

Case Overview:

The case study project I choose is a company called CAP Index, which provides analytic solution help business choose the right amount of security guard service support. Their clients including store, bank, park, business center etc. With the right amount of security, the client will save for expenses on the security or lower the risk of potential crime that might cost business loss. The company use demographic and crime data that across the north America, UK to help company decide how many security forces they need for the business, the analytical technique including optimization, forecasting to help company make correct decision on where and how to spend. I'm interested in this topic since I used to work as an analyst in a security company in my early career and this is great case for me to use what I learned from the course and use my analytical approach for the similar challenge. My overall approach will be using statistical model to determine what variable should be use, what constrains should be considered and the objective function for the Optimization models. After that, I will also delivery my recommendation by use the forecasting output. The subject will be including what data needed, how I'm going to do the model selection and why. As well as what the advantage or disadvantage of the model. In the end, combine with my personal experiences working in the security service company and share my insight in crime and security industry. Original story link: https://www.sas.com/en_us/customers/cap-index.html

Goal:

The goal of the topic is answer how many security guards service should a business have to best meet the security needs and minimized the total expense on security.

Data Required:

Data is the key in this case to produce accurate model and insight for predicting the number of security force. It is important to understand that for each physical business location, the security needs will be determined by its surrounding environment.

- For each business location, the 10 years of crime rate in the area, crime rate by business type
- Business revenue and Insurance claims report
- Average or median income, employment rate, population, housing price
- Average hourly salary for security officers, security equipment and gear costs
- The alcohol or drug regulation in the area.
- The capacity of the business space and the nature of the business, i.e.(Bank, School, or Shopping Mall)
- Historical data of customer volume for the business or similar business in the area.

* Note: Data will not include Race, Gender, and Religion information. This means decision will not made based on these information.

Most data on my list are accessible on the public organization website or available to acquire from several main data vendors; the local police data, and company loss reports across the nation and industries. For example, Fannie Mae and Freddie Mac, Black Knight have almost every data point that related to the housing. On the other hand, company can also use public census report, FBI and government released information which free to download on their official website. Meanwhile, we will get the data of capacity, nature, main business from the client. When I built the model, I'll also consider some regulation for use of the data. So that is also why my above data list has excluded the race, gender, or religion. Even though these can be a great data contribute to the model accuracy. Overall, we would not need a long-term data collection process in this case which will shorten the time to deliver the solution. Another advantage of the data selected are all from official data source. As a result, when we use these data, it won't take us too much time and effort on data validation, cleaning, and it will also reduce random effect of our model. Last but not least, since the government, public agency, and data

vendors always acquire and update their data regularly, we can expect our model can be replicate or reproduce for update or further optimization. The challenge here is to get the business customer volume and traffic. It will be easy to compare with for those chain stores nearby such as how many walk-in customers for Bank of America each day on average. We then will have estimate idea of traffic. For those business such as shopping mall or theme park, which probably not having any similar business nearby, we need to do some research and maybe look at a different city that has similar demographic structure to see traffic volume for the similar business.

Analytic Process:

i. Predict the crime rate in the business location for the next 12 months.

Given	Local police report; FBI information release, national crime map for the last 10 years of crime data; local business loss report case due to crime; Regulation for drug or alcohol use; City level median income, unemployment rate, housing data, median resident age in the area
Use	Exponential Smoothing, Autoregressive Model
To	Predict the crime rate in the next 12 months

In this step, we are using the city level information of historical crime data, the relevant local business loss due to crime reports and the income and the regulation for alcohol and drug to predict and forecast the crime rate for the next year. Since we want to remove the seasonal effect of the crime data, we will use the exponential smoothing first to reduce the seasonal fluctuations. The reason I'm including regulation for drug and alcohol here is because there are many studies associate with crime rate change and influenced by cannabis legalization. So, we should also include these effects when building the model.

ii. Determine risk score for each business type

Given	Nature of the business, size of the business, local business loss report case due to crime, crime rate by industry,
Use	Elbow method and K-Means Clustering
To	Assign each business type a risk score

This step is important for future analysis, based on the business type, we will be able to estimate the security needs by its risk level. The risk score is a measurement of the risk level of given a business type. For instance, 1 being the lowest risk score, and 10 being the highest risk score. For example, in common sense, within a scale of 1 to 10, we know the risk level of shopping mall would be somewhere higher than 6, and the risk level of a baby daycare should be lower than 3. Here we first use the Elbow method to choose the value of K, then use K-Means Clustering to group the business by its crime rate. For example, based on the which cluster the business in, and the number of clusters, assign each business type a risk score.

