FLOPKART, LLC

CHURN ANALYSIS, CUSTOMER LIFETIME VALUE, REVENUE PREDICTION

TEAM 5

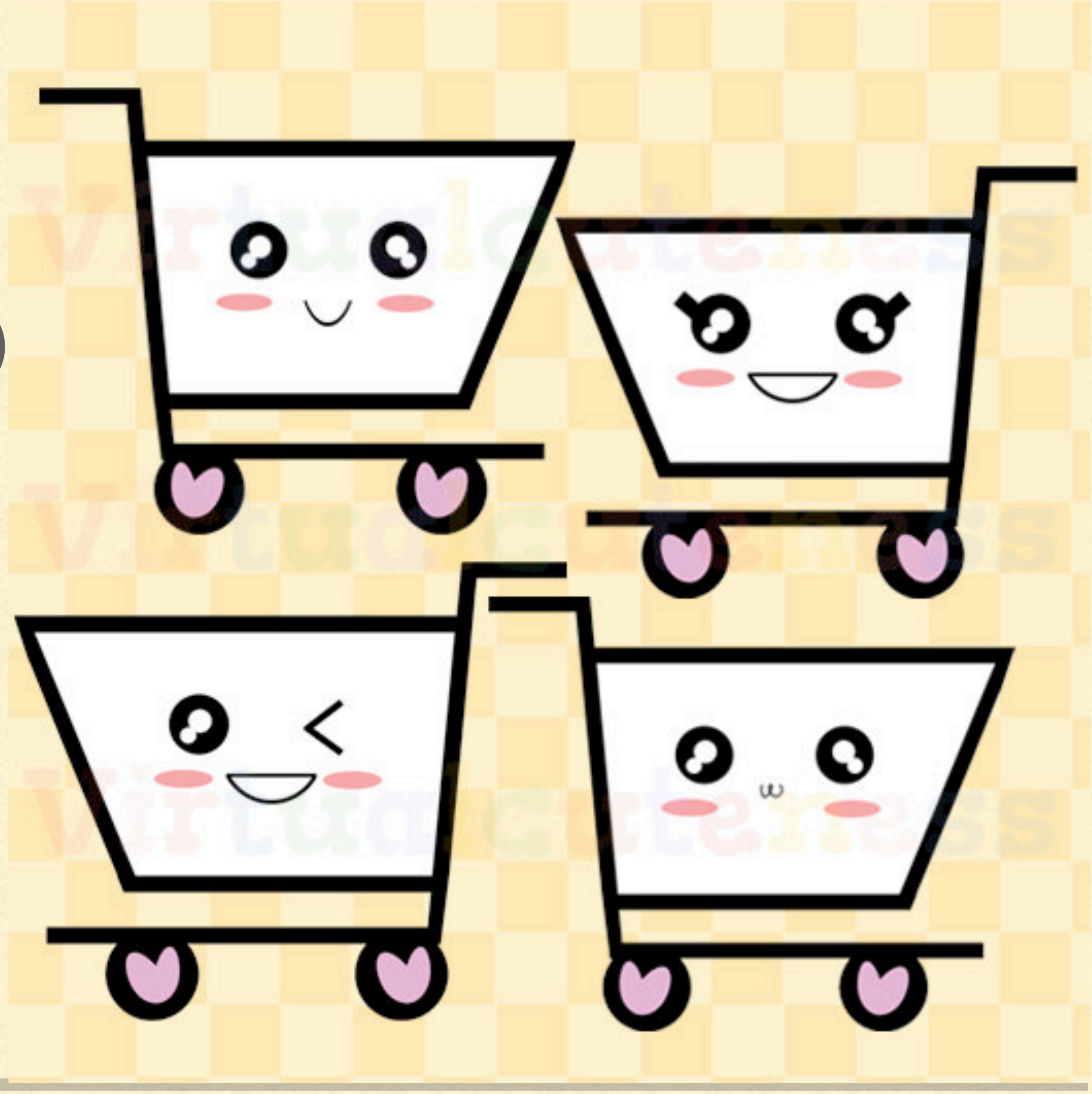
Justin Wang

Daniel Gorham

Pranav Prabhakaran

COMPANY BACKGROUND

FLOPKART :)



