Clipboard interview - Case study : hail-riding pricing

Jordan Porcu

November 2022

Contents

1 Abstract

In this case study, the main goal is to find the best pricing for a hail-riding company. Here are few information we have :

- The ride is between Toledo Airport and Downtown.
- Two main people : "riders" are the ones who need a ride and "drivers" are the ones who conduct riders to destination.
- The study takes place for a full year only.

And here are the few constraints :

- The company charges 30$ (fixed price) for the riders to take the ride.
- The number of rides asked by riders and accepted by drivers are determined following a Poisson distribution.
- The λ factor of the Poisson distribution is equal 1 for the first month a rider enters the study. Then, it becomes the number of accepted rides for the next month, and so forth.
- Only 10,000 riders a year, and 1,000 riders a month

Finally, the goal here is to find the optimal pricing the company has to set for drivers, using information, common sense and data analysis, respecting the constraints.

2 Data exploration

To run our study, we use Python language, and you can the notebook we created here : click me.
We have one .csv file with two features :

- PAY : the pay offered to some drivers
- ACCEPTED : boolean 0 if driver declined, 1 if accepted

The 5 first lines of the data set looks like this :

	PAY	ACCEPTED
0	29.358732	0
1	22.986847	0
2	18.020348	0
3	45.730717	1
4	14.642845	0

Figure 1: Sample of the given data set

Data was clear : no missing values. However, there were few outliers that have been removed. A box plot is shown to get a better understanding of data distribution :

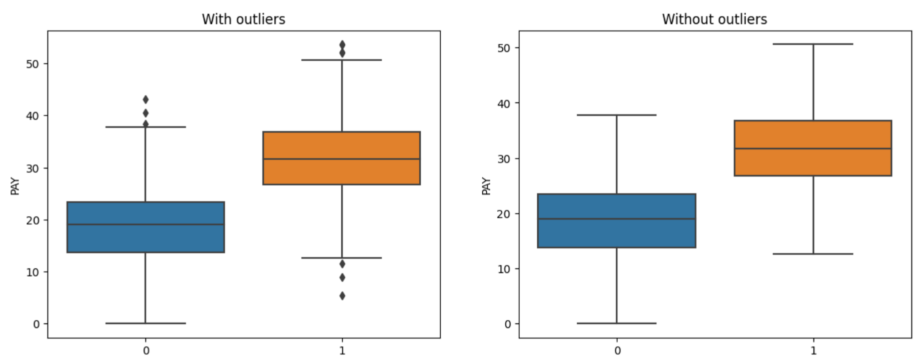

Figure 2: Boxplot of ACCEPTED feature from given data set

From this data set, we can extract the curve defining the probability of riding acceptance per pay offered. Here is the plot : (you can find one for 4 different prices in the appendices)

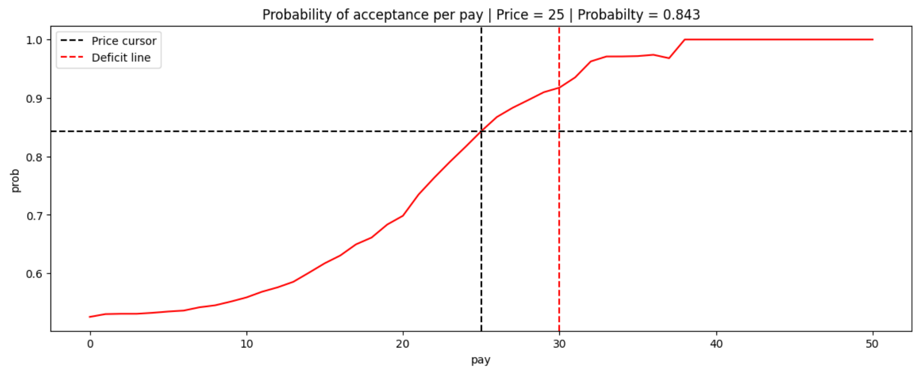

Figure 3: Probability of acceptance per pay offered. Example for price = 25 $

3 Using external data

In this part, we extracted information from internet to study an issue we can encounter : what is a minimum we can decently pay a driver ?

Here is the road map of thinking to find this information :

- Finding the petrol price for the most represented consumption in Ohio
- Finding the distance and time for a ride from Airport to Downtown Toledo
- Finding a mean for miles/gallon rate
- Finding the average salary of Uber driver (main competition) and convert it to see how much an Uber driver earns for this ride
- Convert information into a ride cost for each consumption modes
- Compute a weighted (by proportion) mean of ride cost + average salary

In the end, we have a final data frame looking like this :

	gas	e85	diesel
proportion	9000000	875000	199000
percent	0.893389	0.086857	0.019754
price	4.52	3.97	5.5
travel_price	5.07122	4.454146	6.170732
travel_total_needed	12.07122	11.454146	13.170732
weighted_total	10.784294	0.994876	0.260172

Figure 4: Driver cost data frame

Finally, when we compute the sum of weighted_total and we round it, we find that the minimum price we pay the driver is **12 $**.

