

Optimizing Merage's Online Advertising - SMP Data

Group 9

Sujaya Darke, Sammi Chang, Isha Chaudhari, Song Han, Poonam Ahuja



AGENDA

1

PROJECT SCOPE

2

**DATA, EXPLORATION,
RECOMMENDATION**

3

**OVERALL
CONCLUSIONS**

4

Q&A



PROJECT OVERVIEW





PROJECT SCOPE

Help Merage improve
online advertising
strategy for SMP

Improve Merage
campaigns on **search**
engines and **social media**
platforms

Maximize number of
submitted applications
and overall **clicks** on their
digital advertising
campaigns

KEY QUESTIONS

Can we maximize the number of application submissions?

What platforms are optimal to advertise on?

What types of applicants are more likely to apply?
Which leads increase chances of applying?

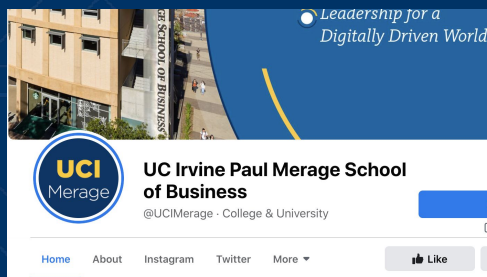
How can we improve click and conversion rates?

Do application submission dates tell us anything about applicant behavior?

DATASETS

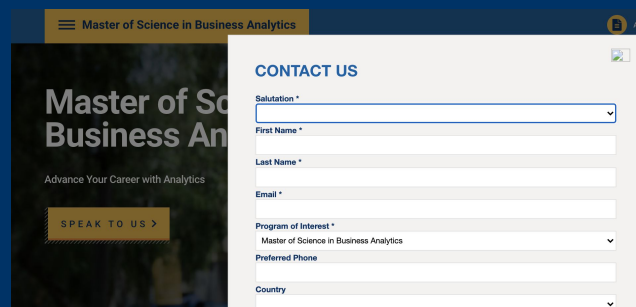
Campaign

- Engagement and performance data
- Variables of interest:
 - Unique clicks
 - CTR



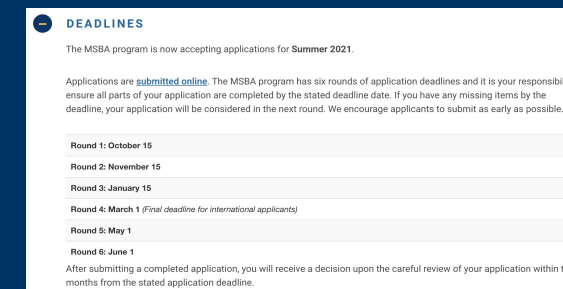
Leads

- Lead source data (web inquiry forms, events)
- Date of first visit
- Gender
- Variables of interest:
 - Stage: Archive, Applicant, Inquiry



Applicants

- Lead source data
- Date of first visit and application submit date
- Residency data
- Program data
- Variables of interest:
 - Time elapsed



CAMPAIGN DATA



Data Prep and Cleaning

- Add Campaign_ID column
- Delete columns with too few values/input
- Filter SMP and add column 'Platform' with text mining function (grep)
- Change columns 'End' and 'Starts' to date format
- Dummy Variables to indicate the delivery of campaigns

A	B	C	D	E	F	G	H	I	J	K
Campaign ID	Campaign Name	Platform	Campaign Delivery	Set Budget	Results	Reach	Impressions	Cost per Results	Amount Spent (USD)	Ends
9	SMP - General (CSU)	Facebook	1	80	29	11148	51279	12.75827586	369.99	9/11/14
22	SMP - General (CSU)	Facebook	1	100	60	23664	71388	6.042	362.52	10/7/14
23	SMP - General (Vang)	Facebook	1	100	77	20907	74412	4.708051948	362.52	10/7/14
24	SMP - General (CSU)	Facebook	1	100	38	16872	58985	9.54	362.52	10/7/14
25	SMP - General (CSU)	Facebook	1	100	29	11651	45582	12.50034483	362.51	10/8/14
26	SMP - General (San J)	Facebook	1	80	64	20112	68271	5.78140625	370.01	10/9/14
27	SMP - General (San D)	Facebook	1	80	45	21044	69710	6.828888889	307.3	10/9/14
29	SMP - General (UC Sa)	Facebook	1	80	29	9364	45422	13.79310345	400	10/13/14
30	SMP - General (Geor)	Facebook	1	80	52	23303	54053	7.69	399.88	10/14/14
31	SMP - General (Univ)	Facebook	1	80	53	22032	83059	7.545849057	399.93	10/14/14
32	SMP - General (CSU)	Facebook	1	80	45	17284	61739	8.885555556	399.85	10/14/14