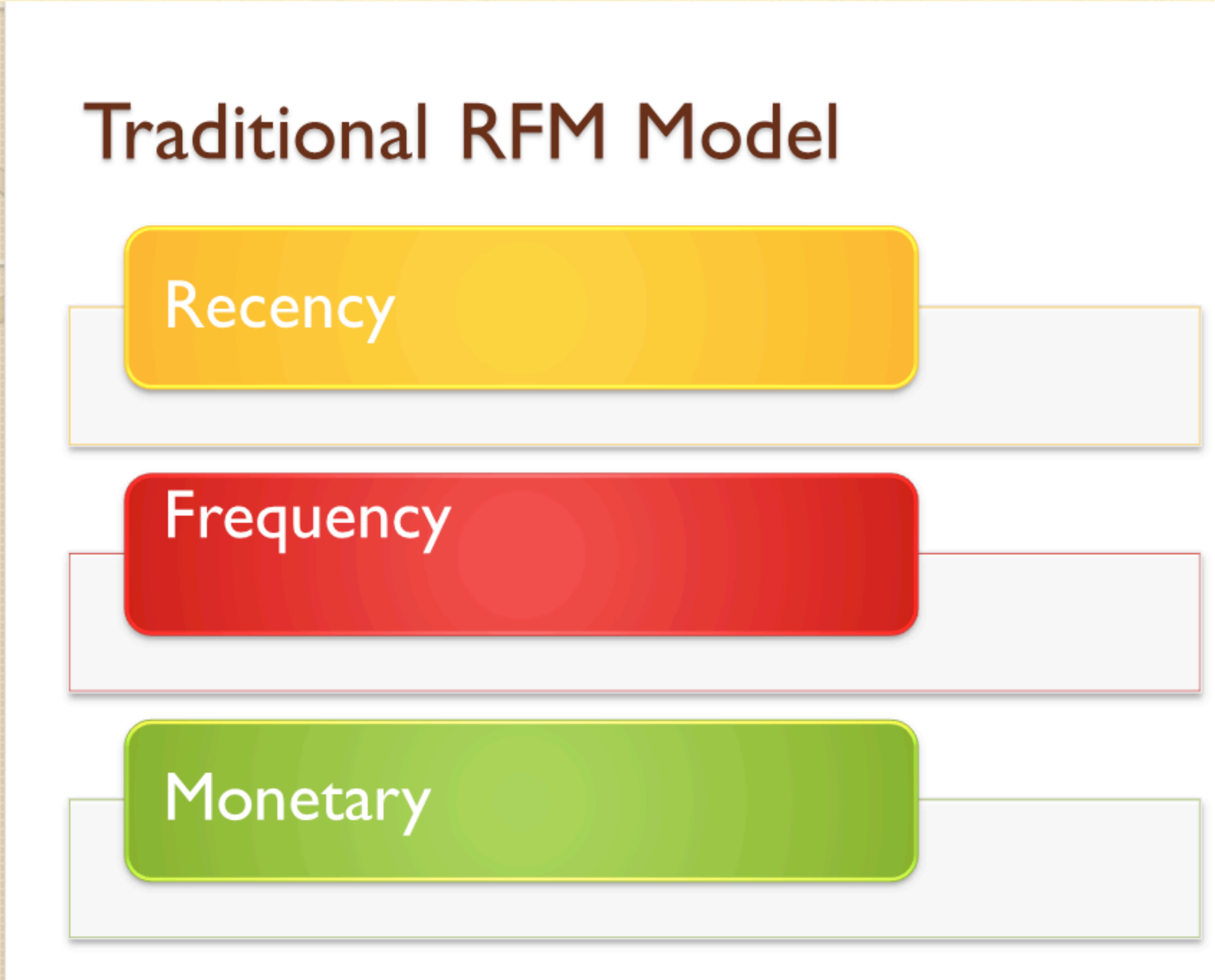
THE CHALLENGE

- FlopKart is an e-commerce company
- Leader in space, but spending aggressively and ultimately not profitable
- Facing increasing competition due to developing economy
- Want to spend money smarter

ACQUISITION VS. RETENTION

- Retaining is much cheaper than acquiring
  - Good: spend more money on high value customers
  - Bad: waste money on customers that have churned or are going to churn soon
- Problem: churn is not clearly defined in e-commerce
  - Cannot be predicted using supervised learning
  - Answer: RFM/CLV

