**What is Eatmarna**

Eatmarna is a governmental application developed by Ministry of Hajj and Umrah for the service of the guests of Rahman,It’s one of the most popular applications in Saudi Arabia, used to enable those wishing to perform Umrah and visit to request the issuance of permits to enter the Two Holy Mosques to perform Umrah, visit and prayers according to the capacity approved by the concerned authorities to ensure the provision of a spiritual and safe atmosphere that achieves health precautionary measures and controls And the regulatory system.

**Problem Definition and Goal**

In twitter Account (@MOHU\_CSC) there are considerable variety in the customers’ opinion of the provided services. These hashtags describe the customer's experience with such as #اعتمرنا

The goal is to figure out the customers registration problems, take their opinion about the Eatmarna application performance and if the customer services team have high responsiveness or not.

**Objective of Analytical Solution**

* In this project, we aim to analyze number of tweets to achieve several objectives, such as:
* Describe the problems related to the registration process.
* Measure Customers’ satisfaction regarding customer services team responsiveness.
* Measure Customers’ satisfaction about the application performance.

**This project contains four steps**

# **Data collection**

I connected to Twitter API, extracting the data I need about Eatmarna, the libraries/tools that we will use in this phase: pandas, and NumPy.

The data collection mechanism extracted tweets to people mentioning (@MOHU\_CSC), searching about specific keywords (تطبيق اعتمرنا) and extract tweets from a hashtag (#اعتمرنا). After observing the data, we have noticed that most of the text of the tweets has been written in the Arabic language. Moreover, the data revealed the key factors and issues contributing to customer satisfaction, which we will further analyze in the subsequent phases.

## **Data Exploration and Investigation**

To ensure accurate analysis of the data and to truly know people's opinions, collected a total of 27870 tweets from different ways of search. For data collection mechanism, I looked up for:

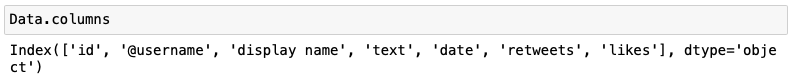
* tweets containing the keyword ("تطبيق اعتمرنا").
* tweets for people mentioning (@MOHU\_CSC).
* tweets that contain specific hashtag (#اعتمرنا).

**The total amount of the Eatmarna dataset:**

Graphical user interface, text, application, email

Description automatically generated

**The columns list of datasets:**

****

**The data type of each column:**

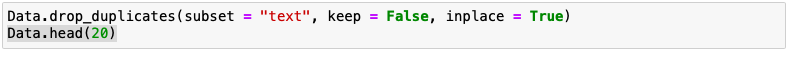
Graphical user interface, text, application

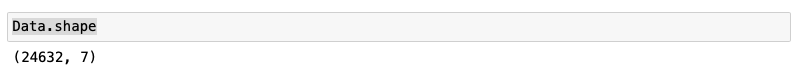
Description automatically generated

# **Data Preprocessing**

## **Remove duplicate tweets**

there are several duplicate tweets that needed to be removed. Pandas’ library was used to fix this issue.

**Applied (drop\_duplicates) method on ‘text’ column for removing the duplicate tweets.**

**The total amount of the dataset after removing the duplicate tweets**

## **Remove mentions, hashtags, lines, taps, and links from tweet text**

To decrease the length of tweets text and keeping only the needed valuable part of data so, we need to remove mentions, hashtags, lines, taps and links from tweets since they are redundant and unneeded for analysis in our case. Pandas library was used for handling the dataset.

**Applied (replace) method on ‘text’ column to remove the hashtags, lines, taps, and links from tweet text by replacing them with empty string.**

Graphical user interface, text, application

Description automatically generated

**appended the result to the ‘text’ column.**

****

## **Remove punctuations**

To make the tweet text more clear, readable, and easier to analyze for a better-quality result, we need to removed punctuations from the tweet text.

**Defined the Arabic punctuation characters and used the English punctuation characters defined method. Then we defined the (remove\_ punctuation) method for removing them from each tweet text in the ‘text’ column.**

Graphical user interface, text, application, email

Description automatically generated

**appended the result to the ‘text’ column.**



## **Remove Arabic stop-words**

Arabic stop words don't add any value to the analysis, for that we remove it using nltk library.

Graphical user interface, text, application, email

Description automatically generated**Imported nltk library. Then downloaded all Arabic stop words. Then took each tweet text from ‘text’ column and checked if it contained any stop word then remove that word from the tweet text.**

## **Tokenize the ‘text’ column**

An empty text will produce some errors in the code, for that first removed all empty cells in ‘text’ column after that:

a. tokenizes “text” column using (tokenize) method from nltk library

b. Then, assigned the result to a new column called “tokens”.”.

Graphical user interface, text, application, email

Description automatically generated

After preprocessing data, the total of data 3478 tweets that satisfy our issues solution, and 8 columns contain the data we will use in our analytical project.

Graphical user interface, application

Description automatically generated with medium confidence

# **Model Planning and Building**

In this phase, I will conduct descriptive and predictive analytics on our dataset to help reach the goal we have introduced in the Problem Definition and Goal.

For descriptive-analytic, I will provide a summary of collected data like the mean, variance, standard deviation, etc. The libraries/tools that we will use in descriptive-analytic: pandas, matplotlib. pyplot, and seaborn.

For predictive analytics, we will choose two suitable models for our data such as navïe bayes, Logistic Regression, etc. In addition, we may apply data mining techniques.

# **Communicate Results**

In the final phase, we will visualize our findings based on the initial hypotheses that I assumed. Moreover, based on the outcome of data analysis, especially from the Model Planning and Building phase, I would give a brief description about if we achieved our goal or not, and then we will give recommendations on how to further the research project.

The libraries/tools that we may use in this phase: matplotlib library.