Objective: Develop a model for predicting customer churn at Cell2Cell

Deliverables:

1. Write-up of 5 pages max answering the following questions, 1.5 line spacing.
2. Pdf or slideshow of a contingency plan explaining incentives to offer to customers.

Constant Variables:

- Marginal cost of providing cell service is $0
- Cell2cell.csv contains:
  - 71,047 customers overall, 40,000 in calibration and 31,047 in validation.
  - 20,000 in calibration churn (50%), 609 in validation churn (1.962%)

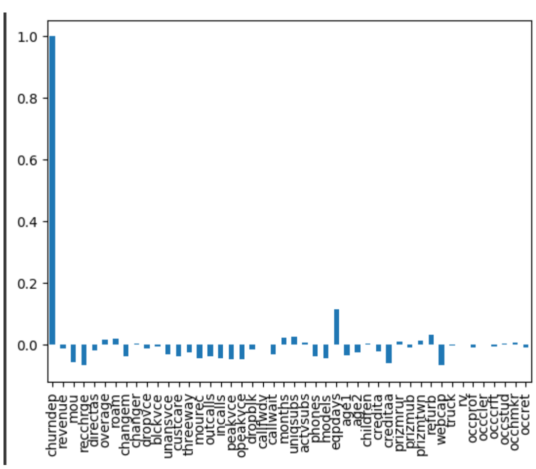
Tips:

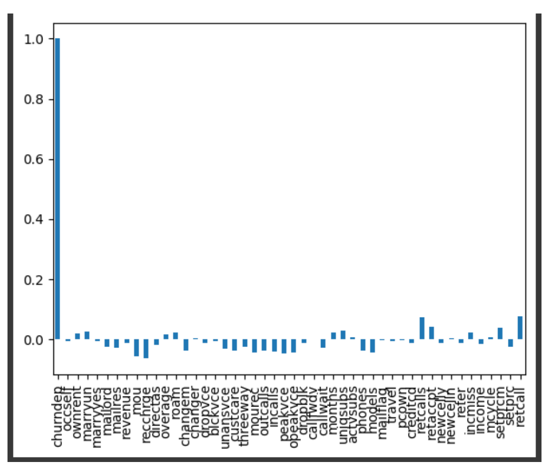
- Logistic regression is best choice for developing this model
- Need at least a lift of 170 (70% higher than average churn probability)

1. **Briefly describe your predictive churn model. How did you select variables to be included in the model?**

We first plotted the direct correlation between each variable and ‘churndep’ in a bar graph. This helped us visualize which variable to potentially focus on.

```
data[ ['churndep', 'revenue', 'mou', 'recchrge', 'directas', 'overage', 'roam', 'changem', 'changer', 'dropvce',
       'outcalls', 'incalls', 'peakvce', 'opeakvce', 'dropblk', 'callfwdv', 'callwait', 'months', 'uniqsubs', 'a
       'eqpdays', 'age1', 'age2', 'children', 'credita', 'creditaa', 'prizmrur', 'prizmub', 'prizmtwn', 'refurb'
       'occprof', 'occcler', 'occcrft', 'occstud', 'occhmkr', 'occret'] ].corr()['churndep'].plot.bar()
```





From there, we then ran a logistic regression on the variables that and looked at their coefficients and p-values to locate statistically significant predictors of ‘churndep’. Finally, we decided as a team to establish a list of 17 variables to base our model off of. We selected these variables for their nature of being relevant, complementary to each other, and having notable statistical significance.

Variables we included to have ~174 lift:
var_list = ['revenue', 'recchrge', 'marryun', 'months', 'uniqsubs', 'eqpdays', 'refurb', 'dropvce', 'actvsubs', 'retcall']

Converted into definitions:

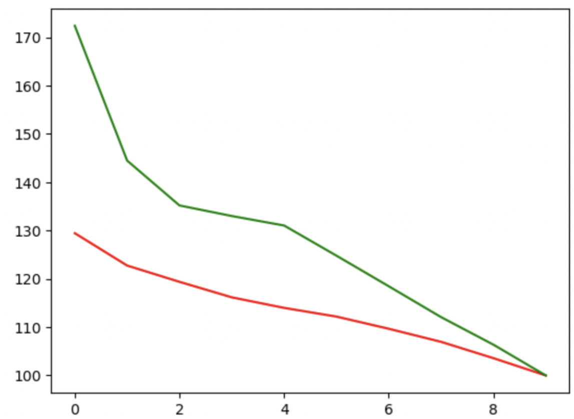
- revenue = Mean monthly revenue
- recchrge = Mean total recurring charge
- marryun = Marital status unknown
- months = Number of months the customer has had service
- uniqsubs = Number of unique subscribers (Number of individuals listed with the account.)
- eqpdays = Number of days of the current equipment
- refurb = Handset is refurbished
- dropvce = Mean number of dropped voice calls
- actvsubs = Number of individuals listed with the account who actively use the service
- retcall = Customer has made made call to retention team

2. **Demonstrate the predictive performance of the model. Is the performance adequate?**

After predicting the churn rate of both the calibration and validation sub-samples, we calculated the lift chart for the validation data. We found out that the lift is more than 70% higher than average probability of churning. Therefore, we strongly believe that our model is adequate.