Exploratory Data Analysis

Instagram advertising is more expensive than FB. If we try to replicate the spend, we can see that for the same spend, we get only half the clicks and impressions.

isFacebook	1							
Sum of Amount Spent (USD)	Sum of Page Likes	Sum of Post Comments	Sum of Post Engagement	Sum of Post Reactions	Sum of Post Shares	Sum of Unique Clicks (All)	Sum of Impressions	
11152.01	32	4	3441	323	4	5579	2,058,825	
isFacebook = IG	0							
Sum of Amount Spent (USD)	Sum of Page Likes	Sum of Post Comments	Sum of Post Engagement	Sum of Post Reactions	Sum of Post Shares	Sum of Unique Clicks (All)	Sum of Impressions	
3791.72	0	1	811	499	9	901	349949	
11148	0	3	2384	1467	26	2649	1,028,850	

Modeling

At a 0.1 alpha level, page likes, reactions, and shares are significant.

Page likes have a positive effect on unique clicks.

A 1% increase in page likes results in a 1.93% **increase** in odds of gaining unique clicks.

A 1% increase in post reactions results in a 3.18% **decrease** in odds of gaining unique clicks.

A 1% increase in post shares results in a 6.3% **decrease** in odds of gaining unique clicks.

```
Call:
glm(formula = UniqueClicks ~ isFacebook + log(Reach) + log(Impressions) +
    log(AmountSpent) + log(PageEngagement) + log(PageLikes +
    1) + log(PostComments + 1) + log(PostEngagement) + log(PostReactions) +
    log(PostShares + 1) + CampaignDuration, data = smp_cd)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-47.110	-15.848	3.155	10.951	46.953

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	316.991	431.408	0.735	0.4686
isFacebook	6.593	17.524	0.376	0.7096
log(Reach)	-6.275	17.237	-0.364	0.7186
log(Impressions)	-32.381	28.975	-1.118	0.2732
log(AmountSpent)	-89.280	62.973	-1.418	0.1673
log(PageEngagement)	2761.840	1021.437	2.703	0.01617
log(PageLikes + 1)	64.799	35.565	1.822	0.0792 .
log(PostComments + 1)	7.861	20.288	0.388	0.6978
log(PostEngagement)	2933.985	1916.211	1.531	0.1370
log(PostReactions)	-14.006	7.536	-1.858	0.0736 .
log(PostShares + 1)	-25.166	13.378	-1.881	0.0704 .
CampaignDuration	3.543	3.188	1.111	0.2759

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> ((exp(coef(logitFB))-1)*100)
      (Intercept)      isFacebook      log(Reach)      log(Impressions)      log(AmountSpent)      log(PageEngagement)      log(PageLikes + 1)
      4799.9087735      1.2195392      -5.3928878      -28.5246969      -18.9246662      -1.8035824      1.9293501
log(PostComments + 1)  log(PostEngagement)  log(PostReactions)  log(PostShares + 1)      CampaignDuration
      0.6659343      53.8364124      -3.1864470      -6.2953206      0.3407656
```

Modeling

If a campaign runs on Facebook, it is likely to get about 95 more clicks than if it were to run on Instagram.

IG base clicks: 90 clicks

FB clicks: 185 clicks

```
Call:
lm(formula = UniqueClicks ~ isFacebook, data = smp_cd)

Residuals:
    Min       1Q   Median       3Q      Max
-131.967  -26.600    2.467   26.033  119.900

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    90.10      17.65    5.103 9.61e-06 ***
isFacebook     95.87      20.39    4.703 3.35e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 55.83 on 38 degrees of freedom
Multiple R-squared:  0.3679,    Adjusted R-squared:  0.3512
F-statistic: 22.11 on 1 and 38 DF,  p-value: 3.347e-05
```

Facebook is better for advertising because we get a higher amount of clicks per campaign as compared to Instagram

Recommendations

Page likes have a positive effect on unique clicks.



