McDonald's – Analysis on Calorie and Nutritional Value

Prepared by: **Haritha Krishnankutty Nair**

CONTENTS

Cross-Industry Standard Process (CRISP-DM)

Problem Statement (Business Understanding)

Relevant Data (Data Understanding)

Data Preprocessing & Data Quality (Data Preparation)

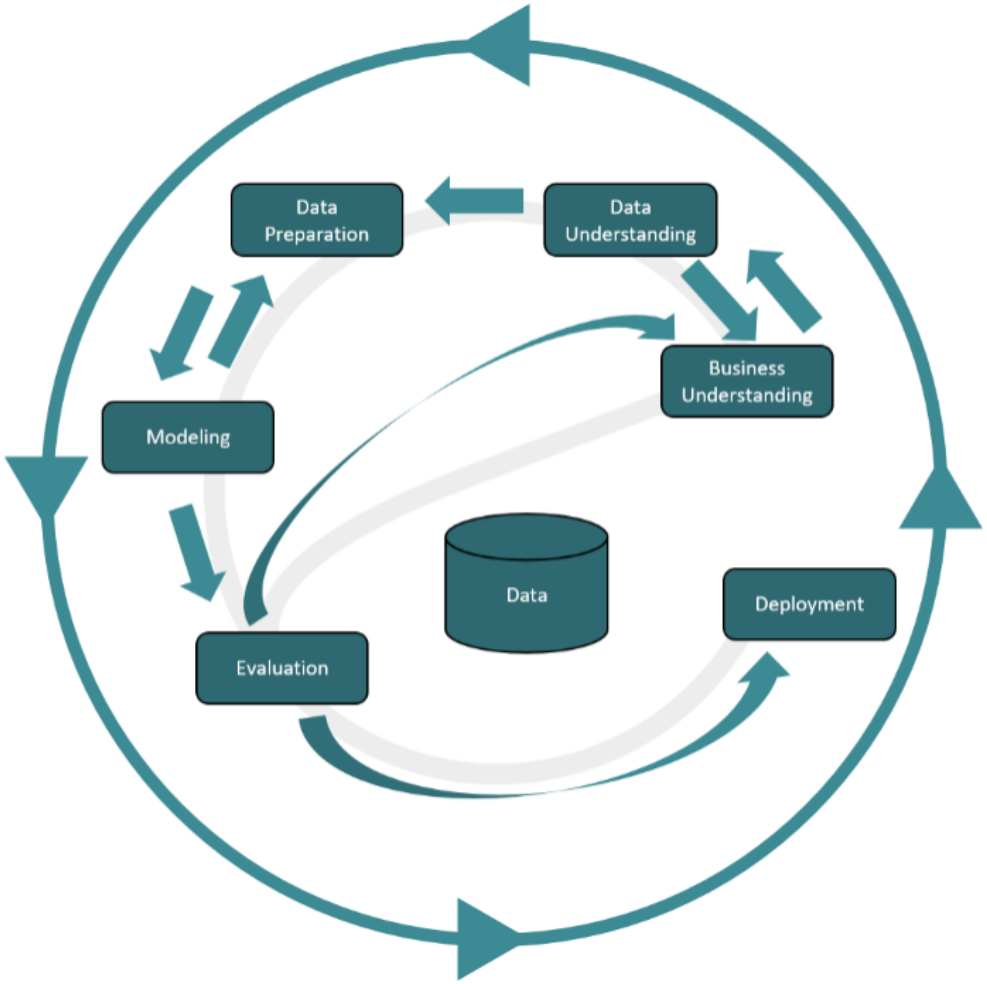
Methodologies/Algorithm’s (Modeling & Execution)

Data Insights (Visualization)

Conclusion and Analysis (Evaluation & Deployment)

Software Packages

References



CROSS-INDUSTRY STANDARD PROCESS (CRISP-DM)

This project primarily focuses on CRISP-DM methodology for creating data insights. The project will be executed in six steps and the methodologies used are as follows:

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

PROBLEM STATEMENT – BUSINESS UNDERSTANDING

- Imagine how convenient it would be if the customers could walk up to a fast food joint and order the appropriate number of items from the menu to meet their nutritional requirements
- This project mainly focuses on analyzing the nutritional value of every item on the US McDonald's menu and thereby helping customers make smart purchases by providing a comparative analysis of the calories and nutrition associated with each menu items that they order

RELEVANT DATA – DATA UNDERSTANDING

- The sample dataset used for analyzing the calorie and nutritional value of the food from the McDonald's menu is obtained from the 'Nutritional Facts for McDonald's Menu' in https://www.kaggle.com/mcdonalds/nutrition-facts

