Draft report update: Aug 15, 2023

Project Contributors: Huan-Che Hsu (216498610)
Duy Duc Nguyen (217558354)
Luan Anh Tu (218277558)
Rojina Saberi (217354051)

Pricing Strategies and Consumer Behavior in McDonald's

Abstract:

This project examines pricing variations among selected products across different McDonald's chains in Toronto. By analyzing factors such as location proximity and consumer preferences, we aim to uncover the drivers behind these pricing strategies. Our study focuses on the correlation between downtown proximity and higher product prices, as well as consumer purchasing patterns. Our research contains Clustering and PCA(Principal Component Analysis) to investigate the fast-food chain pattern.

Introduction:

Our project focuses on analysing the variations in prices among selected products across different McDonald's chains. We will collect data on prices, locations, opening hours, and sector locations in Toronto. With this information, our objective is to investigate the factors that contribute to the differences in pricing strategies adopted by McDonald's for their products.

Problems and Goals:

1. First thesis

The objective of this thesis is to analyse and provide evidence to support the notion that when a McDonald's chain is located closer to or within the downtown area, the prices of its products tend to be higher compared to other areas.

2. Second thesis

By examining data on various combinations of food, we can gain valuable insights into consumer purchasing patterns, enabling us to delve deeper into analyzing consumer preferences.

Motivation for the Project:

The aim is to uncover the business strategies implemented by McDonald's and assess the efficacy of data in guiding businesses' decision-making processes.

The expectation of research result:

In the first thesis, our hypothesis suggests that the analysis of prices will reveal higher results when the chain restaurant is in closer proximity to the main districts. This is attributed to the increased rent and consumption in downtown areas, making it more likely for chains located there to have higher prices.

Additionally, the second thesis proposes that with the appropriate clustering and classification techniques, it is anticipated that popular products such as the Big Mac or fries will exhibit fewer price fluctuations. This suggests that these highly demanded items are likely to maintain relatively stable prices across different locations.

Formalization:

1. First Thesis

We will gather data on various McDonald's locations in the city of Toronto and employ the k-means clustering algorithm to group McDonald's restaurants into clusters based on their sector and their location in major Toronto Neighbourhoods such as Kensington Market, Fashion District and Financial District. We will further analyze the pricing patterns within each cluster using the Naive Bayes classifier. This approach allows us to explore and uncover patterns and similarities among the menu offerings of McDonald's within these neighbourhoods.

2. Second Thesis

We will gather data on the combination of foods in McDonald's locations and apply PCA (Principal Component Analysis), a dimensionality reduction technique that can be useful for preprocessing data with multiple food combinations. It transforms the original data into a lower-dimensional space while retaining the most important information. (combinations like fries, big mac combo, etc.), then transforms those data into different dimensional spaces (graph visualisation) (clustered) based on the customer's buying (number of transactions for each McDonald's Location/Zones) trend. As a result, it will be underlying patterns or combinations of food items that contribute most to the variation in the data.

Data plan:

- id(int) - a unique identifier for each record
- Price(int) - the price number of fast food products displayed on the mobile app
- Item Name(Char) - The item name that is associated with the price.
- Location(String) - The address of specific chains in String format
- District(Char) - The area where the specific chain located
- Sector(Char) - Indicates if the specific chains are located in Uptown, Midtown, or Downtown.
- Open Hour(Char) - This data will be a nominal data type that indicates if the specific chain is open 24 hours or not

For this project, we will gather data using the McDonald's online ordering application, specifically relying on the official prices and locations provided by the application. Additionally, we will collect location information using Google Maps.

Data Collection:

By integrating Google Maps and the online ordering system for McDonald's restaurants in Toronto, we were able to successfully gather information about their costs, locations, hours of operation, and sector regions. The information acquired is thorough and covers an extensive number of locations.

With the help of Kaggle raw data, we are able to collect the estimation of revenue based on the product combination. It plays an important role in terms of applying the PCA for visualizing the customer’s trend and behaviour.