Merage should focus on people who **like** the page rather than engage with the post (share/react)

Facebook gets higher clicks per campaign than Instagram



Continue to advertise in a higher proportion on **Facebook** than on Instagram for **SMP**

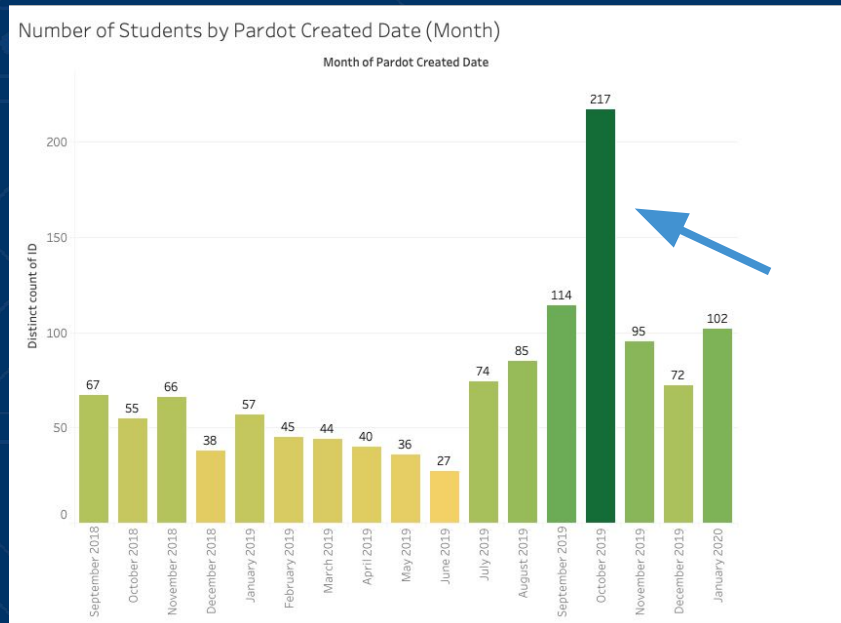
LEADS DATA



Data Prep and Cleaning

- Dummy Variables Transformation: gender, student stage, lead source data
- Mapping the date first visited as variable `time_til_oct`: absolute value difference

lead_webinq	male	stage_archive
1	0	0
1	0	0
1	1	0
1	1	0



Visits to the website spike during October; application window opens in September