- The dataset consists of the following major attributes:
  - Category: This column gives details on the category under which the menu falls like breakfast, beverages, salads, etc.
  - Item: The item attribute includes all the food item included in the menu
  - Serving size: This attribute specifies the quantity of the food item in one serving.
  - Calories: Provides details on the calorie count for one serving of each food item
  - Nutritional value: Provides details on the nutritional values like total fat, sodium, carbohydrates, cholesterol, etc. in one serving of each food item

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	Category	Item	Serving Size	Calories	Calori	Total F	Total F	Saturat	Satura	Trans	Cholest	Cholester	Sodium	Sodiur	Carbohy	Carbc	Dietar	Dieta	Suga	Prote	Vitam	Vitan	Calciu	Iron (% D
2	Breakfast	Egg McMuffin	4.8 oz (136	300	120	13	20	5	25	0	260	87	750	31	31	10	4	17	3	17	10	0	25	15
3	Breakfast	Egg White Delight	4.8 oz (135	250	70	8	12	3	15	0	25	8	770	32	30	10	4	17	3	18	6	0	25	8
4	Breakfast	Sausage McMuffin	3.9 oz (111	370	200	23	35	8	42	0	45	15	780	33	29	10	4	17	2	14	8	0	25	10
5	Breakfast	Sausage McMuffin with Egg	5.7 oz (161	450	250	28	43	10	52	0	285	95	860	36	30	10	4	17	2	21	15	0	30	15
6	Breakfast	Sausage McMuffin with Egg Whites	5.7 oz (161	400	210	23	35	8	42	0	50	16	880	37	30	10	4	17	2	21	6	0	25	10
7	Breakfast	Steak & Egg McMuffin	6.5 oz (185	430	210	23	36	9	46	1	300	100	960	40	31	10	4	18	3	26	15	2	30	20
8	Breakfast	Bacon, Egg & Cheese Biscuit (Regular Biscuit)	5.3 oz (150	460	230	26	40	13	65	0	250	83	1300	54	38	13	2	7	3	19	10	8	15	15
9	Breakfast	Bacon, Egg & Cheese Biscuit (Large Biscuit)	5.8 oz (164	520	270	30	47	14	68	0	250	83	1410	59	43	14	3	12	4	19	15	8	20	20
10	Breakfast	Bacon, Egg & Cheese Biscuit with Egg Whites (Regular Biscuit)	5.4 oz (153	410	180	20	32	11	56	0	35	11	1300	54	36	12	2	7	3	20	2	8	15	10
11	Breakfast	Bacon, Egg & Cheese Biscuit with Egg Whites (Large Biscuit)	5.9 oz (167	470	220	25	38	12	59	0	35	11	1420	59	42	14	3	12	4	20	6	8	15	15
12	Breakfast	Sausage Biscuit (Regular Biscuit)	4.1 oz (117	430	240	27	42	12	62	0	30	10	1080	45	34	11	2	6	2	11	0	0	6	15
13	Breakfast	Sausage Biscuit (Large Biscuit)	4.6 oz (131	480	280	31	48	13	65	0	30	10	1190	50	39	13	3	11	3	11	4	0	8	15
14	Breakfast	Sausage Biscuit with Egg (Regular Biscuit)	5.7 oz (163	510	290	33	50	14	71	0	250	83	1170	49	36	12	2	6	2	18	6	0	10	20
15	Breakfast	Sausage Biscuit with Egg (Large Biscuit)	6.2 oz (177	570	330	37	57	15	74	0	250	83	1280	53	42	14	3	11	3	18	10	0	10	20
16	Breakfast	Sausage Biscuit with Egg Whites (Regular Biscuit)	5.9 oz (167	460	250	27	42	12	62	0	35	11	1180	49	34	11	2	6	3	18	0	0	8	15