Furthermore, in order to enhance the comprehensiveness of our study, we have incorporated a diverse array of menu selections, thereby enriching our dataset and facilitating more robust testing procedures. This deliberate inclusion of a wide range of menu items ensures a comprehensive exploration of pricing patterns and consumer behaviours, bolstering the reliability and validity of our analytical findings.

Pricing Strategies and Consumer Behavior in McDonald's

Project Overview

The research "Pricing Strategies and Consumer Behavior in McDonald's" intends to examine how product prices differ throughout the several Toronto-area McDonald's chains. We attempt to find factors impacting pricing strategies and obtain insights into customer purchase habits by analyzing data on prices, locations, and consumer preferences. This study uses clustering, classification, and Principal Component Analysis (PCA) to find useful data that firms can use to comprehend client preferences and make data-driven decisions.

Data Preprocessing: (both 1st and 2nd Thesis)

The data was gathered, and we ensured its accuracy and consistency by cleaning and preprocessing it. To get the data ready for future analysis, the preprocessing absorbs and handles missing values, normalizes numerical features, and encodes categorical variables

Challenges Overcome

As the project moved forward, we ran into a few issues that needed consideration and management. Dealing with outliers in the obtained data, which had an impact on the reliability of the first clustering results, was one of the main obstacles. However, we were able to enhance the clustering analysis by applying techniques for identifying outliers and enhancing the preprocessing techniques.

For the 2nd Thesis, it was hard to format the raw data as some of the food prices were different and varied based on 5 locations, which was hard to manually find the combination food price... However, we were able to find it and also continue to apply the PCA with specific variants components for clustering.

Evaluation of measures:

First thesis:

- Assignment Step: For each data point, calculate the distance to each centroid. Assign the data point to the cluster whose centroid is closest.
- Update Step: Recalculate the centroids of each cluster by taking the mean of all data points assigned to that cluster.
- Iteration: Repeat the assignment and update steps until convergence criteria are met (e.g., centroids don't change significantly or a maximum number of iterations is reached).
- Output: The algorithm produces K clusters, with each data point assigned to one of these clusters.

Second thesis:

- Principal Component Analysis (PCA): Apply Principal Component Analysis (PCA) to the normalized dataset to reduce its dimensionality. Calculate the covariance matrix of the dataset.
- Dimensionality Reduction: Choose the top 'k' eigenvectors (principal components) that collectively capture a significant portion of the data's variance. Project the original data onto the 'k' selected eigenvectors to obtain a lower-dimensional representation of the data.
- Visualization and Clustering: Visualize the lower-dimensional data points using scatter plots or other graph visualization techniques. Implement a clustering algorithm (e.g., K-Means, DBSCAN) on the lower-dimensional data to identify clusters of McDonald's locations/zones with similar buying trends.
- Interpretation and Insights: Analyze the clusters to identify common customer buying trends, i.e., specific combinations of food items that tend to be purchased together more frequently. Interpret the principal components to understand which original food combinations contribute most to the variance in the lower-dimensional representation.

Demonstration tests:

Utilizing the first thesis for clustering purposes necessitates the preprocessing of the data sourced from McDonald's. Subsequently, the computation of spatial disparities between each restaurant and the downtown center will be conducted, followed by the assignment of price categories to their respective restaurants. It appears that we have acquired pricing data from various sources, followed by the calculation of distances between each restaurant and the central area of Downtown Toronto. This process enables us to assemble the values for the X-axis and Y-axis in our clustering analysis. Consequently, we will create visual representations by mapping these variables onto price points and the corresponding distances to the center of Toronto.

```python
# Data Preprocessing
data['BigMac'] = pd.to_numeric(data['BigMac'])
data['BigMacCombo'] = pd.to_numeric(data['BigMacCombo'])
data['Filet-O-Fish'] = pd.to_numeric(data['Filet-O-Fish'])
data['Filet-O-Fish combo'] = pd.to_numeric(data['Filet-O-Fish combo'])
data['Small Fries'] = pd.to_numeric(data['Small Fries'])
data['Distance from root'] = pd.to_numeric(data['Distance from root'])
```

```python
# Calculate the correlation coefficient of the 4 items with the distance
correlation_bigmac = data['BigMac'].corr(data['Distance from root'])
correlation_bigmaccombo = data['BigMacCombo'].corr(data['Distance from root'])
correlation_filetofish = data['Filet-O-Fish'].corr(data['Distance from root'])
correlation_filetofish_combo = data['Filet-O-Fish combo'].corr(data['Distance from root'])
```

```python
# Print Correlation Coefficients
print(f"Correlation between BigMac and Distance from root: r = {correlation_bigmac:.4f}")
print(f"Correlation between BigMacCombo and Distance from root: r = {correlation_bigmaccombo:.4f}")
print(f"Correlation between Filet-O-Fish and Distance from root: r = {correlation_filetofish:.4f}")
print(f"Correlation between Filet-O-Fish combo and Distance from root: r = {correlation_filetofish_combo:.4f}")
```

```
Correlation between BigMac and Distance from root: r = -0.5265
Correlation between BigMacCombo and Distance from root: r = -0.5920
Correlation between Filet-O-Fish and Distance from root: r = -0.7206
Correlation between Filet-O-Fish combo and Distance from root: r = -0.6972
```

```python
# Extract the relevant columns for clustering
food_items = ['BigMac', 'BigMacCombo', 'Filet-O-Fish', 'Filet-O-Fish combo', 'Small Fries']

# Set up the grid of subplots
fig, axes = plt.subplots(1, len(food_items), figsize=(18, 6))
fig.suptitle('K-means Clustering of Food Item Prices', fontsize=16)

# Perform k-means clustering for each food item and create subplots
for idx, item in enumerate(food_items):
    # Extract the data for the current food item
    food_data = data[['Distance from root', item]]

    # Perform k-means clustering
    k = 3  # Number of clusters
    kmeans = KMeans(n_clusters=k, random_state=0)
    clusters = kmeans.fit_predict(food_data)

    # Add the cluster labels to the DataFrame
    food_data['Cluster'] = clusters
```

Once we've obtained the distances from the root to each distinct restaurant, we can initiate the process of clustering variables. This will allow us to explore the connection between the restaurant distance and the pricing of individual items, along with assessing how this interplay influences the item's price.

Expanding upon the second thesis, our investigation will harness the dataset accessible on Kaggle, specifically centred around McDonald's sales revenue. This dataset will serve as the foundation for our endeavour to unveil customer tendencies, extracting knowledge that facilitates a comprehensive grasp of their predilections and buying behaviours. Given the voluminous data offered by each restaurant, we will implement a random selection process encompassing five distinct establishments, subsequently augmenting the dataset's depth through Principal Component Analysis (PCA).

```python
Res_data = pd.DataFrame(data=df['price'] ,
         columns= ['price'] + ['serve size'] + ['calories'] + ['Protein'] + ['Monthly Consumption'] + ['Revenue'])
Res_data['price'] = df['price']
Res_data['serve size'] = df['serve size']
Res_data['calories'] = df['calories']
Res_data['Protein'] = df['Protein']
Res_data['Monthly Consumption'] = df['Monthly Consumption']
Res_data['Revenue'] = df['Revenue']
# np.array

print(Res_data) # object Res_data
Res_data.info() # Checking to make sure Dtype are all Float


# In[23]:


Res_data.isnull() # make sure all the Data type is not null
print([Res_data])


# In[24]:


from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()

# Scaler to Fit into data :  vào dữ liệu
scaler.fit(Res_data)

# Performing Transform Scale for Res_data (5 columns) :
# Thực hiện transform scale
cd = scaler.transform(Res_data)
```

Results and Findings:

Utilizing our analytical frameworks, we have discerned discernible trends in product attributes that correspond to their geographical placement. Across the majority of graphs, a consistent pattern emerges: as the distance from the central reference point decreases, there is a proportional inclination for prices to be elevated. This observation potentially lends support to the hypothesis that prices of McDonald's items exhibit a positive correlation with their proximity and spatial orientation relative to the center of Toronto.

