

Database, Data Warehouse Technology for U.S.

Chain-Restaurant Menu Item Nutrition

Taoli Zhen

Zhe-Li

Minh-Tam Pham

Chloe Ngo

Foroozan Akhavan

# Abstract

This project aims to help people in choosing the best meals to minimize health risks when dining out. We worked on a dataset that has nutrition facts of each item in chain restaurants, notorious for being unhealthy, from 2008 to 2018. The questions we answered are which restaurants to avoid, what to avoid, and if there were healthy changes in the menus over the years.

To answer these questions, we were required to set up and use a data warehouse cloud server to enable the analysts to analyze the data collected simultaneously. The results will help people living in the USA to have healthy options.

# RoadMap

1. Introduction
2. Methods and Implementation
  - a. Data cleansing and transformation - MySQL, Python
  - b. Remote connectivity and data warehouse optimization - MariaDB, SkySQL, HTAP
  - c. Analytic Tools and packages used - Seaborn, Sklearn, Aggregations and filters
3. Analysis and Interpretation
  - a. Trends study
  - b. Correlation analysis
  - c. interpretation on aggregation results
  - d. Clustering grouping
4. Discussion
5. Learning and Conclusion

# Introduction

- Over 1/3 of Americans dine out
- Government requires chain restaurants to display information on menus
- Are possible for Americans to eat nutritiously?



**Emergency →** **DRIVE THRU**

**WELCOME TO ST. RAY'S MEMORIAL HOSPITAL**  
Now a division of ValueFunHealth™ Hunger Maintenance Organization

**Try our new burger, fries, shake, and angioplasty combo!**  
**\$5.99** with \$5 co-pay

**COMBO TRAYS**  
Burger + Fries + Large Soft Drink . . . . . \$3.99  
Double Burger + Cheese Fries + Shake . . . \$4.99  
Surprise Meal (don't ask) . . . . . \$5.99  
*LIPSIZ any combo for just \$3.99!*

**BURGERS**  
Hamburger . . . . . \$1.99  
Cheeseburger . . . . . \$2.29  
Double Hamburger . . . . . \$2.49  
Double Cheeseburger . . . . . \$2.79  
Triple Hamburger . . . . . \$2.99  
Triple Bypass . . . . . ask  
*Ask about our FREE cholesterol tests*

**SIDES 'N' DRINKS**  
French Fries . . . . . \$ .99 / 1.49 / 1.99  
Soft Drinks . . . . . \$ .99 / 1.49 / 1.99  
Milk Shakes . . . . . \$1.29 / 1.79 / 2.29  
*Free in-network REFERRAL with every order*

**Yes!**

PLEASE DRIVE FORWARD TO PICK UP YOUR ORDER → THANK YOU PLEASE COME AGAIN

# Data description

- Menustat.org
- CSV format and separated by years
- 2018 dataset(latest) as base case

**MenuStat**

[Home](#)

[About](#)

## Interactive Restaurant Nutrition Database

MenuStat is a free nutritional database of foods and beverages served by the nation's largest chain restaurants

Search

Graph

Export

Choose criteria to tailor search results:

Food Categories

All Selected (12) ▼

Restaurants

All Selected (96) ▼

Years

All Selected (9) ▼

Nutrients

All Selected (12) ▼

Search

# MySQL and Python

- Data cleansing and transformation
- Third normal form for the galaxy schema
- Loading data into MySQL

```
dat['r_id'] = dat.groupby(['Restaurant']).ngroup()
restaurant = dat[['r_id', 'Restaurant']].drop_duplicates().sort_values(by=['r_id'])
restaurant.rename(columns={'Restaurant':'r_name'}, inplace=True)
restaurant
```

```
# Fact table, nutrition data
nutrition_facts = dat[['Menu_Item_ID', 'Year', 'r_id', 'cat_id',
                      'Calories', 'Total_Fat', 'Saturated_Fat', 'Trans_Fat', 'Cholesterol',
                      'Sodium', 'Potassium',
                      'Carbohydrates', 'Protein', 'Sugar', 'Dietary_Fiber']]
nutrition_facts.columns = ['item_id', 'year', 'r_id', 'cat_id',
                           'calories', 'total_fat', 'saturated_fat', 'trans_fat', 'cholesterol',
                           'sodium', 'potassium',
                           'carbs', 'protein', 'sugar', 'dietary_fiber']
nutrition_facts = nutrition_facts.drop_duplicates()
#nutrition_facts.astype({'calories':'int64', 'total_fat':'int64', 'saturated_fat':'int64', 'trans_fat':'int64',
#                        'cholesterol':'int64', 'sodium':'int64', 'potassium':'int64',
#                        'carbs':'int64', 'protein':'int64', 'sugar':'int64', 'dietary_fiber':'int64'})
nutrition_facts
#nutrition_facts.to_csv('cleaned_tables_by_year/2018/nutrition_facts.csv', index=False)
```