17	Breakfast	Sausage Biscuit with Egg Whites (Large Biscuit)	6.4 oz (181	520	280	32	49	13	65	0	35	11	1290	54	40	13	3	11	3	18	4	0	8	15
18	Breakfast	Southern Style Chicken Biscuit (Regular Biscuit)	5 oz (143 g)	410	180	20	31	8	41	0	30	10	1180	49	41	14	2	6	3	17	0	2	6	15
19	Breakfast	Southern Style Chicken Biscuit (Large Biscuit)	5.5 oz (157	470	220	24	37	9	45	0	30	10	1290	54	46	15	3	11	4	17	4	2	8	15
20	Breakfast	Steak & Egg Biscuit (Regular Biscuit)	7.1 oz (201	540	290	32	49	16	78	1	280	93	1470	61	38	13	2	8	3	25	10	2	20	25
21	Breakfast	Bacon, Egg & Cheese McGriddles	6.1 oz (174	460	190	21	32	9	44	0	250	84	1250	52	48	16	2	9	15	19	10	10	20	15
22	Breakfast	Bacon, Egg & Cheese McGriddles with Egg Whites	6.3 oz (178	400	140	15	24	7	34	0	35	11	1250	52	47	16	2	9	16	20	2	10	15	10
23	Breakfast	Sausage McGriddles	5 oz (141 g)	420	200	22	34	8	40	0	35	11	1030	43	44	15	2	8	15	11	0	0	8	10
24	Breakfast	Sausage, Egg & Cheese McGriddles	7.1 oz (201	550	280	31	48	12	61	0	265	89	1320	55	48	16	2	9	15	20	10	0	20	15
25	Breakfast	Sausage, Egg & Cheese McGriddles with Egg Whites	7.2 oz (205	500	230	26	40	10	52	0	50	17	1320	55	46	15	2	9	15	21	2	0	20	10
26	Breakfast	Bacon, Egg & Cheese Bagel	6.9 oz (197	620	280	31	48	11	56	0.5	275	92	1480	62	57	19	3	11	7	30	20	15	20	20
27	Breakfast	Bacon, Egg & Cheese Bagel with Egg Whites	7.1 oz (201	570	230	25	39	9	45	0.5	60	20	1480	62	55	18	3	12	8	30	10	15	20	15
28	Breakfast	Steak, Egg & Cheese Bagel	8.5 oz (241	670	310	35	53	13	63	1.5	295	99	1510	63	56	19	3	12	7	33	20	4	25	25
29	Breakfast	Big Breakfast (Regular Biscuit)	9.5 oz (269	740	430	48	73	17	87	0	555	185	1560	65	51	17	3	12	3	28	15	2	15	25
30	Breakfast	Big Breakfast (Large Biscuit)	10 oz (283 g	800	470	52	80	18	90	0	555	185	1680	70	56	19	4	17	3	28	15	2	15	30
31	Breakfast	Big Breakfast with Egg Whites (Regular Biscuit)	9.6 oz (272	640	330	37	57	14	69	0	35	12	1590	66	50	17	3	12	3	26	0	2	10	15
32	Breakfast	Big Breakfast with Egg Whites (Large Biscuit)	10.1 oz (286	690	370	41	63	14	72	0	35	12	1700	71	55	18	4	17	4	26	4	2	10	15
33	Breakfast	Big Breakfast with Hotcakes (Regular Biscuit)	14.8 oz (420	1090	510	56	87	19	96	0	575	192	2150	90	111	37	6	23	17	36	15	2	25	40
34	Breakfast	Big Breakfast with Hotcakes (Large Biscuit)	15.3 oz (434	1150	540	60	93	20	100	0	575	192	2260	94	116	39	7	28	17	36	15	2	30	40
35	Breakfast	Big Breakfast with Hotcakes and Egg Whites (Regular Biscuit)	14.9 oz (423	990	410	46	70	16	78	0	55	19	2170	91	110	37	6	23	17	35	0	2	25	30
36	Breakfast	Big Breakfast with Hotcakes and Egg Whites (Large Biscuit)	15.4 oz (437	1050	450	50	77	16	81	0	55	19	2290	95	115	38	7	28	18	35	4	2	25	30

