【店面】

* 1. quality analysis ： physical catalogue的问题
* - 寄给什么样的用户：
* 我提到了查看和分析浏览记录、搜索记录，购买记录，wish list，以及对网站的一些特殊project （比如新家装修）的使用情况等等。
* 面试官很nice的问了“有没有什么 demographical 的数据可以看的”，就回答了看看年龄什么，中年以上可能有购买需求（有家庭）和购买能力。| 感觉可以具体提前研究一下他们的目标用户群进行准备  
  + - 如果revenue有所提升，你怎么知道是归功于 寄送的 catelogue -> 我答了ab test，不知道是不是瞎扯
  + - 你怎么知道他们下的单是因为看到了catelogue-> promo code

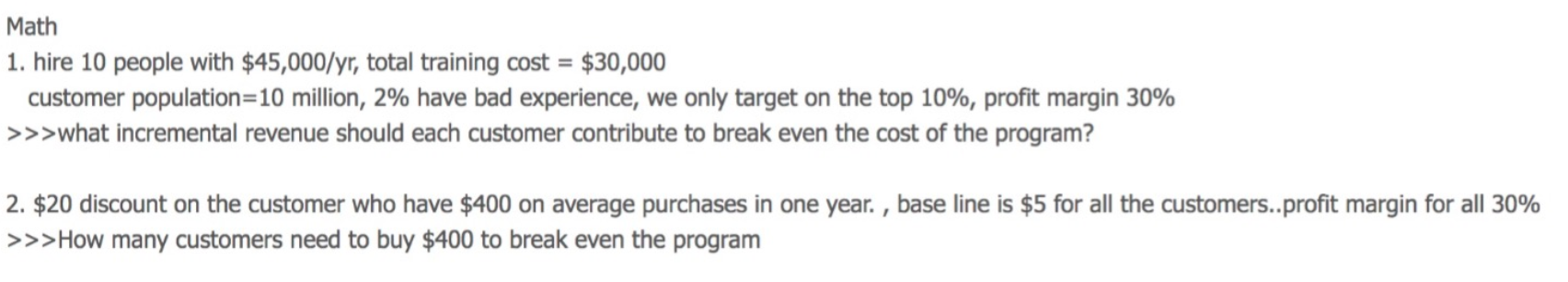
2. quantity analysis

类似之前地里的面经，就是改了该数据，但结果都是很好算的大整数。

* - 25% profit margin, 制造成本每本0.5 （小）， 1.0（大 , 每单价值280 刀( 小)， 260 刀（大）。如果小catelogue 有3%的conversion rate， 使用大本子需要多大的conversion rate （从而获得同样的利润）?
* - 你现在有一笔 90,000 刀的设计费 （固定cost）, 并决定使用大的catelogue， 你需要将catelogue 寄给多少用户，才能获 150,000的利润.
* 面试官人超好，我第二个问题一个加减号反了，算出了小数点，他看出来了，然后引导我改正了。真的非常nice。

3. case study之前也有详细问自己的一个项目。流程和视频面试相似。case study的时候有点出乎意料，并没有问技术问题。cas是一个简化的spreadsheet，上面有用户的页面浏览记录和购买记录。面试上来就询问是否有注意到什么特别的p[att](http://redirect.viglink.com/?key=a1aa544c3b328def412653f9fc432107&u=https%3A%2F%2Fwww.att.com%2Fshop%2Fwireless%2Fdevices%2Fcellphones.html)ern。我自己的兴趣在用户behavior上面，就一直在讨论用户的不同，以及如何根据用户的行为涉及promotion和recommender。之后快结束的时候，面试官才介绍说他们在希望划分不同的商品。划分的依据是商品的独特性、竞争性。举个例子，iphone和其他generic的手机。用户对iPhone的需求比较不容易被其他手机影响，iPhone的价格弹性也不是很大。

4. Poor experience customer   
1)        How to estimate which customer will give poor experience  
  
2） How to prove your method works?  
3)        Among those customers that have a bad experience, we can only reach out to a part of them. Who should we target?



