An analytics interview case study

analyticsvidhya.com/blog/2014/02/interesting-analytics-case-study

Tavish Srivastava Tavish Srivastava, co-founder and Chief Strategy Officer of Analytics Vidhya, is an IIT Madras graduate and a passionate data-science professional with 8+ years of diverse experience in markets including the US, India and Singapore, domains including Digital Acquisitions, Customer Servicing and Customer Management, and industry including Retail Banking, Credit Cards and Insurance. He is fascinated by the idea of artificial intelligence inspired by human intelligence and enjoys every discussion, theory or even movie related to this idea.

February 5, 2014

Introduction

Case study is the most important round for any analytics hiring. However, a lot of people feel nervous with the mention of undergoing a case interview. There are multiple reasons for this, but the popular ones are:

- You need to think on your feet in a situation where there is already enough pressure
- Limited resources available to prepare for analytical case studies. Even with the amount of content available on web, there aren't many analytical case studies which are available freely.

From an interviewer perspective, he is judging the candidate on structured thinking, problem solving and comfort level with numbers using these case studies. This article will take you through a case study. Answer to each question takes you deeper into the same problem.

Make sure you check out the 'Ace Data Science Interviews' course. We have poured all our combined experience of over a decade and hundreds of interviews into this comprehensive and ultimate course. It's a guide you don't want to miss!

Background:

I moved to Bangalore 10 months back. Bangalore is a big city with number of roads tagged as one-way. You take a wrong turn and you are late by more than 20 minutes. Every single day I compare the time taken on different routes and choose the best among all possible combinations. This article takes you through an interesting road puzzle which took me considerable time to crack.

Process to solve:

I have structured this in a fashion very similar to an analytics interview. You will be provided with background at start of the interview, which will be followed by questions. After you have brainstormed / solved a question, you will be presented with additional information which will progress the case further.

If you want to undergo this case in true spirit, just ask one of your friends to take the questions and information (provided in next section) and present them to you at the right time. After all the questions, I have provided asnwers which I expect. You can compare your answers to mine.

Please note that there is no right or wrong answer in many situation and a case evolves in the way the interviewer wants. If you have a different answer / approach, please feel free to post in comments and I would love to discuss them.

Problem statement :

Background : There are two alternate roads I take to hit the main road from my home. Average speed on each of the road comes out around 30 km/hr. Let's call the two roads as road A and road B. Total distance one needs to travel on road A and road B is 1 km and 1.3 km respectively to hit the same point on the main road . Note that, before the two roads split, I see a signal (say Z) which is common to both the roads and hence does not come in this calculation. See figure for clarifications.
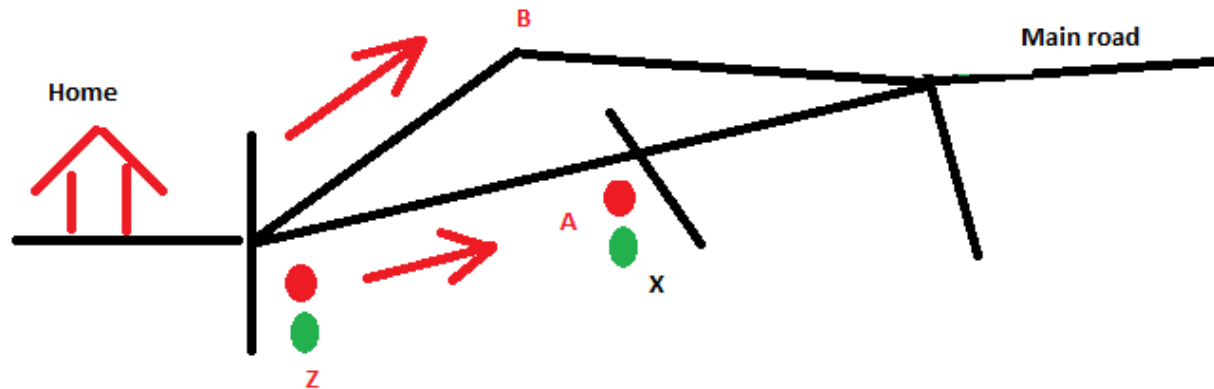
Q1 : What are the possible factors, I should consider to come up with the total time taken on each road?

Q2 : Which road should one take to reach the main road so as to minimize the time taken? And what is the difference in total time taken by the two alternate routes?

Additional information (to be provided after question 2): Recently, one of the junction (say, X) on road A got too crowded and a traffic signal was installed on the same. The traffic signal was configured for 80 seconds red and 20 seconds green. Let's denote the seconds of signal as R1 R2 R3 … G1 G2 G3 . Here, R1 denotes 1 sec after signal switched to red.

Q3 : Does it still makes sense to take road A, or to switch to road B provided the average speed on the road A is still the same except the halt at signal?

Additional information (to be provided after question 3): If I reach the signal at R1, I will be in the front rows to be released once the signal turns green. Whereas, if I reach the signal at R80, I might have to wait for some time even after signal turns green because the vehicles in the front rows will block me for some seconds before I start. Let's take some realistic guesses for the wait time after signal turns green.

R1 – R 10 : 0 sec , R11-R20 : 3 sec , R21 – R60 : 10 sec, R61 – R80 : 15 sec, G1-G15 : 5 sec, G15-G20 : 0 sec

Q4 : Does it still makes sense to take road A, or to switch to road B provided the average speed on the road A is still the same except the halt at signal?