RELEVANT DATA – DATA UNDERSTANDING

Insights are derived from the above dataset

DATA PREPARATION

The following Quality Checks has been performed on dataset:

Missing Data Analysis - to identify whether there are any null values or blanks in the columns/features

Check For Numbers – to identify whether there are any numbers like infinity that does not make sense

Based on the results obtained, decisions would be made on whether there is a need for further data processing techniques like data transformation, data cleaning, etc.

DATA PREPARATION – Execution Codes

- The acquired data set is imported using panda packages

```
menu = pd.read_csv('../input/menu.csv')
```

- Null value checks are carried out

```
# Check for Nulls
print(menu.isnull().any())
print("---------------------------------------")
```

- Data checks like number check are carried out

```
# check for numbers
print(menu.describe())
print("---------------------------------------")
```

DATA PREPARATION - Output

- Initial Missing data analysis is carried out on the data, and we could observe that the dataset available does not consist of any missing information and therefore we could conclude that the data set is perfect for further processing

```
Category                          False
Item                              False
Serving Size                      False
Calories                          False
Calories from Fat                 False
Total Fat                         False
Total Fat (% Daily Value)         False
Saturated Fat                     False
Saturated Fat (% Daily Value)     False
Trans Fat                         False
Cholesterol                       False
Cholesterol (% Daily Value)       False
Sodium                            False
Sodium (% Daily Value)            False
Carbohydrates                     False
Carbohydrates (% Daily Value)     False
Dietary Fiber                     False
Dietary Fiber (% Daily Value)     False
Sugars                            False
Protein                           False
Vitamin A (% Daily Value)         False
Vitamin C (% Daily Value)         False
Calcium (% Daily Value)           False
Iron (% Daily Value)              False
dtype: bool
-----------------------------------
```

DATA PREPARATION - Output

- The standard deviation value shows that there are not many outliers that could affect the training data set
- We could also observe that the dataset does not consists of random numbers like infinity which may potentially affect our analysis

```
         Calories         ...         Iron (% Daily Value)
count   260.000000        ...                   260.000000
mean    368.269231        ...                     7.734615
std     240.269886        ...                     8.723263
min       0.000000        ...                     0.000000
25%     210.000000        ...                     0.000000
50%     340.000000        ...                     4.000000
75%     500.000000        ...                    15.000000
max    1880.000000        ...                    40.000000

[8 rows x 21 columns]

------------------------------------
```

METHODOLOGIES/ ALGORITHMS – MODELING & EXECUTION

- Machine learning algorithms like Decision Trees (CART) will be used to perform the comparative analysis of the nutrients and calories in each item per serving

- **Decision Trees:**
  - Decision trees are used to build classification or regression models in the form of tree structures
  - It breaks down the data to smaller subsets, creating a tree with decision nodes and leaf nodes
  - The leaf node also called the root node represents a classification or decision

  We are using decision trees in this project because it could use both categorical and numerical data

- We are using decision trees in this project because it could use both categorical and numerical data

METHODOLOGIES/ALGORITHMS – MODELING & EXECUTION

Modeling and Execution:

- Store the menu dataset as a string `str(menu)`

```
'data.frame':    260 obs. of  24 variables:
 $ Category                              : Factor w/ 9 levels "Beef & Pork",..: 3 3 3 3 3 3 3
3 3 3 ...
 $ Item                                  : Factor w/ 260 levels "1% Low Fat Milk Jug",..: 76
77 228 229 230 245 12 11 14 13 ...
 $ Serving.Size                          : Factor w/ 107 levels "1 carton (236 ml)",..: 55 54
42 69 69 83 63 72 65 73 ...
 $ Calories                              : int  300 250 370 450 400 430 460 520 410 470 ...
 $ Calories.from.Fat                     : int  120 70 200 250 210 210 230 270 180 220 ...
 $ Total.Fat                             : num  13 8 23 28 23 23 26 30 20 25 ...
 $ Total.Fat....Daily.Value.             : int  20 12 35 43 35 36 40 47 32 38 ...
 $ Saturated.Fat                         : num  5 3 8 10 8 9 13 14 11 12 ...
 $ Saturated.Fat....Daily.Value.         : int  25 15 42 52 42 46 65 68 56 59 ...
 $ Trans.Fat                             : num  0 0 0 0 0 1 0 0 0 0 ...
 $ Cholesterol                           : int  260 25 45 285 50 300 250 250 35 35 ...
 $ Cholesterol....Daily.Value.           : int  87 8 15 95 16 100 83 83 11 11 ...
 $ Sodium                                : int  750 770 780 860 880 960 1300 1410 1300 1420
...
 $ Sodium....Daily.Value.                : int  31 32 33 36 37 40 54 59 54 59 ...
 $ Carbohydrates                         : int  31 30 29 30 30 31 38 43 36 42 ...
 $ Carbohydrates....Daily.Value.         : int  10 10 10 10 10 10 13 14 12 14 ...
 $ Dietary.Fiber                         : int  4 4 4 4 4 4 2 3 2 3 ...
 $ Dietary.Fiber....Daily.Value.         : int  17 17 17 17 17 17 18 7 12 7 12 ...
 $ Sugars                                : int  3 3 2 2 2 3 3 4 3 4 ...
 $ Protein                               : int  17 18 14 21 21 26 19 19 20 20 ...
 $ Vitamin.A....Daily.Value.             : int  10 6 8 15 6 15 10 15 2 6 ...
 $ Vitamin.C....Daily.Value.             : int  0 0 0 0 0 2 8 8 8 8 ...
 $ Calcium....Daily.Value.               : int  25 25 25 30 25 30 15 20 15 15 ...
 $ Iron....Daily.Value.                  : int  15 8 10 15 10 20 15 20 10 15 ...
```

