# Flight Overbooking

Most airlines practice overbooking. That is, they’re willing to make more reservations than the number of seats that they have on an airplane.

Overbooking is an airline’s way of ensuring they have no empty seats left at the time when a flight takes off.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Why do airlines overbook flights?**

The simple answer to why airlines overbook is to maximize profit. Airlines want to avoid empty seats at all costs, and overselling tickets is a great way to do just that.

On any given flight a few passengers are likely to be “no-shows”. If the airline overbooks slightly, it still may be able to fill the airplane.

“no-shows” in the aviation industry are people who did not arrive on time for the flight or did not bother to cancel their flight. There are also people who canceled at the very last moment.

Another reason tickets are oversold is because airlines expect a percentage of people to be coming from **connecting flights**. Flights that could very possibly be delayed or canceled. A 20-minute delay on the first flight could be the difference between boarding and not boarding the connecting flight in many cases.

These types of events may lead to flights with empty seats even if all physical seats have been sold.

* On flights operated by Lufthansa German Airlines 4.9 million passengers did not show up in 2005. This corresponds to 12,500 full Boeing 747s.
* Overbooking allowed Lufthansa to carry more than 570,000 additional passengers. Lufthansa credits the practice of selling more tickets for a flight than there are physical seats for a revenue increase of $105 m in 2005.

In addition, overbooking is not only a significant source of incremental revenue for most airlines but also creates economic benefits to the traveling public like an increased seat availability and reduced overall costs of travel through more efficient use of airline seats.

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## **Passenger** B**umping :**

When a flight is oversold i.e. the number of passengers showing up exceeded the seats on the flight, the airline would pick customers to bump, that is, rebook on a later flight.

If the flight was much later, the bumped passengers are provided with a free meal; if it was the next day, they were provided with an overnight accommodation.

In addition, the airline pays a penalty to each bumped passenger. This depends from country to country as well as airline to airline. It also depends upon how late you will be to the destination.

* In some countries if the airline arranges substitute transportation that is scheduled to get you to your final destination (including later connections) within one hour of your original scheduled arrival time, there is no compensation.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## **Approaching the problem :**

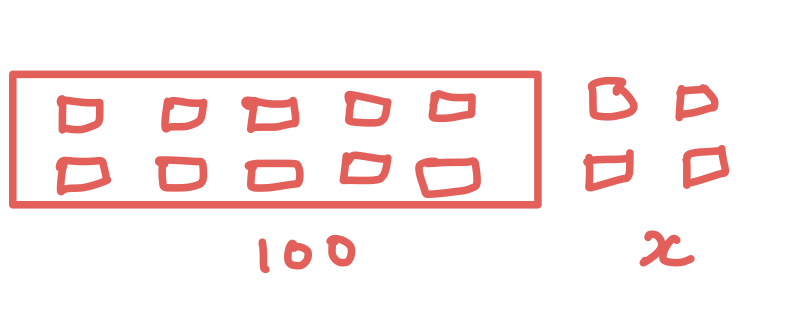
If you are working in the revenue department for an airline and need to find out what is the optimal number of overbooking that can be done for the maximum revenue.

* Your goal is to sell as many tickets as possible.
* If the overbooking is too high the airline will lose money that it will have to pay to each extra passenger that has been bumped.

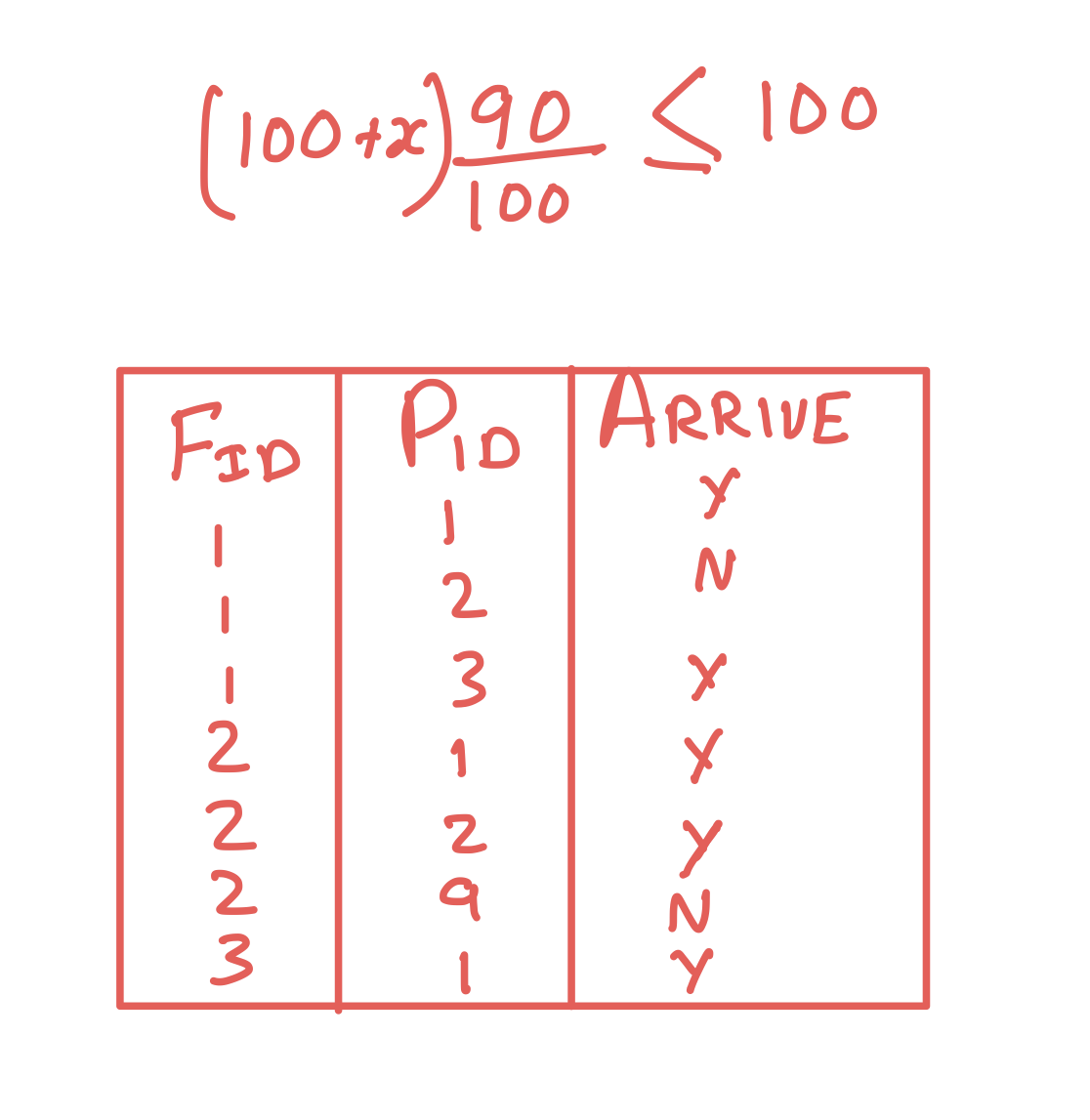
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## Simple probabilistic model :

