

## **Business Problem:**

An important world wide insurance company has recently opened an affiliated company in Mexico. This mexican subsidiary has created a new business process to measure the inherent risk for a person by using their personal information: age, gender, car information, such as model and vehicle type, and their Zip Code risk information. Unfortunately, in Mexican police departments and services has a lack in capture data. For instance, there is not a public vehicular accidents database.

In the researchgate is possible to identify an alternative way to measure the risk per ZipCode. In a couple of papers in which conclusions claims that exist a higher correlation between foursquare and other social media data with car accidents and crimes. So, the company has decided to use a public information from a known city to train a Machine Learning model to qualify wheather a zip code is risky and wheather not. The information needed to train the model is the public New York Police Deapartment accidents database and Foursquare venues and ZipCode radius. Once the model is finished, it will be possible to predict the inherent risk in each zip code for the new mexican market. For example: Guadalajara city.

### **Mexican Subsidiary Stakeholders:**

- 1) **CEO**
- 2) **Chief Marketing Officer**
- 3) **Chief Risk and Compliance Officer**

### **Success Criteria:**

be able to predict the degree of accident risk by zip code using metrics such as number of foursquare venues and radius of the zip code. The accuracy is not important for the insurer at the moment.

