Food Recommendation System

Group No. – 11
Team Members – Shahbaz Aslam, Shivani Bajaj, Ritvik Wadhwa

The company selected for this project is **Fitbit**, a popular health and fitness technology company that specializes in wearable fitness trackers, smartwatches, and accessories.

Plan for designing a new strategic information product

Our strategic information product for Fitbit will be a comprehensive food recommendation system that provides personalized suggestions for healthy food options based on the user's daily calorie limit and preferred food category. Leveraging the dataset provided by the City of New York on nutritional information of food items sold in restaurants, our product will offer a user-friendly experience.

Working of the System

Our food recommendation system for Fitbit allows users to set their daily calorie limit and preferred food category to generate personalized suggestions. Based on the dataset provided by the City of New York, our system curates accurate and relevant recommendations that include detailed nutritional information for each food item. The visually appealing user interface will be accessible via mobile app and web portal and continuously updated with the latest nutritional information. Our product promotes a healthier lifestyle by supporting users in making informed decisions about their nutrition, making it a valuable addition to Fitbit's suite of health and wellness products. Out system also allows users to track their food intake and monitor their progress towards their dietary goals. Users can log their meals and snacks, view their calorie and nutrient intake, and adjust their goals as needed. The system also provides reminders and alerts to help users stay on track and make healthy choices throughout the day.

Furthermore, our food recommendation system is designed to be flexible and customizable, allowing users to adjust their preferences and goals over time. As users' dietary needs and goals change, our system adapts and provides new recommendations to ensure they continue to meet their nutritional requirements.

Overall, our food recommendation system for Fitbit offers a comprehensive and personalized approach to nutrition, empowering users to take control of their health and make informed decisions about their diet. With its user-friendly interface, accurate and up-to-date data, and customizable features, we believe our product will be an indispensable tool for anyone looking to improve their health and wellness.

The Porters' forces that can be addressed through our Information Product are: -

Competitive Rivalry - The food recommendation system can provide Fitbit with a competitive advantage over other health and fitness companies by offering a unique and personalized feature that helps users make informed decisions about their nutrition. This can differentiate Fitbit from its competitors and attract more customers who are looking for comprehensive health and wellness solutions.

Threat of Substitutes - The food recommendation system can serve as a unique value-added service for Fitbit users, making it less likely for them to switch to substitute products or services. By providing personalized food recommendations that align with users' dietary preferences and goals, the system can enhance user loyalty and reduce the likelihood of users seeking alternative solutions for their nutrition needs.

Why strategic, and what makes our product a quality information product

The proposed product is strategic for Fitbit for several reasons:

- The product aligns with Fitbit's mission of promoting healthy lifestyles and wellness by providing users with accurate and actionable nutritional information.
- The product offers a unique value proposition that differentiates Fitbit from its competitors and can be used to attract new customers.
- The product leverages the vast amounts of data available in the City of New York's dataset to provide users with personalized recommendations that are tailored to their specific needs and preferences.
- The proposed product is a quality information product for several reasons:
- The product is based on a comprehensive and reliable dataset provided by the City of New York, which ensures the accuracy and reliability of the information provided.
- The product uses advanced algorithms and machine learning techniques to process the data and provide personalized recommendations to users, which increases the product's effectiveness and relevance.
- The product provides users with actionable information that can help them make informed decisions about their diet and lifestyle choices, which is essential for maintaining good health and wellness.

Expected challenges and learning

- Ensuring the accuracy and reliability of the recommendations provided to users, given the complexity and variability of nutritional information.
- Designing an intuitive and user-friendly interface that makes it easy for users to input their preferences and view the recommended food items.

- Managing the vast amounts of data and ensuring that the system operates efficiently and effectively.
- The development of the proposed product is also expected to offer some learning opportunities, including:
- Developing a deeper understanding of the nutritional information and how it can be leveraged to promote healthy lifestyles and wellness.
- Gaining insights into user preferences and how they can be used to personalize recommendations and improve user engagement.
- Enhancing skills in data analysis, machine learning, and software development.

Data Sets

First Data Set: -

Source – NYC Government Data
See appendix A

The data set was chosen because it contains comprehensive information about menu items from various restaurants, including detailed nutritional content such as calories, fat content, carbohydrates, etc. This data set is relevant for our project, which focuses on analyzing the nutritional impact of restaurant menu items and identifying trends in food choices.

This data set can be used for strategic Intellectual Property (IP) purposes, such as developing proprietary algorithms or models for analyzing restaurant menu item data, creating innovative health or dietary applications, or generating insights for strategic planning in the food industry.

This data set is critical for our project as it provides crucial information about the nutritional content of restaurant menu items, which can significantly impact consumer choices and health outcomes. This data set is essential for projects related to public health, nutrition, or food policy.