Assume the plane has 100 seats, you need to find out the number of extra tickets that can be booked for best results.



You have some historical data that has flight id, passenger id, and the information whether the passenger arrived for that particular flight or not.



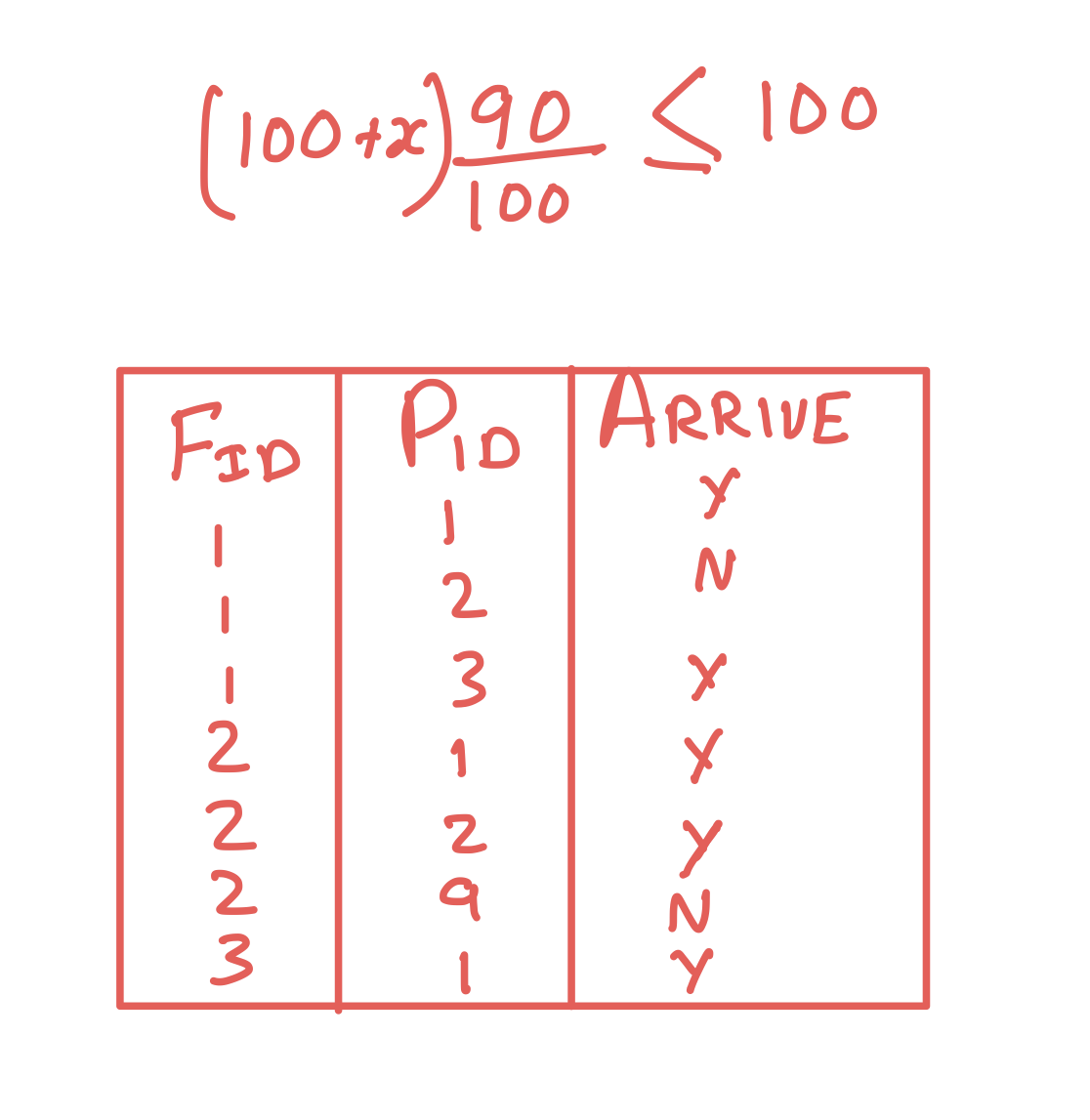
Let's say we have in total 1 lakh rows and in 90k of the cases, the passengers arrived and in 10k they didn't.

