

Coursera Capstone Project for IBM Data Science Specialization - Week 1

By Aravindan Natarajan

1. Introduction

1.1. Background

Dallas city is one of the most populous cities in U.S. and is home to many immigrant populations in Texas after San Antonio and Houston. Furthermore, it is the fourth-largest metropolitan area in the U.S. at 7.5 million people as of 2018 with an estimated population of 7,846,293 residents. Being a metropolitan city, Dallas is also home to many restaurants which serves wide variety of cuisines. Owing to significant number of Indian expatriate population, Dallas City and its nearby Suburbs have handful of Indian restaurants.

If someone from India visits Dallas City for the first time, it will be useful if he/she have some prior information about the Indian Restaurants in Dallas City and how good they are. Moreover, prior information on location of other restaurants and their violation history will help in coming to an informed decision.

So, as a part of this project using the Dallas City Inspection Data and FourSquare API Indian restaurants in Dallas City will be listed, visualized and rated.

1.2. Problem Description

By utilizing the Dallas City restaurants inspection data, Indian Restaurants in Dallas City and their risk category will be Analyzed. Secondly, a classsifier model will be built to predict the risk categories of restrutants. Furthermore, using the foursqure API the ratings of Indian Restaurants in Dallas City will be obtained.

1.3. Target Audience

- People looking to open new restaurants
- Restaurants
- Travellers who love Indian food

2. Data

For this project I will use the following data :

1. Dallas City restaurants inspection data from 2016-2019
 - Data source : <https://www.dallasopendata.com/api/views/dri5-wcct/rows.csv?accessType=DOWNLOAD>
 - Description : This data set contains 37876 rows and 114 coulms with Restaurant Name, Street Name, violation descriptions along with their latitude and longitude. This data will be downloaded and used.

	Restaurant Name	Inspection Type	Inspection Date	Inspection Score	Street Number	Street Name	Street Direction	Street Type	Street Unit	Street Address	Zip Code	Violation Description - 1	Violation Points - 1	Violation Detail - 1	Violation Memo - 1	Violation Description - 2
0	HARVEY'S	Routine	10/03/2016	82	12835	PRESTON	NaN	RD	#306	12835 PRESTON RD #306	75230	*34 Outer door: solid,selfclosing,tightfitting	1.0	228.174 Physical Facilities, Function...	NaN	*29 Concentration of the sanitizing solution s...
1	7-11	Routine	10/03/2016	86	5123	LOVERS	W	LN	NaN	5123 W LOVERS LN	75209	*18 Chemical sanitizer generated onsite, devic...	3.0	228.111 Equipment, Utensils, and Linens. ...	NaN	*03 Food products not maintained at 135oF or a...
2	MI HONDURAS	Routine	10/03/2016	75	10818	DENNIS	NaN	RD	#101	10818 DENNIS RD #101	75229	*14 When to wash hands after handling soiled e...	3.0	228.38 Management and Personnel (d)...	All hand sinks were blocked	*21 RFMS - Not On Site
3	TORTILLERIA LA ESPIGA	Routine	10/03/2016	82	1328	JIM MILLER	N	RD	#104	1328 N JIM MILLER RD #104	75217	*39 In-use utensils, between-use storage. Duri...	1.0	228.68 Food Preventing Contaminatio...	cannot use water containers to keep utensils ...	*20 Grease Trap Tickets
	TMGM @									2323 N				228.35	missing	*22 Handlers...

Fig.1. Snapshot of the Dallas City Restaurants Inspection data loaded into a data frame

This dataset contains most of the information that will be needed for the project such as location information, street name, etc., However, this dataset contains mixed datatypes in all 114 columns and needs extensive cleaning before it can be used for the project.

```

Restaurant Name      object
Inspection Type      object
Inspection Date      object
Inspection Score     int64
Street Number        int64
Street Name          object
Street Direction     object
Street Type          object
Street Unit          object
Street Address       object
Zip Code             object
Violation Description - 1  object
Violation Points - 1     float64
Violation Detail - 1    object
Violation Memo - 1     object
Violation Description - 2  object
Violation Points - 2     float64
Violation Detail - 2    object
Violation Memo - 2     object
Violation Description - 3  object
Violation Points - 3     float64
Violation Detail - 3    object
Violation Memo - 3     object
Violation Description - 4  object
Violation Points - 4     float64
Violation Detail - 4    object

```

Fig.2. A snapshot of datatypes of the columns in the dataframe.

2. Ratings of Indian resturants for selected locality in Dallas City
 - Data source : FourSquare API
 - Description : By using this api we will get all the ratings for Indian restaurants in selected neighbourhood