Second Data Set: -

Source - Fitbit Dataset (Kaggle)
See appendix B

The data set was chosen because it contains data on activities, dates, and calorie burn from Fitbit fitness trackers. This data set is relevant for our project, which focuses on analyzing physical activity patterns, assessing fitness levels, or studying health behaviors.

This data set can be used for strategic IP purposes, such as developing proprietary algorithms or models for analyzing physical activity data, creating personalized fitness or health applications, or generating insights for strategic planning in the fitness or health industry.

This data set is important for our project as it provides data on individual activity levels, which can inform health outcomes and interventions related to health and wellness, fitness tracking, or physical activity interventions.

Both the data sets will be used for various purposes, including statistical analysis, data visualization, machine learning modeling, and generating insights for decision-making in relevant industries or fields. The data sets will help us gain insights into nutritional content, food choices, physical activity patterns, and health behaviors, which will inform strategic planning, policymaking, product development, or marketing strategies.

Assessment of Data Quality

The quality of the data sets will be assessed based on several factors, including data accuracy, completeness, consistency, timeliness, and relevance to the project objectives. Data accuracy will be assessed by comparing the data against trusted sources, verifying data entry processes, and identifying any outliers or inconsistencies. Data completeness will be assessed by checking for missing values, identifying any gaps in data coverage, and evaluating the representativeness of the data. Data consistency will be assessed by examining the format, structure, and coding of the data variables, and checking for any discrepancies or contradictions. Data timeliness will be assessed by evaluating the currency of the data, checking for any outdated or obsolete information, and considering the relevance of the data to the project timeline and objectives.

Plan for Improving Data Quality

To improve the quality of the data sets, several steps will be taken, including data cleaning, data validation, and data enrichment. Data cleaning will involve identifying and correcting errors, inconsistencies, or outliers in the data, and filling in any missing values using appropriate techniques, such as imputation or estimation. Data validation will involve verifying the accuracy and completeness of the data by cross-checking against trusted sources, conducting data audits, and validating data entry processes. Data enrichment will involve enhancing the data with additional information, such as external data sources, data transformations, or data aggregations, to augment the data's relevance and usefulness for the project. These steps will ensure that the data sets used in our project are of high quality and reliable for generating accurate and meaningful insights. Overall, our plan for improving data quality involves thorough data cleaning, validation, and enrichment processes to ensure that the data sets are reliable and relevant for designing our strategic and quality information product. Further processes involved may include data standardization.

Strategic Information Product:

The product has been refined to be a Food recommender system based on calorie inputs .This will act as new product line for Fitbit. Our target final product is a comprehensive food recommendation system that provides personalized suggestions for healthy food options based on the user's daily calorie limit and preferred food category. It will leverage the dataset provided by the City of New York on nutritional information of food items sold in restaurants.

Aim

The product aims to promote a healthier lifestyle by supporting users in making informed decisions about their nutrition, making it a valuable addition to Fitbit's suite of health and wellness products. It will empower users to take control of their health and make informed decisions about their diet.

This food recommendation system will allow users to set their daily calorie limit and preferred food category to generate personalized suggestions. Based on the dataset provided by the City of New York, the system will curate accurate and relevant recommendations that include detailed nutritional information for each food item. The visually appealing user interface will be accessible via mobile app and web portal and continuously updated with the latest nutritional information.

Consumers

The product is aimed at anyone looking to improve their health and wellness by adopting a healthier diet. It is especially relevant for Fitbit's existing user base, which includes health-conscious individuals who are already interested in tracking their fitness and wellness goals. In addition, we can also share these details with food delivery companies so that they can mention this information on their app. This will help their users to make more informed decisions about the food they are going to order, and promote a culture of healthy eating.

Activities done:

1) Exploration of relevant details for the planned recommender system:

Nutrition data: The project involves the use of nutritional data to provide accurate and relevant recommendations for healthy food options. The dataset containing the New York restaurant data, their menu and the nutritional value of the menu items will be apt for the model development.

Personalization: The second dataset contains the dataset of the Fitbit wherein the calories intake of athlete/person has been recorded for the period of 2 months. This calorie intake can be personalized and target calorie for the day tab can be created. At any time of day, based on the calorie consumption till that point of time, a real time food recommendation can be made for the rest of the day. The food recommendation system will be designed to offer personalized suggestions based on the user's daily calorie limit and preferred food category.

We understand that these available datasets will be sufficient to develop the recommender system.

2) Data Cleaning: To ensure the accuracy and relevance of the recommendations, we have conducted data cleaning on the New York restaurant menu data to remove any inconsistencies or errors.