Menu Nutrition Data Schema

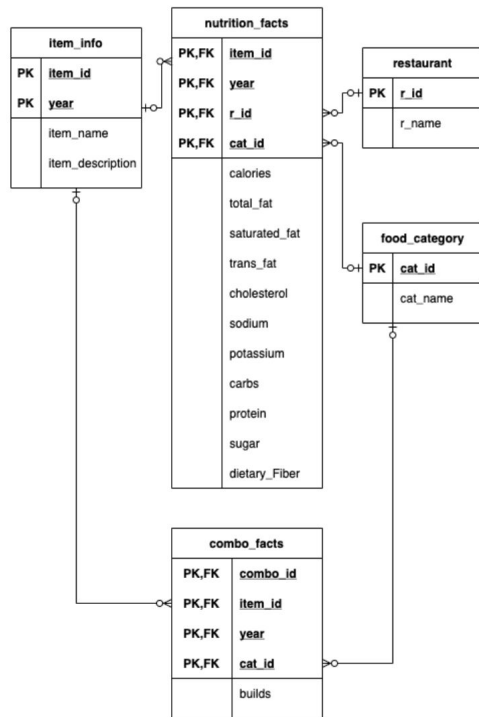


Fig. 1. Galaxy Schema for the Menu Nutrition Data. Serve as a data warehouse schema.

# MariaDB, SkySQL, HTAP platforms

- **MariaDB** is a fork of MySQL - importing a dump file
  - Capability to support a variety of storage engine and better performance in many cases
- **SkySQL** is the cloud service for MariaDB - provide remote connectivity
  - Including 3 platforms:
    - Transaction platform - OLTP
    - Analytics platform - OLAP
    - Hybrid Transaction/ Analytical platform - **HTAP**

In this project, we used the Hybrid Transactional-Analytical Processing Platform to simulate the real-life work environment. We can have a seamless data movement from the transactional database to the analytics data warehouse.

# MariaDB, SkySQL, HTAP platforms

- MySQL connector is compatible with MariaDB
- Set up the replication filter using the `set_htap_replication()` UDF
- Column store uses a ColumnStore engine, need to specify in table creation DDL
- Every row inserted to the row store will automatically replicate to the column store in realtime
- Enables OLAP with real-time data

```
sql_execute(conn,
    sql = '''
        SELECT set_htap_replication(
            'menu_rowstore.[[:word:]]+',
            'menu_rowstore',
            'menu_cstore');
    ;''')
```

```
def create_connection():
    conn = None
    try:
        conn = mysql.connector.connect(
            user="DB00004175",
            password="tTNlXEir[REDACTED]p4koTR",
            host="menu-data-db.mdb[REDACTED]7.db.skysql.net",
            ssl_ca="skysql_chain.pem",
            port=5001)
        conn.autocommit = False
        return conn
    except mysql.connector.Error as e:
        print(f"Error connecting to MariaDB Platform: {e}")
    return conn
```

```
CREATE TABLE IF NOT EXISTS menu_cstore.nutrition_facts
ENGINE = COLUMNSTORE
SELECT * FROM menu_rowstore.nutrition_facts
```



# Analytic tools and packages

- Aggregation:
  - ROLLUP function to get aggregate data for both 2D and 3D cubes
    - Used Pandas pivot function to put the ROLLUP in a cube structure table
  - Pivot table function in Pandas to achieve CUBE-like structure
- Python Package Seaborn:
  - why: 96 restaurants and 12 categories - hard to read and keep track with only CUBE
  - Visualizing the data for us to interpret and compare
- Python Package Sklearn:
  - why: Nutrition patterns - multidimensional factors - hard to set up a solid criteria
  - Clustering - an unsupervised machine learning to identify groups and help segment the restaurants based on their nutrition information.

# Trend study

- Fig.3 shows 7-Eleven, their average calories for all menu items were decreasing over the five-year period.
- Fig. 5 displays a line plot for average calories from protein in all food categories. We can see that there is an upward trend in the Burgers, the Sandwiches, the Appetizers & Sides, and the Soup category. Which is a good indicator of good ingredients.
- Based on the result from Fig. 6 and Fig. 7, there seems to be a correlation between the amount of cholesterol and calories.