In [238]:
```
plt.plot(lift_calibration,'r')
plt.plot(lift_validation,'g')
```

Out[238]: `[<matplotlib.lines.Line2D at 0x7f88396c6d30>]`



3. **What are the key factors that predict customer churn? Do these factors make sense? What is their relative importance?**

Here are the five key factors that have significant impact on the prediction of customer churn and their economic importance:

In [243]:
```
normalized_odds_ratios.sort_values(ascending=False)
```

Out[243]:
```
eqpdays     1.356199
uniqsubs    1.184760
recchrge    1.150681
months      1.140396
retcall     1.138287
actvsubs    1.114904
revenue     1.109747
refurb      1.106502
marryun     1.060802
dropvce     1.017572
dtype: float64
```

- 'Eqpdays': The number of days of the current equipment can be an important factor in preventing customer churn for a telecom company, as outdated or malfunctioning equipment can lead to dissatisfaction and increased likelihood of churn. However, the cost of upgrading or replacing equipment can be significant, and other incentives or solutions may be more cost-effective.
- 'Months': Customers who have been with the company for a longer period of time may be more valuable because they are likely to have higher lifetime value, meaning they are expected to generate more revenue over the course of their relationship with the company.

- ‘Uniqsubs’: The more subscribers listed on an account, the higher the potential revenue for the company, as each subscriber may use the company's services and products, generating more revenue. Additionally, having multiple subscribers listed on an account can increase the stickiness of the account, making it more difficult for customers to switch to a competitor.
- ‘Retcall’: Customers who have reached out to the retention team may be at risk of churning or dissatisfied with their current service, which can lead to lost revenue and increased customer acquisition costs.
- ‘Recchrge’: This metric reflects the average amount that customers are paying for recurring services and can be an indicator of the company's revenue and profitability. Increasing the mean total recurring charge can help the company to generate more revenue from each customer, as long as the prices remain competitive and customers perceive the value of the services they are paying for.

4. **What offers should be made to which customers to encourage them to remain with Cell2Cell? Assume that your objective is to generate net positive cash flow, i.e., generate additional customer revenues after subtracting out the cost of the incentive.**

Possible actions:

1. We can offer customers a new device after 13 months of subscription under Cell2Cell.

```
[55] no_churn['eqpdays'].mean()
     363.2737221935842

[56] churn['eqpdays'].mean()
     421.8327914988597
```

   We can observe that customers who have the same device for over 13 months are more likely to churn than customers who have had their device for 12 months.

```
not_missing = data[data['setprcm'] == 0]

[50] no_churn_no_miss = not_missing[not_missing['churn'] == 0]
     churn_no_miss = not_missing[not_missing['churn'] == 1]

[51] no_churn_no_miss['setprc'].mean()

     82.6513995871987

churn_no_miss['setprc'].mean()

     82.39812948671455
```

We can also observe that the average price of devices for people who churn and don't churn are around 83 dollars so we can offer a 82 dollar device to customers who have had the same device for 12-13 months.

2. Marginal cost of providing additional phone service is stated to be $0 therefore we can offer an additional month of cell service if 7 voice calls are dropped in last 1 month to ensure low churn rates.
3. If the customer had an active subscription with Cell2Cell for 18 months, offer the customer a 15% discounted subscription plan for a 3 month period since we observe that customers tend to churn after the 18th month period.

```
[61] no_churn['months'].mean()

     18.633074269400055

[62] churn['months'].mean()

     19.039012082100054
```

4. If customer calls retention frequently, we can offer them 10% discount on a new device since calling retention usually means the customer is interested in the services provided.
5. If customer has multiple subscription under their account, offer a customized cell service plan since customers are more likely to use cell service if they are under a family/group plan.

5. Assuming these actions were implemented, how would you determine whether they had worked?

Assuming these actions were implemented, we would be able to determine if our implementation had worked if our LTV increased due to the decrease in churn rate. The implementations listed above will hypothetically ensure that the retention rate of customers increase and therefore increase LTV since we made sure our incentive costs are less than our maximum cost of incentive.