<u>Some of the data cleaning steps that were taken are:</u>

1. In initial dataset, there were 65219 rows and 49 columns in our dataset

2. These columns contain integer, float and categorical values.

3. We also observed unique values in each column.

4. The columns Calories_text, Total_Fat_text, Saturated_Fat_text, Trans_Fat_text, Cholesterol_text,Sodium_text,Potassium_text,Carbohydrates_text,Protein_text,Sugar_text,Dietary_Fiber_text,Potassium,Potassium_100g,Serving_Size_text,Serving_Size,Serving_Size_Unit,Serving_Size_household do not provide relevant information for the model deployment. So these columns needed to be dropped.

5. Since the major analysis will revolve around the calories column. The rows that had NaN values in the calorie column were dropped from the dataframe.

6. There were missing values in 19 columns. Since these were mostly related to nutrients, the NaN cells were filled with value zero for analysis.

Upon processing, the dataset was left with 55315 rows and 32 columns. This cleaned dataset was exported to new excel sheet for further use.

The second dataset was clean and it did not require any data pre-processing.

<u>Implementation:</u>

There are several machine learning models that can be used for building a food recommendation system, including Collaborative Filtering, Content-based Filtering, Hybrid Models, Matrix Factorization, and Deep Learning Models.

Collaborative Filtering is based on the idea that users who have similar food preferences in the past will have similar preferences in the future. Collaborative filtering models can be memory-based or model-based, and they are widely used in food recommendation systems.

Content-based Filtering recommends food items to users based on the similarity between the attributes of the food item and the user's preferences. In this approach, the food items are represented by their features such as ingredients, nutritional value, and cuisine type.

Hybrid Models combine collaborative and content-based filtering to provide better recommendations. Hybrid models can overcome the limitations of individual models and provide more accurate recommendations.

Matrix Factorization is used to factorize the user-item rating matrix into two lower-dimensional matrices, representing users and items, respectively. This model is used in collaborative filtering to provide recommendations.

Deep Learning Models such as neural networks can also be used for building food recommendation systems. These models can learn complex relationships between food items and user preferences and provide accurate recommendations.

Choosing the appropriate model for the food recommendation system requires careful consideration of the specific use case and available dataset.

For this project, a suitable model could be Collaborative Filtering, as it takes into account the user's past preferences to make recommendations. The available dataset includes the nutritional data of menu items from various restaurants in New York City and the calorie intake of Fitbit users. Collaborative Filtering can leverage this data to recommend food items that match a user's calorie intake and nutritional preferences.

Collaborative filtering has several advantages over other recommendation systems such as content-based filtering. One of the main advantages is that it can recommend items that the user may not have considered before, based on the preferences of similar users. This can lead to serendipitous discoveries and a more diverse set of recommendations. Additionally, collaborative filtering can handle new users and items without requiring extensive knowledge about the user or item, as it relies on the similarity between users or items rather than their features.
To implement collaborative filtering for a food recommendation system, the following technical steps were taken:

- Data Collection: The first step is to collect data about the users and their food preferences. This data was collected from sources such as New York government website, restaurant menus etc.

- Data Preparation: The next step was to prepare the data for analysis. This involved cleaning the data, removing any irrelevant or duplicate data, and transforming the data into a format that can be used for collaborative filtering.

<u>Following steps can be taken to make the model:</u>

User-Item Matrix: The user-item matrix is a matrix that represents the interactions between users and items. In the context of a food recommendation system, this matrix would represent the rating or preference of each user for each food item. This matrix can be used to calculate the similarity between users or items.

Similarity Calculation: Similarity between users or items can be calculated using various metrics such as cosine similarity, Pearson correlation, or Euclidean distance. These metrics are used to find users or items that are most similar to each other.

Recommendation Generation: Once the similar users or items have been identified, recommendations can be generated for each user based on the preferences of similar users. The recommendations can be ranked based on the similarity score between the users or items.

In terms of the columns that can be used for collaborative filtering, the most important column would be the rating or preference column, which represents the user's opinion or preference for each food item. Other columns such as food category, restaurant name, and location can also be used to enhance the recommendations and provide more relevant suggestions to the user.

Since the existing dataset does not contain food preferences, we cannot directly use collaborative filtering techniques that require user ratings. However, we can use a variant of collaborative filtering known as item-based collaborative filtering.

In item-based collaborative filtering, similarities between items are computed based on the ratings given to those items by users. In our case, we can consider each menu item as an "item" and calculate the similarity between them based on the other features in the dataset.

One way to implement item-based collaborative filtering is to use the cosine similarity measure. The cosine similarity is a measure of the similarity between two non-zero vectors, which ranges from -1 to 1. A value of 1 indicates that the two vectors are identical, while a value of -1 indicates that they are completely dissimilar.

To implement item-based collaborative filtering using cosine similarity, we can follow these steps:

- Convert the menu items into a matrix where each row represents a menu item and each column represents a feature.

- Calculate the cosine similarity between each pair of menu items.

- For a given menu item, find the k most similar items based on their cosine similarity scores.

- Recommend the k most similar items to the user.