4 Annual simulation

Now we have the maximum price we need to set (30$) and the minimum (12$). Following the code you can find in the notebook, we can simulate 6 years with this price range. Here is an example of a final plot that resumes the optimal price :

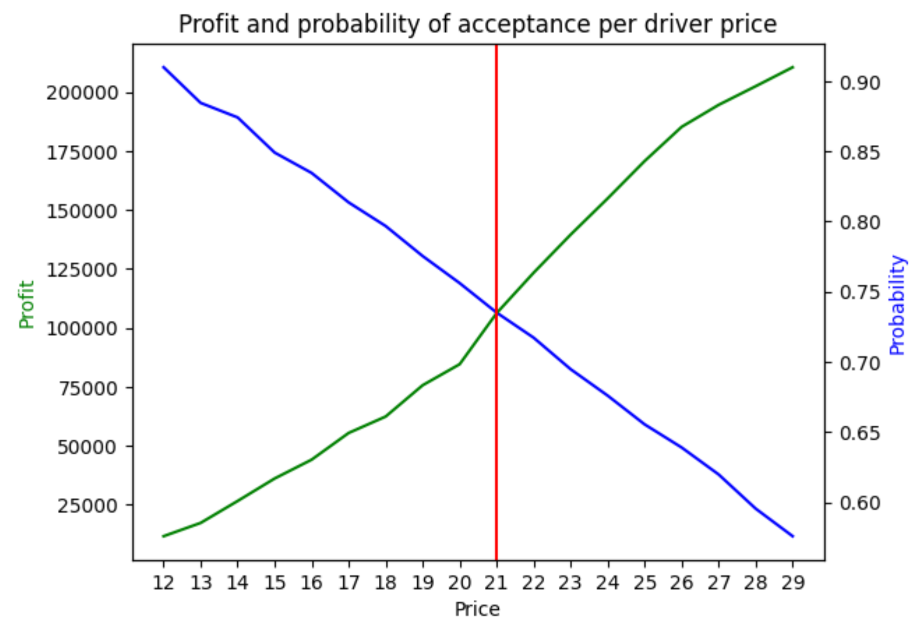

Figure 5: Profit and probability of acceptance per driver price for one year, example for a random year

And now we have results, we can display the mean of the 6 years results :

```
opt_price         21.00
opt_prob           0.73
opt_profit    107119.50
```

Figure 6: Optimal price, probability and annual profit associated

5 Results interpretation

Through our study, we found that if we set our price (for drivers pay) at 21 \$, it has the best balance between the probability the driver will accept the deal, and the profit the company makes with such a price a year. But a question needs to be answered : Why such a balance ? This comes from two major points :

1. The higher the price, the higher the profit. So, if we set a price to 1 \$, the company profit will be huge. And if we set a price to 29 \$ (1\$ profit), it will be very low. It seems obvious, but it makes much more sense considering the second point.

2. The company reputation from riders and drivers is an important key of success, and maintaining profits through the year. If we set price to a low value, profit will increase but not many drivers will accept the ride. Then, how a person will use the company application if most of it's rides are declined ?

To find a compromise between a good profit, and a good reputation, this balance is needed. As we said previously, a driver needs at least 5\$ to make the ride profitable. Then, for a ride, the driver also needs 7\$ more to match the average salary. Considering the given data, with a 21\$ price, 3 drivers out of 4 will accept the ride, and will earn better than the average salary. Because customers satisfaction is important, this price seems to be correct, and still makes the company profitable (around 107,000\$ a year)

However, few points can bias the results :

- Dependencies of Poisson distribution laws
- Only one data set given, with only one feature
- Lack of information of what could make a driver accept or not
- No information of given data (source ? year ? country ? ride path ?)
- Data to have the average salary could have been more precised and not time-spreaded

Finally, even if, intuitionously a 21\$ seems correct, for a 9\$ profit per ride, it needs to be taken with cautious, because of the biases it has. If you have any suggestion on what could make this study better, or even raise a point of a mistake that could be made, don't hesitate to reach me at jordan.porcu@gmail.com.