* What this means is that there is 90% probability that the passenger will show, i.e. 90000/100000.

So how do we use this data? We know that there are 10% chances of not showing up and we have 100 seats.

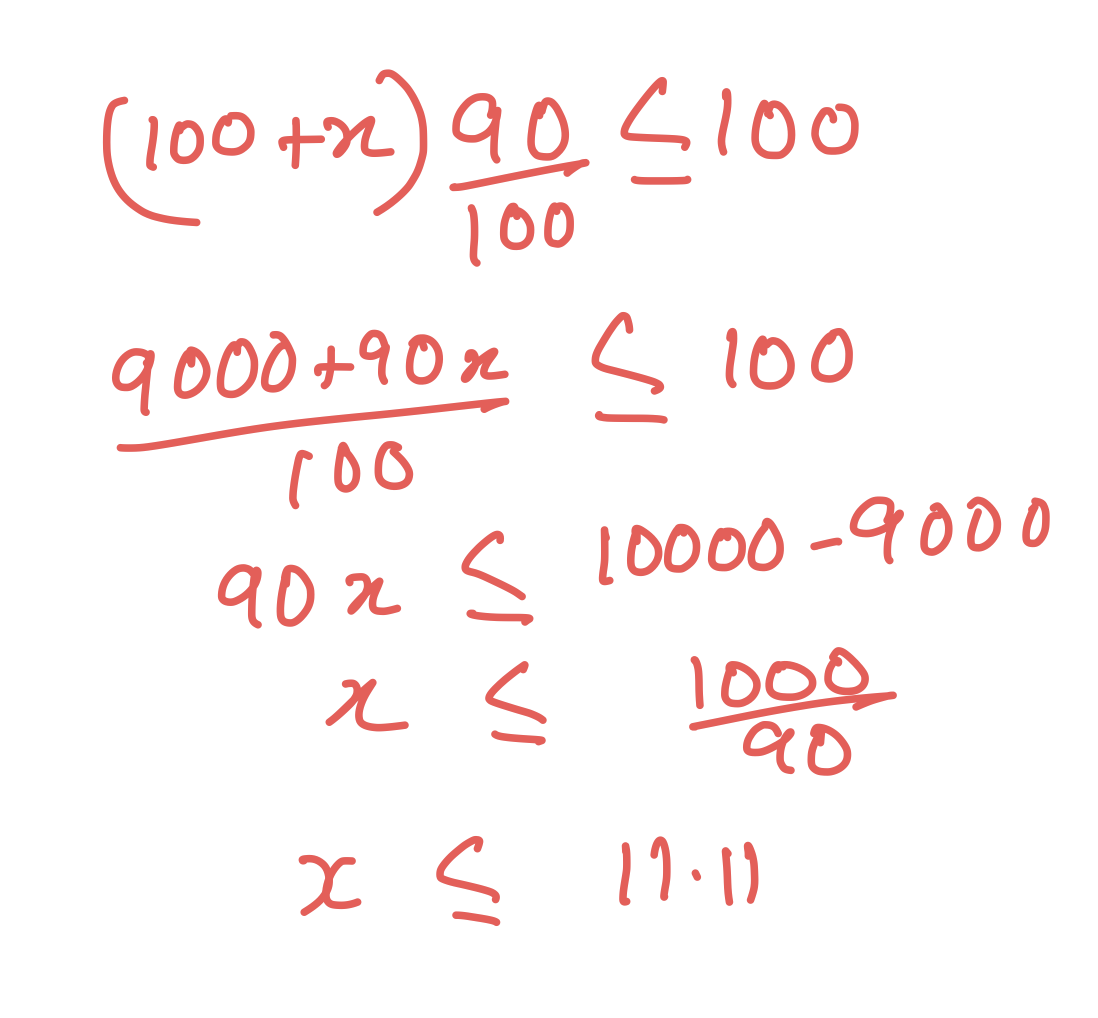
So let's assume that we sell 100+x tickets.

Out of these 100+x tickets how many are expected to show up?



And how many do we want to show up, i.e. less than or equal to 100, because we only have 100 seats.

Now if we solve this equation for x, what do we get?



This was the most simple probabilistic model that tells us based on historical data that if we have 90% chances of passengers arriving we can book additional 11 seats to get a fully booked flight.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## Why does this model sometimes fail?

For all additional passengers that we do not have a seat for, we have to pay them a penalty, so assume 102 passengers arrive, then we have to pay the extra two passengers.

If the penalty for a particular flight is 5000 or if the penalty is 500000, will you sell the same amount of ticket in both the scenarios?

Most probably if the penalty for that particular flight is so high maybe we do not want to sell any additional tickets because the risk is very high.