MODEL DESCRIPTION

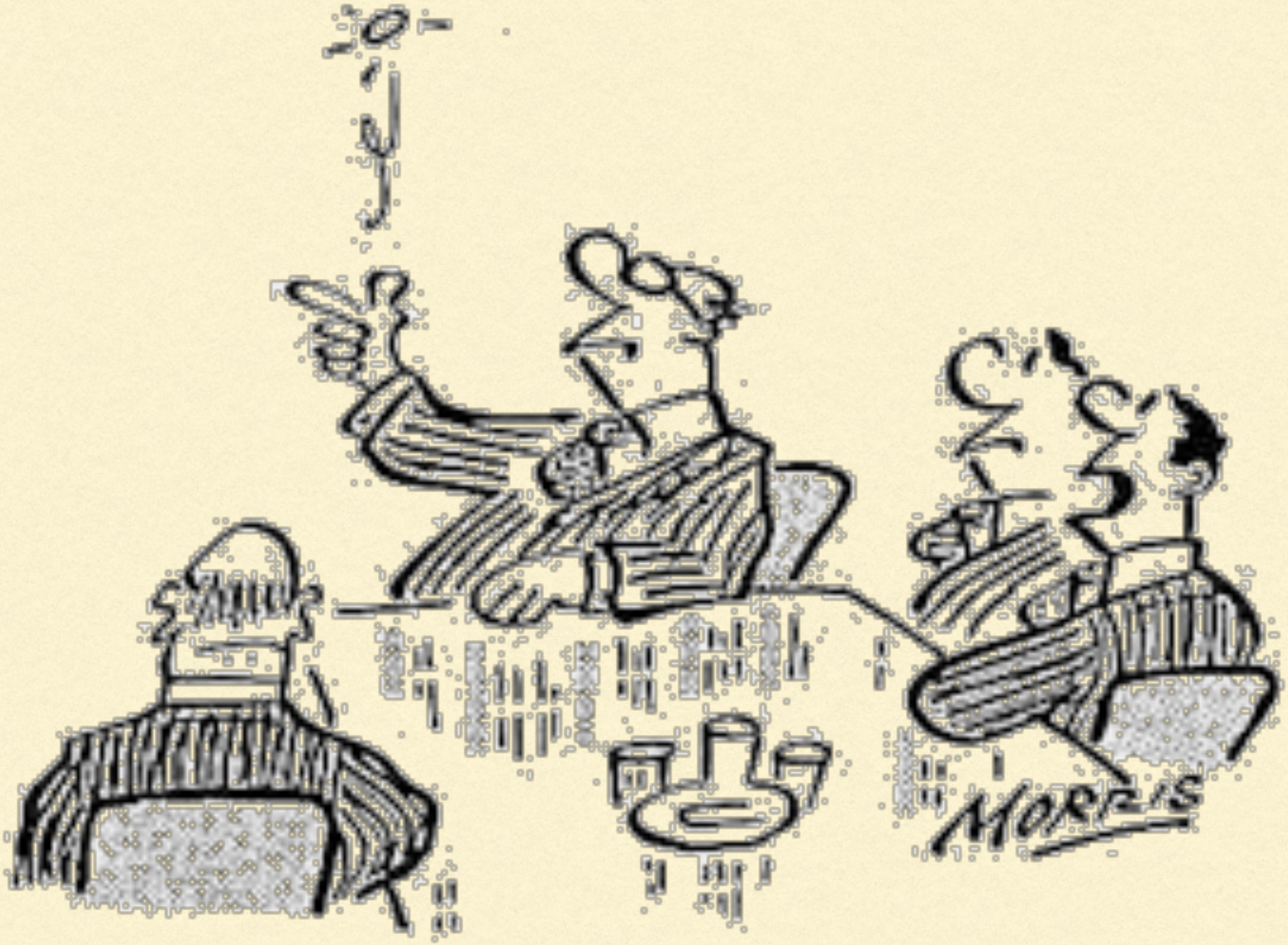


RFM MODEL

- All users assigned a “tuple” score
- Recency - number of years since last purchase
- Frequency - number of purchases to date
- Monetary - total amount spent in each year

CUSTOMER SEGMENTATION

- RFM tuples make up an infinite state space
- Segmentation is necessary
  - e.g. "platinum, gold, silver"
- State space becomes discrete and finite
- A cut-off point for recency is defined as "churn"
  - Probability of returning is very low

