simplilearn | PURDUE UNIVERSITY

Data Analyst Capstone Project

Project 1: Marketing

Problem Statement

A restaurant consolidator is looking to revamp the B2C portal using intelligent automation tech. This requires a different matrix to identify the **star** restaurants and generate recommendations. To make sure an effective model can be achieved, it is important to understand the behavior of the data in hand.

There are 2 datasets, **data** and **Country -Code**. Dataset **data** has 19 attributes and **Country -Code** has two attributes.

You are required to determine certain matrices to identify the **star** restaurants and generate recommendations.

Dataset Snapshot

Dataset: data

Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude
7402935	Skye	94	Jakarta	Menara BCA, Lantai 56, Jl. MH. Thamrin, Thamrin, Jakarta	Grand Indonesia Mall, Thamrin	Grand Indonesia Mall, Thamrin, Jakarta	106.821999	-6.196778
7410290	Satoo - Hotel Shangri-La	94	Jakarta	Hotel Shangri-La, Jl. Jend. Sudirman	Hotel Shangri-La, Sudirman	Hotel Shangri-La, Sudirman, Jakarta	106.8189611	-6.203291667
7420899	Sushi Masa	94	Jakarta	Jl. Tuna Raya No. 5, Penjaringan	Penjaringan	Penjaringan, Jakarta	106.800144	-6.101298

Dataset Snapshot

Dataset: data

Cuisines	Average Cost for Two	Currency	Has Table Booking	Has Online Delivery	Price Range	Aggregate Rating	Rating Color	Rating Text	Votes
Italian, Continental	800000	Indonesian Rupiah(IDR)	No	No	3	4.1	Green	Very Good	1498
Asian, Indonesian, Western	800000	Indonesian Rupiah(IDR)	No	No	3	4.6	Dark Green	Excellent	873
Sushi, Japanese	500000	Indonesian Rupiah(IDR)	No	No	3	4.9	Dark Green	Excellent	605

Dataset Snapshot

Dataset: Country-Code

Country Code	Country
1	India
14	Australia
30	Brazil

Dataset Snapshot

Variable Description

Variable	Description
Restaurant ID	Identification number
Restaurant Name	Name of the restaurant
Country Code	Country code
City	City name
Address	Address of the restaurant
Locality	Short address of the restaurant
Locality Verbose	More detailed address of the restaurant
Longitude	Shows longitudinal location
Latitude	Shows latitudinal location

Dataset Snapshot

Variable Description

Variable	Description
Cuisines	Types of cuisines served
Average Cost for two	Average cost if two people visit the restaurant
Currency	Local currency
Has Table booking	Can we book tables in restaurant? Yes/No
Has Online delivery	Can we have online delivery ? Yes/No
Price range	Categorized price between 1-4
Aggregate rating	Categorizing ratings between 1-5
Rating color	Different colors representing customer rating
Rating text	Different rating like excellent, very good ,good, avg., poor, not rated
Votes	No.of votes received by restaurant from customers.

Project Task: Week 1

Importing, Understanding, and Inspecting Data :

1. Perform preliminary data inspection and report the findings as the structure of the data, missing values, duplicates, etc.

2. Based on the findings from the previous questions, identify duplicates and remove them

Project Task: Week 2

Performing EDA:

1. Explore the geographical distribution of the restaurants and identify the cities with the maximum and minimum number of restaurants
2. Restaurant franchising is a thriving venture. So, it is very important to explore the franchise with most national presence
3. Find out the ratio between restaurants that allow table booking vs. those that do not allow table booking
4. Find out the percentage of restaurants providing online delivery
5. Calculate the difference in number of votes for the restaurants that deliver and the restaurants that do not deliver

Project Task: Week 3

Performing EDA:

1. What are the top 10 cuisines served across cities?
2. What is the maximum and minimum number of cuisines that a restaurant serves? Also, which is the most served cuisine across the restaurant for each city?
3. What is the distribution cost across the restaurants?
4. How ratings are distributed among the various factors?
5. Explain the factors in the data that may have an effect on ratings. For example, number of cuisines, cost, delivery option, etc.

Project Task: Week 4

Dashboarding:

1. Visualize the variables using Tableau to help user explore the data and create a better understanding of the restaurants to identify the ‘’star’’ restaurant

2. Demonstrate the variables associated with each other and factors to build a dashboard