

1. Data Section

Following sources of data are used while executing the Insurance Zip Code Risk Model Project: -

- **Data title: -**
Open Data published by the New York City Police Department under the section **Collisions**.

Main use: Create target to train the risk predictive model.

Type of datasource: -

JSON File by using API Rest

Time Interval: -

From January 2018 until today

Description of the dataset: -

The New York Collisions dataset has the last year crash accidents from the city. It is possible to identify collisions per borough and per ZIP Code.

The dataset was downloaded in JSON Format by using an API Rest query to the New York City open data hub.

It is important to know that we need the collisions resume by Zip Code. The Resume needs the zipcode id, and the number of accidents and a new variable identifying whether a zip code is risky and whether not. The last variable is created by using the two highest deciles to represent a high risk and others.

Source: -

<https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-Collisions-Crashes/h9gi-nx95>

- **Data title: -**
Google Maps Geocoding API

Main use: Get postalcode and create radius variable for both datasets: training and scoring.

Type of data: -

JSON File by using API Rest

Duration: -

N/A

Description of the data: -

Location coordinates obtained by Gmaps API calls.

PostalCode from New York City and Guadalajara City is used to obtain the location coordinates from Google Maps. With Gmaps location coordinates the Radius metric per each Postal Code is created.

Source: -

Google Cloud Platform/ Google Maps

- **Data title: -**

Foursquare venues data

Main use: Get venues information for both training and scoring datasets.

Type of data: -

JSON File by using API Rest

Duration: -

N/A

Description of the data: -

Venues information obtained by Foursquare API calls.

To determine the density of total amenities per each Postal Code and the density of nightlife places.

Source: -

<https://foursquare.com/>

- **Data title: -**

Correos de México: Mexican Post office

Main use: Get postalcodes for the scoring dataset.

Type of data: -

CSV File

Duration: -

N/A

Description of the data: -

PostalCodes for Guadalajara City.

Useful to get location coordinates by using Gmap API and to get Foursquare Venues Information. The Guadalajara City dataframe Will be used to score the risk per postalcode.

Source: -

<https://www.correosdemexico.gob.mx/>