METHODOLOGIES/ALGORITHMS – MODELING & EXECUTION

Modeling and Execution:

- Assigning null values to initialize and eliminate unwanted columns

```
menu$'Total.Fat....Daily.Value.' <- NULL
menu$'Saturated.Fat....Daily.Value.' <- NULL
menu$'Cholesterol....Daily.Value.' <- NULL
menu$'Sodium....Daily.Value.' <- NULL
menu$'Carbohydrates....Daily.Value.' <- NULL
menu$'Dietary.Fiber....Daily.Value.' <- NULL
menu$Item <- NULL
menu$Serving.Size <- NULL
```

- Partition the data using the createDataPartition() function to assign 80% of the total data as training data

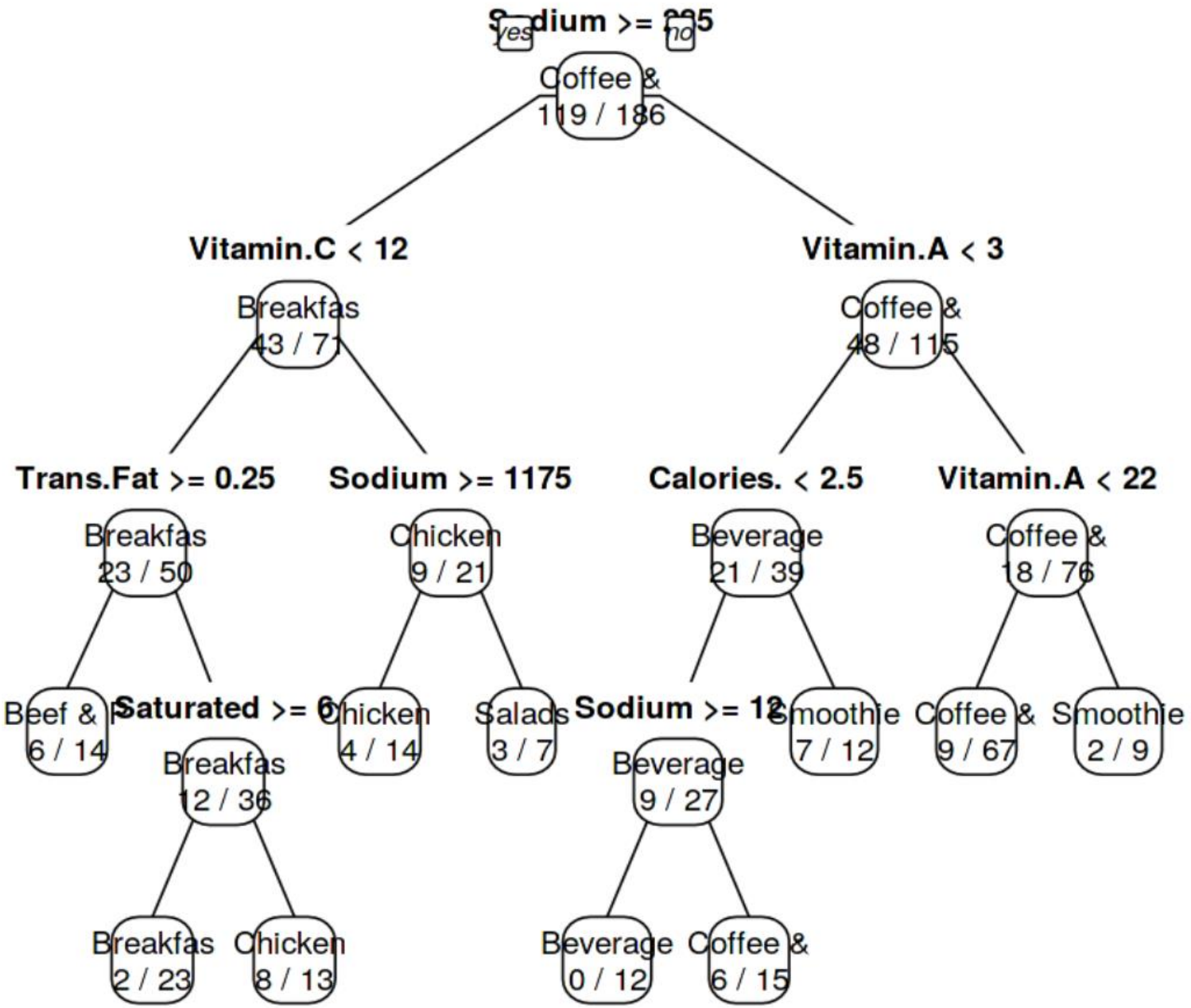
```
set.seed(1)
index <- createDataPartition(menu$Category, p=0.8, list=FALSE)
train <- menu[index,]
test <- menu[-index,]
```

