**Appendix 1- Documentation of Assumptions**

***Data Scope***

1. We assume that all interactions between Zillow and Real Estate Agent are included in dataset.

***Missing Data***

1. Our data excludes all the observations that lead type is null as it is difficult to assign the data to a marketing channel.
2. We remove meeting and phone calls data which do not have Reagent ID.

***Data Filtering***

1. We include leads of all RealAgentID in our analysis no matter if it’s associated with an acquisition or not. Because even if it does not lead to acquisition, it still causes cost.
2. We only consider outbound phone calls. We assume that leads are identified by sales outreach via email or cold calling as first touch point.

***Data Mapping***

1. Each ZUID relates to one acquisition. Since each ZUID is unique, we believe that Zillow generates the unique ZUID for each agent after they purchase the ads.
2. Except Paid Search - Bing, Paid Search - Google, Email, and Organic, all other lead channels are assigned to “Paid Social”.

***Revenue Assumption***

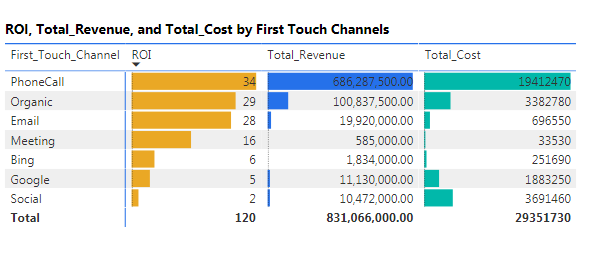
1. We use Typical 2-year NPV of an acquired agent to measure revenue value. Based on the data and given information, we believe that provided NPV is the best parameter to measure revenue.
2. Because the average cost of all the lead channels is $7,500, we decide to assign revenue of $7500 to outbound phone call. In addition, we agree to assign higher revenue of $9000 to meeting because it has highest cost across all channels.

**Appendix 2 - Data Findings**

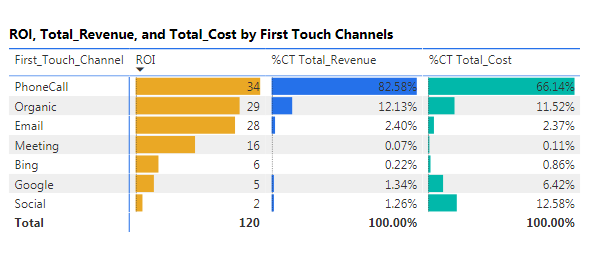
1. Generally, acquisition date must be later than first touch date for a complete lead. However, we found that 56.7% of acquisition dates are earlier than first touch dates (phone call dates, meeting dates, or lead dates). And we infer the reason is that the range of acquisition dates is larger than range of first touch dates in given dataset. Specifically, AcquisionDate on Table AgentAcquisitionDates are from 1/1/2009 to 7/5/2017. PhoneCallDateTime on Table PhoneCalls are from 10/1/2016 to 6/30/2017. SalesMeetingDate on Table SalesMeetings are from 3/13/2017 to 6/30/2017. LeadDateTime on AgentLeads Table are from 10/1/2016 to 6/29/2017.
2. Historical conversion rate calculated from the given data set by marketing lead channels are: Bing 14.92%, Email 24.36% Google 14.13%, Organic 18.53% and Social 8.51%. They are in the consistent order with the given conversion rate of Paid (including Bing, Google and Social), Email and Organic.
3. We find it useful to begin analyzing ROI, percentage of total cost and total revenue generated by first touch channels (Table 1 and 2). The total revenue and total cost are around $831M and $29M, respectively. Phone Call yields the highest revenue ($686M/82.58%) and also the highest cost ($19M/66.14%). Moreover, Meeting Channel has the lowest revenue ($0.59M/0.07%) and the lowest cost ($0.34M/0.11%), resulting from the insufficient data provided. Therefore, the Meeting Channel will not be considered any further. Lastly, we discovered that Social Channel constitutes only 1.26% of total revenue while its cost is relatively high at 12.58% of total cost, indicating that it is the least profitable channel. As a measure of return, we calculated ROI and noticed that compared to Bing, Google and Social, Phone Call, Organic and Email have much higher ROI, making these Channels more profitable.
4. As displayed in Table 3, Phone call and Email yield higher conversion rate - approximately at 25% and Social the lowest at - 8.5%.
5. In Table 4, first we assumed 200k budget for each marketing channel. From there, we calculated estimated leads per lead type by dividing 200k over average cost per lead type, then we get project agent acquired by lead type by multiplying estimated leads per lead type and conversion rate gathered in table 2. We discovered that Phone Call, Organic and Email have more agent acquired while Bing, Google and Social have less agent acquired.
6. As shown in Table 5, we computed the average phone call times after first touch. It appears that Organic and Email requires less phone call times among all channels, suggesting that they are more effective in landing the clients. On the contrary, Google and Social need more follow-up phone calls after first touch, making them less desirable channels to acquire clients.