In our case, the features available in the dataset include the restaurant, food category, calories, total fat, saturated fat, cholesterol, sodium, carbohydrates, protein, sugar, and dietary fiber. We can use these features to compute the similarity between menu items.

However, it's important to note that item-based collaborative filtering has limitations. It may not capture complex user preferences or provide personalized recommendations. Therefore, we may need to supplement this approach with other techniques, such as content-based filtering or hybrid filtering, to improve the recommendations.

Dashboard:

First graph: Total number of restaurants in the dataset.

Second graph: Category count, providing an overview of the variety of cuisine types available. Third graph: Kids meal availability and count, allowing busy parents to quickly pinpoint family-friendly options.

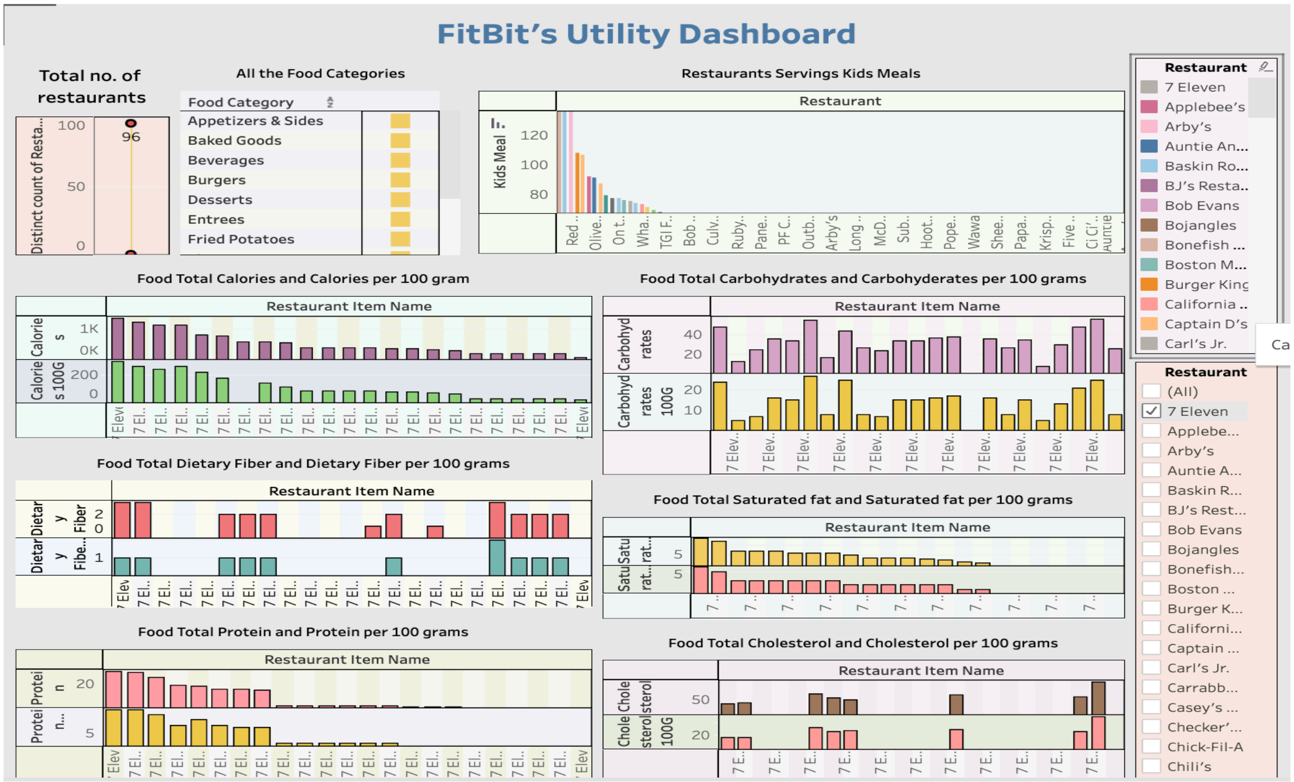
Fourth graph: Calorie content and per 100g, to help health-conscious diners find dishes that fit their dietary goals.