METHODOLOGIES/ALGORITHMS – MODELING & EXECUTION

Modeling and Execution:

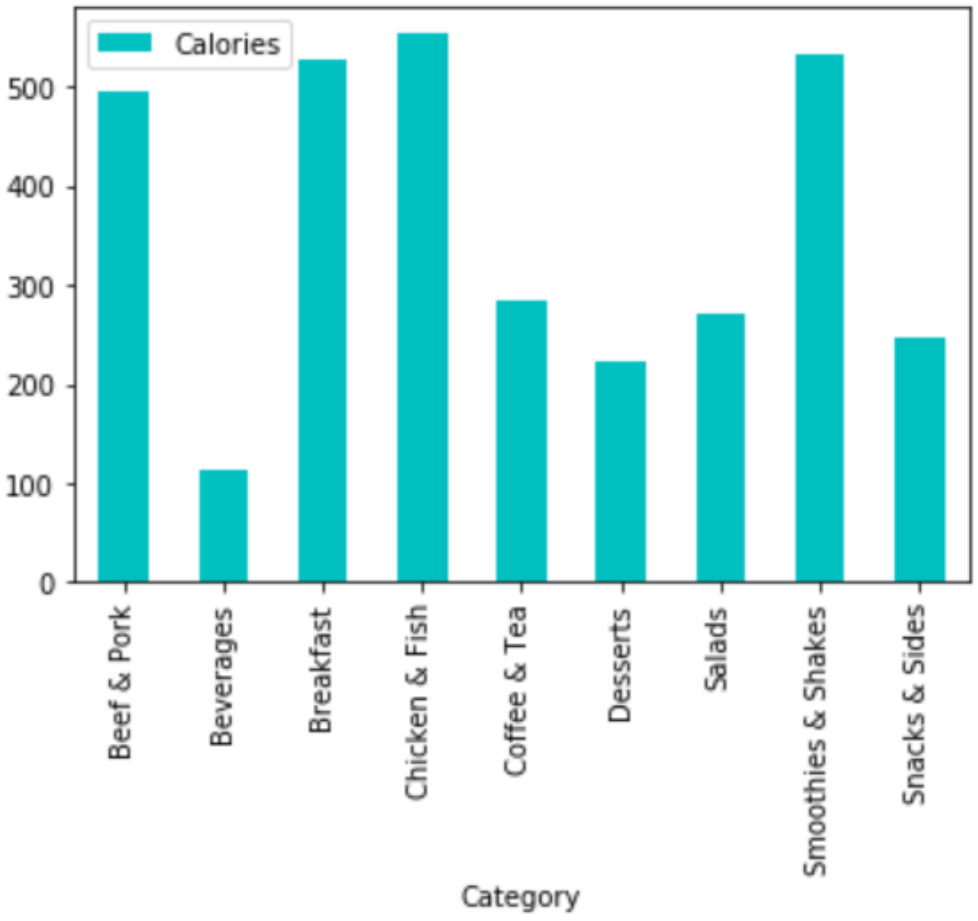
- Applying the below decision tree algorithm to identify which category of food in the menu consists of what nutrients the most

```
library(rpart)
library(rpart.plot)
model<-rpart(Category~., data = Train)
prp(model, type=1, extra = 3, main="Decision Tree")
```