6 Appendices

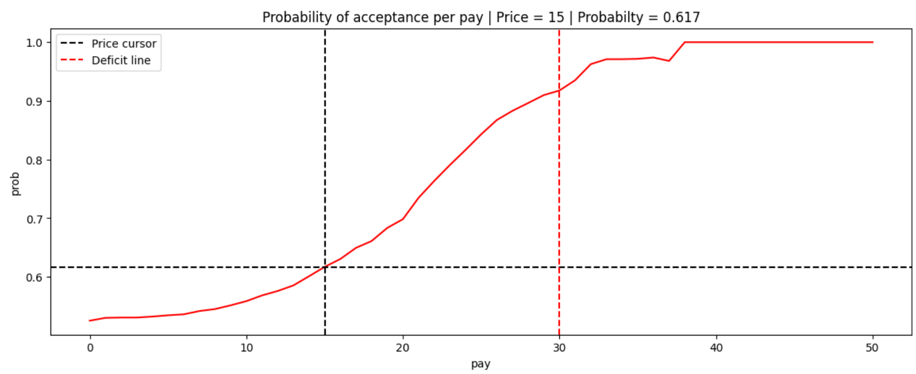

Figure 7: Price = 15 $

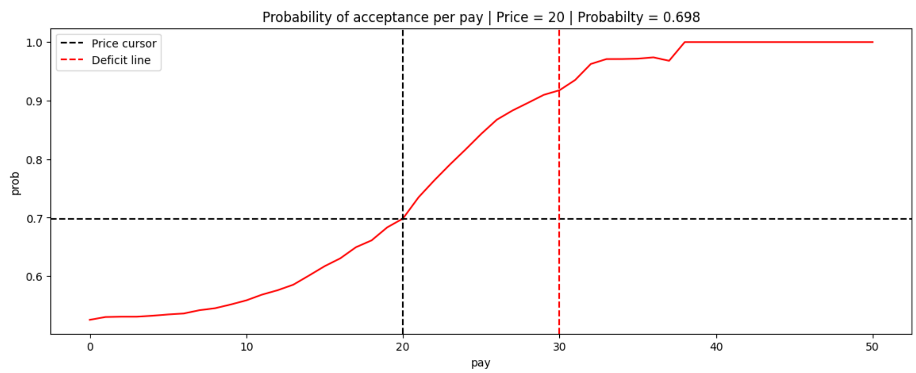

Figure 8: Price = 20 $

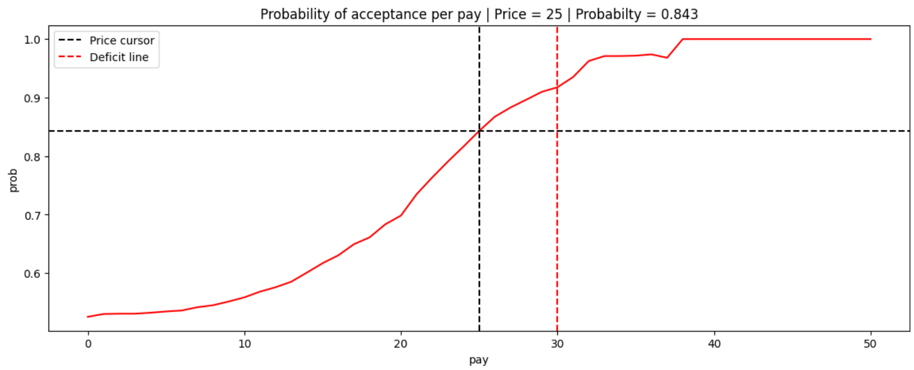

Figure 9: Price = 25 $

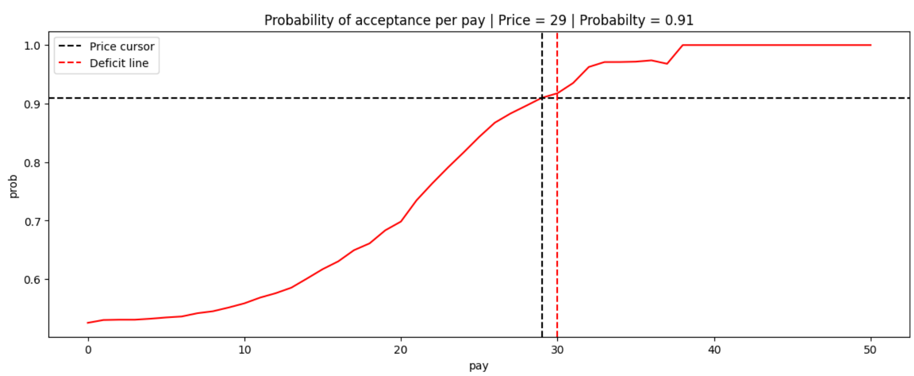

Figure 10: Price = 29 $