Modelling Advertising Conversion Rates: A Spotify Case Study

Yuxiang Hong and Alex Bachrach

Math 111A: Mathematical Modelling

Professor Pearson Miller

June 11, 2025

Abstract

For this project, we applied the concept of the SIR model in an advertising environment. The SIR model is a compartmental model that compares rates of epidemic spread between susceptible, infected, and recovered parties using differential equations. Using data from Spotify, we used this same idea to compare the dynamics of potential customers, active paying customers, and lost customer parties over time. To further model the dynamic changes of the potential customer party, we designed a baseline advertising conversion rate and a baseline customer churn rate to predict how effective Spotify's advertising is on an individual level and the turnover of current customers per day. Modelling customer churn rate required use of python to create a linear regression model based off of the limited estimates we had found, and created a piecewise model reflecting the time inputs. Using this, we then could model performance evaluations for the years 2015-2017. We found that this model was comparably accurate, but further evaluation would require more extensive data, which we did not have the resources to collect. We believe that in the future, this extension of the SIR model could be a functional model to estimate the success rates of a company based on its advertising functionality.

Introduction

By the end of 2023, Spotify had accumulated over 600 million active users, with around a third of them being paid subscribers for Spotify premium (Leu, "Spotify's Monthly Active Users 2015-2024"; Leu, "Spotify's Premium Subscribers 2015-2024"). Not only this, but they had also spent around 1.7 billion U.S. dollars on sales and advertising (Faria). Looking at both of these aspects of their company organization, one can surmise that they are doing something right– it can be seen in the huge amass of paying subscribers for a premium service, and one can assume that there is a link between their massive amounts of advertising spending and the amount of customers they have currently.

Using these two ideas, one can think a bit about the SIR model in the field of epidemiology. Using differential equations, this model separates the dynamics of the spread of a disease into 3 compartmentalized groups: susceptible, infected, and recovered. Susceptible is the amount of people in the population that have not been infected yet, infected is the people in the population that have already been infected, and recovered is the amount of people infected that have gotten over the disease. By using this model, one can analyze how a disease spreads over time, and how many people could possibly decrease in the susceptible group over a period of time ("Technical Explainer: Infectious Disease Transmission Models). This also leaves room for development upon the model, which we used at our disposal.

Recent studies have applied compartmental models, particularly the SIR model from epidemiology, to simulate user dynamics under advertising influences in marketing and information diffusion (Jing et al). Based on this information, in the realm of Spotify (or most other companies of the same model, for this sake) one can view the "susceptible" category as those who have not been exposed to Spotify advertisements and are not customers. The "infected" category can refer to those who are active customers to the service, and the "recovered" category can refer to those who have left the service over time. This gives us, respectively, a group of potential customers, current customers, and lost customers that we can view the dynamics rates of over time, and view how a company gains, retains, or loses subscribers over time.

However, most existing models assume constant advertising effectiveness, and few have incorporated its temporal variations. To address this gap, this project proposes a modified SIR-based model that incorporates time-varying advertising effectiveness. Specifically, we model the variations between weekdays and weekends, where advertising is typically more effective during weekends (Pratama & Sugianto).

Model Formulation

To study the dynamics of customer acquisition and churn rate in Spotify, a subscription-based service company, we develop a model inspired by the SIR model in epidemiology.

The total market size N is divided into three compartments:

- $P(t)$: Number of potential subscribers at time t
- $C(t)$: Number of current active subscribers at time t
- $L(t)$: Number of lost (churned) subscribers at time t

We assume that subscribers move from potential to current subscribers via advertisement conversion and from current to lost subscribers through churn. The transitions are modeled by the following

$$\frac{dP}{dt} = -\alpha(t)P(t)$$
$$\frac{dC}{dt} = \alpha(t)P(t) - \beta(t)C(t)$$
$$\frac{dL}{dt} = \beta(t)C(t)$$

Where:

- $\alpha(t)$: Time-dependent daily advertising conversion rate
- $\beta(t)$: Time-dependent daily churn rate

We define:

$$\alpha(t) = \alpha_0 + A\sin\left(\frac{2\pi t}{7} - \phi\right)$$

α_0 is the baseline advertising conversion rate. This formulation captures the weekly variation in advertising effectiveness, where weekends (typically days 6 and 7) see higher conversion rates. The amplitude A represents the variation in response, and aligns the peak of the sine wave to weekends.