EVALUATION OUTPUT- DECISION TREE

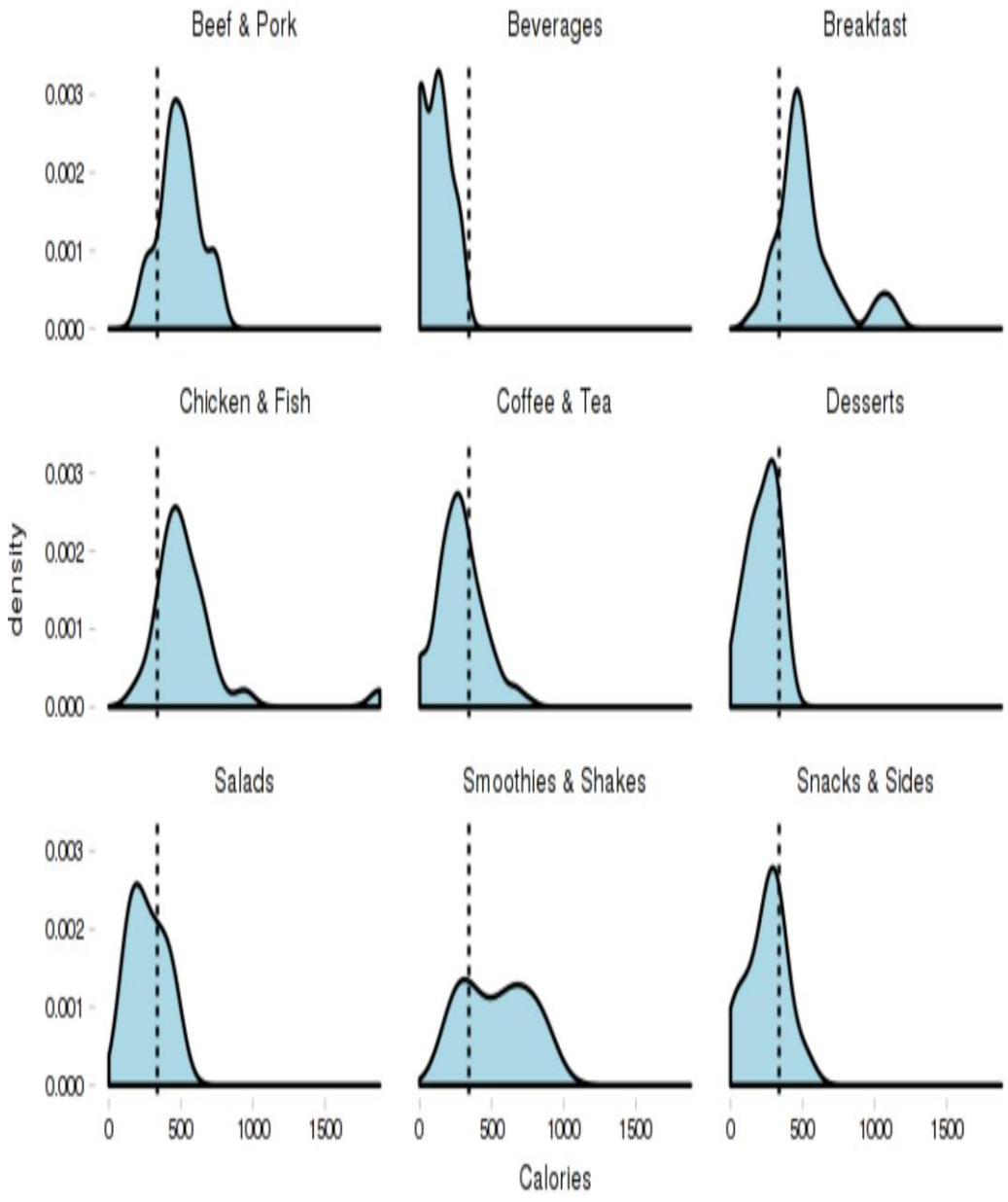
The percentage of accuracy of the prediction by the decision tree is around 74%



DATA INSIGHTS – VISUALIZATION

- To estimate the calories in each category of food, a bar graph has been plotted using seaborn libraries in python