* If the penalty is zero, theoretically we can sell almost an unlimited number of tickets because there is no risk. There will be a problem of brand reputation and customer experience but from just a complete mathematical point of view there is no loss or no risk, right now we will just consider the mathematics side of the problem.
* If the penalty is small then we can sell some additional tickets.
* If the penalty is very large then in that case we may not want to sell any additional ticket.

The above probabilistic model does not take into account the penalty or the risk factor for the booking; it will in all scenarios tell you to book 11 extra tickets.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## How can we improve it?

The first attempt was a good guess but not a great result.

Let's assume the penalty is 10k, i.e. the net amount given back to the user including everything.

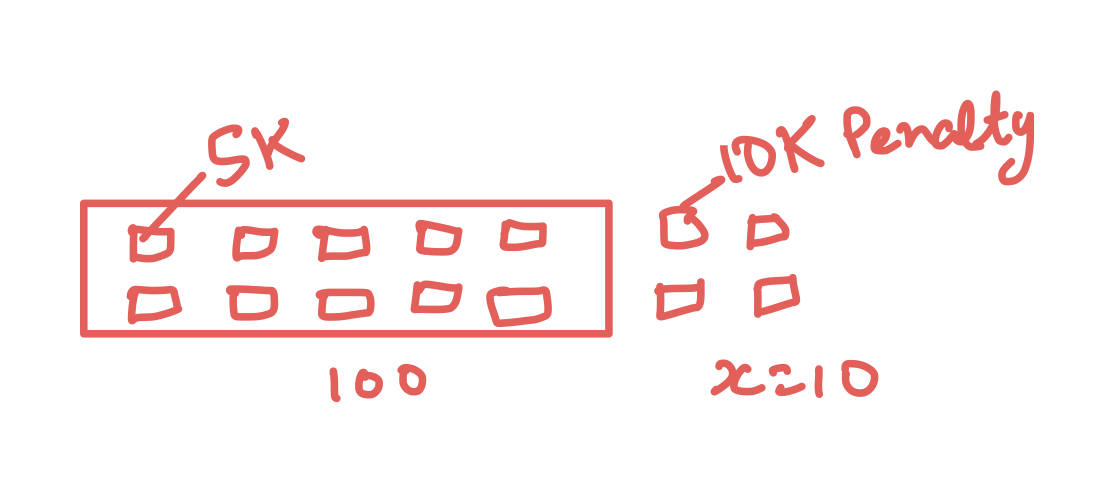
The model should be able to update its recommendation on the basis of the penalty, i.e.

* If the penalty is high it should be able to recommend a lower number of additional tickets.
* If the penalty is low it should recommend a higher number of additional tickets.

We will do the calculations based on 10,000 and then generalize the output for the variables.

* The revenue per seat is 5k
* The penalty is 10k
* Let us also fix the extra number of tickets at 10

For every ticket you sell we get 5,000 and for every extra passenger we lose 10,000.



If zero additional people show up, then

* The revenue is 110\*5000=550000 i.e. five lakh fifty thousand
* No penalty is to be given
* Earning=550000-0

So in this case the additional 50k that we are earning is because of the additional tickets that we sold.

