
104      6526   6520
111      6545   6574
> data.frame("predicted" = cartob.lm.pred[1:12], "actual" = valid.df$Tobacco[1:12])
    predicted actual
31         83     74
34         74     77
35         72     82
48         79     97
66        124    130
69        145    149
77        248    239
```

```
80       222   218
83       193   183
98       221   227
104       282   287
111       106    77
> data.frame("predicted" = carspices.lm.pred[1:12], "actual" = valid.df$Spices[1:12])
   predicted actual
31         57     47
34         46     42
35         45     48
48         63     73
66         83     85
69         75     72
77        121    111
80         98     98
83        109    100
98        154    171
104       100    104
111       141    127
> data.frame("predicted" = carani.lm.pred[1:12], "actual" =
valid.df$Live_Animals[1:12])
   predicted actual
31        131    119
34        193    192
35        228    237
48        170    187
66         87     93
69        208    214
77        241    222
80        230    224
83        239    226
98        280    283
104       314    322
111       315    289
> data.frame("predicted" = carrice.lm.pred[1:12], "actual" = valid.df$Rice[1:12])
   predicted actual
31         24     13
34         19     15
35         18     19
48         18     24
66         19     17
69         19     17
77         36     23
80         36     28
83         39     37
98         42     43
```

```
104        50    48
111        53    46
```

> #Accuracy Check

```
> accuracy(car.lm.pred, valid.df$Wine_and_Beer)
         ME RMSE MAE MPE MAPE
Test set -2   11   8  -0    1
> accuracy(carspices.lm.pred, valid.df$Spices)
         ME RMSE MAE MPE MAPE
Test set -0    7   6  -1    6
> accuracy(carani.lm.pred, valid.df$Live_Animals)
         ME RMSE MAE MPE MAPE
Test set -1   11   9  -1    4
> accuracy(cartob.lm.pred, valid.df$Tobacco)
         ME RMSE MAE MPE MAPE
Test set -1   10   7  -2    7
> accuracy(carrice.lm.pred, valid.df$Rice)
         ME RMSE MAE MPE MAPE
Test set -2    6   5 -10   16
> accuracy(caragri.lm.pred, valid.df$Agricultural_Products)
         ME RMSE MAE MPE MAPE
Test set  2   11   9   0    0
```

Rohan [6:52 PM]
Values obtained From KNN

```
> pred$prediction
Time Series:
Start = 181
End = 192
Frequency = 1
[1] 620 724 706 855 372 866 708 628 706 672 708 465
> predagri$prediction
Time Series:
Start = 181
End = 192
Frequency = 1
[1] 3316 7104 4408 6868 7264 6736 4714 4545 5948 5710 9447 6394
> predtob$prediction
Time Series:
Start = 181
End = 192
Frequency = 1
[1] 141 180 108 114 192 137  71 226 166 292 219 214
> predrice$prediction
Time Series:
Start = 181
End = 192
```

```
Frequency = 1
[1] 38 19 64 42 68 62 20 62 84 40 59 43
> predspices$prediction
Time Series:
Start = 181
End = 192
Frequency = 1
[1]  56  72  86 122 101 126 126 109 103  77 112  86
> predani$prediction
Time Series:
Start = 181
End = 192
Frequency = 1
[1] 147 332 223 335 446 384 288 146 280 271 520 401
```

Results:

carforcast.lm.pred

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Feb 228	572	290	854	141	1004
Mar 228	707	425	990	276	1139
Apr 228	703	421	985	271	1134
May 228	728	446	1010	297	1160
Jun 228	712	429	994	280	1143
Jul 228	713	431	995	282	1145
Aug 228	694	412	976	263	1126
Sep 228	653	370	935	221	1084
Oct 228	744	462	1026	312	1175
Nov 228	721	438	1003	289	1152
Dec 228	689	407	972	258	1121
Jan 229	585	303	867	153	1017

> caragrifor.lm.pred

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Feb 228	6577	3449	9706	1791	11363
Mar 228	7576	4447	10704	2790	12362
Apr 228	7366	4237	10495	2580	12152
May 228	7290	4161	10418	2504	12076
Jun 228	6946	3817	10074	2160	11732
Jul 228	6818	3689	9946	2032	11604
Aug 228	6773	3644	9901	1987	11559
Sep 228	6502	3373	9631	1716	11288
Oct 228	7067	3938	10196	2281	11853
Nov 228	7000	3871	10129	2214	11786
Dec 228	7110	3981	10239	2324	11896
Jan 229	7011	3883	10140	2225	11798

> cartobfor.lm.pred

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Feb 228	158	36	281	-29	346