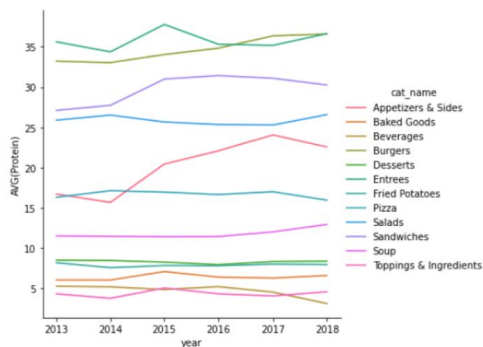


Fig. 5. Line plot for the average protein for all categories from 2013 to 2018.

	year	r_name	avg_calories
0	2013	7 Eleven	284.1429
1	2014	7 Eleven	301.5238
2	2015	7 Eleven	287.8980
3	2016	7 Eleven	303.2778
4	2017	7 Eleven	308.2353
5	2018	7 Eleven	133.3333

Fig. 3. Avg. calories for 7 Eleven from 2013 to 2018.

	Year	2013	2014	2015	2016	2017	2018	All Years
Restaurant	Category							
7 Eleven	All Categories	NaN	NaN	NaN	NaN	NaN	NaN	289.9115
	Appetizers & Sides	66.6667	80.0000	85.0000	80.0000	NaN	NaN	75.7143
	Beverages	99.0000	103.4286	101.0000	130.6667	180.0000	100.0000	122.9556
	Burgers	440.0000	440.0000	440.0000	NaN	NaN	NaN	440.0000
	Entrees	212.5000	213.3333	217.2727	210.0000	NaN	NaN	214.1935
	Fried Potatoes	186.6667	170.0000	170.0000	240.0000	NaN	NaN	185.0000
	Pizza	300.0000	435.0000	435.0000	NaN	NaN	NaN	408.0000
	Salads	NaN	210.0000	225.0000	330.0000	230.0000	NaN	246.0000
All Restaurants	Sandwiches	407.0000	425.2632	414.5000	455.2941	518.3333	300.0000	429.6386
	All Categories	NaN	NaN	NaN	NaN	NaN	NaN	397.0637

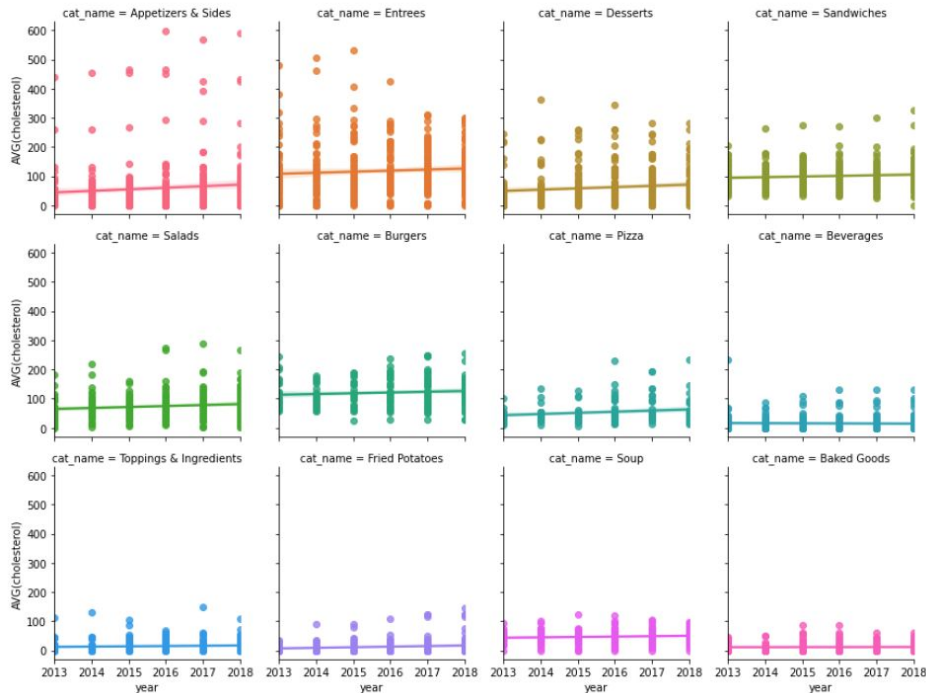
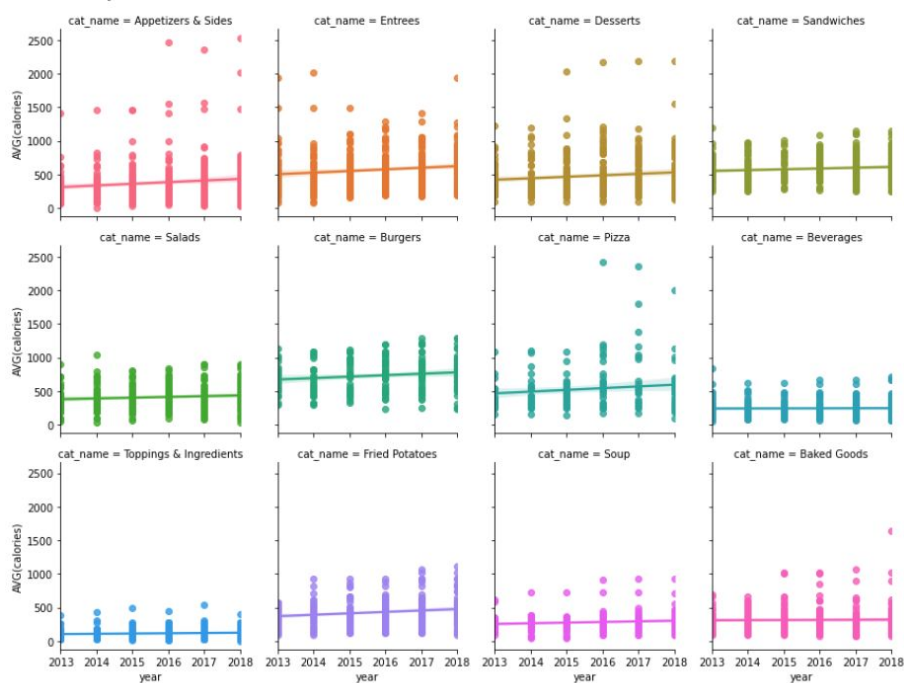
# Aggregation results and Interpretation