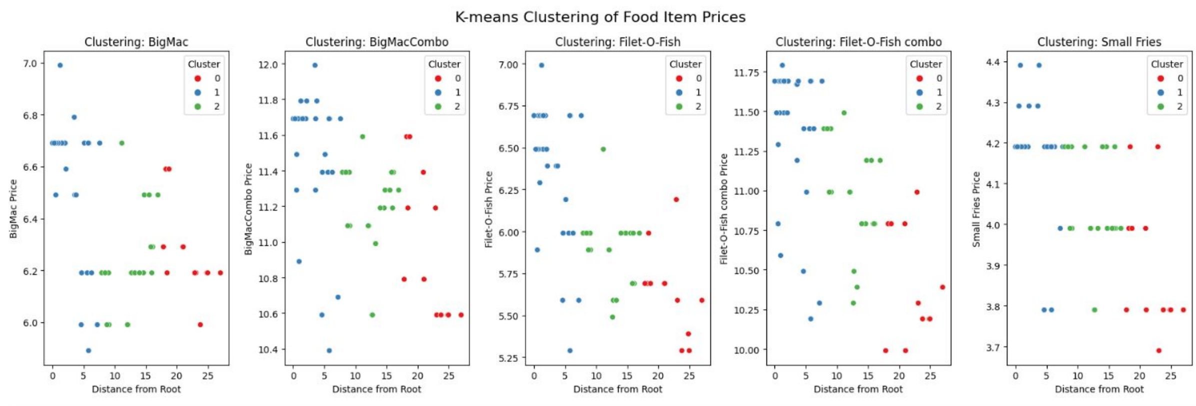

Figure 1: Although individual items exhibit varying outcomes on the scale, a consistent pattern emerges from the graphs, demonstrating a noticeable negative correlation from the upper-left to the lower-right quadrant. Clustering the data reveals that establishments situated in downtown and midtown areas consistently feature higher prices, while uptown restaurants tend to offer comparatively lower average prices. However, it is worth noting that some uptown restaurants also present outliers, with pricing equivalent to that of their downtown and midtown counterparts.

Transitioning to the realm of customer behaviors, our analysis has revealed a lack of substantial correlation between individual menu items and specific restaurant locations. While the dataset presents a scattering of outliers and encompasses diverse data variables, a subtle negative correlation is perceptible from the upper left to the middle bottom quadrant. Despite its seemingly modest nature, this correlation, albeit insignificant, holds potential for analysts to extract insights that could inform enhanced enterprise-level management strategies, thereby facilitating performance augmentation across all restaurants.

Evaluation:

Using our own collected data has been really helpful. It made sure that the information we gathered was just right for our project's goals. We could dig deep into how prices and customer behaviours vary among different McDonald's chains in Toronto. However, it is worth noting that the absence of extensive coding knowledge posed challenges during the clustering phase, which ultimately led to suboptimal results. Still, this project has been a great learning experience. We now understand the importance of having well-organized data, and we've gotten better at understanding complex data. This will definitely help us in the future when we analyze data again.

Conclusion:

The research has made substantial progress in studying price fluctuations and consumer preferences across various McDonald's franchises. We got important insights into factors impacting pricing strategies and found popular food combinations that maintain stable prices through diligent data collecting, preprocessing, and advanced analytical approaches such as clustering and dimensionality reduction. The study's findings highlighted the importance of different areas in determining price discrepancies and provided insight into the influence of location distance on product pricing. Additionally, we can better comprehend customer decisions because of the underlying trends in consumer preferences that our study of customer purchase patterns has shown.

References:

Our Mcdonald price data:

First thesis

2nd Mcdonald Price.xlsx

Second thesis

Thesis 2 - Google Sheets