Building a Better Understanding of Credit Sesame's Customer Base

Julia Donheiser (leader), Naman Agarwal, Ziqi Deng, Vincent Lin

December 2019

Context, Motivation and Approach

- **Credit Sesame**: What it does and how it profits
- **Our Approach**: Investigate user engagement by platform provide suggestions for increasing revenue
- **API Prototype**: An innovative way to assist decision making

Dataset

Our project utilizes three datasets provided by Credit Sesame. We merged all the datasets by `user_id`.

- `User_Profile`: User demomographics and credit profile information for people who registered in July 2018
- `First_Session`: Summary of user engagement during their first session, including key variables such as platform, time spent on website, and "click-actions"
- `30_Day_User_Engagement`: Same as `First_Session`, but detailing user engagement for each user's first 30 days after registering with CS. Each entry summarizes a single session by a unique user.

Exploring the Dataset

Our EDA focused on understanding Credit Sesame's userbase and determining the best way to segment our analysis.

- Credit score distribution
- Percentage of homeowners
- Demographics by login platform

Age

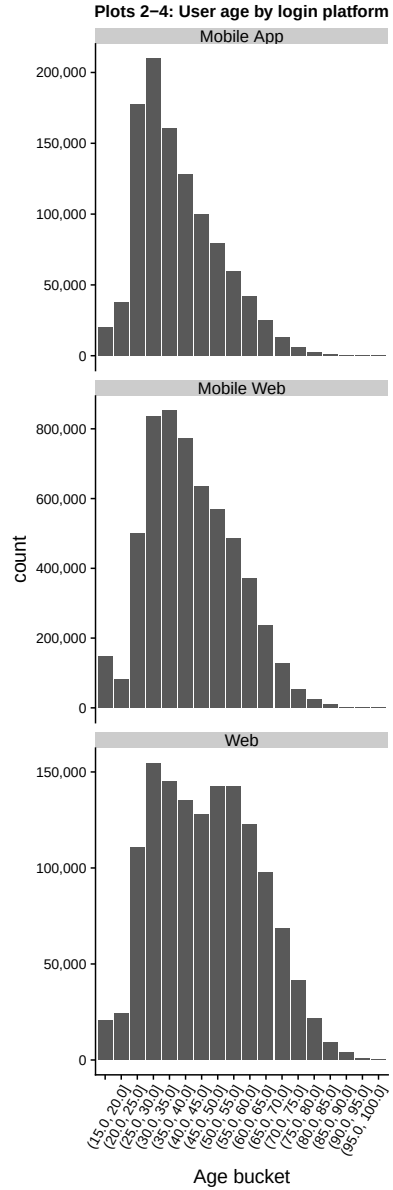


Figure 1: User age by platform

Credit Score Distribution

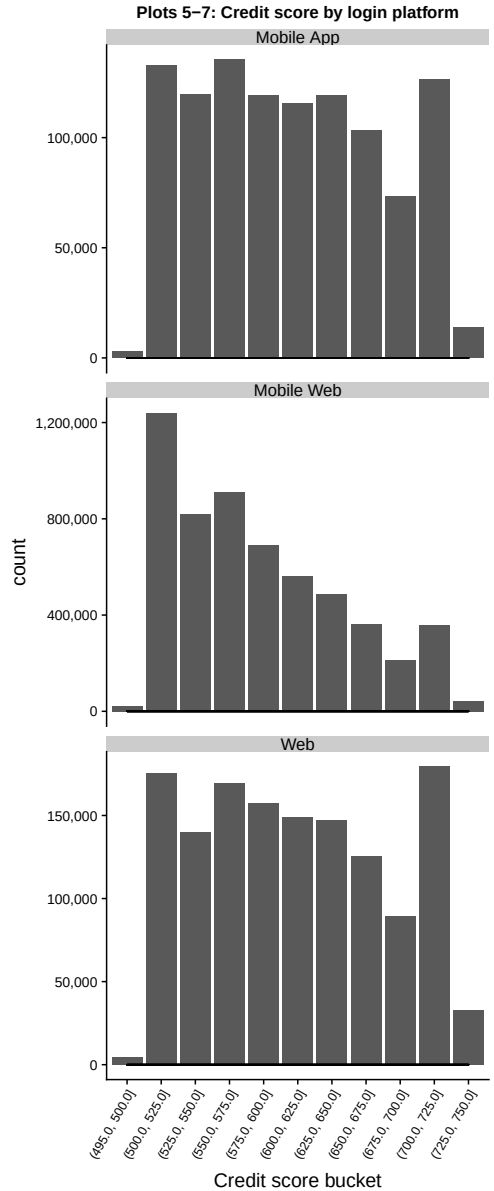


Figure 2: Users' Credit Score Distribution

Homeowner or not?