**Appendix 3**

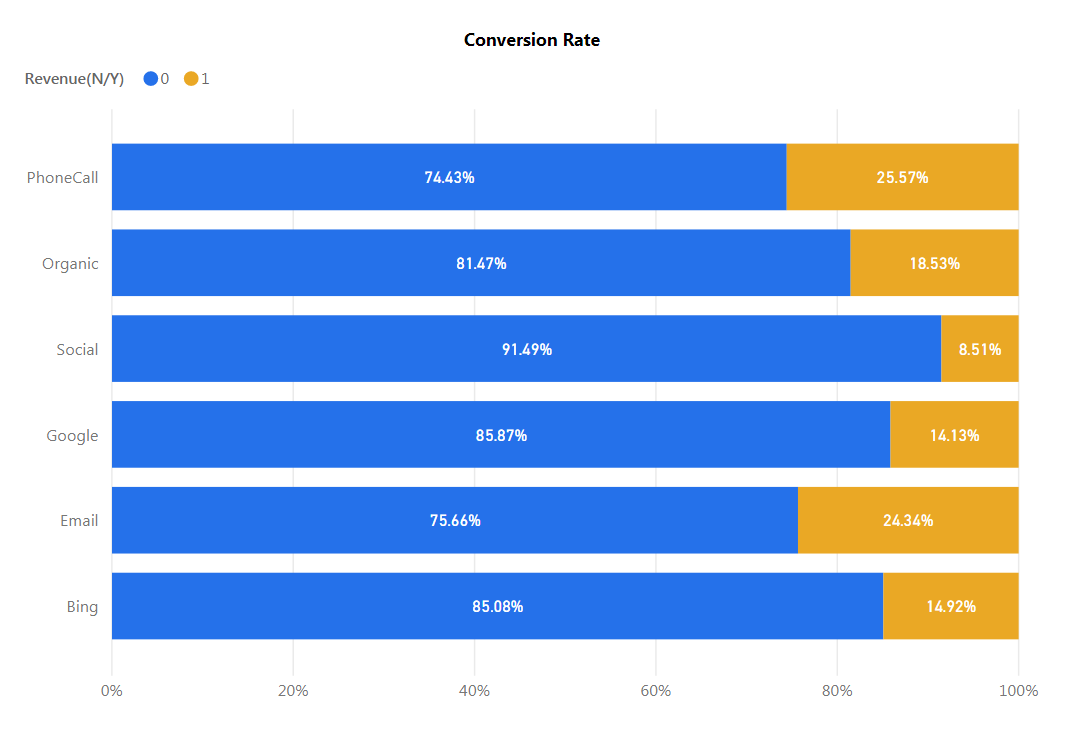
**Table 1**

****

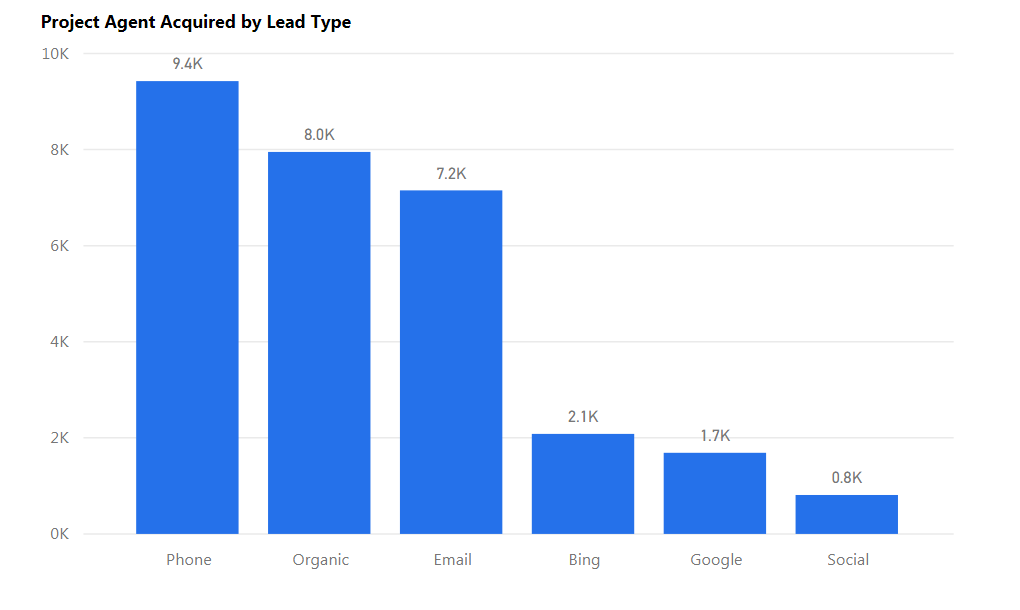
**Table 2**



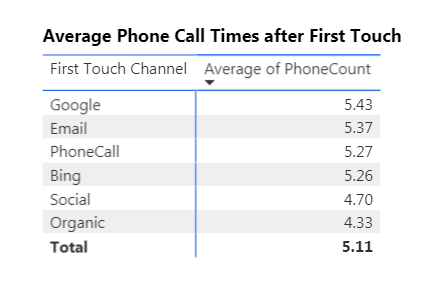
**Table 3**

****

**Table 4**

****

**Table 5**

****

**Table 6**

|  |  |  |
| --- | --- | --- |
| Channels | Before | After |
| Phone Calls | 66.14% | 70% |
| Organic | 11.52% | 17% |
| Email | 2.37% | 6% |
| Meeting | 0.11% | 0.11% |
| Bing | 0.86% | 0.86% |
| Google | 6.42% | 3.5% |
| Social | 12.58% | 2.53% |

**Appendix 4 - SQL Code**

/\*\*Count Phone by REAgentID\*\*/

SELECT REAgentID, SUM(PHONECount) AS PhoneCount FROM (

SELECT REAgentID,

--CONVERT(DATE, PhoneCallDateTime) AS CallDate,

--SUM(TalkTimeMinutes) AS TotalMinute

COUNT(\*) AS PhoneCount

FROM PhoneCalls20161001\_20170228

WHERE REAgentID IS NOT NULL

AND TalkTimeMinutes > 0

--AND PhoneCallType like 'RE Agent Called Sales %'

GROUP BY REAgentID

--, PhoneCallDateTime

UNION

SELECT REAgentID,

--CONVERT(DATE, PhoneCallDateTime) AS CallDate,

--SUM(TalkTimeMinutes) as TotalMinute

COUNT(\*) as PhoneCount

FROM PhoneCalls20170301\_20170630

WHERE REAgentID IS NOT NULL