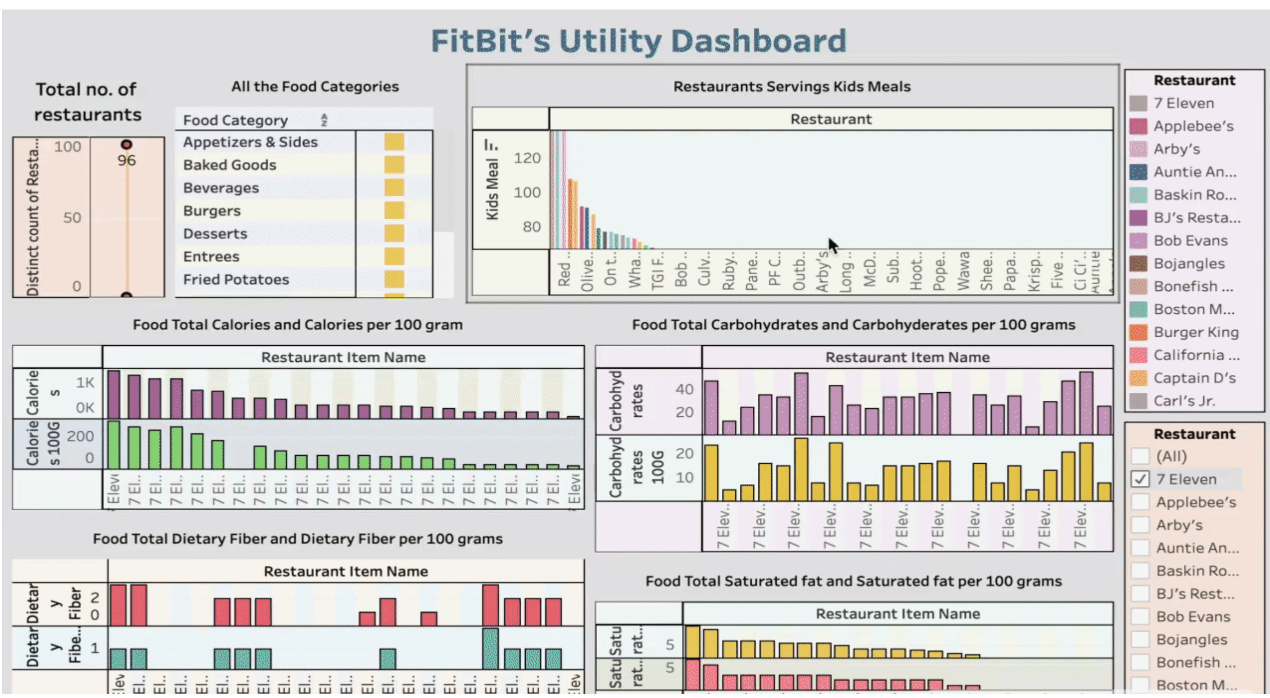
Other graphs: Sodium, Carbohydrates, Saturated Fat, Dietary Fiber, Protein, and Cholesterol, all presented per 100g of food as well, giving you an in-depth look at the nutritional content of dishes at each restaurant.

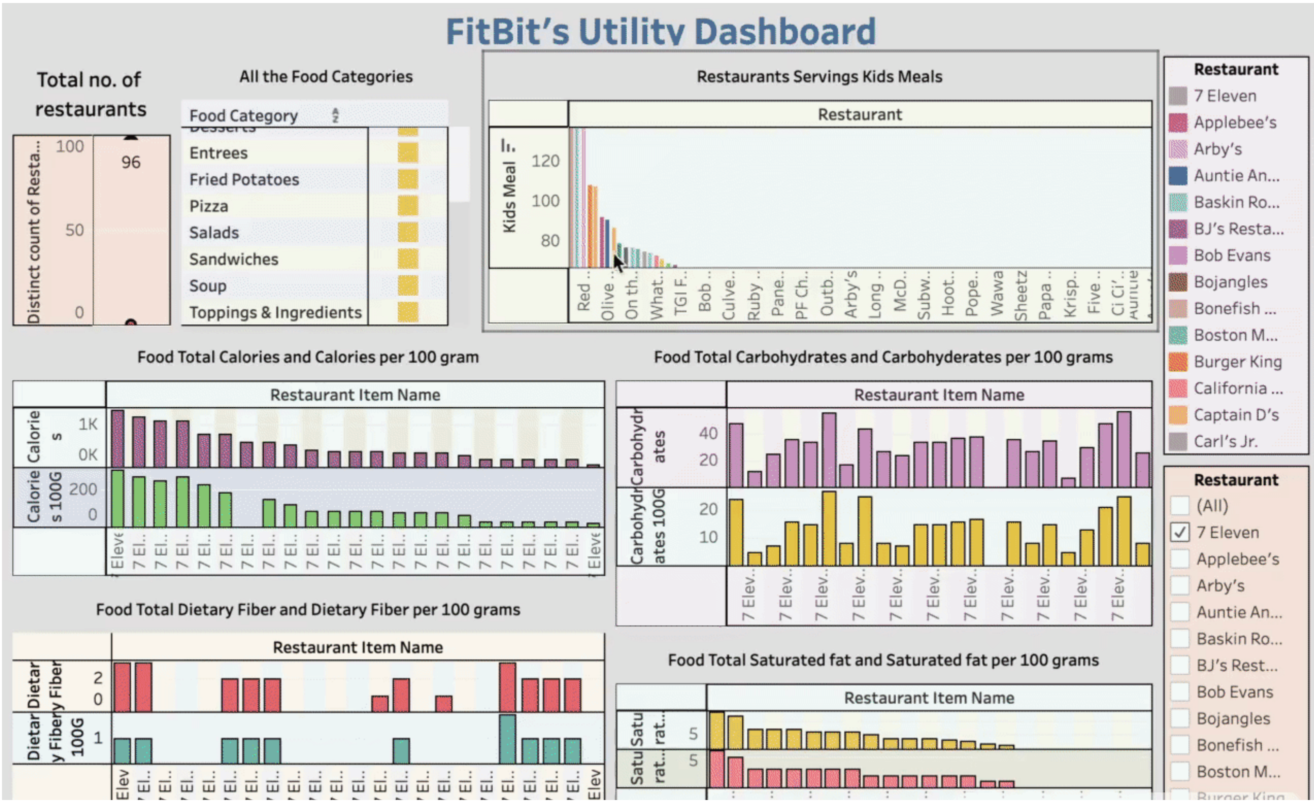
Filter: Easily switch between restaurants to compare and contrast nutritional info.