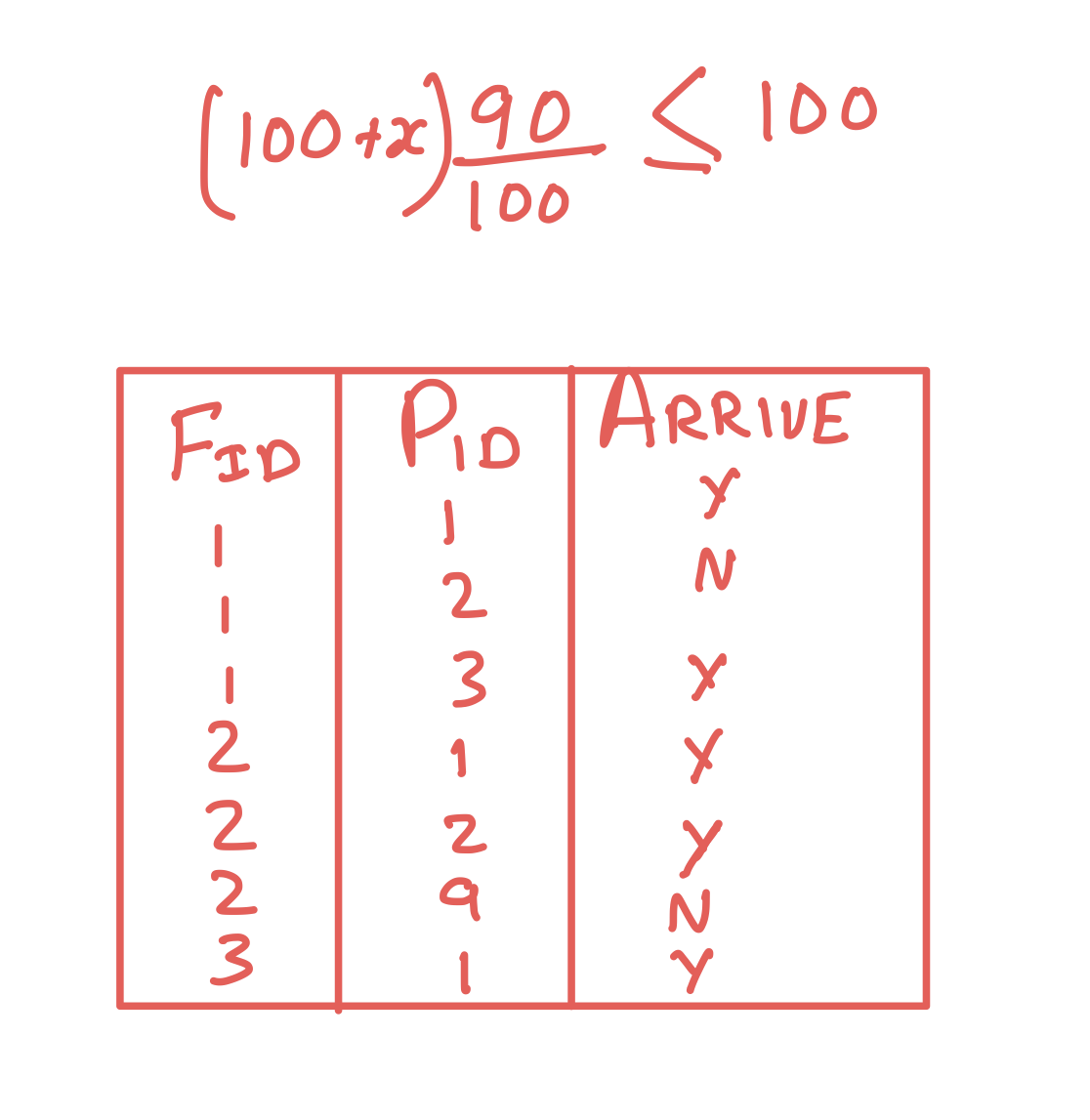
What if 1 extra person shows up. Then,

* The revenue is still 550000
* Penalty =10000
* Earning=550000-10000=540000

So let's see what is the probability for the situation..

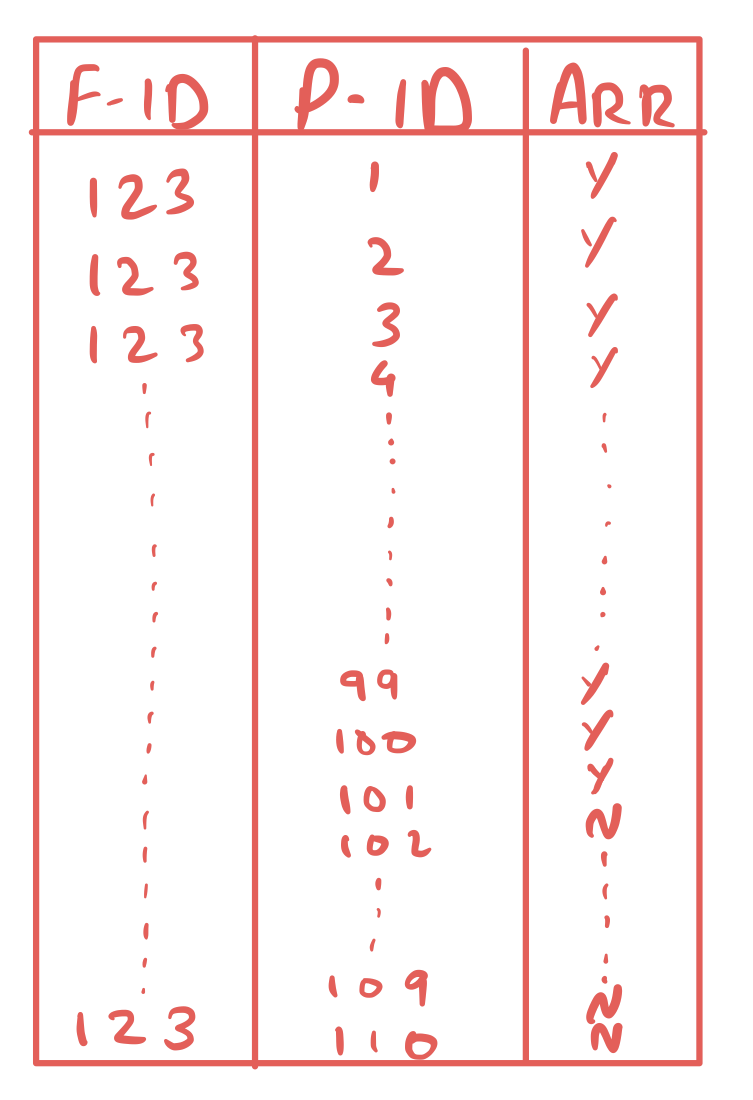
## Probability of one extra person showing up

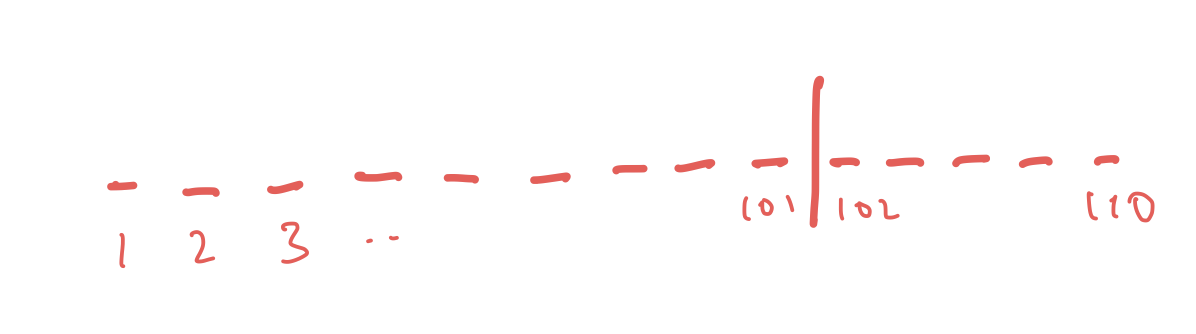
Let's look back to our data



Now assume we have a certain flight-id (assume it to be 123) and for that we sold 110 tickets

So according to the data format we have

￼



Let us write every passenger number 1-110, and we know only 101 people showed up.