Q5: Can you think of a reason, why road A can still be a better choice for reaching junction X in minimum time?

Additional information (to be provided after question 5): The signal Z (before the two roads split) has the exact same cycle as the signal at point X i.e. 90 sec red and 20 sec green. Average speed of any vehicle vary on road A from 25km/hr (heavy traffic) to 30km/hr (light traffic). The signal X is offset from signal Z by 25 seconds. Hence, when it turns green at Z, it is R55 at signal X.

Q6 : Does it still makes sense to take road A, or to switch to road B provided the average speed on the road A is still the same except the halt at signal?

Solution :

Background : There are two alternate roads I take to hit the main road from my home. Average speed on each of the road comes out around 30 km/hr. Let's call the two roads as road A and road B. Total distance one needs to travel on road A and road B is 1 km and 1.3 km respectively to hit the same point on the main road . Note that, before the two roads split, I see a signal (say Z) which is common to both the roads and hence does not come in this calculation.

Question : Which road should one take to reach the main road so as to minimize the time taken? And what is the difference in total time taken by the two alternate routes?

Solution :

Time taken on road A = 1/30 * 60 min = 2 minutes

Time taken on road B = 1.3/30 * 60 min = 2.6 minutes = 2 min 36 sec

Hence, the clear choice is road A. Road B would have taken 36 sec more than road A.

Interviewer tests your comfort with numbers and your confidence with the answer in this step.

Background : Recently, one of the junction (say, X) on road A got too crowded and a traffic signal was installed on the same. The traffic signal was configured for 80 seconds red and 20 seconds green. Let's denote the seconds of signal as R1 R2 R3 … G1 G2 G3 . Here, R1 denotes 1sec after signal switched to red.

Question : Does it still makes sense to take road A, or to switch to road B provided the average speed on the road A is still the same except the halt at signal?

Solution : Let's assume I come to the signal at a random time. Hence, probability of getting to the signal at R1 R2 R3 …or G1 G2 G3 are all equal. Hence, the expected time taken at the signal is :

E(halt time) = (1+2+3+4+…….80)/(80+20) = (80*81)/(100*2) = 32.4 seconds.

Still we see 32.4 sec < 36 sec. Hence, it still made sense to take road A.

Interviewer tests your knowledge of statistics (Calculation of expected value) , approach to the problem and the interpretation of the final results in this step.

Background : Till this point, the solution will look good in books. Lets spice the problem up by ground realities. If I reach the signal at R1, I will be in the front rows to be released once the signal turns green. Whereas, if I reach the signal at R80, I might have to wait for some time even after signal turns green because the vehicles in the front rows will block me for some seconds before I start. Let's take some realistic guesses for the wait time after signal turns green.

R1 – R 10 : 0 sec , R11-R20 : 3 sec , R21 – R60 : 10 sec, R61 – R80 : 15 sec, G1-G15 : 5 sec, G15-G20 : 0 sec

Question : Does it still makes sense to take road A, or to switch to road B provided the average speed on the road A is still the same except the halt at signal?

Solution :.

E(halt time) = {(1+2+3+4+.......80) + 3*10 + 10*40 + 15*20 + 5*15}/(80+20) = 40.15 seconds.

This time the game changes and as 40.15 sec > 36 sec, I will prefer road B over road A.

Interviewer tests how well swiftly you change some of the assumption so as to minimize the added calculations.

Background : Even after making such logical calculation, I noted that in 30 different events, I was commuting more than 25 sec faster on road A compared to road B every single time. I did not change my average velocity on either of the roads. It could have been acceptable in case I found x number of event where A wins and 30 – x where B wins. But A winning every single time was fishy. I was struggling for last 10 days to figure out a valid cause. It struck me today and following is what I figured out:

The signal Z (before the two roads split), which I initially though had nothing to do with the calculation was actually the game changer. Here is how it played a role. This signal had the exact same cycle as the signal at point X i.e. 90 sec red and 20 sec green. Whenever, the two lights have the same cycle, the incidence on signal X is no longer random.

Question : Does it still makes sense to take road A, or to switch to road B provided the average speed on the road A is still the same except the halt at signal?

Solution :

Say, my average speed vary on road A from 25km/hr to 30km/hr. The signal X is offset from signal Z by 25 seconds. Hence, when it turns green at Z, it is R55 at signal X.

Case 1 : (Light traffic) Time taken to cover road A = 2 mins = 120 sec

Reading at X when I reach the signal = R55 + 120 = R75.

Case 2 : (Heavy traffic) Time taken to cover road A = 2 mins 24 sec = 144 sec

Reading at X when I reach the signal = R55 + 144 = G19

Hence, the probability of R1- R74 is zero. And the revised equation of expected time is :

E(halt time) = (5 + 4+ 3+ 2+ 1 + 15*5 + 5*15)/25 = 6.6 sec

Therefore, as 6.6 sec < 36 sec road A always wins on road B.

Thus, the assumption of random events is not always true. Try to figure out all possible factors that can possibly influence the happening of event before making random events assumption.

Interviewer tests your out of the box thinking, questioning your assumption skill and interpretation of results skill in this step.

End Notes:

Did you find the article useful? Share with us any other problem statements you can think of. Do let us know your thoughts about this article in the box below.

In one of the upcoming articles, we will share how an interviewer judges an analyst during a case study.

You can also read this article on Analytics Vidhya's Android APP