Since we are lack of daily churn data, we construct $\beta(t)$ as a piecewise function:

$$\text{year}(t) = \left\lfloor \frac{t}{365} \right\rfloor + 1$$

$$\beta(t) = \begin{cases} 2.14 \times 10^{-6} & \text{if year}(t) = 1 \\ 1.70 \times 10^{-6} & \text{if year}(t) = 2 \\ 1.43 \times 10^{-6} & \text{if year}(t) = 3 \\ \beta_4 & \text{if year}(t) = 4 \\ \beta_5 & \text{if year}(t) = 5 \\ \beta_6 & \text{if year}(t) \geq 6 \end{cases}$$

Where $year(t)$ is a helper function that determines which year we are currently in based on the number of days t since the start of our simulation.

Because our model uses days as the unit of time, we divide t by 365 and take the floor of that value to get the number of full years passed. Then we add 1 because we consider the simulation starting from Year 1, not Year 0. Then, we use the formula

$$1 - (1 - annual\ churn\ rate)^{1/365}$$

to get the daily churn rate for the first three years, which are 2015, 2016, and 2017 since we can only access these three years' data. While, for the remaining three years, we use the linear regression with accessible data to predict the daily churn rate for 2018 ($1.047 * 10^{-6}$), 2019 ($6.917 * 10^{-7}$), and 2020 ($3.367 * 10^{-7}$).

Additionally, we assume a closed system such that:

$$P(t) + C(t) + L(t) = N$$

Below is the detailed unit table for each parameters in our model:

Symbol	Description	Unit
t	Time	days
$P(t)$	Number of potential customers	individuals
$C(t)$	Number of current customers (active subscribers)	individuals
$L(t)$	Number of lost (churned) customers	individuals
N	Total market size	individuals
$\alpha(t)$	Advertising conversion rate	probability/day
$\beta(t)$	Customer churn rate	probability/day
$\alpha 0$	Baseline advertising conversion rate	probability/day

0	Baseline customer churn rate	probability/day
A	Amplitude of weekly advertising effectiveness	probability/day
ϕ	Phase shift to align weekend effect	radians

Mathematical Methods

To analyze the constructed ordinary differential equation (ODE) system, we applied the following mathematical methods:

1. Parameter Transformation:

   Since data we got for advertising conversion rate and churn rate are annual, we converted them into daily rates to fit our time scale (in days):

   $$\alpha_0 = 1 - (1 - annual\ advertising\ conversion\ rate)^{1/365}$$

   and

   $$\beta_0 = 1 - (1 - annual\ churn\ rate)^{1/365}$$

2. Numerical Simulation:

   The system of ODEs is solved using the Runge-Kutta method (through `**scipy.integrate.solve_ivp**` in Python) over a time span of a year.

3. Sensitive Analysis:

   Through the combination of A (amplitude) and (phase shift), we simulated a specific scenario that allows us to study how weekend effects impact the advertising conversion rate and the subscriber growth.

4. Linear Regression:

Due to the lack of the data of annual advertising conversion rate and churn rate after 2018, simple linear regression was applied to predict these data for next three years after 2018 based on the limited available data.

