Dan Costello

Pricing Model Development via Analysis of Spam Call Reports

Introduction

ROBOSTPR develops software which blocks unwanted spam calls. To further explore pricing related to customer location and the quantity of spam this area receives, it is necessary to estimate the rate at which spam calls occur throughout California. We will analyze where spam calls occur, when they occur, and the target demographics of spam callers. This analysis will be used to recommend a pricing strategy.

Dataset

We consider data from two sources, one of which being the FCC's Open Data Website. Data is also sourced from the American Community Survey (ACS) from the US Census Bureau. The FCC and US Census Bureau are federal government entities and reliable. The ACS data contains demographic information for ZIP codes in California while the FCC data details individual spam calls.

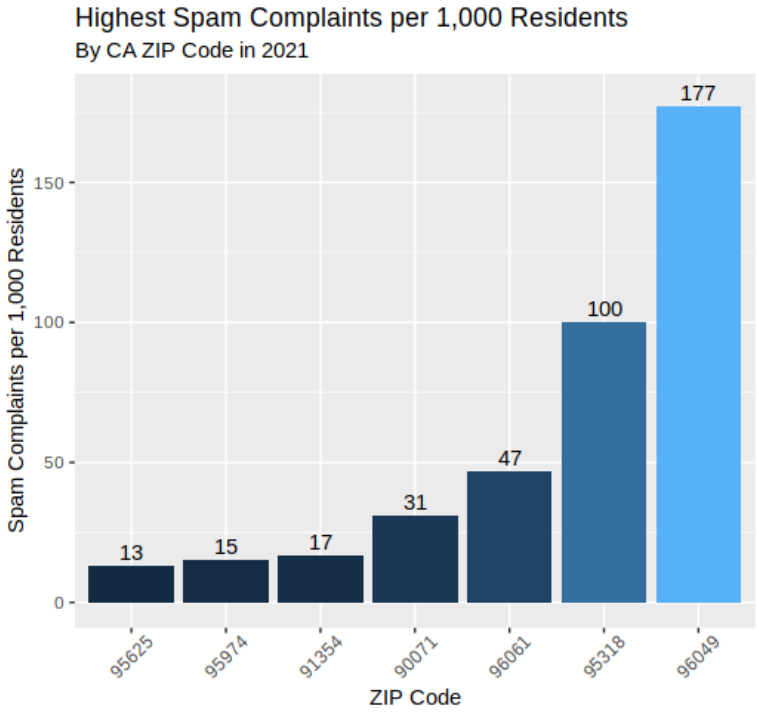
The FCC data contains about 30,000 reports of spam calls in California. This sample provides a reliable estimate of which California cities and ZIP codes receive the most spam calls. Combining the results of FCC data analysis with the ACS dataset will yield demographic information for California cities and ZIP codes that report the most spam to the FCC. We will base a pricing strategy on this demographic information of spam-ridden areas of California.

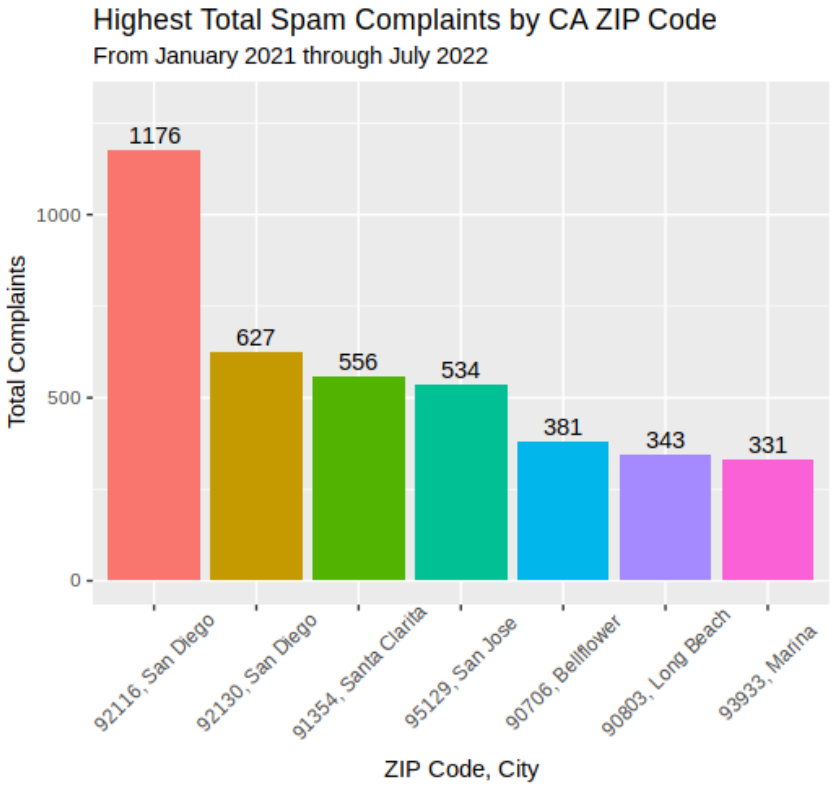
Where

In the FCC data, each report of spam includes the city and ZIP code from which the report was submitted. It is simple to find which ZIP codes reported the greatest quantity of spam complaints but this yields a list of population centers in California. To verify this result, we can also consider the ACS data which includes the population of each ZIP code in California to find where receives the most spam per 1,000 residents. ZIP codes that report the most spam and the most spam per 1,000 residents are most relevant to ROBOSTPR.