5. 问了case 说bank要和wayfair合作推出一张wayfair专属信用卡，可以让顾客买wayfair的时候有更多discount，但只能在wayfair用。问你怎么决定要不要推出这个信用卡？  
  
然后基于这个credit card 做了计算题：10M 老用户，100M潜在 新用户，on average $200/purchase, 20% gross margin, 1% fee paid to the bank (before was 3%), advertising fee $10/new customer; $1/existing customer,   
2 credit card solutions:. From 1point 3acres bbs  
1. $50 off on the first purchase, 5% off on all orders, 30% existing customer and 2% new customer will sign up   
2. $25 off on the first purchase, 5% off on all orders, 25% existing customer and 1% new customer will sign up

6. 背景设定是6个月以前他们公司决定在一部分商品左上方放一个红色的sale的tag，这代表现在这个商品在降价，同时他们也给之前的客户发邮件告诉他们这些商品on sale。每周会换1000种不同的商品。  
第一个问题是现在我们得到一下observational data，how to measure the effects of adding sale tag. 这个问题他会细分问怎么建模，用那些features，有那些concerns等等。还有如果你的模型预测的跟实际并不符合，可能有那些原因。  
第二个问题是用A/B testing得到了很多数据，现在有个问题，用户往往只会看到前三页的产品，那么对于那些销量不好的产品，怎么判断这个sale tag的作用呢。我在这里我扯了好多其他的比如ranking会不会随着sale tag改变，可以用其他方法来获得额外信息，如email点击率啥的，然而并没有卵用。他后来说排名static，而且只关注这个A/B testing的结果。然后我就说可以试试把这些产品先做个clustering之类的，然后用同一类的产品预测。扯了两句之后发现一个小时就剩两分钟了。草草收尾然后我问了俩问题就结束了。

7. 上来先尬聊了几句有的没的开学感觉怎么样的问题，然后直接进入正题，问的是wayfair想做一个campaign关于如何把购物体验不好的客户拉回来，本人无商科marketing背景所以全程有点蒙，大概就是redeem customer? 具体的不太记得了，大概是这样:

1. If we want to do this project, how would you approach customers  
  
2. If we have a limited budget for this project, which types of customers would you focus on  
  
3. How would you evaluate the outcome of this project  
  
4. How would you consider the cost of this project  
  
接着是两道计算，小姐姐讲的很快，有些地方听不清（google hangout还卡顿...) 她讲完之后我一直在和她确认number，但是她人很nice, 会再解释是什么意思  
  
1. To run this project, we need 10 employees, each employee's annual salary is $45000, and the cost of employee training(for all 10 people) is $30000. Wayfair expects to have 10 million customers per year, in which 2% of them will have a bad shopping experience, and we only approach 10% of them. Profit margin is 30%. How much does each redeemed customer need to spend to make ends meet?  
  
2. Wayfair is sending this campaign to 20000 customers. Each redeemed customer will have $20 discount and will make $400's purchase. Profit margin stays as 30%. Wayfair is expecting a $5 base profit for all customers. How many customers do we need to redeem to make that happen?

-------------------------edition 2---------------------------------------

1. If we want to do this project, how would you approach customers--  
这个题我的面试官希望我给出一些能够鉴别是bad experience customer 的feature  
  
2. If we have a limited budget for this project, which types of customers would you focus on  
  
这个我问的是如果他们只想要做一个pilot实验 结合底下数学题 的 top 10 思考一下那些客户可以纳入我们的考察范围 还是考的feature  
3. How would you evaluate the outcome of this project-baidu 1point3acres  
  
A/B test. 这里他们会问评估标准， 标准是我们确定我们从实验组获得的profit可以cover整个project的费用，这个profit是incremental profit（实验组-control组增长的部分）  
4. How would you consider the cost of this project  
  
分部门答一下，背景是他们打电话给这些customer，不是coupon code  
design组 data组 和 customerservice组的employee费用  
  