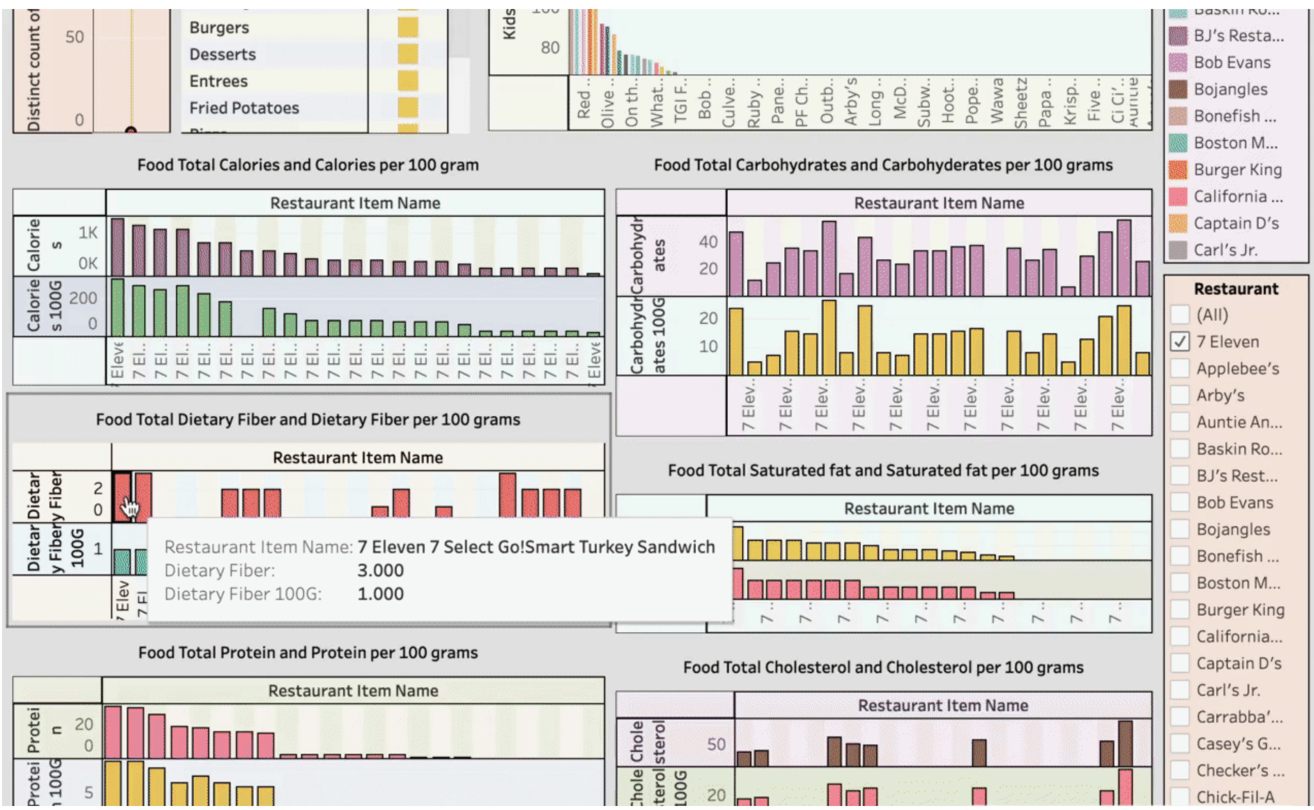
This dynamic dashboard isn't just helpful for customers - it's also a powerful tool for restaurants and food delivery apps. With access to detailed information about their offerings, restaurants can tailor their menus to meet the needs of their customers. Meanwhile, food delivery apps can use this data to provide more comprehensive information to customers, empowering them to make informed decisions about what they eat.