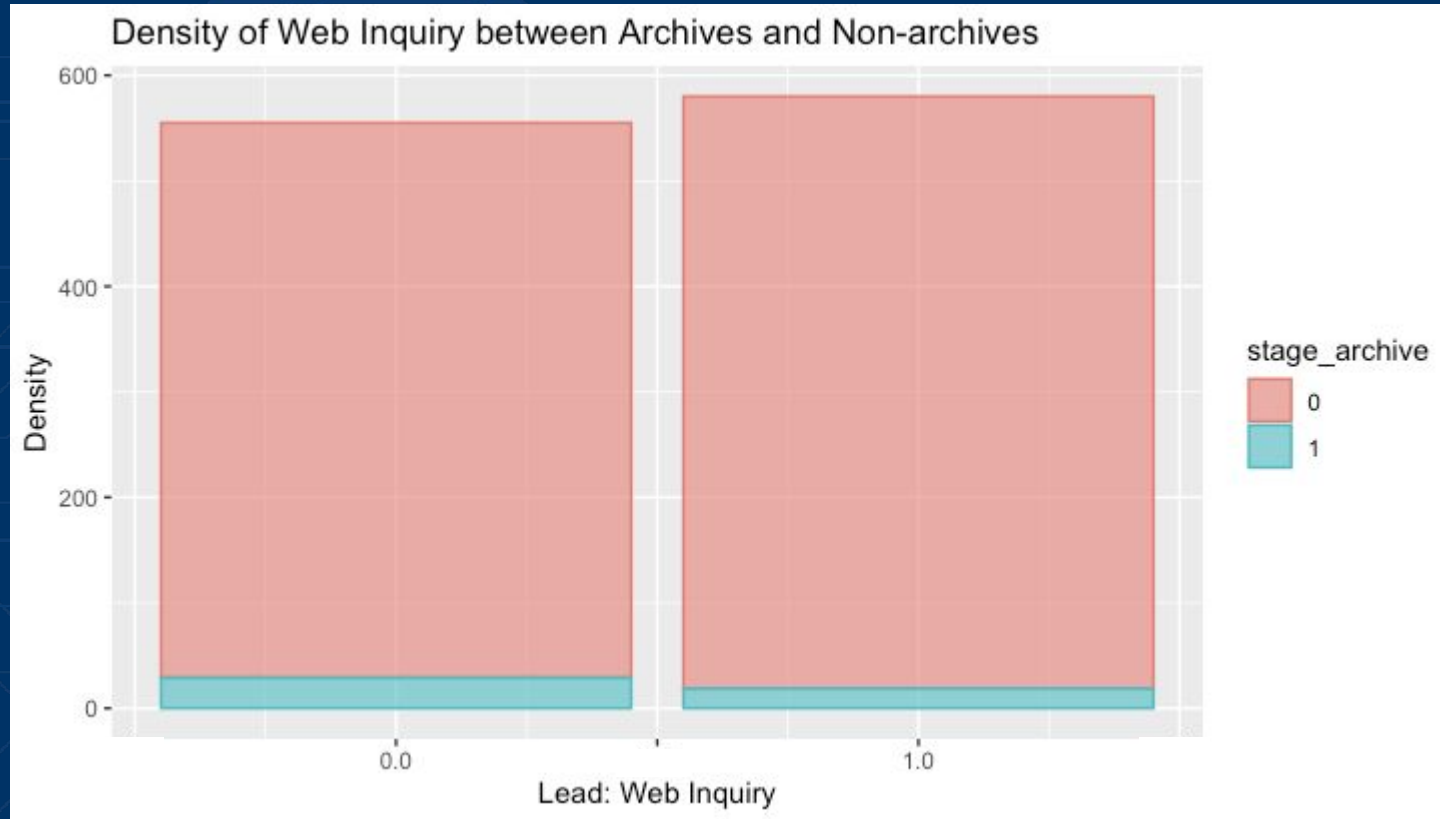
Exploratory Data Analysis: Programs

Number of Students by Program&Stage

Student Sta..	Program			
	MFIN	MIE	MPA	MSBA
Applicant	61	41	37	110
Archive	9	5	5	29
Inquiry	190	138	192	417

MSBA program has more inquiries than other programs, which leads to higher applicant and archive numbers.

Exploratory Data Analysis: Archives



Non-archives are more likely to file a web inquiry (red), and Archives are more likely to not file a web inquiry (blue).

Exploratory Data Analysis: Events



Many people who start an application do not come from an event registration lead.

Modeling

Logistic Regression:

Predicting if they will start an Application

If an applicant fills out a web inquiry form or registers for an event, then they are less likely to start an application than those who don't.

Call:

```
glm(formula = arch_or_appl ~ male + lead_webinq + event_reg +  
     log(time_til_oct), family = binomial(), data = new_data1)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.5308	-1.1249	0.8621	1.0715	2.3066

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.25138	0.17211	1.461	0.1441
male	-0.17253	0.10110	-1.707	0.0879 .
lead_webinq	-0.74238	0.10447	-7.106	1.19e-12 ***
event_reg	-2.82432	0.36266	-7.788	6.82e-15 ***
log(time_til_oct)	0.09796	0.04120	2.378	0.0174 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 2391.3 on 1724 degrees of freedom
Residual deviance: 2227.4 on 1720 degrees of freedom
AIC: 2237.4

Number of Fisher Scoring iterations: 5

Modeling

Logistic Regression:

Predicting if they will submit application or not

When a student fills out a web inquiry form, their odds of applying decreases by 6%

May be because these students are unsure of applying and therefore submit a web inquiry form.

```
> exp(coef(regress_archive_reduced))
```

(Intercept)	male	lead_webinq	log(time_til_oct)
0.4013650	0.8066604	0.4349248	1.4401611

```
Call:
glm(formula = stage_archive ~ male + lead_webinq + log(time_til_oct),
     family = binomial(), data = new_data)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.6807	-1.1261	-0.5133	1.1299	1.6782

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.91288	0.14689	-6.215	5.13e-10	***
male	-0.21485	0.08903	-2.413	0.0158	*
lead_webinq	-0.83258	0.09484	-8.779	< 2e-16	***
log(time_til_oct)	0.36476	0.03729	9.781	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 3009.6 on 2170 degrees of freedom
Residual deviance: 2860.5 on 2167 degrees of freedom
AIC: 2868.5

Number of Fisher Scoring iterations: 4



Recommendations

Inquiry Form & Event Registration
doesn't have positive impact on
submitting application



We recommend Merage to take
a closer look at calls to action &
reactions of those who fill out
forms and attend the event

Females are more likely to apply
even though more males express
more overall interest in programs