iii. Determine business loss range when crime happen

Given	Estimate traffic volume, business historical sales and revenue, loss due to crime, and insurance claims reports, physical asset value of the business
Use	Simulation and maximum likelihood estimation
To	Determine business loss

This is an approach inspired by another simulation class, everything can be simulated, and a simulation-based method will provide an idea what would be the loss caused by certain kinds of crime to a business. For example, the loss when robbery happening in jewelry store or looting happen to the sports store. By using the dollar amount on the insurance claims, and typical crime happen to the business type, we can simulate couple of different disruptive events to get the maximum likelihood estimation of the business loss.

IV. Forecast the demand of security force based on previous steps

Given	Result from previous steps: Predict crime rate in the business location in the future, Risk score of the business type, and potential business loss if crime happen to the business, business size, business type. Seasonal Cycles in Crime
Use	Regression model
To	Determine demand of the security force for the business

This step is putting all the previous output together. Since we have predicted crime rate, we could use the result as predictor variable to build the regression model that estimate the security demands for a given type of business. For example, in a low crime rate peaceful town, a new bank branch near 3,500 square feet will probably less demand of security than a high crime rate district that a bank with only 1,000 square feet demands of security. And for instance, the summer could has a higher crime rate than winter, etc.

V. Optimization on the number of security officer needs

Given	Result from previous steps: security demand for certain type of business, business operation hour, size, security force implement cost, number of business, officer hourly rate
Use	Optimization model for prescriptive analytics Minimize: Business loss (L) and security force cost (C) Number of security officer $n \geq 0$ Security officer hourly rate = s Risk score = R implementation cost = θ Number of the business = N Objective function: Minimize $L \cdot R + \sum (n \cdot s + \theta)$
To	Determine best combination of security officer and equipment

In the last step, an optimization output is desired. A result of the best combination of security force that meet the demand of the business just right. When creating the objective function, there are constraints need to take into consideration. First, the hour of the security officer should be less or equal to 8 hours. Although there is 12-hour schedule also available, but I personally won't recommend since energy and alertness gradually decrease with shift time goes by. Due to the nature of this job, which needs the security dispatch being concentrate to the environment especially during the night. For business type that not need night shift security officer, several security camera and alarm system installation are recommend. Secondly, the total demand should be(ideally) \leq Total security force acquired. This is like call center, we want all the crime to be handle as soon as possible before there is more severe result, but we also don't want to have too much security force than needed which generates extra unnecessary expense. Additionally, we want to add seasonal cycles in crime rate to help determine in different time of the year/day should increase or decrease the security force. Lastly, we also should add θ that represent potential cost of the security force implementation, such as installation, maintenance such as increasing bills of keep light on during night or cloud storage for CCTV footage etc.

Conclusion:

First, I'd like to address there are several limitations for the model. I think even use data science technique to come up with a solution still might not be the ideal solution due to variance of reality situation. According to my work experience in a security service company earlier in my career, I found that the company didn't use data science or statistical model at all. One of the reasons could be because back then data science is not widely being used. The company owner was a retired Marine member and the director was a retired local police department head and most of the security officer are either veteran or previously worked for police force. Most of the time they just used their knowledge to estimate how many security officer shifts and equipment need to be installed, and the business doing great in the state. They probably will come up with a result less than a second just by experiences than I'm building the model here. However, a model without bias and just using the data that is more objective and flexible, and with more data we will keep update and optimize the model to achieve more accurate output in terms of predicting the crime rate or the probability of risk of a certain business type.

Building a model is just a start of using statistical analysis helping business in various type and location with solution prevent crime and reduce business loss. In this case study project, this model did not include the 'Black-Swan' event which is rare to happen but once happen could have severe and widely impact like the COV-19 or Social crisis that resulting looting frequently happening in the San Francisco or New York city like right now. Also due to the nature of the security officer job, company should also increase the life insurance and training investment to the officer which also increase actual costs and legal liabilities.

In fact, model should be better customized with unique business needs, and it is hard to have a dynamic universal model that fits all kinds of business. When focusing on the goal of reduce cost of guarding and equipment, in some extreme cases for example, the loss of shoplifting is less than the cost of security cameras and security forces. In that case, adjustment should be considered to achieve the goal of reducing the cost.