- For the trans fat/fat ratio, by grouping items by restaurant and filtering, we found that **Long John Silver's** and Arby's are the highest in trans fat/fat ratio. Long John Silver's restaurant events have **27.45%** of trans fat in the total fat.
- For the **sugar to total calories ratio**, we find that two of the five worst restaurants are beverage stores: Jamba Juice and Starbucks. **Long John Silver's**, KFC, and Krispy Kreme are also in the top five in the sugar/ calories ratio.
- Restaurants that offer high amounts of fibers are Chipotle, Qdoba, Round Table Pizza, Jamba Juice, and Carrabba's Italian Grill. While **Long John Silver's** is one of the top five restaurants, it offers the least fibers in its items
- Long John Silver's** is the most unhealthy restaurant. And we also find that Applebee's indeed have reduced their unhealthy nutrients over the years.

	year	r_name	cat_name	AVG(calories)	AVG(cholesterol)	AVG(trans_fat)	AVG(Protein)	AVG(carbs)
0	2013	Culver's	Entrees	1943.0000	479.9412	2.0588	135.0000	102.1765
1	2013	Checker's Drive-In/Rallys	Appetizers & Sides	1409.7619	437.8571	0.0000	138.6667	30.2857
2	2013	IHOP	Entrees	880.3302	379.4575	0.8208	35.0283	76.9670
3	2013	Perkins	Entrees	1046.1475	316.7951	0.0492	37.9016	115.6393
4	2013	Krystal	Entrees	449.4444	282.7778	0.5333	17.7222	31.8889
5	2013	Whataburger	Entrees	684.1379	263.7931	0.0000	23.7586	83.5714
6	2013	Dominos	Appetizers & Sides	475.5556	258.6364	0.1818	40.6667	9.2222
7	2013	Bojangles	Entrees	1485.4762	247.4762	0.5238	74.1429	104.7143
8	2013	Perkins	Burgers	1138.0000	245.5000	0.0000	46.6000	80.8000
9	2013	California Pizza Kitchen	Desserts	880.0000	245.1000	0.5000	9.2000	82.0000
10	2013	Denny's	Entrees	674.4872	236.0127	0.0759	34.4684	56.0000
11	2013	Culver's	Beverages	833.7895	234.0526	0.2105	14.7368	101.4737
12	2013	Culver's	Desserts	799.4184	220.5408	1.1837	14.1020	88.3673
13	2013	Perkins	Desserts	1229.5833	215.1042	1.0625	11.6875	130.2917
14	2013	O'Charley's	Entrees	986.6269	209.7761	1.6316	53.8507	60.2388
15	2013	IHOP	Burgers	851.6667	207.5000	2.1667	46.5000	44.8333
16	2013	Jack in the Box	Sandwiches	518.9706	204.3529	0.2059	25.2059	40.6176
17	2013	Denny's	Burgers	990.0000	201.6250	1.7500	55.6250	56.8750