Percentage of Credit Sesame users who are homeowners

is_homeowner	percentage
FALSE	74.54
TRUE	25.46

Figure 3: Homeowner Ratio

User Engagement

- Analyzed longitudinal engagement patterns
- Developed “user funnel chart” and calculated relevant statistics
- Used logistic regression to predit “click-apply” events

Session time

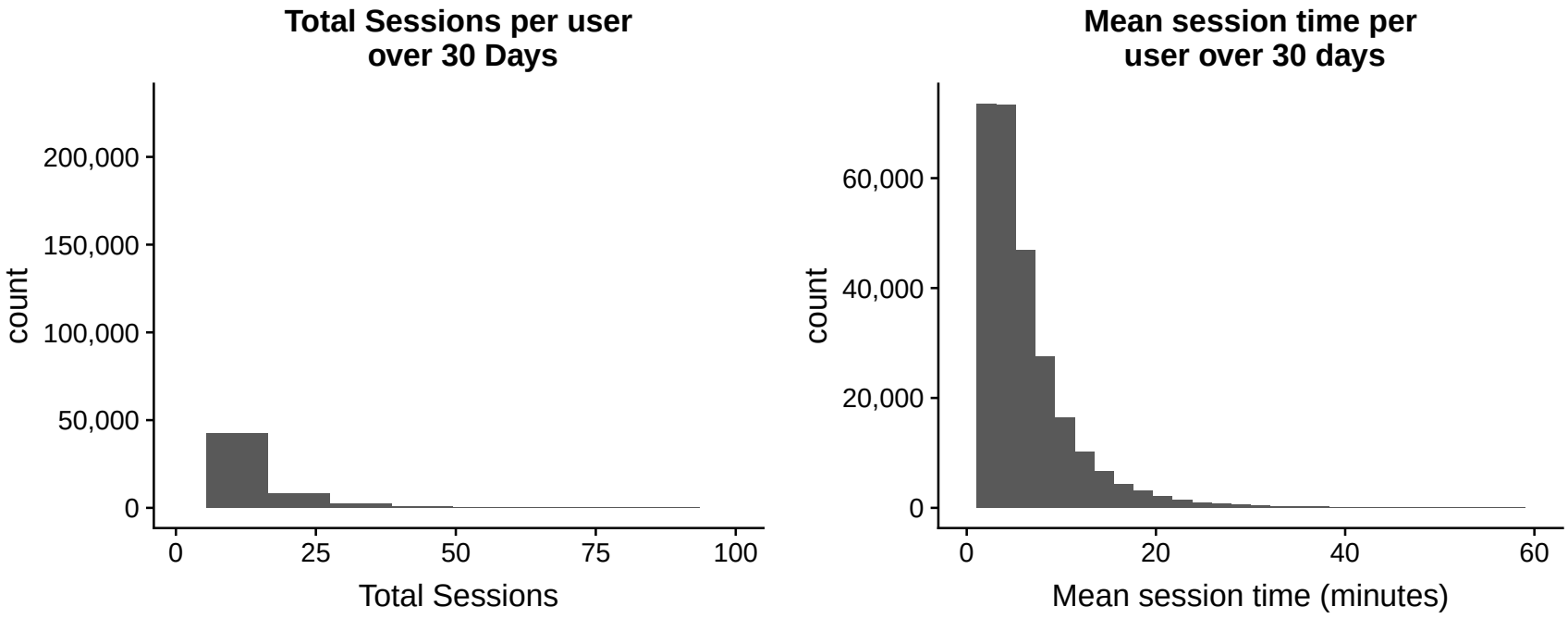


Figure 4: Sessions per user and mean session time

Engagement by Platform

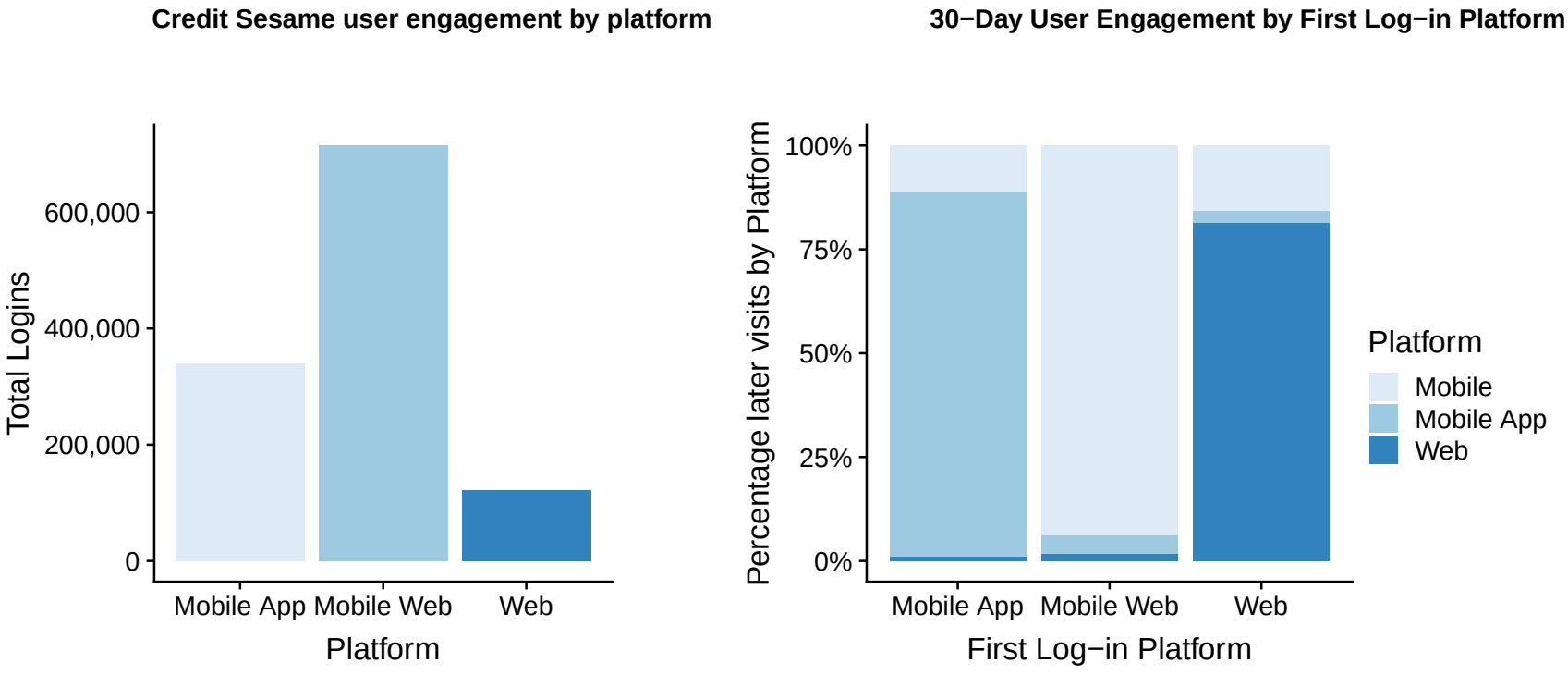
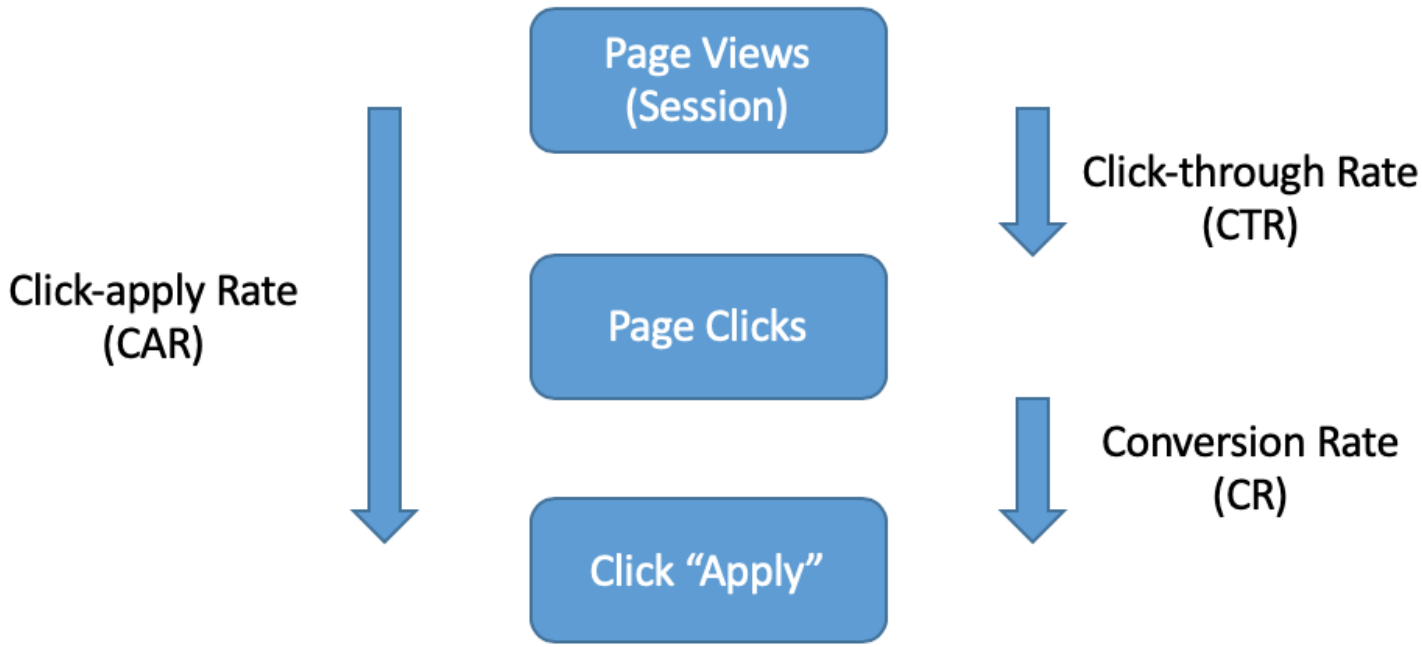