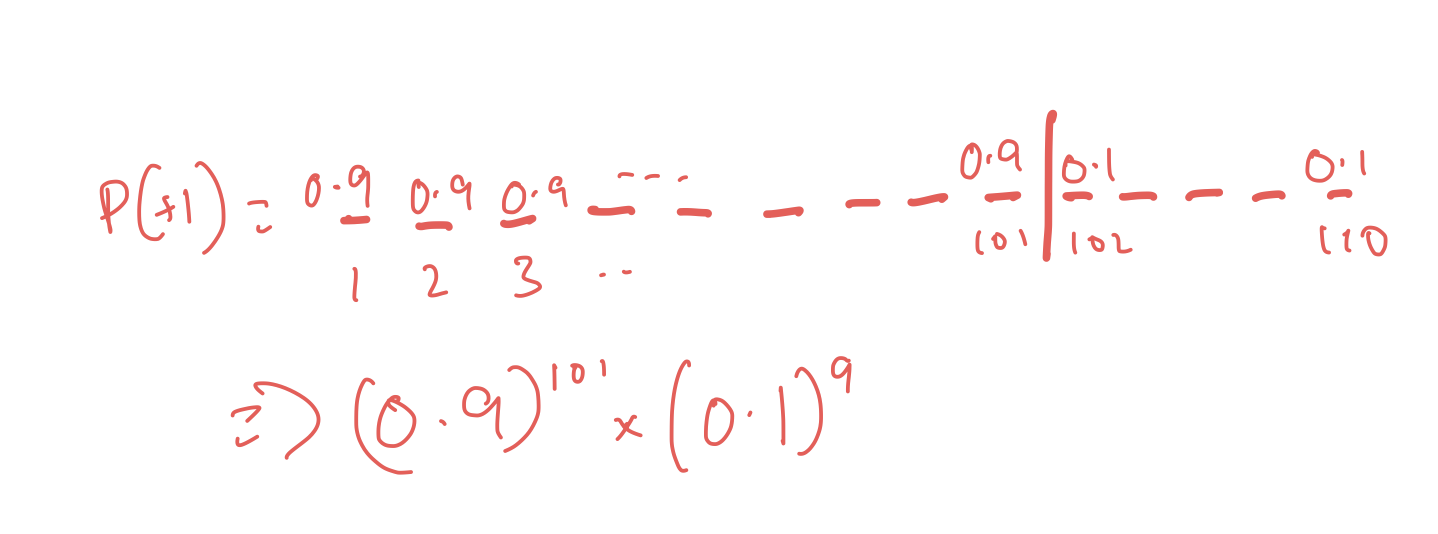
* **For now** let us assume P-ID from 1-100 show up and P-ID 101-110 do not

Now, what is the probability for each of them who are showing up?

From the past data we know that it is 90% and for not showing up is 10%

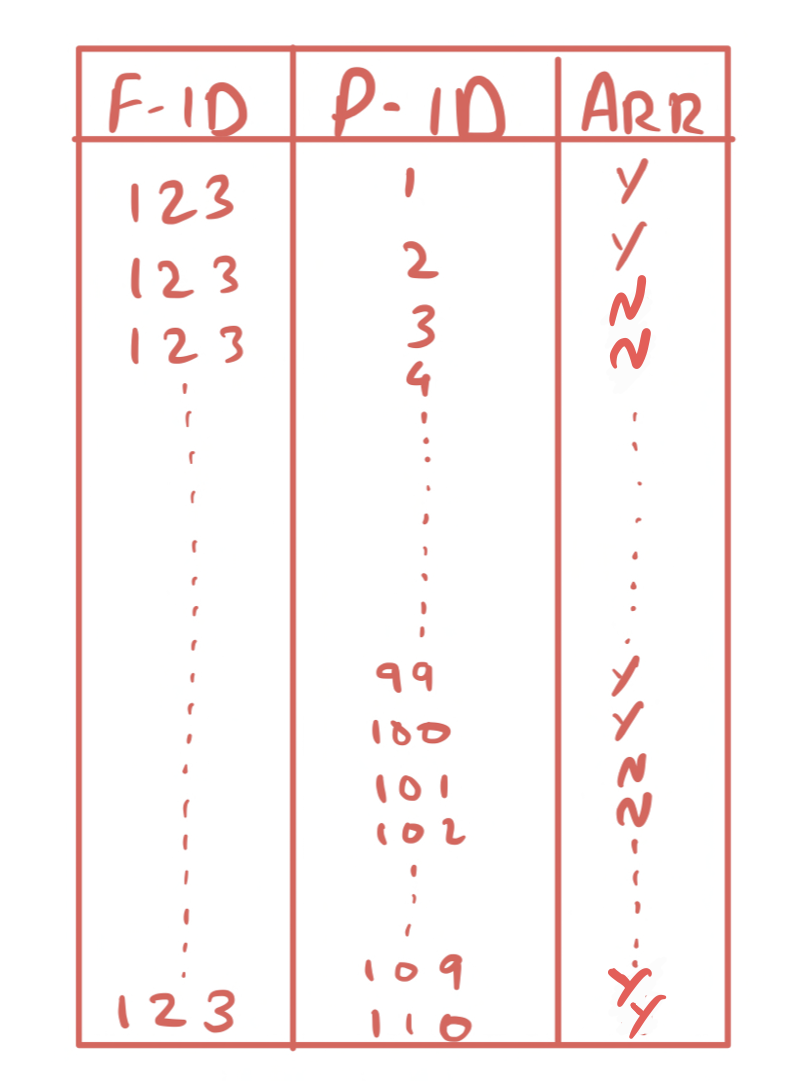
**Lets calculate the probability of these 101 people showing up**