1. To run this project, we need 10 employees, each employee's annual salary is $45000, and the cost of employee training(for all 10 people) is $30000. Wayfair expects to have 10 million customers per year, in which 2% of them will have a bad shopping experience, and we only approach 10% of them. Profit margin is 30%. How much does each redeemed customer need to spend to make ends meet?  
. check 1point3acres for more.  
45000\*10 + 30000 = 10million \* 2% \*10% \*30% \* num of customer  所以num of customer = 80  
2. Wayfair is sending this campaign to 20000 customers. Each redeemed customer will have $20 discount and will make $400's purchase. Profit margin stays as 30%. Wayfair is expecting a $5 base profit for all customers. How many customers do we need to redeem to make that happen?  
这道题有点triky 所以我当时没做出来   
以下是答案 20000\*5 = （400\*30% - 20） \* 所求num 最后应该是1000

8. 一上来先是简短介绍。然后就开始问有关 on sale tag 对 click through rate 的影响。包括如何定义 metric（click through rate），如何设计实验来验证（sale tag 的分布）。先问了对所有商品的测试，然后会问在 product level，和 customer-product level 的测试。Product level 就是建一个模型来预测 click through rate change ，我答了random forest，不过面试官说 linear regression 就好。Customer-product level 需要考虑有哪些 interaction。

# 9. *Wayfair is going to send 2 different catalogs to their*

*customers. One of the catalogs costs 50 cents to make and is 50 pages long. The conversion rate for the catalog is 5% and each customer brings in 315 dollars. The second catalog costs 95 cents to make, is 100 pages long and each customer brings in 300 dollars from it. The profit margin is 30%. What should the conversion rate for the second catalog be to make at least the same amount of profit as the first one.*

*After you find the conversion rate for the second one, there is a second part of the problem. Wayfair is planning to make a new catalog which is going to cost 10 cents more than the 100 page one. The more expensive catalog is going to be sent out to 20% of the customers while the remaining 80% are going to get the 100 page one. Assume the same 30% profit margin and 300 dollar profit from each customer. What should the conversion rate for the new catalog be in order to receive the same profit at the end?*

1. 300\*c(profit)\*0.3(profit margin) - 0.95(cost) = 315\*0.05\*0.3 - 0.5, assuming constant number of customers. this gives c=0.0575, or 5.75%.  
2. Essentially we need to replace 20% of the second catalog with a slightly higher cost catalog. Intuitively we would need a higher conversion rate. The actual proportion of 20/80 doesn't matter.  
300(revenue/customer bought)\*0.0575(customer bought / customer sent)\*0.3(profit/revenue) - 0.95(cost/customer sent) = 300(revenue/customer bought)\*c(customer bought / customer sent)\*0.3(profit/revenue) - 1.05(cost/customer sent)  
this gives me c = 0.0586, or 5.86%.

10. wayfair data analyst interview question:  
  
1. sent a physical advertisement to some people , how can you know it is useful and how to measure its conversion rate:   
  
Ans: check the address and name of the customer, then do the hypothesis test to make sure the advertisement does influence the revenue   
  
2. If you have bedroom catalog, who would you choose to send the advertisement ?   
  
Ans: find the customer who bought pillow or comforter or the customers who bought bed before.  
  
3. Small manual: 75 pages, 1 dollar  ;  Large manual: 100 pages , 50 cents. conversion\_rate(S) = 3%, order\_profit(S) = 280 , order\_profit(L) = 260, order\_value(S) = 25%, order\_value(L) = 25%, if the revenue of both manuals are the same, what is the conversion\_rate(L) ?  (ps. conversion\_rate = # of orders / distribution )  
  
Ans: (280\*0.03\*0.25 - 1) = 260\*x\*0.25 - 0.5  , so x = 0.04

11. sales tag case

主要问题就是，一个商品放在商品列表的第一个位置，然后给它加一个sale的tag，并且给发生过购买动作的用户发送了邮件推荐。问要怎么衡量加这个tag对销售有没有影响。