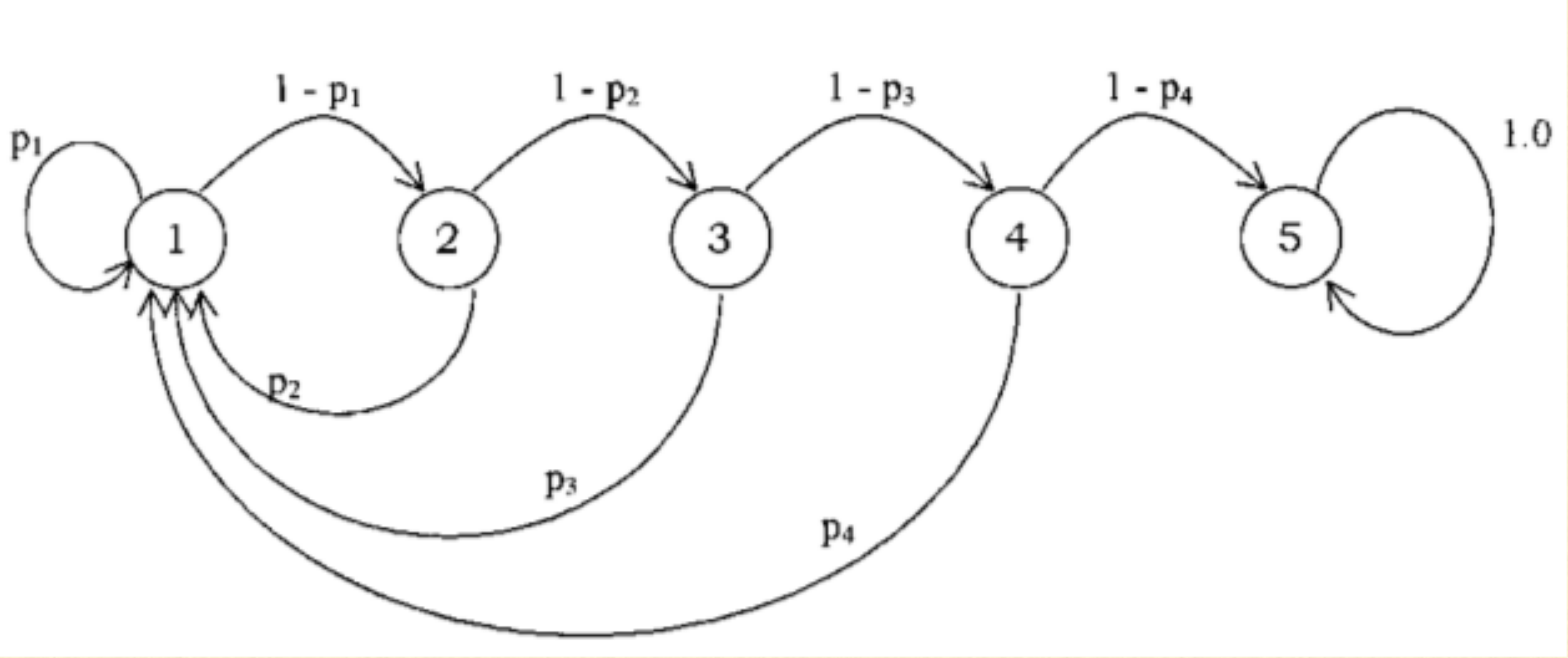

MORRIS

MARKOV MODEL

- Customers transition to different states depending on purchase behavior
- From historical data, transition probabilities may be calculated
  - P(i, j) = probability of transitioning from state i to state j over some time period
- Build matrix P of transition probabilities
  - P is of size nxn where n is number of possible states
- Once a customer is churned, they are considered churned forever
  - Absorbing state: they can only transition back into current state

MARKOV MODEL: ILLUSTRATION

- Markov transitions for recency
- 5 is absorbing state



MARKOV MODEL: PREDICTIONS

- Predict future state of customer base using Markov model
- State vector V_p(t) is defined at time-step t
  - V_p is of size n; n is number of states
  - V_p[i] gives proportion of customers in state i
- V_p(t+1) = V_p(t)P

CUSTOMER LIFETIME VALUE

- Given a customer starting at some RFM state, CLV gives their value to the company over an infinite time frame
    - Why is it valuable?
        - Some customers may not have churned yet, but have negative CLV
        - Should not expend marketing budget on these customers