AND TalkTimeMinutes > 0

--AND PhoneCallType like 'RE Agent Called Sales %'

GROUP BY REAgentID) as a

GROUP BY ReagentID

SELECT t1.REAgentID, PhoneCount, MeetCount FROM [dbo].[PhoneCount] as t1

FULL OUTER JOIN (

SELECT REAgentID,

COUNT(\*) AS MeetCount FROM [dbo].[SalesMeetings]

WHERE REAgentID IS NOT NULL

GROUP BY REAgentID) as t2

ON t1.REAgentID = t2.REAgentID

/\*\*Count Meet by REAgentID\*\*/

SELECT \* INTO FullMeet FROM(

SELECT t1.REAgentID, PhoneCount, MeetCount

FROM [dbo].[PhoneCount] as t1

FULL OUTER JOIN(

SELECT REAgentID,

COUNT(\*) AS MeetCount FROM [dbo].[SalesMeetings]

WHERE REAgentID IS NOT NULL

GROUP BY REAgentID) as t2

ON t1.REAgentID = t2.REAgentID)a;

/\*\*Merge Phone and Meet Count\*\*/

SELECT \* INTO PhoneMeetCount FROM(

SELECT \* FROM FullCall

UNION

SELECT \* FROM FullMeet)a;

/\*\*Count lead count by LeadVendor and REagentID\*\*/

SELECT DISTINCT(CONVERT(VARCHAR, LeadVendor)) FROM AgentLeads

WHERE LeadVendor LIKE 'email'

--email

SELECT \* INTO ZLEmail FROM (

SELECT ReagentID,

COUNT(\*) as EmailCount

FROM AgentLeads

WHERE REAgentID IS NOT NULL

AND leadVendor like 'email'