一开始我说，应该分情况。  
加tag前的时间，统计下有多少人M点进了页面查看商品详细信息，以及最后购买的人数N, 计算转化率r。  
加tag后，统计下有多少人m'点了邮件链接查看商品，以及最后购买的人n’， 计算转化率r‘。  
这样比较看看是否有效。（我后来才想起来就算加tag了也有人是没收邮件直接去看了之后买的，所以忘记考虑这部分了。。  
  
然后小哥就说，这是一种思路，但现在不是要关注转化率。假如我说上周，没tag前，卖了100个。这周，加了tag ，卖了102个。你能判断tag起到作用了吗？  
我说不行，因为有可能两周的销售，有来自相同的顾客，他们的购买操作不会受到tag影响。应该考虑这个因素。  
小哥说，假设顾客全部都不相同。。而且其他因素，价格，评价都不变。我说展示位置呢？回答也不变  
我说可能要AB testing？。。他说是的但这是之后的事情。（其实我并不知道ab testing具体是啥，只是看面经提到过。。  
然后小哥给我提示说，假设还有过去四年的这个商品数据，要怎么用。  
我说，可以计算购买趋势，计算增长率，然后对比加tag和没加的时候的增长率变化。（这里我忘记问过去是否有在这个商品加tag时候的销售数额了。。。  
之后我还提了，说可以计算过去四年销售额的平均，min，max，方差，然后看看加了tag之后这一周的购买量是否在正常波动范围内。如果是就表明加tag没什么效果。。  
然后我又说了，可以在过去四年分段计算平均minmax方差等指标，然后画图用线性回归预测本周的销量。  
  
这时候小哥就开始问我回归模型的设计。。说这个模型你打算考虑什么factor。我说价格，评价，小哥提示说，是不是要加入是否有展示tag标签的因素。。（在这里看来过去四年是有加过tag的了）  
然后我说对，这个属性可以用01变量表示。小哥追问，没错01变量，然后你怎么衡量和销量的关系？我说我先把所有的销量正则化到01区间，然后衡量相关度。不过至于怎么衡量我说我没想出来。。  
小哥问那其他factor你怎么确定对结果有影响？我还是说衡量两个连续变量的相关度，但怎么衡量没想到。。（好蠢啊  
  
还剩不到20分钟了，小哥说case分析就到这吧。然后让我问问题。我问了下他主要做什么，他讲了好多，但中间我有一段时间去搜怎么衡量相关度了没怎么听。. check 1point3acres for more.  
后来等他讲完后我互动了几句后说，刚刚太紧张了没想到衡量方法，我觉得可以用pearson相关系数衡量，小哥好像没有什么特别的语气。

12. promo code

promo code 的pros and cons  
2） 计算题，条件很多，一定要全记下来，不要慌，但是条件真的挺多的， 要分情况讨论求expectation = sum(probability \* values)之类的  
3)   black friday use promo code or low demand days use promo code.

3. 在同一个category里面的两个产品，conversion rate 增加了10%，但一个产品的order value增加，另一个下降，是为什么？  
4. 有一个买三件打八折的活动，我想要keep revenue，需要看哪些数据？（这个正常情况下是一道计算题，我这个时间太短变成了一道神奇的题目）  
5. 我会选择在black friday还是平常发布新产品？

我感觉新品应该是选平时发布，因为黑五的客户都是冲着打折去的，新品不可能一上来就打折，所以没有必要选择黑五

6. 我会选择给black friday给顾客发折扣信息，还是在平常给顾客发送折扣信息？

13. catalog

1. Wayfair 准备给两类客户发送  Catalog.(Catalog是关于Bedroom Furniture)   
一类是 已经购买过的客户。另一类是 有website visiting信息的客户。问你怎么确定target customer list, 才能使revenue最大？  
2. 你已经确定了target customer list,怎么样才能知道这些客户是因为收到这些catalog 才有的购买行为？ 有什么方法能够衡量发catalog确实能给公司带来更多的利润？  
  
3. 公司现在有A, B 两个catalog. 然后算conversion rate。这个问题网上贴子都有 glassdor 也有，就不赘述。总体来说就是(Revenue - cost = profit)  
4. 用第三题的conversion rate以及会给出一个profit的值，算一共要发多少的catalog?