```
Mar 228        174   52  297  -13  362
Apr 228        184   61  306   -4  371
May 228        176   53  299  -12  363
Jun 228        179   57  302   -8  367
Jul 228        165   42  287  -23  352
Aug 228        159   36  281  -29  346
Sep 228        152   29  274  -36  339
Oct 228        156   34  279  -31  344
Nov 228        148   25  271  -40  335
Dec 228        169   47  292  -18  357
Jan 229        159   36  281  -29  346
> carspicesfor.lm.pred
         Point Forecast Lo 80 Hi 80 Lo 95 Hi 95
Feb 228        110   61  159   35  185
Mar 228        106   57  155   31  181
Apr 228        118   69  168   43  194
May 228        120   71  169   45  195
Jun 228         97   48  146   22  172
Jul 228         95   46  144   20  170
Aug 228         97   48  146   22  172
Sep 228         99   50  148   24  174
Oct 228        105   55  154   29  180
Nov 228        111   62  160   36  186
Dec 228        102   53  151   27  177
Jan 229        110   61  159   35  185
> caranifor.lm.pred
         Point Forecast Lo 80 Hi 80 Lo 95 Hi 95
Feb 228        345   93  597  -41  730
Mar 228        396  144  648   11  781
Apr 228        398  146  650   12  783
May 228        398  146  649   12  783
Jun 228        398  147  650   13  784
Jul 228        408  157  660   23  794
Aug 228        398  146  650   13  783
Sep 228        371  119  623  -15  756
Oct 228        399  147  651   14  785
Nov 228        387  135  639    2  773
Dec 228        392  140  644    6  777
Jan 229        367  115  619  -18  753
> carricefor.lm.pred
         Point Forecast Lo 80 Hi 80 Lo 95 Hi 95
Feb 228         41   13   69   -2   84
Mar 228         45   17   73    2   88
Apr 228         39   11   67   -3   82
May 228         40   12   68   -3   83
Jun 228         48   20   76    5   91
```

```
Jul 228        47   19   75    4   89
Aug 228        44   16   72    1   87
Sep 228        49   21   77    6   92
Oct 228        49   21   77    6   91
Nov 228        50   22   78    7   92
Dec 228        41   13   69   -2   83
Jan 229        46   18   74    3   88
```

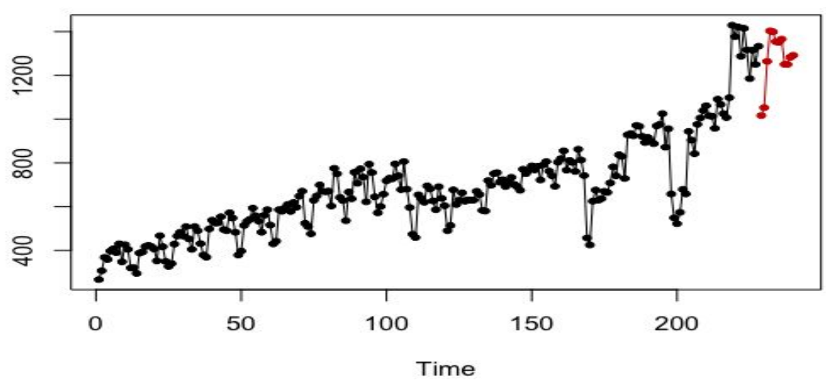
Prediction from KNN