GROUP BY REAgentID) a

--bing

SELECT \* INTO ZLBing FROM(

SELECT ReagentID,

COUNT(\*) as BingCount

FROM AgentLeads

WHERE REAgentID IS NOT NULL

AND leadVendor like 'bing'

GROUP BY REAgentID) a

--Google

SELECT \* INTO ZLGoogle FROM(

SELECT ReagentID,

COUNT(\*) as GoogleCount

FROM AgentLeads

WHERE REAgentID IS NOT NULL

AND leadVendor like 'Google'

GROUP BY REAgentID) a

--Organic

SELECT \* INTO ZLOrganic FROM(

SELECT ReagentID,

COUNT(\*) as OrganicCount

FROM AgentLeads

WHERE REAgentID IS NOT NULL

AND leadVendor like 'Organic'

GROUP BY REAgentID) a

--social

SELECT \* INTO ZLSocial FROM(

SELECT ReagentID,

COUNT(\*) as SocialCount

FROM AgentLeads

WHERE REAgentID IS NOT NULL

AND leadVendor like 'Social'

GROUP BY REAgentID) a

--left join with original agentleads

SELECT \* INTO  fullAgentLeadsCount FROM(

SELECT t4.\*, s.SocialCount FROM(

SELECT t3.\*, o.OrganicCount FROM(

SELECT  t2.\*, g.GoogleCount FROM(

SELECT t1.\*, b.BingCount FROM

(SELECT a.\*, e.EmailCount FROM AgentLeads AS a

LEFT JOIN ZLEmail e ON a.REAgentID = e.REAgentID) AS t1

LEFT JOIN ZLBing b ON  t1.REAgentID = b.REAgentID) AS t2

LEFT JOIN ZLGoogle g ON t2.REAgentID = g.REAgentID) AS t3

LEFT JOIN ZLOrganic o ON t3.REAgentID = o.REAgentID) AS t4

LEFT JOIN ZLSocial s ON t4.REAgentID = s.REAgentID) AS t5

/\*\*Select First Call Date by ReagenID\*\*/

--filter out no ReagentID, TALKTIME = 0

--min date by reagentID,

--only outbound calls

SELECT a.REAgentID, '' as SalesRepID, '' as LeadPlatform, 'PhoneCall' as LeadType, 'PhoneCall' as LeadVendor, min(a.DateTime) as 'DateTime' into fullfirstcall from(

SELECT REAgentID,

--min(SalesRepID) as SalesRepID,

--null as LeadPlatform, 'PhoneCall' as LeadType,

--'PhoneCall' as LeadVendor,

min(PhoneCallDateTime) as 'DateTime'

FROM PhoneCalls20161001\_20170228

WHERE REAgentID is not null

AND TalkTimeMinutes > 0

AND PhoneCallType LIKE '%Sales Rep Called RE Agent%'

GROUP BY REAgentID

UNION

SELECT REAgentID,

--min(SalesRepID) as SalesRepID,

--null as LeadPlatform, 'PhoneCall' as LeadType,

--'PhoneCall' as LeadVendor,

min(PhoneCallDateTime) as 'DateTime'

FROM PhoneCalls20170301\_20170630

WHERE REAgentID is not null

AND TalkTimeMinutes > 0

AND PhoneCallType like '%Sales Rep Called RE Agent%'

GROUP BY REAgentID) a

GROUP BY REAgentID

/\*\*Select First Meeting Date by ReagentID\*\*/

SELECT \* into fullmeeting FROM (

SELECT REAgentID,

'' as SalesRepID,

'' as LeadPlatform,

'Meeting' as LeadType,

'Meeting' as LeadVendor,

min(SalesMeetingDate) as 'DateTime'

FROM SalesMeetings

WHERE REAgentID is not null

GROUP BY REAgentID) a

SELECT \* INTO fullcallmeeting FROM(

SELECT \* FROM fullfirstcall

UNION

SELECT \* FROM fullmeeting)a;

/\*\*Match LeadVendor with LeadType for missing LeadVendor values\*\*/

UPDATE [dbo].[AgentLeads]

SET LeadVendor = LeadType

WHERE LeadVendor IS NULL

/\*\*Group LeadVendor as 5 marketing channels\*\*/

UPDATE[dbo].[AgentLeads]

SET [dbo].[AgentLeads].LeadVendor = 'Google'