The analysis strategy is as follows. We began with the FCC data and removed data points with an NA ZIP code. Then the data is grouped by ZIP code and a total quantity of spam complaints is recorded for each ZIP code. Next, we arrange the data by total quantity of spam in decreasing order. The 7 ZIP codes that reported the most spam are in the bar plot below.

The analysis strategy to find complaints per 1,000 residents is similar. We began by grouping FCC by ZIP code and took their sum to find total complaints by ZIP code. Next, we joined the FCC and ACS data by ZIP code, only considering data from 2021 and excluding data with a population of 0. We found complaints per 1,000 by dividing total complaints by population and multiplying by 1,000. We then arranged the data by complaints per 1,000 in decreasing order. The 7 ZIP codes with the most complaints per 1,000 residents are graphed below.





The bar plots above show the highest total complaints and the highest complaints per 1,000 residents. Both plots have a clear winner, with 92116 in San Diego submitting the most complaints and 96049 near Redding submitting the most per 1,000 residents. The plot of total complaints decreases more slowly than complaints per 1,000 residents.

The plot of total complaints shows that San Diego submits the most complaints by a significant margin, with all of its ZIP codes submitting 2,856 complaints collectively. This is 1,002 more complaints than Los Angeles, with 1854 collectively. Consider the above bar plot per 1,000 residents, specifically the 4 highest observations.

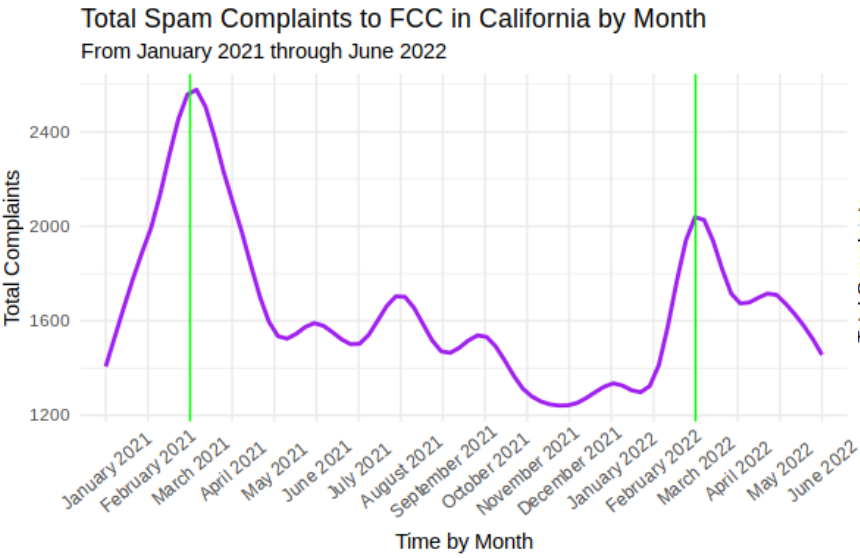
- 96049 is north of Redding and has about 90 people.
- 95318 is in Mariposa county and contains about 220 people.
- 96061 is in Mill Creek, north of Butt Mountain, with only about 21 people.
- 90071 is in a dense part of LA but only has about 126 permanent residents.

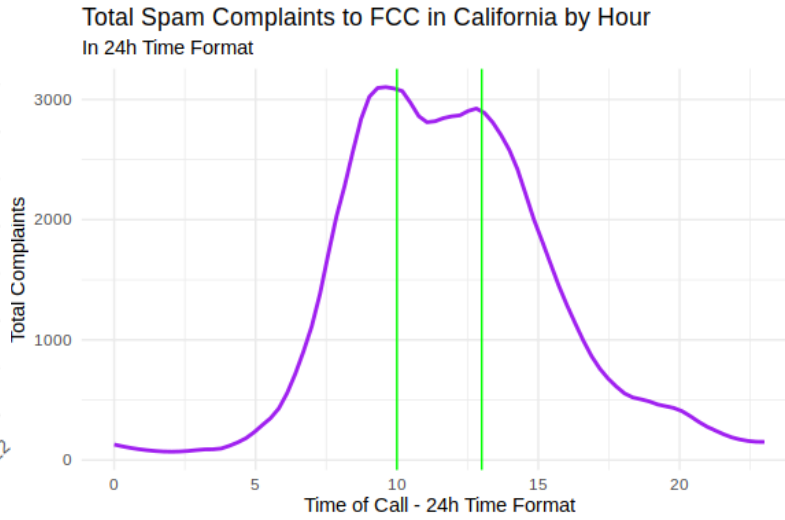
Despite displaying similar data, these plots convey different results. The only overlapping result is 91354 in Santa Clarita. Otherwise, California population centers dominated the plot of total complaints while various obscure regions dominated complaints per 1,000 residents. An estimated 170 complaints per 1,000 residents is not very significant in a ZIP code with only 90 residents. ROBOSTPR should primarily focus on targeting the results of the total plot, specifically California population centers. A secondary consideration of areas with many complaints per 1,000 residents seems to have potential.