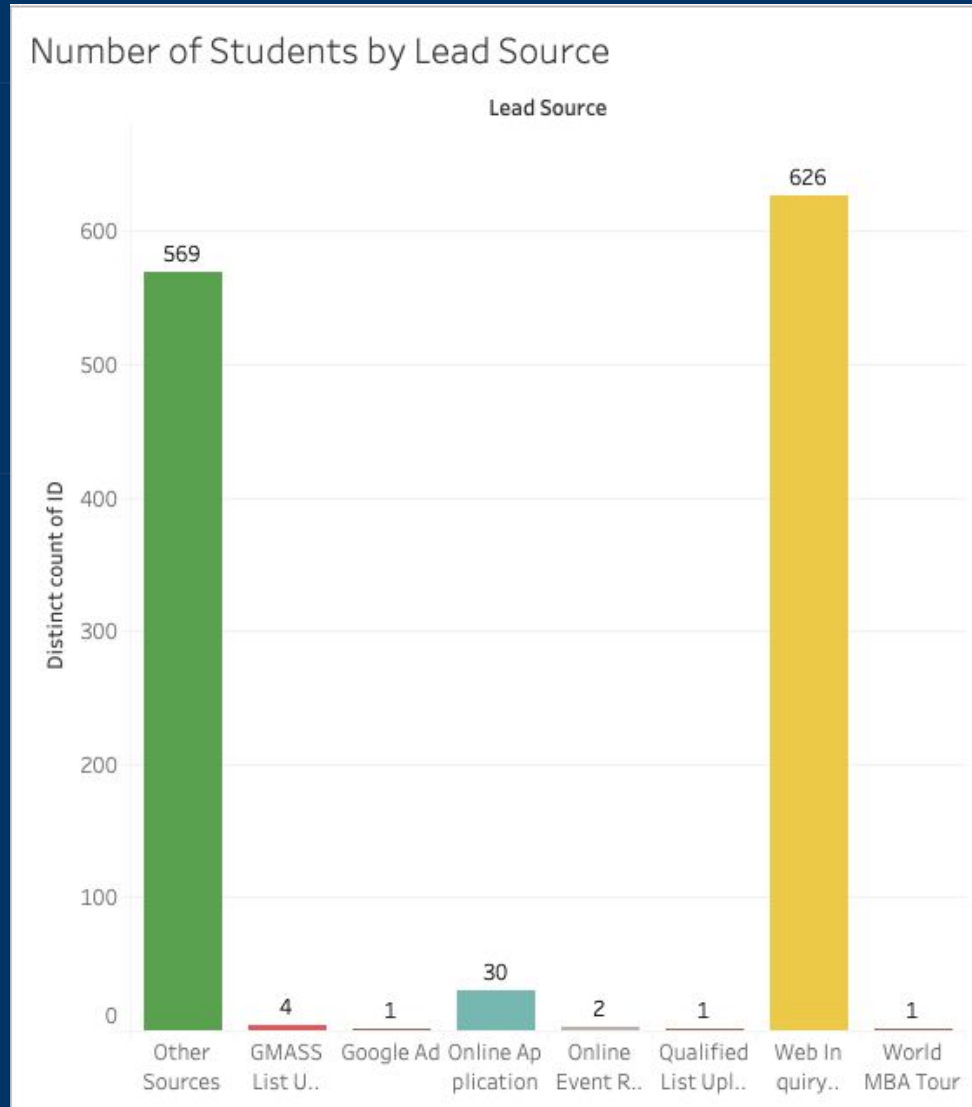
Conduct in depth analysis on
why males drop off in the
conversion funnel as
opposed to females

APPLICANTS DATA



Exploratory Data Analysis

Other sources indicates leads from other sources such as grescoresender, hobsons, gradfair, and gmat.