Figure 5: User Engagement by Platform

User Funnel Chart and Investigation



login_platform	mean_ctr	mean_cr	mean_car
Mobile App	1.81	0.03	0.03
Mobile Web	0.96	0.10	0.09
Web	1.03	0.07	0.06

Figure 6: Funnel Chart and Statistics

Logistic Model

Then we tried to predict click-apply events with the following variables

- Whether someone is a homeowner
- Credit card utilization ratio
- Total tradeline accounts opened in last 6 months
- Total inquiries in the past 6 months

Diagnostics

It didn't go too well.

Summary of Logistic Regression Model

```
##
## Call:
## glm(formula = apply_binary ~ is_homeowner + avg_cc_utilization_ratio +
##     count_tradelines_cc_opened_24_months + count_inquiries_6_months,
##     data = eng_user_join)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.6034  -0.1379  -0.1211  -0.1010   0.9153
##
## Coefficients:
##                                       Estimate Std. Error t value
## (Intercept)                           9.528e-02  8.786e-04 108.449
## is_homeownerTRUE                     -1.013e-02  8.616e-04 -11.761
## avg_cc_utilization_ratio              4.408e-02  1.071e-03  41.174
## count_tradelines_cc_opened_24_months -9.469e-05  2.605e-04  -0.363
## count_inquiries_6_months              5.777e-03  1.428e-04  40.449
##                                      Pr(>|t|)
## (Intercept)                            <2e-16 ***
## is_homeownerTRUE                       <2e-16 ***
## avg_cc_utilization_ratio               <2e-16 ***
## count_tradelines_cc_opened_24_months    0.716
## count_inquiries_6_months               <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.1103198)
##
##     Null deviance: 77719  on 700806  degrees of freedom
## Residual deviance: 77312  on 700802  degrees of freedom
##   (476401 observations deleted due to missingness)
## AIC: 443972
##
## Number of Fisher Scoring iterations: 2
```

Takeaways

- Mobile web users seem to earn Credit Sesame the most revenue
- They are also the users with the worst credit scores, and probably need the most services
- To increase revenue, Credit Sesame should delve more into the demographics of their mobile web users as they are the most money-earning userbase
- External knowledge of housing/auto/insurance policies by state would better contextualize users' need for services

Analytics Dashboard

To streamline this process, we made an analytic dashboard.