```
> pred$prediction
Time Series:
Start = 229
End = 240
Frequency = 1
[1] 1016 1052 1264 1403 1399 1354 1351 1366 1250 1250 1282 1292
> predagri$prediction
Time Series:
Start = 229
End = 240
Frequency = 1
[1] 10612 10284 11384 10918 11616 10842 11062 10336 10020 10632  9889 10792
> predtob$prediction
Time Series:
Start = 229
End = 240
Frequency = 1
[1] 148 158 144 187 136 212 143 200 180 148 208 160
> predrice$prediction
Time Series:
Start = 229
End = 240
Frequency = 1
[1] 41 48 44 54 48 54 54 54 51 52 51 41
> predspices$prediction
Time Series:
Start = 229
End = 240
Frequency = 1
[1] 124 110  96  90  86  80  84  80  91  86  92 114
> predani$prediction
Time Series:
Start = 229
End = 240
Frequency = 1
[1] 595 594 606 564 570 574 538 502 502 488 495 516
```

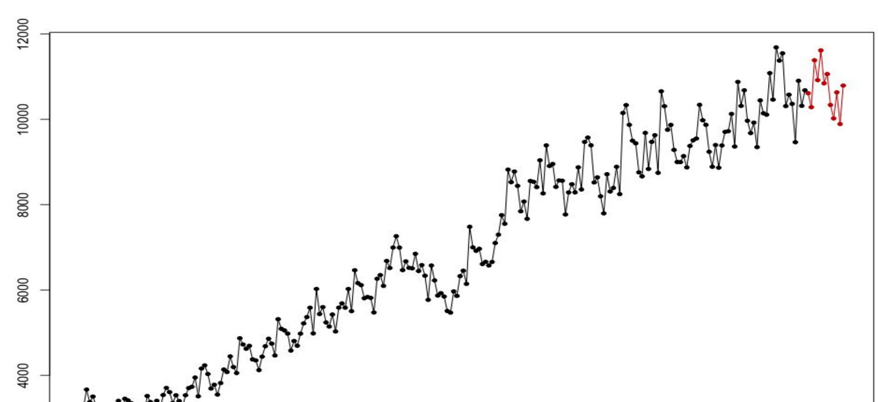
PLOTS:

X-Axis: Number of Months
Y-Axis: Millions of Dollars
Red Points indicate predicted values

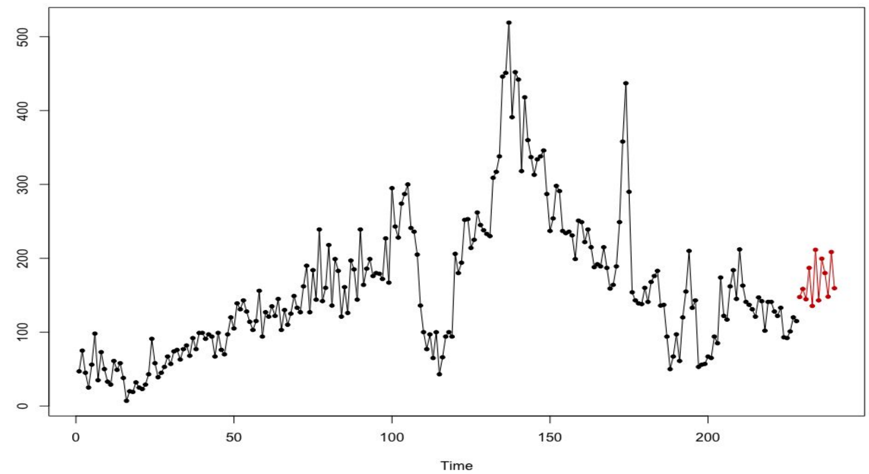
Beer and Wine:



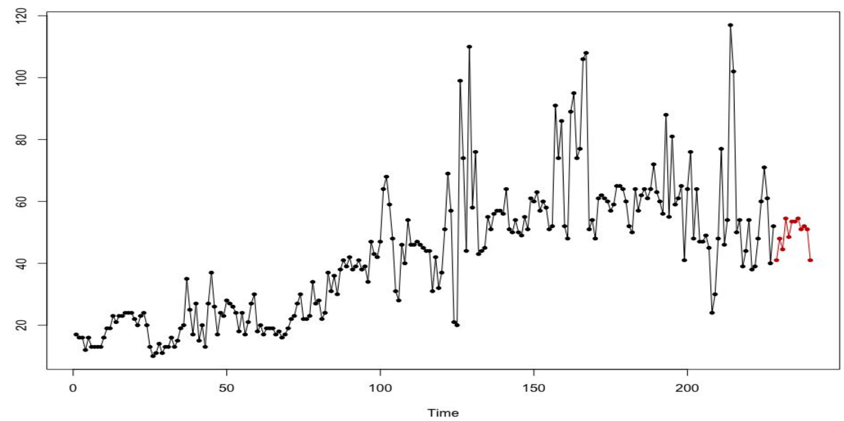
Agriculture Products:



Tobacco:

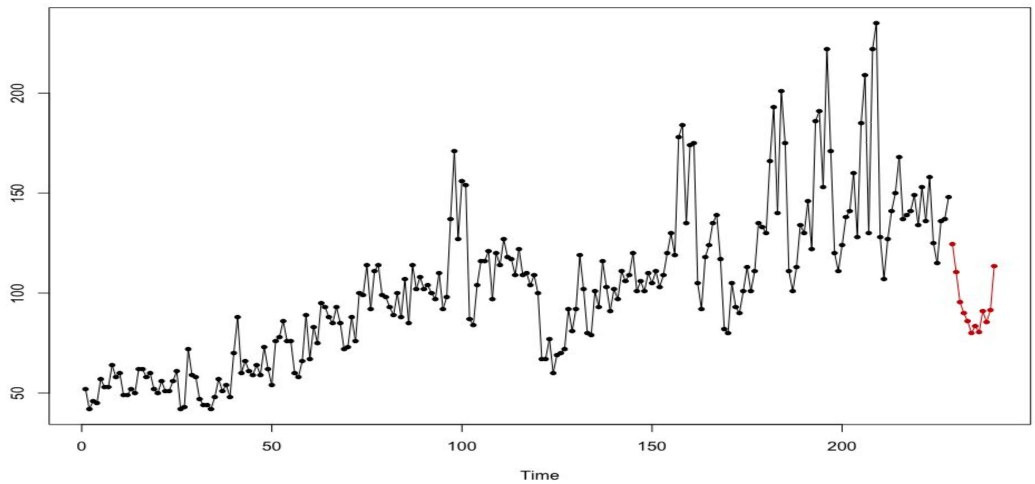


Rice:

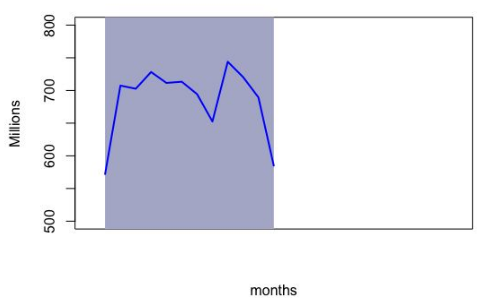


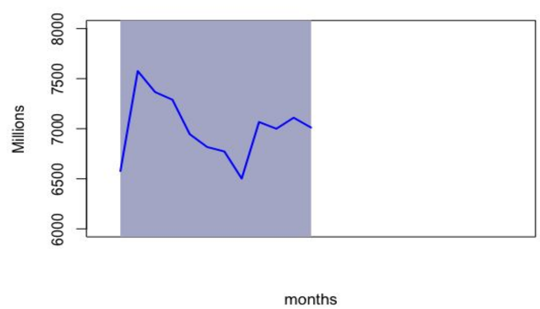
Live Animals:

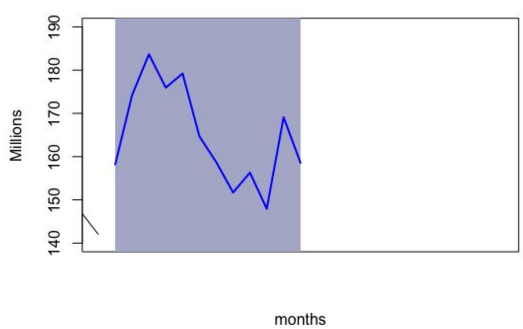
Spices:

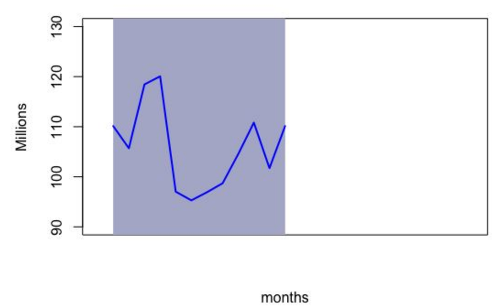