Merage School Web Inquiry form supplies the most leads

Traffic most likely being driven from other sources

SURVIVAL ANALYSIS

- time elapsed
calculated field
- lead source
information
- residency status
- program data

```
# Import Data
data <- read_csv("~/Desktop/277 data clean/applicants/smp_applicants_datacleaned.csv")

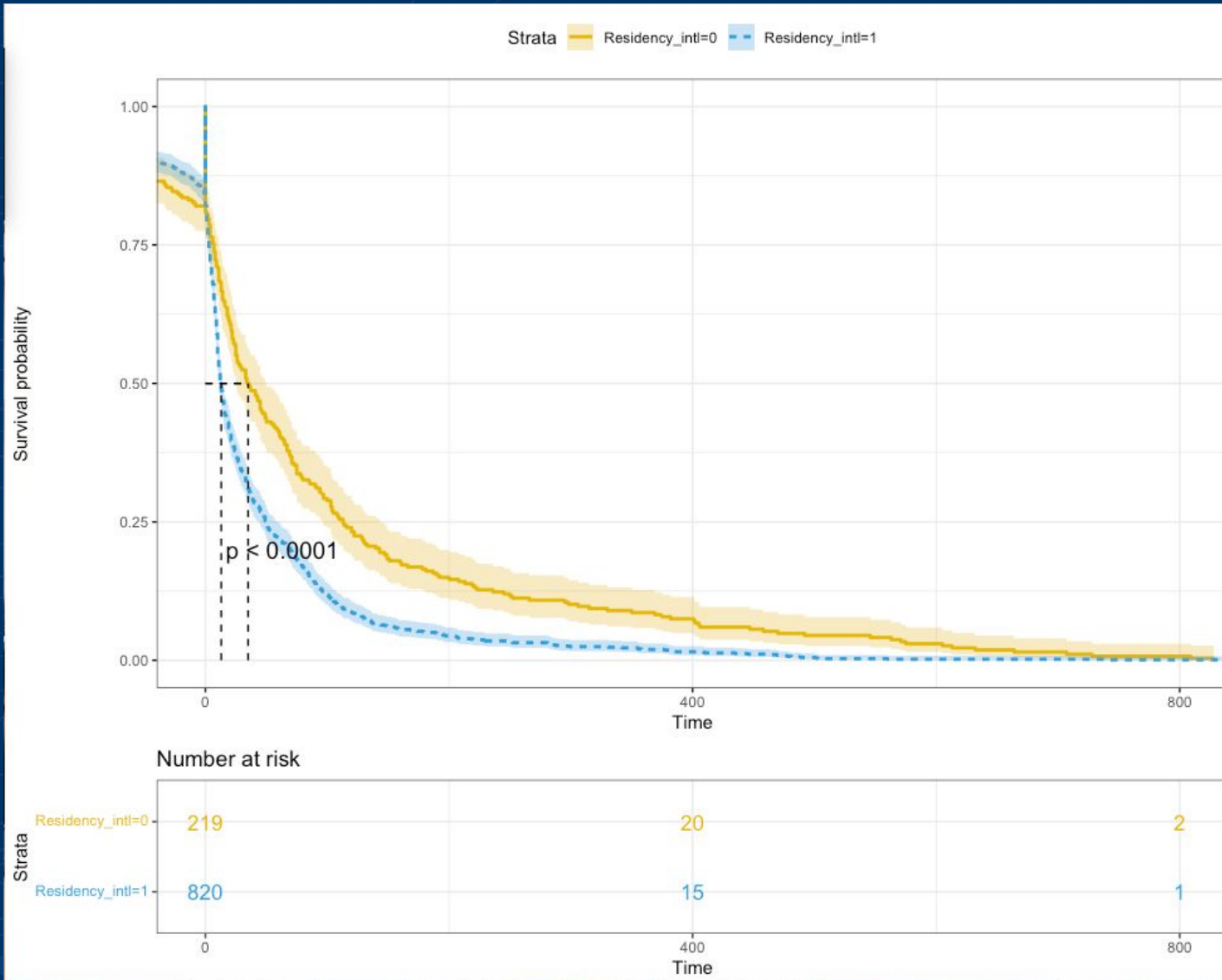
data$status <- rep(1,1239)
fit <- survfit(Surv(time_elapsed, status) ~ Residency_intl, data = data)
print(fit)

summary(fit)
summary(fit)$table

d <- data.frame(time = fit$time,
                n.risk = fit$n.risk,
                n.event = fit$n.event,
