
DIDI

DIDI BUSINESS INTELLIGENCE CHALLENGE

Job position: Business Intelligence

Jesús Alberto Urrutia Camacho

Ciudad de México

27 de junio de 2022

Índice

1. Based on the data and your ideas, plan strategies to double the total restaurant visitors in six months. 3
2. - Imagine that these restaurants are in your city (and not in Japan), what other data would you want to join in order of get more insights to increase the visitors? 3
3. How many channels can you think of downloading a DiDi Rides APP and how will you estimate the quality and cost of each channel? 3
4. We want to build up a model to predict “Possible Churn Users” for DiDi Rides APP (e.g.: no trips in the past 4 weeks). Please list all features that you can think about and the data mining or machine learning model or other methods you may use for this case 4

1. Based on the data and your ideas, plan strategies to double the total restaurant visitors in six months.

I would like to use 2 strategies:

- Based on the modelling techniques, I strongly use variables that are statistical significant in the regression model: Holiday, the type of Western food, the Restaurant, and the day of the week (specially a Saturday)
- Based on the professional expertise, I will hire a specialist in restaurant (because nowadays no one machine or statistical model is better than the human expertise to simulate or predict in long-term situations, specially in complex or chaotic environments, as selling or human interactions).
- Based on my experience as a business model designer, designing statistical experiments of advertising promotions, in order to know what strategy and variable allow to have higher sales. Once the variables are known, design a sales model collaboratively with prototypical customers

2. - Imagine that these restaurants are in your city (and not in Japan), what other data would you want to join in order to get more insights to increase the visitors?

I would like to have information about restaurants and consumers.

- About consumers, it would be pertinent to define the type of consumer profile and its socio-demographic, cultural variables (such as age, income, education, address, work location, family/friend networks with whom they go to eat, partisan affiliation) in order to generate the proto person.
- Regarding the profile of the restaurant and the industry, it would be pertinent to know the economic cycle where it is at the moment, also I strongly suggest to identify the variables that trigger consumption (such as location, type of food, proximity to other restaurants, food ranking, day of the week, perception of location insecurity).

So, in general, there are two factors (restaurant and customer), and for the set of variables, I would run a Structural Equations Model, in order to validate if the proposed variables have consistency in practice, and later, I would run Statistical experiments, in order to compare which variables have more weight in the prediction.

Finally, although I suggested some variables to consider to increase visits, it would be necessary to qualitatively validate if the proposed variables are the best.

3. How many channels can you think of downloading a DiDi Rides APP and how will you estimate the quality and cost of each channel?

In the case of downloading applications, I did a brief documentary research, I found that more than half (57 %) of app downloads are from digital channels. The rest is done through other channels, such as prepaid cards.

Given that two channels are proposed, the way in which quality and price are estimated will be as follows:

- The quality of the channels will depend on 2 indicators: The first is Quantity (in money) of purchase per user given the channel; and purchase frequency per user given channel.
- Regarding the price of the channels, it is proposed that the Commercial team make a proposal of the costs to implement both strategies. Financially, financial indicators are proposed by download channel (indicators such as Return on Investment).

4. We want to build up a model to predict “Possible Churn Users” for DiDi Rides APP (e.g.: no trips in the past 4 weeks). Please list all features that you can think about and the data mining or machine learning model or other methods you may use for this case

I believe that the conceptualization of *churn user* should be defined by the business area, also the definition of the problem to solve is a task for business Area.

I consider that the variables that would allow making the prediction are the following:

- Demographics, net income montly, home ublication, job ublication, risk for leaving our app (as risk in insurance).

Since this is a predictive model, I would choose a nonlinear regression. However, I consider necessary to use a classification model, to know which drivers trigger users to be disabled. A simple model would be a logistic regression, or a cluster model.