CLV: REWARD VECTOR

- First step in CLV is to define reward vector R
  - R[i] gives the value of a customer in state i
  - R = NC - M
    - NC = net contribution (money spent that year)
    - M = cost spent on marketing

$$\mathbf{R} = \begin{bmatrix} NC - M \\ -M \\ -M \\ -M \\ 0 \end{bmatrix}$$

- Only customers with recency 0 have non-zero NC
  - Unchurned customers with recency >= 1 have negative reward
  - Churned customers have no marketing cost and therefore zero reward

CLV: CALCULATION

- Value vector over some time frame T:

$$\mathbf{V}^T = \sum_{t=0}^{T} [(1 + d)^{-1}\mathbf{P}]^t\mathbf{R}$$

- CLV extends time frame to infinity:

$$\mathbf{V} \equiv \lim_{T\to\infty} \mathbf{V}^T$$
$$= \{\mathbf{I} - (1 + d)^{-1}\mathbf{P}\}^{-1}\mathbf{R}$$

CLV: POLICY ADJUSTMENT

- Standard policy is to stop marketing to customers who have churned
- Customers with negative CLV haven't churned yet, but are a net negative to the company over the long term
- Policy is then adjusted to also stop marketing to these customers

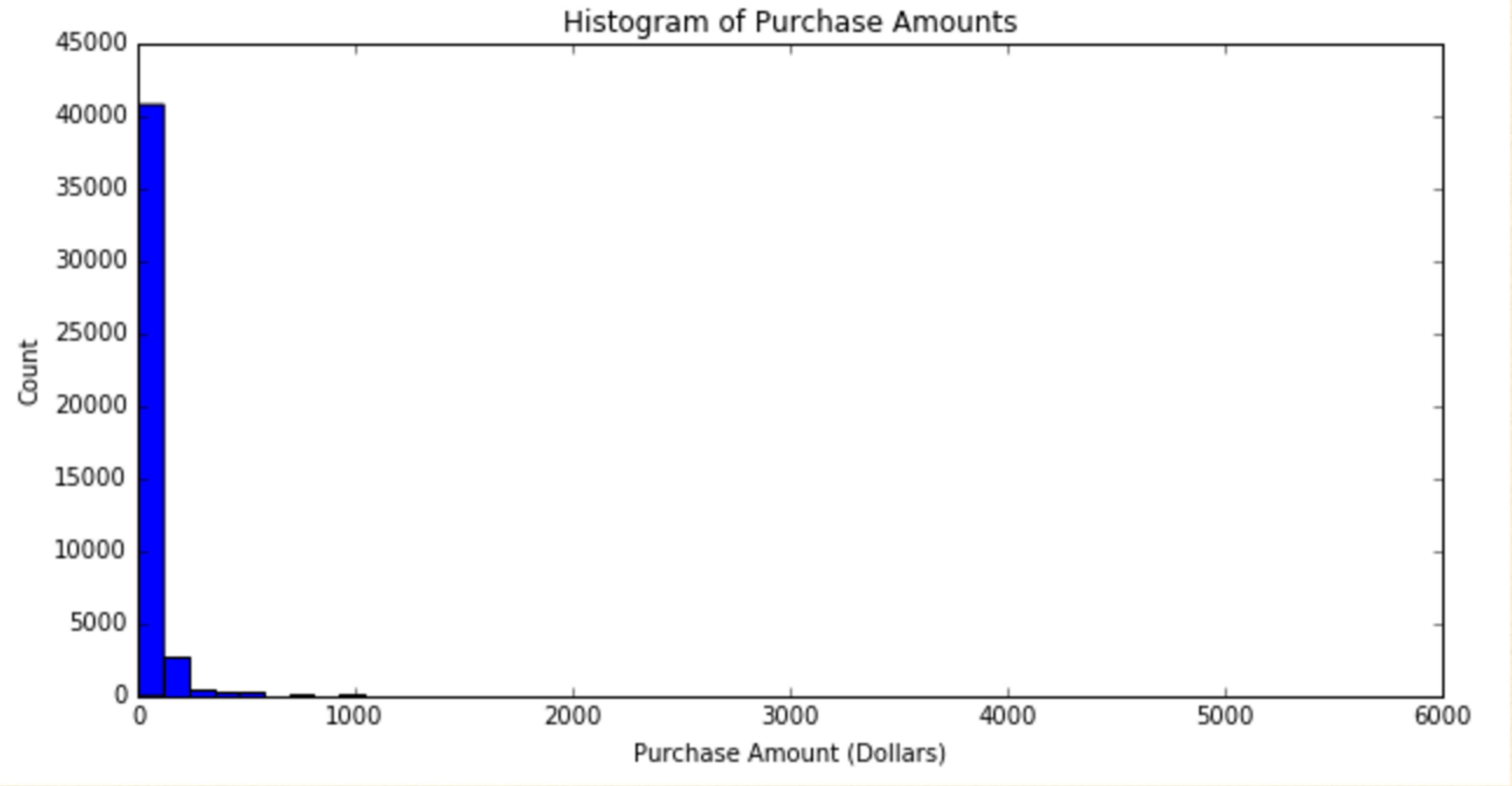
CASE SPECIFICS: FLOPKART

DATA

- Entirety of FlopKart's transactional data
- 44966 records
- Format: cust_id, purchase_amt, date_purchase
- Data from 2005 to 2015