	Year	Restaurant	Category	AVG(calories)	AVG(total_fat)	AVG(trans_fat)	AVG(cholesterol)	AVG(sodium)	AVG(potassium)	AVG(sugar)	AVG(Protein)
0	2013	7 Eleven	Appetizers & Sides	66.6667	2.3333	0.0000	15.0000	370.0000	NaN	0.0000	5.3333
1	2014	7 Eleven	Appetizers & Sides	80.0000	4.0000	0.0000	25.0000	480.0000	NaN	0.0000	6.0000
2	2015	7 Eleven	Appetizers & Sides	85.0000	5.0000	0.0000	17.5000	375.0000	NaN	0.0000	4.5000
3	2016	7 Eleven	Appetizers & Sides	80.0000	4.0000	0.0000	25.0000	480.0000	NaN	0.0000	6.0000
4	2017	7 Eleven	Appetizers & Sides	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...
4855	2014	Zaxby's	Toppings & Ingredients	166.6438	11.8493	NaN	16.2603	369.1370	NaN	7.0548	3.0274
4856	2015	Zaxby's	Toppings & Ingredients	155.5195	10.9091	NaN	14.0390	370.2727	NaN	6.1558	3.0519
4857	2016	Zaxby's	Toppings & Ingredients	151.8182	10.5714	NaN	15.1429	365.5195	NaN	5.6234	3.3896
4858	2017	Zaxby's	Toppings & Ingredients	151.8182	10.5714	NaN	15.1429	365.5195	NaN	5.6234	3.3896
4859	2018	Zaxby's	Toppings & Ingredients	172.7941	14.7353	0.0303	14.0294	653.8235	NaN	5.1471	1.6471

# Avg. Calories Trends & Avg. Cholesterol Trends



# Correlation Analysis

Fig. 8 is the heatmap for the correlation of average nutrition attributes grouped by the restaurant, category, and year. The result from the data warehouse is already included year, restaurant, and food category variances. From the correlation heatmap, calories and total fat have the highest correlation which is 0.93. Protein, carbohydrates, sodium, cholesterol, and trans fat all have a relatively strong correlation with calories between the range of 0.61 to 0.75.

With beverage items included, there is no correlation between sugar and other nutrients except for carbohydrates. One of the reasons for the low correlation between sugar and calories is that the weight of the beverage category is small while beverages have the highest amount of sugar in any given category. We will discuss the relationship of correlation and the implication on other results later in the discussion.

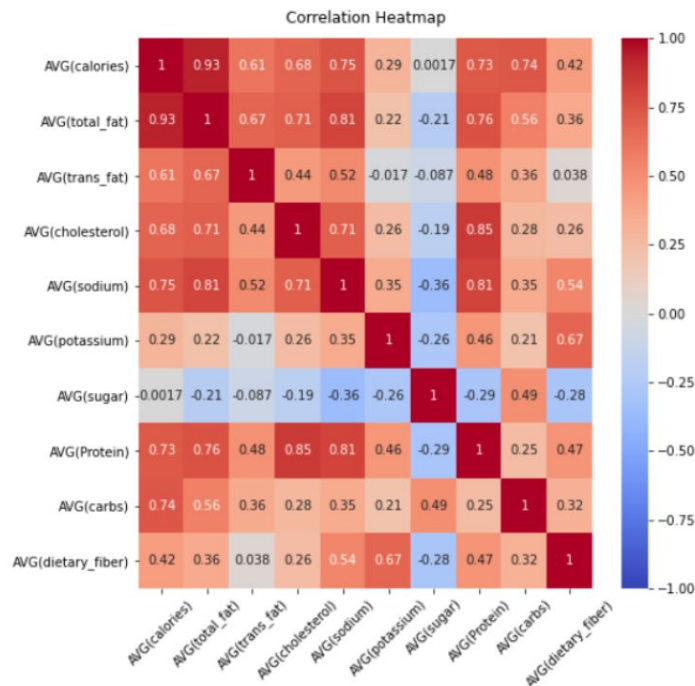


Fig. 8. Correlation heatmap for nutrient features grouped by year, restaurant, and food category.