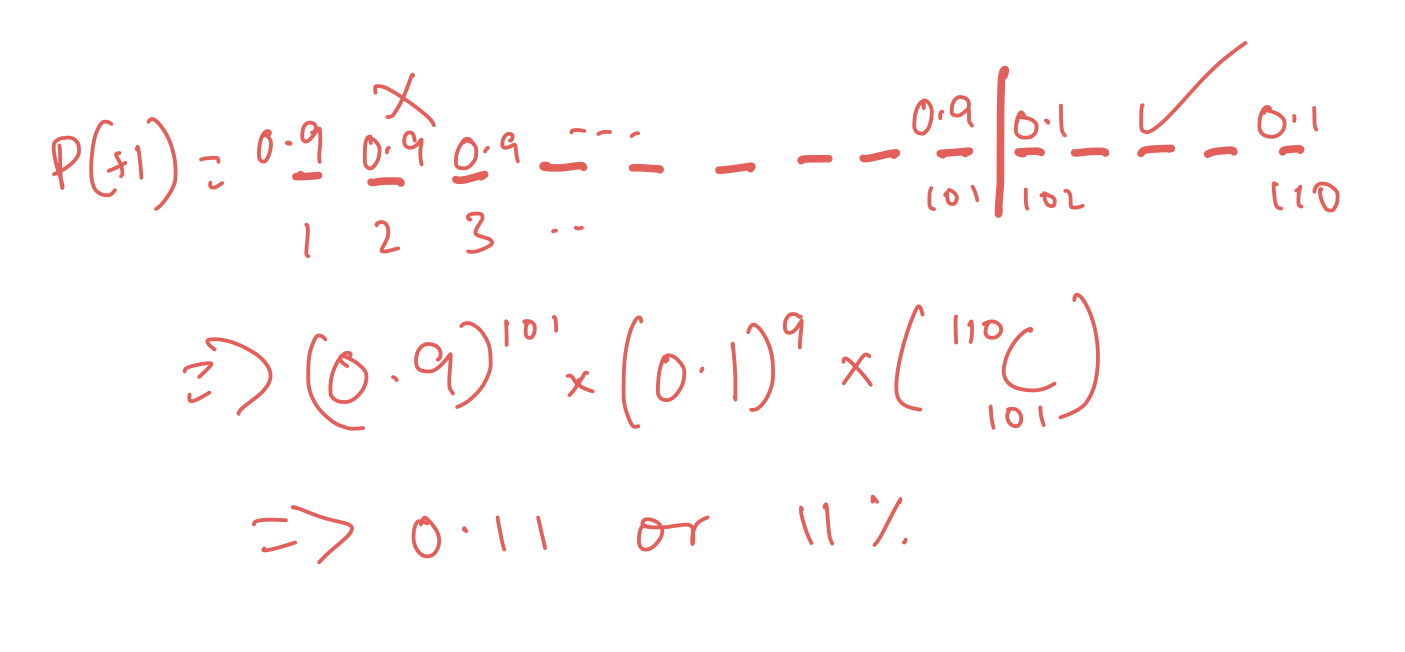
Now for these 101 passengers, we have one extra passenger above our 100 passenger available space, lets see the probability of that happening



So we can multiply all of them together to get the actual probability of these 101 people showing up.

* But there is one more thing that we need to take in consideration, we assumed that P-ID 1-101 showed up
* when calculating these probabilities, which is that it may be the case that ticket number 102 shows up instead of 2 or any other combination of ticket numbers can be the one showing up or not showing up. Above we only selected a single order or combination of 101 people showing up, but it can be very different.
* The probability for all of them is 90% but still they are very different people any of them could be or not be present.





We can keep shuffling them and the order would change, so we also need to get the combination or the number of arrangements possible.

So out of 110 tickets it can be any 101 passenger ids.

* We need to get the combination to get the probability of one extra person showing up, It is not required for the revenue calculation.
* A better way to understand the need for a combination is that instead of considering tickets, consider them as passengers.
  + Passenger 1, Passenger 2 or 3 and so on. Then the situation when Passenger 1 shows up and Passenger 2 does not will be completely different than 2 showing up but 1 doesn’t.

So when calculated, we get that the probability of one additional person showing up is 11 percent.

And the penalty for an extra person showing up is 10000,

But when do we have to pay this penalty?

* In 11% cases only.