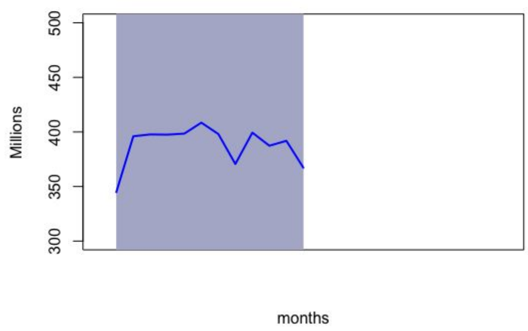
Linear Regression Models:
X-Axis: Number of Months
Y-Axis: Millions of Dollars

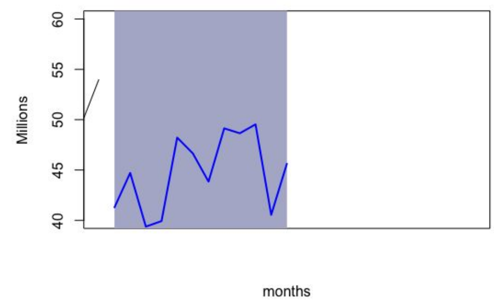












VI. Discussion and Recommendation

The data for the following project could be more lucrative. Providing us with information about the rate at which a item is brought could help us calculate the quantity of the product.
Providing with the country from which the product is brought could help us get an idea about the dependency for an item on a country.

VII. Summary

This data set provides import values of edible products (food and beverages) entering U.S. ports and their origin of shipment. Data are from the U.S. Department of Commerce, U.S. Census Bureau. Food and beverage import values are compiled by calendar year into food groups corresponding to major commodities or level of processing. At least 18 years of annual data are included, enabling users to track long-term patterns. In order to predict the future import trends, we used multiple linear regression as well as K-Nearest Neighbor. In multiple linear regression we used seasonality and trends in order to get optimal results and obtained plots for the same.

Appendix: R Code for use case study

