

## **Problem Statement for Data Competition April 2019**


The city of Los Angeles publishes data on Environment Health inspection and enforcement results from restaurants in the Los Angeles county. These data cover 85 of 88 cities and all unincorporated areas in the LA county.

There are two datasets available: (i) market inspection dataset: contains results of inspection; and (ii) market violations dataset: contains information on health code violations in restaurants. These data were last updated on January 16, 2019. Data dictionaries for the above two data sets are included below. Feel free to supplement the above information with other publicly-available information.

We are interested in understanding the following questions:

- (1) What are the key factors in predicting health “scores” of the restaurants in Los Angeles county?
- (2) What are the most important factors in classifying restaurants into different “grades”?
- (3) Are there any relationships between various types of health code violations and scores/grades of a restaurant?
- (4) Are there any patterns in terms of how health scores of restaurants change over time?

Data dictionary for the inspection data set:

Columns in this Dataset				
Column Name	Description	Type		
ACTIVITY DATE	Date of the inspection	Date & Time		▼
OWNER ID	Represents a specific owner record	Plain Text	T	▼
OWNER NAME	Name of the legal owner	Plain Text	T	▼
FACILITY ID	Represents a specific facility record	Plain Text	T	▼
FACILITY NAME	Name of the facility	Plain Text	T	▼
RECORD ID	Unique ID number for each health program at a fa...	Plain Text	T	▼
PROGRAM NAME	A unique name for each program. There may be m...	Plain Text	T	▼
PROGRAM STATUS	Indicates the status of the program	Plain Text	T	▼
PROGRAM ELEMENT (PE)	The code identifies the type of program	Plain Text	T	▼
PE DESCRIPTION	Describes the type of program	Plain Text	T	▼
FACILITY ADDRESS	Address of the facility	Plain Text	T	▼
FACILITY CITY	City of the facility	Plain Text	T	▼
FACILITY STATE	State of the facility	Plain Text	T	▼
FACILITY ZIP	Zip code of the facility	Plain Text	T	▼
SERVICE CODE	Identifies the service performed	Plain Text	T	▼
SERVICE DESCRIPTION	Describes the service that was performed	Plain Text	T	▼
SCORE	The score of the inspection	Number	#	▼
GRADE		Plain Text	T	▼
SERIAL NUMBER	Unique ID for each inspection	Plain Text	T	▼
EMPLOYEE ID	ID number of the employee who did the inspection	Plain Text	T	▼

Data dictionary for the violations data set:

Columns in this Dataset

Column Name	Description	Type		
SERIAL NUMBER	Unique ID for each inspection	Plain Text	T	▼
VIOLATION STATUS	Identifies the status of the violation	Plain Text	T	▼
VIOLATION CODE	The code of the violation	Plain Text	T	▼
VIOLATION DESCRIPTION	A description of the violation	Plain Text	T	▼
POINTS	The amount of points deducted from the score for a violati...	Number	#	▼