Tableau Dashboard Link


FitBit's Utility Dashboard
Total no. of restaurants
Distinct count of Resta...
100
96
50
0
All the Food Categories
Food Category
Appetizers & Sides
Baked Goods
Beverages
Burgers
Desserts
Entrees
Fried Potatoes
Restaurants Servings Kids Meals
Restaurant
Kids Meal
120
100
80
Food Total Calories and Calories per 100 gram
Restaurant Item Name
1K
0K
200
0
Food Total Carbohydrates and Carbohyderates per 100 grams
40
20
20
10
Food Total Dietary Fiber and Dietary Fiber per 100 grams
Food Total Saturated fat and Saturated fat per 100 grams
Food Total Protein and Protein per 100 grams
Food Total Cholesterol and Cholesterol per 100 grams
Restaurant
7 Eleven
Applebee's
Arby's
Auntie An...
Baskin Ro...
BJ's Resta...
Bob Evans
Bojangles
Bonefish ...
Boston M...
Burger King
California ...
Captain D's
Carl's Jr.
Restaurant
(All)
7 Eleven
Applebe...
Arby's
Auntie A...
Baskin R...
BJ's Rest...
Bob Evans
Bojangles
Bonefish...
Boston ...
Burger K...
Californi...
Captain ...
Carl's Jr.
Carrabb...
Casey's ...
Checker'...
Chick-Fil-A
Chili's


FitBit's Utility Dashboard
Total no. of restaurants
Distinct count of Resta...
100
96
50
0
All the Food Categories
Food Category
Appetizers & Sides
Baked Goods
Beverages
Burgers
Desserts
Entrees
Fried Potatoes
Restaurants Servings Kids Meals
Restaurant
Kids Meal
120
100
80
Food Total Calories and Calories per 100 gram
Restaurant Item Name
Food Total Carbohydrates and Carbohyderates per 100 grams
Food Total Dietary Fiber and Dietary Fiber per 100 grams
Food Total Saturated fat and Saturated fat per 100 grams
Restaurant
7 Eleven
Applebee's
Arby's
Auntie An...
Baskin Ro...
BJ's Resta...
Bob Evans
Bojangles
Bonefish ...
Boston M...
Burger King
California ...
Captain D's
Carl's Jr.
Restaurant
(All)
7 Eleven
Applebee's
Arby's
Auntie An...
Baskin Ro...
BJ's Rest...
Bob Evans
Bojangles
Bonefish ...
Boston M...


FitBit's Utility Dashboard
Total no. of restaurants
96
All the Food Categories
Food Category
Entrees
Fried Potatoes
Pizza
Salads
Sandwiches
Soup
Toppings & Ingredients
Restaurants Servings Kids Meals
Restaurant
Kids Meal
Food Total Calories and Calories per 100 gram
Restaurant Item Name
Food Total Carbohydrates and Carbohyderates per 100 grams
Food Total Dietary Fiber and Dietary Fiber per 100 grams
Food Total Saturated fat and Saturated fat per 100 grams
Restaurant
7 Eleven
Applebee's
Arby's
Auntie An...
Baskin Ro...
BJ's Resta...
Bob Evans
Bojangles
Bonefish ...
Boston M...
Burger King
California ...
Captain D's
Carl's Jr.
(All)


Burgers
Desserts
Entrees
Fried Potatoes
Food Total Calories and Calories per 100 gram
Food Total Carbohydrates and Carbohyderates per 100 grams
Food Total Dietary Fiber and Dietary Fiber per 100 grams
Restaurant Item Name
Restaurant Item Name: 7 Eleven 7 Select Go!Smart Turkey Sandwich
Dietary Fiber: 3.000
Dietary Fiber 100G: 1.000
Food Total Saturated fat and Saturated fat per 100 grams
Food Total Protein and Protein per 100 grams
Food Total Cholesterol and Cholesterol per 100 grams
Restaurant
(All)
7 Eleven
Applebee's
Arby's
Auntie An...
Baskin Ro...
BJ's Rest...
Bob Evans
Bojangles
Bonefish ...
Boston M...
Burger King
California...
Captain D's
Carl's Jr.
Carrabba'...
Casey's G...
Checker's ...
Chick-Fil-A

Conclusion:

In conclusion, the development of a food recommendation system using machine learning models has the potential to revolutionize the way we discover and choose food items from menus. By analyzing user behavior and preferences, these systems can provide personalized recommendations that increase user satisfaction and loyalty, while also benefitting businesses by increasing sales and customer retention.

The collaborative filtering model appears to be a promising approach for this project, given its ability to leverage user ratings and item similarity to make accurate recommendations. While the lack of explicit food preference data in the existing dataset and the user ratings presents a challenge, feature engineering and other techniques can be employed to extract relevant information and improve the model's performance.