```
library(ggplot2)
library(forecast)
library(tsfknn)
library(graphics)
library(grDevices)
library(ISLR)
library(lattice)
library(plyr)
library(readxl)
library(stats)
library(zoo)
library(corrplot)
library(RColorBrewer)
```

```
setwd("~/Desktop/R_Final")

car.df <- read.csv("data_ar.csv")

#View(car.df)

car.df <- car.df[,-c(1,2)]

#View(car.df)

#Original Time Series Data Plotting
plot.ts(car.df[c(1,2,14,18,29,36)],  plot.type = "multiple")

#Correlation Analysis
#For General Selected variables
M <- cor(car.df[c(1,2,14,18,29,36,3,4,5,6,7,16,17)])
corrplot(M, type="full", order="hclust", col=brewer.pal(n=10, name="RdYlBu"), diag =
TRUE)

#For Selected variables
M <- cor(car.df[c(1,2,14,18,29,36)])
corrplot(M, type="upper", order="hclust", col=brewer.pal(n=10, name="RdYlBu"))


#Multiple Linear Regression
#We just selected the first 200 rows for calculating the accuracy
car1.df <- car.df[1:200,]

selected.var <- c(1:36)

train.index <- sample(c(1:200), 180)

train.df <- car1.df[train.index, selected.var]

valid.df <- car1.df[-train.index, selected.var]