When

Each spam report in the FCC dataset includes the time and date when the report was submitted. We must organize the FCC data by data and time. Using the glue package and function, we begin by creating a new column with the year and month of the complaint. Grouping by this column allows us to find the total quantity of complaints submitted in each month. This yields 19 observations, one for each month in the data.

To consider the hour of each report, we use the hour function on the time of issue variable. This outputs an integer corresponding to the hour in a 24h format. We then row bind and summarize to find the total number of observations for each hour. Month and hour data are graphed with ggplot and a smooth line.



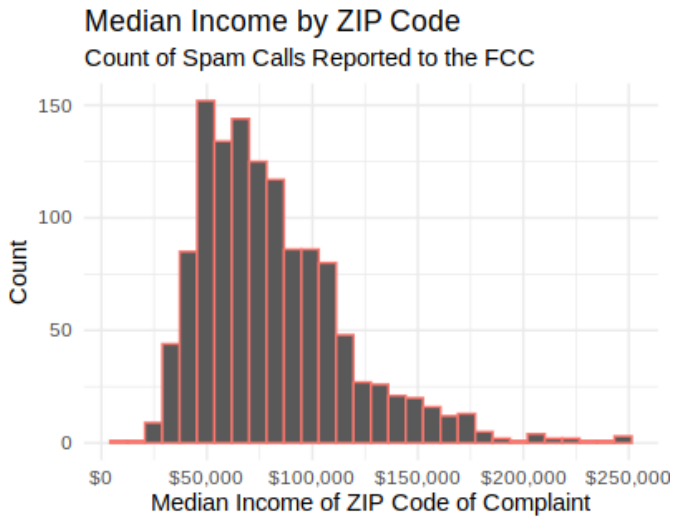


The graph of complaints by hour is about what we expect. Spam is much more common within standard working hours (9am-5pm) than outside these hours. There are two local maximums at 10AM and 1PM. Spam is reported to the FCC most often between 9am and 5pm.

The graph of complaints in each month is more useful. There are several local maximums, most notably in March of 2021 and 2022. This shows that spam is significantly and consistently greater in March than in other months. This will create an increase in demand for spam prevention software that ROBOSTPR should take advantage of.

Demographics

The FCC dataset includes the ZIP code from which each complaint was reported. Also considering the ACS dataset, we have the median income and age from each ZIP code. After joining the datasets, it is straightforward to visualize with a histogram and find the mean of median age and income. Consider the demographic information below.



California Demographic Information						
ZIP	Median Age	Median Income	Av. Income	Total Pop.	Complaints	Per 1,000
91354	37.6	$ 128,308	Above Av.	32544	572	17.6
92116	36.2	$ 74,559	Below Av.	33408	424	12.7
92130	39	$ 155,452	Above Av.	56134	410	7.3
90706	34.6	$ 60,011	Below Av.	77195	379	4.9
90808	42.4	$ 110,625	Above Av.	39330	256	6.5
Total Mean:	38.82	$ 90,520		40,415	15	1.0

See the chart above for the average population, median age, and median income from the ZIP code of each spam report. The 5 ZIP codes with the most spam are also included. The histogram shows that income of spam recipients is skewed to the right, similar to the general population.

Pricing Strategy

Spam call prevention software varies significantly in price, with competitors pricing between $3 and $13 per month. A higher priced model should be considered if ROBOSTPR software outperforms competitors or has significant differentiating features. Otherwise, a reasonable starting price would be 0.005% of the median income in each ZIP code in California, an average of $4.53 per month. For ZIP codes without an available median income, we will use the average price of $4.53. If a customer's ZIP code is unavailable, we will use the average of median incomes of ZIP codes in their area code.

Additionally, the FCC data contains significantly more reports of spam during March than other months. To take advantage of this, each March and April ROBOSTPR should offer a 50% off sale to new customers. This will attract potential customers who have experienced a significant increase in spam calls.

Some areas report spam complaints significantly more. To take advantage of areas with high rates of spam, we should offer an initial 50% discount for three months in the 20 ZIP codes with the greatest quantity of complaints and complaints per 1,000 residents. Consider the following proposed pricing model.

PRICE = (ZIP MEDIAN INCOME * .00005) * $\text{INDICATOR}^{\text{MARCH}}$ * $\text{INDICATOR}^{\text{INITIAL}}$

$\text{INDICATOR}^{\text{MARCH}}$ = .5 if it is March or April, 1 otherwise

$\text{INDICATOR}^{\text{INITIAL}}$ = .5 if the customer's ZIP is in the 20 ZIP codes with the most complaints or complaints per 1,000 residents in the first three months of ROBOSTPR's launch. 1 otherwise.