# Unhealthy Categories and restaurants

	cat_name	cat_id	avg_cal
0	Burgers	3	737.9365
1	Entrees	5	674.4843
2	Sandwiches	9	640.4563
3	Desserts	4	535.3693
4	Salads	8	491.7148
5	Fried Potatoes	6	474.8087
6	Appetizers & Sides	0	452.5349
7	Pizza	7	375.2979
8	Baked Goods	1	320.5402
9	Soup	10	293.3740
10	Beverages	2	277.8308
11	Toppings & Ingredients	11	125.0382

Fig. 9.1. Result of the query which shows average calorie per food category.

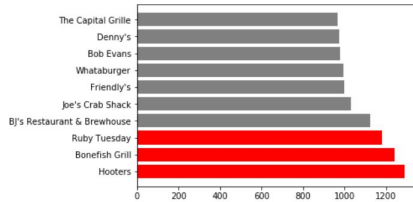


Fig. 9.2. Result of the query which shows 10 restaurants that have higher-calorie burgers.

	cat_name	avg_chol
0	Entrees	175.5359
1	Burgers	108.5269
2	Sandwiches	107.2060
3	Salads	86.5368
4	Appetizers & Sides	77.2685
5	Desserts	70.7206
6	Soup	45.8194
7	Pizza	42.5097
8	Beverages	18.5630
9	Toppings & Ingredients	17.7580
10	Baked Goods	14.9791
11	Fried Potatoes	13.7474

Fig. 14. Result of the query which shows entrees have the highest average cholesterol.

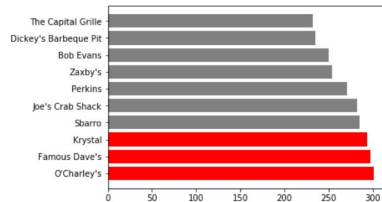


Fig. 10.2. Result of the query which shows 10 top restaurants that have higher cholesterol in their entrees.

	cat_name	cat_id	avg_sodium
0	Sandwiches	9	1741.3864
1	Entrees	5	1522.4357
2	Burgers	3	1360.6340
3	Soup	10	1284.7896
4	Appetizers & Sides	0	1135.5703
5	Salads	8	1066.0680
6	Pizza	7	894.3457
7	Fried Potatoes	6	846.6797
8	Baked Goods	1	431.0031
9	Desserts	4	300.7714
10	Toppings & Ingredients	11	283.7959
11	Beverages	2	127.5020

Fig. 11.1. Result of the query which shows sandwiches have the highest average sodium.

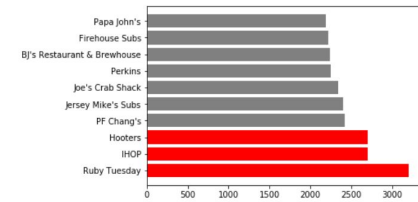


Fig. 11.2. Result of the query which shows 10 top restaurants that have higher sodium in their sandwiches.

# Discussion - Trends in nutrition composition

Fig. 5 showed an upward trend for the average protein content in all categories. This is a sign that chain restaurants made their items more nutritious by increasing the amount of protein (meat) in their food.

Based on Fig.6 and Fig.7, with a simple linear regression line drawn in each time vs average nutrition scatter plot by food category, the trend becomes distinguishable. Six main categories have an identifiable upward trend in the average category while the regression line for the remaining calories was flat. The same pattern also appeared in Fig.7 which is showing trends for the average cholesterol in all categories.

The correlation heatmap showed a strong positive correlation between protein, calories, total fat, cholesterol, and sodium. This further suggested the theory of increasing protein content doesn't make the food healthier. One explanation for this result might be that the protein source is not clean enough (the proportion of fat in meat is high). To combine the findings in the Methods section 3.3 and 3.5, we can say that there isn't a clear trend of chain restaurants making their food healthier.

# Clustering groups

- fetch needed data from data warehouse
- Data cleansing and normalization for machine learning
- Set up inertia function to find number of groups needed  $k = 4$
- Clustering segments restaurants into 4 groups based on nutrients

```
df_sql = sql_execute_show(conn, sql = '''
SELECT b.r_name, AVG(calories), AVG(total_fat), AVG(Protein), AVG(sodium), AVG(sugar), AVG(dietary_Fiber)
FROM menu_cstore.nutrition_facts a
LEFT JOIN menu_cstore.restaurant b ON a.r_id = b.r_id
GROUP BY b.r_name
;''')
```

```
def clean_dataset(df):
    assert isinstance(df, pd.DataFrame), "df needs to be a pd.DataFrame"
    df.dropna(inplace=True)
    indices_to_keep = ~df.isin([np.nan, np.inf, -np.inf]).any(1)
    return df[indices_to_keep].astype(np.float64)
```

```
df_re = df_sql.set_index('r_name')
clean_dataset(df_re)
names = pd.DataFrame(df_re.index)
```

```
scaler = preprocessing.MinMaxScaler()
df_re_normal = scaler.fit_transform(df_re)
```