#Linear Regression for Wine and Beer.

car.lm <- lm(Wine_and_Beer ~., data = car1.df)

car.lm.pred <- predict(car.lm, valid.df)

options(scipen = 999, digits = 0)
```

```
accuracy(car.lm.pred, valid.df$Wine_and_Beer)

#LINEAR REGRESSION FOR AGRICULTURAL PRODUCTS

caragri.lm <- lm(Agricultural_Products ~., data = car1.df)

caragri.lm.pred <- predict(caragri.lm, valid.df)

options(scipen = 999, digits = 0)

accuracy(caragri.lm.pred, valid.df$Agricultural_Products)

#Linear Regression for Rice

carrice.lm <- lm(Rice ~., data = car1.df)

carrice.lm.pred <- predict(carrice.lm, valid.df)

options(scipen = 999, digits = 0)

accuracy(carrice.lm.pred, valid.df$Rice)

#Linear Regression for Tobacco

cartob.lm <- lm(Tobacco ~., data = car1.df)

cartob.lm.pred <- predict(cartob.lm, valid.df)

options(scipen = 999, digits = 0)

accuracy(cartob.lm.pred, valid.df$Tobacco)

#Linear Regression for Live Animals

carani.lm <- lm(Live_Animals ~., data = car1.df)

carani.lm.pred <- predict(carani.lm, valid.df)

options(scipen = 999, digits = 0)

accuracy(carani.lm.pred, valid.df$Live_Animals)

#Linear Regression for Spices
```

```
carspices.lm <- lm(Spices ~., data = car1.df)

carspices.lm.pred <- predict(carspices.lm, valid.df)

options(scipen = 999, digits = 0)

accuracy(carspices.lm.pred, valid.df$Spices)

#Data output comparing Test and Train
data.frame("predicted" = car.lm.pred[1:12], "actual" = valid.df$Wine_and_Beer[1:12])

data.frame("predicted" = caragri.lm.pred[1:12], "actual" =
valid.df$Agricultural_Products[1:12])

data.frame("predicted" = cartob.lm.pred[1:12], "actual" = valid.df$Tobacco[1:12])

data.frame("predicted" = carspices.lm.pred[1:12], "actual" = valid.df$Spices[1:12])

data.frame("predicted" = carani.lm.pred[1:12], "actual" = valid.df$Live_Animals[1:12])

data.frame("predicted" = carrice.lm.pred[1:12], "actual" = valid.df$Rice[1:12])

#Accuracy Check
accuracy(car.lm.pred, valid.df$Wine_and_Beer)

accuracy(carspices.lm.pred, valid.df$Spices)

accuracy(carani.lm.pred, valid.df$Live_Animals)

accuracy(cartob.lm.pred, valid.df$Tobacco)

accuracy(carrice.lm.pred, valid.df$Rice)

accuracy(caragri.lm.pred, valid.df$Agricultural_Products)

#time series forecasting using KNN
library(tsfknn)
predagri <- knn_forecasting(train.df$Agricultural_Products, h = 12, lags = 1:12, k = 2,
msas = "MIMO")

pred <- knn_forecasting(train.df$Wine_and_Beer, h = 12, lags = 1:12, k = 2, msas =
"MIMO")
```

```
predtob <- knn_forecasting(train.df$Tobacco, h = 12, lags = 1:12, k = 2, msas =
"MIMO")

predrice <- knn_forecasting(train.df$Rice, h = 12, lags = 1:12, k = 2, msas = "MIMO")

predspices <- knn_forecasting(train.df$Spices, h = 12, lags = 1:12, k = 2, msas =
"MIMO")

predani <- knn_forecasting(train.df$Live_Animals, h = 12, lags = 1:12, k = 2, msas =
"MIMO")