                n.censor = fit$n.censor,
                surv = fit$surv,
                upper = fit$upper,
                lower = fit$lower)

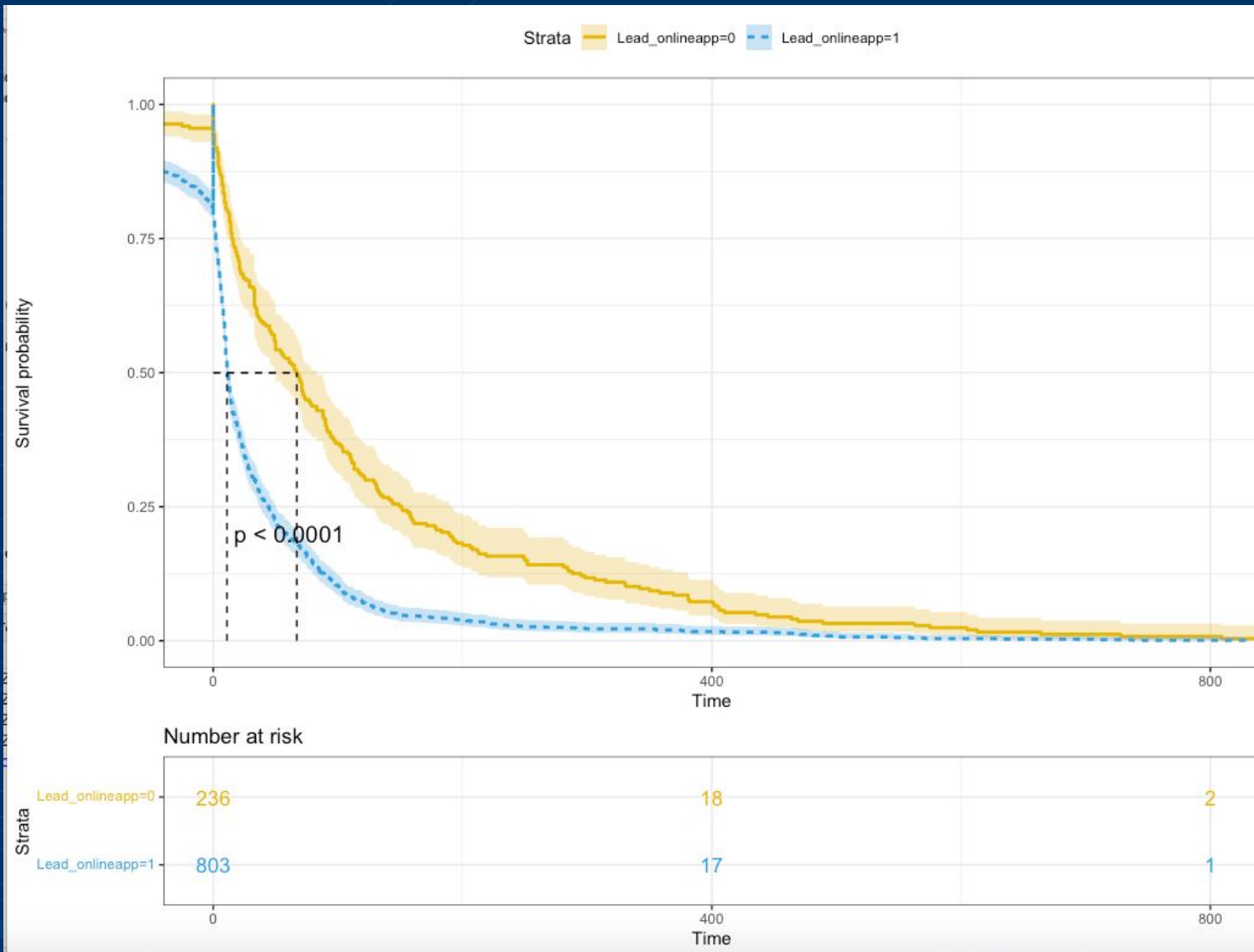
ggsurvplot(fit, data = data,
            pval = TRUE, conf.int = TRUE,
            risk.table = TRUE, # Add risk table
            risk.table.col = "strata", # Change risk table color by groups
            linetype = "strata", # Change line type by groups
            surv.median.line = "hv", # Specify median survival
            ggtheme = theme_bw(), # Change ggplot2 theme
            palette = c("#E7B800", "#2E9FDF"))
```