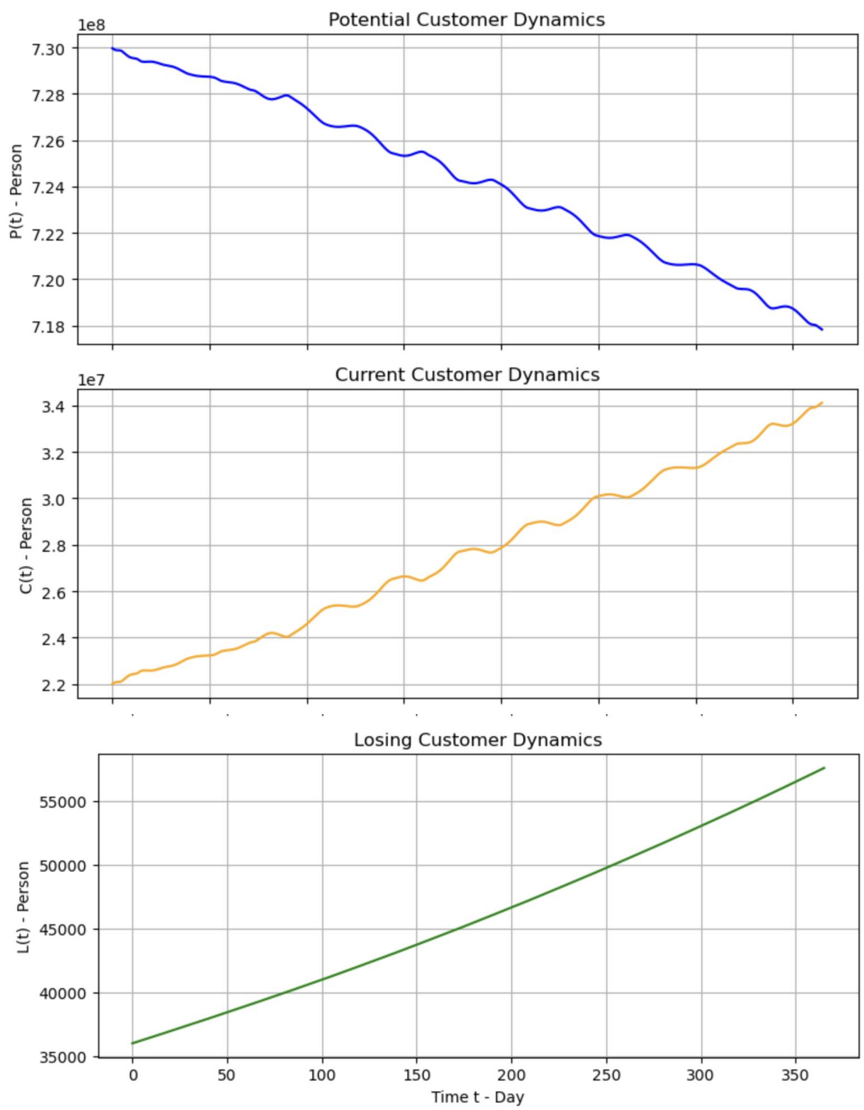
Result

To evaluate the performance of our model, we simulated the customer dynamics over a one-year period using 2015 as our baseline. By inputting known Spotify's data in 2015 (including market size, current subscriber counts, churn counts, advertising conversion rate, and churn rate), we examined how well our model predicts subscriber dynamic behavior and overall market trends.

In applying our mathematical model, we estimate the baseline annual advertising conversion rate α_0 using available public commercial data. We first multiplied the average click-through rate for Spotify advertisements (0.06%) with the conversion rate of spotify of people into active premium subscribers (28.3%), resulting in an estimated annual conversion rate of approximately 1.698% ("Audio Ads Help Primark Drive In-Store Traffic"; Ingham). Since the data we got is annual, we converted it into daily advertising conversion rate, which is $4.7 * 10^{-5}$. Since we have had the annual churn rate for premium subscribers in 2015, which is 0.078% (based on our own calculations), we converted it into daily churn rate ($2.14 * 10^{-6}$).

To reflect the weekend effects on our model, we set the peak of weekend effects between Saturday and Sunday, when t = $\frac{6+7}{2}$. We plugged it into $\frac{2\pi t}{7} - \phi = \frac{\pi}{2}$ to get the value of phase shift ϕ ($\frac{19\pi}{14}$). Based on known data, we adjusted A to be 0.0005 for better reflecting weekend effects. At this stage, we have had all values for needed parameters and initial conditions in

2015 – market size is $7.52 * 10^8$ persons, current premium subscriber counts is $2.2 * 10^7$ persons, and the lost subscriber counts is $3.6 * 10^4$ persons (Leu, “Streaming Music Subscribers Worldwide 2019-2023; Iqbal). We proceeded to numerically solve our ODE system using Python. Specifically, we employed `scipy.integrate.solve_ivp` function with a Runge-Kutta method to simulate the daily customer dynamics over one year and visualized the output to compare with the actual reported Spotify’s data in 2016 from external sources.



Year\Count	Total Market (Person)	Potential Customer (Person)	Current Customer (Person)	Losing Customer (Person)
2015	7.52*10^8	7.30*10^8	2.20*10^7	3.60*10^4
2016		6.94*10^8	3.60*10^7	5.80*10^4
2016 (Model Output)		7.17*10^8	3.41*10^7	5.76*10^4

Relative Error	3.31%	5.28%	0.69%

Overall, our model produced a subscriber dynamics that aligns closely with Spotify's reported 2016 data, which demonstrates a reasonable relative error within acceptable modeling bounds.