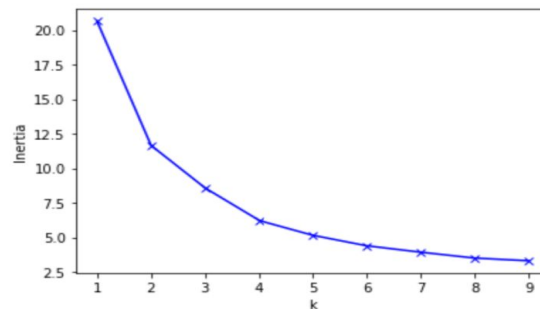


Fig. 13.1. Kmean Model: K vs Inertia.



From the results in the section of aggregation results, it seems like Long John Silver's is the worst restaurant to eat at, especially for a person struggling with type 2 diabetes as well as heart problems. Their food doesn't seem to be nutritious (low fiber, high trans fat and high sugar). In this analysis, we note that Long John Silver's is consistently the least nutritious for you.

With Long John Silver's satisfied the criteria for the worst restaurant, we classified restaurants with similar nutritional patterns as shown in the up right figure. They are the restaurants that people should avoid. Although we determined that there isn't a trend of chain restaurants making their food healthier, there are still restaurants that have signs of making their food healthier such as Applebee's. In the down right figure, we also provided a list of restaurants that have a similar pattern with Applebee's and they are the restaurants we recommend.

	Types	AVG(calories)	AVG(total_fat)	AVG(Protein)	AVG(sodium)	AVG(sugar)	AVG(dietary_Fiber)
r_name							
Auntie Anne's	3	275.5607	5.2865	4.4921	330.4551	33.4180	0.6921
Baskin Robbins	3	457.2594	18.5234	8.7413	207.2902	56.3655	1.5204
Burger King	3	398.2821	17.7964	11.6295	616.6327	27.6312	1.2671
Church's Chicken	3	249.2882	8.3869	5.9079	432.1224	27.3121	1.0487
Culver's	3	474.1866	30.4674	19.1938	520.8213	36.4140	2.0164
Dairy Queen	3	525.8934	19.7327	11.5364	427.1246	58.1372	1.4572
Dunkin' Donuts	3	273.7371	8.2878	6.2484	242.9236	33.6873	0.8429
In-N-Out Burger	3	303.7284	15.5200	11.7850	405.5948	28.6121	1.3100
Jamba Juice	3	299.6158	4.1945	6.8399	133.2113	50.0751	3.7621
KFC	3	244.6993	7.2760	6.5721	409.5972	31.7460	0.8768
Krispy Kreme	3	261.4813	7.4261	6.3331	160.4532	33.6321	0.8313
Long John Silver's	3	215.3468	5.9252	3.1508	291.2043	32.6805	0.6428
McDonald's	3	318.1525	11.4862	10.3669	339.7588	32.2826	1.0867
Panda Express	3	214.9644	6.2894	6.7272	368.8802	24.6987	0.9578
Sheetz	3	223.1426	6.9818	6.5991	224.6046	26.7965	0.6452
Sonic	3	406.9168	16.2108	7.0566	394.4275	46.6661	0.9627
Starbucks	3	260.2133	7.4618	6.4358	145.7592	37.4298	0.9140
Steak 'N Shake	3	438.0960	25.9651	12.9231	591.8237	32.4213	2.1976
Wawa	3	327.9445	12.8564	11.4601	605.0117	25.4090	1.4552
Whataburger	3	434.8670	18.6071	13.5658	691.8955	30.3362	1.6897
White Castle	3	378.2370	13.7935	8.1544	379.6129	46.0859	1.3448