Survival Analysis: Citizenship



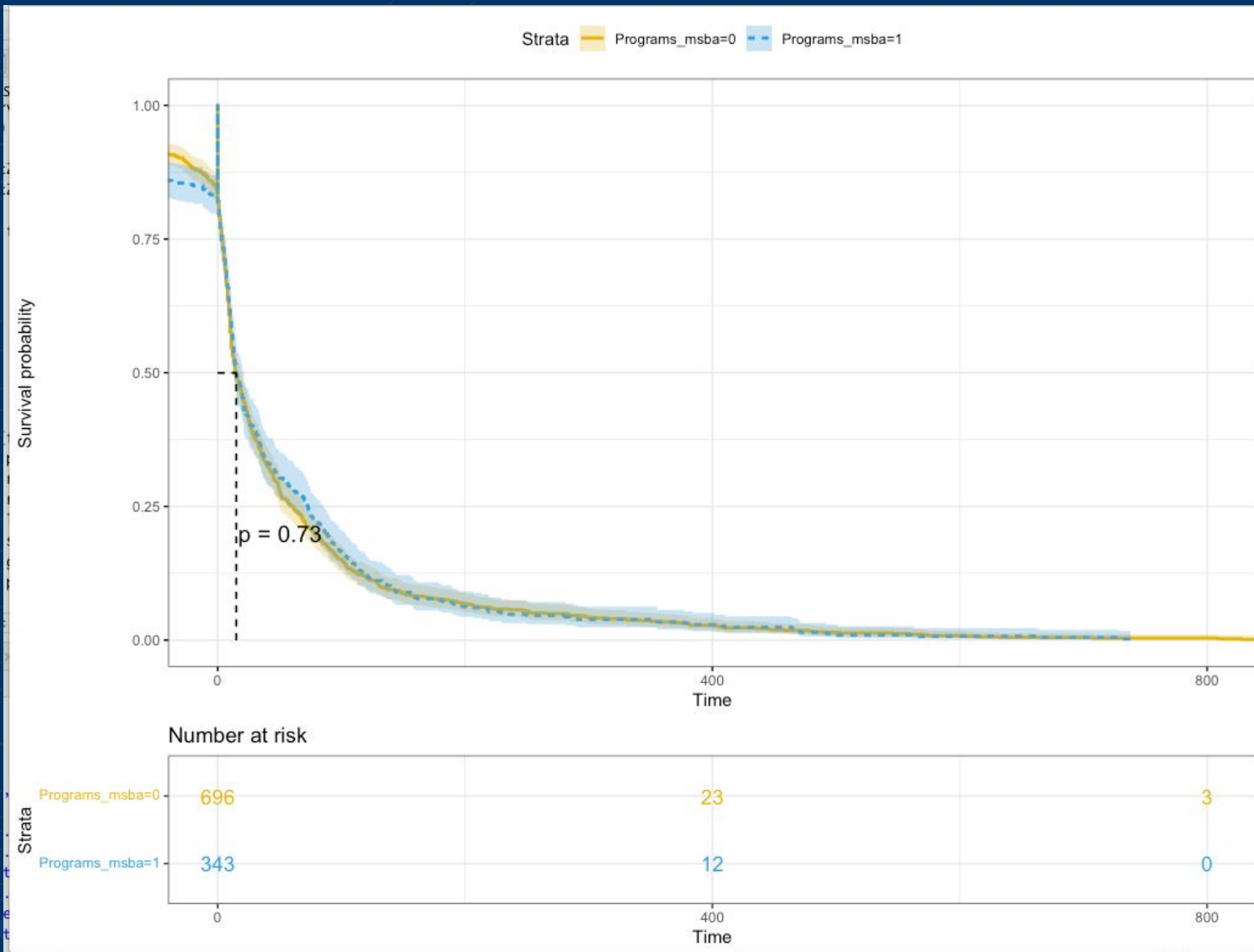
Compared to citizens, international students are more likely to submit applications earlier.

Survival Analysis: Online Application



Compared to applicants from other sources (like gmat, grescoresender, gradfair, etc), Lead_online_applicants (direct applicants) are more likely to submit applications early.

Survival Analysis: Programs



There is no significant difference in submission date between MSBA and non_MSBA.

Modeling

Poisson Regression

At .001 alpha level: All variables are significant.

Lead online applicants have a smaller elapsed time

Very small difference between elapsed time in the programs

```
Call:
glm(formula = time_elapsed ~ Programs_fin + Programs_mpa + Programs_mie +
    Residency_citizen + Lead_onlineapp + Lead_webinquiryform,
    family = poisson(), data = data1)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-20.740	-7.439	-4.721	1.132	52.646

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.590513	0.008935	513.742	<2e-16 ***
Programs_fin	-0.004913	0.009986	-9.505	<2e-16 ***
Programs_mpa	0.117001	0.009266	12.626	<2e-16 ***
Programs_mie	-0.710459	0.017570	-40.437	<2e-16 ***
Residency_citizen	0.679354	0.008171	83.144	<2e-16 ***
Lead_onlineapp	-0.804641	0.008595	-93.618	<2e-16 ***
Lead_webinquiryform	0.137108	0.013572	10.102	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 118716 on 1038 degrees of freedom
Residual deviance: 95723 on 1032 degrees of freedom
AIC: 100932

Number of Fisher Scoring iterations: 6

Recommendations

International Students apply earlier than Citizens



Focus campaign efforts on international students in the beginning rounds and use targeted marketing for residents in later rounds

Direct online applicants have faster rates of submission



Prompt those who have expressed interests from other forms with reminders of submission deadlines

Elapsed time doesn't vary much between programs



No need for Merage to focus campaign efforts based on deadlines

CONCLUSIONS



BUSINESS SOLUTIONS

Focused marketing efforts for Facebook and users who like the Merage FB page

Automate drip mailer communication planner for various lead stages from the conversion funnel

Take a customized effort to improve male conversion rate and to increase applicant numbers for female while retaining conversion rate

Target marketing toward international students for first few submission deadlines and residents for later deadlines

Minimize drop off from lead sources other than “online form” by improving engagement efforts

THANK YOU

Q&A