Discussion and Interpretation

The results generated by our model indicates that a compartmental approach based on the SIR model can capture meaningful patterns in subscriber dynamic behavior over time. By simulating advertising conversion rate and churn rate on a daily basis under the weekend effects, our model closely reflects real-world dynamics – at least during the timeframe for which we had reliable data. The alignment between the simulated 2016 subscriber count and the actual data supports the model's baseline validity.

However, some discrepancies begin to emerge when we simulate into the future. This is largely due to the lack of detailed commercial data on customer size and churn rate, which makes it difficult to validate our churn modeling precisely. Moreover, we also found that our model shows significant relative errors in simulating years after 2019, and the most important reason for this was the impact of Covid-19. Due to the impact of Covid-19, most countries implemented home quarantine systems, which has implicitly increased the public's reliance on and usage of internet products, inevitably significantly boosting the advertising conversion rate compared to before. Meanwhile, due to the public's high reliance on internet products, the subscriber churn rate was also affected accordingly – only a very small portion of subscribers canceled their premium subscription services. Because our model mainly focuses on the weekend effect and does not take into account these important influencing factors that emerged since 2019, the accuracy of the model output would decrease significantly when simulating years after 2019.

Although there are some factors that affect the accuracy of the final output results of the model, our model still successfully captures the dynamic behavior trends of subscribers under weekend effects, which is in line with our initial expectations for this project. With the emergence of more accessible data in the future, our model can further improve accuracy and gradually incorporate new influencing factors to better simulate complex business market behaviors. This will assist many subscription service companies like Spotify in making decisions related to advertising to grow the customer base more effectively.

Conclusions and future work

While our model heeded some accurate results, the major hindrance to its finality was the lack of accessible data to our model. We can surmise, however, that this model will work overall because of its accuracy and its small relative errors in the years tabulated. This is a model that should be able to accurately predict how a company's customer base reacts to advertising, and one can use this information to model dynamic rates of which a company gains, loses, or retains customers over time. One can also use this model in a different company with a similar business model too– while we did not have the time to explore this, we could search for data for freemium-based companies like Youtube, Apple Music, and Zoom to see if it is any more accessible than Spotify.

In the future, moreover, one other idea that could be of use exploring to make our model more robust could be the aspect of how it relates to CPI. We tried this model at the beginning of this quarter, however, the lack of accessible data made it difficult to determine whether the churn rate regression listed earlier had any relation to CPI and a consumer's purchasing power. Being able to estimate how a company’s potential SIR model changes over time with the additional estimate of inflation can give a more insightful look to how well an advertising campaign can run with the onset of positive or negative inflation, affecting a potential consumer's purchasing power.

Overall, if given the ability to use accurate and reflective data of a company’s consumer base (especially that of lost consumers over a period of time), our model could effectively reflect a company’s effectiveness in the realm of advertising.

Works Cited

"Audio Ads Help Primark Drive In-Store Traffic." *Spotify*, Spotify Advertising Team, 27, Oct. 2021, https://ads.spotify.com/en-US/inspiration/primark-ad-studio-case-study/.

Faria, Julia. "Spotify's Sales and Marketing Costs 2013-2023," *Statista*, 14, Aug. 2024, https://www.statista.com/statistics/813757/spotify-sales-marketing-costs/.

Iqbal, Mansoor. "Spotify Revenue and Usage Statistics (2025)," *Business of Apps*, 25, Feb. 2025, https://www.businessofapps.com/data/spotify-statistics/.

Ingham, Tim. "Spotify is Converting More People Into Paying Subscribers Than Ever Before," *Music Business Worldwide*, 31, Jan. 2016, https://www.musicbusinessworldwide.com/spotify-is-converting-more-people-into-paying-customers-than-ever-before/.

Jing, Yi et al. "Improved SIR Advertising Spreading Model and Its Effectiveness in Social Network." *Volume 129*, Procedia Computer Science, 2018, https://www.sciencedirect.com/science/article/pii/S1877050918302679

Leu, Patrick. "Spotify's Monthly Active Users 2015-2024," *Statista*, 21, Mar. 2025, https://www.statista.com/statistics/367739/spotify-global-mau/.

Leu, Patrick. "Spotify's Premium Subscribers 2015-2024," *Statista*, 21, Mar. 2025, https://www.statista.com/statistics/244995/number-of-paying-spotify-subscribers/.