	Types	AVG(calories)	AVG(total_fat)	AVG(Protein)	AVG(sodium)	AVG(sugar)	AVG(dietary_Fiber)
r_name							
NaN	0	459.4577	24.4788	17.7381	845.6825	18.2593	2.3239
Applebee's	0	535.8140	29.0960	23.8558	1262.2796	19.2851	3.3980
BJ's Restaurant & Brewhouse	0	488.9209	22.6404	18.6297	1002.3109	14.1138	3.1049
Bonefish Grill	0	445.8686	26.3080	23.7449	867.4768	9.6942	3.0023
California Pizza Kitchen	0	550.7461	22.8102	18.7803	780.5334	13.7242	4.0460
Carl's Jr.	0	455.2334	24.1069	17.7208	872.0063	16.0521	2.4034
Carrabba's Italian Grill	0	540.9292	29.4325	27.1574	1176.5037	8.9181	6.3316
Checker's Drive-In/Rallys	0	565.7641	26.8247	28.2008	1246.5229	32.2480	2.3455
Chili's	0	560.4523	31.4662	24.8794	1380.7838	12.2538	3.5240
Del Taco	0	427.9030	19.7764	16.0883	787.1680	17.3592	3.7621
Denny's	0	466.8052	24.3677	19.4258	974.7025	14.7853	2.6639
Friendly's	0	532.6335	26.2782	16.4836	823.5714	24.9517	2.2099
Hardee's	0	408.7226	21.0490	15.4217	900.1518	15.0945	2.1629
IHOP	0	524.3899	28.3318	20.1964	1042.0655	18.2119	3.0185
Jason's Deli	0	511.1826	27.3693	25.7436	1165.6545	10.8429	4.4280
Jersey Mike's Subs	0	574.1110	26.5068	32.3962	1671.8651	9.3339	4.4195
LongHorn Steakhouse	0	431.3268	22.6560	25.5308	783.1325	11.4371	1.8379
McAllister's Deli	0	384.2238	17.7285	16.2064	954.1388	8.9572	3.7433
Noodles & Company	0	436.1173	20.0838	16.5251	922.3045	10.5114	2.8600
Olive Garden	0	423.2257	17.9719	19.2368	699.1405	15.5257	3.3487
Outback Steakhouse	0	534.4170	33.3252	34.3965	1022.7305	9.9900	3.1252
Perkins	0	509.5529	26.2869	15.8987	906.1914	19.0613	2.2936
PF Chang's	0	529.0462	22.3746	26.2289	1542.6432	24.4016	4.0592
Quiznos	0	403.4264	38.9949	29.5047	1117.7274	6.3954	2.3553
Red Robin	0	515.6705	23.9758	16.1816	803.5196	28.9303	2.6926
Ruby Tuesday	0	501.7967	27.3628	24.8744	1079.6934	7.9091	2.9740
TGI Friday's	0	545.5464	27.6834	22.4653	1239.0578	24.2707	2.9801
The Capital Grille	0	521.4688	30.0893	29.6514	752.1726	11.7125	2.6314
Zaxby's	0	435.8547	29.8580	21.9276	971.5600	14.0769	2.4705

# Lessons Learned and Conclusion

For aggregate data generated from only one dimension, e.g. Restaurants, there needs to be an exploration in the other dimensions to make sure the data in the later years are consistent. Also, examining 3-dimensional data cubes can be helpful in finding the root cause for missing values. To decide what to do with missing values, we have to drill down to see it in a finer grind level in order to determine what happened. (e.g 7-eleven average calories mistakes)

Despite the fact that restaurants were trying to make their meals and items more nutritious, the accompanying increment in cholesterol, calories, total fat, and sodium didn't make their food any healthier. Regardless, it is still possible to avoid the worst food categories to choose in specific restaurants and we have provided a list of relatively healthier restaurants. More recent data is needed in order to analyze the latest trends in chain-restaurants' menu items nutrition. Furthermore, while the aggregate data didn't show any healthier trends, there can be more in depth research to investigate the nutrition facts in newly introduced items in later years.

# References

*A. Tran, A. Moran, and S. N. Bleich, "Calorie changes among food items sold in U.S. convenience stores and pizza restaurant chains from 2013 to 2017," Preventive Medicine Reports, vol. 15, p. 100932, 2019.*

*MenuStat. [Online]. Available: <http://www.menustat.org/>. [Accessed: 08-May-2021].*

*"Open Source Database (RDBMS) for the Enterprise," MariaDB. [Online]. Available: <https://mariadb.com/docs/>. [Accessed: 08-May-2021].*

*T. Jankov, "MariaDB vs MySQL, a Database Technologies Rundown," Kinsta, 30-Dec-2020. [Online]. Available: <https://kinsta.com/blog/mariadb-vs-mysql/>. [Accessed: 08-May-2021].*

*"Hybrid Transactional-Analytical Processing (HTAP)," MariaDB. [Online]. Available: <https://mariadb.com/docs/multi-node/htap/>. [Accessed: 08-May-2021].*

*Database, Data Warehouse Technology for U.S. Chain-Restaurant Menu Item Nutrition Data Analysis - final project report*

# Q&A

Thank you for your attention!