WHERE [dbo].[AgentLeads].LeadVendor like 'Google%'

UPDATE [dbo].[AgentLeads]

SET [dbo].[AgentLeads].LeadVendor = 'Organic'

WHERE [dbo].[AgentLeads].LeadVendor LIKE 'unknown'

UPDATE [dbo].[AgentLeads]

SET [dbo].[AgentLeads].LeadVendor = 'Bing'

WHERE [dbo].[AgentLeads].LeadVendor LIKE 'Bing%'

UPDATE [dbo].[AgentLeads]

SET [dbo].[AgentLeads].LeadVendor = 'email'

WHERE [dbo].[AgentLeads].LeadVendor LIKE '%mail%'

UPDATE [dbo].[AgentLeads]

SET [dbo].[AgentLeads].LeadVendor = 'email'

WHERE [dbo].[AgentLeads].LeadVendor LIKE 'Outlook'

UPDATE [dbo].[AgentLeads]

SET [dbo].[AgentLeads].LeadVendor = 'Social'

WHERE [dbo].[AgentLeads].LeadVendor NOT LIKE 'Bing%'

AND [dbo].[AgentLeads].LeadVendor NOT LIKE 'Google%'

AND [dbo].[AgentLeads].LeadVendor NOT LIKE 'email%'

AND [dbo].[AgentLeads].LeadVendor NOT LIKE 'organic%'

AND [dbo].[AgentLeads].LeadVendor NOT LIKE 'phone%'

AND [dbo].[AgentLeads].LeadVendor NOT LIKE 'meeting'

/\*\*Add Revenue for successful acquisitions\*\*/

select \*, case when Vendor\_Revenue like 'email' then 8000

when Vendor\_Revenue like 'Organic' then 7500

when Vendor\_Revenue is null then 0

when Vendor\_Revenue like 'Phone' then 7500

when Vendor\_Revenue like 'Meeting' then 9000

else 7000 end as revenue

from(

/\*\*Add Revenue channel and Cost channel columns\*\*/

select m."date" as first\_touch\_date,m.REAgentID,m.Vendor\_Cost,m.PhoneCount,

m.MeetCount,m.emailcount,m.bingcount,m.googlecount,

m.organiccount,m.socialcount,

d.ZUID,convert(date, d.AcquisitionDate) as AcquisitionDate,

case when Vendor\_Cost like 'Organic' and d.ZUID is not null then 'Organic'

when Vendor\_Cost like 'email' and d.ZUID is not null then 'Email'

when Vendor\_Cost like 'Phone%' and d.ZUID is not null then 'Phone'

when Vendor\_Cost like 'Meeting' and d.ZUID is not null then 'Meeting'

when d.ZUID is null then null

else 'Paid' end as Vendor\_Revenue,

from(

/\*\*Map ZUID with REAgentID\*\*/

select k.touch,k."date",k.REAgentID,k.LeadVendor as Vendor\_Cost,

k.Phonecount,k.Meetcount,k.emailcount,k.bingcount,k.googlecount,

k.organiccount,k.socialcount,map.ZUID

from(

/\*\*Attach Sales activity counts and filter first touch records. Filter out records that are unable to trace back vendor type\*\*/

select a.touch,a."date", a.REAgentID, a.LeadVendor,

a.emailcount,a.bingcount,a.googlecount,a.organiccount,a.socialcount,

c.Phonecount, c.MeetCount

from(

/\*\*Assign touch counts and marketing channel counts for each REAgentID\*\*/

select  ROW\_NUMBER() OVER (PARTITION BY Lead.REAgentID

ORDER BY Lead.LeadDateTime ASC) AS touch, convert(date,Lead.LeadDateTime) as 'date', lead.\*, ful.emailcount,ful.bingcount,ful.googlecount,ful.organiccount,

ful.socialcount

from  [dbo].[AgentLeads] Lead

join  [dbo].[fullAgentLeadsCount] ful

on ful.REAgentID = lead.REAgentID) a

left join [dbo].[PhoneMeetCount] c

on a.REAgentID = c.REAgentID

where a.touch = 1

and Leadtype is not null) k

left join [dbo].[AgentIDZUIDLookup] map

on k.REAgentID = map.REAgentID) m

left join [dbo].[AgentAcquisitionDates] d

on d.ZUID = m.ZUID) z

Project profit = agents \* revenue

Projected Agents acquisited = (conversion rate \* budget$)/cost($/lead)

Leads = budget/cost