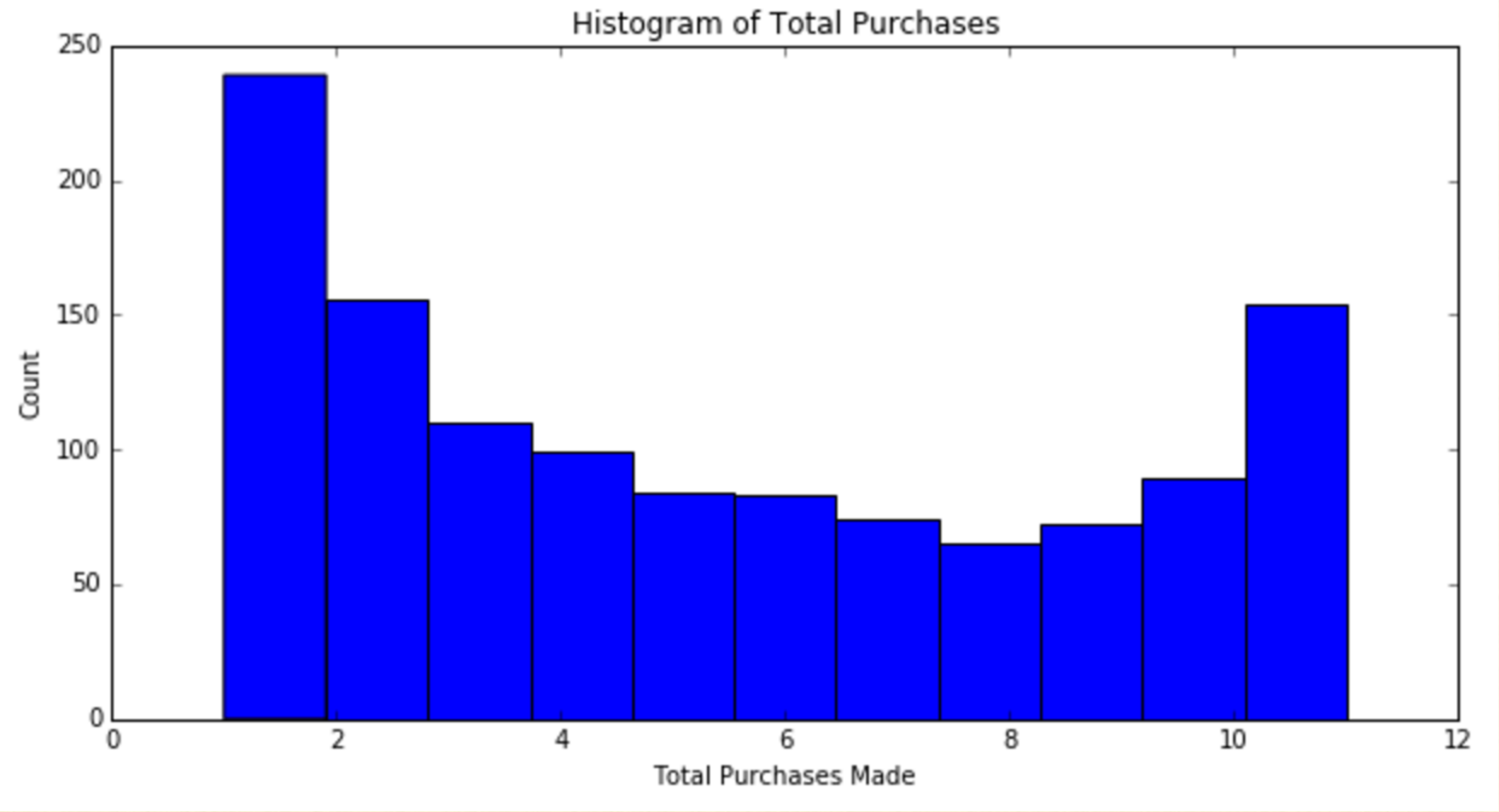
EDA: PURCHASE AMOUNTS

- Purchase amounts extremely right skewed
- Mean: $71
- Median: $40
- Max: $5750
- 90% <= $100



EDA: TOTAL PURCHASES MADE

- Decrease until an uptick at people who purchase every year
- Median: 5



EDA: BOOLEAN MATRIX

- Vast majority of customers only make one purchase per year
- Markov transition period: one year

year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
cust_id											
10	1	0	0	0	0	0	0	0	0	0	0
80	1	0	1	0	1	0	1	0	1	1	1
90	1	1	1	1	1	1	1	1	1	0	0
130	1	0	1	0	0	0	0	0	0	0	0
190	1	1	1	1	1	0	0	0	0	0	0

RECENCY TRANSITION MATRIX

- At R >= 7, probability is < 5%: churn state

	0	1	2	3	4	5	6	7	8	9	10
0	0.716090	0.28391	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.329261	0.00000	0.670739	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.137864	0.00000	0.000000	0.862136	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.086849	0.00000	0.000000	0.000000	0.913151	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	0.074963	0.00000	0.000000	0.000000	0.000000	0.925037	0.000000	0.000000	0.000000	0.000000	0.000000
5	0.035842	0.00000	0.000000	0.000000	0.000000	0.000000	0.964158	0.000000	0.000000	0.000000	0.000000
6	0.057732	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.942268	0.000000	0.000000	0.000000
7	0.041162	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.958838	0.000000	0.000000
8	0.009585	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.990415	0.000000
9	0.020492	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.979508
10	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

SEGMENTATION

- Recency:
  - R = 0 // 1 < R < 7 // R >= 7
- Frequency:
  - F <= 5 // F > 5
- Monetary:
  - M <= 30 // 30 < M <= 50 // M > 50
- (3)(2)(3) = 18 total RFM states
- All states with R >= 7 are absorbing states