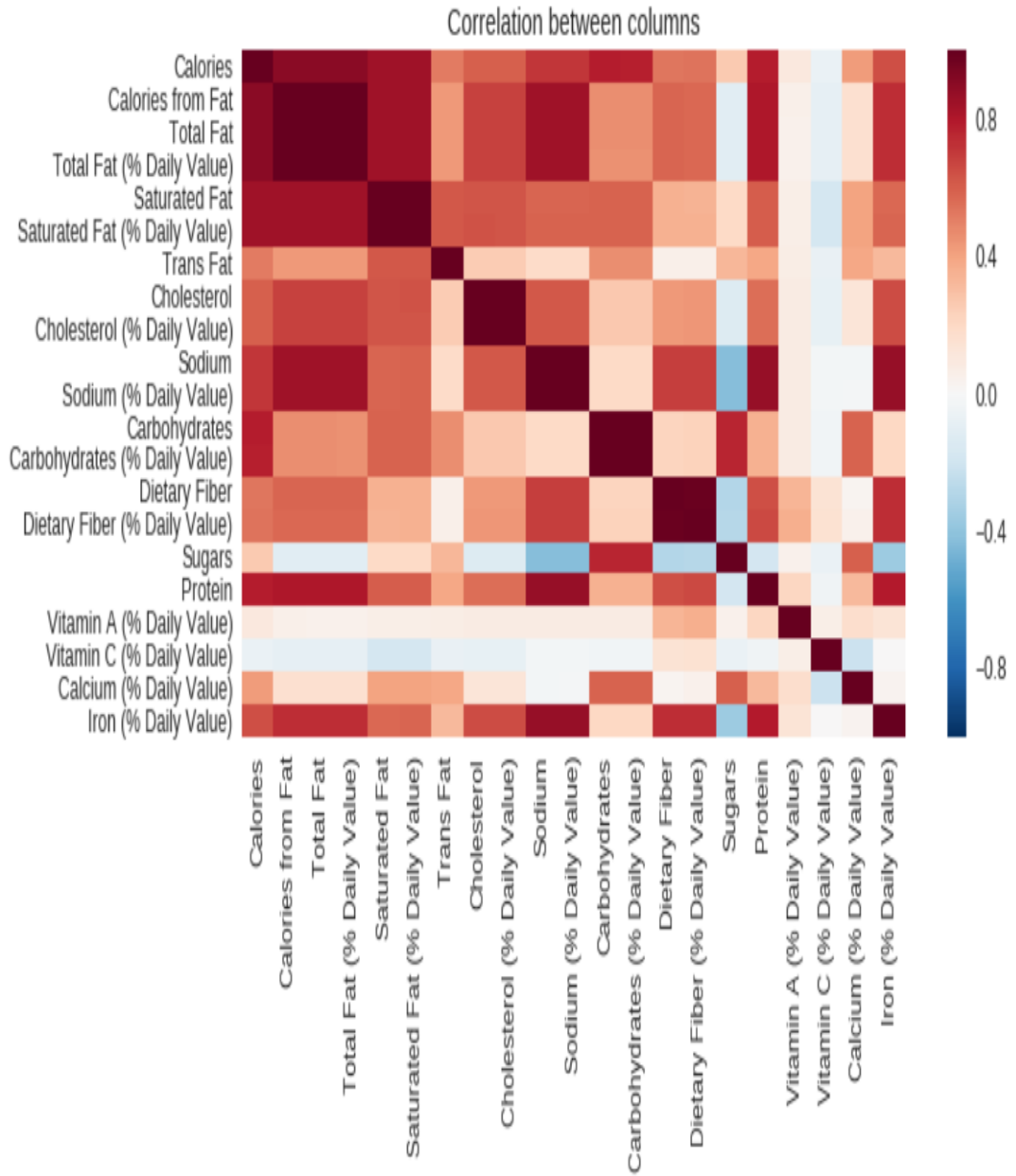
```
import seaborn as sns
cols = ['Calories','Cholesterol','Trans Fat','Sugars','Dietary Fiber']
cm = np.corrcoef(data[cols].values.T)
sns.set(font_scale = 1.5)
hm = sns.heatmap(cm,cbar = True, annot = True,square = True, fmt = '.2f', annot_kws =
{'size':15}, yticklabels = cols, xticklabels = cols)
```



DATA INSIGHTS – VISUALIZATION

- Faceted Density Curve has been plotted to identify the calorie distribution for different categories of food in the menu

```
#Smooth density estimate (faceted) - calories by category
options(repr.plot.height=4, repr.plot.width=6)
ggplot(menu2, aes(x = Calories)) +
  geom_density(fill="lightblue") +
  facet_wrap( ~ Category) +
  geom_vline(aes(xintercept = median(Calories)), linetype = 2) +
  theme_pander(base_size=8)
```



DATA INSIGHTS – VISUALIZATION

- Correlation maps are plotted to identify the correlation between the nutritional values/ calories (in food) with one another

```
fig, ax = plt.subplots(figsize=(11, 5))

sns.heatmap(data.corr(), ax=ax)
ax.set_title('Correlation between columns')
```

CONCLUSION AND ANALYSIS

- Based on the evaluation from the Decision Tree algorithm and the Calorie density plots, we could Suggest the fast food consumers on various insights on their daily nutrition intake as follows
  - Their food's maximum calorie value is found in the food category 'beef & pork' followed by 'Breakfast' and 'Chicken & Fish'
  - The Nutritional value of Sodium for the category 'coffee & tea' is high when compared to other food categories based on the decision tree analysis
- From the correlation plot, we could conclude that the correlation among total fat, sodium and cholesterol is very high indicating that the increase in one will result in the decrease in the other
- Based on these analysis a consumer can customize their daily food intake based on their nutrient preference, also it will be widely useful for health-conscious people who track their daily food intake.

SOFTWARE PACKAGES

The software packages used for this analysis are:

- **Programming**
  - Python Jupyter
  - Spyder
- **Visualization Libraries**
  - matplotlib
  - seaborn
  - ggplot

REFERENCE

1. https://www.kaggle.com/mcdonalds/nutrition-facts

Thank You