- Using Vue for the frontend and R plumber for the backend, streamlined data visualization
- Snapshot of user engagement across platforms, which can influence product management, marketing, and more
- Easy to add new key findings to API, code is reproducible

API Prototype Screenshot

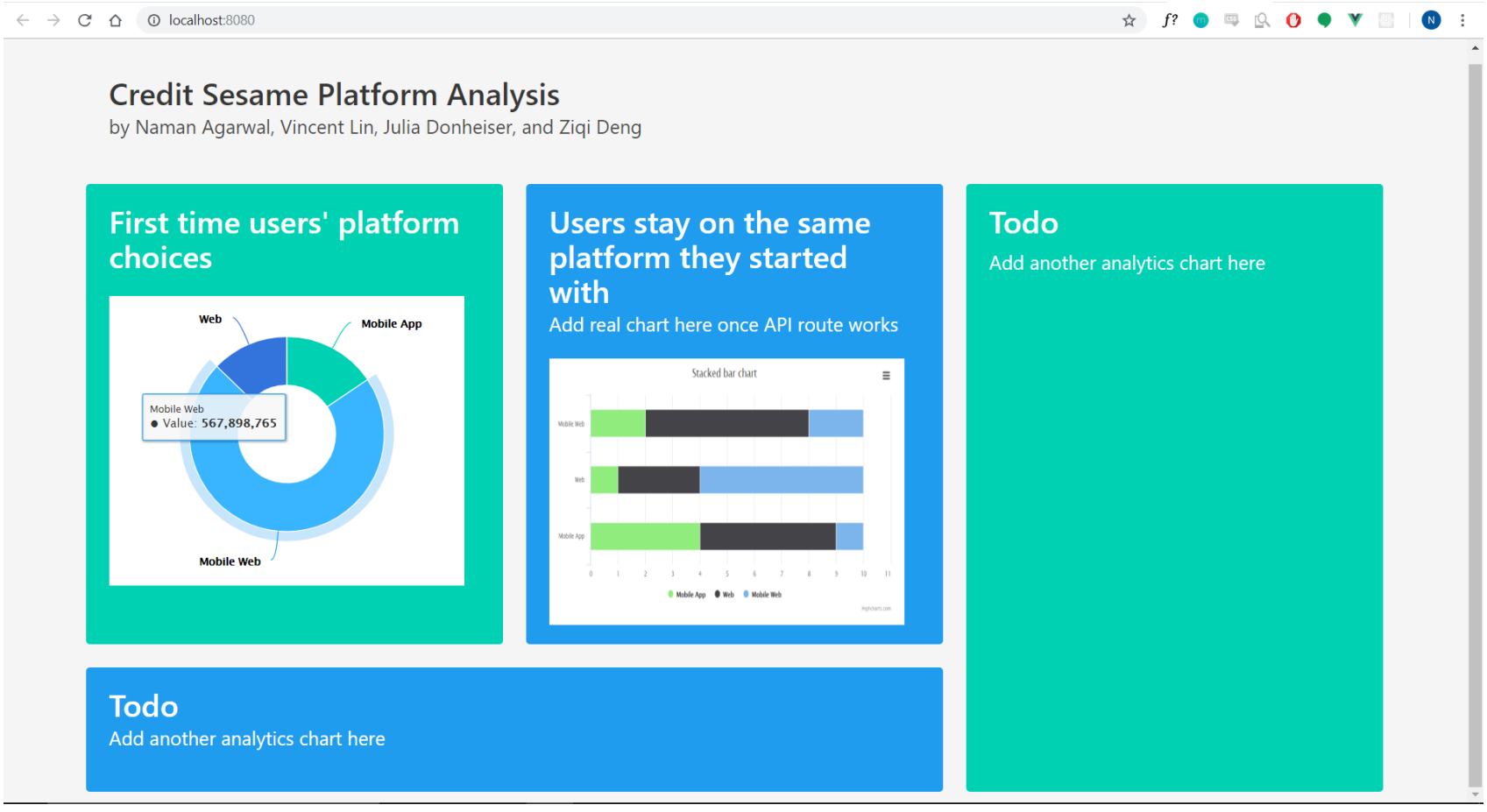


Figure 8: API

Our Conclusion and Advice for Credit Sesame

- Credit Sesame should capitalize on mobile web users, who currently earn them the most revenue
- Determine ways to increase revenue from users who have a low need for outside services