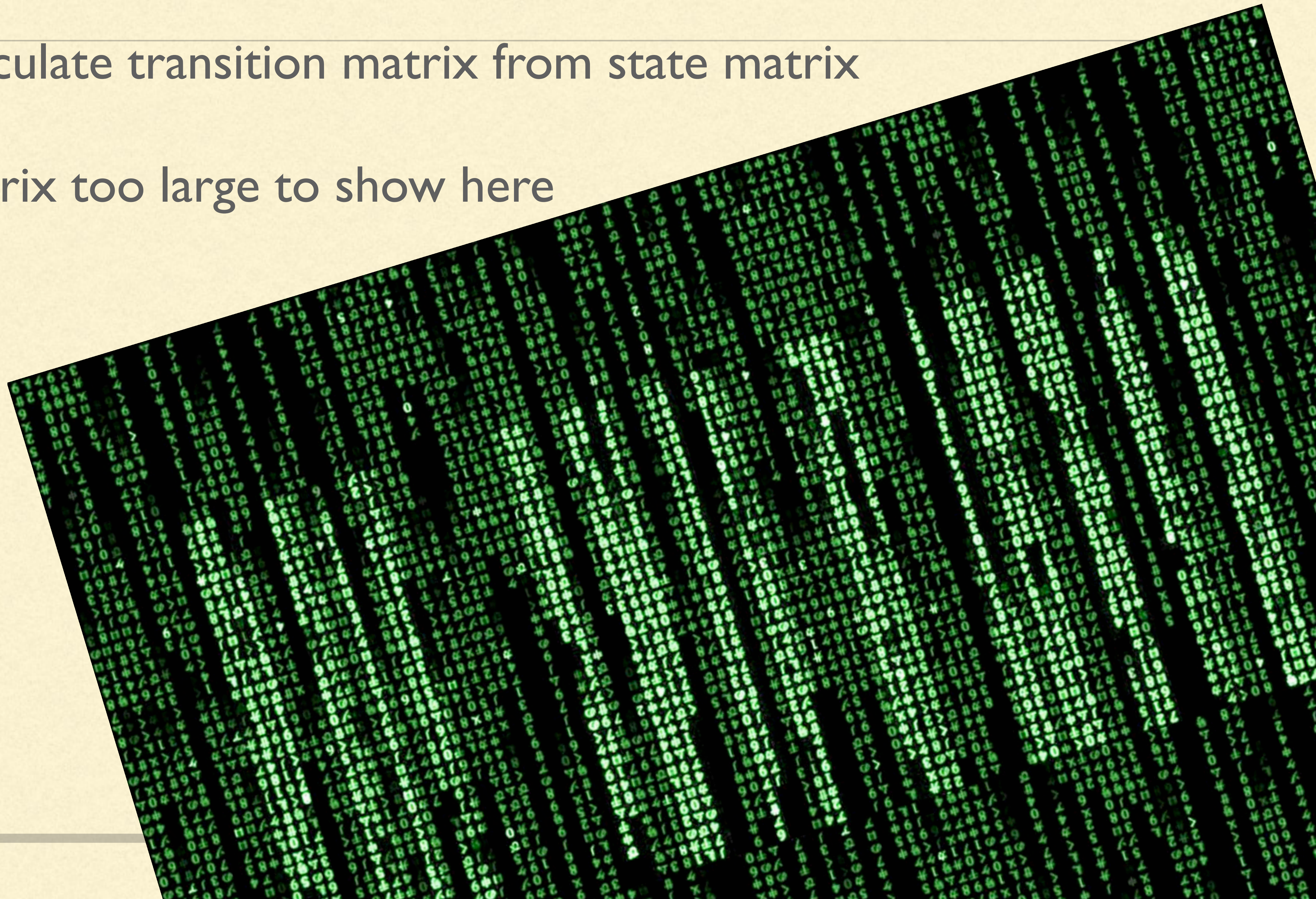
STATE MATRIX

- Calculate RFM tuples from boolean matrix
- Convert to state values using segments

year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
cust_id											
10	6	12	12	12	12	12	18	18	18	18	18
80	5	11	4	10	4	10	4	10	4	1	1
90	4	4	4	4	4	1	1	1	1	7	7
130	5	11	4	10	10	10	10	10	16	16	16
190	5	4	6	4	4	10	10	10	10	10	16

RFM TRANSITION MATRIX

- Calculate transition matrix from state matrix
- Matrix too large to show here



REWARD VECTOR

- M = ~$71
  - mean purchase amount
- C = $25
  - Flopkart's marketing cost per year

```
[[ 46.03911622],
 [ 46.03911622],
 [ 46.03911622],
 [ 46.03911622],
 [ 46.03911622],
 [ 46.03911622],
 [-25.        ],
 [-25.        ],
 [-25.        ],
 [-25.        ],
 [-25.        ],
 [-25.        ],
 [  0.        ],
 [  0.        ],
 [  0.        ],
 [  0.        ],
 [  0.        ],
 [  0.        ]]
```

FLOPKART CLV

- States 9-12 are unchurned but have negative CLV
- Policy implication
  - Stop spending money on people in these states

```
[ 164.97606887],
[ 163.53654628],
[ 132.88523581],
[ 141.27641545],
[ 133.1715939 ],
[  94.75529752],
[  13.41616067],
[  11.0259514 ],
[  -7.17343942],
[  -5.88976893],
[  -4.71694682],
[ -18.15269545],
[   0.        ],
[   0.        ],
[   0.        ],
[   0.        ],
[   0.        ],
[   0.        ]]
```

REVENUE CALCULATION

- Calculate based on previous equation using $T = 10$
- 2144 new customers each year
  - Average observed in historical data
  - Distributed according to observed entrance states

REVENUE: NO POLICY

- Total revenue: $2.6M

2016	117192.674744
2017	213756.539922
2018	274843.715024
2019	292909.959191
2020	295450.852444
2021	292481.817299
2022	287665.24994
2023	282355.970819
2024	277062.491102
2025	271977.620176

REVENUE: WITH POLICY

- **Total revenue: $3.7M**

2016	242931.3269
2017	339495.192079
2018	387466.427388
2019	400117.133611
2020	399549.233561
2021	394034.374733
2022	386729.05203
2023	378889.023419
2024	371044.570756
2025	363435.328919

Any Questions