Leu, Patrick. "Streaming Music Subscribers Worldwide 2019-2023," *Statista*, 4, Nov. 2024. https://www.statista.com/statistics/669113/number-music-streaming-subscribers/.

Pratama, Satrya and Sugianto, Dwi. "Temporal Patterns in User Conversions: Investigating the Impact of Ad Scheduling in Digital Marketing." *Vol. 1 No. 2 (2024): Regular Issue September*, Journal of Digital Market and Digital Currency, 2024, https://jdmdc.com/index.php/JDMDC/article/view/10.

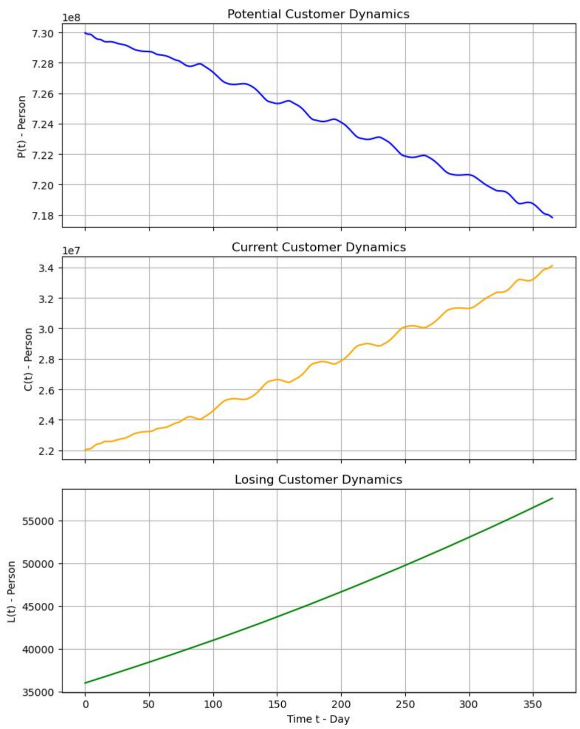
"Technical Explainer: Infectious Disease Transmission Models." *Center for Disease Control*, 6, Jan. 2025, https://www.cdc.gov/cfa-modeling-and-forecasting/about/explainer-transmission-models.html.

Appendices

```python
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
```
```python
from scipy.integrate import solve_ivp
```
```python
# parameter setting
alpha_0 = 4.7e-5#0.000047     # baseline of adv conversion rate
A = 0.00005          # periodic fluctuation
phi = 19*np.pi/14           # fluctuation phase
beta_0 = 2.14e-6 #1.7e-6    # baseline of churn rate
#m = 0#0.000035/365
# ODE system
def ode_system(t, y):
    P, C, L = y
    alpha_t = alpha_0 + A * np.sin(2 * np.pi * t / 7 - phi)
    alpha_t = max(alpha_t, 0)  # should be non-negative
    beta_t = beta_0
    dPdt = -alpha_t * P
    dCdt = alpha_t * P - beta_t * C
    dLdt = beta_t * C
    return [dPdt, dCdt, dLdt]
# initial conditions (data in 2025)
N = 752000000#729964000
C0 = 22000000#36000000
L0 = 36000#58000
P0 = N - C0 - L0
y0 = [P0, C0, L0]
# time setting
t_span = (0, 365)  # time interval 365
t_eval = np.arange(0,366)
```
```python
# Solve ODE
sol = solve_ivp(ode_system, t_span, y0, t_eval=t_eval)
t_vals, P_vals, C_vals, L_vals = sol.t, sol.y[0], sol.y[1], sol.y[2]
```
```python
# Visualization
fig, axes = plt.subplots(3, 1, figsize=(8, 10), sharex=True)
axes[0].plot(t_vals, P_vals, color='blue')
axes[0].set_ylabel('P(t) - Person')
axes[0].set_title('Potential Customer Dynamics')
axes[1].plot(t_vals, C_vals, color='orange')
axes[1].set_ylabel('C(t) - Person')
axes[1].set_title('Current Customer Dynamics')
axes[2].plot(t_vals, L_vals, color='green')
```

```
axes[2].set_ylabel('L(t) - Person')
axes[2].set_title('Losing Customer Dynamics')
axes[2].set_xlabel('Time t - Day')
for ax in axes:
    ax.grid(True)
plt.tight_layout()
plt.show()
print(f"P(365) = {sol.y[0][-1]:,.0f}")
print(f"C(365) = {sol.y[1][-1]:,.0f}")
print(f"L(365) = {sol.y[2][-1]:,.0f}")
```