#Display the Estimates  values
pred$prediction

predagri$prediction

predtob$prediction

predrice$prediction

predspices$prediction

predani$prediction


#Final calculation
car.ts <- ts(car.df$Wine_and_Beer, start = c(1), end = c(228), frequency = 12)
carfor.lm <- tslm( car.ts ~ trend + I(trend^2) + season)
carforcast.lm.pred <- forecast(carfor.lm, h= 12)


caragri.ts <- ts(car.df$Agricultural_Products, start = c(1), end = c(228), frequency = 12)
caragrifor.lm <- tslm( caragri.ts ~ trend + I(trend^2) + season)
caragrifor.lm.pred <- forecast(caragrifor.lm, h= 12)


cartob.ts <- ts(car.df$Tobacco, start = c(1), end = c(228), frequency = 12)
cartobfor.lm <- tslm( cartob.ts ~ trend + I(trend^2) + season)
cartobfor.lm.pred <- forecast(cartobfor.lm, h= 12)


carspices.ts <- ts(car.df$Spices, start = c(1), end = c(228), frequency = 12)
carspicesfor.lm <- tslm( carspices.ts ~ trend + I(trend^2) + season)
carspicesfor.lm.pred <- forecast(carspicesfor.lm, h= 12)
```

```
carani.ts <- ts(car.df$Live_Animals, start = c(1), end = c(228), frequency = 12)
caranifor.lm <- tslm( carani.ts ~ trend + I(trend^2) + season)
caranifor.lm.pred <- forecast(caranifor.lm, h= 12)


carrice.ts <- ts(car.df$Rice, start = c(1), end = c(228), frequency = 12)
carricefor.lm <- tslm( carrice.ts ~ trend + I(trend^2) + season)
carricefor.lm.pred <- forecast(carricefor.lm, h= 12)

#Final Caluclation for KNN
predagri <- knn_forecasting(car.df$Agricultural_Products, h = 12, lags = 1:12, k = 2,
msas = "MIMO")

pred <- knn_forecasting(car.df$Wine_and_Beer, h = 12, lags = 1:12, k = 2, msas =
"MIMO")

predtob <- knn_forecasting(car.df$Tobacco, h = 12, lags = 1:12, k = 2, msas = "MIMO")

predrice <- knn_forecasting(car.df$Rice, h = 12, lags = 1:12, k = 2, msas = "MIMO")

predspices <- knn_forecasting(car.df$Spices, h = 12, lags = 1:12, k = 2, msas = "MIMO")

predani <- knn_forecasting(car.df$Live_Animals, h = 12, lags = 1:12, k = 2, msas =
"MIMO")

#Display the Estimates  values

#Display Prediction from Linear Regression Forecast
carforcast.lm.pred
caragrifor.lm.pred
cartobfor.lm.pred
carspicesfor.lm.pred
caranifor.lm.pred
carricefor.lm.pred

#Display prediction Values from KNN

pred$prediction

predagri$prediction

predtob$prediction
```

```
predrice$prediction

predspices$prediction

predani$prediction

#visualization

#By KNN
plot(pred)

plot(predagri)

plot(predtob)

plot(predrice)

plot(predani)

plot(predspices)

#by Linear Regression

plot(carforcast.lm.pred, ylim = c(500, 800) ,ylab = "Millions", xlab = "months", bty= "o",
xaxt = "n", xlim = c(228,230), main = "", include = 13)

plot(caragrifor.lm.pred, ylim = c(6000, 8000), ylab = "Millions", xlab = "months", bty=
"o", xaxt = "n", xlim = c(228,230), main = "", include = 13)

plot(cartobfor.lm.pred, ylim = c(140, 190), ylab = "Millions", xlab = "months", bty= "o",
xaxt = "n", xlim = c(228,230), main = "", include = 13)

plot(carspicesfor.lm.pred, ylim = c(90,130), ylab = "Millions", xlab = "months", bty=
"o", xaxt = "n", xlim = c(228,230), main = "", include = 13)

plot(caranifor.lm.pred, ylim = c(300, 500), ylab = "Millions", xlab = "months", bty= "o",
xaxt = "n", xlim = c(228,230), main = "", include = 13)

plot(carricefor.lm.pred, ylim = c(40, 60), ylab = "Millions", xlab = "months", bty= "o",
xaxt = "n", xlim = c(228,230), main = "", include = 13)
```