Looking ahead, there are several avenues for future development and improvement of this food recommendation system. For example, incorporating additional data sources such as user demographics, location, and time of day could further enhance the system's personalization capabilities. Additionally, the use of deep learning models such as neural networks could provide even more accurate and nuanced recommendations.

However, there are also challenges to be addressed in implementing a food recommendation system. One of the main concerns is user privacy and data security, as the system relies on collecting and analyzing user data to provide recommendations. Ensuring that user data is protected and used ethically will be crucial to building trust and adoption of the system.
Overall, the development of a fitness centered food recommendation system has the potential to greatly improve the experience for users and benefit businesses. By addressing the challenges and leveraging the latest machine learning techniques, this project can pave the way for more effective and personalized food recommendation systems in the future.

Appendix A

Menu_Item_ID	Year	Restaurant_Item_Na	restaurant	Restaurant_ID	Item_Name	Item_Description	Food_Category	Serving_Size	Serving_Size_Unit
1197	2017	ox 4 Crispy C	lack in the Bo:	42	spy Chicken S	ken Strips, Chi	Entrees	195	g
2120	2017	Box 3 Stuffe	ack in the Bo:	42	tuffed Jalape	eeses, Snack	petizers & Sic	71	g
2121	2017	Box 7 Stuffe	ack in the Bo:	42	tuffed Jalape	eeses, Snack	petizers & Sic	165	g
2148	2017	the Box 1 Eg	ack in the Bo:	42	1 Egg Roll	& Spices, Sna	petizers & Sic	58	g
2149	2017	the Box 3 Eg	ack in the Bo:	42	3 Egg Rolls	& Spices, Sn	petizers & Sic	174	g
3117	2017	e Box Ketchu	ack in the Bo:	42	etchup, Pack	up, Packet, C	ings & Ingred	9	g
3181	2017	e Box Hot Ta	ack in the Bo:	42	lot Taco Sauc	Taco Sauce, C	ings & Ingred	7	g
3564	2017	ne Box 5 Mini	ack in the Bo:	42	5 Mini Churro	rros, Shakes	Baked Goods	86	g
3565	2017	e Box 10 Min	ack in the Bo:	42	0 Mini Churr	urros, Shakes	Baked Goods	172	g
17508	2017	e Box 2 Regu	ack in the Bo:	42	Regular Tac	se, Shredded	Sandwiches	162	g

Sugar	Dietary_Fiber	Calories_100g	Total_Fat_100g	Saturated_Fat_100	Trans_Fat_100g	Cholesterol_100g	Sodium_100g	Potassium_100g	Carbohydrates_100
0	3	289	12	2	0	30	811	281	27
2	1	308	17	6	0	24	1023	128	30
5	3	310	18	6	1	24	1027	128	30
2	2	253	12	3	0	12	547	250	26
6	7	254	13	3	0	13	547	249	26
2	0	111	0	0	0	0	944	0	33
0	0	0	0	0	0	0	857	214	0
12	2	403	21	3	0	7	321	66	49
25	4	403	22	4	0	7	322	66	49
2	5	210	12	4	0	15	444	222	20

Serving_Size_household	Calories	Total_Fat	Saturated_Fat	Trans_Fat	Cholesterol	Sodium	Potassium	Carbohydrates	Protein
4 Strips	563	24	3	0	58	1581	547	53	33
3 Pieces	219	12	4	0	17	726	91	21	6
7 Pieces	511	29	10	1	40	1694	212	49	14
1 Piece	147	7	2	0	7	317	145	15	5
3 Pieces	442	22	5	0	22	952	434	46	16
1 Packet	10	0	0	0	0	85	0	3	0
1 Packet	0	0	0	0	0	60	15	0	0
5 Pieces	347	18	3	0	6	276	57	42	4
10 Pieces	693	37	7	0	12	553	114	84	7
2 Tacos	340	19	6	0.5	25	720	360	33	12

Protein_100g	Sugar_100g	Dietary_Fiber_100	Kids_Meal	Limited_Time_Off	Regional	Shareable
17	0	2	0	0	0	0
8	3	1	0	0	0	0
8	3	2	0	0	0	0
9	3	3	0	0	0	0
9	3	4	0	0	0	0
0	22	0	0	0	0	0
0	0	0	0	0	0	0
5	14	2	0	0	0	0
4	15	2	0	0	0	0
7	1	3	0	0	0	0

Appendix B –

Id	ActivityDay	Calories
1503960366	4/12/16	1985
1503960366	4/13/16	1797
1503960366	4/14/16	1776
1503960366	4/15/16	1745
1503960366	4/16/16	1863
1503960366	4/17/16	1728
1503960366	4/18/16	1921
1503960366	4/19/16	2035
1503960366	4/20/16	1786
1503960366	4/21/16	1775