`![png](output_4_0.png)`



    P(365) = 717,828,153
    C(365) = 34,114,280
    L(365) = 57,567

```python
#churn_rates = [0.00000214, 0.00000170, 0.00000143]
churn_rates = [2.14e-6, 1.7e-6, 1.43e-6]
years = [1, 2, 3]  # representing the 1st, 2nd, and 3rd year
```
```python
import numpy as np
from sklearn.linear_model import LinearRegression
X = np.array(years).reshape(-1, 1)
y = np.array(churn_rates)
model = LinearRegression()
model.fit(X, y)
# predict daily churn rate in future three years
future_years = np.array([4, 5, 6]).reshape(-1, 1)
predicted_churn = model.predict(future_years)
```

print(predicted_churn)
```
    [1.04666667e-06 6.91666667e-07 3.36666667e-07]
```python
```
```

```
[1]: import numpy as np
     import matplotlib.pyplot as plt
     from sklearn.linear_model import LinearRegression
```

```
[2]: from scipy.integrate import solve_ivp
```

```
[3]: # parameter setting
     alpha_0 = 4.7e-5#0.000047     # baseline of adv conversion rate
     A = 0.00005         # periodic fluctuation
     phi = 19*np.pi/14           # fluctuation phase
     beta_0 = 2.14e-6 #1.7e-6    # baseline of churn rate
     #m = 0#0.000035/365

     # ODE system
     def ode_system(t, y):
         P, C, L = y
         alpha_t = alpha_0 + A * np.sin(2 * np.pi * t / 7 - phi)
         alpha_t = max(alpha_t, 0)  # should be non-negative
         beta_t = beta_0
         dPdt = -alpha_t * P
         dCdt = alpha_t * P - beta_t * C
         dLdt = beta_t * C
         return [dPdt, dCdt, dLdt]

     # initial conditions (data in 2025)
     N = 752000000#729964000
     C0 = 22000000#36000000
     L0 = 36000#58000
     P0 = N - C0 - L0
     y0 = [P0, C0, L0]

     # time setting
     t_span = (0, 365)  # time interval 365
     t_eval = np.arange(0,366)
```

```
[4]: # Solve ODE
     sol = solve_ivp(ode_system, t_span, y0, t_eval=t_eval)
     t_vals, P_vals, C_vals, L_vals = sol.t, sol.y[0], sol.y[1], sol.y[2]
```

```
[5]: # Visualization
     fig, axes = plt.subplots(3, 1, figsize=(8, 10), sharex=True)

     axes[0].plot(t_vals, P_vals, color='blue')
     axes[0].set_ylabel('P(t) - Person')
     axes[0].set_title('Potential Customer Dynamics')

     axes[1].plot(t_vals, C_vals, color='orange')
     axes[1].set_ylabel('C(t) - Person')
     axes[1].set_title('Current Customer Dynamics')

     axes[2].plot(t_vals, L_vals, color='green')
     axes[2].set_ylabel('L(t) - Person')
     axes[2].set_title('Losing Customer Dynamics')
     axes[2].set_xlabel('Time t - Day')

     for ax in axes:
         ax.grid(True)

     plt.tight_layout()
     plt.show()
     print(f"P(365) = {sol.y[0][-1]:,.0f}")
     print(f"C(365) = {sol.y[1][-1]:,.0f}")
     print(f"L(365) = {sol.y[2][-1]:,.0f}")
```

```
[6]: #churn_rates = [0.00000214, 0.00000170, 0.00000143]
     churn_rates = [2.14e-6, 1.7e-6, 1.43e-6]
     years = [1, 2, 3]  # representing the 1st, 2nd, and 3rd year
```

```
[7]: import numpy as np
     from sklearn.linear_model import LinearRegression

     X = np.array(years).reshape(-1, 1)
     y = np.array(churn_rates)

     model = LinearRegression()
     model.fit(X, y)

     # predict daily churn rate in future three years
     future_years = np.array([4, 5, 6]).reshape(-1, 1)
     predicted_churn = model.predict(future_years)
     print(predicted_churn)
```

```
[1.04666667e-06 6.91666667e-07 3.36666667e-07]
```