14. device match

果真是很多人都被问到的从不同设备判断是否同一用户的问题。首先她问我，你觉得这个问题为什么对wayfair重要。我就说哦那当然是为了给同一个用户推荐他浏览过的同类产品咯。接下来就问怎么approach这个问题。我就说那先看有没有登录，如果登录了那直接用userid就可以判断。然后如果没登录，可以看设备的ip啊，浏览同类/个产品之间的delay啊什么的。再往下说，就是通过计算这两个设备上的商品浏览记录有多相似咯。这时候我就开始给自己挖坑了。。我一上来就想用推荐系统那套，看各个浏览历史间的similarity用vector distance。对方就有点不满意了，说那有点unsupervised，如果我们已经有数据，怎么把这个问题做成supervised呢。我然后提到做logistic regression，然后就是商品浏览记录的feature engineering。我这块不熟悉，答得磕磕绊绊。最后问我如何一个feature的coefficient是100，怎么interpret这个100.我脑子这时候已经停顿了，把这个想成了linear regression的coefficient。。扯着扯着时间就到了。  
结束前我问了她关于team文化的问题，她说他们ds team大概有80多人，背景非常多样，平时不同项目间合作也很多吧啦吧啦。结束。

15. poor experience

刚面了还是地理一样的题，没有新题，主要就是说现在wayfair有poor experience的客户：

1. 如何找到他们，最好不要通过网页数据。

2. 如何找到他们其中的potential customer

3. 解决这个问题要花多少钱，这个问题面试官会引导

4. 如何确定结果，这个只要A/B testing说了感觉就行

之后是两道计算，解决这个问题需要10个人，年薪每人45000，培训费一共30000，一共有10 Million的客户，2% poor experience, target其中的10%，profit margin 30%问每个人需要给多少钱才能收支持平。

答案是80 45k\*10 + 30k = 10million\*2%\*10%\*30%\*x

第二个就是挽回的客户每个人花400，profit margin 30%,同时有20的优惠券，然后预计是之前的每个人赚五块钱。最少挽回多少客户。

地里有总之答案是5800

【昂赛】  
onsite 分为三个section，会和同你一起面试的另外的同学滚动进行，比如你先第一个section，他就先第二个，轮流来。  
1. 9步模型题，这个和版上说的一模一样，好像是mit早期的一个inventory模型拿来用的，面试官说当时拿来用了一下，结果效果超好，也是个很古老的存在了。需要一点点inventory的知识，我当时虽然准备了一些terminology，可是在面试的时候就没有说那些，觉得有些刻意。。就把自己大概的理解说了一下。比如bundle 售价会比较便宜，相比于late delivery对用户带来的影响，多储存少量的 库存的cost更低等等。  
2. presentation. 1point3acres  
这个很神奇，就是onsite之前收到邮件说，你准备一个你以前遇到过的 most complex 的分析问题，然后在当天present给面试官。可以做slide也可以不做。  
面试官会follow up一些问题，我被问到的有：  
- 你觉得你在处理这个问题的过程中，最大的技术瓶颈是什么  
- 你觉得这个过程中最大的失败是什么  
3. behav  
- 说说你曾经如何一个error （似乎是别人已经检查过但没有检查出来的error）,  
- 如果你现在有两个project要做，ddl相同，你只能完成一个，你怎么处理。如果只能放弃一个你怎么办；如果你想保全两个你怎么办。  
- 如果你选了一个课，但你并不喜欢这个课，或者这个课有你不喜欢的地方，你会作出什么行动 （去改变它？）。  
- 讲述你曾经 “没有足够信息”却不得不做出“决定”的时候。