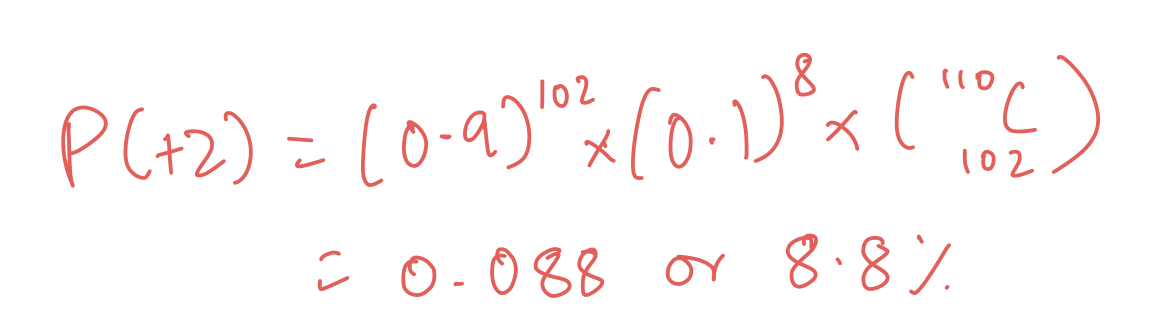
So what is the expected penalty?

The expected penalty is 10k \* 11% i.e. 1100

* So the raw penalty is 10k but the expected penalty is 1100.

**Now, what if 2 additional people show up?**

* We can just use the formula.

****

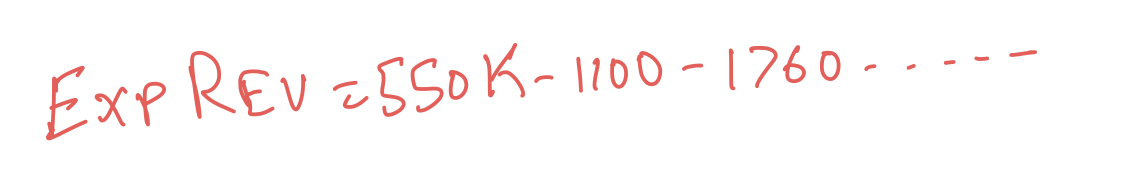
This is 8.8%.

And 8.8% of 20000 is 1760,

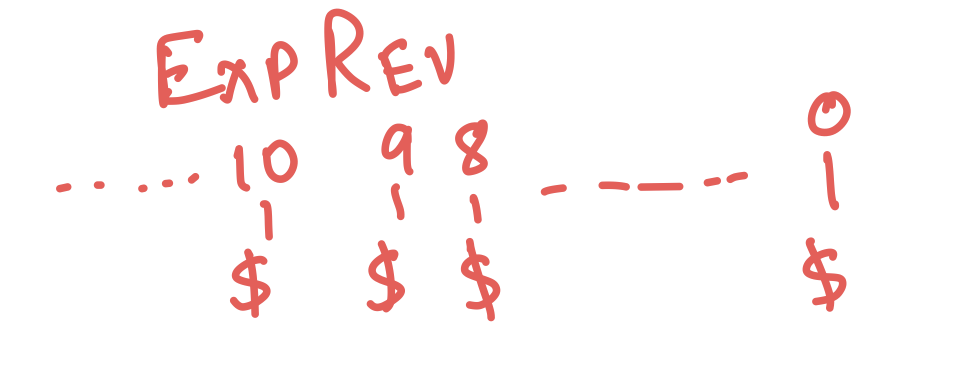
So on for the rest of the seats, the percentage of actually showing up goes down and down further.

For 3 extra people it will be 6.1% and so on.

So the total expected revenue for that flight will be the sum of all those results.



Now that we have calculated the results for 10 additional tickets, we can use the same method to calculate for all other numbers of extra tickets from zero to whatever we want.



**Generalizing the equation -**

Just like we used 10k as our penalty, we can use anything and treat it like a variable and then the results will be updated accordingly.

The last step of our calculation where we multiplied the percentage with the amount is what will change.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## Additional Factors :

Right now as we look at the data, we just look at the overall percentage of arriving at the flights and missing the flight that is 90% and 10%.

But it may not be the case in all situations, which may affect the numbers.

* Not everyone has the same tendency to miss the flights. Some people are more likely to miss a flight than others. Maybe someone is more careless and is late or forgets to cancel the tickets in case of a change of plan, those are more likely to miss the flight than others.
* Factors like traffic around the airport also affect the chances of missing the flights. Sometimes people are stuck in traffic for hours before reaching the airport and many times people do not take into account this time and miss the flights, so such airports will have higher miss rate.
* Factors like time of day are also very important. It is more likely that a person will miss a 4 AM flight than a 4 PM flight.
* Seasonal or factors related to weather are also important. If it starts to snow or rain heavily, it is more likely to miss flight because it will be difficult to reach the airport.
* Ticket price and distance of the flights can also be a factor. People generally try not to miss long flights or costlier flights, whereas if the price is pretty low there are more chances that people will miss.
* Destination airports can also play a role here. People traveling to Bangalore are more likely to be traveling due to business reasons and traveling to Goa may be more likely for leisure purposes.

There are a lot of factors that we can use for our model.

There are a lot of advanced ML techniques that we can use to build better models and utilize all the available features and data.

* We can use customer segmentation for getting groups of such customers who are likely to miss or not.
* We also cannot keep on adding all the data or all the factors to the data. There may be some factors that in reality do not contribute much to the actual result, which we can find through feature importance in machine learning.
* If we want to build a model that updates the recommendations in real time, we can use a data pipeline that can monitor real time data like weather, traffic etc.